

## Encouraging and Utilizing Linked Data from Open Online Courses

Ruchira Koli, Manali Lakhan, Vighnesh Mahadik

Department of Computer Engineering  
DattaMeghe College Of Engineering, Airoli.

Mrs. Nita Patil (Asst. Prof.).

Department of Computer Engineering  
DattaMeghe College Of Engineering, Airoli.

**Abstract** - Access to affordable education to achieve printed and digital literacy helping all learners to acquire knowledge, coping with change, and seeding mindsets for creativity and intellectual curiosity are considered major indicators and measures of quality of life worldwide. The emergence of MOOCs (Massive Open Online Courses) promising new, scalable models that can provide an “education for everyone” has generated a new and broad interest in rethinking learning and education. Frames of reference (identifying underlying assumptions, conceptualizations, and perspectives) are needed to conceptualize the meaning and the implications of MOOCs in the context of rich landscapes for learning. Most of the discussions and analyses about MOOCs have been based on economic perspectives and technological perspectives. This contribution critically assesses MOOCs from a learning sciences perspectives. This paper focuses on integrating all the trending websites which includes Coursera, Udacity and Swayam and searches for the best optimal course that the user requires. Information is retrieved using web crawler with the help of ontology schema.

\*\*\*\*\*

### I. INTRODUCTION -

In Linked data the data is interlinked and can be used through queries which help to retrieval of information based on syntactic, semantic and structural information contained in data. Principles of Linked data are 1. Data is identified using URIs. 2. Provides useful information via open standards such as RDF (Resource Description Framework), SPARQL etc. Massive open online course is aimed at unlimited participation and open access via the web. There is a need to increase interaction between the students and the professors. The aim of this project is to reuse and remixing of resources from different sites which contain courses. This MOOC gives free access for students who require the optimal courses with ease.

In this paper we are going to focus on three main sites which are Udacity, Swayam and Coursera.

Udacity is a for-profit educational organization founded by Sebastian Thrun, David Stavens, and Mike Sokolsky offering massive open online courses (MOOCs). According to Thrun who is founder of Udacity, the origin of the name Udacity comes from the

company's desire to be "audacious for you, the student". While it originally focused on offering university-style courses, it now focuses more on vocational courses for professionals.

Coursera works with universities and other organizations to offer online courses, specializations, and degrees in a variety of subjects, such as engineering, humanities, medicine, biology, social sciences, mathematics, business, computer science, digital marketing, data science, and others. As of October 2017, Coursera had more than 28 million registered users and more than 2,000 courses [1].

SWAYAM is a programme initiated by Government of India and designed to achieve the three cardinal principles of Education Policy viz., access, equity and quality. The objective of this effort is to take the best teaching learning resources to all, including the most disadvantaged. SWAYAM seeks to bridge the digital divide for students who have hitherto remained untouched by the digital revolution and have not been able to join the mainstream of the knowledge economy.

Benefits of MOOC are as follows:

1. It helps students to find a right course.
2. Courses are available to large and diverse audience across the globe.
3. It provides easy access to global resources and promotes sharing of ideas and knowledge.
4. It enhances active learning.
5. MOOC provides the opportunity to learn from world class universities and from renowned instructor.
6. It promotes knowledge sharing in discussion forum.

MOOC Link is a web application which aggregates online courses as Linked Data and utilizes

that data to discover and compare online courseware. In order to incorporate MOOC data into the Linked Data cloud as well as demonstrate the potential of Linked Data when applied to education, we propose to (i) build or extend an RDF which is used to maintain ontology that denotes MOOC properties and relationships (ii) use our ontology to generate Linked Data from multiple MOOC providers and (iii) implement this data in a practical web application that allows users to discover courses across different MOOC providers.

**Objectives-**

1. Represent first class courses from the renowned professor coming from elite institution.
2. Bring the education in the world to the most remote corners of the planet.
3. Support communities among the students participants thereby expanding their intellectual and personal networks.
4. Provide students with insightful feedback in case they went wrong or got stuck in a problem.

**II. RELATED WORK-**

In Linking and Maintaining Quality of Data about MOOCs Using Semantic Computing[2] by DhekneChinmay and Bansal Srividya, focuses on devising strategies to effectively integrate data from multiple sources and publish it as Linked Data, and maintain the quality of Linked Data. It presented the integration of data from various MOOC providers and algorithms for incrementally updating linked data to maintain their quality.

In Implementation of an Efficient Web Crawler to Search Medicinal Plants and Relevant Diseases [3] by PawarNisha and Dr. K Rajeshwari, the Naive Bayes algorithm is used to define whether the current web page is relevant or not related to the medicinal plant information. If the web page is related to medicinal plant information then it is classified as 'Yes' and if the web page is not related to medicinal plant information then it is classified as 'No'.

In MOOC Link: Building and Utilizing Linked Data from Massive Open Online Courses [4] by Kagemann Sebastian and Bansal Srividya, Massive Open Online Courses or MOOCs are online courses accessible to anyone on the web. Hundreds of institutions have joined in an effort to make education

more accessible by teaming up with MOOC providers such as Udacity, Coursera and udemy.

In Information Retrieval in Web Crawling: A Survey [5] by Saini Chandani and Arora Vinay, Ontology is used for structuring and filtering the knowledge repository approach: Ontology based web crawlers use ontological concepts for improving their performance. So, it will be very easy to get relevant data as per the user requirements.

In RCrawler: An R package for parallel for Web Crawling and Scraping [6] by Salim Khalil and Mohamed Fakir, when a crawler can also perform data extraction during crawling it can be referred to as a web scraper. Scraper Scrapes a large amount of data automatically, they turn to external tools to collect URLs or carry out the complete task.

Sr.no	Paper	Year	Techniques	Outcomes
1	Linking and maintaining quality of data about MOOCs using Semantic Computing	2017	SPARQL and RDF	User can search for any course or keywords. SPARQL query engine will search throughout the RDF triples and return a list of courses with an exact or a partial match with the keyword
2	Implementation of an Efficient Web Crawler to Search Medicinal Plants and Relevant Diseases	2016	Web crawler	Efficient web crawler is used to search the web pages for a medicinal plant domain.
3	MOOC Link: Building and Utilizing Linked Data from Massive Open Online Courses	2015	MOOC Link web application	It serves with optimal courses that are available.
4	Information Retrieval in Web Crawling: A Survey	2016	Web crawling ontology	Web crawling
5	RCrawler: An R package for parallel web Crawling and package	2017	Scraping	scrape a large amount of data automatically

**III. EXISTING SYSTEM-**

Though there are various number of courses available on internet, it is not feasible for the user to go through each and every course and analyse whether the course is viable or not. The user cannot differentiate

whether the courses available are by renowned author. In existing system the user has to search through each website for searching a particular course. The user has to visit each and every course and has to compare as in which courses are more optimal for his choice. This procedure is more complicated and time consuming. So there has to be a system which gives optimal solution to the user in an efficient manner.

Despite the increasing prevalence of Linked Data, there are a limited number of applications that implement and take advantage of its capabilities, particularly in the domain of education . The procedure for searching the process is lengthy and time consuming.

#### IV. PROPOSED SYSTEM-

In this proposed system we will be solving the existing problem where the user had to go through all the websites one by one to get the required and satisfactory course. This is solved by aggregating all the websites into one domain. Here we consider open online courses for three websites which are Udacity, Coursera and Swayam.

It provides easy access to global resources and promotes sharing of ideas and knowledge. MOOC also enhances active learning. MOOC provides the opportunity to learn from world class universities and from renowned instructor. It promotes knowledge sharing in discussion forum.

#### V. METHODOLOGY-

1. Maintaining Ontology : RDF Schema (RDFS) is a language for writing ontologies. An ontology is a model of (a relevant part of) the world, listing the types of object, the relationships that connect them, and constraints on the ways that objects and relationships can be combined. It is more like a type of hierarchy.

Example : Suppose there is a course by Bal Krishna Swami on C++ and if the user searches for the author Bal Krishna Swami this book should be displayed.

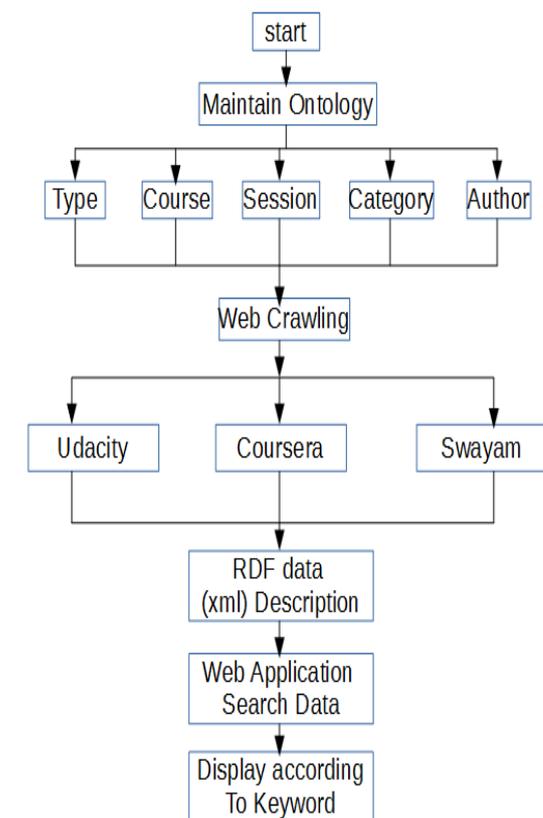
2. RDF Schema : RDF Schema (Resource Description Framework Schema, variously abbreviated as RDFS, RDF(S), RDF-S, or RDF/S) is a set of classes with certain properties using the RDF extensible knowledge representation data model, providing basic elements for the description of ontologies, otherwise called RDF

vocabularies, intended to structure RDF resources. These resources can be saved in a triple-store to reach them with the query language SPARQL.

Example : Suppose Bal Krishna Swami is and author of C++ and also he has published other books on Java, HTML etc then the cross relationship between them is maintained using RDF Schema.

3. Web Crawler : A web crawler (also known as a web spider or web robot) is a program or automated script which browses the World Wide Web in a methodical, automated manner. We retrieve Coursera and Swayam course properties via their course catalog API but use screen scrapers for Udacity. The URLs that are fetched by the web crawler are selected using Xpath selector “//strong/a/@href“ then iterates through each site. .Request is made for the URL selected, calling the parse\_details method.

Example : We consider a site www.udacity.com/computerscience/androidstudio then we parse the URL ( using split() function which is inbuilt) and then find the particular course android studio in subject computer science of all the three websites.The next method, parse\_details,collects attributes from individual course pages using XPath selectors.



When the user searches for a course that query is converted into tokens, these tokens are then searched by categories like author, type, session etc. These categories are maintained using ontology schema. Simultaneously a data set is maintained whose data is collected using web crawler which is used to fetch courses through API's of the websites. So the query searched by the user is searched within this data set and corresponding result is displayed.

## VI. RESULT

Initially, the user wants to search their desired course. There are various categories and sub-categories related to courses. Here, in figure 1, the user enters their desired search as Android.

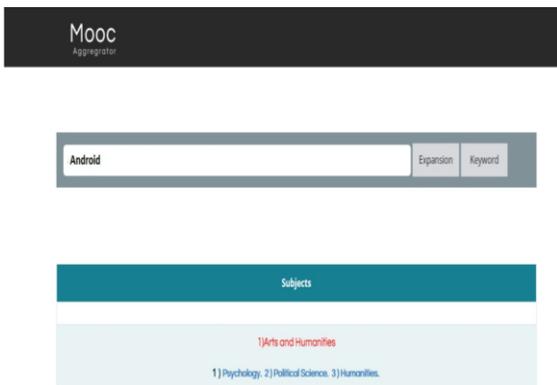


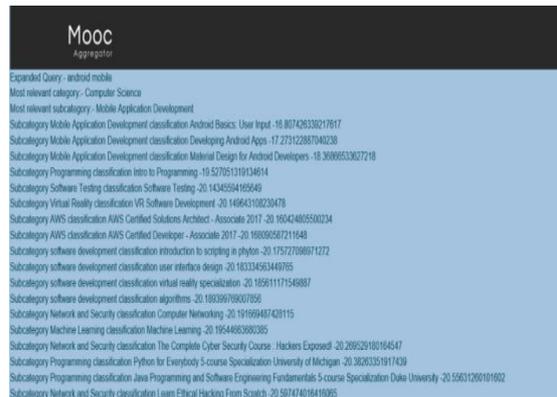
Figure 1

The result obtained for the course related to Android is shown in Figure 2. The result gives the related categories for the word searched along with the website on which the course is available.

Course Title	URL
Material Design for Android Developers	<a href="https://www.udacity.com/course/material-design-for-android-developers--ud892">https://www.udacity.com/course/material-design-for-android-developers--ud892</a>
Android Basics: User Input	<a href="https://www.udacity.com/course/android-basics-user-input--ud836">https://www.udacity.com/course/android-basics-user-input--ud836</a>
Developing Android Apps	<a href="https://www.udacity.com/course/developing-android-apps--ud853">https://www.udacity.com/course/developing-android-apps--ud853</a>

Figure 2

Later the related category and sub-categories of the courses available on the MOOC websites are displayed.



## VII. CONCLUSION-

In this project We describe a prototype implementation of MOOC Link, a web application which utilizes the Linked MOOC Data to allow users to discover and compare similar online courses. Also presented is our approach for collecting and generating Linked Data from three MOOC providers: Coursera, Swayam, and Udacity.

## REFERENCES-

- [1] "Coursera focuses on corporate training programmes for future growth". "Mint". Retrieved 17 October 2017.
- [2] DhekneChinmay, and Srividya K. Bansal. "Linking and Maintaining Quality of Data about MOOCs Using Semantic Computing." Semantic Computing (ICSC), 2017 IEEE 11th International Conference on. IEEE, 2017.
- [3] Pawar Nisha, K. Rajeswari, and Aniruddha Joshi, "Implementation of an efficient web crawler to search medicinal plants and relevant diseases." Computing Communication Control and automation (ICCUBEA), 2016 International Conference on. IEEE, 2016.
- [4] Sebastian Kagemann and Srividya Bansal, "MOOC Link: Building and Utilizing Linked Data from Massive Open Online Courses" IEEE 9th International Conference on Semantic Computing, 2015.
- [5] Saini Chandni and Vinay Arora. "Information retrieval in web crawling: A survey." Advances in Computing, Communications and Informatics (ICACCI), 2016 International Conference on. IEEE, 2016.
- [6] Salim Khalil and Mohamed Fakir, "RCrawler: An R package for parallel for Web Crawling and Scraping" Department of Information, Faculty Of Sciences and Technics, 2017.
- [7] (2017) The Udacity website. [Online]. Available: <https://in.udacity.com/>
- [8] (2017) The Coursera website. [Online]. Available: <https://www.coursera.org/>
- [9] (2017) Swayam website [Online]. Available: <https://swayam.gov.in/>