Bardhan, Susmita. and Dutta, Biswanath (2023). ONCO: an ontology model on MOOC platforms. Knowledge Organization (preprint copy).

ONCO: an ontology model for MOOC platforms

Abstract:

In the process of searching for a particular course on e-learning platforms, it is required to browse through different platforms and it becomes a time-consuming process. To resolve the issue, an ontology has been developed that can provide single-point access to all the e-learning platforms. The modelled ONline Course Ontology (ONCO) is based on YAMO, METHONTOLOGY and IDEF5 and built on the Protégé ontology editing tool. ONCO is integrated with sample data and later evaluated using pre-defined competency questions. Complex SPARQL queries are executed to identify the effectiveness of the constructed ontology. The modelled ontology is able to retrieve all the sampled queries. The ONCO has been developed for the efficient retrieval of similar courses from massive open online course (MOOC) platforms.

Keywords: E-learning, Ontology, Protégé, SPARQL queries, ONCO, MOOC platforms

Paper Type: Research Paper

1. Introduction

An ontology provides a way to organize and structure information, making it easier to find relevant facts and data. Ontologies are used to define the concepts and relationships within a specific domain of knowledge and can be used to represent the structure of the data. Ontology is defined as a formal explicit specification of a shared conceptualization (Studer et al. 1998). Ontology provides an abstract model by identifying its concepts within a specific domain that can be understood by humans and machines (Guarino et al. 1995).

E-learning is gradually becoming a part of every student's life as it provides the facility to learn something new and interesting to an individual. E-learning platforms have made it possible for learners from anywhere in the world to get access to online courses offered by top academicians and universities (Conache et al. 2016). Learning is made hassle-free as one is enabled to get the exact course that one wants to learn based on his interest, cost, level of knowledge and all other factors.

In the information society, the process of learning is continuous. However, it is not possible to enrol for everything in formal education which leads to an online learning boom. But there are certain problems when someone searches for courses online. Firstly, search engines pop up with a lot of massive open online course (MOOC) platforms. Secondly, browsing through each platform provides thousands of similar courses. Finally, comparing courses from one platform to another takes a lot of time. Hence, selecting the course of one's choice is a tedious process. We are mostly dependent on search engines such as Google, Bing, etc. for information retrieval. Considering Google Search can discover courses from MOOC platforms which are based on simple queries such as courses on python, java, music, etc. However, when it comes to the ability to find specific requirements such as courses provided by Harvard University on topics related to python which are of beginner level and similar queries, search engines cannot discover accurate results. Ontology can be employed for semantic-based information retrieval which aims to improve the accuracy of results through query expansion and terms disambiguation resolution (Asim et. al, 2019). Ontology is one of the most common knowledge representation models used extensively in information retrieval as it represents the knowledge in terms of machine-readable, understandable, and processable information hierarchies (Gai et. al, 2015). The objectives of the current study are as follows:

1. To design and development of an ontology that can support the retrieval of courses from various platforms in time when implemented in the backend of a system.
2. To develop an ontology that supports personalization, for instance for finding courses based on one's particular need such as the frontend coding course provided by Harvard University which is meant for the introductory level of knowledge.

The purpose of the study is to construct an ontology that enables the representation of information about courses available in various MOOC platforms based on certain criteria such as level of knowledge, subject, language, university, instructor, etc. for a better understanding and retrieval of courses according to learners' needs. Here, learners' needs can be course language, subject area, level of understanding of a particular course, and so forth. In case, the course is provided in German language and the student lacks proficiency in the language, the language parameter is a need of learners and they can choose the course in any language which is easier for them to understand (Dutta et al. 2009).

The primary contributions of the current work are:

1. A modelled ontology has been developed by identifying the various facets of a course and MOOC platforms.
2. An integrated methodology for ontology has been proposed based on existing methodologies.
3. An open-source knowledge graph has been formulated based on ontology.

E-learning has become an integral part of the student as well as industrialist community. Especially in the post covid era, probably online learning will gain prominence over traditional learning. So, the motivation is to reduce the time and effort in finding the right course for an individual by introducing information retrieval through Ontology. The rest of the paper is organized as follows: section 2 briefly explains the ONCO and its evaluation section 3 discusses the related works; section 4 finally concludes the paper along with future prospects of the study.

2. ONCO

The ONCO is based on three existing methodologies namely YAMO (Yet Another Methodology for Ontology) (Dutta et. al, 2015), METHONTOLOGY (Ferndndez et. al. 1997) and IDEF5 (Integrated Definition for Ontology Description Capture Method) (Perakath, 1994). YAMO is based on the facet-analysis approach proposed by S.R. Ranganathan (Ranganathan, 1967) where facets are aspects of meaning. Methontology is one of the ontology engineering methodologies which focuses on the full development lifecycle from knowledge gathering to maintenance. In the present work, methontology plays a crucial role in shaping ONCO ontology as a continual process. IDEF5 methodology is based on the entity-relationship modelling approach. We have implemented the entity-relationship approach while modelling the ontology. Based on the above-mentioned methodology, the ONCO is developed. The steps followed in the development of the ontology are

1. **Identification of Domain**: The first step of constructing a domain ontology is to select the domain for which modelling work is to be performed. The domain of the present ontology is to model Massive Open Online Course(MOOC) platforms' contents in such a way that can be retrieved from various parameters

2. **Specific Purpose**: Next step is to identify the problem that needs to be solved. In other words, finding the objective and purpose of developing an ontology. The purpose of the study is to identify the courses based on level of knowledge, subject, language, university, instructor, and other parameters. Based on this purpose, some of the competency questions are formed to evaluate the entire model.
   Competency questions: Competency questions help in better understanding and building the ontology. Competency questions are questions for which an ontology can give a specific answer.

    1. What are the courses available?
    2. What are the courses available at different levels?
    3. Show the courses related to Python.
    4. Find the courses along with the sponsored organizations/universities and instructors.
    5. Find the affiliation of a particular instructor.
    6. Find the course provided by a particular instructor and his area of expertise along with the language in which he/she taught.
    7. Find all the courses which have more than 1,00,000 students
    8. Find all the courses having an introductory level which is provided by a particular institution.
    9. Find introductory courses from a particular university on a particular topic.
    10. Find the number of courses available on different subjects

3. **Knowledge Gathering**: In order to understand a problem and find the solution for it. It is required to have knowledge of the domain. For that purpose, the collection of information from various sources to acquire knowledge about the domain in depth. In order to build the ontology, information has been collected from various MOOC platforms. The following are the resources from which information has been collected.
   edX: edX is a MOOC platform developed by two world-class institutions Harvard University and Massachusetts Institute of Technology (MIT) in 2012. It provides high-level courses from various universities and institutions. It runs as a not-for-profit organisation.
   Udemy: Udemy was started in May 2010 as a MOOC platform to provide content on the web by Eren Bali, Gagan Biyani, and Oktay Caglar. The courses are mainly for professionals.
   Coursera: Coursera is a MOOC platform founded in 2012. Daphne Koller and Andrew Ng are the founders of the website. They are from the computer science department of Stanford University. Coursera provides access to online courses and degrees from various universities and companies to anyone, anywhere and anytime.
   IITBX: IITBombayX is an online platform developed by the Indian Institute of Technology Bombay which provides MOOC courses. Above are the main platforms from which most of the information about a course in the MOOC platform has been acquired. However, this is not the exclusive list of platforms that have been searched for understanding the domain. We have gone through other platforms such as Udacity, SWAYAM, Khan Academy, etc. Most of the platforms provide similar information about a course.

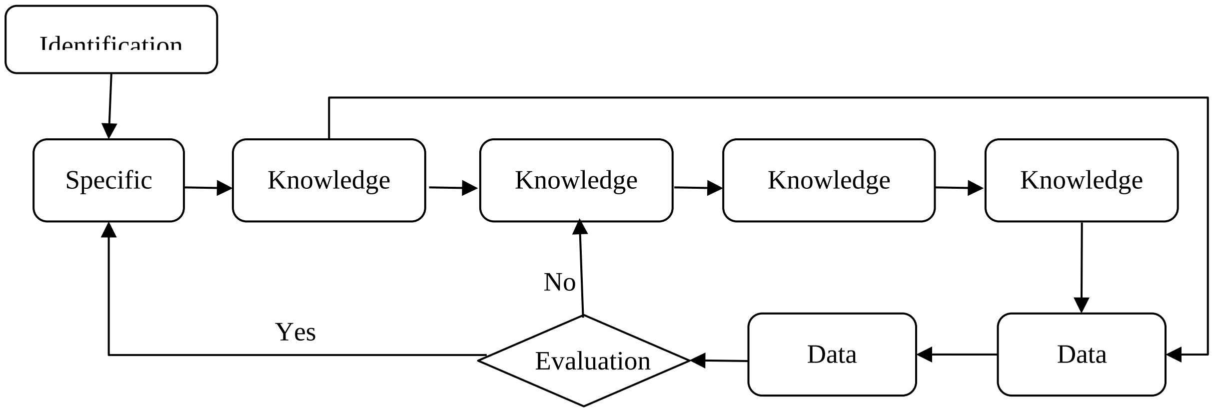


Figure 1: Steps of ONCO development process

4. **Knowledge Analysis**: The acquired knowledge has to be analysed and evaluated to fulfil the purpose of constructing an ontology. Knowledge analysis is the process of analysis of the various resources available. Data are analysed to understand the importance of the information provided by the platforms about a course and related information. The analysis has been done by taking each facet of a course such as course name, course length, and course mode in MOOC. Each course has a different level of knowledge which means the depth or intention of a course. A Beginner-level course gives an overview of basic study materials and learning scope to the learners. An intermediate-level course provides an opportunity to learn another level of depth of the course which can be understood once a student has the basic knowledge of the topic. An advanced-level course offers a deeper intention of a course. Beginner-level, intermediate-level and advanced-level courses are ideas that can be broken down to the level of courses.
   Some terms that are extracted from edX are instructors, institution, level, language, subject, prerequisites, effort, Video Transcript, Course Type, length, title, price (verified track and audit track), enrolled number, and syllabus.
   Some terms that are extracted from Udemy are certification on completion, created by, the organization from where instructor belongs, language, subtitle, rating, number of students rated, certification/ result, syllabus, students, and description.
   Some terms extracted from Coursera are shareable certification, skill, level, duration, flexible deadline, subject, language, subtitle, instructors, a skill you will gain, offered by, syllabus, enrolled number of students, about this course, rating, number of ratings
   Some terms extracted from IITBX are effort, course number, course start and course end.
   To retrieve a course, it is required to have a course name, and the course needs to be taught by someone who can be affiliated with some university or institute or he can independently teach.

5. **Knowledge Identification**: Identification of various related terms of the domain and their features are arranged in a hierarchy. As ontology is for reuse and interoperability, the terms used in the ontology should be commonly shared vocabularies and most frequently used terms by domain experts. Knowledge identification is the process of identifying relationships between concepts, object properties and data properties. In order to provide the standard terms, authors have searched for some ontologies and reviewed some literature. The work of (Dang, Tang, and Li 2019) has been useful for identifying the standards term in the MOOC domain. Each MOOC platform has different terminologies describing the same concepts. A course needs to have a person

who can teach the study materials. The person mentioned can be addressed by various names such as instructor, teacher, creator etc. In order to improve the searchability, we have used synonyms terms under skos:prefLabel, skos:altLabel.

Class: A class defines a set of instances of an entity and the classes are defined using owl:Class.

Class	**Definition**	**Label**
onco:Course	Course number or course code	rdfs:label: Course
onco:Course_ Duration	Duration of the course	rdfs:label: Course Duration, skos:prefLabel: Course Length
onco:Course_ mode	Way of learning the course such as self-paced and instructor-led	rdfs:label: Course Mode
dc:Language	Language in which the course is taught	rdfs:label: Language
dc:Subject	Discipline or domain under which the course is being covered	rdfs:label: Subject
foaf:Agent	Person, Organization or group	rdfs:label: Agent
foaf:Organization	Institution which provides the course and the instructor affiliated to that institution.	rdfs:label: Institution, skos:prefLabel: University
foaf:Person	The person who teaches the course.	rdfs:label: Instructor, skos:prefLabel: teacher, skos:altLabel: creator
onco:Level	The depth of the course	rdfs:label: Level
onco:Mooc_platform	MOOC platforms such as edX, Coursera, Udacity, etc.	rdfs:label: Mooc Platform
onco:Assignment	Assignment are the task to complete to progress in the course.	rdfs:label: Assignment

onco:Outcome	Outcome of completing the course as a form of certification or degree	rdfs:label: Outcome
onco:Skill	Skill acquires after completing the course	rdfs:label: Skill set, skos:prefLabel: Skill
onco:Course_ Material	Study materials has been used in teaching the course and for further reference to students	rdfs:label: Course material, skos:prefLabel: Study material
onco:Syllabus	Outlines of the course that will be taught	rdfs:label: Syllabus
onco:Module	Module is a section of syllabus. Subclass of Syllabus	rdfs:label: Module
onco:Article	Journal or Conference Article	rdfs:label: :Article
onco: Book	Books can be refered by the student for further reference	rdfs:label: Book
onco:Lecturer_ Note	Note prepares by Instructor for students reference	rdfs:label: Lecturer note
onco:Useful link	Useful website link can be refered by the student for further reference	rdfs:label: Useful link, skos:prefLabel: Website Link
onco: Video Material	Primary material for teaching the MOOC course.	rdfs:label: Video Material
onco: Exercise	Subclass of Assignment	rdfs:label: Exercise
onco: Project	Subclass of Assignment	rdfs:label: Project
onco: Quiz	Subclass of Assignment	rdfs:label: Quiz

Table 1: Shows classes used in ontology with the basic definition of each class

6. **Knowledge Modelling:** Modelling of ontology is the structuring of the objects and their relationships between the two entities and structuring the features of an entity.

Knowledge modelling is the process of modelling and structuring the concepts that are developed in knowledge production. It depicts the relationship between entities and their properties which are based on the entity-relationship approach of modelling. Instructor and Institution are class entities. They are related to each other as Instructor has an affiliation with Institution.

Object Properties: Object properties provide a directional connection or relationship between individuals.

Object Properties	**Domain**	**Range**	**Inverse of**
onco:has_affliation	foaf:Person	foaf:Organization	onco:is_affliation_of
onco:has_course_level	onco:Course	onco:Level	onco:is_level_for_course
onco:learning_mode	onco:Course	onco:Course_mode	onco:is_mode_of_learning_for_course
onco:has_duration	onco:Course	onco:Course_length	onco:is_duration_for_course
onco:has_assignment	onco:Course	onco:Assignment	onco:is_assignment_for_course
onco:has_instructor	onco:Course	foaf:Person	onco:teach
onco:has_language	onco:Course	dc:Language	onco:is_language_for_course
onco:prerequisited_knowledge_and_skill	onco:Course	onco:Skill	onco:prerequisite_knowledge_for
onco:has_availability_status	onco:Course	onco:Course_Availablity	onco:is_available_for_course
onco:has_sponsored	foaf:Organization	onco:Course	onco:sponsored_by
onco:has_subject	onco:Course	onco:Discipline	onco:is_subject_for_course
onco:subtitle_language	onco:Course	dc:Language	onco:is_subtitle_language_of_course
onco:has_study_material	onco:Course	onco:Course_material	onco:is_study_material_for_course
onco:has_weekly_hours	onco:Course	onco:Course_Duration	onco:is_weekly_effort_for_course

onco:is_affiliation_of	foaf:Organization	foaf:Person	onco:has_affliation
onco:offered_by	onco:Course	onco:Mooc_platform	onco:offers
onco:acquire_skill	onco:Course	onco:Skill	onco:acquire_skill_from
onco:result	onco:Course	onco:Outcome	onco:is_result_for_course

Table 2: Shows object properties used in graphs with the domain and range of them

Data Properties: Datatype properties relate entities to data values. A datatype property is said to be an example of the built-in OWL class owl:DatatypeProperty. In other words, data property is to describe the attributes of an entity in the form of literals.

Data Properties	**Domain**	**Range**
dc:Description	onco:Course	xsd:string
dc:Identifier	onco:Course	schema:url
onco:end_date	onco:Course	xsd:dateTime
foaf:homepage	foaf:Organization foaf:Person	schema:url
foaf:mbox	foaf:Person	xsd:string
onco:number_of_enrolled_students	onco:Course	xsd:integer
onco:Fee	onco:Course	xsd:integer
onco:number_of_reviews	onco:Course	xsd:integer
onco:rating	onco:Course	xsd:decimal
onco:start_date	onco:Course	xsd:dateTime
onco:summary	onco:Course	xsd:string
onco:certification	onco:Course	xsd:boolean
onco:title	onco:Course	xsd:string

Table 3: Shows data properties used in graphs with the domain and range of them

The ontology is built with the use of the Protégé ontology editing tool. Protégé is developed by Biomedical Informatics Research (BMIR) lab at Stanford University. Protégé is a free and open-source platform that provides tools to construct ontologies along with domain modelling (“Protégé” n.d.). It supports RDF specification and OWL 2.0 which are recommended by the World wide web consortium (W3C).

The structure of the Course has been shown below:

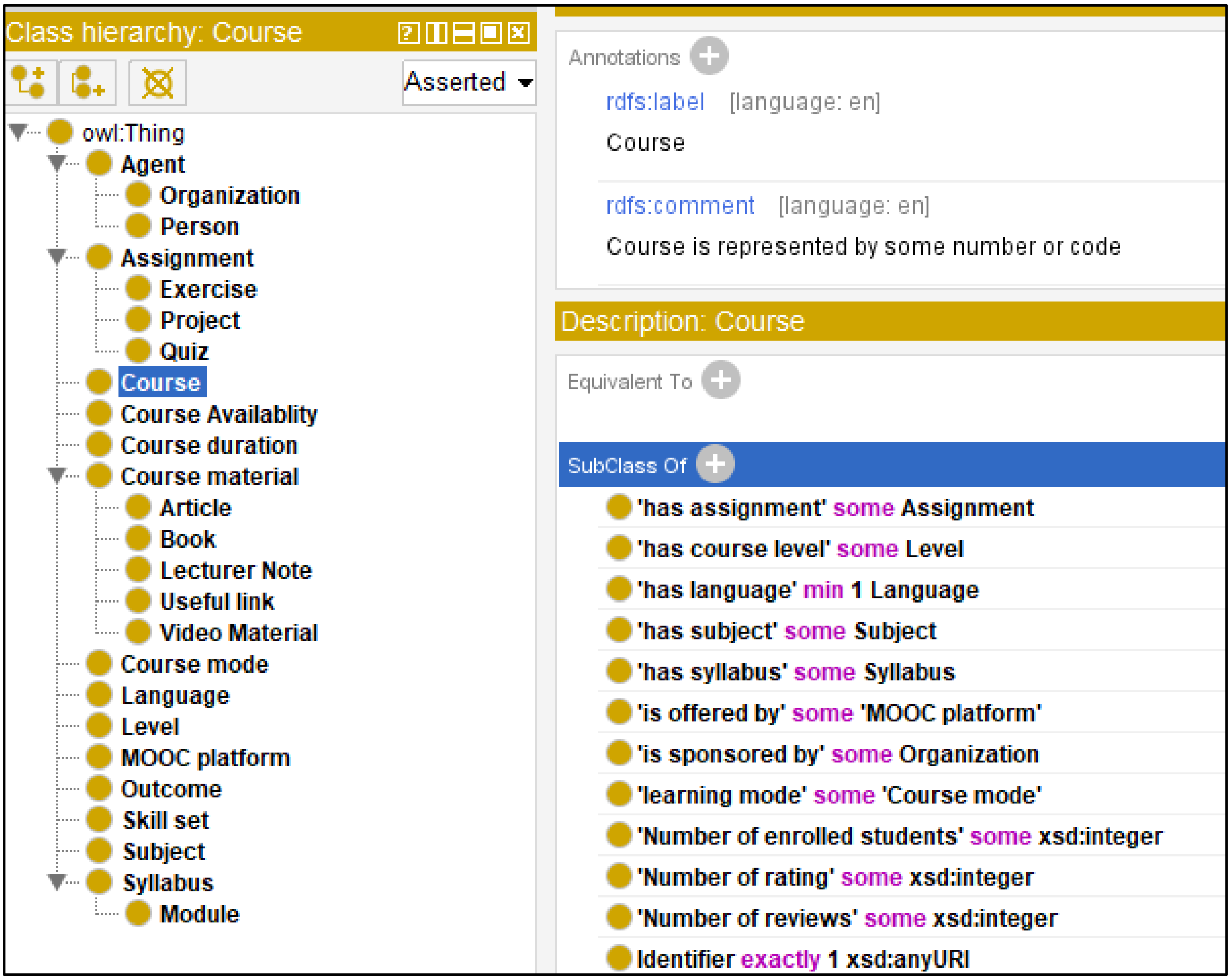


Figure 2: Course class with subClass

7. Data Cleaning: Data collected in the step of Knowledge gathering are being cleaned in order to integrate with the model. An instance of cleaning data can be the price of the course is $100 and the data property is an integer. So, the dollar sign has been removed from the data so that it can be integrated with the data property.

In the current study, we have taken sample data of 7402 courses (edx_courses.csv, udemy_courses.csv) from Kaggle and cleaned them with the use of Open Refine software. To successfully transfer the data into the ontology, it is necessary to clean it thoroughly in accordance with the defined classes and properties.

	A	B	C	D	E	F	G	H	I	J
1	Course	title	summary	number_of_enrolled_students	has_course_type	foaf:Organization	Instructor	has_course_level	has_subject	has_language
2	Course1	How to Lea	Learn esse	124980	Self-paced (	edX	Nina Hunter	Introductory	Education	English
3	Course1						Robyn Belair			
4	Course1						Ben Piscopo			
5	Course2	Programm	This cours	293864	Self-paced (	The University of	Charles Seve	Introductory	Computer	English
6	Course3	CS50's Intr	An introdu	2442271	Self-paced (	Harvard Universit	David J. Mala	Introductory	Computer	English
7	Course3						Doug Lloyd			
8	Course3						Brian Yu			
9	Course4	The Analyt	Through ir	129555	Instructor-l	Massachusetts In	Dimitris Berts	Intermediate	Data Analy	English
10	Course4						Allison O'Hair			
11	Course4						John Silberholz			
12	Course4						Iain Dunning			
13	Course5	Marketing	This cours	81140	Self-paced (	University of Calif	Stephan Sorg	Introductory	Computer	English
14	Course6	Introductic	Learn the (	301793	Self-paced (	University of Ade	Frank Schultr	Introductory	Business &	English
15	Course6						Noel Lindsay			
16	Course6						Anton Jordaan			
17	Course6						John Sing			
18	Course7	Leading Hi	Learn how	32847	Self-paced (	The University of	Tyler G. Okim	Intermediate	Communic	English
19	Course8	The Found	Learn why	67073	Self-paced (	University of Calif	Dacher Keltn	Introductory	Business &	English
20	Course8						Emiliana Simon			
21	Course8						Thomas			

Figure 3: An overview of the cleaned dataset

8. Data Mapping: The process of merging the data with the ontology in order to have instances of entities. Cellfie plugin of protege and JSON rules have been implemented to map or merge the cleaned data with the ontology.

```
Individual: @A*(rdfs:label=(@A*))
Types: Course
Facts: title @B*(xsd:string),
          summary @C*(xsd:string),
          number_of_enrolled_students @D*(xsd:integer),
          Fee @N*(xsd:integer),
          has_syllabus @Q*(xsd:string),
          has_course_mode @E*,
          sponsored_by @F*,
          has_instructor @G*(rdfs:label=(@G*)),
          has_course_level @H*,
          has_subject @I*,
          has_language @J*,
          has_subtitle @K*,
          has_weekly_hours @L*,
          has_duration @M*,
          Identifier @R*
```

Table 4: A glimpse of Json Transformation rule

The ONCO has been visualized using GraphDB tool. GraphDB (“GraphDB™ - Ontotext” n.d.) is a visualization tool for ontology and knowledge graphs. It also allows running SPARQL queries and the results can be represented in the form of tables and various charts such as bar graphs, pie charts, and so on. For the current knowledge graph, GraphDB is used to produce the graphs visually.

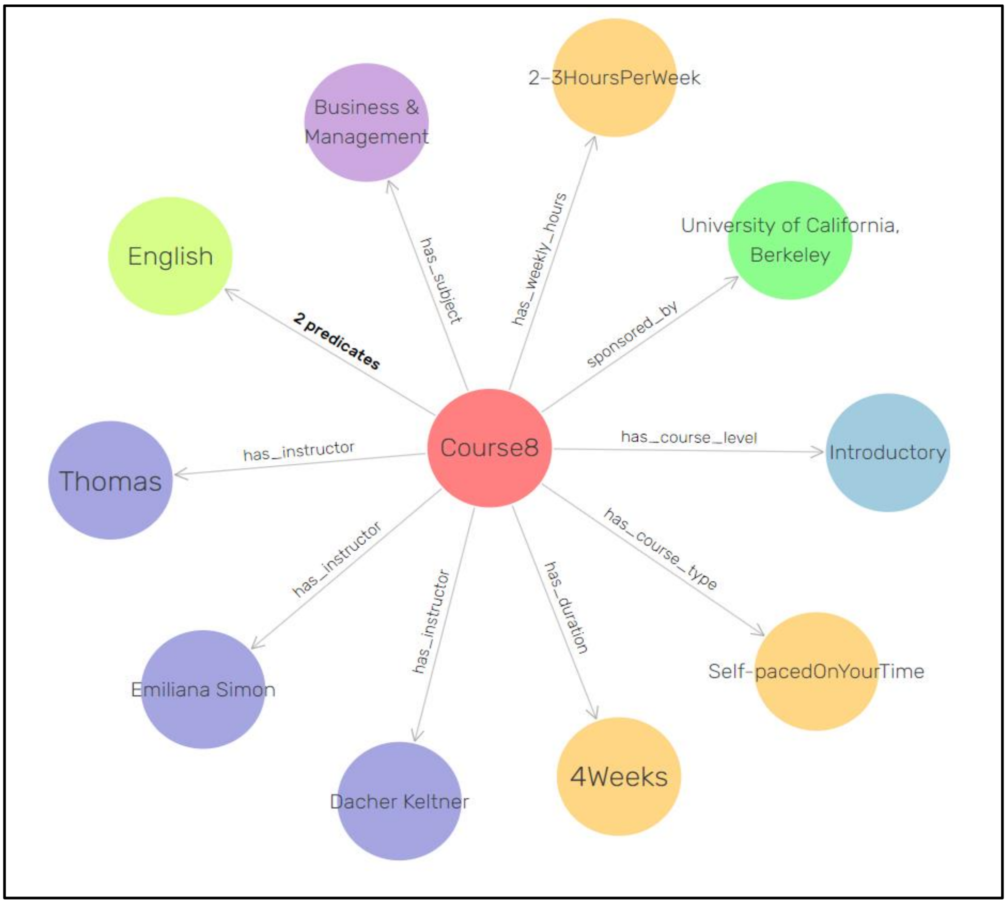


Figure 4: Shows the relationship between an instance of a course and its relationship with other classes.

In figure 4, represents an instance of course along with its object properties. Course8 has 2 predicates which are English which is the same for both subtitle_language and has_language. It has 3 instructors and other properties are shown.

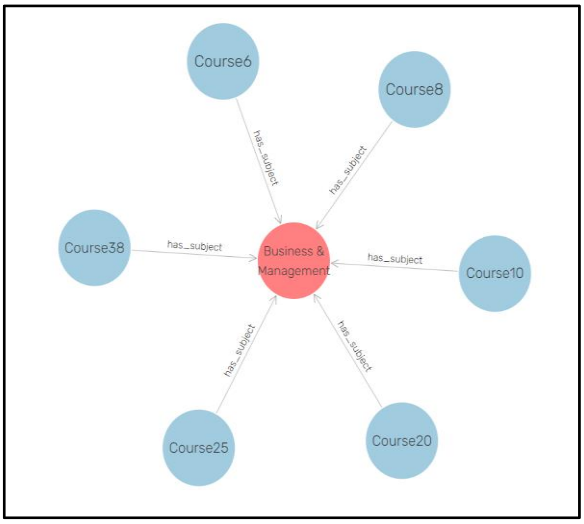


Figure 5: Shows courses available based on a particular subject

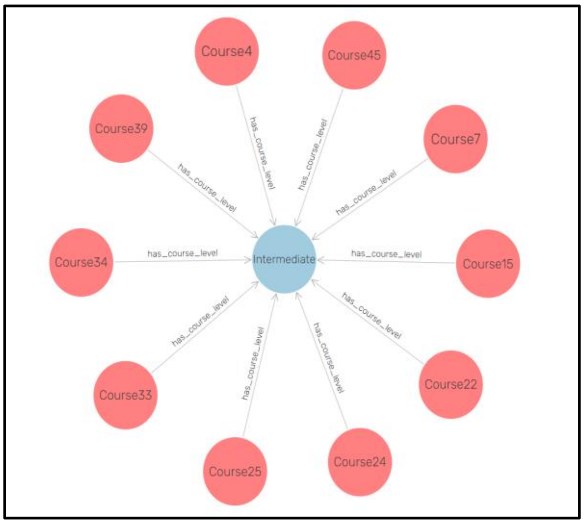
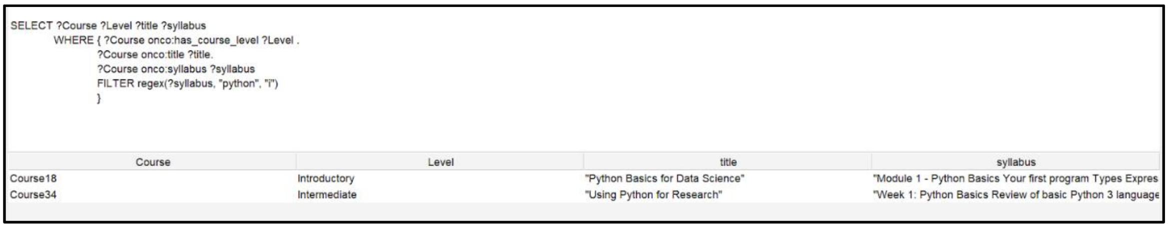


Figure 6: shows the courses having intermediate level

Figure 5 and Figure 6 show all the courses on a particular subject and having the same level of knowledge respectively. Six courses have subject as Business & Management and 10 courses have a course level as intermediate. It can be seen as Course has_subject Business & Management and Course has_course_level Intermediate. The graph shows the object property and data property about a particular course and that particular course is an instance of entity course.

9. Evaluation: Evaluating the quality of the ontology as well as fulfilling the purpose is the final step of developing the ontology. We have implemented the pallet reasoner which is an open-source OWL reasoner. The pallet can be used to infer new information from an OWL ontology based on the relationships between its classes and individuals and to detect inconsistencies in the ontology (Sirin et al, 2007). The constructed ontology has consistency in its structure. The quality of the ontology is dependent on the effective retrieval of course information. In order to identify the efficiency of the ONCO, some of the competency questions are transformed into SPARQL Queries. The queries are given below:
PREFIX is used to avoid repeating the resource IRIs in the query. The following prefixes are required for every query. Prefix onco is the IRI for onco ontology and has used some Dublin core prefixes to build the ontology.
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX dc: <http://purl.org/dc/elements/1.1/>
PREFIX onco:<http://www.isibang.ac.in/ns/onco/>

- Query 1: find courses which have “python” in its syllabus and also show their levels

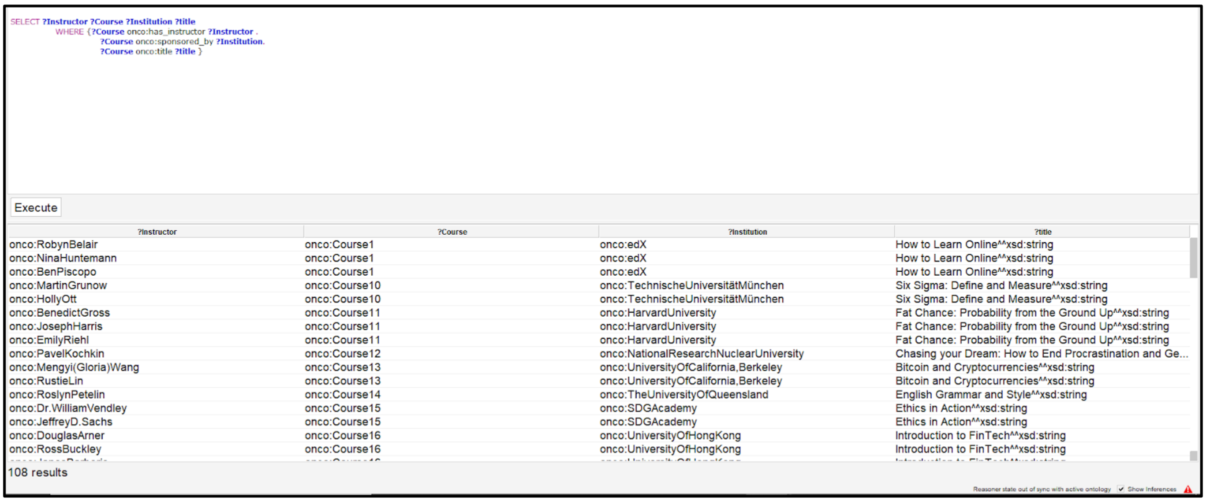


```
SELECT ?Course ?Level ?title ?syllabus
      WHERE { ?Course onco:has_course_level ?Level .
            ?Course onco:title ?title.
            ?Course onco:syllabus ?syllabus
            FILTER regex(?syllabus, "python", "i")
            }
```

Course	Level	title	syllabus
Course18	Introductory	"Python Basics for Data Science"	"Module 1 - Python Basics Your first program Types Expres
Course34	Intermediate	"Using Python for Research"	"Week 1: Python Basics Review of basic Python 3 language

Figure 6: Shows the python courses based on syllabus

- Query 2: retrieve courses with their instructor and institution.



```
SELECT ?Instructor ?Course ?Institution ?title
      WHERE {?Course onco:has_instructor ?Instructor .
            ?Course onco:sponsored_by ?Institution.
            ?Course onco:title ?title }
```

Execute

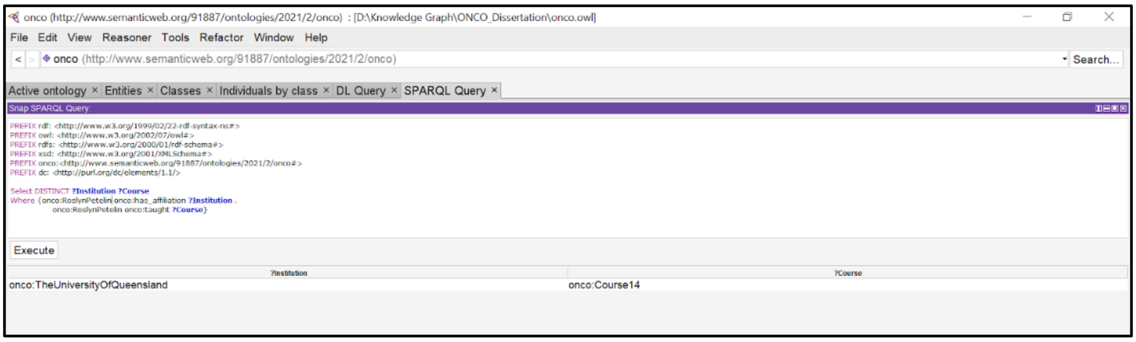
?Instructor	?Course	?Institution	?title
onco:RobynBelair	onco:Course1	onco:edX	How to Learn Online^^xsd:string
onco:NinaHuntemann	onco:Course1	onco:edX	How to Learn Online^^xsd:string
onco:BenPiscopo	onco:Course1	onco:edX	How to Learn Online^^xsd:string
onco:MartinGrunow	onco:Course10	onco:TechnischeUniversitätMünchen	Six Sigma: Define and Measure^^xsd:string
onco:HollyOtt	onco:Course10	onco:TechnischeUniversitätMünchen	Six Sigma: Define and Measure^^xsd:string
onco:BenedictGross	onco:Course11	onco:HarvardUniversity	Fat Chance: Probability from the Ground Up^^xsd:string
onco:JosephHarris	onco:Course11	onco:HarvardUniversity	Fat Chance: Probability from the Ground Up^^xsd:string
onco:EmilyRiehl	onco:Course11	onco:HarvardUniversity	Fat Chance: Probability from the Ground Up^^xsd:string
onco:PavelKochkin	onco:Course12	onco:NationalResearchNuclearUniversity	Chasing your Dream: How to End Procrastination and Ge...
onco:Mengyi(Gloria)Wang	onco:Course13	onco:UniversityOfCalifornia,Berkeley	Bitcoin and Cryptocurrencies^^xsd:string
onco:RustieLin	onco:Course13	onco:UniversityOfCalifornia,Berkeley	Bitcoin and Cryptocurrencies^^xsd:string
onco:RoslynPetelin	onco:Course14	onco:TheUniversityOfQueensland	English Grammar and Style^^xsd:string
onco:Dr.WilliamVendley	onco:Course15	onco:SDGAcademy	Ethics in Action^^xsd:string
onco:JeffreyD.Sachs	onco:Course15	onco:SDGAcademy	Ethics in Action^^xsd:string
onco:DouglasArner	onco:Course16	onco:UniversityOfHongKong	Introduction to FinTech^^xsd:string
onco:RossBuckley	onco:Course16	onco:UniversityOfHongKong	Introduction to FinTech^^xsd:string

108 results

Reasoner state out of sync with active ontology ✔ Show Inferences

Figure 7: shows the courses along with its instructors and institution

- Query 3: find a particular instructor' affiliation.



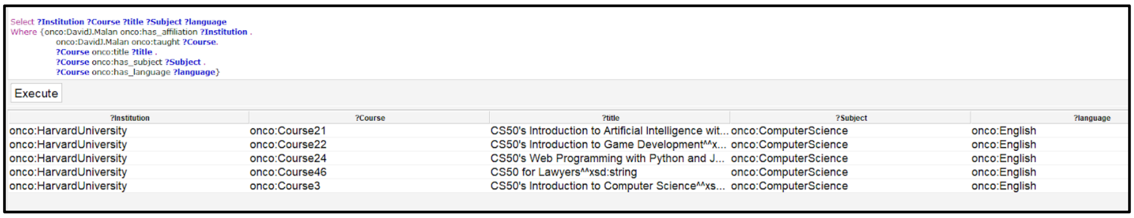

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX onco:<http://www.semanticweb.org/91887/ontologies/2021/2/onco#>
PREFIX dc: <http://purl.org/dc/elements/1.1/>

Select DISTINCT ?Institution ?Course
Where {onco:RoslynPetelin onco:has_affiliation ?Institution .
        onco:RoslynPetelin onco:taught ?Course}
```

?Institution	?Course
onco:TheUniversityOfQueensland	onco:Course14

Figure 8: Shows affiliation of a particular instructor

- Query 4: find the course provided by a particular instructor and his area of expertise.




```
Select ?Institution ?Course ?title ?Subject ?language
Where {onco:DavidJ.Malan onco:has_affiliation ?Institution .
        onco:DavidJ.Malan onco:taught ?Course.
        ?Course onco:title ?title .
        ?Course onco:has_subject ?Subject .
        ?Course onco:has_language ?language}
```

?Institution	?Course	?title	?Subject	?language
onco:HarvardUniversity	onco:Course21	CS50's Introduction to Artificial Intelligence wit...	onco:ComputerScience	onco:English
onco:HarvardUniversity	onco:Course22	CS50's Introduction to Game Development^^x...	onco:ComputerScience	onco:English
onco:HarvardUniversity	onco:Course24	CS50's Web Programming with Python and J...	onco:ComputerScience	onco:English
onco:HarvardUniversity	onco:Course46	CS50 for Lawyers^^xsd:string	onco:ComputerScience	onco:English
onco:HarvardUniversity	onco:Course3	CS50's Introduction to Computer Science^^xs...	onco:ComputerScience	onco:English

Figure 9: Shows the subject interest, language and course provided by a particular instructor

- Query 5: retrieve all the courses which have more than 1,00,000 students.


```
SELECT ?Course ?title ?number
      WHERE { ?Course onco:title ?title.
              ?Course onco:number_of_enrolled_students ?number
              FILTER (?number >= 100000)
              } Order by asc (?number)
```

?Course	?title	?number
onco:Course20	Strategic Social Media Marketing^^xsd:string	103640
onco:Course37	The Health Effects of Climate Change^^xsd:string	110024
onco:Course41	Pyramids of Giza: Ancient Egyptian Art and Archaeology^^xsd:string	123870
onco:Course1	How to Learn Online^^xsd:string	124980
onco:Course48	Christianity Through Its Scriptures^^xsd:string	125051
onco:Course45	Improving Global Health: Focusing on Quality and Safety^^xsd:string	127353
onco:Course4	The Analytics Edge^^xsd:string	129555
onco:Course35	Hinduism Through Its Scriptures^^xsd:string	140782
onco:Course40	Science & Cooking: From Haute Cuisine to Soft Matter Science chemistry^^xsd:string	145011
onco:Course47	Data Science: Visualization^^xsd:string	171258
onco:Course29	Child Protection: Children's Rights in Theory and Practice^^xsd:string	212453
onco:Course18	Python Basics for Data Science^^xsd:string	228809
onco:Course50	Leaders of Learning^^xsd:string	231131
onco:Course33	Principles of Biochemistry^^xsd:string	257600
onco:Course21	CS50's Introduction to Artificial Intelligence with Python^^xsd:string	258006
onco:Course34	Using Python for Research^^xsd:string	262183
onco:Course28	Data Science: Machine Learning^^xsd:string	276209
onco:Course43	Excel: Fundamentos y herramientas^^xsd:string	284279
onco:Course2	Programming for Everybody (Getting Started with Python)^^xsd:string	293864
onco:Course6	Introduction to Project Management^^xsd:string	301793
onco:Course39	Statistics and R^^xsd:string	314586
onco:Course23	Rhetoric: The Art of Persuasive Writing and Public Speaking^^xsd:string	320225

Figure 10: Shows courses having more than 100000 enrolled students

- Query 6: find introductory courses from a particular university on a particular topic.

```
SELECT ?Course ?title ?level
      WHERE { ?Course onco:sponsored_by ?HarvardUniversity .
              ?Course onco:title ?title.
              ?Course onco:has_course_level ?level .
              FILTER regex(?title, "python", "i") |
              FILTER (?level = onco:Introductory)
              }
```

Course	title	level
Course2	"Programming for Everybody (Getting Started with Python)"	Introductory
Course18	"Python Basics for Data Science"	Introductory
Course21	"CS50's Introduction to Artificial Intelligence with Python"	Introductory

Figure 11: Show the courses provided by Harvard University on topics related to python which are of beginner level

- Query 7: identify the number of courses available on different subjects.

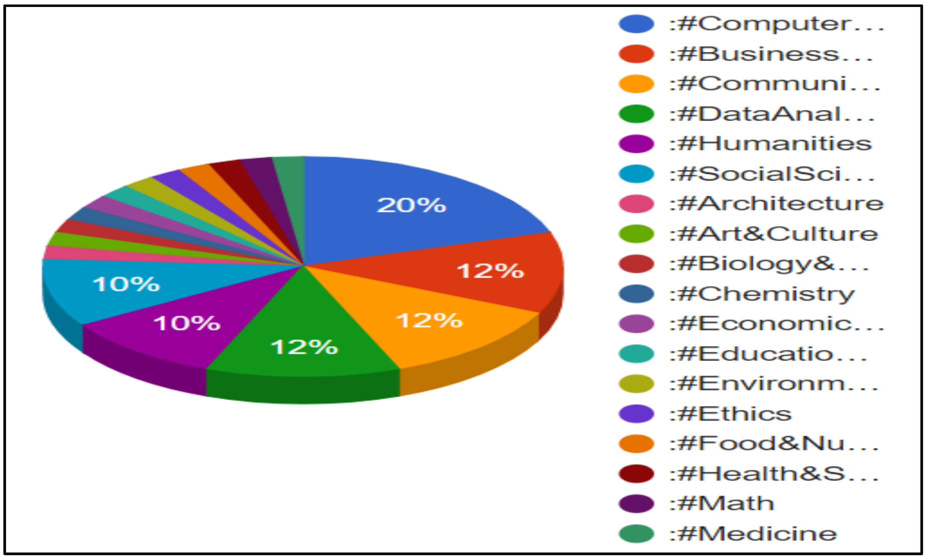


Figure 12: Shows the percentage of the number of courses available on various subjects.

Some of the queries have been shown to provide an idea of what we can expect from the ONCO graph. Since the name and syllabus of the course are a string, a regex filter has been used to identify a particular topic in figure 6. The model can retrieve information about instructors that can be depicted in Figures 8 and 9. Figure 10 shows the arithmetic filters on integers and strings. Figure 11 also shows one of the complex queries by using double filters. Last but not the least, the query is used for information purposes as it shows the count of courses on various available subjects and shows the visualization of the number of courses as a percentage of available courses on various different subjects which are based on the query of counting courses of various subjects in figure 12. Most of the courses are related to the computer science field.

In the process of modelling an ontology, the first step is to identify the domain and the problem that is required to be solved. Next, is to identify the sources of information and brainstorm to identify with precise questions or in other words, competency questions. The modelling is like three planes of work given by Ranganathan in 1967. First, the idea plane where the main process of modelling starts as modelling is a mental process followed by a verbal plane in which the idea is presented in natural language in the form of questions and terms that can be used in the model. Finally, the notational plane is implemented from various shared vocabularies such as foaf, dc, etc.

With the use of protégé, the modelling of class, object property and data property with constraints is developed. A sample data of 7402 courses is merged in the ontology with the use of cellfie tool and rules based on is on OntoGraf, SPARQL query and Span SPARQL query are applied to display and validate the knowledge graph.

3. Related work

El-Ghalayini (2011) constructed an ontology for developing an e-course to describe the concepts of the e-learning domain. An e-Course ontology can improve the quality and effectiveness of e-Learning courses by providing a clear and standardized structure for the course content. Tulasi et al. 2013 developed a generic ontology for the concepts related to the e-learning domain. The current study is focused on the retrieval of courses using different facets of courses from MOOC platforms rather than creating content of courses for e-learning such as Course, Unit, Lesson, and Activity and generalized retrieval of information regarding the e-learning domain.

Sammour et al. (2015) presented an ontology for recommending MOOC courses based on the learner's knowledge of subjects. Harrathi et al. (2017) proposed a hybrid recommender system based on ontology, knowledge, and rules. The system provides recommendations based on learners' and learners' activities in the setting of MOOCs. Rabahallah et al. (2018) also proposed a hybrid approach based on ontology and collaborative filtering to recommend online courses based on learners' activity in MOOC platforms which led to lessening the dropout rate. These studies focused on the recommendation of courses based on learners' existing knowledge and activities in MOOC platforms rather than supporting the free will of students in search of different courses from different domains based on personal requirements.

Dang, Tang, & Li (2019) developed a knowledge graph called MOOC-KG which is based on Curriculum course syllabus ontology. They have constructed the knowledge graph to represent and capture information about MOOCs. They focused more on the Knowledge graph than formulating the ontology. Abdulazeez & Salah (2020) presented an ontology specific to the Coursera MOOC platform with the scope of finding courses according to learners' interests and navigating through the course. They had limited the work to Coursera, some of the classes and subclasses might be irrelevant while implementing for all MOOC platforms and didn't follow any methodology to develop the ontology.

Gupta & Sabitha (2020) identified interaction with the instructor and social environment as the cause of dropout and provided a solution by constructing an ontology to deal with the issue of dropouts in MOOCs. Some learners may not have the perseverance to complete a course that takes a long time to finish. The ability to identify a course with parameters such as course length and course mode can significantly reduce the dropout rate that is considered in the current study.

The current study focuses on the retrieval of relevant courses based on different parameters such as subject area, university, instructor, language, knowledge level and so on.

4. Conclusion and Future prospect

In this present study, an ontology has been constructed to retrieve information related to courses available on various MOOC platforms based on level of knowledge, subject, language, university, instructor, etc. Ontology helps in retrieving and representing knowledge in a well-structured manner.

ONCO can be implemented for educational courses that are delivered online by institutions. The ONCO ontology can be extended to overall e-learning which will include the electronic materials provided by institutions to their students. In the covid 19 pandemic situation, almost every school, college, and university have shifted to online learning. But access to study materials and class recordings is limited to specific students of that particular institution and over time course materials will increase and searchability will be an issue. An institute can use the ONCO ontology in the backend of the platform from where materials are delivered so that students will have better access to the content of the courses.

Future plan is to validate the ontology by utilizing course data from various MOOC platforms. The evaluation of the ontology will be carried out by experts and students engaged in online learning. Further, a website or a system can be built where the model ontology will be implemented in the backend and data can be integrated by implementing various technologies for crawling and cleaning the data automatically.

REFERENCE

Abdulazeez, D. H., & Salah, R. M. (2020). Developing an ontology for retrieving massive open online courses (moocs) information in coursera platform. Journal of Duhok University, 23(1), 103-114.

Asim, M. N., Wasim, M., Khan, M. U. G., Mahmood, N., & Mahmood, W. 2019. "The use of ontology in retrieval: a study on textual, multilingual, and multimedia retrieval." *IEEE Access* 7:21662-21686.

Benjamin Perakath, C. 1994. "The IDEF5 Ontology Description Capture Method Overview." Texas: Knoledge Based Systems Incorporation. 1994.

Conache, Maria, Ramona Dima, and Andreea Mutu. 2016. "A Comparative Analysis of MOOC (Massive Open Online Course) Platforms." *Informatica Economica* 20 (2). https://doi.org/10.12948/issn14531305/20.2.2016.01.

Dang, Furong, Jintao Tang, and Shasha Li. 2019. "MOOC-KG: A MOOC Knowledge Graph for Cross-Platform Online Learning Resources." *2019 IEEE 9th International Conference on Electronics Information and Emergency Communication (ICEIEC)*, July, 596–603. https://doi.org/10.1109/ICEIEC.2019.8784572.

Dutta, B., P Madalli, D. and Prasad, A.R.D., 2009. "Ontology supported personalized eLearning repositories."

Dutta, Biswanath, Usashi Chatterjee, and Devika P. Madalli. 2015. "Yamo: Yet Another Methodology for Large-Scale Faceted Ontology Construction." *Journal of Knowledge Management* 19 (1): 6–24. https://doi.org/10.1108/JKM-10-2014-0439/FULL/HTML.

El-Ghalayini, Haya. 2011. "E-Course Ontology for Developing e-Learning Courses." *Proceedings - 4th International Conference on Developments in ESystems Engineering, DeSE 2011*, 245–49. https://doi.org/10.1109/DeSE.2011.29.

Ferndndez, Mariano, A. Gomez-Perez, and Natalia Juristo. 1997. "Methontology: From Ontological Art towards Ontological Engineering." *Proceedings of the AAAI97 Spring Symposium Series on Ontological Engineering*, 33–40. https://oa.upm.es/id/eprint/5484.

Gai, K., Qiu, M., Jayaraman, S., & Tao, L. 2015. "Ontology-based knowledge representation for secure self-diagnosis in patient-centered teleheath with cloud systems." *2015 IEEE 2nd International Conference on Cyber Security and Cloud Computing*, November, 98-103.

"GraphDB™ - Ontotext." n.d. Accessed July 1, 2022. https://www.ontotext.com/products/graphdb/.

Guarino, N. 1995. "Formal ontology, conceptual analysis and knowledge representation." *International journal of human-computer studies* 43(5-6): 625-640.

Gupta, S, and A. S. Sabitha. 2020. "Designing Ontology for Massive Open Online Courses Using Protégé." *2020 8th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO).*

Harrathi, M, N Touzani, and R Braham. 2017. "A Hybrid Knowledge-Based Approach for Recommending Massive Learning Activities." 2017 IEEE/ACS 14th International Conference on Computer Systems and Applications (AICCSA), 49–54. https://ieeexplore.ieee.org/abstract/document/8308261/?casa_token=NOSoxcCg1bAAAAAA:CfZh-SFmyLtyVyR0CTgAWyC4vdRg93-DRdXCxfU1ka492UnliPp5hKA6Rk5mneYmzXbO9r_zwA.

"Protégé." n.d. Accessed July 1, 2022. https://protege.stanford.edu/.

Rabahallah, K, L Mahdaoui, and F Azouaou. 2018. "MOOCs Recommender System Using Ontology and Memory-Based Collaborative Filtering." *Proceedings of the 20th International Conference on Enterprise Information Systems (ICEIS 2018)* 1: 635–41. https://doi.org/10.5220/0006786006350641.

Ranganathan, S.R. 1967. "Prolegomena to library classification." *Asia Publishing House*, New York.

Sammour, G, A Al-Zoubi, A. Gladun, K. Khala, and J. Schreurs. 2015. "Semantic Web and Ontologies for Personalisation of Learning in MOOCs." *2015 IEEE Seventh International Conference on Intelligent Computing and Information Systems (ICICIS)*, 183–86. https://ieeexplore.ieee.org/abstract/document/7397219/?casa_token=VUrAmnCFt98AAAAA:PVjTtITVQ27cY4kFnj0QH4zMIwskgGih2ZD8Pjk0p_lxp90ty-WHRwHbV2BjzMZksLW3zoGqgA.

Sirin, E., Parsia, B., Grau, B. C., Kalyanpur, A., & Katz, Y. 2007. "Pellet: A practical owl-dl reasoner." *Journal of Web Semantics* 5(2): 51-53.

Studer, R., Benjamins, V.R., Fensel, D. 1998. "Knowledge engineering: Principles and methods." *Data and Knowledge Engineering* 25 (1-2): 161-197.

Tulasi, R. L., Rao, M. S., & Gouda, G. R. 2013. "Study of e-learning information retrieval model based on ontology." *International Journal of Computer Applications* 61(17).