Serendipity a platform to discover and visualize Open OER Data from OpenCourseWare repositories
Nelson Piedra, Jorge López, Janneth Chicaiza, Universidad Técnica Particular de Loja, Ecuador
nopiedra@utpl.edu.ec, {jachicaiza, jorgaf}@utpl.edu.ec
Edmundo Tovar Caro, Universidad Politécnica de Madrid, España
edmundo.tovar@upm.es

Abstract
Facet browsing has become popular as a user-friendly interface to explore data repositories. In this work, the authors apply the Linked Data Design Issues to explore, visualize and use information that is semantically related to open educational resources that are shared via the OpenCourseWare Consortium and OCW Universia.net, two communities of over 280 member institutions worldwide committed to sharing their courses and OER online. Linked data have the potential of create bridges between OCW data silos. To assess the impact of Linked Data in OpenCourseWare, the authors present a faceted tool called Serendipity and a Data Visualization Web interface called Serendipity Maps. Serendipity is an interface of faceted search for open educational content based on Open Educational Resources Data. Additionally, the linked OER and OCW data environment enabled us to show data visualizations.

Keywords
OpenCourseWare, OCW, Open Educational Resources, OER, Open Data, Linked Data, Faceted Browsing, Data Visualization, Open Educational Resources Data, silos of open educational content, search open material

Introduction
In the last years, the amount of Open Educational Resources (OER) on the Web has increased dramatically, especially thanks to initiatives like OpenCourseWare (OCW) and other OER movements. The potential of this vast amount of resources is enormous but in most cases it is very difficult and cumbersome for users (teachers, students and self-learners) to visualize, explore and use this resources, especially for lay-users without experience with search technologies.

Semantic Web technologies and, more precisely, Linked Data are changing the way information is stored and exploited. The term “Linked Data” refers to a set of best practices for publishing and connecting structured data on the Web [1, 3]. Linked data is mainly about publishing structured data in RDF using URIs rather than focusing on the ontological level or inference. The Linked Data Design Issues, outlined by Tim Berners-Lee back in [2], provide guidelines on how to use standardized Web technologies to set data-level links between data from different sources [3]. These Linked Data Design Issues are: (a) Use URIs as names for things, which can be unambiguously identified (e.g. OERs, courses, Moocs, OER creators, OCW providers, knowledge areas.). (b) Use HTTP URIs so that people can look up those names. With the aid of URIs, the corresponding OER data and relevant interconnected data can be dereferenced. (c) When someone looks up a URI, provide useful information, using the standards (RDF, SPARQL) to describe linked OER data, which are machine-readable and repurposed to serve the proposed architecture to enhance integration with reused and interoperated OER data. (d) Include links to other URIs, so that they can discover more entities. Linked Data—particularly data available using open licenses—has an important role to play in information systems and could be a key feature for Open Education based on OER data on the Web of Data.
In [5], authors apply the Linked Data Design Issues to explore, visualize and use information that is semantically related to open educational resources that are accessible via the OCW Consortium. Linked data have the potential of create bridges between OCW data silos. The authors demonstrate that OCW resource metadata can be enriched using datasets hosted by the LinkedOpenData (LOD) cloud. Additionally, the Linked OER and OCW Data environment enabled us to discovery educational resources, and show data visualizations.

**Linked OCW/OER Data Exploitation Model**

Facet browser interfaces provide a convenient and user-friendly way to navigate through a wide range of Open Educational Resources Data collections. The exploitation model pipeline is organized in four stages that data needs to pass through:

1. **RDF data generation:**
   RDF data generation: the raw data of OCW/OER and LOD Cloud, which can be all kinds of information adhering to the LOCWD data model. In [5], authors argue that the generation and use of linked data from heterogeneous open access initiative sources will create new opportunities for OER and OCW initiatives. The collection under study consisted of approximately 8,000 OCWs and 100,000 OERs in the collection of the OCW-Dataset of LOCWD Project. This collection contained standard OER metadata facets, including creator’s names, language, licenses of OCW, OERs, links to LOD Cloud, tags, knowledge areas, universities, countries and dates. These courses have 6340 creators, including organizations and academic personnel linked to one of the 80 higher education institutions from 28 countries. They describe 657 branches of knowledge (and 27588 user-generated keywords and ten of thousands of tags generated automatically, which are used to link and categorize resources. The OER data obtained from the different OCW repositories were cleaned, disambiguated and formalized for later processing according to Linked Data Design Issues and using LOCWD Vocabulary.

2. **Analytical approach:**
   The applications to visualize, explore and use OCW/OER data are generated depending on the data extraction and visual configuration. Technically, at this stage is important define SPARQL statements (SPARQL\(^1\) is a query language designed specifically to query RDF databases, and used for the queries of logical expressions) to extract the data which contain the information required for each application. SPARQL queries are sent from a client app to a service known as a SPARQL end point\(^2\) using the HTTP protocol. The interaction between the client and the endpoint is defined in a machine-friendly protocol\(^3\) that is not intended to be interpreted by humans, so use of SPARQL requires an interface that allows the human-user to enter the queries and to display the results in a meaningful way.

---

\(^1\) SPARQL 1.1 Query Language [http://www.w3.org/TR/sparql11-query/](http://www.w3.org/TR/sparql11-query/)

\(^2\) Definition of SPARQL endpoint [http://semanticweb.org/wiki/SPARQL_endpoint](http://semanticweb.org/wiki/SPARQL_endpoint)

\(^3\) Communication protocol for SPARQL [http://www.w3.org/TR/sparql11-protocol/](http://www.w3.org/TR/sparql11-protocol/)
Serendipity Faceted Search and Serendipity Maps are interfaces developed so that the queries are constructed and launched through web-forms that do not require the human user to have any knowledge of RDF and SPARQL.

3. **Visual abstraction**
   The third stage takes analytical extractions and transforms them into a visual abstraction of an application. The purpose of this stage is to condense the data into a displayable way for particular application (information that is visualizable on the end-user application). Because the linked data on the Web are accessible via services or public query libraries, OCW/OER data can be processed, reused, combined, integrated, and used for several purposes.

4. **Visual representation: Connect LOCWD dataset with visual tools for Consume and display linked OER/OCW data.**
   This stage processes the visualization abstractions in order to obtain a visual representation of end-user applications. In this stage, the result of the process is presented to the user, e.g. faceted search, data visualizations: plot, treemap, map, timeline, etc. The domain of OER and OCW resources described as Linked OpenCourseWare Data holds the potential to move OER and OCW collections out of their silos and therefore opening the data:
   - To leverage the knowledge capital represented by our OCW repositories
   - To enrich our information landscape, to improve visibility
   - To improve ease of discovery open academic resources
   - To improve ease of consumption and reuse of OER & OCW
   - To reduce redundancy in search of OER & OCW
   - To promote innovation and added value to Open Educational Initiatives.

In this work, authors have implemented two services that support the model showed in previous section: (a) Serendipity a faceted search engine, and (b) Serendipity Maps, a service of OER data visualizations. Applications allow explore and interact with the LOCWD data through different ways. These applications allow users to obtain an overview of RDF datasets from OCW and OER, and realize what the data is about: their main types, relationships, key entities and properties. Services are accessible via the following addresses: [http://serendipity.utpl.edu.ec/](http://serendipity.utpl.edu.ec/) and [http://serendipity.utpl.edu.ec/map](http://serendipity.utpl.edu.ec/map)

**Serendipity a Faceted Search for OpenCourseWare Content**
In this section, we present Serendipity ([http://serendipity.utpl.edu.ec/](http://serendipity.utpl.edu.ec/)), the service that enables the faceted exploration of OER in an OpenCourseWare search engine. The main purpose of the service developed is to provide students, teachers and self-learners with an faceted search engine that allow them to find and discover open educational resources related to OCW from OpenCourseWare Consortium (OCWC is a community of higher education institutions and associated organizations committed with global open education) and OCW-Universia.

Serendipity is a faceted search engine based on Semantic Web Technologies. Moving towards a Web of Linked Data, Serendipity provides a service that enables faceted exploration of large OCW and OER collections. The user, when presented with the facets, is likely to discover new
facets of the query that they were not aware of before. When clicking on a facet, they will narrow down their search by expanding the original query with the suggested facet. In this way, the user get more accurate and complete results, since it locates OERs using different metadata and data elements, providing the user with visible options that help clarify and refine the queries. See Figure 1.

Fig. 1 Exploring OCW in an integrated and incremental way, from any of the repositories of institutions that publish OER and OCWs.

**OCW search is not a solved problem**

Search is among the most disruptive and innovative technology of the Internet. Look for information or an item of interest forms an essential part of the Internet, allowing people to navigate on the Web, find information on a specific topic, or as a starting point for business, learning, social relationship, or entertainment. Faceted browsing for semi-structured and structured sources has been proposed in research papers for some years.

Associate with the increasing volume of text-stored data, the keyword-based query mechanism becomes a natural and effective means for users to interact with Web content. Currently, the OCW search engines can process simple queries over enormous, distributed, heterogeneous, and diverse data sets [4, 5]. However, search is not a solved problem. Most queries entered by users leave room for some sort of ambiguity and will lead to a huge and diverse set of search results. As may be expected, the next challenge is to improve the interoperability and usability of OCW database systems in terms of providing search and query interfaces more than structured query languages can, as well as query by examples, queries based on natural language processing, pattern recognition and semantic queries.

The main objectives of faceted navigation are to support flexible navigation through the information space: refining and expanding, provide suggestions of exploration choices at each point in the search process, prevent empty result sets, and provide a sense of control and reduce
confusion in the use. Other objective of the system is to aggregate the knowledge and high-quality content that is available on and off the OCW-Consortium, and to support the user in their quest for information by identifying the most relevant aspects of a query.

**The Serendipity faceted exploration**
The Serendipity faceted exploration is a guided and exploratory search mechanism, which provides an iterative way to refine search results by a faceted taxonomy or a schema of classification. The user, when presented with the facets, is likely to discover new facets of the query that they were not aware of before. When clicking on a facet, they will narrow down their search by expanding the original query with the suggested facet.

![Fig. 2-a. Accessing to a particular course. The first section contains general data about the selected course.](image)

In Serendipity, facets correspond to properties of the OpenCourseWare content. The data are derived by analysis of the text of an item using entity extraction techniques or from pre-existing fields in a database such as author, descriptor, language, and format. Any of the following cases might prompt to a teacher, student or self-learner to use Serendipity faceted search:

- Users need to filter content using multiple category or taxonomy terms at the same time. Users want to combine text searches, category term filtering, and other search criteria.
- Self-learners don't know precisely what they can find on OCW site, or what to search for. Self-learners are trying to discover relationships or trends between OER & OCW.
OCW providers want to clearly show users what subject areas are the most comprehensive on your site. OCW sites have too much content for it to be displayed through fixed navigational structures, but you still want it to be navigable; or users often get empty result sets when searching your site.

Self-learners want to use a faceted classification because a single taxonomic order or a single folksonomy is not suitable or sufficient for OER & OCW content.

The application of the system is best demonstrated with an example. When a user types in a search for a known entity or knowledge area, such as Web, the system shows the user diverse images of the OCW but also gives the user the option to explore different facets of the entity.

- Figure 1 shows the search engine result page when the user queries for web courses. On the right-hand side the traditional image search results are shown. It can be clearly seen that the image search results shown are diverse, due to the broad nature of the query. On the left-hand side there is the list of facets for OCW and OER resources. In this case, a list of prominent resources in Serendipity is displayed. If the user clicks on one of the facets a new set of OER resources is shown, displaying only open courses that correspond to the clicked facet.

- Figures 2 shows the search engine result page when the user has clicked on the "Ontologies and Semantic Web" course.

Serendipity demonstrates following key features: grouping search results by facet; displaying a total number of OER & 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 OER & OCW, and ability to reduce redundancy in search of OER & OCW.

Fig. 2-c. Accessing to a particular course. The third section provides access to OER from the selected course

**Data Visualization**

Applying information visualization techniques to the OER and OCW data helps users to explore large amounts of data and interact with them. Visualizations are useful for obtaining an overview of the datasets, their main types, properties and the relationships between them. Compared to prior information visualization strategies, we have a unique opportunity on the Open OER Data environment generated by LOCWD and Serendipity. The unified RDF data model being prevalent on the LOCWD enables us to bind data to visualizations in an unforeseen and dynamic way. An information visualization technique requires certain data structures to be present. When we can derive and generate these data structures automatically from reused vocabularies or semantic representations, we are able to realize a largely automatic visualization workflow. This will enable users to explore datasets even if the publisher of the data does not provide any exploration or visualization means.

As an important feature of Serendipity, Serendipity MAPs is a data visualization of OERs that allows users visualize data of OCW / OER / MOOC / OEP / OA Projects / OA Repositories from a dataset based on Linked Data technologies (e.g. Figure 3 show Points of Interest, POIs).
Figure 3. OER Data Visualization: Map of Open Educational Repositories distributed in the world and described on LOCWD

Serendipity Maps use icons to represent different categories of data on a map graphically. For example, a point of interest, or POI\(^4\), is an OER specific point location that someone may find useful or interesting. A point of interest specifies, at minimum, the latitude and longitude of the POI, assuming a certain map datum (extracted from dbpedia datasource). An example is a point on the Earth representing the location of the Massachusetts Institute of Technology, or a point on Spain representing the location of an OCW University (see Figure 4).

To conclude this section, we present other data visualizations approach. Figure 5 shows the results of search courses by tag and use geographic information to show courses of universities and social network analysis (SNA) to form networks of collaboration and recommend related tags. The keyword-based query formulation mechanism has to a great extent contributed to the wide acceptance of the Internet. This simple mechanism offered by Web search engines allows users to query collections of OER & OCW information freely and to quickly decide in a particular direction of interest.

\(^4\) A description for the POI is usually included, and other information such as description, number of resources, contact information, language, license or a link to dbpedia/freebase may also be attached.
Acknowledgments
We like to thank our UTPL colleagues, students and members of the Advanced Web Technologies & Knowledge-based Systems Team for their technical support and assistance to solve different problems during development of Serendipity: René Elizalde, Jhonny Zaruma, Andrea Novillo, María del Carmen Cabrera in data extraction process; Manuel Valarezo in development of serendipity prototype based on flamenco (serendipity alpha version); Eduardo Encalada supporting technological infrastructure; Ramiro Ramírez, Rodrigo Saraguro and Marlon Viñan in Data Viz and Web development; Elizabeth Cadme, Charbel Torres and Diana Torres in evaluation of web accessibility. This work was supported by UTPL through internal funding research projects and scholarships awarded by the government of Ecuador through Senescyt. Authors also thank the OCWC and OCW Universia.net for their support.

Conclusions
For effective use, reuse, discovery, distribution and sharing of OER and OCW repositories must provide a declarative query interface that supports complex expressive Web queries. In this paper, authors developed Serendipity to experiment with facet browsing of Open OER Data in a highly heterogeneous and distributed Open Educational semantic web environment. Faceted search and data visualizations are important aspects to visualize, explore, and use Open Educational Resources Data, especially for lay-users without experience with advanced search technologies. The power of Serendipity lies in the ease of combining multiple facets to construct a complex and personalized query. Combining constraints from multiple facets, a user (teacher, student or self-learner) is able to specify relatively complex queries through an intuitive Web interface. Compared to prior data visualization strategies, we have a unique opportunity on the Open OER Data environment generated by LOCWD and Open Data Visualizations technologies. The unified RDF data model being prevalent on the LOCWD enables us to bind data to visualizations in an unforeseen and dynamic way. 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. Linked Data vision enables a new generation of OERs and OCW that can be semantically described, integrated and connected with other data and discoverable sources.
References

License and Citation
This work is licensed under the Creative Commons Attribution License http://creativecommons.org/licenses/by/3.0/. Please cite this work as: Piedra, N. et al. (2014). Serendipity a platform to discover and visualize Open OER Data from OpenCourseWare repositories. In Proceedings of OpenCourseWare Consortium Global 2014: Open Education for a Multicultural World.