

Biswanath Dutta
(Editor)

**Proceedings of International Conference
on**

**Bridging the Gap: Trends in
Library and Information Science**

“Connecting Practice, Education, and Innovation”

LIB-BRIDGE 2026



Biswanath Dutta
Editor



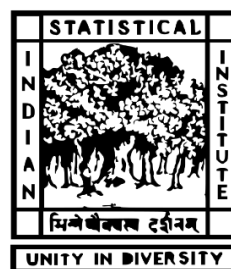
LIB-BRIDGE 2026

Bridging the Gap: Trends in Library and
Information Science

~ Connecting Practice, Education, and Innovation

Conference Proceedings

11–13 March 2026
Bangalore, India



Documentation Research and Training Centre (DRTC)

Indian Statistical Institute, Bangalore Centre
8th Mile, Mysore Road
R. V. College Post
Bangalore – 560059, India

LIB-BRIDGE 2026

Bridging the Gap: Trends in Library and Information Science

Connecting Practice, Education, and Innovation

**Held on
11–13 March 2026**

Organized by

**Documentation Research and Training Centre (DRTC)
Indian Statistical Institute (ISI)
Bangalore, India**

Venue

**Indian Statistical Institute, Bangalore Centre
8th Mile, Mysore Road,
R. V. College Post,
Bangalore – 560059, India**

Editor

**Dr. Biswanath Dutta
Documentation Research and Training Centre (DRTC),
Indian Statistical Institute, India**

Publication Information

All papers included in this volume were presented at **LIB-BRIDGE 2026** and underwent a peer-review process.

Sponsors



Indian Council of Social Science Research (ICSSR), India



The Defence Research and Development Organisation (DRDO), India



Sarada Ranganathan Endowment for Library Science (SRELS), India

Copyright and Legal Disclaimer

© **2026 Documentation Research and Training Centre (DRTC)**, Indian Statistical Institute, Bangalore, India.

All rights reserved. No part of this publication can be reproduced, stored or transmitted in any form or by any means, electronic or mechanical, including photocopying, microfilming and recording or by any information storage and retrieval system, without the prior permission in writing of the Documentation Research and Training Centre (DRTC), Indian Statistical Institute, Bangalore and the copyright owners.

Disclaimer: The opinions expressed in the articles by the authors and contributors are their own. The Editorial Advisory Board, and Editor assume no responsibility for the views expressed herein.

ISBN: 978-93-5619-660-5

Preface

Library and Information Science stands today at a historic crossroads. The profession that once centered on physical collections, catalogues, and reading rooms is now being reshaped by digital media and rapid technological advancement, particularly in the field of artificial intelligence.

Libraries have long served as institutions mediating between knowledge and society. Today, that mediation is increasingly carried out by machines, recommendation systems, and generative models. This transformation compels us to reconsider professional identity, ethical responsibility, and the competencies future librarians must develop.

Organized by the Documentation Research and Training Centre (DRTC), Indian Statistical Institute, Bangalore, **LIB-BRIDGE 2026** (11–13 March 2026) was conceived under the theme “*Bridging the Gap: Trends in Library and Information Science – Connecting Practice, Education, and Innovation.*” The conference aimed to bridge multiple divides: between library tradition and technological disruption, between LIS educators and industry innovators, between metadata specialists and AI researchers, and between the expectations of users and the design of future information systems. A collocated workshop on *Data Management* further underscored the growing importance of research data stewardship, transparency, and responsible reuse.

At the heart of the conference lay a series of guiding reflections:

- ❖ If machines claim to “organize the world’s information,” what distinctive role remains for the library?
- ❖ Can generative AI replace human cataloguers, or does it make their intellectual contribution even more essential?
- ❖ Is the traditional catalogue giving way to semantic metadata, knowledge graphs, and object-centric descriptions?
- ❖ What competencies should LIS curricula incorporate to ensure graduates remain relevant in 2030 and beyond?
- ❖ Above all, *what kind of library profession do we aspire to build for the future?*

LIB-BRIDGE 2026 sought to create a reflective space to engage with these questions. The conference was not merely a forum for presenting research; it was an effort to reinterpret and redefine the library, the librarian, and the very concept of information service in the context of artificial intelligence and emerging technologies.

The thematic tracks reflected these aspirations, encompassing knowledge organization, semantic technologies, linked data, information retrieval, metrics, digital transformation, user experience, ethics, and strategic planning. Collectively, they addressed a central concern: how libraries can remain meaningful social institutions while embracing technological change.

The conference convened experts from diverse domains, including library science, computer science, data analytics, digital humanities, and industry technology sectors. The keynote speakers embodied this interdisciplinary spirit, fostering dialogue across professional and intellectual boundaries. Their insights encouraged participants to envision new directions in library services, user engagement, ethical AI adoption, and curricular reform.

The accepted papers demonstrate both respect for foundational LIS scholarship and a willingness to experiment with emerging paradigms. All submissions underwent a

rigorous blind peer-review process to ensure academic quality and relevance. The resulting proceedings reflect the evolving landscape of LIS, from traditional cataloguing debates to AI-assisted discovery, from metadata frameworks to issues of inclusion and accessibility. Poster presentations further broadened participation, amplifying the voices of emerging researchers and practitioners.

LIB-BRIDGE is envisioned as more than a single event; it seeks to initiate an ongoing dialogue about the identity, responsibility, and future of the library profession. Whether libraries remain primarily physical spaces, evolve into data hubs, or become hybrid knowledge platforms, the conference affirmed a central conviction: *the human values of access, trust, organization, and service must remain paramount in an AI-mediated world*. It is our hope that these proceedings will inspire continued reflection: What should the library become? What should we teach future librarians? May LIB-BRIDGE 2026 serve as a stepping stone toward these answers and toward a renewed, resilient, and imaginative library profession.

We extend our sincere gratitude to everyone including the Chief Patron, Conference Director, Chair, committee members, reviewers, keynote speakers, authors, and participants whose collective efforts shaped this dialogue. Their commitment reflects a shared belief that the future of libraries should be consciously designed rather than passively inherited.

March 2026

Dr. Biswanath Dutta
Documentation Research and Training Centre (DRTC)
Indian Statistical Institute, India

Organization

Chief Patron

Prof. Ayanendranath Basu	Officiating Director, Indian Statistical Institute, India
--------------------------	---

Conference Director

Prof. B. S. Daya Sagar	Centre Head, Indian Statistical Institute, India
------------------------	--

Chair

Prof. Tridip Ray	Professor, Economics and Planning Unit, Indian Statistical Institute, India
------------------	---

Convener

Dr. Biswanath Dutta	Documentation Research and Training Centre (DRTC), Indian Statistical Institute, India
---------------------	--

Programme Advisory Committee

Fausto Giunchiglia	University of Trento, Italy
Carolyn Watters	Professor Emeritus, Dalhousie University, Canada
Michael Shepherd	Professor Emeritus, Dalhousie University, Canada
K. S. Raghavan	Professor (Retd.), Indian Statistical Institute, India
Pijushkanti Panigrahi	University of Calcutta, India
M. P. Satija	Professor Emeritus, Guru Nanak Dev University, India
Claudio Gnoli	University of Pavia, Italy
Deborah Lee	University College London, United Kingdom
Arun Kumar Chakraborty	Netaji Subhash Open University, India
P. S. Mukhopadhyay	University of Kalyani, India
N. V. Satyanarayana	Informatics (India) Ltd., India
Umeshreddy Kacherki	Indian Institute of Science, India

Programme Committee

M. P. Satija	Guru Nanak Dev University, India
Arun Kumar Chakraborty	Netaji Subhash Open University, India
Susmita Chakraborty	University of Calcutta, India
Manoj Kumar Verma	Mizoram University, India
Shashank Sonwane	Dr. Babasaheb Ambedkar Marathwada University, India
Ramesha B	Bangalore University, India
Saptarshi Ghosh	University of North Bengal, India
Bhakti Gala	Central University of Gujarat, India
Jaideep Sharma	Indira Gandhi National Open University, India
Monali Paladhi	Indian Statistical Institute, India
Jiban K. Pal	Indian Statistical Institute, India
Manju Naika	Indian Institute of Technology, India
Bijan Kumar Roy	University of Calcutta, India
Rajesh Das	University of Burdwan, India
Soumen Mallik	Vidyasagar University, India
Syamili C	University of Calicut, India
Prashant Kumar Sinha	Yahoo! India
Barnali Roy Choudhury	Netaji Subhas Open University, India
Usashi Chatterjee	Cardiff University, United Kingdom
Amit Tiwari	University of Delhi, India
Anwesha Bhattacharya	Acharya Jagadish Chandra Bose College, India
Gopalji	Indian Statistical Institute, India
Mayukh Bagchi	University of Trento, Italy
Mohit Garg	Indian Institute of Technology, India
Sagar Gajbe	Indian Statistical Institute, India
Vinit Kumar	Babasaheb Bhimrao Ambedkar University, India
Raj K. Bhardwaj	University of Delhi, India
Udaya Varadarajan	S&P Global, India
Ranjeet Kumar Singh	University of Delhi, India
Preedip Balaji	Indian Institute for Human Settlements (IIHS), India
Narendra Kumar Bhoi	BITS Pilani, India
Subhranshu Sahoo	National Law University, Jodhpur
Abhinav Sharma	Indian Statistical Institute, India
Amisha Anil Jadhav	Indian Statistical Institute, India
Anamika Venugopal K	Indian Statistical Institute, India
Anshul Mishra	Indian Statistical Institute, India
Fathima Nasreen	Indian Statistical Institute, India
Ishita Majumder	Indian Statistical Institute, India

Madhu M	Indian Statistical Institute, India
Prasenjit Das	Indian Statistical Institute, India
Puranjani Das	University of Illinois Urbana-Champaign, USA
Rakendu S	Indian Statistical Institute, India
Rasheedha Sulthana	Indian Statistical Institute, India
Shakeeb Arzoo	CRISIL, India
Vaidehi C Y	Indian Statistical Institute, India
Nishanth Kumar Chiddhala	Indian Statistical Institute, India
Nabina Khamaru	Indian Statistical Institute, India
Vikas Kumar	Indian Statistical Institute, India
Debanjali Bain	Indian Statistical Institute, India
Biswanath Dutta	Indian Statistical Institute, India

Publicity Committee

Debanjali Bain	Indian Statistical Institute, India
Nabina Khamaru	Indian Statistical Institute, India
Ram Vishwas Chourasiya	Indian Statistical Institute, India
Vikas Kumar	Indian Statistical Institute, India
Nishanth Kumar Chiddhala	Indian Statistical Institute, India

Organizing Committee

Biswanath Dutta	Indian Statistical Institute, India
Gopalji	Indian Statistical Institute, India
Sagar Bhimrao Gajbe	Indian Statistical Institute, India
Vikas Kumar	Indian Statistical Institute, India
Debanjali Bain	Indian Statistical Institute, India
Nabina Khamaru	Indian Statistical Institute, India
Nishanth Kumar Chiddhala	Indian Statistical Institute, India
Mohammed Sabith	Indian Statistical Institute, India

Content

Knowledge Organization and Discovery	1
From Strings to Things: The Role of Linked Data in RDA-Coded MARC 21	
<i>Prajith Jona D and Kanagasabai K</i>	3
Design and Development of Yogic Physical Culture Ontology towards Traditional Knowledge Organization	
<i>Sundari S, Rekha R. V., R. Murugesan, K. Tirou Mourougane and N Siva</i>	9
Systematic Literature Review of Crop Cultivation Ontologies using PRISMA with Parametric Approach	
<i>Anjali Chaurasiya, Shilpi Verma and Ankit Gupta</i>	19
What can Computer Vision learn from Ranganathan?	
<i>Mayukh Bagchi and Fausto Giunchiglia</i>	33
An Analytical Study of Web Dewey 2.0 and UDC Online	
<i>Partha Chattopadhyay and Arun Kumar Chakraborty</i>	37
 Reimagining Metadata for a Semantic Web	43
A Rule-Governed Agentic AI Architecture for Intelligent Discovery and Metadata Management for Library and Information Science	
<i>Aravind R. Nair, Vandhana S. P. and Chandhana S. P.</i>	45
 Linked Data and Knowledge Graphs in LIS	57
Readiness of Indian Cultural Heritage Websites for Linked Open Data: An Empirical Study Based on the Five-Star Model	
<i>Anuradha Kumari, Bhakti Gala and Satyam Kumar</i>	59
Designing for Conversational Coherence: A Cross-Modal Analysis of Semantic Drift in Human–AI Interactions	
<i>Aadya Bharti, Rishu, Abhishek Sharma and Anishka Gupta</i>	69
From Content to Concepts: A Semantic Framework for Enhancing Classroom Learning	
<i>Chitra Ghosh, Shubham Mishra and Ram Vishwas Chourasiya</i>	77
Towards the Transformation of Operational Library Data into Linked Data: An Analytical Overview	
<i>Nabina Khamaru and Biswanath Dutta</i>	89
 Information Retrieval in Complex Digital Environments	99
Navigating Complexity: Advanced Approaches to Information Retrieval in Modern	

Digital Ecosystems	
<i>Debkanta Halder and Subarna Kumar Das</i>	101
ChatGPT as a Research Discovery Tool: An Evaluation Using Citations and Altmetric Evidences	
<i>Subhasmita Sankhua and Sadik Batcha M</i>	111
Metrics and Statistical Applications in LIS	117
Mapping the Knowledge Structure of Robotics Research in India (2015–2024): A Scientometric Approach	
<i>Laishram Thoithoiba and Keisham Sangeeta Devi</i>	119
Thematic Evolution of Indian Library and Information Science Research: A Topic Modelling Analysis of DJLIT and ALIS	
<i>Subhayan Chakraborty and Manoj Kumar Verma</i>	127
Mapping the Research Landscape of Artificial Intelligence in Academic Libraries: A Topic Modelling Analysis	
<i>Gulla Bhaskar, Uday Pratap Singh, Sudhanshu Bajpai and Anand Gupta</i>	137
Scientometric Portrait of Bharat Ratna Prof. C. N. R. Rao: A Study	
<i>Tejashwini B, Ananda S. K., Shivakumar H. C. and Akshatha A. K.</i>	147
Technology and Digital Transformation	153
Enhancing AI-based Library Services in Academic Libraries	
<i>Dashrath Singh Nag</i>	155
AI-Based Library Service to Assist Undergraduate Student Projects	
<i>Sheetal S. Sonawane, Anupama G. Phakatkhar and Girish K. Palshikar</i>	161
Metaverse Libraries: Revolutionising Knowledge Dissemination	
<i>Barnali Roy Choudhury</i>	169
Digital Libraries and Repositories: Future Directions	179
Digital Libraries through the Lens of YouTube: Content and Sentiment Insights	
<i>Ramakrushna Baral, Fazlunnisa H., Kratika Paraste, Mounapriya R. and Jidigunta Jaji</i>	181
Digital Workflows and Management Practices of Electronic Resources in Engineering College Libraries: A Pilot Study	
<i>Preethi and Mahabaleshwara Rao</i>	191
Data Management and Research Data Services	199
Illustration of Cases of Research Data Management Services in Libraries in Asia: An Empirical Study	
<i>Prameeta Sharma and Chandrashekara M.</i>	201

A Curriculum Mapping Study of Open Science and Research Data Management in PhD Coursework of LIS in India	
<i>Mayurakshi and Preeti Mahajan</i>	211
Roles and Skills in LIS	219
Reforming LIS Curriculum in Line with NEP 2020: Recognizing Indian Knowledge Systems	
<i>Rajesh Rangappa Aldarathi, Manoj Kumar Sinha and Moirangthem Sajan Singh</i>	221
AI Across the LIS Curriculum in India: An Empirical Syllabus Analysis Using Text Analytics	
<i>Adithya Mohan, Neethusha P. and S. Ravikumar</i>	229
Mapping the Global LIS Landscape: A Cluster-Based Correlation Study of Course Offerings in the USA and India	
<i>John D. Selwin and Fazlunnisa H.</i>	239
User Experience and Access	247
Evidence based Evaluation of Chatbot Services in Academic Libraries	
<i>Varalakshmi V. and Fazlunnisa H.</i>	249
An Exploration of Users' Perception towards AI-Based Library Search Systems at the University of Jammu: A Study	
<i>Anjali Thakur, Meghna Dhar and Vishal Roudhan</i>	261
Sustainable Library Website Evaluation: Significant Parametric Study	
<i>Kreethika. M, D. Gnanasekaran, P. Ganesan and T.M. Kalpana</i>	267
Policy, Ethics, and Social Responsibility	277
Assessing Perceptions of Library Professionals on Academic Integrity in Indian Educational Institutions	
<i>M. Manivannan and Durga Sankar Rath</i>	279
Privacy and Data Ethics in Academic Libraries: Policy Frameworks, Ethical Challenges, and Social Responsibility in the Digital Age	
<i>Kiran Bapu Shewale</i>	289
Responsible and Ethical Use of AI for Publications: Exploring Major Publishers' Guidelines	
<i>P Ganesan, Reshma John, M. Mahalakshmi and A. Pavithra</i>	295
LIS in Practice: Case Studies & Innovation Labs	301
Reimagining Academic Library Spaces for Active Learning: A University Library Case Study	

<i>Kothainayaki S.</i>	303
Patterns and Practices of Usage of E-Resources Among Students of Central University of Rajasthan	
<i>Vineeta Chauhan, Kiran Yadav and Sobhagyawati Gupta</i>	315
Short Papers	327
The Power of Data Analytics and Artificial Intelligence to Enhance Decision-Making through Personalized, Data-Driven Recommendations: A Study	
<i>Asmita Deka and Md. Imran Ahmad</i>	329
The Open-Air Green Library for Academic Sustainability: Improvement Strategies at Gulbarga University Library	
<i>Rajashree Hadapad and Suresh Jange</i>	331
Information Retrieval in Complex Digital Environments: A Library and Information Science Perspective	
<i>Munegowda D. M. and Amrutha R.</i>	335
Reconceptualising Library Spaces towards Human-Centric Design for Cognitive Well-being and Inner Balance	
<i>Nisha S., Balasubramani R., Keerthna M. and Eswar M.</i>	339
Reframing Bibliographic Data: BIBFRAME Adoption in Koha ILS	
<i>Eswar M., Vedashawmya K. S., Nisha S. and Keerthna M.</i>	343
Bridging the Technological Gap in Library Information Retrieval with Integration of Elastic Search in Koha	
<i>Madhu Bala S. S., Marimuthu R. and Jeyapragash B.</i>	345

Knowledge Organization and Discovery

From Strings to Things: The Role of Linked Data in RDA-Coded MARC 21

Prajith Jona D¹ [0009-0003-7223-1000] and Kanagasabai K² [0009-0009-8858-7785]

¹PGDLAN, Indira Gandhi National Open University, New Delhi

² Great Lakes Institute of Management, Chennai

dprajithjona@gmail.com

Abstract. For over half a century, the Machine-Readable Cataloguing (MARC) format has served as the global standard for bibliographic exchange. However, the adoption of Resource Description and Access (RDA) has shifted the cataloguing paradigm from "transcribing text" to "identifying entities and relationships." This paper examines the critical need for Linked Data (LD) integration within MARC 21 to fulfil the potential of RDA. It analyses the benefits for library users, such as enhanced discoverability and contextual search, and for librarians, focusing on metadata sustainability and interoperability. The study concludes that while MARC 21 remains a necessary legacy format, its evolution into a Linked Data-ready structure is essential for the survival of library data in the Semantic Web.

Keywords: MARC21, RDA, Linked Data, Authority Control.

1 Introduction

The transition from AACR2 to RDA marked a fundamental shift in how libraries describe resources. RDA is designed to align with the IFLA Library Reference Model (LRM), focusing on the relationships between Works, Expressions, Manifestations, and Items (WEMI). Despite this modern conceptual framework, most libraries still store this data in MARC 21, a format originally designed for the era of the printed card. This "content-carrier" mismatch creates a bottleneck, where the rich relational data of RDA is flattened into text strings [2]. As noted by the Library of Congress (2024), the integration of Linked Data principles, specifically, the use of Uniform Resource Identifiers (URIs) is no longer a theoretical luxury, but a technical necessity to bridge the gap between legacy systems and the Semantic Web.

2 Methodology

This research utilizes a comprehensive literature review as the primary methodology, synthesizing academic discourse, technical standards, and institutional reports to investigate the intersection of RDA-coded MARC 21 and the Linked Data environment. Prior scholarship consistently underscores the tension between MARC's legacy structure and the relational richness of RDA. [2] traces MARC's gradual evolution toward BIBFRAME, emphasizing the limitations of text-string encoding, while [7] demonstrates how RDA's designators are inherently suited for Linked Data environments. [6] extends this argument by showing that semantic interoperability is not merely a technical upgrade, but a pathway to building a global knowledge space. Together, these studies establish that the "string-to-thing" transition is both conceptually necessary and practically feasible, laying the foundation for this paper's focus on URI integration in MARC 21.

The technical dimension of this methodology involves a systematic review of MARC 21 format updates from the Network Development and MARC Standards Office, focusing on the utility of subfields \$0 and \$1 and field 758. These elements serve as the essential mechanisms for embedding machine-actionable identifiers within legacy textual records [9]. Furthermore, the study incorporates professional findings from the Program for Cooperative Cataloguing (PCC), specifically the LDAC/IMAC URIs Guidance Subgroup Final Report (Program for Cooperative Cataloguing [11]). By synthesizing these pilot reports and task group findings, the research provides evidence on the practical feasibility of automated URI generation and its repeatability across diverse library infrastructures, ensuring the methodology reflects both theoretical ideals and current industry capabilities.

3 RDA and Linked Data

RDA describes "entities" rather than just textual labels. In a Linked Data environment, these entities must be represented by URIs to be machine understandable [3].

3.1 Library standards in Linked Data Era

This is our current reality, a paradigm shift driven by the need for our data to be part of the broader web. We are moving beyond a self-contained catalogue and toward a global network of interconnected information [8].

- RDA (2010s): Resource Description and Access were our successor to AACR2. Its core strength lies in its modularity and its foundation in the FRBR model. While still used with MARC in most places, its true power is realized in a linked-data environment.
- BIBFRAME (2010s): This is the future. “BIBFRAME (Bibliographic Framework)” is an initiative to create a new data model specifically for linked data. It is designed to replace MARC and make catalogue data truly web-friendly, enabling discovery and use by search engines and other web applications in a way MARC21 never could [10].
- IFLA LRM (2010s): A consolidation of earlier FRBR models, the LRM (Library Reference Model) is a single, unified reference model for bibliographic data. It serves as an authoritative conceptual model for RDA and other linked-data projects [12].
- RDA & LRM Alignment (2020s): This is the ongoing work of "re-tooling" RDA. So, it can fully function within the LRM framework. It's about bringing our primary cataloguing rules into full alignment with the new conceptual model.

3.2 The String-to-Thing Transition

The transition from textual strings to identified "things" allows systems to distinguish between homonymous entities (e.g., two authors with the same name) by assigning unique URIs from authoritative sources such as VIAF or Wiki data [13]. This shift is essential for making library entities discoverable via standard search engines and other discovery layers.

4 Benefits

4.1 Discoverability and Metadata Reusability

Linked Data allows library resources to complement their internal data with external sources [1]. This "link-once, use-many" approach reduces the costs associated with traditional authority maintenance and allows library data to participate in the wider Web of Data [5].

4.2 Metadata Sustainability and Professional Workflow

The shift to URI-enriched MARC records addresses the "crisis of sustainability" in traditional cataloguing.

- The "Link-Once, Use-Many" Paradigm: Instead of manually updating textual fields across thousands of records when a heading change (e.g., a name change or a subject update), librarians can update the authoritative entity record once. Any system dereferencing the associated URI will automatically reflect the most current information [5].
- Interoperability and Scaling: Libraries can "harness the collective expertise" of the global community by pulling in metadata from external sources like Wiki data or VIAF. This reduces redundant work and allows staff to focus on unique local collections rather than re-cataloguing standard items.

5 Implementation in MARC 21: The Bridge Strategy

To support the shift toward Linked Data, MARC 21 has evolved to accommodate URIs in both bibliographic and authority environments. This dual-layer approach ensures that both the description of the resource and the identification of the entity are semantically grounded.

5.1 Key MARC Linked Data Elements

- Subfield \$0 (Authority Record Control Number/URI): Points to a controlled identifier for a name or subject.
- Subfield \$1 (Real World Object URI): Points directly to the non-bibliographic entity (e.g., the actual person or place) [9].

- Field 758 (Resource Identifier): Used to uniquely identify the RDA Work or Instance.

5.2 Bibliographic and Authority Enriched Records

The following examples demonstrate how URIs function as connective tissue between different record types.

Enriched Bibliographic Record (MARC 21 Bibliographic). The bibliographic record uses URIs to "dereference" text strings, allowing the catalogue to pull data from external sources.

- Control & Identification
 - LDR (Leader): Encodes the record type (book/text) and processing info. It's machine-readable metadata that sets the stage for how the record is interpreted.
 - 001 (Control number): System-assigned identifier, unique to this record.
 - 008 (Fixed-length data): Encodes critical coded info: date of publication (1976), language (English), country of publication (India), etc. This supports indexing and retrieval.
- Cataloguing Source
 - 040: Identifies who created the record (\$a DLC = Library of Congress) and the cataloguing rules used (\$e rda = Resource Description and Access).
- Access Points
 - 100 (Main entry – Personal name): Authorized access point for Ambedkar, including life dates, role (\$e author), and linked data URI (\$0). This ensures authority control and semantic interoperability.
- Title & Publication
 - 245 (Title statement): Title proper and statement of responsibility (Ambedkar).
 - 264 (Publication statement): Place (Nagpur), publisher (Y. M. Panchbhair), date [1976]. Square brackets indicate supplied information not explicitly printed.
- Physical Description
 - 300: Extent (16 pages), dimensions (23 cm). Helps users understand the physical characteristics.
- RDA Content/Carrier Framework
 - 336 (Content type): "text" (txt) → what the resource is.
 - 337 (Media type): "unmediated" (n) → no device needed to access.
 - 338 (Carrier type): "volume" (nc) → physical book.

Together, these three fields form the RDA triad for describing resources.
- Notes
 - 500 (General note): "Cover title" clarifies the source of the title.
- Subjects
 - 650 (Subject heading): Topical subject "Democracy" with LC authority URI. Provides subject access and controlled vocabulary.
- Relationships
 - 758 (Related work): Records relationship to the *work entity*. Provides authorized access point and URI to the LC Work record, linking this manifestation to its broader intellectual context.

Enriched Authority Record in MARC 21. In this authority record example, URIs point toward the Virtual International Authority File (VIAF) as a central hub for global entity reconciliation.

- Name Authority Record
 - This record establishes the authorized form for B. R. Ambedkar.
- Identity & Rules (Control Fields)
 - LDR & 008: These fixed fields tell the computer this is an Authority record (z), it is a new record (n), and it follows RDA rules (z in position 10 of the 008).
 - 010 & 024: These are the unique IDs. The 024 is the Linked Data URI. It links this record to the global Semantic Web, ensuring that every library in the world knows this is the same "B. R. Ambedkar."
 - 040: Confirms the record was created by the Library of Congress (DLC) using RDA rules (\$e rda).
- The "Authorized" Name (100 Field)
 - 100 1\$a Ambedkar, B. R. ...

```

=LDR 00000nz a2200000n 4500
                                =001 n50051293
=008 240522|n|anznnbabn      |a|ana||| d
=010 \\\$an50051293
=024 7\\$a http://id.loc.gov/authorities/names/n50051293 $2uri
=040 \\\$aDLC$beng$serda$cDLC
=100 1\\$a Ambedkar, B. R.$q(Bhimrao Ramji), $d1891-1956
=370 \\\$aMhow, India $b Bombay, India $c India
=372 \\\$a Law $a Economics $a Politics $a Social reform $2 lcsh
=374 \\\$a Jurist $a Politician $a Social reformer $2 lcsh
=375 \\\$a male
=400 1\\$a Ambedkar, Bhimrao Ramji, $d 1891-1956

```

This is the "Authorized Access Point." Every book written by him, regardless of how his name appears on the cover, must be linked to this specific string of text so that a user searching "Ambedkar" finds everything in one place.

- RDA Person Attributes (3XX Fields)

These fields follow the LRM (Library Reference Model) by defining the attributes of the "Person" entity:

- 370 (Associated Place): Lists his place of birth (Mhow), place of death/activity (Bombay), and country (India).
- 372 (Field of Activity): Identifies his professional domains: Law, Economics, Politics, and Social Reform.
- 374 (Occupation): Identifies his roles: Jurist, Politician, and Social Reformer.
- 375 (Gender): Records the gender (male) as part of the person's identifying data.

- Variant Names (400 Field)

- 400 1\\\$a Ambedkar, Bhimrao Ramji...

This is a "See From" reference. If a user searches for the full name "Bhimrao Ramji Ambedkar," the library system uses this field to automatically redirect them to the authorized heading in the 100 field.

Subject Authority Record. This record defines the topical heading for "Democracy".

- 3XX Fields: These "New RDA" fields (370, 372, 374, 375) allow systems to perform "faceted searching" (e.g., finding all authors who are "Social reformers").
 - 368 (Other Attributes): Classifies the term within the "Social sciences."
 - 550 \$wg (Broader Term): Shows that Democracy is a "type of" Political Science.
 - 550 \$wh (Narrower Term): Shows that Representative Government is a "subset" of Democracy.

5.3 Analysis of Relationship Mapping

By embedding these URIs, the authority record acts as the "source of truth" for the entity's attributes (place of birth, occupation), while the bibliographic record acts as the "event" linking that entity to a specific intellectual work. This separation of concerns is fundamental to the RDA and LRM model.

5.4 RDA-Driven MARC 21 Updates

Authority & Bibliographic Fields for Works and Expressions

- 370 – Associated Place
- 377 – Associated Language
- 380 – Form of Work
- 381 – Other Distinguishing Characteristics
- 382 – Medium of Performance

- 383 – Numeric Designation of Musical Work
- 384 – Key

Together, these fields capture the intellectual and artistic attributes of works and expressions, aligning MARC with RDA's WEMI framework.

Miscellaneous MARC Changes for RDA.

- 250 (Edition Statement) made repeatable.
- 264 introduced for production, publication, distribution, manufacture, and copyright notice.
- Expanded subfields in 033 and 518 for event details.
- Repeatable 040 \$e with "rda" code for descriptive conventions.
- Enhanced 502 dissertation notes and identifiers in 020, 024, 028.

Reconciliation of MARC and RDA Relators. A critical step in aligning MARC with RDA was the reconciliation of relator codes with RDA relationship designators.

- Unified roles: Overlapping MARC relators were harmonized with RDA-preferred terminology.
- Expanded coverage: 38 new codes were added, while 2 were deprecated.
- Semantic precision: Relationships were mapped to FRBR/BIBFRAME levels (work, expression, manifestation, item).
- Global interoperability: Ensures consistent role representation across MARC, RDA, and Linked Data vocabularies.

This reconciliation strengthens interoperability, ensuring cataloguing data remains authoritative and machine-actionable in the Semantic Web.

6 Prospects: Beyond the Bridge

6.1 MARC as a Launchpad for BIBFRAME

Enriched MARC records act as high-fidelity data sources for conversion into the BIBFRAME model. As noted by the Library of Congress (2024), BIBFRAME relies on relationships between resources rather than the "flat" bibliographic records characteristic of the MARC era.

6.2 Challenges to the Future State

Significant hurdles remain, including "system lag" where legacy Integrated Library Systems (ILS) fail to deference URIs, and a "skill gap" as cataloguers transition from entry-based workflows to ontology-based thinking [4].

7 Conclusion

The integration of Linked Data into RDA-coded MARC 21 represents more than a technical upgrade; it is a strategic necessity for the continued relevance of the library profession. By transitioning from "strings to things," libraries can finally unlock the relational power inherent in the RDA framework, transforming static records into dynamic nodes within a global knowledge network. While MARC 21 remains the primary carrier for the immediate future, the consistent application of URIs in subfields \$0 and \$1 provides a crucial bridge that ensures legacy data is not lost during the inevitable migration to BIBFRAME and the Semantic Web.

Ultimately, this evolution empowers libraries to move beyond their institutional silos and participate in the wider information ecosystem. By adopting these identifiers today, librarians ensure that library-curated knowledge remains discoverable, interoperable, and authoritative in a web-dominated information landscape.

Implications. This paper's key contribution lies in demonstrating how RDA-coded MARC 21 can serve as a pragmatic bridge toward Linked Data without abandoning legacy infrastructures. By highlighting the role of URIs in ensuring metadata sustainability, the study offers cataloguers a concrete pathway to future-proof their records. Practically, this means libraries can reduce redundant authority work, enhance discoverability for users, and prepare their data for seamless migration into BIBFRAME and the Semantic Web. In doing so, the profession not

only safeguards its existing investments, but also positions itself as a vital player in the global information ecosystem.

Declaration of AI and AI-Assisted Technologies used in Manuscript Writing Process: Not Applicable

References

1. Akullo, W. (2024). Linked Data adoption in university libraries in Uganda: the potential benefits. *Journal of Electronic Resources Librarianship*, 36(4), 292–304.
2. Bakhturina, T. (2018). From MARC 21 - to BIBFRAME model: The evolution of the machine-readable formats of the U. S. Library of Congress. *Scientific and Technical Libraries*, 3, 62–70.
3. Berners-Lee, T. (2006). *Linked Data Design Issues*. W3C.
4. Chen, Y. N. (2020). A Study on MARC21 Transformation and Application for Linked Data. *Journal of Educational Media and Library Sciences*, 57(1), 35–72.
5. Frank, P., Cannan, J. P., & Ford, K. (2014). *BIBFRAME: Why, What, Who?* Library of Congress Publications.
6. Gaitanou, P. (2024). Linked data for libraries: Creating a global knowledge space, a systematic literature review. *Journal of Information Science*, 50(1), 204–244.
7. Hider, P. (2024). Bibliographic Relationships “in the Links Below”: The RDA Designators’ Applicability to YouTube Video Descriptions. *Cataloging & Classification Quarterly*, 62(5), 491–505.
8. Jona, D. P., Sivagiri, S., Ilayaraja, M., & Pandiyan, B. M. (2025, September 24). Linked Data in Library: The Need for Transition from MARC21 to BIBFRAME. *14th National Conference on Recent Advances in Information Technology*.
9. Library of Congress. (2017). *MARC 21 Update No. 25: \$1 Real World Object URI*. Network Development and MARC Standards Office.
10. Library of Congress. (2024). *BIBFRAME - Bibliographic Framework Initiative FAQs*.
11. Program for Cooperative Cataloging (PCC). (2023). *LDAC/IMAC URIs Guidance Subgroup: Final Report*.
12. Riva, P., Le Boeuf, P., & Žumer, M. (2017). *Definition of a conceptual reference model to provide a framework for the analysis of non-administrative metadata relating to library resources IFLA Library Reference Model A Conceptual Model for Bibliographic Information*. <http://creativecommons.org/licenses/by/4.0www.ifla.org>
13. Serra, L. G. (2020). Person Identifiers in MARC 21 Records in a Semantic Environment. *Cataloging & Classification Quarterly*, 58(5), 1–15

Design and Development of Yogic Physical Culture Ontology towards Traditional Knowledge Organization

Sundar i. S¹[0009-0003-6105-0220] Rekha. R.V.²[0000-0001-5062-5541] R. Murugesan³, K. Tirou Mourougane⁴ and N Siva⁵[0000-0002-9088-4270]

^{1,5} Department of Library and Information Science, Faculty of Science and Humanities, SRM Institute of Science and Technology, KTR, Tamil Nadu, IN.

² Department of Library and Information Science, Pondicherry University, Puducherry, IN

³ Department of Physical Education and Yoga, Pondicherry University Community College, Puducherry, IN.

⁴ Department of Physical Education and Sports, Pondicherry University, Puducherry, IN.

¹us0028@srmist.edu.in

Abstract. This study was conducted to explore the use of ontologies in representing traditional knowledge. The system of Yogic Physical Culture (YPC), a framework established by Swami Kuvalayananda, was selected as the domain of the study. YPC is a selective set of yogic practices of Āsana, Bandha, Kriyā, Mudrā, Prāṇāyāma, the practice of which would regulate physiological functions of the human body. An ontology named “Yogic Physical Culture (YPC) Ontology” was constructed through a three-stage methodology: Planning and knowledge acquisition, iterative ontology construction using OWL in Protégé editor, and the evaluation of the constructed ontology. Domain knowledge was captured via expert interviews and the expert-recommended literature of the domain, then formalised into YPC ontology knowledge base with structured classes, properties and instances representing the profiles of the yogic practices. Competency questions and reasoning tools guided validation, while expert and user feedback ensured consistency and usability. The resulting ontology enables the retrieval of yogic practice information supporting trainers, students, and any users interested in accessing traditional yogic knowledge.

Keywords: Ontologies, Ontology Construction Methodologies, Knowledge base, Traditional Knowledge, Yogic Physical Culture, Asana.

1 Introduction

Ontologies are formal representations of knowledge providing a common understanding of the concepts and relationships across different data sources. In the context of knowledge management, a domain ontology defines (specifies) the concepts, relationships, and other distinctions that are relevant for modelling a domain. The specification takes the form of the definitions of representational vocabulary (classes, relations, and so forth), which provide meanings for the vocabulary and formal constraints on its coherent use [1]. Ontologies allow experts to encode their domain or subject knowledge, and the resulting ontology concepts can then be used along a problem-solving method to answer the questions [2].

This study was conducted to explore the usage of the ontologies to represent traditional knowledge. As the domain of the study, Yogic Physical Culture (YPC), a yogic system formulated by Swami Kuvalayananda, the founder of Kaivalyadhama, the oldest Yoga Research Institute in India, was selected. YPC is a self-contained, independent and complete yogic system of physical culture based on scientific principles of significance that comprises selective practices of Āsana, Prāṇāyāma, Bandha, Kriyā, and Mudrā. In YPC, irrespective of their spiritual value, only their physical utility would be considered [3]. The metadata of a sample set of journal articles that examines the impact of yogic practices to alleviate specific physical conditions were also collected to explore the mapping of the articles against the YPC Ontology.

This paper discusses the combination of methodologies used in various stages of the design and development of the YPC Ontology, starting from knowledge acquisition to ontology construction, ending with the evaluation of the YPC Ontology.

2 Related Studies

Earlier works related to the construction of domain ontologies including traditional knowledge systems were reviewed specifically to consult the methodologies and tools used for knowledge acquisition and construction of domain ontologies.

2.1 Domain Ontologies

The physical activity ontologies through a systematic review of 28 ontologies in the field of physical activity and their assessment based on the OBO Foundry principles of good ontology practice is discussed in a study [4]. The study provided the list of recommendations for the identification, evaluation, and adaptation of ontologies.

Taxonomies and ontologies play a crucial role in knowledge organization within libraries [5]. The authors studied the role of taxonomies and ontologies in building Semantic Web and recommended adopting the semantic search using taxonomies and ontologies to bring about a transformative impact on libraries by offering a more streamlined search for their patrons.

On traditional knowledge systems, a study aimed to develop an Ontology on Traditional Dances (OTD) for preserving the knowledge of Indian traditional dances in the semantic web environment [6]. The OTD knowledge base recorded the concepts of traditional dances of the Rabha tribes of Northeast India.

“Yoga Ontology”, focusing on a few aspects of asana and pranayama, including the props that can be used and the anatomical effects of the practice was published [7].

The process of building Solat ontology from the field of Islamic knowledge was discussed [8]. Constructed ontology aims to answer competency questions related to Solat, or prayer. The authors especially highlighted the significance of using reliable sources and following a structured methodology to construct ontologies with accuracy and relevance in representing knowledge.

The integration of Formal Concept Analysis (FCA) with Protégé to construct an ontology-based Yoga knowledge-sharing platform called Yoga Health Knowledge Management System (YHKMS) was proposed [2]. The work demonstrated that “Yoga Knowledge” ontology can be used to improve the information retrieval system.

On ontology-based systems in Library and Information science, a generalized diagram for the development of Ontology-related Systems (OrS) for Digital Libraries of Scientific Publications (DLSP) utilizing the Semantic Web to address the challenges researchers face in finding information for their research was presented [9].

2.2 Methodologies and Tools

On knowledge acquisition methodologies, engaging domain experts is crucial for establishing appropriate, comprehensive, accurate and clear ontologies [10]. In addition, since it is rare for one person or author to have all the knowledge required, a working paper presented a methodology for knowledge acquisition using well-known literature of the domain as a source of knowledge [11].

An ontology design primarily depends on the purpose and scope of the system for which it is being designed, and there is no one “correct” way or methodology for developing ontologies [12]. Different methodologies for ontology engineering that were used in various domains for ontology construction has been discussed, such as Uschold & King and Gruninger & Fox, Methontology, YAMO, OnTo, Sensus, and so on and it was concluded that Methontology is the most widely used methodology in developing ontology, regardless of the time [13]. On the evaluation of the constructed ontology, ontology engineering still does not have an agreed methodology and standards for ontology evaluation [14].

It is common to merge different methodologies since each of them provides design ideas that distinguish it from the others. This merging depends on the ontology users and ontology goals. In study a systematic process was shared that merged different methodologies successfully in due course of the process for building a government budgetary ontology [15].

An ontology for Greek mythology [16], a rare and underdeveloped ontology for epics and myths, was developed using a combination of methodologies, including the methodology proposed by Gruninger and Fox, the developing cycle from Methontology, and the analytico-synthetic approach from YAMO methodology.

In another study, an ontology was designed that is aimed to visualize the doctoral research output in Library and Information Science and to answer most of the questions related to those theses [17]. The ontology was developed using Methontology methodology and coded using Protégé, an open-source ontology editing software.

Different ontology editors and visualization tools cater to user needs and the types of ontology being constructed. A study on the evaluation and ranking of ontology construction tools provided a detailed analysis of ontology construction tools and their evaluation criteria in choosing the appropriate tool for their needs [18].

In a study, WebProtégé, a lightweight collaborative ontology editor and knowledge acquisition tool for the Semantic Web, which allows for on-line collaborative ontology development with features for web-based content creation and customization for users of varying expertise levels was presented [19].

Visualization plays a crucial role in understanding ontologies, highlighting essential concepts at the OWL level. A study compared the visualization tools Protégé VOWL and Web VOWL for ontology visualization [20].

In summary, review of related studies highlighted that there are well-tested methodology and tools for construction and evaluation of ontologies. The developers have come up with their own approach that suits the purpose of ontology, adopting and combining well-proved approaches for their ontology construction.

3 YPC Ontology Construction Methodology

The complete process of design and development of the YPC Ontology adopted the Methontology framework [21] merged with the Knowledge Engineering methodology [12] as an iterative approach to ontology development [15]. The process was executed in three stages: planning the scope and knowledge acquisition, construction of the ontology iteratively, and the evaluation of the ontology (See Figure 1).

3.1 Planning and Preparation

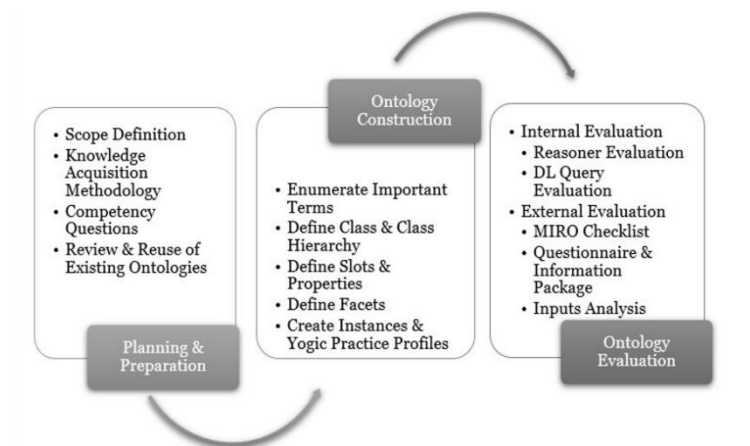
In this planning stage, it is helpful to define the boundaries and benchmarks of the construction in the planning stage as it would help to streamline the activities going forward.

Scope Definition.

- What is included in the scope?
 - Gather knowledge of Āsana practices of YPC by Kaivalyadhama (See Figure 2 for the YPC practices of Kaivalyadhama course in the scope).
 - Construct the YPC Ontology, embedding the knowledge acquired, including concepts, dependencies, relations between concepts and instances.

Fig. 1. Stages of Design and Development of the YPC Ontology

- Demonstrate the retrieval of articles on specific yogic practices.
- Who uses Ontology?



- Trainers, Practitioners, Students and any users who are interested in yogic practices
- What is not included in the scope?
 - Technique descriptions or videos or images of the yogic practice, contraindications of each yogic practice and comparison of practices among various schools of Yoga.

Śavāsana	Ardha Halāsana	Halāsana
Dhanurāsana	Ardha Śhalabhāsana	Śhalabhāsana
Paśchimātānāsana	YogaMudrā	ViparītaKaranī
Bhujangāsana	Ardha Matsyendrāsana	Matsyāsana
Vakrāsana	Mayūrāsana	Uddiyāna
Sarvāṅgāsana	Śīrāsana	Chakrāsana
Nauli	Kapālabhāti	Bhastrikā
Ujjāyī	Supta Vajrāsana	Makarāsana
Hamsāsana	Vṛkṣāsana	Tadāsana
Vajrāsana	Ardha Padmāsana	Padmāsana

Fig. 2. Yogic practices (30) in scope

Competency Questions.

- How do we know that the constructed ontology is correct and serves the purpose?
 - Prepare a list of Competency questions (CQs) in consultation with domain experts. For Example, a few CQs prepared are,
 - What are the yogic practices performed in the ‘Sitting’ position?
 - What are the physiological functions promoted by Bhujangāsana?
 - What are the yogic practices suggested for a person with Diabetes?
 - Which journal articles refer to yogic practices related to stress management?
 - The YPC Ontology should be able to answer each of these CQs to confirm that the ontology is accurate and logically consistent, and the objectives of the study are met satisfactorily.
 - Further, the CQs were also planned to ensure the applicability and usability of the YPC Ontology.

Knowledge Acquisition.

- What methodology is suitable for knowledge acquisition?
 - As per the scope defined, knowledge experts were identified, and a 1-1 interview method was selected as a knowledge gathering tool.
 - Knowledge gathering sessions and follow-up verification sessions were scheduled in consultation with the domain experts.
 - In addition, as suggested by the domain experts, well-established literature from the YPC domain such as ‘Yoga Kosa’ [22], the dictionary of Yoga terms by Kaivalyadhama, were also used for knowledge capture.
- How is the knowledge recorded and validated?
 - After each session with the knowledge experts, the knowledge acquired was documented in detail using an Excel sheet (For example, See Figure 3 for the knowledge capture of 7 practices – Śavāsana, Halāsana, Bhujangāsana, Dhanurāsana, Paśchimātānāsana, Matsyāsana and Bhastrikā).
 - Documented knowledge in natural language is validated in subsequent sessions with the knowledge experts, along with reference to suggested literature.

Practice Name	Śavāsana	Halāsana	Bhujangāsana	Dhanurāsana	Pāśchimātānāsana	Matsyāsana	Bhastrikā
Function	Relaxative	Cultural	Cultural	Cultural	Cultural	Cultural	Prāṇāyāma
Posture	Supine	Supine	Prone	Prone	Sitting	Sitting	Sitting
Description	Śava means the dead body in Sanskrit. This āsana imitates the complete relaxation of the muscles as in the case of a dead person.	Hala means the plough in Sanskrit. In this āsana practice, the body imitates the shape of Indian plough.	In this āsana practice, the body imitates the appearance of hooded snake under irritation	Dhanu means the bow in Sanskrit. In this āsana the body resembles a bow with its string attached to it	In Sanskrit, Pāśchima means stretching the posterior and the root tān means the stretch. This āsana almost stretches the whole body.	In Sanskrit, Matsya means the fish. Practitioner can float on the water for longer duration like a fish using this āsana.	One of the Sahita Kumbhaka (Prāṇāyāma with conscious breathing effort), Bhastrikā starts with performance of the kriyā Kapālabhāti
Also Referred As	Dead Pose, Corpse Pose	Plough Pose	Cobra Pose	Arrow Pose	Posterior-Stretching Pose	Fish Pose	Bellows Breathing
Preparatory Practice	--	Ardha Halāsana	Makarāsana	Bhujangāsana, Śhalabhāsana	Yoga Mudrā	Padmāsana	Kapālabhāti
Complementary Practice	--	Bhujangāsana	Śhalabhāsana	Halāsana	Supta Vajrāsana	Halāsana	Nadi shuddhi
Influences	Nervous System, musculoskeletal system	Spinal Column, Thyroid, Digestive Organs	Spinal Vertebrae, Musculoskeletal of Back, development of abdominal	Abdomen, Pelvis	Spinal Vertebrae, Musculoskeletal of Back, digestive system	Thyroid and thru it whole body and its functions	Respiratory System, Nervous System
Helps to prevent or cure	Stress, fatigue, high blood pressure, neurasthenia	symptoms of old age, Spinal column displacements, Constipation, Dyspepsia (Indigestion), Diabetes, enlarged liver and spleen	Spinal column displacements, Backpain, Flatulence, diabetes	Diabetes, Backpain, flatulence	Diabetes, Dyspepsia (Indigestion), Sciatica, Obesity	symptoms of old age, thyroid related issues	Diseases of mouth, diseases caused by excess of vata, pita or kapha
Promotes	Muscle relaxation, venous blood circulation, mental calmness	Flexible and Healthy Spinal Nerve, Digestion, strong abdominal muscles	Flexible and Healthy Spinal Nerve, Increased Blood Circulation, strengthen the core muscles of body	Digestion, Increased blood circulation, strengthen the core muscles of body	Digestion, Weightloss	Healthy Growth	Mental Calmness, Respiration, Bloodcirculation
Breath Pattern	Rhythmic	Regular	Regular	Regular	Regular	Regular	Kumbhaka
Practice Level	Basic	Intermediate	Basic	Intermediate	Basic	Advanced	Advanced
Practice From ?	Kaivalyadhama course	Kaivalyadhama course	Kaivalyadhama course	Kaivalyadhama course	Kaivalyadhama course	Kaivalyadhama course	Kaivalyadhama course

Fig. 3. Knowledge capture of 7 Yogic practices.

3.2 Ontology Construction: an iterative process

The construction of the YPC Ontology followed a simple knowledge-engineering methodology. The ontology was then formalised using OWL (Web Ontology Language) in Protégé Editor. Also, during formalisation, as a standard design consideration, entity names in OWL were done using CamelCase notation so that multi-word domain names, such as YogicPhysicalCulture, were more legible. The iterative steps of the knowledge-engineering methodology used to construct the YPC Ontology were:

Step 1: Enumerate important terms in the ontology. Important terms that represent the concepts, associations and their characteristics with respect to the practices of YPC were enumerated. Some of the important terms enumerated were: Āsana, Prāṇāyāma, Relaxative, Complementary, Preparatory, Kumbhaka, etc. Following are the significant concepts of the YPC that have been identified as per scope:

- YPC practices (Āsana, Prāṇāyāma)
- Preparatory and complementary Practices
- Alternate names of the YPC practices
- Physiological functions and health conditions
- Research papers

Step 2: Define the classes and the class hierarchy. Using the top-down approach, classes for top-level concepts such as Yogic Physical Culture, Āsana and Prāṇāyāma are created first. Āsana practices were further classified based on the functional effect of the yogic practice and then further classified based on the starting or final position of the āsana when practised. Then, the middle-level concepts, including Practice Level, Posture Classification, Functional Classification, and Breathing Pattern, were added and further expanded.

Step 3: Define the properties of classes—slots. The internal structure of the concepts of YPC was described. For example, as Āsanas were categorized based on their function, performing position or difficulty level, then properties such as HasFunctionalClassification and HasPostureClassification were defined and tagged with the class Āsana.

Step 4: Define the facets of the slots. Embedding the knowledge capture of YPC concepts, the attributes are now assigned facets. For example, The HasFunctionalClassification slot is tagged with the FunctionalClassification class to take values within the range {Cultural, Meditative and Relaxative} under the domain of class YogicPhysicalCulture. Value type slots such as Author Name (Property ArticleAuthor) & Title (Property ArticleTitle) are tagged with the ResearchPaper class to tag the bibliography information of research articles for their instances.

Step 5: Create Instances and Define Profiles of Yogic Practices. The YPC Ontology was formalised in OWL using Protégé Editor for the yogic practices in the scope. A profile for each of the yogic practices with its unique characteristics and the relationship with other yogic practices was now added to the YPC Ontology Knowledge base. For example, as per knowledge acquisition, Paśchimātānāsana is a corrective āsana practice, starting from the sitting position, practised regularly to alleviate health conditions such as diabetes and to promote digestion and weight loss. A profile for Paśchimātānāsana (See Figure 4) was created as follows:

- First, an individual or an instance named *Paśchimātānāsana* that represents the class *SittingCultural* is created.
- For this individual *Paśchimātānāsana*, property assertions are assigned as per the knowledge captured in the form of Property mapping using Protégé editor.

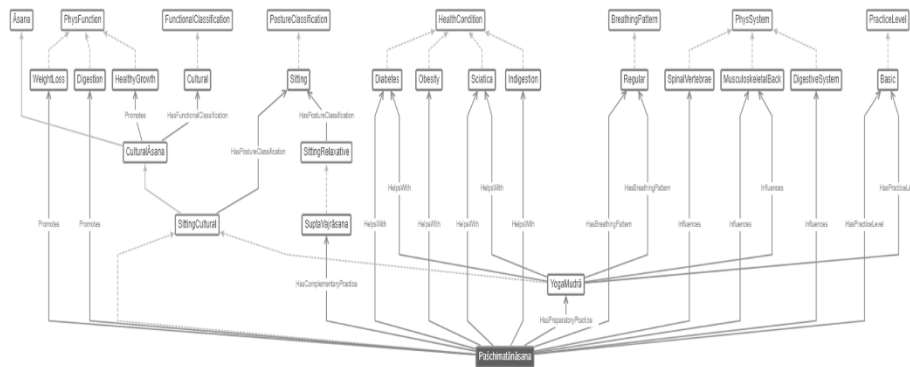


Fig. 4. Paśchimātānāsana Profile as Entity Graph (WebProtégé)

In similar way the knowledge capture of 30 yogic practices of YPC is embedded. Profiles for each yogic practice are created and represented in the knowledge base of the YPC Ontology. The final version of the YPC Ontology is available on the web as a Knowledge Graph in TriplyDB (<https://triplydb.com/sunsri0377/yoga-ypc-onto/>) and as RDF/OWL in Web Protégé.

3.3 Ontology Evaluation

The ontology evaluation ensures that the YPC Ontology serves its purpose effectively and efficiently. Note that tasks at the evaluation stage (See Figure 5) were iterative, at some level, as and when ontology is revised or updated.

Internal Evaluation. For the Consistency check, the Pellet Reasoner, a plug-in of Protégé Editor, was used. Ontology was reviewed automatically by the Reasoner for the consistency, and the listed inconsistencies were corrected till the Reasoner listed zero inconsistencies. After the consistency check, key metrics of ontology are 41 classes, 30 object properties, 7 data properties and 143 individuals

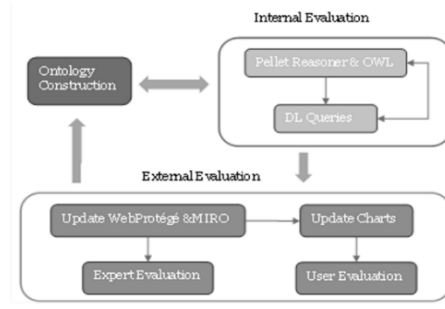


Fig. 5. Tasks in Evaluation Stage

For the accuracy check, manual evaluation was done using CQs defined in the planning stage. The DL Query, a plug-in of Protégé Editor, was used to build and execute queries for each CQ (For example, See Figure 6), validating the results against the knowledge capture.



Fig. 6. CQ: What are the yogic practices suggested for a person with Diabetes?

External Evaluation. The YPC Ontology is evaluated for its quality both by the expert ontology developers and users of the ontology. Ontology was hosted in TriplyDB.com (<https://triplydb.com/sunsri0377/yoga-ypc-onto/>) for evaluation and Ontology experts were requested to evaluate the ontology for consistency, completeness, clarity, generality and conceptualization. The MIRO (Minimal Information for Reporting of an ontology) checklist done according to guidelines recommended by OBO Foundry [23] was used to present the YPC Ontology.

Review feedback agreed on the potential of the YPC Ontology for further development. The conceptualization part of the ontology was also well-received. A set of users (trainers, students, practitioners and general users) were selected and requested to evaluate the usability and applicability of the YPC Ontology for the domain based on CQs. Group sessions were also held with a selective set of users that helped gather feedback and their preferred improvements or additions in the YPC Ontology.

4 Conclusion and Future Developments

As YPC is a traditional knowledge domain, adopting the Methontology framework merged with the knowledge engineering methodology was found to be a very suitable methodology for the design and development of YPC Ontology. Apart from knowledge captured from the experts in YPC domain using the interview sessions, knowledge documented in literature can also be mined for the construction of subject ontologies. However, it had to be verified with domain experts as there may be inconsistencies in the recorded knowledge in the literature.

The YPC Ontology and the knowledge base was constructed in OWL using Protégé Editor. The interoperability of RDF/XML generated by Protégé for YPC Ontology was very useful in porting the ontology to other tools such as WebVOWL, Web Protégé or TriplyDB (<https://triplydb.com/sunsri0377/yoga-ypc-onto/>) (See Figure 7).

Usage of YPC Ontology and its future expansions has defined many of the modelling decisions, such as whether a concept of a domain is a class, a subclass, a slot with restrictions or an individual in the ontology. The significant concepts that were formalized in YPC Ontology are yogic physical culture practices, alternate names of the practice, preparatory and complementary practices of the practice, physiological functions enhanced by the practice, health conditions that can be alleviated by the practice. In addition, the concept of resource tagging of the research articles, for example, from Yoga Mimamsa journal, by the yogic practice also incorporated.

11. Lehmann, A.: A methodology for knowledge acquisition through literature: Building knowledge-based systems for strategic management. *Manuskripte Aus Den Instituten Für Betriebswirtschaftslehre der Universität Kiel*, No. 279 (1991). <https://ideas.repec.org/p/zbw/cauman/279.html>, last accessed 2025/12/25.
12. Noy, N. F., McGuinness, D. L.: *Ontology Development 101: A Guide to Creating Your First Ontology*. (2001). https://protege.stanford.edu/publications/ontology_development/ontology101.pdf, last accessed 2025/12/25.
13. Syamili, C., & Rekha, R. V.: *Ontology engineering methodologies: An analytical study*. (2017). <https://ir.inflibnet.ac.in/handle/1944/2118>, last accessed 2025/12/25.
14. Wilson, S. I., Goonetillake, J. S., Ginige, A., & Walisadeera, A. I.: Towards a usable ontology: The identification of quality characteristics for an ontology-driven decision support system. *IEEE Access*, 10, 12889–12912. (2022). <https://doi.org/10.1109/access.2022.3146331>
15. Brusa, G., Caliusco, M., Chiotti, O.: A process for building a domain ontology: An experience in developing a government budgetary ontology. In: *Australasian Ontology Workshop (AOW 2006)*, Hobart, Australia. *Conferences in Research and Practice in Information Technology*, 72, 7–15 (2006).
16. Syamili, C., & Rekha, R. V.: Developing an ontology for Greek mythology. *The Electronic Library*, 36(1), 119–132 (2018). <https://doi.org/10.1108/EL-02-2017-0030>
17. Syamili, C., & Rekha, R. V.: *Ontology of Library and Information Science Theses of the Central Universities in India. Library Philosophy and Practice (e-Journal)* (2021). <https://digitalcommons.unl.edu/libphilprac/5197>, last accessed 2025/12/25.
18. Murshed, M. A., & Ramanjeet, S.: Evaluation and Ranking of Ontology Construction Tools. (2005, March 1). <https://www.semanticscholar.org/paper/Evaluation-and-Ranking-of-Ontology-Construction-Murshed-Singh/350a0af25518382ecc579368b107729f2b68f67f>, Last accessed 2025/12/25.
19. Tudorache, T., Nyulas, C., Noy, N. F., & Musen, M. A. (2013). WebProtégé: A collaborative ontology editor and knowledge acquisition tool for the Web. *Semantic Web*, 4(1), 89–99. <https://doi.org/10.3233/SW-2012-0057>
20. Al-zebari, A., Zeebaree, S., Jacksi, K., Selamat, A.: ELMS-DPU Ontology Visualization with Protégé VOWL and Web VOWL. *Journal of Advanced Research in Dynamical and Control Systems*, 11, 478–485 (2019).
21. Fernandez, M., Gomez-Pearez, A., & Juristo, N.: *Methontology: From Ontological Art Towards Ontological Engineering*. AAAI. (1997). <https://aaai.org/papers/0005-ss97-06-005-methontology-from-ontological-art-towards-ontological-engineering/>, last accessed 2026/01/19.
22. Kaivalyadhama S.M.Y.M. Samiti (Ed.). *Yoga Kośa (Reprint)*. Kaivalyadhama S.M.Y.M. Samiti. Lonavala, Pune (2017).
23. Matentzoglou, N., Malone, J., Mungall, C., Stevens, R.: MIRO: Guidelines for minimum information for the reporting of an ontology. *Journal of Biomedical Semantics*, 9(1), 6 (2018). <https://doi.org/10.1186/s13326-017-0172-7>

Systematic Literature Review of Crop Cultivation Ontologies using PRISMA with Parametric Approach

Anjali Chaurasiya^{1[0009-0003-7673-7789]} Shilpi Verma^{2[0000-0003-3638-5304]} and Ankit Gupta^{3[0009-0007-2878-1506]}

^{1,2} Babasaheb Bhimrao Ambedkar University, Lucknow, India

³University of Delhi, India

anjali@drtc.isibang.ac.in

shilpi@bbau.ac.in

ankit@drtc.isibang.ac.in

Abstract. This study reviewed existing ontologies of crop cultivation practices to highlight their importance and inspire future research, presenting key findings from comprehensive syntheses that address predominant objectives, scope, methodology, tools, evaluation status, and impact. The study used the PRISMA 2020 guidelines to identify forty crop cultivation ontologies from the Google Scholar database, published between 2005 and 2025, and conducted a detailed parametric analysis, demonstrating a rigorous and systematic approach to the review. Knowledge representation and decision support are the predominant objectives across all selected ontologies. Studies mostly favour cereal and vegetable groups, while maize is the most targeted crop, and pest management and fertilisation are typical cultivation stages modelled in the ontology. The review revealed that Methontology is the most preferred methodology, along with Protégé, an ontology editing tool. Most ontologies utilise metric analysis and an application-oriented evaluation method. HermiT and Pellet are preferred for consistency checking, and OOPS! used for scanning pitfalls. The study reveals that the limited scope is a significant limitation, and expanding this scope is the most critical future direction identified in the corpus.

Keywords: Crop Cultivation, Ontology, SLR, PRISMA, Ontology Review.

1 Introduction

Cultivation is one of the oldest and most basic human activities, ensuring food security, supporting livelihoods, and driving socio-economic development. It is a complex, multistage process that includes preparing the land, planting seeds, fertilising the soil, controlling pests, and harvesting mature crops (Global Forum on Food Security and Nutrition (FSN Forum)), to achieve optimal yield. Since the crops themselves are of various kinds, like food crops, cash crops, fibre, oilseed [63], etc., the cultivation practices of each of these crops vary significantly. These practices are influenced by or depend on various factors, including soil type, climatic conditions, terrain type, local food preferences, resource availability, and farmers' knowledge or experience [11,48]. With the advancement in agricultural science and the increasing adoption of digital technologies, crop cultivation has evolved into a data-intensive domain. The diversity of practices and influencing factors results in highly heterogeneous data [64]. These data are characterised by variations in terminology, data formats, and conceptual interpretations across regions and systems, posing substantial challenges to integration, sharing, and reuse [57].

Ontology is the formal representation of knowledge within a specific domain. It offers a powerful mechanism to overcome these challenges by structuring concepts, relationships, and axioms logically [61]. This structure not only enables efficient data integration and retrieval but also supports advanced reasoning and inferencing, facilitating the extraction of new knowledge and insights. Ontologies provide a shared, common understanding for both people and computer systems and play a critical role in enabling interoperability across applications, supporting knowledge reuse, and powering technologies such as the Semantic Web [23].

Within agriculture, crop cultivation ontologies provide standardised knowledge frameworks for representing the concepts, processes, and relationships inherent to crop cultivation [19]. These ontologies support efficient data integration, information retrieval, and reasoning across diverse sources, like soil databases and climate records, to issue advisories and farmer guidelines. For example, a farmer needs precise information about different aspects to make day-to-day decisions, such as "Which crop variety would suit lime soil that needs minimum irrigation?". In this context, implementing a crop cultivation ontology is essential for accurately addressing such complex questions. Furthermore, such ontologies can support precision agriculture tools, decision-support systems and IoT-based automation and artificial intelligence applications to promote sustainable farming practices.

2 Related Work

Prior surveys have addressed broader applications of ontologies in agriculture, but no studies have reviewed crop cultivation ontologies. For instance, various authors, examined various ontology-based applications specifically developed for the agricultural sector and ontologies formulated by developing nations [32]. The use of ontology and semantic techniques in crop pest and disease management by integrating data from IoT sensors, drones, and satellite images into ontology-based systems to enable real-time decision-making and predictive analytics was reviewed [62]. The application of ontologies to improve performance across multiple domains, including education, biomedicine, and agriculture was reviewed [56]. Semantic resources (e.g., semantic web technologies and ontologies) to address data integration and interoperability in innovative agricultural systems, as well as the integration of understandable artificial intelligence. Only an evaluation framework for selecting an evaluation method for agricultural-domain ontologies was provided [26].

Despite these contributions, no comprehensive analysis exists to examine ontologies that systematically represent crop cultivation practices. Existing ontologies focused on specific crops, regions, or cultivation aspects, with varying objectives, methodologies, tools, and evaluation methods. This fragmentation underscores the need for a systematic review to consolidate findings, assess evaluation rigour, and identify research gaps.

There have been various studies in other domains employing a parametric approach to review ontology development, such as conducting a parametric analysis of flood ontologies [50], and a comprehensive review of data mining ontologies was performed too [51]. This study provides a novel analysis of crop cultivation ontologies in systematic reviews, using the PRISMA 2020 guidelines, followed by a parametric analysis. The following research questions guide the review:

RQ 1: What are the primary objectives, targeted crops and cultivation stages of the published crop cultivation ontologies?

RQ 2: Which methodological frameworks, knowledge sources, and ontology editors have been predominantly used in the design and development of crop cultivation ontologies?

RQ 3: How have these ontologies been evaluated, and what are their reported real-world benefits ?

RQ 4: What are the gap areas in existing research and future directions?

3 Methodology

The study employed the Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA-2020) for a Systematic Literature Review (SLR). The PRISMA guidelines provide a workflow for systematically reporting evidence-based items, making it useful for the critical appraisal of published literature [21]. It is a four-step study, comprising identification, screening, eligibility, and inclusion of records (Fig. 1) [5,37].

Initially, we investigated the AGROVOC multidisciplinary thesaurus to determine the definition and scope of the terms such as "crop" and "cultivation." A simple search query, consisting of the terms "crop", "cultivation", and "ontology", was executed on the Google Scholar database. The abstracts and author keywords of the initial 10 articles were examined to identify the most pertinent terms for this study (related and synonymous terms; enumerated in Table 1), while deliberately excluding specific terms, such as "gene", to ensure that the findings exclusively reflect crop cultivation ontologies rather than broader plant sciences.

Table 1. Related and Synonymous Terms.

Term	Related and Synonymous Terms
Crop	plant
Cultivation	cropping system, farming system, crop cultivation
Ontology	knowledge representation, semantic web
Develop	model, design

3.1 Search Strategy

We selected the Google Scholar database, which allows comprehensive searching of scholarly materials from academic publishers, professional associations, online repositories, institutions, and other websites, including articles, theses, books, abstracts, and conference proceedings. Based on Table 1, three repeated searches were conducted in the Google Scholar database using the advanced search option, with different combinations (Table 2).

Filters were applied to a range of publication years (2005-2025) and to exclude patents and citations available in the advanced search option. The final search was conducted on December 4, 2025. However, the result yielded only a few relevant articles; a total of 80 potential articles were obtained and added to the library after screening the titles and abstracts of the top 500 results for each query. Applying the inclusion-exclusion criteria (A, B, H, I, J, and K of Table 3) resulted in 53 articles for full-text download. A targeted full-text review of these papers against the remaining criteria (C, D, E, F, G, L, and M in Table 3), excluding 13 articles, resulted in a final corpus of 40 articles for core analysis.

Table 2. Queries with Responses.

Query	Results
crop cultivation ontology "crop cultivation" OR "cropping system" OR "farming system" -gene*	4,760
crop cultivation ontology* develop* OR design* OR model* OR "crop cultivation" OR "farming system" OR "cropping system" – gene*	17,500
ontology OR "knowledge representation" OR "semantic web" AND crop OR plant AND "crop cultivation" OR "farming system" OR "cropping system" -gene*	2,970

Table 3. Inclusion-Exclusion Criteria.

Inclusion	Exclusion
A. Study published during 2005-2025	H. Book chapter = 2
B. Conference and journal papers	I. Not available in full text = 19
C. Papers dealt with ontologies on crops or any specific aspect of the cultivation process (seed selection, soil preparation, ... to... harvesting).	J. Not written in the English language = 3
D. Developed ontologies should benefit farmers	K. Grey Literature = 2 (Thesis) + 1 (Report)
E. Papers mention the process of ontology development	L. Ontologies that are beyond the cultivation process, for example, Gene Ontology and Food Track & Trace ontology (Pizzuti et al., 2014)
F. Papers provided the ontologies expressed in Web Ontology Language (OWL), Resource Description Framework Schema (RDFS), or any other language	M. Papers discussing only conceptual or theoretical frameworks.
G. Papers discuss at least the top-level entities and attributes	

3.2 Analysis

To systematically address the RQs, we conducted a detailed parametric content analysis on the final corpus of selected ontology papers. This approach is established in the field of structured ontology reviews, as demonstrated by prior research that applied a parametric lens to analyse a set of domain ontologies. Our analytical framework was synthesised from existing parameters related to ontology description and documentation [8,22] and ontology review [50]. Despite these, we developed and integrated several novel parameters to capture domain-specific and practical dimensions critical to this study. For example, crop group, target crop (inspired by Crop Group Tables – IR-4 Project), knowledge sources, evaluation results, benefits, limitations, and future directions. All parameters, along with their descriptions and precise coding rules, are catalogued [10]. Coding rules were developed and refined through discussion among the research team and using an AI language model, DeepSeek (<https://www.deepseek.com/en/>). Subsequently, to validate the framework, sample data was extracted from 5 papers, which led to adjustments to address any ambiguities. Once the final protocol was finalised, the extracted data was compiled and analysed using Microsoft Excel. The software enables the quantitative and qualitative synthesis of parametric data by applying analytical techniques, such as frequency analysis, to identify patterns and gaps across parameters. Furthermore, Lucid Chart (https://lucid.app/documents#/home?folder_id=recent) and Microsoft Excel were used for data visualisation.

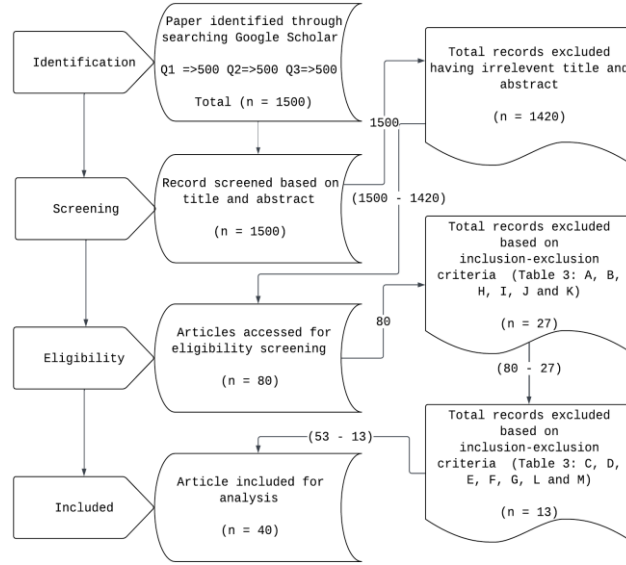


Fig. 1. Literature Identification Using PRISMA.

4 Result and Discussion

4.1 RQ 1

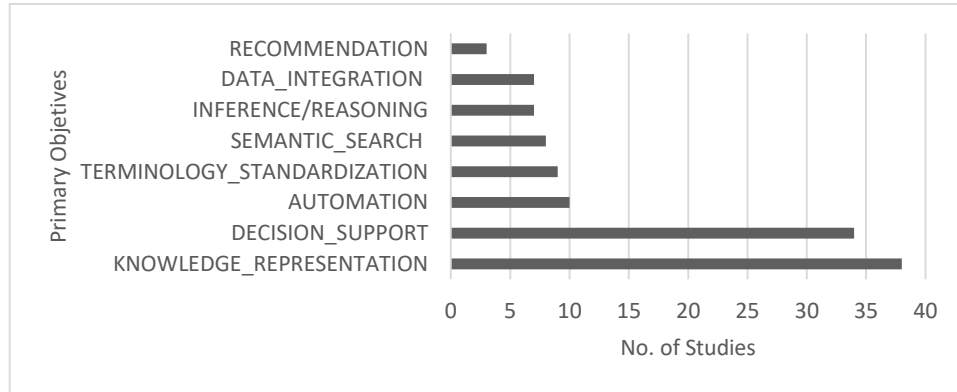


Fig. 2. Frequency of core objectives across the selected ontology studies.

Primary objectives. Analysis of selected crop cultivation ontologies, identifying 9 primary objectives. Their frequencies reveal a clear hierarchy of research focus. Analysis shows that an ontology may have multiple objectives. Fig. 2 shows that knowledge representation and decision support are the most prevalent objectives, appearing in 38 and 34 studies, respectively. This dual focus addresses both the complexity of crop cultivation knowledge and the external demand for tools that enable farmers to understand the complex environment and decision support. Other important goals include automation, standard terminology, and semantic searching, which highlight critical, ongoing issues in achieving interoperability across data sources and systems. On the other hand, inference/reasoning, data integration, recommendation and knowledge management indicate emerging applications and future expansion for precision and intelligent farming systems.

Targeted Crop. The analysis of the corpus highlights a strong and diverse research focus across crop groups, emphasising the breadth of current studies (Fig. 3). While 8 ontologies focus on multiple crop groups [24, 47], the remaining target a single crop group [1,13,14,29]. Cereals and vegetables are the most frequently modelled groups, each represented in 15 studies, highlighting their prominence. Maize (modelled in 7 ontologies) is the most prominent crop among all crop species (Fig. 3), followed by wheat (5) and rice (5). Vegetables are the most diverse crop group, where 14 individual species are modelled. Tomatoes (represented in 4 ontologies) are the most preferred vegetables, closely followed by soybeans (3); the remaining vegetables are modelled in a single ontology. The fruit group is modelled in 8 studies, with notable examples including grapes (4), strawberries, bananas,

and olives. The 'Other' category primarily encompasses commercial crops, including cotton (2), coffee (2), and tobacco (2), as well as flowers (2) and herbs (1). While 6 ontologies did not specify a particular crop, 2 mentioned a crop group; these demonstrate more generic, task-oriented ontologies engineered for cultivation stage(s), irrespective of the crop.

Significant attention to cereal crops such as maize, rice, and wheat supports the optimisation of the staple food system. Conversely, ontologies for high-value horticulture crops, such as fruits and vegetables, and for industrial crops, such as coffee and cotton, reflect a parallel research focus on enhancing management and profitability in specialised sectors. Factors like economic importance, data availability, and research tradition influence the dominance of specific crops in ontology research. Focusing on major staples and high-value cash crops ensures relevance and facilitates funding and collaboration. However, many nutrition-rich but less commercially dominant crops, such as millets, remain underrepresented, creating a notable gap. Consequently, the study suggests an important avenue for future research and the comprehensive semantic representation of these "orphan crops", with the aim of building a more equitable and nutrition-sensitive semantic resource in the agricultural domain.

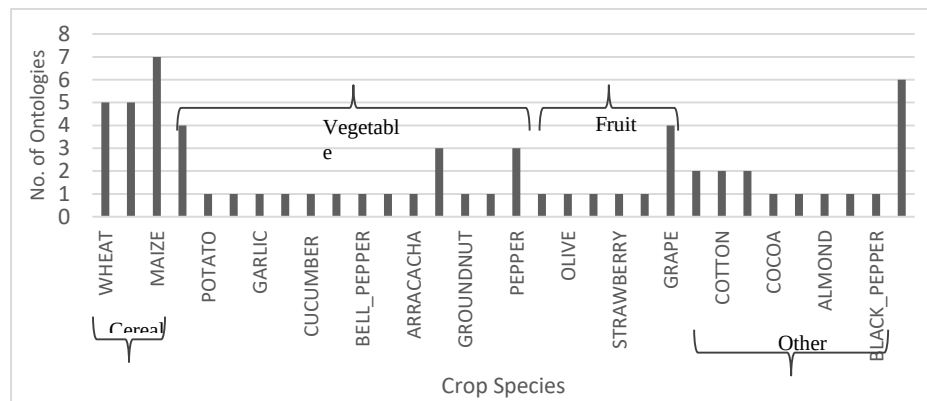


Fig. 3. Distribution of crop-specific focus across the reviewed ontologies.

Cultivation Stages. The analysis reveals the scope of the reviewed ontologies, based on criteria such as activity type, crop lifecycle coverage, and focus on specific cultivation stages, with a focus on crop protection and resource management practices.

As shown in Fig. 4, pest management is the most frequently addressed stage (represented in 22 ontologies), followed by fertilisation (20) and irrigation (16). A pronounced focus on knowledge-intensive stages underscores their critical status, where precise intervention has a direct and significant impact on yield quality, quantity, and cost, as well as on the environment. Other stages, such as soil preparation (modelled in 14 ontologies), planting (15), disease management (15), harvesting (12), and post-harvesting (12), are also modelled significantly. The strong representation of these stages indicates a shift towards utilising ontologies as a knowledge base for decision support and recommendation systems, as well as the automation of repetitive tasks. In contrast, seed selection (modelled in 9 ontologies) receives moderate attention, while activities such as crop identification, sowing, and growth management receive less attention (each with $n \leq 3$). The 'Other' (in 3 ontologies) category indicates a minor aspect or supporting activity, such as land use, pesticide life cycle, and soil type. Some authors do not specify any cultivation stage, as it represents knowledge about pesticides intended to make information accessible and interoperable [40].

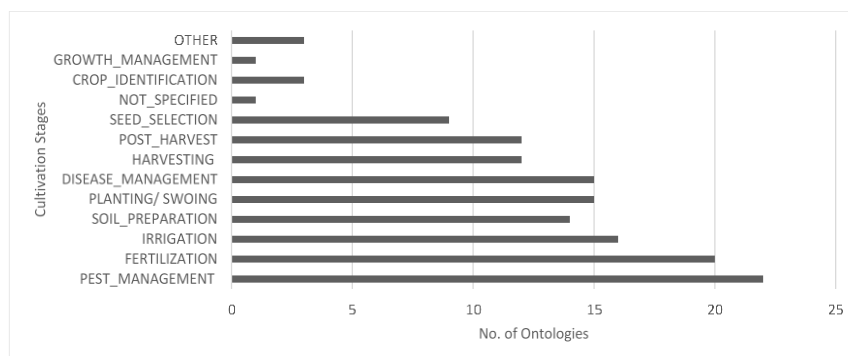


Fig. 4. Frequency of modelled cultivation stages across the selected ontologies.

Methodological Framework. A methodology framework provides the structured, established, and reliable approaches to guide the ontology development (OD) process. In the corpus, only 22 ontologies (55%) explicitly referenced a formal methodology. Among these, the majority (17 ontologies) employed a single framework; for example, ‘Methontology’, ‘Ontology Development 101’, and ‘LOT’ [55, 24, 44]. The remaining 5 ontologies have adopted a hybrid approach by integrating two or more methodologies, like ‘Grüninger & Fox’, ‘Methontology’, and ‘FAO guidelines’ [1]. Some authors employed both ‘TOVE’ and ‘Methontology’ [34], while others used ‘NEON’, ‘LOT’, and ‘SAMOD’ [19].

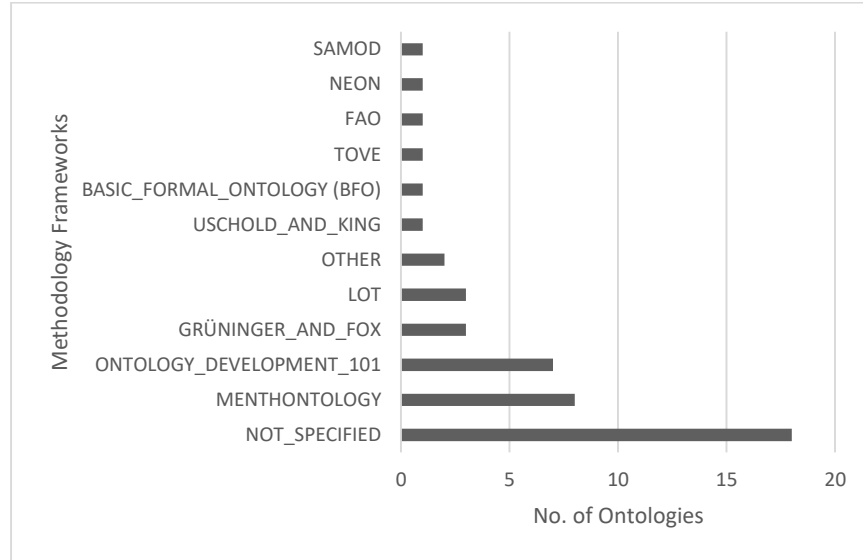


Fig. 5. Preferred ontology development methodologies in the reviewed studies

As illustrated in Fig. 5, ‘Methontology’ (referenced in 8 ontologies) [50] and ‘Ontology Development 101’ (7) are the predominant frameworks among the reviewed ontologies, underscoring their widespread acceptance. ‘LOT’ and ‘Grüninger & Fox’ were each employed in 3 ontologies. The infrequent use of foundational ontologies like BFO and domain-specific guidelines from FAO indicates a significant gap in adopting structural and interoperability practices that are not yet mainstream. Furthermore, specialised frameworks like ‘SMOD’, ‘NEON’, ‘TOVE’, and ‘Uschold and King’ were each used in only a single ontology, highlighting their limited adoption despite their potential niche value. The category ‘Other’ encompasses methodologies not listed in the code rules (e.g., the OBO Foundry framework) or entirely novel ontology development processes (e.g., the six-step iterative ontology engineering principle) [2,3].

The high frequency of ‘Not Specified’ (18 studies) is a critical finding, highlighting a significant gap in transparency of reporting and methodological rigour. This omission hinders the reproducibility and cumulative advancement of the field. The co-dominance of ‘Methontology’ and ‘Ontology Development 101’ indicates a bimodal community for semantic modelling in the agricultural domain, comprising both ontology engineers and domain experts. The use of ‘BFO’ and ‘FAO’ demonstrates the practices for achieving interoperability and integration with a wide array of domain-specific semantic knowledge resources.

The study recommends explicit documentation of the adopted methodology to ensure transparency and standardisation, and to support the use of hybrid approaches. For instance, integrating the competency question-driven scoping approach of ‘Grüninger & Fox’ with the detailed implementation steps of ‘Methontology’ would enhance the methodological robustness and flexibility required for complex domains, such as agriculture.

Knowledge Sources and Acquisition Method: The knowledge sources and acquisition methods determine the authenticity, completeness, and accuracy of an ontology that directly affect the system’s reasoning, inference, and decision-making. The analysis shows that while some rely solely on existing literature [12], most studies cite multiple knowledge sources. Fig. 6 shows that the existing literature (e.g., academic papers, textbooks) is the most relied-upon source (cited in 32 ontologies), followed by domain experts (25), including working professionals, researchers, and farmers. In contrast, farmers’ experience is involved infrequently. Government guidelines in 21 studies [58] underpinned the standardisation process, such as an initiative led by the South Korean government for vertical farms. The integration of databases (16) and existing ontologies (13) significantly promotes reusability and data-driven modelling. However, direct input from sensor data (6) is infrequent.

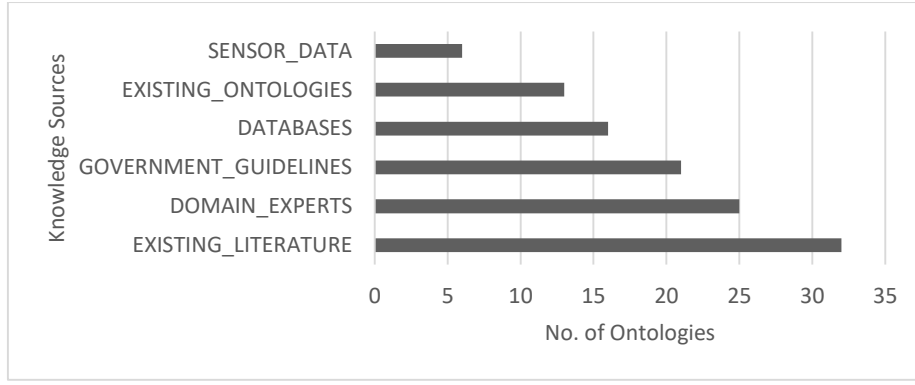


Fig. 6. Knowledge sources referenced across the reviewed ontologies.

The emphasis on existing literature and government guidelines indicates that ontological research in this domain aims to formalise established, codified knowledge. This approach is practical for creating reliable, evidence-based models, but may lag cutting-edge fields or local tacit knowledge [15]. Recognising the value of indigenous knowledge and local practices can foster a sense of respect and relevance for practitioners. The significant role of domain experts is a positive indicator of relevance and validated knowledge. However, the under-utilisation of sensor data represents the gap between ontology research and IoT-enabled systems for real-time data processing in precision farming.

The selection of a knowledge acquisition method depends on the format of the source material (e.g., literature review or text extraction for textual documents), the nature of the target knowledge holders (e.g., interviews or workshops for domain expert groups), and the available budget. Fig. 7 illustrates the literature review (34) as the predominant method, reflecting the foundational reliance on codified, published knowledge within the domain. To structure new knowledge and reuse existing semantic and structured data, ontology integration (13) and database curation (11) were key technical methods. Among human-centric approaches, interview (11) is the most frequently used technique for eliciting tacit knowledge from practitioners, surpassing workshop (3) and questionnaire (3). Only two ontologies utilise text extraction techniques for automated or semi-automated concept harvesting. Some authors employed JRE and Naive Bayes [20], and some used KEA and Stanford Parser [12]. Their limited adoption highlights a missed opportunity to improve the efficiency and scalability of knowledge acquisition.

Findings suggest prioritising methods, such as crowdsourcing, to seamlessly integrate indigenous farmers' knowledge and high-frequency, real-time sensor data into ontological frameworks. It would support dynamic, adaptive ontologies that reflect real-world variables and enable real-time decision-making.

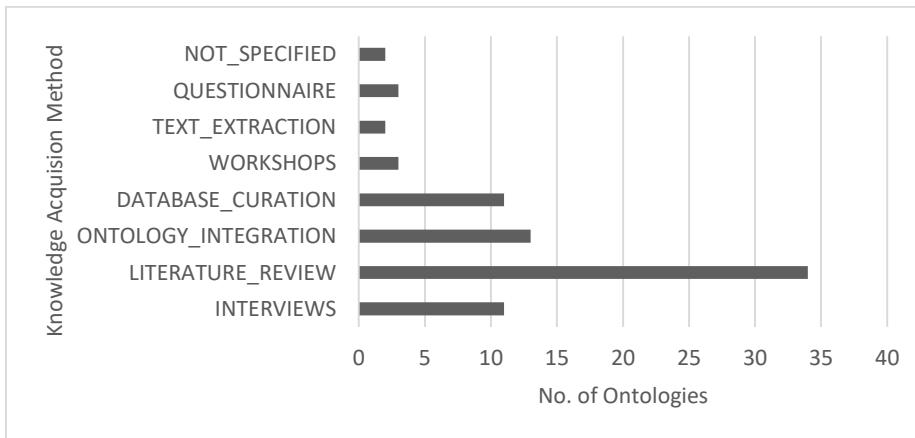


Fig. 7. Knowledge acquisition methods employed in the reviewed ontologies.

Ontology Editor. Fig. 8 presents the distribution of preferred ontology editing tools among the reviewed studies. Protégé, an open-source software, is the most preferred editor [50] and is used in 80% (32) of the studies. Its popularity is attributed to its extensibility via plugins, robust support for the Web Ontology Language (OWL), an intuitive interface, and a large user community. All other tools, including freeware such as Hozo, commercial options such as Topbraid Composer, and domain-specific vocabulary platforms such as AGROVOC Concept

Server Workbench, were used in a single study each. Five studies (12.5%) do not explicitly mention any ontology editor, however the screenshots used inside the articles often indicate an informal adaptation of Protégé. The widespread use of Protégé, supported by tools such as OWLReady2 and Chowlk Converter [35], demonstrates that accessible, well-supported tools can empower users and foster confidence in ontology development.

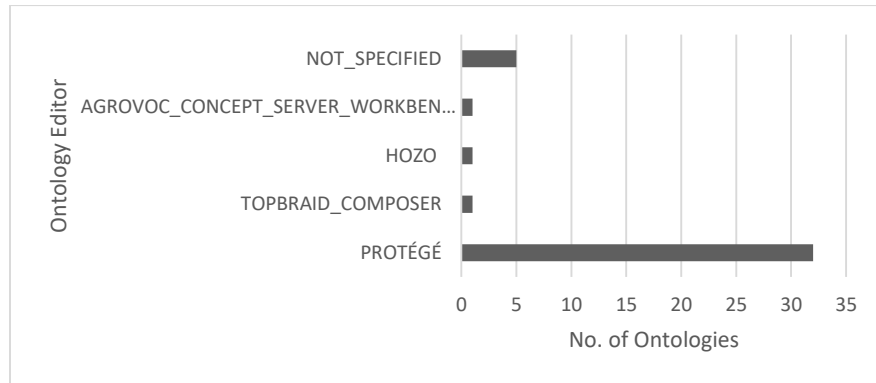


Fig. 8. Preferred ontology editing tools among the selected studies

4.2 RQ 3

Evaluation Status. A systematic, multifaceted evaluation is paramount to ensure the correctness, completeness, and practical utility of an ontology. This study reveals strong assessment within the field: 90% (36) of studies employ one or more approaches. However, 4 ontologies (10%) are not evaluated, highlighting a significant gap in methodological accountability that may impact the overall reliability of current practices.

Analysis highlights diverse combined approaches applied for evaluation, with 16 studies (40%) using more than 3 approaches. However, some authors conducted the most comprehensive evaluation, integrating all six approaches. Conversely, 9 ontologies (22.5%) relied on only one method, which limits the assessment depth. Fig. 9 illustrates the choice of individual approach among selected studies [29]. Metrics, including both quantitative performance and qualitative outcome narratives, were the most prevalent approach, featured in 25 ontologies (62.5%). Application-based evaluation (52.5%) is used to test ontologies within a functional system. The foundational approaches, such as competency question fulfilment (47.5% of ontologies), consistency checking by the reasoner (42.5%), and expert review (37.5%), were also widely utilised. However, challenges such as resource intensiveness and subjectivity can affect their consistency and reliability.

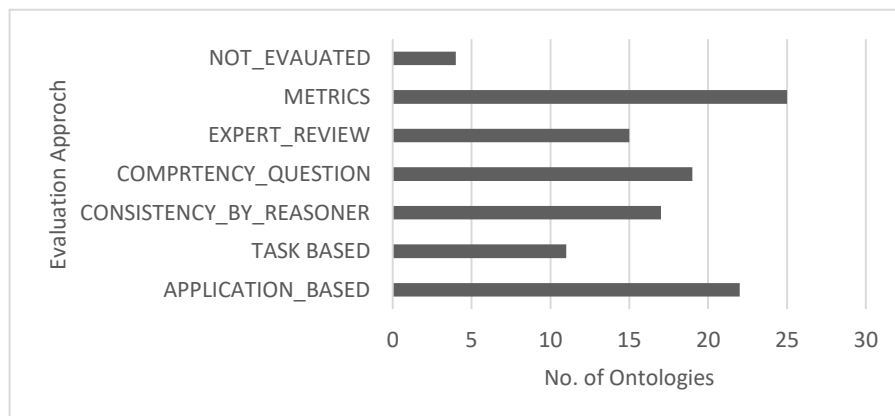


Fig. 9. Evaluation approaches employed in the reviewed ontologies.

Tools Used. Analysis reveals that automated reasoners are primarily used to verify consistency and perform inference. HermiT (employed in 11 ontologies) and Pellet (employed in 10) are the most prevalent automated reasoners, often used individually or in combination [50,51]. Their widespread adoption is facilitated by integrating them as plugins into the dominant ontology editor, Protégé. Other reasoners, such as Racer and ELK, were rarely used. Beyond these core reasoners, other specific validation and analysis tools are also employed. For example, OOPS! (OntOlogy Pitfall Scanner): a web-based, environment-independent tool for detecting potential pitfalls or syntax modelling errors was significantly leveraged [19,36]. Furthermore, some studies utilise tools outside the

Semantic Web stack to improve rule-based reasoning. For example, some researchers used Drools, a Java-based business rules management system [2,3,54]. JESS (Java Expert System Shell), a Java-based rule engine, was referenced by some, [29,53] while some authors used Weka, an open-source machine learning software, to analyse and identify relationships within a structured database [14,42]. Other tools, including RDFS Validator, Allegro-Graph 6.1, and YOLOv8, were each referenced in a single study.

Evaluation Result and Benefit Reported. Analysis of the evaluation results reveals that 85% (34) of the studies reported results in either qualitative or quantitative form, while 15% (6) did not. However, two of these state that they are under evaluation, without presenting any findings [38,54]. The reported results/benefits can be consolidated into two categories:

Quantitative and Technical Validation Outcome. The quantitative measure presents metrics that validate the correctness and accuracy of the ontology. A significant subset of reviewed studies referenced standard classification and information retrieval metrics. For instance, the precision score ranged from 0.70 to 0.98 [20,47], the recall ranged from 0.64 to 0.96 [47,59], and the F-measure ranged from 0.716 to 0.95 [20,47]. Other specific accuracy metrics for different tasks are also reported, such as for recommendations (for zone: 0.85, crop: 0.76, and fertiliser: 97.1%) [47,65], and location mapping (100%) [40]. However, formal validation through competency questions and reasoners yielded strong results with 100% recall on a test query [59], 89% (80/89) from a fertilisation system [65], and the successful execution of all 44 competency questions in one study [2,3]. Numerous studies reported logical consistency using reasoners such as HermiT and Pellet, with no contradictions or inconsistencies [19,24]. The technical validation is often supported by expert review, which confirms the practical accuracy of the outputs, with a suitability of 97.1% for fertiliser recommendations [65].

Qualitative and Real-World Benefits. Practical benefits are captured through application and task-based evaluation, with improvements reported when ontologies are implemented as a knowledge base. The most frequently reported benefit is the reduction in task completion time from over 15 minutes to 1.5 minutes [25]. Successful system automation has been achieved to control irrigation and fertilisation in a hydroponic setup [35], providing a precise watering schedule through sensor networks [45]. Furthermore, an ontology enhances the retrieval of cultivation standards through semantic search [7,34], improving the quality of large language model outputs [39] and addressing interoperability challenges. For example, PestOn achieved over 99% alignment of active ingredients across disparate pesticide registries [40].

4.3 RQ 4

Limitations and Future Directions Reported in the Corpus. A rigorous analysis of the limitations identified by authors reveals the current challenges within the domain. All reported limitations across 29 studies are grouped into seven primary themes: limited scope, data issue, modelling issue, evaluation gap, performance issue, practical implementation issue, and evaluation gap.

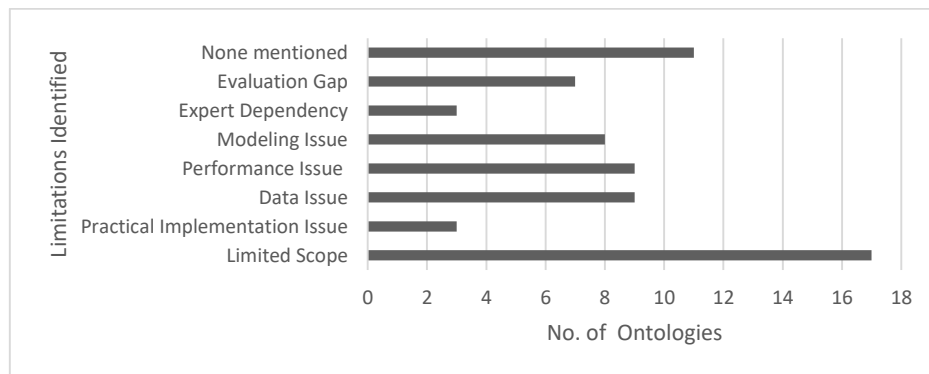


Fig. 10. Documented limitations within the reviewed ontologies.

Fig. 10 shows that a limited scope is the most common problem reported in 17 (42.5%) studies. Limited scope manifests as ontologies constrained to a single crop or region, leading to fragmented research and non-interoperable solutions, while also revealing a deep focus on a specific concept/problem. For example, some studies are limited to the Thai Orchid [43] or focus exclusively on citrus in the mountainous region [65]. Furthermore, many systems are designed for a narrow functional scope, supporting only factoid and list-type questions [59] or excluding key cultivation stages, such as post-harvest processes [39].

Data-related and performance issues were the most significant challenges, each reported in 9 (22.5%) studies. It encompasses dependencies on limited initial dataset (s) [47] or concerns about data quality, such as human

errors in source datasets [44]. Performance limitations include algorithmic inaccuracies, where performance varies across sub-domains [20], and rule-based inconsistencies, where multiple pests share common symptoms.

Modelling issues and evaluation gaps are reported in 8 (20%) and 7 (17%) studies, respectively. These challenges included a lack of alignment with foundational ontologies, such as BFO [18] and technical inconsistencies, including naming conflicts between modules [19].

Evaluation shortcomings underscored the research's transitional phase, necessitating real-world validation [39] and addressing the constrained testing scale [45]. Moreover, difficulties in practical implementation and reliance on experts are infrequently cited. Practical implementation encompasses issues that limit the system's application of an ontology [12,45], such as limited public availability and concerns about reliance on experts, including the lack of validation from domain expertise [29,38].

Future directions reported in the reviewed studies outline a coherent, forward-looking research agenda that directly addresses the field's limitations. These recommendations were synthesised into five strategic themes for analysis.

Expanding scope: Ontology development is a continuous process. As the domain expands, ontologies need to be extended in parallel to broaden their applicability and remain relevant in the future. It entails the inclusion of a broader range of crops [29,60], various cultivation stages [39], distinct geographical regions, and diverse regulatory frameworks [40].

Improving interoperability: To enhance integration within the broader semantic web, an ontology should align with established top-level and domain ontologies and standards. For example, some authors recommended aligning C3PO with the BFO model published on AgroPortal [18], and some suggested improving technical interoperability through standards like XTM [1].

Integrating advanced computational techniques: A review of studies reveals a clear consensus on leveraging modern AI and data science techniques, including automated knowledge extraction from text and ML/NLP integration for rule discovery [47], enhancing system intelligence, and real-time inferencing [28].

Focusing on practical implementation: Numerous studies emphasise the real-world application of ontology-based systems that have a user-friendly design, encompassing a Java-based GUI [2,3] and mobile-based web applications [59], as well as creating natural language interfaces to answer farmers' questions in their native language. Furthermore, integrating ontologies with IoT platforms and sensor data enables real-time, data-driven operations, such as crop monitoring and irrigation scheduling [43,65].

Comprehensive evaluation: The reviewed studies placed significant importance on the hybrid evaluation approach, merging technical assessment with a strong focus on pilot testing and end-user experience analysis to assess both agronomic and practical impacts. For example, conducting pilot studies and validation was suggested [39]; deploying the ontology-integrated system at the experimental farm was emphasised [44]; and evaluations based on orchid specialists and farmer surveys were highlighted [43].

5 Limitations and way forward

Several limitations must be kept in mind when reading the study's findings. The scope of the study was limited to only the Google Scholar database with the top 500 relevant papers for each three-round query. For this study, we manually identified literature using inclusion-exclusion criteria, so there may be a chance of missing important studies that could influence the results. As this study is preliminary, it will be expanded by collecting literature from other databases and integrating additional parameters to assess availability, interoperability, and reuse.

6 Conclusion

This review used the PRISMA 2020 guidelines and a parametric approach to analyse existing literature, aiming to answer four questions comprising key findings of comprehensive syntheses with predominating objectives, scope, methodology, tools, evaluation status and impact. The study laid the foundation for reviewing crop ontologies using a parametric approach. The finding highlights maize as the most targeted crop, methontology as the most preferred methodology, and protégé as the most used ontology editor. The primary constraints revealed by the analysis include fragmented research on specific crops, neglect of 'orphan crops' and farmers' local knowledge in ontology modelling and a lack of methodological transparency and of comprehensive practical evaluation through farmers' experiences.

To advance the field, future research must prioritise: a) developing more holistic and interoperable ontologies, covering the full cultivation lifecycle aligned with top-level ontologies like BFO, b) adopting and properly documenting hybrid methodology frameworks, c) bridging the gap between knowledge models and real-world data by

incorporating farmers local knowledge and, d) employing comprehensive evaluation by integrating technical methods along with end-user experience.

7 Acknowledgement

Raw data and analysis were published in FigShare, a data repository accessible via DOI: 10.6084/m9.figshare.31146349.

Declaration of AI and AI-Assisted Technologies used in Manuscript Writing Process: This study utilised Grammarly and QuillBot for language assistance, the DeepSeek AI language model for drafting coding rules and Zotero for reference management.

References

1. Aminu, E. F., Oyefolahan, I. O., Abdullahi, M. B., & Salaudeen, M. T. (2019). An OWL Based Ontology Model for Soils and Fertilizations Knowledge on Maize Crop Farming: Scenario for Developing Intelligent Systems. *2019 15th International Conference on Electronics, Computer and Computation (ICECCO)*, 1–8. <https://doi.org/10.1109/ICECCO48375.2019.9043214>
2. Aminu, E. F., Oyefolahan, I. O., Abdullahi, M. B., & Salaudeen, M. T. (2022a). MaConto: A robust maize crop ontology based on soils, fertilisers and irrigation knowledge. *Intelligent Systems with Applications*, 16, 200125. <https://doi.org/10.1016/j.iswa.2022.200125>
3. Aminu, E. F., Oyefolahan, I. O., Abdullahi, M. B., & Salaudeen, M. T. (2022b). Modeling Competency Questions-Based Ontology for the Domain of Maize Crop: SIMcOnto. In J. K. Mandal, R. Buyya, & D. De (Eds.), *Proceedings of International Conference on Advanced Computing Applications* (Vol. 1406, pp. 751–763). Springer Singapore. https://doi.org/10.1007/978-981-16-5207-3_61
4. Aree Thunkijjanukij, Asanee Kawtrakulj, Supamard Panichsakpatana, & Uamporn Veesommai. (2009). Ontology Development: A Case Study for Thai Rice. *Agriculture and Natural Resources*, 43(3), 594–604.
5. Ashiq, M., Jabeen, F., & Mahmood, K. (2022). Transformation of libraries during Covid-19 pandemic: A systematic review. *The Journal of Academic Librarianship*, 48(4), 102534. <https://doi.org/10.1016/j.acalib.2022.102534>
6. Ashiq, M., Rehman, S. U., Safdar, M., & Ali, H. (2021). Academic library leadership in the dawn of the new millennium: A systematic literature review. *The Journal of Academic Librarianship*, 47(3), 102355. <https://doi.org/10.1016/j.acalib.2021.102355>
7. Bangkhomned, W., & Payakpate, J. (2020). Applying Ontology Knowledge Representation Technology and Semantic Searching Methods to Support the Production of High Quality Longan Fruit. In K. J. Kim & H.-Y. Kim (Eds.), *Information Science and Applications* (Vol. 621, pp. 601–612). Springer Singapore. https://doi.org/10.1007/978-981-15-1465-4_59
8. Biswanath, D., Durgesh, N., & Gautam, S. (2015, September 5). MOD: Metadata for Ontology Description and Publication. *Proceedings of the International Conference on Dublin Core and Metadata Applications*. <https://doi.org/10.23106/DCMI.952136974>
9. Chaudhary, S., Bhise, M., Banerjee, A., Goyal, A., & Moradiya, C. (2015). Agro advisory system for cotton crop. *2015 7th International Conference on Communication Systems and Networks (COMSNETS)*, 1–6. <https://doi.org/10.1109/COMSNETS.2015.7098701>
10. Chaurasiya, A., Verma, S., & Gupta, A. (2026). *Table consisting of Parameters for Reviewing Crop Cultivation Ontologies* (p. 17470 Bytes) [Dataset]. figshare. <https://doi.org/10.6084/M9.FIGSHARE.31146259>
11. Choden, T., & Ghaley, B. B. (2021). A Portfolio of Effective Water and Soil Conservation Practices for Arable Production Systems in Europe and North Africa. *Sustainability*, 13(5), 2726. <https://doi.org/10.3390/su13052726>
12. Chougule, A., Jha, V. K., & Mukhopadhyay, D. (2016). Using IoT for integrated pest management. *2016 International Conference on Internet of Things and Applications (IOTA)*, 17–22. <https://doi.org/10.1109/IOTA.2016.7562688>
13. Chougule, A., Jha, V. K., & Mukhopadhyay, D. (2019a). Decision support for grape crop protection using ontology. *International Journal of Reasoning-Based Intelligent Systems*, 11(1), 24. <https://doi.org/10.1504/IJRIS.2019.098051>
14. Chougule, A., Jha, V. K., & Mukhopadhyay, D. (2019b). Decision support for nutrition management of grapes using ontology based on decision trees. *International Journal of Information and Decision Sciences*, 11(3), 234. <https://doi.org/10.1504/IJIDS.2019.101993>
15. Collins, H. (2010). *Tacit and Explicit Knowledge*. University of Chicago Press. <https://doi.org/10.7208/chicago/9780226113821.001.0001>
16. Comment view | Global Forum on Food Security and Nutrition (FSN Forum). (n.d.). <https://www.fao.org/fsnforum/comment/7649>
17. Crop Group Tables – IR-4 project. (n.d.). <https://www.ir4project.org/fc/crop-grouping/crop-group-tables/>
18. Darnala, B., Amardeilh, F., Roussey, C., & Jonquet, C. (2021, September 11). Crop Planning and Production Process Ontology (C3PO), A New Model to Assist Diversified Crop Production. *CEUR Workshop Proceedings*. 2nd Integrated Food Ontology Workshop, held at JOWO 2021, Bolzano, Italy.

19. Darnala, B., Amardeilh, F., Roussey, C., Todorov, K., & Jonquet, C. (2023b). C3PO: A crop planning and production process ontology and knowledge graph. *Frontiers in Artificial Intelligence*, 6, 1187090. <https://doi.org/10.3389/frai.2023.1187090>
20. Deepa, R., & Vigneshwari, S. (2022). An effective automated ontology construction based on the agriculture domain. *ETRI Journal*, 44(4), 573–587. <https://doi.org/10.4218/etrij.2020-0439>
21. Dehdarirad, H., Ghazimirsaeid, J., & Jalalimanesh, A. (2020). Scholarly publication venue recommender systems: A systematic literature review. *Data Technologies and Applications*, 54(2), 169–191. <https://doi.org/10.1108/DTA-08-2019-0135>
22. Dutta, B., Toulet, A., Emonet, V., & Jonquet, C. (2017). New Generation Metadata Vocabulary for Ontology Description and Publication. In E. Garoufallou, S. Virkus, R. Siatry, & D. Koutsomiha (Eds.), *Metadata and Semantic Research* (Vol. 755, pp. 173–185). Springer International Publishing. https://doi.org/10.1007/978-3-319-70863-8_17
23. Fensel, D., Van Harmelen, F., Horrocks, I., McGuinness, D. L., & Patel-Schneider, P. F. (2001). OIL: An ontology infrastructure for the Semantic Web. *IEEE Intelligent Systems*, 16(2), 38–45. <https://doi.org/10.1109/5254.920598>
24. Gawich, M., Lahoud, C., Baazaoui, H., & Jomaa, I. (2024). Towards an Ontology-Driven System For Building and Farming Greenhouses. *Procedia Computer Science*, 246, 3208–3217. <https://doi.org/10.1016/j.procs.2024.09.319>
25. Goldstein, A., Fink, L., Raphaeli, O., Hetzroni, A., & Ravid, G. (2019). Addressing the 'Tower of Babel' of pesticide regulations: An ontology for supporting pest-control decisions. *The Journal of Agricultural Science*, 157(6), 493–503. <https://doi.org/10.1017/S0021859619000820>
26. Goldstein, A., Fink, L., & Ravid, G. (2021). A Framework for Evaluating Agricultural Ontologies. *Sustainability*, 13(11), 6387. <https://doi.org/10.3390/su13116387>
27. Grati, R., Fattouch, N., & Boukadi, K. (2025). Ontologies for Smart Agriculture: A Path Toward Explainable AI—A Systematic Literature Review. *IEEE Access*, 13, 72883–72905. <https://doi.org/10.1109/ACCESS.2025.3563202>
28. Helfer, G. A., Da Costa, A. B., & Barbosa, J. L. V. (2025). SoilBR-Onto: An Ontology for Soil Fertility Management and Classification in Brazil. *Applied Ontology*, 20(2), 181–196. <https://doi.org/10.1177/15705838251315571>
29. Hifza Afzal. (2021). An Ontology-Driven Decision Support System for Rice Crop Production. *Journal of Applied and Emerging Sciences*, 11(01).
30. Hlomani, H., & Stacey, D. (2014). Approaches, methods, metrics, measures, and subjectivity in ontology evaluation: A survey. *Semantic Web Journal*, 1(5), 1–11.
31. Institute of Electrical and Electronics Engineers (Ed.). (2015). *Proceedings, 2015 International Conference on Man and Machine Interfacing (MAMI): 17-19 December 2015, Bhubaneswar, Odisha, India*. International Conference on Man and Machine Interfacing, Piscataway, NJ. IEEE.
32. Jebaraj, J., Sathiaselvan, J. G. R., & Department of Computer of Science Bishop Heber College, Trichy, Tamilnadu, India. (2017). An Exploratory Study on Agriculture Ontology: A Global Perspective. *International Journal of Advanced Research in Computer Science and Software Engineering*, 7(6), 202–206. <https://doi.org/10.23956/ijarcse/V7I6/0148>
33. Khan, F. A., Adamu Abubakar, Marwan Mahmoud, Mahmoud Ahmad Al-Khasawneh, & Ala Abdulsalam Alarood. (2019). Cotton Crop Cultivation Oriented Semantic Framework Based on IoT Smart Farming Application. *International Journal of Engineering and Advanced Technology (IJEAT)*, 8(3). https://www.researchgate.net/profile/Fazeel-Khan/publication/331500877_Cotton_Crop_Cultivation_Oriented_Semantic_Framework_Based_on_IoT_Smart_Farming_Application/links/5c7d61d9299bf1268d390750/Cotton-Crop-Cultivation-Oriented-Semantic-Framework-Based-on-IoT-Smart-Farming-Application.pdf
34. Li, D., Kang, L., Cheng, X., Li, D., Ji, L., Wang, K., & Chen, Y. (2013). An ontology-based knowledge representation and implement method for crop cultivation standard. *Mathematical and Computer Modelling*, 58(3–4), 466–473. <https://doi.org/10.1016/j.mcm.2011.11.004>
35. Lin, Y.-J., & Tu, Y.-M. (2025). A Scalable IoT-driven Smart Agriculture System: Ontology-based Inference and Automation for Hydroponic Farming. *Sensors and Materials*, 37(3), 883. <https://doi.org/10.18494/SAM5410>
36. Malik, N., Hijam, D., & Sharan, A. (2021). Ontology based knowledge representation: Case study from agriculture domain. *International Journal of Knowledge-Based and Intelligent Engineering Systems*, 25(1), 97–108. <https://doi.org/10.3233/KES-210055>
37. Mamta, & Kumar, V. (2024). A systematic review of library service quality studies: Models, dimensions, research populations and methods. *Journal of Librarianship and Information Science*, 56(2), 534–546. <https://doi.org/10.1177/09610006221148190>
38. Mawardi, S. Y., Abdel-Aziz, M. H., Omran, A. M., & Mahmoud, T. A. (2013). An Ontology-Driven decision support system for wheat production. *International Journal of Computer Science and Telecommunications*, 4(8), 11–12.
39. Medeiro Alves, L. F., De Oliveira, J. M. P., Bonacin, R., & De Franco Rosa, F. (2025). An Ontology of Tobacco Production: Enriching Large Language Model-based Decision Support. *Revista de Informática Teórica e Aplicada*, 32(3), 102–111. <https://doi.org/10.22456/2175-2745.146658>
40. Medici, M., Dooley, D., & Canavari, M. (2022). PestOn: An Ontology to Make Pesticides Information Easily Accessible and Interoperable. *Sustainability*, 14(11), 6673. <https://doi.org/10.3390/su14116673>
41. Easwaran, S. & Thottupuram, Fr. Rubin. (2011). Farm-Agro Ontology formation: A black pepper model. *International Journal of Research and Reviews in Information Technology*, 01(01).
42. Murali, E., & Anuncia, S. M. (2023). An Ontology-based Knowledge Mining Model for Effective Exploitation of Agro Information. *IETE Journal of Research*, 69(11), 7856–7873. <https://doi.org/10.1080/03772063.2022.2058629>

43. Nattapong, K., Wirapong, C., & Marut, B. (2020). Ontology-based big data analysis for orchid smart farming. *Library and Information Science Research E-Journal*, 29(2). <https://doi.org/10.32655/LIBRES.2019.2.2>
44. Nguyen, Q.-D., Roussey, C., Poveda-Villalón, M., De Vaulx, C., & Chanet, J.-P. (2020). Development Experience of a Context-Aware System for Smart Irrigation Using CASO and IRRIG Ontologies. *Applied Sciences*, 10(5), 1803. <https://doi.org/10.3390/app10051803>
45. Panawong, N., & Namahoot, C. S. (2017). Cultivation of Plants Harnessing an Ontology-Based Expert System and A Wireless Sensor Network. *Journal of Telecommunication, Electronic and Computer Engineering*, 9(2–3), 109–113.
46. Raad, J., & Cruz, C. (2015). A Survey on Ontology Evaluation Methods: *Proceedings of the 7th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management*, 179–186. <https://doi.org/10.5220/0005591001790186>
47. Riaño, M. A., Rodriguez, A. O. R., Velandia, J. B., García, P. A. G., & Marín, C. E. M. (2022). Design and application of an ontology to identify crop areas and improve land use. *Acta Geophysica*, 71(3), 1409–1426. <https://doi.org/10.1007/s11600-022-00808-5>
48. Salomon, J. M., Tulin, A., & Monderondo, J. M. (2014). Indigenous Knowledge, Agricultural Practices and Adaptation in the Marginal Uplands: The Case of Brgy. Linao, Inopacan, Leyte. *Annals of Tropical Research*, 30–47. <https://doi.org/10.32945/atr36s3.2014>
49. Sasmita Pani & Jibitesh Mishra. (2015). Building semantics of E-agriculture in India. *Proceedings, 2015 International Conference on Man and Machine Interfacing (MAMI)*. International Conference on Man and Machine Interfacing (MAMI), Piscataway, NJ.
50. Sinha, P. K., & Dutta, B. (2020). A Systematic Analysis of Flood Ontologies: A Parametric Approach. *KNOWLEDGE ORGANISATION*, 47(2), 138–159. <https://doi.org/10.5771/0943-7444-2020-2-138>
51. Sinha, P. K., Gajbe, S. B., Debnath, S., Sahoo, S., Chakraborty, K., & Mahato, S. S. (2022). A review of data mining ontologies. *Data Technologies and Applications*, 56(2), 172–204. <https://doi.org/10.1108/DTA-04-2021-0106>
52. Sofi Defianti, Ahmad Ashari, & Danang Lelono. (2022). Ontology based knowledge modelling for indonesian rice varieties. *Journal of Theoretical and Applied Information Technology*, 100(23). <https://www.jatit.org/volumes/Vol100No23/3Vol100No23.pdf>
53. Stoyanova-Doycheva, A., Doychev, E., Ivanova, V., Valkanov, V., & Tabakova-Komsalova, V. (2022). Event Ontology about Wheat Cultivation. *IFAC-PapersOnLine*, 55(32), 206–210. <https://doi.org/10.1016/j.ifacol.2022.11.140>
54. Stoyanova-Doycheva, A., Madanska, S., & Bilyanov, S. (2023). Development of an Ontology for Bulgarian Soil Types. *2023 International Conference Automatics and Informatics (ICAI)*, 378–382. <https://doi.org/10.1109/ICAI58806.2023.10339031>
55. Suarez, C., Griol, D., Figueroa, C., Corrales, J. C., & Corrales, D. C. (2022). RustOnt: An Ontology to Explain Weather Favorable Conditions of the Coffee Rust. *Sensors*, 22(24), 9598. <https://doi.org/10.3390/s22249598>
56. Suganya, G., & Porkodi, R. (2018). Ontology Based Information Extraction -A Review. *2018 International Conference on Current Trends towards Converging Technologies (ICCTCT)*, 1–7. <https://doi.org/10.1109/ICCTCT.2018.8551071>
57. Sullivan, K., McLaughlin, J., O'Meara, C., McDonnell, K., & Kehoe, C. (2024). Open-Source Tools and Supports to Advance Data Interoperability in the Agriculture Domain. *2024 IEEE International Conference on Omni-Layer Intelligent Systems (COINS)*, 1–6. <https://doi.org/10.1109/COINS61597.2024.10622152>
58. Taehyung Kim, NamJin Bae, MyeongBae Lee, Changsun Shin, Jangwoo Park, & Yongyun Cho. (2013). A Study of an Agricultural Ontology Model for an Intelligent Service in a Vertical Farm. *International Journal of Smart Home*, 07(04)
59. T.Cynthia & P. Calduwel Newton. (2016). A Semantic Based Answering Technique Using Ontology in MCC. *International Journal in IT and Engineering*, 04(11).
60. Umar, U., Sardjono, T. A., & Kusuma, H. (2024). Smart Ontology-Based System for Recommending Practices in Melon Cultivation. *IEEE Access*, 12, 162204–162216. <https://doi.org/10.1109/ACCESS.2024.3487288>
61. Uschold, M., & Grüninger, M. (1996). Ontologies: Principles, methods and applications. *The Knowledge Engineering Review*, 11(2), 93–136. <https://doi.org/10.1017/S0269888900007797>
62. Usip, P. U., Udo, E. N., Fakiyesi, T. J., & Olowu, T. C. (2025). Ontology and semantic techniques in agricultural crop pest and disease management: A systematic literature review. *World Journal of Applied Science & Technology*, 16(2), 202–207. <https://doi.org/10.4314/wojast.v16i2.202>
63. UN. (2010). World Programme for the Census of Agriculture 2010. In *APPENDIX 3*. https://www.fao.org/fileadmin/templates/ess/documents/world_census_of_agriculture/appendix3_r7.pdf
64. Williams, S. C. (2012). Data Practices in the Crop Sciences: A Review of Selected Faculty Publications. *Journal of Agricultural & Food Information*, 13(4), 308–325. <https://doi.org/10.1080/10496505.2012.717846>
65. Zhong, X., Wang, Y., Wen, X., & Liao, J. (2022). An Ontology-Based Automation System: A Case Study of Citrus Fertilization. *International Journal on Semantic Web and Information Systems*, 18(1), 1–22. <https://doi.org/10.4018/IJSWIS.295946>

What can Computer Vision learn from Ranganathan?

Mayukh Bagchi¹[0000-0002-2946-5018] and Fausto Giunchiglia¹[0000-0002-5903-6150]

¹DISI, University of Trento, Italy.

{mayukh.bagchi, fausto.giunchiglia}@unitn.it

Abstract. The Semantic Gap Problem (SGP) in Computer Vision (CV) arises from the misalignment between visual and lexical semantics leading to flawed CV dataset design and CV benchmarks. This paper proposes that classification principles of S.R. Ranganathan can offer a principled starting point to address SGP and design high-quality CV datasets. We elucidate how these principles, suitably adapted, underpin the *vTelos* CV annotation methodology. The paper also briefly presents experimental evidence showing improvements in CV annotation and accuracy, thereby, validating *vTelos*.

Keywords: Visual Classification, S.R. Ranganathan, Computer Vision, Dataset Design.

1 Introduction

The SGP [1], the unavoidable misalignment between visual information encoded in images and linguistic labels used to describe them, is an established bottleneck in designing CV benchmark datasets. We argue that the root cause of the SGP is an unprincipled annotation process that allows arbitrary subjective classification decisions while designing such benchmarks and results in CV datasets (e.g., ImageNet [2]) with systematic flaws [3]. The arbitrariness stems from an unspecified many-to-many mapping between images and labels [4-6], where an image can justify multiple labels and a label can describe visually diverse images (see, Figure 1, for examples of different categories of images including single-object, multi-object and mislabelled images, reproduced from [7]). These flaws, in turn, are critical as they can undermine the reliability and generalizability of multimodal Artificial Intelligence (AI) systems, e.g., Vision Language Model (VLMs), trained and evaluated on such data.

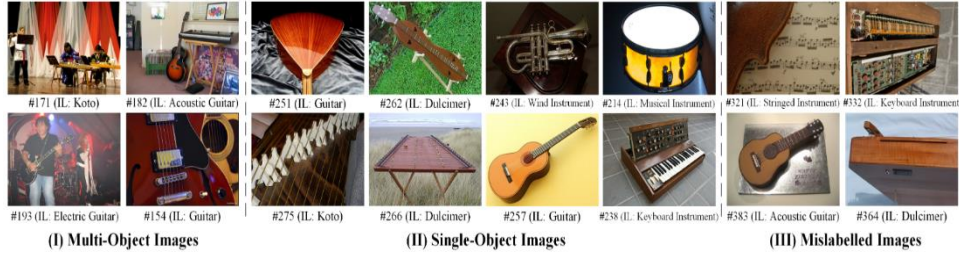


Fig. 1. Samples from ImageNet Musical Instrument Sub-Hierarchy (IL: ImageNet Label)

We posit a shift from ad-hoc labeling-based CV annotation to principled classification-based CV annotation. To that end, this paper advances the proposal that the CV dataset design can find a tried-and-tested starting point in the principles of library classification as advanced by S.R. Ranganathan [8, 9]. The analytico-synthetic classification paradigm, originally developed for organizing knowledge in libraries, provides a stratified process that characteristically decouples perception, conception, lexicalization and notation, while, at the same time, keeps them functionally interlinked [10,11]. This process offers a base for aligning visual semantics (derived from visual properties) with lexical semantics (derived from word meanings) [12, 13], thereby, absolving the SGP.

In the above context, we demonstrate how Ranganathan's principles can be adapted into a principled classification-based CV annotation methodology termed *vTelos* [7]. To that end, *vTelos* enforces a one-to-one alignment between images and labels by grounding the annotation process in visual properties and lexical-semantic definitions. The remainder of the paper is organized as follows. Sections (2) and (3) elucidates this cross-disciplinary adaptation in terms of library classification and visual classification, respectively. Section (4) presents experimental evidence that *vTelos* significantly improves CV annotation and accuracy. Section (5) concludes the paper.

2 Ranganathan's Theory of Classification

The theory of classification advanced by S.R. Ranganathan offers a tried-and-tested framework for systematic knowledge organization [14-16]. To that end, the analytico-synthetic paradigm of classification was developed to manage the complexity of multidimensional (bibliographic) knowledge by isolating the process of classification into distinct yet functionally linked stages. This structured approach, in addition to the body of classificatory guiding principles ('canons'), provided a powerful model to transform bibliographic thought content into well-defined classification categories.

A key characteristic of this paradigm is the stratification of classification work across four planes, namely, the Pre-Idea Stage, the Idea Plane, the Verbal Plane and the Notational Plane. The Pre-Idea stage concerns the initial perception and recognition of entities. It is where pure percepts aggregate into compound percepts to finally form basic concepts. In CV terms, this stage aligns with the extraction of visual properties and the formation of substance concepts from images (or videos). Subsequently, the Idea Plane concerns organizing these recognized concepts into a classification hierarchy based on their (inherent) characteristics. This is achieved via the cumulative application of canons of classification of the Idea Plane, wherein, concepts, at each level, are defined by a general category (genus) and specific distinguishing features (differentia). For CV data, this stage translates to building a visual classification hierarchy using visual properties.

The final two planes of work bridge perception with language and notation. To that end, the Verbal Plane assigns standardized linguistic labels to the nodes of the classification hierarchy (visual hierarchy, in the case of CV), thereby, creating a lexical (subsumption) hierarchy. Finally, the Notational Plane disambiguates these linguistically labeled concepts by attaching unique identifiers (e.g., classification number) resolving issues of synonymy and polysemy. The entire classification process across the planes of work is governed by laws and canons. In particular, the Law of Local Variation ensures that a classification is adapted for a specific context and purpose. This principle is key to CV's need for context-aware and egocentric object recognition benchmarks.

3 Visual Classification

In visual classification the Pre-Idea Stage of Ranganathan corresponds to the Substance Concept Recognition phase. In this phase, a CV system, much like a human observer, encounters visual data from a cumulation of partial views (e.g., of a musical instrument). To that end, it extracts low-level visual properties (pure percepts, as from Ranganathan) which, over multiple encounters, aggregate into compound percepts, thereby forming (stable) substance concepts, e.g., photos of stringed instrument, stringed instrument with six taut strings (see, Figure 1). The process is computationally supported by a cumulative memory, namely, a dynamic knowledge base [17] that incrementally refines concepts with new visual evidence, thereby, providing a mechanism for persistent and evolving CV object identity.

The output of this recognition phase feeds is the input to the next phase of Visual Classification. In this phase, the (isolated) substance concepts are collected and organized into a purpose-specific visual subsumption hierarchy. This is implemented by applying the definition of (visual) genus-differentiae to the identified visual properties. The genus represents a shared visual characteristic that groups visual concepts (e.g., has taut strings), while the differentia provides the specific visual property that distinguishes them within that group (e.g., has six taut strings versus has thirteen taut strings). By iteratively applying this logic, this phase archives a visual classification hierarchy. For example, a root visual concept like musical instrument is classified into stringed instrument and wind instrument based on the genus of sound production, and stringed instrument is further classified into guitar and koto based on the differentia of string counts. This hierarchical construction is based on visual affordances and it ensures the organization is grounded in visual perception rather than linguistic labels.

The transition from a visual hierarchy to a lexical-semantic hierarchy is implemented in the next phase of Linguistic Classification. The nodes of the visual subsumption hierarchy, each representing a substance concept defined by visual properties, is labeled for human and machine reasoning. To that end, it involves mapping each node to a word or synset from a natural or domain language (often, exploiting multilingual lexical-semantic resources [18, 19]). For instance, the substance concept defined by the visual differentia has six strings is assigned the label guitar. This step creates a lexical subsumption hierarchy out of the visual classification hierarchy. Notice that this process must account for: (i) lexical gaps, wherein, a language may lack a direct label for a visual concept, and (ii) egocentric purpose, where the chosen label may reflect a specific user's context. The alignment, therefore, is definitional, i.e., the linguistic label's gloss must linguistically encode the visual genus and differentia that defined the node, ensuring a one-to-one mapping between the visual and lexical concept at that hierarchical level.

The final phase of Conceptual Classification introduces a layer of formal disambiguation essential for computational reasoning in the context of CV-driven applications. This is important as language is inherently ambiguous

due to linguistic phenomena like synonymy and polysemy. To eliminate this ambiguity, each linguistically labeled concept in the lexical hierarchy is assigned a unique supralingual identifier (again, exploiting multilingual lexical-semantic resources). For example, the classification concept labeled guitar might be assigned an identifier like 1278956. This identifier is independent of any particular language, thereby ensuring that the concept remains distinct whether it is referred to as guitar or six-stringed instrument. This phase finalizes the transformation of a perceptual substance concept into a formal, machine-actionable classification concept, ready for use in CV-driven knowledge graphs, databases and reasoning systems.

To operationalize this adapted process as a real-world annotation pipeline, we developed the vTelos methodology and a redefinition of CV annotator roles. vTelos structures the CV annotation task around four key design choices: C1) Object Localization, C2) Visual Classification, C3) Label Generation, and C4) Label Disambiguation. Notice that differently from mainstream CV annotation practice, in vTelos, the linguistic choices (C3 and C4) are made prior to the visual ones (C1 and C2). This means the space of possible labels, each with a clearly defined meaning from a multilingual lexical-semantic resource, is established which creates a controlled vocabulary. The annotator then selects labels based on how well definitions match the visual properties of the localized object.

This workflow necessitates splitting the traditional annotator role into two specialized functions borrowed from knowledge organization: the Classificationist and the Classifier. The classificationist is the domain expert who performs a priori work of linguistic choices C3 and C4. They define the lexical-semantic hierarchy and ensure each label's gloss provides clear linguistic genus-differentia. The classifier, who could be a human annotator or a (semi-)automated system [20], then performs the perceptual work of visual choices C1 and C2. To that end, they localize objects in images and, by matching visual properties to the pre-defined linguistic differentia, assign the appropriate label from the classificationist's schema. This separation of linguistic and visual choices ensure that subjective interpretation is minimized during annotation. The classifier does not choose a label based on intuition, but executes a match between visual evidence and a pre-established semantic definition. The vTelos methodology, therefore, enforces a principled, one-to-one mapping between visual and lexical semantics, directly tackling the many-to-many mappings at the heart of the SGP.

4 Experimental Evidence

We applied the vTelos methodology to a subset - the musical instrument hierarchy - of ImageNet. The experiment [7] was designed to compare mainstream ad-hoc label-based annotation against the principled classification-based annotation. To that end, the subset was used to design a ground truth dataset and annotators were divided into groups: (i) one group annotated images via ad-hoc selection from ImageNet category labels, while (ii) the second group performed annotation using the vTelos methodology before assigning a final label. The results demonstrated an 18% improvement in inter-annotator agreement for (ii) over (i) [7]. This indicates a reduction in subjective judgment as annotators converge more reliably with respect to visual properties than labels. Further, when the full vTelos process including object localization was implemented, agreement amongst expert annotators reached near-perfect levels. This confirms that structuring the CV annotation task with principled classification choices absolves SGP ambiguity.

The impact on CV accuracy was equally encouraging. Several state-of-the-art convolutional neural networks, including ResNet and VGG, were trained on the dataset re-annotated following the vTelos methodology. In comparison to models trained on the original ImageNet labels, these models showed improvement in CV classification accuracy, with gains of up to 23% for some architectures [7]. Additionally, an ablation study confirmed that the key driver of this improvement was the Visual Classification phase itself, rather than merely the isolation of single objects. The models effectively learned to associate the linguistically grounded differentia with the corresponding visual features, as evidenced by attention heatmaps that highlighted relevant image regions [7]. This case study provided concrete evidence that adopting principles from Ranganathan's library science directly addresses the SGP and produces high-quality CV datasets.

5 Conclusion

In summary, the CV community can learn a key lesson from Ranganathan, namely, that principled classification is foundational for high-quality CV dataset design.

Declaration of AI and AI-Assisted Technologies used in Manuscript Writing Process: Not Applicable

References

1. Smeulders, A. W., Worring, M., Santini, S., Gupta, A., & Jain, R. (2002). Content-based image retrieval at the end of the early years. *IEEE Transactions on pattern analysis and machine intelligence*, 22(12), 1349-1380.
2. Deng, J., Dong, W., Socher, R., Li, L. J., Li, K., & Fei-Fei, L. (2009). Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*. IEEE.
3. Tsipras, D., Santurkar, S., Engstrom, L., Ilyas, A., & Madry, A. (2020). From imagenet to image classification: Contextualizing progress on benchmarks. In *the International Conference on Machine Learning* (pp. 9625-9635). PMLR.
4. Giunchiglia, F., Bagchi, M. (2021). Millikan + Ranganathan - from perception to classification. In: *5th Cognition and Ontologies (CAOS), 12th International Conference on Formal Ontology in Information Systems (FOIS)*.
5. Giunchiglia, F., & Bagchi, M. (2022a). Object Recognition as Classification via Visual Properties. In *Knowledge Organization across Disciplines, Domains, Services and Technologies* (pp. 87-100). Ergon-Verlag.
6. Giunchiglia, F., & Bagchi, M. (2022b). Representation Heterogeneity. In: *Formal Models of Knowledge Diversity (FMKD), Joint Ontology Workshops (JOWO), Jönköping, Sweden* (Vol. 1613, p. 0073).
7. Giunchiglia, F., Bagchi, M., & Diao, X. (2023b). A semantics-driven methodology for high-quality image annotation. In: *26th European Conference on Artificial Intelligence (ECAI)*, 372, 851-858.
8. Ranganathan, S. R. (1967). *Prolegomena to library classification*. Asia Publishing House (New York).
9. Ranganathan, S. R. (1989). *Philosophy of library classification*. Sarada Ranganathan Endowment for Library Science (Bangalore, India).
10. Bagchi, M. (2022). A Diversity-Aware Domain Development Methodology. In *PhD Symposium, 41st International Conference on Conceptual Modeling* (Vol. 3211, pp. 1-11). CEUR-WS.
11. Bagchi, M. (2025). *Language and Knowledge Representation A Stratified Approach* (Doctoral dissertation, University of Trento, Italy).
12. Giunchiglia, F., Bagchi, M., & Diao, X. (2023a). Aligning visual and lexical semantics. In: *International Conference on Information - iConference -* (pp. 294-302). Cham: Springer Nature Switzerland.
13. Giunchiglia, F., Diao, X., & Bagchi, M. (2023c). Incremental image labeling via iterative refinement. In *2023 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)* (pp. 1-5). IEEE.
14. Giunchiglia, F., Dutta, B., & Maltese, V. (2014). From knowledge organization to knowledge representation. *Knowledge Organization*, 41(1), 44-56.
15. Giunchiglia, F., Bagchi, M., & Das, S. (2024). From Knowledge Organization to Knowledge Representation and Back. *Annals of Library and Information Studies*, 71(1), 137-149.
16. Giunchiglia, F., & Bagchi, M. (2024). From knowledge representation to knowledge organization and back. In: *International Conference on Information - iConference -* (pp. 270-287). Cham: Springer Nature Switzerland.
17. Erculiani, L., Bontempelli, A., Passerini, A., & Giunchiglia, F. (2023). Egocentric hierarchical visual semantics. In *HHAI 2023: Augmenting Human Intellect* (pp. 320-329). IOS Press.
18. Giunchiglia, F., Batsuren, K., & Bella, G. (2017). Understanding and Exploiting Language Diversity. In *IJCAI* (pp. 4009-4017).
19. Giunchiglia, F., Batsuren, K., & Alhakim Freihat, A. (2018). One world-seven thousand languages. In: *International Conference on Computational Linguistics and Intelligent Text Processing* (pp. 220-235). Cham: Springer Nature Switzerland.
20. Erculiani, L., Giunchiglia, F., & Passerini, A. (2020). Continual egocentric object recognition. In *24th European Conference on Artificial Intelligence (ECAI)*. IOS Press BV.

An Analytical Study of Web Dewey 2.0 and UDC Online

Partha Chattopadhyay¹ [0000-0001-5790-2767] and Arun Kumar Chakraborty² [0000-0002-4630-9260]

¹Hiralal Bhakat College, Nalhati, Birbhum, WB-731220

²Dept. of Library & Information Science, Netaji Subhas Open University (NSOU), Kolkata 700064

¹chattopadhyaypartha224@gmail.com

²akchakraborty@wbnsou.ac.in, akchakraborty2009@gmail.com

Abstract. The present study examines web-based library classification schemes available on digital platforms. With the rapid advancement of digital technologies, classification in web-based systems has become as significant as its printed counterparts. This paper focuses on an analytical study of Web Dewey 2.0, developed from the 23rd edition of the Dewey Decimal Classification (DDC), and UDC Online, based on the UDC IME edition. The databases of both classification schemes were accessed and analyzed to understand their features, functionality, and currency in a web environment. The study reveals that Web Dewey 2.0 is more frequently updated and provides a more dynamic and user-oriented interface for subject organization and retrieval compared to UDC Online. The findings highlight the growing importance of web-based classification schemes in enhancing subject access and information retrieval in modern library systems.

Keywords: Web-Based Library Classification Scheme; Web Dewey 2.0; UDC Online; Digital Library Classification; Subject Retrieval

1 Introduction

The library classification is an art for systematic arrangement of various subjects of universe of knowledge logically so that it can locate any particular item quickly and easily in a library. There are some popularly used library classification schemes. These are DDC (Dewey Decimal Classification), UDC (Universal Decimal Classification), Colon Classification (CC) etc. Web-based library classification schemes play a pivotal role for multiple searching by terms and numbers and have user-friendly browsing option. These are Web- Dewey 2.0 (DDC 23rd Edition) and UDC online. These make it possible to understand the potentiality of Library Classification to improve subject retrieval. It promotes interest in classification as an organizing and retrieval device for information sources by the growth in usage of the internet and World Wide Web (www).

1.1 Advantages of Web-based classification schemes

- Easy access of schedules and tables
- Access of related terms by using different keywords
- Online retrieval of class numbers
- Provides work area for building numbers
- Uses of online retrieval of subject heading list, dictionaries and glossaries
- Access through hyperlinks
- Multiple keywords searching & browsing
- Facilitates OPAC
- Boolean searching

1.2 Disadvantages of Web-Based Classification Schemes

Despite their numerous advantages, web-based classification schemes also face certain limitations:

- Dependence on technological infrastructure such as computers, internet connectivity, and compatible software
- Technological barriers faced by users with limited digital literacy
- Language barriers, as most web-based classification schemes are predominantly available in English
- Restricted access due to subscription-based or licensed platforms

- Requirement of continuous technical support and system maintenance

2 Objectives of the Study

The present study has the following objectives:

- To identify the major web-based library classification schemes
- To collect and examine information related to web-based library classification systems
- To access and analyze the databases of selected web-based classification schemes, namely Web Dewey 2.0 and UDC Online

3 Methodology

The study is based on an analytical approach using secondary data sources. Data were collected from the Internet by searching relevant literature and resources using the term “web-based library classification schemes.” For the purpose of this study, Dewey Decimal Classification (DDC) and Universal Decimal Classification (UDC) were selected. The web-based databases of Web Dewey 2.0 and UDC Online were accessed and examined to analyze their features, functionality, and scope in a digital environment.

4 Literature Review

The transition from print-based library classification schemes to web-based and digital environments has been widely discussed in Library and Information Science literature. Scholars have emphasized that traditional classification systems, while conceptually robust, must evolve to remain effective in an increasingly digital and networked information landscape [5]. The emergence of web-based classification tools such as Web Dewey and UDC Online represents a significant shift in how bibliographic control and subject access are managed in modern libraries.

Several studies have examined the development and evolution of the Dewey Decimal Classification (DDC) in the digital era. Chan and Zeng (2006) highlighted that the online versions of DDC provide enhanced flexibility, faster updates, and improved linking capabilities compared to print editions. Web Dewey, derived from the 23rd edition of DDC, has been recognized as a dynamic tool that supports continuous revision, relative indexing, and interoperability with metadata standards such as MARC and Dublin Core [10]. Researchers have also noted that Web Dewey facilitates improved subject analysis and browsing through hierarchical displays and built-in notes, making it more user-friendly for cataloguers and information professionals [17].

In contrast, the Universal Decimal Classification (UDC) has been widely studied for its analytico-synthetic structure and suitability for complex subject representation. Studies by Slavic (2008) and Broughton (2015) emphasized that UDC’s faceted approach allows greater precision in subject representation, particularly in multilingual and interdisciplinary contexts. With the introduction of UDC Online, efforts have been made to provide web-based access to the UDC Master Reference File (MRF), enabling global usage and regular updates. However, researchers have pointed out that while UDC Online ensures authoritative control and consistency, its interface and navigational features are relatively less intuitive compared to Web Dewey [7].

Comparative studies between DDC and UDC in digital environments remain limited, though existing literature suggests notable differences in update frequency, usability, and functional design. Satija and Comaromi (2012) observed that DDC’s editorial infrastructure allows for quicker responsiveness to emerging subjects, especially in science and technology. On the other hand, UDC revisions often undergo more centralized control, which may affect the speed of updates in the online environment [21]. This difference has implications for subject currency and retrieval effectiveness in web-based library systems.

Recent research has also focused on the role of web-based classification schemes in enhancing information retrieval and subject access. Studies indicate that online classification tools improve consistency in cataloguing, support linked data initiatives, and enable better integration with discovery systems [28]. Web-based interfaces allow classifiers to navigate schedules, notes, and relationships more efficiently, thereby reducing classification errors and improving retrieval outcomes.

Overall, the literature underscores the growing importance of web-based classification schemes in contemporary librarianship. While Web Dewey 2.0 is frequently praised for its dynamic updates and user-oriented design, UDC Online is valued for its structural depth and analytical strength. The existing body of research provides a

foundation for the present study, which seeks to analytically examine and compare Web Dewey 2.0 and UDC Online in terms of features, functionality, and effectiveness in a web-based environment.

5 Feature-wise Analysis of Web Dewey 2.0 and UDC Online

An analysis of selected database features of Web Dewey 2.0 and UDC Online was carried out to examine their search capabilities, functional tools, and user-support mechanisms in a web-based environment. The assessment is limited to observable features available within the respective online databases.

Table 1. Database Features of Web Dewey 2.0 and UDC Online

Features	Web Dewey 2.0	UDC Online
Flexible Search	Yes	Yes
Truncation Search	Yes	Yes
Boolean Search	Yes	Yes
Character Masking / Wildcards	Yes	No
Mathematical Operators	No	Yes
Links to LCSH	Yes	No
Links to OPAC	Yes	No
User Notes	Yes	No
View Records	Yes	Yes
Outline Display	Yes	Yes
Moving Arrows	No	Yes
Stop Lists	Yes	No
Clear Search Option	Yes	Yes
Delete Option	No	Yes
Number of Available Parameters	11	09

6 Findings

The analysis indicates that both Web Dewey 2.0 and UDC Online provide core search functionalities such as flexible searching, truncation, Boolean operations, record viewing, and outline display. These features support efficient navigation and retrieval within web-based classification databases.

Web Dewey 2.0 demonstrates stronger integration with library tools through the availability of links to Library of Congress Subject Headings (LCSH), OPAC connectivity, user notes, wildcard searching, and stop lists, which facilitate enhanced subject access and classification control.

UDC Online, however, incorporates mathematical operators, moving arrows, and delete functions, reflecting its analytico-synthetic structure and emphasis on notational manipulation. While both systems support essential classification tasks, Web Dewey 2.0 offers a higher number of functional parameters, suggesting broader support for interactive classification and subject retrieval in a web environment.

7 Analysis Based on Essential Qualities

An analysis was conducted to examine selected essential qualities of Web Dewey 2.0 and UDC Online with reference to indexing structure, number-building facilities, and availability of supporting tools such as glossaries and dictionaries.

Table 2. Essential Qualities of Web Dewey 2.0 and UDC Online

Features	Web Dewey 2.0	UDC Online
Index	Print: Relative Index (Alphabetical)	Print: Alphabetical Index

Features	Web Dewey 2.0	UDC Online
	Web: 09 searchable and 05 browsable indexes	Web: 02 indexes searchable for terms and UDC class numbers
Built Numbers	Availability of approximately 10,000 built numbers	Facility provided for number building
Glossary / Dictionary	Glossary	UDC Dictionary

8 Analysis from the Perspective of Versions

The availability of different versions of classification schemes determines their usability across varied library environments. The following table presents the formats in which Web Dewey 2.0 and UDC Online are available.

Table 3. Availability of Different Versions

Features	Web Dewey 2.0	UDC Online
Print Version	Yes	Yes
LAN-Based Version	No	Yes
CD-ROM Version	Yes	Yes

9 Conclusion

The study reveals that both Web Dewey 2.0 and UDC Online exhibit several similarities as web-based library classification schemes, particularly in terms of multiple search strategies, number-building facilities, and support for searching and browsing of terms and class numbers. These features contribute to making classification activities more user-friendly in digital environments.

However, Web Dewey 2.0 demonstrates certain advanced features, such as extensive indexing options, a large number of pre-built class numbers, and stronger integration with subject access tools, which enhance its functionality in comparison to UDC Online. Despite these differences, both schemes play a significant role in improving subject organization and retrieval in web-enabled library systems.

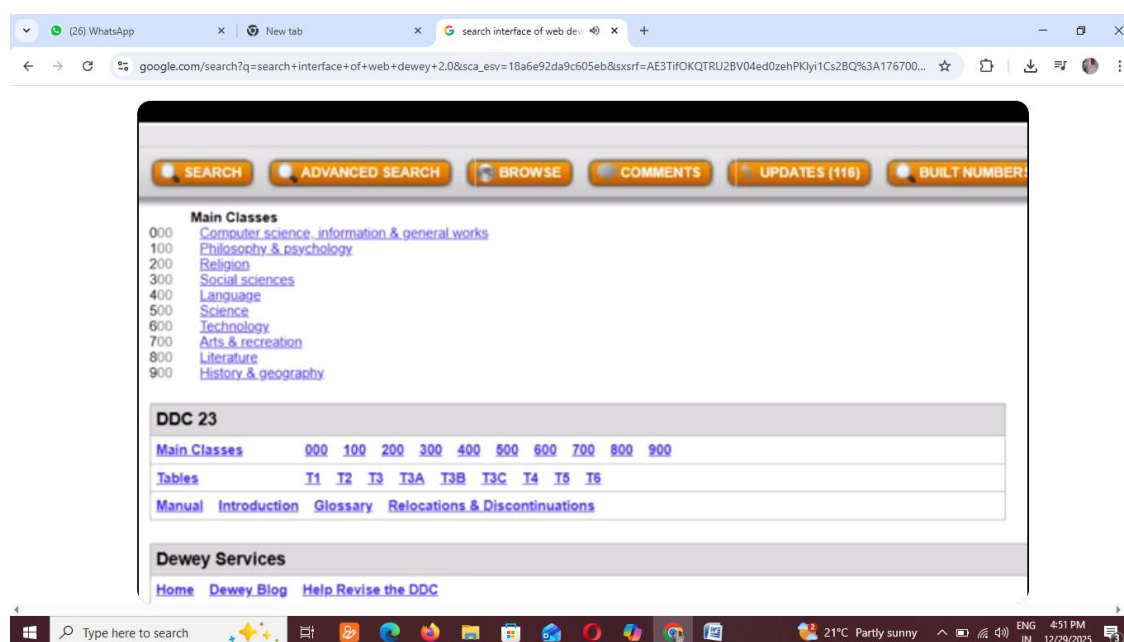
Declaration of AI and AI-Assisted Technologies used in Manuscript Writing Process: Not Applicable

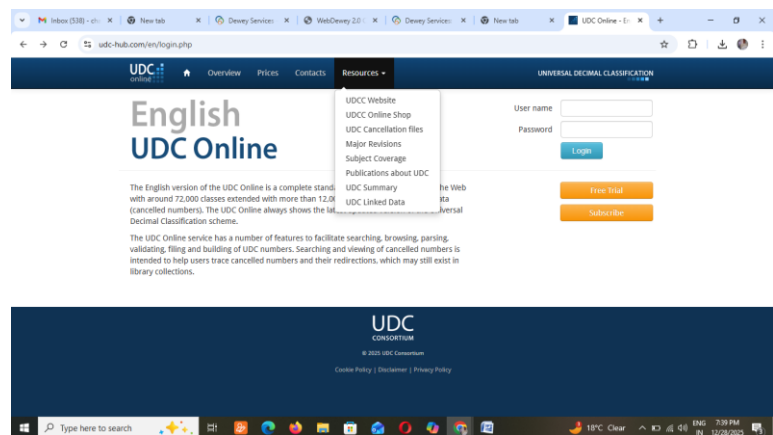
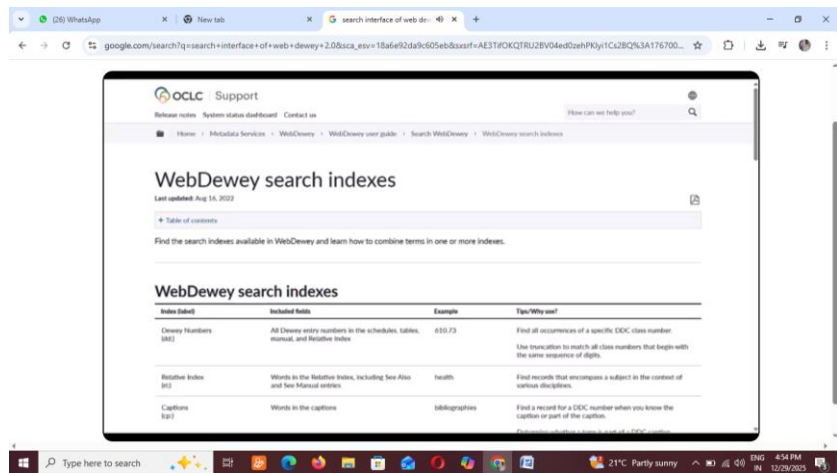
References

1. Aagaard, H., Conradi, E., & Mengel, T. (2016). Leveraging the Dewey Decimal Classification for online subject access: Three use cases of WebDewey Search. IFLA World Library and Information Congress. Retrieved from <https://repository.ifla.org/items/b56f28a7-f5e9-43f4-8f2c-1e71b0db5f56>
2. Broughton, V. (2015). Essential classification (2nd ed.). Facet Publishing.
3. Chan, L. M., & Zeng, M. L. (2006). Metadata interoperability and standardization—A study of methodology. D-Lib Magazine, 12(6). <https://doi.org/10.1045/june2006-chan>
4. Chudamani, K. S. (2025). Classification model for libraries in the digital environment. Retrieved from <https://ir.inflibnet.ac.in/bitstreams/772e31ad-dbe4-41e7-b364-8f92b0827622/download>
5. Hjørland, B. (2013). User-based and cognitive approaches to knowledge organization: A theoretical analysis of the research literature. Knowledge Organization, 40(1), 11–27.
6. Khan, M. T. M. (2005). Dewey Decimal Classification. Shree Publishers.
7. Kumar, A., & Sharma, J. (2019). Web-based classification systems and their usability in digital libraries. International Journal of Library and Information Studies, 9(2), 45–52.
8. Kumar, P. S. G. (2003). Knowledge organization, information processing, and retrieval: Theory. B. R. Publishing.
9. LIS Academy. (2024). The evolution of library classification: From print to digital systems. Retrieved from <https://lis.academy/organising-and-managing-information/evolution-library-classification-print-digital/>
10. Mitchell, J. S., & Beall, J. (2014). Dewey Decimal Classification: Principles and applications (4th ed.). OCLC.
11. Mitchell, J. S. (2003). Dewey Decimal Classification and Relative Index (22nd ed.). OCLC.

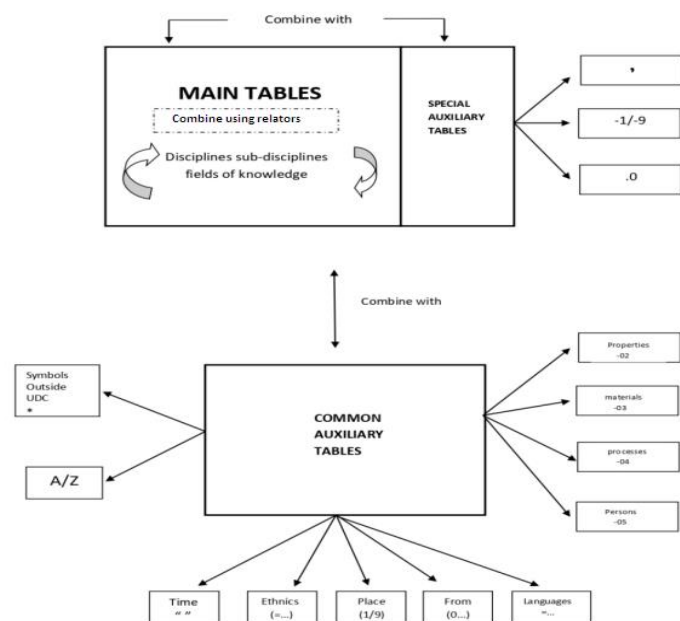
12. OCLC. (n.d.-a). Dewey documentation. Retrieved December 19, 2025, from <http://www.oclc.org/support/documentation/dewey/>
13. OCLC. (n.d.-b). Dewey glossary. Retrieved December 19, 2025, from <http://www.oclc.org/support/documentation/glossary/dewey/>
14. OCLC. (n.d.-c). UDC-DDC conference presentation. Retrieved December 18, 2025, from <http://www.oclc.org/dewey/news/conferences/udcddc.ppt>
15. OCLC. (n.d.-d). Dewey updates and enhancements. Retrieved December 19, 2025, from <http://www.oclc.org/dewey/updates/enhancements/>
16. OCLC. (n.d.-e). WebDewey: The online Dewey Decimal Classification tool. Retrieved from <https://www.oclc.org/en/dewey/webdewey.html>
17. Satija, M. P. (2017). The theory and practice of the Dewey Decimal Classification system (2nd ed.). Chandos Publishing.
18. Satija, M. P., & Comaromi, J. P. (2012). Introduction to the Dewey Decimal Classification. Forest Press.
19. Satyanarayana, V. V. V. (1998). Universal Decimal Classification: A practical primer. Ess Ess Publications.
20. Slavic, A. (2008). Faceted classification: Management and use. *Axiomathes*, 18(3), 257–271.
21. Slavic, A., & Cordeiro, M. I. (2014). UDC in a multilingual environment. *Knowledge Organization*, 41(1), 30–38.
22. UKOLN. (n.d.). The role of classification schemes in Internet resource description and discovery: UDC perspective. Retrieved from https://www.ukoln.ac.uk/metadata/desire/classification/class_3.htm
23. UDC Consortium. (n.d.-a). Seminar presentation. Retrieved December 21, 2025, from <http://www.udcc.org/seminar07/presentations/bsi.pdf>
24. UDC Consortium. (n.d.-b). UDC bibliography. Retrieved December 20, 2025, from <http://www.udcc.org/bibliography.htm>
25. UDC Online. (n.d.). UDC Online. Retrieved December 21, 2025, from <http://www.udconline.net/>
26. Vizine-Goetz, D. (n.d.). Using library classification schemes for Internet resources. OCLC Office of Research and Special Projects. Retrieved from <https://webdoc.sub.gwdg.de/ebook/aw/oclc/man/colloq/v-g.htm>
27. Zeng, M. L., & Qin, J. (2016). Metadata (2nd ed.). Neal-Schuman.

Appendices





Structure of UDC



Reimagining Metadata for a Semantic Web

A Rule-Governed Agentic AI Architecture for Intelligent Discovery and Metadata Management for Library and Information Science

Aravind R Nair ¹[0009-0003-6951-6742], Vandhana S P ²[0009-0006-5599-4307] and Chandhana S P ³[0009-0008-1502-5119]

¹ Department of Library, Sai University, Chennai

² Library, Chanakya University, Bangalore

³ Library, ICAR-Central Institute of Fisheries Technology, Kochi

{aravindrna05, vandhanasp1997@, chandhana024}@gmail.com

Abstract. Recent developments in artificial intelligence, particularly large language models, have made AI accessible beyond specialized technical communities and closer to everyday users. Libraries and librarians are increasingly adopting AI based approaches in service delivery, following similar transformations observed across other industries. Continuous progress in AI has resulted in the emergence of intelligent agents, defined as autonomous systems capable of planning actions and executing tasks with minimal human intervention. This paper focuses on the design and development of an Agentic AI system intended specifically for intelligent information retrieval and discovery services within library environments. The study further proposes a localized approach to metadata classification that can be independently designed, customized, and maintained by individual libraries according to their institutional needs and collection characteristics. By integrating retrieval intelligence with flexible metadata practices, the proposed framework aims to improve discovery effectiveness and user support. The paper concludes by outlining future scope for extending Agentic AI through the integration of additional functional modules and intelligent robotic applications to support digital library operations environments globally.

Keywords: Agentic AI, Artificial Intelligence, Canonical Classification of Metadata, Discovery Service, Information Retrieval, Large Language Model.

1 Introduction

Artificial intelligence (AI) is a broad area of computer science that tends to create intelligent devices to deal with the kind of work that is usually done by humans. The effect of such technology on Library and Information Science (LIS) is colossal. The ability to automate at least the simpler tasks like sorting out books and returning the information one seeks at the minimal time possible can make libraries smarter with the help of AI. It relies on applying such tools as natural language processing (NLP) to learn more about the questions and provide more accurate answers. Libraries rely on AI to enhance the services they provide, including book recommendations or assistance even after normal working hours. Nevertheless, numerous libraries do not have enough money, lack technical expertise, or find it challenging to make the AI systems effective. Simple chatbots were the early types of AI used in libraries.

Nowadays, there are more sophisticated systems, such as Agentic AI, which can work on complex tasks by themselves. The agentic AI is an intelligent assistant, which coordinates its actions in its own plan of actions. Most importantly, it can rectify its own errors and strive to reach the objectives without the critical involvement of humans. This system is not merely a question answering service, but it will deal with entire tasks e.g. research project planning or collection inventory management. This technology is already very popular not only within LIS. As an illustration, the Agentic AI can keep track of the vital signs of a patient in the healthcare sector, while it can identify fraud in real-time transactions in the financial sector. This has been an effective system in LIS that can enhance searches, anticipate what users will require and even develop correct summaries of articles. This assurance of efficiency and improved personalization to users is high. Nevertheless, there is also one significant problem that should be addressed regardless of all these advantages: occasionally, AI generates incorrect data also known as hallucinations. Thus, the paper examines the creation of this Agentic AI system to refine the experience of users seeking knowledge within academic libraries.

2 Background of Study

Artificial intelligence (AI) is an ambiguous and broad research field in computer science that deals with the development of smart systems that replicate the functions of humans and contain important implications on the

profession of librarianship. AI is considered one of the key forces of the development of present-day society, and it is a central part of the digitalization of the university sphere and literally promotes the creation of intelligent library spaces [1]. The adoption of AI methods can increase the workings of a core library by accepting robotization of indexing and classifying, utilizing natural language processing (NLP) in finding databases more efficiently, and applying smart data analysis in managing collections and circulation. Through such changing technologies, libraries are trying to increase the general service standards to their patrons, increase access to content and position themselves to be able to take advantage of the mind-based computing applications. Although the adoption of AI is inevitable as it is acknowledged, many academic libraries face substantial obstacles to its extensive use, which are mostly associated with a lack of financial capacity, the lack of technological or human resources, and challenges associated with the implementation and maturity of the systems [2]. First implementations were custom chatbots such as ChatGPT API powered Aisha of Zayed University Library, successfully providing quick and effective reference assistance during non-operational hours. Even though these systems offer potential providing cost effective and dedicated support, they often face some issues like “hallucination” (generating non-existent information) [3].

The analysis conducted by Lo [4] revealed the consequences of the report on AI published by the U.S. Department of Education which suggests that AI-based chatbots and tools should be included in the scope of the library services and operations, such as the existing online catalogue and databases, to improve and enhance them. Kaushal and Yadav [5] explored how AI-based chatbots play the experience component in academic libraries and found that they can offer various services, act as a triage tool in solving simple and repetitive requests, simplify search results, and generate personalized recommendations, etc. Gupta [5] identified certain examples of implemented Generative AI applications, which are the result of experimental usage in the library field and include personalization of student reading lists, assistance with metadata tags, research paper summarization to promote access to information and its discovery. Panda and Kaur [7] described the large-scale usage of the SheetGPT tool, both in managing budgets or monitoring circulation rates, but also giving user services, such as translation and reference services to users. As an example of a more basic application of AI to LIS research itself, [8] used machine learning to categorize and study the frequency and type of reported research methods in scholarly articles over 31 years. Although the study [9] did not avoid ethical matters, the speaker admitted that AI systems typically applied in automated processes in library and information industry include recommendations, search results and decision making. These AI solutions speed up research processes in many areas of LIS with the help of AI based literature review solutions or even offer round-the-clock information services to customers. These applications reflect that AI is assisting in advanced administrative automation as well as user-friendly information services in the changing LIS realm.

3 What is an Agentic AI

Agentic Artificial Intelligence (Agentic AI) refers to autonomous, goal-driven systems that can operate independently for long periods with little human guidance. This represents a major step forward from traditional AI, generative AI, and passive tools by focusing on independence, flexibility, and goal-focused thinking. Unlike standard AI or typical large language models (LLMs), which mainly respond to specific prompts, Agentic AI systems perform complex, multi-step tasks. They break down broad goals into smaller ones, create actionable plans, and actively pursue them. These systems use external tools to complete tasks in order, track progress, link dependent steps, recover from mistakes, and automate processes from start to finish. For example, they can autonomously handle activities like travel planning, booking flights, and optimizing itineraries. Designed for flexibility, Agentic AI adjusts its strategies in real time, learns continuously from experiences, and thinks adaptively in changing and unpredictable situations. Additionally, these systems combine foundation models with planning and reasoning modules, along with memory and reflection mechanisms. This setup allows them to keep context, evaluate themselves repeatedly, and improve their plans and actions, enabling consistent independent performance across complex tasks [10].

3.1 How is Agentic AI useful?

Agentic AI is highly useful because it represents a qualitative leap in artificial intelligence, enabling systems to pursue complex goals autonomously with minimal human intervention. By demonstrating adaptability and advanced decision-making, Agentic AI enhances human action, expands the application area of AI, and increases productivity across multiple sectors [11].

3.2 Agentic AI is implemented in the following areas:

Healthcare Monitoring and Diagnostics: The system independently identifies patient deterioration through steady tracking of vital signs and autonomously notifying providers. Financial Market Analysis and Algorithmic Trading: The system dynamically adjusts trading strategies based on real-time market data to optimize outcomes during high volatility with minimum human interaction. Autonomous Customer Support in E-Commerce: The e-agent adapts to past interactions and user inputs to provide individualized, context-responsive support based on individual user needs and preferences. Smart Manufacturing and Predictive Maintenance: The system computes the predicted time to failure of machines and optimizes maintenance activities proactively to maximize operational availability. Disaster Management: The system autonomously analyzes real-time environmental data (e.g., during a flood) and dynamically adjusts strategies, reallocating resources like rescue teams and medical supplies [11].

4 Current examples in Agentic AI

Academic libraries have historically adopted new technologies, firstly implementing Artificial Intelligence (AI) in the form of simple role-based chatbots to answer frequent, common inquiries like library hours. These early systems were limited and struggled with any complicated query that didn't fit into their predefined scripts. The development of generative AI and large language models (LLMs) has greatly increased system capabilities, enabling more sophisticated conversational agents that understand context and generate natural, human-like responses. The integration of Agentic AI or autonomous AI which goes beyond just answering questions. These agents are designed to perform complex objectives on their own without any human assistance and execute a multi-step sequence of actions until the goal is accomplished. This enables them to offer proactive, personal assistance in libraries while in other sectors, this advanced automation is used for critical functions such as real time fraud detection in finance, supply chain logistics rerouting and so on [12].

Currently these agents are being used in several crucial sectors, which are as follows.

- a) **Finance:** Agentic AI is used for risk prevention by analysing large amounts of transaction data in real time to detect unusual patterns which might signal fraud. Moreover, these systems assist investment management, trade execution, and market data processing at the best times with minimal human supervision [13].
- b) **Healthcare:** Agents help to coordinate complex administrative workflows, appointment scheduling, insurance eligibility checks. They also improve clinical effectiveness by using NLP to analyse unstructured clinical note extraction key medical insights to enhance diagnostic accuracy [14].
- c) **Supply Chain Optimisation:** Agents continuously monitor the real time data such as inventory level and logistics and autonomously reroute the shipping when unexpected disruptions are detected [14].
- d) **Human Resource (HR):** Agentic systems manage every aspect of the employee lifecycle, from analyzing past hiring patterns to coordinating the entire onboarding process that includes parsing resumes, interview scheduling and account provision [14].
- e) **Customer Experience and IT Support:** Agents provide sophisticated support chatbots which streamline issue resolution for customers. In IT, these agents can perform diagnostic, reset passwords on their own and prioritize support tickets in IT, escalating only the most complex issues to human support [14].
- f) **Education:** Agentic AI is used to create more engaging and personalized user experience by analyzing user preferences and behaviors. These systems can offer book recommendations, customized learning materials, research path suggestions. This transforms library systems into more personalized service hubs. Moreover, automate grading of assignments and exams help the educators by saving time and ensuring consistency in evaluation [13].
- g) **Construction:** Agents are used for optimizing project scheduling and ensuring site monitoring and safety [13].
- h) **Retail:** Agents are used for personalized shopping experience and autonomous inventory management to minimize stock outs and excess shock [13].
- i) **Manufacturing:** These systems are used for predictive maintenance, monitoring health equipment to anticipate failures and schedule upkeep, thereby reducing downtime [13].

5 Adaptation of Agentic AI in LIS

Artificial Intelligence (AI) has already transformed Library and Information Science (LIS) through direct applications like chatbots for reference services, cataloguing automation, and personalized recommendations which

are seen in Aisha system at Zayed University Library which seamlessly offers reference services and personalized recommendations [3]. Moreover, AI enhances library operations directly or indirectly by predictive analytics for collection development, usage pattern analysis, and accessibility features like text-to-speech. Machine learning and natural language processing simplify inventory control and customer service, and such problems as hallucinations continue to occur during early adoption. Academic libraries use AI in technical services, in circulation through RFID, and 24/7 query service. In general, these AI tools increase productivity and user engagement while complementing human librarians rather than replacing their place.

Agentic AI refers to autonomous systems which perform multi tasks by using reasoning and tools. It's already implemented in healthcare for patient monitoring and diagnostic, in finance for fraud detection, in retail for personalized shopping and autonomous inventory management, in manufacturing for predictive maintenance, in education for personalized learning, and in supply chain management for logistic optimization. There is no existing literature which covers agentic AI in LIS, which presents a new opportunity.

Libraries can also implement agentic AI by integrating existing AI use cases by combining chatbots and retrieval augmented generation (RAG) to perform advanced research, predictive analysis to weed collections, and NLP to support multilingual users. Agents could perform automate scheduling of events, generate annotated bibliographies from databases, and identify gaps anticipating user research needs. This single-task AI is scaled to proactive and multi-tool workflow based on library data.

Through this study, we have designed agentic AI prototype specially designed for information retrieval and discovery services in LIS, enabling autonomous query understanding, adaptive retrieval strategies, and context-aware result synthesis.

This agentic framework is the first autonomous innovation in LIS, which serves a new pathway to scalable autonomous library service that enhances discovery, efficiency and user empowerment within academic ecosystems.

6 System Architecture

6.1 User Interaction Layer

This topmost layer serves as the stakeholder for external interaction between the system and its end users. Libraries can implement their own user interfaces including websites, mobile applications, chatbots, OPAC Plugins, Learning management System extensions etc. The system communicates via REST APIs, endpoints, web sockets, Command Line Interfaces to third party applications.

The responsibility of ULI includes:

- Accept user queries.
- Support multiple interaction modes: natural language, structured queries
- Communicates securely with the API Gateway layer
- Display personalized recommendations, alerts.

6.2 API Gateway and Access Control Layer

This layer acts as the “front door” or the primary access point to the complete system entry and exit of data flow and its control. It is responsible for Token-based authentication, API key management, enforcing role-based access control for admin, librarian and user, rate limiting and logging request metadata. The API gateway ensures that only authorized users and applications can interact with the agent and data pipelines.

REST APIs used by with CURD Feature:

- End User Input/Output API: This API handles user queries, returns search results, recommendations, and explanations, and serves as the primary interface for all user-facing interactions.
- User Profile configuration API: This API lets users update their preferences, saved topics, alert settings, and personalization data that guide the agent's behavior.
- User API: This exposes operations like authentication, saved searches, alerts, and personal settings so users can manage their own data.
- Admin Panel API: This API features administrative controls for managing ingestion, LLM configuration, embeddings, vocabulary, and system settings.
- Log API: This API provides secure access to system logs such as errors, agent reasoning steps, and performance data for monitoring and debugging.

6.3 Cataloguing Metadata Model Engine (CMM Engine)

This engine ensures metadata consistency for the agent. It is the most important component for the agent as well as librarians to configure for managing the ingestion of metadata. This layer performs as per the rule-based Canonical Classification of Metadata to store the data on the database.

Canonical Classification of Metadata (CCM). This is a rule-based engine which helps the system to catalogue and assign data to its database fields. The canonical classification refers to classifying and assigning the data based on nature (eg: contributor/create/author or title/heading/caption or etc.). As mentioned, each publisher or data provider has a different format of metadata (MARC 21, Dublin Core, MODS, BIBFRAME etc.) and the system cannot identify each API response field to map with the field of the database. That means the system cannot diagnose the metadata standard followed by each API. So, the CMM engine will give the institution or the admin the freedom to design their own metadata standard and map it to the API response. This ensures reduction of conflicts in metadata standards followed by publisher or data provider. The system will be independent of metadata standards. At the time of installation of Agentic AI, one of the major steps is to configure the institute level metadata standards, based on the configuration, a database will be automatically created.

Design of institutional metadata.

```
<Records>
<!-- A single record keyed by "Field code" -->
<Record>
  <FieldCode>CODE_VALUE</Field Code>
  <!-- A list of fields that belong to this Field code -->
  <Fields>
    <Field>
      <Name>FIELD_NAME_1</Name>
      <Description>FIELD_DESCRIPTION_1</Description>
    </Field>
  </Fields>
</Record>
```

Field Mapping. Configure the flow of response from the API response to its chosen schema. Using CCM, each data element of the response API is mapped to the database schema.

The usual metadata formats include:

- MARC 21 - Machine Readable Catalogue
- Dublin Core
- MODS - Metadata Object Description Scheme
- BIBFRAME - Linked Data for Bibliographic Resources
- All these are mapped to institutional Schema.

Value Transformation. While building the schema CMM engine also defines rules for type conversion like date formats. Also helps to normalize the subject headings and language normalization with CMM rules.

Schema Enforcement. With the help of CMM system, we can achieve metadata integrity by

- Presence of required fields: System checks whether all mandatory fields are filled in before allowing the record to be saved or processed.
- Data type validation: It ensures each field contains the correct type of value, like numbers in numeric fields, dates in date fields, and text where text is expected.
- Controlled vocabulary enforcement: It forces users to select values only from approved lists or authority terms, preventing inconsistent or unauthorized terminology.
- Rule driven correction: This automatically fixes common errors or formatting issues based on predefined rules, reducing manual cleanup and improving data quality.

6.4 Agent Core Layer

This layer acts as the core entity for the system’s reasoning, automated coordination and management of multiple computer systems, applications, and services. Simply, it is a reasoning and orchestration engine. The layer is responsible for doing intent analysis, semantic explanation, retrieval strategy planning and response generation. Each unit in the layer has their own responsibility and roles to achieve the goal of the agent. Basically, the intelligent unit for the agent is the agent’s core layer.

Intent Analysis. It analyses and identifies the user query and the requirements of the user.

- Search query: The system interprets the user’s request and retrieves the most relevant records based on keywords, filters, and intent.
- Full-text processing: The engine analyzes the entire document content to extract concepts, entities, and context that improve retrieval accuracy.
- Alert generation: The system notifies users automatically when new content matches their saved topics, keywords, or research interests.
- User profile update: The platform refines the user’s preferences and behavior data based on their searches, interactions, and selected alerts.
- Analytics or reporting: The system produces visual or statistical summaries that show patterns in usage, searches, metadata, or system performance.

Semantic Expansion. Expands the query with related concepts using:

- Embedding similarity: The system finds semantically related terms by comparing vector embeddings and adds those concepts to strengthen the user’s query.
- Controlled vocabularies (LCSH, MeSH): The engine expands the query using standardized subject terms from authority lists to improve precision and consistency.
- Ontology mappings: The system identifies related concepts through hierarchical or semantic relationships defined in ontologies and adds them to the query.
- User profile context: The query is enriched with terms that match the user’s past behavior, interests, and domain preferences to deliver more personalized results.

This enables deeper discovery rather than simple keyword matching.

Retrieval Strategy Planning. Agent decides the optimal retrieval path:

- Full-text search in local DB: The agent triggers a direct text search across your stored documents when exact or phrase-level matching is most effective.
- Vector similarity search: The agent uses embedding-based retrieval when the query needs semantic understanding rather than keyword matching.
- Keyword indexing: The agent falls back to traditional indexed fields when structured metadata can answer the query faster and more accurately.
- API calls to external sources (if local data is missing): The agent queries external databases or services when your local repository doesn’t contain enough information to satisfy the request.
- Hybrid fusion retrieval: The agent blends result from multiple retrieval methods and ranks them to produce the strongest, most context-aware answer.

This strategy is dynamically selected based on input type, user behavior, and system capabilities.

Response Generation. The agent synthesizes:

- Search results: The agent merges and refines the retrieved items into a clean, ranked output for the user.
- Metadata summaries: The agent condenses key metadata fields into short, readable summaries that highlight the most important details.
- Recommendations: The agent generates suggestions for related documents, keywords, or next steps based on patterns in the retrieved data.
- Explanations: The agent provides short justifications describing how and why the results or recommendations were produced.
- This uses both deterministic rules and LLM-based natural language generation.

6.5 Data Ingestion Layer

This layer is responsible for the entire data pipeline from external sources. It supports scheduled harvesting or on demand harvesting and real time harvesting of metadata.

- Scheduled API Fetch: Supports time-based retrieval of metadata from external sources like e-resources database, ILMS, institutional repositories etc. with the help of APIs. It encourages timely data synchronization and keeps the system relevant.
- One-time fetch: One-time fetch facilitates data seeding, migration of datasets from external sources, specific data ingestion etc.
- Normalization Pipeline: Standardizes raw metadata into a neutral, predictable format before passing it to the CMM engine.

6.6 Storage Layer

The Storage Layer consists of databases handled by the Agent to retrieve the information needed to process the user queries. This layer comprises two databases.

- a) SQL Database - Stores harvested data and other information
- b) Vector Database - It helps in indexing, querying high-dimensional vector embeddings and ranking

a) The SQL Database consists of the following tables:

- CMM - Varies depending upon how the admin created the metadata schema
- Synonym_bank: Stores synonym expansions generated by LLM.
- Ontology_mappings: Maps keywords to their corresponding ontological concepts.
- Controlled_vocab: For LCSH, MeSH that supports indexing and interoperability.
- User_profiles: Stores personalized information related to users.
- User_Alerts: Supports Current Awareness Services (CAS) and Selective Dissemination of Information (SDI) by enabling timely configuration of personalised user alerts to these services.
- Ingestion_logs: collects data and stores from various workflows and records them for future purposes.
- User_logs: Collects user queries, interactions and patterns for better system enhancement.
- Llm_configuration: For Storing API keys

b) Vector Database

- Keyword_embeddings: Used for synonym generation, query expansion, and SDG matching.
- User_profile_embeddings: Allows personalized retrieval.

- Combined Architecture Summary (How they work together)

Table 1. Division of Responsibilities Between MySQL and Vector Database Components in the Proposed System Architecture

MySQL handles	Vector DB handles
▪ Metadata ▪ Logs ▪ Alerts ▪ Controlled vocab ▪ Synonyms ▪ Ontology ▪ User profiles ▪ LLM configuration	▪ Semantic search ▪ Query expansion via embeddings ▪ User personalization embeddings ▪ Hybrid fusion ranking

In addition to data storage, Storage Layer is also responsible for Clustering and Contextual recommendation. It also captures search history, query patterns, language preferences, subject interests, saved items, API calls, processing failures, records ingested, timestamps, and system actions. This ensures transparency and maintainability.

6.7 Large Language Model (LLM) Layer

This is the “reasoning module” plugged into the system. It is model agnostic which means it is independent of any specific machine learning or system model. The admin has the flexibility to replace, switch or integrate different models without replacing the core structure of the system. The module helps the admin to choose the model from hugging face models [15] for LLM integration.

Hugging face. Hugging Face is a prominent platform and ecosystem in the field of Large Language Models (LLMs), primarily known for providing inference mechanisms via its Transformer library. This library offers simple, easy-to-use, and concise APIs, particularly through components called pipelines, which abstract the complex code necessary to generate output based on input data (inference) for various tasks, such as text generation. Furthermore, Hugging Face hosts a vast array of resources, including hundreds of models that can be utilized for inference. The availability of open-source LLMs through platforms like Hugging Face enables developers to create AI-based solutions while maintaining control over elements like privacy and compliance, providing governance and ownership of the model deployment process. Hug Inference engines like the Hugging Face Transformers library (HFpl) are frequently studied in evaluations concerning LLM inference performance, where factors like hyperparameter optimization including tuning the batch size or the number of GPUs are analyzed to maximize throughput [16].

Model like

- Llama 3
- Mistral
- OpenAI GPT
- Gemini
- Falcon
- Hugging Face quantized models
- Locally hosted models

These are models that are responsible for:

- Metadata extraction
- Query understanding
- Contextual enrichment
- Summarization
- Embedding generation
- Reranking assistance

The LLM layer never replaces the agent core. It augments it.

6.8 Admin and Diagnostic Layer

This provides operational control and system transparency.

Command Line Interface (CLI). As mentioned earlier, the system will communicate to the outer world via either API or CLI. Here the CLI helps the admin to configure, test and diagnose the agent.

Admins can:

- Configure LLMs: Admins can update model settings, temperature, API endpoints, and model selection directly from the command line.
- Test ingestion: Admins can run a CLI command that simulates or triggers metadata ingestion to verify source connectivity and parsing.
- Run validation checks: A CLI tool scans stored records for missing fields, data-type issues, and vocabulary violations before they reach production.
- Rebuild embeddings: Admins can regenerate all documents or keyword embeddings when models are upgraded or content changes.
- Manage API keys: A secure CLI interface lets admins create, rotate, or revoke API keys used by collectors, clients, or integration services.
- View logs: Admins can quickly inspect real-time logs or filtered log categories (errors, ingestion, agent steps) through terminal commands.

Monitoring and Logs. Also, the agent will collect the logs which include:

- Error traces: The system captures full stack traces of failures to help admins quickly identify and fix issues.
- Slow query logs: The monitoring layer records queries or retrieval operations that exceed performance thresholds.
- Agent reasoning logs: The system logs how the agent made decisions, including retrieval strategies, ranking choices, and rule triggers.

- Performance metrics: Key system metrics such as latency, ingestion time, vector search speed, and API throughput are continuously tracked.

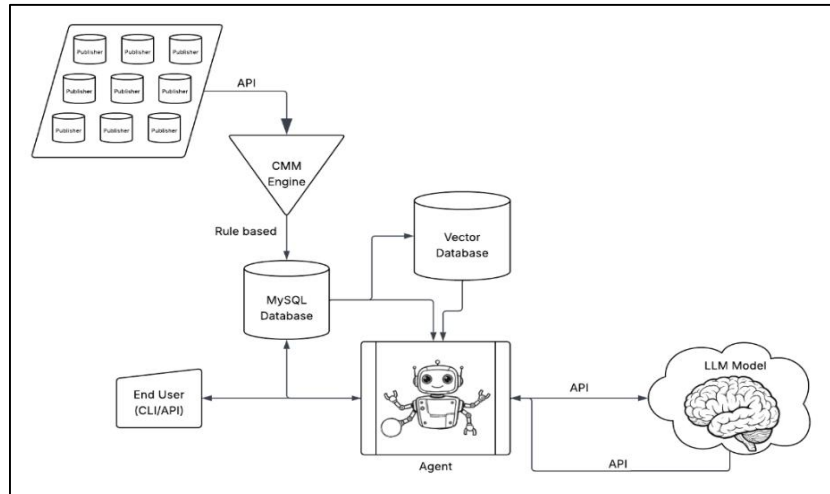


Fig. 1. High-Level Architecture of the Agentic AI-Driven Library Discovery System Showing Data Ingestion, Metadata Management, Vector Search, and LLM Interaction

The complete Agent/system is compiled as a python SDK which is hosted as standard Python package via PyPI repository.

As mentioned earlier, the current agent is only designed for information retrieval, but the agent can be designed with more features that can perform several tasks which can be done in a library.

Workflow. Install the Agent as SDK from PyPI repository

- Pip installs an intelligent search-agent
- System asks for CMM engine configuration to create fields
- System will ask for API connection and its configuration to connect the publisher database
- Map the API response with database
- Configure Admin email
- Ask about the LLM model to be added from huggingface. Based on the user selection system, it will automatically download and start configuring the system including the secret key. The system will also check the connection strength.
- One-time data fetching from all API of publisher and store on the database.
- Asks for the configuration of daily scheduled injection of API response.
- The system will prompt the API URLs and its Key that is used to communicate with Agent.

7 Why Agentic AI, not Intelligent Retrieval Software

Retrieval software is only a functional component of the system. Agentic AI is an autonomous worker that uses retrieval, reasoning, planning, and decision-making to produce intelligent outcomes.

- Agentic AI makes decisions: It chooses whether to run keyword search, vector search, ontology expansion, external API lookup, or hybrid fusion.
- Retrieval software does not reason: It cannot explain why it has selected one method over another.
- Agentic AI uses reasoning traces: It logs the decision path, evaluation steps, rule triggers, and ranking logic.
- Retrieval software is static: It cannot adapt to user preferences, profile data, search patterns, or domain context.
- Agentic AI adapts continually: It learns from user behavior, refines the profile, updates embeddings, and adjusts strategies.

Further feature integration. In this paper, information retrieval and discovery services are enabled as the feature of agents. But as mentioned in literature, AI has been used in multiple ways in libraries like, information retrieval, cataloging assistant, classification assistant, chatbot etc. The same feature can be added, and the Agent will be

capable of multiple tasks. These agents can communicate to the real world with multiple connectors, sensors etc. This can achieve the implementation of mobile and non-mobile robots.

Agentic AI in Robotics. Agentic AI is implemented in Robotics with the integration of foundation models (FMs), such as Large Language Models (LLMs), Vision-Language Models (VLMs), Diffusion Models, and Robotics Foundation Models (RFMs). They help in developing autonomous robotic agents. These integrated models unlock robotic autonomy by enabling the agents to independently understand and reason about semantic contexts, plan actions, physically interact with their surroundings, and adapt to novel or untrained tasks. This agentic capability is systematically applied across three fundamental elements of robotic autonomy: perception and situational awareness (e.g., semantic understanding), decision making and task planning (e.g., task decomposition and validation), and motion planning and control (e.g., trajectory and control command generation) [17].

Achieving context engineering. AI agents are autonomous software programs endowed with human-like intelligence, capable of learning from data and executing actions, while contextually aware systems are implemented to meticulously capture the essence of past interactions, including context, user intent, and conversational cadence. The implementation of these systems significantly enhances efficiency by avoiding redundant information, while the agents' adaptability ensures they maximize their effectiveness and relevance in dynamic, real-world scenarios [18]. The key emerging trend is the advancement toward multimodal agents that integrate diverse sensory inputs and leverage Large Language Models (LLMs) to achieve contextual, situational, and embodied intelligence, serving as a crucial vehicle for attaining Artificial General Intelligence (AGI) goals [19].

8 Conclusion

Agentic AI is an intelligent system that can independently plan and adjust the workflow and operate in complex environments with minimal human supervision. The scattered, direct and indirect use of AI in library by librarian for multiple works and services are increasing day by day. The multi-tasking and being intelligent using AI models enable Agentic AI to help to bring all those tasks under one umbrella. This helps to build an intelligent agent for libraries to enhance the user support and reduce the stress of work. The future of advanced libraries is AI enabled libraries, and if it is an agentic AI then a single system can perform multiple tasks at the same time. The paper discussed how an agent designed for intelligent information retrieval and discovery services should be. The paper also solves conflicts in different metadata used by different publishers using CCM Model. The key achievements of the paper are design and architecture of intelligent agents for information retrieval and discovery services, also discussing other features that can be integrated into the same agent to be complete library assistant. This paper also leaves a scope of study on how the same agentic AI can be used to build mobile and non-mobile robots in libraries.

Declaration of AI and AI-Assisted Technologies used in Manuscript Writing Process: Not Applicable

References

1. Abdulwahid, A. H., Pattnaik, M., Palav, M. R., Babu, S. B. G. T., Manoharan, G., & Selvi, G. P. (2023). Library Management System Using Artificial Intelligence (pp. 1–7). <https://doi.org/10.1109/iconstem56934.2023.10142904>
2. Huang, Y. (2024). Exploring the implementation of artificial intelligence applications among academic libraries in Taiwan. *Library Hi Tech*, 42(3), 885–905. <https://doi.org/10.1108/lht-03-2022-0159>
3. Lappalainen, Y., & Narayanan, N. (2023). Aisha: A custom AI library chatbot using the ChatGPT API. *Journal of Web Librarianship*, 17(3), 37–58. <https://doi.org/10.1080/19322909.2023.2221477>
4. Lo, L. S. (2023). An initial interpretation of the U.S. Department of Education's AI report: Implications and recommendations for academic libraries. *The Journal of Academic Librarianship*, 49(5), 102761. <https://doi.org/10.1016/j.acalib.2023.102761>
5. Kaushal, V., & Yadav, R. (2022). The Role of Chatbots in Academic Libraries: An Experience-based Perspective. *Journal of the Australian Library and Information Association*, 71(3), 215–232. <https://doi.org/10.1080/24750158.2022.2106403>
6. Gupta, V. (2025). From hype to strategy: Navigating the reality of experimental strategic adoption of AI technologies in libraries. *Reference Services Review*, 53(1), 1–14. <https://doi.org/10.1108/rsr-08-2024-0042>
7. Panda, S., & Kaur, N. (2023). Revolutionizing language processing in libraries with SheetGPT: An integration of Google Sheet and ChatGPT plugin. *Library Hi Tech News*, 41(5), 8–10. <https://doi.org/10.1108/lhtn-03-2023-0051>

8. Zhang, C., Tian, L., & Chu, H. (2023). Usage frequency and application variety of research methods in library and information science: Continuous investigation from 1991 to 2021. *Information Processing & Management*, 60(6), 103507. <https://doi.org/10.1016/j.ipm.2023.103507>
9. Saeidnia, H. R. (2023). Ethical artificial intelligence (AI): Confronting bias and discrimination in the library and information industry. *Library Hi Tech News*. <https://doi.org/10.1108/lhtn-10-2023-0182>
10. Bandi, A., Kongari, B., Naguru, R., Pasnoor, S., & Vilipala, S. V. (2025). The rise of Agentic AI: A Review of Definitions, Frameworks, Architectures, Applications, Evaluation Metrics, and Challenges. *Future Internet*, 17(9), 404. <https://doi.org/10.3390/fi17090404>
11. Acharya, D. B., Kuppan, K., & Divya, B. (2025). Agentic AI: Autonomous Intelligence for Complex Goals—A Comprehensive Survey. *IEEE Access*, 13, 18912–18936. <https://doi.org/10.1109/access.2025.3532853>
12. Narayanan, N. (2024). Beyond chatbots: The emergence of autonomous AI agents in libraries. In K. P. Vijayakumar, K. P. Sadasivan, M. G. Sreekumar, & K. G. Sudhier (Eds.), *Navigating the information landscape: Trends and innovations in libraries. A Festschrift in honour of Prof. K. A. Isaac* (pp. 230–252). Prof. K. A. Isaac Society for Library and Information Science.
13. Ghonghe, S., & Gosavi, A. (2025). Agentic AI [eBook]. Konverge AI. <https://konverge.ai/pdf/Ebook-Agentic-AI.pdf>
14. Stryker, C. (2025). What is agentic automation? IBM. <https://www.ibm.com/think/topics/agentic-automation#257779832>
15. Hugging Face. (2025, December 8). Hugging Face – The AI community building the future. <https://huggingface.co/>
16. Martinez, M. (2025). The Impact of Hyperparameters on Large Language Model Inference Performance: An Evaluation of vLLM and HuggingFace Pipelines. *Proceedings of the 33rd ACM International Conference on the Foundations of Software Engineering*, 1672–1678. <https://doi.org/10.1145/3696630.3728704>
17. Jang, D., Cho, D., Lee, W., Ryu, S., Jeong, B., Hong, M., Jung, M., Kim, M., Lee, M., Lee, S., & Choi, H. (2024). Unlocking Robotic Autonomy: A Survey on the Applications of Foundation Models. *International Journal of Control Automation and Systems*, 22(8), 2341–2384. <https://doi.org/10.1007/s12555-024-0438-7>
18. Durgaprasad, K. V. V. B., Abozibid, H. K., Hawas, J. N., & Krishna, P. B. S. (2024). AI agents and conversation system. In *Proceedings of the 2024 International Conference on Augmented Reality, Intelligent Systems, and Industrial Automation (ARIIA)* (pp. 1–7). IEEE. <https://doi.org/10.1109/ARIIA63345.2024.11051848>
19. Jain, R. C. (2024). Multimodal agents: From vision to reality. *IEEE Multimedia*, 31(4), 4–16. <https://doi.org/10.1109/MMUL.2024.3485253>

Linked Data and Knowledge Graphs in LIS

Readiness of Indian Cultural Heritage Websites for Linked Open Data: An Empirical Study Based on the Five-Star Model

Anuradha Kumari¹ [0009-0000-1828-1697], Bhakti Gala² [0000-0002-3512-9395] and Satyam Kumar³ [0009-0007-6669-6548]

^{1,2}Central University of Gujarat, Vadodara, Gujarat, India

³Central Library, IIT Ropar, Panjab, India

anuradhamishra174@gmail.com

Abstract. The speedy growth of digitisation of Cultural Heritage (CH) resources has increased the need for data practices that ensure seamless data integration, enhanced data discovery, improved interoperability, and reuse. This study assesses selected Indian Government Cultural Heritage websites using Tim Berners-Lee's five-star schema for Linked Open Data (LOD) proposed in 2010. A descriptive and qualitative research methodology was adopted, involving a systematic examination of the availability of data on websites; their structure and machine readability; the presence of CH data in a non-proprietary format; publication of data using open standards, and linking of CH data to other LOD.

The findings reveal that these websites have a substantial amount of CH data available and that the data are provided in a structured format, using non-proprietary open formats for publishing the data, thereby fulfilling the first star, and partially complying with the second and third stars of the five-star schema of LOD. The websites neither publish the data using open standards like RDF or SPARQL nor link to other LOD, nor hence do not follow the fourth and fifth stars. Overall, the study concludes that the Indian CH websites are at a nascent stage of adopting LOD practices. An emphasis on structured data publication using open standards, interlinking with global organisations that are already using LOD, is required to move towards the higher levels of the five-star schema. Implementing these practices will significantly improve the visibility, reusability, and scholarly value of Indian CH Data within the global open data ecosystem.

Keywords: Linked Open Data (LOD), Five-Star Open Data Schema, Linked Open Data; Digital Cultural Heritage, Indian Cultural Heritage Websites, Data Interoperability and Reuse.

1 Introduction

Cultural heritage links individuals to share social and cultural values, traditions, beliefs, and religious practices, and is reflected in the ways of life developed by communities and transmitted across generations (Kumar & Nair, 2021). Cultural Heritage (CH) data is characterised by syntactic and semantic heterogeneity, multilingualism, rich semantics with extensive interlinking, and is produced in a distributed and open manner by museums, libraries, archives, media organisations, and individual contributors [11]. In the last decade, India has invested heavily in digitization initiatives to capture CH content from cultural communities across the nation. The goal of these initiatives is not only to make culture more accessible to the people but also to present CH information in a way that enables their use and creative reuse. CH institutions, including libraries, archives, and museums (LAMs), are actively exploring innovative methodologies to effectively engage and educate their diverse user communities [19]. The most effective means of disseminating such information is to do so in accordance with Linked Open Data (LOD) principles. Linked Data is a set of best practices for publishing structured data on the Web coined by Tim Berners-Lee (2006). LOD comprises both Linked Data and Open Data [21], which means data is linked and released under an open license, hence LOD is a powerful amalgamation of the two concepts related to Data.

In 2010, Tim Berners-Lee proposed a five-star schema for LOD, which includes:

- ☆ Data is available on the Web, in whatever format.
- ☆☆ Available as machine-readable structured data (i.e., not a scanned image).
- ☆☆☆ Available in a non-proprietary format (i.e, CSV, not Microsoft Excel).
- ☆☆☆☆ Published using open standards from the W3C (RDF and SPARQL).
- ☆☆☆☆☆ All of the above and links to other LOD.

India's rich tapestry of indigenous cultures faces increasing pressures from globalization and modernization, with digital technologies offering effective pathways for their systematic preservation, documentation, and revi-

talization [24]. Digitizing cultural heritage ensures sustained public accessibility to these resources without restrictions, while also providing comprehensive and consolidated documentation [15]. Emerging technologies, such as Linked Data for digital cultural heritage, present both new challenges and opportunities for memory institutions in modeling their digital collections for open access [29]. The crucial role of knowledge management and knowledge innovation empowers Linked Data projects and services [28]. The redefinition of the traditional data models for CH resources into data models based on semantic technology has been a major activity of the CH community [9]. Applying LOD concepts to the CH domain has become an active field with many libraries, museums, and archives currently exploring ways to convert their data into the RDF format to develop new interfaces, and thus providing a richer experience for users of CH websites [6]. The current study has adopted the following objectives:

- To identify Indian CH websites compliant with the LOD five-star principles.
- To explore the possible solutions for implementation by the CH websites to adhere to LOD five-star principles.

2 Literature Review

Galleries, Libraries, Archives, and Museums, known as CH institutions, work collaboratively to preserve and disseminate the CH of the community, society, or nation [22, 23, 13]. Linked open data (LOD) is an organised, well-connected feature of the emerging paradigm for data available on the Web, making it easily accessible through the web in a well-defined manner to people [5, 27]. Tim Berners Lee (2006) established the five-star principles to measure the open access maturity level of LOD. These principles are intended to assess how well data are integrated into the web, identifying the formats and technologies used to create open-access data on the web [10], and further describing the mandate for openness [19]. In the future, the challenge of accessing CH Linked Open Data will grow as more CH institutions share their data to enhance collection usability and synergy, empowering and transforming historic organisations [1][6][18]. Experts emphasize that achieving LOD compliance will require not only technical measures, but also organizational readiness and alignment with open data policies to progress towards full adherence to LOD five-star principles [11].

Despite India's rich CH, little effort has been made to integrate resources from galleries, libraries, archives, and museums [8]. Research indicates that many national CH portals in India publish digitized data using basic web format. In India, the National Data Sharing and Accessibility Policy (NDSAP) promote open and machine-readable data, but does not explicitly mandate LOD, which partially explains the limited adoption among CH websites (Government of India, 2012).

This study advances existing work by shifting the focus from general discussions of LOD adoption to an empirical assessment of five-star Linked Open Data compliance across Indian government cultural heritage websites. By evaluating the extent to which selected websites meet each level of the five-star schema, the study provides concrete evidence of India's current LOD maturity in the cultural heritage sector. By doing so, the research contributes to a clearer understanding of practical LOD readiness at the national level. It provides actionable insights for enhancing semantic interoperability and data reuse across Indian cultural heritage institutions.

3 Methodology

This study employs a descriptive research design, supported by qualitative content analysis, to evaluate how Indian CH websites align with the five-star LOD Principles. Using purposive sampling, six Government of India websites were selected based on two criteria: first, they belonged to the official government domain, and second, they contained substantial CH content, such as museum collections, manuscripts, archives, archaeological resources, or digital catalogues. In the initial phase of the study, more than six Government of India (GoI) cultural heritage websites were shortlisted for evaluation. Each website was systematically examined to verify the presence of digitised cultural heritage data, resulting in the selection of six websites that met the inclusion criteria. The final sample included both aggregator platforms and individual institutional websites. Aggregator portals were prioritised as they provide wider coverage of cultural heritage resources; for instance, the National Portal and Digital Repository of Indian Museums aggregates content from eleven museums through direct hyperlinks, whereas the National Mission for Manuscripts, despite its aggregator role, does not provide explicit links to the contributing repositories. The list of selected websites has been reflected in the table 1.

Table 1. Selected websites with Indian CH data

Sr.No	Organisation	Official Website
1	National Mission for Manuscripts	https://www.namami.gov.in/
2	National Mission on Monuments and Antiquities	http://nmma.nic.in/nmma/indexAction.do
3	NDLI — Cultural Heritage	https://ndl.iitkgp.ac.in
4	IGNCA Digital Repository	https://ignca.gov.in/r
5	National Portal and Digital Repository for Museums of India	https://museumsofindia.gov.in/repository
6	India Culture	https://indianculture.gov.in/

These websites contain paintings, manuscripts, philately, decorative art, beads, arms and armour, terracotta, sculpture, ornaments, drawing, and many more types of CH objects. Data were collected by systematically visiting each website, examining digital resources, and then extracting relevant elements into a structured Excel sheet. Each website was assessed using established indicators derived from the five-star Linked Data model. The collected data were analysed with the overall aim of determining the extent to which these CH websites adhere to the five-star schema of LOD.

The first star was assessed by verifying the presence of digitised cultural heritage artefacts or metadata accessible through the website. Whereas the second star was evaluated by checking whether the data were provided in structured, machine-readable forms rather than scanned or unstructured files. The third star was assessed by examining whether the data were published in open, non-proprietary formats that support reuse. Fourth, were evaluated by inspecting the source code or data access mechanisms for standard semantic formats and verifying the presence of links to external LOD resources, with the fifth star requiring fulfillment of all preceding criteria with linking to other Linked Data. The authors operationalized the compliance criteria as follows in Table 2:

Table 2. Operational Definitions

<i>Data availability</i>	“Yes”, if the data is publicly available (metadata and/or CH object are displayed on the website), and “No”, if none were found, and “Partial” if any one of the above is on the website.
<i>Machine-readable structured data</i>	We view machine readability in two respects. One is whether it can be accessed through a web browser (metadata is embedded in the HTML of the search query or listing), and the second is whether there is an API to access the above data. Information about structured data was deciphered by gathering information about the standard followed (for e.g., MARC, Dublin Core, METS, EAD, schema.org, etc.). If the website adheres to all conditions, it was documented as “Yes,” and “No”, if none are fulfilled, and “Partial”, if one condition is met. E.g., refer to the link https://www.pandulipatala.nic.in/manus/search-view?id=Nzk1
<i>Available in a non-proprietary format</i>	“Yes”, if the data was available in non-proprietary file formats (available as HTML, PDF, JPG, CSV, JSON, XML, or RDF), and “No”, if none of the conditions were fulfilled, whereas “Partial”, if the data fulfilled any one condition.

<i>Published using open standards from the W3C (RDF and SPARQL).</i>	“Yes”, if the website’s HTML page is using open standards like RDF and SPARQL to publish the data on the web. “No”, if they are not using an open standard to publish the data, as mentioned above.
<i>All of the above and links to other LOD.</i>	“Yes”, if satisfying all four stars and linking to other LOD. “No”, if not satisfying the initial four stars and linking to other LOD.

4 Findings and Data Analysis

This section presents the analysis of the data collected by the researchers to assess the adherence of selected CH websites to the Five-Star LO, as reflected in Table 3.

Table 3. Adherence of the five-star schema for LOD by selected CH websites of India

Organization	*	**	***	****	*****
National Mission for Manuscripts	Yes	Partial	Partial	No	No
National Mission on Monuments and Antiquities	Yes	Partial	Partial	No	No
NDLI Cultural Heritage	Yes	Partial	Partial	No	No
IGNCA Digital Repository	Yes	Partial	Partial	No	No
National Portal and Digital Repository for Museums of India	Yes	Partial	Partial	No	No
India Culture	Yes	Partial	Partial	No	No

4.1 ★ Data is available on the Web, in whatever format

All six websites included in the study provide digital data and therefore satisfy the first star of the five-star LOD model, which requires data to be available on the web. The data across these websites is primarily available in the form of images, while the National Mission for Manuscripts (NMM) provides content in PDF format, whereas the Indira Gandhi National Centre for the Arts (IGNCA) and the Indian Culture website, which serves as a major resource for arts documentation and cultural understanding and documents diverse forms of CH, offer data in multiple formats, including images, text files, audio, and video files. This diversity of formats reflects varying institutional objectives and stages of digital maturity among the selected CH websites. When it comes to the license, all these websites keep the copyright to themselves and do not provide the data under an open License, which is an important aspect of LOD. Globally and in developing countries like India, barriers to “open” CH data include legal and economic barriers such as complex rights clearances, high digitization and management costs, and institutional reluctance to relinquish control, which together limit interoperability and reuse [25][26].

4.2 ☆☆ Available as machine-readable structured data (i.e., not a scanned image).

An examination of the selected websites' source codes and metadata structures of data available indicates that the data is in a partially structured format, but not in a machine-readable format. Consequently, these websites partially satisfy the second star of the LOD five-star model. The absence of standardised structured data formats, such as CSV, XML, and JSON, limits machine interpretability and interoperability, thereby constraining further reuse and linkage of the data.

4.3 ☆☆☆ Available in a non-proprietary format (i.e., CSV, not Microsoft Excel)

The data is primarily presented in a partially structured format. The websites use HTML (Hyper Text Markup Language), which is an open standard for publishing data on the web, but is not in a reusable format. These websites contain the data mainly in image and textual form and use mainly PDF, JPG, PNG, etc., which are non-proprietary file formats. Hence, partially fulfilling the third star also.

4.4 ☆☆☆☆ Published using open standards from the W3C (RDF and SPARQL)

These websites do not comply with the fourth star of the Linked Data principles. This was verified through manual inspection of the websites' source code, which revealed the absence of RDF or SPARQL. Hence, these websites are not using open standards from the W3C.

4.5 ☆☆☆☆☆ All of the above and links to other Linked Open Data.

As we move forward, these websites comply with the first, second, and third stars of Linked Data, but they start failing from adhering to the fourth star of LOD Schema. Since the fifth star requires fulfilling all the stars, including linking to other LOD and so on, for both cases, this data is available on these websites, which do not link to any other LOD.

5 Possible Solutions

As a possible solution to address the absence of compliance with the fifth star of Linked Data, the datasets of these websites can be interlinked with established LOD resources. In this study, the National Digital Library of India (NDLI) is considered as a reference example to demonstrate such an approach. Meaningful linking can be achieved at least at the levels of title, subject, and the creator's name, thereby enhancing semantic interoperability and contextual richness. For instance, the title of a piece of work can be linked to authoritative bibliographic datasets such as the Library of Congress and Virtual International Authority File. To illustrate, Ignited Minds, a well-known motivational book by former Indian President and renowned scientist Dr. A.P.J. Abdul Kalam, can have its title and subject aligned with the corresponding Library of Congress record, while the author's name can be linked to a persistent identifier in the Virtual International Authority File (VIAF). Such linkages not only smooth the path of disambiguation and global discoverability but also allow integration of local CH data within the broader LOD ecosystem.

VIAF aggregates numerous national name authority files in multiple languages and scripts into a single name authority service hosted by OCLC, and it is considered an early example of LOD [7]. Linking the names of authors with this platform will lead the users to different libraries, and it will basically connect all the authors' names and their work available in libraries situated in different parts of the world. The different works of Dr A.P.J. Abdul Kalam are available in different libraries across the world. As shown, the availability of the discussed author's name in Figure 1, if the author's name is linked to VIAF, then it will link it to different works by the author in different libraries that are contributing to VIAF.

Abdul Kalam, A. P. J. (Avul Pakir Jainulabdeen), 1931-2015. | Speeches. Selections

VIAF ID: 177161050 (Work)

Permalink: <http://viaf.org/viaf/177161050>

Fig. 1. A. P. J Abdul Kalam's name authority link record VIAF records (Source: <https://viaf.org/en/viaf/177161050> accessed on 26th January, 2026)

Same as linking the name of Creator/authors, the title of the data, as the availability of title can be seen in both the organisations, NDLI reflected in figure 2, as well as the Library of Congress Catalogue shown in figure 3, can be connected to the Library of Congress, since the organisation is already using LOD technology. By establishing a link between the NDLI title record and the corresponding Library of Congress record, the work can be unambiguously identified despite variations in language, script, or edition. This linkage refines discoverability by enabling users and systems to retrieve related information about Ignited Minds from various international knowledge centres. Furthermore, it supports data reuse and integration, allowing researchers to link NDLI resources with other CH collections, bibliographic datasets, and scholarly platforms.

Same as the author's name and the title of the work, the subject, as we can see the availability in LC in figure 3, which can be linked with the Library of Congress to make the data more accessible and retrievable for libraries across the world. The above illustration and proposed possible solutions show how the title of any work available on websites, its author, and its subject can be linked to other websites that are already using LOD technology.

Ignited Minds Unleashing The Power Within India

Content Provider	Digital Library of India
Author	Kalam, A. P. J. Abdul
Description	Custodian Institution: Sni, Vetapalem.
Page Count	224
File Format	PDF
Language	English
Publisher	Viking By Penguin Books
Publisher Date	2002-01-01
Requires	PDF viewer Plugin
Access Restriction	NDLI
Subject Keyword	Social Sciences • Viking By Penguin Books • Politics • A P J Abdul Kalam
Content Type	Text
Resource Type	Book

Fig. 2. Scope for linking title, author and subject on NDLI (Source: https://www.ndli.gov.in/ch_document/dli/dli_ndli/201496?e=0|ignited%20minds%20unleashing%20the%20power%20within%20india|| accessed on 26th January, 2026)



Fig. 3. LC Catalogue showing linked title and subject of “Ignited Minds (Source: <https://search.catalog.loc.gov/instances/3c516393-2bb7-5fbd-b0a5-4a633843e65e?option=keyword&query=Ignited%20Minds> accessed on 26th January, 2026)

6 Limitations and Future Work

The study is limited by a small sample size of six Indian government websites due to the lack of CH websites fulfilling the criteria of open data availability. The authors acknowledge the limitation of using binary judgment in data presentation and propose to extend the study using graded or weighted scoring models in the future to capture varying levels of LOD maturity across the cultural heritage websites. The evaluated websites were not directly compared with established Linked Open Data-enabled platforms such as Europeana or the Getty Research Institute, which represent mature implementations of the fourth and fifth stars through standardised ontologies, persistent identifiers, and extensive external linking. Accordingly, this is acknowledged as a limitation of the present study. Future research can undertake comparative analyses with such international platforms to more clearly illustrate advanced LOD compliance and inform improvements in Indian cultural heritage portals. More Indian government websites that contain any government data can be explored with the same set of schema in the future. The CH websites of India can be explored with frameworks like FAIR (Findable, Accessible, Interoperable, Reusable), CARE (Collective Benefit, Authority to Control, Responsibility, Ethics), and an open data ecosystem to get a more in-depth understanding of their current state of publishing the data on the web and of their alignment with the broader open science movement.

7 Conclusion

The study evaluated selected Indian government CH websites to examine their compliance with the five-star schema of Linked data proposed by Tim Berners-Lee to understand the readiness and preparedness of Indian CH institutions for LOD. All the websites comply with the first star, partially fulfil the second, and the third star, whereas all the websites start failing to comply with the schema from the fourth star onward and do not publish their data using open standards from the W3C (RDF and SPARQL). Overall, the finding suggests that the Indian CH institutions are at the nascent stage of their Linked Data journey. A key insight of the study is that India already has a presence of rich metadata, authoritative institutions, and national aggregation portals. The gap is not in the content but in the absence of data modelling with the missing LOD layer. Although the National Data Sharing and Accessibility Policy (Government of India, 2012) and the subsequent Implementation Guidelines (2015) promote Open Data and set the foundation for the government’s Open Government Data (OGD) Policy; they do not explicitly mention Linked Open Data. The OGD (data.gov.in) platform in India already adheres to open data formats, and although NDSAP does not mandate LOD, the CH websites could benefit by adopting the LOD principles. The overall compliance of these websites with the five-star schema of LOD will not only improve the visibility and use of Indian CH data but will also help researchers, digital humanists, and policy makers in building a more interconnected and sustainable Cultural Heritage data ecosystem. The establishment of India’s indigenous Digital Heritage Stack by In heritage Foundation marks a decisive move toward a standards-driven, interoperable, and Linked Open Data-ready ecosystem, enhancing metadata consistency and semantic integration for sustainable cultural heritage research [14].

Declaration of AI and AI-Assisted Technologies used in the Manuscript Writing Process: During the preparation of this manuscript, the author used ChatGPT (GPT-5.2, OpenAI) solely for language correction, including grammar, clarity, and academic tone.

References

- Alexiev, V. (2018). Museum linked open data: Ontologies, datasets, projects. *Digital Presentation and Preservation of Cultural and Scientific Heritage*, 8, 19–50. <https://doi.org/10.55630/dipp.2018.8.1>
- Al-Khatib, B., & Ali, A. A. (2021). Linked data: A framework for publishing five-star open government data. *International Journal of Information Technology and Computer Science*, 13(6), 1–15. <https://doi.org/10.5815/ijitcs.2021.06.01>
- Berners-Lee, T. (2006). Linked data – Design issues. World Wide Web Consortium. <https://www.w3.org/DesignIssues/LinkedData.html>
- Bhardwaj, S., Pavani, P. R., Fukey, L. N., Jyoshna, C., Valluri, J. Y., & Ibadilullayeva, B. N. (2024). Examining the development of historical computing tools for visualizing cultural heritage data. In *Proceedings of the 2024 International Conference on Advances in Computing, Communication and Materials (ICACCM)* (pp. 1–6). IEEE. <https://doi.org/10.1109/ICACCM61117.2024.11059130>
- Bizer, C. (2009). The emerging web of linked data. *IEEE Intelligent Systems*, 24(5), 87–92. <https://doi.org/10.1109/MIS.2009.102>
- Candela, G., Escobar, P., Carrasco, R. C., & Marco-Such, M. (2019). A linked open data framework to enhance the discoverability and impact of cultural heritage. *Journal of Information Science*, 45(6), 756–766. <https://doi.org/10.1177/0165551518812658>
- Chou, C. (2025). Linking challenges for pseudonyms: The studies of VIAF, ISNI, Wikidata, Meridian, and beyond. *Cataloging & Classification Quarterly*, 63(4), 195–239. <https://doi.org/10.1080/01639374.2025.2502329>
- Dutta, A., & Mukhopadhyay, P. (2022). Towards unified retrieval system for GLAM institutions in India: Designing a prototype for biblio-cultural information space. *Annals of Library and Information Studies*, 69(1), 75–91. <https://doi.org/10.56042/alis.v69i1.58292>
- Freire, N., Voorburg, R., Cornelissen, R., Valk, S. D., Meijers, E., & Isaac, A. (2019). Aggregation of linked data in the cultural heritage domain: A case study in the Europeana network. *Information*, 10(8), Article 252. <https://doi.org/10.3390/info10080252>
- Gunawan, D., & Amalia, A. (2016). The implementation of open data in Indonesia. In *Proceedings of the 2016 International Conference on Data and Software Engineering (ICoDSE)* (pp. 1–6). IEEE. <https://doi.org/10.1109/ICoDSE.2016.7936164>
- Hyvönen, E. (2012). Publishing and using cultural heritage linked data on the semantic web. Springer. <https://doi.org/10.1007/978-3-031-79438-4>
- Hallo, M., Luján-Mora, S., Maté, A., & Trujillo, J. (2016). Current state of linked data in digital libraries. *Journal of Information Science*, 42(2), 117–127. <https://doi.org/10.1177/0165551515594729>
- Harisanty, D., Obille, K. L. B., Anna, N. E. V., Purwanti, E., & Retrialisca, F. (2024). Cultural heritage preservation in the digital age: Harnessing artificial intelligence for the future—A bibliometric analysis. *Digital Library Perspectives*, 40(4), 609–630. <https://doi.org/10.1108/DLP-01-2024-0018>
- Inheritance Foundation. (2025, November 19). India finally has its own digital heritage stack — And it's built for researchers, ASI, and cultural institutions. Medium. <https://medium.com/@inheritage.foundation/india-finally-has-its-own-digital-heritage-stack-and-its-built-for-researchers-asi-and-af63a67bd006>
- Kara, G. K. T. (2021). Developing a sustainable cultural heritage information system. *Library Hi Tech News*, 38(6), 17–20. <https://doi.org/10.1108/LHTN-08-2021-0053>
- Kumar, G. T. K. (2023). Sustainable preservation and accessibility to cultural heritage in India. *Library Hi Tech News*, 40(2), 12–14. <https://doi.org/10.1108/LHTN-02-2022-0029>
- Kumar, G. T. K., & Nair, R. R. (2021). Conserving knowledge heritage: Opportunities and challenges in conceptualizing cultural heritage information system (CHIS) in the Indian context. *Global Knowledge, Memory and Communication*, 71(6–7), 564–583. <https://doi.org/10.1108/GKMC-02-2021-0020>
- Marcondes, C. H. (2018). Culturally relevant relationships: Publishing and connecting digital objects in collections of archives, libraries, and museums over the web. In F. Ribeiro & M. E. Cerveira (Eds.), *Challenges and opportunities for knowledge organization in the digital age* (pp. 539–553). Ergon Verlag. <https://doi.org/10.5771/9783956504211-539>
- Marden, J., Li-Madeo, C., Whysel, N., & Edelstein, J. (2013). Linked open data for cultural heritage: Evolution of an information technology. In *Proceedings of the 31st ACM International Conference on Design of Communication* (pp. 107–112). ACM. <https://doi.org/10.1145/2507065.2507103>
- Mustillo, T. (2025). Building a FAIR, linked, and open data ecosystem: Innovation cascades in political science. *European Political Science*, 24(4), 810–825. <https://doi.org/10.1057/s41304-025-00520-0>
- Ontotext. (n.d.). What are linked data and linked open data? Retrieved December 26, 2025, from <https://www.ontotext.com/knowledgehub/fundamentals/linked-data-linked-open-data/>
- Pandey, R., & Kumar, V. (2023). A survey of the metadata element sets used for digital art objects in the online collections of the museums of India. *DESIDOC Journal of Library & Information Technology*, 43(4), 226–233. <https://doi.org/10.14429/djlit.43.04.19228>

23. Pandey, R., & Kumar, V. (2025). Web accessibility of Indian cultural heritage information: An analytical study of museum websites. *Journal of Cultural Heritage Management and Sustainable Development*, 15(3), 651–663. <https://doi.org/10.1108/JCHMSD-08-2021-0153>
24. Singh, B., & Kumar, A. (2024). Examining the impact of digitalization on indigenous cultural heritage preservation in India. *Enigma in Cultural*, 2(1), 43–57. <https://doi.org/10.61996/cultural.v2i1.59>
25. UNESCO. (n.d.). Open data. <https://www.unesco.org/en/open-solutions/open-data>
26. UNESCO. (n.d.). Tangible and intangible heritage. <https://ich.unesco.org/en/tangible-and-intangible-heritage-00097>
27. Tsoukalas, C., Dervos, D., Martinez-Gil, J., & Aldana-Montes, J. F. (2012). TheMa: An API for mining linked datasets. In *Proceedings of the 16th Panhellenic Conference on Informatics* (pp. 449–453). IEEE. <https://doi.org/10.1109/PCi.2012.58>
28. Zhang, L. (2022). Empowering linked data in cultural heritage institutions: A knowledge management perspective. *Data and Information Management*, 6(3), Article 100013. <https://doi.org/10.1016/j.dim.2022.100013>
29. Ziku, M. (2020). Digital cultural heritage and linked data: Semantically-informed conceptualisations and practices with a focus on intangible cultural heritage. *LIBER Quarterly*, 30(1), 1–16. <https://doi.org/10.18352/lq.10315>

Designing for Conversational Coherence: A Cross-Modal Analysis of Semantic Drift in Human-AI Interactions

Aadya Bharti¹[0009-0009-2104-7311], Rishu²[0009-0005-7642-2829] Abhishek Sharma³[0009-0009-4070-4129], Anishka Gupta⁴[0009-0006-3024-098X]

¹⁻⁴ National Institute of Design, Kurukshetra, Haryana 136131
aadya.b@nidh.ac.in

Abstract. Large Language Models (LLMs) are increasingly embedded in information access systems, learning environments, and collaborative knowledge work. Yet, their generative flexibility also gives rise to semantic drift, whereby conversational trajectories progressively deviate from the original communicative intent. This study presents a comparative analysis of conversational drift in human-only, human-AI voice-based, and AI text-based interactions across both discussion-oriented and learning-centered settings. Drawing on transcript annotation, semantic-embedding similarity metrics, and qualitative coding, we find that human-only discussions exhibit the greatest degree of drift, largely driven by associative reasoning processes and socio-emotional dynamics. Human-AI interactions display a moderate level of drift, while structured, prompt-engineered exchanges remain tightly constrained with minimal topical divergence. These results demonstrate that conversational coherence is strongly mediated by interaction modality and the structuring of prompts. In response, we introduce FOCUS (Framework for Organized Conversational Understanding with Safeguards), a design framework that integrates topic-anchoring embedding's, a dedicated prompt-engineering agent, relevance-gating mechanisms, and continuous drift-monitoring loops to sustain topic fidelity while allowing for productive exploratory dialogue. This framework supports the development of reliable, transparent, and pedagogically aligned conversational AI systems and advances emerging priorities in library and information science (LIS) related to AI-enabled information access, user experience design, and responsible knowledge mediation.

Keywords: **Keywrds:** Semantic Drift; Large Language Models; Human-AI Interaction; Systems Design.

1 Objective of the Study

The objective of this study is to systematically investigate how semantic drift arises, develops, and varies across human-only, human-AI, and prompt-engineered conversational settings in both discussion and learning environments. Specifically, the study aims to

- Quantify the extent and types of conversational deviation in each condition
- Identify modality-based differences between voice and text interactions with LLMs
- Analyze the cognitive and structural factors that contribute to drift in instructional and dialogic contexts

By combining computational measures of semantic similarity with qualitative discourse analysis, the study seeks to provide an empirically grounded account of how AI participation affects conversational coherence. Ultimately, the research aims to derive design implications for developing conversational AI systems that are more reliable, topic-anchored, and aligned with pedagogical goals.

2 Methodology

This study adopts a comparative experimental design to analyze semantic drift as a function of conversational modality and AI integration, aligned with the broader theme of designing for conversational coherence. Two conversational environments-Discussion and Learning-are each examined under three interaction conditions: (1) human-only baseline, (2) human-AI hybrid via voice interaction, and (3) AI-mediated text interaction supported by structured prompt engineering. This design enables systematic comparison of drift patterns within and across modalities, isolating the effects of AI participation, interaction mode, and prompt structure on conversational coherence. Participants were undergraduate design students with varied exposure to AI tools, ensuring diversity in reasoning styles and conversational behavior.

2.1 Discussion Environment

The discussion environment investigates how coherence is maintained or eroded during open, multi-perspective dialogue on complex social or technical issues. Three conditions are implemented:

- **Human-Only Discussion:** Groups of three participants engage in a 15-20 minute debate on a predefined controversial topic (e.g., climate policy, algorithmic bias, or online censorship). A moderator passively records the interaction and documents deviations, interruptions, and topic shifts without intervening. This condition serves as a baseline for naturally occurring drift in human conversation.
- **Human + Voice-AI Discussion:** Three human participants hold a live debate while concurrently interacting with AI agents (e.g., GPT, Claude, and Gemini) that respond via voice. The AIs act as parallel discussants, contributing in real time. This condition examines how voice-based AI participation influences drift dynamics, including whether AI contributions amplify, constrain, or redirect emerging topic shifts.
- **Text-Based AI Discussion:** A single human participant debates the same topic with an LLM through text-based interaction. The conversation follows a standardized, prompt-engineered script designed to narrow scope, reinforce the central theme, and support coherence-preserving dialogue moves (e.g., refocusing questions, summarization prompts). This condition assesses the degree to which structured prompting can mitigate semantic drift relative to spontaneous, multi-party voice interactions.

2.2 Learning Environment

The learning environment focuses on semantic drift in instructional and inquiry-driven exchanges, where coherence is closely tied to pedagogical alignment:

- **Human-Human Teaching:** One participant assumes the role of instructor and explains a specified topic (e.g., principles of renewable energy, fundamentals of probability) to peers, who ask clarifying questions. Drift is defined as movement away from core learning objectives, including unnecessary digressions, over-extended tangents, or abandonment of key concepts.
- **Human Learning from Voice AI:** Individual learners engage in natural, voice-based dialogue with AI tutors (e.g., GPT, Claude, and Gemini). The conversational trajectory is examined for AI-induced digressions, learner-initiated off-topic queries, and repair moves (if any) that re-establish coherence. This condition highlights how voice-based AI instruction shapes the balance between exploration and topical focus.
- **AI-Text Teaching with Prompt Engineering:** Learners interact textually with LLMs using structured, pedagogy-informed prompts (e.g., Socratic questioning sequences, stepwise scaffolding, explicit recap prompts). This condition evaluates how prompt design can function as an architectural constraint on semantic drift, promoting stable topic anchoring while still enabling elaboration and deeper understanding.

2.3 Data Collection

For each session, the following data sources are collected to support both quantitative and qualitative analysis of conversational coherence:

- Complete audio or text transcripts of all interactions
- Moderator observation logs noting:
 - Topic shifts
 - Tangential insertions
 - Repetitions and conversational loops
 - Contradictory or incoherent statements
 - Computational markers of semantic drift, calculated using Sentence-BERT embedding's to quantify deviation from an initial topic-seed embedding over time
 - Participant reflections from post-session questionnaires assessing perceived focus, coherence, and clarity of the interaction

2.4 Analysis Framework

A mixed-methods analytical framework is employed to link computational measures of drift with discourse-level explanations:

2.5 Semantic Drift Quantification:

Semantic drift is operationalized as the cosine distance between each conversational turn and the topic-seed embedding. A longitudinal Drift Score is computed across turns to trace moment-to-moment divergence and recovery, enabling comparison of coherence trajectories across conditions.

2.6 Comparative Drift Analysis:

Drift rates and patterns are compared across:

- Human-only vs. human-AI vs. AI-text conditions
- Discussion vs. learning environments
- Voice vs. text modalities

Drift is categorized into types-such as scope broadening, domain shifting, causal detours, and emotionally driven divergence-to characterize how conversations move away from their intended focus and how these movements differ by modality and AI involvement.

2.7 Qualitative Discourse Coding:

Transcripts are coded for thematic domains (e.g., policy, technical detail, anecdotal evidence, ethics, personal experience) to contextualize quantitative drift measures. Coding is used to identify triggers of semantic drift (e.g., examples, questions, emotional reactions, AI-generated suggestions) and to surface coherence-preserving strategies (e.g., summarizing, reframing, redirecting) that may inform the design of conversational safeguards.

2.8 Ethical Considerations

The study adheres to established ethical standards for human-AI interaction research. All participants provide informed consent and are explicitly informed when AI systems are involved in the conversation. Transcripts are anonymized before analysis, and all data are stored and managed in accordance with institutional confidentiality and data protection guidelines. Particular care is taken to communicate the experimental nature of the AI interactions and the focus on conversational coherence, ensuring transparency and minimizing potential harm or misunderstanding.

3 Literature Review

Recent progress in Large Language Models (LLMs) has changed how people search for information, learn new topics, and hold conversations. Surveys of LLMs show how these models have grown from basic statistical tools into powerful transformer-based systems that can reason, plan, and respond in context [1][3]. These abilities make conversations richer, but they also increase the chance that the model may slowly drift away from the main topic. Even the GPT-4 report notes that although the model is more accurate and aligned, it can still produce off-topic or hallucinated responses that lead to drift over longer exchanges [2].

When speech is added, conversations become even more fluid. Work on AudioPaLM shows that combining text and voice creates more natural interactions, but also makes it harder for systems to stay focused because voice responses are faster, spontaneous, and less edited [4]. This insight is important for our comparison between voice-based interactions and structured text-based prompts.

Semantic drift is not unique to AI; it also occurs in normal human conversations. Liber man’s study shows how people naturally shift topics as they negotiate meaning, follow associations, and respond emotionally or socially [6]. This supports our human-only baseline, where drift is seen as a normal part of dialogue.

Research on knowledge systems and machine learning provides tools to measure and correct drift. Ontology studies describe different types of semantic change, while Yu et al. show that drift can be detected and compensated for using embedding’s, similar to how we measure conversation drift in our study [5].

Studies in conversational AI also highlight that maintaining coherence is still a major challenge. Work from the Alexa Prize shows that even advanced systems struggle to stay on topic in open-ended dialogue [7]. Design research further stresses that human-AI interactions are sociotechnical systems where coherence, alignment, and user expectations must all be managed together [8].

However, existing work does not compare drift patterns across human-only, human-AI voice, and prompt-engineered text interactions, nor does it evaluate both learning and discussion settings. Our study fills this gap by examining how drift changes across modalities and contexts, and what this means for designing more stable, focused, and user-aligned conversational systems.

4 Discussion

The comparative analysis of conversational behavior across human-only, human-AI, and structured human-AI interactions reveals clear differences in how semantic drift emerges, develops, and is constrained. Overall, the findings show that drift is not just a linguistic phenomenon, but the result of cognitive habits, social dynamics, and modality-specific features that shape how topics shift over time.

In human-human conversations, semantic drift appeared as a natural and recurrent pattern. Participants frequently moved into adjacent or loosely related domains, driven by associative reasoning, personal relevance, and narrative expansion. For example, a discussion on the technical differences between petrol and diesel engines [9] expanded into geopolitical fuel policies, personal stories about vehicle ownership, and new automotive technologies. In a learning-focused session on forensic identification [9], curiosity led to extended detours into molecular biology and genetics, at times overshadowing the core forensic content. These patterns are consistent with prior discourse research that characterizes human conversation as dynamic, fluid, and socially co-constructed.

By contrast, human-AI conversations showed greater topical stability. The AI often functioned as an anchor, re-centering the dialogue on the main theme. In learning-oriented interactions, the AI’s instructional tendencies supported sequential structure and coherence. In the contextual inquiry session [9], for instance, the conversation remained closely aligned with methodological principles, and elaborations reinforced rather than displaced the learning objective. Even in discussion-oriented dialogues on topics such as exam systems or AI’s impact on creativity, the AI preserved thematic alignment, engaging in conceptual expansion without drifting into irrelevant territory.

The structured-prompt human-AI condition imposed the strongest constraint on drift. Here, the AI followed predefined subtopics, producing focused, domain-specific responses with minimal deviation. Modality also mattered: voice-based exchanges allowed slightly more tangential movement than text, but AI-driven corrections helped restore focus. Across conditions, drift served different roles, enriching social connection in human-only talk, and enabling controlled conceptual broadening in human-AI dialogue, highlighting how AI participation actively shapes conversational coherence.

5 Results

The study revealed a consistent empirical gradient of semantic drift across conditions: human-only interactions showed the highest drift, followed by human-AI dialogues, with structured human-AI (prompt-engineered) interactions exhibiting the least drift. This pattern held across both discussion and learning environments.

In human-human conversations, drift was frequent and wide-ranging. Topics shifted rapidly and in multiple directions, often triggered by personal anecdotes, socio-political connections, emotional reactions, spontaneous analogies, and chained contributions from multiple speakers. In the petrol-diesel discussion, more than half of the transcript consisted of tangential content. In the forensic identification session, about 40% of turns moved into domains outside forensic science [9]. On average, noticeable drift occurred every 4-6 turns, frequently resulting in extended digressions before the conversation returned to the main topic.

In human-AI interactions, drift was substantially reduced and conversations displayed higher topical discipline. The AI often helped to restore focus, reframe questions to keep within scope, and avoid personal or multi-domain tangents unless explicitly invited. Learning-oriented sessions were particularly coherent: in the contextual inquiry conversation, fewer than 5% of turns diverged, and those that did remained closely related to the core methodological theme [9]. Discussion-oriented sessions (e.g., on exam systems or creative labor) showed some drift, but these deviations generally served to deepen or extend the main argument. Drift intervals increased to roughly every 12-15 turns.

The structured prompt human-AI condition exhibited minimal to near-zero drift. The AI followed the predefined subtopics and sequence, produced no domain-irrelevant expansions, and maintained high semantic alignment with the topic throughout. No anecdotal or socially driven drift was observed.

INTERACTION TYPE	DRIFT LEVEL	WHAT CAUSED THE DRIFT	MODALITY SPECIFIC FEATURES	FREQUENCY OF TOPIC SHIFTS	HOW RECOVERY IS MADE	WHAT DOES IT MEAN
HUMAN - HUMAN DISCUSSIONS	8-9	Personal opinions, emotional reactions, social dynamics, shared memories, narrative expansion, associative thinking.	Natural speech patterns, spontaneous interruptions, overlapping dialogue, emotional pauses, tangential humor, multi-speaker contributions.	Every 4-6 turns, approximately 40% of conversations venture into tangential content.	organic return through shared emotions, deliberate moderator intervention, or natural fatigue of the topic of diversion.	Natural, socially-driven conversations meander organically—this is how humans naturally communicate.
HUMAN - AI VOICE INTERACTION	4-6	Learner initiated off-topic queries, exploratory questions, multi-domain connections, AI suggestion of related concepts.	Concurrent parallel discussion, AI voice creates tonal stability, natural speech behavior, real-time responsiveness.	Drift occurs through every 12-15 turns, controlled elaboration around core topic, fewer than 5% diversion.	AI re-centers dialogue through subtle reprimand, acknowledgment of tangent with redirection, or explicit boundary setting questions.	AI voice-based systems help maintain topical coherence while still allowing exploratory conversation.
HUMAN - AI TEXT (structured prompts)	0.5-2	User attempts to broaden scope beyond prompt boundaries, AI strictly adheres to predefined subtopics and sequence.	Predetermined topic architecture, keyword constraints, sequential prompt scaffolding, minimal emotional language.	Near zero drifts, no drift intervals typically beyond 20+ turns.	system regenerates on-topic responses and flags off-topic requests.	Strict prompting constraints ensure high coherence but may reduce natural engagement.

Table 1. Summary of results table

6 Proposed Framework

This study proposes FOCUS (Framework for Organized Conversational Understanding with Safeguards) as a design solution for reducing semantic drift in AI-mediated conversations while still allowing useful exploration and creativity. FOCUS is conceived as a layered, feedback-based system that (1) anchors the topic, (2) defines semantic boundaries, (3) monitors drift in real time, and (4) keeps humans in control of how strictly coherence is enforced. Figure 2 shows an overview of the system map.



Table 2. Overview of the FOCUS conversational control workflow

In library and information science (LIS) contexts, FOCUS can be applied to AI-assisted reference services, information literacy instruction, research consultations, and digital learning platforms, where maintaining topical focus, transparency, and user trust is critical to effective knowledge mediation and service delivery.

FOCUS starts with the user’s natural, unstructured input. Users are not required to write precise or technical prompts. Instead, a Prompt Engineering Agent interprets the query and formalizes it. This agent extracts a Core Topic Statement (CTS), identifies key subtopics, and generates a structured prompt that serves as the “semantic blueprint” for the conversation. In this way, vague or broad inputs are translated into a clear, machine-readable scope; any missing information is prompted to enter.

Once the CTS are established, a Boundary Construction process converts the topic and its key terms into a semantic boundary. This boundary represents the allowed topic space for the conversation. It includes both what is relevant and what should be avoided, enabling the system to distinguish between useful elaboration and unintended drift. This approach reflects familiar library practices such as scope definition, subject classification, and controlled vocabularies.

During the interaction, the Conversation Engine generates candidate responses that pass through a Relevance Gate. The gate computes similarity between the candidate response and the boundary embedding. If the response falls outside a predefined threshold, the system intervenes by either regenerating a more on-topic reply, flagging

the potential drift and asking the user whether to continue or refocus, or routing the off-topic content to a Peripheral Archive. The Peripheral Archive stores creative digressions separately so they are preserved without derailing the main thread.

FOCUS also includes a Recap and Revalidation Loop. At intervals, the system summarizes the conversation, reports a simple coherence indicator, and invites the user to tighten, relax, or revise the topic scope. Users can see and optionally edit the engineered prompt and boundary settings, supporting transparency, explain ability, and low-friction control over conversational coherence.

In library and information science (LIS) contexts, FOCUS can be applied to AI-assisted reference services, research consultations, information literacy instruction, and digital learning or discovery platforms, where maintaining topical focus, transparency, and user trust is essential for effective knowledge mediation. The framework is designed for seamless integration into existing library workflows, allowing gradual adoption without disrupting established practices. Its modular structure enables libraries to implement selected components based on institutional needs and capacity. Importantly, librarians remain central to the process, using FOCUS as a supportive tool that enhances professional judgment rather than replacing it, thereby reinforcing the library's role as a trusted and structured information environment in AI-supported services.

FOCUS offers a practical and human-centered approach to managing conversational coherence. It supports focused learning and discussion while respecting the exploratory nature of information seeking, reinforcing the role of libraries as trusted, well-structured knowledge environments in an AI-enabled landscape.

7 Limitations

This study has several limitations. First, the sample size is modest and the range of conversational contexts is relatively narrow, which may limit the extent to which the findings represent broader patterns of human communication. Second, the measurement of drift is based on embedding-derived similarity scores and qualitative coding; both approaches can simplify or miss nuanced, contextual, and pragmatic forms of divergence. Third, the observed AI behaviors are specific to the particular models, versions, and voice pipelines used at the time of the study, making the results sensitive to future system updates. Finally, human factors such as prior knowledge, emotional state, and conversational goals were not systematically controlled and may have influenced drift patterns. Taken together, these constraints position the study as exploratory and suggest that the results should be interpreted with caution rather than as fully generalizable.

8 Conclusion

This study shows that semantic drift varies systematically across human-only, human-AI, and structured prompt-based interactions, forming a clear gradient of conversational coherence. Human-only dialogues display the greatest divergence, shaped by associative thinking, personal relevance, and social dynamics. Human-AI interactions exhibit greater topical stability: the model's relevance-oriented generation helps maintain focus while still allowing controlled elaboration around the core theme. Structured prompt interactions demonstrate the highest coherence, with explicit scope-setting and predefined subtopics effectively constraining off-topic movement and producing minimal measurable drift. These results indicate that conversational coherence is not an inherent property of AI systems alone, but a function of interaction design, modality, and prompt structure. By clarifying how different configurations influence drift trajectories, the study provides an empirical basis for designing AI-mediated communication systems that better balance creative exploration with sustained topical fidelity, particularly in educational and information-seeking settings.

Acknowledgements

We thank the members of the Design Technology Club - Lakshita Yadav, Om Singh, Mayank Kumar, Harit Saraswat, Prakriti Saxena, Shruti Kushwaha, Siya Garg, Abhishek Sharma, Rishu, Anishka Gupta, Himani Gupta and Aadya Bharti, for their participation and contributions to the study.

Declaration of AI and AI-Assisted Technologies used in Manuscript Writing Process:

ChatGPT (GPT-5.2), Google Gemini 3, Claude, and Perplexity Pro were used both as research tools and writing-support aids. These models were employed to conduct AI-mediated conversational sessions for the study and to

assist with language refinement, including grammar, clarity, coherence, and academic tone. All study design, analysis, and conclusions were determined by the authors.

References

1. Naveed, H., et. al (2025). A comprehensive overview of large language models. *ACM Transactions on Intelligent Systems and Technology*, 16(5), Article 106. <https://doi.org/10.1145/3744746>
2. OpenAI. (2023). *GPT-4 technical report*. arXiv. <https://doi.org/10.48550/arXiv.2303.08774>
3. Zhao, W. X., Zhou, K. et. al (2023). A survey of large language models. arXiv. <https://doi.org/10.48550/arXiv.2303.18223>
4. Rubenstein, P. K., Asawaroengchai, et. al. (2023). AudioPaLM: A large language model that can speak and listen. arXiv. <https://doi.org/10.48550/arXiv.2306.12925>
5. Yu, L., Twardowski, et. al (2020). Semantic drift compensation for class-incremental learning. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 6982-6991. <https://doi.org/10.1109/CVPR42600.2020.00700>
6. Liberman, K. (2012). Semantic drift in conversations. *Human Studies*, 35(2), 263-277. <https://doi.org/10.1007/s10746-012-9225-1>
7. Ram, A., et. al (2017). Conversational AI: The science behind the Alexa Prize. 1st Proceedings of Alexa Prize (Alexa Prize 2017). Amazon.
8. Abbasi, A., et. al (2024). Pathways for design research on artificial intelligence. *Information Systems Research*, 35(2), 441-459. <https://doi.org/10.1287/isre.2024.editorial.v35.n2>

From Content to Concepts: A Semantic Framework for Enhancing Classroom Learning

Chitra Ghosh¹ [0009-0004-2881-0846], Shubham Mishra² [0009-0002-7599-8111],
Ram Vishwas Chourasiya³ [0009-0007-2561-5334]

¹Department of Library and Information Science, University of Delhi, Delhi

²Department of Library and Information Science, Panjab University, Chandigarh

³Pillai College of Engineering, New Panvel (Autonomous), Navi Mumbai
{chitra, shubham, ramvishwas}@drtc.isibang.ac.in

Abstract. The rapid expansion of digital and hybrid learning environments has intensified the need for systematic knowledge organization that supports efficient information access and meaningful comprehension. Despite the abundance of instructional resources, classroom learning materials often remain weakly structured at the semantic level, limiting learners' ability to recognize conceptual relationships, go through knowledge spaces effectively, and engage in deeper learning processes. This paper proposes a conceptual framework for enhancing classroom learning through semantic technologies, emphasizing the role of structured knowledge representations in enriching teaching and learning practices. The framework integrates semantic components such as ontologies, controlled vocabularies, and knowledge graphs to organize instructional content, establish explicit conceptual linkages, and enable contextual retrieval of learning resources. By semantically mapping curriculum elements and classroom materials, the model supports improved content discovery and cognitive connectivity for learners. Drawing on interdisciplinary literature from Library and Information Science, education, and semantic web research, the study demonstrates how semantic enrichment can transform conventional learning materials into interconnected knowledge spaces that facilitate conceptual clarity, adaptive learning pathways, and enhanced information retrieval. The paper highlights the strategic role of LIS professionals in designing and implementing semantic layers within educational ecosystems and contributes to reimagining knowledge organization practices for more interactive, inclusive, and intellectually coherent classroom learning environments.

Keywords: Semantic Technologies, Knowledge Organization, Ontologies, Knowledge Graphs, Classroom Learning, Instructional Design, Library and Information Science, Pedagogy

1 Introduction

The rapid transformation of educational environments, driven by technological advancements and evolving learner expectations, has fundamentally reshaped how knowledge is created, accessed, and consumed. While classrooms, today, are increasingly supported by digital tools and content-rich resources, instructional materials often remain fragmented, loosely structured, and semantically shallow. This disconnects limits learners' ability to perceive conceptual relationships, transfer knowledge across contexts, and engage in deeper cognitive processes, essential for meaningful learning. As educational systems expand in scale, complexity, and interdisciplinary, the need for systematic knowledge organization within classroom learning has become increasingly critical.

Semantic technologies such as ontologies, controlled vocabularies, semantic annotations, and knowledge graphs, offer powerful mechanisms for structuring information in ways that reflect conceptual meaning rather than mere content aggregation. Although these technologies have been extensively explored in digital libraries, information retrieval, and online learning platforms, their potential to reorganize classroom centric instructional content and curricula remains underutilized. Existing educational implementations largely emphasize post-learning analytics or platform-based delivery, with limited attention to semantic structuring at the instructional planning and curriculum design stages.

This paper argues that effective classroom learning requires more than access to information; it demands semantically coherent knowledge environments that explicitly represent conceptual relationships, learning intent, and pedagogical alignment. By integrating principles of knowledge organization with instructional design, semantic enrichment can transform traditional teaching materials into interconnected learning ecosystems that support conceptual continuity, interdisciplinary integration, and adaptive learning pathways.

This model proposes a Semantic-Enhanced Classroom Learning Framework that re-conceptualizes classroom learning as a layered semantic system. The framework integrates intra and inter-subject semantic linking, intent driven content structuring, and pedagogically aligned learning elements through formal semantic models. Central

to this approach is the collaborative role of Library and Information Science (LIS) professionals and educators, combining expertise in knowledge organization and pedagogy to ensure both semantic rigor and instructional relevance.

1.1 Problem Statement

Classroom learning increasingly relies on a growing volume of digital instructional materials viz. lecture slides, textbooks, videos, research articles, learning objects, and online resources. However, these materials are often organized in unstructured or loosely structured formats that limit meaningful discovery, contextual understanding, and conceptual linkage. Students struggle to connect related concepts across courses, identify prerequisite knowledge, or retrieve resources aligned with their learning needs. Educators, too, lack systematic tools to present content in a semantically enriched manner that supports deeper comprehension and personalized learning pathways.

While semantic technologies such as ontologies, controlled vocabularies, and knowledge graphs offer powerful mechanisms for structuring knowledge and connecting ideas, their application within classroom environments remains underexplored. The absence of a coherent semantic framework leads to fragmented learning experiences, reduced discoverability of educational content, and limited support for the learners-centered pedagogy.

There is a need to develop a conceptual framework that leverages semantic structures to enhance classroom learning by improving content organization, facilitating concept level navigation, and enabling context aware information access. Such a framework can bridge the gap between educational content and learner needs.

1.2 Objectives of the Study

- To conceptualize a semantic framework for organizing and enhancing classroom learning materials using structured knowledge representations such as ontologies controlled vocabularies and knowledge graphs.
- To identify gaps in current classroom content organization practices that limit concept-level understanding, contextual linking and effective resource discovery for learners.
- To analyze how semantic technologies can improve information access, personalization, and cognitive navigation within teaching learning environments.
- To explore the role of LIS professionals in designing and implementing semantic structures that support academic instruction and enrich digital learning ecosystems.
- To propose a model that integrates semantic annotation, conceptual mapping, and contextual retrieval to facilitate deeper learning and improve pedagogical effectiveness.
- To highlight potential applications, benefits, and challenges of adopting semantic frameworks in classroom settings, providing insights for future research and practice.

1.3 Scope of the Study

This study focuses on developing a conceptual framework that integrates semantic technologies to enhance classroom learning experiences. The scope includes:

- Semantic Technologies: Exploration of ontologies, controlled vocabularies, taxonomies, and knowledge graphs as tools for structuring classroom content.
- Classroom Learning Materials: Consideration of typical instructional resources such as lecture notes, reading materials, digital learning objects, course modules and supplementary references.
- Teaching Learning Context: Focus on academic settings where structured knowledge representation can support both instructors and learners.
- Role of LIS Professionals: Examination of how LIS expertise in metadata, knowledge organization, and information retrieval contributes to educational environments.
- Framework Development: Formulation of a conceptual, dynamic model that shows how the semantic annotation, conceptual linking, and contextual retrieval can enrich learning.

The study does not involve full system development, implementation, or quantitative evaluation. Instead, it offers a theoretical and practice-oriented model supported by existing research.

2 Literature Review

The rapid expansion of digital classroom learning environments has increased the demand for structured approaches to organizing educational content. Semantic technologies such as ontologies, semantic annotations, and knowledge graphs have been widely explored as mechanisms for transforming unstructured learning resources into machine-interpretable and conceptually connected knowledge structures [5]. Prior studies demonstrate that semantic enrichment enhances content discoverability, contextual navigation, and learner comprehension by explicitly modeling conceptual relationships and dependencies [13]. However, much of this research remains focused on online platforms, learning management systems, and MOOCs, with comparatively limited attention to classroom-centric semantic organization where curriculum structure and instructional sequencing play a dominant role [17].

Ontologies have emerged as foundational instruments for semantic modeling in educational systems, offering formal representations of domain concepts, hierarchies, and relationships [1]. Ontology-based course models have been shown to support the alignment of learning objectives, instructional content, and assessment components, thereby improving conceptual coherence [22]. In parallel, ontology-driven learning analytics frameworks have been proposed to semantically interpret learner interactions and assessment outcomes [17]. While these approaches advance post-instructional analysis, they largely overlook the role of semantic modeling in pre-instructional design, such as curriculum mapping, concept dependency identification, and instructional planning [20][21].

The semantic annotation has been identified as a key mechanism for linking educational resources to conceptual models by enriching learning materials with metadata tied to ontology concepts [4]. When combined with knowledge graphs, annotated content can be organized into relational structures that represent conceptual dependencies, prerequisite pathways, and cross-domain linkages. Studies report that such relational navigation supports cognitive linking and personalized learning experiences [3]. Nevertheless, existing implementations are predominantly technology-driven and seldom incorporate pedagogical intent or differentiated instructional purposes into semantic models [14][15].

The growing emphasis on multidisciplinary and interdisciplinary education further exposes limitations in existing curricular designs, where conceptually related topics across subjects are often taught independently without explicit coordination. Semantic technologies provide mechanisms for representing shared principles and cross-domain relationships through inter-subject mappings [9].

Within this context, the role of Library and Information Science (LIS) professionals has been increasingly acknowledged due to their expertise in metadata standards, controlled vocabularies, and knowledge organization systems [6]. However, their contribution is often confined to technical support roles rather than integrated pedagogical collaboration. Recent studies suggest that involving LIS professionals in semantic design processes can enhance interoperability and conceptual consistency across educational systems.

Existing literature demonstrates substantial progress in applying semantic technologies to education but reveals persistent gaps related to classroom-centric semantic structuring, integration of learning intent, inter-subject conceptual alignment, and the systematic involvement of LIS expertise. Addressing these gaps requires a pedagogically grounded semantic framework that integrates curriculum design, instructional planning, and knowledge organization [11].

2.1 Gaps Identified in the Literature

The review reveals several gaps:

- Limited conceptual frameworks that integrate semantic technologies directly into classroom teaching & learning processes.
- Scarce research connecting LIS knowledge organization practices with pedagogy and instructional design.
- Emphasis on technical implementations while neglecting practical, educator-friendly models.
- Insufficient exploration of collaborative roles between LIS professionals and educators in semantic structuring of learning materials.

These gaps justify the need for a framework that bridges semantics, pedagogy, and LIS expertise, precisely the focus of the present study.

3 Methodologies

This study adopts a conceptual research design aimed at developing a framework for enhancing classroom learning through semantic technologies. As the objective is to propose a theoretical model rather than conduct empirical testing, the methodology focuses on a systematic analysis, synthesis, and conceptual mapping.

3.1 Research Design

A qualitative, conceptual design was employed to explore how semantic tools and knowledge organization principles can be integrated into classroom learning environments. This approach allowed the study to analyze theories, models, and practices from LIS, educational technology, semantics, and instructional design.

3.2 Methodological Steps

The methodological process adopted in this study was structured into five interrelated stages to ensure theoretical rigor and conceptual clarity.

The first stage involved an extensive exploration of the existing literature across multiple intersecting domains, including semantic technologies such as ontologies, knowledge graphs, and linked data. This was further associated with knowledge organization practices within Library and Information Science (LIS), educational ontologies and learning analytics, and user-centered learning and pedagogical approaches. Through a systematic synthesis of these works, several recurring themes emerged, notably semantic enrichment of learning resources, concept mapping, contextual and concept-based retrieval, and the evolving instructional role of LIS professionals. These themes provided the intellectual foundation for the development of the proposed conceptual model.

In the second stage, insights derived from the literature were critically examined to identify gaps and unmet conceptual needs within current classroom learning environments. This analysis revealed persistent challenges, including inadequacies in the organization of instructional content, limited accessibility at the concept level, weak semantic alignment across instructional materials, and the absence of integrative frameworks, that meaningfully connect LIS expertise with classroom pedagogy. These deficiencies were instrumental in defining the functional and conceptual requirements that the proposed framework seeks to address.

The third stage focused on mapping semantic elements to core learning processes. Key semantic components namely ontologies, controlled vocabularies, semantic annotations, and knowledge graphs were systematically aligned with essential learning activities such as instructional content design, educational resource organization, student navigation pathways, and contextual discovery of learning materials. This mapping exercise established the foundational relationships between semantic technologies and pedagogical processes, thereby operationalizing the role of semantics within the learning environment.

Building upon the thematic synthesis and gap analysis, the fourth stage entailed the formulation of the conceptual framework. The resulting model is structured around three interdependent layers: the content layer, encompassing instructional materials, curriculum components, and learning resources; the semantic layer, comprising ontologies, concept hierarchies, semantic relationships, and annotations; and the learning interaction layer, which includes user navigation pathways, context-aware retrieval mechanisms, and personalized access to resources.

The final stage involved conceptual validation of the proposed framework through logical analysis. This validation process included checks for internal coherence, assessment of logical consistency among the framework's constructs, and evaluation of alignment with the established principles in LIS, semantic technologies and pedagogy. Through this process, the framework was affirmed to be theoretically robust and contextually relevant for application in contemporary learning environments.

4 Semantic-Enhanced Classroom Learning Frameworks (SECLF)

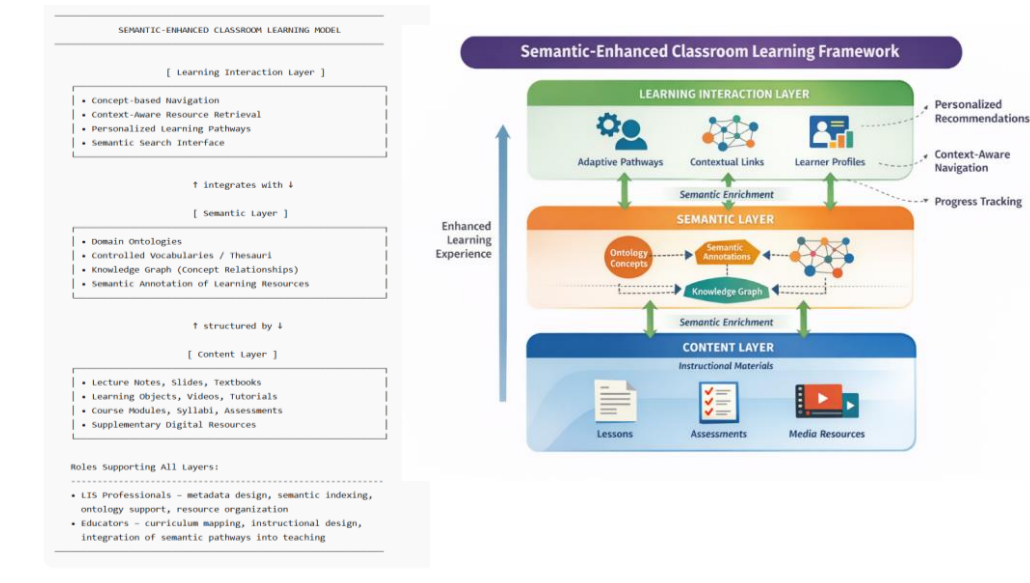


Fig. 1. Diagrammatic Representation of SECLF

The proposed framework integrates semantic technologies with instructional content and learner interaction to enhance classroom learning experiences. It is structured across three interconnected layers—Content Layer, Semantic Layer, and Learning Interaction Layer—supported by the active contribution of LIS professionals and educators. Together, these layers create a unified environment where concepts, resources, and learning pathways are semantically enriched to improve understanding, retrieval, and engagement.

The relevance of this layered semantic approach can be illustrated through existing educational semantic models such as the Hyper Physics project developed at Georgia State University. Hyper Physics demonstrates how subject knowledge, particularly in physics, can be organized as an interconnected network of concepts rather than as isolated topics, using concept maps to visualize hierarchical and associative relationships among principles, laws, and applications. By enabling learners to navigate concepts semantically across related themes, prerequisites, and applications, Hyper Physics exemplifies how semantic structuring deepens conceptual understanding and supports exploratory learning. In a similar manner, the proposed SECLF extends this idea beyond a single subject by systematically integrating controlled vocabularies, ontologies, and knowledge graphs at the curriculum level, thereby enabling semantic coherence across subjects, instructional content, and learning interactions.

4.1 Semantic-Enhanced Classroom Learning Framework (SECLF) Diagram

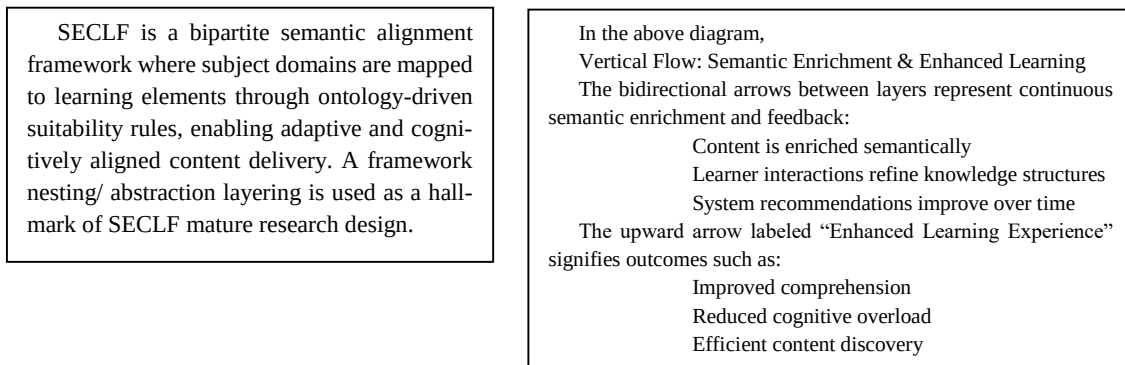


Fig. 2. Textual Representation of SECLF

4.2 Explanation of the Framework

The diagram presents a three-layered Semantic-Enhanced Classroom Learning Framework designed to improve teaching-learning processes through semantic technologies. The framework shows how raw educational content is semantically enriched to support personalized, context-aware, and adaptive learning experiences.

Content Layer (Base Layer). The Content Layer represents the foundational educational resources used in classroom and digital learning environments. This layer acts as the input source for semantic processing. It includes all instructional materials used in the classroom such as Lessons (textual learning materials, lecture notes, textbooks, course modules); Assessments (quizzes, assignments, tests); Media Resources (videos, presentations, tutorials, digital learning objects, simulations). At this stage, the content exists largely in an unstructured or semi-structured form, making it difficult for systems to understand relationships between concepts or learner needs. Therefore, this layer alone cannot support personalization or intelligent retrieval.

The purpose of this layer is to identify, collect, and prepare classroom materials that will be semantically enriched. LIS professionals can contribute to metadata creation, structuring, and organizing diverse content formats.

Semantic Layer (Core Intelligence Layer). This layer introduces structured semantic elements that transform raw educational content into a meaningful knowledge network. The Semantic Layer is the core of the framework and introduces intelligence into the learning environment through semantic enrichment. It consists of the following components: Semantic Annotations, Controlled Vocabularies and Thesauri, Ontology Concepts, and Knowledge Graph.

Role of the Semantic Layer. Collectively, these components transform unstructured or loosely structured instructional content into structured, connected, and intelligent learning knowledge. The Semantic Layer serves as the bridge between raw educational resources and adaptive, learner-centered instructional systems, enabling meaningful knowledge organization, contextual access, and enhanced cognitive engagement. The Semantic Layer comprises two interrelated semantic dimensions of learning:

- i. Semantic Structuring of Subject Knowledge
- ii. Semantic Alignment of Learning Elements and Content Representation.

i. Semantic Structuring of Subject Knowledge. The first methodological component focuses on how subject knowledge is organized and sequenced to promote semantic understanding and cognitive coherence. Through intra- & inter-subject semantic linking, the framework supports a more coherent curriculum structure in which learning progresses logically across interconnected concepts, thereby strengthening conceptual clarity and fostering sustained cognitive development within and across the discipline.

a) Intra-Subject Semantic Linking (Conceptual Coherence)

Intra-subject semantic linking refers to the explicit identification and representation of conceptual relationships within a single subject domain. Traditional curricula often present topics in isolation, assuming learners will implicitly infer conceptual dependencies across modules and semesters. Such fragmentation can hinder conceptual continuity and increase cognitive load. This framework addresses this limitation by:

- Identifying core concepts, sub-concepts, and prerequisite relationships within a subject
- Representing these relationships using ontologies and concept hierarchies
- Enabling learners to navigate between semantically related concepts based on conceptual proximity rather than chronological placement

A practical example can be observed in engineering curricula, where conceptually similar topics are frequently distributed across different semesters under separate course titles. For instance, foundational concepts of Linear Algebra introduced in a first-year engineering mathematics course are often revisited in later semesters within subjects such as Control Systems, Signal Processing, or Machine Learning, without explicit semantic linkage. By semantically mapping these overlapping concepts—such as matrices, eigenvalues, and vector spaces—within a unified subject ontology, the framework enables these related topics to be taught concurrently or in a closely coordinated manner within the same semester. This alignment promotes deeper conceptual understanding, reduces redundancy, and allows students to immediately apply theoretical foundations to advanced engineering applications.

b) Inter-Subject Semantic Linking (Multidisciplinary Integration)

Inter-subject semantic linking captures conceptual overlaps and dependencies across different subjects, supporting multidisciplinary and integrative learning. Many academic programs teach related concepts across different papers or semesters without explicit coordination, resulting in fragmented understanding. The methodology proposes:

- Identifying shared concepts, laws, and principles across subjects
- Mapping these shared concepts semantically, even when they appear under different terminologies
- Coordinating the timing and instructional emphasis of conceptually identical or related topics across semesters or courses

A suitable real-life example can be observed in an engineering curriculum involving Probability and Statistics and Quality Control / Industrial Engineering. Concepts such as probability distributions mean and variance and normal distribution are typically taught in a statistics course during an early semester. The same concepts reappear later in a different subject when students learn Statistical Process Control (SPC), control charts, and process capability analysis.

Although these topics rely on the same underlying statistical principles, they are often taught months or years apart, and the connection between them is left implicit. Through inter-subject semantic linking, the statistical foundations and their engineering applications can be taught in a coordinated manner or explicitly cross-referenced through semantic mappings. This enables students to immediately recognize how abstract statistical concepts are applied in real engineering contexts, improving knowledge transfer, retention, and multidisciplinary understanding.

Outcome: If these concepts are brought together, it will help students to improve their cognitive thinking in semantically connected concepts saving their time and mature their growth in all possible related concepts in that particular area of study.

Semantic role: Ontologies and knowledge graphs serve as unifying structures that reveal hidden conceptual connections, enabling coherent curriculum design and reducing cognitive dissonance.

ii. Semantic Alignment of Learning Elements and Content Representation. The second methodological component addresses how learning content is delivered, focusing on the semantic suitability of learning elements and content granularity for different topics and learning objectives.

a) Semantic Mapping of Learning Elements to Topics

Different concepts demand different modes of representation. By semantically annotating topics with pedagogical requirements, the framework ensures that learning elements are selected based on conceptual needs rather than convenience. Topic-Element Semantic Matching

- Statistics semantically aligns with graphs, field studies, and simulations because concepts are probabilistic and data-driven.
- Law aligns with case studies and field observations, reflecting precedent-based reasoning.
- Mathematics benefits from games and puzzles, which encode abstract semantics through logic patterns.

This sub-layer introduces semantic rules that associate learning topics with the most appropriate learning elements, such as:

Table 1. Semantic Alignment Matrix — Subjects × Learning Elements

✓ = High Semantic Alignment

● = Moderate

— = Low relevance

Subject ↓ / Learning Element →	Lec- ture Note s*	PPT*	Video	Graphs	Charts	Maps / Atlas / Globes	Lab	Case Study	Field Study	Simula- tion	Games / Puzzles
Mathe- matics	✓	✓	●	●	●	—	—	—	—	✓	✓
Statistics	✓	✓	✓	✓	✓	—	✓	●	✓	✓	●
Physics	✓	✓	✓	✓	✓	—	✓	●	●	✓	●

Chemistry	✓	✓	✓	✓	✓	—	✓	◌	◌	✓	—
Engineering	✓	✓	✓	◌	◌	—	✓	✓	◌	✓	—
Computer Science	✓	✓	✓	—	◌	—	✓	✓	—	✓	◌
Medical Sciences	✓	✓	✓	✓	✓	—	✓	✓	✓	✓	◌
Law	✓	✓	✓	—	◌	✓	—	✓	✓	—	◌
Management Studies	✓	✓	✓	◌	✓	—	—	✓	✓	◌	—
Social Sciences	✓	✓	✓	✓	✓	✓	—	✓	✓	—	—
Library & Information Science (LIS)	✓	✓	✓	✓	✓	✓	◌	✓	✓	✓	◌
Environmental Studies	✓	✓	✓	✓	✓	✓	◌	✓	✓	✓	—
History	✓	✓	✓	◌	✓	✓	—	✓	✓	◌	✓
Geography	✓	✓	✓	✓	✓	✓	◌	✓	✓	✓	✓

*Lecture Notes and PowerPoint Presentations are treated as complementary learning elements rather than mutually exclusive categories, where lecture notes provide detailed conceptual exposition and reference material, while presentations offer structured, visual scaffolding to support in-class explanation and discussion. The above table is not pedagogical intuition; it is semantic suitability modeling.

b) Semantic Structuring of Content Based on Learning Intent

The methodology further differentiates content representation based on learning intent, recognizing that understanding, retention, and assessment require different semantic structures. Rather than treating learning materials as universally applicable, the framework aligns content formats with the pedagogical purpose they are intended to serve. These may include: Conceptual Understanding and Exploration (e.g., brainstorming sessions, introductory learning, concept clarification); Revision, Reinforcement, and Summative Evaluation (e.g., exam preparation, final assessments, recall-oriented learning); Skill Development and Formative Evaluation (e.g., practice-based learning, continuous assessment, competency building).

This table lists learning elements as semantic carriers, not just teaching tools.

Table 2. Semantic Mapping of Learning Elements to Learning Objectives

Learning Element	Semantic Role	Best-Suited Cognitive Purpose
Lecture Notes (Textual)	Concept definition, terminology standardization	Conceptual grounding
PPT Slides	Hierarchical structuring of concepts	Structured understanding
Video Lectures	Contextual explanation, real-world linkage	Deep comprehension
Info graphics	Concept summarization and visual cognition	Quick recall
Graphs & Charts	Pattern recognition, trend analysis	Analytical thinking
Concept Maps	Relationship visualization	Semantic linking
Simulations	Dynamic concept interaction	Applied understanding
Lab Experiments	Empirical validation of theory	Experiential learning
Case Studies	Contextual application	Decision-making
Field Studies	Environmental and situational learning	Real-world cognition

Clinical Trials	Evidence-based reasoning	Inferential learning
Maps / Globes / Atlas	Spatial representation, geospatial relationships	Spatial reasoning & contextual understanding
Problem Sets	Skill reinforcement	Procedural learning
Games / Puzzles	Logical association	Engagement & reasoning
Discussion Forums	Knowledge negotiation	Critical thinking
Question Banks	Exam-oriented semantic compression	Memorization
Flash Cards	Keyword association	Recall efficiency
MOOCs / LMS Modules	Modular learning	Self-paced cognition

Semantic principle: Each element carries different representational semantics (explanatory, procedural, experiential, or evaluative).

Semantic role: Semantic annotations encode learning intent, allowing systems to recommend the most appropriate content format for specific learning goals like:

- Understanding → Video, PPT, discussion
- Application → Labs, simulations, case studies
- Retention → Notes, flashcards, info graphics

Outcome: By explicitly encoding learning intent within the semantic structure of content, the framework enables adaptive selection and presentation of learning materials, ensuring that instructional resources align with pedagogical objectives at each stage of the learning process.

Semantic Integration Layer. Both dimensions ((i) & (ii)) are unified through a semantic integration layer, which:

- Links curriculum structure with learning elements
- Uses knowledge graphs to connect concepts, subjects, learning elements, and learner profiles
- Supports contextual retrieval and adaptive learning pathways

This layer enables learners to explore content not linearly, but conceptually, facilitating higher-order cognitive processes such as synthesis, application, and transfer of knowledge.

Learning Interaction Layer (User-Centric Layer). This layer operationalizes the semantics to enhance the learner's experience. The Learning Interaction Layer focuses on how learners and instructors interact with the semantically enriched system. Key components include:

Learner Profiles. This module maintains dynamic profiles containing:

- Learner preferences
- Learning pace and performance
- Prior knowledge and interests

Contextual Links. Based on the knowledge graph, learners receive

- Related topics
- Prerequisite concepts
- Supplementary learning materials

This ensures context-aware navigation through learning content.

Adaptive Pathways. Using semantic reasoning, the system generates:

- Personalized learning paths
- Adaptive content sequencing
- Targeted recommendations

This layer enables personalized recommendations, progress tracking, and adaptive learning experiences.

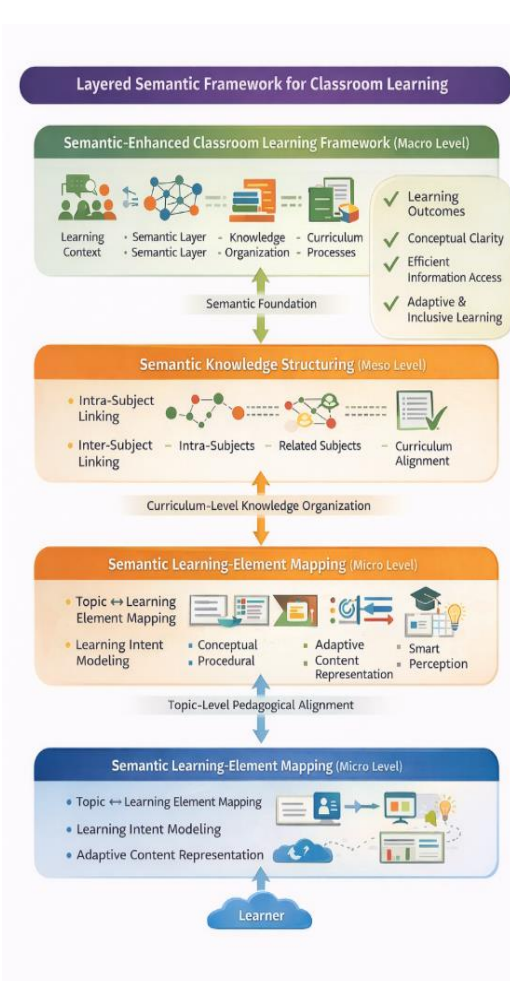


Fig. 3. Layered View of SECLF Diagram

Conceptual Integration Flow (of Semantic Layer, i.e., (i) & (ii))

The layered view of the Semantic-Enhanced Classroom Learning Framework (SECLF) visually depicts how semantic principles flow from strategic curriculum design to instructional practice. At the macro level, the framework establishes a semantic foundation by organizing learning contexts, curriculum processes, and learning outcomes through semantic layers supported by knowledge organization mechanisms. This foundation enables coherent conceptual structuring across the learning ecosystem.

Building upon this base, the meso-level layer plays a pivotal integrative role by operationalizing semantic coherence through intra- and inter-subject linking. Here, related concepts within a subject and across disciplines are semantically connected, enabling curriculum-level alignment and synchronized sequencing of conceptually related topics. This middle layer acts as a semantic bridge, translating high-level curriculum organization into structured subject relationships that can be pedagogically enacted.

The subsequent micro-level layer further refines these semantic connections by mapping topics to appropriate learning elements and modeling learning intent. This ensures that semantically linked concepts are delivered using instructional formats aligned with cognitive objectives, such as conceptual understanding, skill development, or assessment readiness. Together, the middle two layers ensure that semantic structuring is not only embedded in curriculum design but is also transformed into context-aware, learner-centric instructional delivery, making the framework both conceptually robust and pedagogically actionable.

Integrative Role of LIS Professionals and Educators. The framework positions LIS professionals and educators as complementary stakeholders in the design and implementation of semantic-enhanced learning environments. LIS professionals contribute expertise in metadata standards, ontology and controlled vocabulary development, semantic indexing, and the systematic organization of learning resources. Educators, in turn, provide pedagogical input through curriculum mapping, identification of core concepts, validation of semantic relationships, and integration of semantic pathways—such as concept maps—into instructional practices. The framework positions both LIS professionals and educators as central actors:

Table 3. Role of Key Stakeholders in SECLF

Stakeholder	Contribution
LIS Professionals	Metadata standards, ontology development, semantic indexing, resource organization
Educators	Curriculum mapping, identifying core concepts, validating semantic relationships

This collaboration ensures structural integrity in knowledge organization and pedagogical relevance in classroom application, enabling effective semantic enrichment of teaching and learning processes.

5 Conclusions

This study proposes a conceptual framework for enhancing classroom learning through the integration of semantic technologies grounded in the principles of Library and Information Science. By synthesizing insights from semantics, knowledge organization, and educational technology, the framework demonstrates how structured meaning—expressed through ontologies, semantic annotations, and knowledge graphs—can enrich instructional content and improve students' ability to understand, navigate, and relate concepts within the curriculum.

The proposed three-layer architecture (Content Layer, Semantic Layer, and Learning Interaction Layer) offers a structured way to semantically enrich classroom materials and to support meaningful learning pathways. The analysis highlights that semantic enrichment not only enhances conceptual clarity but also enables personalized access, cross-concept linking, and deeper cognitive engagement. The framework positions LIS professionals and educators as key collaborators in designing semantically coherent learning environments.

While the model remains conceptual, it lays a strong foundation for future empirical research and system development. Further studies may focus on implementing the framework in real classroom settings, developing domain-specific ontologies, evaluating student learning outcomes, or integrating the model into digital learning platforms. Overall, this work reaffirms the potential of semantics to bridge instructional content with learner cognition, contributing to more intelligent, connected, and student-centered learning ecosystems.

Declaration of AI and AI-Assisted Technologies used in Manuscript Writing Process: ChatGPT-5.2 used for grammatical assistance and spelling checks.

References

1. Alomari, J., Hussain, M., Turki, S., & Masud, M. (2015). Well-formed semantic model for co-learning. *Computers in Human Behavior*, 51, 821–828. <https://doi.org/10.1016/j.chb.2014.09.047>
2. Atzler, J. K. (2011). Twist in the list: Frame semantics as vocabulary teaching and learning tool. <http://hdl.handle.net/2152/ETD-UT-2011-05-2752>
3. Bittencourt, I. I., Costa, E., Silva, M., & Soares, E. (2009). A computational model for developing semantic web-based educational systems. *Knowledge-Based Systems*, 22(4), 302–315. <https://doi.org/10.1016/j.knosys.2009.02.012>
4. Carmichael, P., & Martinez Garcia, A. (2009, September 30). Semantic Technologies to Support Teaching and Learning with Cases: Challenges and Opportunities. 1st International Workshop on Semantic Web Applications for Learning and Teaching Support in Higher Education (SemHE'09) (30/09/09). <https://eprints.soton.ac.uk/268051/>
5. Collazos, C. A., & Garcia, R. (2007). Semantics-supported cooperative learning for enhanced awareness. *International Journal of Knowledge and Learning*, 3(4/5), 421. <https://doi.org/10.1504/IJKL.2007.016703>
6. Fraihat, S., & Shambour, Q. (2015). A Framework of Semantic Recommender System for e-Learning. *Journal of Software*, 10(3), 317–330. <https://doi.org/10.17706/jsw.10.3.317-330>
7. Gasmí, H., & Bouras, A. (2018). Ontology-Based Education/Industry Collaboration System. *IEEE Access*, 6, 1362–1371. <https://doi.org/10.1109/ACCESS.2017.2778879>
8. Huang, W., Webster, D., Wood, D., & Ishaya, T. (2006). An intelligent semantic e-learning framework using context-aware Semantic Web technologies. *British Journal of Educational Technology*, 37(3), 351–373. <https://doi.org/10.1111/j.1467-8535.2006.00610.x>
9. Iatrellis, O., Kameas, A., & Fitsilis, P. (2019). EDUC8 ontology: Semantic modeling of multi-facet learning pathways. *Education and Information Technologies*, 24(4), 2371–2390. <https://doi.org/10.1007/s10639-019-09877-4>
10. Icoz, K., Sanalan, V. A., Ozdemir, E. B., Kaya, S., & Akar, M. A. (2015). Using Students' Performance to Improve Ontologies for Intelligent E-Learning System. *Educational Sciences: Theory and Practice*, 15(4), 1039–1049
11. Jensen, J. (2019). A systematic literature review of the use of Semantic Web technologies in formal education. *British Journal of Educational Technology*, 50(2), 505–517. <https://doi.org/10.1111/bjet.12570>
12. Khmour, T. (2020). A semantic assessment framework for e-learning systems. *International Journal of Knowledge and Learning*, 13(2), 110. <https://doi.org/10.1504/IJKL.2020.106648>
13. Khedr, A. E., Idrees, A. M., & Alsheref, F. K. (2019). A Proposed Framework to Explore Semantic Relations for Learning Process Management. *International Journal of E-Collaboration (IJeC)*, 15(4), 46–70. <https://doi.org/10.4018/IJeC.2019100104>
14. Lee, J. (2017). Design of Teaching-Learning Plan Based on Instructional Strategy Ontology of Direct Instruction Model. *Journal of Knowledge Information Technology and Systems*, 12(1), 59–67. <https://doi.org/10.34163/jkits.2017.12.1.006>
15. Li, J., Gaurav, A., & Chui, K. T. (2023). Enhancing Class Management in Chinese Schools Through Semantic Web Technologies. *International Journal on Semantic Web and Information Systems*, 19(1), 1–14. <https://doi.org/10.4018/IJSWIS.328527>
16. Muñoz-Carpio, J. C., Cowling, M., & Birt, J. (2018). Framework to Enhance Teaching and Learning in System Analysis and Unified Modelling Language. 2018 IEEE International Conference on Teaching, Assessment, and Learning for Engineering (TALE), 91–98. <https://doi.org/10.1109/TALE.2018.8615284>
17. Noura, A., Cheniti-Belcadhi, L., & Braham, R. (2019). An ontology-based framework of assessment analytics for massive learning. *Computer Applications in Engineering Education*, 27(6), 1343–1360. <https://doi.org/10.1002/cae.22155>
18. Senior Lecturer, Institute/Organization: University of Technology and Applied Sciences, Salalah, Oman, & Eragamreddy, N. (2024). Examining How Semantics Influences Communication in the Classroom. *International Journal of Current Science Research and Review*, 07(05). <https://doi.org/10.47191/ijcsrr/V7-i5-96>
19. Sherimon, P. C., Vinu, P. V., & Krishnan, R. (2011). Enhancing the learning experience in blended learning systems: A semantic approach. *Proceedings of the 2011 International Conference on Communication, Computing & Security - ICCCS '11*, 449. <https://doi.org/10.1145/1947940.1948032>

20. Sila, C., Belo, O., & Barros, V. (2018). Methodology for The Development of an Ontology Network on The Brazilian National System for the Evaluation of higher Education (OntoSINAES). *JISTEM - Journal of Information Systems and Technology Management*, 15, e201815001. <https://doi.org/10.4301/S1807-1775201815001>
21. Syuhadak, Inayatur Rosyidah, Muh Zamroni, Usfiyatur Rusuly, Yesi Darasita, & Faisal Mahmoud Adam Ibrahim. (2025). Pedagogical Approach to Developing Linguistic Competence through Contextual Semantics in Classical Texts. *Izdiyar : Journal of Arabic Language Teaching, Linguistics, and Literature*, 8(2). <https://doi.org/10.22219/jiz.v8i2.35909>
22. Wang, L., Han, W., & Xi, Z. (2025). Exploring Semantic Web Tools in Education to Boost Learning and Improve Organizational Efficiency. *International Journal on Semantic Web and Information Systems (IJSWIS)*, 21(1), 1–27. <https://doi.org/10.4018/IJSWIS.370315>
23. Wei, W. (2024). Enhancing Vocational Training through the Application of Semantic Analysis in Linguistics: A Case Study of Local Normal College Students in Hebei. *Eurasian Journal of Applied Linguistics*, 10(2), 138–152.

Towards the Transformation of Operational Library Data into Linked Data: An Analytical Overview

Nabina Khamaru^{1[0009-0007-7772-1243]} and Biswanath Dutta^{2[0000-0003-3059-8202]}

^{1,2} Documentation Research and Training Centre (DRTC), Indian Statistical Institute, Bangalore, India

¹ Department of Library and Information Science, Calcutta University, Kolkata, India

{nkhamaru,bisu}@drtc.isibang.ac.in

Abstract. This paper aims to examine the transformation of the routine operational library data (e.g., acquisition, circulation, and cataloguing, etc.), which are currently stored in silos within the Library Management System (LMS), into a web-interoperable environment. This will enable semantic integration with external information resources, enhance data discoverability across platforms, and support evidence-based decision-making in libraries. The study also suggests the current limitations and interoperability challenges of the traditional MARC (library-specific) format and highlights the need for an RDF-based knowledge graph infrastructure for enabling machine-readable library data. It adopted a three-phase methodology, including collection, analysis, and interpretation, and each phase was executed following sequential steps. A total of 97 peer-reviewed articles were identified, screened, and selected for in-depth analysis. This analytical process categorized the operational library data, inspected existing linked data (LD) techniques, vocabularies, and publishing practices. Overall, this work demonstrates the importance and requisites for embedding library data into the web ecosystem, thereby promoting the transition of the traditional RDBMS into graph-based management system.

Keywords: Library Management System, operational library data, RDBMS, linked data, vocabularies, knowledge graph.

1 Introduction

Libraries utilize Library Management Systems (LMS), also known as Integrated Library Systems (ILS), to systematically record, manage, store, process, and automate their core operational functionalities, collectively referred to as housekeeping operations. These functionalities include selecting, ordering, acquiring, classifying, cataloguing, locating, lending, reserving, maintaining, distributing information and resources [1]. LMS facilitates libraries to access their collections, enhance operational effectiveness, and enable connectivity to external information resources [2]. Conventionally, LMS are relational database management systems (RDBMS) that store data generated from routine library operations. They store this data in a distributed manner across integrated modules, which support various functional areas of the library [3]. For example, the circulation module manages item check-in and check-out, hold requests, interlibrary loans, renewals, patron records, and fine management; whereas the acquisition module handles resource ordering, accessioning, vendor information, and invoicing.

Hence, LMS aims to seamlessly integrate and centralize multiple core library functions within a unified system [4]. To retrieve information from the LMS, mechanisms such as indexing and metadata standards (e.g., Machine-Readable Cataloguing (MARC)) are employed to support efficient record description, discoverability, and interoperability across diverse systems. However, library data encoded as MARC records represent a “library-specific” format, which restricts its interaction and exchange with external sectors [5]. Additionally, [6] note that although MARC facilitates the transition of library data into the digital environment, it largely replicates the traditional catalogue card in electronic form, thereby limiting web accessibility and discoverability. Thus, MARC standards lock library data, limiting the sharing of metadata beyond the library community [7]. Due to these limitations, studies indicate a significant shift from traditional ILS to new generation cloud-based systems, for example, Worldshare by OCLC, Ex Libris by Alma, and others [8, 9]. In recent times, libraries are increasingly adopting linked data (LD) initiatives at both the individual and organizational levels to make their collections more accessible and interconnected on the web, thereby reducing information silos and improving user accessibility [10, 11]. For instance, the Bibliographic Framework (BIBFRAME) initiative links the library's bibliographic records to the web [12]. [13] proposed a shift from library-specific record structure to the Resource Description Framework (RDF) model to enhance shareability, interoperability, and reusability of library data. A linked dataset (datos.bne.es) was created for the National Library of Spain's authority and bibliographic catalogues to convert them into RDF format [14]. These studies collectively highlight that library data, which is predominantly represented

in MARC format and stored in traditional RDBMS, needs to be transformed into web-reliable formats (e.g., RDF, XML, JSON-LD, etc.) to enable library metadata to interact, interconnect, and link with the web.

So, as an initial step, the current study aims to identify and categorize the various types of library data generated through routine operational activities, examine initiatives undertaken by libraries to convert such data into LD, and identify gaps and requirements for developing a system that integrates library data into the LD environment. This study focuses on the operational data generated through routine library activities (e.g., acquisition, circulation, cataloguing, etc.) that are stored in siloes within the LMS tools. It emphasizes the need to convert such data into a web-interoperable format for enhancing discoverability and enabling integration with external information resources in the online environment. This will facilitate improved library services, evidence-based decision-making, and meaningful interpretation (both syntactic and semantic) of library data, while enabling users to seamlessly explore and access resources across the web. Moreover, this also advocates a shift from traditional RDBMS towards graph-based LMS. The main contributions of this paper are:

- To systematically identify and categorize the diverse data generated during routine library operations.
- To present an analytical overview of LD techniques applied in library practices.
- It reveals notable gaps in existing practices and underscores the need to design and develop an RDF-based knowledge graph (KG) infrastructure to enable integration of library operations.

The rest of the paper is organized as follows: Section 2 presents studies that emphasize the significance and prospects of converting library data to LD. Section 3 details the research method, which includes gathering relevant information, analysing the current state, and interpreting the study’s findings. Section 4 outlines the results and insights drawn from the current work. Section 5 summarizes the study and draws its conclusion.

2 Related Work

This section provides a synthesis of relevant research studies that examine the significance, need, and possibility associated with transforming existing library data into LD entities. It will support cohesive and harmonized integration of library data across diverse systems within the web ecosystem. Various databases consulted for the literature review include Google Scholar, Scopus, ACM Digital Library, IEEE Xplore, and Semantic Scholar. The process followed for literature identification and selection is detailed in Step 3 of the Methodology section.

Over the last decade, the rapid expansion and growing heterogeneity of information objects have prompted libraries to adopt new metadata approaches [15]. Presently, libraries are actively transitioning their metadata from MARC to LD format. This transformation involves adopting RDF and ontologies to transition from a “document-centric to data-centric” model, which benefits metadata openness, discoverability, and interoperability across library datasets [13]. [16] emphasized that major organizations (like OCLC, Library of Congress) and national libraries are converting their MARC records to LD. [17] demonstrated that MARC elements can be successfully transformed to LD using the tool named eXtensible Catalogue (XC) that facilitates batch conversion of legacy metadata. Additionally, to support future data management workflow, OCLC is continuously developing robust and interactive LD infrastructure that can maintain interoperability across diverse resources and metadata providers [18].

In line with the evolving landscape, libraries are emerging with various products and services to enable LD creation and usage. For instance, the Linked Data for Production (LD4P) project transformed libraries’ technical services to LD workflow in six institutes [19]. [20] discussed that the German National Library created universally open LD for global library exchange. [21] noted that libraries are embracing Linked Open Data (LOD) projects, like BIBFRAME and Digital Public Library of America (DPLA). “BiographySampo” was created by the Finnish National Library as a LOD and semantic portal [22]. “BIBFLOW” guided individual libraries in planning and implementing LD operations [23]. Some libraries have already adopted library linked data, such as the British Library¹, the Library of Congress², and the National Libraries of Sweden³, France⁴, Germany⁵, Spain⁶, Hungary⁷, and Finland⁸. Recent research (though white papers) and blog posts present that libraries are actively incorporating

¹ <https://bl.natbib-lod.org/>

² <https://id.loc.gov/>

³ https://youtu.be/eY2eAKS_OrU?si=L2yyRd5vNurbxzhb

⁴ <https://data.bnf.fr/>

⁵ <https://www.dnb.de/>

⁶ <https://datos.bne.es/inicio.html>

⁷ https://mek.oszk.hu/html/lod_eng.html

⁸ <https://data.nationallibrary.fi/>

LD as a core component in the library management sector. For example, LIBRIS (“Swedish Union Catalogue”) leveraged LOD to free library data from rigid formats [24], the Swedish National Library planned to implement an LD integrated library system (using FOLIO and VuFind) by 2025 [25], Ex Libris underlined LOD’s potential in next-generation LMS [26], [27] pointed towards using RDF data models for integrating library data on the web, [28] investigated the usage of LODRefine for conversion of Koha catalogue into LD, and few open source library systems started publishing their bibliographic records on the web using the schema.org vocabulary [29].

While existing studies emphasize the prospect and potential of library data transformation into LD, the present study examines routine operational library data to support the development of a graph-based infrastructure for such transformation.

3 Methodology

This study follows a systematic approach to identify, analyze, and examine existing research on transforming routine library data into LD entities for web publication, interoperability, and connectivity. The overall process is divided into three phases (collection, analysis, and interpretation), with each phase further subdivided into a series of steps. The methodology adopts an analytical process that involves gathering relevant information, systematically analyzing the data, and interpreting the outcomes to draw meaningful inferences and conclusions. Consequently, this paper studies and classifies the types of library data routinely produced during operational activities, explains how they are managed using automated LMS modules, reviews individual and organizational efforts to convert library data into web-interactive, compatible, and interoperable data, and suggests the requirements for integrating library data into the LD ecosystem. Fig. 1 provides a detailed overview of the methodology followed. The methods are further detailed below:

3.1 Phase I: Collection

In this phase, the research problem is articulated, the scope of the study is defined, and the aspects addressed to resolve the issue are clearly outlined. Moreover, relevant studies are methodologically identified and compiled based on certain inclusion and exclusion criteria listed in step 3.

Step 1: Problem Statement

This paper aims to understand and examine the feasibility of transforming library data into LD and integrating it with the web for broader accessibility and extended outreach. To achieve this objective, we first need to understand how library data is categorised and managed by LMS tools. Then we need to analyze the recent initiatives undertaken by both individual and collaborative libraries to transform their data into LD. Based on this analysis, we will be able to identify limitations and outline the prerequisites for the effective integration of library data on the web. Thereby, the broader problem is further addressed through specific research questions (RQs):

RQ1. What types of data are generated through routine library operations?

RQ2. How can library operational data be systematically identified and categorized?

RQ3. What LD techniques are currently applied in library practices?

RQ4. What gaps and limitations exist in the application of LD within library operations?

RQ5. How can a web-interoperable infrastructure be developed to support the integration of library operational data?

Step 2: Scope and Coverage

This study solely focuses on the operational data generated through routine library activities, such as acquisition, cataloguing, circulation, serials control, etc., and managed using various LMS tools. Through studies, we have observed that existing research on “LD and libraries” primarily emphasized bibliographic (cataloguing) and authority data, with limited weightage to holdings and serials data. Meanwhile, other categories of operational data largely remain unexplored, which is addressed in the current study.

Step 3: Literature Selection

The key terms used for literature search include: “types of library data”, “libraries and linked data”, “LD for libraries”, and “LD in libraries”. Table 1 presents the criteria considered essential while selecting the relevant literature. The various academic databases consulted for the research are listed in Section 2, Related Work.

Table 1. Criteria for inclusion and exclusion of literature.

Inclusion criteria	Exclusion criteria
Peer-reviewed journal articles and conference papers.	Papers with no accessible full-text.

Publications between the period 2000 and 2025	Publications that discussed archives, museums, galleries, or other cultural organizations.
Articles written only in the English language.	Studies that do not address the management of library data.
Articles that discussed LD in the context of libraries.	Studies that involve library data managed through systems (e.g., digital repositories, discovery solutions) other than LMS tools.

Primarily, we found a total of 245 articles. After removing duplicates, checking the abstracts, and shortlisting based on the criteria outlined in Table 1, a total of 97 articles were ultimately selected for the study.

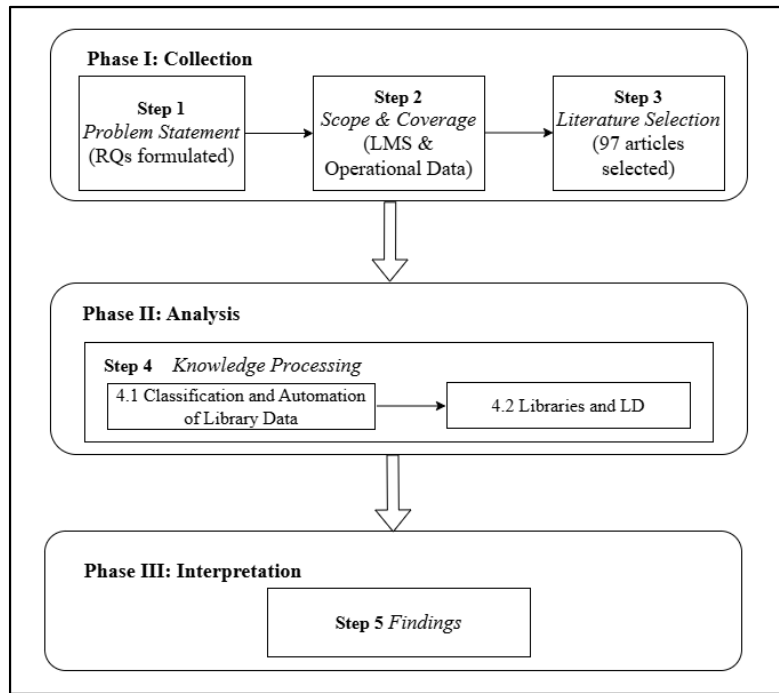


Fig. 1. Overview of the Methodology followed.

3.2 Phase II: Analysis

This phase focuses on analysing the accumulated knowledge to examine operational library data and their automation within LMS tools. It aims to understand the prevailing practices in library data formats and storage. This part also helps to outline the various initiatives, programs, and projects undertaken to enable linked library data.

Step 4: Knowledge Processing

4.1. Classification and automation of library data

Based on a thorough review and understanding, we analysed that the LMS follows a structured architecture consisting of presentation, application, data, and integration layers. The data layer hosts the relational database that stores all records (e.g., library resource details, user information, library transaction history, fines and overdue, etc.) generated through core library functions (referred to as housekeeping operations), which include acquisition, processing, use, and maintenance of library resources [30]. These functions are implemented through individual modules that collectively enable automation of the LMS workflow [31]. This clear understanding of the backend structure is essential for the effective transition of a traditional LMS to a web-interoperable system. Fig. 2 below details the complete process.

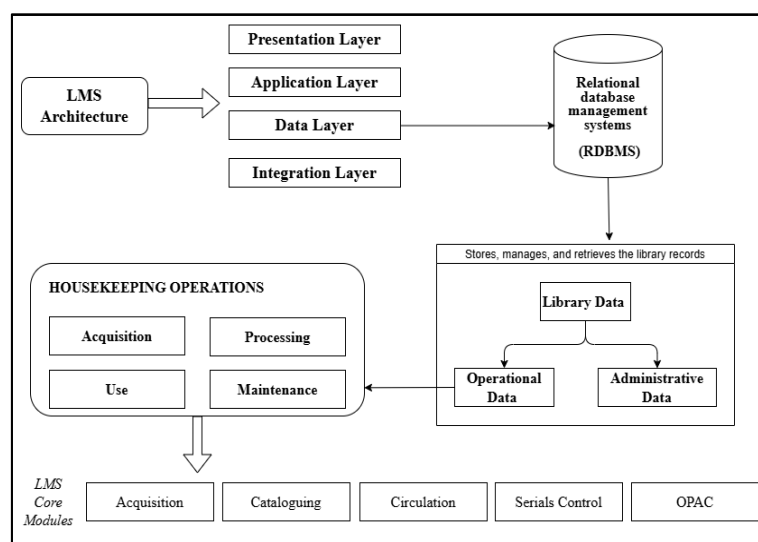


Fig 2. LMS and its data management

4.2. Libraries and LD

We examined various projects and initiatives undertaken by individual libraries (e.g., National Library of Sweden) and library organizations (e.g., OCLC) for the application of LD in library operations. This study indicates that “LD and libraries” remains an active and ongoing area of research. Following this, in the Findings section, we have further enlisted some of these predominant LD applications in the context of libraries.

3.3 Phase III: Interpretation

In this phase, we inspect the trends in the transformation of library data into LD and interpret the system-level changes that need to be adapted for enabling seamless conversion and interlinking of library data for web interoperability and accessibility. Hence, it comprehends the prerequisites and necessities for integrating operational library data into the LD environment.

Step 5: Findings

Ultimately, this step provides an explanatory solution to the broader research problem by addressing every individual RQ. It aids in deriving and comprehensively presenting insights gained through analysis of the selected literature.

4 Findings

This section addresses each research question (RQ) formulated in Step 1 of the *Methodology* and supports it with evidence from the relevant literature. All proposed RQs are dealt with step-by-step, based on the authors’ understanding of the reviewed literature. This part substantiates the suggested solutions by drawing on insights obtained from the analysis and interpretation of the selected literature. This process streamlines the study by clearly identifying the types of library data involved, how they are conventionally managed, and what measures are required in the current digital context to ensure adaptability and interoperability within the web ecosystem. Presented below are the detailed solutions derived for each RQ:

RQ1. *What types of data are generated through routine library operations?*

Upon analysis of the library data generated through various housekeeping operations as discussed in Step 4.1 of the Methodology section, the study identifies five primary data categories, such as acquisition, cataloguing (bibliographic), circulation, serial control, and others (e.g., patron data, usage data, system and policy data, etc.).

RQ2. *How can library operational data be systematically identified and categorized?*

In any LMS tool, the library operations are managed through a modular architecture, where each core library function (e.g., circulation, acquisition, OPAC, etc.) is handled by an individual module. Each module is integrated with the others to enable the LMS to function as a unified system [32]. Here, we reviewed relevant literature and

blog posts that examine the architectural setup of the LMS tools to identify and categorize the operational library data. This is further illustrated in Fig. 2.

RQ3. What LD techniques are currently applied in library practices?

Although the potential of LD applications in libraries has been recognized, it remains an area of ongoing research. For example, several well-known LD projects have been undertaken by libraries in different countries, including Finland's LDF.fi project (<https://www.ldf.fi/project.html>) and the Linked Library Data (LKD) project (<https://www.kiwi.fi/spaces/LKD/pages/298647995/Linked+Library+Data+project+LKD>), the Organization Name Linked Data (ONLD) project by North Carolina State University (<https://www.lib.ncsu.edu/projects/organization-name-linked-data>), CONTENTdm by OCLC (<https://www.oclc.org/research/areas/data-science/linkddata/contentdm-linked-data-pilot.html>), UC LD project (<https://libraries.universityofcalifornia.edu/doc/ucldp/>) by the University of California Libraries, Cambridge OPen METadata (COMET) project (<https://librarytechnology.org/pr/15508/university-of-cambridge-and-oclc-research-collaborate-on-open-metadata-project>) jointly by University of Cambridge and OCLC Research, etc. Moreover, various LOD working groups are actively supporting libraries in LD initiatives. A few of them are listed here as follows: LIBER LOD Working Group (<https://libereurope.eu/working-group/liber-linked-open-data-working-group/>), IGeLU (<https://igelu.org/products-and-initiatives/communities-of-practice/linked-open-data/>) LOD Working Group by Ex Libris, etc.

From the perspective of practical application, several predominantly discussed and adopted LD services in libraries include BIBFRAME (<https://bibframe.org/>), Web Dewey LD (https://help.oclc.org/Metadata_Services/WebDewey/WebDewey_Linked_Data) by OCLC, WorldCat (<https://search.worldcat.org/>), EBSCO BiblioGraph (<https://about.ebsco.com/products/bibliograph>), Universal Decimal Classification (UDC) summary LD (<https://udcdata.info/>), as well as authoritative sources, such as VIAF (<https://viaf.org/en>), FAST (<https://id.worldcat.org/fast/961166>), GeoNames (<https://www.geonames.org/>), etc.

RQ4. What gaps and limitations exist in the application of LD within library operations?

From the literature review, it is evident that distinct vocabularies (a.k.a ontologies) are designed and developed for managing different library data. It also reveals that while numerous ontologies exist for bibliographic and authority data, only a few studies have highlighted the importance of holdings and serials data [33–35]. However, there remains a lack of vocabulary for representing, analyzing, and managing other library data, such as acquisition, circulation, digital assets, patron information, interlibrary loans, etc.

Research shows that libraries globally use varied data elements, diverse ontologies, and vocabularies for describing resources. [36] highlighted the use of various data models for semantic web integration; [37] noted schema differences; MarcOnt [38] integrated MARC21, BibTeX, and Dublin Core for interoperability; [39] detailed OCLC's LD efforts; [14] documented Spain's National Library's MARC21-to-RDF conversion (using IFLA ontologies, RDA, Dublin Core, etc.); [40] compared bibliographic records using multiple ontologies (e.g., BIBO, FaBio, BIBFRAME, schema.org, MADSRDF/MODSRDF). These variations in schema pose challenges related to standardization and interoperability. Due to space constraints, the authors could highlight only a few gaps and limitations in this paper.

RQ5. How can a web-interoperable infrastructure be developed to support the integration of library operational data?

The shift from MARC to LD in library metadata management has enhanced resource description. However, attempts to align native models (like RDA, BIBFRAME, FRBR, etc.) with MARC have resulted in ontological disparities, unique extensions, repurposing schemas, incompatible data models, and different choices of data elements [41–46]. These studies emphasize the need for a standardized, comprehensive KG data model to enable seamless library data integration across applications, thereby making library metadata web-interoperable. Additionally, the model should be developed using LD practices and implemented in a web-compatible RDF structure. This will also support graph databases as a promising alternative to traditional RDBMS for library data management.

For instance, the EBSCO FOLIO (<https://about.ebsco.com/academic-libraries/products/ebsco-folio>) is a modern and flexible LMS that employs LD techniques to enhance interoperability across platforms and support the application of new innovative services. At present, this collaborative Library Services Platform (LSP) supports various core library functionalities, including acquisition, circulation, cataloguing, and electronic resource management across domains. Few North American institutions are currently using FOLIO LSP for their library operations.

Table 2. Presents a concise overview of the corresponding answers for each of the proposed RQs.

Sl. No.	Research Question (RQ)	Finding/Solution
RQ1.	<i>What types of data are generated through routine library operations?</i>	Acquisition, Circulation, Cataloguing, Serial Control, and others.
RQ2.	<i>How can library operational data be systematically identified and categorized?</i>	Understanding the modular architecture of LMS, where each module handles core library functionalities (e.g., circulation, acquisition, OPAC, etc.) and the metadata elements involved in each process.
RQ3.	<i>What LD techniques are currently applied in library practices?</i>	Several LD applications are BIBFRAME, WorldCat, Web Dewey LD, UDC summary LD, EBSCO BiblioGraph, etc., and authority sources like VIAF, FAST, etc. Certain libraries have already adopted library LD, such as the National Library of Sweden, followed by the Library of Congress, the British Library, and the National Libraries of France, Germany, Spain, Hungary, and Finland. Various LD projects (refer to section 4 <i>Findings</i>) and initiatives (e.g., EBSCO FOLIO, National Library of Sweden's LD integrated library system using FOLIO and VuFind, LIBRIS, etc.).
RQ4.	<i>What gaps and limitations exist in the application of LD within library operations?</i>	Varied data elements, diverse ontologies, and vocabularies for describing resources. Lack of vocabulary for managing different types of library data, except for bibliographic and authority records.
RQ5.	<i>How can a web-interoperable infrastructure be developed to support the integration of library operational data?</i>	Need to develop a KG data model following LD practices and based on web-reliable RDF structure, which makes library metadata web-interoperable. Promote the adoption of a graph-based management system as an alternative to traditional RDBMS.

5 Discussion and Conclusion

Overall, this paper investigates the types of data generated through routine library operations, the systematic identification and organization of such data, the LD practices currently used by libraries, the gaps and limitations encountered in their implementation, and the need to develop a web-interoperable infrastructure to enable the integration of library operational data. As reported by [47], reliance on traditional RDBMS constraints libraries and library systems to fully unlock the potential of their data. The LMS stores library records in silos within its database, preventing data integration with external platforms and the web, thereby limiting the data's full potential. Thereby, this study explores the system changes and requirements involved in the transformation of routine library data from library-specific MARC records into a web-compatible format like the RDF structure. It is also influenced by recent efforts undertaken by individual and organizational libraries to publish their library metadata as linked entities, thereby enabling their integration and interoperability within the web ecosystem. Hence, this work underscores the need to move from traditional RDBMS to graph based system for library data storage, management, and retrieval.

This study is confined to library operational data and excludes administrative data, as these are institution-specific and often sensitive. Moreover, the electronic resource management (ERM) system that supports the management of library resources is beyond the scope of this study and will be examined in future work.

Declaration of AI and AI-Assisted Technologies used in Manuscript Writing Process: Not Applicable

References

1. Alexander, J.O.: Library Applications. In: Encyclopedia of Information Systems. pp. 55–76. Elsevier (2003).
2. Kochtanek, T.R., Matthews, J.R.: Library Information Systems: From Library Automation to Distributed Information Access Solutions. Libraries Unlimited (2002).

3. Gajbhiye, S., Patrange, M., Kumbhare, B., Laxane, Y.B.: Design and Implementation of an Automated Library System for Modern Academic Environments. *Indian J. Comput. Sci. Technol.* 160–166 (2025). <https://doi.org/10.59256/indjst.20250402019>.
4. Chow, A.S., Bucknall, T.: 3 - Customized fashion: finding the right fit. In: *Library Technology and User Services*. pp. 33–79. Chandos Publishing (2012).
5. Andresen, L.: After MARC – what then? *Library Hi Tech.* 22, 40–51 (2004). <https://doi.org/10.1108/07378830410524486>.
6. Schreur, P.E., Carlson, A.J.: Bridging the Worlds of MARC and Linked Data: Transition, Transformation, Accountability. *Ser. Libr.* 78, 48–56 (2020). <https://doi.org/10.1080/0361526X.2020.1716584>.
7. Kumar, S., Ujjal, M., Utpal, B.: Exposing MARC 21 Format for Bibliographic Data As Linked Data With Provenance. *J. Libr. Metadata.* 13, 212–229 (2013). <https://doi.org/10.1080/19386389.2013.826076>.
8. Wilson, K.: The Knowledgebase at the Center of the Universe. *Ser. Libr.* 72, 111–117 (2017). <https://doi.org/10.1080/0361526X.2017.1320869>.
9. Yang, S.: From integrated library systems to library management services: time for change? *Library Hi Tech News.* 30, 1–8 (2013). <https://doi.org/10.1108/LHTN-02-2013-0006>.
10. Miller Presenter, E., Westfall Recorder, M.: Linked Data and Libraries. *The Serials Librarian.* 60, 17–22 (2011). <https://doi.org/10.1080/0361526X.2011.556427>.
11. Williams, J.: From silo to the web: Library cataloging data in an open environment. In: *Adv. Libr. Adm. Organ.* pp. 175–191. Emerald Group Holdings Ltd. (2021). <https://doi.org/10.1108/S0732-067120210000042013>.
12. Gonzales, B.M.: Linking libraries to the web: Linked data and the future of the bibliographic record. *Inf. Technol. Libr.* 33, 10–22 (2014). <https://doi.org/10.6017/ital.v33i4.5631>.
13. Alemu, G., Stevens, B., Ross, P., Chandler, J.: Linked Data for libraries: Benefits of a conceptual shift from library-specific record structures to RDF-based data models. *New Libr. World.* 113, 549–570 (2012). <https://doi.org/10.1108/03074801211282920>.
14. Vila-Suero, D., Villazon-Terrazas, B., Gomez-Perez, A.: Datos.bne.es: A library linked data dataset. *Semantic Web.* 4, 307–313 (2013). <https://doi.org/10.3233/SW-120094>.
15. Alemu, G., Stevens, B.: An Emergent Theory of Digital Library Metadata: Enrich then Filter. Chandos Publishing (2015).
16. Wang, Y., Yang, S.Q.: Linked Data Technologies and What Libraries Have Accomplished So Far. *International Journal of Librarianship.* 3, 3–20 (2018). <https://doi.org/10.23974/ijol.2018.vol3.1.62>.
17. Bowen, J.: Moving library metadata toward linked data: Opportunities provided by the extensible catalog. In: *Proc. Int. Conf. Dublin Core Metadata Appl.* pp. 44–59 (2010).
18. Washington, A., Mixer, J.: Linked Data in Library Workflows. *NASIG Proceedings.* 38, (2024). <https://doi.org/10.3998/nasig.6729>.
19. Schreur, P.E., Lorimer, N.: Linked data in libraries' technical services workflows. In: Koutsomihia D., Garoufallou E., Siatiri R., and Virkus S. (eds.) *Commun. Comput. Info. Sci.* pp. 224–229. Springer Verlag (2017). https://doi.org/10.1007/978-3-319-70863-8_21.
20. Hannemann, J., Kett, J.: Linked data for libraries. In: *Proc of the world library and information congress of the Int'l Federation of Library Associations and Institutions (IFLA)* (2010).
21. Mitchell, E.T.: Chapter 1. The Current State of Linked Data in Libraries, Archives, and Museums. *Library Technology Reports.* 52, 5–13 (2016).
22. Tamper, M., Leskinen, P., Hyvönen, E., Valjus, R., Keravuori, K.: Analyzing biography collections historiographically as Linked Data: Case National Biography of Finland. *Semantic Web.* 14, 385–419 (2022). <https://doi.org/10.3233/SW-222887>.
23. Smith, M., Stahmer, C.G., Li, X., Gonzalez, G.: BIBFLOW: A Roadmap for Library Linked Data Transition. (2017).
24. Wennerlund, B., Berggren, A.: Leaving Comfort Behind: a National Union Catalogue Transition to Linked Data. Presented at the Data intelligence in libraries: the actual and artificial perspectives (2017).
25. Library Technology Guides, L.T.G.: The National Library of Sweden makes strategic decision to implement a new Library Service Platform, <https://librarytechnology.org/pr/29785>, last accessed 2025/02/13.
26. Kaschte, A.: Linked open data on its way into next generation library management and discovery solutions. *JLIS.it.* 4, 313–313 (2013). <https://doi.org/10.4403/jlis.it-5492>.
27. Knight, F.: Break On Through to the Other Side: The Library and Linked Data. *Librarian Publications & Presentations.* (2011).
28. Mandal, S., Saha, S.: Moving Koha library catalogue into linked data using the LODRefine. *Libr. Hi Tech News.* 39, 12–16 (2022). <https://doi.org/10.1108/LHTN-12-2021-0092>.
29. Scott, D.: Seeding Structured Data by Default via Open Source Library Systems. In: Presutti, V., d'Amato, C., Gandon, F., d'Aquin, M., Staab, S., and Tordai, A. (eds.) *The Semantic Web: Trends and Challenges*. pp. 659–674. Springer International Publishing, Cham (2014). https://doi.org/10.1007/978-3-319-07443-6_44.
30. IGNOU.: Unit 3 Housekeeping Operations.
31. Gangu, B.T., Reddy, E.R.: Automating the House-keeping operations Experience At The IGM Library University OF Hyderabad. (1996).
32. Chow, A.S., Timothy, B.: *Emerging technology trends in libraries*. Chandos Publishing (2012). <https://doi.org/10.1016/B978-1-84334-638-8.50006-0>.

33. Fallgren, N., Lauruhn, M., Reynolds, R.R., Kaplan, L.: The Missing Link: The Evolving Current State of Linked Data for Serials. *Ser. Libr.* 66, 123–138 (2014). <https://doi.org/10.1080/0361526X.2014.879690>.
34. Han, M.-J.K., Cole, T.W., Lampron, P., Sarol, M.J.: Exposing library holdings metadata in RDF using schema.org semantics. In: Malta M.C., Borsetti Gregorio Vidotti S.A., and Malta M.C. (eds.) *Proc. Int. Conf. Dublin Core Metadata Appl.* pp. 41–49. Dublin Core metadata initiative (2015).
35. Zapounidou, S.: Studying conceptual models for publishing library data to the semantic web. In: Manolopoulos Y., Kamps J., Tsakonas G., Iliadis L., and Karydis I. (eds.) *Lect. Notes Comput. Sci.* pp. 652–655. Springer Verlag (2017). https://doi.org/10.1007/978-3-319-67008-9_63.
36. Zapounidou, S., Sfakakis, M., Papatheodorou, C.: Highlights of Library Data Models in the Era of Linked Open Data. In: *Commun. Comput. Info. Sci.* pp. 396–407. Springer Verlag (2013). https://doi.org/10.1007/978-3-319-03437-9_38.
37. Park, H., Kipp, M.: Library Linked Data Models: Library Data in the Semantic Web. *Cataloging & Classification Quarterly*. 57, 261–277 (2019). <https://doi.org/10.1080/01639374.2019.1641171>.
38. Kruk, S.R., Synak, M., Zimmermann, K.: MarcOnt: integration ontology for bibliographic description formats. In: *Proceedings of the 2005 international conference on Dublin Core and metadata applications: vocabularies in practice*. pp. 1–5. Dublin Core Metadata Initiative, Madrid, Spain (2005).
39. Fons, T., Penka, J., Wallis, R.: OCLC's Linked Data Initiative: Using Schema.org to Make Library Data Relevant on the Web. *ISQ*. 24, 29 (2012). <https://doi.org/10.3789/isqv24n2-3.2012.05>.
40. Jett, J., Nurmikko-Fuller, T., Cole, T.W., Page, K.R., Downie, J.S.: Enhancing scholarly use of digital libraries: A comparative survey and review of bibliographic metadata ontologies. In: *2016 IEEE/ACM Joint Conference on Digital Libraries (JCDL)*. pp. 35–44 (2016).
41. Bigelow, I., Sparling, A.: Control or Chaos: Embracing Change and Harnessing Innovation in an Ecosystem of Shared Bibliographic Data. *JLIS.it*. 13, 67–85 (2022). <https://doi.org/10.4403/jlis.it-12735>.
42. Kettouch, M.S., Luca, C.: LinkD: element-based data interlinking of RDF datasets in linked data. *Computing*. 104, 2685–2709 (2022). <https://doi.org/10.1007/s00607-022-01107-z>.
43. Samples, J., Bigelow, I.: MARC to BIBFRAME: Converting the PCC to Linked Data. *Cat. Classif. Q.* 58, 403–417 (2020). <https://doi.org/10.1080/01639374.2020.1751764>.
44. Schreur, P.E., Carlson, A.J.: Bridging the Worlds of MARC and Linked Data: Transition, Transformation, Accountability. *Ser. Libr.* 78, 48–56 (2020). <https://doi.org/10.1080/0361526X.2020.1716584>.
45. Suominen, O., Hyvönen, N.: From MARC silos to Linked Data silos? o-bib. *Das offene Bibliotheksjournal / Herausgeber VDB*. 4, 1–13 (2017). <https://doi.org/10.5282/o-bib/2017H2S1-13>.
46. Tosaka, Y.: Analyzing Library Metadata for Web-Based Metadata Reuse Services: A Case-Study Examination of WorldCat.org and RefWorks. *Journal of Library Metadata*. 10, 257–275 (2010). <https://doi.org/10.1080/19386389.2010.524864>.
47. Kent, A.: *Encyclopedia of Library and Information Science: Volume 45 - Supplement 10: Anglo-American Cataloguing Rules, Second Edition to Vocabularies for Online Subject Searching*. CRC Press (1989).

Information Retrieval in Complex Digital Environments

Navigating Complexity: Advanced Approaches to Information Retrieval in Modern Digital Ecosystems

Debkanta Halder^{1[0009-0007-7772-1243]} and Subarna Kumar Das²

^{1,2} Department of Library and Information Science, Jadavpur University

¹debkantah.lis.rs@jadavpuruniversity.in

²skd_ju2002@yahoo.co.in

Abstract. Information retrieval in complex digital environments presents significant challenges as organizations increasingly manage vast, heterogeneous data repositories spanning structured databases, unstructured documents, multimedia content, and multilingual sources. This report examines contemporary approaches to retrieval across intricate information landscapes, integrating semantic search technologies, neural ranking models, and knowledge graphs with traditional indexing methodologies. The study synthesizes current literature on information retrieval system architectures, evaluates critical challenges including scalability, data heterogeneity, multilingual processing, and query understanding, and explores emerging solutions leveraging artificial intelligence and machine learning. Findings reveal that hybrid retrieval systems combining dense vector embeddings with sparse term-based approaches yield superior performance compared to single-modality techniques. The research identifies essential evaluation metrics necessary for systematic assessment of retrieval effectiveness. Recommendations emphasize the integration of deep learning models such as BERT transformers for contextual understanding, implementation of federated search architectures for distributed collections, and development of relevance feedback mechanisms for continuous system improvement. Future research directions highlight the convergence of information retrieval with generative AI, necessitating new frameworks for instruction-following retrieval models and integration with large language models. This investigation demonstrates that effective information retrieval in complex environments requires multifaceted approaches combining advanced indexing techniques, sophisticated ranking algorithms, and adaptive learning mechanisms.

Keywords: Semantic Search, Knowledge Graphs, Natural Language Processing, Distributed Systems, Vector Embeddings, Query Understanding.

1 Introduction

The digital information landscape has undergone transformative changes over the past two decades, evolving from relatively simple, text-based repositories to complex, multifaceted ecosystems encompassing diverse data types, sources, and user contexts. Contemporary digital environments, as examined in this study, specifically encompass institutional repositories managing scholarly outputs, digital library collections preserving cultural heritage materials, federated discovery platforms aggregating resources across multiple institutions, archival systems maintaining historical records, and organizational knowledge bases supporting enterprise information needs. These systems are characterized by unprecedented scale, variety, and velocity of information creation and dissemination. For instance, modern institutional repositories such as DSpace and EPrints manage millions of scholarly articles, theses, and datasets across thousands of institutions globally, while digital library platforms like Europeana aggregate over 50 million cultural heritage items from more than 3,700 institutions. Organizations and individuals face mounting challenges in locating relevant information amidst this abundance, making effective information retrieval systems more critical than ever for libraries, archives, and information centers.

Information retrieval, defined as the process of obtaining information resources relevant to an information need from a collection, has traditionally relied on keyword matching and statistical relevance models such as TF-IDF and BM25. However, these conventional approaches often prove inadequate when confronted with the complexities inherent in modern digital library environments. Such complexities include semantic ambiguity, where identical terms carry different meanings across disciplines (for example, 'cell' in biology versus cellular technology), heterogeneous data formats spanning bibliographic metadata, full-text articles, archival finding aids, images, and structured datasets, distributed information sources requiring coordinated search across multiple institutional repositories and digital collections, and diverse user populations including researchers, students, librarians, and the general public with varying information literacy levels, expertise in search formulation, and domain-specific needs.

The rise of scholarly big data has amplified these challenges exponentially. Academic institutions now contend with petabytes of research data generated from laboratory instruments, social media archives, digitized historical collections, and institutional publication outputs. For example, the CERN Large Hadron Collider generates approximately 30 petabytes of research data annually that requires sophisticated discovery mechanisms, while digital humanities projects produce vast textual corpora requiring advanced natural language processing for meaningful access. This data explosion necessitates retrieval systems capable of processing and indexing massive volumes while maintaining acceptable response times (typically under one second for interactive queries) and retrieval quality measured through precision and recall metrics. Furthermore, the globalization of digital information demands systems that can bridge linguistic and cultural barriers, enabling effective cross-lingual information access across library collections serving diverse international user communities.

Recent advances in artificial intelligence, particularly machine learning and natural language processing, have opened new avenues for addressing these challenges within library and information science contexts. Specifically, transformer-based language models such as BERT (Devlin et al., 2019) enable semantic understanding of user queries and document content beyond keyword matching, while dense retrieval methods using learned vector representations (Karpukhin et al., 2020) allow for similarity-based search that captures conceptual relationships. For instance, a researcher searching for 'climate change mitigation strategies' in an institutional repository can retrieve relevant documents discussing 'carbon reduction policies' or 'greenhouse gas abatement' even without exact keyword matches. Neural ranking models can capture semantic relationships beyond simple keyword matching, while deep learning architectures enable automated feature learning from raw bibliographic and full-text data. Recent work on retrieval-augmented generation (Jiang et al., 2025) demonstrates how large language models can enhance document discovery by combining retrieval with natural language understanding. These technological developments, combined with distributed computing frameworks enabling federated search across institutional boundaries and cloud infrastructure supporting scalable indexing operations, provide foundations for next-generation information retrieval systems designed specifically for complex digital library and archival environments.

This report synthesizes research findings from computational information retrieval, digital libraries, and scholarly communication domains, analyzes contemporary methodologies with critical evaluation of their applicability to institutional settings, and provides evidence-based recommendations for practitioners managing discovery systems and researchers advancing retrieval technologies. The investigation encompasses theoretical foundations from information science, practical implementations in production library systems, and future directions for this rapidly evolving field at the intersection of library science and artificial intelligence.

2 Objectives

- To comprehensively examine and evaluate approaches to information retrieval within complex digital environments, identifying effective methodologies, persistent challenges, and opportunities for advancement.
- To review and synthesize current theoretical frameworks and models applicable to information retrieval in complex digital settings, with particular attention to recent advances in neural information retrieval and hybrid approaches.
- To evaluate the effectiveness of emerging technologies, particularly neural networks, machine learning, and semantic web technologies, in addressing retrieval challenges within library and information science contexts.
- To assess the role of distributed and federated architectures in enabling scalable information retrieval across heterogeneous data sources including institutional repositories, digital libraries, and archival systems.

3 Literature Review

3.1 Recent Advances in Neural Information Retrieval

The past five years have witnessed a substantial transformation in information retrieval through neural approaches. Hambarde and Proenca (2023) provide a comprehensive survey of recent advances, categorizing developments into dense retrieval methods using learned embeddings, cross-encoder reranking architectures, and hybrid systems combining neural and lexical approaches. Their critical analysis reveals that while pure neural methods achieve strong performance on benchmark datasets, hybrid approaches demonstrate superior robustness across diverse query types and domains, particularly relevant for heterogeneous library collections where queries range from known-item searches to exploratory information seeking.

Karpukhin et al. (2020) introduced Dense Passage Retrieval (DPR), which encodes queries and documents into dense vector representations learned through contrastive learning. This approach addresses a fundamental limitation of traditional lexical matching by capturing semantic similarity even when queries and relevant documents share no common terms. For institutional repositories serving interdisciplinary research, this capability proves essential as researchers often formulate queries using terminology from their own discipline while seeking relevant work published under different terminological conventions. However, critical evaluation reveals that DPR requires substantial labeled training data and computational resources for fine-tuning, presenting practical barriers for smaller institutions with limited technical infrastructure.

The integration of retrieval systems with large language models represents a particularly significant recent development. Jiang et al. (2025) examine retrieval and structuring augmented generation, demonstrating how retrieval systems can provide factual grounding for language model outputs while language models can enhance retrieval through query reformulation and result reranking. Their work on retrieval-augmented generation shows promising applications for digital library reference services, enabling more sophisticated question answering from institutional collections. However, the computational requirements and potential for hallucination in generated responses require careful implementation considerations for production library systems where accuracy and verifiability remain paramount. This critical assessment suggests that while retrieval-augmented generation offers exciting possibilities, libraries must balance innovation with traditional values of information accuracy and source attribution.

3.2 Semantic Search and Knowledge Graphs in Library Contexts

Semantic search technologies aim to understand query intent and document meaning beyond surface-level keyword matching, though early work by Guha et al. (2003) on Semantic Search highlighted challenges in leveraging structured metadata and ontological relationships for improved retrieval. More recently, knowledge graphs have emerged as powerful tools for enhancing search in library and archival contexts through explicit representation of entities (authors, works, subjects), relationships (authorship, citation, aboutness), and factual knowledge. Singhal's (2012) description of Google's Knowledge Graph illustrated practical applications of semantic technologies at web scale, enabling entity-centric search, factual question answering, and disambiguation of ambiguous queries relevant to reference librarianship.

Research by Xiong et al. (2017) demonstrated knowledge graph applications in academic search, showing how entity embeddings and relationship modeling improve retrieval effectiveness. However, critical evaluation reveals significant limitations for library implementation. Knowledge graphs require extensive manual curation or sophisticated entity extraction systems, both representing substantial ongoing resource investments. Coverage limitations pose particular challenges as knowledge graphs typically focus on prominent entities while library collections include many niches scholarly works and archival materials not represented in general knowledge bases. Furthermore, maintenance requirements intensify as collections grow and evolve, raising questions about long-term sustainability for institutions without dedicated knowledge organization staff. These practical constraints suggest that knowledge graphs may be most viable for focused special collections or collaborative efforts where multiple institutions share curation responsibilities rather than as comprehensive solutions for all library discovery needs.

3.3 Distributed and Federated Architectures for Library Collections

As library collections increasingly span multiple institutional repositories and digital platforms, distributed and federated search architectures have become essential infrastructure. These systems enable unified search interfaces across heterogeneous collections while respecting institutional autonomy and avoiding costly centralized indexing. Federated search protocols such as Z39.50 and SRU have long served library consortia, though contemporary implementations increasingly leverage modern web APIs and search aggregation frameworks.

However, critical analysis reveals significant technical and organizational challenges in federated search deployment. Query latency increases substantially when searching multiple collections sequentially, degrading user experience unless sophisticated parallel query execution and timeout handling are implemented. Result merging presents additional complexity as different collections use different relevance ranking algorithms, metadata schemas, and quality standards, making cross-collection relevance comparison problematic. Furthermore, collection selection mechanisms that determine which repositories to query for particular information needs require ongoing maintenance as collection coverage evolves. These challenges suggest that federated search, while enabling important discovery scenarios, demands substantial technical expertise and organizational coordination beyond the

capabilities of many individual libraries. Collaborative development of shared federated search platforms may offer more sustainable approaches than institution-specific implementations.

4 Methodology

4.1 Research Design

This investigation employed a mixed-methods research design combining comprehensive literature analysis with empirical evaluation of retrieval systems and methodologies. The approach integrated multiple evidence sources to provide triangulated insights into information retrieval in complex digital environments. The research encompassed both qualitative examination of theoretical frameworks and quantitative assessment of system performance.

The methodology was structured in four interconnected phases: systematic literature review, comparative system analysis, performance evaluation, and synthesis of findings. This phased approach enabled progressive refinement of understanding while maintaining rigorous standards for evidence collection and analysis.

4.2 System Selection and Characterization

Comparative analysis focused on representative systems embodying different approaches to information retrieval in complex environments. Selection criteria included architectural diversity, application domains, scale of deployment, and documentation availability. Systems examined included:

- Traditional inverted index systems employing BM25 ranking with enhancements for large-scale deployment.
- Neural retrieval systems utilizing deep learning models for semantic matching.
- Hybrid systems combining traditional retrieval with neural reranking.
- Distributed search architectures for federated retrieval across multiple collections.
- Semantic search systems leveraging knowledge graphs and entity recognition.
- Domain-specific retrieval systems optimized for specialized content types.

Each system was characterized according to its architecture, indexing approach, ranking methodology, query processing capabilities, scalability characteristics, and support for complex retrieval scenarios, including multi-modal content, cross-lingual search, and personalization.

5 Findings

The primary findings confirm the necessity of hybrid methods, the critical role of vector efficiency algorithms, and the profound challenge posed by non-uniform data distribution in large-scale CDEs.

5.1 Performance analysis of Sparse vs. Dense vs. Hybrid retrieval in CDEs

The analysis confirms that while deep neural models provide significant performance improvements in capturing semantic relevance, they cannot operate effectively in isolation within a CDE. Dense retrieval excels at generating conceptual connections, but it consistently requires augmentation from sparse methods. Systems that implement hybrid retrieval—combining semantic matching with keyword-based scoring (like BM25)—demonstrate superior overall performance, especially where specific facts or technical terminology are involved. The non-determinism inherent in pure semantic models, where slight variations in query phrasing can undermine accuracy, necessitates the robustness of lexical matching, particularly in domains requiring high reliability and accountability, such as technical or legal environments [13] [16]. The relationship between these models and their relevance to CDE implementation is summarized below.

Table 1. Comparative Analysis of Neural Ranking Models for Library Discovery Systems

Model Type	Example Model	Mathematical Basis	Primary Strength in CDEs	Limitation
------------	---------------	--------------------	--------------------------	------------

Set-Theoretic	Standard Boolean	Logic and Set Operations	High precision in standardized / legal queries	Lacks ranking capability; no semantic inference
Algebraic	Vector Space Model (VSM)	Linear Algebra (Vectors, Matrices)	Relevance ranking via similarity (e.g., cosine)	Struggles with synonymy (lexical mismatch)
Probabilistic	Okapi BM25	Statistical Inference / Bayes' Theorem	Excellent baseline effectiveness with sophisticated term weighting	Requires relevance feedback; suffers from sparsity issues at a massive scale
Semantic / Probabilistic	PLSA / LDA	Mixture Decomposition / Latent Class Model	Captures latent concepts; statistically principled approach	Computational complexity for decomposition

5.2 Scalability results using Approximate Nearest Neighbor (ANN) Search

The ANN search technology is mandatory for achieving practical latency in massive-scale dense retrieval systems. The use of ANN algorithms dramatically reduces query time, a finding critical for enabling real-time applications such as recommendation engines and dynamic RAG pipelines. The choice between specific ANN algorithms (e.g., high-recall HNSW versus memory-efficient LSH) constitutes an essential technical decision that directly dictates the system’s effectiveness and cost profile [14] [15].

Crucially, the inherent memory intensity of storing high-dimensional embeddings—a limiting factor for scaling BM25 indices as well—is significantly mitigated by Vector Quantization (VQ). VQ, through methods like Product Quantization, effectively compresses the index size, making large-scale vector databases economically feasible. Furthermore, the introduction of Streaming VQ enables immediate index updates, resolving a traditional latency bottleneck in dynamic CDEs and ensuring that the retrieval process consistently accesses the most current information [1].

Table 2. Architectural and Algorithmic Solutions for Scalability in Complex Digital Environments

CDE Challenge	Architectural Solution	Algorithm / Technique	Benefit
Massive Data Volume / Ingestion	Distributed Indexing / Sharding	MapReduce Framework	Horizontal scalability; fault tolerance
Semantic Matching Speed (Latency)	Dense Retrieval Pipeline	Approximate Nearest Neighbor Search (ANN) (HNSW)	Fast query processing in high-dimensional space
System Resilience / Maintainability	Microservices / Loose Coupling	Event-Driven Architecture	Independent scaling; asynchronous, non-blocking communication
Index Size / Computational Cost	Vector Embedding Storage	Vector Quantization (PQ, Streaming VQ)	Compression of high-dimensional vectors; real-time index updates

5.3 Efficiency Trade-Offs in Neural Ranking Architectures

The pursuit of effectiveness through deep neural models is fundamentally constrained by their substantial resource and energy consumption. Neural Information Retrieval introduces a complex trade-off that extends beyond simple execution time and into areas of long-term economic and environmental sustainability [8].

The deployment of these energy-hungry models necessitates a systematic re-evaluation of efficiency standards. The IR community is actively developing new metrics and benchmarking frameworks to holistically assess the cost-benefit ratio, focusing on dimensions like resource consumption and energy efficiency. Strategies such as multi-stage rankers and cost-aware probabilistic data structures are essentially highly sophisticated engineering

solutions designed to manage this intrinsic trade-off, aiming to extract high effectiveness gains while minimizing the corresponding expenditure of resources [2].

Table 3. The Efficiency-Effectiveness Trade-Off in Modern Retrieval Paradigms

Paradigm	Effectiveness Metric Focus	Typical Efficiency Challenge	Mitigation Strategy
Sparse Retrieval (BM25)	Keyword Recall / Precision	Memory-intensive for very large indices	Optimized inverted file structure; hardware optimization
Dense Retrieval (BERT / Embeddings)	Semantic Relevance (MAP / nDCG)	High computational cost for inference/training; large index size	Vector Quantization (PQ), Multi-stage ranking
Hybrid Retrieval	Optimized Blend (Semantic + Lexical)	System complexity; overhead of two-stage indexing	Learned weighting schemes and optimized fusion strategies

6 Discussion

6.1 Neural Ranking Models for Library Applications

Table 1 presents a critical comparison of neural ranking models with demonstrated applicability to digital library and institutional repository contexts. The models were selected based on four criteria: peer-reviewed publication in major information retrieval or natural language processing venues ensuring methodological rigor; availability of pre-trained models or documented training procedures enabling practical implementation; demonstrated effectiveness on academic or document retrieval tasks relevant to library applications; and computational feasibility for deployment in institutional settings. This selection specifically excludes proprietary commercial systems and models requiring training data or computational resources beyond typical library technical capabilities. Furthermore, the selection emphasizes models with documented performance on English-language academic corpora while acknowledging that multilingual retrieval capabilities vary significantly across models and represent an important area for future evaluation.

However, this selection is not exhaustive and reflects specific constraints rather than comprehensive coverage of all neural ranking approaches. Models requiring custom hardware, such as specialized tensor processing units were excluded despite potentially superior performance, as most libraries lack such infrastructure. Similarly, models requiring millions of labeled training examples were excluded as few libraries possess resources for generating such training data. This pragmatic selection approach prioritizes practical deployability over theoretical effectiveness, though it potentially excludes cutting-edge approaches that may become viable as technical capabilities evolve. Future work should reassess model selection as library technical infrastructure advances and as model training becomes more accessible through transfer learning and few-shot learning approaches.

Critical analysis of these models reveals several important considerations for library implementation:

- First, computational requirements vary dramatically, with cross-encoder models requiring orders of magnitude more inference time than dual-encoder approaches. This suggests that two-stage architectures using efficient first-stage retrieval followed by expensive reranking offer more practical deployment paths than attempting to apply complex models to entire collections.
- Second, training data requirements present significant barriers, as most models achieve optimal performance only when fine-tuned on domain-specific labeled data that libraries typically lack. Transfer learning from general-domain pre-training partially addresses this challenge, but performance gaps between in-domain and out-of-domain applications remain substantial.
- Third, model selection involves trade-offs between effectiveness and interpretability. Neural models achieve superior performance on aggregate metrics but provide little insight into why particular documents rank highly, potentially hindering user trust and making system debugging difficult. This opacity contrasts with traditional ranking functions like BM25, where scoring components have clear semantic interpretations.

Libraries implementing neural ranking must therefore consider whether effectiveness improvements justify reduced interpretability, particularly for applications where result explanation serves important information literacy objectives.

6.2 Scalability and Performance in Large-Scale Library Systems

Scalability remains a fundamental challenge as digital library collections grow to millions of items. Traditional inverted index structures enable efficient keyword search but face limitations when incorporating complex semantic features or dense neural representations. For example, approximate nearest neighbor search in high-dimensional embedding spaces, while enabling semantic similarity matching, requires specialized indexing structures such as hierarchical navigable small world graphs or product quantization that introduce trade-offs between search accuracy, index size, and query latency. Libraries implementing dense retrieval must critically evaluate whether the semantic understanding benefits justify the computational complexity and infrastructure costs compared to traditional approaches.

Real-time indexing presents additional scalability challenges for institutional repositories receiving continuous deposits of new scholarly works. Bin et al. (2025) examine streaming vector quantization for real-time index updates in large-scale recommendation systems, demonstrating approaches for incremental embedding computation without full reindexing. However, their focus on recommendation rather than retrieval raises questions about applicability to library discovery, where different evaluation metrics and user expectations apply. Libraries must assess whether the complexity of real-time neural index updates provides sufficient value over traditional batch indexing approaches that, while introducing some delay in new content availability, offer simpler implementation and more predictable resource requirements.

6.3 Data Heterogeneity and Format Complexity

Modern digital libraries encompass heterogeneous data types, including bibliographic metadata in various schemas (MARC, Dublin Core, MODS), full-text scholarly articles in multiple formats (PDF, HTML, XML), digitized historical documents with varying OCR quality, multimedia materials (images, audio, video), and structured research datasets. Each format requires specialized processing pipelines for feature extraction and indexing, and users expect unified search interfaces spanning all formats. This heterogeneity challenges traditional retrieval approaches designed for homogeneous text collections.

Furthermore, metadata quality varies substantially across and within collections. Legacy catalog records may contain minimal subject indexing while contemporary materials benefit from rich descriptive metadata, creating retrieval effectiveness disparities that disadvantage older materials. Digitized archival materials often have only collection-level descriptions rather than item-level metadata, limiting discoverability. These quality issues necessitate retrieval systems robust to missing or inconsistent metadata, suggesting that hybrid approaches combining metadata-based search with content-based retrieval (for example, through full-text indexing or image analysis) may prove more effective than approaches relying solely on curated metadata.

6.4 Query Understanding and Intent Recognition

Understanding user search intent remains challenging despite advances in natural language processing. Library users formulate queries ranging from precise known-item searches seeking specific works, to exploratory searches investigating unfamiliar topics, to complex research questions requiring synthesis across multiple sources. Traditional retrieval systems typically apply identical ranking algorithms regardless of query type, potentially mismatching system behavior with user needs. For instance, known-item searches benefit from high-precision ranking emphasizing exact metadata matches, while exploratory searches may benefit from diversity and novelty in result presentation.

Query ambiguity and under specification further complicate intent understanding. Users often express information needs using incomplete or ambiguous terminology, assuming systems will infer intended meaning from context. For example, a query for 'Washington' might seek materials about George Washington, the state of Washington, Washington D.C., or any of numerous other entities and topics. While query disambiguation techniques using knowledge graphs or search history can help resolve ambiguity, these approaches require substantial infrastructure and may introduce errors when disambiguation predictions are incorrect. Libraries must therefore balance the potential benefits of sophisticated query understanding against the risks of misinterpretation and the complexity of implementation.

7 Recommendations

Based on the critical synthesis of current research and practical considerations for library contexts, this study offers the following evidence-based recommendations for information professionals implementing or improving retrieval systems:

- Adopt hybrid retrieval architectures combining traditional keyword-based methods with neural semantic search rather than pursuing exclusively neural approaches. This strategy leverages the complementary strengths of each paradigm while mitigating weaknesses. For institutional repositories and digital libraries, implementing BM25 for first-stage retrieval followed by BERT-based reranking for the top results balances efficiency with semantic understanding. This approach proves particularly valuable for heterogeneous collections where some queries benefit primarily from keyword matching (for example, known-item searches with specific titles or authors) while others require semantic understanding (for example, conceptual searches across disciplinary boundaries). Libraries should critically evaluate if neural components provide sufficient effectiveness improvements to justify added complexity rather than assuming neural methods universally outperform traditional approaches.
- Implement federated search capabilities to enable unified discovery across distributed institutional repositories, digital library platforms, and external databases serving library users. However, recognize that federated search introduces substantial technical complexity requiring specialized expertise. Libraries lacking in-house technical staff should consider collaborative federated search platforms developed and maintained by library consortia rather than building institution-specific implementations. When implementing federated search, prioritize user experience through parallel query execution, timeout handling for slow repositories, and transparent indication of which collections were searched. Critical attention to result merging algorithms that account for different relevance ranking schemes across collections is essential for providing coherent result sets.
- Develop multilingual retrieval capabilities for libraries serving diverse linguistic communities or managing multilingual collections. Implementation should begin with an assessment of actual language coverage needs based on collection analysis and user demographics rather than attempting to support all languages. For high-resource languages with substantial user populations, multilingual language models like mBERT or XLM-R provide viable solutions. However, for low-resource languages, particularly indigenous languages in archival collections, these general-purpose models often perform inadequately. In such cases, libraries should consider developing targeted parallel corpora or partnering with community organizations to create linguistic resources rather than relying solely on commercial multilingual systems that may not adequately represent specialized vocabularies.
- Establish rigorous evaluation frameworks incorporating both system-centered metrics (precision, recall, mean average precision, and normalized discounted cumulative gain) and user-centered assessments through usability testing and user satisfaction surveys. System-centered metrics enable comparison with research benchmarks and tracking of system improvements over time, but they do not capture all aspects of search experience relevant to library users. User-centered evaluation should specifically assess different user populations, including novice searchers who may struggle with query formulation and expert researchers with specialized information needs. Libraries should critically examine whether their evaluation frameworks actually reflect institutional priorities and user needs rather than uncritically adopting metrics from information retrieval research that may emphasize different objectives.
- Invest in metadata quality improvement and standardization as foundational work supporting all retrieval approaches. While neural methods can sometimes compensate for poor metadata through content-based indexing, high-quality, consistent metadata improves effectiveness for both traditional and neural retrieval. For digitized archival materials and special collections, metadata enhancement through item-level description, subject indexing, and named entity tagging provides substantial retrieval benefits. However, metadata enhancement requires ongoing resource allocation and organizational commitment. Libraries should critically prioritize metadata improvement efforts focusing on high-use collections and materials where metadata gaps most severely limit discovery, rather than attempting comprehensive enhancement across all holdings.
- Develop organizational capacity for continuous learning and adaptation as retrieval technologies evolve. This includes establishing mechanisms for monitoring emerging research through selective engagement with information retrieval conference proceedings and journals, evaluating new technologies for applicability to institutional contexts through pilot projects and controlled experiments rather than wholesale system replacements, and implementing feedback loops incorporating search log analysis, user studies, and relevance

judgment collection to identify system weaknesses and improvement opportunities. Libraries should recognize that effective retrieval systems require ongoing refinement rather than one-time implementation, necessitating sustained technical and organizational investment.

8 Future Research Directions

Several promising research directions warrant investigation by the library and information science community. The integration of retrieval systems with large language models for conversational search and question answering presents opportunities for more natural and interactive information access interfaces. However, critical research is needed to address challenges of factual accuracy, source attribution, and information literacy implications when search results are mediated through generated text rather than presented as direct document links. Libraries implementing such systems must carefully consider how to maintain traditional values of authority, provenance, and verifiability in conversational search contexts.

Personalization and context-aware retrieval, long-standing information retrieval research themes, deserve renewed attention in library contexts. User modeling techniques leveraging search history, borrowed materials, and explicitly indicated preferences could enable more tailored discovery experiences. However, personalization raises important privacy considerations and potential concerns about filter bubbles limiting exposure to diverse perspectives. Libraries must critically examine whether personalization benefits justify privacy trade-offs and develop appropriate governance frameworks for responsible personalization implementation.

Accessibility of advanced retrieval technologies for diverse user populations requires systematic investigation. Much neural information retrieval research focuses on effectiveness for average users with moderate search expertise, potentially overlooking the needs of users with disabilities, varying digital literacy levels, or specific information-seeking challenges. Research examining how different retrieval approaches serve diverse user populations, including the design of appropriate evaluation frameworks and metrics for accessibility assessment, would advance more inclusive discovery systems.

Finally, the environmental impact of increasingly computationally intensive neural retrieval methods warrants critical examination. Training large language models consumes substantial energy, and inference costs for neural reranking scale with collection size and query volume. As libraries consider adopting neural retrieval approaches, research on energy-efficient model architectures, sustainable deployment strategies balancing effectiveness with environmental impact, and methods for assessing cost-benefit trade-offs would support more informed decision-making aligned with institutional sustainability commitments.

9 Conclusion

Information retrieval in complex digital environments presents multifaceted challenges requiring sophisticated technical solutions informed by a deep understanding of user needs and organizational contexts. The research synthesized in this report demonstrates that contemporary retrieval systems must transcend traditional keyword-matching paradigms to address semantic complexity, heterogeneous data sources, massive scale, and diverse user populations.

Ultimately, effective information retrieval in complex digital environments requires integration of advanced technologies with sound design principles, organizational capabilities, and commitment to serving user information needs. Success demands technical expertise, user empathy, and strategic vision recognizing that retrieval systems serve broader organizational missions of knowledge access and utilization.

Declaration of AI and AI-Assisted Technologies used in Manuscript Writing Process: Not Applicable

References

1. Bin, X., Cui, J., Yan, W., Zhao, Z., Han, X., Yan, C., Zhang, F., Zhou, X., Wu, Q., & Liu, Z. (2025). Real-time indexing for large-scale recommendation by streaming vector quantization retriever. No. arXiv:2501.08695. arXiv. <https://doi.org/10.48550/arXiv.2501.08695>
2. Clarke, C. L. A., Kolla, M., Cormack, G. V., Vechtomova, O., Ashkan, A., Büttcher, S., & MacKinnon, I. (2008). Novelty and diversity in information retrieval evaluation. Proceedings of the 31st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, 659-666. <https://doi.org/10.1145/1390334.1390446>

3. Devlin, J., Chang, M.W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics, 4171-4186. <https://doi.org/10.18653/v1/N19-1423>
4. Guha, R., McCool, R., & Miller, E. (2003). Semantic search. Proceedings of the Twelfth International Conference on World Wide Web - WWW '03, 700-709. <https://doi.org/10.1145/775152.775250>
5. Hambarde, K. A., & Proenca, H. (2023). Information retrieval: Recent advances and beyond. IEEE Access, 11, 76581-76604. <https://doi.org/10.1109/ACCESS.2023.3295776>
6. Jiang, P., Ouyang, S., Jiao, Y., Zhong, M., Tian, R., & Han, J. (2025). Retrieval and structuring augmented generation with large language models. Proceedings of the 31st ACM SIGKDD Conference on Knowledge Discovery and Data Mining V.2 (KDD '25), 6032-6042. <https://doi.org/10.1145/3711896.3736557>
7. Karpukhin, V., Oguz, B., Min, S., Lewis, P., Wu, L., Edunov, S., Chen, D., & Yih, W. (2020). Dense passage retrieval for open-domain question answering. Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP), 6769-6781. <https://doi.org/10.18653/v1/2020.emnlp-main.550>
8. Mitra, B., & Craswell, N. (2018). An introduction to neural information retrieval. Foundations and Trends in Information Retrieval, 13(1), 1-126. <https://doi.org/10.1561/15000000061>
9. Nogueira, R., & Cho, K. (2020). Passage re-ranking with BERT. No. arXiv:1901.04085. arXiv. <https://doi.org/10.48550/arXiv.1901.04085>
10. Singhal, A. (2012, May 16). Introducing the Knowledge Graph: Things, not strings. Google. <https://blog.google/products/search/introducing-knowledge-graph-things-not/>
11. Voorhees, E. M., & Harman, D. K. (2005). TREC: Experiment and evaluation in information retrieval. MIT Press. <https://aclanthology.org/anthology-files/pdf/J/J06/J06-4008.pdf>
12. Xiong, C., Power, R., & Callan, J. (2017). Explicit semantic ranking for academic search via knowledge graph embedding. Proceedings of the 26th International Conference on World Wide Web, 1271-1279. <https://doi.org/10.1145/3038912.3052558>
13. *The Ultimate Guide to Connected Common Data Environments*. (2025). FGM; 12d Synergy Pty Ltd. <https://www.12dsynergy.com/connected-common-data-environment-guide/>
14. *What is approximate nearest neighbor (Ann) search?* (2025). MongoDB. <https://www.mongodb.com/resources/basics/ann-search>
15. *What is approximate nearest neighbor (Ann) search in IR?* (2025). Zilliz; Milvus. <https://milvus.io/ai-quick-reference/what-is-approximate-nearest-neighbor-ann-search-in-ir>
16. *What is the difference between sparse and dense retrieval?* (2025). Milvus. <https://milvus.io/ai-quick-reference/what-is-the-difference-between-sparse-and-dense-retrieval>

ChatGPT as a Research Discovery Tool: An Evaluation Using Citations and Altmetric Evidences

Subhasmita Sankhua¹ [0009-0009-3202-6540] and M. Sadik Batcha² [0000-0002-8533-1073]

^{1,2} Department of Library and Information Science, Annamalai University, Chidambaram, India

{subhasmitasankhua27, batchadlis}@gmail.com

Abstract. Generative AI tools are being used extensively by researchers and students for various academic purposes. However, their ability for research discovery is still not well explored. This study attempts to investigate the ability and accuracy of ChatGPT as a research discovery tool, with reference to the retrieval of articles related to ‘AI hallucination’. The study involves a comparison of two data sets with 25 papers in each. While one set was selected from the Dimensions AI database, based on citation count, the other one was retrieved from ChatGPT, using a specific prompt. The comparison between these two datasets revealed significant differences in citation count and Altmetric score. It was observed that the first set has received 1633 citations in total, whereas the second set has only 98 citations. The mean Attention score is 50 and 1.16 for both sets, respectively. 68% of the papers suggested by AI are OA, whereas the proportion of OA in the first set is only 10%. Only a single paper overlap was found in both sets. Although 24 AI-recommended articles are real, still 64% of the papers contain incorrect author names, and one paper does not exist at all. The findings revealed that ChatGPT has recommended many papers that do not possess any citations and Altmetric score.

Keywords: AI hallucination; ChatGPT; Dimensions AI; Generative AI; Research Discovery

1 Introduction

Scientists and researchers document their work and disseminate it among the scholarly community. Students and authors often refer to existing literature to find evidentiary support for their own work. Not only this, but going through any domain literature gives an idea about the developments, advancements and discoveries associated with the field. For years, readers have been relying on the authentic databases, gateways, publisher sites, etc., to find relevant and reliable publications. However, technology has reshaped the way of information seeking and resource discovery among the scientific community which involves a shift from the old manual searching to the availability of electronically-backed tools. The most talked about technology of the present time is Artificial Intelligence.

The introduction of Artificial Intelligence (AI) has revolutionised all sectors, influencing people from all strata of society. Especially Generative AI, which can create texts, images, videos and much more, has been used by all [1]. One of the most popular Generative AI models is ChatGPT, which is now familiar to all types of users [2]. It has left its imprint in various fields, with substantial assistance to the users in diverse ways. Authors from multiple fields have come forward and demonstrated the possible integration of it in their respective fields [3]. For instance, there can be a possible implementation of ChatGPT in global warming, where it can be used in climate change prediction, model evaluation, and policy recommendations [4]. Possible use cases have also been proposed in the field of public health [5]. Moreover, it has also been considered for research discovery by various authors [6], [7].

Despite its multiple positive use cases, AI models come with some negative counterparts as well, like data fabrication, falsification, etc. [8]. ‘AI hallucination’ is a recently introduced term, which refers to the incorrect and misleading information generated by AI models caused by various reasons like biased and insufficient training data and inefficient assumptions [9]. The present study intends to examine the accuracy and ability of ChatGPT for research discovery in publications related to AI hallucinations.

2 Review of Literature

The use of AI, especially LLMs, has reshaped various aspects of scholarly investigations, including the ways of discovering and utilising the resources. Multiple studies conducted using these models have provided valuable insights into this topic. In their study, Walters and Wilder have attempted to evaluate the hallucinations associated with ChatGPT 3.5 and ChatGPT 4. Their study concentrated more on fabricated bibliographic citations. 55% of

GPT-3.5 citations were found to be fabricated, while ChatGPT-4 had 18% fabricated citations [8]. Another study by Glickman and Zhang has emphasised the usage of Generative AI tools for research discovery and summarisation. Highlighting the abductive research discovery capabilities of LLMs, their study reveals the improvement of AI tools for summarising and extracting key points over time [6]. Despite several studies conducted in the present area, the assimilation of Altmetrics and the accessibility aspect is still not well explored. The present study aims to fill the same gap.

3 Research Questions

RQ 1: To what extent do the ChatGPT-recommended papers on AI hallucination align with the set of established scholarly works?

RQ 2: Is there an accessibility bias in AI-led research discovery?

RQ 3: How accurately does ChatGPT provide the bibliographic metadata?

4 Methodology

This study attempted to compare two sets of papers in the domain of AI hallucination, published in the year 2025. For obtaining the relevant data set, two data sources have been selected: Dimensions and ChatGPT. Dimensions is a comprehensive database that provides an exclusive collection of scholarly work, enriched with citations and Altmetric Attention Score [10]. A controlled and limited search query was used to retrieve the relevant results related to AI hallucination: ("AI Hallucination*" OR "Artificial Intelligence Hallucination*" OR "LLM Hallucination*" OR "Large Language Model Hallucination*" OR "AI confabulation" OR "GenAI hallucination*" OR "hallucinating AI"), for the year 2025 and the publication type was limited to only articles and review articles. 223 publications were identified initially which were again rearranged according to the descending order of their citations and 25 articles were chosen after manual review. To ensure the comprehensiveness and to minimise selection bias, the primary search query was executed in an additional database, which is OpenAlex. It is a free-to-access database that has a robust collection [11].

For the second set of papers, a natural language prompt was used to retrieve 25 key papers using ChatGPT 5.2 [12]. The query is: List the 25 most influential and key research articles or review articles related to "AI hallucination" or "LLM hallucination" or "GenAI hallucination", published in the year 2025, based on your internal knowledge and data. Provide the Full Title, first author, DOI, and Source title of those publications.

The analysis was done using MS Excel. The complete flow of the study is presented in the figure1.

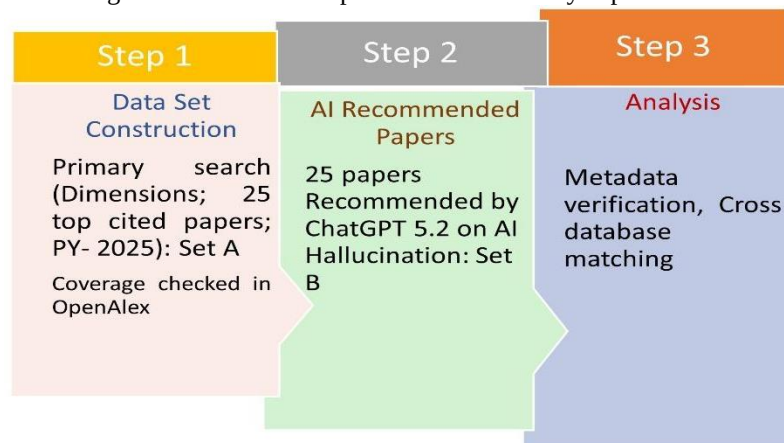


Fig. 1. Flow of the study

5 Results

Table 1. Comparative summary of Impact Indicators

Metric	Dimensions (Top-Cited Set) Set A	ChatGPT (Recommended Set) Set B
No. of papers	25	25

Total citations	1633	98
Mean Citations	65.32	3.92
Total AAS	1250	29
mean AAS	50	1.16
OA	10 (40%)	17 (68%)

AAS: Altmetric Attention Score, OA: Open Access

Table 1 highlights the state of some selected parameters associated with both sets of papers. The data shows a clear contrast between the key papers identified from dimensions on the basis of citations and the ones recommended by ChatGPT. A huge difference in the number of citations and AAS has been observed. The papers in Set A have received 1633 citations collectively, whereas the total citation count of the documents in Set B is only 98. Similarly, the total AAS for set A is 1250, and for set B, it is only 29. The difference among the means is also quite striking. The mean AAS for Set A is almost 50 times that of Set B. On the other hand, Set B consists of 68% OA papers, while in Set A, only 10% of the documents are OA.

Table 2: The Overlap Analysis

Paper Title	Author	citations	AAS
“My AI is Lying to Me”: User-reported LLM hallucinations in AI mobile apps reviews	Rhodes Massenon	6	5

The paper “My AI is Lying to Me”: User-reported LLM hallucinations in AI mobile apps reviews, with 6 citations and AAS 5, is identified as the singular overlap between both sets. The presence of a single common paper demonstrates the limited coalescence between the citation-based identification and an AI recommendation. This again speaks about the diverse selection mechanisms of LLMs, which do not align with scientific scrutiny. Moreover, this paper is OA, which again indicates the higher probability of OA publications being chosen. A 4% overlap suggests that LLMs do not majorly coincide with the scholarly metrics while identifying a seminal work in the field. On the other hand, papers with more citations and no AAS did not appear in this list, while this paper with a moderate number of citations and significant AAS has been selected anyway. Set B has 2 overlapped papers with the OpenAlex set, which include the above one and ‘HalluLens: LLM Hallucination Benchmark’.

Table 3. Metadata Integrity and Hallucination Assessment

Category	Frequency	Percentage
Valid Record (Exists in Dimensions or in another site)	24	96
Incorrect Attribution (Real paper, wrong author)	16	64
Non-Existent Record (Model hallucination)	1	4
Broken or incorrect DOI	1	4

The Metadata and hallucination assessment of the ChatGPT-recommended papers reveals evident reliability concerns. While a major proportion of articles (96%) could be accessed on either Dimensions or the publisher’s site, about 64% of articles carried wrong author names, although they possessed the real DOI and title. Whereas 4% of papers fall under the non-existent category. One paper could not be identified either through DOI or title search, exemplifying the existence of model hallucination.

Table 4. Visibility classification of ChatGPT-recommended papers

Visibility Category	Criteria	No. of papers	%
High visibility	Both Citations and AAS present	5	20
Moderate visibility	Either citations or AAS present	9	36
Low/Invisible papers	Neither Citations nor AAS present	11	44

Table 4 highlights the visibility categories of Set B papers, where all 25 papers have been grouped under 3 categories. There are only 5 papers (20%) that possess both citations and AAS. Another 9 papers (36%) are categorised as moderately visible papers that have either the citations or the AAS. Whereas the rest 44% neither have any

citations nor any online attention. This signifies that nearly half of the AI-recommended papers lack both traditional scholarly engagement and public attention. This suggests that although AI is able to provide real papers, it does not really take into account the other essential parameters.

6 Discussion

The present study provides empirical and descriptive insights into the research discovery capabilities of ChatGPT compared with the metrics-based selection approach. Results suggest that ChatGPT does not select truly influential papers either in terms of citations or AAS. It generated a list of papers that are mostly OA with minimal citations and AAS. The findings also indicate the accessibility bias of ChatGPT, where it selects papers that are mostly free to read [13].

The metadata assessment exemplifies that although this version of ChatGPT identifies real works, it fails to locate and provide the correct associated metadata, which raises concerns. This was also observed in previous versions [14]. The previous versions of ChatGPT have been seen to produce fabricated results and data [15].

The limited overlap between the citation-based papers and the AI-recommended papers implicates the diverse selection strategies of AI models, which do not align with the scholarly consensus or with the online attention. However, ChatGPT can show improved discoverability and retrieve papers more effectively when provided with the abstract of the paper [16].

The papers recommended by ChatGPT for the present study do not represent the core literature of the domain. Although there are some instances where AI models have generated a good set of papers in the concerned fields, the present study does not align with those [17].

7 Limitation

The present study used the two data sources only - Dimensions and OpenAI ChatGPT5.2. The data was restricted to the year 2025, and only articles and review articles were included, obtained by the mentioned query and prompt.

8 Conclusion

This study critically examined the implementation and usefulness of ChatGPT as a research discovery tool, focusing primarily on publications related to AI hallucination. The findings illuminated the discovery capability of ChatGPT, with a drastic difference between the papers that are essential in a scholarly parlance and those that are suggested by the AI. The articles, recommended as key papers in Set B, possess low citation counts and Attention Score as compared to Set A. The difference in metrics is quite significant. Moreover, ChatGPT recommended those papers that are mostly free to read and are easily accessible, while completely ignoring their impact and importance. Although, on a positive note, the hallucination is limited, the occurrence of incorrect author names does make it unreliable. However, the assistance of ChatGPT as a sole tool for research discovery is not suggested.

Declaration of AI and AI-Assisted Technologies used in the Manuscript Writing Process: ChatGPT 5.2 was used for grammatical refinements.

References

1. "What is Generative AI? | IBM." Accessed: Dec. 27, 2025. [Online]. Available: <https://www.ibm.com/think/topics/generative-ai>
2. S. K. Singh, S. Kumar, and P. S. Mehra, "Chat GPT & Google Bard AI: A Review," in *2023 International Conference on IoT, Communication and Automation Technology (ICICAT)*, Jun. 2023, pp. 1–6. doi: 10.1109/ICICAT57735.2023.10263706.
3. D. Kalla, N. Smith, F. Samaah, and S. Kuraku, "Study and Analysis of Chat GPT and its Impact on Different Fields of Study," Mar. 01, 2023, *Social Science Research Network, Rochester, NY*: 4402499. Accessed: Dec. 27, 2025. [Online]. Available: <https://papers.ssrn.com/abstract=4402499>
4. S. S. Biswas, "Potential Use of Chat GPT in Global Warming," *Ann Biomed Eng*, vol. 51, no. 6, pp. 1126–1127, Jun. 2023, doi: 10.1007/s10439-023-03171-8.
5. S. S. Biswas, "Role of Chat GPT in Public Health," *Ann Biomed Eng*, vol. 51, no. 5, pp. 868–869, May 2023, doi: 10.1007/s10439-023-03172-7.

6. M. Glickman and Y. Zhang, "AI and Generative AI for Research Discovery and Summarization," Mar. 26, 2024, *arXiv*: arXiv:2401.06795. doi: 10.48550/arXiv.2401.06795.
7. M. M. Rahman and Y. Watanobe, "ChatGPT for Education and Research: Opportunities, Threats, and Strategies," *Applied Sciences*, vol. 13, no. 9, p. 5783, Jan. 2023, doi: 10.3390/app13095783.
8. W. H. Walters and E. I. Wilder, "Fabrication and errors in the bibliographic citations generated by ChatGPT," *Sci Rep*, vol. 13, no. 1, p. 14045, Sep. 2023, doi: 10.1038/s41598-023-41032-5.
9. "What are AI hallucinations?," Google Cloud. Accessed: Dec. 27, 2025. [Online]. Available: <https://cloud.google.com/discover/what-are-ai-hallucinations>
10. "Dimensions AI | The most advanced scientific research database." Accessed: Dec. 27, 2025. [Online]. Available: <https://www.dimensions.ai/>
11. "OpenAlex: The open catalog to the global research system | OpenAlex." Accessed: Jan. 25, 2026. [Online]. Available: <https://openalex.org/>
12. OpenAI, "2025 AI Hallucination Papers," ChatGPT.
13. I. Dergaa and H. B. Saad, "Artificial Intelligence and Promoting Open Access in Academic Publishing Intelligence artificielle et promotion de l'accès libre dans la publication académique," *LA TUNISIE MEDICALE*, vol. 101, 2023.
14. K. Metze, R. C. Morandin-Reis, I. Lorand-Metze, and J. B. Florindo, "Bibliographic research with large language model ChatGPT-4: instability, hallucinations and sometimes alerts," *Clinics (Sao Paulo)*, vol. 79, Jan. 2024, doi: 10.1016/j.clinsp.2024.100409.
15. M. Naddaf, "ChatGPT generates fake data set to support scientific hypothesis," *Nature*, vol. 623, no. 7989, pp. 895–896, Nov. 2023, doi: 10.1038/d41586-023-03635-w.
16. M. Yamasaki, Y. Tomiura, and T. Shimizu, "Investigation of ChatGPT Use in Research Data Retrieval," in *Leveraging Generative Intelligence in Digital Libraries: Towards Human-Machine Collaboration*, D. H. Goh, S.-J. Chen, and S. Tuarob, Eds., Singapore: Springer Nature, 2023, pp. 36–40. doi: 10.1007/978-981-99-8085-7_4.
17. S. Insuk *et al.*, "How Well Do ChatGPT and Claude Perform in Study Selection for Systematic Review in Obstetrics," *J Med Syst*, vol. 49, no. 1, p. 110, Sep. 2025, doi: 10.1007/s10916-025-02246-4.

Metrics and Statistical Applications in LIS

Mapping the Knowledge Structure of Robotics Research in India (2015-2024): A Scientometric Approach

Laishram Thoithoiba¹ [0009-0007-2422-9351] and Keisham Sangeeta Devi² [0000-0003-1125-4375]

^{1,2} Department of Library and Information Science, Manipur University, Imphal – 795003, India

{laishramthoithoiba777, sangeetakeisham}@gmail.com

Abstract. The study examined Indian robotics research output, comprising 1373 records indexed in the Web of Science database from 2015 to 2024, to understand India's growth rate, global share, emerging themes, co-citation networks, and collaborative patterns of research publications within the study domain. The study's findings revealed that India accounted for 15.9% of the annual growth rate, which was faster than the global rate (14.8%), and 2.37% of the global share during the period. Keyword analysis revealed that 'path planning', 'robotics', and 'robotic surgery' are the most dominant conceptual themes. Pandey, A., the most prolific cited author, and "Zadeh La, 1965, Inform Control, v8, p338, doi 10.1016/s0019-9958(65)90241-X", the most cited reference, were observed in co-citation analysis. Collaborative research publications secured 33.43% of international co-authorship, and the Indian Institute of Technology (IIT) system was the most influential, contributing across India. The study provided a clear and comprehensive mapping of the conceptual, intellectual, and social structures of robotics research in India over the last decade.

Keywords: Robotics, Robotics research, India, Scientometric analysis, Web of Science

1 Introduction

Robotic technology is rapidly advancing its application in various sectors, including manufacturing, construction, space exploration, warfare, healthcare, agriculture, and education. The term 'robotics' was coined and first used by the Russian-born American scientist and writer Isaac Asimov (1920 -1992). "Robotics is an interdisciplinary sector of science and engineering dedicated to the design, construction, and use of mechanical robots" [1]. The Robotics Institute of America (RIA) defines "A robot as a reprogrammable multifunctional manipulator designed to move material, parts, tools, or specialised devices through variable programmed motions for the performance of a variety of tasks" [2]. With the integration of artificial intelligence (AI) and machine learning (ML), robots can interact independently with their environment and perform more complex tasks [3].

The paper aims to systematically explore the knowledge structure of robotics research publications in India over the last decade (2015-2024), utilising scientometric techniques. The study analyses the growth rate of publications (global vs. India), authors' keywords, co-citation of cited references and cited authors' co-citation, and collaborative works among authors and institutions within the dataset retrieved from the WoSCC database. Such analyses help to examine the conceptual, social, and intellectual structure of research, which will aid in advancing future research in the study domain.

2 Review of Literature

Many scholars have conducted research on robotics from various perspectives. However, only a few researchers have carried out bibliometric or scientometric studies on robotics research. Among such studies, a comprehensive bibliometric analysis of Indian robotics research output from 2007 to 2016, highlighted that India's research growth rate at 24.84% is faster than the global average (7.71%) [4]. In a bibliometric study, Dadvar and Habibian [5] discussed global research in response to robotics, which focuses on robots designed for search, rescue, and disaster response in both natural and artificial disasters. Aghimien et al. [6] revealed that research in robotics and automation emphasises systems and designs of robots, construction and automation, industrial robots, and information systems and control of robots. A bibliometric review of soft robotics research revealed that the United States leads in publications and citations, followed by China and Italy [7]. Wang et al. [8] visualised the research landscape to understand research trends and hotspots in the field of robotics and surgery, utilising the Web of Science database. In a bibliometric and systematic literature review study, Manuela et al. [9] provided a compre-

hensive overview of workers' concerns and benefits towards the acceptance of human-robot collaboration in industry settings. Atzeni et al. [10] quantitatively analysed 64 scientific papers from the Scopus database on collaborative robots and their usage in the logistics 4.0 environment. Evjemo [11] discussed the role of humans and robots in smart factories, their relationship to Industry 4.0, and the promotion they make to connect technologies. The usefulness and effects of social robots in improving mental health in older adults have low internal and external validity [12]. In a study, Anwar et al. [13] reviewed 147 studies systematically to investigate educational robotics and its impact on the academic and social skills of young learners in various fields of STEM education. Basumatary et al. [14] analysed data on the topic of robotic technology adoption in libraries, finding a 12.93% annual growth in robotics research in libraries. Asemi et al. [15] discussed the application of expert systems, AI and robots in libraries for various library services and management. An analysis found that, over the decade from 2012 to 2021, the USA led in publication number, and China ranked highest at the institutional level on the publication of AI in libraries [16].

From the above literature review, it becomes clear that no researchers have conducted a scientometric analysis in this emerging field of robotics in India over the last decade. Therefore, this study highlights the need for conducting comprehensive scientometric research to map the structure of knowledge in the study domain using WoSCC database data, thereby addressing this research gap. The study findings will also be beneficial to researchers and key stakeholders in the field.

3 Objectives of the study

- To examine Global publications vs. Indian publications in robotics research during 2015-24.
- To analyse the author keywords for understanding the conceptual structure of the study.
- To determine author and institutional research collaborations to visualise the social structure.
- To analyse the co-citation of cited authors and cited references to comprehend the intellectual research structure.

4 Methodology and Data

The study employed a scientometric analysis to explore the structure of robotics research publishing during the last decade, from 2015 to 2024, as indexed in the Web of Science (WoS) database. The data for the study were retrieved on December 3, 2025, from the Web of Science Core Collection (WoSCC) databases. Title searches (TI) are the most accurate as they focus on concise, author-defined titles, enabling researchers to determine directly relevant documents and save time [17]. So, the title search (TI) technique was preferred for the study to ensure high precision in assessing relevant documents, using the keywords "robot*" or "robotic*" in the "keyword tag" to collect data for both global and Indian research outputs. The study included articles (Document Type) and excluded early access documents of this type, limiting the period to 2015-2024. A variety of tools, such as Biblioshiny (Bibliometrix R package, v 5.2.1; Aria and Cuccurullo, 2025), VOSviewer (v1.6.20; van Eck & Waltman, 2023), and Microsoft Excel (2019), are used to analyse and visualise the collected data [18,19].

4.1 Data collection

Initially, 57864 of global research output were retrieved using the refined search strategy: "robot*" (Title) OR "robotic*" (Title) and Article (Document Types) and Early Access (Exclude – Document Types) and 2015 or 2016 or 2017 or 2018 or 2019 or 2020 or 2021 or 2022 or 2023 or 2024 (Final Publication Year). Secondly, the search strategy is again refined as: "robot*" (Title) OR "robotic*" (Title) and Article (Document Types) and Early Access (Exclude – Document Types) and 2015 or 2016 or 2017 or 2018 or 2019 or 2020 or 2021 or 2022 or 2023 or 2024 (Final Publication Year) and INDIA (Countries/Regions) to collect the Indian publication record. The search resulted in 1373 records and was exported in (.txt) format from the WoS database and converted into (.csv) format for further analysis.

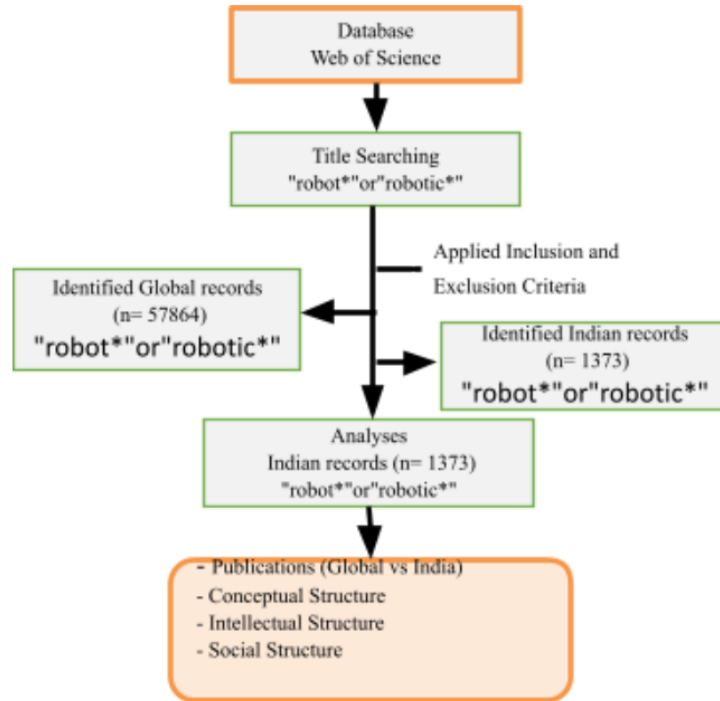


Fig. 1. Research framework

5 Analysis and Interpretation

A total of 57,864 records were identified in the Web of Science Core Collection (WoSCC) database for global publications in robotics research over the 10 years from 2015 to 2024. India accumulated 1373 records, a surge from 60 in 2015 to 227 in 2024. India shared 2.37% of global output during the years in the field. India secured a faster annual growth rate of 15.9% compared to the 14.8% growth rate of international research output in the field during the same years. The R2 value (0.9723) of global publications and the R2 value (0.9077) of Indian publications showed a strong and sustained growth of scholarly publications in both cases over the study time. (Table 1. Figure 1). India's five-year share increased from 2015-19 (2.16%) to 2020-24 (2.48%). However, its declining average citation per publication (ACPP) during the period 2020-24 may be due to a citation time lag. Overall, this analysis suggests that Indian research performance in robotics is moving positively towards the global research productivity pattern.

Table 1. Global and Indian Research Output in Robotics, 2015-2024

Publication Year	Global		India				
	TP	Cum. TP	TP	Cum. TP	TC	ACPP	%
2015	28 59	285 9	60	60	5 9	0.98	2.1
2016	31 61	602 0	59	119	6 0	1.02	3.7 6
2017	36 07	962 7	83	202	5 3	0.64	5.6
2018	41 01	137 28	82	284	5 3	0.65	6.9 3
2019	49 19	186 47	118	402	4 2	0.36	8.1 7
2020	58 70	245 17	113	515	4 4	0.39	8.7 7
2021	68 00	313 17	179	694	3 1	0.17	10. 21

2022	81 44	394 61	233	927	2 5	0.11	11. 38
2023	84 89	479 50	219	1146	2 7	0.12	13. 5
2024	99 14	578 64	227	1373	2 6	0.11	13. 85
2015-19	TP = 18647		TP = 402		2 67	0.66	2.1 6
2020-24	TP = 39217		TP = 971		1 53	0.16	2.4 8
2015-24	TP = 57864		TP = 1373		4 20	0.31	2.3 7

*TP = Total publication, Cum. = Cumulative, ACPP = Average citation per publication

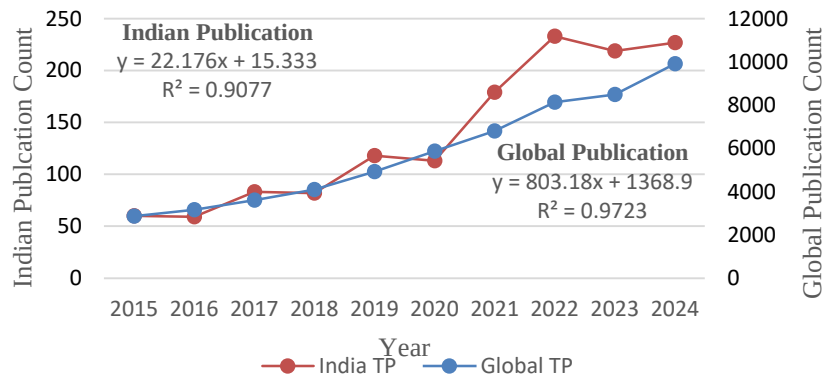


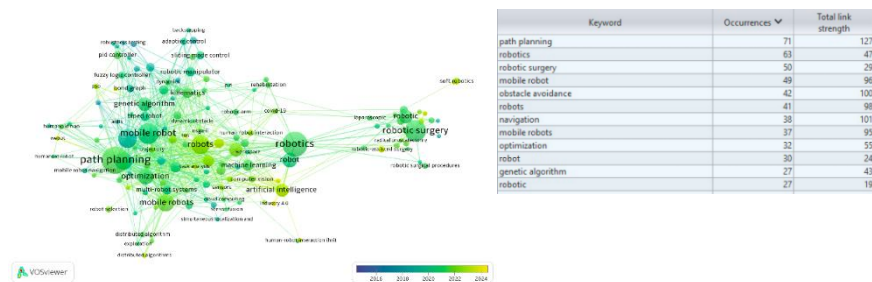
Fig. 2. India vs Global: Comparative Publications Growth Rates in Robotics, 2015-2024

5.1 Conceptual structure of robotics research

The conceptual structure of a research publication on a particular area can be assessed by analysing the keywords used by the author in their papers. The predominant keywords assigned by the authors in their paper reflect the focused theme of the study within a specific discipline [14]. Figure 3 visually displays the most prominent keywords used by the authors in robotics research. The VOSviewer software is used to visualise the keywords, employing a full-counting technique with a minimum keyword occurrence of 5. Out of 4793 keywords found in 1373 documents, 139 keywords met the threshold for analysis. The bubbles of different colours represented the prominent keywords, and each bubble size indicated the frequency of their appearance. The strength of co-occurrence links between nodes indicates the volume of research publications featuring the presented keywords. In the analysis, the keyword "path planning" is found to be the most frequently appearing keyword with 71 occurrences (OC) and 127 total link strengths (TLS), followed by "robotics" (OC=63, TLS=47), "robotic surgery" (OC=50, TLS=29), etc., as illustrated in the figure.

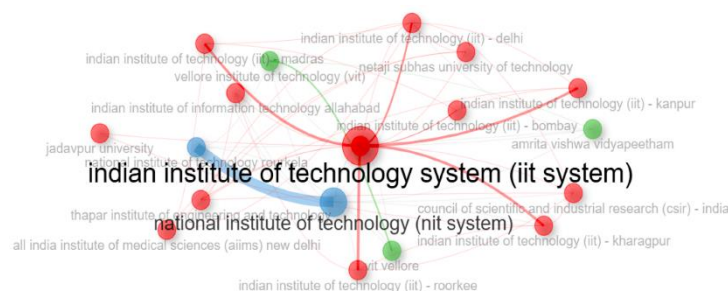
5.2 Social structure of research publications

The social structure of robotics research can be examined by analysing the collaboration between authors, organisations, and countries. This study examines the co-authorship patterns among authors and institutions to explore how researchers from different regions have collaborated on robotics research.



Collaboration networks of authors. Figure 4 displays a visualisation of the author collaboration pattern using the biblioshiny parameter. Circular node size represents the number of co-authors. Larger nodes show authors with a higher number of collaborative works with other authors within the dataset. Edge thickness denotes the strength of the collaboration between researchers. Different colours of the nodes and edges in the network exhibit distinct communities. Nodes of the same colour indicate clusters of researchers working closely on similar research topics. In the figure, 'Parhi, Dayal R.', 'Bhandari, Mahendra', 'Ahlawat, Rajesh', 'Rawal, Sudir', etc, are visibly the most contributed authors in the collaboration network. Authors such as 'Bhandari, Mahendra', 'Dasgupta, Prokar', 'Porpiglia, Francesco', and 'Abaza, Ronney' in the figure show the strength of the collaboration between authors. These authors are the key players who facilitate research across different subfields and institutions in the field. This network analysis reveals the strength and frequency of research collaboration of authors within the WoS dataset.

Fig. 4. Collaboration network of authors



5.3 Intellectual Structure

A co-citation analysis helps researchers explore intellectual connections between influential articles within a subject and facilitates visualising the intellectual research structure of a field. It can aid in identifying hotspots of study, highlighting the frontline of research, and also helps bring out high-impact, transformative discoveries [20, 21].

Co-citation analysis of cited authors. Figure 6 illustrates the network visualisation of cited authors in research publications on robotics research in India. The VOSviewer software is used to visualise the cited authors, employing a method of complete counting and a minimum of 10 citations per author. Out of 29860 authors, 94 authors have met the threshold for analysis. The selected top authors are identified based on their overall link strength and are represented by circles or bubbles of different colours. The largest circle nodes represent the most cited authors in the particular research publications. Pandey, A., is the most prolific cited author among the 29860 authors with 78 total citations (TC) and 658 total link strengths (TLS), followed by Mohanty, Pk. (TC = 74, TLS = 513), Parhi, Dr. (TC = 70, TLS = 655), etc., as shown in the figure. This analysis suggests a strong relationship between the two documents, likely due to their studies being closely related or sharing the same topic within a discipline. It is solely based on how frequently two papers are cited together.

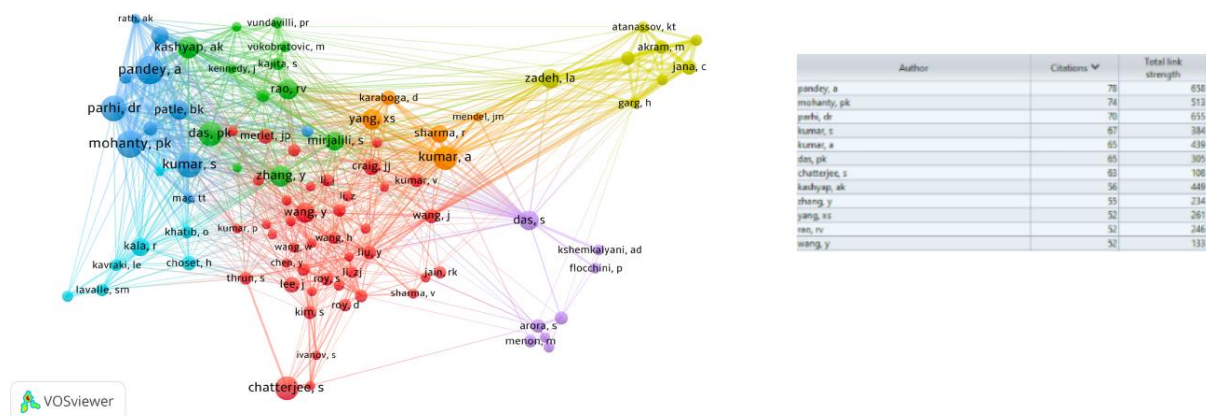


Fig. 6. Co-citation network of cited authors.

Co-citation analysis of cited references. The total number of 1373 documents has 42616 cited references. Out of the total cited references, 63 references meet the threshold for analysis. The visualisation (Fig. 7) is performed using a full-counting technique and includes a minimum of 10 citations of cited references. The figure shows that the article, "Zadeh La, 1965, Inform Control, v8, p338, doi 10.1016/s0019-9958(65)90241-X" is the most cited reference with 36 total citations (TC) and 60 total link strengths (TLS), followed by "Dindo D, 2004, Ann Surg, v240, p205, doi 10.17116/hirurgia2018090162, 10.1097/01.sla.0000133083.54934.AE" with 24 total etc., as shown in Figure 7.

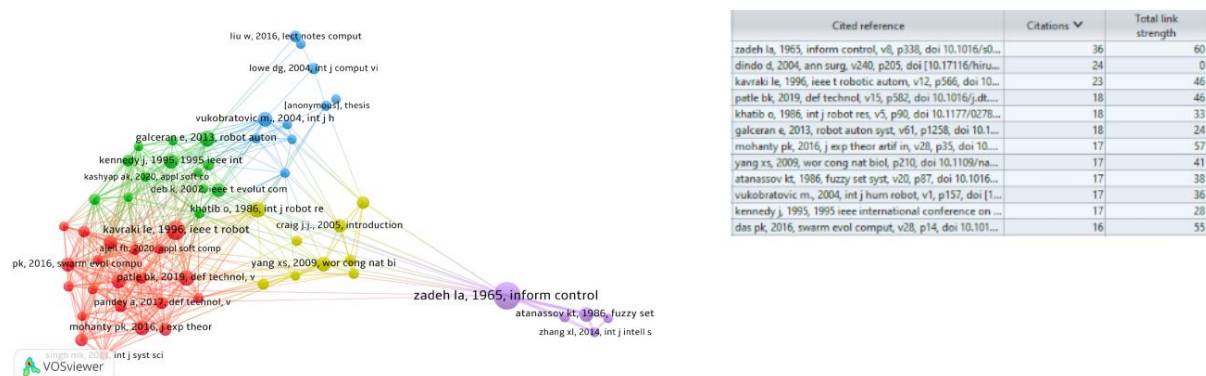


Fig. 7. Co-citation network of cited references.

6 Discussion and Conclusion

Initially, the study identified 57864 records in the WoSCC database after applying the inclusion and exclusion criteria to the search strategy for robotics research from 2015 to 2024. Of 57864 data, a total of 1373 data were exported after refining to the country, India, for the present analysis. Although 2.37% of India's share in global outputs, its annual growth rate (15.9%) is faster than the global average (14.8%). It suggests that the robotics research in India is closely paralleling the global publication trend. In the conceptual structure (Fig. 3), the appearance of keywords such as path planning, robotic surgery, motion control, and mobile robots reflects India's focus on robotic research in engineering applications, the healthcare sector, and autonomous systems over the last decade. Collaboration network of authors' visualisation (Fig.4) shows that 'Parhi, Dayal R.', 'Bhandari, Mahendra', 'Ahlawat, Rajesh', 'Rawal, Sudir', etc, are the most prolific authors in robotics research in the Indian context. Institutional collaboration analysis (Fig. 5) reveals the key role of the Indian Institutes of Technology (IITs) and National Institutes of Technology (NITs) in shaping the robotics research landscape in India. The limited participation of small or private institutions, such as VIT Vellore and Amrita Vishwa Vidyapeetham, suggests an unequal distribution of research capacity across India. The co-citation analysis of authors and references indicates intellectual linkages within the study domain. Pandey, A., Parhi, D.R., and Mohanty, P.K. emerge as the most cited authors with strong citation count and total link strength. Zadeh's (1965) work on fuzzy logic theory is the most co-cited reference, suggesting the ongoing significance of classical computational intelligence frameworks in modern robotics research. Overall, the analysis results demonstrate substantial growth and thematic diversification in Indian robotics research. However, India needs to emphasise high-impact research, balanced institutional participation, and maximise global collaboration to strengthen its performance on international platforms.

Over the last decade, India has made remarkable progress in robotics research output, outpacing the global average growth rate. However, the declining citation impact over the past five years suggests that India needs funding agencies and research institutes to ensure high-quality research and to encourage global collaborative research. The study not only highlights the current state of robotics research in India but also provides valuable implications for researchers, engineering professionals, medical professionals, and policymakers to comprehend research trends, identify gaps, and guide future advancements in robotics. The study is limited to Indian publications from 2015 to 2024 indexed in the Web of Science database. The search keywords "robot*" or "robotic*" can be used in the abstract or keyword tags to include relevant research papers that include robotics in the keywords or abstract. The study has not gone in-depth to trace the thematic evolution, nor has it conducted an altmetric analysis of the study domain, which would yield significant findings. Moreover, researchers can analyse data from different databases, such as Google Scholar, Dimensions, Scopus, etc., which will yield different results.

Declaration of AI and AI-Assisted Technologies used in Manuscript Writing: Not Applicable

References

1. Built In. (2024). Robotics: What are robots? Robotics definition & uses. Built In. Retrieved October 30, 2025 from <https://builtin.com/robotics>
2. Ajwad, S. A., & Iqbal, J. (2015, March). Emerging trends in robotics—a review from applications perspective. In International Conference on Engineering and Emerging Technologies (ICEET).
3. Licardo, J. T., Domjan, M., & Orehovac ki, T. (2024). Intelligent robotics—A systematic review of emerging technologies and trends. *Electronics*, 13(3), 542. <https://www.mdpi.com/2079-9292/13/3/542>
4. Gupta, B. M., & Dhawan, S. M. (2018). Robotics Research in India: A Scientometric Assessment of Indian Publications Output during. *Journal of Scientometric Research*, 7(2), 84-93. <https://doi.org/10.5530/jscires.7.2.14>
5. Dadvar, M., & Habibian, S. (2022). Contemporary research trends in response robotics. *ROBOMECH Journal*, 9(1), 9. <https://doi.org/10.1186/s40648-022-00221-z>
6. Aghimien, D. O., Aigbavboa, C. O., Oke, A. E., & Thwala, W. D. (2020). Mapping out research focus for robotics and automation research in construction-related studies: A bibliometric approach. *Journal of Engineering, Design and Technology*, 18(5), 1063-1079. <https://www.emerald.com/insight/1726-0531.htm>
7. Bao, G., Fang, H., Chen, L., Wan, Y., Xu, F., Yang, Q., & Zhang, L. (2018). Soft robotics: Academic insights and perspectives through bibliometric analysis. *Soft robotics*, 5(3), 229-241. <https://doi.org/10.1089/soro.2017.0135>
8. Wang, H., Zhang, J., Xie, N., & Liu, J. (2025). Visualised knowledge map of robot application in surgical field: a bibliometric analysis. *Journal of Robotic Surgery*, 19(1), 350.
9. Concetta Manuela, L. F., Antonio, G., Giada, L. S., & Rosa, M. (2025). Toward acceptance of human-robot collaboration in industrial settings: a bibliometric and systematic literature review. *The International Journal of Advanced Manufacturing Technology*, 139(5), 2139-2160.
10. Atzeni, G., Vignali, G., Tebaldi, L., & Bottani, E. (2021). A bibliometric analysis on collaborative robots in Logistics 4.0 environments. *Procedia Computer Science*, 180, 686-695. <https://doi.org/10.1016/j.procs.2021.01.291>

11. Evjemo, L. D., Gjerstad, T., Grøtli, E. I., & Sziebig, G. (2020). Trends in smart manufacturing: Role of humans and industrial robots in smart factories. *Current Robotics Reports*, 1, 35-41. <https://link.springer.com/article/10.1007/s43154-020-00006-5>
12. Araujo, B. S. D., Fantinato, M., Marques Peres, S., Melo, R. C. D., Batistoni, S. S. T., Cachioni, M., & Hung, P. C. (2022). Effects of social robots on depressive symptoms in older adults: a scoping review. *Library hi tech*, 40(5), 1108-1126. <https://doi.org/10.1108/LHT-09-2020-0244>.
13. Anwar, S., Bascou, N. A., Menekse, M., & Kardgar, A. (2019). A systematic review of studies on educational robotics. *Journal of Pre-College Engineering Education Research (J-PEER)*, 9(2), 2. <https://doi.org/10.7771/2157-9288.1223>.
14. Basumatary, B., Yuvaraj, M., Verma, N. K., & Verma, M. K. (2023). Mapping of the selected literature on robotic technology applications in libraries based on Scopus database: a subjective computational review. *Library Hi Tech*. <https://doi.org/10.1108/LHT-02-2023-0031>
15. Asemi, A., Ko, A., & Nowkarizi, M. (2021). Intelligent libraries: a review on expert systems, artificial intelligence, and robot. *Library Hi Tech*, 39(2), 412-434. <https://doi.org/10.1108/LHT-02-2020-0038>.
16. Borgohain, D. J., Bhardwaj, R. K., & Verma, M. K. (2024). Mapping the literature on the application of artificial intelligence in libraries (AAIL): a scientometric analysis. *Library Hi Tech*, 42(1), 149-179. <https://doi.org/10.1108/LHT-07-2022-0331>.
17. Gusenbauer, M. (2022). Search where you will find most: Comparing the disciplinary coverage of 56 bibliographic databases. *Scientometrics*, 127(5), 2683-2745. <https://doi.org/10.1007/s11192-022-04289-7>
18. Aria, M., and C. Cuccurullo. (2017). Bibliometrix: An R-Tool for Comprehensive Science Mapping Analysis. *Journal of Informetrics*, 11 (4), 959–975. <https://doi.org/10.1016/j.joi.2017.08.007>
19. van Eck, N.J. and Waltman, L. (2017). Citation-based clustering of publications using CitNetExplorer and VOSviewer, *Scientometrics*, 111:1053-1070. <https://doi.org/10.1007/s11192-017-2300-7>.
20. White, H.D. and Griffith, B.C. (1981). Author co-citation: a literature measure of intellectual structure. *Journal of the American Society for Information Science*, 32(3):163-171. <https://doi.org/10.1002/asi.4630320302>.
21. Calabretta, G., Durisin, B. and Ogliengo, M. (2011). Uncovering the intellectual structure of research in business ethics: a journey through the history, the classics, and the pillars of Journal of Business Ethics. *Journal of Business Ethics*, 104, 499-524. <https://doi.org/10.1007/s10551-011-0924-8>.

Thematic Evolution of Indian Library and Information Science Research: A Topic Modelling Analysis of DJLIT and ALIS

Subhayan Chakraborty^{1[0009-0004-6161-5102]} and Manoj Kumar Verma^{2[0000-0002-3009-3258]}

^{1,2}Department of Library and Information Science, Mizoram University, Aizawl, 796004, India,

subhayan@drtc.isibang.ac.in

manojdlis@mzu.edu.in

Abstract. Journals play a very significant role in shaping a domain's research landscape. The use of computational techniques like topic modelling has been growing to understand scholarly trends. This study aimed to assess the latent research themes in the last ten years (2015-2024) of India's two prominent LIS journals indexed in Scopus using LDA topic modelling. The journals are DESIDOC Journal of Library & Information Technology (DJLIT) and Annals of Library and Information Studies (ALIS). The study applied LDA on the titles and abstracts of the articles. The results reveal that DJLIT predominantly focused on ICT-driven and technology-oriented LIS research. Whereas, ALIS was largely metrics- and library service-oriented. The study also showed that metrics and research evaluation-based topics received closer academic attention. The study also indicated that the thematic similarity between the two journals was limited. It provides insights into the evolving intellectual landscape of Indian LIS research and offers valuable implications for researchers, editors, and policymakers.

Keywords: Topic modelling, LDA, DJLIT, ALIS, Indian LIS journals, Research trends

1 Introduction

Scholarly communication in Library and Information Science (LIS) has expanded in both volume and scope over the last few decades (Chang et al., 2015; Järvelin & Vakkari, 2022; Li et al., 2019). Additionally, the research themes have evolved. The field has moved from a dominant focus on library services and information retrieval toward areas such as scholarly communication, information seeking and behavior, social media, data science, and big data-driven information systems (Ma & Lund, 2021). Academic journals are the primary channels through which these developments are recorded and disseminated (Safón & Docampo, 2023).

Journals are considered a "reliable collective knowledge base" that communicates scholarly information (Gu & Blackmore, 2016). Journal-level analyses indicate disciplinary priorities and research directions in a domain (Bence & Oppenheim, 2004). Therefore, tracking how research themes within specific journals change helps clarify shifts in a research domain. Traditional qualitative or manual content analyses are rich in detail, though they have scalability limits (McGrath, 2012).

Topic modeling has become a prominent tool to address these challenges. It is a text mining technique that infers latent topics as groups of co-occurring words. It is used to map research topics and intellectual structures in large literatures (A. et al., 2023; Jelodar et al., 2019). Latent Dirichlet Allocation (LDA) is the most popular probabilistic topic modelling technique. It automatically uncovers hidden thematic structures within a text (Alghamdi & Alfalqi, 2015; Ambrosino et al., 2018). The present study aimed to apply LDA topic modelling on two prime Indian LIS journals, i.e., DESIDOC Journal of Library & Information Technology (DJLIT) and the Annals of Library and Information Studies (ALIS).

The DESIDOC Journal of Library & Information Technology (DJLIT) and the Annals of Library and Information Studies (ALIS) are the most prominent, Scopus-indexed LIS journals. DJLIT is a bimonthly LIS journal (Majumdar, 2023; Negi, 2021). It is published from India by the Defence Scientific Information & Documentation Centre (DESIDOC) (Kumar & Moorthy, 2011). It was originally launched in 1980 as the DESIDOC Bulletin. It evolved through the DESIDOC Bulletin of Information Technology and was renamed in 2008 as DJLIT (Suri, 2018). The main themes include ICT applications in libraries, library services, and bibliometrics (Batcha et al., 2025). ALIS is a quarterly journal in LIS (D. K. Gupta & Satija, 2023). It is published from India by CSIR-NIScPR (formerly CSIR-NISCAIR/INSDOC) (Sakhaka, 2024). Founded in 1954 as Annals of Library Science, it was renamed Annals of Library Science and Documentation in 1964. It took its current title in 2001 (Mahesh, 2017). Its main research themes include bibliometrics, ICT in LIS, user studies, and LIS education (Prieto-Gutiérrez & Segado-Boj, 2019). Both are long-standing, high-visibility Indian LIS journals.

Multiple studies have applied topic modeling on either global LIS journal sets or single LIS journals. On the other hand, studies have mostly relied on traditional methods, such as content analysis, citation analysis, and keyword analysis. Recent works have used LDA to compare DJLIT, ALIS, and the Journal of Scientometric Research, but primarily aimed to identify dominant topics and broad longitudinal trends. However, these studies do not incorporate the integration of topic-level temporal dynamics with citation impact. Addressing this gap would offer a comprehensive thematic evolution of Indian LIS scholarship at the journal level.

2 Literature Review

Several studies have been reviewed to have a comprehensive understanding of the trends in LIS research, methods applying topic modelling, and their application in mapping Indian LIS research.

2.1 Trend Studies in Indian LIS Research

Several Indian LIS trend studies used other techniques instead of topic modelling and found similar topic trends. (Dora & Anil Kumar, 2017) found bibliometrics/scientometrics as the most popular research area after analyzing 3,713 articles from 14 Indian LIS journals. The study adopted co-occurrence analysis, keyword frequency analysis, and the burst detection method. (N. Gupta & Chakravarty, 2021) performed a bibliometric mapping of 2,483 LIS documents in the Indian Citation Index. They found bibliometrics as the leading subject, followed by information technology, user studies, digital libraries, and library services. (Patra, 2021) mapped Library and Information Science (LIS) research in Indian universities via LISTA. The paper confirmed that bibliometrics is the most significant area.

2.2 Topic-Modelling of LIS: Comparative and Contextual Studies

Multiple studies applied topic modelling to uncover latent research themes in LIS. (Saha & Ghosh, 2023) conducted a cross-national topic-modeling of 734 articles from two Indian LIS journals (ALIS; DJLIT) and two international (College & Research Libraries; Liber Quarterly). The results showed increasing cross-cultural adaptation in research. (Figuerola et al., 2017) performed topic modelling on LISA records and found 19 high-level topics grouped into four areas (processes, information technology, library, and specific application areas). (Han, 2020) used topic modelling to uncover hidden patterns in LIS articles. The results showed information retrieval and bibliometrics analysis as persistently dominant and traditional library topics faded after 2000. (Kurata et al., 2018) applied LDA on 1,648 LIS full-text articles in five core journals. The study found 30 latent topics in LIS research. (Thakuria & Deka, 2024) analysed open access LIS journal articles using topic modelling and found that research on traditional topics is declining. Furthermore, it found that scholarly communication and information systems are dominant topics in open-access LIS journals.

2.3 Core Topic-Modeling Applications in Indian LIS

Recent studies have been increasingly using topic modeling to map Library and Information Science research in India. (Majhi & Mukherjee, 2024) applied LDA on three Scopus-indexed Indian LIS journals, i.e., ALIS, DJLIT, and JSCIRES. The study traced the rise and decline of topics in LIS. (Lamba & Madhusudhan, 2019b) attempted to uncover latent topics in DJLIT using LDA. They identified 50 core topics, with 26 unique ones. (Miyata et al., 2020) performed topic modelling on the full text of 1648 articles from five LIS journals. The study claimed that topic diversity has decreased over time, and the Internet has become the premise of research in LIS. (Lamba & Madhusudhan, 2019a) applied author–topic modeling to 393 full-text DJLIT articles, assuming one dominant topic per author and article. The study identified 16 core topics, matched with “subject expert” authors. (Lamba & Madhusudhan, 2018) conducted LDA on LIS theses in India’s Shodhganga ETD repository. The findings revealed five core thesis topics, such as information literacy, user studies, scientometrics, library resources, and library services.

2.4 Research Gaps

The review of the existing literature revealed the landscape of applying topic modelling on LIS articles, especially in the Indian context. However, a limited number of attempts have been made to make comparative topic analysis

of Indian LIS journals. Moreover, no studies have yet focused on combining bibliometric analysis and topic modelling to uncover latent topics in Indian LIS journals. The present study aimed to fill this gap by applying several bibliometric techniques and LDA topic modelling to comprehensively evaluate research themes and the evolution of two Indian LIS journals.

3 Objectives

- To analyze the chronological growth of publications and citations in DJLIT and ALIS in the last ten years (2015-2024).
- To identify the dominant research topics present in DJLIT and ALIS through Latent Dirichlet Allocation (LDA)-based topic modeling.
- To examine the evolution of research topics in both journals by analyzing topic-wise distribution in the last ten years.
- To compare the thematic focus of DJLIT and ALIS to understand their research priorities.
- To analyze topic-wise citation impact to understand scholarly attention.

4 Methodology

The present study applied LDA topic modeling to the titles and abstracts of articles published in DJLIT and ALIS in the last ten years (2015-2024). The purpose was to identify latent research themes and analyze their temporal evolution and citation impact. Figure 3 shows the methodological pipeline for the study.

4.1 Data Collection

The bibliographic data for the study were retrieved from the bibliographic database, Scopus. Two different search queries were used to prepare datasets for DJLIT and ALIS:

SRCTITLE ("DESIDOC Journal of Library and Information Technology") AND PUBYEAR > 2014 AND PUBYEAR < 2025

SRCTITLE ("Annals of Library and Information Studies") AND PUBYEAR > 2014 AND PUBYEAR < 2025

For each record, metadata such as title, abstract, publication year, authors, affiliations, keywords, and citation counts were extracted.

4.2 Data Preprocessing

Multiple text preparation and preprocessing steps were followed using Python. First, text in the title and abstracts was merged to prepare a consolidated textual dataset. Punctuations, URLs, special characters, numbers, etc., and noises were removed. Next, Lowercasing, Tokenization and Stopword removal were performed using NLTK. Finally, Words were reduced to their base form using WordNet lemmatization.

4.3 Topic Modelling and Visualization

The Gensim library was used to train LDA on the DJLIT and ALIS datasets. The coherence score (c_v) was calculated for each model. The purpose was to determine the optimal number of topics. The topic number with the highest c_v was selected for further analysis. For DJLIT, the number of topics was 5 (See Figure 1), while for ALIS, it was 10 (See Figure 2).

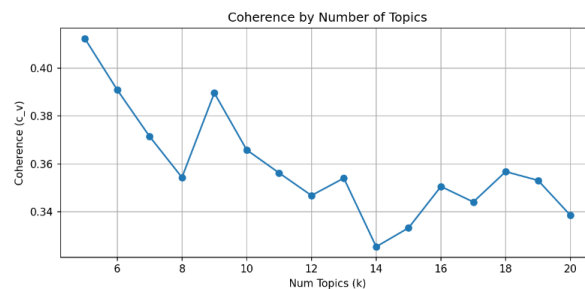


Fig. 1. Topic Coherence Scores for DJLIT

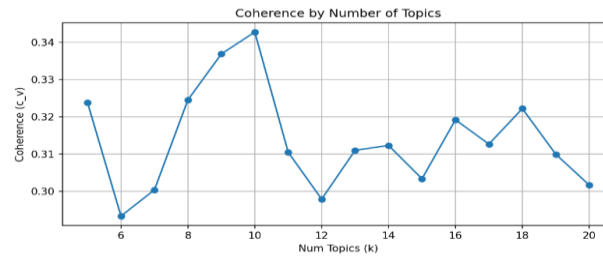


Fig. 2. Topic Coherence Scores for ALIS

Each topic was interpreted based on its top-ranked keywords. The study used pyLDAvis for interactive topic exploration. Matplotlib and Seaborn were used for different visualizations.

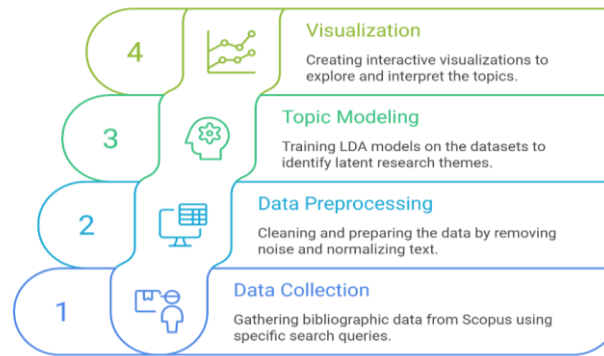


Fig. 4. Research Methodology Workflow

5 Data Analysis and Interpretation

5.1 Topic Distribution Overview

LDA identified five distinct topics from the corpus of DJLIT and ten distinct topics from the corpus of ALIS, respectively. Tables 1 and 2 illustrate the top twenty keywords for each topic. They also represent the topic labels assigned manually based on the keywords. The findings showed that ICT-based topics and technology-related topics dominated the corpus of DJLIT. Whereas the ALIS corpus was mostly metrics-oriented. Figures 3 and 4 show the topic-wise distributions in both DJLIT and ALIS. Figure 4 revealed that the most papers of DJLIT were published on topic 0 (Digital Libraries, Knowledge Management, and Research Practices). Similarly, topic 7 (User-Centric Library Services and Information Systems) was the most prominent research topic in ALIS (See figure 5).

Table 1. Top Five Research Topics in DJLIT

topic_id	top_keywords	Label
0	data, research, information, digital, library, knowledge, management, india, system, also, institution, paper, analysis, review, humanity, academic, based, open, plagiarism, literature	Digital Libraries, Knowledge Management, and Research Practices
1	e-book, library, use, reading, information, student, user, university, social, mobile, technology, online, resource, learning, science, patent, access, health, term	E-Resources, User Reading Behaviour, and Technology-Enabled Learning
2	library, service, university, information, resource, user, website, use, web, e, academic, professional, india, using, technology, collection, need, also, ict, tool	ICT-Based Library Services
3	student, information, research, university, library, among, respondent, data, per, used, cent, faculty, level, use, questionnaire, article, science, literacy, significant, scholar	Information Literacy and Library Use

4	research, journal, publication, citation, paper, author, science, article, published, per, analysis, university, country, india, year, output, global, indian, cited, cent	Scholarly Publishing, Citation Analysis, and Research Output Trends
---	--	---

Table 1. Top Ten Research Topics in ALIS

topic_id	top_keywords	Label
0	journal, article, citation, open, access, information, title, number, science, published, research, indian, indexed, university, publication, year, top, database, word, data	Open Access Publishing
1	book, library, paper, information, research, analysis, published, indian, subject, number, science, india, also, university, development, literature, impact, present, journal, national	Library Literature and Scholarly Communication in India
2	social, information, data, source, knowledge, field, change, research, term, rangathan, web, semantic, use, paper, open, present, medium, based, scientific, used	Knowledge Organization and Semantic Web
3	research, used, publication, data, lotka, paper, software, funding, law, education, library, literature, citation, scientist, authorship, factor, author, using, physic, doctoral	Scientometric Laws, Research Evaluation, and Funding Analysis
4	research, publication, citation, analysis, academic, paper, top, science, review, author, country, journal, e, social, university, researcher, global, impact, bibliometric, digital	Bibliometric Impact and Global Academic Research Trends
5	resource, e, university, indicator, growth, h, bradford, library, index, law, article, model, pattern, price, india, paper, citation, provided, group, based	Library Resource Distribution Patterns
6	paper, research, publication, science, author, collaboration, analysis, citation, india, classification, year, published, output, institution, productivity, highest, number, journal, scientometric, period	Research Collaboration and Institutional Productivity Analysis
7	library, information, service, user, university, system, student, model, use, based, need, resource, repository, also, collection, quality, knowledge, technology, paper, india	User-Centric Library Services and Information Systems
8	data, research, science, information, development, sustainable, paper, management, policy, open, lanka, sri, sharing, metric, knowledge, practice, resource, national, literacy, based	Information Policy and Knowledge Management
9	library, university, use, professional, student, service, digital, data, resource, india, li, research, e, questionnaire, academic, paper, analysis, librarian, used, found	Digital libraries and user studies

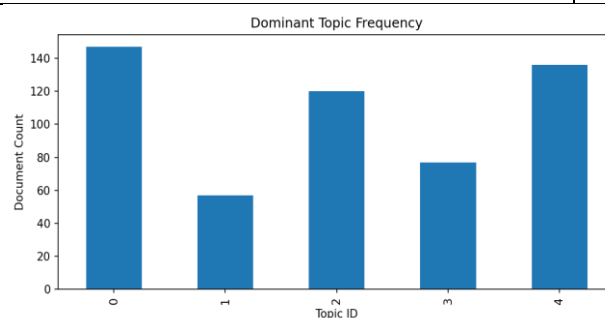


Fig. 5. Topic-wise Distribution of Papers in DJLIT

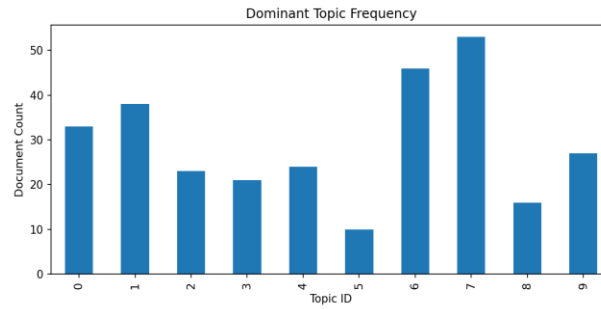


Fig. 6. Topic-wise Distribution of Papers in ALIS

5.2 Temporal Evolution of Topics

Two-line charts were constructed to examine the evolution of themes (See Figures 6 and 7). Figure 5 shows that topic 4 was the most consistent, topic 0 was the most emerging, and topic 2 was the most declining in DJLIT. Figure 6 shows that topic 1 declined the most, topic 6 was the most consistent, and topic 4 was the most emerging in ALIS.

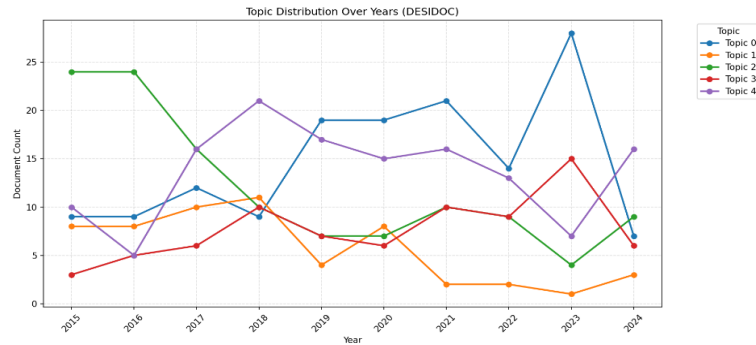


Fig. 7. Chronological Evolution of Research Themes in DJLIT

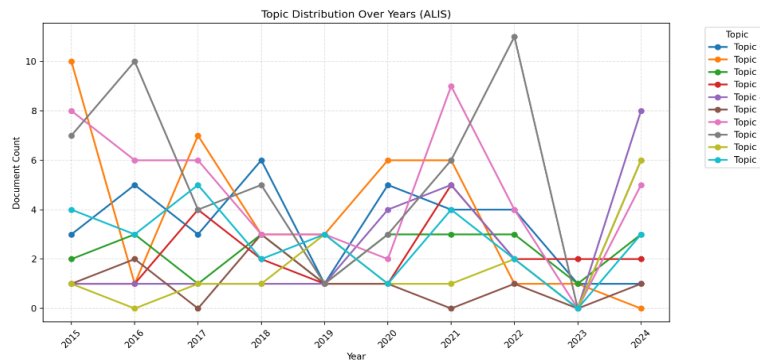


Fig. 8. Chronological Evolution of Research Themes in ALIS

5.3 Citation Impact across Topics

Citation analysis across topics revealed a variation in scholarly impact. Figures 8 and 9 show the average number of citations received by each research topic in DJLIT and ALIS, respectively. Figure 8 shows that among DJLIT topics, topic 4 (Scholarly Publishing, Citation Analysis, and Research Output Trends) received the highest mean citations. Whereas, topic 3 (Information Literacy and Library Use) showed the least mean citation. Similarly, Figure 9 illustrates that topic 4 (Bibliometric Impact and Global Academic Research Trends) received the highest mean citation among the topics in ALIS. Topic 8 (Information Policy and Knowledge Management) received the lowest mean citation. This indicates that metric studies are more academically visible than traditional library-oriented research topics in both journals.

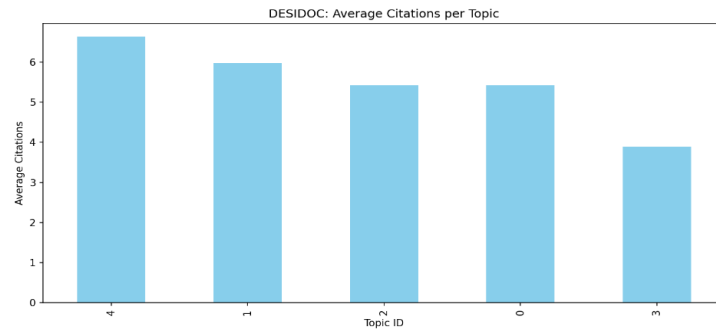


Fig. 9. Citation Impact across Topics in DJLIT

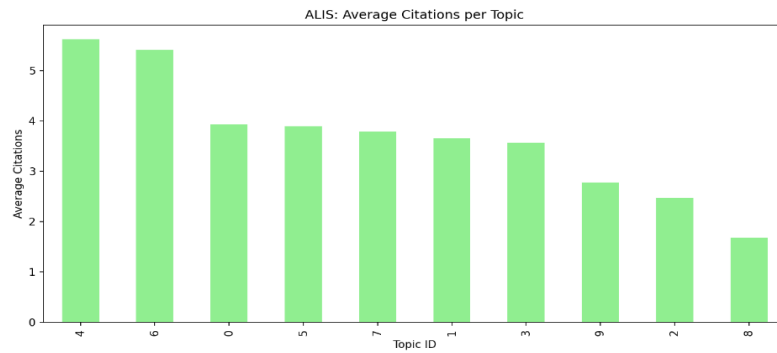


Fig. 10. Citation Impact across Topics in ALIS

5.4 Comparative Interpretation

The pattern of research topics and citations revealed distinct editorial and research orientations of the two journals. DJLIT primarily emphasized technology and metrics-based topics. It promoted data-intensive LIS studies. In contrast, ALIS mostly engaged with metrics and library-oriented topics. It addressed practical and institutional LIS concerns. Figure 10 represents a heatmap of the topic similarity of DJLIT and ALIS. The comparison was conducted using the cosine similarity of topics. The higher the similarity score, the more yellowish the colour of the cell. The figure illustrates that the topics of both journals were mostly distinct.

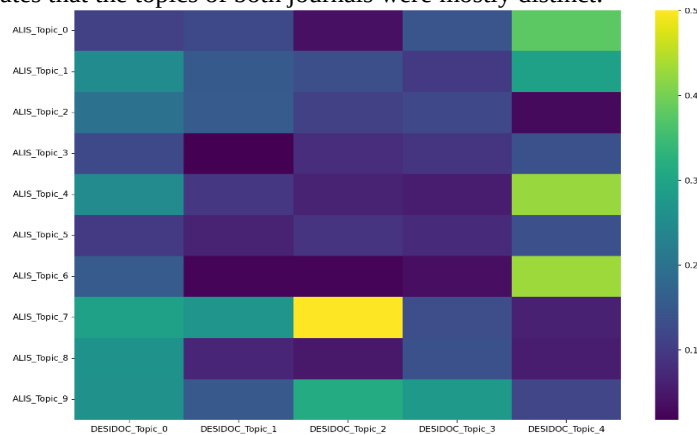


Fig. 11. Topic Similarity Heatmap of ALIS and DJLIT based on Cosine Similarity

5.5 Interactive HTML Visualization

The library, pyLDAvis, was used to explore the topics in an interactive HTML page (see Figures 11 and 12). Each circle represents one topic. The size of the circles represents the frequency of the topics. The position of the circles shows how different the topics are from each other. For each topic, the contribution of the top keywords can be explored. However, only the static views are represented here.

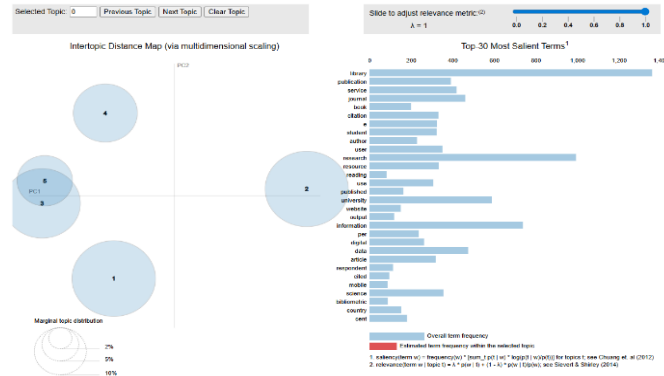


Fig. 12. Interactive HTML Page of Topic Distribution in DJLIT

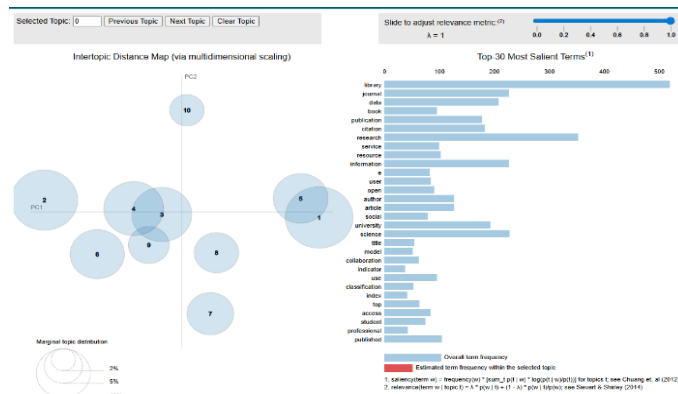


Fig. 13. Interactive HTML Page of Topic Distribution in ALIS

6 Discussion

The present study analyses the research topics and their trends in articles published between 2015 and 2024 in two prominent Indian LIS journals, i.e., DJLIT and ALIS. It applied Latent Dirichlet Allocation (LDA) to perform topic modelling on the titles and abstracts of those articles. The study found 5 core latent topics in DJLIT and 10 in ALIS. The findings revealed that the research topics in DJLIT were mostly surrounding technology and metrics. On the other hand, the orientation of research in ALIS was on library services, user studies etc. The most dominant topic in DJLIT was topic 0 (Digital Libraries, Knowledge Management, and Research Practices) (See figure 4). Figure 5 shows that topic 7 (User Centric Library Services and Information Systems) is the most prominent in ALIS. The analysis of temporal evolution revealed that traditional LIS concerns moved to more metrics-driven and analytical research in these two journals. The citation impact analysis further indicated the scholarly visibility of research themes. While Scholarly Publishing, Citation Analysis, and Research Output Trends received the highest average citation in DJLIT, Bibliometric Impact and Global Academic Research Trends received the highest in ALIS. This suggests that bibliometric and scientometrics studies were in commanding positions in these two journals. The comparative interpretation showed that despite both journals having focused on emerging LIS themes, their granular themes have been different. Moreover, topic visualizations were provided through pyLDA-vis.

7 Limitations

The study incorporates a few limitations. Firstly, it applied topic modelling only on the titles and abstracts of the articles. In depth analysis of the full text articles might capture more informed conceptual structures. Furthermore, the study focuses on two Indian LIS journals only. This might not make the findings generalizable to the entire LIS research landscape in India. Future studies are encouraged to address these limitations for a more comprehensive and robust understanding of Indian LIS research.

8 Conclusion

The present scenario of LIS research in India has been explored in this present study based on two Indian journals. The study applied LDA topic modelling to uncover the latent thematic structure of articles published in the DESIDOC Journal of Library & Information Technology (DJLIT) and the Annals of Library and Information Studies (ALIS) over the past ten years (2015-2024). The findings reveal clear thematic distinctions between the two journals. While DJLIT focused on technology-oriented and data-intensive studies, ALIS were predominant in metrics and library services. The results also highlight the shift in research objectives, digital transformation, and research evaluation practices. The results suggest that journals might focus on diversifying their scope. They can incorporate underrepresented research topics such as community engagement, indigenous knowledge systems, digital inclusion, ethical dimensions of AI in libraries, and sustainability-oriented LIS services. Furthermore, it was evident that metric-oriented topics attracted closer scholarly attention. The findings of the study contribute to an understanding of the evolving Indian LIS research practices. Moreover, it potentially guides future researchers, editors, and policymakers who aim to work on or for DJLIT and ALIS. The study suggests further extensions, including incorporating additional journals, conducting full-text analysis, or applying advanced topic modelling techniques.

Declaration of AI and AI-Assisted Technologies used in Manuscript Writing: Not Applicable

References

1. A., A. M. G., Robledo, S., & Zuluaga, M. (2023). Topic Modeling: Perspectives From a Literature Review. *IEEE Access*, 11, 4066–4078. <https://doi.org/10.1109/ACCESS.2022.3232939>
2. Alghamdi, R., & Alfalqi, K. (2015). A Survey of Topic Modeling in Text Mining. *International Journal of Advanced Computer Science and Applications*, 6(1). <https://doi.org/10.14569/IJACSA.2015.060121>
3. Ambrosino, A., Cedrini, M., Davis, J. B., Fiori, S., Guerzoni, M., & Nuccio, M. (2018). What topic modeling could reveal about the evolution of economics. *Journal of Economic Methodology*, 25(4), 329–348. <https://doi.org/10.1080/1350178X.2018.1529215>
4. Batcha, M. S., Jahina, S. R., & Ahmad, M. (2025). *Publication Trend in DESIDOC Journal of Library and Information Technology during 2013-2017: A Scientometric Approach* (Version 1). arXiv. <https://doi.org/10.48550/ARXIV.2511.04082>
5. Bence, V., & Oppenheim, C. (2004). The role of academic journal publications in the UK Research Assessment Exercise. *Learned Publishing*, 17(1), 53–68. <https://doi.org/10.1087/095315104322710278>
6. Chang, Y.-W., Huang, M.-H., & Lin, C.-W. (2015). Evolution of research subjects in library and information science based on keyword, bibliographical coupling, and co-citation analyses. *Scientometrics*, 105(3), 2071–2087. <https://doi.org/10.1007/s11192-015-1762-8>
7. Dora, M., & Anil Kumar, H. (2017). COLLNET: An empirical analysis of the research trends in the field of library and information science in India –2004-2015. *COLLNET Journal of Scientometrics and Information Management*, 11(2), 361–378. <https://doi.org/10.1080/09737766.2017.1317959>
8. Figuerola, C. G., García Marco, F. J., & Pinto, M. (2017). Mapping the evolution of library and information science (1978–2014) using topic modeling on LISA. *Scientometrics*, 112(3), 1507–1535. <https://doi.org/10.1007/s11192-017-2432-9>
9. Gu, X., & Blackmore, K. L. (2016). Recent trends in academic journal growth. *Scientometrics*, 108(2), 693–716. <https://doi.org/10.1007/s11192-016-1985-3>
10. Gupta, D. K., & Satija, M. P. (2023). Annals of Library and Information Studies: Some reflections and future directions. *Annals of Library and Information Studies*, 70, 52–58. <https://doi.org/10.56042/alis.v70i2.4093>
11. Gupta, N., & Chakravarty, R. (2021). Research visualization of Indian LIS research using VOSviewer and Bibliometrix. *Library Hi Tech News*, 38(8), 6–8. <https://doi.org/10.1108/LHTN-10-2021-0076>
12. Han, X. (2020). Evolution of research topics in LIS between 1996 and 2019: An analysis based on latent Dirichlet allocation topic model. *Scientometrics*, 125(3), 2561–2595. <https://doi.org/10.1007/s11192-020-03721-0>
13. Järvelin, K., & Vakkari, P. (2022). LIS research across 50 years: Content analysis of journal articles. *Journal of Documentation*, 78(7), 65–88. <https://doi.org/10.1108/JD-03-2021-0062>
14. Jelodar, H., Wang, Y., Yuan, C., Feng, X., Jiang, X., Li, Y., & Zhao, L. (2019). Latent Dirichlet allocation (LDA) and topic modeling: Models, applications, a survey. *Multimedia Tools and Applications*, 78(11), 15169–15211. <https://doi.org/10.1007/s11042-018-6894-4>
15. Kumar, M., & Moorthy, A. L. (2011). Bibliometric Analysis of DESIDOC Journal of Library and Information Technology from 2001-2010. *DESIDOC Journal of Library & Information Technology*, 31(3), 203–208. <https://doi.org/10.14429/djlit.31.3.989>

16. Kurata, K., Miyata, Y., Ishita, E., Yamamoto, M., Yang, F., & Iwase, A. (2018). Analyzing library and information science full-text articles using a topic modeling approach. *Proceedings of the Association for Information Science and Technology*, 55, 847–848. <https://doi.org/10.1002/pra2.2018.14505501143>
17. Lamba, M., & Madhusudhan, M. (2018, September 26). METADATA TAGGING OF LIBRARY AND INFORMATION SCIENCE THESES: SHODHGANGA (2013-2017). *ETD 2018: Beyond the Boundaries of Rims and Oceans Globalizing Knowledge with ETDs*. ETD 2018: Beyond the Boundaries of Rims and Oceans Globalizing Knowledge with ETDs, National Central Library, Taipei, Taiwan. <https://doi.org/10.5281/zenodo.1475795>
18. Lamba, M., & Madhusudhan, M. (2019a). Author-topic modeling of DESIDOC Journal of Library and Information Technology (2008-2017), India. *Library Philosophy and Practice*, 2593.
19. Lamba, M., & Madhusudhan, M. (2019b). Mapping of topics in DESIDOC Journal of Library and Information Technology, India: A study. *Scientometrics*, 120(2), 477–505. <https://doi.org/10.1007/s11192-019-03137-5>
20. Li, P., Yang, G., & Wang, C. (2019). Visual topical analysis of library and information science. *Scientometrics*, 121(3), 1753–1791. <https://doi.org/10.1007/s11192-019-03239-0>
21. Ma, J., & Lund, B. (2021). The evolution and shift of research topics and methods in library and information science. *Journal of the Association for Information Science and Technology*, 72(8), 1059–1074. <https://doi.org/10.1002/asi.24474>
22. Mahesh, G. (2017). Tour of Indian LIS through the pages of *Annals of Library and Information Studies (ALIS)*. *Library Herald*, 55(2), 127–139. <https://doi.org/10.5958/0976-2469.2017.00009.4>
23. Majhi, D., & Mukherjee, B. (2024). Analysing Library and Information Science Articles Using Topic Modeling Approaches. *DESIDOC Journal of Library & Information Technology*, 44(2), 114–123. <https://doi.org/10.14429/djlit.44.2.19312>
24. Majumdar, S. (2023). A nine-year (2012-2020) study of publication delay in an open access Library and Information Science journal. *Malaysian Journal of Library and Information Science*, 28(1), 35–50. <https://doi.org/10.22452/mjlis.vol28no1.3>
25. McGrath, J. M. (2012). Systematic and Integrative Reviews of the Literature: How Are They Changing Our Thoughts About Practice? *Journal of Perinatal & Neonatal Nursing*, 26(2), 193–195. <https://doi.org/10.1097/JPN.0b013e3182629c7d>
26. Miyata, Y., Ishita, E., Yang, F., Yamamoto, M., Iwase, A., & Kurata, K. (2020). Knowledge structure transition in library and information science: Topic modeling and visualization. *Scientometrics*, 125(1), 665–687. <https://doi.org/10.1007/s11192-020-03657-5>
27. Negi, D. S. (2021). Desidoc Journal of Library and Information Technology (DJLIT) (2016-2020): A Bibliometric Study. *Journal of Information Management*, 8(2), 83–90. <https://doi.org/10.5958/2348-1773.2021.00012.6>
28. Patra, S. K. (2021). *Library and Information Science Research in Indian Universities: Growth, Core Journals, Keywords and Collaboration Patterns* (Version 1). arXiv. <https://doi.org/10.48550/ARXIV.2107.13238>
29. Prieto-Gutiérrez, J. J., & Segado-Boj, F. (2019). Annals of Library and Information Studies: A Bibliometric Analysis of the Journal and a Comparison with the Top Library and Information Studies Journals in Asia and Worldwide (2011–2017). *The Serials Librarian*, 77(1–2), 38–48. <https://doi.org/10.1080/0361526X.2019.1637387>
30. Safón, V., & Docampo, D. (2023). What are you reading? From core journals to trendy journals in the Library and Information Science (LIS) field. *Scientometrics*, 128(5), 2777–2801. <https://doi.org/10.1007/s11192-023-04673-x>
31. Saha, M., & Ghosh, S. (2023). Topic Modelling in Library and Information Science from the Primary Data: Swing in Thrust Areas. *International Journal of Information Science and Management (IJISM)*, 21(3), 19–34. <https://doi.org/10.22034/ijism.2023.1977569.0>
32. Sakhaka, S. M. (2024). Content Analysis of Annals of Library and Information Studies during 2002 – 2022. *International Journal For Multidisciplinary Research*, 6(3), 1–12. <https://doi.org/10.36948/ijfmr.2024.v06i03.19059>
33. Suri, A. (2018). Commemorative Issue of DESIDOC Journal of Library and Information Technology on DRDO@60. *DESIDOC Journal of Library & Information Technology*, 38(1), 3. <https://doi.org/10.14429/djlit.38.1.12477>
34. Thakuria, A., & Deka, D. (2024). A decadal study on identifying latent topics and research trends in open access LIS journals using topic modeling approach. *Scientometrics*, 129(7), 3841–3869. <https://doi.org/10.1007/s11192-024-05058-4>

Mapping the Research Landscape of Artificial Intelligence in Academic Libraries: A Topic Modelling Analysis

Gulla Bhaskar ¹[0009-0002-7223-161X], Uday Pratap Singh ²[0009-0002-9172-468X], Sudhanshu Bajpai ³[0009-0007-5141-6283] and Anand Gupta ⁴[0009-0007-9623-0335]

¹ College of Defence Management, Secunderabad, India,

^{2,3} Indian Institute of Technology (BHU), Varanasi, India,

⁴ Indian Institute of Management, Bodhgaya, India,

thebhaskargulla@gmail.com

upsingh.lib@itbhu.ac.in

lakssbajpai@gmail.com

anand@iimb.ac.in

Abstract. This study tries to find out artificial intelligence (AI) applications in academic libraries through topic modelling and objective is to identify themes, patterns in the Library Science field. A corpus of 673 Scopus-indexed articles was analysed using Latent Dirichlet Allocation (LDA). The data was uploaded on Orange Data Mining (Version 3.39.0) (ODM). Prior to analysis text preprocessing was carried done. The analysis conducted through ODM gave ten distinct thematic clusters. Important themes include generative AI and large language models, machine learning, information literacy etc. Temporal correlation analysis shows a systematic decline in dominant topics ($r = -0.13$ to -0.23) alongside showing stability in peripheral themes, suggesting there is a shift in the field. This study seems to be the first AI research in academic libraries. It is trying to show how topic modelling applications remain limited in Library science, although there is widespread adoption in other disciplines.

Keywords: Artificial Intelligence, Academic Libraries, Topic Modelling, Latent Dirichlet Allocation, Machine Learning, Generative AI.

1 Introduction

Artificial intelligence (AI) has changed the shape of academic libraries. Since universities are moving toward digital teaching environments. Integrating AI into library services requires technical knowledge and skills to manage complex workflows effectively. Applications such as intelligent search, discovery tools and personalised user services have redefined the role of libraries and library functions. The adoption of AI in libraries have led to shift libraries from traditional repositories to actively supporting knowledge discovery and use.

1.1 How AI Has Shaped Academic Librarianship

AI has changed the face of academic librarianship across various fronts. AI search platforms have moved libraries from traditional catalogue based searching systems to personalised based access to resources. Due to Generative AI usage librarians' way of literature searching and research synthesis has changed. It has transformed librarianship in a drastic and constructive way. Helping end users in availing services at a faster pace.

Integration of AI in libraries has led librarians to learn new professional skills. Areas such as machine learning, natural language processing have gained tremendous response from the librarian community. This usage of technology tells us how librarianship has evolved over time. Working with AI is a need of the hour, else the library will in no time be turned into a storage house.

To fully understand these ongoing transformations, there has to be some systematic approach or methods which can process a large volume of literature and find patterns which one can not do with traditional techniques. Topic modelling and computational text analytics have effectively shown their effectiveness and strength in mapping research across disciplines. In applied sciences, topic modelling has examined argumentation structures in research articles (Tanev et al., 2024). In healthcare innovation, Cobelli and Blasi (2024) tells how combining bibliometric techniques with Latent Dirichlet Allocation (LDA) enables the discovery of patterns of technological adoption and research evolution. Similarly, Sonk and Tunger (2024) utilised Orange Data Mining for trend anal-

ysis in futures research, suggesting and guiding the potential of machine-learning tools in domain analysis. Scientometric studies have further added to the utility of topic modelling in mapping global research contributions (Natarajan et al., 2024).

Within library and information science (LIS), however, computational text mining remains comparatively underdeveloped or less utilised, with limited work having been done. Some studies have explored these approaches, but research in this area remains scattered and limited in scope. The existing LIS literature primarily focuses on descriptive narratives, tool-based discussions, or case studies of specific AI implementations. These works provide valuable insights but miss the point of how AI research has shaped academic librarianship.

This methodological gap is particularly important given the rapid growth of publications on artificial intelligence in library and information science. Traditional manual literature reviews are no longer sufficient to provide a comprehensive overview of the research domain. Topic modelling offers a systematic and scalable alternative, enabling the identification of key themes, the visualisation of relationships between topics, and finding emerging research areas. Orange Data Mining provides a user-friendly approach for researchers without programming knowledge, integrating preprocessing, LDA, multidimensional scaling, and visualisation tools, makes it ideal for LIS scholars (Sonk & Tunger, 2024).

This study analyses AI research in academic libraries using topic modelling. By applying LDA through Orange Data Mining to 673 Scopus-indexed documents published till November 2025, we try to: (i) identify dominant thematic clusters, (ii) find patterns and topic evolution, (iii) analyse inter-topic relationships, and (iv) find emerging research and underexplored areas. The findings will inform institutional policy, technology planning, professional development, and future research agendas.

2 Literature Review and Research Gap

Topic modelling is a powerful computational method for discovering latent thematic structures in large textual corpora across diverse disciplines. In applied sciences, Tanev et al. (2024) employed topic modelling to examine structures in management research, while Cobelli and Blasi (2024) demonstrated how combining bibliometric techniques with Latent Dirichlet Allocation (LDA) enables the discovery of technological adoption patterns in healthcare innovation. Sonk and Tunger (2024) successfully used Orange Data Mining for trend mining in futures research, showing a path for machine-learning tools for researchers without programming knowledge. Scientometric applications have further demonstrated the utility of topic modelling, with Natarajan et al. (2024) mapping academic support for Sustainable Development Goal 7. These cross-disciplinary applications demonstrate topic modelling as an analytical approach that shows validity across contexts and increasing reach through user-friendly usage.

In academic library contexts, recent research has begun mapping AI applications. Islam et al. (2025) revealed exponential growth in AI-library publications, from 8 papers in 2010 to 66 by 2022, identifying AI integration in library services, data mining, and user-based services as central themes. Practical implementations have focused on conversational AI systems. Chase (2024) finds out that libraries can develop chatbots with minimal technical knowledge or expertise. The emergence of large language models has shown new areas for reference services.

Topic modelling in other fields have effectively shown emerging technologies and interdisciplinary research. Entezami et al. (2025) used AI-assisted topic modelling to map media literacy research. He finds out how computational methods can reveal hidden patterns in evolving fields. In healthcare, Talebi et al. (2024) applied machine learning for predictive diagnostics, showcasing AI's analytical power in complex domains. Pachlore and Chakkarwar (2023) combined LDA with BERT-based models for opinion mining, illustrating the complementary strengths of traditional topic modelling and modern neural approaches.

Despite widespread adoption elsewhere, topic modelling applications within library and information science (LIS) remain limited. Verma and Yuvaraj (2023) analysed green library literature, while Majhi and Mukherjee (2024) examined Indian LIS journals and has given an idea how topic modelling can be implemented in libraries. These studies demonstrate feasibility but remain limited to specific sub-domains rather than comprehensive mapping of AI applications in academic libraries.

2.1 Research Gap

Although AI applications have been used in healthcare (Talebi et al., 2024), innovation management (Cobelli & Blasi, 2024), and futures research (Sonk & Tunger, 2024), LIS lacks a comprehensive topic modelling-based mapping of AI research in academic libraries. Existing LIS studies remain limited in scope in the context of AI in academic libraries, focusing on single sub-themes rather than the full thematic landscape. Specifically, current studies lack: (i) comprehensive thematic mapping of AI in academic libraries; (ii) detailed analysis of how AI

themes emerge and evolve; (iii) examination of inter-topic relationships through correlation and network mapping; and (iv) full utilisation of topic modelling approaches proven effective in other disciplines. This study addresses these gaps by providing a computationally grounded synthesis of global research on AI applications in academic libraries.

3 Methodology

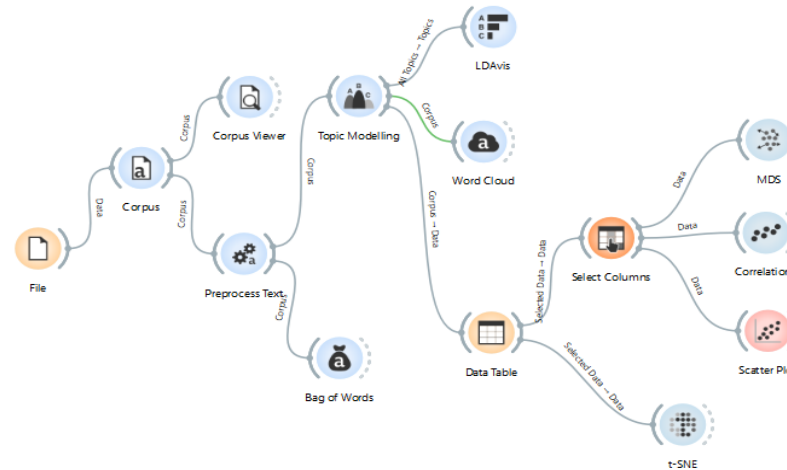


Fig.1. Orange Data Mining Structure

3.1 Research Design

This study uses a quantitative approach, using Latent Dirichlet Allocation (LDA) topic modelling to examine 673 documents on artificial intelligence in academic libraries. It combines text mining with bibliometric analysis and usage to provide comprehensive domain mapping

3.2 Data Collection

Data was downloaded from Scopus on November 26, 2025 in CSV format, using the search query: TITLE-ABS-KEY ("Artificial Intelligence" AND "Academic Librar*"). Scopus was used because it has broad coverage and high-quality indexing standards. It is noteworthy that Scopus is widely used in bibliometric and topic modelling studies, making the results comparable with existing research.

The "Librar*" wildcard includes library, libraries, and librarian. No restrictions were made on document type or publication year, yielding 673 documents. No restriction was made and from evolution till date AI theme was considered. Prior studies in LIS and allied domains have shown that datasets ranging from 300–1000 documents are sufficient to generate stable and interpretable topics. Therefore, the selected sample size is adequate to achieve the study's objectives.

3.3 Data Preprocessing

Data preprocessing is one of the important part for topic modelling. Metadata includes titles, abstracts, and author keywords, which were exported in CSV format. The preprocessing pipeline consisted of the following steps:

- i. Lowercasing: All text was converted to lowercase to ensure consistency. This also prevents the same word from being treated as different tokens due to capitalisation differences.
- ii. Stop word removal: Common English words (e.g., "the," "is," "and," "of") were removed using standard stop word lists to reduce noise and improve topic coherence and some were manually removed.
- iii. Punctuation removal: All punctuation marks, special characters, and numerical digits were removed from the text to focus the analysis on meaningful lexical content.
- iv. Tokenisation: The cleaned text was segmented into individual words (tokens), creating a bag-of-words representation.

3.4 Topic Modelling

Latent Dirichlet Allocation (LDA) was chosen because it is a well-established, and proven. LDA is a widely accepted probabilistic topic modelling technique for identifying hidden thematic patterns. One of the important features is that it represents each document as a combination of multiple topics, which is particularly appropriate for interdisciplinary areas. LDA generates probabilistic topic distributions that allow for clear and intuitive interpretation of both document–topic and topic–word relationships. This interpretability makes LDA especially suitable for exploratory analysis in Library and Information Science.

LDA was implemented using Orange Data Mining (v3.39.0), as it has user friendly interface, integrated workflows, and proven effectiveness (Sonk & Tunger, 2024). Multiple iterations were tested varying topic numbers ($k = 5, 7, 10, 12, 15$). The optimal 10-topic model was selected based on semantic coherence, minimal overlap, thematic coverage, and quantitative metrics: Log Perplexity (39.969), indicating a reasonable statistical fit, and Topic Coherence (0.616), reflecting moderate to good interpretability. The model selection was further validated using topic coherence and log perplexity measures, ensuring reliability. Also 10 gives better and distinct topics idea.

3.5 Methodological Justification: Use of Abstracts

This study has used abstracts, titles and keywords, instead of full texts for clear methodological and practical reasons. Full-text access to 673 articles from multiple publishers would require subscription fees which is not technically viable. Abstracts are readily available through Scopus metadata exports. Methodologically, abstracts are well-established proxies for full texts in bibliometric and topic modelling research. Abstract provides information about the articles in a concise way. It summarises key objectives, methods, and findings.

4 Results and Discussion

The LDA model achieved a log perplexity of 39.969 and a topic coherence of 0.616 by using ODM. The log perplexity value indicates reasonable statistical fitness, suggesting the selected number of topics has an acceptable balance between model complexity and predictive accuracy. The coherence score of 0.616 shows a moderate to good quality of topics, indicating that the extracted topics are consistent, meaningful, and suitable for qualitative interpretation in the context of AI research in academic libraries. Together, these metrics confirm the usage and reliability of the topic modelling results for thematic analysis in the context of AI in academic libraries.

4.1 Dominant Thematic Clusters

The LDA analysis using ODM provides ten distinct topics or themes representing major research themes in AI and academic libraries. Table 1 presents representative articles which have the highest probability score for each topic, demonstrating the model's ability to identify highly representative documents for each thematic cluster. (Table 1&2)

Table 1. Representative Articles for Each Topic

Topic	Representative Article	Year	Probability
Topic 1	Wisdom Study Open a New Era: Theoretical Innovation and Value Reconstruction of Library Science in China	2025	0.87
Topic 2	Exploring the Use of Generative Artificial Intelligence in Systematic Searching: A Comparative Case Study	2025	0.85
Topic 3	The Role of Artificial Intelligence in the Prediction, Identification, Diagnosis and Treatment of Perinatal Depression and Anxiety	2025	0.87
Topic 4	Document Clustering and Labelling for Research Trend Extraction and Evolution Mapping	2020	0.78
Topic 5	Navigating the Digital Landscape: A Survey of Technology Integration Among Academic Business Librarians	2025	0.85
Topic 6	Towards Digital Information Literacy Guidelines for African Libraries	2025	0.82
Topic 7	A Systematic Literature Review on Plant Disease Detection: Motivations, Classification Techniques, Datasets	2023	0.82

Topic 8	Impact of Conversational and Generative AI Systems on Libraries: A Use Case Large Language Model	2024	0.85
Topic 9	Artificial Intelligence in the Library: Gauging the Potential Application and Implications for Contemporary Library Services in Nigeria	2023	0.85
Topic 10	A 2010s-Era MLS Library Reference Assignment Meets 2025-Era AI: An Assessment of the Suitability of AI Tools	2025	0.82

Based on keyword analysis and examination of high-probability documents, the topics were interpreted as:

Table 2. Topic Interpretations and Keywords

Topic	Interpretation	Keywords
Topic 1	Machine Learning and Natural Language Processing Innovation	machine, innovation, language, natural, systematic, protocol
Topic 2	ChatGPT and Bibliometric Mapping	chatgpt, science, systematic, chatbots, mapping, bibliometric
Topic 3	Chatbot Technology and Diagnostic Applications	factors, systematic, innovative, chatbot, diagnosis, machine, techniques
Topic 4	Health Interventions and Healthcare Transformation	systematic, health, healthcare, interventions, evolution, patients, transforming
Topic 5	Ethical Systems and Evidence-Based Business Applications	systems, emerging, business, ethics, machine, integrity, evidence, advancements
Topic 6	Information Literacy and African LIS Development	literacy, perceptions, africa, lis, generative, south, policy, professionals
Topic 7	Intelligent Engineering and Classification Techniques	intelligent, engineering, practice, science, advancements, classification, techniques
Topic 8	Generative AI in Scholarly Communication	generative, conference, scholarly, prediction, strategies, exploration, search
Topic 9	User-Centered Smart Systems and Ethical Assessment	user, smart, potential, ethical, nigeria, public, identification, machine
Topic 10	Decision Support and Reference Services	era, perspectives, decision, search, reference, medical, scholarly, landscape

4.2 Visualisation Analysis: t-SNE and MDS

To visualise the thematic structure and inter-topic relationships, both t-distributed Stochastic Neighbour Embedding (t-SNE) and Multidimensional Scaling (MDS) were used to have better understanding. These models provide high-dimensional topic distributions into two-dimensional representations, enabling researchers to better understand visualisation of topic proximity and clustering patterns.

t-SNE Analysis. The t-SNE visualisation revealed distinct spatial clustering of related topics, with nodes coloured according to their probability scores ranging from 0-0.1 (dark blue) to 0.8-0.9 (yellow-green).

Topic 1: Machine Learning and Natural Language Processing Innovation appeared distributed across multiple regions with moderate concentration in the left-centre area, this showcase how its foundational role is spanning across multiple researches.

Topic 2: ChatGPT and Bibliometric Mapping and Topic 8: This tight clustering indicates strong thematic coherence around conversational and generative AI systems, confirming these topics represent the emerging paradigm of user-facing AI applications in the field of Library Science and services.

Topic 3: Chatbot Technology and Diagnostic Applications and Topic 4: Health Interventions and Healthcare Transformation are positioned adjacent in the right-centre region, their spatial proximity confirming complementary healthcare-oriented focus. Usage of AI in health care is clearly showcased.

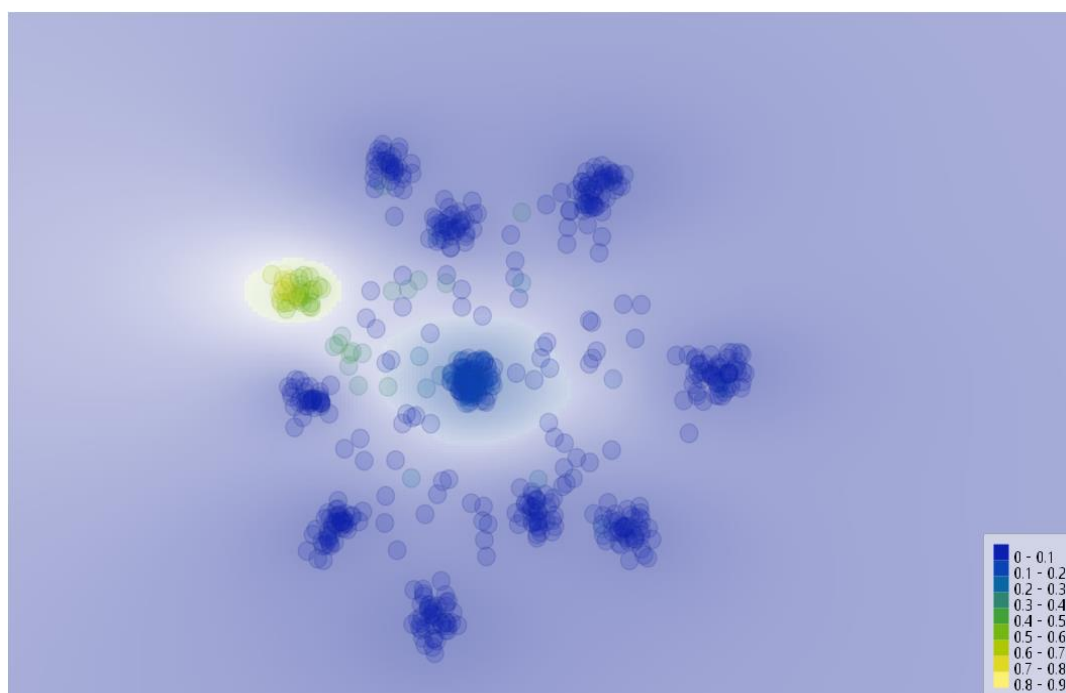


Fig.2. t-SNE of Topic-7 Distribution

Topic 5: Ethical Systems and Evidence-Based Business Applications occupied a middle position with connections to both technical and application-oriented clusters, suggesting its bridging role between technological implementation and organisational contexts. Focusses on ethics in usage of AI.

Topic 6: Information Literacy and African LIS Development appeared in the left peripheral region, its isolated position reflecting a specialised regional and professional development focus distinct from mainstream technical AI research. Country wide usage is showcased.

Topic 7: Intelligent Engineering and Classification Techniques occupied the most isolated position in the lower-left region, appearing in green-yellow colouring distinct from the blue cluster of other topics. This extreme spatial separation indicates that Topic 7 represents a fundamentally different research trajectory focusing on technical engineering applications with minimal thematic overlap with other AI-library research areas.

Topic 8: User-Centred Smart Systems and Ethical Assessment and Topic 10: Decision Support and Reference Services formed a cohesive cluster in the lower-right region, consistent with their shared focus on user-facing applications and service delivery. Basically it is user oriented ethical based topics.

MDS Analysis. MDS analysis confirmed the relationships and results observed in t-SNE. The MDS results showed that Topics 3, 4, 9, and 10 form a central themed cluster, which are very consistent with their high inter-correlations and gave emphasis on applied AI systems in library services. Topics 2 and 8 too maintained their proximity, being in the limelight as the forefront of generative AI research.

Most notably, Topic 7 remained spatially distant from all other topics in both visualisations, confirming its status as a distinct research trajectory. Topic 6 similarly occupied peripheral space, reflecting its specialised focus on regional LIS development and professional contexts.

The convergence of findings across t-SNE and MDS validates the robustness of the identified topic structure and confirms the presence of both highly interconnected core topics (3, 4, 9, 10) representing applied AI services, an emerging generative AI cluster (2, 8), and distinct peripheral research trajectories (6, 7) representing specialised domains within AI-library scholarship.

4.3 Word Cloud Analysis

Most of the prominent terms in word cloud are generative, systematic, machine, ChatGPT, literacy, science, and innovation which indicate a shift toward advanced generative AI applications and natural language processing. Health-related terms (health, medical, diagnosis, healthcare, and patients) highlight the interdisciplinary expan-

sion and relation into health sciences support, with chatbots and diagnostic applications emerging as key implementation areas. Methodological terms (bibliometric, mapping, scholarly, systematic, classification) reflect the growing use of AI for research evaluation and intelligent engineering approaches in the field of Library science. Geographic references (Africa, Nigeria, South) demonstrate global engagement beyond Western contexts, particularly in information literacy and LIS professional development. That has influenced other parts of the world as well. Ethics-related terms (ethics, ethical, integrity) underscore growing concerns about the responsible implementation of AI.

4.4 Temporal and Correlation Analysis

Spearman's correlation analysis reveals distinct patterns. The most dominant cluster among other clusters (Topics 3, 7, 9, 10) shows strong intercorrelations ($r = 0.50\text{--}0.56$) but have negative temporal correlations: Topic 7 ($r = -0.23$), Topic 3 ($r = -0.18$), Topic 10 ($r = -0.17$), Topic 9 ($r = -0.13$). This shows a decline in established paradigms such as chatbot diagnostics, user-centred design, and reference services.

Topic 5 (Ethical Systems and Business Applications) is serving as a bridge, exhibiting moderate correlations ($r = 0.48\text{--}0.51$) and a temporal decline ($r = -0.16$), suggesting transitional concerns about business ethics and the application of evidence-based practices. Topics 4 (Health Interventions) and 8 (Generative AI in Scholarly Communication) formed a secondary cluster ($r = 0.40\text{--}0.50$), with Topic 8 showing near-zero temporal change ($r = -0.09$), indicating stability in scholarly communication applications.

Peripheral topics (1, 2, 6) exhibited weaker intercorrelations ($r = 0.27\text{--}0.46$). Topic 6 (Information Literacy and African LIS Development) demonstrated a near zero temporal correlation ($r = -0.001$) with weak intercorrelations, representing an independent and stable trajectory. Topics 2 (ChatGPT and Bibliometric Mapping, $r = -0.08$) and 8 ($r = -0.09$) showed temporal stability.

4.5 Discussion

Paradigmatic Transition. The temporal-correlation analysis shows significant disciplinary bifurcation. The prominent cluster (Topics 3, 7, 9, 10) showed strong intercorrelations ($r = 0.50\text{--}0.56$) but there has been a significant decline: Topic 7 ($r = -0.23$), Topic 3 ($r = -0.18$), Topic 10 ($r = -0.17$), and Topic 9 ($r = -0.13$). This pattern shows that established research areas, such as chatbot diagnostics, user-centred systems, and reference services, are declining as foundational questions have been clearly addressed.

Where as peripheral topics demonstrated stability—Topic 6 ($r = -0.001$), Topic 2 ($r = -0.08$), and Topic 8 ($r = -0.09$) indicate a transition toward emerging applications in conversational AI, bibliometric mapping, and scholarly communication. There are considerable changes over time.



Fig.3. Word Cloud

Generative AI and Conversational Systems as Emerging Paradigms. Topics 2 and 8, focusing on ChatGPT, bibliometric mapping, and generative AI in scholarly communication, show temporal stability ($r = -0.08$, $r = -0.09$) and tight clustering in t-SNE visualisation, which represent emerging paradigms. Studies like "Exploring the Use of Generative Artificial Intelligence in Systematic Searching" (probability = 0.850) emphasise AI shift from chatbot applications to more sophisticated generative systems. There is a change in libraries from basic interfaces such as ChatGPT toward the integration of large language models into their usage.

Healthcare Applications and Methodological Isolation. Topics 3 and 4 are focusing on chatbot diagnostics and health interventions (probabilities = 0.871, 0.775) making libraries as contributing partners in healthcare.

Topic 7 spatial isolation in t-SNE visualisation reveals its fundamentally different nature. Despite focusing on intelligent engineering and classification techniques, this topic operates as a distinct domain from mainstream AI library applications can be seen in figure 2.

Global Literacy and Ethical Dimensions. Topic 6 high stability ($r = -0.001$) combined with keywords (literacy, Africa, LIS, South, Policy, Professionals) suggests that information literacy and professional development represent persistent challenges that transcend technological shifts, particularly in African contexts. AI implementation requires context-sensitive strategies that acknowledge infrastructural and educational disparities in libraries to address this.

Topic 5 focus on ethics, integrity this highlight concerns about the responsible adoption of AI in organisational contexts. Its moderate decline ($r = -0.16$) suggests that researchers are more aware than before.

Topic 9, with a declining trajectory ($r = -0.13$) and a focus on ethical assessment and public implications, may indicate that practice is advancing faster than governance structures, creating implementation-regulation gaps that require attention.

Machine Learning Innovation and Decision Support. Topic 1, with an emphasis on machine learning and innovation protocols (probability = 0.869), indicates ongoing foundational work in core AI technologies. Topic 10 focuses on decision support and reference services, showcasing AI tools as integral to librarian workflows but there is a decline. This highlights the need for a trained professional.

5 Limitations

This study has some limitations that should be considered when interpreting the results. First, the analysis is restricted to Scopus-indexed publications, due to which relevant work from non-indexed journals and conference proceedings can be rejected by default. Second, the focus on English-language publications may overlook important contributions from non-English-speaking regions, say in European countries. Third, the study relies on titles, abstracts, and keywords rather than full texts, which is mainly due to copyright and access constraints. Full-text analysis could provide deeper methodological and contextual insights. Finally, although the dataset includes publications up to November 2025, the fast-evolving nature of AI means that some emerging themes may not be included due to publication and indexing delays. These limitations contextualise the findings and highlight opportunities for future research refinement rather than undermining the study's validity.

6 Conclusion

This study has used Latent Dirichlet Allocation topic modelling which was implemented using Orange Data Mining to analyse 673 Scopus-indexed documents on artificial intelligence applications in academic libraries. The analysis provided a satisfactory result (log perplexity: 39.969; coherence: 0.616) and focused on ten distinct thematic clusters that represent the field's intellectual structure.

As per study and analysis, there were three most important patterns that were noticed which are:

- First, generative AI and large language models are emerging as a central research focus, showing a drastic shift from automation focused applications toward conversational systems in a more efficient and effective way. Topics such as ChatGPT and conversational AI demonstrate temporal stability.
- Second, temporal-correlation analysis revealed a decline in dominant topics alongside the stability of peripheral themes, indicating a movement from foundational questions toward specialised applications such as LLM.
- Third, interdisciplinary usage in health sciences, growing global research diversity, and increasing attention to ethical and professional dimensions indicate that AI research extends beyond technological implementation toward broader institutional and societal considerations. AI and ethics have been major discussion topics among researchers.

Library administrators or Librarians should be more focused and prioritise generative AI infrastructure and professional development for their staff. The decline of foundational topics suggests established implementation

models exist to inform strategic planning, which is the need of the hour. In today's rapidly evolving world, professional development must encompass technical competencies, critical evaluation skills, and ethical frameworks, as well as information literacy. Policy frameworks must have a balance between innovation and algorithmic transparency for better ethics during the research process, data privacy, and equitable access are too necessary topics.

7 Future Research Directions

Future research should employ dynamic topic modelling to track topic evolution across time. Comparative database analysis and multilingual studies would address coverage biases. Mixed-methods integration, such as combining computational text mining with qualitative expert interviews in the library science field, would validate findings and give a better idea while capturing practitioner perspectives not reflected in published literature. Inclusion of bibliometric analysis will also help in better understanding this area as it has majorly evolved seven years back.

Declaration of AI and AI-Assisted Technologies used in Manuscript Writing Process: Used Claude AI Sonnet 4.5 (Free Plan) in manuscript writing process for grammar and spelling check.

References

1. Casillano, N. F. B. (2022). Discovering sentiments and latent themes in educational technology. *International Journal of Information and Education Technology*, 12(4), 345-352.
2. Chase, M. (2024). Academic libraries can develop AI chatbots for virtual reference services with minimal technical knowledge and limited resources. *Evidence Based Library and Information Practice*, 19(2), 136-138. <https://doi.org/10.18438/ebliip30523>
3. Cobelli, N., & Blasi, S. (2024). Combining topic modeling and bibliometric analysis to map technological innovation in healthcare. *European Journal of Innovation Management*, 27(9), 127-149.
4. Entezami, Z., Davis, C. H., & Entezami, M. (2025). An AI-assisted topic model for media literacy research. *Media Literacy and Academic Research*, 8(1), 15-31.
5. Islam, M. N., Ahmad, S., Aqil, M., Hu, G., Ashiq, M., Abusharhah, M. M., & Saky, S. A. T. (2025). Application of artificial intelligence in academic libraries: A bibliometric analysis and knowledge mapping. *Discover Artificial Intelligence*, 5, Article 59. <https://doi.org/10.1007/s44163-025-00295-9>
6. Majhi, D., & Mukherjee, B. (2024). Analyzing LIS articles using topic modeling approaches. *DESIDOC Journal of Library & Information Technology*, 44(2), 114-123.
7. Natarajan, R., Verma, M. K., & Muthuraj, S. (2024). Mapping global academic support for sustainable development: A scientometric analysis. *Journal of Scientometric Research*, 13(1), 285-297.
8. Pachlore, A., & Chakkarwar, V. (2023). Opinion mining and tweet analysis using LDA and BERT-based models. In *ACVAIT 2022 Proceedings* (pp. 660-673). Springer.
9. Sonk, M., & Tunger, D. (2024). Trend mining with Orange data mining toolkit in futures research. *European Journal of Futures Research*, 12, Article 6. <https://doi.org/10.1186/s40309-024-00231-5>
10. Talebi, R., Azimian, A., Rahmani, K., & Afshar, S. (2024). Machine-learning-based classifiers for predicting metastasis in colorectal cancer. *Frontiers in Artificial Intelligence*, 7, Article 1285037.
11. Tanev, S., Keen, C., Bailetti, T., & Hudson, D. (2024). Topic modeling of management research: Analyzing argumentation structures in academic discourse. *Applied Sciences*, 14, Article 3277.
12. Verma, M. K., & Yuvaraj, M. (2023). Topic modeling analysis of green library literature: A computational approach. *Journal of Information & Knowledge Management*, 22(2), 97-104.

Scientometric Portrait of Bharat Ratna Prof. C.N. R. Rao: A study

Tejashwini B.¹, Ananda S.K.², Shiva Kumar H. C.³ and Akshatha A. K. ⁴

¹ RV Institute of Management, India
tejashwinib.rvim@rvei.edu.in

² EMPRI, Hasiru Bhavan, Bangalore, India
ananda.tthr@gmail.com

³RV Institute of Management, Bangalore, India
shivakumar.rvim@rvei.edu.in

⁴ RNS Institute of Technology, Bangalore, India
akshatha.ak@rnsit.ac.in

Abstract. The present study deals with Scientometrics analysis of 143 research publications by Prof. CNR Rao for the period of ten years i.e. 2015 to 2024. The study focuses on publication pattern over the years, authorship pattern, citation analysis, h-index and Altmetrics of Prof. CNR Rao obtained by official webpage of Jawaharlal Nehru Centre for Advanced Scientific Research (JNCASR). The study reveals that there is decrease in single authored articles i.e. 4.89% vs Multi authored articles 95.11%. Prof. CNR Rao has published highest number of articles in year 2015 with 20.27% and least number of articles in the year 2021 with 1.39%.

Keywords: Scientometrics, CNR Rao, Authorship pattern, Collaborations, Altmetrics, h-index.

1 Introduction

Scientometric portrait of scientists helps in identification of individual contribution to the field of the study. It also helps to identify the growth of research publications by year-wise, Country-wise and pattern of collaboration with other authors. A prominent scientist Prof. CNR Rao who is contributing to the field of Solid State Chemistry, Atomic Layer Deposition & Pulsed Laser Deposition, Two Dimensional Materials & Catalysis event at age of 91 years who has received Bharat Ratna has to be celebrated in every aspect of scientific research. His contribution to the field of Chemistry has inspired & will be inspiring many millions of Scientists, Researchers, students and even layman's in their day to day Life.

2 About C.N.R. Rao

C.N.R. Rao was born on 30 June 1934, Bangalore, India, received the M.Sc. degree from Banaras in 1953, Ph.D. from Purdue in 1958, D.Sc. from Mysore University in 1960. He is Honorary President and Linus Pauling Research Professor at the Jawaharlal Nehru Centre for Advanced Scientific Research.

He is also an Honorary Professor at the Indian Institute of Science. His main research interests are in solid state and materials chemistry. He is an author of over 1774 research papers and 54 books. He has received honoris causa doctorate degrees from 84 universities including Purdue. Prof. C.N.R. Rao was awarded the Bharat Ratna in 2014 for his significant contributions to chemistry and materials science, becoming the third scientist to receive India's highest civilian honour. Among the various medals, honours and awards received by him, mention may be made of the Marlow Medal of the Faraday Society (1967), Bhatnagar Prize (1968), Padma Shri (1974), Royal Society of Chemistry (London) Medal (1981), Padma Vibhushan (1985), Honorary Fellowship of the Royal Society of Chemistry, London (1989), Blackett Lectureship of the Royal Society (1991), Einstein Gold Medal of UNESCO (1996), Linnett Professorship of the University of Cambridge (1998), Centenary Medal of the Royal Society of Chemistry (2000), Hughes Medal of the Royal Society for original discovery in physical sciences (2000), Karnataka Ratna (2001), the Order of Scientific Merit (Grand-Cross) from the President of Brazil (2002) and the Somiya Award of the International Union of Materials Research (2004).

3 Objectives and Methodology of the Study

Bharat Ratna Prof. C.N.R. Rao's publications has been taken from official website of Jawaharlal Nehru Centre for Advanced Scientific Research for the Scientometric analysis which excludes reviews, book chapters, editorials

etc. concentrating only on research articles. Lots of Scientometric studies of Prof. C N R Rao have been carried out in the past i.e. 1956 to 1975 by Gadad & Ravi (2018), 1957-2013 by Ahirwal & Pathak (2015), from 2010-2017 by Gadad, Ravi, Savannuru (2019), Mulla et.al (2010) studies bibliometric analysis of Prof. C N R Rao literature from 1989 to 2008 to study research output, there is gap in earlier studies which can be observed in these articles, even our study is restricted for the period of ten years i.e. 2015-2024. The data from updated website has been taken till December -15th 2025.

The total numbers of research articles are 143 papers including 2 papers published in 2025. The data available from Jawaharlal Nehru Centre for Advanced Scientific research covers all the articles of Prof. C N R Rao along with Altmetric data and citations received by article i.e. year-wise. The past studies don't investigate these criteria of Altmetrics, updated citations, hence our study covers extensive articles from 1957 to 2025 for h-index, citations, google scholar data and Altmetrics.

The objectives of the Study are:

- To study year wise publications of Bharat Ratna Prof. C.N.R. Rao
- To analyse growth pattern of research publications
- To understand authorship pattern and degree of collaboration in publications
- To study citation and h-index of Bharat Ratna Prof. C.N.R. Rao
- To study Altmetrics of Bharat Ratna Prof. C.N.R. Rao publications.

4 Data Analysis and Interpretation

Table 1. Publication wise

Sl. No.	Publications	Total In Numbers	Percentage
1.	Research articles	1500	83.79
2.	Book chapter	82	4.58
3.	Conference proceedings	37	2.06
4.	Reviews	62	3.46
5.	Editorial	40	2.23
6.	Books	31	1.73
7.	Letters	17	0.94
8.	Notes	15	0.83
9.	Short survey	6	0.33
	Total	1790	100

Table -1 shows publication wise distribution of articles published by Prof. CNR Rao during the period of 1957 to 2025 i.e. 1790 and more than 1500(83.79) research articles, 82 (4.58) book chapters followed by conference proceedings with 37(2.06) articles. Reviews are 62(3.46) followed by editorial articles with 40(2.23), 31(1.73) books, followed by letters with 17(0.94), notes with 15(0.83) and lastly short surveys with 6(0.33) which supports that his major part of communications are research articles and less are short surveys.

Table 2. Year-Wise Publications

Sl. No.	Years	No. of. Articles	Percentage
1.	2015	29	20.27
2.	2016	23	16.08
3.	2017	17	11.88
4.	2018	19	13.28
5.	2019	16	11.18
6.	2020	14	9.79
7.	2021	2	1.39
8.	2022	8	5.59
9.	2023	8	5.59
10.	2024	5	3.49
11.	2025	2	1.39
	Total	143	100

Table-2 shows year-wise publications of Prof. C N R Rao during the period of 2015-2025. Highest number of articles were published during 2015 with 20.27% articles and lowest number of articles were published during 2021 with 1.39%.

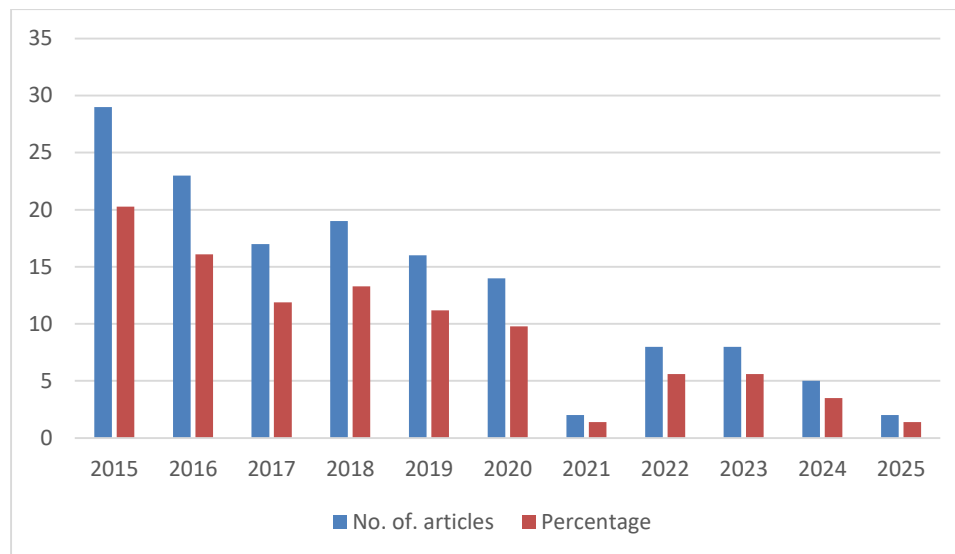


Fig.1. Year-wise publication of Prof. C N R Rao

Table 3.Citations & Cross reference Citations

Sl. No.	Years	Citations	Percentage	Cross Reference Citations	Percentage
1	2015	1607	32.99	1527	32.24
2	2016	1027	19.07	999	21.09
3	2017	929	19.07	941	19.86
4	2018	411	8.43	404	8.53
5	2019	486	9.97	450	9.50
6	2020	184	3.77	188	3.96
7	2021	34	0.69	34	0.71
8	2022	91	1.86	92	1.94
9	2023	77	1.58	76	1.60
10	2024	24	0.49	25	0.52
11	2025	0	0	0	0
	Total	4870	100	4736	100

Table-3 shows citations & cross reference citations received for articles by Prof. C N R Rao published during 2015 -2025. Highest number of articles i.e. 29 and highest number of citations i.e. 1607 are received in 2015, over the years there is decrease in number of citations. Least number of citations are received in 2024 & 2025.

Figure 2. Citation Analysis

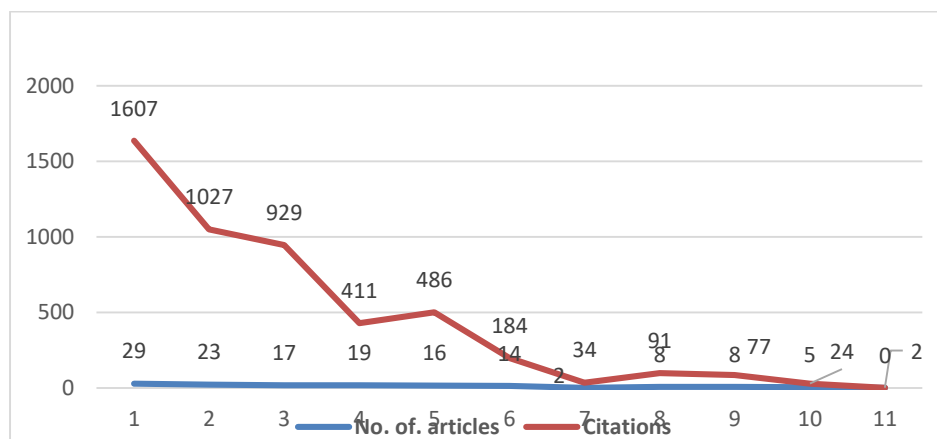


Fig.2. Citation analysis of articles published by Prof. CNR Rao during 2015 to 2025.

Table 4. Authorship Pattern

Year	1 Author	2 Authors	3 Authors	4 Authors	5 Authors	6 Authors	More Than 6 Authors	Total
2015	1	2	7	7	5	6	1	29
2016	1	2	3	6	5	5	1	23
2017	0	1	3	5	1	3	4	17
2018	1	1	1	4	4	4	4	19
2019	2	1	0	1	2	8	2	16
2020	2	1	1	6	1	2	1	14
2021	0	0	0	1	0	1	0	2
2022	0	2	1	3	1	1	0	8
2023	0	1	0	2	3	2	0	8
2024	0	0	1	2	1	1	0	5
2025	0	1	1	0	0	0	0	2
Total	7	12	18	37	23	33	13	143
Pert%	4.89	8.39	12.58	25.87	16.08	23.07	9.09	100

Table-4 shows authorship patterns of publications by Prof. C N R Rao during 2015 to 2025. Single authored articles are 4.89% and Multiple authored articles 95.11%. Among multi-authored articles 4-authored articles are highest with 25.87% and lowest are 2-authored articles with 8.39%.

Table 4. Degree of Collaboration

Sl. No.	Year	Single	Multi-authored	Degree of Collaboration
1	2015	1	28	0.96
2	2016	1	22	0.75
3	2017	0	17	0.58
4	2018	1	18	0.62
5	2019	2	14	0.48
6	2020	2	12	0.41
7	2021	0	2	0.06
8	2022	0	8	0.27

9	2023	0	8	0.27
10	2024	0	5	0.17
11	2025	0	2	0.06
	Total	7	136	0.95
	Percentage	4.89	95.11	

Table -5 depicts degree of collaboration in research articles published by Prof. C N R Rao during 2015 to 2025. The degree of collaboration is 0.95 for research articles published by Prof. C N R Rao. The degree of collaboration is calculated using the formula proposed by Subramanyam (1983) and has been computed as under:

$$C = Nm / (Nm + Ns)$$

C: Degree of Collaboration (value typically between 0 and 1)

Nm: Number of multiple-authored papers

Ns: Number of single-authored papers

Table 6. Citations

Sl. No	Type	Total In Numbers
1	Citations	106815
2	Cross –reference citations	96344
3	H - index	151

Table-6 shows citations data obtained from the vidwan webpage which depicts that for 1707 articles, 106815 citations were received from 1957 to 2025. 96344 cross reference citations were received which in total gives H-index of 151 for Prof. C N R Rao.

Table 7. Altmetrics

Sl. No	Type	Total in Numbers
1	Google scholar	283
2	Facebook	176
3	X(Twitter)	984
4	G+	193
5	Cross reference	69666

Table-7 depicts Altmetrics data of Prof. CNR Rao, which shows in google scholar Prof. CNR Rao has 283 citations or occurrences. 176 Facebook, 984 twitter and G+ with 193 but Cross reference is 69666 which is highest among Altmetrics data obtained from Vidwan web page.

Table 8. Google Scholar

Sl. No	Type	Total in Numbers	Total in Numbers (IN 2020)
1	Citations	124080	27090
2	h-index	165	77
3	I10–index	1214	60

Table-8 shows google scholar data of Prof. CNR Rao, 124080 citations were received in total and 27090 citations were received till 2020. The H-index of Prof. CNR Rao is 165 and 77 in the year 2020. I10 –index is 1214 and 60 in the year 2020.

5 Findings

- Prof. CNR Rao has published 143 research articles during 2015 to 2025 and more than 1790 publications from 1957 to 2025.
- Highest number of articles were published during 2015 with 29(20.27%) and less number of articles i.e. 2(1.39%) in 2021.

- iii. A Total of 4807 citations are received during 2015 to 2025 and the highest number of citations i.e. 1607 are received in 2015 and 24 citations are received during 2024.
- iv. Single authored articles are 7(4.89%) and multi-authored articles are 136(95.11%).
- v. The Degree of collaboration is 0.95 which proves highest number of articles are produced by collaboration.
- vi. Altmetrics and google scholar data shows that Prof. C N R Rao has an h-index of 151 with more than 106815 citations.

6 Conclusion

Prof. C N R Rao has published more than 1790 articles over the period of 1957 to 2025 and still he is publishing extensively at the age of 91 years. Many Scientometrics studies have been carried early with time constraint, even our study is limited to ten years of publications, an extensive study can be carried out which includes all of his publications till date. His most preferred channel of communication is research articles i.e. more than 1500 and less preferred mode of communications is short surveys. Over the years it can be observed that single authored articles are decreasing and multi-authored articles are increasing and citations of articles are more compared to Altmetrics data, this might be because his communications are more in scientific journals and social media access to these are very much restricted. Prof. C N R Rao's legacy continues to grow over time and his contributions have to be celebrated in every way.

Declaration of AI and AI-Assisted Technologies used in Manuscript Writing: Not Applicable

References

1. Ahirwal, Ashish Kumar & Pathak, Sandeep Kumar (2015). Scientometric portrait of Bharat Ratna Professor C N R Rao: A Study. *IJESR*, 3(11), 54-80.
2. Eleventh COLLNET Meeting, October 19 – 22, 2010, University of Mysore.
3. Gadad Raju & Ravi B. (2018). Publication productivity of Bharat Ratna Prof. C N R Rao during 1956-1975: A Scientometric Portrait. *IJLS*, 8(2), pp282-287.
4. Gadad Raju, Ravi, B. & Savanur, Kiran (2019). Highly cited paper of Bharat Ratna Prof, C N R Rao: Scientometric attempt. *Asian Journal of Information Science & Technology*. 9(1). pp91-94.
5. Gadad, Raju. & Ravi, B. (2016). Nobel Laureate TC Sudof: A Scientometric Portrait. *Journal of Research in Librarianship*, 3(6), pp. 11-18.
6. Gadad, Raju. & Ravi, B. (2017). Nobel Laureate Prof. M.B. Moser: A Scientometric Portrait. *ISST Journal of Advances in Librarianship*, 8(2), pp. 18-26.
7. Jawaharlal Nehru Centre for Advanced Scientific Research | IRINS. (2025). *Irins.org*. <https://jncasr.irins.org/>
8. K.R. Mulla & et. ella (2010) A STUDY ON BIOBIBLIOMETRICS ANALYSIS OF PROF. C. N. R. RAO 'S RESEARCH PUBLICATIONS in Sixth International Conference on Webometrics, Informatics and Scientometrics &
9. Kalyane, V.L. & Sen B.K. (1996) Scientometric portrait of Nobel laureate Pierre-Gilles de Gennes. *Malaysian Journal of Lib.Infor.Sci.* 1(2), pp.13-26.
10. Kulkarni, Manoj (1997) Bibliometric Portrait of Scientist working on yoga. *IASLIC Bulletin*, 42(4), pp.185-88.
11. Nagarjuna, P., & Javali, S. I. (2015). Highly Cited Paper from Indian Institute of Science, Bengaluru, India during 2001-14: A Scientometrics Analysis Case Study. In 10th International CALIBER-2015. HP University and IIAS, Shimla, Himachal Pradesh, India March 12-14,
12. Sangam, S L. et al. (2006). Scientometric portrait of Prof. Peter John Wyllie. *Scientometrics*, 66(1), 43-53

Technology and Digital Transformation

Enhancing AI-based Library Services in Academic Libraries.

Dashrath Singh Nag ^[0009-0001-4810-1594]

Central Library, AISECT University, Jharkhand
dashrathnag3@gmail.com

Abstract. Academic libraries are undergoing a profound transformation driven by artificial intelligence (AI). From automated cataloguing and discovery tools to conversational agents and predictive analytics, AI promises to make library services more efficient, personalized, and accessible. However, implementation raises technical, ethical, and organizational challenges, including bias, privacy, transparency, and staff skill gaps. This essay reviews contemporary literature and policy guidance, surveys current AI applications in academic libraries, analyses benefits and risks, and proposes a practical framework and targeted recommendations for ethically and effectively enhancing AI-based services in academic libraries. Key recommendations include stakeholder-driven needs assessment, ethical risk assessment aligned with professional principles, phased pilots with robust evaluation metrics, staff capacity building, and transparent communication with users. Citations and references are provided in APA style.

Keywords: Artificial Intelligence, Academic Libraries, Machine Learning, Semantic Search, Black Boxes.

1 Introduction

Academic libraries have historically adapted to technological change from card catalogues to online public access catalogues (OPACs) and discovery layers. The latest wave of transformation is propelled by Artificial Intelligence (AI) technologies such as natural language processing (NLP), machine learning (ML) and large language models (LLMs). These technologies can automate routine tasks (e.g., metadata generation), enhance discovery (semantic search and recommendation), provide 24/7 user assistance (chatbots/conversational agents), and support research workflows (data curation, literature mapping). Yet, integrating AI must be done carefully to safeguard intellectual freedom, privacy, and equity while preserving librarians' professional values. This work synthesizes recent research and professional guidance; then proposes actionable steps that academic libraries can take to enhance AI-based services responsibly.

2 Literature review and policy context

A growing body of research examines AI applications, impacts and risks in library settings. Systematic reviews of chatbots in libraries highlight both the rapid uptake of conversational agents and persistent concerns about privacy, accessibility, and professional identity [6]. Empirical studies of chatbots in academic libraries point to improvements in user access and reference efficiency but note adoption barriers and the need for careful design and evaluation [5]. Broader reviews on AI in academic libraries describe benefits such as increased operational efficiency, personalized services, and improved accessibility, while flagging issues such as algorithmic bias and lack of staff training [7][6].

Professional bodies and international organizations have released guidance to help libraries navigate AI responsibly. The International Federation of Library Associations and Institutions (IFLA) published an "Entry Point to Libraries and AI" to help librarians assess opportunities and risks and to promote ethically aligned adoption. Similarly, research library organizations have developed guiding principles for AI emphasizing transparency, accountability, and user privacy. UNESCO's Recommendation on the Ethics of AI provides a global framework relevant to public-interest institutions, including libraries. These documents stress that AI adoption in libraries should align with core library values and human rights.

3 Current AI applications in academic libraries

Academic libraries are implementing AI across multiple service areas. Representative categories include:

3.1 Discovery and Search Enhancement

AI-driven semantic search, entity extraction, and recommender systems help users find relevant literature beyond keyword matching. These systems improve recall and relevance ranking by understanding context and relationships among concepts. (Research literature and case studies show improved discovery, but caution about opaque ranking algorithms.)

3.2 Chatbots and Virtual Reference

Conversational agents provide 24/7 basic reference, FAQs, navigation help, and transactional services (renewals, holds). Studies indicate chatbots can reduce staff load on routine queries while increasing user satisfaction when well-designed and transparently limited. However, concerns about privacy, accessibility, and managing complex references remain.

3.3 Automated Metadata and Digitization Workflows

ML models assist in automatically generating descriptive metadata, classifying content, OCR correction, and entity recognition for digitized collections. This accelerates processing backlogs and enhances the discoverability of archival material.

3.4 Research Support and Scholarly Communication

AI tools assist researchers with literature reviews (topic modelling), data cleaning/curation, and bibliometric analysis. Libraries increasingly offer training and platforms to mediate researchers' use of AI tools responsibly.

3.5 Analytics and Collection Management

Predictive analytics inform acquisition decisions, usage forecasting, and resource optimization. When used carefully, analytics can improve resource allocation but must respect privacy and avoid reinforcing inequitable resource distributions.

4 Benefits of AI-based services

AI has several compelling benefits for academic libraries:

- **Improved efficiency:** Automation of repetitive tasks (cataloguing, metadata enrichment) frees staff for higher-value work (reference, instruction, and outreach).
- **Enhanced user experience:** Personalized recommendations, semantic search, and chat assistance can shorten discovery time and make resources more accessible.
- **Scalability:** Conversational agents and automated pipelines enable libraries to serve larger, more dispersed user populations (e.g., distance students) without proportional staffing increases.
- **Support for research outcomes:** AI-enabled services (topic modelling, citation mapping) accelerate literature review and interdisciplinary discovery, supporting the institution's research mission.

These benefits make a strong case for thoughtfully integrating AI across library services, but realising them requires governance, evaluation, and alignment with ethical standards.

5 Risks and challenges

Despite potential, AI adoption raises significant risks:

5.1 Bias and Fairness

AI models trained on biased corpora can perpetuate and amplify inequities in discovery, classification, and recommendation. Libraries must anticipate dataset biases, especially in historical and culturally diverse collections. (Professional guidance stresses assessing systems for fairness.)

5.2 Transparency and Explainability

Many AI systems (particularly deep learning and proprietary LLMs) act as “black boxes.” Lack of explainability undermines user trust and complicates problem resolution when errors or hallucinations occur. IFLA guidance recommends transparency about system capabilities and limitations.

5.3 Privacy and Data Protection

Conversational logs, personalization signals, and usage analytics may contain sensitive information. Libraries must balance analytic value with legal and ethical obligations to protect patron privacy. UNESCO and library principles emphasize privacy-preserving approaches.

5.4 Intellectual Freedom and Academic Integrity

Automated filtering and algorithmic ranking can inadvertently suppress minority viewpoints or prioritize certain publishers, impacting intellectual diversity. Additionally, the use of generative AI in scholarship raises questions about attribution and authenticity. Recent attention to AI’s role in academic publishing underscores the need for disclosure and policies.

5.5 Technical and Resource Constraints

Implementing AI entails costs (infrastructure, licensing), skill requirements, and ongoing maintenance. Smaller libraries or those in resource-constrained institutions may face barriers to meaningful adoption.

5.6 Quality and Trustworthiness of AI Outputs

LLMs can “hallucinate” producing plausible but incorrect information. Libraries must design safeguards to prevent and correct misinformation in services relying on generated content.

6 A practical framework for enhancement

To enhance AI-based services responsibly, academic libraries should adopt a structured, phased framework that integrates ethical assessment, stakeholder engagement, technical validation, and iterative evaluation. Below is a recommended framework:

6.1 Strategic alignment and needs assessment

- Start with institutional priorities teaching, learning, and research goals, not technology for its own sake. Conduct user needs assessments (surveys, focus groups) to identify high-impact use cases (e.g., 24/7 reference, backlog metadata enrichment).
- Map legal and policy constraints (data protection laws, institutional AI policies). Align potential AI projects with library values (equity, access, privacy).

6.2 Ethical risk assessment and governance

Conduct an ethical impact assessment for each proposed AI service addressing bias, privacy, transparency, and accountability. Use checklists and reflective questions drawn from IFLA’s Entry Point and comparable guidelines to structure assessment. Establish an AI governance group with librarians, IT, legal counsel, faculty, and student representation.

6.3 Pilot projects with measurable outcomes

Run small, time-boxed pilots for prioritized use cases (e.g., chatbot for FAQs, automated metadata enrichment in a specific collection). Define success metrics (accuracy, user satisfaction, time saved, privacy incidents), collect baseline data, and compare outcomes. Use pilots to learn about integration costs and unintended effects before scaling.

6.4 Technology selection and procurement

Prefer solutions that allow transparency (open-source models where possible), offer privacy-preserving options (on-premises deployment or strong contractual data protections), and provide explainability features. Procurement should include requirements for auditability and vendor disclosure about training data and bias-mitigation practices.

6.5 Human-in-the-loop design

Keep librarians at the centre of workflows. Design systems so humans can review, correct, and contextualize AI outputs especially for reference answers, metadata edits, and recommendations. Human oversight reduces the risk of hallucinations, ensures ethical judgment, and preserves professional roles.

6.6 Capacity building and change management

Invest in staff training (AI literacy, data ethics, and model evaluation). Libraries should offer internal workshops and partner with computer science or data science departments. Build cross-functional teams combining librarianship and technical skills.

6.7 Continuous monitoring, evaluation, and transparency

Implement logging (with privacy safeguards), periodic audits for bias and accuracy, and user feedback channels. Publish clear user-facing explanations of what AI does, what data it collects, and how to challenge incorrect outputs. This transparency builds trust and meets professional obligations.

7 Case examples and lessons learned

Several studies and institutional reports provide practical lessons:

- **Chatbot implementations:** Systematic reviews show that chatbots can handle a significant portion of routine queries, improve accessibility and reduce staff workload but success hinges on iterative design, domain-specific training data, and clear handoff to human librarians for complex queries. Continuous performance monitoring is essential to catch errors and maintain accuracy.
- **Automated metadata projects:** Projects that applied ML for metadata enrichment reduced backlog volumes and improved discoverability. However, oversight was required to correct algorithmic misclassifications, and metadata quality improved most when human cataloguers reviewed AI suggestions rather than accepting them blindly.
- **Policy-led approaches:** Institutions that adopted principled AI governance (drawing on ARL principles and IFLA guidance) were better positioned to articulate procurement requirements and address privacy and equity concerns during procurement and deployment phases.

These examples reinforce the framework's emphasis on pilots, human oversight, governance, and evaluation.

8 Recommendations for academic libraries

Based on the literature, policy guidance, and implementation lessons, the following concrete recommendations are proposed for academic libraries seeking to enhance AI-based services:

8.1 Adopt an AI policy aligned with library values

Use ARL, IFLA, and UNESCO documents to craft an institutional policy that addresses transparency, privacy, bias mitigation, and vendor accountability. Publish a concise, user-facing summary to build trust.

8.2 Prioritize high-impact, low-risk pilots

Begin with projects that deliver measurable user benefits and limited privacy exposure (e.g., public FAQs chatbot, internal metadata enrichment for non-sensitive collections).

8.3 Insist on procurement transparency

When acquiring third-party AI services, require vendors to disclose training data sources, provide model cards or documentation on limitations, and agree to privacy clauses that limit retention and reuse of patron data.

8.4 Implement human-in-the-loop workflows

Ensure AI outputs are reviewed by librarians where errors could cause harm (reference replies, archival descriptions). Use AI as an assistive tool that augments professional judgment.

8.5 Build staff capacity

Offer continuous professional development in AI literacy, data ethics, and evaluation methods. Partner with academic departments to host joint workshops and create cross-disciplinary teams.

8.6 Measure outcomes and publish findings

Define clear KPIs for pilots (accuracy, response times, user satisfaction, privacy incidents). Share outcomes, both successes and failures to contribute to the profession's evidence base.

8.7 Engage users and stakeholders

Conduct outreach to explain AI services, obtain consent where appropriate, and provide channels to challenge AI-generated content. Involve faculty and students in governance to ensure alignment with academic needs.

8.8 Monitor broader AI ecosystem risks

Stay informed about developments in AI research integrity (e.g., concerns about AI-generated scholarship) and update library policies for scholarly communication, plagiarism detection, and authorship statements.

9 Conclusion

AI offers substantial opportunities to enhance academic library services, improving efficiency, accessibility, and research support. However, those gains come with important responsibilities: to manage bias, protect privacy, preserve intellectual freedom, and maintain trust. Libraries should pursue a cautious, values-driven approach: align AI projects with institutional goals, conduct ethical risk assessments, pilot and evaluate carefully, and keep librarians central to design and decision-making. By adopting transparent governance, investing in capacity building, and publishing outcomes, academic libraries can harness AI to extend their service mission while upholding core professional values.

Declaration of AI and AI-Assisted Technologies used in Manuscript Writing Process: Not Applicable

References

1. Aboelmaged, M. (2025). *Conversational AI Chatbots in library research*. Journal of Information Science. [Article details]. SAGE Journals
2. Association of Research Libraries. (n.d.). *Research libraries guiding principles for artificial intelligence*. ARL. Retrieved from <https://www.arl.org/resources/research-libraries-guiding-principles-for-artificial-intelligence/> Association of Research Libraries
3. IFLA. (2025). *Entry Point to Libraries and AI*. International Federation of Library Associations and Institutions. Retrieved from <https://www.ifla.org/news/release-of-ifla-entry-point-to-libraries-and-ai-document/> IFLA+1
4. IFLA FAIFE. (2020). *IFLA Statement on Libraries and Artificial Intelligence*. IFLA Repository. Retrieved from <https://repository.ifla.org/items/8c05d706-498b-42c2-a93a-3d47f69f7646/> repository.ifla.org+1
5. Kaushal, V. (2022). *The Role of Chatbots in Academic Libraries*. Journal of Academic Librarianship. <https://doi.org/10.1080/24750158.2022.2106403> Taylor & Francis Online
6. Kisilowska-Szurmińska, M. (2025). *Artificial intelligence in academic libraries – a tool or partner?* Library Quarterly. [Article details]. ScienceDirect

7. Mallikarjuna, C. (2024). *An analysis of integrating artificial intelligence in academic libraries*. SSRN. Retrieved from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4898532 SSRN
8. Ngulube, P. (2025). *Integrating artificial intelligence-based technologies 'safely' into academic libraries*. Journal of Librarianship and Information Science. [Article details]. Taylor & Francis Online
9. Saeidnia, S. (2023). *Ensuring ethical intelligence: Guiding the integration of AI in modern libraries*. Zenodo. Retrieved from <https://zenodo.org/records/13903781/files/Ensuring%20Ethical%20Intelligence%20--%20Guiding%20the%20Integration%20of%20AI%20in%20Modern%20Libraries.pdf> Zenodo
10. UNESCO. (2021). *Recommendation on the ethics of artificial intelligence*. United Nations Educational, Scientific and Cultural Organization. Retrieved from <https://www.unesco.org/en/artificial-intelligence/recommendation-ethics> UNESCO
11. Wired. (2023). *Use of AI is seeping into academic journals—and it's proving difficult to detect*. Wired. Retrieved from <https://www.wired.com/story/use-of-ai-is-seeping-into-academic-journals-and-its-proving-difficult-to-detect/> WIRED
12. Yan, R., et al. (2023). *Chatbots in libraries: A systematic literature review*. White Rose Repository. Retrieved from <https://eprints.whiterose.ac.uk/id/eprint/204750/1/EFI-230045%20%281%29.pdf> White Rose Research Online

AI-Based Library Service to Assist Undergraduate Student Projects

Sheetal S. Sonawane¹[0000-0002-8733-8340] and Anupama G. Phakatkar²[0000-0002-9856-3586] and Girish K. Palshikar³[0000-0003-3625-6705]

¹⁻³ Pune Institute of Computer Technology, Pune 411043, India
{sssonawane, agphakatkar, gkpalshikar}@pict.edu

Abstract. Undergraduate students do a project in their final year where they learn to solve real-life problems by learning and applying sophisticated techniques. After coming up with a clear and concise problem statement, the next step is doing a focused literature survey: identify the most relevant books and papers, and build an understanding of the core concepts and techniques relevant to the project. While large Language Models (LLMs) can be used directly to identify references, they suffer from severe hallucinations, generating 80-90% fake references. We propose a novel library information service to get the UG students started on the journey of literature survey, by recommending a gentle, graded path of few highly relevant, high-quality, and really published books, survey papers and research papers. We associate each suggested reference to core technical concepts in their project and also generate a concise explanation as to what they can learn from each reference. We validate these ideas on real-life UG project descriptions. In line with pedagogical expectations from UG student projects, our goal is not to automatically construct a comprehensive literature review for the project, nor identify research gaps, nor generate summaries of suggested references. We leave these tasks for UG students, as it is part of their learning process.

Keywords: Undergraduate Student Projects, Project-based Learning, AI-based Library Services, Natural Language Processing, Artificial Intelligence.

1 Introduction

Undergraduate (UG) (e.g., B.Tech., B.E.) students in colleges and universities often have to do a project in their final year as part of their degree requirements. Doing the project work (hopefully) teaches the students to understand and solve real-life problems by learning and applying sophisticated techniques [1], [2], [3], [4]. The work also teaches the students important skills (e.g., team-work, technical communication, project management) and processes (e.g., problem formulation, literature survey, solution building, user interactions and documentation), that are useful in practice. Often the first step in the project is coming up with a clear and concise problem statement: what is the problem to be solved, motivation, benefits, challenges, expected outcomes, deliverables and sometimes initial ideas about the solution approach. The next step is doing a focused literature survey: identify the most relevant books and papers, build an understanding of the core concepts and techniques and their relevance to the project.

In recent years there has been an explosion in the number and complexity of published papers across a large number of journals, conferences and other venues. UG students sometimes get overwhelmed, particularly when working under time pressure, by the task of understanding and building a coherent mental picture of the existing literature useful for their project.

With the advent of Large Language Models (LLM), the task of literature survey is getting rapidly automated [5], [6], [7], [8], [9]. Several AI-based techniques have been developed to provide novel services to library users [10], [11], [12], including literature survey. Most such tools are expensive or not adaptable for UG students' requirements. Moreover, such approaches generate not just a list of references but also a complex technical review of the literature as well, which makes them better suited for post-graduate or doctoral students and experienced researchers. We propose a novel library information service with the goal of getting the UG students started on the journey of learning to do literature survey, by suggesting a gentle, graded path of a few highly relevant, high-quality books, survey papers and research papers. We generate short explanations (justifications) that clearly state the relevance of the suggested references to the given project description (why the students should look at that reference). We also annotate the given project descriptions by inserting the suggested references at appropriate points. Our goal is *not* to use LLMs to automatically construct a comprehensive literature review for the project, nor do we identify gaps nor generate summaries of the suggested research papers. Pedagogically, learning to read and understand research literature is important for the UG students' learning, we want these students to do that task on their own. Moreover, often the project description is not frozen initially, and the scope may change. In

that case, the relevant literature may change as well. Our approach makes it easy for the students to re-generate references (along with annotations and explanations), if the project description changes.

Our approach is explicitly aimed at assisting UG students in the *initial* stage of their projects. UG projects are of different kinds: system (prototype) building, theoretical research, modelling and evaluation etc. UG projects have shorter project duration, UG students have relatively weak technical backgrounds, as well as less experience in reading research paper. Hence the goals of literature survey in UG projects are quite different as compared to post-graduate or doctoral research or more advanced professional research. The emphasis is less on breadth of coverage of the research literature and more on gathering the following: A *few* survey papers or books that provide broad technical background to UG students necessary for their project – this includes terminology, core concepts, important techniques, methodology etc.; and Research papers that are highly specific and relevant to their project i.e., those which address either exactly the same or a very similar problem. Ideally, the suggested references should be as recent as possible and published in highly reputed journals or conferences. Emphasis is on suggesting only a *few* papers that are highly relevant.

Few papers from (1) and (2) are often sought by the UG students and their guides. The notion of relevance of a reference with the given project description is crucial. Given this scope, in this paper, we address the following problems: Given a short textual description D for a UG project⁹, generate a list of k (k is user-specified) most relevant papers or books, taking into account the needs and constraints of literature survey for UG students as stated above. Ideally, the output should include a clear explanation or justification of the relevance of the paper for their project. Create a dataset of project descriptions (initial synopsis) for UG projects, created at the beginning of the project, along with title, abstract, keywords, domain, and list of actual references cited in the final project report. Automatically associate (insert) each reference with the given project description D (+ project title). We do this by associating each suggested reference with a core technical concept in the project description. We use an LLM to generate a concise explanation (1-2 sentences) that summarizes what concepts and techniques students can learn from each suggested reference and how these are relevant to their project. We use human annotations to evaluate the “quality” of the suggested references and their justifications, based on the relevance, adequacy, recency and importance to the given project description D .

2 Example

Fig. 1 shows a sample project description, where some key technical terms are in bold. We tried to directly use Gemini LLM to generate references for a given project description, but its response was needlessly complicated (despite experimentation with many prompts). Gemini attempted a complete technical formulation of the project, whereas the prompt sought only the references relevant to the project. Moreover, almost 80-90% references suggested by Gemini were either imaginary (hallucinatory) or what Gemini rather coyly called “conceptual” i.e., based on combining parts (e.g., title, authors, publication venue) of multiple references. Note that since LLMs are probabilistic generative models, every run even for the same project description yields different results.

3 Dataset

We report results on a small dataset of 14 UG project descriptions. We do have a larger dataset of project descriptions. The format of the data about each project is: <Year, Title, Description, List-of-References>. The projects span the following business domains: healthcare, finance, document processing, education, fashion, news, agriculture, human resources, tourism, retail, and security. Table 1 shows the summary statistics for the dataset. Fig. 2 shows the word cloud obtained from project descriptions dataset.

Efficient Vaccine Distribution Planning using IoT
Vaccine distribution planning is crucial for **children's immunization** and **prevention of infectious diseases** from spreading. In this project, we want to design **optimization algorithms** so as to maximize **demand fulfilment**, reduce wastage and minimize cost and time of **vaccine transportation**. We will consider **cold chain constraints** and number and capacity of the vehicles needed for the distribution. Another goal is to explore the role of **IoT** to improve the **vaccine supply chain efficiency**.

Fig. 1. Example project description.

⁹ We focus on UG (4th year B.Tech.) students in Computer Science, Computer Engineering and Information Technology departments in our Institute.

Table 1. Summary Statistics for the Dataset.

	#Sentences	#Words
Min	3	86
Max	14	378
Average	8.07	221.1
Std. Dev.	2.92	83.4
First Quartile	6.25	164.5
Median	8.5	258.5
Third Quartile	9	263



Fig.2. Word Cloud for project descriptions.

4 Solution Approach and Results

Fig. 3 shows the steps in our solution approach.

4.1 Identification of Key Technical Terms from Project Description

First, we use an LLM10 to identify domain-specific technical terms (technical terms, in short) in the given project description. Technical terms are part of the subject nomenclature, and have a clear, precise and specialized meaning within that domain. Such technical terms form a common, public vocabulary that helps people working within that domain communicate effectively. Amortization, capital gain, depreciation, return on investment are examples of technical terms in the finance domain. We focus on identifying technical terms in closely related domains such as: computer science, computer engineering, information technology, software engineering. The reason is that while many UG student projects do happen to solve problems in different real-life business domains (such as tourism, healthcare, agriculture etc.), the focus of the UG students is almost always on building related software solutions.

We consider technical terms which are only in the form of a noun phrase (NP), although we have also considered noun or verb phrases that specify a task (activity, action, work item) that the student will undertake as part of the project. We used Gemini to identify technical terms in the given project description. Our prompt to Gemini included several guardrails i.e., rules to ensure only simple, easy-to-understand technical terms were extracted, such as the following:

- extract only noun phrases (NP);
- prefer less complex NPs (e.g., those that are not “connected” by prepositions), such as vaccine transportation over cost and time of vaccine transportation;
- prefer longer NPs i.e., avoid an NP if it occurs within another NP;

¹⁰ We have used gemini-2.5-flash LLM in this paper.

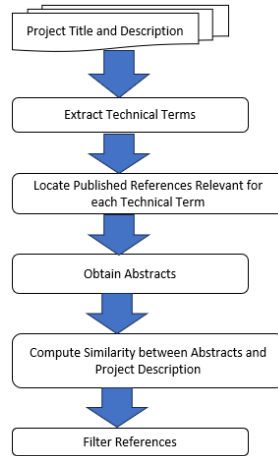


Fig.3. Steps in our solution approach.

- strip adjectives that occur at the beginning of the NP, if such adjectives are too generic, such as *intelligent*, *AI-driven* etc.
- avoid single word NPs, if the word is too generic, e.g., *system*, *process*, *solution*, *application* etc.;
- avoid an NP if it does not belong to the domain of *computer science*, or *computer engineering*, or *information technology* or *software engineering*.

We asked Gemini to rank the technical terms in terms of their relevance to the given project description, and output only top k terms (we used $k = 6$). As an alternative, we computed the relevance as the cosine similarity between the two 384-dimensional embedding vectors¹¹: one for the project description and the other for a particular technical term. Removing technical terms below a specified threshold, we obtained results very similar to those given by Gemini. Top 6 technical terms extracted and ranked by Gemini from the project description in Fig. 1 are: IoT, optimization algorithms, vaccine distribution planning, vaccine supply chain efficiency, cold chain constraints, vaccine transportation. Technical terms such as children’s immunization and prevention of infectious diseases were not extracted by Gemini (as instructed) because they fall outside our domains of focus (mentioned earlier). Gemini (correctly) did not extract technical terms like supply chain because it is part of another, longer technical term vaccine supply chain efficiency. Also, Gemini correctly did not extract single word technical terms, such as wastage, as this word is too generic. We observed that on rare occasions Gemini did not adhere to these rules.

We manually evaluated the technical terms extracted from the 14 project descriptions. Taking into account the project title and project description, we labelled each extracted technical term as Correct, Partly-Correct and Incorrect. A technical term was labelled as Correct (Incorrect), if it (not) represented an important and core concept in the project. A technical term was labelled as Partly-Correct if it is somewhat important to the project, but was not really a core concept. From the 14 project descriptions in our dataset, 74 technical terms were extracted, out of which 72 (97.3% precision) were manually labelled as Correct, 1 as Partly-Correct and 1 as Incorrect.

4.2 Generating Reference List

We initially used Gemini to identify most relevant references for each of the technical terms that we had extracted in the previous step from the given project description, instructing Gemini to make sure that they are suitable for UG students. We also asked Gemini to generate a succinct, clear justification for how UG students can use each reference to learn about the technical concept. Unfortunately, we found that 80-90% of the references identified by Gemini were non-existent, after we implemented several post-generation checks, such as reachability of URL link provided with the reference, and trying to locate the reference in another database. Several attempts at revising the prompt using advanced prompting strategies, such as Chain-of-Thought, and metacognitive prompting [13], did not reduce the hallucination level much.

¹¹ In this paper, we have used the sentence transformer model **all-MiniLM-L6-v2** to compute embedding vectors for a sentence or paragraph.

Next, we tried to programmatically use Google Scholar and SCOPUS to locate relevant references, but this failed, probably because of their anti-bot measures.

Finally, we settled on the DBLP database (<https://dblp.org>) of 8.2 million references, mainly for Computer Science. We used the Python package **dblp** to remotely (over the Internet) search the DBLP database through our Python program. While the DBLP API search interface allows a program to send simple queries to DBLP, DBLP does not generate any justification for why the reference was relevant to the given project description. For a given query, only a few details for each reference were returned by the search interface: *<Type, Title, Authors, Publication-Venue, Year, Link>*. *Type* refers to whether the reference is a book, journal article, conference paper etc. *Link* refers to a URL, usually a DOI link, or the journals or conference's or book publisher's website, where additional details (e.g., abstract) about the reference are expected to be available. For a given query, DBLP returns a ranked list of 30 references, ranked using their own relevance score; the relevance score is not returned. DBLP sorts the selected 30 references using the year of publication (latest first). Unfortunately, abstracts for *none* of the papers were accessible through the DBLP API (probably for copyright reasons). To obtain references for a given project description *D*, we follow these steps:

- i. Send each technical term *T* extracted previously as a query to DBLP as a query through the search API. Collect all references (30 or less) together for all technical terms for *D* as a list *R*.
- ii. For each reference in *R*, use the *Link* and try to locate the abstract from that website. If this does not succeed (often because of anti-bot measures implemented on the publisher's website), we next try to obtain the abstracts from the public-domain CrossRef database¹², using the **habanero** Python package, which succeeds in getting abstracts for 30-40% of the references. Else, we ask Gemini to locate the abstract. Gemini mostly refuses to give the actual abstract (probably for copyright reasons) and instead *generates* an abstract, from the title and from other data it has about the reference.
- iii. Compute the cosine similarity of the project title with reference title using the Sentence Transformer embeddings, as explained earlier.
- iv. Compute the cosine similarity of the project description with the abstract of the reference, using the Sentence Transformer embeddings, as explained earlier.
- v. Remove those references from *R* having a low similarity with project title (below a given threshold). Ensure at least one reference is retained for each technical term.
- vi. Remove those references from *R* having a low similarity with project description (below given threshold), keeping at least one reference for each technical term.
- vii. Sort the references on both the similarities mentioned above (e.g., sorting on their sum). From this sorted list, for each technical term, output at most 2 books (if present in *R*), at most 2 review / survey papers (if present in *R*) and remaining research papers, so as to reach a count of at most 5 references for each technical term.

A. Ifeanyichukw et al, *Exploring artificial intelligence-based distribution planning and scheduling systems' effectiveness in ensuring equitable vaccine distribution in low-and middle-income countries*. **Discov. Artif. Intell.**, 2025.

Abstract (from DOI Link): During the COVID-19 pandemic, the global issues of vaccine access and equity, particularly in low-and middle-income countries (LMICs), came to the forefront. Simultaneously, there was notable advancement in artificial intelligence (AI) and its potential applications in vaccine distribution and scheduling. In response to these developments, we gathered insights, lessons, and perspectives to inform future strategies for AI-based distribution planning and scheduling systems effectiveness in ensuring equitable vaccine distribution in LMICs. We conducted a scoping review...

Similarity of Title with Project Title: 0.61

Similarity of Abstract with Project Description: 0.40

Fig.4. One reference identified for technical term `vaccine distribution planning` for the project description in Fig. 1.

Fig. 4 shows one of the output references identified for a technical term `vaccine distribution planning` for the project description in Fig. 1. Since each reference is associated with a specific technical term, it is now easy to:

- associate (insert) the reference to an appropriate point in the project description, such as the first mention of the technical term;
- understand the relevance of the reference to the project, namely, to get a better understanding of the concepts and techniques associated with that technical term;
- judge relevance of the reference for the project, using associated similarity scores.

¹² www.crossref.org provides an open infrastructure for research objects like papers, authors etc.

From the 14 project descriptions in our dataset, 74 technical terms were extracted, and a total of 375 references were identified (5.1 references per technical term, on average). Table 2 shows summary statistics of the similarity of the abstract of each reference with project title and project description. Fig. 5 shows histograms for these two similarities.

Table 2. Summary Statistics for the Reference Similarity with Project.

	SimWithProjectTitle	SimWithProjectDescription
Min	0.046	0.110
Max	0.980	0.802
Average	0.331	0.420
Std. Deviation	0.145	0.117
First Quartile	0.238	0.337
Median	0.301	0.407
Third Quartile	0.392	0.475

An important question is: does using a Gemini-generated abstract seriously affect the quality of our system’s results? We manually checked the actual abstracts and Gemini-generated abstracts and they match well; we did not see much hallucination in the generated abstracts. To check this more rigorously, we created a small dataset of randomly selected references (out of those recommended by our system) having Gemini-generated as well as original abstracts (we manually collected the latter). We computed similarities of these abstracts with each other as well as their similarities with project title and project description. Table 3 shows the results. The abstracts have very high similarities with each other, and their similarities with project titles and project descriptions are quite close to each other. A reasonable conclusion is that using Gemini-generated abstracts does not significantly reduce the quality of results.

We manually marked each reference as Acceptable or Not-Acceptable, by taking into account the project description as well as the reference’s abstract. A reference was labelled Not-Acceptable, if it was not relevant or at best only weakly relevant to the project, or if the reference was too difficult for UG students for getting a basic understanding of the concepts or techniques relevant for that project. About 56% to 60% of these references were manually labelled as Acceptable, and the rest as Not-Acceptable. Two main reasons for Unacceptability were: reference was either not too relevant for the project or technically too complex for UG students.

Table 3. Comparison of Original Abstract vs. Gemini-Generated Abstracts for 6 References.

	s0	s1	s2	s3	s4
Reference 1	0.7424	0.3726	0.3996	0.4149	0.4433
Reference 2	0.8587	0.4027	0.4247	0.4529	0.4706
Reference 3	0.7749	0.0681	0.1307	0.2212	0.2093
Reference 4	0.7898	0.1537	0.1765	0.2759	0.3020
Reference 5	0.6341	0.4120	0.6339	0.4490	0.6354
Reference 6	0.7302	0.4268	0.4774	0.5890	0.5961

s0: similarity between original and Gemini-generated abstract

s1 (s2): similarity between original (Gemini-generated) abstract and project title

s3 (s4): similarity between original (Gemini-generated) abstract and project description

We also used Gemini to generate a 1-2 sentence justification for each reference: what concepts and techniques the UG student would learn using this reference and in what way they are relevant to the project. We omit the details for lack of space.

4.3 Another Approach

For comparative analysis, we used chatgpt-4o (instead of Gemini) in two ways.

- **Approach 1:** Exactly as in the approach described earlier, we used ChatGPT to first extract core technical terms from the given project description, and then to suggest references for each such technical term.
- **Approach 2:** We prompted ChatGPT to directly suggest relevant references for the given project description, without extracting core technical terms.

Two-sample t-test shows that average embedding-based cosine similarity of suggested references to project description in Approach (1) (0.707) is statistically significantly better than Approach (2) (0.682) (Fig. 6). Approach (1) is also better for students to understand where and in what way to use the suggested reference, since it

is associated with a technical term. However, despite several attempts at prompt engineering, ChatGPT hallucinates, generating a large percentage of imaginary references. This severely limits the utility of these approaches, whereas the earlier approach does generate real, published references that are relevant to the given project description.

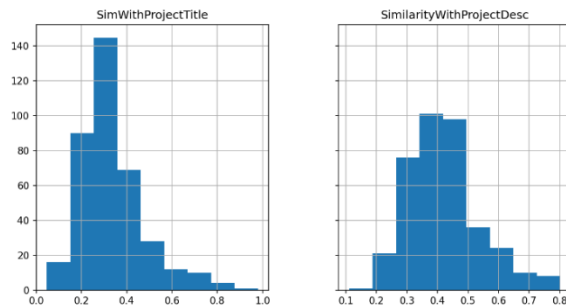


Fig. 5. Histograms of similarities of reference abstracts.

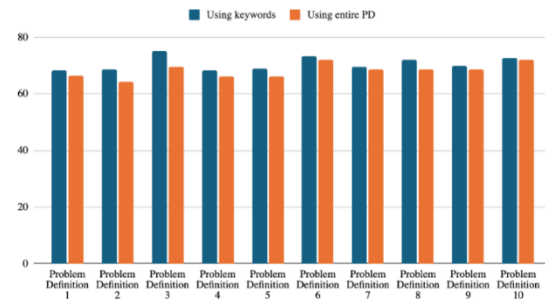


Fig. 6. Average reference similarity to project descriptions using Approaches (1) and (2).

5 Conclusions and Further Work

Experiential learning is important in higher education, and UG projects provide an opportunity to students to learn advanced techniques and use them to solve real-life practical problems. In this paper, our goal was to build an AI-based “intelligent” library service to provide assistance to UG students by generating a gentle, graded path consisting of a few highly relevant, high-quality books, survey papers and research papers for their project. We associated each suggested reference to core technical concepts in their project and also generated a concise explanation as to what they can learn from each reference. We validated these ideas on real-life UG project descriptions.

While quality of the generated references is found to be good (subjectively speaking), our techniques and results can be improved further. For future work, we will work on improving the search and reference selection strategies. Currently, DBLP automatically prefers latest references, but we are looking into a better, more “intelligent” selection criterion, such as mixing recency, “popularity” (number of times the reference was cited) and “utility” (i.e., extracting and using concepts and techniques from the reference and devise direct measures for the “usefulness” of the reference for the project). We plan to conduct a user study to collect feedback and evaluation of the suggested references from UG project guides as well as UG students. We are exploring integrating other sources of published literature, such as Google Scholar and SCOPUS. Finally, this work is an initial part of a wider system envisaged to support the students through all stages of the project; e.g., using AI to review, critique and suggest improvements in the system architecture or high-level design of the project.

Declaration of AI and AI-Assisted Technologies used in Manuscript Writing: Not Applicable

References

1. Barker L.: Student and Faculty Perceptions of Undergraduate Research Experiences in Computing. *ACM Transactions on Computing Education* 9(1), 1–28 (2009).
2. Brewer, G., Dewhurst, A.M., Doran D.: Undergraduate Research Projects: Practice and Perceptions. *Psychology Learning & Teaching*, 11(2), 208–217 (2012).
3. Lopatto D.: Survey of Undergraduate Research Experiences (SURE): First findings. *Cell Biology Education* 3, 270–277 (2004).
4. Guo, P., Saab, N., Post, L.S., Admiraal, W.: A review of project-based learning in higher education: Student outcomes and measures. *International Journal of Educational Research* 102:101586 (2020).
5. Baumbach, L.: Transforming literature screening: The emerging role of large language models in systematic reviews. In: *National Academy of Sciences of the United States of America*, 122(2), e2411962122 (2025).
6. Antu, S.A., Chen, H., Richards, C.K.: Using LLM (Large Language Model) to Improve Efficiency in Literature Review for Undergraduate Research. In: *Workshop on Empowering Education with LLMs –Next-Gen Interface and Content Generation*, Tokyo, Japan (2023).

7. Thode, L., Iftikhar, U., Mendez, D.: Exploring the use of LLMs for the selection phase in systematic literature studies. *Information and Software Technology*, 184:107757 (2025).
8. Qi, R., Li, W., Lyu, H.: Generation of Scientific Literature Surveys Based on Large Language Models (LLM) and Multi-Agent Systems (MAS). In: *Natural Language Processing and Chinese Computing (NLPCC)*. LNCS 15363. Springer (2024).
9. Kasanishi, T., Isonuma, M., Mori, J., Sakata, I.: SciReviewGen: a large-scale dataset for automatic literature review generation. In: *Findings of the Association for Computational Linguistics*, pp. 6695–6715, Toronto, Canada (2023).
10. Ayinde, L., Ebiefung, R., Oladokun, B.D.: Adoption of artificial intelligence in academic libraries: A systematic review of current practices, challenges, and research opportunities. *Journal of Academic Librarianship* 52(1):103185 (2026).
11. Kulkanjanapiban, P., Silwattananusarn, T., Lambovska, M.: Research on AI-driven innovations and services in academic libraries: A bibliometric and systematic literature review. *Journal of Data and Information Science*, 10(4), 146-196 (2025).
12. Pawde, A., Apte, S., Ingale, K., Apte, M., Palshikar, G.K.: Identifying Emerging Information Needs of Library Users using Data Mining. *Journal of Indian Library Association*, 57(4), 155–164 (2021).
13. Wang, Y., Zhao, Y.: Metacognitive Prompting Improves Understanding in Large Language Models. In: *NAACL-HLT (Vol. 1: Long Papers)*, pp. 1914–1926, Mexico City (2024)

Metaverse Libraries: Revolutionising Knowledge Dissemination

Barnali Roy Choudhury^[0000-0002-1773-5279]

Department of Library and Information Science, Netaji Subhas Open University
barnaliroychoudhury@wbnsou.ac.in

Abstract. Metaverse libraries offer exciting opportunities for transforming the way knowledge is shared and accessed. By integrating virtual reality, augmented reality, blockchain technology, and artificial intelligence, these innovative platforms effectively break down the barriers of physical space and time. This paper highlights their key features, such as immersive three-dimensional collections, collaborative virtual environments, personalised AI-driven assistants, and gamified educational experiences, all designed to enhance learning. The research focuses on various constructive inquiries, including the factors that encourage the adoption of these libraries, user behaviour patterns, the integration of technology, and important equity challenges. These challenges encompass accessibility, privacy, digital disparities, copyright management, and the need for an ethical framework. The existing literature reveals a promising progression from conceptual ideas to tangible applications, as seen in initiatives like the Community Virtual Library. These projects demonstrate remarkable global engagement while addressing critical issues related to inclusivity and data security. In a nutshell, metaverse libraries are paving the way for a more democratized approach to spatial learning, reshaping the role of librarians to that of curators within expansive virtual knowledge ecosystems. This evolution presents an exciting path forward for education and information sharing in the digital age.

Keywords: Metaverse, Metaverse Libraries; Augmented Reality; Virtual Reality, Mixed Reality, Avatars, Block Chains; Digital Libraries, Artificial Intelligence.

1 Introduction

Metaverse libraries are on the verge of revolutionising the way knowledge is shared and accessed, offering a glimpse into the future of learning. By tapping into the immersive power of virtual environments, these libraries are not only enhancing traditional educational experiences but also forging unparalleled opportunities for collaboration and discovery among users. The new era of knowledge dissemination is dawning, promising a more dynamic, engaging, and interactive experience for everyone involved. The shift toward remote work and the necessity for virtual communication that emerged during the COVID-19 pandemic have significantly accelerated the evolution of the metaverse. This intricate digital ecosystem is composed of diverse virtual environments, each boasting unique features, interactive elements, and functionalities that cater to various learning styles and preferences.

2 Definitional Analysis

The term "metaverse" originates from the Greek prefix "μετά" (meta), meaning "beyond" or "above." Neal Stephenson popularised it in his visionary 1992 novel "Snow Crash," which envisions a future where individuals can fully immerse themselves in a boundless virtual realm, highlighting the immense potential for connection, creativity, and exploration within this digital landscape. The concept of the metaverse, as part of the broader technology of virtual worlds, began to take shape in the early 2000s with online games such as Second Life and World of Warcraft. These platforms enabled users to craft their own avatars.(Ali et al., 2023; Suh & Ahn, 2022) and interact in rich, vibrant virtual worlds, paving the way for innovative social interactions and collaborative experiences that lay the groundwork for today's metaverse libraries.

3 Literature Review

The literature on metaverse libraries has experienced significant growth since 2022, transitioning from initial conceptual explorations of virtual reality within academic contexts to more empirical studies that investigate usability, adoption processes, and the transformative potential of these digital environments.(Amzat & Adewojo, 2023; Anna et al., 2023; Johan et al., 2023; Sağsan, 2024) . Early research primarily emphasises the architectural

features of these libraries, showcasing immersive 3D collections, intelligent AI chatbots, the integration of blockchain technology for NFTs, and seamless cloud integration. (CANNAVO et al., 2025; Chow et al., 2023; Guo et al., 2024; Sağsan, 2024; Tella et al., 2023). These elements frame metaverse libraries as vibrant hubs for global collaboration, allowing users to interact and share knowledge beyond the constraints of physical locations.

In contrast, more recent studies have shifted their focus toward user-centric evaluations. They include in-depth case analyses of notable platforms like Chilgok Public Library and Community Virtual Library. While these investigations demonstrate a high level of user immersion and engagement, they also reveal several challenges, including navigation difficulties, privacy concerns, and the presence of digital divides that may hinder equitable access. Metaverse libraries leverage virtual reality and immersive digital spaces to transform traditional knowledge access into interactive, global experiences. The review highlights intriguing gaps in the current literature, especially when it comes to the Metaverse and its potential impact on libraries.

4 History of the Metaverse: From Early Virtual Worlds to Immersive Technology

The concept of the Metaverse has evolved significantly over several decades, transitioning from simple virtual environments to complex, immersive experiences. This journey encompasses four main aspects: early virtual worlds, the emergence of the modern metaverse, advancements in technology and the emergence of immersive platforms. These broader categories contain sub-eras segregated based on technological interventions.

4.1 Early Virtual Worlds

- **1970s: Text-Based Adventures:** The earliest forms of virtual worlds can be traced back to text-based games like ADVENTURE (also known as *Colossal Cave Adventure*), which allowed players to explore a digital world through text input.
- **1980s: MUDs:** The invention of Multi-User Dungeons (MUDs), which combined role-playing with real-time interaction among players, marked a pivotal step. Players could communicate, explore, and interact with each other in a shared text-based environment.
- **1990s: 3D Online Worlds:** The rise of 3D graphics paved the way for more visually immersive experiences. Games like Habbo Hotel and Ultima Online introduced avatars and richer environments, allowing users to socialise and create communities.

4.2 The Dawn of the Modern Metaverse

- **2000s: Second Life and Virtual Economies:** Second Life launched in 2003, allowing users to create and trade virtual goods, influencing the development of virtual economies. It became a platform where users could build their digital identities and engage in a multitude of activities.
- **2006: Virtual Reality Movement:** The concept of virtual reality (VR) began to gain traction with devices like the **Oculus Rift** being funded in 2012. This represented a shift towards more immersive experiences.

4.3 Advancements in Technology

- **2010s: Mainstream Attention:** The advent of **mobile technology** and social media platforms further popularised virtual worlds. Games like **Fortnite** began integrating social spaces where players could connect and explore together, enhancing the idea of a Metaverse.
- **2016: Pokémon GO:** This augmented reality game marked a significant moment in mainstream adoption of AR/VR technology, blending the digital and physical worlds and demonstrating the potential of immersive experiences.

4.4 The Emergence of Immersive Platforms

- **2020 and Beyond: The Current Metaverse Landscape:** Major tech companies like Facebook (now Meta) began investing heavily in the development of the Metaverse. Initiatives included the creation of virtual reality platforms such as Horizon Worlds that support social interaction, creativity, and commerce in immersive spaces.

- **NFTs and Blockchain:** The introduction of Non-Fungible Tokens (NFTs) has revolutionised ownership within the Metaverse, enabling true ownership of digital assets and creating new revenue streams for creators.

Following the above-mentioned historical background, this study may create a flowchart of the Metaverse (Fig. 1), driven by gaming, social media, and digital economies.

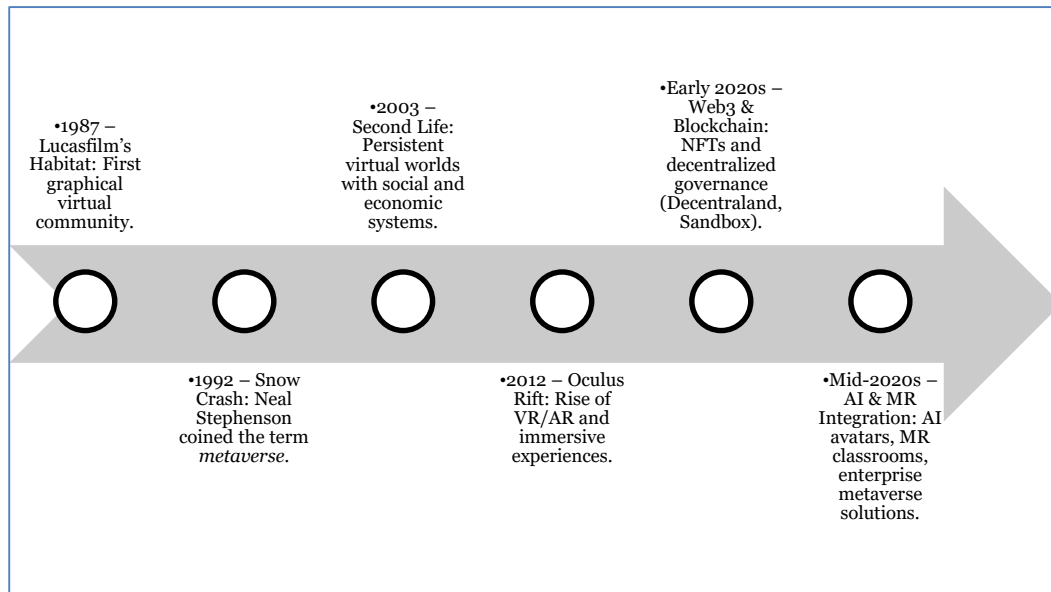


Fig. 1. Flowchart of the Metaverse

The next section will reveal the key components of metaverse technology, which may explain the situation for incorporating such technology in libraries to extend their services for immersive learning experiences.

5 Key Components of the Metaverse

Metaverse is a combination of technologies are as follows:

5.1 Virtual Reality (VR)

Users can fully immerse themselves in simulated environments with the help of VR headsets. This technology opens up exciting avenues for experiences, from engaging gaming sessions to productive virtual meetings.

5.2 Augmented Reality (AR)

By overlaying digital information onto our real-world surroundings, AR enriches daily interactions. This innovation can be easily accessed through devices like smartphones and smart glasses, making it highly versatile.

5.3 Avatars

Users have the opportunity to create customizable avatars that reflect their individuality. This feature fosters personal expression and enhances social engagement within virtual spaces.

5.4 Blockchain Technology

As a key component of the metaverse, blockchain ensures secure transactions and verifies ownership of digital assets, including virtual real estate and non-fungible tokens (NFTs). This technology promotes trust and transparency in the digital landscape.

5.5 Social Interaction

The metaverse fosters meaningful social connectivity, allowing users to participate in events, collaborate on projects, or simply enjoy time with friends in a virtual environment. This sense of community and collaboration can significantly enrich our social experiences. The metaverse is supported by five key components that work together to create a sophisticated and immersive technological environment. These components encompass advanced virtual reality (VR) systems, which provide users with an engaging and lifelike experience; augmented reality (AR) that blends digital elements with the physical world; robust blockchain technology that ensures secure transactions and ownership of digital assets; high-speed internet infrastructure facilitating seamless connectivity; and artificial intelligence (AI) algorithms that personalize user interactions and enhance the overall experience. Together, these elements lay the foundation for a dynamic and interactive metaverse where users can communicate, create, and explore in ways never before possible.

6 Key Technologies of the Metaverse

The Metaverse relies on a combination of advanced technologies to create immersive experiences that blend the digital and physical worlds. This section will give a bird's-eye view of these key technologies:

6.1 Virtual Reality (VR)

VR technology uses headsets like Oculus Rift and HTC Vive to create fully immersive digital environments to facilitate the rich experience of gaming, training simulations, virtual meetings, and educational experiences.

6.2 Augmented Reality (AR)

AR overlays digital information onto the real world through devices like smartphones and AR glasses for Interactive gaming navigation, and industrial applications for maintenance and repair. Examples of the same are Pokémon GO Microsoft HoloLens.

6.3 Mixed Reality (MR)

MR combines elements of both VR and AR, allowing real and virtual objects to interact in real time to create collaborative workspaces where users can interact with 3D digital content in a shared physical space.

6.4 Blockchain Technology

A decentralised ledger system that supports secure transactions and asset ownership, enabling true ownership of digital assets through Non-Fungible Tokens (NFTs) and facilitating virtual economies.

6.5 Artificial Intelligence (AI)

AI enhances user experiences through algorithms that enable human-like interactions and personalised content to provide virtual assistants, procedural content generation, and realistic NPC (non-playable character) behaviours.

6.6 Spatial Computing

Combines AR, VR, and IoT (Internet of Things) to create interactive 3D spaces that respond to user actions vis., such as real-time data visualisation, smart cities, and location-based services.

6.7 5G Connectivity

Next-generation mobile networks providing high-speed internet with low latency to enhance VR/AR applications' responsiveness, enabling seamless interactions in immersive environments.

6.8 3D Modelling and Animation

By using software tools like Blender and Maya, anyone can design everything from virtual avatars to complex landscapes in the Metaverse, creating detailed virtual assets and environments.

6.9 Cloud Computing

Provides the necessary computing power and storage for hosting and rendering virtual environments to facilitate large-scale multiplayer interactions and data-heavy applications.

6.10 User Interfaces (UI) and Experience (UX) Design

Focuses on creating intuitive interfaces for navigation and interaction within the Metaverse for enhancing user engagement through well-designed controls and feedback systems in virtual spaces.

The convergence of these technologies creates a robust framework for developing the Metaverse. As these technologies continue to evolve, they will enable more engaging, interactive, and connected experiences, ultimately shaping the future of how we interact with digital content and one another.

The sections mentioned above have outlined various opportunities for integrating advanced technologies within the expansive realm of libraries. It is essential to explore how these innovations can be effectively applied in dynamic library environments. By examining the current trends, we can identify practical ways to enhance library services and resources, making them more accessible and engaging for users. This exploration encompasses everything from digital cataloguing and electronic resources to virtual reality experiences and artificial intelligence tools that can personalise the user experience. (Amzat & Adewojo, 2023; Ma et al., 2023; Yuan et al., 2025). Let's delve into the specifics of how these technologies can transform traditional library functions and meet the evolving needs of our communities.

7 Libraries in the Digital Age

Transition: From physical books to digital repositories. Libraries have transformed significantly in the digital age, adapting to technological advancements and changing user needs. (Zhang et al., 2023). They now serve not only as repositories of physical books but also as digital hubs that offer a wide range of services and resources.

7.1 Digital Collections and Accessibility

Digital Libraries: Many libraries have established extensive digital collections, offering e-books, audiobooks, and online databases. This expansion allows users to access materials remotely, breaking down geographical barriers. Access to eBooks, journals, and databases.

Accessibility Enhancements: Digital platforms have made it easier for individuals with disabilities to access information. Features such as screen readers and adjustable display settings enhance the user experience.

7.2 Technological Integration

E-Learning Platforms: Libraries are increasingly partnering with educational institutions to provide online courses and resources, bridging the gap between students and educational content.

Interactive Services: Innovative tools like virtual reality (VR) and augmented reality (AR) are being integrated into library experiences. These technologies create immersive learning environments, making information more engaging.

7.3 Community Engagement and Collaboration

Public Programs: Digital literacy programs and workshops are now common, aiming to equip users with essential digital skills. Libraries often hold community events encouraging collaboration and knowledge-sharing.

Collaborations: Libraries frequently collaborate with local organisations, schools, and tech companies to enhance offerings like coding boot camps or maker spaces, thus fostering innovation and creativity.

Libraries are evolving to keep pace with the digital landscape, ensuring that they stay relevant and essential for community education, information access, and cultural preservation. These institutions are no longer just places

for books; they have transformed into vibrant centres for learning and collaboration in the digital age, intending to democratize access to information. The shift from traditional libraries to a presence in the metaverse is not a coincidence; it results from various factors, including both timely inclusions and exclusions. The picture below may give a concept of the entire journey of shifting from traditional to metaverse libraries.

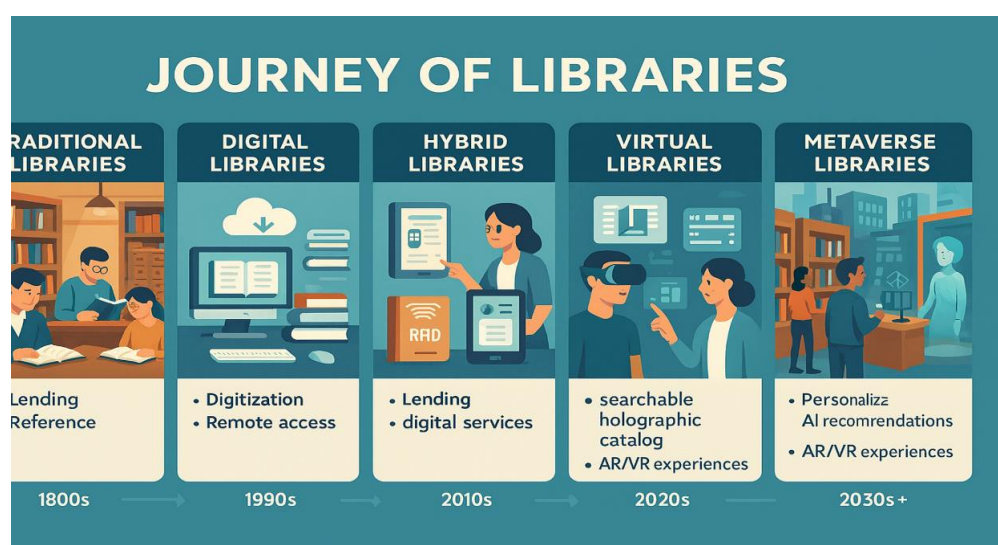


Fig.2. Journey of Libraries

8 Metaverse Libraries

Table 1. Technology and Tools

Technology	Tools and devices
Virtual Spaces and Interactivity	3D Digital Environments: Metaverse libraries provide interactive 3D spaces where users can explore various collections, attend events, and engage in activities. These environments mimic physical libraries but offer much more in terms of accessibility and user interaction. User Avatars: Users can create avatars to navigate these virtual libraries, enabling a more personal and engaging experience. This aspect allows individuals to interact with other users and librarians in real-time.
Enhanced Access to Resources	Vast Collections: Metaverse libraries can host extensive digital collections that include e-books, academic journals, multimedia resources, and even interactive learning modules. Users can access these materials anytime, without the physical limitations of traditional libraries. Global Collaboration: By functioning in the metaverse, libraries can collaborate with global institutions, allowing users to access international resources and participate in cross-cultural exchanges. AR/VR/MR combine real and virtual elements for collaborative learning—users interact with holographic books and avatars in shared spaces.
Community and Collaboration	Virtual Events: Metaverse libraries can host a variety of events, including author talks, workshops, and discussions, where users from around the world can participate without needing to travel. Study Groups and Learning Communities: Users can join study groups facilitated by the library, allowing for collaborative learning and engagement. These groups can focus on specific subjects or projects, fostering a sense of community. Mixed Reality (MR) combines real and virtual elements for collaborative learning—users interact with holographic books and avatars in shared spaces.

Educational Innovations	Immersive Learning Experiences: Utilising augmented reality (AR) and virtual reality (VR), metaverse libraries can offer immersive educational experiences, such as virtual field trips or simulations, enriching the learning process. Gamification: Incorporating game-like elements can make learning more engaging. For example, users could earn points or badges for completing certain tasks or attending events.
Secure Access, Ownership, And Traceability Of Digital Assets	Block Chains ensures secure access, ownership, and traceability of digital assets like rare manuscripts, NFTs, and user credentials.

Sources: Springer College & Research Libraries

9 Benefits of Metaverse Libraries

Metaverse libraries offer a range of advantages that enhance both the user experience and service delivery, transforming library functions for the digital age. (Yuan et al., 2025).

9.1 Resource Optimisation:

Metaverse libraries have the potential to enhance accessibility by generating accurate metadata and providing content in multiple languages through translation services. This innovative approach allows the library team to operate from digital offices within the metaverse, fostering collaboration and community engagement. By integrating acquisition, organisation, and outreach efforts in shared spatial environments, we can create a more inclusive and resourceful experience for all users.

9.2 Enhanced Accessibility

- Global Reach: Users can access resources from anywhere in the world, eliminating geographical barriers.
- 24/7 Availability: Metaverse libraries operate continuously, allowing users to explore collections and access resources at any time.

9.3 Immersive Learning Experiences

- Interactive Learning: Users can engage with materials in 3D environments, making learning more engaging and effective.
- Experiential Education: Incorporating virtual and augmented reality can facilitate unique educational experiences, such as virtual field trips and simulations.

9.4 Community Building

- Social Interaction: Users can connect through avatars, fostering community and collaboration despite physical distances.
- Event Hosting: Libraries can host events, such as workshops and discussions, in virtual spaces, attracting diverse audiences.

9.5 Resource Diversity

- Extensive Collections: Metaverse libraries can offer a vast array of digital resources, including e-books, academic journals, videos, and immersive content.
- Collaborative Resources: Partnerships with global institutions allow access to a wider range of materials and expertise.

9.6 User-Centric Services

- Personalisation: Users can tailor their experiences based on interests and needs, selecting specific areas of focus within the library.

- **Engagement Tools:** Features like gamification can encourage participation and exploration of library resources.

9.7 Innovative Technology Integration

Emerging Tech: Metaverse libraries can integrate cutting-edge technologies, such as artificial intelligence, to provide personalised recommendations and enhance user interaction.

Flexible Learning Environments: Users can engage in various formats, from traditional reading to interactive workshops, catering to diverse learning preferences.

Metaverse libraries are at the forefront of a new era in information access, blending technology, community, and education to create enriching experiences for users. They hold the potential to redefine the traditional library model for modern needs.

In India, a few libraries have started to incorporate this technology. Among them, N. L. Dalmia Institute of Management Studies and Research (Maharashtra), Central Library, Rajiv Gandhi University (Arunachal Pradesh), University of Petroleum and Energy Studies (UPES), Dehradun and AIIM, New Delhi developed a Metaverse Library model integrating VR for immersive scholarly interaction. Students can explore interactive environments beyond physical constraints, enhancing engagement and research ijrpls.in. It uses encryption, authentication, and firewalls to protect virtual library environments. (Sinha, 2023). In a global perspective, Decentral and Virtual Library (Global), Second Life's Virtual Libraries (USA & Europe), Springer Nature MR Classrooms (Europe), and MIT Media Lab (USA) use a blockchain-based metaverse platform hosting virtual libraries and museums. Users can access digital books, attend lectures, and even purchase NFTs representing rare manuscripts. Also, experimenting with blockchain-secured digital archives ensures authenticity and traceability of rare documents in decentralised metaverse environments. To summarise, the research highlights that India is concentrating on addressing geographic challenges, boosting academic participation, and safeguarding cybersecurity, while the global emphasis is on blockchain ownership, collaborative mixed-reality classrooms, and decentralised governance (ijrpls.in, INFLIBNET, IRMA-International.org).

10 Challenges and Considerations

Metaverse libraries face significant challenges in accessibility, equity, privacy, copyright, and technical-ethical domains that must be addressed for sustainable adoption. Accessibility requires inclusive design features like customizable avatars, voice navigation, and low-bandwidth AR alternatives to accommodate users with disabilities or varying hardware capabilities. Equity efforts focus on bridging the digital divide through subsidised VR access and training programs for underserved regions, preventing exclusion of low-income or rural populations. Privacy concerns demand robust data protection in immersive environments, including anonymised avatars and GDPR-compliant tracking to safeguard biometric and behavioural data (Qi, 2022). Copyright issues involve managing digital rights for virtual replicas via blockchain-NFT systems, ensuring fair use while compensating creators in shared spaces. Technical and ethical considerations encompass scalable infrastructure, digital literacy initiatives, and AI-driven content moderation to combat misinformation and harassment.

11 Future Impact

- **AI Integration:** Utilising advanced artificial intelligence to offer personalised recommendations that cater to individual preferences, while also employing virtual assistants to optimise tasks and boost productivity across multiple platforms (Amzat & Adewojo, 2023).
- **Immersive Learning:** Creating engaging educational settings through realistic simulations and compelling interactive storytelling (Olmedo-Pagés et al., 2025; Yang, 2023), which captivate learners and enhance their critical thinking and problem-solving abilities (Johan et al., 2023; Olmedo-Pagés et al., 2025; Qi, 2022).
- **Global Collaboration:** Fostering a culture of knowledge exchange that goes beyond geographical limits, allowing professionals and researchers from various fields to work together on innovative initiatives, exchange best practices, and drive global advancement (Ajani et al., 2023; Wang et al., 2020).
- **Sustainability:** Adopting strategies designed to lessen the use of physical resources, such as reducing waste, supporting digital alternatives, and promoting sustainable practices within organisations to aid environmental conservation and the welfare of future generations (Owusu & Ofori-Owusu, 2024).
- **Haptic Feedback Devices:** Offering users tactile interactions that further enhance the immersive experience.

- **Interoperability:** Guaranteeing smooth experiences across different platforms and environments, enabling users to easily transition between various Metaverse experiences (Olmedo-Pagés et al., 2025) (Al-kfairy et al., 2024).

The Metaverse revolutionises digital interaction by offering immersive spaces where users engage through avatars. It significantly enhances delivery services, user engagement, and the targeted dissemination of information, all within a 24/7 virtual learning environment. Patrons have the opportunity to explore collections directly through their avatars. Intelligent avatars provide crucial reference assistance, ensuring users can easily locate resources that meet their needs. Leveraging user experience, learning goals, and search history, these avatars deliver tailored resource recommendations at precisely the right time, pace, and place. Users can step beyond traditional reading and dive into 3D simulations of ancient civilisations, virtual labs, and AI-driven knowledge landscapes. Moreover, by preserving cultural heritage digitally—encompassing manuscripts, artefacts, and oral histories—in a 3D Metaverse, it ensures interactive access for future generations. This powerful approach dismantles barriers to equitable access, overcoming challenges related to physical space, disabilities, or socio-economic circumstances through the effective utilisation of Metaverse libraries.

12 Conclusion

The evolution of the Metaverse, transitioning from rudimentary text-based games to sophisticated, immersive platforms, marks a significant milestone in the landscape of technological innovation. As cutting-edge technologies continue to advance and converge, the opportunities for even more enriched and transformative virtual experiences expand exponentially. The future of the Metaverse holds the promise of redefining our perceptions of reality, how we connect, and our engagement with digital content. Metaverse libraries symbolise a revolutionary approach to knowledge, transcending the traditional notion of merely storing books. They become dynamic spaces where learning transforms into a collaborative, immersive journey—one that is both vivid and enduring. These libraries skillfully blur the lines between the realms of imagination and reality, inviting every reader to step into the role of an explorer, navigating through a landscape of ideas and insights. Forget the concept of digital repositories; Metaverse libraries are vibrant ecosystems for learning, rich with interactivity and security. They foster an environment where knowledge is not just acquired but experienced, making education a participatory adventure that resonates deeply and lingers long after the experience ends.

Declaration of AI and AI-Assisted Technologies used in Manuscript Writing Process: Grammarly version 1.2 for spelling, grammar check, and accurate sentence construction.

References

1. Amzat, O. B., & Adewojo, A. A. (2023). Metaverse-infused academic libraries: a glimpse into the future. *Library Hi Tech News*, 40(10), 17–19. <https://doi.org/10.1108/LHTN-10-2023-0187>
2. Anna, N. E. V., Harisanty, D., & Ismail, N. (2023). Libraries on metaverse, do they exist? *Library Hi Tech News*, 40(6), 1–2. <https://doi.org/10.1108/LHTN-02-2023-0019>
3. CANNAVÒ, A., ARRIGO, G., VISCONTI, A., De LORENZIS, F., & LAMBERTI, F. (2025). Designing social immersive virtual environments for the Metaverse: The case study of MetaLibrary. *Virtual Reality and Intelligent Hardware*, 7(3), 279–298. <https://doi.org/10.1016/j.vrih.2025.04.002>
4. Chow, Y. W., Susilo, W., Li, Y., Li, N., & Nguyen, C. (2023). Visualisation and Cybersecurity in the Metaverse: A Survey. *Journal of Imaging*, 9(1). <https://doi.org/10.3390/JIMAGING9010011>
5. Guo, Y., Yuan, Y., Li, S., Guo, Y., Fu, Y., & Jin, Z. (2024). Applications of metaverse-related technologies in the services of US urban libraries. *Library Hi Tech*, 42(5), 1477–1495. <https://doi.org/10.1108/LHT-10-2022-0486>
6. Johan, R. C., Ardiansah, Damayanti, & Suhardini, D. (2023). Systematic Literature Review on Metaverse Library. *Khi-zanah Al-Hikmah*, 11(2), 327–334. <https://doi.org/10.24252/KAH.V11I2CF2>
7. Sağsan, M. (2024). Metalibrarians as Residents Beyond the Second Life: An Overview of Contradictory Concepts on “The Metaverse.” *Bilgi Dünyası*, 25(1), 32–52. <https://doi.org/10.15612/BD.2024.736>
8. Tella, A., Ajani, Y. A., & Ailaku, U. V. (2023). Libraries in the metaverse: the need for metaliteracy for digital librarians and digital age library users. *Library Hi Tech News*, 40(8), 14–18. <https://doi.org/10.1108/LHTN-06-2023-0094>
9. Yuan, Y., Wu, J., Yu, Y., He, C., & Jin, Z. (2025). Positive and Negative Topics of Metaverse Library from China: A Sentiment-Topic Analysis. *Libri*, 75(3), 297–310. <https://doi.org/10.1515/LIBRI-2024-0128>
10. Amzat, O. B., & Adewojo, A. A. (2023). Metaverse-infused academic libraries: a glimpse into the future. *Library Hi Tech News*, 40(10), 17–19. <https://doi.org/10.1108/LHTN-10-2023-0187>

11. Anna, N. E. V., Harisanty, D., & Ismail, N. (2023). Libraries on metaverse, do they exist? *Library Hi Tech News*, 40(6), 1–2. <https://doi.org/10.1108/LHTN-02-2023-0019>
12. CANNAVÒ, A., ARRIGO, G., VISCONTI, A., De LORENZIS, F., & LAMBERTI, F. (2025). Designing social immersive virtual environments for the Metaverse: The case study of MetaLibrary. *Virtual Reality and Intelligent Hardware*, 7(3), 279–298. <https://doi.org/10.1016/j.vrih.2025.04.002>
13. Chow, Y. W., Susilo, W., Li, Y., Li, N., & Nguyen, C. (2023). Visualisation and Cybersecurity in the Metaverse: A Survey. *Journal of Imaging*, 9(1). <https://doi.org/10.3390/JIMAGING9010011>
14. Guo, Y., Yuan, Y., Li, S., Guo, Y., Fu, Y., & Jin, Z. (2024). Applications of metaverse-related technologies in the services of US urban libraries. *Library Hi Tech*, 42(5), 1477–1495. <https://doi.org/10.1108/LHT-10-2022-0486>
15. Johan, R. C., Ardiansah, Damayanti, & Suhardini, D. (2023). Systematic Literature Review on Metaverse Library. *Khi-zanah Al-Hikmah*, 11(2), 327–334. <https://doi.org/10.24252/KAH.V11I2CF2>
16. Sağsan, M. (2024). Metalibrarians as Residents Beyond the Second Life: An Overview of Contradictory Concepts on “The Metaverse.” *Bilgi Dünyası*, 25(1), 32–52. <https://doi.org/10.15612/BD.2024.736>
17. Tella, A., Ajani, Y. A., & Ailaku, U. V. (2023). Libraries in the metaverse: the need for metaliteracy for digital librarians and digital age library users. *Library Hi Tech News*, 40(8), 14–18. <https://doi.org/10.1108/LHTN-06-2023-0094>
18. Yuan, Y., Wu, J., Yu, Y., He, C., & Jin, Z. (2025). Positive and Negative Topics of Metaverse Library from China: A Sentiment-Topic Analysis. *Libri*, 75(3), 297–310. <https://doi.org/10.1515/LIBRI-2024-0128>

Digital Libraries and Repositories: Future Directions

Digital Libraries through the Lens of YouTube: Content and Sentiment Insights

Ramakrushna Baral ¹[0009-0000-0718-0065], Fazlunnisa H ²[0000-0001-9780-4520], Kratika Paraste ³[0009-0000-8682-7583],

Mounapriya R ⁴[0009-0007-0430-8685] and Jidigunta Jaji ⁵[0009-0000-9703-6330]

^{1,2} Department of Library and Information Science, University of Madras

³ Rajiv Gandhi National Aviation University

⁴Symbiosis Institute of Business Management, Bengaluru

⁵Library and information assistant trainee, IISER TVM

ramakrushnabaral1715@gmail.com

Abstract. With the rapid expansion of digital information and online learning, digital libraries are crucial for equitable, remote access to knowledge. YouTube has become a popular platform for sharing educational content and public opinions through comments. Despite many digital library videos on YouTube, little research has explored how this content is structured, engaged with, and perceived. This study analyses YouTube videos on “Digital Library” to understand content features, audience engagement, and public sentiment. Metadata and comments were collected on 15th January 2025 using the YouTube Data API v3 and Webometric Analyst. Of 553 videos, 216 were analysed. Content analysis examined growth trends, video features, and key discussion themes, while sentiment and emotion analysis used Excel and the syuzhet package of R-studio based on the NRC Emotion Lexicon. Results show a steady rise in digital library content, especially recently. Sentiment analysis reveals mostly positive emotions like trust, anticipation, and joy, indicating favourable perceptions. The study provides insights to improve digital library communication and engagement.

Keywords: Digital Library, YouTube, Sentiment Analysis, Content Analysis, Webometric Analyst

1 Introduction

In the current digital age, social media platforms like YouTube have undoubtedly become one of the most significant platforms for sharing information and providing users with access to a wide range of educational materials. Beyond amusement, it has evolved into a powerful information tool that helps users learn new things, experiment with creative methods, and stay up-to-date with the latest trends in their area of interest. The fifth fundamental law of library science debates the importance of adapting libraries to the changing times [18]. In the age of rapid global transformation, libraries have shifted from traditional to Digital Libraries (DL). Digital Libraries are platforms that allow users to access digital resources without constraints on space or time. The services of DL play a vital role in the modern digital education system, particularly by making knowledge more accessible to everyone, regardless of location. As repositories of organised knowledge, digital libraries primarily rely on technology-driven outreach to engage users in a digital environment. YouTube complements this mission by serving as a widely accessible pathway where users can engage with library services, learn about digital library resources, and enhance their digital literacy skills [4]. As platforms like YouTube, which play an increasingly significant role in disseminating educational content, become essential, it is important to understand how digital library-related videos engage audiences. Analysing YouTube content, due to its interactive and user-driven nature, offers valuable insights into both the information shared and users’ sentiments and perceptions of digital library-related content [4,7]. This approach is highly useful for evaluating viewers’ emotional and intellectual states, understanding how they engage with the videos for learning, and assessing how these interactions are reflected in the comments [2,7]. This study aims to explore the characteristics and engagement of YouTube videos about digital libraries, using a combination of metadata analysis and sentiment analysis. Sentiment analysis is the technical term used for opinion mining. It eventually detects the impact and influence of specific content or commodities in the market by categorising them as positive or negative. Sentiment analysis of comments allows for assessing how audiences respond to specific phenomena. Similarly, topic analysis helps in understanding different aspects of a particular phenomenon.

2 Literature Review

YouTube is the second most popular search platform after Google. It started emerging as a learning tool in the early 2010s alongside the growth of MOOCs and educational channels offering structured instructional content. Its crucial role as a medium for distributing academic content became especially clear during the COVID-19 pandemic, when the global education system transitioned to online learning [17]. Research has underscored YouTube's effectiveness in facilitating online education, highlighting its accessibility, user-friendliness, and support for active and collaborative learning, parental control options, and the capacity to choose content suitable for different age groups [17, 23]. Research studies analysing educational YouTube channels, such as e-Gyankosh, IGNOU, and e-PG Pathshala, found that they were well received by audiences [20], and similar studies found them useful among learners with disabilities [25]. Controlled experiments and systematic reviews have consistently demonstrated that YouTube-based instructions yield positive learning outcomes, with students finding it an effective, engaging, and motivating platform [12, 24, and 28]. Various studies have noted that to better understand audience engagement, sentiment analysis is used to assess public opinion and emotion through text-based interactions, such as comments and reviews, on digital platforms like YouTube and X. Sentiment analysis combines text mining, natural language processing, and machine learning to determine whether opinions expressed are positive, negative, or neutral [11,20]. With the rapid growth of online media, sentiment or emotion analysis has become widely used in data-driven domains, such as product reviews, public opinion monitoring, and social media analytics [15, 26, 30]. Within the information science context, Rout et al. analysed YouTube comments on research ethics and misconduct, Deori et al. studied responses to videos on Koha and DSpace, and Sharma et al. examined NDLI App reviews, revealing that only one-third of users were satisfied [5,21,22]. Research on digital library awareness and use reveals both benefits and obstacles. Digital libraries provide easy access, integrated search tools, and quick downloads, but face issues like underutilisation, insufficient training, and digital literacy gaps. Studies show moderate awareness levels, yet actual usage remains low. Faculty and researchers are generally more skilled in utilising digital libraries compared to students, who often prefer print or need more support [1,8,10,14] . Although no specific analysis has focused on YouTube content about digital libraries, previous research offers useful insights. Parabhoi and Saha analysed YouTube comments on Koha software, examining intention, subjectivity, emotion, and word frequency to gauge audience perception and involvement, while Hoffmann used automatic sentiment analysis on U.S. presidential campaign speeches to explore emotional language employed by Trump and Clinton [9,16]. These studies highlight the advancing sophistication of sentiment analysis methods that go beyond basic lexicon models. Building on these techniques, this study aims to analyse YouTube videos about digital libraries using content and sentiment analysis to understand viewers' perceptions and engagement with these educational resources.

3 Research Gap

While digital libraries are becoming more widespread in academic and public domains, there is a limited understanding of how users perceive and discuss these services on open social media platforms. YouTube, a top video-sharing platform, functions both as an information resource and a space for public feedback, where users express opinions, learning experiences, and expectations. Analysing viewers' interests, opinions, content characteristics, and public reactions is crucial, as these insights help library professionals and content creators to evaluate public awareness, uncover information needs, and improve digital library outreach and engagement. This study is driven by the lack of user-focused, sentiment-based analyses of YouTube content specifically related to digital libraries.

4 Objectives

The major objective of the study is to understand how viewers on YouTube perceive and engage with content related to the concept of "Digital Library". Some other specific objectives are:

- To determine the characteristics, such as resolution and release licenses, used by the YouTube channels for posting YouTube content on the "Digital Library".
- To quantify the number of videos on "Digital Library", and assess the growth pattern.
- To investigate viewer sentiments of the selected videos to gain insight on the user perception of digital libraries.
- To identify major themes on the topic using topic analysis, to understand the contextual focus area of content creators.

5 Methodology

The search term “Digital Library” was used on 15th January 2025 to identify videos posted on YouTube. Using Webometric Analyst 4.0 with YouTube API v3, metadata related to the keyword “Digital Library” was extracted. The primary search initially provided us with the details of the 553 videos, including their metadata, such as video ID, channel ID, channel title, video title, description, duration, dimension, and definition, view count, like count, comment count, and license type. The data was initially collected in text files, which were then converted into an Excel spreadsheet. This spreadsheet was refined by removing short videos under 30 seconds, leaving 216 videos for analysis. Comments related to these videos were retrieved using their unique IDs, totalling 3,749 comments in various languages. Only comments in English were used for the final analysis. The study examined both emotions and sentiments conveyed in these comments. Quantitative metrics such as view, like, and comment counts helped identify the most engaging and popular videos, while an NRC lexicon-based sentiment analysis approach assessed viewers’ emotions and attitudes. The “syuzhet” package in R-Studio, which utilises the NRC Emotion Lexicon, classified comments into eight basic emotions- joy, trust, anger, fear, surprise, sadness, anticipation, and disgust and two main sentiments: positive and negative. Additional R libraries, such as “word_Cloud”, were used to identify the most common topics in the dataset’s titles and descriptions.

6 Findings and Discussions

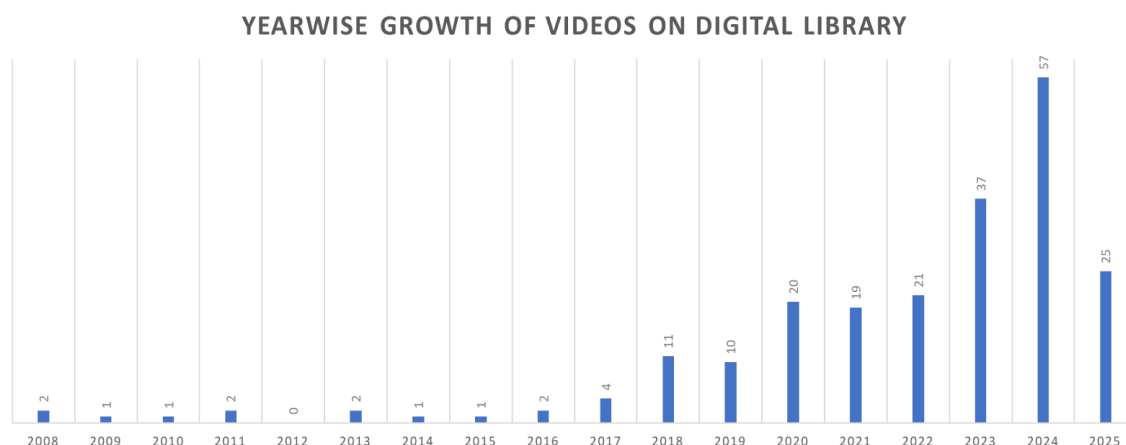


Fig. 1. Growth of YouTube Videos on the Digital Library

6.1 Yearly Growth of YouTube Content on Digital Library

Figure 1 depicts the timeline of the growth of YouTube content related to “DL” from 2008 to 2025. It is observed that the first video on DL was uploaded to YouTube in 2008. The minimum number of videos on DL was uploaded during the initial 10 years. Since 2018, the number of videos on the concept has steadily grown. The highest number of videos was uploaded in 2024. It is expected that in 2025, the numbers will also increase, as data was collected up to January 2025, by which time 25 videos had been uploaded on the concept. It is also observed that 92% of videos uploaded by YouTubers on the concept are in HD, while the remaining 8% are in SD. Similarly, 98% of videos are uploaded under a YouTube Licence, and 2% are uploaded under a “Creative Commons” Licence. The notable rise in Digital Library-related videos in the last five years can be linked to several factors. The post-pandemic boost in online education, universities’ greater focus on digital resources, students preparing for competitive exams via online platforms, and government initiatives like promoting the National Digital Library of India (NDLI) all played a role in increasing content creation. Future research could further explore these factors by examining policy and institutional data.

6.2 Channel Productivity using Lotka’s Law

Lotka’s law specifies that a small number of contributors produce a large quantity of works, while most produce only a few [13]. This pattern has been widely observed in scientific publishing and can also be applied to digital

content creation, such as YouTube videos on digital libraries. It is important to note from Table 1 that most channels have created only a single video on DL, and a small number have posted multiple videos on DL. Therefore, the observed distribution aligns with the general principle of Lotka's Law.

Table 1. Lotka's Law of Channel Productivity

Number of Channels	Number of videos on DL
1	5
1	4
4	3
9	2
177	1

Table 2. Top-ranked videos on Digital Library Based on Like Count, View Count and Comment Count

S.No	Title	Channel Title	Like Count (Rank)	View Count (Rank)	Comment Count (Rank)
1	Create your own digital library	TREX LABS	5186 (1 st)	95707 (6 th)	555 (1 st)
2	मेरी लाइब्रेरी में कितना पैसा लगा (*How much money was invested in my library) Library opening cost 2023	Ladka Banaras Wala	4785 (2 nd)	490766 (1 st)	510 (2 nd)
3	National Digital Library of India	NDL India	2879 (3 rd)	284741 (3 rd)	89 (8 th)
4	2023 में लाइब्रेरी open करना है तो सब कुछ समझें (*If you want to open a library in 2023, understand everything) how to open library in 2023	Ladka Banaras Wala	2857 (4 th)	194313 (5 th)	332 (4 th)
5	Yugavandna library gandhinagar no. 1 library in india one of the best library for upsc student	YUGVANDANA LIBRARY	2508 (5 th)	201385 (4 th)	114 (5 th)
6	2025 library business plan in india सुनहरा मौका 0% इन्वेस्टमेंट (*Golden Opportunity with 0% Investment)	Np Speaks { Nilesh Patle }	2315 (6 th)	286512 (2 nd)	345 (3 rd)
7	Digital Library Punjab Bihar DSSB NTA-NET (Library Science)	Library Science By Target Abhi	1707 (7 th)	-	109 (6 th)
8	Aliflaila: Connecting Kids with Books Pakistan's First Kids Digital Library	AlifLaila Kids Digital Library	935 (8 th)	-	-
9	Library business plan in india library business plan in hindi library business plan	Shobhit Mishra	854 (9 th)	88139 (7 th)	84 (9 th)
10	How to start a small library business library business investment modern library business ASK	Anoop Singh Kamboj	765 (10 th)	53399 (9 th)	-
11	How to Access and Use National Digital Library of India	NDL India	-	64281 (8 th)	-
12	Brewster Kahle: A digital library, free to the world	TED	-	47355 (10 th)	-
13	Perpustakaan Digital dengan Google Site Digital Library Google Site	Sunsun GTSN	-	-	96 (7 th)
14	The internet archive wants to be a digital library for everything sunday today	TODAY	-	-	77 (10 th)

Note: The table lists relevant YouTube videos for the study. Titles originally in Hindi are retained in their original form, followed by their English translations in parentheses for clarity. Like, view, and comment counts are presented along with their respective ranks.

6.3 Prolific videos by Likes, Comments, and Views

The appeal of social media platforms like YouTube is that they allow their users to share their appreciation and emotions for content by liking posts or videos, which acts as a sign of recognition and support for creators' efforts. Table 2 features 14 videos that include the top 10 most liked, viewed, and commented videos on misinformation. The top 10 liked videos are ranked by the number of likes received by content on "Digital Library". Among all the content, the video titled "Create Your Own Digital Library" from the YouTube channel "TREX LABS" stands out, with the highest number of likes (5,186) for its concept. The second position is held by "मेरीलाइब्रेरीमेंकितनापैसालगा| Library opening cost 2023", uploaded by the channel titled "Ladka Banaras Wala", which has received 4,785 likes. Interestingly, in third place is the "National Digital Library of India" video from the channel "NDL India", the official YouTube channel of the National Digital Library of India, followed by "2023 मेंलाइब्रेरीopen करनाहैतोसबकुछसमझे| How to Open Library in 2023", which garnered an impressive 2,857 likes, placing it fourth overall.

Similarly, viewers search for videos on popular social media platforms like YouTube based on their information needs. The number of views a video has received is regarded as a benchmark that determines its importance and popularity. Table 2 also shows the titles and channels of the top 10 most viewed videos on DL on YouTube, along with the channel names, view counts, and the percentage of total views for each video. Notably, the video titled "मेरीलाइब्रेरीमेंकितनापैसालगा| Library opening cost 2023" from the channel "Ladka Banaras Wala" has received the highest number of views. The video titled "2025 Library Business Plan In India || सुनहरामौका 0% इन्वेस्टमेंट", uploaded by the channel "Np Speaks {Nilesh Patle}", ranks second with a total of 286512 views. In third place, the video titled "National Digital Library of India", uploaded by the channel "NDL India", has been viewed by 284741 viewers; another video from the same channel, titled "YUGAVANDNA LIBRARY GANDHINAGAR |NO. 1 LIBRARY IN INDIA | ONE OF THE BEST LIBRARIES FOR UPSC STUDENT", is the fourth most viewed video on the topic DL.

Another interesting feature of social media platforms is their interactivity, allowing for instant feedback through likes or, even better, comments. Viewers can comment on the content or the concept to express their feelings, sentiments, and emotions. Broadly, these feelings can be positive, negative, or neutral. Table 2 also displays the top 10 most commented videos on YouTube at the time of data collection. It is noted that the video titled "Create Your Own Digital Library" has the highest number of comments. Similarly, the video titled "मेरीलाइब्रेरीमेंकितनापैसालगा| Library opening cost 2023" has the second-highest number, with 510 comments. Additionally, the video uploaded by the channel "Np Speaks { Nilesh Patle }" titled "2025 Library Business Plan In India || सुनहरामौका0% इन्वेस्टमेंट" holds the third position in comments.

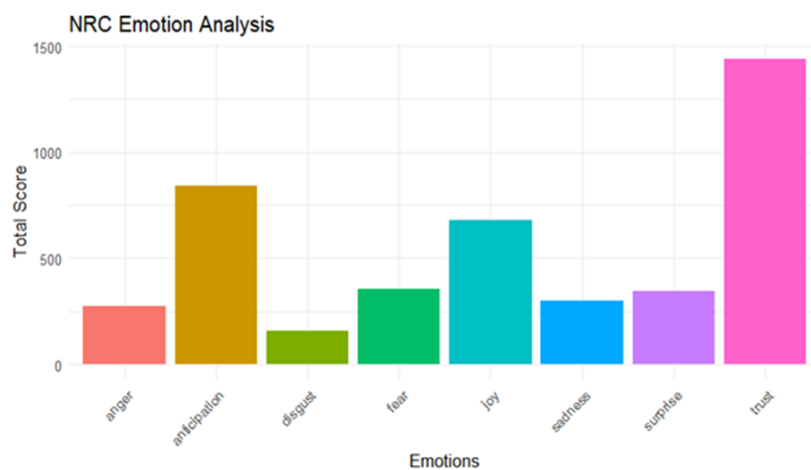


Fig. 2. Distribution of emotions in comments

6.4 Exploring relationships among prolific videos

The video "मेरीलाइब्रेरीमेंकितनापैसालगा| Library opening cost 2023" by Ladka Banaras Wala has the highest views at 490,766 and ranks second for both likes (4,785) and comments (510). Similarly, "2025 Library Business Plan In India || सुनहरामौका0% इन्वेस्टमेंट" by Np Speaks has 286,512 views, making it second in views, but it has only 2,315 likes, placing it sixth in likes and 345 comments, which is third in comment count. The video "Create Your

Own Digital Library” by TREX LABS has the highest number of comments (555) and likes (5,186), but ranks sixth in views with 95,707. Meanwhile, “2023 में लाईब्रेरी open करना है तो सब कुछ समझे” by Ladka Banaras Wala holds the fourth position in comments (332), fourth in likes (2,857), and fifth in views (194,313). At least five of the top 10 videos focus on the financial aspects of digital libraries, including setup costs, business planning, and investment opportunities. The high view count of such videos indicates that a large audience is interested in starting or managing digital libraries as businesses rather than solely as academic or educational resources. The video “National Digital Library of India” by NDL India ranks third in views (284,741) and third in likes (2,879), despite being eighth in comments (89), demonstrating that official digital library platforms attract substantial audiences. It is also clear that “How to Access and Use National Digital Library of India” is the eighth most viewed (64,281), showing that users seek guidance on navigating digital library platforms.

6.5 Sentiment Analysis

Sentiment analysis was included in this study to capture viewers’ emotional and attitudinal reactions to “digital library” related content, which metadata alone cannot reveal. Although metrics like view counts and likes reflect popularity, sentiment analysis offers a more nuanced understanding of users’ perceptions regarding the usefulness, reliability, and relevance of digital libraries. Recognising emotions such as trust, anticipation, anger and sadness is especially valuable for library professionals aiming to enhance digital services and communication.

Viewers’ emotions on DL videos. Sentiment analysis enables researchers to grasp the emotions linked to a specific phenomenon. In this study, we used the syuzhet package with the NRC Lexicon to analyse YouTube comments across eight emotion categories: anger, fear, anticipation, trust, surprise, sadness, joy, and disgust [6,19]. Trust emerged as the most common emotion (1439), followed by anticipation (842) and joy (677), as illustrated in Figure 2. These positive emotions dominate the overall sentiment pattern among viewers, aligning with other research showing trust, joy, and anticipation as prevalent in positive content [19, 29]. Viewers generally responded to the digital library videos on YouTube with confidence and positive engagement, as reflected by the sentiments of trust, anticipation, and joy representing confidence, expectancy, and satisfaction, respectively. Negative emotions such as fear (354), surprise (345), sadness (301), anger (274), and disgust (156) were less frequent, indicating limited negative reactions. Emotionally, surprise often indicates recognition or enjoyment of new information, while sadness, anger, and disgust tend to reflect disappointment or negative judgments about the content. This pattern is consistent with previous sentiment studies; for example, Vosoughi et al. found that trust, anticipation, and joy are linked to truthful or positive content, whereas fear, disgust, and surprise are more common with negative or false content [29]. Overall, the viewer sentiment pattern reflects a predominantly positive response, characterised by trust, joy, and anticipation, with relatively few negative emotions like sadness, anger, or disgust.

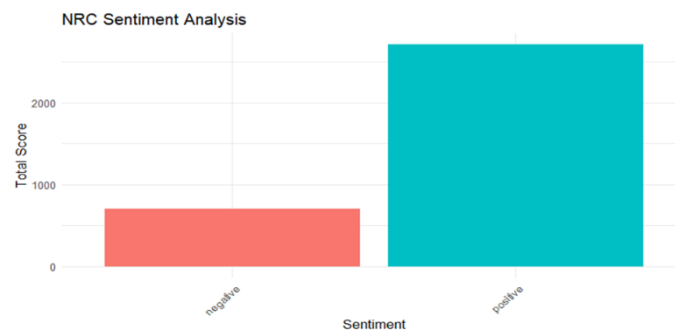


Fig. 3. Distribution of viewers sentiments

Viewers’ sentiments on DL videos. Figure 3 illustrates how viewers who watched videos about digital libraries (DL) on YouTube felt overall, with feelings divided into positive and negative sentiments. The analysis, carried out with the syuzhet package and the NRC Lexicon, indicates that positive sentiment is predominant, with 2,714 instances compared to 707 negatives. This significant variation indicates that most viewers appreciated the content on the Digital Library. This type of sentiment distribution demonstrates that people not only like the videos but also support the ideas and information shared by content creators. Similar patterns have been observed in earlier studies, suggesting that educational and informational videos often generate more positive sentiment due to their perceived value and relevance [27]. The dominance of the positive sentiments in Figure 3 can be ascribed to the

accessibility, utility, and increasing relevance of digital libraries in academic and public domains. Digital libraries are also often associated with providing information accessible to all, which can make people feel more empowered and trusting [3].

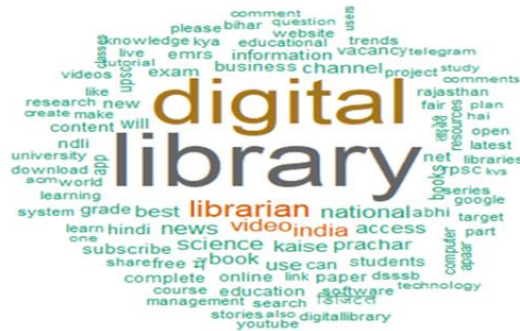


Fig. 4. Word cloud of frequently occurring words

6.6 Topic Analysis

All the titles and descriptions of the observed videos are combined to provide a brief understanding of the topics that content creators have posted on YouTube. The word cloud in Figure 4 represents the most frequently occurring words in the titles and descriptions of the selected YouTube videos. The individual terms appearing in the “title” and “description” are combined prior to preparing the word cloud. These words are fed into the “wordcloud” library of R Studio software to prepare the word cloud. The thickness and size of the words in Figure 4 represent the frequency with which words are used by the content creators in their video titles and descriptions. The most prominent words in the word cloud include “digital” and “library”, confirming that these videos are primarily on the concept of digital libraries. Other frequently occurring words, such as “librarian,” “national,” “India,” “video,” “science,” “university,” “students,” and “research,” indicate that there is a strong association with academic, research, and informational contexts. Words like “book,” “learning,” “education,” “exam,” “content,” “subscribe,” “system,” and “website” suggest that users and content creators discuss digital libraries in terms of educational resources and online learning platforms. The presence of terms like “NDLI” signifies there is discussion about the “National Digital Library of India”. We also observed that terms like “DSSSB,” “PSC,” “UPSC,” “government,” “vacancy,” and “business” are present in the word cloud, which shows that many YouTube videos are related to career opportunities, government initiatives, and professional development through digital library services. The inclusion of terms such as “free,” “open,” “Google,” “download,” and “link” suggests a focus on accessibility and resource sharing within digital library platforms.

7 Conclusion

The study has focused on the phenomenon of YouTube content related to Digital Libraries through both content and sentiment analysis. The findings of this work provide a diverse understanding of how DL is perceived, engaged with, and presented by viewers on YouTube. The analysis of 216 relevant videos and viewers’ comments reveals interesting patterns in audience engagement, emotional reactions, and the thematic focus of content creators. This highlights the relationship between the Digital Library concept and user participation. The sentiment analysis results suggest that there are many positive responses from viewers towards the content, compared to negative sentiments on the videos related to this phenomenon. Among all eight emotions, trust, anticipation, and joy are the most prominent positive emotions. These findings indicate that audiences perceive these contents as reliable, beneficial, and useful for the learning environment. The presence of negative emotions such as sadness, anger, and disgust at minimal levels signals low dissatisfaction among viewers. Furthermore, engagement metrics for the most popular videos show that both individual channels, like Ladka Banaras Wala, and institutional channels, such as NDL India, play significant roles in promoting DL among viewers. Similarly, some videos focus on library business models, setup costs, and investment opportunities, highlighting the business aspect of digital library discourse.

The growth pattern of content related to the Digital Library illustrates the evolution of digital library discourse on YouTube. It is identified that there was a significant increase in the number of videos uploaded from 2018 onwards, after a period of minimal activity. This aligns with the digital environment of the post-pandemic era, where digital education has become essential. Similarly, applying Lotka’s Law to YouTube content creators reveals interesting parallels between scholarly publications and YouTube video productions. It is noted that a small

number of channels have generated multiple videos on digital libraries, offering an opportunity for library professionals to produce more content and promote the Digital Library. Topic analysis of videos also aids in understanding various related themes of the Digital Library. The word cloud indicates that the discourse on YouTube is primarily centred on academic and informational contexts. Additionally, business, entrepreneurship, government, and career themes suggest that YouTube has become not only a repository of digital literacy resources but also a bridge connecting information science with real-world applications.

Declaration of AI and AI-Assisted Technologies used in Manuscript Writing Process: Not Applicable

References

1. Alokluk, J. A., & Al-Amri, A. (2021). Evaluation of a Digital Library: an Experimental study. *Journal of Service Science and Management*, 14(01), 96–114. <https://doi.org/10.4236/jssm.2021.141007>
2. Bindhumol, M., Singh, T., & Patra, P. (2024). Sentiment Analysis using YouTube Comments. 2022 13th International Conference on Computing Communication and Networking Technologies (ICCCNT), 1–7. <https://doi.org/10.1109/icccnt61001.2024.10724166>
3. Candela, L., Castelli, D., Fox, E. A., & Ioannidis, Y. (2010). On Digital Library foundations. *International Journal on Digital Libraries*, 11(1), 37–39. <https://doi.org/10.1007/s00799-011-0063-y>
4. Cho, A. (2013). YouTube and academic libraries: Building a digital collection. *Journal of electronic resources librarianship*, 25(1), 39–50.
5. Deori, M., Kumar, V., & Verma, M. K. (2021). Analysis of YouTube video contents on Koha and DSpace, and sentiment analysis of viewers' comments. *Library Hi Tech*, 41(3), 711–728. <https://doi.org/10.1108/lht-12-2020-0323>
6. Fijačko, N., Greif, R., Štiglic, G., Kocbek, P., & Abella, B. S. (2023). Sentiment Analysis of Social Media Users' Emotional Response to Sudden Cardiac Arrest during a Football Broadcast. *JAMA Network Open*, 6(6), 8–11. <https://doi.org/10.1001/jamanetworkopen.2023.19720>
7. Faizi, R., El Fkihi, S., Rudneva, M., & Chiheb, R. (2020). Using Sentiment Analysis To Examine Students' feedback on Youtube Educational Videos. In *ICERI2020 Proceedings* (pp. 9010–9013).
8. Hindagolla, B. (2017). The use of digital library by art undergraduates in sri lanka: a case study at university of peradeniya. *PEOPLE: International Journal of Social Sciences*, 3(2), 2260–2278. <https://doi.org/10.20319/pijss.2017.32.22602278>
9. Hoffmann, T. (2018). “Too many Americans are trapped in fear, violence and poverty”: a psychology-informed sentiment analysis of campaign speeches from the 2016 US Presidential Election. *Linguistics Vanguard*, 4(1). <https://doi.org/10.1515/lingvan-2017-0008>
10. Lama, N. P. B., & Hamal, N. J. B. (2022). Undergraduate Students' Awareness, Usage and Attitude towards Digital Library in Kathmandu. *Interdisciplinary Journal of Management and Social Sciences*, 3(2), 68–77. <https://doi.org/10.3126/ijmss.v3i2.50263>
11. Lamba, M., & Madhusudhan, M. (2021). Sentiment Analysis. In *Text Mining for Information Professionals* (pp. 191–211). https://doi.org/10.1007/978-3-030-85085-2_7
12. Liu, D., & Luo, J. (2021). College learning from classrooms to the internet: Adoption of the YouTube as supplementary tool in COVID-19 pandemic environment. *Education and Urban Society*, 54(7), 848–870. <https://doi.org/10.1177/00131245211062516>
13. Lotka, A. J. (1926). The frequency distribution of scientific productivity. *Journal of the Washington academy of sciences*, 16(12), 317–323.
14. Namugera, L. (2014). Users' awareness, perceptions and usage of Makerere library services in the main and selected branch libraries. *Qualitative And Quantitative Methods In Libraries*, 3(3), 741–758. Retrieved from <https://www.qqml-journal.net/index.php/qqml/article/view/183>
15. Nanli, Z., Ping, Z., Weiguo, L., & Meng, C. (2012b). Sentiment Analysis: a literature review. In *2012 International Symposium on Management of Technology (ISMOT), Hangzhou, China, 2012*, (pp. 572–576). <https://doi.org/10.1109/is-mot.2012.6679538>
16. Parabhoi, L., & Saha, P. (2018). Sentiment analysis of YouTube comments on koha open source software videos. *International Journal of Library and Information Studies*, 8(4), 96–102.
17. Rahmatika, R., Yusuf, M., & Agung, L. (2021). The Effectiveness of Youtube as an Online Learning Media. *Journal of Education Technology*, 5(1), 152–158. <https://doi.org/10.23887/jet.v5i1.33628>
18. Ranganathan, S. R. (1931). *The five laws of library science*. Madras Library Association (Madras, India) and Edward Goldston (London, UK).
19. Rechowicz, K. J., & Elzie, C. A. (2024). The use of artificial intelligence to detect students' sentiments and emotions in gross anatomy reflections. *Anatomical Sciences Education*, 17(5), 954–966. <https://doi.org/10.1002/ase.2273>
20. Rout, L., & Acharya, S. (2024). Is YouTube an informational source for Indian educational resources? Analysis of YouTube videos with particular reference to e-GyanKosh and e-Pgpathshala. *Open Learning: The Journal of Open, Distance and e-Learning*, 1–21. <https://doi.org/10.1080/02680513.2024.2427815>

21. Rout, L., Khilar, P. P., & Rout, B. (2023). Does YouTube promote research ethics and conduct? A content analysis of Youtube Videos and analysis of sentiments through viewers comments. *Accountability in Research*, 31(8), 1024–1043. <https://doi.org/10.1080/08989621.2023.2192404>
22. Sharma, R., Gulati, S., Kaur, A., & Chakravarty, R. (2021). Users' Sentiment Analysis toward National Digital Library of India: a Quantitative Approach for Understanding User perception. *Library Philosophy and Practice* (e-journal). <https://digitalcommons.unl.edu/libphilprac/6372>
23. Shoufan, A., & Mohamed, F. (2022). YouTube and Education: A scoping review. *IEEE Access*, 10, 125576–125599. <https://doi.org/10.1109/access.2022.3225419>
24. Syafiq, A. N., Rahmawati, A., Anwari, A., & Oktaviana, T. (2021). Increasing Speaking Skill through YouTube Video as English Learning Material during Online Learning in Pandemic Covid-19. *ELSYA Journal of English Language Studies*, 3(1), 50–55. <https://doi.org/10.31849/elsya.v3i1.6206>
25. Tahat, K. M., Al-Sarayrah, W., Salloum, S. A., Habes, M., & Ali, S. (2021). The influence of YouTube videos on the learning experience of disabled people during the COVID-19 outbreak. In *Studies in systems, decision and control* (pp. 239–252). https://doi.org/10.1007/978-3-030-77302-1_13
26. Tan, K. L., Lee, C. P., & Lim, K. M. (2023). A survey of sentiment analysis: Approaches, datasets, and future research. *Applied Sciences*, 13(7), 4550. <https://doi.org/10.3390/app13074550>
27. Thelwall, M., Buckley, K., Paltoglou, G., & Cai, D. (2010). Sentiment Strength Detection in Short Informal Text Mike. *Journal of the American Society for Information Science and Technology*. <https://doi.org/10.1002/asi>
28. Trabelsi, O., Souissi, M. A., Scharenberg, S., Mrayeh, M., & Gharbi, A. (2022). YouTube as a complementary learning tool in times of COVID-19: Self-reports from sports science students. *Trends in Neuroscience and Education*, 29, 100186. <https://doi.org/10.1016/j.tine.2022.100186>
29. Vosoughi, S., Roy, D., & Aral, S. (2018). News Online. *Science*, 1151(March), 1146–1151.
30. Yue, L., Chen, W., Li, X., Zuo, W., & Yin, M. (2018). A survey of sentiment analysis in social media. *Knowledge and Information Systems*, 60(2), 617–663. <https://doi.org/10.1007/s10115-018-1236-4>

Digital Workflows and Management Practices of Electronic Resources in Engineering College Libraries: A Pilot Study

Preethi¹ [0000-0001-9461-8052] and Mahabaleshwara Rao² [0000-0001-9070-091 3]

¹Department of Library and Information Science, Manipal Academy of Higher Education,
Manipal -576 104, Karnataka, India
preethi.dlismpl2024@learner.manipal.edu

²Department of Library & Information Science and KMC Health Sciences Library, Manipal Academy of Higher Education,
Manipal – 576 104, Karnataka, India
m.rao@manipal.edu

Abstract. In modern academic libraries, the core function of e-resource management involves several lifecycle stages, including planning, implementation, use, maintenance, and, ultimately, renewal or cancellation. To use e-resources efficiently, libraries must understand their lifecycle and the stages involved. This pilot study examines the digital workflow and management practices of e-resources of engineering college libraries located in coastal Karnataka. Data were collected using a structured questionnaire from experienced librarians in five engineering college libraries in coastal Karnataka. Additionally, library visits and past studies on the e-resource lifecycle, best practices and case studies provided contextual depth. The findings reveal that 80% of respondents considered user recommendations, faculty suggestions, and resource usage to be the most influential criteria for evaluating and subscribing to e-resources. For access to subscribed resources, all libraries (100%) utilise both off-campus and on-campus access methods, and 80% of the users use library webpages, OPACs, or mobile apps to access the resources. The study highlights that, for accessing e-resources, libraries do not utilise modern discovery tools, SSO (Single Sign-On)- based access systems, or social media platforms; instead, they still rely on traditional processes, indicating a need for technological upgrades. 80% of the libraries review resource usage statistics quarterly through COUNTER reports, which also serve as the primary basis for renewal and cancellation decisions. These results provide insights into current e-resource management practices at similar institutions and lay the foundation for broader regional research, highlighting strengths, areas for improvement, and strategies to achieve strategic enhancements in the engineering college libraries of the region.

Keywords: E-Resources Lifecycle Management, Electronic Resources Management System (ERMS), Dakshina Kannada, Udupi

1 Introduction

In the present digital age, academic libraries have undergone significant development, and traditional print materials are now supplemented or replaced by electronic resources including e-books, e-journals, online databases, e-theses and dissertations, e-newspapers, digital images, multimedia resources and more. E- resources are essential as they are more up-to-date and provide users with the latest information. They can be accessed anywhere, at any time, without any geographical limitations. Multiple users can access resources simultaneously. Therefore, users are increasingly preferring digital resources rather than using print resources because of their convenience and ease of access.

E-resources significantly benefit all stakeholders by reducing publication and distribution costs, enabling rapid access, being eco-friendly, and saving physical space. Moreover, they pose significant barriers, including high subscription costs, the need for specific technical platforms, compatibility issues, authentication requirements, and the need for users to have sufficient IT skills.

The e-resources lifecycle is a widely accepted framework for acquiring and managing e-resources. Pesch introduced a model of the Electronic Resource Management (ERM) lifecycle through a study that outlines the various stages of an e-journal's life [1]. He guides the viewer through each step of the e-journal lifecycle. Pesch's model was expanded by Emery and Stone, who ultimately defined six components, or Techniques for Electronic Resource Management Systems (TERMS) that comprise the cyclical e-resource management process.

TERMS 1: Exploring new content for purchase/addition

TERMS 2: Acquisition and negotiation

TERMS 3: Implementation and access

TERMS 4: Ongoing evaluation and promotion

TERMS 5: Annual review

TERMS 6: Cancellation and replacement review

The digital workflow or lifecycle of e-resources in academic libraries is a continuous process encompassing administrative and support management, acquisition management, access management, evaluation and preservation, and review for replacement and cancellation. This study examines the digital workflow and life-cycle management of e-resources in engineering college libraries, highlighting current mechanisms and challenges in the digital environment.

2 Literature Review

The literature review aims to understand research and opinions on the management of digital resources in academic libraries. It helps researchers develop a concept and expand their knowledge of electronic resource management and its workflow. Several investigations have been conducted to identify the ERM lifecycle and its process in academic library environments.

ERM has been conceptualised as a workflow distributed among library staff, encompassing six key stages: experimentation, negotiation, licensing, storage, usage monitoring, and renewal/de-selection [2]. A lifecycle-based approach further identifies six key terms, including: investigation of materials; procurement and licensing of content; implementation; troubleshooting; evaluation and preservation; and review of cancellations and replacements [3]. Recent studies have also identified various planning, staffing, and preservation practices for e-resource management. This highlighted the importance of adequate planning for efficient management, appropriate staffing provisions (including specialised librarians), and the implementation of diverse preservation techniques, such as migration, born-digital preservation, and recapturing [4].

Several studies identified challenges across the different stages of the electronic resource management lifecycle. At the acquisition stage of the ERMS life cycle and identify the challenges that library professionals face, including budget issues, insufficient staff training, lack of standard procedures, and vendor negotiations, which can be addressed with dedicated ERMS software [5]. Moving towards the negotiation stage, it has been emphasised that libraries and vendors or service providers should have a mutual understanding, and where librarians should be assertive but not aggressive to ensure favourable licensing terms [6]. Administrative management also plays a crucial role in ensuring seamless access, problem tracking, budget management and vendor communication [7].

Studies conducted in university libraries in Maharashtra and observed challenges in managing their e-resources effectively. Significant challenges include the lack of awareness about the life cycle of e-resources, the absence of standard procedures, collection development and preservation policies, and negotiation skills [8]. Additionally, using MS Excel or in-house tools for e-resource management leads to inconsistencies in workflow and management [8]. Similarly, it has been reported that, to manage e-resources properly, libraries do not follow a fixed protocol, and the majority are unaware of the six steps of the TERMS model [9].

The identification and trial stages of e-resource management have also been discussed in the literature. Identifying suitable e-resources commonly includes user recommendations, input from faculty and research scholars, trial offers, vendor demonstrations, discussion lists, and publisher catalogues, among others [10]. In trial access, the library may obtain access to e-resources for a specific period (such as a week or a couple of months) to determine their suitability for purchase [10].

International studies also highlight challenges in terms of user support and infrastructure. A case study of the University of Zimbabwe provides valuable insights into the use of e-resources in developing countries. Although many educational institutions in Zimbabwe have made e-library provisions, student adoption of these resources remains minimal. Several factors are recognised across the e-resource life cycle, including poor marketing strategies, students' lack of knowledge of resources, and high data charges from Internet Service Providers (ISPs) [11]. Compared to the study in India, where ERMS implementation primarily focuses on vendor systems and cost constraints, the South African survey emphasises that process documentation and staff training are equally necessary for effective e-resource management [12].

To address these challenges, several scholars recommend increasing the technical skills of library staff and users through continuous training programs, user-centric support services, collaborative partnerships with faculty and researchers to understand their needs, the implementation of accessibility initiatives, and the analysis of usage data [13]. Low usage of e-resources has also been attributed to inadequate promotion, irrelevant content, and limited user awareness, further underscoring the need for strategic user engagement and outreach efforts [14].

The literature review reveals a significant gap in comprehensive lifecycle management studies specific to engineering college libraries, particularly in coastal Karnataka. This study aims to fill this gap by developing a

structured framework to understand the current mechanisms in engineering college libraries and to provide targeted directions for future improvements in resource management.

3 Objectives

This study aims to provide a complete analysis of the digital workflow and lifecycle management of e-resources in the selected engineering college libraries of coastal Karnataka.

4 Methodology

A pilot study was conducted with experienced librarians from five engineering colleges of Dakshina Kannada and Udupi Districts of Coastal Karnataka, i.e., the NMAM Institute of Technology, Nitte, Shri Madwa Varidha Institute of Technology (SMVIT), Bantakal, St Joseph Engineering College (SJEC), Mangalore, Canara Engineering College (CEC), Mangalore, and Alvas Institute of Technical Education (AITE), Moodbidri.

A structured questionnaire was designed to align with the study's objectives, including multiple-choice questions to collect quantitative insights and open-ended questions to elicit professional viewpoints on digital workflow and management practices for e-resources in their library. The draft questionnaire was validated by subject experts (teachers & librarians) in library and information science. Based on their feedback, a revised questionnaire has been distributed to librarians to identify how engineering college libraries manage e-resources. Additionally, library visits, past studies on the e-resource lifecycle and its workflow, best practices, and case studies were reviewed to provide context for the findings. The gathered data/responses were analysed via MS Office tools, including tables and graphs.

5 Data Analysis & Findings

5.1 Criteria for evaluating e-resources before subscription

Table 1 presents the criteria that libraries use to evaluate e-resources before subscription. The results indicate that in the decision-making process, user recommendations, faculty suggestions, and the use of e-resources are considered the most influential factors, each reported by 80% of the respondents. In comparison, previous experience (60%) and experience with publisher/vendor demos and trials (40% each) have a relatively moderate impact.

Table 1. Evaluation of e-resources before subscription

Evaluation of e-resources before subscription	Response & Percentage
User recommendations	4 (80%)
Faculty suggestions	4 (80%)
Usage of the e-resources	4 (80%)
Previous experience	3 (60%)
Publisher/vendor demo	2 (40%)
Experience during trial access	2 (40%)

5.2 Pre-agreement considerations for finalising e-resource subscriptions

When subscribing to e-resources before the agreement, all the libraries carefully reviewed the license agreements, which were considered the most important, as shown in Table 2. Additionally, the contract terms and conditions, including clarification regarding the annual renewal tenure and purchase orders, indicated that purchase orders were part of their pre-agreement considerations.

Table 2. Pre-agreement considerations for e-resource subscriptions

Pre-agreement Considerations	Response & Percentage
Licence Agreement of the e-resources	5 (100%)

Contract terms and conditions	4 (80%)
Clarification of annual renewal tenure	4 (80%)
Purchase order	2 (40%)

5.3 Key factors considered for the negotiation of full-text e-resources

Table 3 indicates that all the libraries focused on subject coverage and full-text access to e-resources (e-books, e-journals, e-magazines, full-text databases, etc.) to negotiate resources. While 80% of the respondents considered access to the back file, 60% sought discounts. Only 20% consider minimum pricing for negotiating e-resources.

Table 3. Key factors considered for the negotiation of full-text e-resources

Negotiation Factors	Response & Percentage
Maximum subject coverage and full-text e-resources (e-books, e-journals, e-magazines, full-text databases, etc.) in the subscription package	5 (100%)
Long-term perpetual access to full-text content (back file)	4 (80%)
Maximum discount offer	3 (60%)
The minimum price formula (Lowest Price Guarantee)	1 (20%)

5.4 Methods of providing access to e-resources

After the acquisition and negotiation phase, the next step is to provide access to the subscribed e-resources to its users. Libraries use various methods to access their resources. The data reveal that 80% of the surveyed libraries offer access to subscribed e-resources through their library websites, online public access catalogues (OPAC), and mobile applications. However, no libraries have been reported using discovery tools for access provision (Table 4). This finding indicates that nearly all libraries employ traditional methods to access their e-resources.

Table 4. Methods of providing access to e-Resources

Access Provision	Response & Percentage
Through the library website	4 (80%)
Integrated with OPAC	4 (80%)
Through the Mobile app	4 (80%)
Use of discovery tools	0

5.5 Existing Access Modes and Authentication Systems

Table 5 indicates that all surveyed libraries (100%) reported providing e-resource access to their users via both on-campus and remote access. None of the libraries use Single Sign-On (SSO) or access via credentials. They still rely on basic IP-based authentication and have not adopted modern login systems, such as SSO or OpenAthens.

Table 5. Modes of access to e-resources

Access Mode	Response & Percentage
Yes, both on-campus and remotely	5 (100%)
Yes, only on-campus (IP)	0
Log in with SSO (Single Sign On)	0
Credential-Based Login	0

5.6 Methods used by libraries to create awareness about subscribed e-resources

To increase awareness of subscribed e-resources, all libraries reported using notice boards and their library or institutional websites to inform users about these resources. A significant majority (80%) organise workshops or webinars. Moreover, 60% of librarians utilize library blogs and email communication, whereas 40% use social media platforms, such as WhatsApp, to reach users (Table 6).

Table 6. Methods used by librarians to increase their awareness of e-resources

Awareness Creation	Response & Percentage
Displaying on the notice board	5 (100%)
Published on the library website/institutional website	5 (100%)
Organising workshops/webinars	4 (80%)
Posting updates on the library blog	3 (60%)
Sending an Email notification to the users	3 (60%)
Sharing updates via social media like WhatsApp	2 (40%)

5.7 Methods used for usage tracking and statistics

All the surveyed libraries reported collecting usage data of their e-resources. Table 7 shows that most libraries (80%) used the most common method for usage statistics, COUNTER reports. Additionally, 60% of the respondents used portal websites to track usage. A smaller proportion of libraries collected usage data through user feedback surveys and in-house data usage systems (20%).

Table 7. Methods used for usage tracking and statistics

Usage Tracking	Response & Percentage
Usage statistics (COUNTER reports)	4 (80%)
E-resource portal or websites	3 (60%)
User feedback surveys	1 (20%)
Own data usage system	1 (20%)

5.8 Criteria considered for renewal of subscribed e-resources

Table 8 indicates that all librarians (100%) reported that the use of e-resources was the primary criterion for renewing their subscriptions. A large proportion (80%) also considered recommendations from department heads, faculty members, and library users. Sixty percent cited the availability of the fund, whereas 40% of the libraries considered factors such as satisfaction with access provision by publishers, pricing/access to back files, packaging, and content availability.

Table 8. Criteria considered for the renewal of e-resources

Criteria for Renewal of E-Resources	Response & Percentage
Usage of the e-resources	5 (100%)
Recommendations by the head of the department or faculty	4 (80%)
Recommendations by the library user	4 (80%)
Availability of the fund	3 (60%)
Satisfaction with access provision by the publishers	2 (40%)
Pricing structure for back files and content available	2 (40%)
Statutory body requirement	1 (20%)

5.9 Criteria for cancellation or replacement of e-resources

According to Table 9, 80% of the libraries consider cancelling or replacing e-resources because their usage is lower than the prices, whereas 60% cite budget problems and a lack of vendor support. 40% of the libraries considered the lack of support from the IT team to be a reason for the cancellation or replacement of resources.

Table 9. Criteria for cancellation or replacement of e-resources

Criteria for Cancellation or Replacement of E-Resources	Response & Percentage
Low Return on Investment (ROI) in usage	4 (80%)
Lack of support from the vendor	3 (60%)
Budget problem	3 (40%)
Lack of support from the IT team	2 (40%)

6 Discussion

The pilot study shows how engineering college libraries of Dakshina Kannada and Udupi districts manage the lifecycle of e-resources, from selection to renewal or cancellation, which aligns closely with the TERMS model stages.

The findings strongly emphasise user-centric, evidence-based practices in the management of e-resources across academic libraries. According to the study, when evaluating e-resources before subscription, the primary criteria considered were recommendations from the faculty and users, as well as e-resource usage. This finding aligns with TERMS 1.

The study highlights the strong emphasis on licensing agreements and contractual terms in negotiations, connecting TERMS 2 acquisition and negotiation, which is consistent with Harris's previous research (2000). It states that negotiation practice focuses primarily on maximum subject coverage, and its full-text access provision, also includes backfile access.

The study's findings relate to implementation and access, which connect to TERMS 3, indicating that libraries primarily provide access to subscribed resources on both on-campus and off-campus platforms. Notably, the complete absence of discovery tools and the Single Sign-On (e.g., OpenAthens) system highlights a significant gap in modern digital infrastructure. This observation supports Wadekar and Nagarkar's (2018) findings that many libraries still rely on basic tools such as MS Excel or internal systems rather than adopting advanced ERM solutions.

To increase users' awareness or promote subscribed resources, libraries primarily use traditional methods such as noticeboards, websites, and workshops/webinars. In line with TERMS 4 & Bentil's (2022) study, social media platform use remains low, indicating that libraries have yet to capitalise on the popular digital platforms students use entirely.

For annual review and decision-making (TERMS 5 & 6), usage tracking through COUNTER reports is a commonly used method and is considered the primary criterion for renewal. Further study also found that low usage, budget constraints, and a lack of vendor or IT support primarily guided cancellation decisions.

Overall, the study highlights the strengths and gaps in current digital workflows for managing e-resources. The objective is to establish a well-defined structure that enhances the understanding of current practices and supports future improvements in resource management. Although libraries are firmly committed to managing e-resources effectively, there is a clear need to adopt a standardised electronic resource management framework, enhance negotiation skills, and utilise modern tools (discovery and SSO-based access systems) for better resource access. Addressing these areas can significantly improve the efficiency and impact of electronic resource management in academic libraries.

This study encompasses comprehensive life-cycle management of e-resources in the region, unlike previous studies that focus solely on the acquisition stage or the use of e-resources.

7 Limitations of the Study

The sample size of this study was limited to five institutions in Coastal Karnataka, which may not fully represent the diversity of digital workflows and electronic resource management practices across the region. The data relied on self-reported responses from librarians, which may have introduced content bias and may not fully reflect the actual practices. Additionally, the study's sample is limited to librarians, excluding samples from end-users of the library, such as faculty and students.

8 Conclusion

The pilot study identifies positive aspects and areas for improvement in the management of the e-resources used by engineering college libraries across coastal Karnataka. Libraries subscribe to e-resources based on user and faculty recommendations, as well as resource usage. Usage reports determine the final evaluation, renewal, or cancellation of resources. However, the lack of modern discovery services and tools underscores the need for technological improvements and emphasises the importance of the ERMS framework. Improving negotiation strategies with vendors and leveraging digital technologies can enhance access to resources.

Future research should increase the sample size to include more regional institutions and gather user-based surveys, thereby laying the groundwork for broader regional investigations of all engineering colleges in coastal Karnataka. This should include a user study on the effectiveness and usability of digital workflows in electronic resource management and provide clear guidelines for implementing ERMSs within the proposed standardised

framework. Longitudinal analyses of workflow evolution and cross-regional comparisons may also provide deeper insights into emerging trends and practices in electronic resource management.

Declaration of AI and AI-assisted technologies used in the manuscript writing process: The authors acknowledge the use of AI tools (Grammarly Premium, AJE Rubriq, and ChatGPT by OpenAI) to assist in editing and improving the writing style of this article. The AI tool was used solely for language enhancement and not for creating the content or references of the manuscript. The authors carefully reviewed, edited, and revised all the AI-assisted texts and have taken full responsibility for the content and accuracy of this article.

References

1. Pesch, O.: Library standards and e-resource management: A survey of current initiatives and standards efforts. *Serials Librarian*, 55(3), 481–486 (2008). <https://doi.org/10.1080/03615260802059965>.
2. Breeding, M.: Helping you buy: Electronic resource management systems. *Computers in Libraries*, 28(7), 6–96 (2008).
3. Emery, J., & Stone, G.: Electronic resource management in a post-plan world. *Insights: The UKSG Journal*, 34. (2021). <https://doi.org/10.1629/uksg.554>
4. Obidike, N. A., & Mole, A. J. C.: Managing Electronic Collection Resources (ECRs): Nigerian University Libraries Perspective. 19(2), 13-30. (2024).
5. Deshpande, A., & Waray, P.: Acquisition of Electronic Resources in Academic Libraries: Challenges and Opportunities. *International Journal of Creative Research Thoughts*, 12(2), 258–263 (2024).
6. Harris, L.: Licensing, Copyright Law, Technology, and Education. *Copyright & New Media Law Newsletter*, 4(2), (2000). Retrieved from <https://www.copyrightlaws.com/copyright-licensing-books/>.
7. Farmer, L. S.: The life cycle of digital reference sources. *The Reference Librarian*, 50(2), 117-136. (2009). <https://doi.org/10.1080/02763870902755957>
8. Wadekar, P. P., & Nagarkar, S. P.: Current Practices in the Management of Online Databases at University Libraries in Maharashtra State, India. *Library Management*, 39(8/9), 569–582 (2018). <https://doi.org/10.1108/LM-08-2017-0073>.
9. Dhiwar, G. D.: Management of electronic resources by the management institute in Pune City: a survey. *Library Management*, 42(1–2), 93–108 (2021). <https://doi.org/10.1108/LM-01-2020-0009>.
10. Joshipura, S.: Selecting, acquiring, and renewing electronic resources. *Electronic Resource Management in Libraries: Research and Practice*, IGI Global, 48-70 (2008).
11. Mawere, T., & Sai, K.: An investigation on e-resource utilisation among university students in a developing country: A case of Great Zimbabwe University. *South African Journal of Information Management*, 20(1), (2018). <https://doi.org/10.4102/sajim.v20i1.860>.
12. Abrahams, R., & Ngulube, P.: Using lean concepts to enhance electronic resource workflows and operations at an academic library in South Africa. *Journal of Electronic Resources Librarianship*, 36(3), 177–189 (2024). <https://doi.org/10.1080/1941126X.2024.2383519>.
13. Mommoh, R. L., & Emmanuel, V. O. Electronic information resources management in academic libraries. *Journal of Library Services and Technologies*, 6(2), 136–154 (2024). <https://doi.org/10.47524/jlst.v6i2.137>.
14. Bentil, W., Liew, C. L., & Chawner, B.: The management and the usage of electronic resources in academic libraries: A bi-directional relationship. *Information Development*, 38(1), 114-124 (2022). <https://doi.org/10.1177/0266666920983600>.

Data Management and Research Data Services

Illustration of Cases of Research Data Management Services in Libraries in Asia: An Empirical Study

Prameeta Sharma¹ [0000-0003-3712-6566] and Chandrashekara M.² [0000-0002-4431-0102]

¹PG and Research Department of Library & Information Science
Nirmala College for Women (Autonomous)
Coimbatore - 641018, Tamil Nadu.

²DOS in Library and Information Science,
University of Mysore, Mysuru,
Karnataka-570006

{prameeta2, chandra.oum}@gmail.com

Abstract. The present study aims to illustrate the status of research data management (RDM) services in libraries in selected Asian countries. Institutions were selected randomly based on research intensity and the availability of publicly accessible information. The study empirically observes the penetration and expansion of research data management in selected institutions across these countries, with a specific focus on the types of services, viz., training and guidance on Data Management Plan (DMP) support, institutional data repositories, metadata standards, data preservation, and policy on the ethical reuse of data. The study applies a two-pronged method and descriptive design, combining a systematic review of literature and content analysis of academic library websites. The study also examines the alignment of RDM services in selected Asian countries with FAIR data principles and Open Science frameworks. Findings positively supported the research questions and concluded with a diverse & moderate level of maturity in RDM services in libraries.

Keywords: Research Data Management, Open Science, Research Data Services, FAIR Data, DMP.

1 Introduction

Libraries have a significant vulnerability when it comes to verifying the reality of evolution, given the shift in the education system. As an informal centre to support teaching and learning, it gradually transforms with the delivery of need-based services, designed specifically to serve the user community in real-time. The core area of functionality for library services is deeply rooted in innovation and the facilitation of patrons' desired needs. In true essence, terminologically, it is a perennial channel for the growth, development, and communication of scientific research.

Research is inherently hierarchical and an independent field of study that demands refinement in the system for its growth and development. Each research project, with its unique data and findings, makes a contribution to the expansion and maturation of the specific discipline within the universe of knowledge. The infinite, multidimensional, and ever-expanding growth of the universe of knowledge and its propagation in human society is benchmarked by the quality of research and innovation. Data, the smallest component in the life cycle of scientific research, is a comprehensive and interconnected bridge for all knowledge that has ever been created in the past or will be created in the future. Hence, managing the data in a systematic and standard order is the crux of good research and is a much-needed activity to be practiced during the process of research.

RDM covers a series of activities and processes associated with the data's lifecycle which entails collection and/or creation of the data, as well as their storage, security, preservation, recovery, exchange and reuse, all taking into account technical issues, ethical and legal considerations, and the governance frameworks developed for these data [1]. Since data creation is the primary task of all research, data management is a mandatory process to be practiced in compliance with the demands of the funders, sponsors or according to the policy of the organization. [2] libraries have provided leadership in RDM, particularly in advocacy and policy development, but service development is still limited. Data management in research and the role of libraries vis-à-vis librarians is a way more than just providing any anticipatory or reference service. At the outset, libraries have a granular role to play to bridge the gap and support the data management process all throughout the research data life cycle. In fact, it is a systematic approach to provide managerial skills and techniques to the budding researcher during their process of research and also post-research.

2 Need and Significance of the Study

The study is significant from the perspective of RDM service maturity and aims to track ongoing progress, specifically focusing on the types of services delivered in libraries. Empirically, it will escalate the growth of research data management in libraries for benchmarking at the local level. The purpose of this study is to explore the current status of RDM services and to draw conclusions based on concrete, verifiable evidence from the selected libraries in Asia. Therefore, the need and significance of the study are deeply rooted in substantiating the spread of RDM and its similar implementations.

3 Objective of the Study

The primary objective of the study is to trace the ladder of growth and propagation of research data management services in libraries in Asia. The basic objectives set forth to be achieved are:

- To analyse the research data management conceptualisation in scholarly literature specifically from Asia.
- To identify research data management (RDM) services offered in selected Asian countries library.
- To explore the typology of research data management services rendered in library in selected Asian countries.
- To compare patterns, trends and gaps in RDM Service across countries and Institutions in Asia.

4 Scope and Limitation of the Study

This study investigates Research Data Management (RDM) Services in academic and research libraries across seven selected Asian Countries, with a focus on their alignment with FAIR data principles and Open Science policies. The geographical scope includes India, China, Japan, Singapore, Malaysia, Taiwan, and South Korea, representing diverse levels of RDM maturity and policy development. The study examines key library-based services such as RDM training and guidance, Data Management Plan (DMP) support, institutional data repositories, metadata, data preservation, and policy advocacy.

The study is limited by its selective country and institutional coverage, which may not fully represent the heterogeneity of the Asian region. Data collection relies primarily on publicly available documents and institutional websites. Additionally, the study does not directly assess researcher survey or service effectiveness, and findings should be interpreted as a descriptive mapping rather than an evaluative assessment. The data were collected through the rich literature review and content analysis of the websites of the selected universities and institutions only.

5 Methodology

The study undertook a qualitative descriptive design combining a systematic literature review and content analysis of academic library websites. Libraries and national centres were selected purposively based on visibility of RDM services on official websites, policy relevance (Open Science, FAIR, funder mandates), research intensity, the availability of publicly accessible RDM documentation and regional representativeness within Asia. No university rankings are used as indicators for the selection of the same. Sample size is indicative of RDM initiatives only. The study is not exhaustive but limited by its coverage only to illustrate the services for benchmarking at the regional and institutional level, in collaboration with RDM services. The selected literature published on RDM work in the Asian countries was a key input for the sampling and defining the scope of the study.

The literature review is indicative, as only the publications directly relevant to library-based RDM were included. The relevant literature was searched on the Google Scholar search engine by using the keyword search such as research data management, research data services RDM in Asia, RDM policies and FAIR Data. Content analysis of the websites was conducted on the official websites of the selected academic libraries using predefined RDM-related keywords such as research support section, institutional data repositories, Lib Guides and RDM policy documents. The data was collected in the month of November 2025. In this paper, the methodology for coding was according to [3] “each of the sections for typology of services encompasses a series of subcategories, 16 in total that were reviewed in the analysis of the websites and coded as an existing service based on their presence in the offer of services of the academic library examined”. The following protocol was used to code the information:

- The coding of the fields was absolute based on the Yes/No designation, without judging the quantity or quality of the services offered.
- Only the information appeared on the library's website (the information itself or information that provides a link to another resource that expands on it) and not in the sections of other units of the university analysed, such as the research findings transfer office or research support units unrelated to the library.
- The information appeared on the library's institutional website, was identified and not any informal channels, such as blogs, or social platforms, such as Twitter, Instagram was not be taken into account.

6 Literature Review

Research data management (RDM), major types of RDM services provided by academic libraries in South Asian and Southeast Asian countries were data repository, data management training, maintaining Web resources, data study and analysis, and promoting awareness of reusable data sources [4]. Research activity produces valuable data and Research Data Management Services are vital services provided in academic library and are focused on the management, archiving, processing and reuse of critical research data [5]. The emergence of data driven research and demands for the establishment of Research Data Management (RDM) has created interest in academic institutions and research organisations globally [6]. According to [7] Research Data Service at a local level in Chinese Universities are in their infancy. There is more evidence of activity in developing data repositories than support services and little development of local policy. University libraries in Myanmar are developing research data management services (RDMS) to support and boost the country's research and development and to support the development of RDMS, libraries should enact three different roles, namely, participative role, educational and consultative role, and facilitative role [8]. [9] discussed in their study about the research data management and related services offered by South Asian countries' academic libraries and The study's findings revealed that 83.6% of institute libraries provide research data management services to their users and opined that institutes or funding organizations support staff to attend conferences and workshops on research data management, library professionals have to join MOOC to take courses related to research data services, Institute or professionals conduct in-house staff workshops and presentations. [10] their studies on data produced in agricultural sciences and provides information about types of data produced, on understanding the benefits from open data, provided perception and attitude of scholars towards sharing and reusing research data. The perceptions of agricultural researchers to data sharing were generally positive, but respondents reported a number of fears including lack of time, data security and losing control over intellectual property, possible misinterpretation and misuse of data and they mentioned the interest in library assistance for the need of training in research data management. [11] discussed on the issues of aggregated data for API communities followed by suggestions on how health sciences librarians can support and promote better practices for data disaggregation.

Research Data Management (RDM) in the Peking University Library (PKUL) has the Peking University Open Research Data Repository (PKU-ORDR) project and it has helped numerous researchers to support the entire research life cycle and enhanced Open Science (OS) practices on campus, as well as impacted the national OS movement in China through various national events and activities hosted by the PKUL [12]. Managing research data is important as it further showcases research impact and research output of institutions. Robust and effective RDM practices will also help to ensure integrity and transparency of research findings and academic and research libraries will continue to play an integral role in RDM [13]. According to [14] Malaysian academic libraries provide RDM services based on their related or basic skills which are bibliographic management tools, institutional repository and openness of research data rather than impacted services to support RDM, such as data analysis, data citation, data mining or data visualisation services. Although the librarians were aware of RDM and their roles in research data services, the progress of practicing and implementation of the RDM services still has not been fully delivered to support the main RDM elements. This study concludes LIS faculties having adequate skills and knowledge on data curation and preservation in Asian region but the most preferable data analysis tool is Microsoft Excel and the result shows that generated data is mostly stored on personal computers and laptop hard disks [15]. RDM has emerged as a new addition to library research support services. The more recent literature clearly established the pivotal role of libraries and librarians in developing and managing RDM services [16]. According to [17] four core themes in data management functions emerged "Cataloging and Metadata Management," "Data Services and Support," "Research Data Management," and "Systems Management and Maintenance." Over time, these themes have evolved from distinct roles to a more balanced distribution and the steady growth of "Data Services and Support" and "Research Data Management" reflects the adaptability and strategic adjustments of academic libraries in response to the rapidly changing information landscape. Libraries take a dominant and central role in research data management and further their capacity as curators accordingly, University libraries are expected to assemble capabilities, to manage and provide research data for sharing and reusing

efficiently [18]. [10] explored librarians' familiarity with RDM concepts, the extent of RDM services provided, and the challenges faced in implementing these services by conducting a survey among 93 librarians from 10 leading universities in Moldova. The data was collected on the awareness, involvement, and the institutional support available for RDM. The findings revealed a varying degree of familiarity and engagement with RDM services among librarians, highlighting the need for enhanced training and institutional policies to support effective RDM practices. [19] conducted a systematic review of the landscape of research data management in academic libraries and explored that Open Science entails research reproducibility with an emphasis on data sharing and reuse by concluding that research data management (RDM) is an essential asset in research institutions for supporting open science. According to [20] explored the overall architecture and design scheme of the university library data center [21] and puts forward a series of optimization strategies for the data centre to enhance the "vitality" of the data, let the data "speak", and provide support for the library's intelligent decision-making.

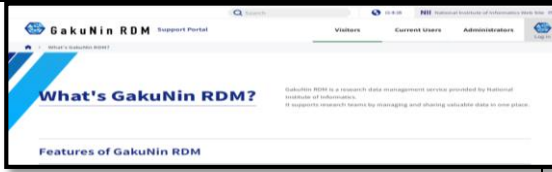
7 Data Illustration and Findings

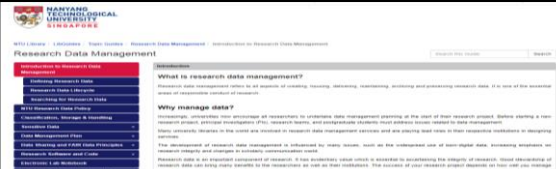
Content analysis of the selected library websites of the Asian Countries represents the different nomenclature for the RDM services offered. The services are varied in nature and cover the RDM research data life cycle support paradigm. The details are described below.

Table 1. RDM Services in selected Asian Countries Library

S.No.	Institution / Library	Program or Service Name	Country / Region
1	National Institute of Informatics	GakuNin RDM	Japan
2	Kyoto University Library	RDM & DMP Support	Japan
3	Hokkaido University Library	Institutional Repository	Japan
4	Nanyang Technological University Library	RDM LibGuide	Singapore
5	National University of Singapore Libraries	Research Data Management	Singapore
6	NTU Library	DR-NTU (Data)	Singapore
7	University of Hong Kong Libraries	Research Data Services	Hong Kong
8	HKU Libraries	DMPTool@HKUL	Hong Kong
9	Chinese University of Hong Kong Library	CUHK Research Data Repository	Hong Kong
10	INFLIBNET Centre	Research Data Management Initiatives	India
11	Jawaharlal Nehru University Library	Institutional Repository	India
12	University of Delhi Library System	Research Data Management and Web Guides	India
13	Universiti Malaya Library	Research Data Management Support	Malaysia
14	Universiti Malaya	UM Research Data Repository & UM Research Data Policy	Malaysia
16	National Taiwan Normal University Library	Research Data Management Guide	Taiwan
17	National Taiwan University Library	Research Data Repository	Taiwan
18	Korea Institute of Science and Technology Information (KISTI)	National Research Data Services	South Korea

Table 2. Website Interface of RDM Services in Library

S.No.	RDM Interface	Organisation/Country/URL
1		National Institute of Informatics Japan https://support.rdm.nii.ac.jp/en/about/

2		Kyoto University Library, Japan https://rdm.kyoto-u.ac.jp/en/
3		Hokkaido University Library, Japan https://eprints.lib.hokudai.ac.jp/
4		Nanyang Technological University Li- brary, Singapore https://libguides.ntu.edu.sg/rdm
5		National University of Singapore Libraries, Singapore https://libguides.nus.edu.sg/rdm
6		NTU Library, Singapore https://researchdata.ntu.edu.sg/
7		University of Hong Kong Libraries, Hong Kong https://hub.hku.hk/researchdata/
8		University of Hong Kong Libraries Hong Kong https://hub.hku.hk/researchdata/#DMP
9		Chinese University of Hong Kong Library, Hong Kong https://www.lib.cuhk.edu.hk/en/research/data/
10		INFLIBNET Centre, India https://hrd.inflibnet.ac.in/DataCite-GAF/team.php
11		Jawaharlal Nehru University Library, India http://lib.jnu.ac.in/research-guide

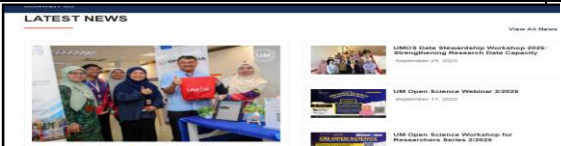
12		University of Delhi Library System, India https://duls.du.ac.in/rss/rdma_wg
13		Universiti Malaya Library, Malaysia https://researchdata.um.edu.my/
14		National Taiwan University Library, Taiwan https://www.lib.ntnu.edu.tw/english/academic_guide/guide_03_01.jsp
15		Korea Institute of Science and Technology Information (KISTI), South Korea https://www.kisti.re.kr/eng/rnd/pageView/250

Table 3. Typology of RDM Services in Library-I

RDM Service Type	Japan	Singapore	Hong Kong
RDM Guidance & Training	1	1	1
Data Management Plan (DMP) Support	1	1	1
Institutional Repository & Data Publishing	1	1	1
Metadata & FAIR support	1	1	1
Secure Storage & Archiving	1	1	1
Legal, Ethical & Sensitive Data Advice	1	1	1
Licensing & Citation Advice	1	1	1
Policy Development & Advocacy	1	1	1

Yes-1

No-0

Table 4. Typology of RDM Services in Library-II

RDM Service Type	India	Malaysia	Taiwan	South Korea
RDM Guidance & Training	1	1	1	1
Data Management Plan (DMP) Support	0	1	1	1
Institutional Repository & Data Publishing	0	1	1	0
Metadata & FAIR support	1	1	1	0
Secure Storage & Archiving	0	1	1	1
Legal, Ethical & Sensitive Data Advice	1	1	1	1
Licensing & Citation Advice	1	1	1	1
Policy Development & Advocacy	1	1	1	1

Yes-1

No-0

Table 5. FAIR-Aligned Gaps, Maturity Levels and Policy Implications in RDM Services

Fair Dimension	Key Gap Identified	Indicative Evidence From Websites	RDM Maturity Level	Policy Implication(s)
Findable	Limited PID and metadata visibility	Sparse DOI/metadata statements	Basic - Developing	Mandate PIDs & metadata standards
Accessible	Unclear access and persistence policies	Missing access protocol descriptions	Developing	Standardise access & retention policies
Interoperable	Weak standards and machine-readability	Few references to schemas/vocabularies	Basic	Promote interoperability & skills
Reusable	Inadequate licensing and provenance	Licenses and reuse guidance often absent	Developing - Basic	Enforce reuse & licensing policies

8 Thematic Comparison of Patterns, Trends and Gaps in RDM Service Provision

The website content analysis reveals a concise, notable, and common baseline of RDM services across the selected libraries in the study. The observed predictable patterns, emerging current trends and the significant persistent gaps have been identified, observed, and described significantly in the context of the available services rendered in libraries.

Table 6. Representing the comparative description encapsulated and structured thematically based on Braun and Clarke's six-stage framework. [22]

Theme(s)	Pattern(s)	Trend(s)	Gap(s)
Service Strategy	RDM Guidance Pages	Shift From the Static guidance to Service oriented models	Limited Details On Data Preservation
Service Structure	Institutional Repositories	DMP Consultation & Training Workshops	Sensitive Data And An Uneven Institutional RDM Maturity
Service Governance	Basic Data Sharing Information.	Alignment With The Fair Principles And Open Science With National Research Policies	Variability In Legal And Ethical Guidance
Service Challenges	RDM As A Part Of Broader Research Support Services	Persistent Identifiers, Metadata Standards And Open Repositories	Lack of Explicit References to Machine- Actionable Metadata, Controlled Vocabularies and Domain Standards.
Service Stewardship Bottlenecks	RDM Ecosystems With Clearly Articulated Service Portfolios	Collaborative Support With IT And Research Division.	Interoperability Standards And Long-Term Stewardship
Service Sustainability	Scholarly Communication And Open Science Initiatives.	Enhance Service Visibility And Accessibility And The Adoption Of Libguides	Legal And Ethical Guidance Especially For Sensitive Data

9 Derived Implications of the Study

The impact of research findings is imperative to plan, design and develop the standard research data services in higher education institutions. The key implications of the study are:

- The study is very coherent to boost the similar scale of growth for innovation in the research support services in libraries.

- The proposition from the patterns observed across the libraries can be utilized to develop similar atmosphere of serving the research community.
- The emanating trends across the libraries can be applied as an immediate blueprint for plan-in-action to enhance the library services.
- The gaps observed in the study across the libraries where the insufficient and lack of comprehension is indicated, can be exhibited as a domain for crucial area of practical improvement.
- The key competencies of librarians and apprehensive timely initiatives towards the management of data in libraries are of significant value for benchmarking the equivalent services design and delivery.

10 Conclusion

The study has undertaken a two-pronged method to draw inferences about the RDM services' maturity level in the selected Asian countries' libraries. Positive inferences were observed from the broad spectrum of growth of research data management services in the selected university libraries in the Asian countries. Literature review positively validated the various services from basic to moderate levels, designed and provided to the researchers. The overall website evidence indicates a transitional stage of RDM service development in libraries, with more focus on growing policy and implementation of FAIR-compliant services. The penetration of RDM in research support services in libraries is commendable, and it empirically justifies the research with a moderate level of growth in the paradigm of new service support for data-driven research of the concerned research community. Within these boundaries, the study has been pertinent in bringing forth the current status of RDM services empirically; however, it is a partial exploration of RDM services, focusing on the maturity levels and the growth trajectory within certain constraints and limitations.

References

1. Cox, A. M., & Pinfield, S. (2014). Research data management and libraries: Current activities and future priorities. *Journal of Librarianship and Information Science*, 46(4), 299–316. <https://doi.org/10.1177/0961000613492542>
2. Cox, A., Kennan, M., Lyon, L., & Pinfield, S. (2017). Developments in research data management in academic libraries: Towards an understanding of research data service maturity. *Journal of the Association for Information Science and Technology*. 68(9), 2182–2200. 10.1002/asi.23781.
3. Martin-Melon, R., Hernández-Pérez, T., & Sara Martínez Cardama. (2023). Research data services (RDS) in Spanish academic libraries. *The Journal of Academic Librarianship*, 49(4), 102732–102732. <https://doi.org/10.1016/j.acalib.2023.102732>
4. Sinha P, Subaveerapandiyan, A. & Sinha MK (2025), "Research data management services in academic libraries: a comparative study of South Asia and Southeast Asia". *Global Knowledge, Memory and Communication*, Vol. 74 No. 3-4 pp. 777–793, doi: <https://doi.org/10.1108/GKMC-01-2023-0033>
5. Singh, R., Bharti, S., & Madalli, D. (2022). Evaluation of Research Data Management (RDM) services in academic libraries of India: A triangulation approach. *The Journal of Academic Librarianship*, 48, (6), 102586. <https://doi.org/10.1016/j.acalib.2022.102586>
6. Mushi, G. (2021). Research data management and services: Resources for different data practitioners. *IASSIST Quarterly*, 45(3-4). <https://doi.org/10.29173/iq995>.
7. Huang, Y., Cox, A., & Sbaffi, L. (2020). Research data management policy and practice in Chinese university libraries. *Journal of the Association for Information Science and Technology*, 72, 4, 493 - 506. <https://doi.org/10.1002/asi.24413>
8. Zar Kyaw, A. T., Zhou, L., & Wang, Z. (2025). User requirements for research data management services in Myanmar university libraries: A case study of the University of Yangon Library. *Information Development*, 0(0). <https://doi.org/10.1177/02666669241309015>
9. Yidavalapati, J., Sinha, P., & Subaveerapandiyan, A. (2021). Research Data Management and Services in South Asian Academic Libraries. *Library Philosophy and Practice (e-journal)*. 6457. <https://digitalcommons.unl.edu/libphilprac/6457>
10. Lupu, V., Țurcan, N., & Cujba, R. (2025). Moldovan academic librarians' perception on research data management. *IFLA Journal*, 51(1), 109–127. <https://doi.org/10.1177/03400352241304118>
11. Bhakta S. (2022). Data disaggregation: the case of Asian and Pacific Islander data and the role of health sciences librarians. *Journal of the Medical Library Association : JMLA*, 110(1), 133–138. <https://doi.org/10.5195/jmla.2022.1372>
12. Nie, H., Luo, P., & Fu, P. (2021). Research Data Management Implementation at Peking University Library: Foster and Promote Open Science and Open Data. *Data Intelligence*, 3(1), 189–204. https://doi.org/10.1162/dint_a_00088
13. Ho, R. C. Y., Wong, S. N., Chia, P., Tang, C., & Ng, M. T. T. (2025). Research data management services in academic libraries to support the research data life cycle: A systematic review. *An Annual Review of Information Science and Technology (ARIST) paper. Journal of the Association for Information Science and Technology*. 1–29. <https://doi.org/10.1002/asi.70008>.

14. Amanullah S.W & Abrizah A (2023). The landscape of research data management services in Malaysian academic libraries: librarians' practices and roles". *The Electronic Library*, 4, 1. 63–86. 10.1108/EL-06-2022-0135
15. Maurya, A., & A., S. (2022). Research Data Preservation Practices of Library and Information Science Faculties. *DESIDOC Journal of Library & Information Technology*, 42, 4. 259-264, DOI : 10.14429/djlit.42.4.17538.
16. Sheikh, A., Malik, A., & Adnan, R. (2023). Evolution of research data management in academic libraries: A review of the literature. *Information Development*, 41(2), 305-319. <https://doi.org/10.1177/02666669231157405> .
17. Yuan, Y., Idaya, A., Noorhidawati, A., & Wang, G. (2025). Trends and changes in academic libraries' data management functions: A topic modeling analysis of job advertisements. *The Journal of Academic Librarianship*, 51 (2). 103017. <https://doi.org/10.1016/j.acalib.2025.103017>
18. Johnson, M. Jing, C. Daniel,W. Angela, M.(2021).Research Librarians' Experiences of Research Data Management Activities at an Academic Library in a Developing Country. *Data and Information Management* (5), 4.412-424. <https://doi.org/10.2478/dim-2021-0002>.
19. Boté-Vericad, J.-J., & Healy, S. (2022). Academic libraries and research data management: a systematic review. *Vjesnik Bibliotekara Hrvatske*, 65(3), 171–193. <https://doi.org/10.30754/vbh.65.3.1016>.
20. Liu, L.,Guo, X., Wang, Li., & Li, Xianhua.(2024). Practice on the Construction Library Data Center: A Case Study at Take Wuhan University of Technology. In *Proceedings of the ICISE-2023 8th International Conference on Information Systems Engineering 2023* (pp.42–47). Association for Computing Machinery. <https://doi.org/10.1145/3641032.3641035>
21. Viorica Lupu, Rodica Cujba, and Vera Sobetchi (2023). The Attitudes of Agricultural Researchers Towards Data Sharing: Case Study of the Republic of Moldova. In *Proceedings of the Central and Eastern European eDem and eGov Days 2023 (CEEeGov '23)*. Association for Computing Machinery, New York, USA, 84–93. <https://doi.org/10.1145/3603304.3603345>
22. Braun V. & Clarke V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77–101. <https://doi.org/10.1191/1478088706qp063oa>

A curriculum mapping study of Open Science and Research Data Management in PhD coursework of LIS in India

Mayurakshi¹ [0009-0001-0793-6501] and Preeti Mahajan² [0000-0003-2155-665X]

^{1,2} Panjab University, Chandigarh

¹ mayurakshi2605@gmail.com

² ipreeti2001@yahoo.com

Abstract. The current trend of Open Science promotes modern research practices. Research data management (RDM) is an integral part of Open Science. It involves the systematic planning, collecting, organizing, preserving, and sharing of research data. Both Open Science and RDM have the ability to reform the roles of researchers and information professionals in the present world. Library and Information Science (LIS) education is crucial for preparing scholars for data-intensive research environments, especially at the initial stage of research. The aim of this study is to assess the nature and extent of incorporation of Open Science and RDM in PhD coursework offered by LIS departments in Indian universities. The study has adopted a content analysis approach which involves analysing the content of PhD coursework documents that are publicly available online. A curriculum mapping framework was employed to assess conceptual and practical inclusion of these key concepts in research coursework syllabi. The findings revealed that while various curricula include some elements of RDM and scholarly communication, Open Science is rarely addressed. There is very limited inclusion of FAIR principles and practical skill development of the concepts. The study also highlighted gaps in LIS education at research level and advocated for the need of systematic, practice-oriented integration of Open Science and RDM concepts in PhD coursework.

Keywords: Research Data, Management (RDM), Open Science, LIS Education, Curriculum Analysis, PhD Coursework.

1 Introduction

The contemporary academic environment has noticed the transformation of traditional research into research that promotes accessibility, transparency, and reproducibility. Open Science and Research Data Management lie at the heart of contemporary research, enabling the systematic planning, preservation, sharing, and usage of research data. Data management is interconnected with Open Science and reproducibility [1]. As research is becoming data-intensive and policy-driven, competencies related to Open Science and RDM have become essential for researchers as well as library and information professionals. Open Science can be defined as “a global movement that aims to make scientific research and its outcomes freely accessible to everyone” (Center for Open Science, n.d.). Open Science is often used as an umbrella term for a wide range of practices, such as peer-review, open access, open methodologies, open data, and research data management. RDM refers to the process of documenting, storing, and preserving research data throughout the research lifecycle [5].

Libraries and Library and Information Science (LIS) professionals play a crucial role in the Open Science ecosystem. Academic libraries support data management planning and open access publishing. They are also involved in defining policies for institutional repositories, and promoting research ethics. As a result, it is the responsibility of LIS programs to equip researchers and library professionals with the skills and knowledge required to engage meaningfully with Open Science practices.

Existing studies from around the world have shown that the concepts of Open Science and RDM are being increasingly incorporated into LIS education. However, most of the focus has been on professional training initiatives, postgraduate programs, or continuing education programs. Whereas less attention has been given to research programs, particularly at the PhD coursework stage. These courses provide researchers the opportunity to systematically introduce and reinforce Open Science and RDM competencies before they start their independent research journey. This is the stage where researchers become primary producers of research data. They are involved in making decisions that directly affect data quality, ethical compliance, and long-term reuse. Although previous studies have highlighted the essence of RDM and the Open Science in educational programs, the PhD curriculum still remains under-examined, especially in the Indian context.

The main aim of the study is to examine how Open Science and RDM components are integrated in PhD coursework offered by LIS departments in central and state universities of India. The study also attempts to identify existing gaps in various curricula and inconsistencies across different universities. The findings of the study will contribute to ongoing discussions on refining the curriculum and improving researchers' readiness to handle and share data. The study will also discuss the evolving role of LIS education in promoting Open Science.

2 Review of Literature

Several studies have been conducted in the past years that emphasise the importance and inclusion of Open Science and RDM education in various subject areas.

Borghi & Van Gulick [1] highlighted the importance of data management alongside research reproducibility and Open Science concepts which can help researchers to identify the benefits at the laboratory level. The authors proposed that RDM should be incorporated directly into research workflows rather than treated as a post-publication activity. Similarly, author [6] explored the evolving role of library professionals in supporting Open Science and RDM and the impact of the 'data revolution' on academic research. The paper advocated for librarians to acquire new skills and adopt a more 'upstream' role so that they can offer comprehensive support for data lifecycle management and promote open, FAIR data practices.

Tammara & Casarosa [8] investigated the data curator education and training programs with regard to the role of data scientists and information professionals in the research lifecycle. The investigation involved conducting various workshops and events. These workshops mainly addressed the concerns of teachers and researchers about digital libraries and data curation. The findings supported the fact that there is an urgent need for proper training and education in data management.

Similarly, (Gonzalez-Velez et al., 2022) in their study, reviewed the progress of an European postgraduate program that focused on Open Science and RDM. Special emphasis was given to FAIR principles during the investigation. A comprehensive survey was conducted among researchers across Europe. The aim of the survey was to analyse and map necessary skills and training required for Open Science and RDM. The authors also considered key factors like different target groups and existing knowledge levels while analysing the data.

Faber et al. [2] launched a new initiative, PhD on Track, to help new researchers kick start their re-search journey. It was primarily a web-based resource that is designed to provide guidance to researchers for different research practices. This included searching for and reviewing scholarly literature, scientific and academic writing, sharing research data, and publishing the data and results. Likewise, authors [4] in their study, developed a research data management curriculum framework for undergraduate and graduate science students. The framework aimed to address the challenges that students faced while managing large datasets, which were often produced during their research projects. This project was also supported by librarians as they applied their information skills and helped in creating education tools for data management.

From the reviewed literature, it can be concluded that Open Science and RDM are central to the present research environment. Library professionals play a pivotal role in supporting collaborative research practices. Studies from around the world have emphasised the need for doctoral-level training in Open Science and RDM, which is supported by structured curriculum and practical engagement. There is a significant lack of empirical research on PhD coursework curriculum discussion in the context of Indian universities.

3 Objectives of the Study

The main objectives of the study are:

- To identify the presence of Open Science and RDM related concepts in PhD coursework of LIS departments in Indian universities,
- To assess the depth and coverage of Open Science and RDM competencies in PhD coursework using a curriculum mapping framework,
- To compare the incorporation of Open Science and RDM components across different categories of universities,
- To highlight the major gaps and inconsistencies in including Open Science and RDM in research programs.

4 Methodology

4.1 Research Design

The study follows a content analysis approach based on a descriptive, document-based research design. This research design is adopted to examine the integration of Open Science and RDM components in LIS PhD coursework through a curriculum-mapping study.

4.2 Scope of the Study

The study is limited to the PhD coursework syllabi in Indian universities offering LIS doctoral programs. Undergraduate and post-graduate courses were excluded from the analysis since the focus of the study is to evaluate the familiarity and preparedness of new researchers with the concepts, tools and technologies of Open Science and RDM. PhD coursework is essential for reinforcing research-related skills and competencies.

4.3 Selection Criteria

A purposive sampling strategy was used to select LIS departments according to the NIRF ranking 2025 and based on the following criteria:

- The department offers a PhD programme in Library and Information Science.
- The PhD coursework syllabus is publicly available on the official website of the university.
- Universities were divided into two categories- central universities and state universities.

12 universities were selected according to the given criteria. To maintain a constructive and non-evaluative focus, universities were anonymised in the main analysis and the list of universities is provided in the appendix for transparency.

5 Curriculum Mapping Framework

The curriculum mapping framework adopted by the paper aims to evaluate the integration of Open Science and RDM within academic curricula. The framework consists of six fundamental components - Open Science Foundation, RDM Fundamentals, FAIR Principles, Open Access and Scholarly Communication, Policy and Ethics, and Practical Application.

Table 1. Components of the Curriculum Mapping Framework

Component	Focus
Open Science Foundation	Conceptual Understanding
RDM Fundamentals	Data Lifecycle Competencies
FAIR Principles	Standards
Open Access & Scholarly Communication	Dissemination Practices
Policy, Ethics & Legal Aspects	Responsible Research Practices
Practical Application	Applied Skills Development

Table 2. Indicators of the Curriculum Mapping Framework

S.no.	Components	Indicators
1	Open Science Foundation	Open Science/Open Research Mentioned
		Transparency/Reproducibility Concepts
		Open Research Workflows
2	Research Data Management (RDM) Fundamentals	Research Data Lifecycle
		Data Collection and Documentation
		Data Storage and Preservation
		Data Analysis

		Data Sharing Practices
3	FAIR Principles	Fair Principles Explicitly Mentioned
		Metadata for Research Data
		Interoperability/Reuse
4	Open Access & Scholarly Communication	Open Access Models
		Institutional/Data Repositories
		Copyright & Licensing (CC)
		Research Visibility & Impact
5	Policy, Ethics & Legal Aspects	Research Ethics related to Data
		Privacy, Consent & Sensitive Data
		Legal/Policy Frameworks
6	Practical Application	Data Management Plans (DMP)
		Use of Repositories

Each component is divided into clearly defined indicators that enable systematic, content analysis. Table 2 represents the different indicators of the components of the framework.

1. **Open Science Foundation:** The first component of the framework assesses the conceptual grounding and understanding of the broader principles of Open Science within the academic context. The indicators under this component evaluate the inclusion of foundational ideas like Open Science, Open Research, Principles Of Transparency and Reproducibility and Research Workflows.
2. **RDM Fundamentals:** The second component of the framework aims to raise awareness regarding the foundational concept of RDM. It evaluates the inclusion of core technical skills that are required for managing research data effectively. Indicators of this component cover the entire research data lifecycle, from data collection and documentation to storage, preservation, analysis, and sharing.
3. **FAIR Principles:** This component mostly focuses on the quality and reusability of research data. These factors are guided by the FAIR principles. The FAIR (Findable, Accessible, Interoperable, Reusable) principles aim to make research data more discoverable, by both machines and humans. The indicators of this component assess the explicit mention of FAIR principles, interoperability and metadata for research data.
4. **Open Access & Scholarly Communication:** The fourth component of the framework deals with the awareness and practical implication of Open Access Publishing models, copyright and licensing information, and role of institutional and data repositories. Research visibility and impact is also included as an indicator since it represents evolving dissemination practices and gaining reach among the readers.
5. **Policy, Ethics & Legal Aspects:** Responsible research conduct is an important part of ensuring ethical research practices. Indicators of this component include ethical issues that are related to research data, privacy issues, and handling sensitive data. Legal and policy frameworks, standards and guidelines at institutional, national, or inter-national levels are also considered.
6. **Practical Application:** The final component evaluates the applied and experiential learning aspects of Open Science and RDM. Indicators of this component assess the training of researchers in Data Management Plans (DMPs) and hands-on use of repositories. This component ensures the reinforcement of theoretical knowledge through practice, enabling skill development and real-life applicability.

6 Data Collection

The syllabi of the twelve universities were downloaded through official websites. The indicators of the components of the framework were mapped against the contents of PhD coursework curricula of selected universities. The presence of each indicator was coded through a three-point ordinal scale (0-2). If the indicator was not mentioned in the syllabus, it was coded as 0. If the indicator was covered in the syllabus as a topic or a module, it was coded as 1. Practical-oriented inclusion of the component was coded as 2.

7 Results and Findings

Table 3. Curriculum Mapping of PhD coursework of Central and State Universities against the indicators of the framework.

University		OS Founda- tion (6)	RDM Fun- damentals (10)	FAIR Prin- ciples (6)	OA & Scholarly Communi- cation (8)	Policy, Eth- ics & Legal Aspects (6)	Practical Application (4)	Total Score (20)
Central Universities	University 1	0	3	0	5	5	0	13
	University 2	0	4	0	0	0	0	4
	University 3	0	3	0	5	3	0	11
	University 4	0	4	1	4	1	1	11
	University 5	0	3	1	6	3	0	13
	University 6	0	2	0	4	4	0	10
State Universities	University 7	0	3	0	1	1	0	5
	University 8	0	3	0	5	4	0	12
	University 9	0	2	0	2	0	0	4
	University 10	0	3	0	0	0	0	3
	University 11	0	3	1	1	0	0	5
	University 12	0	3	0	1	1	0	5

The findings of the study are shown in Table 3. The findings revealed that the elements of Open Science and Research Data Management competencies were unevenly represented in different PhD coursework syllabi. Moreover, their depth, coverage, and practical orientation varied considerably across universities. Surprisingly, none of the universities included the principles of Open Science explicitly in their coursework.

There was no mentions of open research workflows or the concepts of transparency and reproducibility. RDM fundamentals showed a noticeable presence among all universities, with special focus on data collection, documentation, and analysis. However, rather than being a separate module, RDM fundamentals were mostly embedded with traditional research methodology instruction. Only one university coursework included data preservation whereas there was no explicit mention of data sharing practices in any coursework. FAIR principles and related discussions were among the least represented components. Elements of FAIR terminology, interoperability or reusability were rarely discussed. Three universities included limited implicit references to metadata particularly for research data. It was mostly embedded with metadata as a generic concept. Open Access and scholarly communication emerged as a moderately represented component in the findings.

The majority of the universities' syllabi consisted of topics /modules on scholarly publishing, open access models, institutional repositories, and copyright information. Practical engagement with repositories or licensing practices was rarely evident in documented coursework. Ethics, policy, and legal aspects were a common component across most PhD curricula, especially in the form of research ethics, plagiarism, and academic integrity. Although, data-specific ethical issues like privacy, consent for data sharing or legal considerations in open data were less prominently articulated. The analysis indicated an almost non-existent emphasis on applied or practice-oriented OS–RDM training in PhD coursework. Although several syllabi mentioned tools, repositories, or contemporary research practices, explicit requirements for hands-on activities, such as preparing data management plans or engaging with data repositories, were uncommon. Most components of the framework were addressed at a theoretical or conceptual level and there was limited evidence of structured practical exercises incorporated in the coursework documentation.

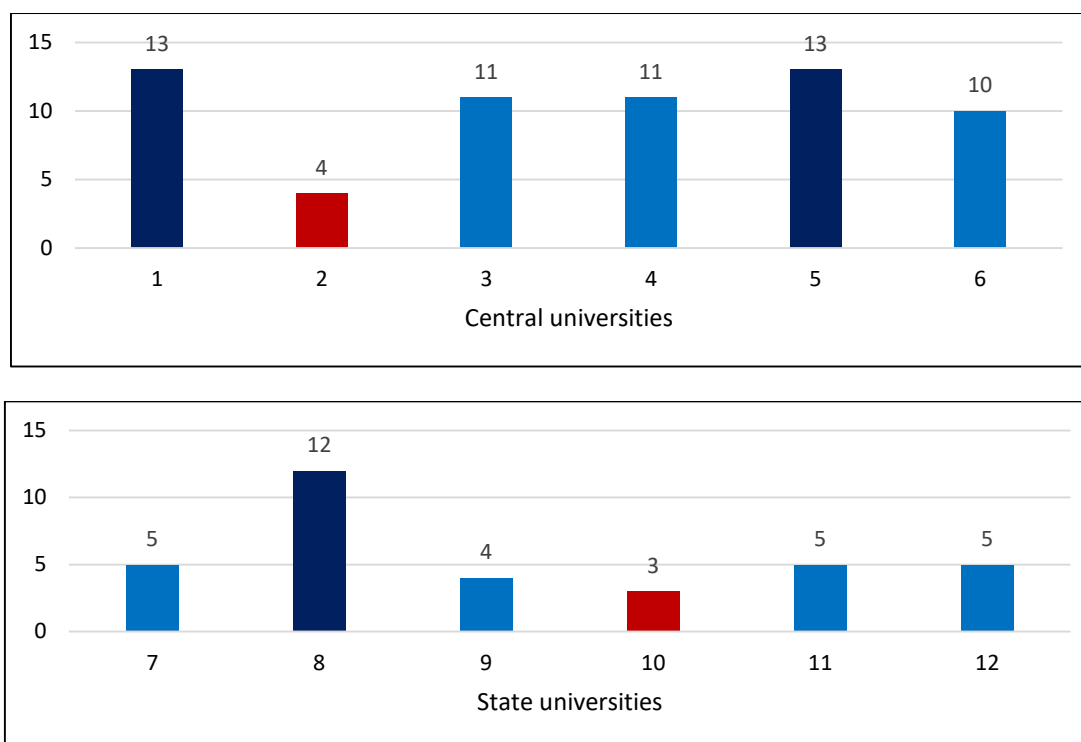


Fig. 1. Graphical representation of total scores attained by central universities vs state universities in Curriculum Mapping

Figure 1 represents the difference between central universities and state universities. While central universities showed better integration and preparedness of Open Science and RDM competencies in their PhD coursework curricula, there was a significant scope of improvement in case of state universities.

8 Conclusion and Discussion

This study involved a curriculum-mapping approach to analyse the integration of Open Science and RDM in PhD coursework offered by LIS departments in Indian universities. The key findings of the study highlighted the absence of Open Science as an explicit topic or module in almost all the coursework documents. RDM concepts were widely present, but they were mostly embedded through traditional research methodology. Almost all syllabi included foundational research and data-related concepts. Surprisingly, alignment with FAIR-oriented data practices was minimal in the LIS PhD coursework of Indian universities. Ethical concerns and considerations related to data were mentioned but not sufficiently contextualized for Open Science and data-driven research.

Looking at the references drawn from the research, it can be concluded that LIS departments in India that offer PhD programmes are not completely equipped to teach Open Science and RDM competencies. Their syllabi lack systematic and practice-oriented integration of these concepts. The results also show that the doctoral coursework is often more focused on theoretical understanding rather than practical application. This leaves a significant scope for improvement in the curriculum. PhD coursework is the final training phase of a researcher before they embark on their independent research journey. Therefore, it is essential to include the theoretical and practical understanding of these concepts during coursework so that it aids in enhancing research quality, researchers' preparedness and long-term data reuse.

The study provides empirical evidence from curriculum documents and contributes to ongoing discussions on renewing and reforming LIS curriculum and doctoral education in India. It is the current need of the hour for both central and state universities to revisit their LIS curricula and look for improvements. Integration of Open Science and RDM into LIS PhD coursework will help in aligning LIS education with modern research practices.

9 Limitations of the Study

This study has certain limitations. Firstly, the study focuses solely on PhD coursework and does not take into account undergraduate (BLIS) or postgraduate (MLIS) programs. Secondly, the analysis is based on publicly available PhD coursework documents. As a result, the study has captured documented curricula rather than actual

discussions in the classroom, informal training activities, or guidance provided by the supervisor that may enhance Open Science and RDM skill development. And lastly, due to purposive sampling, the scope of the study is limited to institutions with accessible documentation.

10 Future Scope

This study primarily involved content analysis of PhD coursework syllabi. The next step towards future research involves a longitudinal curriculum mapping study of Open Science and RDM integration across undergraduate (BLIS), postgraduate (MLIS) and PhD programs. This will provide a broader overview of the presence of Open Science and RDM in LIS education and a comparative analysis will help universities to understand how to reframe the syllabus. Future studies can also adopt a mixed-methods approach, combining curriculum analysis with student surveys, faculty interviews, or classroom observations to reflect actual teaching practices and learning outcomes related to Open Science and RDM. The doctoral coursework can be complemented by improving research support services, conducting training workshops, and refining institutional policies regarding data curation.

Declaration of AI and AI-Assisted Technologies used in Manuscript Writing Process: Not applicable

References

1. Borghi, J., & Van Gulick, A. (2022). Promoting Open Science Through Research Data Management. *Harvard Data Science Review*. <https://doi.org/10.1162/99608f92.9497f68e>
2. Faber, H. C., Kavli, F. A., Grote, M., & Låg, T. (2018). PhD on Track: Open Science as a new track. *Septentrio Conference Series*, 1. <https://doi.org/10.7557/5.4526>
3. GonzalezVelez, H., Dobre, C., SanchezSolis, B., Antinucci, G., Feenan, D., & Gheorghe, D. (2022). Open Science and Research Data Management: A FAIR European Postgraduate Programme. *2022 IEEE International Conference on Big Data (Big Data)*, 2522–2531. <https://doi.org/10.1109/BigData55660.2022.10020931>
4. Piorun, M., Kafel, D., Leger-Hornby, T., Najafi, S., Martin, E., Colombo, P., & LaPelle, N. (2012). Teaching Research Data Management: An Undergraduate/Graduate Curriculum. *Journal of eScience Librarianship*, 46–50. <https://doi.org/10.7191/jeslib.2012.1003>
5. *Research Data Management*. (n.d.). NNLM. Retrieved January 8, 2026, from <https://www.nlm.gov/guides/data-glossary/research-data-management>
6. Rice, R. (2019). Supporting Research Data Management and Open Science in Academic Libraries: A Data Librarian's View. *Mitteilungen Der Vereinigung Österreichischer Bibliothekarinnen Und Bibliothekare*, 72(2), 263–273. <https://doi.org/10.31263/voebm.v72i2.3303>
7. Science, C. for O. (n.d.). *What is open science?* Retrieved November 10, 2025, from <https://www.cos.io/open-science>
8. Tammara, A. M., & Casarosa, V. (2014). Research Data Management in the Curriculum: An Interdisciplinary Approach. *Procedia Computer Science*, 38, 138–142. <https://doi.org/10.1016/j.procs.2014.10.023>

Appendix

List of universities selected for the content analysis

Central University

1. University of Delhi
2. Banaras Hindu University
3. Aligarh Muslim University
4. Babasheb Bhimrao Ambedkar University
5. Mizoram University
6. Central University of Tamil Nadu

State Universities

7. Jadavpur University
8. Panjab University
9. University of Kashmir
10. University of Mysore
11. Maharshi Dayanand University
12. Kerala University

Roles and Skills in LIS

Reforming LIS Curriculum in Line with NEP 2020: Recognizing Indian Knowledge Systems

Rajesh Rangappa Aldarthy¹[0000-0002-7266-0601], Manoj Kumar Sinha²[0000-0002-2145-0006] and Moirangthem Sajan Singh³[0009-0002-2227-7353]

¹ Department of Library and Information Science, Assam University, Silchar, 788011, India, rajesh.aldarthy@gmail.com, rajesh.rangappa.aldarthy@aus.ac.in

² Department of Library and Information Science & Dean, Swami Vivekananda School of Library Sciences, Assam University, Silchar, 788011, India, mksinha1965@gmail.com, manoj.kumar.sinha@aus.ac.in

³ Department of Library and Information Science, Assam University, Silchar, 788011, India, sajanmoirangthem@gmail.com

Abstract. India is renowned for its cultural heritage and rich traditional learning. Educational institutions such as Gurukul schools, Nalanda, Vikramashila, and Takshashila universities were renowned for their academic reputation worldwide. These universities had maintained huge libraries. Accounts of Chinese travellers, such as Fa-Hien, Hieun-Tsang, and I-Tsi, highlight the importance of these university libraries. However, the genesis of formal LIS education in India began with Raja Sri Sayaji Rao Gaekwad III of Baroda in 1911 and the appearance of Asa Dickinson as a librarian at Punjab University in 1915. After independence, the first education commission, under the chairmanship of Dr. S. Radhakrishnan, recommended the proper functioning of university libraries. Since full-fledged Library and Information Departments were established at the University of Delhi, Aligarh Muslim University, and Madras University. The Library Committee (1957), the Review Committee on Library Science Education (1965), and the UGC Curriculum Development Committees (1990 and 2001) recommended the improvement and standardization of the LIS syllabus. Since the 1980s, it has expanded in scope and coverage. This study examines the LIS curriculum of central and state universities of North-Eastern India to provide a framework for the inclusion of Indigenous Knowledge Systems (IKS) into the LIS syllabus in alignment with the National Education Policy (NEP) 2020. The finding suggests that there is no reflection of the Indian knowledge system in the currently adopted curriculum. Therefore, revamping the syllabus is necessary in line with the NEP 2020's policy recommendations regarding the Indian knowledge system.

Keywords: LIS Education-India, LIS Curriculum, NEP-2020, Indian Knowledge System (IKS), Revamping of Curricula, Central Universities, and State Universities, North-East India.

1 Introduction

Libraries are the gatekeepers of knowledge and also serve as an aid to education and the acquisition of knowledge. During ancient and medieval India, libraries were regarded as important centres of learning and a valuable source of knowledge. Indian ancient libraries have a long history associated with the education system, and scholarship. Accounts of Chinese travellers, such as Fa-Hien, Hieun-Tsang, and I-Tsi, mentioned the importance of Nalanda, Vikramashila, and Takshashila university libraries. However, formal education in Library and Information science in India began with Raja Sri Sayaji Rao Gaekwad III of Baroda. Various universities, such as the University of Delhi, Aligarh Muslim University, and Madras University, are the first universities to run a full-fledged LIS department. The LIS curriculum of Indian universities follows universal principles, theories, techniques, classification and cataloguing, information organization, management, and information technologies. However, there is no reflection of integrating specialised units on Indian information systems, socio-cultural information needs, or indigenous knowledge systems.

Meanwhile, the National Education Policy (NEP) 2020, a replacement of the 1986 National Education Policy, aims to transform the education system from school level to higher education in India [1]. NEP 2020 advocated Indian universities to integrate India's indigenous knowledge, traditions, languages, and culture into curricula across disciplines. According to Nakata and Langton [2] "the increasing global recognition of Indigenous knowledge as distinct, legitimate, valuable and vulnerable systems of knowledge raises a range of issues that pose challenges for the library and information services (LIS) sector". As libraries are custodians of cultural memory and community knowledge, integration of Indian Knowledge Systems (IKS) into LIS syllabus will make library education relevant to the Indian knowledge system in line with NEP 2020.

2 Literature Review

Since the colonial period, indigenous knowledge system and traditional learning of community were systematically marginalised by the western centric philosophy and practice [1]. The education systems of postcolonial societies are often deeply rooted in colonial epistemology, structures, and languages. Inherited systems of colonialism tend to marginalize indigenous ways of knowing and traditional learning. Thus, the call for decolonized education is a new urgency worldwide [3]. Recognition of indigenous knowledge is increasing in education globally, but there are certain limitations [4]. As indigenous knowledge and traditional practices are rapidly disappearing and often remain undervalued, misused, or inadequately protected by current systems [5]. Academic libraries are considered as neutral spaces at the heart of the university, but hegemony of colonial recognition structure still exists [6]. Roles of libraries, archives, and museums are critical. Extension of LIS education in preserving indigenous knowledge through proper training, ethical community collaboration, and protection of intellectual property rights are crucial in the digital age [7]. In the context of Indian education, recently the National Education Policy (NEP-2020) encouraged the inclusion of Indian Knowledge Systems in secondary and higher education levels [2]. It also promotes India's intellectual traditions along with holistic and multidisciplinary learning [8]. Indigenous knowledge in LIS education remains marginal, often treated as secondary or optional content [9]. According to Lilley & Paringata overseas LIS related programmes provide necessary technical skills for professional library service, however, also most likely do not include significant amount of local content [10]. Hence, Libraries are agents of educational change, rather than static archives, but its role is underexplored [11]. While digitization projects like the Digital Library of India initiated by the government of India in preserving IKS, application to teaching and learning is limited [12]. Although NEP 2020 provide a strong mandate to incorporate the Indian knowledge system to reform Indian education systems, there is need of effective execution and continuous adaptation [13]

3 Objectives

- To examine provision of National education Policy 2020 on adaptation of Indian Knowledge Systems in the Indian education system.
- To identify the IKS related component from Library and Information Science syllabus of selected central and state universities in North-Eastern India.
- To assess the alignment between NEP 2020 mandates and LIS curriculum design with respect to IKS and curriculum.
- To propose curricular reforms framework to align LIS education with NEP 2020 and Indian Knowledge Systems.

4 Methodology

This study used qualitative approach to examine the incorporation of Indian Knowledge Systems within Library and Information Science (LIS) syllabus. No computational or automated techniques such as Natural Language Processing (NLP) were employed for data collection, textual analysis, or interpretation. The researchers examined thirteen state and central universities in North-Eastern India recognized by the University Grants Commission (UGC). Eleven universities were considered for analysis because these universities offered BLIS or MLIS programs, and their official library and information science syllabus were accessible on their websites. Any university that did not offer these programs or didn't publish their curriculum publicly were excluded from analysis. The analysis was conducted in two stages, with AI assistance limited exclusively to visualisation purposes. In the first stage, the National Education Policy (NEP) 2020 was examined manually to identify explicit and implicit policy provisions related to Indian Knowledge Systems (IKS) in higher education. Researchers identified the sections and clauses relevant to IKS elements such as indigenous knowledge, traditional knowledge, cultural heritage, community-based learning, local knowledge, and digital preservation. These elements were thematically coded to extract NEP 2020 directives relevant to curriculum design and reform. In the second stage, LIS curriculum documents of selected institutions were analysed for the presence of IKS-related components, including dedicated courses, embedded units or modules, indigenous or community-based fieldwork, and representation of IKS in prescribed readings. The study also proposes a framework to integrate IKS in the LIS syllabus in alignment with NEP-2020. An AI-assisted generative tool (Google Gemini) was used only after the analytical stages, exclusively for the visual representation of the proposed conceptual framework. The framework's structure, components, and interpretation were entirely author-developed, and the AI tool did not contribute to data analysis, coding, or interpretation.

5 Result

5.1 Position of NEP 2020 on Indian Traditional Knowledge

The National Education Policy (NEP) 2020 released by the Government of India on 29 July 2020 provided a blueprint for an ideal future education system in India. This policy also advocated the integration of Indian Knowledge Systems (IKS) into curricula. NEP-2020 emphasizes that education must be rooted in India's intellectual traditions while remaining globally relevant. Table 1 maps the key clauses of NEP 2020 that advocate the inclusion of Indian traditional knowledge within curricular frameworks across school and higher education.

Table 1. Position of NEP 2020 on Indian Traditional Knowledge

NEP 2020 Section / Clause	Policy Focus	Key Provision Related to ITK
Introduction & Vision of NEP 2020	Philosophical foundation of education	Recognises India's ancient knowledge traditions (<i>Jnana, Pragyaa, Satya</i>) and institutions such as Takshashila, Nalanda, and Vikramashila as models of multidisciplinary knowledge creation
Para 4.27	School Education – Curriculum	Mandates inclusion of Indian Knowledge Systems, including tribal, indigenous, and traditional knowledge, across disciplines such as mathematics, astronomy, medicine, agriculture, architecture, linguistics, and philosophy
Paras 4.28–4.29	Values, ethics, and cultural rootedness	Advocates embedding Indian ethical traditions, cultural narratives, local context, and indigenous practices into teaching-learning processes
Para 4.16	Linguistic and cultural diversity	Promotes learning about Indian languages, scripts, oral traditions, and cultural unity through curricular projects
Para 11.4	Multidisciplinary learning	Calls for holistic, multidisciplinary education integrating Indian arts, humanities, sciences, and knowledge traditions
Para 11.9	Indian Knowledge Systems in higher education	Explicitly recommends offering courses on Indian Knowledge Systems and integrating them into mainstream curricula
Para 12.3	Experiential and community-based learning	Emphasises field-based learning, internships, and community engagement as core pedagogical approaches
Para 13.4	Teacher capacity building	Encourages faculty development programmes to include Indian knowledge traditions and culturally rooted pedagogy
Para 17.2	Research on Indian knowledge traditions	Encourages research funding for Indian languages, culture, arts, and indigenous knowledge systems
Para 22.1–22.3	Cultural and knowledge preservation	Advocates systematic promotion, documentation, and transmission of Indian languages, arts, and cultural knowledge
Para 23.6	Technology for knowledge preservation	Promotes use of digital tools to preserve cultural heritage, manuscripts, and traditional knowledge

For instance, Section 4.27 of NEP 2020, mandates the inclusion of Indian Knowledge Systems, including tribal knowledge, indigenous and traditional ways of learning, in teaching and learning processes across disciplines. Paras 4.28–4.29 argues for treating traditional knowledge as a legitimate epistemic resource, to be examined scientifically and analytically, rather than relegating it to the status of folklore or cultural symbolism. Further, Section 4.29 stresses the importance of local and regional languages, cultures, and traditional practices in curriculum design. NEP 2020 also promotes experiential and field-based learning through engagement with communities in

para 4.6 and 4.7. Universities are encouraged to introduce dedicated courses on Indian Knowledge Systems across disciplines (paras 11.4, 11.9)

The National Research Foundation was proposed to support research on Indian languages, arts, culture, and indigenous knowledge systems (para 17.2). In addition, the policy advocates the use of digital technologies for the documentation, preservation, and dissemination of Indian traditional knowledge (para 23.6). Overall, NEP-2020 provides a framework for the integration of Indian Knowledge Systems into higher education curricula.

5.2 Indian Knowledge Systems in LIS Curricula

To examine the inclusion of Indian Knowledge Systems (IKS) in LIS education, the curricula of eleven selected central and state universities in North-Eastern India were analysed using a structured evidence framework. The analysis focused on identifying explicit curricular indicators such as IKS, including indigenous, tribal, traditional, and community-based knowledge practices.

Table 2. LIS programme availability and integration of Indian Knowledge Systems across selected North-Eastern universities.

University	Type	LIS Course	Dedicated IKS Course	IKS Unit / Module	Indigenous / Tribal Fieldwork	IKS in Pre-scribed Readings
Assam University, Silchar	Central	Yes	No	No	No	No
Dibrugarh University	State	yes	No	No	No	No
Guwahati University	State	yes	No	No	No	No
Cotton University	State	yes	No	No	No	No
Mizoram University	Central	yes	No	No	No	No
North-Eastern Hill University (NEHU)	Central	yes	No	No	No	No
Tripura University	Central	yes	No	No	No	No
Manipur University	Central	yes	No	No	No	No
Nagaland University	Central	NA	--	--	--	--
Rajiv Gandhi University (Arunachal Pradesh)	Central	NA	---	--	---	--
Sikkim University	Central	yes	No	No	No	No
Assam Women's University	State	yes	No	No	No	No

Five indicators were used to assess the presence of IKS in the curricula:

- existence of a dedicated IKS course
- inclusion of IKS-related units or modules within existing courses
- provision for indigenous or tribal fieldwork
- engagement with community-based knowledge practices
- inclusion of IKS-related content in prescribed readings

Table 2 presents central and state universities in North-East India offering Library and Information Science (LIS) programmes. Universities such as Assam University, Dibrugarh University, Guwahati University, Cotton University, Mizoram University, North-Eastern Hill University, Tripura University, Manipur University, Sikkim University, and Assam Women's University currently run LIS courses. In contrast, Nagaland University and Rajiv Gandhi University do not offer LIS programmes. However, none of universities offer IKS-specific units or modules within their LIS syllabi.

The prescribed reading lists do not reflect the inclusion of IKS related reading materials across the syllabus which indicated that there is no substantive engagement with Indigenous Knowledge Systems (IKS). Units titled Foundations of Library and Information Science, Library, Information and Society, or Role of Libraries in Society appear consistently within the syllabus. These courses address the library as a social institution, the role of libraries in education and national development, professional ethics, library legislation, library movements in India, and the Five Laws of Library Science.

The syllabi also include information society, information flow in society, library extension services, and the relationship between libraries and education, culture, research, and recreation. History of libraries in India, and the development of public library systems, along with the role of national and international bodies such as UNESCO, IFLA, RRRLF, and INFLIBNET. However, there is no explicit engagement with Indian Knowledge Systems within these courses.

Beyond the binary identification of explicit IKS-related modules, in-depth examination of syllabus content revealed that Indian Knowledge Systems (IKS) are not implicitly embedded within existing LIS courses. Units such as Library, Information and Society, Role of Libraries and Information Centres in Society, History and Development of Libraries in India, Public Library Systems, and Museums and Archives acknowledged the social, cultural, and historical roles of libraries. For instance, units on the history of libraries in India describe the development of libraries during ancient civilizations and the evolution of public library systems. However, these units do not engage with indigenous knowledge traditions, oral knowledge systems, tribal learning practices, or community-based modes of knowledge transmission. Other units related to Museums and Archives focus on institutional preservation, collection management, and archival functions, but do not address community archives, oral histories, or ethical issues surrounding the documentation of indigenous knowledge. Likewise, public library-related units emphasise extension services and literacy programmes, yet to integrate local or tribal knowledge practices as knowledge resources.

5.3 Development of an IKS-Centric LIS Curriculum Framework

Drawing on an analysis of LIS curricula across selected central and state universities, the study proposes the following curricular reforms to align LIS education with the vision of NEP 2020 and the integration of Indian Knowledge Systems (IKS). Fig. 1 presents a conceptual framework for aligning Library and Information Science (LIS) education with the National Education Policy (NEP) 2020 through the systematic integration of Indian Knowledge Systems (IKS). At the centre of the model is the idea of IKS-centric LIS education, which reflects NEP 2020's epistemic plurality and curricular decolonisation and surrounding six are the interrelated components.

First, introduction of Indian Knowledge Systems as a mandatory paper or module in BLIS OR MLIS course. This module discussed on indigenous knowledge, oral traditions, community-based knowledge practices and traditional knowledge and such a module would provide conceptual responses directly to NEP 2020. This would provide students with a foundational understanding of indigenous knowledge, oral traditions, and community-based knowledge systems. This would equip LIS professionals with conceptual awareness of culturally rooted knowledge forms and their documentation needs.

The second component focused on embedding IKS within existing core LIS courses rather than being treated as an optional theme. For example, knowledge organization of indigenous knowledge, such as information sources and services can incorporate indigenous, vernacular, and community sources. This approach promotes integration indigenous knowledge into areas such as knowledge organisation, information sources and services, preservation and conservation, and digital libraries. This would enable LIS professionals to apply standard LIS principles to indigenous and vernacular knowledge contexts.

Third, community-based experiential learning should be made compulsory within LIS programmes. Internships, fieldwork, and documentation projects should extend beyond formal library settings to include community libraries, tribal cultural institutions, oral history initiatives, and local heritage repositories. Such engagement enhances ethical sensitivity, participatory documentation skills, and community collaboration, strengthening libraries' roles in cultural preservation. This component will also help LIS professionals to learn how to work with communities to document, preserve, and protect their indigenous knowledge.

The fourth component proposes an elective or laboratory-based course on oral history and community archives. This component focuses on practical skills such as oral documentation methods, participatory archiving, ethical safeguarding, and the creation of community archives. This would enhance oral tradition documentation method, ethical archiving, and participatory knowledge preservation. This component will prepare LIS graduates to support community-driven archival initiatives and safeguard intangible cultural heritage.

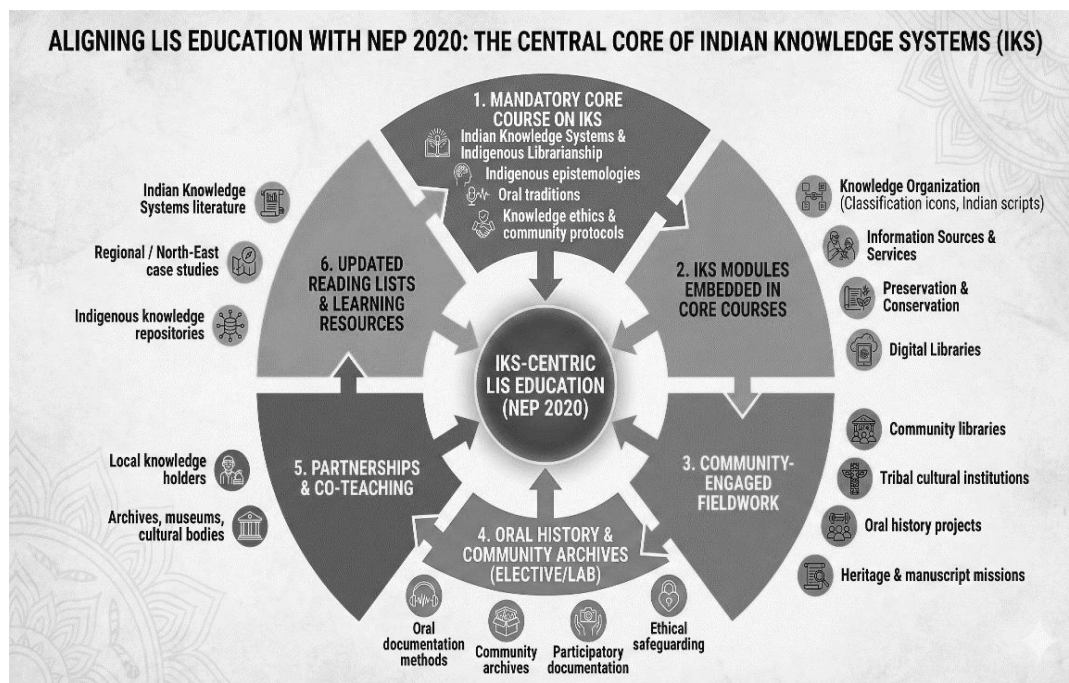


Fig. 1. Framework for Reforming LIS Curriculum in Alignment with NEP 2020 and Indian Knowledge Systems

The fifth component introduces collaboration with local expert members and institutions such as archives, and museums. This component encourages collaboration with local knowledge holders, archives, museums, and cultural institutions. This partnership will enable Library and Information Science (LIS) professionals to work with communities to document, organize, and preserve indigenous and traditional knowledge. This will enable LIS professionals to record indigenous knowledge in communities' own voices, respect community consent and ownership, and avoid misuse or misrepresentation of such knowledge.

The sixth component focuses on the inclusion of learning materials related to the Indian Traditional Knowledge System in LIS curricula. Material such as books, articles, case studies, policy documents, and digital resources that reflect indigenous knowledge, tribal, and traditional knowledge practices of India. Inclusion of such materials into prescribed reading lists will help LIS students to understand Indian knowledge traditions in their cultural and social contexts. It also supports the LIS profession in responsibly preserving, documenting, and promoting traditional knowledge, while contributing to the protection and continuity of India's cultural heritage.

Overall, this framework provides a conceptual foundation for the inclusion of the Indian knowledge system in LIS education in aligned NEP 2020 visions. This framework also provides a comprehensive curriculum design by demanding meaningful changes in teaching practices, field engagement, and learning materials.

6 Discussion

Traditional and indigenous knowledge are useful for people in making informed decisions. It exists in the forms of traditional norms, practices, language, stories, folk-tales and passed down orally from one generation to another generation. Colonization, urbanization, and formal education systems disrupt the traditional way of living and learning. Another important factor is the death of knowledge holders without passing their knowledge to younger generations. Developing policies and programmes to preserve indigenous knowledge is crucial for safeguarding traditional knowledge and practices. Through the education system, inclusion of indigenous knowledge in curricula will promote research, aids to preserve and conserve endangered knowledge in the digital era. The government of India on 29 July 2020 released the National Education Policy (NEP) 2020 and in the same year Indian Knowledge Systems Division (IKS Division) was established by the Ministry of Education in October 2020 to

promote interdisciplinary and transdisciplinary research on all aspects of Indian Knowledge Systems (IKS). NEP-2020 promotes the inclusion of India's traditions and intellectuals in the existing education system as an integral component in both school and higher education curricula (NEP 2020, paras 4.27–4.29, 11.9). NEP 2020's para 12.3 encourages the experiential and community-based learning and in para 17.2 to provide research support for Indian knowledge traditions through the National Research Foundation. It also highlights the necessity of digitisation for conservation, preservation and dissemination of traditional knowledge (para 23.6). The study reveals that within LIS curricula of central and state universities of North East India the integration of IKS remains limited to non-existent. Foundational units of LIS such as Foundations of Library and Information Science, Library, Information and Society, or Role of Libraries in Society appear are common across syllabi. However, assimilation of IKS related units or modules are absent from these syllabi. Reading lists or references given within the syllabus related to IKS are absent which reflect a gap. As per the National Education Policy (NEP) 2020, Indian Knowledge Systems should be included in the LIS syllabus.

7 Conclusion

Indigenous knowledge and traditional practices are rapidly disappearing and facing risk of extinction and endangered from various communities across the globe due to colonisation and modernisation. Various organizations and government bodies have begun discussions on how to conserve and preserve indigenous knowledge and traditional practices. The government of India established the Indian Knowledge Systems Division (IKS Division) to support the conservation, preservation, and promotion of the Indian knowledge system. Libraries are also key stakeholders in conservation and preservation of it. Therefore, integrating Indian Indigenous Knowledge Systems (IKS) related modules or units into the LIS curriculum is essential in the digital age. Library and Information Science (LIS) curriculum in North-Eastern India remains unaddressed. Curricula tend to adopt conventional LIS frameworks, no areas for incorporating Indian indigenous knowledge. Integrating Indian Knowledge Systems into LIS education also involves several challenges. These include the limited availability of learning resources, as much indigenous knowledge is transmitted orally. Insufficient funding for documentation, curriculum development, and community-based activities is another significant issue. Ethical considerations are also important when engaging with indigenous communities, particularly the need for informed consent, respect for community ownership of knowledge, and the avoidance of misuse or misrepresentation. In conclusion, integrating Indian Knowledge Systems into LIS education is an essential need in a diverse society like India. This would contribute to a more inclusive and context-sensitive evolution of LIS education in India.

Declaration of AI and AI-Assisted Technologies used in Manuscript Writing Process: Not Applicable

References

1. Government of India. (2020). *National Education Policy 2020*. Ministry of Education. https://www.education.gov.in/sites/default/files/NEP_Final_English_0.pdf
2. Nakata, M., Byrne, A., Nakata, V., & Gardiner, G. (2005). Indigenous knowledge, the library and information service sector, and protocols. *The Australian Library Journal*. <https://doi.org/10.1080/00048623.2005.10721244>
3. Yesmin, S., & Peter, C. (2025). Decolonizing education systems for inclusive development. In *Rethinking Development: Ideas, Institutions, and Impact in a Changing World* (pp. 12-23). Indian Institute of Industrial and Social Research
4. Das, Rajib Kumar, Kalita, Khagen & Singha, Anuradha (2025). The role of Libraries in Digitizing Indian Knowledge Systems for Technology-Enhanced Teacher Education. *EPRA International Journal of Multidisciplinary Research (IJMR)*, 190–195. <https://doi.org/10.36713/epra20532>
5. Omarsaib, M., Naicker, N., & Rajkoomar, M. (2023). Integration of indigenous knowledge into library and information science teaching practices: A systematic review of the global literature. In *Digital Preservation and Documentation of Global Indigenous Knowledge Systems*. <https://doi.org/10.4018/978-1-6684-7024-4.ch016>
6. Crilly, J. (2024). Diversifying, decentering and decolonizing academic libraries. *New Review of Academic Librarianship*. <https://doi.org/10.1080/13614533.2024.2392309>
7. Oyelude, Adetoun A. 2023. "Indigenous Knowledge Preservation as a Sign of Respect for Culture: Concerns of Libraries, Archives and Museums." *Insights the UKSG Journal* 36 (October): 21. <https://doi.org/10.1629/uksg.628>.
8. Bhat, S., & Girish, A. (2024). Evaluating the promise and pitfalls of India's National Education Policy 2020. *SAGE Open*. <https://doi.org/10.1177/21582440241279367>
9. Adams, A. (2023). Indigenous knowledge preservation as a sign of respect for culture: Concerns of libraries, archives and museums. *Insights*, 36, 21. <https://doi.org/10.1629/uksg.628>

10. Lilley, S., & Paringata, T. (2014). Kia whai taki: Implementing Indigenous knowledge in the Aotearoa New Zealand library and information management curriculum. *The Australian Library Journal*. <https://doi.org/10.1080/00048623.2014.908498>
11. Chigwada, J., et al. (2024). Librarians' role in the preservation and dissemination of indigenous knowledge. *IFLA Journal*. <https://doi.org/10.1177/03400352231217270>
12. Patel, Damyanti, and Leo Appleton. 2024. "Diversifying, Decentering and Decolonizing Academic Libraries." *New Review of Academic Librarianship* 30 (2–3): 107–11. <https://doi.org/10.1080/13614533.2024.2392309>
13. Ashokkumar, T., Russel Raj, T., Rajadurai, A., Abishini, A. H., & Anchani, A. H. (2025). Analyzing the impact of the new educational policy 2020: A comprehensive review of India's educational reforms. *Evaluation and Program Planning*, 108, 102515. <https://doi.org/10.1016/j.evalprogplan.2024.102515>

AI Across the LIS Curriculum in India: An Empirical Syllabus Analysis Using Text Analytics

Adithya Mohan ¹[0009-0008-3849-9085], Neethusha P. ²[0000-0001-6329-0541] and S. Ravikumar ³[0000-0002-2317-1596]

¹⁻³Central University of Tamil Nadu

¹aadithyaaajeesh@gmail.com

²neethushapanachikkalofficial@gmail.com

³ravikumar@acad.cutn.ac.in

Abstract. This article reports a study of AI in LIS curricula at 104 universities across India through systematic, text-analytic means. Based on AI literacy and "AI across the curriculum" theoretical frameworks, coverage was assessed relative to 5 subcategories of AI: foundations, data analytics, language technologies, library applications, and ethics/governance. An AI dictionary of 109 terms allowed for the implementation of quantitative data to assess the presence, prevalence, and topicality of AI themes within BLIS and MLIS course offerings to relative significance across the subcontinent. Ultimately, AI themes are present in 100% of the 71 effective syllabi assessed, yet coverage is superficial—and dissociation/coverage mostly exists in "buzzwords"—machine learning, natural language processing, generative AI—inclusive courses dependent upon technology. For example, the technical aspects of "Foundations" and "ML & Analytics" are rated higher than the practical findings of "Library Applications" and "Ethics & Governance". Thus, findings indicate a disassociated polarized difference between contemporary academic offerings and 21st century skills necessary for knowledge-rich environments. Such a "digital divide" renders graduates without adequate AI competencies to pursue their next steps (professionally or academically) with certainty. Recommendations for curriculum development include implementing AI literacy basics integrated into beginner courses with enhanced capabilities through applied data analytics in advanced courses with a more humane approach to bias in algorithms and privacy considerations. Ultimately, the study justifies a replicable text-analytic approach for longitudinal study and future international comparative analysis.

Keywords: Artificial Intelligence (AI); AI Literacy; LIS Curriculum; Text Analytics; Curriculum Development; Indian Universities.

1 Introduction

Artificial Intelligence (AI) is one of the latest transformative developments across various domains and fields, including Library and Information Science (LIS) and librarianship. With almost all professions being impacted by AI developments concerning professionals' duties, whether or not that field of work involves AI components, all occupations can feel AI's presence as increasingly, discovery, organization, and dissemination involve AI developments. Therefore, the need for LIS professionals to be aware of and skilled in operations involving AI is paramount. Thus, the growing need for AI literacy in higher education courses across the disciplines is a natural progression.[17][11]

In particular, the notion of "AI across the curriculum" supports the awareness of AI concepts, ethics, technologies, etc., in multiple disciplines instead of solely Computer Science departments.[13] This is especially true for LIS education, where AI would only boost the librarianship experience in an increasingly nuanced twenty-first-century information ecosystem. Future LIS professionals will need to ensure and tolerate human-computer interactions for basic literacy competencies and employment demands. In addition, there are many basic components of AI—and accompanying ethics—that future professionals should be aware of—machine learning, natural language processing (NLP), fairness, accountability, transparency and ethics (FATE).[10]

Unfortunately, there is little literature to date on where AI fits in LIS education syllabi in Indian curricula. Therefore, considering the gradual emergence of AI concepts into syllabi over years will reflect how well this new phenomenon is encouraged based on how important it is to LIS curricula. In this thesis, a systematic text-analysis study will assess the presence, degree and relevance of AI concepts in Indian LIS syllabi. The goal of this research is to reveal findings about how AI impacts Indian LIS syllabi based on qualitative data collection from previously published syllabi. Ultimately, the goal is to obtain findings that will support changes to LIS program curricula for the best information acquired to serve India's LIS professional educational and vocational needs.

The study's objectives are to comparative analyze the LIS syllabi of Indian universities for AI implementation in professional educational and vocational needs. This study aims to answer the following research questions:

- What is the level of AI implementation in the LIS courses of universities in India?
- What are the differences among Indian universities within LIS courses in terms of current standing of AI integration?
- What are the other trends of curriculum structure that emerge to provide professional educational and vocational needs.?

In conclusion, these answers will benefit an improved understanding of LIS coursework throughout the nation and how they can be better suited to provide LIS graduates with a more prepared approach to the AI era.

2 Literature Review

LIS education in India has expanded from its foundational roots under pioneers like W.A. Borden and Dr. S.R. Ranganathan into a vast academic network. Academic libraries are undergoing a major transition driven by changing user expectations and digital resources [20]. This growth is evidenced by the existence of over 230 institutions offering courses ranging from certificates to doctoral degrees across 31 states. A research [2] provides a Vision 2021 for the field, they are advocating for the integration of Web 3.0 technologies and trying to shift from the technical focus to the service-oriented competencies.

The central theme of the recent studies revealed that there is a transition from traditional cataloguing to digital management. Works [9] identified that the term digital era means a period which requires a new proposition which addresses the proliferation of LIS schools that often lack faculty and infrastructure. Author [6] points out that while the practical components of ICT are gaining much importance nowadays, there is a significant gap in how these are analysed across the various universities. Similarly, another work [7] emphasizes a persistent tension between the teaching core subjects of LIS and necessary technology-related topics, noting that many programs still rely on outdated methods like chalk and talk methods rather than the advanced teaching techniques and methodologies.

A recurring concern is the parity gap between the university syllabus and the skills which are required for modern employability. Another work [3] critically evaluate this gap, they are suggesting that unfortunately the current curriculum often fails to meet the demands of the market. Works such as, [15] and [12] further argued that the wheels of Indian LIS education are moderately slowed by a lack of uniformity and the failure to update the content which reflect the needs of a competitive and technology-driven job market. Author [14] emphasizes that the employers are increasingly dissatisfied with the current practical skills of the graduates, who often lack the professional skills and managerial abilities which are required for handling the complex information needs of the current global society.

Another work, [18] related empirical studies show that the AI topics like Machine Learning and NLP are beginning to appear in the syllabus, their presence are often seemingly present through the buzzwords rather than the deep practical integration. Authors [4] identify that the Digital Humanities act as an emerging interdisciplinary area of study in Indian LIS curricula, though it is still in its initial stage with a limited technological expertise being offered. Papers such as, [19] and [8] suggest that to move beyond the core LIS skills towards transformative learning and 21st-century has a competency such as critical thinking, media literacy, and AI ethics which prepare the LIS graduates to fight the Web 3.0 environment.

The researchers are consistently identifying several barriers to the progress, including an acute shortage of a well experienced teaching staff and a lack of adequate infrastructure for the students. Muthu et al. (2015) highlight that many librarians fear technology due to receiving training before the advent of digital scholarship. To address these issues, some papers [5] and [16] recommend a radical overhaul of curriculum strategies, favouring the Learning Outcomes-based Curriculum Framework (LOCF) and interdisciplinary collaborations to ensure that LIS professionals remain relevant in the age of AI and data science.

3 Methodology

A systematic text-based approach was adopted to study artificial intelligence in Indian university-based LIS program curricula. Data collection, corpus construction, and text-analytic quantitative approach are performed in a three-step process.

3.1 Data Collection and Corpus Creation

In compiling this data, the course curricula for BLIS, MLIS and other applicable courses were scrutinized through the websites of 104 Indian universities. Thus, 71 course curricula were selected for further study. These course curricula were the unstructured data which, after, became a constructed corpus of comparable course descriptions for the purposes of structured analysis.

3.2 Text Analysis and AI Dictionary

Secondary analysis of the larger text was performed through a custom AI dictionary where the custom dictionary contained 109 AI-related words. Through a rule-based classification system, course syllabi were coded for presence, intensity (i.e., count), and relevance of AI-informed offerings.

3.3 Categorization of AI Subtopics

To operationalize the relative integration of AI in LIS, the AI-related terms were categorized into five subtopics:

- Foundations: Core concepts and principles of AI.
- ML & Analytics: Machine learning algorithms, data analysis, and predictive modelling.
- NLP: Natural Language Processing, including text mining and computational linguistics.
- Library Applications: The application of AI technologies in library services and operations.
- Ethics & Governance: Ethical considerations, fairness, accountability, transparency, and governance of AI.

This categorization allowed for a nuanced analysis of the specific areas of AI that are being integrated into LIS curricula.

4 Results

From the quantitative text analysis conducted on 71 syllabi, results suggest a uniformly defined, yet differentially implemented idea of AI across the LIS curriculum in Indian universities. Thus, this section of the study presents findings from the key occurrences, distribution and subject/topic frequency of the AI-related findings within the assessed syllabi.

Overall, AI Integration. Relative to each other, 100% of the 71 syllabi address something AI-related as it's showing up somewhere. Yet AI coverage, based on other syllabi, is more limited. They mentioned AI-related terms a minimum of 2 times per syllabus and a maximum of 451, meaning average coverage was 159.3 and median coverage was 144 (see Table 1). Thus, the relative value between a minimum of 2 and maximum of 451 means Indian LIS programs either have similar coverage and focus on the subject or enjoy different aspects of coverage for varying reasons (see Fig 1).

Table 1. Summary Statistics of AI Mentions

Metric	Value
Total Syllabi Analysed	71
Syllabi with AI Content	71
Percentage with AI Content	100%
Mean AI Mentions per Syllabus	159.28
Median AI Mentions	144
Min AI Mentions	2
Max AI Mentions	451

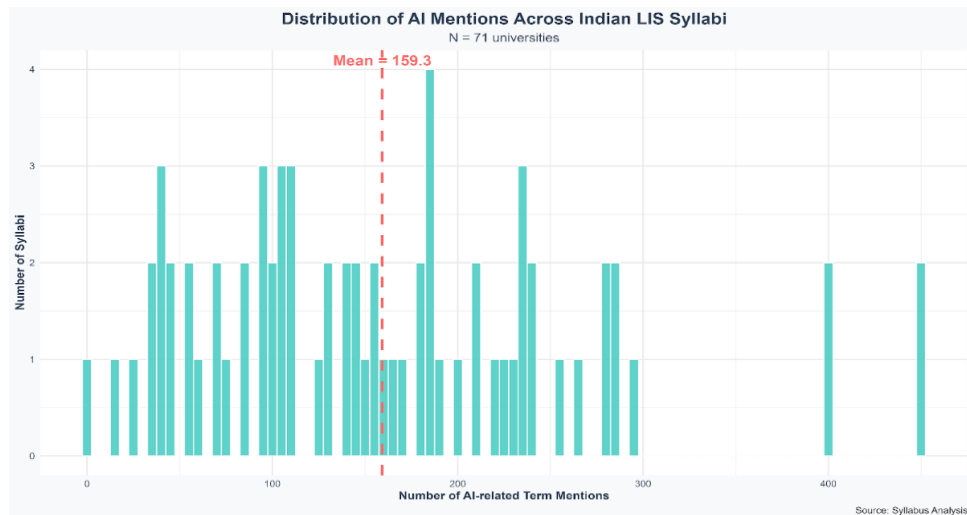


Fig. 1. Distribution of AI Mentions Across Indian LIS Syllabi. [Image failed to load: Distribution of AI Mentions Across Indian LIS Syllabi]

4.1 AI Integration by Category

The analysis of AI integration across the five predefined categories reveals a strong emphasis on the technical aspects of AI, particularly ML & Analytics and Foundations. As shown in Fig 2, 100% of syllabi with AI content included terms related to ML & Analytics, followed closely by Foundations (98.6%) and NLP (97.2%). In contrast, Library Applications (95.8%) and Ethics & Governance (88.7%) were present in a slightly smaller percentage of syllabi.

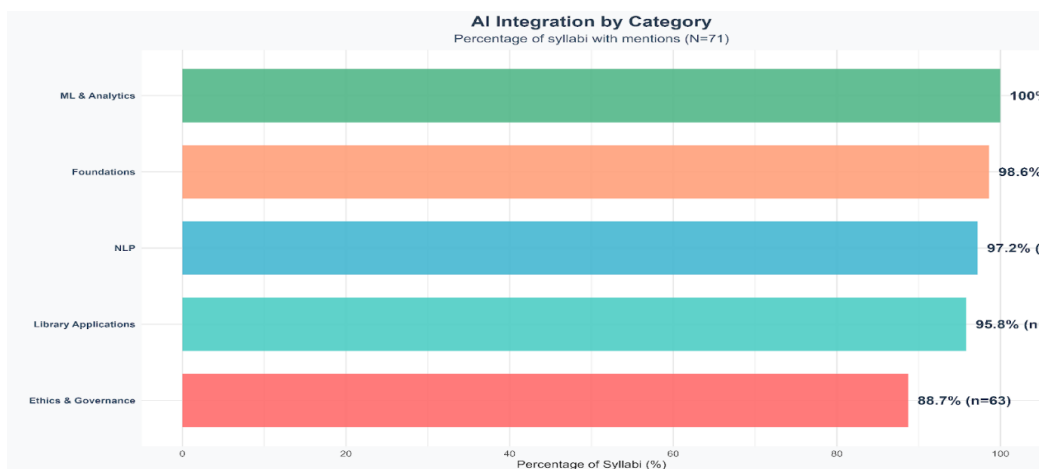


Fig.2. AI Integration by Category.

In terms of the volume of mentions (Fig 3), Foundations was the most prominent category with 4,990 total mentions, followed by ML & Analytics (2,916 mentions) and NLP (1,774 mentions). The categories of Ethics & Governance (848 mentions) and Library Applications (781 mentions) received significantly less attention, suggesting that the practical and ethical dimensions of AI are less emphasized in the current LIS curricula.

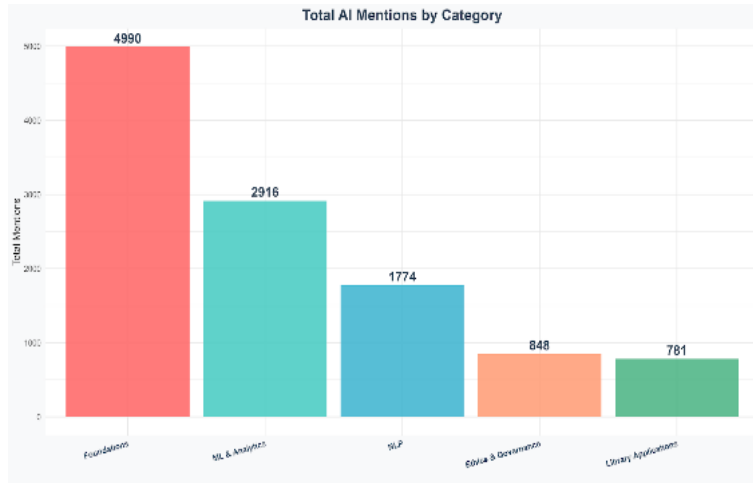


Fig. 3. Total AI Mentions by Category

Table 2. AI Integration Statistics by Category

Category	Syllabi Count	Percentage	Total Mentions	Mean Mentions
Foundations	70	98.6%	4990	70.28
ML & Analytics	71	100%	2916	41.07
NLP	69	97.2%	1774	24.99
Library Applications	68	95.8%	781	11.00
Ethics & Governance	63	88.7%	848	11.94

4.2 University and State-Level Analysis

The analysis identified a concentration of high AI integration in a few universities. Fig 4 presents the top 20 universities by the total number of AI-related term mentions. SNTD Women's University and the University of Burdwan lead the list, each with 451 mentions.

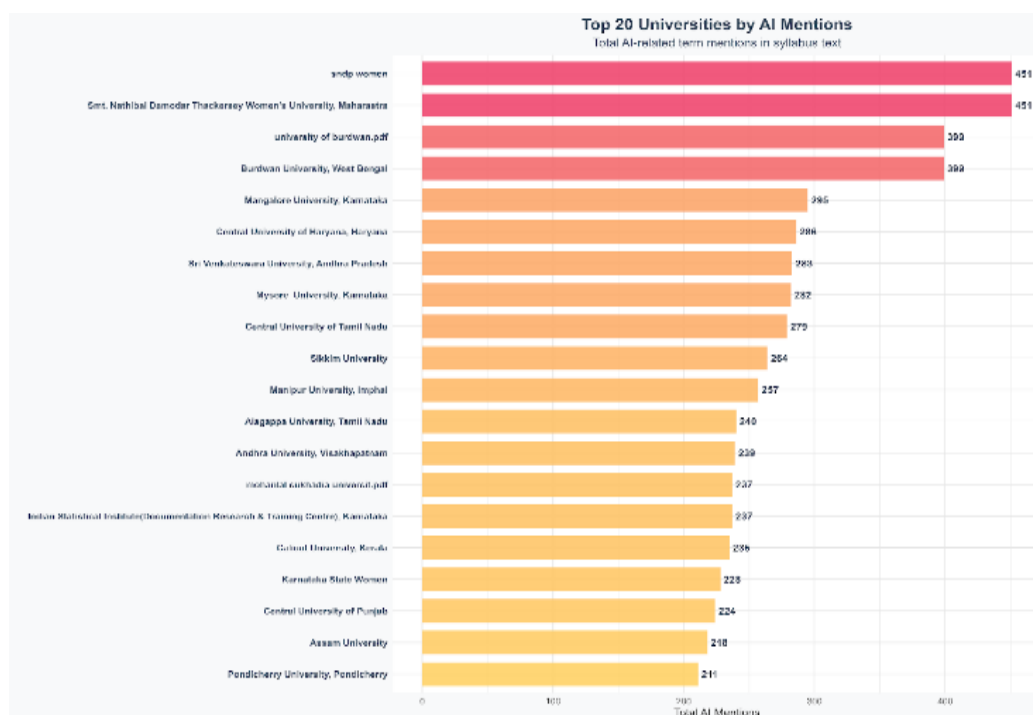


Fig.4. Top 20 Universities by AI Mentions.

The heatmap in Fig 5 provides a more granular view of the top 30 universities, illustrating their relative strengths across the five AI categories. This visualization highlights that universities with high overall AI mentions tend to have a strong focus on Foundations and ML & Analytics.

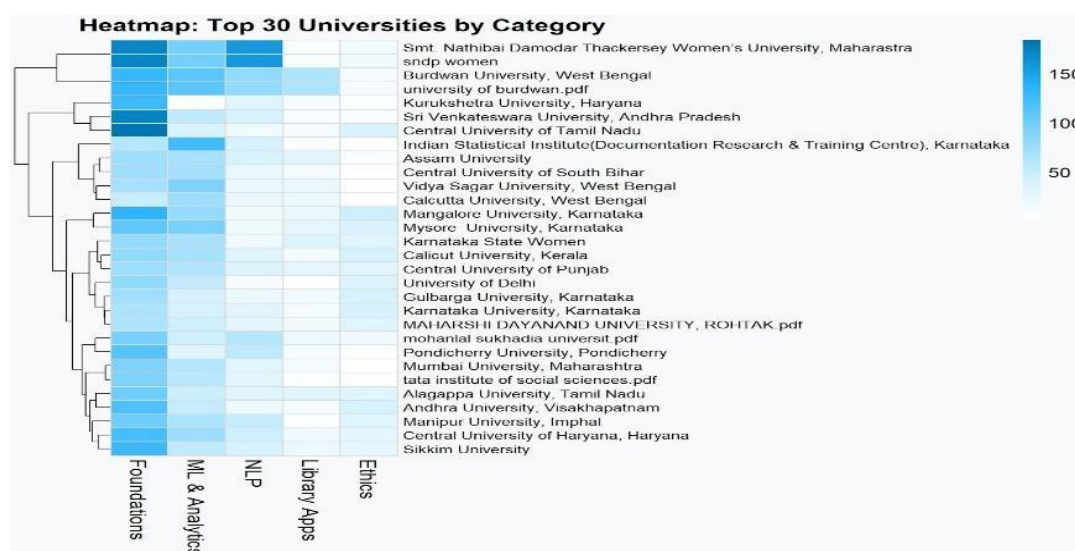


Fig. 5. Heatmap of Top 30 Universities by Category.

A state-wise analysis (Fig 6) shows that states with two or more universities in the dataset all have 100% AI integration in their LIS syllabi, although the depth of integration varies.

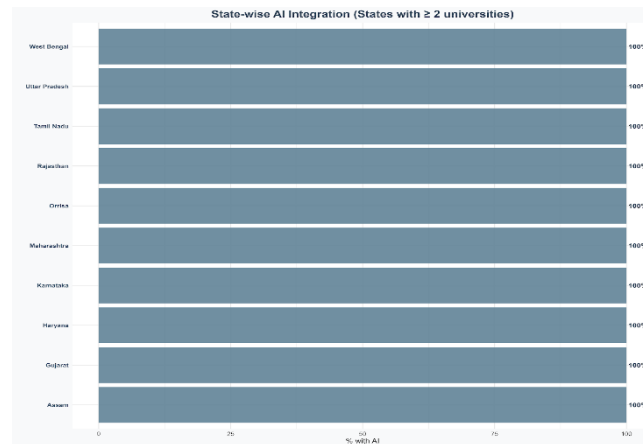


Fig. 6. State-wise AI Integration (States with ≥ 2 universities).

5 Discussion

All in all, these findings provide an overview of the current standing of AI integration across the LIS curricula of Indian universities. The fact that 100% include AI suggests almost universal acknowledgment and appreciation for the positioning of AI as an asset. However, with such drastic differentiation between inclusion levels and topical direction, it begs the question: how adequately prepared are students to graduate into today's AI-driven informational landscape?

Moreover, it's encouraging to know there's an emphasis on Foundational Components and Technical Foundations, as it seems that LIS educators are getting to the crux of the matter with direction most relevant for the aspects that drive the majority of AI creation. However, the relative undervaluation of Library Applications and Ethics & Governance is concerning. Most of the time, those areas that present the practical application of AI in a library setting and those areas that critically assess the perceived values (and costs) are, at best, combined into crucial LIS courses more than less.

Moreover, such findings create an interesting gap with applications. Only four institutions boast better contributor status in relation to these findings, yet at least more than two institutions acknowledge the application of AI, more than others, if not to support this variable, to where it's a good thing to get a head start entering the professional space. Yet simultaneously, it creates a digital divide in which those at one university might be ahead or behind their peers at other institutions when entering the profession. Regardless, with 69% boasting either "Very High" or higher integrated status, it's not an overwhelmingly large amount, but a substantial proportion boasting equity of access to this support variable should be appreciated.

These findings suggest a larger discourse surrounding AI access in LIS programs.[17][11] The fact that a LIS graduate boasts adequate access to AI facilitation in access is necessary for 21st-century student success let alone advanced educated access to ethical concerns surrounding AI.[10] However, the "Low" access afforded to AI's ethical concern is disconcerting—are LIS educators undervaluing the ethical cost of AI, or are they just out of touch with LIS curriculum? Current findings like generative AI and critical literacy are few and far between for value. It seems LIS curriculum has not caught up with what 21st-century librarian prospects should be prepared to face.

6 Conclusion and Future Work

This study is the first of its kind systematic, text-analytic exploration of AI as implemented in LIS curricula across India and reveals an emerging, uneven relative status of development. It's a positive note that AI has been identified across all explored syllabi; however, there lacks detail in extensive responsible governance and practical applications. Furthermore, there is a focus on theoretical construction for the employment of AI more than there is a need for hands-on implementation in the library setting and ethical/governance measures that should dominate such decisions in information professional's endeavours.

Therefore, this content analysis suggests that LIS graduates are currently unprepared to adequately meet the AI-influenced information landscape's demands. Therefore, this author recommends a comprehensive approach to improvement including:

- Foundational Approaches to AI Literacy: AI literacy should be established as a foundational topic for all LIS students. Thus, it should not be treated as its separate topic; instead, it should be embedded into LIS-required coursework. For example, courses on information organization, reference services and collection management could contribute to students' understanding of AI applications from a practical LIS perspective.
- Comprehensive Approach to AI: In addition, as previously noted, LIS programs should accompany an AI literacy component with a comprehensive course focused on AI and data analytics, with practical implementation within libraries and information centers combined with ethical and social considerations.[1]
- Critical Perspectives on AI: Students should be taught to develop humanistic perspectives toward engaging with and developing AI tools—which means curricular ideas that acknowledge algorithmic bias, privacy of information, and the role of the library in teaching its constituents about AI literacy.
- Inter-Departmental Collaboration: LIS departments should be open to collaboration with other departments (computer science, etc.) when possible, to gain practical exposure to interdisciplinary courses that help cement understanding outside of LIS-only concerns.

Future research should include a comparative text analysis based on more Indian institutions and globally to see trends over time for more similar implementations or text-analytic perspectives or generalized qualitative research through interviews with LIS educators and professionals both there and abroad.

Ultimately, AI implementation into LIS curricula should be considered an ethical operating overhaul as an ethical operating one of technological sensibility for the betterment of all. Champions of equity, AI will make the world a more humanized place.

Declaration of AI and AI-Assisted Technologies used in Manuscript Writing Process: Not Applicable

References

1. Borenstein, J., & Howard, A. (2021). Emerging challenges in AI and the need for AI ethics education. *AI and Ethics*, 1(1), 61–65. <https://doi.org/10.1007/s43681-020-00002-7>
2. Chadha, R. K. (2019). *LIS Education in India: Vision 2021*. INFLIBNET Centre, Gandhinagar. <https://ir.inflibnet.ac.in/bitstream/1944/2306/1/27.pdf>
3. Chakraborty, S., & Sarkhel, J. K. (2009). LIS education in India: An appraisal of the parity between the syllabus and the market demands. *World Library and Information Congress: 75th IFLA General Conference and Assembly at Milan, Italy, by International Federation of Library Associations (IFLA), Aug, 23–27*. <https://www.ifla.org/past-wlic/2009/86-chakraborty-en.pdf>
4. Das, S., Tripathi, M., & Das, A. K. (2023). Components of Digital Humanities in Indian LIS Curricula: A Conceptual Analysis. *DESIDOC Journal of Library & Information Technology*, 43(4), 208–218.
5. Eqbal, M., Sohail, N., & Sohail, M. (2010). *A critical evaluation of current status and curriculum development strategy of LIS education in India*. <https://www.semanticscholar.org/paper/A-critical-evaluation-of-current-status-and-of-LIS-Eqbal-Sohail/cb15c54947662414ea5face352782b77219cc23e>
6. ISLAM, S. E. (2023). ICT Practical Components in LIS curriculum: Trend and Advancement Analysis of Selected Indian Universities. *Journal of Indian Library Association Now Available at Hhttps://Journal. Ilaindia. Net/*, 59(4). <https://www.ilaindia.net/jila/index.php/jila/article/view/2060>
7. Kaur, T. (2015). Challenges and Concerns for Library and Information Science (LIS) Education in India and South Asia. *Journal of Education for Library and Information Science*, 56(s1), 6–16. <https://doi.org/10.3138/jelis.56.s1.6>
8. Khanchandani, V. (2019). LIS Education in India: Emerging Trends and Challenges. *Library Herald*, 57(3), 315. <https://doi.org/10.5958/0976-2469.2019.00018.6>
9. Mahapatra, G. (2006). *LIS education in India: Emerging paradigms, challenges and propositions in the digital era*. <https://repository.arizona.edu/handle/10150/106109>
10. Memarian, B., & Doleck, T. (2023). Fairness, Accountability, Transparency, and Ethics (FATE) in Artificial Intelligence (AI) and higher education: A systematic review. *Computers and Education: Artificial Intelligence*, 5, 100152.
11. Ng, D. T. K., Leung, J. K. L., Chu, S. K. W., & Qiao, M. S. (2021). Conceptualizing AI literacy: An exploratory review. *Computers and Education: Artificial Intelligence*, 2, 100041. <https://doi.org/10.1016/j.caeai.2021.100041>
12. Ramasamy, K. (2017). What Slows Down the Wheels of Indian LIS Education and How to Make it Move Faster? The Million Dollar Question. *Science and Humanities*, 253. https://www.gtnartscollege.ac.in/NAAC/dvv-clarification/3.4.3_journal/2017-2018/407.pdf#page=267
13. Sanusi, I. T., Oyelere, S. S., Vartiainen, H., Suhonen, J., & Tukiainen, M. (2023). A systematic review of teaching and learning machine learning in K-12 education. *Education and Information Technologies*, 28(5), 5967–5997. <https://doi.org/10.1007/s10639-022-11416-7>
14. Singh, D. J. (2018). *Library and Information Science Teachers' Perspective on Changes and modifications in the Existing LIS Education and Curricula*.

15. Singh, J., & Shahid, S. M. (2010). Changing needs of library and information science curricula in India. *Library Philosophy and Practice*, 357, 1–8.
16. Singh, K. P., & Shastri, D. K. (2021). Library and Information Science Education in India: Growth, Development, Problems and Prospects. *International Journal of Digital Content Management*, 2(3), 45–61.
17. Southworth, J., Migliaccio, K., Glover, J., Glover, J. N., Reed, D., McCarty, C., ... & Thomas, A. (2023). Developing a model for AI Across the curriculum: Transforming the higher education landscape via innovation in AI literacy. *Computers and Education: Artificial Intelligence*, 4, 100127. <https://doi.org/10.1016/j.caeai.2023.100127>
18. Tamrakar, D. K., Sonkar, S. K., Saraf, S., Kumar, S., Patel, S., & Mishra, R. K. (n.d.). *INTEGRATING AI INTO LIS EDUCATION: MLIS CURRICULUM ANALYSIS OF CENTRAL UNIVERSITIES OF INDIA*. Retrieved December 28, 2025, from https://www.researchgate.net/profile/Dhruv-Tamrakar/publication/397655197_INTEGRATING_AI_INTO_LIS_EDUCATION_MLIS_CURRICULUM_ANALYSIS_OF_CENTRAL_UNIVERSITIES_OF_INDIA/links/6919ed2ed3a2c801362bb09c/INTEGRATING-AI-INTO-LIS-EDUCATION-MLIS-CURRICULUM-ANALYSIS-OF-CENTRAL-UNIVERSITIES-OF-INDIA.pdf
19. Thapa, N. (2021). Building competencies in LIS students—focusing beyond core LIS and ICT skills: What LIS educators have to say. *International Journal of Research in Library Science*, 7(1), 229–242.
20. Yadav, A. K., & Gohain, R. R. (2015). Growth and development of LIS education in India. *SRELS Journal of Information Management*, 52(6), 403–414.

Mapping the Global LIS Landscape: A Cluster-Based Correlation Study of Course Offerings in the USA and India

John D Selwin ¹[0009-0007-9192-4549] and Fazlunnisa H ²[0000-0001-9780-4520]

^{1,2}Department of Library and Information Science, University of Madras
johndselwin98@gmail.com
fazul_8383@yahoo.com

Abstract. The rapid evolution of the digital information landscape has mandated a significant transformation in Library and Information Science (LIS) education, requiring professionals to transition from traditional custodians to facilitators of digital governance and data analytics. This study addresses the global adaptation of these changes through a rigorous cross-national comparative analysis of LIS curricula in the United States and India. Utilizing a purposive sampling approach, data were collected from 50 leading universities (25 per nation) and categorized into ten distinct thematic cluster terms to define the core areas of modern LIS education.

Statistical alignment was evaluated using the Pearson Product-Moment correlation coefficient (r) to measure the linear relationship and volume-based consistency of course offerings between the two nations. Results indicate a strong overall global alignment ($r = 0.758$), suggesting that LIS programs in both countries are modernizing in tandem despite vast geographic and structural differences. Significant linear synchronization was observed in clusters such as "Library Management & Administration" and "Emerging Trends," indicating that as these subjects grow in the USA, they tend to increase in India in a predictable fashion. However, while the USA offers nearly double the total course volume with a focus on elective specialization, Indian curricula maintain deeper technical rigor in foundational knowledge structuring. These findings underscore the necessity of a "glonacal" perspective that balances global technological pressures with local institutional agency. The study advocates for a standardized global competency framework to ensure uniform digital fluency across the evolving information ecosystem.

Keywords: LIS Education; Library Professional; Curriculum; India; United States of America; Pearson Product-Moment; Correlation.

1 Introduction

The landscape of information access and dissemination has undergone a profound transformation, driven by digital innovation and the burgeoning role of artificial intelligence. Consequently, Library and Information Science (LIS) professionals have transitioned from traditional custodians of physical collections to dynamic facilitators of digital literacy, metadata management, and ethical information stewardship [11,12]. As the information environment becomes increasingly complex, LIS education must evolve beyond conventional boundaries to embrace interdisciplinary competencies in data analytics and digital governance [9]. Despite the global nature of these technological shifts, the degree to which LIS curricula have adapted varies across different geographic and academic contexts. This study addresses this gap by conducting a rigorous cross-national comparative analysis of LIS programs. The research focuses on 25 selected universities from the USA and 25 from India, representing a diverse spectrum of LIS education in two of the world's most significant academic landscapes.

To facilitate a systematic comparison, the myriad of course offerings from these 50 institutions were categorized into ten distinct cluster terms, defining the core thematic areas of the modern LIS curriculum. By employing a correlation-based statistical method, this paper investigates the alignment between the USA and Indian educational models. The primary objective is to evaluate how these programs prepare graduates for a digitally-native ecosystem, identifying regional strengths and curricular gaps to propose actionable reforms for the global LIS community.

2 Literature Review

Library and Information Science (LIS) education is undergoing a transformative shift, driven by rapid advancements in technology, growing interdisciplinarity, and evolving user needs. A detailed review of the literature highlights several critical developments and persistent challenges in the design and delivery of LIS curricula globally, with a particular focus on India.

Authors [3] conducted a comprehensive analysis of the statistical reports from the Association for Library and Information Science Education (ALISE) spanning 1997 to 2020. Their study revealed that LIS programs in North America have increasingly diversified, now offering specialized concentrations and certificate options. Emerging courses include data science, digital humanities, and health informatics, reflecting the changing demands of the information profession. While core requirements have been reduced, foundational subjects such as knowledge organization, research methods, and fundamental LIS concepts remain central. This trend is evident across both iSchools and non-iSchools, indicating a growing commitment to interdisciplinarity and readiness for evolving information landscapes.

In contrast, Authors [5] examined the landscape of LIS education in India, uncovering systemic issues such as inconsistency in course structures, inadequate faculty and infrastructure, and a widening gap between the number of trained LIS professionals and industry demands. The transition from traditional to digital libraries necessitates a reorientation of LIS education towards IT competencies. Despite discussions around integrating new technologies and multidisciplinary approaches, India's curriculum reform remains stagnant, with the last major UGC committee meeting on LIS education held in 2001. This disconnect underscores the urgency for curriculum modernization. The importance of social and cultural dimensions in LIS education is underscored by later research [6], which analysed 84 syllabi from 19 ALA-accredited MLIS programs in the U.S. and Canada. Their study found that while diversity-related topics, especially concerning age groups such as children and young adults, were moderately represented, critical areas such as gender, disability, and ethnicity were underexplored. They advocate for continual updates to LIS curricula to foster inclusive and socially responsive information services, ensuring graduates are equipped to serve complex and diverse user communities.

The authors provide a foundational comparison between Pearson's Product-Moment (r) and Spearman's Rank (ρ). [7] Their work is pivotal for research involving "Global Course Distribution," as academic data often contains outliers (e.g., a few "super-hub" universities). They demonstrate that while both coefficients measure association, Spearman is significantly more robust when the data do not meet the assumption of normality.

Authors [10] serves as a vital guide for interpreting these coefficients by providing a standardized framework for classifying relationship strength, which is essential for justifying the "spread" of courses across borders. By moving beyond subjective interpretation, this framework ensures that international education trends are analysed through a lens of statistical reliability rather than mere observational patterns. Applying these thresholds enables a robust comparison between disparate geographical markets, clarifying whether the distribution of curricula is driven by strong systemic links or incidental correlations. Ultimately, [10] criteria act as a methodological safeguard, ensuring that institutional strategies for international growth are anchored in empirical evidence rather than overstated results.

Authors [2] offer a model for curriculum standardization, drawing on syllabi from 39 Korean LIS departments and 28 international universities. They developed standard course content for six core subjects: Introduction to LIS, Information Organization, Information Services, Library Management, Information Retrieval, and Field Studies. These proposed syllabi, validated by academic and professional experts, integrate domestic and global perspectives as well as job market analysis. The study serves as a blueprint for harmonizing core course delivery to ensure consistent training of LIS professionals across institutions.

Authors [8] introduce the "glonacal agency heuristic" as a multidimensional framework to analyse the intersecting forces of globalization, national policy, and local institutional action. They argue that traditional research often suffers from "methodological nationalism," failing to account for how universities function simultaneously across global, national, and local planes. By conceptualizing the university as a "glonacal" actor, the authors highlight that the "spread" of academic courses across borders is not merely a top-down global imposition but a result of complex agency where local institutions and national states actively negotiate global market pressures. This heuristic is particularly relevant for justifying international expansion, as it suggests that the distribution of curricula is a strategic response to a multi-scalar environment where institutions exercise agency to maintain relevance in a globally competitive landscape.

Authors [6] offer a critical "reality check" for comparative higher education research, warning against "methodological nationalism" and the oversimplification of complex national datasets. He argues that the "daily life" of research is often compromised by inconsistent secondary data, such as global rankings, which frequently lack the contextual nuance required for high-level comparison. This perspective directly supports the application of robust statistical tools like Spearman's Rank Correlation to mitigate the impact of "noisy," non-parametric data found in international contexts. By integrating Teichler's framework, this study acknowledges that while numerical coefficients identify trends in global course distribution, these correlations must be interpreted within the specific cultural and structural constraints of the "glonacal" landscape, balancing quantitative "spread" with institutional agency.

A deeper look into curriculum transformation in the digital era is offered by [7], who identified emerging paradigms and challenges for LIS education in India. He stresses that the digital revolution demands new competencies among LIS professionals, particularly in ICT, digital content management, and online service delivery. However, this shift is hindered by resistance to change, a lack of IT-trained faculty, and poor infrastructure. Author

[7] proposes reforms such as enhanced research orientation, integration of e-learning, and collaboration with industry to modernize LIS education and make it more responsive to technological shifts.

This literature review reveals that Library and Information Science (LIS) education is evolving to meet technological demands by integrating data science and digital humanities with traditional core subjects. While regional challenges such as outdated curricula and resource gaps remain, a standardized correlation framework provides a rigorous justification for how these courses spread globally. By balancing global market pressures with local institutional agency, universities can ensure that curriculum expansion is based on statistical reliability rather than mere observation. Ultimately, internationalizing the curriculum transforms growth into a purposeful response to the global knowledge economy, creating a future-ready, interdisciplinary LIS model.

3 Objectives

The current study attempts to:

- Categorise the courses offered by select universities from India and the USA under ten cluster terms
- Quantify and compare the course distribution of the select universities.
- Identify and assess Significant, Moderate, and Low synchronization levels across LIS course clusters.

4 Methodology

Data for this study were extracted from university websites, course catalogues, program brochures, and archived syllabi, then organized into a structured MS-Excel database capturing course titles, classifications (core vs. elective), credit structures, and thematic focuses. A purposive sampling approach was employed to select 50 universities split equally between the USA and India based on their academic reputation, accreditation by recognized bodies (ALA for the USA; UGC for India), the availability of comprehensive public documentation, and a diverse range of program orientations.

Table 1. List of selected Universities offering LIS Programme - India

Sl. No.	University / Institution	Year of Establishment (DLIS)
1	Banaras Hindu University (BHU)	1941
2	University of Mumbai	1943
3	University of Delhi	1946
4	University of Madras	1960
5	DRTC (Bangalore)	1962
6	University of Mysore	1965
7	Andhra University	1967
8	Dr. Babasaheb Ambedkar Marathwada	1967
9	Kurukshetra University (KUK)	1969
10	University of Kalyani	1970
11	Dr. Harisingh Gour Vishwavidyalaya	1971
12	Samrat Vikramaditya Vishwavidyalaya	1971
13	University of Jammu	1971
14	Bangalore University	1973
15	Annamalai University	1979
16	North Eastern Hill University (NEHU)	1979
17	Utkal University	1981
18	Bharathidasan University	1985
19	University of Calicut	1985
20	Devi Ahilya Vishwavidyalaya (DAVV)	1993
21	Assam University	1998
22	Mizoram University	2002
23	Alagappa University	2006
24	Pondicherry University	2007
25	Bharathiar University	2008

Table 2. List of selected Universities offering LIS Programme - USA

Sl. No.	University / Institution	Year of Establishment
1	University of Illinois at Urbana-Champaign	1893
2	Syracuse University	1896
3	Emporia State University (Kansas)	1902
4	University of Washington	1911
5	Dominican University	1930
6	University of North Carolina at Chapel Hill	1931
7	University of Kentucky	1932
8	Louisiana State University	1932
9	St. John's University	1937
10	The Catholic University of America	1939
11	Southern Connecticut State University	1946
12	Florida State University	1947
13	University of Texas at Austin	1948
14	Indiana University Bloomington	1951
15	University of California, Los Angeles (UCLA)	1958
16	Long Island University (LIU)	1959
17	University of Rhode Island	1963
18	University of Maryland	1965
19	University at Buffalo	1966
20	University of Iowa	1967
21	University of Arizona	1968
22	Chicago State University	1968
23	University of Alabama	1970
24	University of South Florida	1970
25	The University of Southern Mississippi	1976

To facilitate a standardized comparative analysis, courses were categorized into ten thematic clusters, and the number of courses offered in that cluster is mentioned in Table 3.

This framework was developed through a hybrid process: a deductive stage where foundational categories were derived from ALA Core Competencies and UGC Model Curricula, followed by an inductive stage where course descriptions were reviewed to integrate emerging themes such as Artificial Intelligence, Data Science, and Digital Curation into the cluster definitions.

Table 3. Quantified list of identified course clusters

Cluster Code	Cluster Name	Number of Courses
ICT	Information Communication Technology (ICT)	262
KOM	Knowledge Organization and Management (KOM)	244
IPS	Information Products & Services (IPS)	228
LMA	Library Management & Administration (LMA)	214
ILS	Information & Library Society (ILS)	190
INS	Information Systems (INS)	156
ETD	Emerging Trends (ETD)	118
IRS	Information Retrieval System (IRS)	98
RMS	Research Methodology & Statistics (RMS)	88
LLN	Library Legislation (LLN)	40

To determine the overall degree of alignment between the USA and Indian LIS educational models, a statistical measure was applied to the aggregated cluster frequency:

- Pearson Product-Moment Correlation (r): Used to measure the linear relationship between the volume of course offerings in each cluster across both countries. It calculates how much the variables change together relative to how much they change separately.

5 Results and Discussion

5.1 Comparative Analysis of LIS Course Clusters

The distribution of Library and Information Science (LIS) courses across the ten identified clusters in the USA and India reveals significant parallels in pedagogical priorities, alongside distinct regional specializations. Table 1 summarizes the frequency (N) and relative percentage distribution of courses for both cohorts.

Table 4. Distribution and Comparison of LIS Course Clusters

Cluster Term	USA (N)	USA (%)	India (N)	India (%)
Information & Library Society (ILS)	146	13.53	44	7.87
Knowledge Organization (KOM)	132	12.23	112	20.04
Information Products & Services (IPS)	157	14.55	71	12.70
Information Retrieval (IRS)	60	5.56	38	6.80
Library Management (LMA)	146	13.53	68	12.16
Information Communication & Technology (ICT)	166	15.38	96	17.17
Research Methodology (RMS)	52	4.82	36	6.44
Information System (INS)	95	8.80	61	10.91
Emerging Trends (ETD)	91	8.43	27	4.83
Legal Legislation (LLN)	34	3.15	6	1.07
Total	1079	100.00	559	100.00

This table provides a comparative breakdown of the LIS curriculum across 50 universities (25 USA, 25 India), categorized into ten thematic clusters. The values represent the raw number of courses (N) and their relative weight as a percentage (%) of the total curriculum in each country.

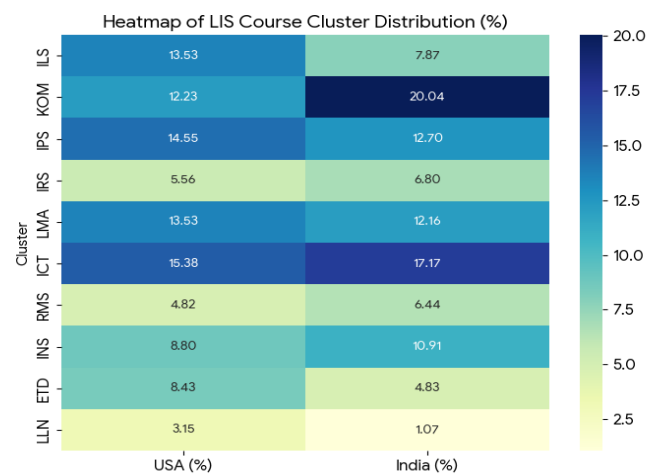


Fig. 1. Heatmap of LIS Course Cluster

The total number of courses offered in the 25 USA universities (1079) is nearly double that of the 25 Indian universities (559). This suggests that USA programs typically offer a wider variety of elective specializations per degree, whereas Indian programs tend to follow a more streamlined or standardized credit structure.

Figure 1 visually captures the thematic landscape of LIS education, highlighting both shared global standards and distinct regional specializations through a series of "hotspots" and "cold zones." At the forefront of the Indian curriculum is a significant concentration in Knowledge Organization and Management (KOM), which at 20.04% represents the study's most intense hotspot and underscores a deep-rooted specialization in foundational technical skills like cataloguing and classification. In contrast, the USA displays a more diversified distribution across several clusters, including Information & Library Society (ILS), Information Products & Services (IPS), and Library Management & Administration (LMA), reflecting a broader, multidisciplinary approach to LIS education. Despite these regional variations, Information Communication Technology (ICT) emerges as a universal high-priority

zone for both nations, statistically identifying it as the dominant contemporary pillar of the field globally. Conversely, Library Legislation (LLN) consistently appears as a "cold zone" in both regions, indicating its status as a necessary but minor niche within the academic curriculum. Furthermore, a visible intensity gap in Emerging Trends (ETD) reveals that modern competencies such as AI and data curation are currently more integrated into Western programs, suggesting that while both countries are technology-driven, the USA leads in the institutionalization of nascent informatics-based subjects.

Table 5. Correlation between the course cluster with respect to USA and India

Sl. No.	Cluster Term	Pearson (<i>r</i>)	Interpretation
1	Information & Library Society (ILS)	−0.109	No Real Correlation
2	Knowledge Organization And Management (KOM)	−0.137	No Real Correlation
3	Information Products & Services (IPS)	0.265	Slight Correlation
4	Information Retrieval System (IRS)	−0.048	No Correlation
5	Library Management & Administration (LMA)	0.401	Noticeable Correlation
6	Information Communication Technology (ICT)	−0.038	No Correlation
7	Research Methodology & Statistics (RMS)	0.083	No Real Correlation
8	Information System (INS)	0.083	No Real Correlation
9	Emerging Trends (ETD)	0.486	Noticeable Correlation
10	Library Legislation (LLN)	0.323	Slight Correlation

The interpretation of the correlation is done based on the scale given below:

- 0.00 – 0.19: No Real Correlation (No significant relationship).
- 0.20 – 0.39: Slight Correlation (Negligible pattern observed but not reliable).
- 0.40 – 0.59: Noticeable Correlation (Emergence of reliable pattern).
- 0.60 – 1.00: Strong Correlation (Significant relationship).

The correlation analysis reveals a strong positive relationship between the distribution of LIS course clusters in the USA and India.

- Pearson Correlation Coefficient (*r*): 0.758, This indicates a high degree of linear correlation. It suggests that as the number of courses in a specific cluster, increases in the USA, it tends to increase in India as well. The two countries generally emphasize similar core areas of LIS education.

The individual cluster study reveals a more nuanced picture than the aggregate country-level correlation. While the overall country-level correlation is high, this table reveals that at the institutional level, the two nations demonstrate significant alignment in some areas while remaining structurally independent in others.

5.2 Significant Synchronization

- Emerging Trends (ETD): Shows the highest level of synchronization with a moderate linear correlation of 0.486, suggesting Indian programs are mirroring the growth of modern subjects like AI and Data Curation found in the USA.
- Library Management & Administration (LMA): Demonstrates consistent alignment (0.401) as a standardized component of LIS education in both regions.
- Library Legislation (LLN): Despite lower raw numbers, this cluster shows moderate synchronization (0.323), indicating a shared level of importance in the curriculum hierarchy.

5.3 Less Synchronization

- Information Products & Services (IPS): Exhibits only a low positive correlation (0.265), indicating some shared emphasis but significant local variation.
- Research Methodology & Statistics (RMS) & Information Systems (INS): Both show weak synchronization (0.083), suggesting that these foundational areas are structured according to specific institutional needs rather than a unified global model.

5.4 Ranking Consistency

- While the aggregate data suggests high overall alignment ($r = 0.758$), specific high-frequency clusters like Information Communication Technology (ICT) and Knowledge Organization and Management (KOM) show a surprising lack of consistency at the institutional level.
- Although both countries prioritize these as vital core subjects, they do not follow a predictable linear ranking pattern across individual universities.

6 Conclusion

This study underscores a significant global alignment in Library and Information Science (LIS) education, as evidenced by the strong positive correlation ($r = 0.758$) between the curricula of leading universities in the USA and India. While the USA offers a more diversified and high-volume course structure, both nations have successfully institutionalized Information Communication Technology (ICT) as a core pillar. The high degree of synchronization in Library Management & Administration (LMA) and Emerging Trends (ETD) suggests that LIS programs are modernizing in tandem, transcending geographic and economic boundaries to meet the demands of the digital era. However, distinct regional priorities persist, with Indian curricula remaining more deeply specialized in Knowledge Organization and Management (KOM), whereas USA programs lean toward user-centric Information Literacy (ILS) and informatics-based specializations.

Moving forward, LIS educators in both regions should strive for a "hybrid" model that balances these strengths. Indian programs could benefit from integrating more multidisciplinary "Emerging Trends" and user-outreach modules to match the agility of the Western model. Conversely, USA programs can look to the Indian framework to reinforce the technical rigor of knowledge structuring. Future research should focus on the development of a standardized global competency framework, ensuring that LIS graduates, whether in a developed or developing economy, possess a uniform baseline of digital fluency, ethical stewardship, and management expertise to lead in the evolving global information ecosystem.

Declaration of AI and AI-Assisted Technologies used in Manuscript Writing Process: Not Applicable

References

1. Alajmi, B. M., & Alshammari, I. (2020). Strands of diversity in library and information science graduate curricula. *Malaysian Journal of Library and Information Science*, 25(1), 103–120. <https://doi.org/10.22452/mjlis.vol25no1.6>
2. Choi, Sang-Ki, Ahn, In-Ja, Noh, Younghee, & Kim, J.-S. (2013). A Comparative Study on Curriculum Contents and Teaching Methods Based on the Syllabi of Library and Information Science in Korea and Foreign Universities. *Journal of the Korean Society for Library and Information Science*. <https://doi.org/10.4275/KSLIS.2013.47.2.223>
3. Chung, E., Schalk, J., & Schalk, J. (2024). How have LIS school curricula evolved over the past twenty years? *The Canadian Journal of Information and Library Science*, 45(1). <https://doi.org/https://doi.org/10.5206/cjilsrscib.v45i1.14192>
4. Hauke, J., & Kossowski, T. (2011). Comparison of values of pearson's and spearman's correlation coefficients on the same sets of data. *Quaestiones Geographicae*, 30(2), 87–93. <https://doi.org/10.2478/v10117-011-0021-1>
5. Khanchandani, V. (2021). LIS Education in India: Emerging Trends and Challenges LIS Education in India: Emerging Trends and Challenges. *Library Herald, January 2019*. <https://doi.org/10.5958/0976-2469.2019.00018.6>
6. Kosmützky, A., & Nokkala, T. (2014). Challenges and trends in comparative higher education: An editorial. *Higher Education*, 67(4), 369–380. <https://doi.org/10.1007/s10734-013-9693-x>
7. Mahapatra, G. (2006). LIS education in India: Emerging paradigms, challenges and propositions in the digital era. *Singapore Journal of Library & Information Management*, 35, 26–33. <http://arizona.openrepository.com/arizona/handle/10150/106109>
8. Marginson, S., & Rhoades, G. (2002). Beyond national states, markets, and systems of higher education: A glonacal agency heuristic. *Higher Education*, 43(3), 281–309. <https://doi.org/10.1023/A:1014699605875>
9. Markey, K. (2004). Current Educational Trends in the Information and Library Science Curriculum. *Journal of Education for Library and Information Science*, 45(4), 317. <https://doi.org/10.2307/40323877>
10. Mukaka, M. M. (2012). Statistics corner: A guide to appropriate use of correlation coefficient in medical research. *Malawi Medical Journal*, 24(3), 69–71.
11. Pawley, C. (2019). History in the library and information science curriculum: Outline of a debate. *Information and Culture*, 54(1), 223–238.
12. Raju, J. (2020). Future LIS Education and Evolving Global Competency Requirements for the Digital Information Environment: An Epistemological Overview. *Journal of Education for Library and Information Science*, 61(3), 342–356. <https://doi.org/10.3138/JELIS.61.3.2019-0088>

User Experience and Access

Evidence based Evaluation of Chatbot Services in Academic Libraries

Varalakshmi V ¹[0009-0004-0620-8707] and Fazlunnisa H ²[0000-0001-9780-4520]

^{1,2} University of Madras, Chennai, Tamil Nadu

varshasuriya.vs@gmail.com

fazul_8383@yahoo.com

Abstract. Academic libraries have always been the beacons of technology adoption to serve the user community more efficiently; AI tools are not stranger things to LIS professionals. The trick lies in being able to identify, evaluate and adopt the apt tool to suit the needs of the academic users – especially to meet the expectations of 24/7 digital-native students. The present study evaluates one such AI-based tool – the library chatbot, based on meticulously developed criteria ranging across 8 categories with 15 queries each totaling to 120 queries. This evaluation led to a revelation among the 6 chosen chatbots that are active in the current academic scenario which include the *Aisha*, *Scottai Library chatbot*, *Queensland Library Chatbot*, *T-Rex*, *Lehman Lightning Bot* and *Seri Library digital assistant*. This study evaluates the chatbots based on the developed criteria and highlights the pros and cons of the existing chatbots. The 120 queries spread across 8 categories will help to effectively evaluate any AI-based chatbot before its implementation into an academic library website.

Keywords: Chatbot in libraries; Artificial Intelligence; AI chatbot; Comparison of Chatbots; Virtual reference assistant; Library Chatbot

1 Introduction

Library and Information Centres have never shied away from embracing new technologies to enhance efficient user services to the user community. Academic libraries are the foremost sectors to be influenced by the progressive technological flux, transforming the relationships between users, information professionals and the resources. Ever since Tim Berners-Lee introduced emails for communication which catapulted the dot-com era, prompted librarians to explore the concept of personalised services. Propelled by waves of digital intercession, services such as ‘Ask the Librarian’ and other similar virtual synchronous chat and messaging platforms featured as we stepped into the 21st century. This opened the doors of the library to offer real-time user assistance, beyond the confines of the library walls to cater to the needs of the digital natives who care less about face-to-face interactions and also demand 24/7 access to information.

Academic libraries face a huge challenge in reaching out to the student-user community consisting of millennials and GenZ, who are more connected on social media platforms such as X, Facebook, Instagram and many others and would rather not set foot inside the library. To mitigate this issue and to bond with the young and dynamic users, AI integrated products and services have been introduced at libraries. We are witnessing transformation from screen-based virtual assistants to humanoid robots and ‘think and respond’ chatbots. This has been documented by Takashi [19], highlighting the installation of robots like Pepper, Koro, Sota, An-San, Uta-San, HAL, and Thouser at libraries in Japan. These were trained to provide services which include – responsive answers, collection search, reservation management, gesture-based interaction, and even book transport assistance. These AI-based services exemplify the seamless integration of AI and replacement of manual labour in repetitive and routine operations. Conversational interfaces such as chatbots have emerged as pivotal AI applications in academic library setup, especially connecting the user with the right and authentic documents. Studies indicate such implementations as Libby at the University of Pretoria, Beepbot at Guelph Public Library, Pepper at Roanoke County Library, Cobot at Chulalongkorn University, Bizzy at the University of Oklahoma, Kingbot at San Jose State University Library, Lehman Lightning Bot at Leonard Lief Library, and T-rex at the University of Calgary Library, which demonstrate the adoption and implementation of AI-based chatbots in institutional setups [1]. Further, Fatouh [3], has found that chatbots are capable of streamlining information retrieval, automating reference queries, and customising user needs, serving as an adjunct to reference librarians in handling the sheer volume of routine queries. We understand the chatbots are not mere technological novelties but a mode of connecting and conversing with the ever-demanding needs of the academic community.

The father of Library Science in India, Dr. S.R. Ranganathan, always focused on providing timely and authentic services to the user community, achieved through the reference desk, once the cornerstone of user-supportive services. It's time to implement and empower users with authoritative information services/resources using AI-

powered chatbots that are sufficiently integrated with library resources. This will enhance user conversations powered by large language models (LLMs), Machine Learning algorithms, and, most importantly, natural language search using NLP (Natural Language Processing).

Numerous studies indicate that libraries are evolving technologically, using appropriate ICT tools to enhance user services, ranging from M-OPAC and QR codes to virtual reference services [12]. In the contemporary landscape, post-millennial users increasingly rely on AI-driven tools such as ChatGPT for academic tasks, including idea generation, summarisation, and assignment support, thereby expecting seamless, round-the-clock assistance [13]. Chatbots, as intelligent conversational agents, are uniquely positioned to meet these expectations by automating routine queries and providing personalised support. Kuhail and colleagues [10] chart early implementations of chatbots in academic libraries, revealing the diversity of operational platforms from web, mobile, to desktop interfaces. While these deployments demonstrate the potential of chatbots to supplement traditional reference services, they also highlight the need for systematic evaluation to ensure alignment with user demands and institutional objectives.

It is practically impossible to assess chatbot effectiveness through a single evaluative lens. Maroengsit and colleagues [14] emphasise that the practical assessment of conversational agents requires multiple metrics tailored to specific chatbot types and purposes, as performance dimensions vary across disciplines and contexts. Indicators developed in prior literature, such as usability, accuracy, responsiveness, personalisation, and integration with library systems, provide a robust framework for evaluating chatbot functionality. [9, 17, 18]

Therefore, a comprehensive study is warranted to assess existing chatbots in academic libraries against these literature-derived indicators. Such an evaluation will not only identify strengths and limitations but also guide the selection of appropriate chatbot models for future development. Libraries need to ensure that implementing chatbots goes beyond novelty and evolves into a sustainable, user-centred service paradigm. This emphasises the need for a closer examination of the existing literature to identify the evaluation metrics discussed in the subsequent section.

2 Literature review

Chatbot integration symbolises the transfiguration of academic libraries, extending services beyond physical walls and revamping traditional reference practices to dialogic, user-centred conversational interfaces. The trajectory of Chatbots in academic libraries has been noted in several studies [2, 5, 9]. They emphasise the anchoring significance of chatbots, which can automate repetitive tasks, enhance user engagement, and provide 24/7 reference services [1]. These implemented chatbots need to be assessed to provide effective services to the user community. The assessment of chatbots has been evaluated using various metrics, including quantitative, qualitative, domain-specific, and comparative approaches. Holmes and colleagues [7] conducted usability testing of a healthcare chatbot using metrics such as the System Usability Scale, User Experience Questionnaire, and Chatbot Usability Questionnaire. Similarly, Gunnam and colleagues [4] assessed the performance of cloud-based Heterogeneous chatbot systems using quantitative metrics, including response time, throughput, connection loss, and load testing. They compared it with that of on-premises server-based chatbots and cloud-based chatbots. Extending this line of evaluation metrics is the study by Kabakuş & Dogru [8], which compared the large language model chatbots such as ChatGPT and Google's Bard (formerly known as Google Gemini). Alternative evaluation metrics, such as content evaluation (assessing response context quality), user satisfaction evaluation (measuring human user's satisfaction), and functional assessments, are categorised from a comprehensive survey of chatbot evaluation methods by Maroengsit and colleagues [14]. A comparative analysis of four selected chatbots, including ChatGPT, Perplexity, Chatsonic, and Google Bard, was conducted by Halboub and colleagues [6] to gather information about oral cancer. They used the DISCERN tool and the Patient Education Materials Assessment Tool (PEMAT), an online calculator tool to assess the quality of the chatbot's responses and the understandability and actionability of the answers to the query. A similar assessment is carried out by Ponzo and colleagues [16] who evaluate large language model chatbots by measuring their accuracy, completeness, reproducibility, and consistency. Møller and colleagues [15] designed a new instrument to measure user satisfaction with AI chatbots, particularly in customer support contexts, using a 40-item scale that encompasses eight comprehensively synthesised constructs.

This study is the first of its kind to compare existing library chatbots through a qualitative analysis, posing questions to each chatbot and evaluating their responses accordingly. The chatbots were selected from the literature [11, 2, 1]. The queries to be posed for the chatbot were inspired by the categories and subcategories of services and items requested by chatbots in the analysis by Kane [9] in the University library Chatbot, ANTsvers. Following his study is the work by Sharma and colleagues [17] in which the authors compared the performance of eight chatbots based on the answers they provided to a set of predefined questions. Some evaluation categories were drawn from the study by Singh & Namin [18], which evaluated grammatical correctness, content accuracy, task

completion time, error rates, response matching, and other factors. Only those mentioned were suitable for a qualitative study and were selected for this study.

3 Objective

This paper aims to systematically assess select academic library chatbots against evaluative criteria developed through an extensive literature review, with the goal of identifying a model chatbot to guide future chatbot design and implementation in academic libraries.

4 Methodology

The study employs a five-step sequential process (Fig.1) to evaluate and compare select academic library chatbots as discussed below:

Step 1: A total of 6 active academic library chatbots were selected using a purposive sampling approach based on fulfilment of any one of the following criteria: a. deployment on an academic library website; b. evidence of active use on the respective website; c. documented in peer-reviewed literature. Accordingly, three chatbots (Library chatbot of University of Queensland, Scottai library chatbot of Carnegie Mellon University, Seri, Library digital assistant of Universiti Malaya (UM) were selected through direct observation of publicly accessible academic library websites, while other 3 - Aisha of Zayed University Library, T-Rex of University of Calgary and Lehman Lightning Bot of Lehman college were identified from prior studies [1, 2, 11]. Chatbots used solely for vendor demonstrations, non-academic contexts, or inactive deployments were excluded.

Step 2: Determine and derive 8 evaluative categories and 3 subcategories each to test the functionality of the chatbots, as represented in Fig.1.

Step 3: To effectively evaluate the chatbots, a set of 5 queries, each under 24 sub-categories, is developed, totalling 120 queries (Annexure 1), with a binary scoring system.

Step 4: The chatbots were accessed through the library website. The chatbots were evaluated using a questionnaire with 120 closed-ended questions; responses were scored in an MS Excel worksheet. The responses to the queries posed to the chatbot were evaluated and scored at 1 point for a relevant response and 0 points for an irrelevant/incomplete answer.

Step 5: The select chatbots are compared based on their scores and their performance patterns are analysed. A visual representation of these analysis helps to identify and choose the ideal chatbot model for an academic library setup.

5 Results and Interpretations

The data collected using a structured questionnaire are analysed and visually represented to evaluate the chatbots based on 8 identified categories. The findings are discussed below:

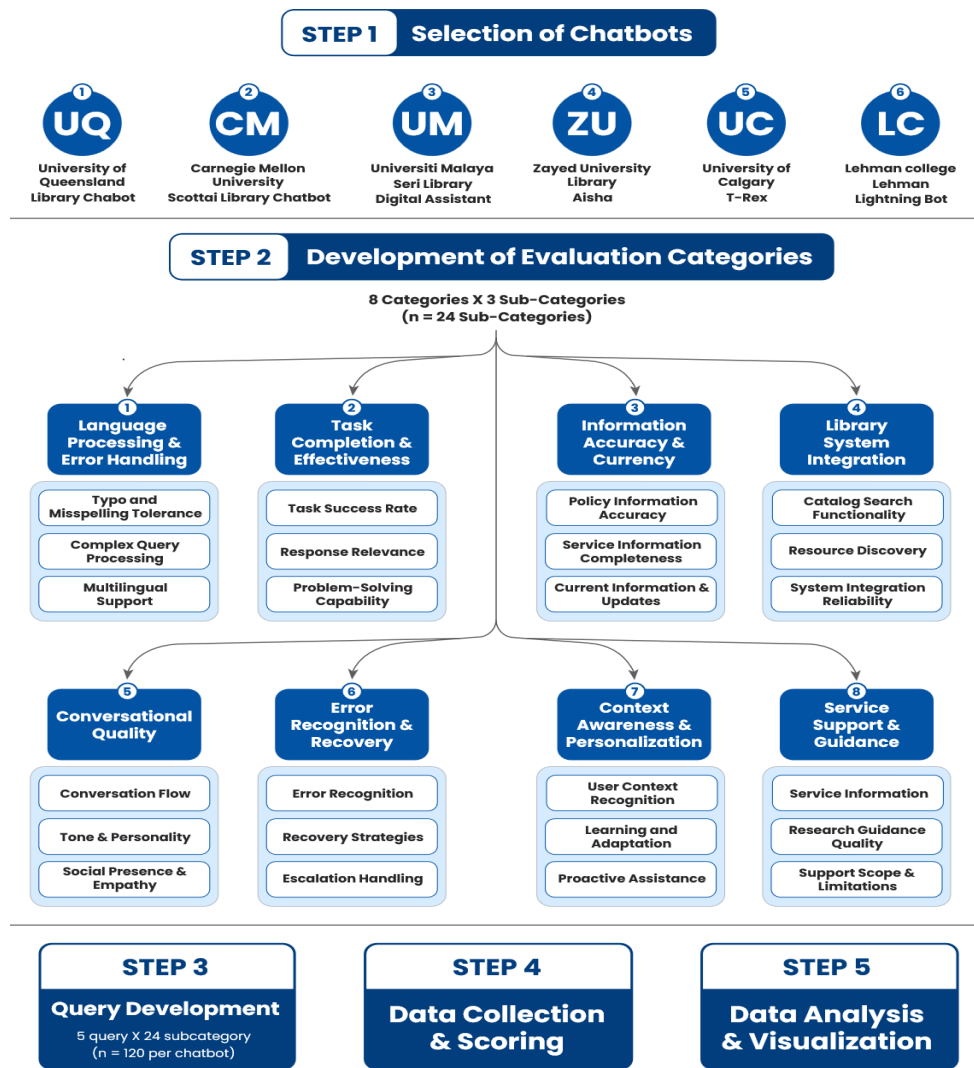


Fig. 1. Methodology

5.1 Language Processing & Error Handling

Evaluation under this category is based on 15 close ended queries developed for the 3 identified sub categories as defined in Table 1.

Table 1. Subcategories for category 1- Language Processing & Error Handling

S.No.	Subcategory	Evaluation Goal
1.1	Typos and Misspelling Tolerance	To assess the chatbot's ability to interpret user-generated natural language queries that may contain typographical errors, complex, and multi-lingual queries.
1.2	Complex Query Processing	
1.3	Multilingual Support	

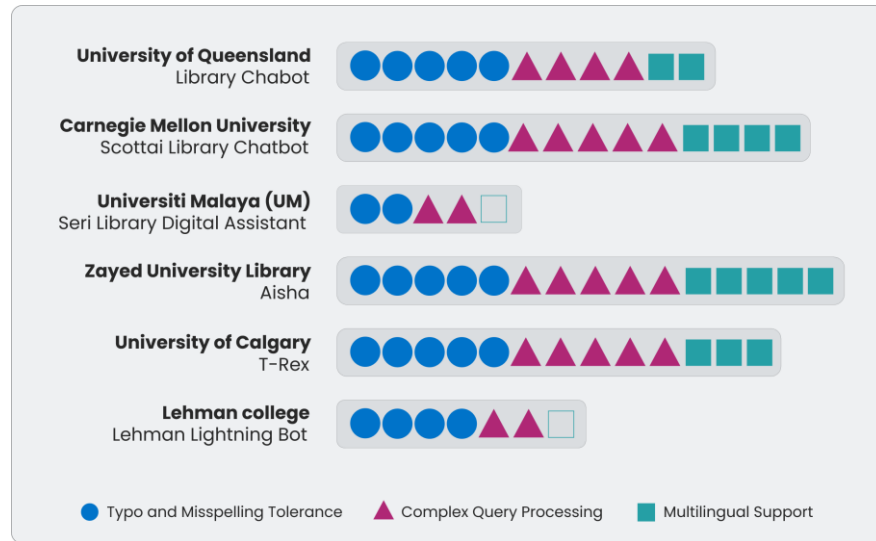


Fig. 2. Language Processing & Error Handling

It is observed from the above Fig.2 that 5 out of 6 chatbots are well equipped to handle typos and misspellings as well as complex queries, the difference though is in the multilingual approach. Aisha, Zayed University Library Chatbot is found to be the only chatbot that supports multiple language user's queries, followed by the Scottai Library Chatbot of the Carnegie Mellon University Library, whereas the other chatbots provide nil or limited language support.

5.2 Task Completion & Effectiveness

Table 2 indicates the evaluation goals under this category:

Table 2. Subcategories for category 2 -Task completion & Effectiveness

S.No.	Subcategory	Evaluation Goal
2.1	Task Success Rate	Examines the chatbot's effectiveness in completing the requested task while providing appropriate output for the pertinent query.
2.2	Response Relevance	
2.3	Problem-Solving Capability	

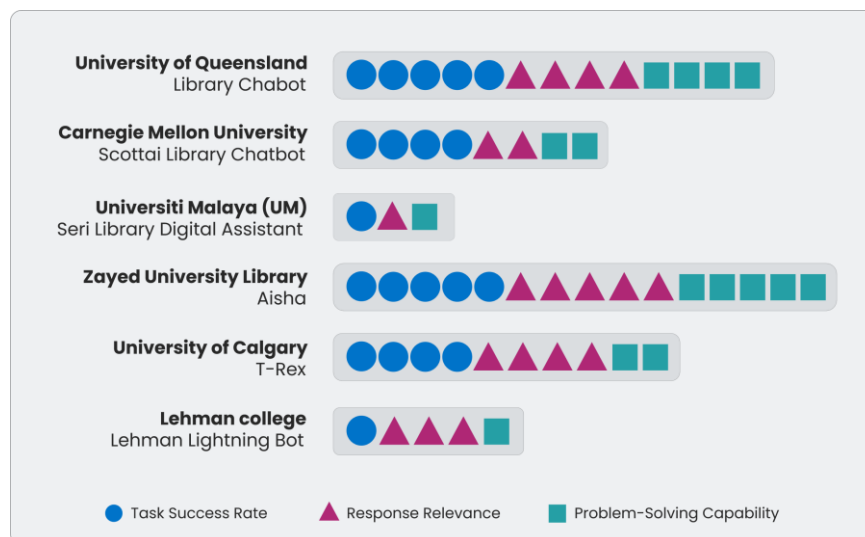


Fig. 3. Task Completion & Effectiveness

As observed from Fig. 3, Aisha and Library Chatbot are found to be the only two academic library chatbots that can handle / complete any task with effectiveness. Coming close are the T-Rex and Scottai Library Chatbot. The Lehman Lightning Bot and Seri Library Digital seem to be underperformers in this category.

5.3 Information Accuracy & Currency

Sub-categories to evaluate the currency and correctness of information are evaluated based on the chatbots' ability to deliver timely and authentic information as discussed in table 3.

Table 3. Subcategories for category 3 - Information Accuracy & Currency

S.No.	Subcategory	Evaluation Goal
3.1	Policy Information Accuracy	Captures the chatbot's standard to provide accurate, up-to-date information and happenings at universities for users, mimicking the digital extension of a reference librarian.
3.2	Service Information Completeness	
3.3	Current Information & Updates	

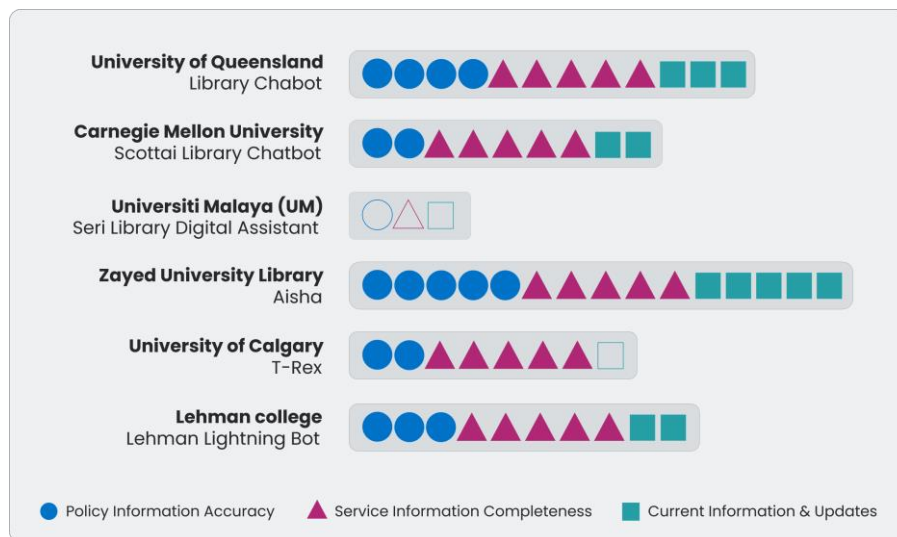


Fig. 4. Information Accuracy & Currency

In stark contrast to Aisha (15/15), the Seri Digital Assistant (0/15) does not offer updated information, indicating its static nature. All other chatbots offer accurate and updated information except T-Rex, which true to its name is extinct when it comes to providing current information and updates.

5.4 Library System Integration

Any assistive technology fails if it is not connected and integrated with its technical source of information, to evaluate the chatbots on its integration with the library software, the following subcategories are employed (Table 4)

Table 4. Subcategories for category 4 - Library System Integration

S.No.	Subcategory	Evaluation Goal
4.1	Catalogue Search Functionality	Assesses the chatbot's extent of integration with the technical core library systems, such as catalogues, databases, and OPAC.
4.2	Resource Discovery	
4.3	System Integration Reliability	

The technical integration of chatbots with their library system, including integration with OPAC, searching resources across different databases, and providing stable responses to user search queries, is assessed through three subcategories: Catalogue Search Functionality, Resource Discovery, and System Integration Reliability. (Fig. 5)

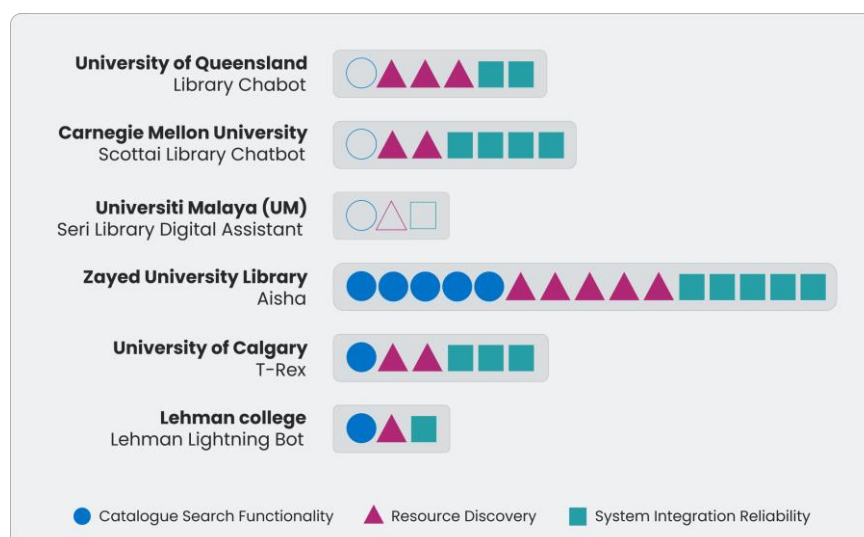


Fig. 5. Library System Integration

As observed previously, Aisha is the only chatbot scoring a perfect 15/15 for its library integration whereas all the other chatbots lack catalogue integration, are deficient with resource search, and indicate minimal system integration reliability.

5.5 Conversational Quality

The chatbot's interaction quality with the user is analysed to determine if it can have a natural conversation flow mimicking human interaction. The chatbot is analysed for its conversation flow, tone, and personality, as well as social presence and empathy with the user. (Table 6)

Table 5. Subcategories for category 5- Conversational Quality

S.No.	Subcategory	Evaluation Goal
5.1	Conversation Flow	Examines whether the chatbot can imitate a human tone, providing the user with an experience as if there is another human on the other end to answer their queries.
5.2	Tone & Personality	
5.3	Social Presence & Empathy	

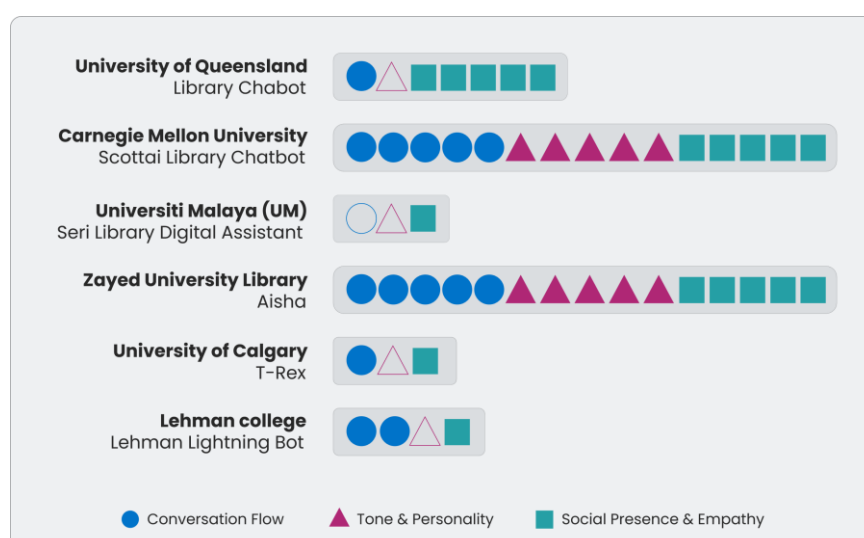


Fig. 6. Conversational Quality

The response analysis reveals 3/6 chatbots to score a perfect 15/15 - Scottai Library Chatbot of Carnegie Mellon University Library, and Aisha of Zayed University Library; while the other 3 – Seri Library Digital Assistant of Universiti Malaya, T- Rex of University Calgary and Lehman Lightning Bot of Lehman college present scope for a lot of improvement. The Library Chatbot of the University of Queensland scores a perfect 5 on social presence and can work on its tone and flow of conversations.

5.6 Error Recognition & Recovery

The subcategories to analyse the chatbot’s capability to identify, acknowledge its limitations, level of responsiveness and its ability to escalate out-of-scope queries to human librarians are presented in the table 6 below

Table 6. Subcategories for category 6 - Error Recognition & Recovery

S.No.	Subcategory	Evaluation Goal
6.1	Error Recognition	Inspects the chatbot’s capability to identify and acknowledge its conversational constraints and adapt measures to restore coherence in user conversations.
6.2	Recovery Strategies	
6.3	Escalation Handling	

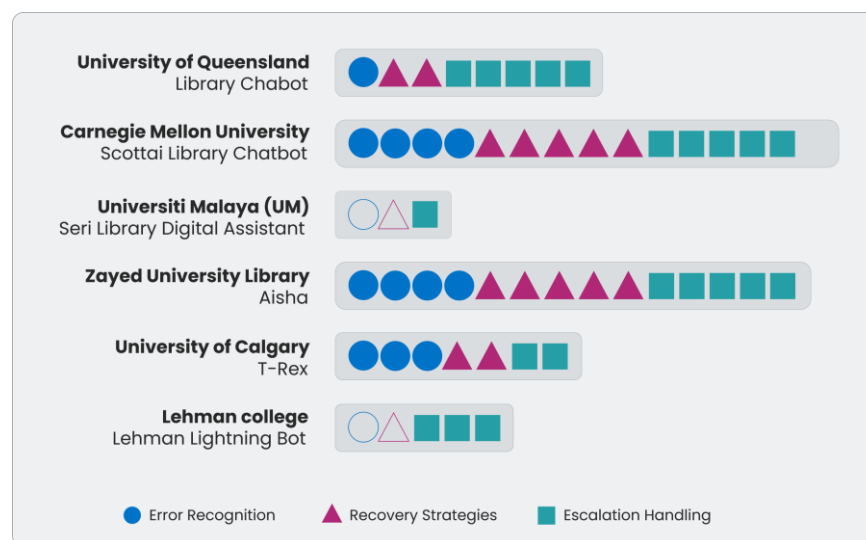


Fig. 7. Error Recognition & Recovery

As observed previously, 3/6 chatbots score 14/15 - Scottai Library Chatbot of Carnegie Mellon University Library, and Aisha of Zayed University Library exhibit the same level of recognising an error, recovering from it, and escalating it to a human librarian. The other four chatbots such as the Library Chatbot of the University of Queensland, Seri of University Malaya Library, the T-Rex of the University of Calgary Library and the Lightning Bot of Lehman College Library were able to escalate to human librarians but were unable to score best in the recognising error and recovering it.

5.7 Context Awareness & Personalisation

The analysis of the chatbot’s ability to handle user’s specific queries, recognise personalised needs, and provide tailored responses is carried out in this category.

Table 7. Subcategories for category 7- Context Awareness & Personalisation

S.No.	Subcategory	Evaluation Goal
7.1	User Context Recognition	Determines the chatbot’s adeptness towards users’ specific queries, recognises personalised needs and provides tailored responses accordingly.
7.2	Learning and Adaptation	
7.3	Proactive Assistance	

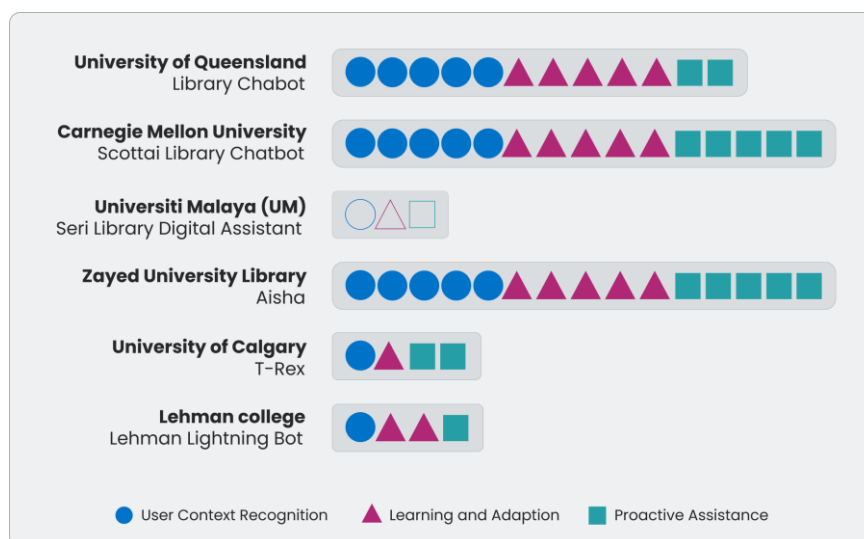


Fig. 8. Context Awareness & Personalisation

It is observed that, Scottai Library Chatbot of Carnegie Mellon University Library and Aisha of Zayed University Library scored a perfect 15/15 while the Library Chatbot of the University of Queensland was nearly at its best. The other 3 chatbots Seri of University Malaya Library, the T-Rex of the University of Calgary Library, and the Lightning Bot of Lehman College Library were observed as the underperformers.

5.8 Service Support & Guidance

The information provided by the chatbot regarding service support, research, its scope and functional boundaries is analysed and discussed in this section. (Fig. 9)

Table 8. Subcategories for category 8- Service Support & Guidance

S.No.	Subcategory	Evaluation Goal
8.1	Service Information	Evaluates the chatbot's role in providing comprehensive library service information and offering support to research-related queries.
8.2	Research Guidance Quality	
8.3	Support Scope & Limitations	

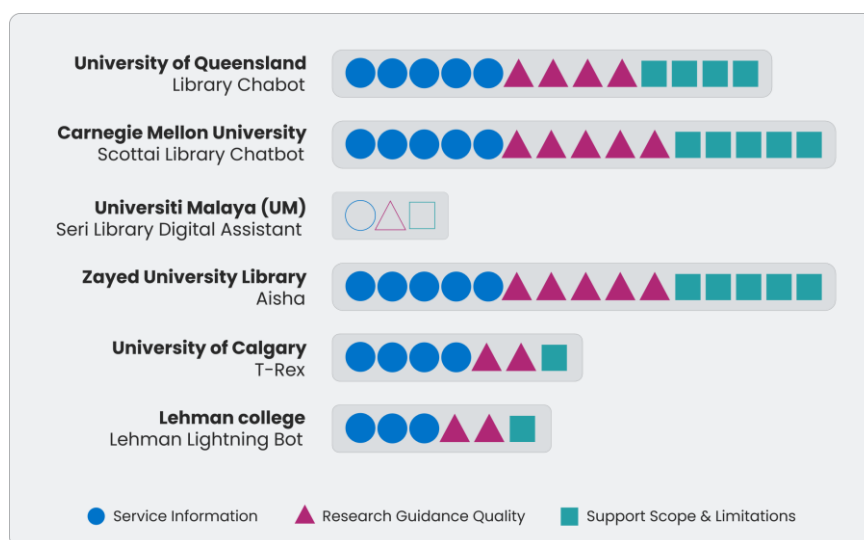


Fig. 9. Service Support & Guidance

In stark contrast to Scottai Library Chatbot of Carnegie Mellon University Library and Aisha of Zayed University Library (15/15), the Seri of Universiti Malaya Library (0/15) does not offer any information about service support, research support and its limitations. The other chatbots such as the Library Chatbot of the University of Queensland, the T-Rex of University of Calgary Library and the Lightning Bot of Lehman College Library exhibited minimal level of service and research support and acknowledge its limitations.

6 Discussion

Table 9. Score comparison of Selected Chatbots

Rank	Chatbots	Language Processing & Error Handling	Task Completion & Effectiveness	Information Accuracy & Currency	Library System Integration	Conversational Quality	Error Recognition & Recovery	Context Awareness & Personalization	Service Support & Guidance	Total 120
1	Aisha	15	15	15	15	15	14	15	15	119
2	Scottai Library chatbot	14	8	9	6	15	14	15	15	96
3	Library Chabot	11	13	12	5	6	8	12	13	80
4	T-Rex	13	10	7	6	2	7	4	7	56
5	Lehman Lightning Bot	6	5	10	3	3	3	4	6	40
6	Seri	4	3	0	0	1	1	0	0	9

Table 10. Salient Features of the Chatbots

Sl. No	Chatbot	University Name	Country	Salient Features
1	Aisha	Zayed University Library	United Arab Emirates	Advanced language processing, high task completion, high conversational quality, context awareness, personalisation
2	Scottie Library Chatbot	Carnegie Mellon University	United States	High conversational quality, effective error recovery, strong context awareness, reliable service guidance
3	Library Chatbot	University of Queensland	Australia	Effective task completion, moderate language processing, reasonable information accuracy
4	T-Rex	University of Calgary	Canada	Basic language processing, limited task effectiveness, minimal conversational quality, weak personalisation
5	Lehman Lightning Bot	Lehman College (CUNY)	United States	Basic information delivery, limited conversational interaction, basic service guidance
6	Seri (Library Digital Assistant)	Universiti Malaya	Malaysia	Very limited language processing, minimal task completion, negligible conversational ability

This paper systematically evaluates 6 chosen academic library chatbots against carefully developed 8 criteria with 120 queries, developed through extensive literature review. The evaluation criteria using a binary scoring system (yes/no) have been successful in identifying among the existing chatbots an exemplar model that can be adopted by academic libraries (Table 9). The salient features of each chatbot have been observed and listed in Table 10. Aisha chatbot of Zayed University emerges as the exemplar model with an excellent score of 119 out of 120 queries, which is testimonial to its ability to handle queries predominant among academic library users. Coming a not so close second and third, are the Scottai Library Chatbot (96) of Carnegie Mellon University Library exhibit adequate performance. By benchmarking strengths and weaknesses across dimensions such as language processing, system integration, conversational quality, and service support, the study not only highlights best practices but also underscores critical pitfalls, thereby offering evidence-based insights for practitioners, developers, and information professionals committed to advancing effective, user-centered chatbot solutions in the academic context. It is proposed that this study will serve as the benchmark for any chatbot to be developed to suit the needs of the academic users.

References

1. De Sarkar, T. (2023). Implementing robotics in library services. *Library Hi Tech News*, 40(1), 8–12. <https://doi.org/10.1108/LHTN-11-2022-0123>
2. Ehrenpreis, M., & DeLooper, J. (2022). Implementing a Chatbot on a Library Website. *Journal of Web Librarianship*, 16(2), 120–142. <https://doi.org/10.1080/19322909.2022.2060893>
3. Fatouh, A. (2024). Investing of Chatbots to Enhance the Library Services. *American Journal of Information Science and Technology*, 8(1), 15–21. <https://doi.org/10.11648/j.ajist.20240801.12>
4. Gunnam, G. R., Inupakutika, D., Mundlamuri, R., Kaghyan, S., & Akopian, D. (2024). Assessing Performance of Cloud-Based Heterogeneous Chatbot Systems and A Case Study. *IEEE Access*, 12(June), 81631–81645. <https://doi.org/10.1109/ACCESS.2024.3397053>
5. Guy, J., Pival, P. R., Lewis, C. J., & Groome, K. (2023). Reference Chatbots in Canadian Academic Libraries. *Information Technology and Libraries*, 42(4), 1–8. <https://doi.org/10.5860/ital.v42i4.16511>
6. Halboub, E., Abdulsamad, R., Hakami, M., Nasser, K., Khalu, A., Ali, S., Hakami, H., & Nasser, M. (2025). Assessment of quality, understandability, actionability, and readability of responses of selected chatbots to the top searched queries about oral cancer. *Digital Dentistry Journal*, 1(February). <https://doi.org/10.1016/j.ddj.2025.100008>
7. Holmes, S., Moorhead, A., Bond, R., Zheng, H., Coates, V., & McTear, M. (2019). Usability testing of a healthcare chatbot: Can we use conventional methods to assess conversational user interfaces? *ECCE 2019 - Proceedings of the 31st European Conference on Cognitive Ergonomics: "Design for Cognition,"* 207–214. <https://doi.org/10.1145/3335082.3335094>
8. Kabakuş, A. T., & Dogru, İ. (2024). The Battle of Chatbot Giants: An Experimental Comparison of ChatGPT and Bard. *Uluslararası Muhendislik Araştırma ve Geliştirme Dergisi*, i. <https://doi.org/10.29137/umagd.1390083>
9. Kane, D. (2019). Analyzing an Interactive Chatbot and its Impact on Academic Reference Services. *Recasting the Narrative: Program Book, April 10-13 ...*, 481–492. <https://dash.lib.uci.edu/stash/dataset/doi:10.7280/D1P075>
10. Kuhail, M. A., Alturki, N., Alramlawi, S., & Alhejori, K. (2023). Interacting with educational chatbots: A systematic review. In *Education and Information Technologies* (Vol. 28, Issue 1). Springer US. <https://doi.org/10.1007/s10639-022-11177-3>
11. Lappalainen, Y., & Narayanan, N. (2023). Aisha: A Custom AI Library Chatbot Using the ChatGPT API. *Journal of Web Librarianship*, 17(3), 37–58. <https://doi.org/10.1080/19322909.2023.2221477>
12. Madhusudhan, M., & Dar, S. A. (2017). Mobile information services and initiatives in university libraries: A new way of delivering information. *DESIDOC Journal of Library and Information Technology*, 37(2), 109–118. <https://doi.org/10.14429/djlit.37.2.11116>
13. Madhusudhan, M., Pandey, P., & Bharti, R. K. (2025). Use of ChatGPT by the Library and Information Science Students at the University of Delhi. *DESIDOC Journal of Library and Information Technology*, 45(2), 85–91. <https://doi.org/10.14429/djlit.45.2.20570>
14. Maroengsit, W., Piyakulpinyo, T., Phonyiam, K., Pongnumkul, S., Chaovalit, P., & Theeramunkong, T. (2019). A survey on evaluation methods for chatbots. *ACM International Conference Proceeding Series, Part F1483*, 111–119. <https://doi.org/10.1145/3323771.3323824>
15. Møller, C. G., Ang, K. E., de Lourdes Bongiovanni, M., Khalid, M. S., & Wu, J. (2025). Metrics of Success: Evaluating User Satisfaction in AI Chatbots. *ICAAI 2024 - Conference Proceedings of the 2024 8th International Conference on Advances in Artificial Intelligence*, 168–173. <https://doi.org/10.1145/3704137.3704182>
16. Ponzio, V., Rosato, R., Scigliano, M. C., Onida, M., Cossai, S., De Vecchi, M., Devecchi, A., Goitre, I., Favaro, E., Merlo, F. D., Sergi, D., & Bo, S. (2024). Comparison of the Accuracy, Completeness, Reproducibility, and Consistency of Different AI Chatbots in Providing Nutritional Advice: An Exploratory Study. *Journal of Clinical Medicine*, 13(24). <https://doi.org/10.3390/jcm13247810>
17. Sharma, M., Verma, S., & Sahni, L. (2020). Comparative Analysis of Chatbots. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.3563674>
18. Singh, S. U., & Namin, A. S. (2025). A survey on chatbots and large language models: Testing and evaluation techniques. *Natural Language Processing Journal*, 10(August 2024), 100128. <https://doi.org/10.1016/j.nlp.2025.100128>
19. Takashi, H. (2019). Robotics and artificial intelligence technology in Japanese libraries. *Ifla Wlic 2019, August*, 1–12. <http://library.ifla.org/id/eprint/2695/1/s08-2019-harada-en.pdf>

An Exploration of Users' Perception towards AI-Based Library Search Systems at the University of Jammu: A Study

Anjali Thakur ¹[0009-0001-6666-4307], Meghna Dhar ²[0009-0008-8296-700X] and Vishal Roudhan ³ [0009-0004-7239-0391]

¹⁻³ Department of Library and Information Science, University of Jammu, Jammu & Kashmir UT, India
at0198941@gmail.com
dhar_meghna@rediffmail.com
vishal991388@gmail.com

Abstract. Artificial intelligence (AI) has significantly revolutionized several sectors and wherein libraries exemplify this paradigmatic shift with unparalleled acuity. AI powered technologies, like machine learning algorithms and natural language processing, enable the libraries to automate their routine tasks which includes retrieval and dissemination of the information, cataloguing and classification of the data (books, periodicals, journals etc.) and other miscellaneous tasks. These technologies exclusively help one to modernize all the library operations thereby making them more accurate, efficient and relevant. This study evaluates how the evolving nature of Artificial Intelligence (AI) enhances the experiences of the users within the academic libraries. The main aim of this study is to understand the perception of the users of the University of Jammu towards the use of AI-based search systems of the library. Further, the study explores various applications of AI in the library settings. Additionally, this paper identifies the key difficulties or challenges and the opportunities of integrating the facility of AI within the infrastructure of university libraries. This paper concludes by underscoring the key importance of purposefully integrating AI into the libraries. Such integration should complement traditional services, thereby enhancing user satisfaction and staff efficiency.

Keywords: Artificial Intelligence, AI-powered Technologies, Information Retrieval, Libraries, Learning Experience, Awareness Programmes.

1 Introduction

The integration of Artificial Intelligence (AI) is currently driving a fundamental shift in library science, moving institutions from traditional storage repositories to dynamic, user-centered information hubs.

Libraries have the potential to automate everyday tasks, streamline operations and achieve accuracy effectively with the help of AI-powered technologies. Artificial intelligence (AI) refers to the development of systems that can perform tasks typically requiring human intelligence such as learning, problem-solving, and decision-making. The day by day increasing implementation of Artificial Intelligence in libraries is driven by its potential to modernize traditional practices, improve efficiency, and enhance user satisfaction. Despite its unlimited benefits, AI implementation introduces significant concerns related to privacy of the data, professional ethics and its influence on the responsibilities of library personnel. To define Artificial Intelligence (AI) in a few sentences is way too difficult because the field is broad, vast and complex. So instead of giving a strict definition, it's more helpful when we explain the boundaries and the key ideas related to AI. At a very basic level, AI is the study of computers and software that have the capability to solve the problems usually handled by humans.

In recent times, AI has become a very rich, interdisciplinary field of study which involves various innovations and ideas. In today's time, people expect more than just computers playing games or making predictions, they imagine robots doing all the tasks just like humans in all situations. And because of these expectations, AI has pushed its boundaries and redefined its intelligent behavior. Most often Intelligence is understood as the ability of one to understand, adapt, learn and apply skills. Intelligence basically is shaped by human experience, imagination, memory, emotions and perception. For example, to recognize a face from decades ago or to understand complex ideas, reflects human intelligence. In intelligence, knowledge plays a vital role and AI also helps one to learn more about how natural intelligence works. AI researchers create models by studying how the brain functions, that advance the field. Although for this AI researchers might have to take many more years to fully understand the brain. It is not just traditional computing or solving simple problems, AI goes further by mimicking human-like reasoning and understanding. It is not limited to psychology, the mind, or the language. To understand AI better let's have a look on several definitions of Artificial Intelligence (AI) specifically in the context of libraries, drawn from expert sources and adapted for clarity are discussed below:

UNESCO (2022) explains AI in libraries as “machines that imitate features of human intelligence like learning, reasoning, and problem-solving to support services such as information retrieval, recommendation systems and digital preservation”. JISC (2022) describes AI in libraries as “theories and techniques developed to allow computer systems to perform tasks normally requiring human or biological intelligence, including personalized services and automated metadata generation”. OECD (2020) states that “AI is a machine-based system that can, for a given set of human-defined objectives, make predictions, recommendations or decisions influencing real or virtual environments such as guiding users to relevant library resources”. These definitions of an AI show how AI in libraries is not just about automation, but about enhancing access, personalization, and efficiency in a rapidly evolving digital information landscape.

2 Review of Related Literature

As more research is now illustrating how AI is changing library services across multiple areas, Barsha and Munshi (2024) in their document talked about using AI in libraries, especially in developing countries [1]. Their paper shows that AI can help in organizing the knowledge, finding the information, improving the services of the users and managing the resources more efficiently. AI-powered tools can automate each and every task in the library like cataloging books, providing recommendations and improving how people search for information. However, there are some major challenges which include lack of awareness among the users, weak technology infrastructure, financial limitations and insufficient trained professionals. The paper emphasized that, to successfully use AI in libraries, countries need better training programs, investment in technology and international support. Jyoti and Kumar (2024) in their document summarizes various important findings in the paper which strongly support the transformative role of artificial intelligence within the academic and public library settings [4]. It emphasized how AI is enhancing resource discovery and user engagement. They find out that the virtual assistants are emerging as key tools and which offer continuous support, customized responses, and language translation and side by side also influencing the evolution of librarian roles. Additionally, AI is also supporting the digitization and preservation of delicate library materials through their technologies such as image enhancement, optical character recognition, 3D scanning and predictive maintenance. The study also stressed on the need to incorporate ethical policies related to privacy, data protection and algorithmic fairness which ensures that the AI applications in libraries are implemented responsibly and sustainably. Lulu- Pokubo and Okwu (2024) in their study examined how Nigerian librarians understand and use an AI in their libraries [5]. It focused on their awareness, benefits, needed applications and the challenges in the adoption of an AI. Researchers surveyed 203 librarians and found out that they are highly aware of AI technologies. They realized that AI, as a helpful tool for reducing errors, speeding up tasks and improving knowledge discovery. Librarians believed that AI is mostly needed for cataloging, classification, circulation and reference services. The study also provided insights into AI’s potential in Nigerian libraries and simultaneously highlighted the need for proper training and strategic implementation of it. Marzuki et al. (2025) in their study explored the user behavior and personalization in digital libraries by identifying six key activities that are searching, personalization, content interaction, collaboration, browsing collections and multimedia engagement [6]. However, challenges like privacy risks, algorithmic bias and scalability issues were noted, highlighting concerns about the data security and equitable access and put considerable focus on privacy, fairness and inclusivity. The study highlighted the importance of balancing innovation with ethical safeguards which ensured digital libraries to remain equitable and effective. University News (2025) paper highlighted that the era of Agentic AI now takes the action and makes choices on its own instead of just showing insights. It evaluates that it enhances the efficiency of the libraries by reducing repetitive tasks like data cleaning, allowing data professionals to focus on business strategy and anomaly detection. It highlighted that as organizations adopt AI, then advanced analytics professionals will play a pivotal role in harnessing its potential for predictive insights, streamlining workflows and driving innovation. It further suggested that the companies must invest in AI training and ethical considerations to ensure responsible implementation and with its ability to redefine strategic planning, Agentic AI is reshaping the future of data-centric decision-making and business growth.

3 Research Gap

While existing studies have extensively explored the applications and benefits of AI in library services, including automation, resource discovery and user engagement, several gaps remain. Most research focuses on technological capabilities and broad implementation challenges, such as infrastructure and training needs, without deeply investigating the user perspective on AI-powered library search systems. Specifically, there is limited empirical evidence on users' awareness levels, their agreement or acceptance of AI integration, and their primary concerns

regarding privacy, data security and ethical implications. Additionally, few studies thoroughly examine the critical factors users consider important when interacting with AI-based search tools in libraries. This gap highlights the need for research that centers on user attitudes and experiences to inform more user-centric and ethically responsible AI adoption in library settings particularly in University of Jammu.

4 Objectives of the Study

The following are the objectives of the study:

- To assess users' awareness towards the application of AI-based search systems.
- To evaluate the extent of user agreement with the adoption of AI-powered library search systems.
- To identify the primary concerns of users regarding the use of AI-powered library search systems.
- To determine the importance of various factors considered by users while utilizing AI-powered library search systems.

5 Research Methodology

The proposed study was undertaken to explore in detail the User's Perception towards AI-Based Library Search Systems at the University of Jammu. In order to methodically investigate user perceptions, experiences and expectations towards AI- powered search tools in the library under study, the quantitative approach was adopted. The data of this study was collected both from primary and secondary sources which included books, reports, journals, theses, online sources etc. The primary data was collected through field work. In this study, wherever possible, observation will be made by participating and interacting with the users of various libraries under study. A semi-structured questionnaire was framed for collecting data and carrying out the research. About 106 questionnaires were filled out by the users of the library under study. The data so collected was analysed and tabulated with the help of various mathematical and statistical techniques. For data analysis, Microsoft Word was used for qualitative feedback. Microsoft Excel was employed for quantitative processing. It was used for creating tables and applying descriptive statistics like percentages and frequency counts.

6 Data Analysis and Interpretation

Data analysis is to convert the raw data that reports the issues in the research, into meaningful knowledge. Here in this section, we enumerate the various analysis techniques applied to gather the data from the questionnaire through Google Forms distributed to 127 students and the scholars of University of Jammu. Out of which only 106 questionnaires were received back duly filled including 74 Students and 32 Research Scholars. The data is analyzed using simple statistical tools.

6.1 Awareness about the application of AI-Based Search System

The respondents were asked about their awareness towards the application of AI Based library search system. Their response is shown in Table 1.

Table 1. Awareness about the application of AI-Based Search System

Option	Response Number	Percentage
Yes	72	67.9%
No	34	32.1%

As shown in Table 1, it is calculated that about 72 (67.9%) users were aware of the AI based search system utilized by University of Jammu Library while 34 (32.1%) respondents point out that they were not familiar with it. The findings suggested that most of the users were informed about the AI-enabled search system used in the university library while nearly one-third remained unaware about its presence or functionality.

6.2 Extension of user agreement with the statements regarding AI -powered library search system.

Regarding the AI-based library search system, the respondents were asked about how much they agreed with the following statements. Their responses are indicated in Table 2.

Table 2. Extension of user agreement with the statements regarding AI-powered library search system.

Statement	No. of Respondents				
	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
AI-powered search systems make finding of information easier.	13 (12.3%)	3 (2.8%)	19 (17.9%)	42 (39.6%)	29 (27.4%)
AI-powered search systems will provide more relevant search results.	6 (5.7%)	5 (4.7%)	33 (31.1%)	37 (34.9%)	25 (23.6)
AI-powered search will understand queries better	9 (8.5%)	6 (5.7%)	28 (26.4%)	38 (35.8%)	25 (23.6%)
AI-powered search systems will introduce bias into search results.	8 (7.5%)	10 (9.4%)	46 (43.4%)	25 (23.6%)	17 (16.1%)
AI-powered search systems will reduce the need for librarians.	14 (13.2%)	19 (17.9%)	32 (30.1%)	23 (21.7%)	18 (17.1%)

From Table 2, it is clear that majority of respondents 71 (42 (39.6%) Agreed + 29 (27.4%) Strongly Agreed) i.e., 67% believe that AI-powered search systems make finding of information easier, while only 16 (3 (2.8%) Disagreed + 13(12.3%) Strongly Disagreed) i.e., 15.1% expressed disagreement with the statement. The positive perception towards AI's ability was because of the reason that it quickly scans vast amounts of data and presents relevant results thereby saving users significant time and efforts compared to manual searches Regarding relevance of search results, 62 respondents (37 (34.9%) Agreed + 25(23.6) Strongly Agreed) i.e., 58.5% supported the claim, with 33 (31.1%) remaining Neutral. The ability to handle complex queries was also rated positively by 63 respondents (38 (35.8%) Agreed + 25 (23.6%) Strongly Agreed) i.e., by 59.4%. The reason being that AI systems can interpret multi-faceted questions and provide detailed answers that traditional search methods may not produce. In terms of potential drawbacks, 25 (23.6%) respondents Agreed and 17 (16.1%) Strongly Agreed that AI may introduce bias showing clear concern among 42 (39.7%) of respondents. The idea that AI might reduce the need for librarians received mixed views. Only 18 (17.1%) Strongly Agreed, while 32 (30.1%) stayed Neutral and 33 disagreed to some extent. Trust in AI recommendations revealed cautious optimism, with 43 respondents (30(28.3%) Agreed + 13 (12.3%) Strongly Agreed) i.e., 40.6% expressing confidence, though 44 (41.5%) remained Neutral. This was due to users' growing familiarity with AI benefits A strong majority of 63 (59.5%) respondents agreed that these systems improve research efficiency and enhance learning experiences.

6.3 Primary concerns of users about using an AI-powered library search system.

The respondents were asked about their primary concerns regarding using an AI-powered library search system. Their responses are indicated in Table 3

Table 3. Primary concerns of users about using an AI-powered library search system.

Option	Responses Number	Percentage
Privacy of my search queries	62	58.5%
Accuracy of results	68	64.2%
Bias in recommendations/results	37	34.9%
Over-reliance on technology	39	36.8%

Lack of human oversight	33	31.1%
Ethical implications of AI	42	39.6%
Difficulty in understanding how it works	22	20.8%
No specific concerns	23	21.7%
Others	13	12.3%

From Table 3, it is indicated that about 68 (64.2%) of respondents were concerned about the accuracy of the results, 62 (58.5%) were concerned about the privacy of their search, 42 (39.6%) concerned about ethical implications of AI, 39 (36.8%) on over-reliance on technology, 33(31.1%) on lack of human oversight, 37 (34.9%) on bias in recommendations/results while 23 (21.7%) had no specific concerns, 22 (20.8%) were concerned on the difficulty in understanding how it works and 13 (12.3%) of respondents had other concerns.

6.4 Important factors while using AI-powered library search system by the users

The respondents were asked about the important factors they consider while using an AI- powered library search system. Their responses are indicated in Table 4.

Table 4. Important factors while using AI-powered library search system by the users

Options	Not Important	Slightly Important	Moderately Important	Very Important	Extremely Important
Speed of results	9 (8.5%)	16(15.1%)	16 (15.1%)	36(34%)	29(27.3%)
Relevance of results	8 (7.5%)	15(14.2%)	17 (16%)	34(32.1%)	32(30.2%)
Comprehensiveness of results	2 (1.9%)	22(20.8%)	14(13.2%)	42(39.6%)	26(24.5%)
Ease of use/user- friendly	8 (7.5%)	15(14.2%)	17 (16%)	36 (34%)	30(28.3%)
Availability of filters	5 (4.7%)	19(17.9%)	13(12.3%)	42(39.6%)	27(25.5%)
Visual presentation of results	8 (7.5%)	18(17.1%)	14(13.2%)	40(37.7%)	26(24.5%)

From Table 4, it is indicated that the most highly rated factor is Relevance of results, with the majority marking it as either Very Important or Extremely Important. Speed of results and Ease of use/user-friendliness also rank prominently, showing a strong preference for fast and user-friendly experiences. Features like Availability of filters/refinement options and Comprehensiveness of results received balanced attention across moderate to high importance levels. Meanwhile, Visual presentation of results was considered less critical, with a larger portion assigning it low or moderate importance. The most highly rated factor is the relevance of results, as users prioritize finding accurate and pertinent information quickly. Speed of results and ease of use are also important, reflecting a preference for efficient and accessible search experiences. Features like filters and comprehensiveness help users narrow down and ensure thorough coverage of information. Visual presentation is considered less critical since users focus more on content quality than on design.

7 Findings

The aim of the study was to investigate user's perceptions and preferences towards the AI based library search system at the University of Jammu where the data was gathered through questionnaires and analyzed using basic statistical techniques. Based on the analysis of this study following are the findings:

- The majority of respondents, 67%, believed that AI-powered search systems make finding of information easier. The users appreciate the practical advantages of AI in search and research, while some respondents were aware of its limitations and the importance of human oversight.
- Most users of the university 70 (66%) were engaged with the AI search system regularly, though 20 (18.9%) respondents reported never using it and indicated potential for increased awareness or improved usability
- Most of the participants showed high levels of confidence in the positive impact of AI on information retrieval and academic tasks, though concerns around bias and human displacement remain notable.
- The majority of the respondents 68 (64.2%) were mainly worried about how correct AI-based search results. At the same time, 22 (20.8%) were concerned about the difficulty in understanding how the AI-powered library system works.
- Most of the users 32 (30.2%) emphasized relevance of results as key aspects of an effective AI-powered library search system. While only 2 (1.9%) of users felt the importance of comprehensiveness of results.

8 Recommendations

Based on the analysis of this study following are the recommendations

- Increase Awareness and Adoption of AI-powered Library Search: To increase adoption of the AI in the university, awareness programs (hands-on training sessions, informative guides, and clear communication about the system's features and functionalities) should be organized to educate users about the capabilities and benefits of AI search tools.
- Enhance Relevance and Usability of AI Search: The university should have major focus on improving the relevance and precision of search results and also optimize the user interface for seamless access and task completion.
- Address Concerns around Bias, Privacy, and Ethics: The university must develop policies and guidelines to ensure that its AI-powered system operates with transparency and fairness.

9 Conclusion and Future Directions

The study evaluates how students, faculty and research scholars at the University of Jammu use AI-based library search systems. This paper highlights extensively the growing role of AI in libraries where automation, personalization and intelligent search tools are reshaping user experiences. This descriptive survey of 107 respondents found that a majority (60.5%) agreed that AI improved their research efficiency and enhances the learning, appreciating the features like faster searches and personalized recommendations. At the same time, they showed concerns about privacy, fairness, and reduced human involvement indicating the need for careful oversight. The survey also revealed that younger users of the University dominated the sample and suggested strong digital literacy and regular exposure to AI technologies. Overall, while users of the University of Jammu expressed confidence about using AI's benefits, and they also highlighted the importance of balancing the evolving technology with human expertise so as to ensure its ethical, fair and effective integration in the academic libraries. The present study leads to the following scope for future research.

- A comparative study of AI -based library search system in Academic and Special libraries.
- Library Professional attitudes towards AI-based search system in various University Libraries.

Declaration of AI and AI-Assisted Technologies used in Manuscript Writing Process: Not Applicable

Reference:

1. Barsha, S., & Munshi, S. A. (2024). Implementing artificial intelligence in library services: A review of current prospects and challenges of developing countries. *Library Hi Tech News*, 41(1), 7–10. <https://doi.org/10.1108/LHTN-07-2023-0126>
2. Desai, A. (2025). The emerging era of agentic AI: Redefining roles in data analytics. *University News*, 63(20).
3. Jain, P. (2022). Transforming libraries with artificial intelligence: Benefits and challenges. Elsevier.
4. Jyoti, & Kumar, P. (2024). Reshaping the library landscape: Exploring the integration of artificial intelligence in libraries. *IP Indian Journal of Library Science and Information Technology*, 9(1), 29–36. <https://doi.org/10.18231/j.ijlsit.2024.005>
5. Lulu-Pokubo, E. P., & Okwu, E. (2024). Librarians' awareness towards the use of artificial intelligence technologies for sustainable library services. *Business Information Review*, 42(1), 20–28. <https://doi.org/10.1177/02663821241310042>
6. Marzuki, M., Azero, S. F. Z., Zamzuri, N. A. A. M., & Kadir, M. R. A. (2025). A systematic literature review of user behavior and personalization in digital libraries. *International Journal of Research and Innovation in Social Science*, 9(1). <https://doi.org/10.47772/IJRISS.2025.9010372>
7. Sharma, P., & Gupta, R. (2020). AI and chatbots in library services: A new era of user interaction. *Information Technology and Libraries*, 39(3), 212–226. <https://doi.org/10.6017/ital.v39i3.12363>

Sustainable Library Website Evaluation: Significance Parametric Study

Kreethika. M ¹[0009-0006-6537-6764], D. Gnanasekaran ²[0000-0003-4999-3210], P. Ganesan ³[0000-0003-1764-9095] and

T.M. Kalpana ⁴[0000-0002-1321-7992]

^{1,2}Kalasalingam Academy of Research and Education, Anand Nagar, Tamil Nadu – 626126

³Alagappa University, Alagappa Puram, Karaikudi, Tamil Nadu - 630 003.

⁴Indian Statistical Institute, Nelson Manickam Road, Aminjikarai, Chennai, Tamil Nadu-600029

¹kreethika15@gmail.com

²gsekard@gmail.com

³gharieni2003@yahoo.com

⁴varatharajkalpana1971@gmail.com

Abstract. Sustainability has become a major concern for academic institutions like universities and colleges as their services increasingly shifted towards digital platforms. This study proposes a SEE approach to evaluate sustainability of library websites by systematically identifying, categorizing and measuring key website based on checklist parameters. This approach ensures the data feasibility through Web Content Accessibility Guidelines (WCAG) and Guidelines for Indian Government Website (GIGW). From the data evaluated from the universities on information availability with respect to enlisted parameters found that major focuses were on cost effectiveness. Networking & performance ranked second with parameters that captures impact of storage of digital information and its impact on environment. The iterated parameter proves environmental sustainability is high (24.32) enables user to access content and least in maintenance efficiency. With align to universities Alagappa University (ALU) secured first with all (three sustainable measures. Anna University (AU) secured second and Annamalai University ANU) with least in consideration of sustainable impacts. This paper highlighted that by applying parametric evaluation model, libraries can identify high impact areas for improvement and balance technological innovations and sustainable resource use globally.

Keywords: Library website, Parametric approach, Sustainability measures, WCAG, GIGW.

1 Introduction

Website acts as gateway to the organization and people of today explores the services and products on their desktop. In the same way, for a library user, library website is the starting point to explore the sources, services and infrastructure facilities, necessitates a comprehensive library website (Vinod Kumar and Bansal, 2014). The developments of Information and Communication Technology (ICT) have substantially changed the functioning of library services and content rich website is judged based on the resources, services and facilities provided to the users (Devi and Verma, 2018). Exponential information growth due to technological advancements forces the library to up-to-date the website and to meet sustainable development goals in terms of quality in education, reducing inequalities, responsible consumption and climate action that denotes the amalgamation of environmental protection, social responsibility and economic efficiency for the betterment of today and tomorrow's generation. Sustainable Development Goals (SDGs) adopted by United Nations in 2015 is a call for the world to end poverty, protect the planet and people to enjoy the peace and prosperity by 2030. These seventeen goals are integrated and action in one area will affect the outcomes in others, and that development must balance social, economic, and environmental sustainability. The United Nations SDGs addressing the current trends of climate change, digital transformation and inclusivity that provides a guiding framework to increase in adopting green practices, renewable energy resources, digital carbon monitoring and socially responsible to align with sustainable globally. Hence, library websites perform these global goals at the institutional levels.

1.1 Sustainability Concepts (SC) in library websites

Sustainability in libraries means ensuring environmental responsibility, economic feasibility and social equality of all library functions to the library users of current and future generations. It should deliver inclusive, and accessible resources through digital forms and economically feasible for information services without negotiating

environmental or social equity. Sustainability has emerged as a guiding principle in higher education in India that encompasses library websites as essential gateway for digital services beyond physical spaces and they have direct implications for sustainable web-based services by adopting lightweight web design, renewable energy-based hosting, and accessibility standards. The (Ministry of Education, 2023) Government of India conducted survey in 2021-2022 in which it has revealed that that India has more than 1,168 universities, 45473 colleges, 12,002 standalone institutions, with Gross Enrolment ratio (GER) demands for digital academic services [1]. Developing sustainable and accessible library website is essential which support various user populations in making policy, monitoring, and improving the quality of higher education that aligns with national policies like National Education Policy (NEP 2020) with align to UN SDGs assessment.

The United Nation (UN) had developed 17 Sustainable Development Goals (SDG) for putting end to poverty and ensuring the prosperity in 2030.

Among 17 SDGs, certain goals can collectively contribute to web-based library services that aligns with web content accessibility compliances for higher education. The growth of nation will be possible by ensuring sustainability measure to standard and improve sustainability for long-term effectiveness and efficiency to minimize the adverse impact.

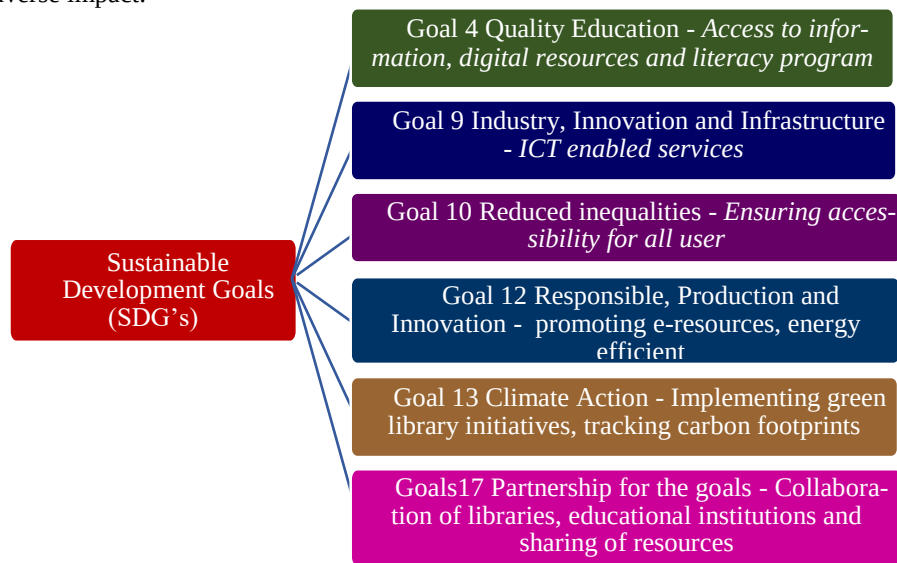


Fig.1. Sustainable Development Goals for digital Libraries

Figure 1, depicts the library websites services in align with SDGs that can be enhanced through accessible web design of Web Content Accessibility Guidelines (WCAG) and GIGW (Guidelines for Indian Government Websites) by (World Wide Web Consortium, 2018) on inclusive content, energy resources, digital practices, optimized navigations and user-centric services. These practices lead to improved accessibility, reduced digital inclusions, enhanced user satisfaction and long-term digital resilience [2].

2 SEE (Social, Environmental, Economic) Approach

This study proposes an integrated SEE based framework with defined indicators to evaluate the sustainability of academic library websites in terms of social, environmental, economical.

- Social Sustainability in library website focuses on socially inclusive, equitable and accessible resources with standard accessibility to all users without geographical or social discretion. e.g. Accessibility compliances, Mobile responsiveness, Multilingual content, Inclusive design
- Environmental Sustainability of library website can be gained by energy efficiency with low carbon emission and promote eco-friendly practices for green libraries. E.g. Green web hosting, lightweight design, Carbon emission, Content Delivery Network.
- Economic Sustainability in library websites helps to adhere cost effective, reliable, capable and shared usage of resources. e.g. Open-source automation tools, Digital preservation, Cost efficiency.

These dimensions widely cover sustainability concepts which are figured as Social (SOC), Economic (ECO) and Environmental (ENV) that covers Content (CN), Usability & Accessibility (UA), Design & Functionality

(DF), User engagement & Quick Links (UEQL) and Networking & Performances (NP) which are parameters widely used in library website evaluation studies.

Table 1. Indicative list of Sustainability Vs. library website parameters & Guidelines

SEE Approach	Library Website Parameter	WCAG (International Accessibility Focus)	GIGW (Indian Government Standards Focus)
Economic	About Vision and Mission	Font visibility, semantic language link/text for images	Trustworthiness of content, metadata for indexing, regular and mandatory updates.
Economic	E-Resources	Responsive design, useful links, e-learning portals, link/text for external sources	Data compatibility with open standards, regular Open standards for data formats, periodic assessment of content, authorized and authenticated sources.
Economic	Policies such as rules, website policies, and security policies	Clarity of content, readability, structure, accessibility format, With tagged	Compliance with legal frameworks.
Environmental	Multimedia information	Encryption, transcription of data, captions, audio and video.	Multilingual facility, authorized mode of platform by the government, web compliance.
Environmental, Social	Library catalog, OPAC/WEB OPAC	Descriptive information for effective search accessibility and functionality with color contrast.	Authenticity of data privacy and security with Secure Sockets Layer (SSL) with an institutional system, interoperability with the network.
Environmental, Social	Design and Layout	Design based on Colour contrast, font resizable, mobile friendliness	Maintaining standard templates, uniformity in look for the organization Guidelines.
Social	Current Information, such as news and events	Correct way for notifications, alternative text/link for event, images and captions for audio and video.	Regular updates, an archival facility, and the genuineness of information.
Social, Economic	User Support service, such as FAQ, Ask the Librarian	Accessibility through a common language with text/link for labels, messages, and Audio	Required feedback mechanism. Regular reviewing, safe and secure submission forms.
Social, Economic	Organogram, such as Staff, committee members	Structured format for screen reader access, link/text for information about them.	Standardization of Information about contacts, roles, and organization visibility.
Social, Economic,	Privacy and Security	Robust code for accessibility and compatibility across browsers	SL security is mandated with regular audits in terms of compliance with CERT security guidelines.
Social Economic	Homepage & Navigation	Title headings, Screen Reader, Keyboard navigation	Simple, reliable access, bilingual language, and organization logo

Table 1, serves as mapping table of International and national guidelines of (Guidelines, 2025) and that links SEE dimensions with specific library website parameters for structured content analysis and scoring [3].

There are similar approaches that aligns to SEE likely (TBL) Triple Bottom Line focused on (People-Planet-Profit) as 3P's coined by John Elkington in 1994 and ESG (Environmental-Social-Governance) initiated by United Nation in 2004. Therefore, the SEE framework is based on similar to triple bottom line approach that directly relevant to developing a sustainable academic library website.

3 Objectives of the study

- To identify and analyze SEE approach of sustainability based on academic library websites.
- To examine the impact of sustainable concepts that cater to WCAG and GIGW standards to build an effective library website
- To quantitatively analyze 5 select university library websites on SEE approach based on ranking methods.

4 Literature Review

4.1 Sustainability Frameworks and Eco-friendly Practices

The literature identified for the study comprises of both conceptual and empirical approaches. Chowdhury (2014, 2015) used a conceptual and analytical methods to develop sustainability framework and estimate the environmental and energy cost of digital libraries. These studies emphasized the Life Cycle Assessment (LCA) and Information Communication Technology (ICT) energy cost reduction for sustainable digital libraries based on three-dimensions such as social, environmental and economic [4-5]. Gajbhiye (2024) focused on integration of eco-friendly practices in library operations including energy efficient designs and community engagement. These practices give awareness programs on sustainability development [6]. Verma (2025) utilized text mining and modelling over scholarly articles with 15 keywords highlighted the signaling growth patterns in keywords analysis in open resources and community support programs [7]. Kalpana & Gopalakrishnan (2013) proposed a FEEL framework focused for financial sustainability, renewable energy implementation, eco-friendly practices and open access resources. This study emphasized innovation strategies of fund raising, use of solar energy, green library concepts to ensure long-term sustainable measures of academic libraries [8].

4.2 Evaluation of Library Websites and Web-based Sustainability Services

Haridasan & Firdaus (2021) adopted a content analysis method using a checklist to evaluate 42 Indian University library websites from NIRF Ranking to assess the availability of web-based services, resources and social media tools supporting sustainability [9].

4.3 Accessibility, Inclusivity, and Policy Perspectives

Ara et al. (2025) proposed a framework that provides detailed accessibility, including overall scores along with disability types and guidance for further manual or expert testing, with several limitations of existing solutions for web content [10]. Asok & Rekha (2025) highlighted policies for stakeholders, policymakers, web developers, and library administrators to bridge the accessibility gaps. The study helps to minimize the violations and create an inclusive website [11].

4.4 Website Design, Development, and User Experience

Kumaran & Subangi (2017) elaborated the principles by describing the design and development of well-structured library website with user friendly approach that eases resources sharing and enhances library services [12]. Kim (2011) explored the key factor that affects the design of university library website highlighting the reputation of institutional forces and user feedback to enhance website effectiveness and efficiency [13]. Feng (2011) proposed a structure for university library mobile website for analyzing the compatibility of existing mobile sites and its functionalities that aimed to assess the improving accessibility to resources based on user experience [14].

4.5 Frontend Technologies and Performance

Kaur, G. & Tiwari, R. G. (2023) evaluated popular frontend framework such as React, Angular and Vue focusing on usability, performances, and community engagement to facilitate developers in developing website [15].

5 Research Gap

Based on the literature review, the research gap highlighted that the absence of an integrated structured framework or standards for developing exclusive library website that links Sustainability WCAG compliance, library service concepts, social inclusivity, operability, adaptability and institutional accountability. The lack of standardized sustainability metrics that paid attention to environmental effects like energy efficiency and carbon footprint and the dearth of research from developing country contexts further underscore the need for a comprehensive and context-specific sustainability framework for academic library websites.

6 Scope of the study

From the UGC website, 22 state universities were identified in which library website of 5 select universities represented in alphabetical order namely Alagappa University (ALU), Anna University (AU), Annamalai University (ANU), Bharathiar University (BRU) and Bharathidasan University (BDU), were analysed for the present study.

7 Limitation of the study

The select 5 universities represent diverse institutional types, geographic distribution and digital levels ensuring comparative evaluation. This study is focused on connecting the concepts of sustainability with maximum number of library website parameters with adherence to national GIGW, international WCAG standards based on library scenario and further analysed with parameters and its significance.

8 Research Methodology

The website of 5 select state university libraries were selected based on purposive sampling and checklist of parameters were prepared on the basis of previous evaluation and a web assessment criterion done by 1. Ambika & Ganesan (2020) [16], 2. Ashrafi-Rizi et al., (2014) [17], 3. Devi & Kumar (2023) [18], 4. Devi & Verma (2018) [19], 5. Jangwal & Madhusudhan & Soni (2024) [20], 6. Kumar et. al., (2009) [21] 7. Samrgandi, (2020) [22], 8. Silvis et al. (2019) [23], 9. Vinod Kumar & Bansal (2014) [24], A structured framework is uniformly employed through all the parameters for evaluation and marked as two Scales No (1) and Yes (2). The collected data have been presented in tabular form and calculated for the purpose of universities library website.

There are 282 parameters which were separated based on five concepts and the repetition was removed. By removing repetition, 227 parameters were taken for study and iterated are widely used, denoted as “Most significant, “Sub-significant, In-significant”. The data were assessed using Statistical Package for Social Sciences (SPSS) (ver. 26) package and Microsoft Excel. The analysis and interpretation of the data collected through the checklist were followed for the study.

9 Data analysis and Interpretation

Table 2. Library website concepts, Sub Concepts and No. of parameters Vs. Sustainability Concepts

Con-cepts	Sub-Con	Soc	Env	Eco	Total	Mean	Std. Dev.	Rank
CN	3	31	31	36	98	43.17	2.887	1
UA	7	16	4	9	29	12.78	6.028	4
DA	5	15	1	3	19	8.37	7.572	5
UEQL	6	25	1	4	30	13.22	13.077	3
NP	6	6	37	8	51	22.47	17.349	2
Total	27	93	74	60	227			

From table 2, it can be inferred that total of 227 parameters for the library website was identified. It can be seen that content places first with the highest list of parameters that provides the vast information to the users about library & its collections and services. Followed by which, Networking & Performances (5) stands in second position with 51 parameters that defines library web hosting, infrastructure and website management and performances. User Engagement & Quick Links (4) are featured to connect with social media, community engagement and events. Usability & Accessibility (2) and Design & Functionality (3) were given less importance to Accessibility & Compliance, Navigation with features and user-friendly approach. Social sustainability has highest number of parameters (93) followed by Environmental sustainability (74) and Economic sustainability (60) being library is a social service entity. Economical aspects should be given least importance compared to social and environmental issues.

Table 3. Iteration of parameters vs Sustainability Concepts

Parameters	Soc	Env	Eco	Total Value	Mean	Std. Dev.
None (0)	39	17	25	81	35.68	11.136
Mean (0)	41.94	22.97	41.67			
One (1)	16	14	4	34	14.98	6.429
Mean (1)	17.20	18.92	6.67			
Two (2)	7	7	7	21	9.25	0.000
Mean (2)	7.53	9.46	11.67			
Three (3)	6	6	8	20	8.81	1.155
Mean (3)	6.45	8.11	13.33			
Four (4)	8	12	5	25	11.01	3.512
Mean (4)	8.60	16.22	8.33			
Five (5)	17	18	11	46	20.26	3.786
Mean (5)	18.28	24.32	18.33			
Total	93	74	60	227		

Table 3, depicts the most significant parameters marked based on “Yes” (2 marks) available in all 5 universities taken for the study. Sub-significant parameters were listed based on “Yes” denoted by above 1 and below 5 universities. The parameters that were not marked by any of the universities were stated as insignificant parameters.

Table 4. Ranking of Universities with Sustainability Concepts (SC)

SC	SOC				ENV			ECO				
	Total Score	SCORE %	Std. Dev.	Rank	Total Score	SCORE %	Std. Dev.	Rank	Total Score	SCORE %	Std. Dev.	Rank
ALU	128	28.19	0.487	2	119	26.21	0.492	1	91	20.04	0.504	1
AU	122	26.87	0.466	4	111	24.45	0.503	2	86	18.94	0.500	2
ANU	122	26.87	0.466	4	110	24.23	0.503	3	75	16.52	0.437	5
BRU	135	29.74	0.500	1	107	23.57	0.500	4	83	18.28	0.490	3
BDU	123	27.09	0.470	3	107	23.57	0.500	4	82	18.06	0.486	4
Max. Score	186	100			148	100			120	100		

Parameters specified under WCAG and GIGW taken for sustainability of library website study be considered as

Let $\{C_i : 1 \leq i \leq 3\}$ denotes the 3 concepts namely Social, Environment and Economic.

— For Concept C1 mean of parameters of Social sustainability is defined as

$$\{\kappa_1, \kappa_2, \dots, \kappa_{93}\}$$

— For Concept C2 mean of parameters of Environmental sustainability are defined as

$$\{\omega_1, \omega_2, \dots, \omega_{74}\}$$

— For Concept C3 mean of parameters of Economic Sustainability are defined as

$$\{\mu_1, \mu_2, \dots, \mu_{60}\}$$

For all parameters marked for different concepts collectively P_i

Parameters were marked based on “Yes” (2 marks) if available and marked as “No” (1 mark) if not available in all 5 universities taken for the study.

Parameters P_i collective defined as

$$P_i = \left\{ \sum_{\kappa=1}^{93} + \sum_{\omega=1}^{74} + \sum_{\mu=1}^{60} \right\} = 227$$

University library website be defined as

$$\tilde{U}_j = \{ \tilde{U}_1, \tilde{U}_2, \dots, \tilde{U}_3 \} \text{ where } P_i \in \tilde{U}_j$$

we mean that the parameter P_i is available in the university library website

Table 4, represents comparative assessment of sustainability across three concepts Social, Environmental, Economic directs the total score gained by each university associated to the total score (454), its percentage, mean, standard deviation. The ranking was specified based on highest mean and low standard deviation.

From the table 4, out of (186) total score, Bharathiar University reveals the strongest performance in social sustainability with total score of (135), indicating high level of inclusivity, accessibility and community engagement that reflects through its library website and practices. Followed by Alagappa University and Bharathidasan University with (128) and (123) score to balance social initiatives and user focused digital services. Whereas, Anna University and Annamalai University shows equal score (122), comparatively giving lesser importance in socially inclusiveness.

Under environmental sustainability (148), Alagappa University secures (119) scores, signifies its activeness in implementation of green library awareness content such as energy efficient hosting, and optimized digital resources. Anna University (111) follows closely with Annamalai University (110) reflects divergence in environmental measures. While Bharathiar and Bharathidasan University shares equal score of (107) denotes less contribution in implementation of environmental sustainability and / or not revealed in the library website.

In economic sustainability (120), Alagappa University scores (91) indicating strong financial sustainability through adopting library management systems, circulation process and digital repositories that reduce operation costs and long-term expenditure. Anna University secures (86) shows moderate economic efficiency and slightly limits depends on vendor-based system and cost effectiveness. Bharathiar University and Bharathidasan places slight variations with (83) and (82) scores of partial and average implementations of cost saving digital measures but indicating a need to further optimization through automation and digital integration. The advancement of shared infrastructure or cloud-based repositories can significantly enhance financial efficiency. Whereas, Annamalai University with lowest score in minimal implementation of digital transformations. The lower performance implies a higher in operational cost structure and insufficient in financial practices.

10 Findings

From the study, it can be seen that sustainability-based website parameters across five concepts reveals significant variations in performances across the three sustainability concepts that shows effective and efficiency of depictions in library website based on inclusivity, eco-friendly and cost effectiveness.

From Table 2, Content revealed the highest contributor to overall sustainability with mean value (43.17%) indicates most library website giving importance to content in rich with accurate and relevant information that aligns with three sustainable concepts.

Networking & Performances rank second with (22.47%) focusing on collaborative networks, web performances and interoperability that emphasis both social visibility and environment efficiency through digital infrastructure in Table 2.

User Engagement & Quick Links places third in Table 2, (13.22%) highlighting the efforts made by libraries to promote user interaction, feedback mechanism and accessibility tools that encourage social inclusion and user satisfaction.

Usability & Accessibility and Design & Functionality rank fourth and fifth in Table 2. It was found to be lowest mean score (12.78% and 8.37%) and higher in standard deviation. Their importance for sustainability is inconsistent in web standards, accessibility compliance and optimization across libraries

Based on the library websites' iteration of parameters divided into three major categories, namely significant parameters, insignificant parameters, and sub-significant parameters.

From the analysis, it was noticed that out of 227 parameters, only 46 parameters are significantly used. The total score of 20.26% plays an important role in creating a library website indicated in Table 3, such as print resources, digital resources, Library Infrastructure, Browser Compatibility etc.

Table 3, the sub-significant values indicate that 44.05% of the parameters are used based on an adaptive situation for developing websites such as social media, accessibility standards, Assistive technologies, Research Gateway, open resources, software updating etc.

According to Table 3, the insignificant values indicate that 35.06% of the parameters are of less importance for developing websites, such as Infographics, AI Assistants, web policies etc.

Overall, Table 2&3, depicts the social sustainability (40.96%) shows the highest value among all 5 concepts in terms of institutional commitment towards inclusive user-centric approach and community engagement. Whereas,

environmental and economic scored (32.59% and 26.43%) that is need of hour to emphasis on green libraries with energy efficient design and cost-effective technologies.

From the table 4, it can be seen that Alagappa University, followed by Anna University dominates in Economic and Environmental sustainability parameters respectively. Bharathiar University tops in Social Sustainability followed by Alagappa University.

11 Conclusion

Libraries as knowledge hub plays vital role in advancing sustainability by providing seamless access to information, promoting resource sharing and adopting eco-friendly practices. Digital libraries are moving towards electronic services, digital repositories and green infrastructure to minimize environmental impact, lifelong learning and social inclusion. Initiatives such as digitization, energy efficient buildings and outreach programs contributes by reducing ecological footprints and creating knowledge societies in sustainable way. This paper suggests the libraries to perform well in socially sustainable while delivering content that strengthen the environmental and economic way by adopting eco-friendly web hosting, quick responsive design with less carbon emission of accessing and transferring data and digital preservation practice to achieve holistic sustainable environment. Universities ranked by agencies like NIRF, AICTE, NAAC etc. consider library infrastructure, services and website maintenance as one of the key parameters for evaluation based on WCAG and GIGW guidelines. Therefore, to score more universities should update maximum information and services at par with global standards.

Declaration of AI and AI-Assisted Technologies used in Manuscript Writing Process: Not Applicable

References

1. Ministry of Education. (2023). *All India Survey on Higher Education (AISHE) 2021–22 report* (Technical report). Government of India, Department of Higher Education. <http://aishe.gov.in>
2. World Wide Web Consortium. (2018). *Web Content Accessibility Guidelines (WCAG) 2.1* (Technical report). W3C. <https://www.w3.org/TR/WCAG21/>
3. Guidelines India. (2025). *Guidelines for Indian Government Websites (GIGW) 3.0* (Technical report). Ministry of Electronics & Information Technology <https://guidelines.india.gov.in/>
4. Chowdhury, G. (2014). Sustainability of digital libraries: A conceptual model and a research framework. *International Journal on Digital Libraries*, 14(3–4), 181–195. <https://doi.org/10.1007/s00799-014-0116-0>
5. Chowdhury, G. G. (2015). How to improve the sustainability of digital libraries and information Services? *Journal of the Association for Information Science and Technology*, 67(10), 2379–2391. <https://doi.org/10.1002/asi.23599>
6. Gajbhiye, C. K. (2024). Green Libraries and Sustainability Initiative. *International Journal of Innovative Science and Research Technology*, 9(12), 284–285. <https://www.ijisrt.com/assets/upload/files/IJISRT24DEC413.pdf>
7. Verma, P. K., Trivedi, S. K., & Biswas, S. (2025). Exploring trends and gaps in sustainable e-library research: A text mining analysis. *Digital Library Perspectives, ahead-of-print*. <https://doi.org/10.1108/DLP-07-2024-0110>
8. Kalpana, T. M., & Gopalakrishnan, S. (2013). Self-Sustainability in academic libraries in the digital era. In *Advances in library and information science (ALIS) book series* (pp. 47–67). <https://doi.org/10.4018/978-1-4666-4070-2.ch004>
9. Haridasan, S., & Firdaus, S. (2021). Application of web-based sources and services for sustainable development in university libraries. *Library Philosophy and Practice (e-journal)*, Article 4878. University of Nebraska–Lincoln. <https://digitalcommons.unl.edu/libphilprac/4878>
10. Ara, J., Sik-Lanyi, C., Kelemen, A., & Guzsvinecz, T. (2025). An inclusive framework for automated web content accessibility evaluation. *Universal Access in the Information Society*, 24(4), 1581–1607. <https://doi.org/10.1007/s10209-024-01164-5>
11. Asok, A. R. A., & Rekha, R. V. (2025). Digital inclusion in higher education: A web content accessibility evaluation of best Asian university library websites. *The Journal of Academic Librarianship*, 103120. <https://doi.org/10.1016/j.acalib.2025.103120>
12. Kumaran, M., & Subangi, M. C. (2017). Design and development of library website: A case study. *ScieXplore International Journal of Research in Science*, 4(2), 50. <https://doi.org/10.15613/sijrs/2017/v4i2/184637>
13. Kim, Y.-M. (2011). Factors affecting university library website design. *Information Technology and Libraries*, 30(3), 176–188. <https://doi.org/10.6017/ital.v30i3.1768>
14. Feng, Q., & Chen, H. (2011). Design a mobile website for university library. *Proceedings of the 2011 International Conference on Computer Science and Network Technology*, 2, 1191–1194. IEEE Xplore. <https://doi.org/10.1109/ICCSNT.2011.6015054>
15. Kaur, G., & Tiwari, R. G. (2023). Comparison and analysis of popular frontend frameworks and libraries: An evaluation of parameters for frontend web development. In *Proceedings of the 2023 4th International Conference on Electronics*

- and Sustainable Communication Systems (ICESC), 1067–1073. IEEE Xplore. <https://doi.org/10.1109/ICESC57686.2023.10192987>.
16. Ambika, C. A., & Ganesan, P. (2020). Content analysis of NAAC-accredited engineering college library websites of Bangalore, India. *International Journal of Scientific & Technology Research*, 9(4), 2629–2630.
 17. Ashrafi-Rizi, H., Nasajpour, M., Soleymani, M., Shahrzadi, L., & Hassanzadeh, A. (2014). Evaluation of the quality of the college library websites in Iranian medical universities based on the Stover model. *Journal of Education and Health Promotion*, 3(1), 17. <https://doi.org/10.4103/2277-9531.145931>
 18. Devi, R., & Kumar, S. (2023). Accessibility evaluation of selected library websites in Delhi-NCR using an automated tool. *Annals of Library and Information Studies*, 70, 215–225. <https://doi.org/10.56042/alis.v70i4.2814>
 19. Devi, Ksh. K. and Verma, M.K. (2018). Content analysis-based evaluation of library websites: a case study. *Annals of Library and Information Studies*, 65, p.239-251.
 20. Jangwal, K., Madhusudhan, M., & Soni, S. (2024). Beyond the pages: Exploring Web 2.0 proficiencies in LIS professionals. *International Journal of Library Information Networks and Knowledge*, 9(1), 82–85.
 21. Kumar, B. T. S., Raj, K. R. P., Naik, A. S., & Reddy, R. (2009). Content analysis of Indian Institute of Management library websites: An analytical study. In *7th International CALIBER 2009* (pp. 194–201). INFLIBNET Centre. <https://caliber.inflibnet.ac.in/caliber2009/CaliberPDF/25.pdf>
 22. Samrgandi, N. (2020). User experience of academic library websites. *International Journal of Computer Science and Mobile Computing*, 9(11), 57–69. <https://doi.org/10.47760/ijcsmc.2020.v09i11.006>
 23. Silvis, I. M., Bothma, T. J. D., & Beer, K. J. W. D. (2019). Evaluating the usability of the information architecture of academic library websites. *Library Hi Tech*, 37(3), 566–590. <https://doi.org/10.1108/LHT-07-2017-0151>
 24. Vinod Kumar and Bansal, J. (2014). Qualities of Library Website: Evaluating Library Websites of New IITs. *International Journal of Information Dissemination and Technology*, 4 (4), p.283-288.

**Policy, Ethics,
and Social Responsibility**

Assessing Perceptions of Library Professionals on Academic Integrity in Indian Educational Institutions

M. Manivannan ¹ [0000-0003-0805-3927] and Durga Sankar Rath ² [0000-0002-0951-3908]

¹ Central Library, Indian Institute of Technology Kharagpur, Kharagpur – 721302, West Bengal, India.
vannan.mm@gmail.com

² Dept. of Library and Information Science, Vidyasagar University, Midnapore – 721301., West Bengal, India
dsrath@mail.vidyasagar.ac.in

Abstract. Academic integrity is at the heart of higher education, and combating plagiarism is critical to maintaining the credibility of academic institutions. This study highlights the widespread efforts in Indian academia to address plagiarism, particularly through the adoption of robust institutional policies and plagiarism-detection systems. Descriptive research seeks to explore the level of plagiarism-checking services provided to students/research scholars in higher education institutions and universities in India. The research methodology outlines a systematic approach to collecting and analyzing survey data from library professionals across 186 higher education institutions in India, including institutes, universities, and colleges. The analysis of the survey data on academic integrity policies across Indian academic institutions highlights several key findings and reflects the positive impact of measures to curb plagiarism. The majority of library professionals surveyed agree that the incidence of plagiarism has decreased over the past 10 years. Nevertheless, HEIs must continuously improve policy adherence, especially in educational institutions, and raise awareness to achieve long-term success. By fostering a culture of integrity in their academic work, Indian educational institutions can make meaningful contributions to knowledge and meet global standards.

Keywords: Plagiarism, Academic Integrity, Higher Education, UGC, Research support

1 Introduction

The increasing significance of ethical issues, such as academic integrity, in nations like India marks the advancement of higher education. Due to online information and digital technology, institutions, universities, and colleges are struggling to implement strong policies that ensure a sustainable education. The University Grants Commission (UGC) has been at the forefront of this movement by enacting the Ensuring Academic Integrity in Higher Educational Institutions and Prevention of Plagiarism Regulations (2018) (UGC, 2018). These guidelines not only articulate a robust framework for defining, detecting, and penalizing plagiarism, but also mandate the use of institutional policy. ICT-based plagiarism detection tools, and the widespread education of students and faculty in the principles of ethical scholarship.

As the problem of plagiarism in academia and research continues to increase, addressing this issue requires an understanding of the organizational behavior of institutions. When it comes to the policies and practices of different educational institutions, there is a desire to foster an atmosphere of truthfulness and responsibility, which has led to the emergence of systems of academic integrity. Adhering to the guidelines set by the UGC and utilizing professional plagiarism detection software suggests that academic integrity can be somewhat variable (UGC, 2018). These elements are crucial within the policies of educational institutions, where they guide and ensure adherence to ethical standards among researchers in universities and colleges.

Although numerous technologies have been established to combat plagiarism on Indian educational institutions. A significant knowledge gap remains concerning how plagiarism prevention systems are implemented, perceived, and adopted by various stakeholders across India's educational institutions, including the college and university system. Most of the previous studies and articles reviewed (including policy and technical aspects) evaluated plagiarism detection systems in terms of their capabilities and applications. However, very few studies have examined the real problems associated with the implementation of Plagiarism Detection Software (PDS), including the perspectives of students and researchers directly affected by it. Additionally, there is a lack of empirical research on the viewpoints of library professionals. They are usually the "facilitators" of academic integrity and the gatekeepers to the academic resources that libraries procure.

There is also limited research on the extent to which national policies are implemented by individual institutions. The effectiveness of education and training initiatives on academic integrity. And whether the increase in enforcement of plagiarism policies at the educational institutional level has resulted in a decrease in plagiarism at the author and researcher level.

This research will employ a quantitative survey methodology targeting library professionals at higher education institutions (HEIs) across India. The results of this research will help to identify the relationship between practice and policy. It will provide insights into the strengths and weaknesses of current policies on plagiarism and its detection. It will also provide recommendations for action for educational administrators. This paper will be organized to include a review of relevant regulations and technological advancements in efforts to combat plagiarism. A detailed description of the research design and methods used to collect the data, a presentation of the results of the survey data, and finally, a discussion of the implications of these results in the context of larger discussions of academic integrity in India.

2 Literature Review

This portion of study explores the challenges of academic integrity, particularly plagiarism, within Indian academia. Many studies highlight the importance of stronger institutional oversight, enhanced peer review systems, and the implementation of academic integrity policies. Addressing these issues is essential for improving research quality and promoting a culture of academic honesty.

A challenge facing educational institutions such as HEIs and universities worldwide is defining and sustaining academic integrity. A universally accepted definition, the themes of academic integrity generally emphasize honesty, integrity, and responsibility within the academic community (Tripathi & Patel, 2021); academic integrity principles were introduced in recent years (Khositashvili, 2024). Bell (2018) conceptualizes plagiarism education in library learning as fostering prosocial behaviors among students, synthesizing literature from various academic subjects, and recommending best teaching practices that emphasize socializing students into academic cultures.

Pandita and Singh (2019) critically measure against plagiarism in higher educational institutions in India and emphasize the need for academic integrity. The study found that forced research or academic writing, which promotes undesirable practices such as plagiarism, the demand for ghostwriters, and predatory journals, thrives. Gunton (2022) examines a case study that emphasizes the vital role of academic librarians in managing academic integrity strategies, as well as an online learning activity in Australia's education sector. Promoting awareness of ethical behavior while supporting learning, teaching, and research beyond their traditional roles. Prashar et al. (2024) address academic dishonesty among higher education students (294) in medical, education, psychology, and social science research fields with the help of a longitudinal quasi-experimental study and finds unintentional plagiarism is prevalent due to limited ethical awareness.

Academic plagiarism is increasingly becoming a challenge to academic integrity worldwide (Nabee et al., 2020), unintentional plagiarism (El Bairi et al., 2024) is still on the rise, particularly among postgraduate international students at Australian universities, English writing skills are insufficient to avoid unintentional plagiarism in a new academic setting (Fatemi & Saito, 2020). Nabee et al. (2020) investigate academic plagiarism among students at South African universities and assess students' perceptions and predictors of plagiarism practices. Findings show a high understanding of plagiarism but moderate plagiarism levels and recommend continuous plagiarism avoidance training. On the contrary, (Krokosz & Ferreira, 2019) say that citation training does not prevent plagiarism practices effectively. Krokosz's study examines students' perceptions of plagiarism with a sample of 3623 students from the University of Sao Paulo, and findings reveal a gap between theoretical and practical knowledge of plagiarism. Smedley et al. (2020) find knowledge of plagiarism improved, but the students struggled with paraphrasing skills among nursing students. Awareness and knowledge about academic integrity is growing among Babasaheb Bhimrao Ambedkar university research scholars and students. The university library supports activities against academic integrity breaches (Singh & Rai, 2021).

In a related study, Alua et al. (2023) examine students' understanding of plagiarism and assess the impact of PDS (Turnitin) on plagiarism avoidance. Many students reported challenges, particularly in interpreting Turnitin originality reports. Meanwhile, Pradhan and Kumar (2023) investigate plagiarism in the research reports of PHD students. They identify the reasons for plagiarism and propose an alternative action plan. Their mixed-method study findings reveal a general prevalence of plagiarism and its causes. The researchers recommend training and awareness programs, establishing ethics committees, and implementing anti-plagiarism software for the academic community. The literature review highlights the complex issues of research quality and academic integrity in the Indian academic context. Plagiarism, exacerbated by structural factors such as ineffective performance appraisal criteria and insufficient regulatory frameworks, has been a consistent theme across studies. To address these concerns, critical approaches include implementing plagiarism-detection methods and enforcing UGC anti-plagiarism policies. Additionally, organizational monitoring, strong peer review processes, and a cultural shift towards authentic information sharing are essential for ensuring long-term improvements.

3 Objective

The goal is to evaluate the role of plagiarism detection systems in upholding academic integrity. This evaluation also includes examining views on the plagiarism control measures instituted in India and assessing opinions on the allowable limits of the similarity index prescribed by the UGC for theses, dissertations, or other academic papers.

- To explore the effectiveness and impact of the plagiarism detection system.
- To understand opinions on the regulations for plagiarism implemented in India.
- To understand opinions on the similarity index level set by UGC to theses/dissertations and papers/articles.
- To understand the benefits towards fundamental principles of writing a thesis and dissertation use of plagiarism detection tools/software.
- To understand the opinions of library professionals on whether plagiarism has decreased in the past ten years.

4 Research Gap

Although many institutions of higher education deploy plagiarism detection systems and procedures and, in some cases, participate in addressing the problem of plagiarism, there seems to be a disparity in the scholarship on the appropriate evaluation of such action and the application of the existing intellectual property laws. Even as scholarly work on these mechanisms has been undertaken, various literature reviews outlining their functionality have already existed. However, how they are ergonomically used by students and in academic writing in India is scanty. The opinion towards plagiarism and its regulation and the prescribed similarity index level by the University Grants Commission (UGC) has not been thoroughly studied. Knowing library professionals is essential for assessing the acceptance and application of these regulations.

Lastly, shared knowledge and some literary works do not affirm that plagiarism cases have decreased in the past ten years; well-structured, thorough qualitative research is absent. Looking at the magnitude of this drop and the reasons for any observed change has not been practiced to the fullest. Addressing these gaps might help policymakers, educators, and students improve academic practice and integrity in Indian higher education.

5 Research Methodology

The study will employ a quantitative approach to collect and analyze survey data from library professionals. A structured questionnaire, meticulously designed as the primary data collection instrument, will be utilized for this research. This instrument draws on the work of Jilani and Ahmad (2021), with significant modifications for this study. We will gather primary data by distributing this structured survey to library professionals at higher education institutions and universities across India. The questionnaire will be developed through a thorough literature review and expert consultations to ensure validity and reliability. Participants will be informed of the study's purpose, the voluntary nature of their participation, and the confidentiality of their responses. All collected data will be kept confidential and anonymized to protect the identities of participants.

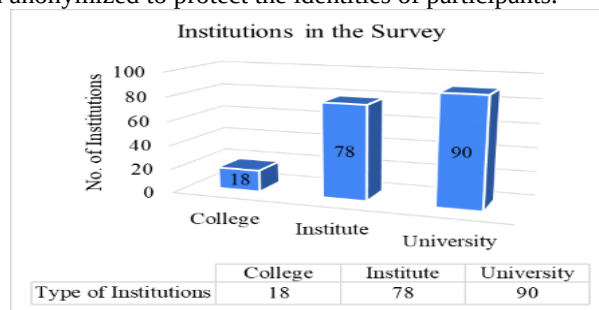


Fig. 1. Survey Participants

The survey is administered over a defined period to allow adequate response time while ensuring timely data collection. A starting e-mail with the questionnaire's link was sent to library professionals in February 2024, and a further three follow-up reminder e-mails were sent to individuals who had surveys in progress or had not yet

started responding. The designated survey open period was February 2024 through July 2024. The findings interpret common themes, trends, and variations among library professionals across different types of institutions. Findings may primarily reflect the perspectives of library professionals in higher education institutes in India and may not be generalizable to other contexts. A total of 281 questionnaires has been sent to randomly selected library professionals working in HEIs across India. This research methodology outlines a systematic approach to collecting and analyzing survey data from library professionals in 186 (66%, n = 281) (refer Figure 1) higher education institutes, universities, and colleges in India.

6 UGC Regulations 2018

The University Grants Commission (Promotion of Academic Integrity and Prevention of Plagiarism in Higher Education Institutions) Regulations, 2018 (UGC, 2018) recommend institutional mechanisms to eliminate plagiarism and maintain academic integrity in higher education institutions (HEIs) in India. Any research and academic work carried out by a researcher or a student reflects the degree to which originality and academic integrity are embraced within the HEI's academic principles. The UGC recognizes plagiarism as a critical issue threatening academic work and research credibility in educational institutions and universities. As part of the solution, educational institutions and universities must implement a mechanism or ICT-based tools to detect plagiarism and promote awareness among stakeholders in the educational system. The mandate of the UGC ensures that academic works and outputs adhere to academic integrity and intellectual originality. These regulations define plagiarism as an unethical academic practice that involves using someone else's work without proper citation. To combat this, higher education institutions should encourage using plagiarism detection software, develop clear policies on plagiarism, and conduct training on proper citation practices.

The regulations categorize four levels of plagiarism in non-core areas, such as up to 10%, 10%-40%, 40%-60%, and 60%, above similarities found. For non-core areas of academic work, the Academic Misconduct Panel (AMP) shall be asked to resubmit the articles/papers within six months for the similarities 10% - 40%, to resubmit the articles/papers within one year but not exceeding eighteen months for the similarities found 40% - 60%, and registration for that course to be cancelled for the similarities found above 60%. Penalties for researcher and faculty of HEI: AMP bans the research scholar or faculty from publishing any work for one year, two years and three years if the similarity index of the manuscript exceeds 10%-40%, 40%-60% and 60%, respectively. The regulations strictly adhere to a zero-tolerance policy in the core areas of the research work, including the abstract, summary, hypothesis, observations, results, conclusions, and recommendations.

In addition, the regulations emphasize the importance of regular awareness training, seminars, and workshops to educate research scholars and academicians about academic integrity. This regulatory framework promotes ethical academic practices and helps prevent unethical behaviors such as plagiarism, ultimately raising the standard of higher education in India.

7 Result Analysis

7.1 Policy on Academic Integrity/Plagiarism

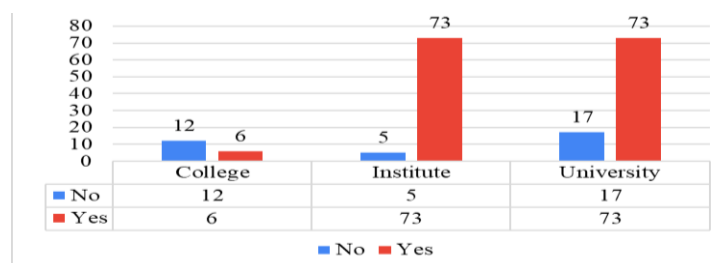


Fig. 2. Policy on Academic Integrity/Plagiarism

In Figure 2, Survey responses to the question "Does your institute follow any Policy on Academic Integrity/Plagiarism?" were collected from library professionals across institutions. The data reveals that 152 out of 186 educational institutions have policies on academic integrity or plagiarism, showing 82% (n=186) adherence. This includes colleges (6), universities (73) and institutes (73). Only 18% of institutions do not follow or have any

policy on academic integrity/plagiarism. The finding indicates that educational institutes and universities maintain their standards than colleges.

7.2 Adoption of the UGC Regulations 2018

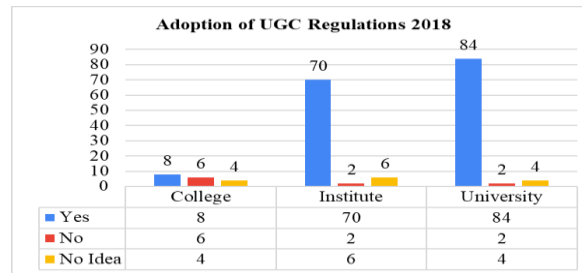


Fig. 3. Adoption of UGC Regulations 2018

In Figure 3, the survey results indicate that 87% (n = 186) of educational institutions have implemented the UGC Regulations 2018 regarding academic dishonesty nationwide. Only 8% of respondents indicated that they have no knowledge of the regulations' adoption. This result shows that institutes and universities are more compliant with the UGC regulations than colleges.

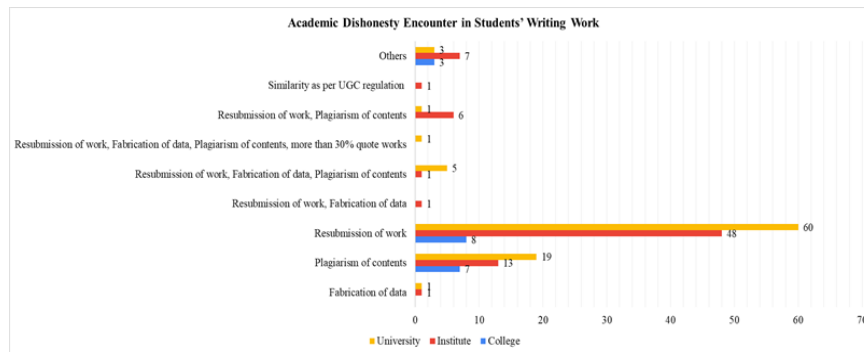


Fig. 4. Academic Dishonesty Encounter in Students' Writing Work

7.3 Academic Dishonesty Encounter in Students' Writing Work

This result highlights various forms of academic dishonesty in academic writing work across educational institutions. The most common form is reporting high numbers of respondents, such as resubmissions of work, from universities (60), institutes (48), and colleges (8). Text similarity is also notable, particularly in universities (19), institutes (13), and colleges (7) (see Figure 4). Less common forms of dishonesty are in lower numbers, such as similarity as per UGC regulations, data fabrication, and resubmission of work with fabrication or plagiarism.

7.4 Plagiarism by Student/Research Scholars

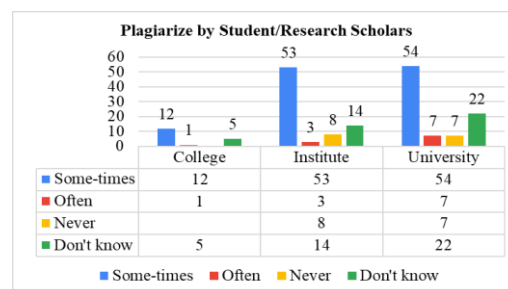


Fig. 5. Plagiarize by Student/Research Scholars

In Figure 5, the data show the frequency of plagiarism among students and research scholars from various institutes, universities, and colleges. The data reveal that 64% (n = 186) of the respondents admit to "sometimes" engaging in plagiarism, indicating a significant number. The proportion of those who often plagiarize is relatively low, with only 11 respondents, which is 6%, and the option "never plagiarize" is 8%. A significant number of respondents are unsure whether they plagiarize, with 22% unsure. This result suggests occasional plagiarism is common, but frequent and intense denial remains less prevalent.

7.5 Possibilities to Avoid Plagiarism

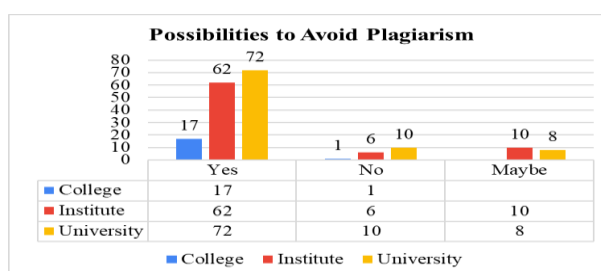


Fig. 6. Possibilities to Avoid Plagiarism

Figure 6 highlights perceptions of their ability to avoid plagiarism, with 81% (n=186) of respondents believing they can. Of the respondents, 9% agreed that avoiding plagiarism is impossible, and 10% were uncertain of the conditions. This result shows that most library professionals believe it is possible to avoid plagiarism.

7.6 Decreasing Plagiarism in the Past Ten Years

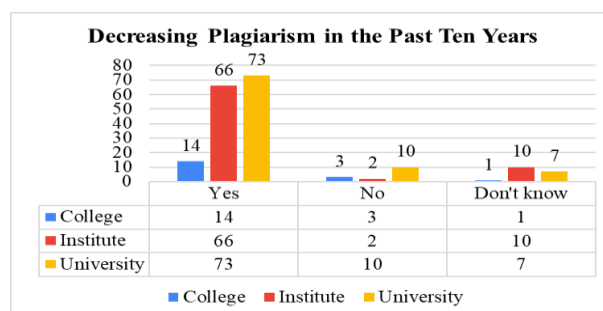


Fig. 7. Decreasing Plagiarism in the Past Ten Years

In Figure 7, the result reflects library professionals' perceptions of whether plagiarism has declined over the past decade. Most 82% (n=186) of respondents believe that plagiarism has decreased. A few respondents believe that plagiarism has not decreased; only 8% and 10% are unaware. This result shows that most respondents feel plagiarism has decreased over the last ten years.

7.7 Benefits of Plagiarism Detection Software/Tool

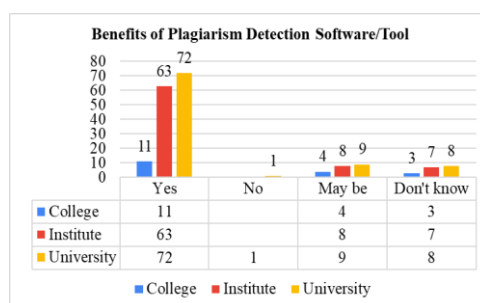


Fig. 8. Benefits of PDS Tool

In Figure 8, most educational institutions (79%; n=186) accept the overwhelming consensus on the advantages of using plagiarism-detection software. This result shows that the majority, especially institutes and universities, recognize the significant benefits of plagiarism detection software. However, there remains some uncertainty or lack of awareness in some cases.

7.8 Impact of Implementing the Plagiarism Detection System

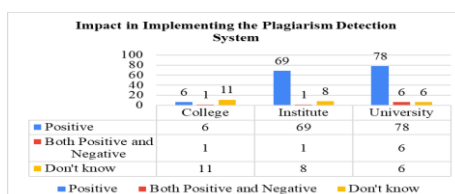


Fig. 9. Impact of Implementing the PDS

In Figure 9, in educational institutions, 82% (n=186) predominantly accept the positive impact of implementing a plagiarism detection system. A small number of 8 respondents accept both positive and negative impacts, and only 25 respondents (13%) are unaware of the impact of implementing the plagiarism detection system. Implementing plagiarism detection systems is seen as having a strong positive impact, particularly in institutes and universities.

7.9 Similarity Index Level would be in Theses/Dissertations (Opinion)

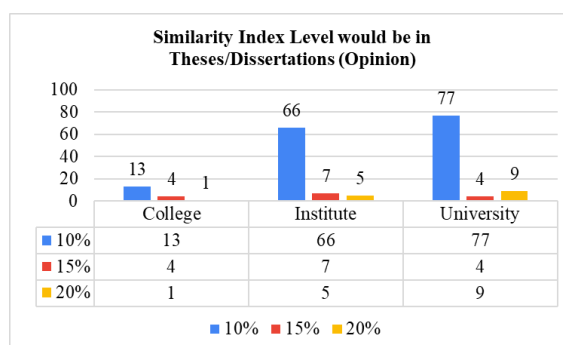


Fig. 10. Similarity Index Level would be in Theses/Dissertations (Opinion)

An opinion of library professionals on the perceived acceptable similarity index level is that 156 out of 186 respondents (see Figure 10) (84%) show interest in the 10% similarity index level for research theses/dissertations. Only 12 and 15 respondents show interest in the 15% and 20% similarity index levels, respectively. This result shows a strong preference for maintaining a low similarity index across research theses/dissertations at all academic institutions, emphasizing the importance of originality in research work.

7.10 Similarity Index Level would be in Research Papers/Articles (Opinion)

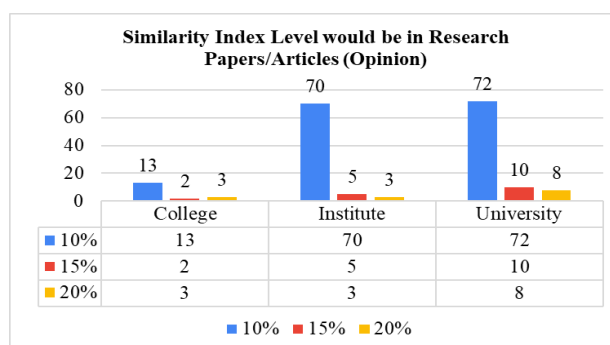


Fig. 11. Similarity Index Level would be in Research Papers/Articles (Opinion)

In Figure 11, library professionals' opinions on the acceptable similarity index level indicate that 155 out of 186 respondents (83%) are interested in a 10% similarity index for research papers/articles. A few respondents, 17 and 14, show interest in the 15% and 20% similarity index levels, respectively. This result shows a strong preference for maintaining a low similarity index in the research theses/dissertations, particularly in institutes and universities, indicating a commitment to original research.

8 Findings

Several important findings are established by conducting the survey data analysis to identify library professionals' perceptions on academic integrity policies within Indian educational institutions.

- While most (82%) universities and institute have academic integrity or plagiarism policies, colleges still trail.
- Additionally, 87% (n=186) of the respondents confirm that their institutions have adopted the UGC Regulations 2018 related to academic dishonesty.
- The analysis identifies high levels of research students' resubmission of writing, both in institutes and universities. The highest rate of reported plagiarism was occasional plagiarism, with 64% (n = 186) of respondents acknowledging it.
- The responses from this study indicate that there is strong evidence of the positive impacts of measures to reduce plagiarism. Most (81%, n = 186) of the library professionals believe that plagiarism could be prevented. And, 82% (n=186) agreed that plagiarism rates have decreased over the last decade.
- Nearly all of the (94%, n=186) educational institutions provide free plagiarism detection services through their library and the services are offered at no charge to students. And, 79% of the respondents acknowledge the value of utilizing plagiarism detection tools.
- Most (82%, n=186) of the respondents also agreed that implementing plagiarism detection systems has a positive effect on the reduction of plagiarism, especially in institutes and universities.
- Nearly 83% (n=186) of the library professionals responded that they would accept a 10% threshold as an acceptable percentage for theses and research articles, demonstrating an institutional-wide commitment to originality in academic research.

9 Conclusion

By methodically collecting the diverse experiences and viewpoints of Library Professionals, this study provides new evidence to the Academic Integrity discussion in India. Institutional administrators can gain practical insights from the findings to improve existing frameworks, address gaps in policy execution and invest in long-term integrity building programs. In conclusion, this study advocates for an inclusive, collaborative effort to ensure that ethics in academic background will continue to be a fundamental component of India's Higher Education, which is experiencing rapid growth.

Declaration of AI and AI-Assisted Technologies used in Manuscript Writing Process: I would like to acknowledge the use of Grammarly and QuillBot, which assisted in checking and improving the grammar and

clarity of this document. Its feedback and suggestions were instrumental in refining the language to ensure it is both accurate and professionally presented.

References

1. Alua, M. A., Asiedu, N. K., & Bumbie-Chi, D. M. (2023). Students' Perception on Plagiarism and Usage of Turnitin Anti-Plagiarism Software: The Role of the Library. In *Journal of Library Administration* (Vol. 63, Issue 1, pp. 119–136). <https://doi.org/10.1080/01930826.2022.2146445>
2. Basic, Z., Kruzic, I., Jerkovic, I., Buljan, I., & Marušić, A. (2019). Attitudes and Knowledge About Plagiarism Among University Students: Cross-Sectional Survey at the University of Split, Croatia. In *Science and Engineering Ethics* (Vol. 25, Issue 5, pp. 1467–1483). <https://doi.org/10.1007/s11948-018-0073-x>
3. Bell, S. (2018). Addressing student plagiarism from the library learning commons. In *Information and Learning Science* (Vol. 119, Issues 3–4, pp. 203–214). <https://doi.org/10.1108/ILS-10-2017-0105>
4. Chokoraia, T. (2023). The problem of plagiarism in the process of establishing academic integrity culture at Georgian higher educational institutions. In *Quality in Higher Education* (Vol. 29, Issue 1, pp. 86–101). <https://doi.org/10.1080/13538322.2022.2100619>
5. Chomintra, M. (2023). Assessing the use of critical literacies in mis/disinformation literacy instruction. *The Journal of Academic Librarianship*, 49(5), 102762. <https://doi.org/10.1016/j.acalib.2023.102762>
6. El Bairi, K., El Kadmiri, N., & Fourtassi, M. (2024). Exploring scientific misconduct in Morocco based on an analysis of plagiarism perception in a cohort of 1,220 researchers and students. In *Accountability in Research* (Vol. 31, Issue 2, pp. 138–157). <https://doi.org/10.1080/08989621.2022.2110866>
7. Fatemi, G., & Saito, E. (2020). Unintentional plagiarism and academic integrity: The challenges and needs of postgraduate international students in Australia. In *Journal of Further and Higher Education* (Vol. 44, Issue 10, pp. 1305–1319). <https://doi.org/10.1080/0309877X.2019.1683521>
8. Gunton, L. (2022). Librarians as Drivers of Academic Integrity for Student Success at University. In *Journal of the Australian Library and Information Association* (Vol. 71, Issue 2, pp. 156–170). <https://doi.org/10.1080/24750158.2022.2063101>
9. Jilani, G., & Ahmad, N. (2021). Role of Librarians and Information Scientists in Plagiarism Control A Study of NIRF Ranked Engineering Institutions Ranked in 2020. *DESIDOC Journal of Library & Information Technology*, 41(03), 206–212. <https://doi.org/10.14429/djlit.41.03.16716>
10. Kampa, R. K., Padhan, D. K., & Ahmad, F. (2020). Perceptions of students and researchers of food technology towards plagiarism: A case study. In *DESIDOC Journal of Library and Information Technology* (Vol. 40, Issue 6, pp. 369–374). Defense Scientific Information and Documentation Centre. <https://doi.org/10.14429/djlit.40.6.16036>
11. Krokosz, M., & Ferreira, S. M. S. P. (2019). Perceptions of graduate students at the University of São Paulo about plagiarism practices in academic works. In *Anais da Academia Brasileira de Ciencias* (Vol. 91, Issue 2, pp. 1–17). <https://doi.org/10.1590/0001-3765201920180196>
12. Nabee, S. G., Mageto, J., & Pisa, N. (2020). Investigating predictors of academic plagiarism among university students. In *International Journal of Learning, Teaching and Educational Research* (Vol. 19, Issue 12, pp. 264–280). <https://doi.org/10.26803/ijlter.19.12.14>
13. Pandita, R., & Singh, S. (2019). Regulations to Prevent Plagiarism in Higher Education in India A Critical Appraisal. *DESIDOC Journal of Library & Information Technology*, 39(2), 67–73. <https://doi.org/10.14429/djlit.39.2.14115>
14. Pradhan, T. K., & Kumar, A. (2023). Plagiarism in the Research Reports of Indian Doctoral Students: Causes and a Remedial Action Plan. In *Center for Educational Policy Studies Journal* (Vol. 13, Issue 3, pp. 185–212). University of Ljubljana. <https://doi.org/10.26529/cepsj.1494>
15. Prashar, A., Gupta, P., & Dwivedi, Y. K. (2024). Plagiarism awareness efforts, students' ethical judgment and behaviors: A longitudinal experiment study on ethical nuances of plagiarism in higher education. In *Studies in Higher Education* (Vol. 49, Issue 6, pp. 929–955). <https://doi.org/10.1080/03075079.2023.2253835>
16. Singh, M. P., & Rai, S. (2021). Knowledge, awareness and attitude towards academic integrity among the scholarly society of Babasaheb Bhimrao Ambedkar University, Lucknow: A study. In *Library Philosophy and Practice* (Vol. 2021, pp. 1–16).
17. Smedley, A., Crawford, T., & Cloete, L. (2020). An Evaluation of an Extended Intervention to Reduce Plagiarism in Bachelor of Nursing Students. In *Nursing Education Perspectives* (Vol. 41, Issue 2, pp. 106–108). <https://doi.org/10.1097/01.NEP.0000000000000492>
18. Issue 4, pp. 281–304). <https://doi.org/10.1080/08989621.2022.2120390>
19. Tripathi, S., & Patel, D. (2021). Academic integrity: Looking beyond plagiarism. *Annals of Library and Information Studies*, 68(2), 190–197. Scopus.
20. University Grants Commission. (2018). *Promotion of Academic Integrity and Prevention of Plagiarism in Higher Educational Institutions Regulations*. Government of India. https://www.ugc.ac.in/pdfnews/7771545_academic-integrity-Regulation2018.pdf

Privacy and Data Ethics in Academic Libraries: Policy Frameworks, Ethical Challenges, and Social Responsibility in the Digital Age

Kiran Babu Shewale ^[0009-0001-6478-3863]

Librarian, Amrutvahini College of Pharmacy, Sangamner,
Affiliated to Savitribai Phule Pune University, Pune, Maharashtra, India

kiranshewale89@gmail.com

Abstract. The rapid digitization of academic libraries has transformed traditional information services into data-intensive, technology-mediated environments. While digital platforms, analytics, and artificial intelligence have enhanced service personalization and efficiency, they have simultaneously intensified concerns related to user privacy, data ethics, and institutional accountability. Libraries, historically regarded as trusted custodians of information and defenders of intellectual freedom, now face complex ethical dilemmas involving user data collection, surveillance risks, third-party vendor practices, and algorithmic decision-making.

This paper examines privacy and data ethics in academic libraries through the lens of policy frameworks, ethical principles, and social responsibility. Using a systematic literature-based analytical approach, the study reviews international ethical standards, library privacy principles, and emerging data governance practices relevant to the Library and Information Science (LIS) domain. It further analyzes the implications of digital infrastructures, licensed electronic resources, and data-driven services on user autonomy and confidentiality.

The paper proposes a conceptual framework that integrates privacy-by-design, ethical data stewardship, and institutional responsibility as core components of contemporary library governance. The study argues that embedding ethical data practices into library policies, professional training, and service design is essential for sustaining user trust and reinforcing the social mission of libraries in the digital age. The findings contribute to ongoing discussions on responsible innovation in LIS and offer policy-oriented recommendations for academic libraries navigating evolving technological and ethical landscapes.

Keywords: Privacy, Data Ethics, Academic Libraries, Information Policy, User Data Protection, Social Responsibility.

1 Introduction

Academic libraries have historically been regarded as socially responsible institutions committed to intellectual freedom, equitable access to information, and the protection of user privacy. Privacy has been recognized as a foundational ethical principle in librarianship because unrestricted access to information is possible only when users are free from surveillance and monitoring [2]. In traditional print-based environments, privacy concerns were largely limited to the confidentiality of circulation records and reference interactions.

The rapid digitization of academic libraries has fundamentally transformed this context. Contemporary libraries operate within complex digital ecosystems that include electronic journals, discovery platforms, authentication systems, learning analytics tools, and artificial intelligence-driven services. These systems routinely collect, process, and store detailed user data such as search histories, access logs, and usage patterns [9]. While such data are often used to improve service quality and personalization, they simultaneously introduce ethical risks related to surveillance, data misuse, and loss of user autonomy.

International professional bodies such as International Federation of Library Associations and Institutions have consistently emphasized that privacy is essential to intellectual freedom and democratic participation. According to [8], the erosion of privacy in library environments undermines users' ability to explore controversial or sensitive topics freely. Similarly, scholars argue that digital surveillance can create a "chilling effect," discouraging users from seeking information openly [15].

The growing reliance on third-party vendors for electronic resources and analytics platforms further complicates privacy governance in academic libraries. Vendor-controlled systems often operate under commercial data practices that may conflict with library ethics, reducing institutional control over user data [12]. At the same time, emerging data protection laws and ethical AI frameworks require libraries to reassess their policies and responsibilities in managing user information responsibly.

Despite increasing awareness, several studies indicate that many academic libraries lack comprehensive data governance policies that explicitly address consent, transparency, data minimization, and ethical accountability [3]. This paper addresses this gap by examining privacy and data ethics in academic libraries through the lenses of policy frameworks, ethical challenges, and social responsibility in the digital age.

2 Research Methodology

This study adopts a qualitative, conceptual, and literature-based analytical approach. Relevant scholarly literature, ethical guidelines, and policy frameworks in the LIS domain were systematically reviewed to identify key themes related to privacy, data ethics, and social responsibility in academic libraries. Insights from these sources were synthesized to develop the proposed Privacy–Data Ethics–Social Responsibility (PDSR) conceptual framework [6].

3 Review of Literature: Privacy and Data Ethics in LIS

3.1 Privacy as a Core Library Value

Privacy has long been embedded in the ethical foundations of librarianship. The American Library Association defines privacy as the right of individuals to control access to their personal information and emphasizes its importance in protecting intellectual freedom [2]. This position is reinforced by IFLA, which asserts that confidentiality of library use is a prerequisite for the free exercise of the right to information [8].

Scholarly literature supports this normative stance. Stachokas [12] argues that privacy is context-dependent and that violations occur when information flows deviate from socially accepted norms. In library contexts, such deviations can arise when digital systems collect user data beyond what is necessary for service provision. Researchers have highlighted that while libraries traditionally resisted surveillance practices, digital infrastructures have weakened their direct control over data [15].

3.2 Data Ethics in Academic Libraries

Data ethics extends beyond legal compliance to address moral responsibilities associated with data practices. [7] define data ethics as the study of ethical issues related to data generation, processing, and governance, including concerns of fairness, transparency, and accountability. In academic libraries, data ethics becomes particularly relevant in areas such as learning analytics, recommender systems, and usage-based assessment tools.

Jones and Salo [10] emphasize that libraries increasingly rely on user data to demonstrate value and impact, often without fully considering the ethical implications of such practices. Studies show that users are frequently unaware of the extent to which their data are collected and analyzed, raising concerns about informed consent and power imbalances between institutions and users [3].

The ethical use of learning analytics in academic libraries has been addressed by the Association of College and Research Libraries through its Learning Analytics Toolkit. The toolkit emphasizes responsible engagement with learning analytics by highlighting principles such as transparency in data collection, informed user consent, protection of student privacy, and alignment with core library values. It serves as a practical guide for academic librarians to balance data-driven decision-making with ethical responsibility and user trust [1].

3.3 Surveillance and Vendor-Controlled Platforms

A significant body of LIS research has examined the ethical challenges posed by third-party vendors. Stachokas [12] documents how licensed electronic resources often engage in extensive data collection, including tracking user behavior across platforms. Such practices can conflict with core library values of confidentiality and user trust.

Zimmer [15] warns that libraries risk becoming “privacy facilitators” for commercial surveillance if they fail to critically evaluate vendor technologies. Scholars argue that ethical responsibility does not end at procurement but extends to ongoing assessment of vendor policies, contract negotiations, and advocacy for privacy-preserving technologies [9].

3.4 Social Responsibility and Ethical Governance

From a social responsibility perspective, libraries are ethical actors within democratic societies. UNESCO [14] emphasizes that access to information must be balanced with respect for human rights, including privacy and dignity. LIS scholars argue that protecting user data is inseparable from libraries' commitments to inclusivity, social justice, and trust-building [8].

Recent literature highlights the need to integrate data ethics into LIS education and professional training. According to Cox et al. [5], ethical literacy is becoming as essential as technical competence for librarians working in data-intensive environments. Ethical governance frameworks thus position librarians as mediators who balance innovation with moral responsibility.

4 Conceptual Framework: Privacy–Data Ethics–Social Responsibility Model

To address the complex ethical challenges of data-driven library environments, this paper proposes a Privacy–Data Ethics–Social Responsibility (PDSR) Framework for academic libraries. The framework conceptualizes privacy protection not as an isolated technical issue but as an ethical and social responsibility embedded across the entire data lifecycle.

At the core of the framework lies user data, encompassing personal identifiers, usage logs, search histories, and research behavior. Surrounding this core are three interdependent dimensions: privacy, data ethics, and social responsibility. Privacy focuses on confidentiality, consent, and data minimization, ensuring that user information is collected only to the extent necessary for service provision [2]. Data ethics emphasizes transparency, fairness, accountability, and respect for user autonomy in data practices [6]. Social responsibility situates library data governance within broader commitments to intellectual freedom, equity, and democratic participation [8].

The framework aligns with the principle of privacy-by-design, which advocates embedding privacy safeguards at the earliest stages of system design rather than treating them as post hoc compliance measures [4]. In academic libraries, this implies ethical evaluation of technologies before adoption, continuous assessment of vendor practices, and active communication with users about data use. By integrating ethical reasoning with institutional accountability, the PDSR framework offers a holistic approach to responsible data governance in LIS.

The proposed PDSR framework illustrates how user data governance in academic libraries progresses through interconnected layers of privacy, data ethics, and social responsibility, leading to ethical library governance [2, 8].

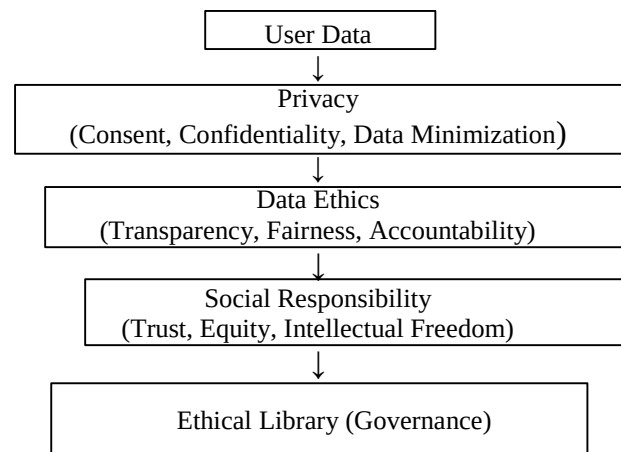


Fig. 1. Privacy–Data Ethics–Social Responsibility (PDSR) Framework for Academic Libraries

5 Privacy Challenges in Academic Libraries

5.1 Data Collection and User Surveillance

Academic libraries increasingly rely on digital platforms that generate extensive user data. Authentication systems, discovery tools, and analytics software routinely capture granular information about user behavior, often without explicit user awareness (Jones, 2020). Such practices raise concerns about surveillance and the erosion of

informational privacy, particularly when data are retained for extended periods or repurposed for assessment and reporting.

Zimmer [15] argues that even well-intentioned data collection can produce a “chilling effect,” discouraging users from exploring sensitive or controversial topics. This is particularly problematic in academic contexts, where intellectual exploration is central to teaching and research.

5.2 Third-Party Vendors and Loss of Control

One of the most significant privacy challenges arises from libraries’ dependence on third-party vendors for electronic resources and platforms. Vendor systems often collect data beyond the library’s direct control and may operate under commercial privacy policies that conflict with professional library ethics [12]. Studies show that user data may be shared across platforms or used for profiling and marketing purposes, undermining user trust [9].

Libraries frequently lack the bargaining power to negotiate privacy-friendly contracts, resulting in ethical tensions between access to information and protection of user rights.

5.3 AI, Analytics, and Algorithmic Opacity

The adoption of AI and machine learning technologies introduces additional ethical complexities. Recommendation systems, predictive analytics, and learning analytics tools rely on large datasets and opaque algorithms. Floridi et al. [6] caution that algorithmic systems can perpetuate bias, reduce transparency, and shift decision-making away from human oversight. In library contexts, such systems may influence what information users see, raising ethical questions about neutrality, fairness, and accountability.

6 Data Ethics and Policy Landscape

6.1 Ethical Principles in LIS

Ethical data governance in libraries is grounded in principles such as informed consent, transparency, proportionality, and accountability. The American Library Association emphasizes that libraries should clearly inform users about what data are collected and how they are used [2]. Similarly, [8] stresses that ethical responsibility extends to advocating for user privacy even when external pressures encourage data exploitation.

Floridi and Taddeo [6] highlight that ethical data practices require proactive moral reasoning rather than mere legal compliance. For libraries, this means evaluating not only what is lawful but also what is ethically justifiable.

The NISO Consensus Principles on Users’ Digital Privacy provide a comprehensive framework for ethical data governance in library, publisher, and software provider systems. These principles stress transparency, data minimization, user control, security safeguards, and accountability in handling user data. For academic libraries, the NISO principles function as a key resource for developing effective data management strategies that protect users’ digital privacy while supporting legitimate analytical and service needs [11].

6.2 Global and Indian Policy Context

Globally, data protection frameworks such as the European Union’s General Data Protection Regulation (GDPR) have reshaped expectations around consent, data minimization, and user rights. Scholars note that GDPR has influenced library data practices worldwide by establishing privacy as a default standard [13].

In the Indian context, the Digital Personal Data Protection Act, 2023 represents a significant step toward formalizing data protection obligations. While not library-specific, the Act emphasizes lawful processing, purpose limitation, and accountability, all of which are directly relevant to academic libraries. Researchers argue that Indian libraries must translate these legal requirements into sector-specific policies and ethical guidelines aligned with LIS values [13].

7 Libraries as Socially Responsible Data Stewards

Academic libraries occupy a unique position as trusted intermediaries between users and information systems. Their social responsibility extends beyond service efficiency to the protection of fundamental rights such as privacy and intellectual freedom. UNESCO [14] underscores that access to information must be balanced with respect for human dignity and human rights.

By adopting ethical data governance frameworks, libraries can reinforce public trust and demonstrate leadership in responsible innovation. Cox et al. [5] argue that librarians increasingly function as ethical mediators who translate abstract principles into practical policies and everyday decision-making. Training in data ethics, transparency in communication, and user-centered policy design are therefore essential components of socially responsible librarianship.

7.1 Role of Key Stakeholders

Ethical data governance in academic libraries involves multiple stakeholders. Library users are entitled to transparency, informed consent, and protection of their privacy. Library staff are responsible for implementing ethical policies, evaluating technologies, and ensuring responsible data practices. Vendors play a critical role by designing systems that support privacy, transparency, and accountability. Collaboration among these stakeholders is essential for effective and ethical data governance in academic libraries. [12, 5].

8 Best Practices and Policy Recommendations for Ethical Data Governance in Academic Libraries

Academic libraries must move beyond reactive compliance models and adopt proactive, ethically grounded data governance strategies. Based on the reviewed literature and policy frameworks, this study proposes the following best practices to strengthen privacy protection and ethical responsibility in academic libraries.

8.1 Privacy-by-Design in Library Systems

Embedding privacy safeguards at the design stage of library systems is critical for ethical data governance. Privacy-by-design emphasizes minimizing data collection, limiting retention periods, and ensuring secure data storage from the outset [4]. Libraries should evaluate digital platforms and tools for privacy implications prior to adoption and avoid systems that rely on excessive user tracking or opaque data practices.

Such an approach aligns with professional library ethics, which prioritize user confidentiality over operational convenience [2]. Ethical technology assessment committees within libraries can play a key role in institutionalizing privacy-conscious decision-making.

8.2 Transparency and Informed Consent

Transparency is a foundational ethical principle in responsible data stewardship. Academic libraries should clearly communicate what types of user data are collected, the purposes of data processing, and the duration of data retention. Research indicates that lack of transparency undermines user trust and weakens informed consent [3].

Libraries should adopt plain-language privacy notices and ensure that consent mechanisms are meaningful rather than symbolic. Such practices are consistent with international ethical guidelines that emphasize user autonomy and accountability [8].

8.3 Ethical Management of Vendor Relationships

Given the pervasive role of third-party vendors, libraries must extend ethical scrutiny beyond institutional boundaries. Contractual agreements with vendors should include explicit privacy clauses, restrictions on secondary data use, and clear accountability mechanisms [12]. Scholars argue that libraries have a moral obligation to advocate for privacy-respecting technologies, even when market options are limited [15].

Collective action through consortia and professional associations can strengthen libraries' negotiating power and promote sector-wide ethical standards.

8.4 Capacity Building and Ethical Literacy

Ethical data governance requires sustained investment in professional development. Librarians must be equipped with the knowledge and skills to critically evaluate data practices, interpret policy frameworks, and engage with ethical dilemmas in everyday work [5]. Integrating data ethics into LIS curricula and continuing education programs is therefore essential.

Training initiatives should emphasize ethical reasoning alongside technical competencies, reinforcing the librarian's role as a socially responsible data steward.

9 Conclusion and Future Scope

The digital transformation of academic libraries has intensified the ethical significance of privacy and data governance. As libraries increasingly adopt data-driven services, analytics, and AI-based tools, traditional commitments to user confidentiality and intellectual freedom face new and complex challenges. This paper has argued that privacy and data ethics must be understood not merely as compliance requirements but as core expressions of libraries' social responsibility.

Through a systematic review of ethical principles, policy frameworks, and scholarly literature, the study highlights the tensions between technological innovation, vendor dependency, and ethical accountability in academic libraries. The proposed Privacy–Data Ethics–Social Responsibility framework emphasizes the need for integrated, values-driven approaches to data governance that prioritize user rights, transparency, and trust.

Looking ahead, future research may empirically examine user perceptions of privacy in academic libraries, evaluate the effectiveness of ethical governance models, and explore the ethical implications of emerging technologies such as generative AI and learning analytics. For practitioners and policymakers, the findings underscore the importance of embedding ethical reflection into library strategy, policy development, and professional education.

By reaffirming their role as trusted, socially responsible institutions, academic libraries can lead by example in shaping ethical information environments that respect privacy, promote equity, and sustain democratic values in the digital age.

Declaration of AI and AI-Assisted Technologies used in Manuscript Writing Process: Not Applicable

References

1. Association of College and Research Libraries (ACRL): Learning Analytics Toolkit. ACRL, Chicago (2017).
2. American Library Association: Privacy: An interpretation of the Library Bill of Rights. American Library Association (2019).
3. Briney, K.: Data management for researchers. Pelagic Publishing (2019).
4. Cavoukian, A.: Privacy by design: The 7 foundational principles. Information and Privacy Commissioner of Ontario (2011).
5. Cox, A.M., Pinfield, S., Smith, J.: Moving a brick building: UK libraries coping with research data management as a 'wicked' problem. *Journal of Librarianship and Information Science* 51(3), 646–661 (2019).
6. Floridi, L., Taddeo, M.: What is data ethics? *Philosophical Transactions of the Royal Society A* 374(2083), 20160360 (2016).
7. Floridi, L., Cowls, J., Beltrametti, M., et al.: AI4People—An ethical framework for a good AI society. *Minds and Machines* 28(4), 689–707 (2018).
8. International Federation of Library Associations and Institutions: IFLA statement on privacy in the library environment. IFLA (2021).
9. Jones, K.M.: The ethics of data analytics in academic libraries. *Journal of Librarianship and Information Science* 52(2), 499–510 (2020).
10. Jones, K.M., Salo, D.: Learning analytics and the academic library: Professional ethics commitments at a crossroads. *College & Research Libraries* 79(3), 304–323 (2018).
11. National Information Standards Organization (NISO): NISO Consensus Principles on Users' Digital Privacy in Library, Publisher, and Software Provider Systems. NISO, Baltimore (2015).
12. Stachokas, G.: Reclaiming the library: Ideas and models for service change. ALA Editions (2018).
13. Tzanou, M.: Data protection as a fundamental right next to privacy? *International Data Privacy Law* 7(2), 88–99 (2017).
14. UNESCO: keystones to foster inclusive knowledge societies. UNESCO Publishing (2015).
15. Zimmer, M.: The third-party problem: Privacy and cloud computing in libraries. *Library Quarterly* 84(4), 447–468 (2014).

Responsible and Ethical Use of AI for Publications: Exploring Major Publishers' Guidelines

P. Ganesan ¹[0000-0003-1764-9095], Reshma John ² [0009-0006-3443-6859],

M. Mahalakshmi ³[0000-0002-3851-9601] and A. Pavithra ⁴[0009-0008-0311-8860]

¹ Department of Library and Information Science, Alagappa University, Karaikudi – 630 003

² Sree Chitra Tirunal Institute for Medical Science and Technology (SCTIMST), Trivandrum

^{3,4} Research Scholar, Department of Library and Information Science, Alagappa University, Karaikudi – 630 003

gharieni2003@yahoo.com

(reshmajohn994, maha27.mlakshmi, pavithraaukkdi}@gmail.com

Abstract. AI in recent years has brought significant changes in academic research and is considered as an indispensable tool for the researchers. The application of AI in research spreads across different disciplines, and is used for various reasons such as processing large volumes of data, optimizing complex processes, speed, efficiency and so on. On one hand, AI benefits academic research if used responsibly, otherwise, it brings complex ethical issues which affects accountability, responsibility, transparency, and trustworthiness of the research. For responsible use of AI for publication, the authors have analyzed five major publishers namely Elsevier, Springer Nature, Wiley, M, Multidisciplinary Digital Publishing Institute (MDPI) and Taylor and Francis and global organizations. From the analysis, it is found that the majority of the publishers mandate authors to take responsibility for use of AI, and policies for rights, use of images, and disclosure statements. No uniformity and inconsistency found among the publishers warranted a common guideline for AI use. Suggestive measures were provided based on the literature review and analysis of the publisher's policies.

Keywords: AI guidelines; Responsible use; Artificial Intelligence; Ethical Guidelines; Academic Integrity.

1 Introduction

Artificial Intelligence (AI) is rapidly transforming society in various aspects, including academic research. Generative AI has the capacity to generate research ideas, and write research papers in different disciplines. The increased use of AI in research has become a significant trend in support of publication activities. The reason for the use of AI is that it can process large volumes of data quickly, automate redundant routine tasks, and generate valuable insights. However, it is important to consider potential risks of misuse and emergence of security incidents during the use. Lack of clear guidelines may pose confusion over the users on what is to be used and not to be used. Therefore, principles and guidelines of what to use becomes crucial. This will enable the researchers to use the AI responsibly and in accordance with guidelines framed [1].

Several organizations across the globe have realized the seriousness of AI's impact on research and started developing guidelines for the authors and publishers to set the standards of accountability and to promote the human input and mechanical processes used for bringing out the research work [2]. Mondal, Mondal & Juhi added further that if AI is used responsibly and transparently, it will not be considered as unethical writing [3]. On the other hand, it enhances efficiency and quality of the publications without compromising integrity. If used properly, AI writing assistants can be a valuable tool to support the pursuit of research.

Although Resnik and Hosseini affirmed that use of AI benefits science and society, they have also cautioned about the complex ethical issues brought by AI which affects accountability, responsibility, trustworthiness, transparency, fairness, objectivity and other important values in research [4]. Kovari remarked that the outcome of the results from AI generated tools would be conventional, lack logic and facts, originality and would also lack critical analysis [5]. Moreover, incorrect references and misstated authors, titles, and other publication details need careful double-checking especially in professional context such as journalism and software development. The study highlighted the researchers, practitioners, and policy makers to critically analyze the situation and engage with the changes and make serious efforts to improve the technology by identifying the strengths and limitations of existing

technology, which will foster an environment for ethical and responsible use of AI effectively. In this aspect, this study has been undertaken to study the AI policy guidelines of some reputed publishers and consolidated findings were given in the study.

2 Review of Literature

The related studies offer a comprehensive review of the earlier studies conducted across the world and provide authors' views on ethical practices and importance of unified ethical guidelines and issues faced by the users of different types. Agar conducted a study to explore the awareness, and ethical use of AI among the research scholars and found that research scholars have basic understanding and involvement, but found shortcomings in training, lack of familiarity with more advanced options to improve their research methods [6]. Well-rounded training programs, tackling of ethical issues, critical thinking and more focus on ethical practices to address the challenges of privacy, bias and authorship, were the suggestions provided. A review study on research ethics and issues regarding use of AI by authors and reviewers by Kim revealed that use of AI in scholarly publication has become common and utilizing the same will bring potential risks in academic integrity, accuracy and originality of the published research output [7]. Many of the articles which are reviewed have not permitted AI to be listed as an author, and human authors were responsible for fact-related issues and academic integrity. Moreover, the reviewers found difficulty in identifying the AI generated content and required AI detection tools for submission and review process. Authors suggested a unified AI related policy to the authors, readers, reviewers, editors and publishers to ensure transparency and concluded that addressing the research ethics associated with AI will ensure research integrity, which will prevent falsification, fabrication and plagiarism in their scientific writing. Ateriya and others explored ethical issues of AI in academic writing and stated that AI provides innovative opportunities in writing research papers [8]. However, the authors also discussed substantial complex and diverse ethical issues which involve authorship, intellectual property rights, data privacy, academic integrity and bias. The work has proposed a set of ethical standards for conducting research based on the analysis of recent literature and case studies. Studies by Alsamhori and Alnaimat stressed ethical use of AI in research by ensuring academic honesty and moral standards [9]. Developing clear guidelines for the future by incorporating clarity of using AI systems, responsible utilization of Machine Learning, and training stakeholders on technological competence were the major concerns and remarks that success of implementation of such policy will be possible through teamwork involving scientists and engineers, teachers, individuals and public servants. Students and teachers' attitudes towards use of AI in higher education was assessed by Lien and others and they found that some have used the technology, whilst the other half have little or no experience in use of technology. Both the respondents have expressed concern over research integrity, with stronger worries and reservations among the respondents who use less [10]. Both the respondents also have no sufficient training. Scepticism in use of AI warrants clearer guidelines and better training. Study results of Vanecek, Yorulma and Dobrovska among students in Czechia (Cz), Turkey (TR) and Slovakia (SK) revealed that students appreciated AI benefits in accessibility and time-saving. However, their limited engagement with advanced features. Ethical concerns such as misuse, bias and over-dependence on AI were prominent warrants needed for AI literacy and institutional ethical practices [11]. Lee asserts that most ethical guidelines remain abstract and ambiguous, and at present, there are no consensus ethical practices, suggesting a standardized universally accepted AI framework for AI use [12].

3 Objectives of the study

The major objectives of the study are:

- To discuss the need and importance of ethical guidelines in relation to use of AI for academic research
- To review various studies which have focused the importance of ethical guidelines for responsible use of AI in research work
- To explore and analyze the AI ethical policy developed by prominent publishers

4 Methodology

This study examines the AI policy for the authors who submit their publications to the reputed publishers. For this study, the authors have examined the reputed publishers' who publish articles and books. Among the publishers, the authors have chosen five major publishers namely Elsevier, Springer Nature, Wiley, Multidisciplinary Digital Publishing Institute (MDPI) and Taylor and Francis [13]. These publishers collectively produce thousands of

journals and books covering different disciplines. The authors have closely examined the publishers' ethical guidelines for AI and analyzed results are discussed in the discussion section. In this study, the authors have checked the AI ethical guidelines for authors and restricted them to journal articles. While exploring the AI guidelines from these publishers, it is found that Elsevier, Springer Nature, Wiley, and Taylor & Francis websites are periodically updated, whereas, MDPI was last updated by August 29, 2025. Data for the study has been explored during the month of December and no recent revision was found among the surveyed websites. Across the globe, thousands of publishers are bringing out their publications and following different guidelines for publication. Within the limited period, it will be difficult to explore many publishers to identify their guidelines for AI use which is the reason for the authors for selecting few prominent publishers.

4.1 Elsevier

Elsevier encourages the authors in using AI tools for preparing manuscripts, but supports the application of AI with human oversight and control. It fixes accountability and responsibility on authors for the content generated for their work which includes carefully reviewing, accuracy, impartiality and comprehensiveness of the content generated by AI, editing and adapting and ensuring author's original contribution, ensuring use of AI tools or sources, and ensuring the safeguards of data privacy, intellectual property rights and terms and conditions for maintaining their data, inputs and unpublished manuscripts. Restriction is placed for duplicate images or images which are copyrighted, real people, others identifiable products or brands, or likeness of individual voice. Other conditions like authors to ensure their content right when they use AI for service and subsequent publication. Authors need to disclose the use of AI including the tool, purpose and their oversight while submitting the manuscript and permits to use AI for grammar, spelling, punctuation etc. Publishers do not allow AI tools to be listed as author or co-author and each author takes responsibility for accuracy or integrity and originality. Publishers do not allow AI or AI-assisted tools to create or modify images and adjustments such as brightness, contrast, or color are allowed which do not obscure or eliminate the information present in the original and also requests the authors to submit image forensic tools or specialized software to identify suspected image irregularities. Exception is only for use of AI or AI-assisted tools for research design or methods and to be described in the methods section. It does not allow for production of artwork such as for graphical abstract [14].

4.2 Springer Nature

Springer Nature puts accountability to the authors and not for the Large Language Model (LLMs). However, it requests the authors to properly document the use of AI in the methods section or in some other suitable area if the method section is not available. Whereas, use of LLM or other AI tools for "AI assisted copy-editing", declaration is not needed. In all cases, human accountability plays an important role for the final version of the text and agreement from authors for originality of the works. Springer Nature journals do not permit AI generated images for publication with some exceptions which will be decided with case -by-case reviews [15].

4.3 Wiley

Wiley requests the authors to carefully review the terms and conditions or licences associated with the technology. The authors need to ensure that AI will not claim ownership for their content as it may interfere with authors' right or journal's right or Wiley's right to publish the material. Authors should use AI only as a companion and authors should take complete responsibility for all the content, claims, citations, references and analysis. The final works should reflect their expertise and originality while adhering to Wiley's ethical standards. Authors must disclose AI use while submitting material. The guidelines say authors should not use AI technology which restricts their or Wiley or party's use of material for which the authors should review terms and conditions for clauses such as 'ownership', 'data reuse' or 'opt out'. Authors must use AI technology which aligns with privacy, confidentiality, and compliance obligations and authors need to adhere to the terms and conditions with Wiley [16].

4.4 Multidisciplinary Digital Publishing Institute (MDPI)

MDPI Supports transparent and ethical use of AI in scholarly publishing. If the author uses AI generative tools, the author needs to disclose the use of AI in the acknowledgement section followed by the detailed explanation in materials and methods section. For text polishing, disclosure is not necessary and authors are the sole respon-

sible persons who should take full responsibility for the manuscript, including AI-generated content, and the manuscript must reflect originality, accuracy, and ethical integrity. The Generative AI tools or LLMs cannot be listed as authors [17].

4.5 Taylor and Francis

Authors should take responsibility for the originality, validity and integrity of the content and should support the use of AI tools which ensures data security, confidentiality, and copyright protection for idea generation and exploration, language improvement, interactive online search with LLM-enhanced search engines, literature classification and coding assistance. Authors are responsible for the content created by the author and also suggest the authors to consult the editor for clarifications. Their author guidelines do not permit Generative AI to be listed as an author. It also urges the author to acknowledge the use of Generative AI tools such as its name, how it is used and reason for using it. The publishers do not permit the use of Generative AI for creation and manipulation of images and files, or original research data. Whereas, they allow AI-assisted technologies in the research process with human oversight and transparency [18].

5 Discussion

Treating AI as an author is like taking a new turn as they cannot be used as authors since they cannot be morally or legally responsible for research work. Developing AI policy in this digital age is a Herculean task as AI tools are evolving fast. For the past few years, publishers and authoritative organizations across the world have developed different and sometimes contradictory policy for use of AI in publication. Disclosure of AI tools promote credit among the research community and establishes fairness, honesty and transparency. Moreover, disclosure of AI promotes accountability which means shouldering the responsibility for the problem to prevent in future. Honestly, disclosing AI use facilitates demarcation between human and machine contribution. Disclosure promotes reproducibility [19]. Disclosing AI tools such as what, when, where and how will foster trustworthiness among the researchers. In some places, such as grammar check, plagiarism, suggesting synonymous terms, formatting, improving existing phrases, disclosure of AI is not necessary as this will not affect overall results. From the analysis of publishers' AI policy, there are no uniform guidelines and inconsistencies could be found with the publishers. From the study, out of five publishers, except Wiley, others have not allowed AI to be listed as authors, which is deviating from the study results of Kim revealing that the publishers have allowed to list AI as author and his findings are that the human authors are the ones who are responsible for content generation and academic integrity, and the findings are coinciding with our results [6].

Table 1. Publishers AI Policy in a nutshell

Guidelines	Elsevier	Springer Nature	Wiley	MDPI	Taylor and Francis
Authorship	No Authorship	No authorship	NA	No Authorship	No Authorship
AI for preparing manuscript	Supports, but with human oversight and control	NA	Companion not human replacement	NA	Supports with human oversight
Responsibility and accountability for content	Authors	Editing and content creation and accountable for final version	Authors	Authors	Authors
Disclosure Statement	Disclosure statement upon submission. Separate AI declaration statement; Research Process in methods section	Methods section, otherwise separate part.	copy-editing in the acknowledgement section. Content generation – in methods. Visual content – in figure	Disclose in Acknowledgement section and detailed information in materials and methods section	Either in the acknowledgement or methods section.
Terms and conditions of AI	Safeguards of data privacy, confidentiality, intellectual	Grammar, spelling, Punctuation, tone,	Terms of use, or any other conditions or li-	NA	Sufficient safeguards and protection of IPR,

	property and other rights	readability and style.	cences. Privacy, confidentiality and compliance obligation.		confidentiality and security
Images	Images and artworks are not allowed. Adjustments without changes in picture and cover arts with permission allowed.	Not allowed, but exceptions are made on a case-by-case basis.	NA	NA	Not allowed
Rights	Authors only grant right to use their materials and no right to train AI tool on their materials	NA	No ownership to AI for author's content and AI should not restrict authors own, Wiley's or other parties' use.	NA	NA
No Declaration	Grammar, spelling and punctuation,	Grammar, spelling. Punctuation, tone, readability and style.	-	Text polishing such as grammar, spelling, formatting and sentence structure	-

The Committee on Publication Ethics (COPE) also joined with the World Association of Editors (WAME) and Journal of the American Medical Association Network (JAMA Network) to state AI tools cannot be listed as an author for the paper [20]. The committee also suggests the authors to disclose the use of AI for writing manuscript, creating images, or for collecting and analyzing data in the materials and methods section and states that authors are fully responsible for the content of the manuscript produced by AI tool. International Committee of Medical Journal Editors (ICMJE) states the authors are to disclose use of AI-assisted technologies in the cover letter and in appropriate sections if available and AI should not be listed as authors as they cannot take responsibility for accuracy, integrity and originality of the submitted works. Therefore, the guidelines put responsibility on the authors for the works that include use of AI-assisted technology [21]. It was also found that the authors can use AI as a companion, while complete responsibility lies with authors in terms of accuracy, originality and academic integrity. Guidelines warrant the authors to use the AI which ensures data security, confidentiality and copyright protection. Overall, the majority of the publishers fix the responsibility to the authors in terms of accuracy, confidentiality, copyright etc. and ensure terms and conditions such as the data privacy, security, copyright etc. The guidelines also request the authors to disclose the use of AI and are least bothered with the use of AI-assisted copy-editing. In most cases, the guidelines are not allowing AI-generated images, but allows for adjustments such as brightness, contrast or color etc. Finally, most of the guidelines framed by the publishers were found to be more or less similar with little variations here and there. Global associations and major publishers across the world need to bring out a common guideline as publishers have different opinions on use of AI and inconsistencies found.

Through literature, it is explored that the academic users have basic understanding and some have no experience demanding AI literacy and institutional ethical practices. Few studies have discussed the importance of developing guidelines for use of AI systems, machine learning, and training stakeholders and proposed a set of ethical guidelines for conduct of research. The studies have emphasized intensive training programs on AI literacy and institutional ethical practices. As of now, there are no institutional level ethical guidelines for AI use and publishers also maintain their own AI policy and update. Providing intensive AI training, creating institutional level ethical guidelines for AI use and unified publishers' guidelines will enlighten the academic users and use AI ethically and responsibly.

6 Conclusion

Establishing trust and transparency is paramount in research publications. Developing AI ethical policy and educating the academic users will ensure academic integrity. Most of the guidelines for AI use were found to be the same. In India, University Grants Commission (UGC), New Delhi has developed ethical guidelines for researchers to use the research information ethically. As per the UGC guidelines, the universities have framed guidelines to ensure academic integrity. However, guidelines for responsible use of AI were not found and latest plagiarism detection software and AI-detection software were able to identify the AI-generated content. According to Lee,

most ethical guidelines were found to be abstract and ambiguous, suggesting standardized universally accepted AI framework. Proper guidelines for both research work and publication will stop the academic users from falling prey to AI. It is also very important that monitoring and continuous development, in tune with adjusting or refining policies appropriately, becomes the need of the hour. Finally, though AI tools applications are overwhelming, they should be judiciously used as a substitute for humans, otherwise it will erode human thinking, expertise and evaluation.

Declaration of AI and AI-Assisted Technologies used in the Manuscript Writing Process: Not Applicable

References

1. Syifa, A.F. (2024). Ethics in the Age of AI: Principles and Guidelines for Responsible Implementation in workplace. *International Journal of Advanced Technology and Social Sciences (IJATSS)*, 2 (2), 237-242. <https://doi.org/10.59890/ijatss.v2i2.1398>.
2. Chetwynd, E. (2024). Ethical Use of Artificial Intelligence for Scientific Writing: Current Trends. *Journal of Human Lactation*, 40 (2), 211-215. DOI: 10.1177/08903344241235160.
3. Mondal, H., Mondal, S., & Juhi, A. (2024). Responsible use of Generative Artificial Intelligence for Research and Writing: Summarizing ICMJE guideline. *Indian Journal of Orthopaedics*, 58, 1504-1505.
4. Resnik, D.B., and Hosseini, M. (2025). Disclosing artificial intelligence use in scientific research and publication: when should disclose be mandatory, optional, or necessary? *Accountability Research*. <https://doi.org/10.1080/08989621949>.
5. Kovari, A. (2025). Ethical use of ChatGPT in education—Best practices to combat AI-induced plagiarism. *Frontiers in Education*, 9, Article 1465703. <https://doi.org/10.3389/educ.2024.1465703>
6. Agar, M.S. (2025). Exploring Awareness, Usage, and Ethical Concerns of AI in Academic Research: A Perspective Study. *National Journal of Education*, XXIII (2)
7. Kim, S. (2024). Research ethics and issues regarding the use of ChatGPT-like artificial intelligence platforms by authors and reviewers: a narrative review. *Science Editing*, 11 (2), 96-106. <https://doi.org/10.6087/kcse.343>.
8. Ateriya, N., Sonwani, N.S., Thakur, K.S., Aravind Kumar., & Verma, S.K. (2025). Exploring the ethical landscape of AI in academic writing. *Egyptian Journal of Forensic Sciences*, 15 (36). <https://doi.org/10.1186/s41935-025-00453-1>.
9. AlSamhori, A.D.F., & Alnaimat, F. (2024). Artificial Intelligence in Writing and Research: Ethical Implications and Best Practices. *Central Asian Journal of Medical Hypotheses and Ethics*, 5 (4). <https://doi.org/10.47316/cajmhe.2024.5.4.02>
10. Lien, S., Skogli, E.W., Abamosa, J.Y., Glomdal, S., & Filukova, P/ (2025). Experience and attitude toward the use of artificial intelligence in higher education. *Universitetsforlaget*, 48 (1), 1-14. <https://creativecommons.org/licenses/4.0>.
11. Vanecek, D., Yorulma, Y.I., & Dobrovska, D. (2025). A holistic exploration of student attitudes toward AI use in higher education: an international comparison. *Cogent Education*, 12 (1). <https://doi.org/10.1080/2331186X.2025.2571691>.
12. Lee, N. (2026). Development of AI ethics guidelines model based on AI life cycle. *AI and Ethics*, 6 (8). <https://doi.org/10.1007/s43681-025-00851-0>.
13. Scilit. (2025). Scilit Rankings. <https://www.scilit.com/rankings>
14. Elsevier. (2025). Generative AI Policies for Journals. <https://www.elsevier.com/en-in/about/policies-and-standards/generative-ai-policies-for-journals>
15. Springer. (2025). Journal Polices. <https://link.springer.com/brands/springer/journal-policies>
16. Wiley. (2000-2025). Using AI tools in your research. <https://www.wiley.com/en-in/publish/article/ai-guidelines/>
17. International Open University Libraries. (2025). Publisher Policy on AI: MDPI. <https://libguides.iou.edu.gm/c.php?g=1482669&p=11059956>
18. Taylor & Francis. (2025). AI Policy. <https://taylorandfrancis.com/our-policies/ai-policy/>
19. Resnik, D.B., & Hosseini, Md. (2025). The ethics of using artificial intelligence in scientific research: new guidance needed for a new tool. *AI and Ethics*, 5, 14991521. <https://doi.org/10.1007/s3681-024-00493=8>
20. Committee on Publication Ethics (COPE). (2025). COPE focus on artificial intelligence. <https://publicationethics.org/cope-focus/cope-focus-artificial-intelligence>
21. ICMJE. (2025). Recommendation for the Conduct, Reporting, Editing, and Publication of Scholarly Work in Medical Journals. <https://www.icmje.org/icmje-recommendations.pdf>

LIS in Practice: Case Studies & Innovation Labs

Reimagining Academic Library Spaces for Active Learning: A University Library Case Study

Kothainayaki S¹ [0000-0002-0503-9217]

¹ Anna University, Chennai – 600 025, Tamil Nadu, India
skothai@annauniv.edu

Abstract. library is widely recognized as the heart of an educational institution, playing a crucial role in supporting learning, teaching, and research. Traditionally, libraries were viewed primarily as repositories of books and physical information sources. In recent years, however, this concept has undergone a significant transformation. Driven by technological advancements and the need for stronger community engagement, libraries are shifting towards more user-focused environments. They are progressively evolving into vibrant learning spaces that enable collaboration, innovation, and equitable digital access, thereby reshaping their role in the modern educational landscape. Physical spaces that were once dedicated solely to book storage and reading halls have been transformed into aesthetically pleasing areas, green spaces, discussion zones, collaborative workspaces, and more. This study examines, through a case study approach, the suitability of educational library spaces at the University Library, Anna University, in supporting contemporary student learning needs. It also explores innovative strategies for redesigning library spaces, highlighting recent developments at the University Library, Anna University.

Keywords: Transformative library spaces, Learner spaces, Academic Libraries, University Libraries, Library Spaces.

1 Introduction

Academic libraries play an essential role in the higher education ecosystem by supporting teaching, learning, and research activities. They serve as intellectual hubs where students, faculty, and researchers access a wide range of scholarly resources, including both print and digital materials. Over the years, the mission of academic libraries has expanded beyond the preservation and dissemination of information to include digital literacy training, research data management, and the creation of collaborative learning environments.

Space management has become one of the most significant challenges for academic libraries in the digital age. Traditionally, library spaces were primarily designed to store large collections of printed materials. However, the evolving needs of users and the transformation of libraries into collaborative learning hubs have created new demands on physical infrastructure. This study examines the spatial characteristics of academic libraries and their role in facilitating student learning. The primary aim is to evaluate how effectively academic library spaces meet the contemporary learning needs of students. The investigation emphasizes the importance of rethinking the library as a dynamic learning environment and understanding students' spatial preferences in relation to informal learning spaces (ILS).

The study focuses on the University Library at Anna University as a case study, highlighting how the library's physical spaces are transformed into learner-centric spaces. The findings provide significant insights into the transformation of physical spaces in accordance with students' spatial preferences for academic library environments. These insights can guide the design and enhancement of more effective and user-centred library spaces in higher education institutions.

2 About University Library, Anna University

The University Library was established in 1978 and gained autonomy in December 2000. The total area of the library building is approximately 4,122 sq.m. It is housed in a three-story building located on the eastern side of the main campus. It plays a pivotal role in augmenting the University's mission of disseminating knowledge by providing academic & research information, as well as technology-oriented library services with infrastructure facilities to undergraduate, postgraduate students, Research Scholars, and Faculty Members of Anna University. The University Library is located on three campuses: the College of Engineering Guindy (CEG) campus, the

Madras Institute of Technology (MIT) campus, and the School of Architecture and Planning (SAP) campus. The University Library has a collection of approximately 2.78 lakh books and 48,000 journal back volumes.

3 Literature Review

Flores R. M. et al. (2021) conducted a study aimed at guiding librarians in better understanding and planning library spaces, with a particular emphasis on spatial segregation of library functions and the effective facilitation of resources to enhance service delivery.

Lotfy M.W. et al. (2022) indicated that libraries are undergoing a profound transformation, evolving from static repositories of knowledge—often described as the “warehouse model”—into dynamic environments that foster creation, collaboration, and social learning, commonly referred to as the “community hub model.” In alignment with this paradigm shift, the University Library has transitioned from a traditional storehouse model to a vibrant learning ecosystem featuring makerspaces and collaborative learning areas. Following Meena J. (2022), the library now offers clearly demarcated quiet zones for individual study, collaborative and social zones with flexible and movable furniture to support group work, and dedicated browsing zones.

Sain and Negi (2023) emphasized the need to redesign libraries by transitioning from a place-based model to a platform-based paradigm, while not overlooking the critical importance of creating welcoming, functional, and inspiring physical spaces. They further argued that by integrating the strengths of both physical and digital environments, academic libraries can develop a comprehensive and inclusive ecosystem that addresses the diverse needs of 21st-century users.

4 Methodology

4.1 Data Sources

Primary data were obtained from:

- Library usage statistics, including daily gate counts and service utilization logs (study spaces, digital resource access, Wi-Fi usage).
- User feedback, collected from students and faculty through open-ended responses.
- Secondary data included institutional reports and internal library records relevant to infrastructure and service enhancements.

4.2 User Feedback through Interviews

This study employs a predominantly qualitative mixed-methods approach to investigate the influence of innovative library services and spatial redesign on user engagement and satisfaction within the University Library. Quantitative usage data and qualitative user feedback were systematically collected. Semi-structured interviews were conducted to capture in-depth perceptions and experiential insights. An interview guide was designed to focus on themes such as:

- Perceptions of library ambience and space utilization
- Effectiveness of collaborative and silent study areas
- Accessibility and usefulness of digital resources and ICT facilities
- Overall satisfaction and frequency of library use following service enhancements.

Participants were selected through purposive sampling to ensure representation from undergraduate students, graduate students, research scholars, and faculty members. Interviews were conducted on library premises during normal working hours. Interviews were conducted in person within the library premises, each lasting approximately 15–30 minutes. Responses were documented through written notes and audio recordings (with informed consent), ensuring confidentiality and voluntary participation.

5 Redefining Library Space Designs

5.1 Kolams at the Library Entrance – Aesthetic Space

Incorporating kolams—traditional South Indian floor art—at the library entrance (Figure 1) not only reflects cultural identity but also strengthens the connection between the academic environment and local heritage. The presence of kolams enhances the aesthetic value of the entrance, transforming it into a vibrant and meaningful threshold that can have a positive impact on students' moods, reducing stress and creating a supportive atmosphere conducive to learning. By integrating kolams into library design, the institution celebrates cultural continuity while simultaneously enriching the spatial experience of its users.

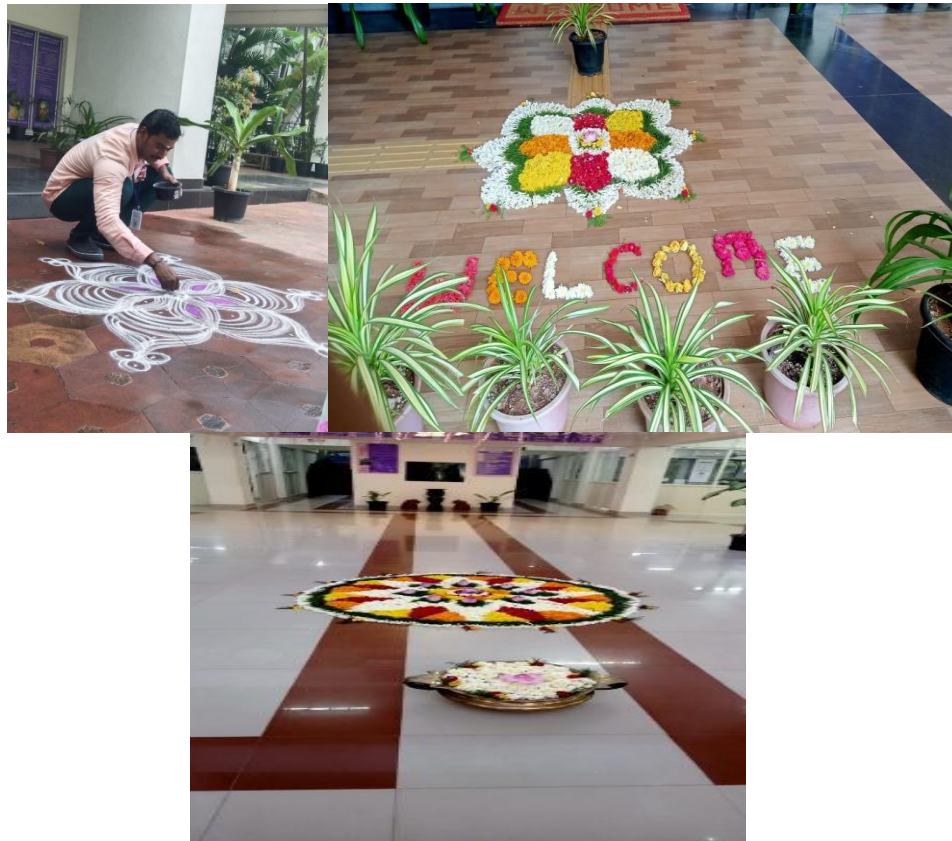


Fig. 1. Aesthetic Spaces – Library Entrance

5.2 Library Reception Space – Welcoming ambience

An aesthetically pleasing entrance not only welcomes visitors but also establishes a positive first impression that reflects the library's identity and mission. The University Library boasts a beautifully designed entrance (Figure 2) that conveys a warm and inviting atmosphere, thereby encouraging users to explore the diverse services and spaces offered within.

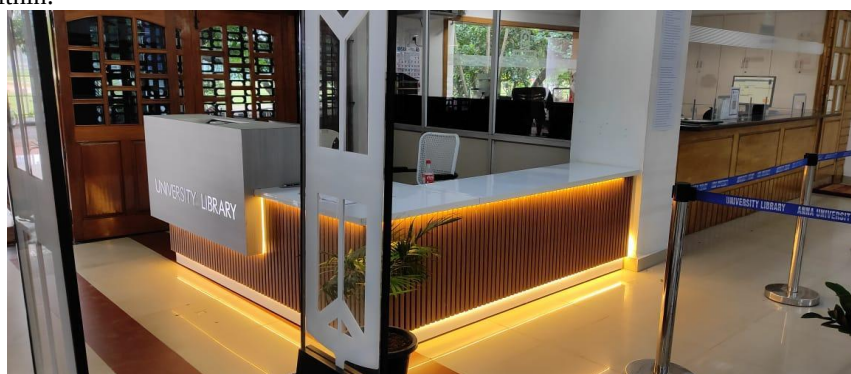


Fig. 2. Library Entrance

5.3 Personal Belongings Space

The library is equipped with wooden storage racks designed to securely accommodate the personal belongings of more than 100 students. These racks help maintain a clutter-free environment while ensuring convenience and safety for users during their library visits.

5.4 Divine Space Corner

This designated space promotes inner calm and positive values, contributing to a peaceful atmosphere that supports concentration, ethical reflection, and holistic development within the academic setting.

5.5 Thirukkural Corner

The Thirukkural Corner in the University Library has been developed as a cultural and value-oriented space to promote the reading of Thirukkural and appreciation of Tamil literature. A Thirukkural couplet is written daily on a whiteboard (Figure 3) to integrate classical Tamil wisdom into the contemporary learning environment, thereby bridging traditional Tamil heritage with modern academic practice.

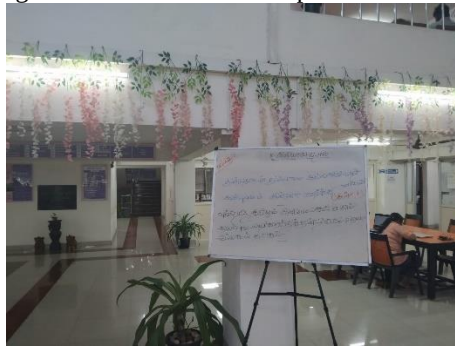


Fig. 3. Thirukkural Corner

5.6 Green Space

Potted plants placed at the entrance, and within the library (Figure 4) create a welcoming green environment that enhances the aesthetic appeal. This natural element contributes to a cool, calm, and soothing atmosphere, promoting relaxation and improving the overall user experience.



Fig. 4. Green Environment outside and inside the library

5.7 Own Book Reading Space

Students are permitted to bring their own books into the library, a policy that not only increases user footfall but also supports independent learning. By offering an air-conditioned, quiet, and distraction-free environment (Figure 5), the library enhances both reading comfort and concentration, encouraging students to make effective use of the space for academic success.

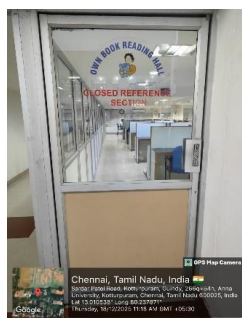


Fig. 5. Own Book Reading Hall

5.8 Digital Reading Space

Students have the option to borrow tablets from the library to browse and access the University's extensive range of e-resources inside the Digital Reading Hall (Figure 6). Additionally, they are encouraged to bring their own laptops for academic projects, class assignments, and other learning-related activities. These provisions facilitate seamless access to digital information and promote the effective use of technology within the library environment.

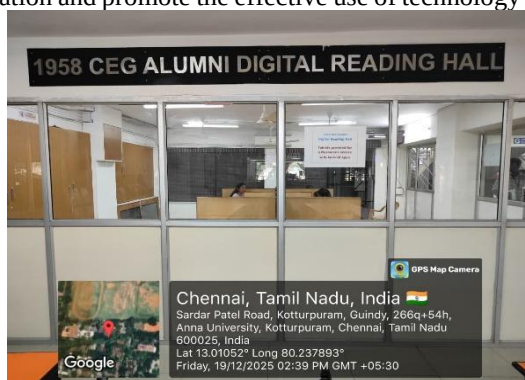


Fig. 6. Digital Reading Hall

5.9 Tamil Literature and Nobel Laureates' Books Section

To help students cope with academic stress and to encourage learning beyond the prescribed curriculum, the library provides a well-utilized space that promotes reading for pleasure and personal growth. The dedicated Tamil collection enables students to explore language, literature, and cultural heritage, while the Nobel Laureates' Books Section (Figure 7) offers access to globally acclaimed works, enriching their knowledge and broadening their intellectual perspectives.



Fig. 7. Tamil Literature Reading Hall and Nobel Laureates' Books Section

5.10 Career Guidance section

The library space also supports students in planning their future academic and professional pathways. The Career Guidance Section (CGS) of the University Library (Figure 8) provides a comprehensive collection of resources to assist students in selecting suitable careers or pursuing higher studies in their areas of interest. This section is

well-equipped with preparation materials for various competitive examinations, including GATE, GRE, GMAT, CAT, SAT, TOEFL, IELTS, UPSC, RRB, SSC, BANK, and TNPSC, among others. These resources help students enhance their readiness for national and international opportunities and play a vital role in shaping their career success.

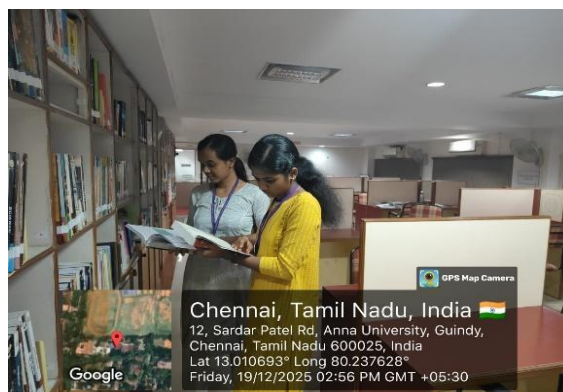


Fig. 8. Career Guidance Section

5.11 Reading Halls as Discussion Spaces

The air-conditioned reading halls (Figure 9) located on all three floors of the library function not only as quiet study spaces but also as collaborative zones for project discussions, assignments, group studies, and peer learning activities. These versatile environments support interactive engagement and provide students with a conducive setting to explore subject interests, share knowledge, and enhance their academic performance through collaboration.



Fig. 9. Reading Hall as Discussion Space

5.12 Newspapers / Magazines Reading Space

Magazines and newspapers subscribed to by the University Library are maintained as a separate collection (Figure 10), providing users with convenient access to current affairs and general knowledge resources. This section supports not only individual reading but also group discussions on recent developments, enabling students to stay informed and continuously update their awareness of national and global events.



Fig. 10. Newspaper Reading Hall

5.13 Ergonomic Design

Ergonomically designed chairs and individual study carrels are available for users who prefer isolated and focused study environments. In addition, open tables and modular seating arrangements support collaborative activities such as group projects and peer discussions. Lounge-style seating and relaxed reading spaces (Figure 11) are also incorporated to create a more informal atmosphere for leisure reading and mental relaxation.

This flexibility allows students to choose seating that best matches their physical comfort and study posture, whether they are reading, writing, using laptops, or interacting with peers. By offering such diverse seating configurations, the library ensures that users can find a space that aligns with their academic tasks and personal preferences, ultimately enhancing productivity and overall well-being. This thoughtful approach to spatial design fosters an inclusive and user-centred environment that supports effective learning experiences.



Fig. 11. Ergonomically Designed Carrels

5.14 Aesthetic Environment

The use of decorative elements, vibrant colours, and motivational imagery—such as inspirational quotes displayed across different areas of the library—helps in creating a visually engaging and stimulating environment (Figure 12). These design features contribute to a sense of creativity and comfort, encouraging users to feel more relaxed and motivated while studying. Such aesthetic enhancements not only improve the overall ambience of the library but also support positive psychological effects, including reduced stress and improved focus. By integrating thoughtful visual elements, the library fosters a welcoming atmosphere that enhances both the user experience and the learning process.

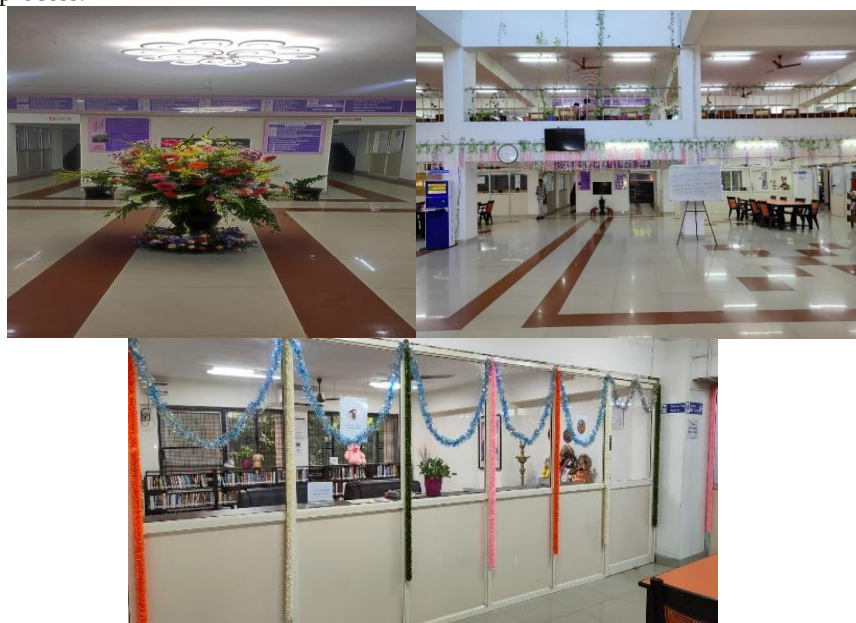


Fig. 12. Aesthetic appeal inside the library

5.15 Collaborative Learning Zones and Technology-Driven Spaces

The Digital Knowledge Centre in the University Library (Figure 13) serves as an essential technology-enabled learning hub. The Centre provides comfortable seating and seamless access to electronic resources. It offers dedicated computer workstations that provide seamless access to the University's vast collection of e-resources, including e-books, e-journals, online learning platforms, and digital repositories. This centre supports students in a variety of academic and career-building activities. The space is frequently used for attending virtual interviews, engaging in self-grooming and skill-enhancement programs, and participating in career-oriented online courses and training modules.



Fig. 13. Digital Knowledge Centre

The zone is thoughtfully designed to encourage collaborative learning, where students work together on digital assignments, research projects, and knowledge-sharing activities. By integrating technology with flexible learning opportunities, the Digital Knowledge Centre plays a vital role in enhancing digital literacy, increasing student confidence, and supporting their readiness for future professional environments.

Technology-driven spaces in the University Library are designed to support modern learning needs by integrating digital tools, high-speed connectivity, and access to electronic resources. These spaces enable students to engage in technologically enhanced learning activities such as browsing scholarly databases, participating in virtual courses, and accessing cloud-based services necessary for research and academic work.

5.16 Digital Content Creation Space

This section facilitates the production of learner-centric online tutorials on key syllabus-based topics. The content developed is integrated into classroom instruction, enabling blended learning, improving topic comprehension, and supporting diverse learning styles among students.

5.17 Multifunctional spaces

Collaborative learning is strongly supported by the multifunctional design of the University Library. Designated group study areas, project spaces, and interactive zones encourage students to actively engage in teamwork, peer discussions, and shared academic tasks. These dynamic spaces (Figure 14) transform the library into a vibrant social hub where knowledge exchange and cooperative learning thrive. As a result, the multifunctional areas of the library become the epicenter of collaborative educational activities, fostering creativity, communication skills, and collective problem-solving among students.



Fig. 14. Multifunctional Space

5.18 Community Engagement

The Seminar Hall in the University Library (Figure 15) serves as a versatile venue for hosting various academic events and programs, including workshops, seminars, and guest lectures. This facility enables the library to actively engage with the academic community and create opportunities for knowledge sharing and skill development. By offering a dedicated space for interactive events, the library enhances its role as a learning and cultural hub, facilitating continuous education beyond traditional library services.



Fig. 15. Seminar Hall

5.19 Reprography Space

The Reprography Section of the University Library (Figure 16) provides facilities for photocopying selected pages from books, journals, and other reference materials. This service is extensively utilised by students for academic study, assignments, and project work, enabling them to access required content conveniently while ensuring the safe handling and preservation of library resources.



Fig. 16. Reprography

6 Continuing Need for Physical Library Spaces in the Digital Era

Advances in digital technologies and the proliferation of electronic resources have fundamentally changed information access in university libraries. While digital libraries offer convenience, remote accessibility, and extensive coverage of scholarly content, they do not eliminate the need for physical library spaces. Instead, physical and digital library services function in a complementary and interdependent manner within the university ecosystem.

Physical library spaces provide a structured academic environment that is critical for focused learning and intellectual engagement. Unlike digital platforms, which are often accessed in informal or distraction-prone settings, physical libraries offer dedicated quiet zones, reading halls, and study areas that promote concentration, deep reading, and critical thinking. For many students—particularly undergraduates—such environments are essential for academic discipline and effective learning.

From a collection perspective, printed books continue to play a vital role in supporting curriculum-based teaching and learning. Core textbooks, reference materials, and prescribed readings are often best accessed in print, especially for sustained study. While e-resources provide quick access to current research, physical collections offer stability, permanence, and curated quality, whereas web-based information can be transient, inconsistent, or unreliable.

Physical libraries also address issues of equity and access. Not all students have personal access to digital devices, high-speed internet, or conducive study spaces. University libraries bridge this digital divide by providing shared access to both physical and electronic resources, computing facilities, and reliable connectivity within a supportive learning environment. In this sense, physical library spaces ensure inclusivity and equal opportunity in higher education.

In addition, physical library spaces play a significant role in fostering collaborative and social learning. Group study rooms, discussion areas, and maker spaces enable peer learning, interdisciplinary interaction, and knowledge co-creation—activities that are difficult to replicate fully in virtual environments. The physical presence of professional library staff further enhances learning through personalized research assistance, information literacy training, and academic support.

From an institutional perspective, the university library serves as a symbolic and functional hub of academic life. It is frequently visited by dignitaries, external scholars, and walk-in users, and is showcased during accreditation, ranking, and official institutional visits. A well-designed physical library reflects the academic stature, intellectual culture, and commitment of the university to quality education and research.

In conclusion, while digital resources have expanded the reach and efficiency of information access, they cannot substitute the multifaceted role of physical library spaces. In a university library setup, physical and digital environments are not competing entities but complementary components of a holistic learning ecosystem. The continued investment in physical library spaces, alongside digital infrastructure, is therefore essential for supporting teaching, learning, research, and community engagement in higher education.

7 Discussion

The library's operating hours (8:30 a.m. to 9:00 p.m.) are considered highly convenient for students. Users also highlight the library's flexibility, physical comfort, spatial connectivity, and access to digital technologies as key factors contributing to its overall usability. Additionally, the library operates throughout the year, closing only on a limited number of holidays, which further enhances its accessibility and convenience. Users note that the library's spaces support a wide range of student interactions and function as a collaborative learning hub in addition to fulfilling traditional functions. They find the library highly conducive to group work, and its effective ventilation, aesthetic appeal, and green, serene environment encourage students to remain and work for extended periods.

The library offers seamless, uninterrupted Wi-Fi connectivity, facilitating the efficient completion of academic projects by students. As a result, the presence of such high-quality facilities has significantly enhanced student engagement, leading to a marked, multi-fold increase in library footfall. Accordingly, the purposeful reconfiguration of library spaces to address the learning requirements of the user community plays a pivotal role in transforming the academic library into a collaborative learning hub, thereby reinforcing the institution's overall learning ecosystem. The library environment is enhanced through the effective use of natural lighting, ergonomic furniture, and vibrant colour schemes. To support digital learning, the facility is equipped with ample charging stations and high-speed Wi-Fi, enabling uninterrupted access to e-resources. The introduction of self-checkout kiosks has further streamlined services, allowing library staff to focus on delivering personalized user support.

Overall, the University Library has evolved from a predominantly silent reading space into an active, collaborative learning hub. This transformation emphasizes user-centric knowledge creation and is supported by a flexi-

ble, modular design that redefines the contemporary academic library space. Increased footfall and user satisfaction: Users reported greater willingness to spend extended periods in the library, attributing this change to improved space utilization, accessibility, and study-friendly environments. The increase in footfall observed after the library space redesign aligns with qualitative interview findings that emphasize comfort, flexibility, and functionality as key determinants of library use. These results corroborate earlier studies that position spatial design as a critical factor in transforming academic libraries into active learning hubs.

8 Conclusion

The study underscores the importance of rethinking the library as a learning hub, with spaces intentionally designed to support diverse student learning activities in order to enhance both effectiveness and user appeal. It further suggests that underutilized print collections may be relocated to off-site storage, where they can be optimally managed and retrieved on demand. Such a strategy enables a reallocation of library space from traditional collection storage to learning-oriented environments for student use.

The introduction of Own Book Reading Spaces, Green Spaces, and Collaborative Spaces within the University Library has notably increased user footfall, thereby providing students with an academic ambience conducive to focused work, meaningful engagement with peers, and access to flexible spaces that meet varied learning needs. Despite these improvements, the study's findings indicate that the addition of a refreshment area within the University Library would further enhance the overall user experience.

Declaration of AI and AI-Assisted Technologies used in Manuscript Writing Process: ChatGPT-5 for spelling and grammar check, Grammarly for spellcheck.

References

1. <https://library.annauniv.edu>
2. Flores R.M. et al. (2021). Library space design framework: a conceptual analysis. *Asia-Pacific Social Science Review*, 21(2): 240-249.
3. Lotfy M.W. et al. (2022). Academic libraries as informal learning spaces in the architectural educational environment. *Ain Shams Engineering Journal*, 13(6), 1-11. <https://doi.org/10.1016/j.asej.2022.101781>
4. Meena J. (2022). Library Spaces and Architecture: Designing for the Future. *ShodhKosh: Journal of Visual and Performing Arts*, 4(2), 1102-1109. <https://doi.org/10.29121/shodhkosh.v4.i2.2023.2192>.
5. Sain S.K. and Negi A. (2023). What's new in 21st-century academic libraries: Redesigning libraries from "place" to "platform". *International Journal of Innovative Science and Research Technology*, 8(8): 1348

CONTRIBUTORS. Dr. S. Kothainayaki is working as Assistant University Librarian in the University Library, Anna University, Chennai, Tamil Nadu, India. She has researched her library and written this paper. Her research interests include Webometrics, Best Practices, E-Resources, Makerspaces, Automation of Libraries, and User Satisfaction Studies.

Patterns and Practices of Usage of E-Resources among Students of the Central University of Rajasthan

Vineeta Chauhan¹[0009-0001-4101-7955], Kiran Yadav²[0000-1111-2222-3333] and Sobhagyawati Gupta³[0000-0003-2355-5239]

¹Sri Karan Narendra Agriculture University, Jobner, Rajasthan, India
inchargelibrary.codst@sknau.ac.in

²IGNOU
kiranyadav542001@gmail.com

³Central University of Rajasthan
shobha.library@curaj.ac.in

Abstract. This study investigates the utilization of online research journals by the students of the University of Rajasthan, Central Library. Understanding user awareness, preferences, satisfaction levels, and difficulties in accessing e-resources is essential for enhancing library services, as academic research increasingly relies on digital resources. This case study examines demographic information, library visitation frequency, resource preferences, and satisfaction with available e-resources using a questionnaire-based survey of 100 students. Despite a persistent reliance on printed materials, key findings show a high level of satisfaction with overall library services and a notable preference for online research journals among digital scholarly tools. The study provides important insights for improving the accessibility and promotion of electronic collections by identifying the main sources of users' knowledge of and acquisition of e-resources.

Keywords: Academic research, User satisfaction, Information access, Open Access journals, Bibliographic databases, Online databases.

1 Introduction

Electronic resources such as e-journals, e-books, and online databases have become indispensable to higher education and research. Compared to print resources, e-resources enable faster access, efficient searching, and improved retrieval of scholarly information. The rapid growth of internet technologies and digital publishing has further expanded access to both subscription-based and open access scholarly content through academic libraries. In this digital environment, academic libraries play a critical role in facilitating access to electronic information resources. Understanding users' awareness, preferences, and satisfaction with e-resources is essential for effective library planning and service improvement. The present study examines the patterns and practices of e-resource usage among students of the Central University of Rajasthan, with a focus on awareness, preference, satisfaction, and perceived impact on academic work.

2 Objectives

The main objectives of this study are as under:

- To study the awareness about digital scholarly tools among students of University of Rajasthan.
- To improve and increase development sources related to use of e-resources.
- To remove the barriers and difficulties faced by the user while accessing digital scholarly tools.
- To understand satisfaction and availability of e – resources.
- To study the users' satisfaction level towards the information and data available through digital scholarly tools.

3 About the Library of Central University of Rajasthan

Together with the founding of the Central University of Rajasthan, the Central Library was established in 2009. The library currently occupies a semi-permanent structure that is roughly 1326 square feet in size. A hybrid collection of learning resources is available at the library in both print and electronic formats. The University's teach-

ing and learning operations are supported and centered by the library. It acts as a knowledge base that the university community may access thanks to fast intranet and internet connection. The library features a well-established cyber library with 15 networked PCs connected to a LAN, complete with air conditioning and Wi-Fi. This allows users to access thousands of digital scholarly tools that are subscribed through e-Shod Sindhu as well as internet resources like NPTEL and e-PG Pathsala, among others. There is a reading hall at the Central Library that can accommodate over 400 students. KOHA Open-Source Library Management Software has been used to automate all of the library's regular tasks. Additionally, the library used RFID technology to improve circulation services and prevent theft. The CURAJ campus-wide network makes the OPAC service accessible around-the-clock. Science, social science, linguistics, humanities, engineering, management, sports science, and yoga are among the subjects covered by the library's hybrid collection of print and electronic learning resources. The University's teaching and learning operations are supported and centered by the library. It acts as a knowledge base that the university community may access thanks to fast intranet and internet connection. The library features a well-established cyber library with 15 networked PCs connected to a LAN, complete with air conditioning and Wi-Fi. This allows users to access thousands of digital scholarly tools that are subscribed through e-Shod Sindhu as well as internet resources like NPTEL and e-PG Pathsala, among others. There is a reading hall at the Central Library that can accommodate over 400 students. KOHA Open-Source Library Management Software has been used to automate all of the library's regular tasks. Additionally, the library used RFID technology to improve circulation services and prevent theft. The CURAJ campus-wide network makes the OPAC service accessible around-the-clock.

4 Methodology

A questionnaire survey method was used to conduct the study. Undergraduate (UG) and postgraduate (PG) students constituted the study population. Data was collected using both online and offline modes. The collected data were analyzed using MS Excel, and the results were presented in the form of tables and diagrams prepared with MS Excel and MS Word.

5 Literature Review

Various studies searched and identified for the literature review related to the present study. A brief review is presented below of the relevant studies to understand the topic and identify the need of the present study. Dadzie (2005) examined the access and use of electronic resources at Ashesi University College, Ghana, through a campus-wide questionnaire survey. The study revealed high levels of general computer and internet usage, supported by the institution's strong IT infrastructure. However, despite this favorable environment, the use of scholarly electronic databases was relatively low. The primary reason identified for this gap was inadequate awareness and promotion of available library e-resources. The study highlighted the need for systematic information literacy initiatives and recommended integrating information competency instruction into the academic curriculum. Overall, the research provides valuable insights into user behavior, infrastructural influence, and the critical role of library communication strategies in enhancing effective utilization of electronic resources in academic institutions. Jamal (2007) investigated the use of digital scholarly tools among students of the Faculty of Engineering & Technology at Aligarh Muslim University, Aligarh. The study analyzed the extent of awareness, access patterns, and frequency of use of digital resources such as e-journals, databases, and online information tools. The findings indicated that while students actively used digital resources for academic and research purposes, limitations related to inadequate training, lack of systematic orientation, and infrastructural constraints affected optimal utilization. The study emphasized the need for user education programs and improved access facilities to enhance effective use of digital scholarly tools. Overall, the paper highlights the importance of information literacy and institutional support in maximizing the academic benefits of digital resources in engineering education. Kumar (2015) examined the use and impact of electronic resources in academic libraries, focusing on their role in supporting teaching, learning, and research activities. The study highlighted increased dependence on e-resources such as e-journals, e-books, and online databases due to their ease of access, timeliness, and wide coverage. The findings indicated that electronic resources significantly enhanced users' academic productivity and information-seeking efficiency. However, challenges such as lack of user awareness, inadequate training, and technical issues were identified as barriers to optimal utilization. The study emphasized the need for effective user education programs and proactive library support to maximize the impact of electronic resources in academic environments. Liu (2006) explored users' perceptions, preferences, and patterns of use of print and electronic resources, focusing on how format influences information-seeking behavior. The study found that electronic resources were generally

preferred for quick searching, current awareness, and convenience, while print resources continued to be valued for in-depth reading and sustained engagement. User preferences were shown to vary according to task type, discipline, and purpose of use. The study concluded that print and electronic resources serve complementary roles rather than being mutually exclusive. The findings provide important insights for academic libraries in developing balanced collection strategies that align with user needs and evolving information practices. Omeluzor et al. (2012) examined the extent to which postgraduate students in a Nigerian university use digital scholarly tools for academic purposes. The study revealed that while students were aware of various digital tools for accessing scholarly information, their level of effective utilization remained moderate due to challenges such as inadequate ICT infrastructure and limited training. The authors emphasized the need for improved digital literacy programs and institutional support to enhance the effective use of scholarly tools in postgraduate research. Tenopir (2003) provided a comprehensive overview of research on the use and users of electronic library resources. The study highlighted a steady increase in the use of electronic resources across academic communities, while also identifying variations in usage patterns based on discipline and user skills. The author emphasized the changing role of libraries in facilitating access, training, and effective use of electronic resources in the digital environment. Thanuskodi (2012) investigated the use of electronic resources by students and research scholars of the Faculty of Arts at Annamalai University. The study found that e-resources were widely used for academic and research purposes, particularly for accessing journals and reference materials. However, issues such as lack of awareness of advanced search techniques and inadequate training were identified as major constraints, highlighting the need for user education and orientation programmes in academic libraries. Tyagi and Singh (2012) examined the use of online research journals by faculty members of selected engineering colleges in Uttar Pradesh. The study revealed a high level of dependence on online journals for teaching and research activities, owing to their timely access and wide coverage. However, constraints such as limited access to subscribed resources and inadequate awareness of available databases were reported, underscoring the need for improved library support and resource promotion.

On the basis of the literature review done for the present study, it was found that the use of electronic resources has become an integral part of academic and research activities in higher education institutions. The above studies also revealed common challenges such as limited awareness of available e-resources, inadequate information literacy skills, infrastructural constraints, and insufficient user training. We understood that after examining and evaluating the present scenario related to e-resources and services, improvements and better decision making is possible for libraries and administrators.

Therefore, to examine students' engagement with electronic resources in Central University of Rajasthan and understanding how students of CURAJ access, use, and perceive e-resources is essential for assessing the effectiveness of existing library services and digital infrastructure, this study was decided to conduct. Moreover, identifying usage patterns, preferences, and constraints will help in designing targeted user orientation programmes, improving resource selection, and enhancing research and learning support services. The findings of the study are expected to assist library professionals, university administrators, and policymakers in strengthening e-resource services and promoting their effective utilization for academic excellence.

6 Open Educational Resources

The following Open e Resources are available in the library

6.1 e-PG Pathshala

e-PG The MHRD started Pathshala, which offers high-quality, interactive e-content based on the curriculum in 70 courses across the fields of social sciences, arts, fine arts and humanities, natural and mathematical sciences, linguistics, and languages.

6.2 Virtual Labs

The Ministry of Human Resource Development (MHRD) in India is in charge of the Virtual Labs project, which is part of the National Mission on Education through Information and Communication Technology (NMEICT). Twelve institutes are working together on this initiative, and IIT Delhi is the coordinating institute. The people who will benefit are: All students and faculty members at science and engineering colleges who don't have access to decent labs or instruments, and researchers at different institutes who can work together and share resources.

6.3 Swayam Prabha

Using the GSAT-15 satellite, the SWAYAM PRABHA group of 32 DTH channels broadcasts high-quality educational programs 24 hours a day, seven days a week. NPTEL, IITs, UGC, CEC, IGNOU, NCERT, and NIOS all provide the content. The web portal is run by the INFLIBNET Centre.

6.4 Vidya-Mitra

Vidya-Mitra is an online learning site for all the e-content projects created by the MHRD's National Mission on Education through Information and Communication Technology (NME-ICT). The portal lets users search and browse all of the online information. This makes it easy for learners to find the materials they need, such as audio and video lessons, text lessons, and multimedia-rich materials, all in one place.

Table 1. Library purchased e-books

S.N.	Platform/package name	URL	No. of Titles
1	Elsevier	http://www.sciencedirect.com/	57
2	Taylor & Francis	http://www.tandfebooks.com/	91
3	Cambridge University Press	https://www.cambridge.org/core/	935
4	Pearson	https://ebookcentral.proquest.com/lib/curaj/search.action	304
5	Wiley	https://onlinelibrary.wiley.com/	1313
6	Oxford University Press	https://academic.oup.com/	1493
7	Sage e-vidya UPSC collection	https://evidya.sagepub.in/evidya/library	93
8	DeGruyter	https://www.degruyter.com/	126
NDL e-Resources			
1	World e-Book Library	http://community.worldlibrary.in/	
2	South Asia Archive (SAA)	http://www.southasiaarchive.com/	

Table 2. E-journals through One Nation One Subscription (ONOS) – Implemented from January 2025

S. N.	Resource/Publisher Name	URL	No. of Titles
1	AAS Science	https://www.aaas.org/science-journals	1
2	ACM Digital Library	http://dl.acm.org/	158
3	American Chemical Society Journals	http://pubs.acs.org/	87
4	American Institute of Aeronautics and Astronautics (AIAA)	http://www.aiaa.org/journals/	9
5	American Institute of Physics Journals	https://pubs.aip.org/	28
6	American Mathematical Society Journals	http://www.ams.org/journals	9
7	American Physical Society-ALL	http://publish.aps.org/browse.html	15
8	American Society for Microbiology Journals	https://www.asm.org/index.php/asm-journals	25
9	Annual Reviews Journals	http://arjournals.annualreviews.org	51
10	ASCE Journals	http://ascelibrary.org/	36
11	ASME Journals	http://asmedigitalcollection.asme.org/	35
12	Bentham Science Journals	http://benthamscience.com/	118
13	BMJ Journals	https://journals.bmj.com/home	36
14	Cambridge University Press Journals	https://www.cambridge.org/core	442
15	Cold Spring Harbor Laboratory Press Journals	https://www.cshlpress.com/	8
16	Elsevier Science Direct Journals	http://www.sciencedirect.com/	2387

17	Emerald Publishing Journals	https://www.emerald.com/insight/	311
18	ICE Publishing Journals	http://www.icevirtuallibrary.com	34
19	IEEE Journals	http://ieeexplore.ieee.org/	210
20	IndianJournals.com	http://indianjournals.com/	258
21	Institute of Physics Journals	http://iopscience.iop.org/journals	74
22	Lippincott Williams & Wilkins (Wolters Kluwer) Journals	https://oce.ovid.com/	305
23	Oxford University Press Journals	https://academic.oup.com/journals/	375
24	Project Muse	http://muse.jhu.edu/	731
25	Sage Publishing Journals	http://journals.sagepub.com/	988
26	SPIE Digital Library	https://www.spiedigitallibrary.org/	11
27	Springer Nature Journals	http://link.springer.com/	2404
28	Taylor & Francis Journals	https://www.tandfonline.com/	2548
29	Thieme Journals	https://www.thiemeconnect.com/products/all/home.html	51
30	Wiley Journals	https://onlinelibrary.wiley.com/	1333
Total			13078

Table 3. E-Resources access through INFLIBNET

S. N.	E-Resources	URL	No. of Titles
1	J Gate Plus (JCCC)	https://jgatenext.com/search	-
2	Web of Science (Database)	https://www.webof-science.com/wos/woscc/basic-search	-

Table 4. 7.2 E-journals and Databases subscribed through University Library

S. N.	Platform/package name	URL	No of Titles
1	EBSCO Business source elite	http://search.ebscohost.com/	4432
2	CMIE PROWESS Database	https://prowessiq.cmie.com/	-
3	EPWRF Database	https://epwrfits.in/	-
4	Sci Finder Database	https://scifinder-n.cas.org	-

7 Data Analysis and Interpretation

A thorough examination of the study is necessary to juxtapose the theoretical framework with practical applications, the variations of which may serve as a foundation for enhancements. With this in mind, and to meet the evaluation criteria that may serve as the foundation for the study's aims, an effort has been made to categorize the different respondents based on certain features gathered from them via a questionnaire. Tables and graphs show them. At the end of this paper, you will find a copy of the questionnaire that was given to 100 people. All of the math and number interpretations are for 100%.

7.1 Age of Students

Table 5. Age of Students

Criteria	Frequency	Percentage
18-25	50	58%
25-30	17	20%
30-40	14	16%
>40	4	5%

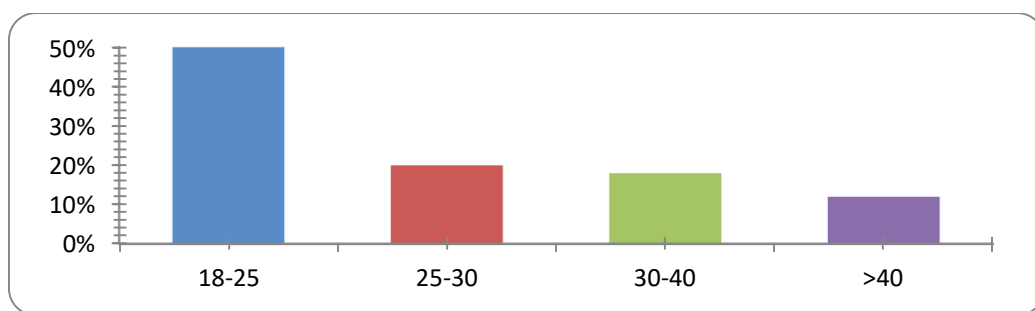


Fig. 1. Age of Students

Analysis: The study in the chart above shows that 58% of the people who answered are between the ages of 18 and 25. 20% of the people who answered were between the ages of 25 and 30, and 16% were between the ages of 30 and 40, and 4% are over 40 years old.

7.2 Gender of Students

Table 6. Gender of Students

Criteria	Frequency	Percentage
Female	60	70%
Male	25	30%

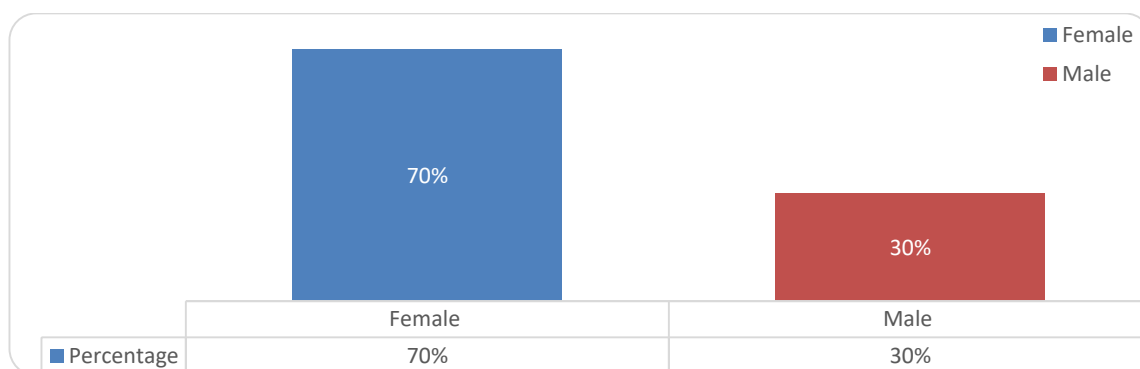


Fig. 2. Gender of students

Analysis: According to the above table, 30% of respondents are men and 70% of respondents are women at the same university.

7.3 Courses of Students

Table 7. Courses of Students

Criteria	Frequency	Percentage
Bachelor	73	86%
Masters	3	4%
Others	9	10%

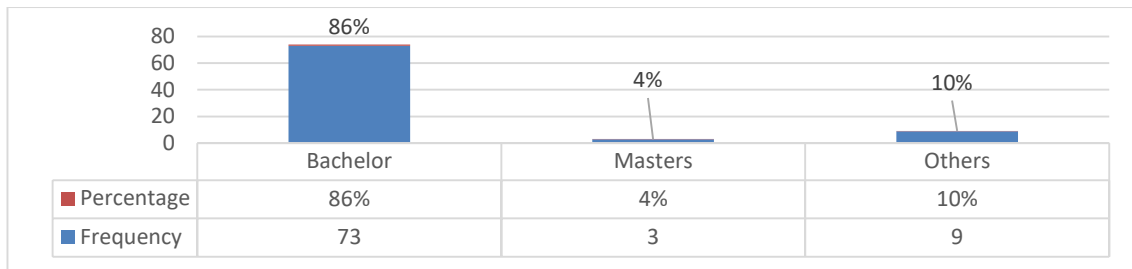


Fig. 3. Courses of Students

Analysis: According to the above diagram, 86% of respondents claimed to have completed a bachelor's degree. 10% of the respondents have taken other courses, and 4% have enrolled in a master's program.

7.4 Frequency of Students visits in Library

Table 8. Frequency of Students visits in Library

Criteria	Frequency	Percentage
Once a day	34	40%
Twice a day	27	32%
Once a week	14	16%
Once in 15 days	10	12%

Analysis: According to the pie chart above, 40% of respondents said they go to the library once a day. Thirty-two percent said twice a day, and sixteen percent said once a week. According to 12% of those surveyed, once every 15 days.

7.5 Availability of study material in the library

Table 9. Availability of study material in the library

Criteria	Frequency	Percentage
Yes	68	80%
No	17	20%

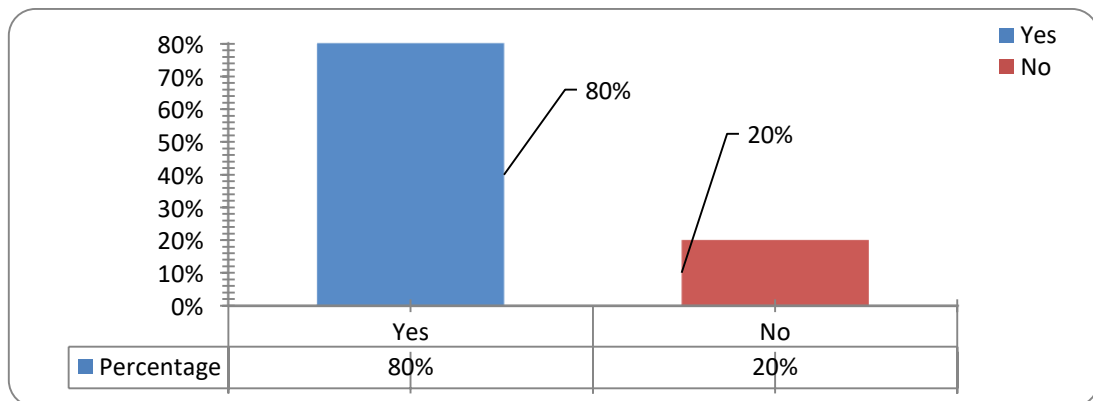


Fig. 4. Availability of study material

Analysis: According to the above chart, 80% of respondents expressed satisfaction with the study materials available in the library. Twenty percent of those surveyed expressed dissatisfaction with the library's study materials.

7.6 Students Preferred resources

Table 10. Students Preferred resources

Criteria	Frequency	Percentage
Printed Material	60	70%
E-resources	25	30%

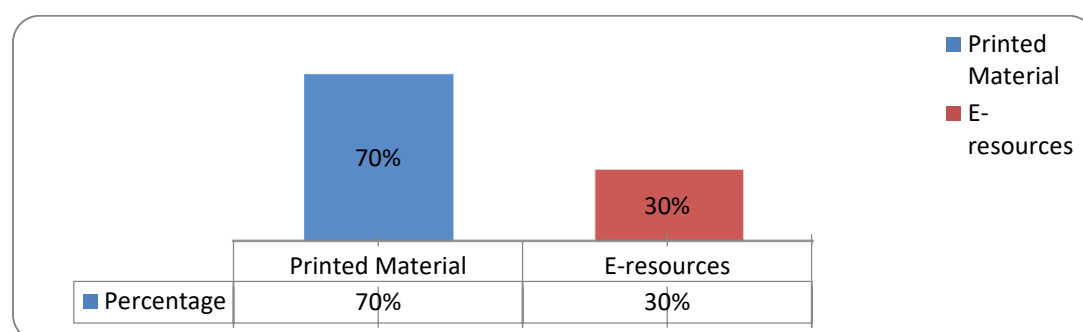


Fig. 5. Students Preferred resources

Analysis: According to the pie chart above, 30% of respondents preferred electronic resources, while 70% of respondents preferred printed materials.

7.7 Satisfaction Level of Students by the internet facilities or services given by the library

Table 11. Satisfaction Level of Students by the internet facilities or services given by the library

Criteria	Frequency	Percentage
Yes	77	90%
No	8	10%

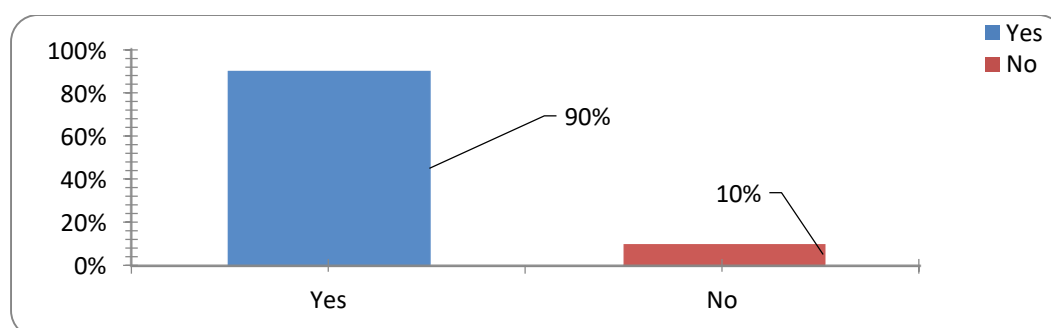


Fig. 6. Satisfaction Level of Students

Analysis: According to the above chart, 90% of respondents were happy with the library's internet resources and services, while 10% disagreed with the aforementioned claim.

7.8 E-resources preferred by the students

Table 12. E-resources preferred by the students

Criteria	Frequency	Percentage
E-Journal	36	42%
E- Books	25	30%
Websites	14	16%
Other	10	12%

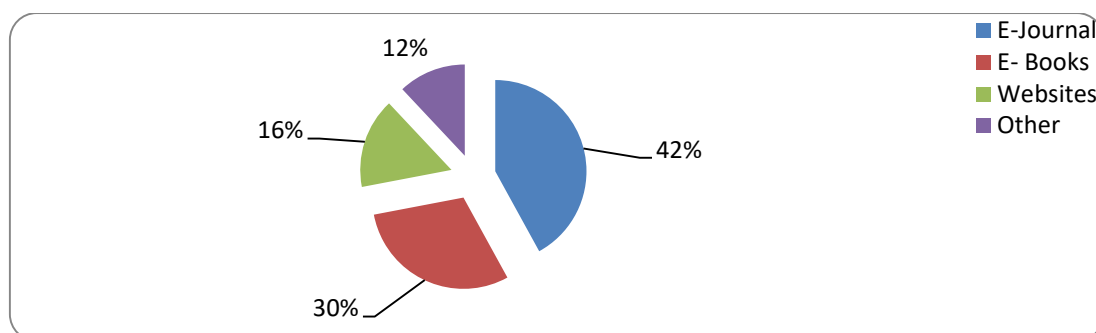


Fig. 7. E-Resources preferred by the students

Analysis: According to the pie chart above, 42% of respondents said they preferred e-journal resources, 30% said they preferred e-books, 16% said they preferred websites, and 12% said they preferred other documents.

7.9 Status of other Digital Scholarly Tools Provided by the library

Table 13. Status of other Digital Scholarly Tools Provided by the library

Criteria	Frequency	Percentage
Excellent	51	60%
Good	26	30%
Not Good	5	6%
Bad	3	4%

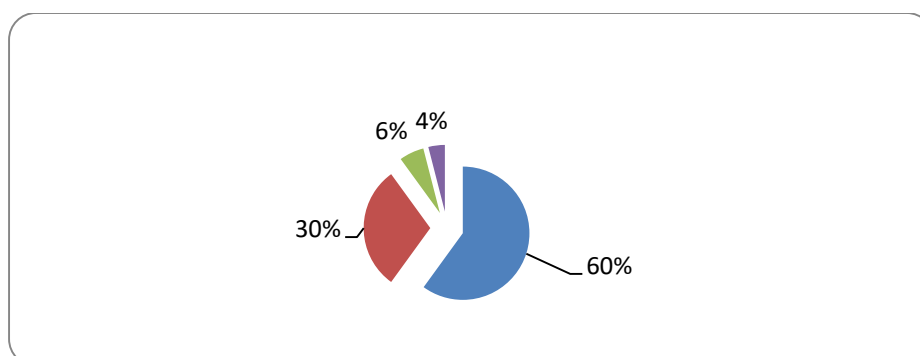


Fig. 8. Status of other Electronic Resources

Analysis: According to the pie chart above, 60% of respondents believe that the Library of CURaj's other digital scholarly tools, such as CDS and PDF, are in excellent condition, 30% believe that the library's collection of scholarly tools is good, 6% of the 85 participants are dissatisfied, and 3% of students believe that the library's other digital scholarly tools, such as CDS and PDF, are bad.

7.10 Impact of E-Resources on students of library of Central University of Rajasthan

Table 14. Impact of E-Resources on students of library of Central University of Rajasthan

Criteria	Frequency	Percentage
Highly Agree	42	50%
Agree	17	20%
Undecided	10	12%
Disagree	9	10%
Highly Disagree	7	8%

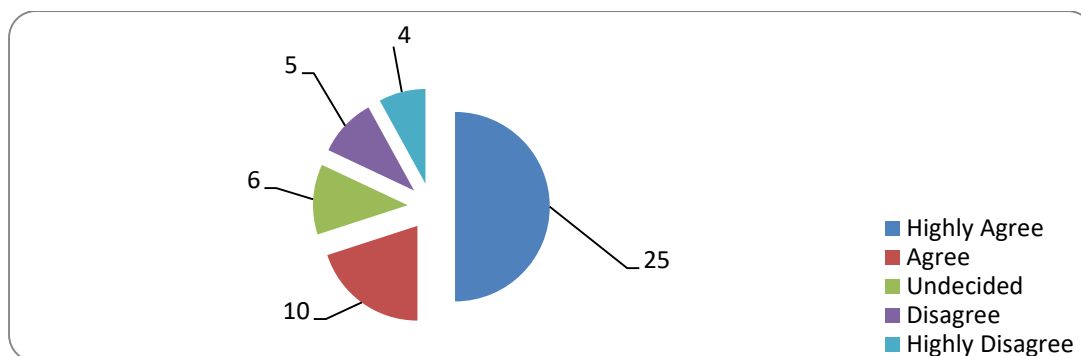


Fig. 9. Impact of E-Resources on students

Analysis: According to the pie chart above, 50% of respondents strongly agree that the impact of e-resources on Central University of Rajasthan library students is growing in the present, 20% agree that e-resources have a significant impact, and 12% are unsure of the impact. In addition, 10% disagree with the impact of e-resources, and 8% strongly disagree with the same.

7.11 Knowledge about E-Resources

Table 15. Knowledge about E-Resources

Criteria	Frequency	Percentage
Library notice board	39	46%
From faculty	22	26%
Library website	14	16%
Other sources	10	12%

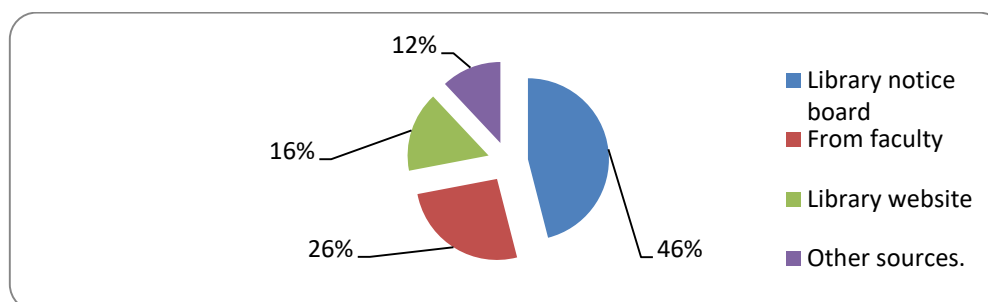


Fig. 10. Knowledge about E-Resources

Analysis: According to the above chart, 46% of respondents claimed that the library notice board was how they learned about e-resources. Twelve percent of the students are aware of e-resources from other sources, while 26 percent of the respondents learned about them from the faculty and 16 percent from the library website.

7.12 Expertise to use the e- resources effectively from

Table 16. Expertise to use the e- resources effectively from

Criteria	Frequency	Percentage
Procedures given on the site	37	44%
Library website	26	30%
Trial and error	14	16%
Guidance from other users	5	6%
Self-thought	3	4%

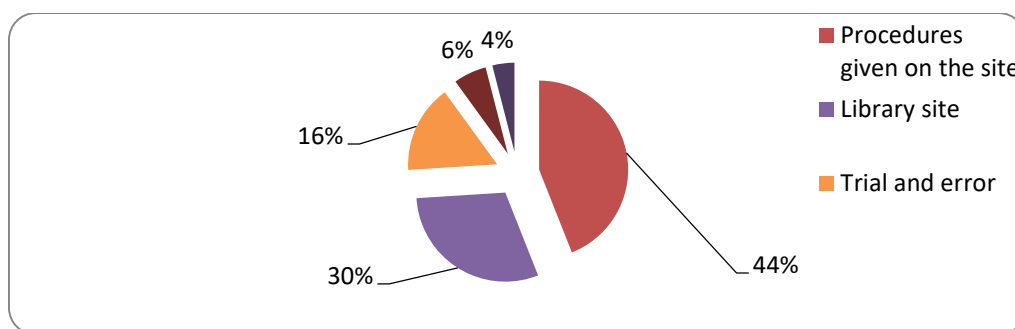


Fig. 11. Expertise to use the e- resources

Analysis: According to the above chart's analysis, 44% of students have mastered the site's instructions for using e-resources. According to 30% of the respondents, they learned from the library's website, 16% through trial-and-error, and 6% through experience using the e-resources efficiently. 4% do this on their own.

8 Major Findings

- Female students constitute a higher proportion of respondents than male students.
- Most users belong to the 18–25 age group and are enrolled in undergraduate programs.
- Printed materials continue to be widely used; however, e-journals are the most preferred electronic resources.
- A high level of satisfaction is reported with internet facilities and e-resource services provided by the library.
- Library notice boards and faculty members are the primary sources of awareness about e-resources.

9 Conclusion

The issue of quality in higher education has become paramount now and to achieve this goal for a library to develop itself with a rich information collection and build a best collection of digital scholarly tools and information services for its users. We are living in a digital world. The evidence is everywhere. The Central Library of University of Rajasthan have developed their own data archives, subject gateways to offer access to back volumes to identify the past research work completed and to hub on present day study trends in order to shift towards a dazzling future. By the optimistic role of digital scholarly tools, the libraries are not only boosting their efficiency in providing electronic services but also gaining the trust of their patrons in doing so. The emergence of digital scholarly tools has opened new vistas. The academia has been exploring the digital scholarly tools and information services to its maximum for the developments of its academic and research needs. The study of the three institutes reveals that the digital scholarly tools classed the traditional once due to its various features and this electronic format of information resources has become inevitable for the academic world. The implementation of e-resources proves accurate to the age-old standard that 'Every reader should get information at any time'. The use of e-resources is helpful to ensure exhaustive and pinpointed information. The e-resources provide various search options to the user and library managers. Using e-resources enable the library to save space of library and time of the users. E-resources are useful for libraries as well as for the users of the society who are starving to get a variety of information through the globe. The developments in the information and Communication Technology services available in the present made wonderful changes in the library operations. Its advantages are for technocrats, usage of the electronic products improve the knowledge of users. E-mails and RSS alerts carry the information for the individual to become aware of the user. Enhancement in Infrastructure like high-speed network, WI-FI in the campus, LAN portals at various rights to use points in the campus and in departments can be prepared to improve the practice effectively.

Despite these positive developments, the study has certain limitations. It is confined to a limited number of institutes, which may restrict the generalization of the findings to other types of universities or regions. The analysis is primarily based on available digital tools and user responses during the study period, and therefore may not fully capture rapidly evolving technologies or changing user behaviors. In addition, factors such as budget constraints, varying levels of digital literacy among users, and technical challenges related to infrastructure were not examined in depth.

Future research may address these limitations by extending the study to a larger and more diverse set of institutions, including universities with different academic orientations and resource capacities. Longitudinal studies

could provide deeper insights into changing usage patterns and the long-term impact of digital scholarly tools on academic performance and research productivity. Further work may also explore user training programs, cost-benefit analyses of e-resource investments, and the integration of emerging technologies such as artificial intelligence and data analytics to enhance library services. Such studies would contribute to a more comprehensive understanding of the role of digital scholarly tools in strengthening quality higher education.

Declaration of AI and AI-Assisted Technologies used in Manuscript Writing Process: Not Applicable

Reference.

1. Ansari, M. A., & Zuberi, N. A. (2010). Impact of ICT on library and information services. *DESIDOC Journal of Library & Information Technology*, 30(2), 5–11. <https://doi.org/10.14429/djlit.30.2.297> last accessed 2025/1
2. Barman, R., & Singh, M. (2009). User studies in academic libraries: A review. *DESIDOC Journal of Library & Information Technology*, 29(6), 3–10.
3. Central University of Rajasthan Library. (2022). *Central University of Rajasthan Library (Central Library services and infrastructure)*. Retrieved August 06, 2025, from <https://www.curaj.ac.in/library> last accessed 2025/10
4. Chowdhury, G. (2010). *Introduction to modern information retrieval* (3rd ed.). London: Facet Publishing.
5. Castells, M. (2010). *The rise of the network society* (2nd ed.). Chichester, UK: Wiley-Blackwell.
6. Dadzie, P. S. (2005). Electronic resources: Access and usage at Ashesi University College. *Campus-Wide Information Systems*, 22(5), 512–519.
7. INFLIBNET Centre. (2020). *Vidya-Mitra: Integrated e-content portal*. Gandhinagar: INFLIBNET Centre. Retrieved from <https://vidyamitra.inflibnet.ac.in> last accessed 2025/10
8. Jamal, A. (2007). Use of digital scholarly tools by the students of Faculty of Engineering & Technology, Aligarh Muslim University, Aligarh. *Annals of Library and Information Studies*, 54(4), 227–232.
9. Knowledge Librarian. (2017). Use and users of electronic library resources. *Knowledge Librarian: An International Peer-Reviewed Bilingual E-Journal of Library and Information Science*, 4(4), 110–116.
10. Kumar, P. S. G. (2004). *Library and information science*. B.R. Publishing Corporation.
11. Kumar, V. (2015). Use and impact of electronic resources in academic libraries. *Journal of Library and Information Science*, 5(1), 45–52.
12. Liu, Z. (2006). Print vs. electronic resources: A study of user perceptions, preferences, and use. *Information Processing & Management*, 42(2), 583–592. <https://doi.org/10.1016/j.ipm.2004.12.002> last accessed 2025/1
13. Mansell, R., & When, U. (1998). *Knowledge societies: Information technology for sustainable development*. Oxford: Oxford University Press.
14. Ministry of Education, Government of India. (2021). *SWAYAM Prabha: 32 DTH educational channels*. New Delhi: MoE. Retrieved from <https://www.swayamprabha.gov.in> last accessed 2025/7
15. Madhusudhan, M. (2008). Use of digital scholarly tools by research scholars of Kurukshetra University, Kurukshetra: A case study. *Library Herald*, 46(2), 165–177.
16. Omeluzor, S. U., Bamidele, I., Ezinwanyi, U. C., & Ukwuoma, S. N. (2012). The use of digital scholarly tools by post-graduate students in a Nigerian university. *International Journal of Library and Information Science*, 4(4), 69–78.
17. Saba, C. S. (2024). Information and communication technology (ICT) and economic development. *Journal of the Knowledge Economy*, 15(2), 567–589. <https://doi.org/10.1007/s13132-023-01571-8> last accessed 2025/8
18. Satpathy, S. K., & Rout, R. K. (2008). Challenges and opportunities for libraries in the digital environment. *IASLIC Bulletin*, 53(1), 22–29.
19. Suber, P. (2012). *Open access*. MIT Press.
20. Tenopir, C. (2003). Use and users of electronic library resources: An overview and analysis of recent research. *Council on Library and Information Resources*.
21. Tenopir, C., & King, D. W. (2000). *Towards electronic journals: Realities for scientists, librarians, and publishers*. Special Libraries Association.
22. Thanuskodi, S. (2012). Use of e-resources by the students and researchers of faculty of Arts, Annamalai University. *International Journal of Library Science*, 1(1), 1–7. <https://doi.org/10.5923/j.library.20120101.01> last accessed 2025/7
23. Tyagi, S., & Singh, B. (2012). Use of online research journals by faculty members of selected engineering colleges of Uttar Pradesh. *International Journal of Library and Information Science*, 4(1), 1–8.
24. UNESCO. (2004). *Integrating ICTs into education: Lessons learned*. UNESCO Asia and Pacific Regional Bureau for Education.
25. University Grants Commission (UGC). (2021). *E-resources and e-journals through INFLIBNET*. New Delhi: UGC. Retrieved from <https://www.inflibnet.ac.in> last accessed 2025/1
26. World Bank. (2002). *Information and communication technologies: A World Bank group strategy*. Washington, DC: World Bank.

Short Papers

The Power of Data Analytics and Artificial Intelligence to Enhance Decision-Making through Personalized, Data-Driven Recommendations: A Study with Reference to College Libraries and Digital Literacy Initiatives in University Libraries of Assam

Asmita Deka¹ and Md. Imran Ahmad²

¹ Hari- Gayatri Das College, Azara, Guwahati- 781017
dekaasmita2619@gmail.com

² University of Science and Technology - [USTM], Ri-Bhoi Dist., Meghalaya- 793101

Abstract. The rapid expansion of data analytics and artificial intelligence (AI) is transforming academic libraries from traditional repositories into intelligent, user-centred knowledge hubs. Globally, libraries are increasingly adopting AI-enabled systems such as recommender tools, chatbots, usage-analytics dashboards, and automated services to enhance personalization and support evidence-based decision-making. In India, and particularly in the northeastern state of Assam, academic libraries play a critical role in reducing the digital divide and promoting digital literacy among students and faculty. However, the integration of AI and data analytics in this regional context remains uneven and under-explored. This study examines how AI and data analytics, when combined with digital literacy initiatives, enhance decision-making processes through personalized, data-driven recommendations in college and university libraries of Assam.

The study is grounded in the conceptual premise that AI-driven data analytics enables libraries to move from intuition-based to evidence-based decision-making. By analysing user data such as circulation records, OPAC searches, and e-resource usage, libraries can generate personalized recommendations that improve user engagement and academic outcomes. At the same time, digital literacy initiatives act as a mediating factor by enabling users to understand, interpret, and apply these recommendations effectively. While international literature highlights the benefits of AI in libraries (Tripathi & Kumar, 2020; IFLA, 2022), there is limited empirical research that integrates AI-based decision-making and digital literacy in the context of Assam. This study aims to address this gap in the regional literature.

This study adopts a descriptive- analytical research design. Primary data were collected through a structured questionnaire administered to library professionals working in selected college and university libraries of Assam. Using purposive sampling, 80 questionnaires were distributed, of which 60 valid responses were received (75% response rate). The instrument included closed-ended and Likert-scale items covering awareness of AI and data analytics, levels of adoption, perceived impact on decision-making, and the role of digital literacy initiatives. Secondary data were gathered from policy documents, institutional reports, and scholarly literature. Data were analysed using descriptive statistics such as percentages and mean scores to identify patterns and trends.

The findings indicate a high level of conceptual awareness but only moderate levels of practical implementation. About 78% of respondents reported awareness of AI applications in libraries, and 65% were familiar with data analytics for library management. However, only 42% indicated that their libraries currently use AI-based tools such as chatbots or virtual reference services, and 39% reported the use of personalized recommendation features in OPACs. Usage analytics and e-resource statistics were the most widely adopted tools (68%), suggesting that libraries are more comfortable with basic analytics than advanced AI applications.

Importantly, 81% of respondents agreed that data analytics supports evidence-based decision-making in areas such as collection development and subscription management, while 84% believed that personalized recommendations improve user engagement. Digital literacy initiatives were also seen as crucial: 73% felt that training programmes on e-resources, research tools, and information literacy significantly enhance users' academic decision-making capabilities. These results suggest a positive relationship between awareness, adoption, and perceived usefulness, even though the depth of implementation remains limited.

The study also identifies key challenges to effective integration. These include inadequate technological infrastructure, shortage of skilled manpower, limited funding, and lack of clear institutional policies for AI adoption. Ethical concerns related to data privacy, algorithmic bias, and user

consent were also noted, highlighting the need for explicit data-governance frameworks. The findings reinforce the argument that technology alone is insufficient; without strong digital literacy programmes, users may not be able to interpret or trust AI-generated recommendations.

This study contributes by presenting an integrated perspective that links AI-driven data analytics with digital literacy as complementary forces in strengthening library decision-making. While the findings are context-specific to Assam and based on a modest purposive sample, they offer valuable insights for policymakers, librarians, and academic administrators working in similar regional and resource-constrained environments. Future research may apply inferential statistics or structural models to empirically test relationships between awareness, adoption, and impact.

In conclusion, AI and data analytics have significant potential to transform academic libraries into responsive, personalized, and evidence-driven institutions. When aligned with robust digital literacy initiatives, these technologies enhance both professional and user-level decision-making. Strategic investment in infrastructure, continuous capacity building, and the development of ethical data-governance policies is essential for ensuring that libraries in Assam become future-ready knowledge hubs in the digital age.

Keywords: Artificial Intelligence in Libraries; Data Analytics; Personalized Recommendations; Digital Literacy; Evidence-Based Decision-Making; Academic Libraries; Assam.

References:

1. Association of College and Research Libraries. (2016). *Framework for information literacy for higher education*. ACRL.
2. Baker, D., & Evans, W. (2019). *Digital information strategies: From applications and content to libraries and people*. Chandos Publishing.
3. Deka, A. (2025). *Aligning libraries with India@2047 developmental agenda for a future-ready libraries: Vision 2047*. In M. K. Verma, F. Chanchinmawia, & D. Gupta (Eds.), *Artificial intelligence in libraries: From Indigenous wisdom to Vision 2047* (Vol. II, Chapter 17). Mizoram University. ISBN 978-81-992907-6-1
4. IFLA. (2022). *Artificial intelligence and libraries: Opportunities and challenges*. International Federation of Library Associations and Institutions.
5. International Federation of Library Associations and Institutions. (2022). *Artificial intelligence and libraries: Opportunities and challenges*. IFLA.
6. INFLIBNET Centre. (2023). *E-resource usage and digital initiatives in Indian universities*. Gandhinagar.
7. Tripathi, M., & Kumar, S. (2020). Artificial intelligence in libraries: A systematic review. *Library Philosophy and Practice*, 1–18.

The Open-Air Green Library for Academic Sustainability: Improvement Strategies at Gulbarga University Library

¹Rajashree Hadapad and ²Suresh Jange

^{1, 2} Gulbarga University, Kalaburgi, Karnataka

¹rsmangalgi@gmail.com

²sureshjange@gmail.com

Abstract.

Background: The Green libraries aim to make the library sustainable by adopting these practices to reduce dependency on non-renewable energy resources and use environmentally friendly resources efficiently by adopting innovative technology and creating awareness to ensure air regulations, control impurities, reduce room temperature to keep the room warm and cold according to the seasons.

Related Work: The people from all over the world are trying to incorporate 'greening' in to their act for sustainability. Green Marketing, Green IT are the prominent and evident examples and libraries are not exception to it. The term sustainability denotes with social, economic and ecological aspects.

The academic sustainability exists found to be treated all over the world. In India Lalhlmpuii (2017) studied on Mizoram University Library, which is one of the first Ecofriendly University in Country. Applin (2009) discussed about the responsibilities shouldered by many colleges and Universities to creating program to address the issues of Global warming. Hule, Latimer and Warner (2013) compiled case studies from European countries, Asia, Australia, and United States of America. The case study comprehensively covers the idea of best practices on green libraries. Antonelli (2018) point out that creation of green libraries was approaching a tipping point of sustainability by generating green library movement. The innovation consists green library building. Enhance existing facilities. Provide green library services. And the author observed the oldest article on green libraries was dated in the 1990's. And study focus on nationwide rekindling of interest in a green environmental movement. The Website of New Mexico State University (2007) explained American College and University Presidents Climate Commitment (ACUPCC) which boasted a membership of over 560 colleges and universities from fifty states for the effective neutralize greenhouse applications in research and educational efforts to promote green living.

Methodology: The research questions in this study are:

RQ1: What is the GUK Library's academic sustainability strategy for supporting the green library concept?

RQ2: What are the impediments and facilitators to implementing the open-air green library concept in the GUK Library through the academic sustainability process?

This study employs a case study Method using a qualitative approach, and the research was conducted from July-August 2025. The data collection method used was responsive sampling, Interview, Observation and Documents.

Results and Discussion: Gulbarga University Open Air Green Library's Academic sustainability strategies support Research and other academic activities in the university environment by attracting students, Research scholars, and other library users, and changing the library's perception in the community, which should help it realize the green library concept.

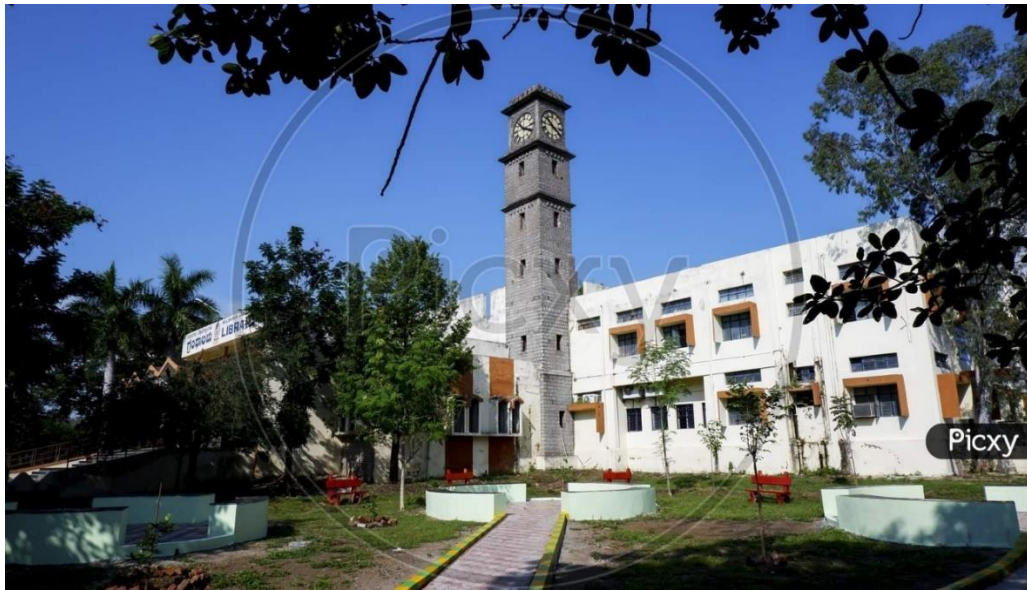


Fig. 1. Gulbarga University Library



Fig. 2. Impact of Open-Air Green Library on Academic Sustainability and Reading culture



Fig. 3. Impact on Reading Culture

From the above figure the following are the significant results.

- Daily time spending for reading time increased which indicates strong positive shift towards more enjoyable reading environment with less restrictive procedures and also shows the increase in engagement and accessibility.
- This observation suggest that open-air green libraries strongly promote reading culture than conventional indoor reading culture.
- Frequency of visit to library (per week) frequency almost doubled which shows the increase in academic accessibility and engagement.
- Stress Management were reduction of stress increased 20% to 50%. Students and other readers reported feeling more relaxed and focused for leading their better academic engagement.
- These finding states that exposure with natural elements will enhance the academic sustainability.
- 35% reduction in energy conservation was observed to decreased on artificial lighting and air conditioning and highlights the eco-friendly and sustainable nature of open-air green libraries.

Recommendations and Conclusion: On ‘Going Green’ The University Grant Commission should be taking steps to promote green libraries thought.

- Governmental financial supports ensure for establishment and maintenance of such type of libraries.
- The library should arrange strategic things and planning sessions to promote reading habits for the development of academic sustainability.
- Every University/College contributes and libraries made efforts to convert green library wherever possible.
- There are always two major challenges are facing any institutional libraries such as budgeting problem. Problem in book selection and another are problems from Dust, Moisture, and Fungus. It requires special care in changing seasons. And one of the most difficult issues is ‘Global Warming’ by encouraging ‘Green Library ‘Concept University contributing with its best for overcoming with these issues.

References

1. Barnes, L. (2012). Green Building as a Sustainability educational Tool. Library Hi Tech. 3(3). <http://doi.org/Retrieved on 28.08.2025>
2. Thakur Geetanjali Sing (2024). Green Library: How are became Green Library. https://library.iitd.ac.in/inter_n_presentations/Green_Library.pdf> Retrieved on 28.08.2025
3. <https://www.ifla.org/g/environment-sustainability-and-libraries/ifla-green-library-definition/>
4. 202004131501352746ranjana_sriv_lib_sc_Green_Library.pdf Retrieved on 30.08.2025
5. Thomas,Raysh (2016). GREEN LIBRARIES: INDIA VS INTERNATIONAL SCENARIO: Scholarly Research Journal for Interdisciplinary Studies. 4(37).
6. Anthonelli M (2008). The Green Library: Movement: An Overview and Beyond, Electronic Green Journal.1-27 Retrieved from <http://escholarship.org/uk/litem>.
7. Hauke, P., Latimer, K and Werner k U (2013). The Green Library: The Challenges of environmental sustainability. Netherlands: International Federation of Library Associations and Institutions. Retrieved from <http://www.repository.library.ifla.org/browse>
8. Lalhlimpuii (2017). Proposing Innovative Services in University Libraries of North East India: A case study on Mizoram University Library System. IASLIC Bulletin, 62(3) .83-191.
9. New Mexico State University (2007). American College and President’s Climate Commitment (ACUPCC). Retrieved from <http://sustainability.nmsu.edu/acupcc>

Information Retrieval in Complex Digital Environments: A Library and Information Science Perspective

¹Munegowda D M. and ²Amrutha R

¹Dravidian University, Srinivasa vanam, Kuppam, Andhra Pradesh

²Vijayanagara College, Hospet, Karnataka

¹munegowdadml1993@gmail.com

²amrutharmlisc@gmail.com

Abstract.

Introduction: The landscape of Information Retrieval (IR) has undergone a seismic shift from bibliographic systems to vast, algorithmically mediated digital ecosystems. These environments are defined by the volume, variety, and velocity of multimedia data, making relevance a complex, contextual judgment. While computer science advances in AI and machine learning power this transformation, a purely technocentric approach risks amplifying bias, misinformation, and inequity. This article argues that the human-centric principles of Library and Information Science (LIS)—organization, ethical access, and user empowerment—are foundational to developing IR systems that are not only efficient but also effective, just, and trustworthy.

Related Work: Our analysis builds upon and connects several key strands of literature:

- **Foundational IR Models:** The limitations of traditional models (Boolean, Vector Space) in complex, dynamic environments are well-documented (Salton & McGill, 1986; Baeza-Yates & Ribeiro-Neto, 2011).
- **User-Centered LIS Theory:** We integrate seminal LIS frameworks that center the user, such as Dervin's (1992) Sense-Making, Kuhlthau's (1991) Information Search Process, and Bates' (1989) Berry picking model.
- **Critical & Ethical Studies:** The analysis engages with critical scholarship on algorithmic bias and power (Noble, 2018), socio-technical systems (Suchman, 2007), and information ethics (Floridi, 2019).
- **Knowledge Organization (KO):** The indispensable role of KO systems, metadata, and linked data in providing semantic structure for modern IR is emphasized (Zeng & Qin, 2021).

This article synthesizes these strands to explicitly position LIS not as a supplementary field, but as a core mediating discipline in contemporary IR.

Methodology: This study employs a systematic critical literature review and conceptual analysis. The review mapped theoretical and practical intersections between LIS and IR in complex digital contexts (2000-2023), with selective inclusion of earlier seminal works. Searches were conducted in LISA, Scopus, Web of Science, and ACM Digital Library using keyword combinations (e.g., "information retrieval" AND "complex digital environment*", "algorithmic bias" AND "information access"). The conceptual analysis was used to synthesize findings and construct a novel integrative framework.

Results: Our analysis yielded three core results, conceptualized in the proposed framework and role analysis below:

- Identification of Critical Gaps: Traditional IR models fail to address the socio-technical challenges of modern digital environments: algorithmic opacity, multimedia content, dynamic relevance, and ethical concerns like bias and inequality.
- The Integrative Socio-Technical Framework for IR: We propose a four-layer framework where LIS principles form the essential mediating core (Fig. 1).
- Evolution of LIS Professional Roles: The analysis defines the shift from end-users to essential stakeholders in the IR lifecycle, as summarized in the table below (Table 1).

Discussion and Conclusion: This article achieves its objectives by demonstrating that the challenges of IR in complex digital environments necessitate a renewed LIS engagement. The proposed

framework operationalizes the unique value of LIS, making its expertise the vital link between user context and technical capability.

Discussion: The findings reaffirm that technical efficiency alone is insufficient. Effective and just IR requires the deliberate integration of professional LIS principles—organization, mediation, and ethical stewardship—into the fabric of system design, evaluation, and governance. This has direct implications for LIS education (requiring AI literacy and ethics), practice (expanding roles into design and auditing), and policy (advocating for transparent, equitable systems).

Conclusion: For IR to be truly effective and just in complex digital landscapes, it must be guided by the foundational values and expertise of the LIS field. The future of human-centered information retrieval depends on the seamless blending of advanced computational techniques with the enduring LIS commitment to organization, equitable access, and user empowerment.

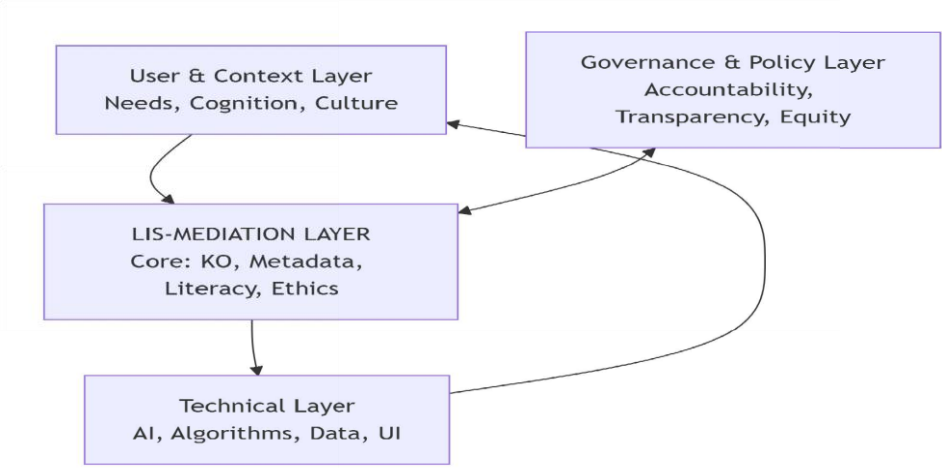


Fig. 1. The proposed Integrative Socio-Technical Framework, positioning the LIS-Mediation Layer as the central engine for contextual and equitable IR.

Traditional Role	Evolving/Emerging Role	Core Contribution
Search Intermediary	System Co-Designer & UX Architect	Informs design via KO expertise & user behavior models.
Collection Organizer	Data Steward & Metadata Engineer	Structures complex digital objects for AI- driven retrieval.
Reference Librarian	Algorithmic Accountability Auditor	Evaluates systems for bias, fairness, and transparency.
Instructor	Digital & AI Literacy Educator	Empowers users to critically navigate algorithmic systems.

Table 1. Traditional, emerging role and core contributions

Keywords: Information Retrieval, Digital Environments, Library and Information Science, Artificial Intelligence, User-Centered Design, Information Literacy, Ethics, Socio-Technical Systems.

References

1. Baeza-Yates, R., & Ribeiro-Neto, B. (2011). Modern information retrieval (2nd ed.). AddisonWesley.
2. Bates,M. J. (1989). The design of browsing and berrypicking techniques for the online search interface. Online Review, 13(5), 407–424.
3. Bawden,D., & Robinson, L. (2022). Introduction to information science (3rd ed.). Facet Publishing.
4. Dervin,B. (1992). From the mind's eye of the user: The sense-making qualitative-quantitative methodology.
5. J. D. Glazier & R. R. Powell (Eds.), Qualitative research in information management (pp. 61–84). Libraries Unlimited.
6. Floridi,L. (2019). The ethics of information. Oxford University Press.

7. Hjørland, B. (2002). Domain analysis in information science: Eleven approaches—traditional as well as innovative. *Journal of Documentation*, 58(4), 422–462.
8. Lin, J., Nogueira, R., & Yates, A. (2021). Pretrained transformers for text ranking: BERT and beyond. *Synthesis Lectures on Human Language Technologies*, 14(4), 1–325. 5.
9. Noble, S. U. (2018). *Algorithms of oppression: How search engines reinforce racism*. New York University Press. 6.
10. Zeng, M. L., & Qin, J. (2021). *Metadata* (3rd ed.). ALA Neal-Schuman.

Reconceptualising Library Spaces towards Human Centric Design for Cognitive Well-being and Inner Balance

Nisha S.¹, Balasubramani R.², Keerthna M.³ and Eswar M.⁴

¹⁻⁴ Department of Library and Information Science,
Bharathidasan University, Tiruchirappalli – 620024.

¹ nishasiva2112@gmail.com

Abstract.

Introduction: The contemporary world is deeply shaped by continuous digital engagement, influencing how individuals learn, work, communicate, and even think. While digital technologies have enhanced access to information and improved efficiency, excessive reliance on screens and constant connectivity has resulted in cognitive overload, reduced attention span, emotional fatigue, and diminished opportunities for reflection. In academic and professional environments, individuals often struggle to maintain concentration, inner balance, and mental clarity due to uninterrupted digital exposure. In this context, the need for intentional digital disengagement has become increasingly important for sustaining cognitive well-being and inclusive human development. Short periods of conscious withdrawal from digital interfaces allow individuals to reset attention, regulate emotions, and re-engage with tasks more meaningfully. Libraries are uniquely positioned to respond to this emerging challenge. Traditionally regarded as repositories of knowledge, libraries are now evolving into dynamic learning and experiential spaces. Library spaces should be designed for user comfort and flexibility to enhance well-being and concentration of the users, as the environment influences emotional and cognitive responses.

Related Work: Jagmohan Meena. (2023) examines the evolution of library spaces, highlighting flexibility, inclusivity, collaboration, and the need to balance innovation with traditional roles. Husaini et al. (2022) present academic libraries as therapeutic landscapes supporting student well-being. Bossaller et al. (2020) emphasise sensory and spatial design for concentration in distraction-heavy environments. Bech-Petersen (2020) advocates human-centred, co-created library design. Pyati (2019) reconceptualises libraries as contemplative civic spaces promoting cognitive well-being.

Methodology: This study adopts a conceptual and design-oriented approach grounded in human-centric principles to reconceptualise library spaces beyond digital dependence. The methodology involves an analytical examination of contemporary digital dependency and its implications for cognitive well-being, followed by a structured exploration of how libraries can serve as supportive environments for intentional digital disengagement. Based on this analysis, the study proposes a typology of digital-free library spaces aligned with academic, creative, social, and well-being needs. The approach emphasises spatial functionality, user experience, and non-digital engagement practices rather than technological intervention.

Proposed Conceptual Framework (Results): This paper reconceptualises library spaces using a human centric design approach. It emphasises the importance of digital-free environments that support focused learning, creative thinking, reflection, and emotional restoration. The study examines how digital dependency influences contemporary human behaviour and explains why libraries are suitable spaces for intentional digital disengagement. To address this need, the paper proposes a range of digital-free library spaces designed to support diverse academic, creative, and personal activities. Digital Silence Spaces provide uninterrupted environments for deep reading, critical thinking, handwritten analysis, and silent intellectual work. Research Writing and Academic Spaces support sustained scholarly engagement, enabling researchers and students to focus on long-form writing, conceptual synthesis, and reflective analysis without digital distractions. Analog Maker spaces encourage hands-on learning through physical creation, promoting experiential understanding and problem-solving skills through material engagement. Creative Spaces promote artistic ex-

pression and imaginative thinking through activities such as sketching, creative writing, visual journaling, and collaborative art practices. Storytelling and Narrative Spaces support oral traditions, autobiographical writing, and shared narratives, strengthening communication skills and preserving personal and community experiences. Art and Culture Incubation Spaces function as platforms for cultural expression, heritage interpretation, and traditional art practices, enabling intergenerational knowledge exchange and cultural continuity. Social and Collaborative Spaces facilitate face-to-face interaction, group study, peer discussions, and collective problem-solving, reinforcing social learning and intellectual cooperation. Well-Being and Relaxation Spaces prioritise emotional and physical restoration through mindfulness practices, gentle movement, quiet rest, yoga, meditation and reflective activities. Green and Sustainable Library Spaces integrate natural elements into learning environments, encouraging ecological awareness, environmental reflection, gardening and sensory restoration. Human Library which can also be considered as one among these digital free spaces enable knowledge exchange through direct human interaction, where individuals share lived experiences, promote empathy, and challenge stereotypes through structured conversations.

Discussion and Conclusion: Beyond spatial design, the paper highlights the role of libraries as facilitators of digital-free cognitive practices. By encouraging print-based engagement, handwritten work, silence literacy, embodied learning, slow reading, and voluntary screen-free routines, libraries can actively support sustained attention, reflective thinking, and cognitive sustainability. Such practices reinforce the library's role as a sanctuary of pause, where stillness and contemplation are socially valued. Library professionals play a central role in planning, implementing, and managing these human-centric spaces. Their responsibilities extend beyond custodial functions to include user-centred planning, space facilitation, policy development, and advocacy for well-being-oriented services. Through participatory design, flexible management strategies, and ethical consideration of user needs, professionals can ensure that digital-free spaces remain inclusive, functional, and responsive.

The implementation of human-centric, digital-free library spaces offers significant benefits, including improved concentration, reduced cognitive fatigue, enhanced creativity, emotional regulation, and strengthened social connection. These benefits extend across students, researchers, and working professionals by supporting healthier learning rhythms and balanced intellectual engagement. While challenges related to space availability, resources, and user adaptation may arise, these are manageable when weighed against the long-term cognitive, emotional, and social advantages. Overall, the paper demonstrates that reconceptualising library spaces towards human-centric, digital-free environments reinforces the library's relevance in a digitally intensive society. By supporting cognitive well-being and inner balance, libraries can evolve into restorative, inclusive, and meaningful spaces that contribute to balanced learning, reflective scholarship, and holistic human development.

Keywords: Digital Dependency, Cognitive Well-being, Human Centric Library Design, Digital-Free Library Spaces, Inner Balance

References

1. Al Badayneh, D. (2013). Human behaviour: when and where virtual society meets physical society. *European Journal of Science and Technology*, 9(1), 105-110.
2. Ashiq, M. (2024). Industry 5.0 – the human centric solution and personalization: Challenges and opportunities for libraries. *Library Hi Tech News*. <https://doi.org/10.1108/LHTN-10-2024-0176>
3. Bech-Petersen, S. (2020). Design thinking as a framework for innovation in libraries. *Bibliothek Forschung Und Praxis*, 44(1), 7–11. <https://doi.org/10.1515/bfp-2020-0009>
4. Bossaller, J., Oprean, D., Urban, A., & Riedel, N. (2020). A happy ambience: Incorporating ba and flow in library design. *The Journal of Academic Librarianship*, 46(6), 102228. <https://doi.org/10.1016/j.acalib.2020.102228>
5. Bussell, H. (2021). Students experience cognitive and emotional responses to academic library spaces. *Evidence Based Library and Information Practice*, 16(3), 149–151. <https://doi.org/10.18438/ebliip29988>
6. Chen, S., Ebrahimi, O. V., & Cheng, C. (2025). New perspective on digital well-being by distinguishing digital competency from dependency: Network approach. *Journal of Medical Internet Research*, 27, e70483. <https://doi.org/10.2196/70483>
7. Cox, A. (2023). Factors shaping future use and design of academic library space. *New Review of Academic Librarianship*, 29(1), 33–50. <https://doi.org/10.1080/13614533.2022.2039244>
8. Denardis, L. (2020). *The internet in everything: Freedom and security in a world with no off switch*. Yale University Press. <https://doi.org/10.2307/j.ctvt1sgc0>

9. Fallin, L. (2016). Beyond books: The concept of the academic library as learning space. *New Library World*, 117(5/6), 308–320. <https://doi.org/10.1108/NLW-10-2015-0079>
10. Garoufali, A., & Garoufallou, E. (2024). Transforming libraries into learning collaborative hubs: The current state of physical spaces and the perceptions of Greek librarians concerning implementation of the “Learning Commons” model. *Global Knowledge, Memory and Communication*, 73(6/7), 828–852. <https://doi.org/10.1108/GKMC-04-2022-0086>
11. Husaini, H., Mohamed Shuhidan, S., Che Daud, S., & Priyanto, I. (2022). Library therapeutic landscape support for covid-19 pandemic-related student well-being in academic libraries. *Environment-Behaviour Proceedings Journal*, 7(SI10), 191–198. <https://doi.org/10.21834/ebpj.v7iSI10.4121>
12. Hynes, M. (2021). The challenges to our innate cognitive abilities and mental well-being. In M. Hynes, *The Social, Cultural and Environmental Costs of Hyper-Connectivity: Sleeping Through the Revolution* (pp. 55–70). Emerald Publishing Limited. <https://doi.org/10.1108/978-1-83909-976-220211004>
13. Jagmohan Meena. (2023). Library spaces and architecture: Designing for the future. *ShodhKosh: Journal of Visual and Performing Arts*, 4(2). <https://doi.org/10.29121/shodhkosh.v4.i2.2023.2192>
14. Johnston, N. (2022). Planning a library of the future: Incorporating the student voice through user perspectives. *Journal of Library Administration*, 62(5), 581–601. <https://doi.org/10.1080/01930826.2022.2083438>
15. Kim, E.-J. (2021). Suggestions regarding universal design direction for user-centered library space and furniture plans. *Research Institute of Human Ecology*, 25(2), 159–172. <https://doi.org/10.36357/johe.2021.25.2.159>
16. Lyon, A. R., Brewer, S. K., & Areán, P. A. (2020). Leveraging human-centered design to implement modern psychological science: Return on an early investment. *American Psychologist*, 75(8), 1067–1079. <https://doi.org/10.1037/amp0000652>
17. Meyer, A., et al., (2024). The evolving role of information professionals in navigating places, spaces, and nurturing new discourses in the in-between. *Proceedings of the Association for Information Science and Technology*, 61(1), 795–798. <https://doi.org/10.1002/pra2.1104>
18. Müller, K. W., & Scherer, L. (2022). Exzessive Nutzungsmuster und Internetnutzungsstörungen: Effekte auf die psychosoziale und kognitive Entwicklung von Jugendlichen. *Praxis der Kinderpsychologie und Kinderpsychiatrie*, 71(4), 345–361. <https://doi.org/10.13109/prkk.2022.71.4.345>
19. Pyati, A. K. (2019). Public libraries as contemplative spaces: A framework for action and research. *Journal of the Australian Library and Information Association*, 68(4), 356–370. <https://doi.org/10.1080/24750158.2019.1670773>
20. Schmitt, J. B., Breuer, J., & Wulf, T. (2021). From cognitive overload to digital detox: Psychological implications of telework during the COVID-19 pandemic. *Computers in Human Behavior*, 124, 106899. <https://doi.org/10.1016/j.chb.2021.106899>
21. Thomas, N. M., Choudhari, S. G., Gaidhane, A. M., & Quazi Syed, Z. (2022). ‘Digital wellbeing’: The need of the hour in today’s digitalized and technology driven world! *Cureus*. <https://doi.org/10.7759/cureus.27743>
22. Vanden Abeele, M. M. P. (2020). *Digital wellbeing as a dynamic construct*. <https://doi.org/10.31219/osf.io/ymtaf>
23. Vanden Abeele, M. M. P., & Nguyen, M. H. (2022). Digital well-being in an age of mobile connectivity: An introduction to the Special Issue. *Mobile Media & Communication*, 10(2), 174–189. <https://doi.org/10.1177/20501579221080899>

Reframing Bibliographic Data: BIBFRAME Adoption in Koha ILS

Eswar M.¹, Vedashawmya K S.², Nisha S.³ and Keerthna M.⁴

¹⁻⁴ Department of Library and Information Science,
Bharathidasan University, Tiruchirappalli – 620024.

liseswar.bdu@gmail.com

Abstract.

Introduction: Koha is one of the widely used and highly regarded integrated library systems in the library world, valued by global communities for its flexibility and open-source foundation. Originally established in 1999 by Katipo Communications for the Horowhenua Library Trust in New Zealand, the system was brought to life by a development team led by Chris Cormack. Today, Koha clearly demonstrates collaborative innovation; however, it faces a crucial challenge how to connect its traditional cataloging foundations with the future of the web. At the core of this challenge lies MARC21, the long-standing standard for library cataloging. For more than fifty years, MARC21's structured fields have been the backbone of how libraries manage collections and share bibliographic data. In Koha, MARC21 is not merely a format but the foundation of key functionalities such as circulation and the Online Public Access Catalog (OPAC).

Related Work: Because MARC21 was developed long before the emergence of the modern web, it has limited compatibility with web search engines and linked data environments. To address these limitations, the Library of Congress introduced BIBFRAME (Bibliographic Framework Initiative) as a successor to MARC21. BIBFRAME shifts library metadata from a record-based model to an entity-relationship model based on the Resource Description Framework (RDF). Unlike MARC21, which represents data as discrete records, BIBFRAME structures information as interconnected entities Works, Instances, and Items linked through Uniform Resource Identifiers (URIs). Prior studies highlight BIBFRAME's potential to enhance interoperability, discoverability, and machine-readability of library data within the semantic web.

Methodology: This study presents a MARC21 to BIBFRAME conversion workflow (Fig. 1) using Koha as a case study. The process begins with the extraction of bibliographic and authority data from the Koha database in MARCXML or ISO 2709 formats. XSLT-based conversion tools provided by the Library of Congress, along with middleware such as LOD Refine, are used to map MARC fields to corresponding BIBFRAME classes. The workflow further enriches data by reconciling textual strings with external authority files such as VIAF and LCNAF, assigning persistent URIs.



Fig. 1. MARC21 to BIBFRAME conversion Workflows

Results: The conversion workflow transforms static MARC21 records into dynamic, linked bibliographic entities. The resulting BIBFRAME data enables improved visibility of library resources on the web, allowing records to appear in external search engines and cultural heritage platforms. The enriched metadata demonstrates enhanced interoperability and semantic clarity without disrupting Koha's existing MARC-based operations.

Discussion and Conclusion: Despite the advantages of BIBFRAME, MARC21 remains critically important for Koha users. Koha's internal architecture is fundamentally record-centric, and a complete transition to an entity-based model would require significant technical and financial investment. Additionally, MARC21 continues to be the dominant international standard, and catalogers

are extensively trained in its use. Since many BIBFRAME tools remain experimental, an abrupt replacement of MARC21 is neither feasible nor necessary. Instead, a phased hybrid approach is recommended, where MARC21 continues to support daily cataloging while BIBFRAME functions as a linked-data publication layer.

References

1. Casalini, M. (2017). Implications of BIBFRAME and linked data for libraries and publishers. Roll With the Times, or the Times Roll Over You. Charleston Conference. <https://doi.org/10.5703/128828431644>.
2. Dutta, B., & Bhuvaneshwari, V. (2023, October). Towards standardizing the library circulation metadata. In Research Conference on Metadata and Semantics Research (pp. 117-131). Cham: Springer Nature Switzerland.
3. Kim, M., Chen, M., & Montgomery, D. (2021). Moving toward bibframe and a linked data environment. In S. S. Hines (Ed.), *Advances in Library Administration and Organization* (pp. 131–154). Emerald Publishing Limited. <https://doi.org/10.1108/S0732-067120210000042011>.
4. Kroeger, A.J. (2013). The Road to BIBFRAME: The Evolution of the Idea of Bibliographic Transition into a Post-MARC Future. *Cataloging & Classification Quarterly*, 51, 873 - 890.
5. McCallum, S.H. (2017). BIBFRAME Development.
6. Pal, M. (2025). Role of Librarians in the Era of Digital Transformation. *International Journal of Advanced Research and Multidisciplinary Trends (IJARMT)*, 2(3), 952-962.
7. Park, J.-R., Brenza, A., & Richards, L. (2020). Bibframe linked data: A conceptual study on the prevailing content standards and data model. In K. Okoye (Ed.), *Linked Open Data—Applications, Trends and Future Developments*. Intech Open. <https://doi.org/10.5772/intechopen.91849>.
8. Samples, J., & Bigelow, I. (2020). Marc to bibframe: Converting the pcc to linked data. *Cataloging & Classification Quarterly*, 58(3–4), 403–417. <https://doi.org/10.1080/01639374.2020.1751764>.
9. Schreur, P. E., & Carlson, A. J. (2020). Bridging the worlds of marc and linked data: Transition, transformation, accountability. *The Serials Librarian*, 78(1–4), 48–56. <https://doi.org/10.1080/0361526X.2020.1716584>.
10. Steele, T. (2019). What comes next: understanding BIBFRAME. *Libr. Hi Tech*, 37, 513-524.
11. Xu, A., Hess, K., & Akerman, L. (2018). From MARC to BIBFRAME 2.0: Crosswalks. *Cataloging & Classification Quarterly*, 56, 224 - 250.
12. Zavalin, V. I., & Miksa, S. D. (2021). Collecting and evaluating large volumes of bibliographic metadata aggregated in the WorldCat database: A proposed methodology to overcome challenges. *The Electronic Library*, 39(3), 486–503. <https://doi.org/10.1108/EL-11-2020-0316>.

Bridging the Technological Gap in Library Information Retrieval with Integration of Elastic Search in Koha

Madhu Bala S.¹ [0000-0002-0883-9375], Marimuthu R.² [0000-0001-9758-1582], and Jeyapragash B.³ [0000-0001-5779-7127]

^{1,3} Bharathidasan University, Tiruchirappalli, Tamil Nadu, 620024, India.

² PSG College of Technology, Coimbatore, 641004, India.

madhubaladlisc@gmail.com

rmm.lib@psgtech.ac.in

bjeyapragash@bdu.ac.in

Abstract.

Introduction: The library transferred to a digital information retrieval system from the conventional catalogue card system long ago. In this era of technological advancement, significantly changes the accessibility of information. The library adopts various technologies to effectively complete tasks such as purchasing, cataloguing, remote access, information retrieval and other library activities. Modern technologies like artificial intelligence, deep learning, augmented reality, edge computing and big data have made their influence in every sector, including library and information sciences. In library multiple systems were used for information retrieval, including card catalogue system, Online Public Access Catalogue (OPAC), several search engines and indexing frameworks, were used namely Lucene search library, solr full text search and elastic search. Elasticsearch was adopted by Koha in response to the need for enhanced user search experience and search performance in library. Koha is an open-source software for library management; it uses zebra search engine as its default search engine. Libraries can improve relevance and adoptability in information retrieval by bridging the technological gap between traditional search engines and contemporary discovery searches through the integration of Elasticsearch. This article focuses library information retrieval, specifically the adaptation of elastic search in the library management system Koha to accelerate the information retrieval in the library. Both zebra and elastic searches appeared similar from the user perspective, the difference is evident when it comes to the relevance and accuracy of the information retrieved from the search results. Since the Koha version 19.11 supports Elastic Search along with zebra search engine. The later updated versions of Koha include additional Elasticsearch-specific system preferences, like “ElasticsearchBoostFieldMatch”, it enable to find exact search term the user entered and “ElasticsearchCrossField” to search from all the searchable indexes and “ESPreventAutoTruncate” finalises which elastic indexes are not to be auto-truncated in search results. This can handle large-scale data with near real-time data indexing, retrieval and advanced query capabilities. Elasticsearch indexes and processes the data in JSON format, which can easily handle text as well as numbers. RESTful API enables flexible integration between Koha and Elasticsearch from any programming language seamlessly by providing HTTP endpoints. This will support Koha to fetch and parse information quickly.

Related Work: Gupta, M. (2023): The study highlights the modular design and advantages for bibliographic control of the open-source ILS Koha, which transforms cataloguing and classification processes in academic libraries. The article describes general library management procedures were improved by implementing Koha. Neupane, R. (2023): The study highlights Koha's versatility and role in contemporary library automation. The study investigates customization of koha usability in libraries. It concludes that altering the Koha interface greatly increased user satisfaction and access to its features, demonstrating the significance of modifying ILS interfaces for improved user experience like Elasticsearch integration. Kulkarni, A. M., Pandiyani, M., & Patankar, G. P. (2023): The article emphasizes Koha's versatility and feature set while examining its usefulness as an open-source integrated library system. It discusses the requirements of contemporary libraries and delicately reinforces the importance of sophisticated search features.

Methodology: The study adopts an experimental approach to explore the features of the elastic search an advanced search engine integrated with the koha Library management system. The information were collected from official koha manuals for version 24 also from research and research

articles. The study highlights the benefits of elastic search integration with koha to improve information retrieval system in library.

Results: Elasticsearch indexes and retrieves new records immediately after cataloguing, enabling near real-time search. Features like fuzzy searches, wildcard searches and cross-field searches strengthen the search experience apart from simple exact keyword matching. Full-Text and metadata search support information retrieval from structured metadata, such as title and author, which are well-defined data. And also unstructured metadata fields to enhance search coverage, from the fields 500 notes, 505 table of contents, 520 abstract/ summary. In faceted search user can narrow down the search results from predefined metadata categories like title, author, and subject. Search engine configuration in Koha administration manages indexes, facets, weightage and their mapping to MARC fields. Numbering the weights based on the predefined patron search pattern. The common command language can be used in Elastic Search for advanced query building. Browse Search is configured to improve search precision, relevance ranking and browsing efficiency within the OPAC. It will enhance user contribution over information retrieval. Zebra search can provide results of similar-sounding words, although it is not well defined to the specific term the user searched, it shows all possible results. The browse search supports users in accessing relevant information. Three field-based searches are available as title, author, and subject. There are three levels of keyword matching; exact search retrieves record that match text precisely by the exact keyword of the item indexed. Fuzzy and really fuzzy search match the user keyword with typos, misspellings, and variations in searching terms for faster retrieval. The major difference visible is that the basic search just goes through the title and provides results similar to that example; if searching for “physics”, it also lists “physical education”. The study expresses the importance and recommends library professionals to upgrade to elastic search for better information retrieval.

Discussion and Conclusion: Integration of elastic search with koha will be a potential technological bridge gap in information retrieval in koha. It includes advanced search features, relevance ranking and improved performance. Elastic search in koha staff interface and koha OPAC will be modernize the search capacities providing user centric discovery experience. Elasticsearch's advanced full-text search features allow Koha to move beyond simple keyword retrieval and into intelligent discovery.

Keywords: Elastic Search, Koha, Library Search, OPAC, Information Retrieval

References

1. Antelman, K., Lynema, E., & Pace, A. K. (2006). Toward a twenty-first century library catalog. *Information Technology and Libraries*, 25(3), 128–139.
2. Barman, D., Dutta, A., & Mukhopadhyay, P. (2025). Full-Text eBook retrieval system in an Integrated library management system: Designing an open- source technical framework. *Journal of Information and Knowledge*, 223–234. <https://doi.org/10.17821/srels/2025/v62i4/171706>
3. Garg, M., Hasan, N., & Gupta, A. (2023). Implementation of Koha in Managing the E-resources of the Library. *Journal of Information and Knowledge*, 60(2), 127–132.
4. Gupta, M. (2023). Using Koha for Cataloguing and Classification: A Case Study. *International Journal of Multidisciplinary Innovation and Research Methodology*, 2(2), 51–56.
5. Koha Searching Manual (latest) – Official Koha documentation on search behavior when using Elasticsearch. Koha Community. <https://koha-community.org/manual/25.05/en/html/searching.htm>
6. Neupane, R. (2023). Customization of KOHA Integrated Library System (ILS) Interface for Better Usability. *Access: An International Journal of Nepal Library Association*, 2(01), 245–256.
7. Nyambaka, S. B., & Mutwiri, C. (2023). Usability of KOHA OPAC by Undergraduate Users for Information Retrieval with Regard to Usability Testing in Technical University of Kenya Library. *International Journal of Current Aspects*, 7(1), 37–50. <https://doi.org/10.35942/ijcab.v7i1.302>
8. Ritesh Kumar. (2024). Query Optimization in Elasticsearch: A Comparative Analysis of Ranking Strategies. *International Journal for Multidisciplinary Research*, 6(1). <https://doi.org/10.36948/ijfmr.2024.v06i01.38173>
9. Stephens, M. K. (2022). Review of Koha Library Software. *TCB Technical Services in Religion & Theology*, 30(2):7–9.
10. Vidhya, R., & Vadivu, G. (2016). Research Document Search using Elastic Search. *Indian Journal of Science and Technology*, 9(37). <https://doi.org/10.17485/ijst/2016/v9i37/102108>
11. Wang, S. (2025). Elastic Architecture Search for Efficient Language Models. 2025 IEEE International Conference on Multimedia and Expo (ICME), 1–6. <https://doi.org/10.1109/icme59968.2025.11210235>

12. Wang, S. (2025). Impact of smart library construction on information retrieval efficiency. *Journal of Intelligence and Knowledge Engineering*, 3(2), 36–42. <https://doi.org/10.62517/jike.202504206>
13. Yadav, G., & Hemlata. (2024). KOHA LMS (Library Management System) Managing in the Tagore Library of Chitkara University, Himachal Pradesh: A study. *International Journal of Research in Library Science*, 10(3), 133–142. <https://doi.org/10.26761/ijrls.10.3.2024.1782>

About the Editor: Dr. Biswanath Dutta is an Associate Professor and Former Head of the Documentation Research and Training Centre (DRTC) at the Indian Statistical Institute (ISI), Bangalore. He holds degrees in Physics and Library and Information Science from Vidyasagar University and ISI Bangalore, respectively, and earned his Ph.D. from the University of Pune, with a research focus on Semantic Web technologies.

Dr. Dutta has held postdoctoral positions at the University of Trento, Italy, and has also been associated with Dalhousie University, Canada. He has actively collaborated on European Union-funded research projects and currently leads both internally and externally funded research initiatives. In addition, he offers consultancy services to industry in the areas of data modeling, knowledge representation, and semantic systems.

He has published over 100 scholarly articles and authored four books. His research interests span ontology engineering, knowledge graphs, metadata frameworks, data management, data science, and semantic technologies. Further details about his academic and research contributions are available at: <https://sites.google.com/view/bdutta/>



Documentation Research and Training Centre
Indian Statistical Institute
Bangalore, India
<http://drtc.isibang.ac.in>

ISBN: 978-93-5619-660-5