Libraries and Librarianship in the Era of *Renewed* Artificial Intelligence

Biswanath Dutta

Assistant Professor

DRTC, Indian Statistical Institute (Bangalore Centre)

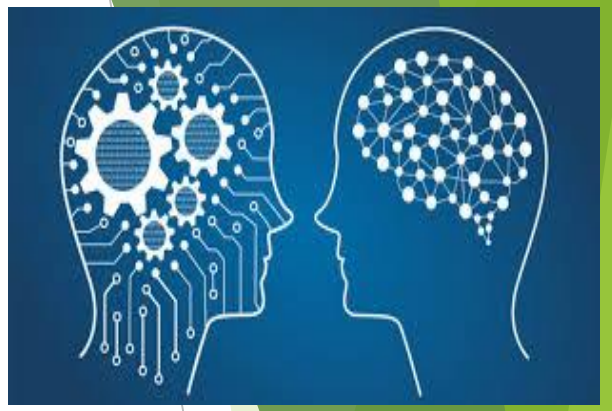
Bangalore, INDIA

Email: bisu@drtc.isibang.ac.in

Artificial Intelligence(AI) is pervasive

- Apple's Siri, Amazon's Alexa, Driverless car
- Google knows what we want to know based on what we search
- Google knows what is on our calender or what is in our email
- System that is capable of alerting us on when to leave for an appointment

The challenges for libraries



- Increasing demand and expectations of the users
  - Complex queries
    - (e.g. “Give me documents about a factory in England established by Richard Arkwright during industrial revolution”)
  - Varieties of collections
  - Expectations of receiving **speedy** and **Smart** information services (especially when they are surrounded by smart tools like Siri, Alexa, Google Assistant, etc.)
  - Increased specialization in research
- Increasing demand of the parent organization
- Limited resources
  - Challenge in utilizing the resources (e.g., the budget, human resources) in a **smart** way

The goal of the talk

- Is AI a threat to libraries and librarianship?
- Can we take advantage of AI in improving the library users experience? If yes, how?
- Can libraries contribute in any means to the creation of AI? If yes, how?

Rest of the presentation: Highlights

- Artificial Intelligence (AI), its purpose and some real world applications, AI concerns
- Opportunities
  - DERA: from Knowledge Organization (KO) to Knowledge Representation (KR) and vice versa and their convergence
  - Other opportunities

Artificial Intelligence

- It “*is often used to describe machines (or computers) that mimic “cognitive” functions that humans associate with the human mind, such as "learning" and “problem solving”* ” (Russell and Norvig [18])

- The goal is to design “intelligent” machines that can work and react more like humans.

AI Approaches

- Logic and rule-based
  - Knowledge representation (KR), logic enabled KR language and rules
- Machine learning
  - Pattern-based

AI types

- **Analytical**
  - Based on cognitive intelligence, learn from the past experience to inform future decisions.
- **Human inspired**
  - Based on cognitive and emotional intelligence, understand and consider the human emotions in decision making.
- **Humanized**
  - Based on all types of competencies, e.g., cognitive, emotion, social intelligence.

AI for various purposes

- **Content organization** and making accessible the large collections of information (e.g. Google Life Tags)
- **Complex query search** (instead of a mere keyword based search) and retrieval (e.g., Google Talk to Books, Semantic Scholar by Allen Institute for Artificial Intelligence)
- **Content moderation** (e.g., Facebook's AI language processing system for filtering out the spam and abusive comments from user's newsfeeds; deep neural networks to identify particular objects in a photo and pick out particular characteristics of the people in the photo to create a caption that a text-to speech engine can then read aloud for users with low visibility)
- **Content generation** (e.g. short story, narratives, news reporting)

AI for various purposes (contd…2)

- **Content evaluation** (e.g., neural networks developed by Disney and the University of Massachusetts Boston that can evaluate short stories **to predict which stories will be most popular**)

- **AI in Education** (e.g., IBM Teacher Advisor With Watson (https://teacheradvisor.org/) to build personalized lesson plans)

- **Many more applications across the domains**: healthcare, automotive, law, military, economics, predictive policing, etc.

Real World Applications

- Google's Life Tags (https://artsexperiments.withgoogle.com/lifetags/)
- Google's Talk to Book (https://books.google.com/talktobooks/)
- GeoDeepDive (http://i.stanford.edu/hazy/geo/)
- Ross: AI attorney (https://searchenterpriseai.techtarget.com/definition/artificially-intelligent-attorney-AI-attorney)
- …

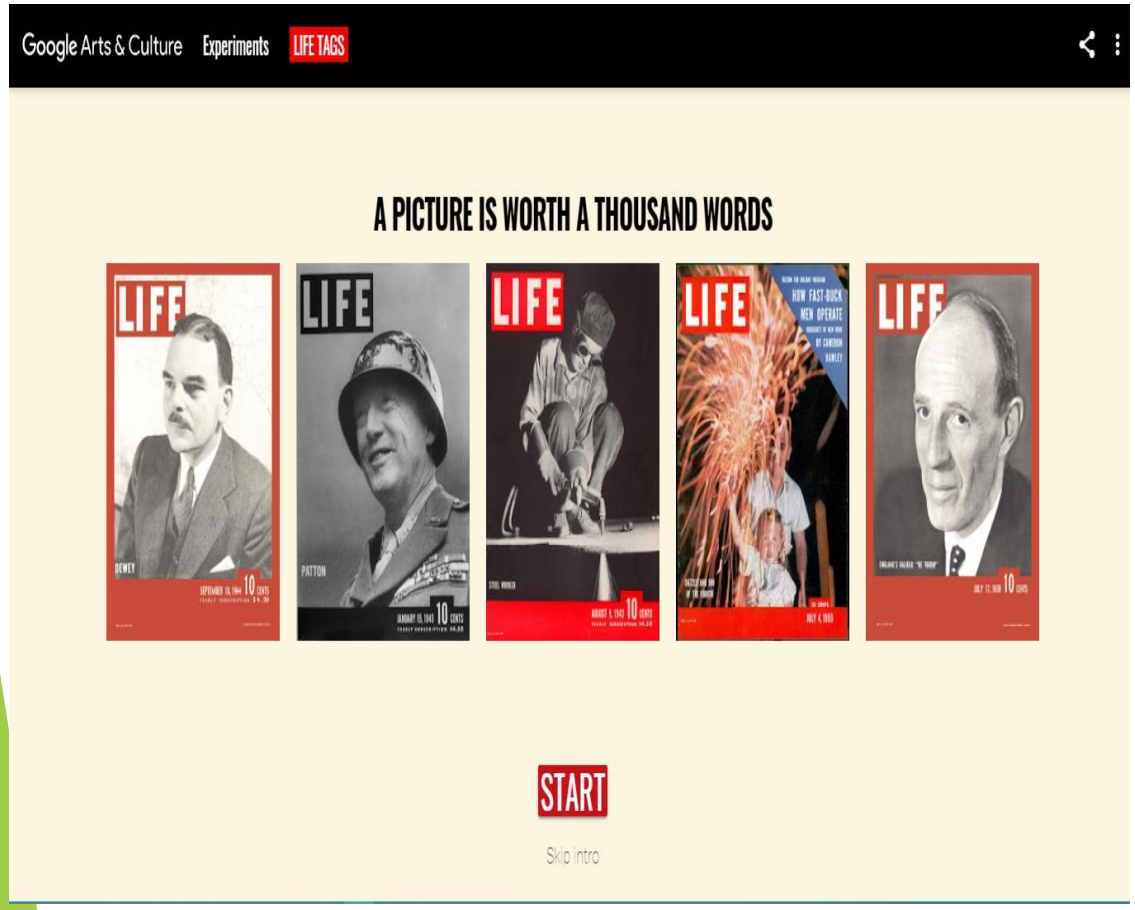
Life Tags

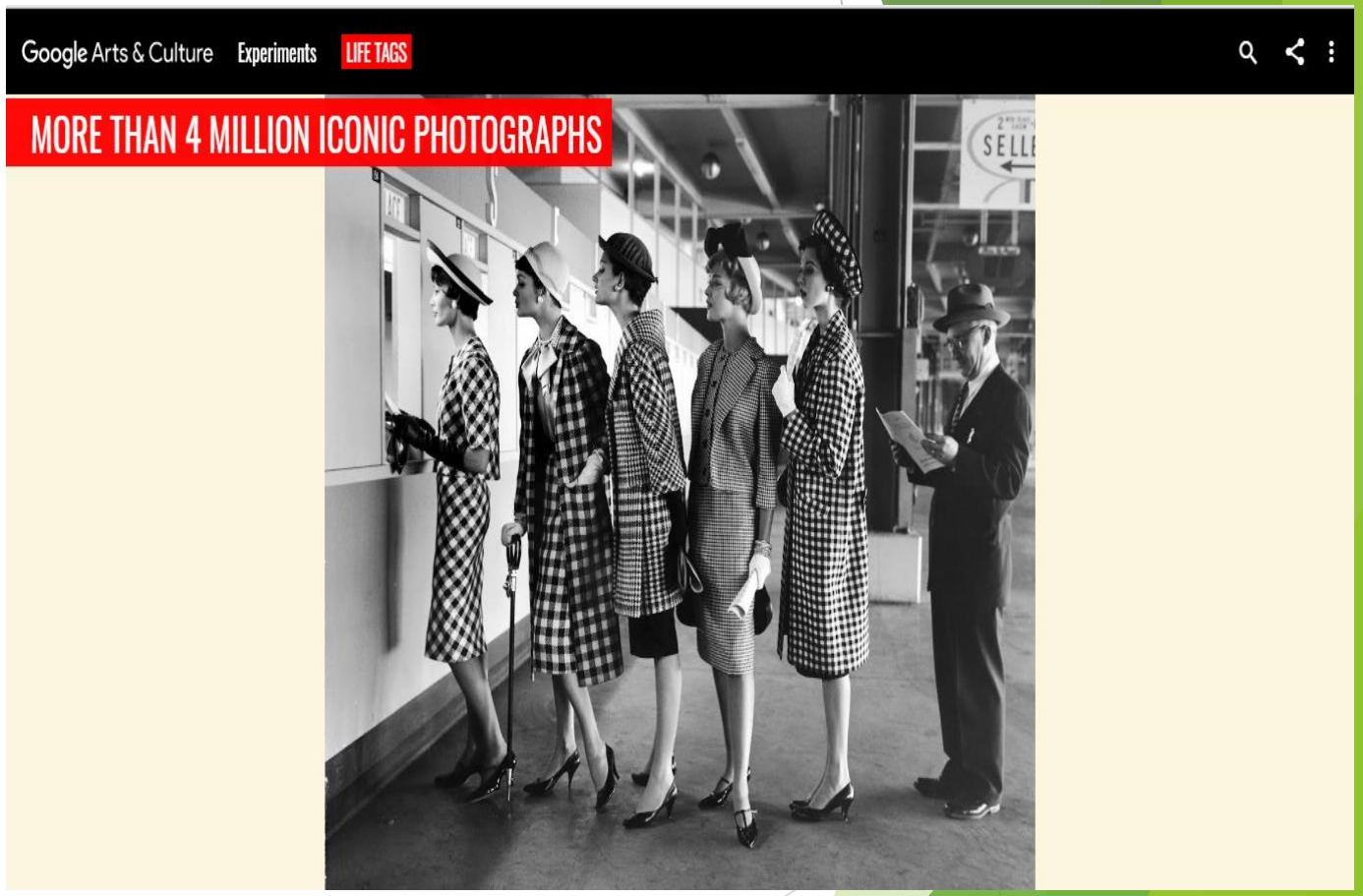
- Life Tags was created by Gael Hugo as part of the Arts & Culture Experiments Collection from “Experiments with Google”

- Life Tags uses AI technology to intelligently sort through, analyze, and tag over **4 million photos** from LIFE Magazine’s publicly available archives.

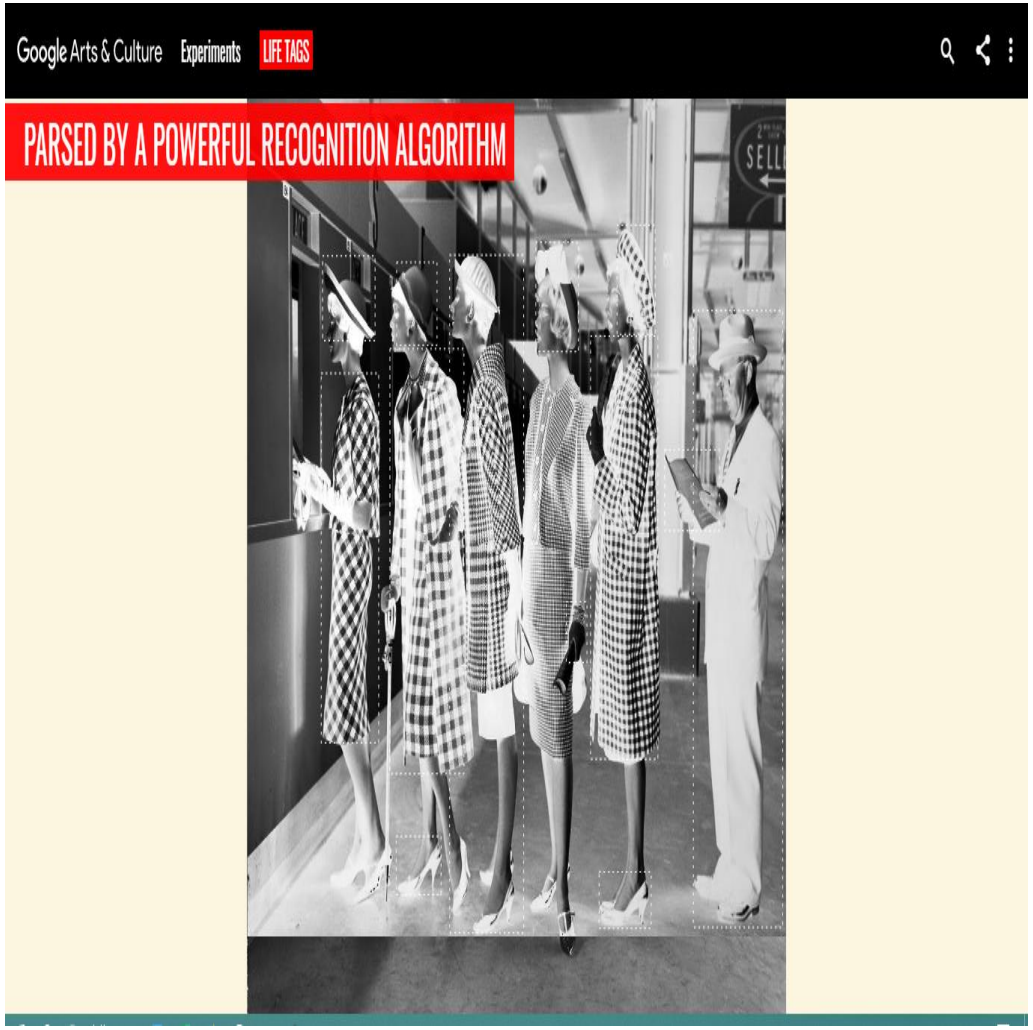
[1]

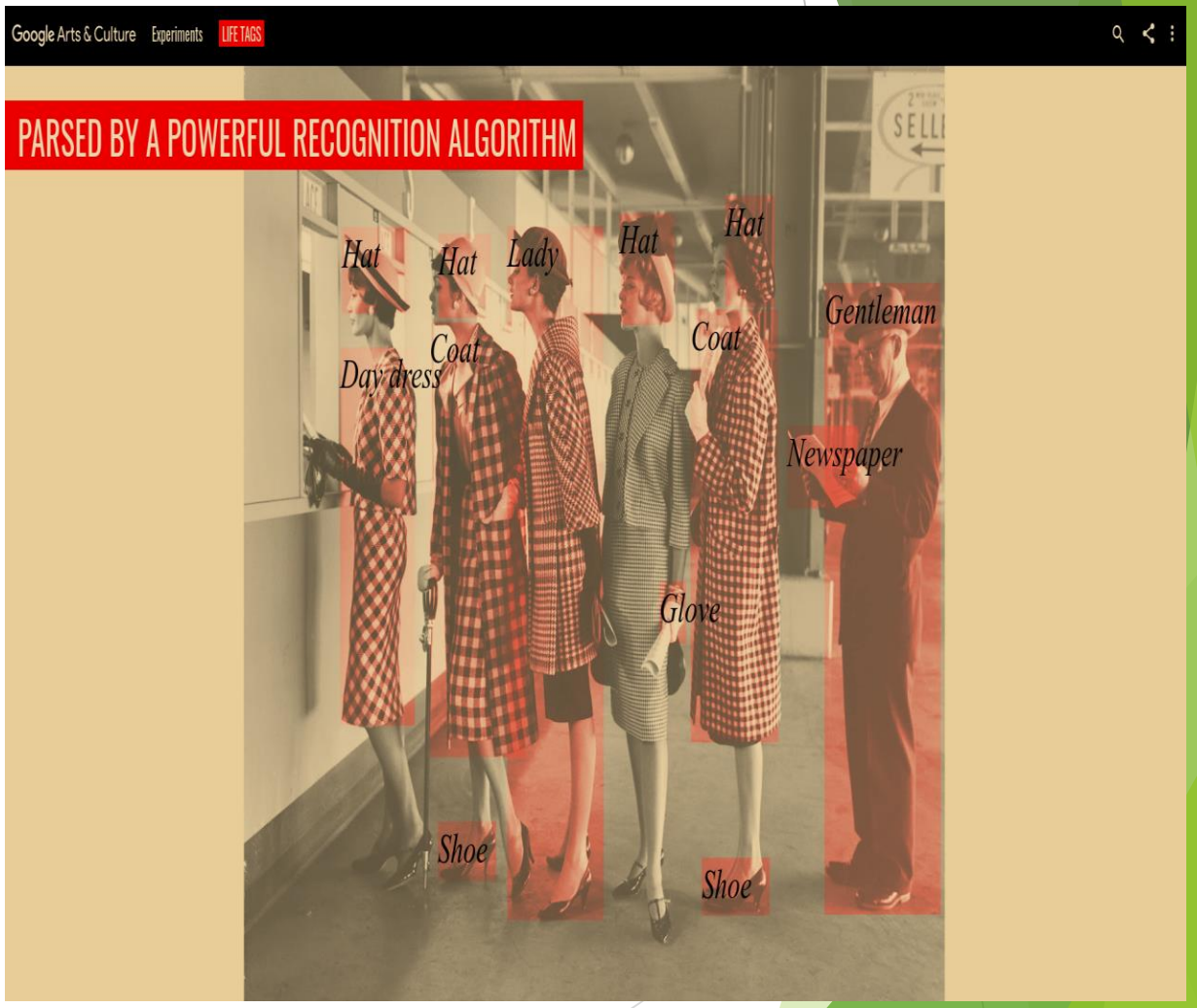
How does Life Tags work?



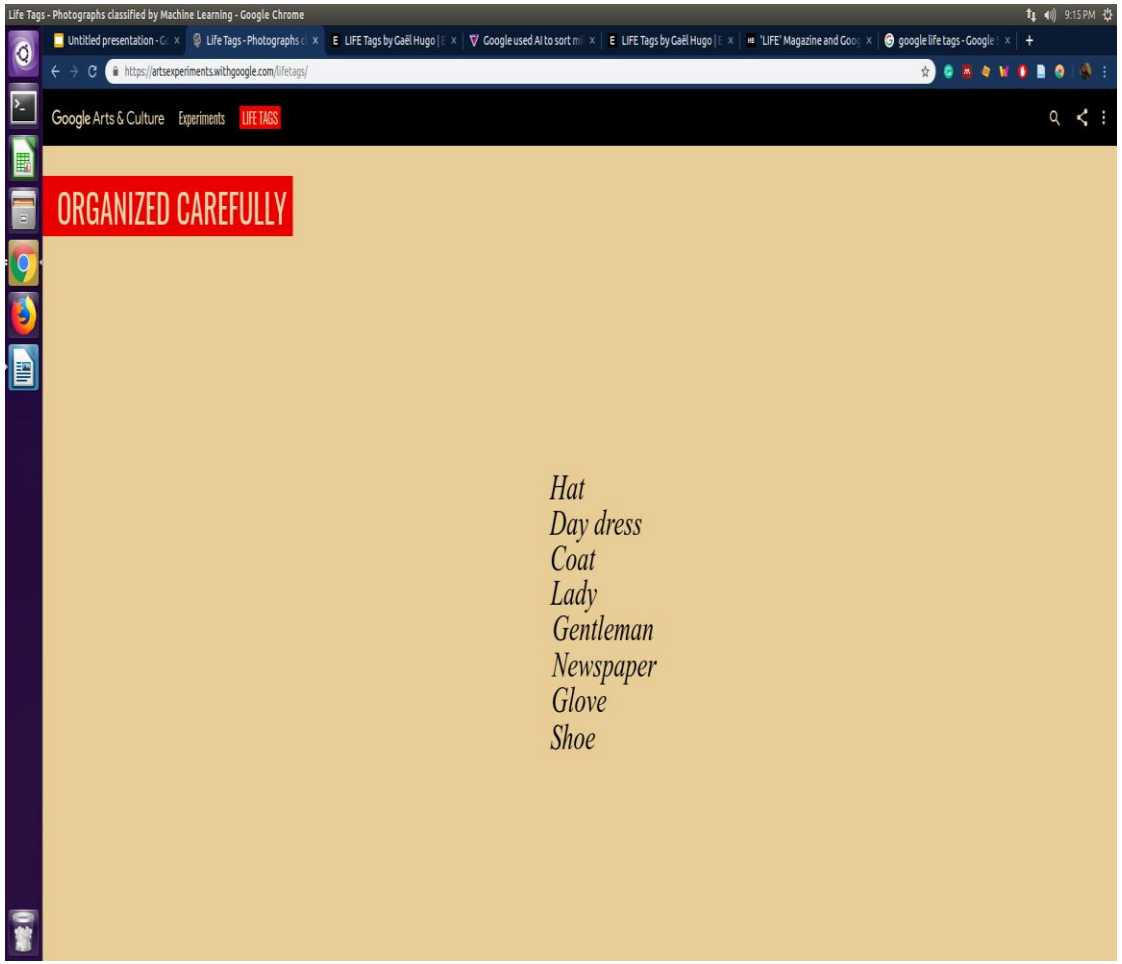


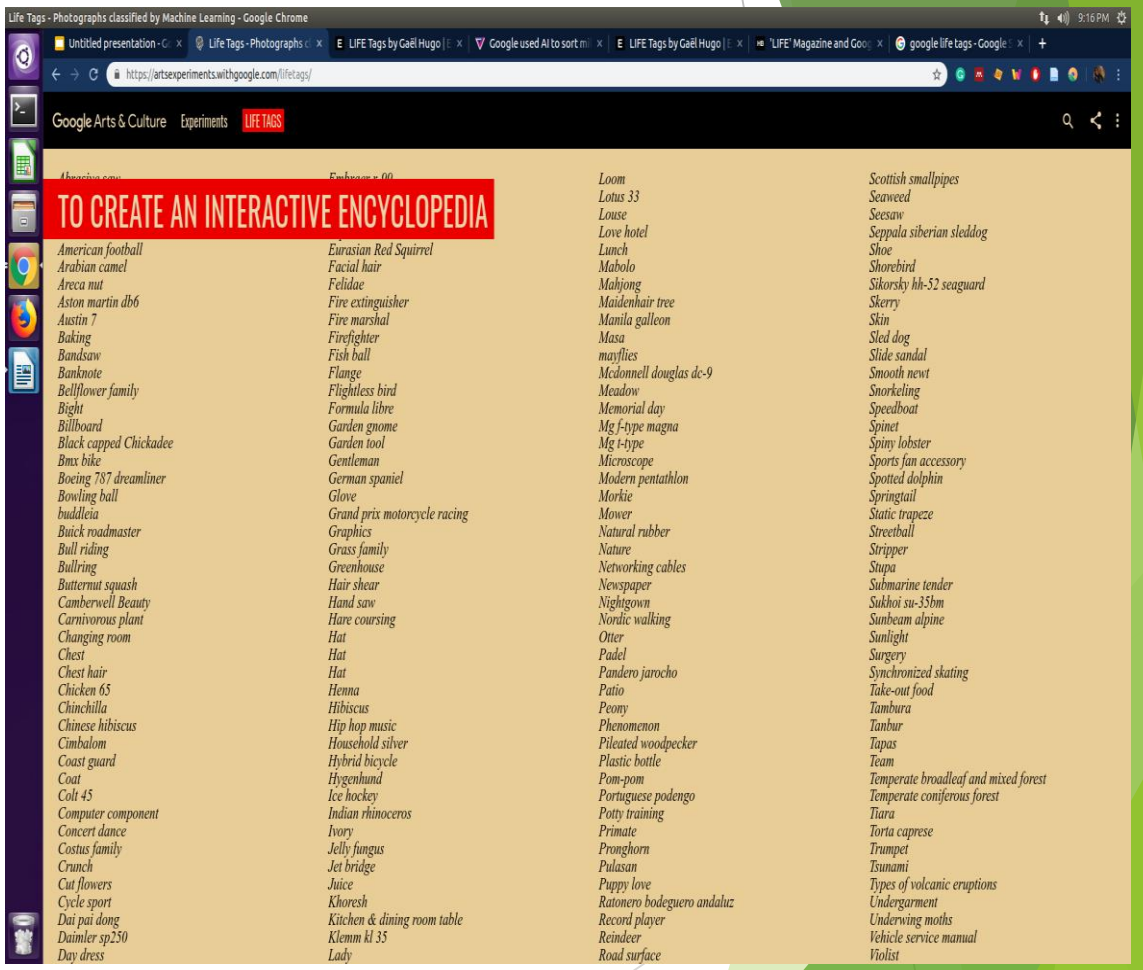
How does Life Tags work?





How does Life Tags work?





Google Arts & Culture Experiments LIFE TAGS

Cane

Cane is any of various tall, perennial grasses with flexible, woody stalks, and more specifically from the genus Arundinaria. Scientifically speaking, there are either of two genera from the family Poaceae. The genus Arundo is native from the Mediterranean Basin to the Far East. The genus Arundinaria is a bamboo (Bambuseae) found in the New World.

cane (1) **cane corso** (21) **cane toad** (15)

cani cross (2) canidae (35255) canis (255)

canis lupus tundrarum (65) canna family (4)

canna lily (5) cannabis (2) canning (22)

cannoli (1) cannon (1363)

Canoe

A canoe is a lightweight narrow watercraft, typically pointed at both ends and open on top, propelled by one or more seated or kneeling paddlers facing the direction of travel using a single-bladed paddle. In International Canoe Federation nomenclature used in some European countries such as the United Kingdom the term canoe refers to kayaks, while canoes are called Canadian canoes. Canoes are used for racing, whitewater canoeing, touring and camping, freestyle, and general recreation.

canoe (3699) **canoe birch** (378)

canoe polo (4) **canoe slalom** (243)

canoe sprint (242)

canoeing (1883) canola (5)

Canopy

canopy (257) **canopy bed** (139)

canopy walkway (7)

cantaloupe (6) cantilever bridge (661)

canvasback duck (2) canyon (1113)

canyoning (6) cap (9181) capacitor (6)

cape (341) capelin (12) caper family (1)

capoeira (58) capsule (7)

burrowing rodents of the family Geomyidae. About 35 species of gophers live in Central and North America. They are commonly known for their extensive tunneling activities. Gophers are endemic to North and Central America. The name "pocket gopher" on its own may be used to refer to any of a number of genera within the family.

gopher (74) **gopher snake** (5)

gopher tortoise (57) **gopher tortoise** (34)

gordon setter (24) gospel music (31)

goth subculture (615)

GOTHIC ARCHITECTURE (2788)

gothic fashion (196) gourd (65)

government (1687) government agency (101)

gown (48047) graduated cylinder (10)

graduation (2228) graffiti (202) grain (139)

PAINT (339)

Paintball

Paintball is a game developed in the 1980s in which players eliminate opponents from play by hitting them with dye-filled, breakable, oil and gelatin paintballs, or pellets, usually shot from a carbon dioxide or compressed air (Nitrogen) powered "paintball marker". The game is regularly played at a sporting level with organized competition involving major tournaments, professional teams, and players.

paintball (15) **paintball equipment** (10)

paintball marker (2)

painted turtle (11) painter (128)

painting (60667) paisley (36) pajamas (58)

palace (5324) pale smartweed (1)

pall-bearer (194) pallet jack (2) pallone (666)

PALM TREE (6799)

tonbak (3) tongue (82)

tongue-and-groove pliers (6) tonkinese (20)

Tool

A tool is any physical item that can be used to achieve a goal, especially if the item is not consumed in the process. Tool use by humans dates back millions of years, and other animals are also known to employ simple tools. Tools that are used in particular fields or activities may have different designations such as "instrument", "utensil", "implement", "machine", "device," or "apparatus".

tool (458) **tool accessory** (53)

tool and cutter grinder (27)

toolbox (4) toolroom (3389) tooth (3384)

tooth brushing (66) toothbrush (8)

toothpick (1) top (8) torball (1) torch (3)

torch lily (3) torii (177) tornado (11)

tornjak (45) torpedo boat (380) torta (5)

torta caprese (3) torte (487) tortoise (552)

tosa (4) tote bag (30) totem (32)

totem pole (79) toucan (24)

Touch

touch football (355)

touch football (american) (64)

touch rugby (13)

touchdown (34) touchpad (3)

tour bus service (18) touring car (15)

touring car racing (4) tourism (59802)

tourist attraction (7070) tournament (23235)

tow truck (25) towed water sport (1296)

towel (35) tower (12916) tower block (8638)

town (616543) town square (11437)

townsend's warbler (1)

Life Tags: search for “Tug of War”

Talk to Books

- It is to interact with the books.
- Against a question or a statement, the AI algorithms look for conversational responses at every sentence in over **100,000 books**.
- The response sentence is shown in **bold**, along with some of the text that appeared next to the sentence for context.

Talk to Books

Talk to Books

Is apple good for health?

Go!

Passages from books

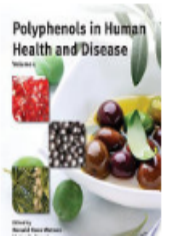
Filters

Showing results from all categories

...every day is probably a smart idea. **Apples are chock-full of nutritional goodies, from pectin, a soluble fiber that helps lower cholesterol, to quercetin, a natural chemical that appears to reduce the risk of heart disease, Alzheimer's, Parkinson's and prostate and lung cancers.** And every time you crunch an apple,... (view in book)

from The Looneyspoons Collection
by Janet Podleski, Greta Podleski
Hay House, 2012 • Recreation and Leisure

In numerous epidemiological and cohort studies, apples have been associated with a decreased risk of cardiovascular disease.36,37 Regular consumption of fruits and vegetables with fiber, including apples, as part of a healthy diet may help in the prevention of chronic disease and maintenance of good health. Further... (view in book)



Filter books by category

- [ ] Arts
- [ ] Current events
- [ ] Fiction
- [ ] History and Biographies
- [ ] Literary Criticism and Collections
- [ ] Medicine and Health
- [ ] Philosophy and Self-help
- [ ] Recreation and Leisure
- [ ] Reference

Go

Talk to Books

BUY EBOOK - ₹2,431.66

Get this book in print ▼

★★★★★
1 Review
Write review

The Looneyspoons Collection

By Janet Podleski, Greta Podleski

"Apples are chock full o Go

About this book

▸ My library

▸ My History

Books on Google Play

Terms of Service

Pages displayed by permission of Hay House, Inc.

Result 1 of 1 in this book for "Apples are chock full of nutritional goodies from pectin a soluble fiber that helps lower cholesterol to quercetin a natural chemical that appears to reduce the risk of heart disease Alzheimer's Parkinson's and prostate and lung cancers"

Recipe Tip Toasting nuts before using them in recipes intensifies their flavor. To toast nuts, place them in a dry skillet over medium heat. Shake the pan often and toast nuts for 4 to 5 minutes, until fragrant. Cool before using. Because nuts have a high fat content, they go rancid quickly. The best way to store shelled nuts is in an airtight container in the refrigerator or freezer.

FOOD BITE

Waldorf salad, originally a mixture of apples, celery and mayonnaise, was invented in 1893 by Oscar Tschirky, a maître d' at the Waldorf Astoria Hotel in New York City.

Good to the Core

You know what they say, "Two apples a day gets the doctor's OK!" Well, that's not exactly what they say, but eating two apples every day is probably a smart idea. Apples are chock-full of nutritional goodies, from pectin, a soluble fiber that helps lower cholesterol, to quercetin, a natural chemical that appears to reduce the risk of heart disease, Alzheimer's, Parkinson's and prostate and lung cancers. And every time you crunch an apple, you may be doing your eyes a favor, too. One of the flavonoids in apples actually protects against cataracts. So what "they're" really saying is that regular daily consumption of apples keeps the cardiologist, the oncologist and the ophthalmologist at bay!

GeoDeepDive

- GeoDeep Dive (http://i.stanford.edu/hazy/geo/) is a tool for geologists designed using machine learning.
- The **goal** is to extract data about rock formations that is buried in the text, tables, and figures of journal articles and web sites, sometimes called dark data.
- Its infrastructure can be repurposed on other data sources to build our own applications [see https://github.com/UW-Deepdive-Infrastructure/app-template/wiki].
    - VideoClip

[2]

ROSS: an AI attorney

- ROSS is a **legal expert system** that applies AI technologies to replicate and improve upon the abilities of a human legal research assistant
- It is built on IBM's Watson cognitive computing platform.
- It depends on self-learning systems that use data mining, pattern recognition and natural language processing to mimic the way the human brain works.
- ROSS can mines data from billion text documents, analyze the information and provide precise responses to complicated questions.
- It **supports natural language queries**.

[3]

Should the libraries be exploring how these kinds of tools can be put into use in improving the library information activities and services?

Some AI concern

- Increasing concern of unemployment
- Sometimes things may go terribly wrong
  - E.g., Google's photo application labelled black people as gorillas [17]
  - Microsoft Tay, a chatter bot, caused subsequent controversy when the bot began to post worst racist sexist and other sorts of offensive tweets through its Twitter account [16]
  - Pizza robot: dough vs. baby

Libraries and librarianship: Leveraging AI

- For better information retrieval
- Better information services (advanced SDI (**adaptive??**))
- Cataloguing and organizing our collections
- Designed smart subscription module analyzing and understanding the **real need of the users**
- Reference services
  - AI chat bot to assist the reference librarians to provide a better service
- Leveraging AI for recommendation systems
- Helping the scholars in finding the right venue (e.g. journal) for publishing their works
- Smart user assistive systems
  - E.g., user orientation, in museum in object description in story telling manner
- Smart surveillance

[5]

AI sub-problems

- Logic
- Learning
- Natural Language Processing
- Perception
- Motion, manipulation
- **Knowledge representation**
- …

DERA: from Knowledge Organization (KO) to Knowledge Representation (KR) and vice versa and their convergence

Based on our **earlier works** in [11, 12] and the presentation available here https://slideplayer.com/slide/14716570/

What is KR in AI?

- It is a medium of human expression about the world.
- It enables an entity to determine consequences by thinking rather than acting, i.e., by reasoning about the world rather than taking action in it.
- It is a medium for pragmatically efficient computation, i.e., the computational environment in which thinking is accomplished.

[19, 6]

KR strength

- KR has developed very powerful and expressive techniques which via the use of **ontologies** support queries by any entity property.
- KR is concerned with the **development of ontologies** describing the relevant entities of a domain in terms of their basic properties.
  - This enables an effective communication and information exchange, as well as **automated reasoning**.

[7, 8]

KR issues [there are many]

KR has failed as it lacks of appropriate entity specification methodologies.

KO and its strength

- KO as a process aims to organize the knowledge in the form of classification systems which are used to represent knowledge in documents/ things.
- Historically the KO approach has scaled as it follows for the classification, indexing and search of millions of books (though at very high costs of training and maintenance).
- Several methodologies have been developed for the construction and maintenance, often centralized, of controlled vocabularies.
- Faceted approach is known to have great benefits in terms of quality and scalability of the developed resources.

[9, 10, 11]

KO: make information available

[10]

Searching for books manually

Searching for books electronically with OPAC

Searching for books by document properties

[10]

KO Limitations: Limited reasoning

- The relations between the terms are implicit and/or linguistic
- Not intended to represent the fine granular level of human or machine activities

[10, 11]

KO Limitation: lack of expressiveness

Search by document properties
Give me documents with author “Nash, David” and subject “marble sculpture”

Search by entity properties
Give me documents about wood sculptures written by an artist born in Italy

KO Limitation: lack of formalization

Subject: Buonarroti, Michelangelo

Subject: sculpture - Renaissance

Michelangelo the Italian artist? When and where he was born? What are his most famous works?

Do you mean sculpture the form of art?

Do you mean Renaissance the historical period? When and where exactly?

KO and KR: to complement each other

KO to KR

- Assist in knowledge acquisition in KR processes
  - Remodelling KOS to develop knowledge bases
  - Using KOS to build relation vocabularies and rules
- Assist with entity specification methodology

KR thought in KO

- Complex queries and the availability of digital data and information calls for radical innovations in KO practices

"The inter-influence between KO and KR will extend the power of human intelligence to enrich and enhance artificial intelligence"

[10]

From documents to entities

Class: statue
Author: Michelangelo
Date of creation: 1504
Matter: marble
Height: 5.17 m

The Statue of David, one of the most renowned works of the Renaissance (source: Wikipedia).

From KO to KR

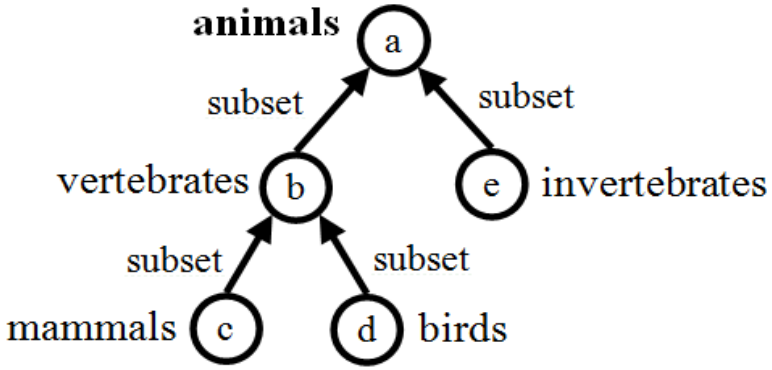

KR
search by any entity property
KO
search by
title, author,
subject

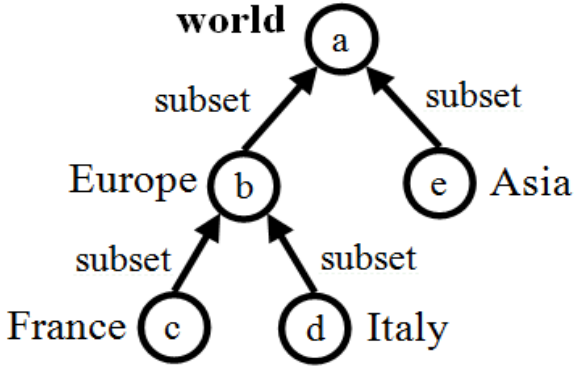
KO and KR: the convergence

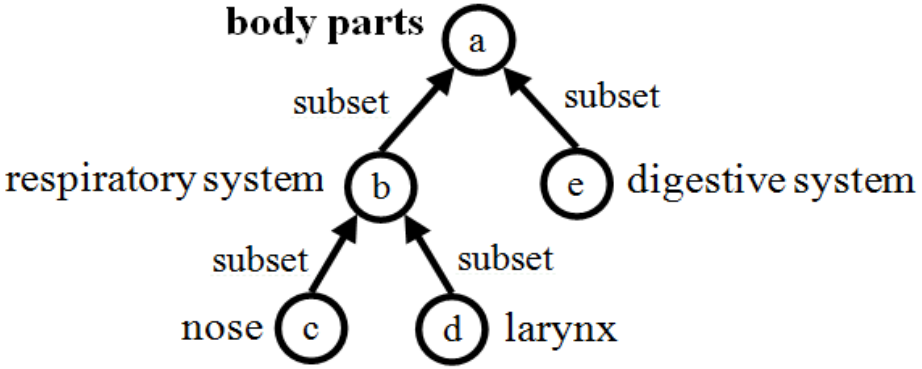
Ontology

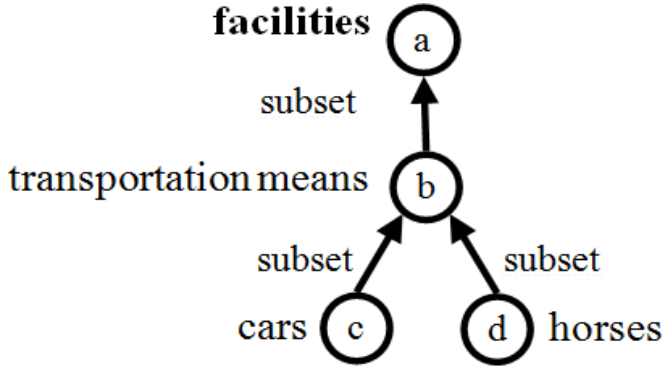
- For modelling
- For semantics
- For reasoning of knowledge

Classification Ontologies

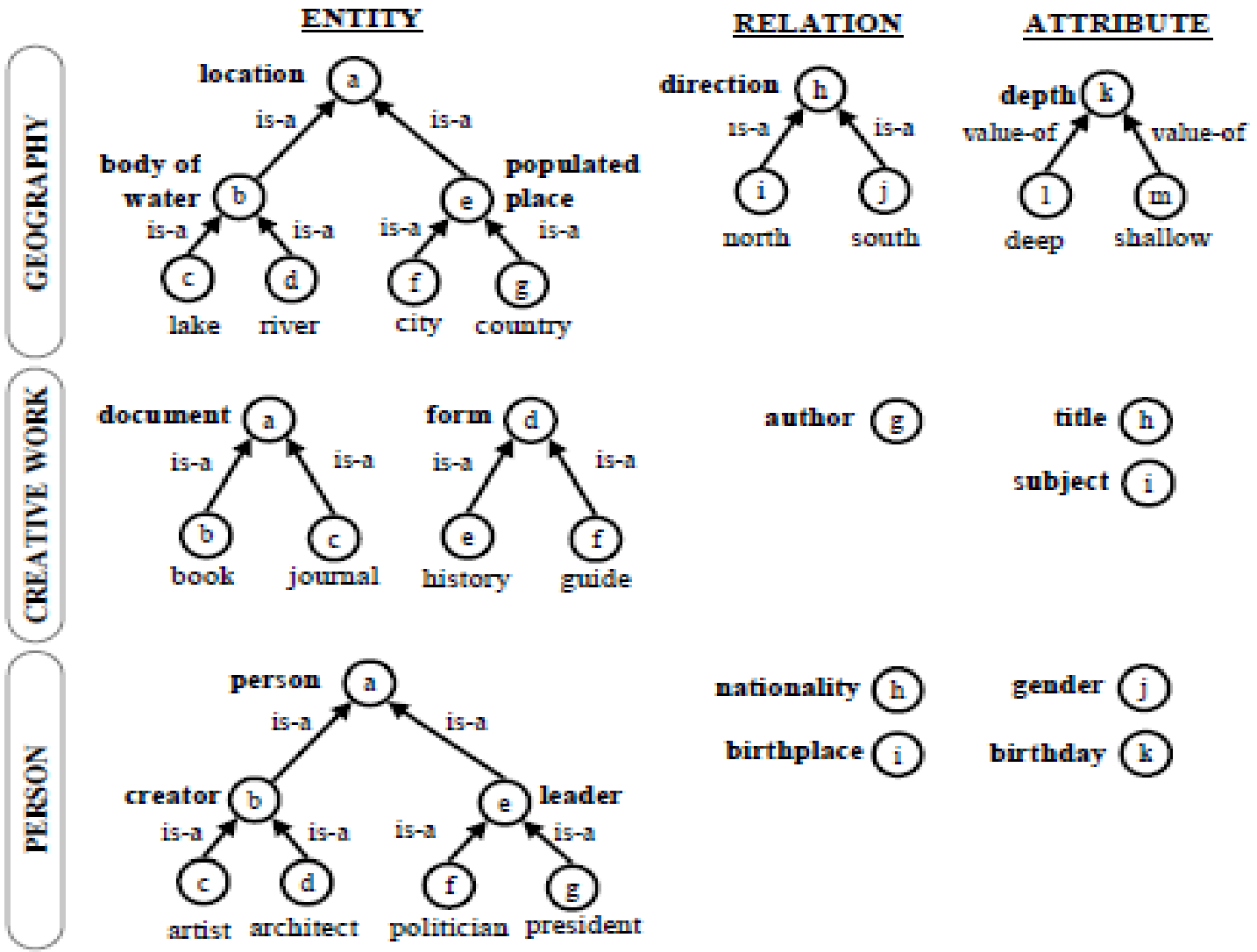




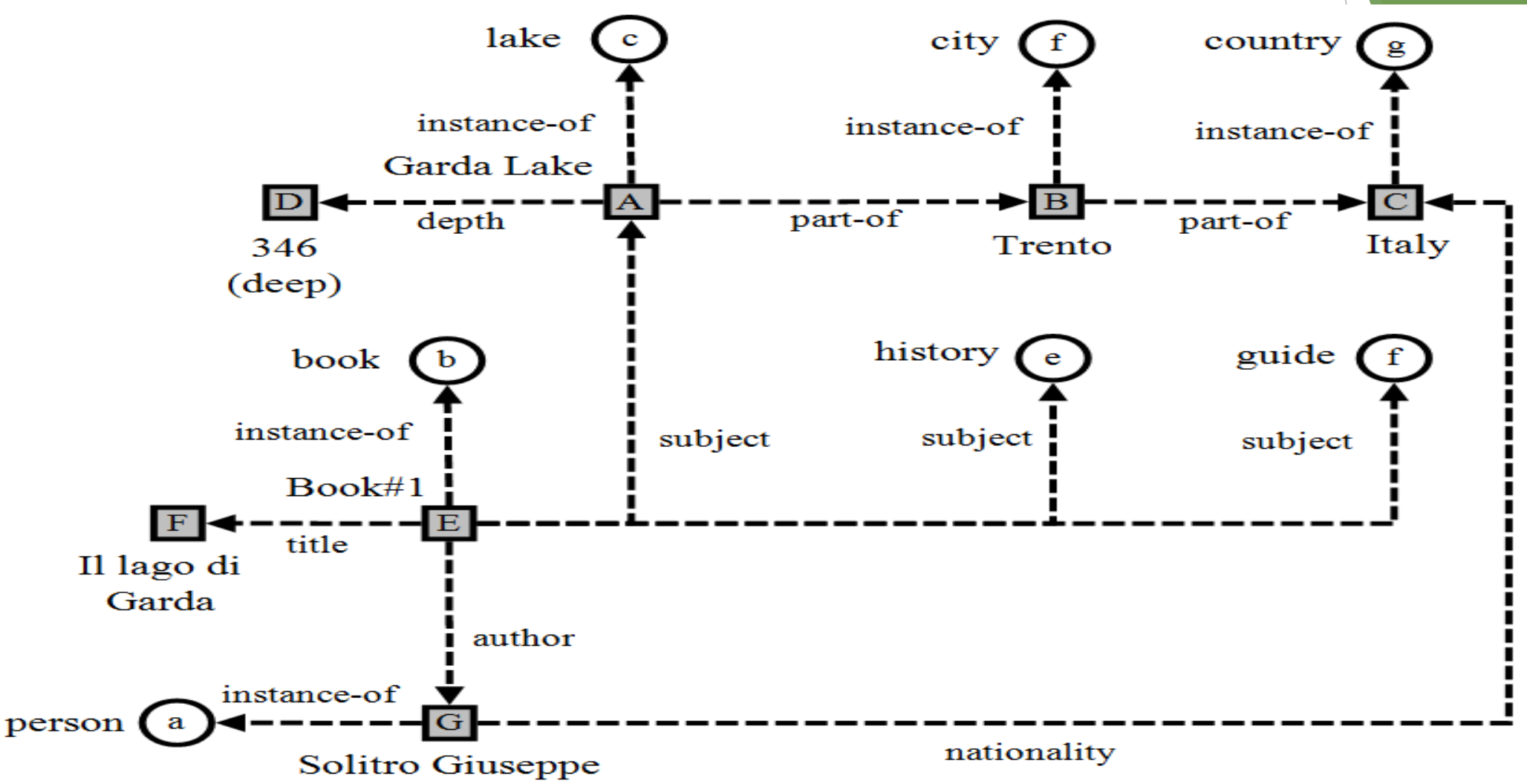




Descriptive Ontologies: intensional knowledge



Descriptive Ontologies: extensional knowledge



Give me documents about any lake with depth greater than 100 written by Italians

DERA

DERA is faceted as it is inspired by the principles and canons of the faceted approach by Ranganathan

DERA is a KR approach as it models entities of a domain (D) by their entity classes (E), relations (R) and attributes (A)

$$D = \langle E, R, A \rangle$$

[11, 12]

Descriptive Ontologies in DERA (I)

Step 1: Identification of the atomic concepts

(E) watercourse, stream: a natural body of running water flowing on or under the earth

Step 2: Analysis

a body of water
a flowing body of water
no fixed boundary
confined within a bed and stream banks
larger than a brook

Descriptive Ontologies in DERA (II)

Step 3: Synthesis.

Body of water
- (is-a) Flowing body of water
  - (is-a) Stream
    - (is-a) Brook
    - (is-a) River
- (is-a) Still body of water
  - (is-a) Pond
  - (is-a) Lake

Step 4: Standardization.

(E) stream, watercourse: a natural body of running water flowing on or under the earth

Descriptive Ontologies in DERA (III)

Step 5: Ordering

Terms and concepts in the facets are ordered

Step 6: Formalization

Descriptive ontologies are translated into Description Logic formal ontologies, e.g.,:

river $\sqsubseteq$ stream
lake $\sqsubseteq$ body-of-water
north $\sqsubseteq$ direction
river (Volga)
length (Volga, 3692)

DERA

- DERA facets have *explicit semantics* as they are modeled as descriptive ontologies
- DERA facets inherits all the nice properties of the faceted approach, such as robustness and scalability
- **DERA allows for a very expressive document search by any entity property**
- DERA allows for automated reasoning via the formalization into Description Logics ontologies

Summary

The usefulness of moving from KO to KR

- KO is methodologically very strong, but limited in expressivity as, by employing *classification ontologies*, it only supports queries by document properties.
- **KR, by employing *descriptive ontologies*, supports queries by any entity property**.

We propose the DERA faceted KR approach

- DERA is faceted as it allows the development of high quality and scalable descriptive ontologies
- DERA allows modeling relevant entities of the domain and their E/R/A properties and enables automated reasoning.
- It supports a highly expressive search of documents exploiting entity properties.

Opportunities for library professionals

- Can help in making available the **high quality data** for machine to use
- Can help in incorporating the library principles of **privacy** and **ethics**
  - Gender bias, racism, etc. are just part of the larger ethical concerns around AI
  - This ensures the data used to train AI is inclusive and diverse
- Can help the AI researchers on how to prepare "Information literate" AI
  - This may help in building a **learned machine**.
- Can share the ideas of how we encode information which allows the contextual information retrieval
- Can help the programmers to find the best data for their algorithms to learn on

- Can help the communities in understanding AI technology and its uses
- Can closely work with the AI and machine learning scientists/scholars to define and solve AI related library problems.

Our ongoing research in the area

- Ontology learning
- Data integration
- Narrative medicine

- Algorithm selection support system

References

1. Hillen, B. (2018). Google 'LIFE Tags' uses AI to organize LIFE magazine's 4 million-photo archive. https://www.dpreview.com/news/4243494233/google-life-tags-uses-ai-to-organize-life-magazine-s-4-million-photo-archive
2. Zhang, Ce & Govindaraju, Vidhya & Borchardt, Jackson & Foltz, Tim & Ré, Christopher & Peters, Shanan. (2013). GeoDeepDive: Statistical Inference using Familiar Data-Processing Languages.
3. Rouse, M. Artificially-intelligent attorney (AI attorney). https://searchenterpriseai.techtarget.com/definition/artificially-intelligent-attorney-AI-attorney
4. Artificial intelligence. http://www.ala.org/tools/future/trends/artificialintelligence
5. What happens to libraries and librarians when machines can read all the books? https://chrisbourg.wordpress.com/2017/03/16/what-happens-to-libraries-and-librarians-when-machines-can-read-all-the-books/
6. Knowledge representation and reasoning. https://en.wikipedia.org/wiki/Knowledge_representation_and_reasoning
7. Berners-Lee et al. (2001). Berners-Lee, T., Hendler, J., Lassila, O. (2001). The semantic web. Scientific American, 284 (5), 28-27.
8. Berners-Lee et al. (2001). Berners-Lee, T., Hendler, J., Lassila, O. (2001). The semantic web. Scientific American, 284 (5), 28-27.

References (contd…2)

9. Bouquet et al. (2004). Bouquet, P., Giunchiglia, F., Harmelen, F. van, Serafini, L. and Stuckenschmidt, H. (2004). Contextualizing ontologies. Journal of Web Semantics, 1(4), 325-343.
10. Jian Quin (2019). Paradigmatic Similarities in Knowledge Representation between KO and AI. In ISKO-UK Conference, London, UK.
11. Giunchiglia, F., Dutta, B. and Maltese, V. (2014). From Knowledge Organization to Knowledge representation. In ISKO Knowledge Organization. Vol. 41, no. 1, pp 44-56.
12. Giunchiglia, F. and Dutta, B. (2011). DERA: a Faceted Knowledge Organization Framework. Available online: http://eprints.biblio.unitn.it/archive/00002104/
13. Smith, D. (2017). Living with an AI: a glimpse into the future. https://scholarlykitchen.sspnet.org/2017/03/22/living-with-an-ai-a-glimpse-into-the-future/
14. Tay, A. (2017). How libraries might change when AI, Machine learning, open data, block chain & other technologies are the norm. http://musingsaboutlibrarianship.blogspot.com/2017/04/how-libraries-might-change-when-ai.html
15. Artificial intelligence, revealed. https://engineering.fb.com/ml-applications/artificial-intelligence-revealed/
16. After racist tweets, Microsoft muzzles teen chat bot Tay. https://money.cnn.com/2016/03/24/technology/tay-racist-microsoft/index.html
17. Artificial Intelligence's White Guy Problem. https://www.nytimes.com/2016/06/26/opinion/sunday/artificial-intelligences-white-guy-problem.html
18. Russell, Stuart J.; Norvig, Peter (2009). Artificial Intelligence: A Modern Approach (3rd ed.). Upper Saddle River, New Jersey: Prentice Hall.
19. Davis, R., Shrobe, H. and Szolovits, P. What is a Knowledge Representation? AI Magazine, 14(1):17-33, 1993.

Thank you for listening to me!!!

Contact

Biswanath Dutta
bisu@drtc.isibang.ac.in
dutta2005@gmail.com