Using linked open data to improve the search of open educational resources for engineering students

Nelson Piedra, Janneth Chicaiza, Jorge López
Departamento de Ciencias de la Computación
Universidad Técnica Particular de Loja, UTPL
Loja, Ecuador
nopiedra@utpl.edu.ec; jachicaiza@utpl.edu.ec;
jalopez2@utpl.edu.ec

Edmundo Tovar
Facultad de Informática
Universidad Politécnica de Madrid, UPM
Madrid, España
etovar@fi.upm.es

***Abstract*— In this paper, authors apply the Linked Data Design Issues to describe and retrieve information that is semantically related to open educational resources related to the Engineering Education, that are accessible via the OCW Higher Institutions. Linked data have the potential of create bridges between OCW data silos. To assess the impact of Linked Data in OCW, the authors present an interface of faceted search for open educational content. The authors demonstrate that OCW resource metadata related to engineering open courses can be consumed and enriched using datasets hosted by the LinkedOpenData cloud.**

Keywords*— *OpenCourseWare; Open Educational Resources; OCW; OER; Linked Data; Faceted Search Engine; Serendipity; LOCWD

I. INTRODUCTION

The key idea the Open Educational Resources (OER) movement is that open educational content should be maximally shared [1]. Heterogeneity leads to problems of interoperability and accessibility of open content among institutions and within them. The lack of interoperability shows some disadvantages in the discovery, reuse, re-mix and adaptation of OER. OER Community must find a way to exchange quick and easy access to open educational materials.

Materials in OCW repositories are not usually described by metadata. Heterogeneity leads to problems of interoperability and accessibility of open content among institutions and within them. The lack of interoperability shows some disadvantages in the discovery, reuse, re-mix and adaptation of OCWs. OCW Community must find a way to exchange quick and easy access to OER.

Considerable work has been devoted to increase the interoperability between Learning Object Repositories that rely on different metadata schemas e.g. IEEE LOM. However, learning object metadata is typically not linked across repositories and not is possible navigate or interoperate between different data sources available on the Web. In this work, this problem is addressed through Linked Data by that describes how linked data has been integrated to data extracted from OCW repositories to navigate OCW resources.

Based on the perspective of Linked Open Data, free open OCW data also fosters interoperability and creates a basis on which the use, re-use, remix, and adaptation of open educational tools or commercial applications can be built more easily. In Section 2, we describe the OpenCourseWare domain, and the notions of Linked Data and Linked OpenCourseWare Data Vocabulary, which ensures that an OER of type OCW can be safely discovered and reused. In Section 3, we present Serendipity, our implementation as well as an experimental evaluation of it. Finally, in Section 4, authors present the conclusions.

II. THE OPENCOURSEWARE DATASET

A. Data Source and Coverage

There is not a standardized way to implement OCW initiatives. The internal organization, structure and technological infrastructure of an OCW project are diverse, and respond to the vision of each university. Current OCW initiatives are Silos of OER. These silos of OER have no way to link to a particular item, and so hinder the free flow of information [8]. In this respect, OCW data is locked away in independent data silos, making it much less useful than it could be.

In this work, the data source is provided by higher education institutions associated to OCW Consortium (www.ocwconsortium.org/) and/or OCW – Universia (ocw.universia.net/).

Our dataset contains data about the main OCW concepts: (i) OER, the OER are teaching, learning, and research resources that reside in the public domain or have been released under an intellectual property license that permits their free use or re-purposing by others. (ii) OpenCourseWare, OCW is a type of OER.; (iii) repositories that contains OER and OCW courses; (iv) educational organizations; (v) users, creators or authors; (vi) branch of knowledge to which an OER belongs; and, (vii) open licenses or similar license that generally allows more liberal use, reuse, redistribution, re-mix and adaptation than a traditional copyrighted work.

B. Use of Linked Data from OCW content

Semantic Web technologies and, more precisely, Linked Data are changing the way information is stored and exploited. [5, 6]. The term “Linked Data” refers to a set of best practices for publishing and connecting structured data on the Web [4]. In summary, the Linked Data Design Issues, outlined by Tim Berners-Lee back in 2006, provide guidelines on how to use

standardized Web technologies to set data-level links between data from different sources [3, 7].

The motivation behind creating the Linked OpenCourseWare Data (LOCWD) is threefold [9]: *Firstly*, data about OCW resources such as those stored in the Serendipity system are important for Educational decision makers, teachers, self-learners and other interested persons. We are particularly interested in courses related to engineering (see Table I and Table II). *Secondly*, although there are several open educational datasets, none provide the level of details that LOCWD does, nor are they available as linked data. And, *thirdly*, the current technology solutions used by OCW and OER sources are limited to offering data via diverse Web pages (not as raw data, but as embedded data within web pages) and very few services through APIs.

The selected open educational content were converted to Linked Data using the LOCWD vocabulary [9]. The resources described in Linked Data/RDF were stored in a RDF-Store. At this point, each resource was identified by a URI with a dereferencing option, and thus display the results retrieved as Linked Data.

TABLE I. OPEN COURSES RELATED TO ENGINEERING (EXTRACT)

University	*OCW quantity*
Massachusetts Institute of Technology, MIT	704
Universidad Carlos III de Madrid, UC3M	148
Universitat Politécnica de Catalunya, UPC	95
The Open University, OU	87
Universidad Politécnica de Madrid, UPM	56
Korea University, KU	38
Universidad de Alicante, UA	38
Universidad Politécnica de Valencia, UPV	33

TABLE II. OCW - KNOWLEDGE AREAS RELATED TO ENGINEERING

Knowledge Area related to Engineering	*Universities*
Electrical Engineering, Technological and Computer Sciences	MIT, UC3M, UPM, UPV, UPC, OU, KU, UA
Mathematics & Physics	MIT, UC3M, UPM
Mechanical Engineering	MIT, UC3M, UPC
Chemical Engineering	MIT, UC3M, UPC
Materials Engineering	UC3M, UPC, UPV

III. CURRENT USAGE

It's difficult to develop tools for consume data from multiple OCW silos. Searching OCW/OER across multiple silos means invoking each one's user interface, and receiving the results in separate groups. This severely impedes the development of applications to OER/OCW that wish to combine data from different sources. However, the use of linked data approach on OCW repositories provides the framework for its evolution into a more interoperable and integrated system to sharing, connecting and discovering data and metadata of OCW initiatives.

The authors have personalized a faceted search engine that consumes data from LOCWD: Serendipity. The first version of Serendipity is based on flamenco.berkeley.edu. With Serendipity we explore the potential of LOCWD in supporting discovering process to give assistance to use, reuse and remix OER and OCW resources. See fig 1.

A. *Serendipity a Faceted Seach for OpenCourseWare Content*

Serendipity is an interface of faceted search accessible from http://serendipity.utpl.edu.ec/. It is based on data extracted from OCW sites. Serendipity provides a search interface for allowing users to browse OpenCourseWare content in such a way that they can rapidly get acquainted with the scope and nature of the content, and never feel lost in the data. This interface exposes OCW metadata in such a way that users can build their queries as they go, refining or expanding the current query, with results automatically reflecting the current query. This interface also combines free-text search, it avoids complex search forms.

Faceted search, also called faceted navigation, is a technique for accessing content organized according to a faceted classification system, allowing users to explore a collection of information by applying dynamic and multiple filters. A faceted classification system classifies each information element along multiple explicit dimensions, enabling the classifications to be accessed and ordered in multiple ways rather than in a single, pre-determined, taxonomic order.

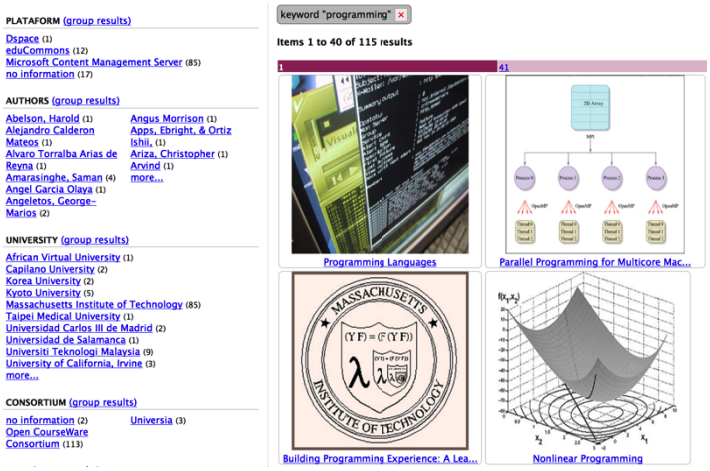

Fig. 1. **Explore OCW in an integrated and incremental manner**, from any of the repositories of institutions that publish OpenCourseWare.

B. *Data and Facets in Serendipity*

In Serendipity, facets correspond to properties of the OpenCourseWare content.

Any of the following cases might prompt to a teacher, student or self-learner to use Serentipity faceted search: (i) Users need to filter content using multiple category or taxonomy terms at the same time. (ii) Users want to combine text searches, category term filtering, and other search criteria. (iii) Self-learners don't know precisely what they can find on OCW site, or what to search for. (iv) Self-learners want to clearly show users what subject areas are the most comprehensive on your site. (v) Self-learners are trying to discover relationships or trends between OCW. (vi) OCW sites has too much content for it to be displayed through fixed navigational structures, but you still want it to be navigable. (vii) Self-learners want to use a faceted classification because a single taxonomic order or a single folksonomy is not suitable or sufficient for OCW content. (viii) Users often get empty

result sets when searching your site. (ix) In cases that "advanced" search forms are not fun to use.

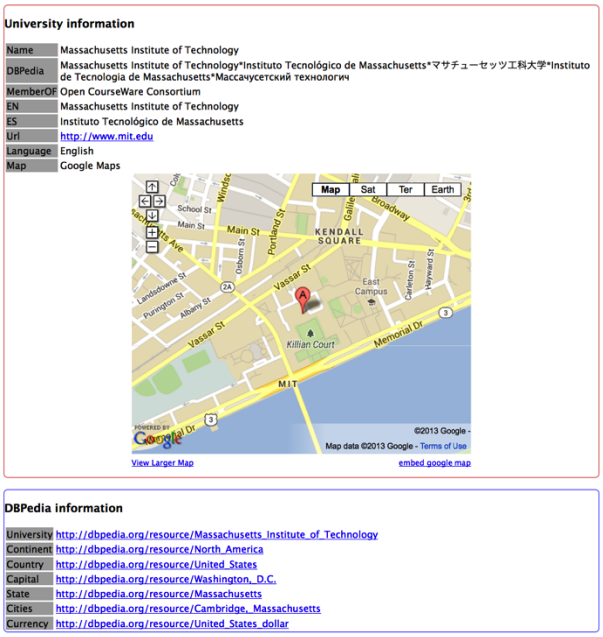

Fig. 2. Link from current OCW in Serendipity To Other LinkedData Source

C. *Navigation of OCW engineering courses*

In the search of OCW courses related to engineering, Serendipity demonstrates the following key features: Grouping search results by facet, displaying a total number of OCW per facet value, refining search results by facet value, update of the facet menu based on refined search criteria, displaying of the search criteria in a Bread Crumbs (navigation guides), ability to exclude the chosen facet from the search criteria, ability to improve ease of discovery open academic resources, ability to improve ease of consumption and reuse of OCW. ability to reduce redundancy in search of OCW, and Connect OCW Data with LOD data. Querying DBPedia, authors obtained additional information about universities such as name in different languages, label, comment, latitude, and longitude. From Geonames were extracted data about locations like continent, country, capital, city or state. To find and create external links, in this work were made directly SPARQL queries (see Fig 2). With Serendipity, we demonstrated that OCW resource data can be enriched using datasets hosted by the LinkedOpenData cloud.

We have verified that the data published in Serendipity is consistent and corresponds to the information obtained in various OCW sites and OCW Consortiums.

IV. CONCLUSIONS

The key idea the OER and OCW movement is that open educational content should be maximally shared. Any open educational data initiative should focus on providing data access permissions so that: provide non-discriminatory access/use/reuse/create derivate works/adapt/share to raw data, information and knowledge about the educational resource.

A lot of open educational resources ends up focussed on usage of open licenses, but not enough attention is paid to structure. Simply putting educational resources online under an open license is obviously not enough. Doing open educational resources data well depends on releasing key datasets in the right way. Moreover, with the proliferation of OCW/OER sites it has become increasingly hard to discovery and track what is happening in Open Educational Context.

In summary, we have shown that our implementation provides better results than other similar systems. Besides, we have also shown that our implementation, which benefits from linked data technologies, can be more appropriate to deal with interoperability and integration of OCW repositories.

ACKNOWLEDGMENT

This work has been partially funded by Madrid Regional Government through the project eMadrid (Ref:S2009/TIC-1650) 'Research and Development for e-learning technologies', Universidad Técnica Particular de Loja and scholarship provided by the "Secretaría Nacional de Educación Superior, Ciencia y Tecnología" of Ecuador (SENESCYT). Finally, authors thank the OCW Consortium, OCW-Universia and technical personal of TAWSBC on UTPL.

REFERENCES

[1] Hilton, J., Wiley, D., Stein, J., & Johnson, A. (2010). The Four R's of Openness and ALMS Analysis: Frameworks for Open Educational Resources. Open Learning: The Journal of Open and Distance Learning, 25(1), 37-44.

[2] Tuomi, I. (2006) Open Educational Resources: What Are They and Why Do They Matter? Report for the OECD Centre for Educational Research and Innovation. 5 October 2006.

[3] Bizer, C. (2009). The Emerging Web of Linked Data. *Intelligent Systems, IEEE*, *24*(5), 87 –92.

[4] Bizer, C., Heath, T., & Berners-Lee, T. (2009). Linked Data - The Story So Far. *International Journal on Semantic Web and Information Systems*, *5*(3), 1–22.

[5] Hausenblas, M. (2009). Exploiting Linked Data to Build Web Applications. *Internet Computing, IEEE*, *13*(4), 68 –73.

[6] Bechhofer, S., Buchan, I., De Roure, D., Missier, P., Ainsworth, J., Bhagat, J., Couch, P., et al. (2012). Why linked data is not enough for scientists. *Future Generation Computer Systems*, in press.

[7] Passant, A., Laublet, P., Breslin, J. G., & Decker, S. (2009). A URI is Worth a Thousand Tags. *International Journal on Semantic Web and Information Systems*, *5*(3), 71–94.

[8] Tovar, E., Piedra, N., Chicaiza, J., Lopez, J., & Martinez-Bonastre, O. (2012). OER Development and Promotion. Outcomes of an International Research Projecton the OpenCourseWare Model. *Journal of Unviersal Computer Science*, *18*(1), 123–141.

[9] Piedra, N., Tovar, E., Colomo-Palacios, R., López, J. & Chicaiza, J. (2013). Consuming and producing linked open data: the case of Opencourseware. Emerald EarlyCite, (PROG-Jul-2012-0045.R2), for publication in 2013.