IRACST - International Journal of Computer Science and Information Technology & Security (IJCSITS), ISSN: 2249-9555
Vol. 3, No.2, April 2013

Digital Library Service Integration of Educational Videos Using Linked Data and Semantic Web

Surbhi Nayyar
Dept. of Computer Science and Engineering
SRM University
Chennai, India

S.Nagadevi
Dept. of Computer Science and Engineering
SRM University
Chennai, India

***Abstract*— For e-learning, educational multimedia resources are very useful. Educational organizations are creating number of video resources to offer the best services to their students. The main aim of this paper is to design a platform to search all the multimedia educational video resources using linked data. Ontology based information retrieval is also defined with semantic web. Video resources are being annotated semantically using vocabularies defined in the Linked Data Cloud. Semantically annotated data will provide more meaningful information. Users are allowed to browse semantically linked video resources with enhanced web information from different online resources. In this paper, ranking of video resources is also included to make search easy.**

Keywords—semantic web; distance learning; linked data; annotation; ontology; ranking

I. INTRODUCTION

E-learning activities are becoming popular now-a-days. Digital educational video resources play a key-role in distance learning environment. Different departments of a university manage their video resources separately because they are provided by different authors. Students from one university do not have permission to access the video resources of a particular course available in other university. To efficiently use and search all related educational resources, they should be linked together. More accurate description of videos should be done. Semantic annotation of video resources should be performed so that these can be linked to other web resources using linked data. A semantic-based video searching browser is provided for searching videos [8]. Also, we suggested a new ranking algorithm to find related resources which could be used in a semantic web based search engine.

II. RELATED WORK

With annotation more information will be generated for video resources, which will help students to view complete picture of the topic. These annotations will then be linked to linked data cloud using vocabularies. Linked data cloud is built in order to link the educational video resources with different sources. This linked data is published on web by using restful web services [4]. All the linked data should be readable by machines. In this cloud, data are identified by URIs which contains useful information about the data.

A semantic web based layer is defined which will convert video content features into semantic web standards such as RDF [9]. Ontology based languages are used like RDFS which helps to link relationships between objects. Videos that are being searched using ontology and annotations contain meaningful information as compared to syntactic searching.

We also define a ranking algorithm for sorting videos search results in such a framework. A graph will be designed to show ranking of videos.

III. ARCHITECTURE

Fig.1 shows the overall architecture of annotation and semantic search which consists of linked data cloud containing all annotated linked data.

Administrator will manage uploading of videos which will be done by tutors, domain experts as well as searching of videos which is to be done by users. Tutors can upload as well as annotate the videos to add more meaningful information such as adding text, making changes in timeline, removing noise.

Linked open data cloud can have different information such as geographical locations etc. Semantic search of videos done by users will help them to find related videos both inside and outside linked data cloud.

Modules consist of:

- Upload module in which server stores educational resources such as video, PDF, e-book. Domain experts and tutors can login with valid username and password to server and upload their PDF and video files in the server storage.
- Annotation module in which tutors can modify the video according to requirements to make it more informative.
- User search module in which user can login to the server system to search database for collecting PDF and video files. The administrator will provide user specified query oriented results.

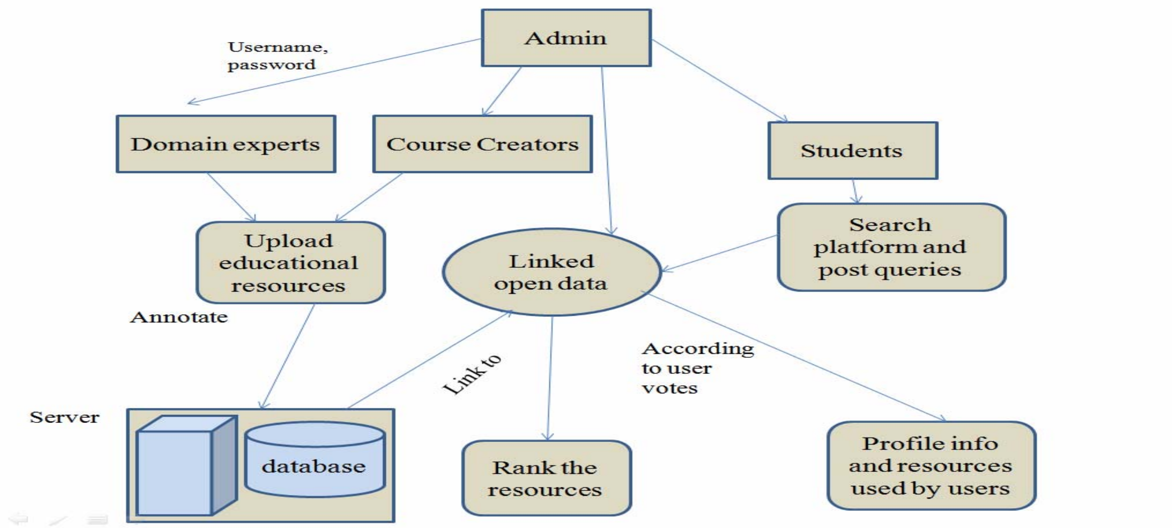

Fig. 1 Overall Architecture to annotate and search educational video resources with linked open data cloud.

- Rank module in which administrator can rank the videos to make search easy.
- Query post module in which users can post their queries to tutors if they have any doubts. In reply tutors will give the solution.

IV. Annotation

With annotation we can modify videos according to user requirements. After pausing video, we can add annotations to instants or durations on the timeline. Videos can be converted to any format like .flv, .avi, .wmv etc. Following features can be added by using a video editor.

1) Text: We can add simple text, animation text, titles and can change its font, size etc.

2) Trim: Duration of videos can be adjusted by changing the end time of videos.

3) Speed: It can be increased or decreased according to requirements.

4) Color: We can change the color of video. In addition we can adjust brightness, contrast and its temperature.

5) Audio Settings: In this removal of noise from video can be done. We can equalize, normalize and even pitch can be shifted.

6) Stabilization: It can be done by adjusting compensation between low and high bars.

7) Transitions: We can apply transitions like fade, rotate, flip etc.

8) Special Effects: like sepia, warmify can also be given.

A. Annotation Ontology

One of the important components of semantic web is ontology. Ontology languages proposed in literature can be classified as XML based or Non-XML based. Among the XML based languages, we have RDF [9], SHOE, OWL [10], DAML [5], XOL etc. and for Non-XML based, we have Unified Modeling Language UML, Cycl [15], Ontolingua [1], etc. This ontology reuses a number of RDF vocabularies.

- Friend-of-a-Friend [6], for identifying users and their accounts.
- Dublin Core, in its RDFS form for metadata like the author, video's title, annotation time.
- Timeline ontology [16], for identifying temporal instants and durations on the video timeline.

The predominant query language for RDF graphs is SPARQL [7]. It is an SQL-like language.

An example of a SPARQL queries to show country capitals in Africa using a fictional ontology.

PREFIX abc: <null: //sparql /example Ontology# >

SELECT? Capital? Country

WHERE {

?x abc: cityname? Capital;

abc: isCapitalOf? y

?y abc: Countryname? Country;

abc: isInContinent abc: Africa }

B. Linked Data approach For Annotation

Linked open data [3] has gained much popularity over the past couple years. It is also called as web of data. Using linked data principles [14] many large-scale datasets have been published.

- All items in a dataset should be identified using URIs;
- All URIs should be dereferenceable: using HTTP URIs [11] allows looking up an item identified through an URI;
- When looking up an URI, it leads to more data.
- Links to URIs in other datasets should be included.

In linked data cloud each vocabulary is accurately defined and controlled. It consists of a unique URI so that it can be distinguished from other vocabularies. With this no conflict can occur between different vocabularies and meanings.

Different vocabularies which describe the same thing are linked using the owl: sameAs property as an equation definition. A number of semantic annotations are used to build relationships between vocabularies such as rdfs: subclassof and rdfs: seeAlso. When any vocabulary is applied to an annotation, the related vocabularies are associated with the annotation. With this approach we can find a good number of related educational resources.

Some linked data services are:

- GeoNames [2] that provides RDF descriptions of geographical features worldwide. It is used to identify named locations using a keyword search.
- DBpedia that contain millions of concepts described by billion triples, including abstracts in different languages.
- Zemanta [17] that provides analysis of natural language text and entities returning their URIs to linked data.

V. SEMANTIC WEB

The current standards for web services provide only syntactic description of resources. With semantic web [13], a semantic layer on top of the web services is provided which facilitates automated discovery, invocation, composition and interoperation of services within an open environment.

In semantic web, resources and services over the web will have well-defined meanings as it attaches semantic information to contents. It enables the automated processing of web content by machines. An approach towards semantic search is shown in fig.2. The annotations are semantically matched to other annotated educational resources from web in order to search any resource.

Users can use semantic search tool in order to interact with web specifying website or documents to get educational video

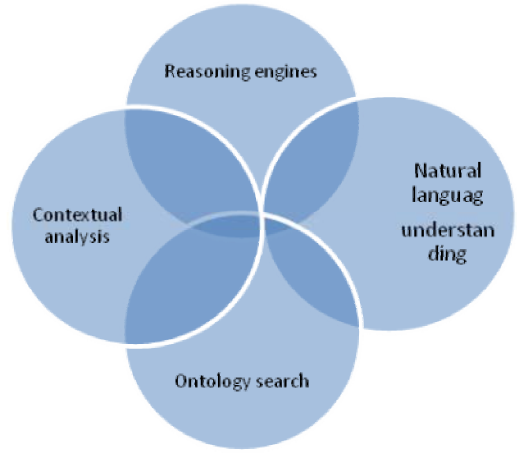

Fig. 2 Approach towards semantic search

resources. Queries are generated to search web. This technique supports searching videos by automatically analyzing documents. The result will also contain relevant learning resources.

It uses tree-structure of ontology based semantic annotations. For example, if http://dbpedia.org/page/india, owlsameAs http://www.freebase.com/view/en/india, and then any video annotated with either of these two URIs will be related to the other. With this searching results will be more accurate. Linked videos from different educational resources are shown in fig. 3.

Some linked data services that are applied to search annotated video resources with semantic search are:

- Sindice [8] that provides services such as keyword-based searching for linked data with websites whose content is marked up with RDF, micro formats, micro data and more.
- Swoogle that employs a system of crawlers to discover RDF documents and HTML documents with embedded RDF content.
- WorldHistory that provides API access to retrieve the information about people, events, places and genealogy in history.

VI. RANKING ALGORITHM

As the number of videos over web is huge, the query results cannot be expected to fit in one page. A ranking algorithm should be there to present the user with top results. For this we define a new ranking algorithm which is based on the Google Normalized Distance [12]. According to relative importance videos will be sorted.

A video is converted to its RDF format and is considered as a collection of child resources. Each resource has a set of

Fig. 3 Linked videos from different educational resources

properties called as descriptors. The child resources could be Shots or Frames or Scenes or Descriptors.

Distance measure between two descriptors D1 and D2 is calculated which are the leaf nodes of the RDF tree and do not have any sub children.

GND (D1, D2) is the Google Normalized Distance [12] between textual content of descriptors D1 and D2.

Distance (D1, D2) = GND (D1, D2) (1)

Now, distance measure for resources is calculated that have one or more children.

Let D1 and D2 be two resources having N1 and N2 children respectively and the child (D1, i) represent the i^{th} child of D1. The distance between D1 and D2 is defined as the sum of distances between its children.

Distance (D1, D2)
=

$$\sum_{i,} \min\left(\sum_{j,} \mathbf{Distance}(\mathbf{Child}(\mathbf{D1},\mathbf{i}),\mathbf{Child}(\mathbf{D2},\mathbf{j}))\right)$$

where i=1 to N1 and j=1 to N2 (2)

We now calculate the distance measure between videos for ranking. If there are N videos, we calculate a NxN matrix where A (i, j) is the distance between the i^{th} and j^{th} video.

A graph is created where videos represent the vertices and the edges represent the distance between two videos given by the following equation 3.

$$VR(i) = (1-d) + \sum_{j=0\ to\ N-1} \frac{VR(j)A(i,j)}{\sum_{k=0\ to\ N-1} A(j,k)}$$

(3)

In equation 3, d is the probability of jumping from one vertex to another, empirically calculated as 0.85. The linear array VR of size N will give ranking of videos. The matrix VR gives the ranking of each of the video. Videos will be arranged according to their ranking. From matrix A we can find related videos. In fig. 4, a graph is generated for ranking of videos.

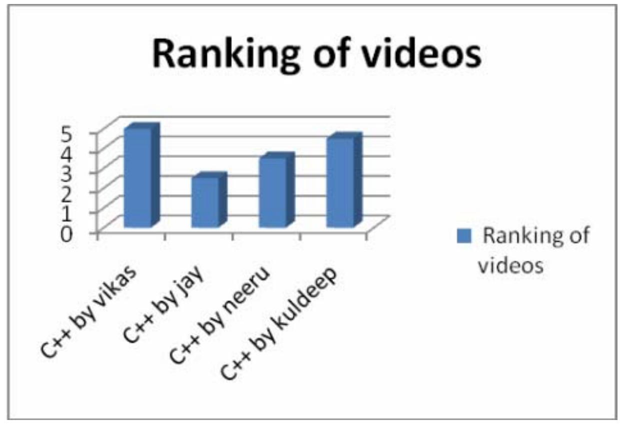

Fig. 4 Ranking of videos

CONCLUSION

This paper defines use of linked data technology and its services in distance learning. Annotation is performed over videos using ontology to make it more informative. Different videos from various sources can be linked and used by many users using semantic search technology. Semantic web will provide relevant data only with value-added information.

Ranking of video resources is also included to make search easy. This will reduce searching time. Linked data vocabulary will help in finding related videos. Further research work will include security of linked data cloud and adding more functionality to annotation.

REFERENCES

[1] A. Farquhar, R. Fikes, and J. Rice, "The ontolingua server: A tool for collaborative ontology construction," International journal of human-computer studies, 1996.

[2] B. Vatant and M. Wick, “Geonames Ontology,” http://www. geonames.org/ontology, 2006.

[3] C. Bizer, T. Heath, and T. Berners-Lee, “Linked Data—The Story So Far,” IJSWIS, Special Issue on Linked Data, 2009.

[4] C. Pautasso, O. Zimmermann, and F. Leymann, “Restful Web Services vs. ‘Big’ Web Services: Making the Right Architectural Decision,” Proc. 17th Int’l Conf. World Wide Web (WWW ’08), pp. 805-814, 2008.

[5] "DAML Home Page". Series "DAML Home Page" Sep, 09 [last accessed on 2010 March 2010]; Available from: http://www.daml.org.

[6] D. Brickley and L. Miller, “FOAF Vocabulary Specification 0.91, ” http:// xmlns.com/foaf/spec, 2007.

[7] E. Prud'hommeaux and A. Seaborne, "SPARQL Query Language for RDF," technical report, World Wide Web Consortium, Jan.2008.

[8] E. Oren, R. Delbru, M. Catasta, R. Cyganiak, H. Stenzhorn, and G. Tummarello, "Sindice.com: A Document-Oriented Lookup Index for Open Linked Data," Int'l J. Metadata Semantics and Ontologies, vol. 3, pp. 37-52, Nov. 2008.

[9] G. Klyne, J.J. Carroll, and B. McBride, "Resource Description Framework (RDF): Concepts and Abstract Syntax," http://www.w3.org/TR/rdf-concepts/ , RDF Core Working Group, W3C Recommendation 10 February 2004.

[10] "OWL 2 Web Ontology Language: Structural Specification and Functional- Style Syntax, W3C Working Draft" . Series," W3C Working Draft [last accessed on March 2010]; Available from: http://www.w3.org/TR/2008/WD-owl2-syntax-20080411/.

[11] R. Fielding, J. Gettys, J. Mogul, H. Frystyk, L. Masinter, P. Leach, and T. Berners-Lee, "Hypertext Transfer Protocol – HTTP/1.1," http://tools.ietf.org/html/rfc2616, IETF Network Working Group, Request for Comments: 2616, June 1999,

[12] Sanjay Ghemawat, Howard Gobioff, Shun-Tak Leung GFS, The Google File System, 19th ACM Symposium on Operating Systems Principles, Lake George, NY ,October, 2003.

[13] T. Berners-Lee, J. Hendler, and O.Lassila, "The Semantic Web," Scientific Am. Magazine, 2001.

[14] T. Berners-Lee, "Linked Data, Design Issues," http://www.w3.org/DesignIssues/LinkedData.html, 2009

[15] "The Syntax of CycL". Series "The Syntax of CycL"[last accessed on March 2010]; Available : http://www.cyc.com/cycdoc/ref/cyclsyntax. html.

[16] Y. Raimond and S.Abdallah, "The Timeline Ontology," http://purl.org/NET/c4dm/timeline.owl#, 2007.

[17] Zemanta API Overview, Stand 13.5, http://www.zemanta.com/api, 2009.