International Conference On Emanations in Modern Technology and Engineering (ICEMTE-2017) ISSN: 2321-8169
Volume: 5 Issue: 3 230 - 233

Survey on Application of Semantic Web on Massive Open Online Course (MOOC) using Linked Data- part 1

Anjali D.Dwivedi
Student, M.E. Computer Engineering
Shree L.R.Tiwari College of engineering, Mumbai University
Mumbai, India
danjali319@gmail.com

Varsha Bhosale
Vice Principle and Associate professor, Information Technology
Vidiyalankar Institute of technology, Mumbai University
Mumbai, India
varsha.bhosale@vit.edu.in

Abstract-The term Linked Data is best practices for connecting and publishing structured data on the Web. Over a past few years these practice have been adopted by increasing number of data providers, leading to the creation of a data space containing billions of assertions- the Web of Data. Linked Data is an emerging trend used by the top companies for promoting their own means of marking up semantic data, connecting and publishing data on Web. Despite the increasing generality of Linked Data, there are a limited number of applications that take advantage of its capability, particularly in the domain of online/open education on web. In this project we are using Semantic web technologies to create a semantic data model for online/open educational data, specifically about Massive Open Online Courses (MOOCs), connecting and publishing this data as linked data on the Web. Data from various open/online courses educational providers is collected, integrated and published as Linked Data. We present a web portal that utilizes the data to discover and compare open courseware

Keywords—*Linked data; Semantic web; Mooc;*

*****

I. INTRODUCTION

Linked Data include using the Web to create typed links between data from different sources [1]. Source data may vary in size, location, subject-matter and how compatibly it is structured. Typed links generated from this data create define properties explicitly and uniformity in an effort to make data machine readable form. This is opening up opportunities for applications that were previously impractical such as Berners-Lee's intelligent agents [2].

Linked Data is relatively not evaluated in the domain of education. Although there are several data models for structuring educational data as well as repositories adopting these models, Linked Data-driven educational applications are far and few between. As a result, initiatives such as the LinkedUp Challenge have surfaced to encourage innovative applications focused on open educational data [3].

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 edX and Courser. Delivering course content through videos and lectures as well as traditional materials such as problem sets and reading, MOOCs encourage interactivity between student and professors around the world by way of graded assessment and discussion forums.

Coursera, a leading MOOC provider, offers a RESTful API [6] for most information associated with their course catalog. This includes properties such as a course's instructor, title and syllabus details. Although Coursera's course catalog data is easily accessible as JSON, there is no option to retrieve and use it in a Linked Data format such as the Resource Description Framework (RDF). Moreover, there is no Linked Data available for MOOCs or an ontology that denotes properties distinct to MOOCs.

II. OBJECTIVE

The objective of this research is to develop a Web Portal that uses the concept of linked data and ontology to develop a semantic web data model for education. This web portal will collect data from different MOOC (massive open online course) providers, integrate these data and published them, so that it will be easy for student, researcher or people to search for different courses, syllabus, topics, lectures, videos and even they can compare same topic from different course providers and select which one is better to read.

III. LITERATURE REVIEW

Linked Data practices and principle have been adopted by a growing number of data providers, resulting in the creation of a global data space on the Web that contains billions of RDF triples. Linked Data has the potential to enable a regime change in how data is utilized and access. The success of Web APIs has shown the efficiency of applications that can be created by merging or mashing up content from different Web data sources. However, mash up developers face the challenge of scaling their development approach beyond predefined data silos. In contrast, Linked Data understand the view of evolving the Web into a global data space, allowing applications to function on top of an disconnected set of data sources, via standardized access mechanisms [1].

In [2] The Semantic Web involves publishing in languages specifically designed for data: Web Ontology Language (OWL), Resource Description Framework (RDF), and Extensible Markup Language (XML). HTML describes the links between them. XML, RDF and OWL by contrast, can describe arbitrary things such as meeting, people or airplane parts. These technologies are combined in order to provide descriptions that replace and supplement the content of Web documents. The machine-readable descriptions provide content managers to add meaning to the content, i.e., to describe the overall structure of the knowledge we have about that

particular content. In this way, a machine can process knowledge itself using processes similar to human inferences and deductive reasoning, thereby obtaining meaningful results and helping computers to gather information automatically. Semantic Web technology is dependent on ontology as a tool for modeling an abstract view of the contextual semantic and real world analysis of documents. Therefore, the success of the semantic Web depends on the proliferation of ontologies, which requires easy and fast engineering of ontology and avoidance of a knowledge acquisition bottleneck [3] by defining shared and common domain theories, ontology helps both machine and people to communicate precisely to support the exchange of semantics. Ontology language editors help to build semantic Web. Hence, the fast and cheap construction of domain specific ontology is crucial for the success of the semantic Web.

In [4] MOOCs have attracted thousands of people from all over the world; however, closer inspection of the learner demograraphics suggests that the majority of people served by these courses have achieved a university education. As such, at present, they online courses seem to be only increasing access to higher education instead of broadening access. With the changes in higher education sector funding and sternness measures MOOCs seem to be fulfilling the needs of “knowledge workers” in professional development and upgrading skills development. Therefore, at present MOOCs seem to be continuously serving the professional development sector. MOOCs can also provide value for leisure learners.

In [10] a new and efficient technique is used to assign a page rank for domains on particular web portal. In this technique clustering is used to assign page rank to the different domains on the basis of links connected to domain. For this purpose a KDD dataset with links are used. In previous techniques a page rank algorithm, weighted page rank algorithm, hyper induced topic search algorithm. But these algorithms are not efficient to provide better link analysis thus an Enhanced page rank over domain (EPRAD) technique is created and used.

IV. PROBLEM STATEMENT

Linked Data involves using the Web to create typed links between data from different sources [1]. Typed links produced from this data create uniformity and define properties explicitly in an effort to make data easier to read for machines. This is opening up opportunities for applications that were previously impractical such as Berners-Lee's intelligent agents [5].

Linked Data is relatively unexplored in the domain of education. Although there are several data models for structuring educational data as well as repositories adopting these models [6], Linked Data-driven educational applications are far and few between. As a result, initiatives such as the LinkedUp Challenge have surfaced to encourage innovative applications focused on open educational data [7].

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. Delivering course content through lecture videos as well as traditional materials such as readings and problem sets, MOOCs encourage interactivity between professors and students around the world by way of discussion forums and graded assessments.

V. PROPOSED SOLUTION FOR THE ABOVE PROBLEM

A lot of work has been done in information retrieval domain of knowledge but no work has been done for efficient course discovery and comparison from MOOC. Thus, this work proposes that the concepts of semantic web used can be used for implementation of semantic search in MOOC. It is basically exploratory work, thus initially, the causes of problems of search from MOOC have been explored and we will create ontology for our web portal, for easy exchange information between program and user.

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 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 allow user to discover courses among different MOOC provider.(iv) categorization of MOOC(v)adding dictionary to search result(vi)mapping of relevant course from other provider(vii)developing item pipeline for crawlers(viii)summaries for more robust data(ix)enabling user profile ,course track, automatic website update.

VI. RANKING ALGORITHM

A Page rank algorithm is used to rank pages in our domain. A ranking is a relationship among a set of items such that, for any two points, the first is either 'ranked lower than', 'ranked higher than', or 'ranked equal to' the second. The rankings themselves are totally merged. With help of Clustering, ranking operations to estimate the occurrence of the target or data items. In [10] a paper proposed to put a value to the ranking of overall design of database. To optimize the effect of K-mean clustering algorithm the ranking functions have introduces new opportunities.

1. Ranking methods are needed to provide a way to search like data search and relevant records is a popular function of database, to get knowledge. That`s why, it need to improve effectiveness and rank the more relevant student marks by a using ranking method and to improve search effectiveness. At final, related answers will be delivered for a given keyword query by better K-means clustering method and ranking method strategy and creating index. Then apply this Ranking method with because this method is likewise causing the property to obtain relevant records. So it is also helpful for creating clusters of all data having similar properties within bunch of data.

2. Weighted Page, a web graph technique were introduced where the technique take the weight of the page rank from the web pages hyperlinks. There are number of algorithms devised which are based on link analysis. Weighted Page Content Rank Algorithm is proposed to give the output to the user based on its high output and search get by using the weighted hyperlink from the web engine search result. Weighted Page Content Rank Algorithm is a score based on which the web pages provide a score of its weight and visiting. This algorithm employs web structure mining. This mining employs the no. of time a page is visited and at the same time no. of pages linked to the current visiting page. It is based on the number of out-links and in-links on the page.

A. EPRAD ALGORITHM

This technique is based on analysis of previously work for mining web dataset, In the existing technique like hyper induced topic search, weighted page rank, page rank algorithm are not able to perform efficient solution for the page rank. Thus Enhanced page rank algorithm (EPRAD) technique is proposed. In the existing technique page rank algorithm make use of link structure to determine the importance of web page.

In Page rank algorithm back links used as input parameter. Page rank not provides efficient result Because of relevancy of the resultant pages. So overcome this problem weighted rank is used that divide the rank value of a page evenly among its out link page. Each page gets a value based on its out-links and in-links. In existing techniques search function is performed on bunch of data and no clustering is performed. Thus Enhanced page rank algorithm technique is proposed [10] where page rank algorithm used for assign rank. In efficient page rank algorithm technique KDD (knowledge description database) dataset is used which contain number of links. And then clustering is performed on the KDD dataset bases of protocol, and then mining is performed over domain.

Step1. Load database containing links (Where no. of links =599).

Step2. Finding K-mean variable.

Step3. Performing K-mean technique on cluster on the basis of protocol.

Step4. Mining the data on the basis of their domain.

Step5. Performing page rank for single link.

Step6. Apply EPRAD over other domains.

EPR (p) = limit of $\sum$ m=0 to i, $\sum$ pr (p)/n Where n= no. of links found on domain , i=no. of link found on database

Step7. EPR, If (n=found) n= no. of links over database

EPRAD= 1-9.

Step8. Else if (n≠ found)

EPRAD=0.

Step9. Else (if there is no connection available) EPR = - Output: EPR (P) Rank Score.

This technique reduced searching time of people/student on our web portal by placing the courses , lectures, videos, notes or information on the top of our website that have been mostly referred by other people/Students. This technique reduces searching time and redundancy and provides efficient result.

VII. PROPOSED MODEL

The Let us finally discuss the proposed framework figure 2. There is various consideration level of proposed framework thus these are being proposed in context with certain already existing framework not only proposed but also implemented. On the basis of framework proposed in [12], the framework has been proposed that explains the layers into better modules and in a sequence

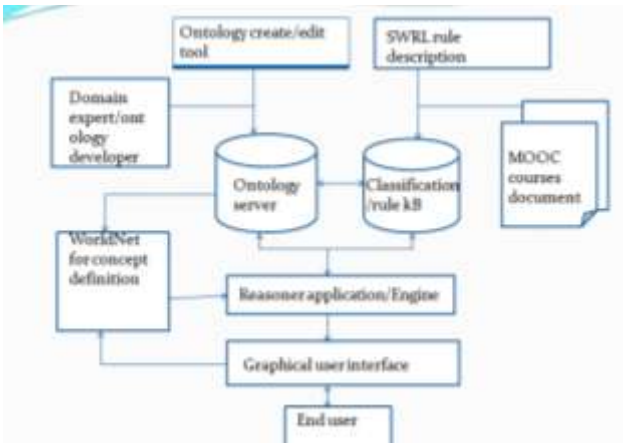

Figure 1:- The Proposed Framework

It clearly depicts that MOOC document will be used for annotations and that will be directly linked with Semantic Web Rule Language that will be used to give us knowledge base consisting of classification and rules. At the same level, ontology developers along with mutual consultation with domain experts will develop domain ontology. Such consultation is mandatory because proper definition of concepts is ultimate goal and no second thought can be taken into consideration. The same care has been taken into account that concepts have only been defined as defined in open online courses such complex architecture.

The figure 2 architecture is being proposed based on [11]. It is evident that query will be processed in the following order Figure 2, that when the user request a query. Then if it is simple keyword based query that it can be bypassed from WordNet, the sole purpose is not to complicate the process into many definition, as not required also less time will be consumed. Then parser will be used as the query can be broken down into token and parsing can be done. Next step is Reasoner, as inference for Reasoner is based on ontologies then thus Reasoner is linked with ontologies data/knowledge base which were created on the basis of open online course text. As we want to do search the concepts from other open online course sources on the basis of ontologies of

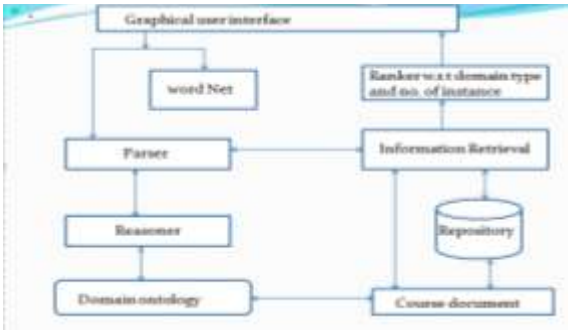

Figure 2:- The query processing for proposed semantic web system

MOOC thus other documents can be annotated for semantic search from those documents. As the number of documents is more the ranking has been proposed to rank the result on simple logic of frequency of occurrences in all those documents present in our web portal

VIII. BACKGROUND

A. Simple Protocol and RDF Query Language

RDF Query Language or simple protocol or SPARQL is an RDF query language that allows users to modify and retrieve data stored in RDF format [7]. SPARQL is used as our application's query language in order to retrieve data to populate web pages with course information.

B. Resource Description Framework

The most notable model for Linked Data is the Resource Description Framework (RDF), which encodes data as predicate, subject, predicate and object triples [7]. The object and subject are both Uniform Resource Identifiers (URIs), while the predicate inform how the subject and object are related to each other, using a URI. For the purposes of this paper, Linked Data is presented as XML/RDF, XML syntax for RDF, which is also used in the development of our application.

C. Linked Data Principles

[1] Tim Berners-Lee established a set of rules for publishing data on the Web so that this data can becomes part of a global space in which every resources are connected. The rules are as follows:

1. Use URIs as names for things.
2. Use HTTP URIs so that people can look up those names.
3. When someone looks up a URI, provide useful information (RDF, SPARQL).
4. Include links to other URIs, so that they can discover more things.

These principles by Tim provide basis knowledge for contributing to a Linked Data in which a many form of datasets from different fields in global space of human knowledge are interconnected. Our project aims to work on these principles.

D. Linked Education Data

Many models have been develop for structuring educational data, among the most popular are the IEEE (LOM) Learning Object Metadata and Specification and Sharable Content Object Reference model (SCORM). LOM is converted in XML and includes nine categories with sub-elements that hold data. An RDF binding for LOM exists [9]. SCORM is an extensive technical standard, typically encoded in XML, that defines how educational content should delivered, packaged and how learners navigate between various parts of an online course [11].

In 2013, the Learning Resource Metadata Initiative (LRMI) specification was added into Schema.org's vocabulary for tagging educational content. The properties added in this adoption introduced fields for online course details including the type of time required, learning resource and so on. While there is significant overlap between LRMI's additions to Schema.org, Schema.org's Creative Work properties and MOOC course details like those provided in Coursera's API, several crucial missing data fields such as course difficulties, syllabus details and predicates linking courses to other objects, make it necessary to extend the vocabulary for MOOC data.

E. Building Ontologies

Ontology is common vocabulary for researcher who needs to share information in a domain. We develop ontology to share common understanding of structure of information among software agent and people to separate domain knowledge from the operational knowledge and to analyze domain knowledge.

IX. SIMULATION ENVIRONMENT

A computer with minimum configuration of Windows 7 with 4 GB RAM and software like SPARQL, XML, APACHE TOMCAT SERVER, FUSEKI SERVER, SCRAPPY TOOL, JSON needs to be installed on the computer for the project work.

X. EXPECTED RESULTS

We will build a web portal that utilizes linked data and semantic web ontology that will collect, integrate and published data from different online educational course providers. It will provide user with all the features mention above in proposed system, so that user can discover and compare similar online courses and efficiently take advantage of open online education.

REFERENCES

[1] Linked Data - The Story So Far Christian Bizer, FreieUniversität Berlin, Germany Tom Heath, Talis Information Ltd, United Kingdom Tim Berners-Lee, Massachusetts Institute of Technology, USA. Special Issue on Linked Data, International Journal on Semantic Web and Information Systems (IJSWIS). http://linkeddata.org/docs/ijswis-special-issue (2011)

[2] ScientificAmerican:TheSemanticWebhttp://www.sciam.com/print_version.cfm?articleID=00048144-10D2-1C70-84A9809EC588EF21 May 17, 2001

[3] Role of Ontology in Semantic Web DESIDOC Journal of Library & Information Technology, Vol. 31, No. 2, March 2011, pp. 116-120

[4] Massive Open Online Courses Tharindu Rekha Liyanagunawardena School of Systems Engineering, University of Reading, Whiteknights, Reading, RG66AY, UK;Academic Editor: Satoshi P. Watanabe Received: 13 November 2014 Published: 28 January 2015

[5] Pappano, L. (2012). The Year of the MOOe. The New York Times, 2(12),2012. Chicago

[6] Coursera Catalog API. (n.d.). - Cow'sera Technology. Retrieved July 22, 2014, from https:lltech.coursera.org/app-platform/catalog/

[7] Klyne, G., & Carroll, J. J. (2006). Resource description framework (RDF): Concepts and abstract syntax.

[8] X. H. Zhang and G. S. Ying, "Ontology-based semantic retrieval system," in Proc. International Conference on Wireless Communication, Networking and Mobile Computing, pp. 1-4, October 12-14, 2008

[9] W. D. Fang, L. Zhang, Y. X. Wang, and S. B. Dong, "Towards a semantic search engine based on ontologies," in Proc. of the Fourth International Conference on Machine Learning and Cybernetics, Guang Zhou, vol. 3, pp. 1913-1918, 18-21 August 2009

[10] An Enhanced Page Rank Algorithm over Domain , Preetibala Deshmukh ,Vikram Garg International Journal of Computer Applications (0975 – 8887) Volume 139 – No.1, April 2016

[11] https://www.csee.umbc.edu/courses/771/papers/ieeeIntelligentSystems/ontologyLearning for Semantic web, Alexander Maedche and Steffen Staab, University of Karlsruhe

[12] X. H. Zhang and G. S. Ying, "Ontology-based semantic retrieval system," in Proc. International Conference on Wireless Communication, Networking and Mobile Computing, pp. 1-4, October 12-14, 2008.