Editorial

Salvador Sanchez-Alonso and Enayat Rajabi, Guest editors

Welcome to the Bulletin of the IEEE Technical Committee on Learning Technology, Volume 16, Number 4, December 2014 issue. This special issue discusses how the Linked Data approach can be applied for sharing, reusing and enriching the eLearning resources on the Web and what advantages it can bring to the eLearning stakeholders.

The first paper in this issue presents a solution to one of the main problems in Open Educational Resources repositories, which is the multiplicity of norms, standards and application profiles that preclude efficient search for resources within multiple repositories. The authors built a Linked data OER repository manager called COMETE, relying on semantic web techniques, and largely compatible with the new ISO-MLR standard. The new platform was based on RDF and provides more intelligent search capabilities within the Web of data both for designers who are building online environments such as MOOCs, or for students who should be equipped with friendly tools to choose resources, activities and co-learners suited to their needs.

In the second paper, the authors compare the performance of two distinct RDF native database implementations (4store and Jena Apache) in the specific context of a Semantic Learning Repository by evaluating the time of uploading the RDF data to the databases, and the response time for running the SPARQL queries. The results show that 4store performed better than Apache Jena for all the evaluated scenarios.

The third paper reports on the reflection of learning activities and revealing hidden information based on tracking user behaviors with Linked Data. The authors present a case study on usage of semantic context modeling and creation of Linked Data from logs in educational systems like a Personal Learning Environment (PLE) with focus on reflection and prediction of trends in such systems. Their case study demonstrates the application of semantic modeling of the activity context, from data collected for over two years from a widget-based system at the Graz University of Technology. The authors modeled the learning activities using domain ontologies, and query them using semantic technologies. The proposed approach offers easy interfacing and extensibility on the technological level as well as a fast insight on trends in e-learning systems.

In the forth paper, Carpani and his team present a Semantic Web based collaborative platform, so-called OpenFING, for video annotations. Using this platform, students and teachers can easily create video fragments, add annotations, as the the tool provides searching mechanism over structured metadata.

The last paper shows a novel approach to enable automatically matching MOOCs learning outcomes with learners' needs according to specific skills and competencies. The authors exploited novel developments published by the European Commission (the ESCO taxonomy) and developed a pilot application where candidates can create their professional profile by submitting their skills and enrich their profile by receiving supplementary content from the web of data. This pilot application demonstrates the feasibility and opportunities that derive from linking concepts from the ESCO taxonomy with skills on candidate profiles and learning outcomes of open digital resources. More specifically, the use of such an application can significantly benefit formal, informal and lifelong learners in developing appropriate competences that will increase their qualifications.

We sincerely hope that the issue helps in keeping you abreast of the current research and developments in Learning Technology. We also would like to take the opportunity to invite you to contribute your own work to the Bulletin of the IEEE Technical Committee on Learning Technology if you are involved in research and/or in the implementation of any aspect of advanced learning technology. For more details, please refer to the author guidelines at http://www.ieeetclt.org/content/authors-guidelines .

Deadline for submission of next issue:
30 January 2015

COMETE – An Educational Search Engine on the Web of Linked Data

Gilbert Paquette, Alexis Miara and Frédéric Bergeron
LICEF Research Center, Télé-université, Montreal, Canada
gilbert.paquette@licef.ca; alexis.miara@licef.ca; frederic.bergeron@licef.ca

***Abstract*— We present here a new Open Educational Resource (OER) repository management tool called COMETE. The evolution of research and practice in the field of OER repositories is moving from OER metadata stored in relational databases to RDF-based descriptions of resources stored in triple stores. COMETE provide the necessary translation from DC, LOM and other metadata schema to an RDF meta-model that offers more intelligent search capabilities, both for designers who are building online environments such as MOOCs, or for students who should be equipped with friendly tools to select resources, activities and co-learners suited to their needs.**

***Index Terms*—open educational resources (OER); resource description framework (RDF); resource repositories; semantic web; web of data.**

I. Introduction

In the last three years, our research on educational resource management moved from Learning object metadata repositories managed by the PALOMA tool [1], to the use of ontology-based annotations within the TELOS system [2] and, finally to the use of semantic technologies for the Web of data [3,4]. The result in a mature tool, COMETE, that is being used in the colleges of Quebec for educational resource referencing and search. The present paper describes the principles and the architecture of the COMETE system and its relation to the ISO-MLR standard [5].

The paper is organized into four sections. In section II, we introduce the notion of an educational resource repository and of a resource manager. Section III presents a recent evolution of e-learning standards, norms and application profiles resulting in the publication of the ISO standard on Metadata for Learning Resources (ISO-MLR) based on the RDF and the Web of data. In the fourth section we provide the principles and the architecture of COMETE, our RDF-based OER manager and we illustrate various kinds of search methods available in COMÈTE.

II. Open educational resources repositories

The term "Open educational resources" was first coined at UNESCO's 2002 Forum on Open Courseware and defined as "teaching, learning and research materials in any medium, digital or otherwise, that reside in the public domain or have been released under an open license that permits no-cost access, use, adaptation and redistribution by others with no or limited restrictions" Ten year later, UNESCO held in Paris an international OER congress on 20-22 June 2012 where the Paris OER Declaration was issued, recommending that States, within their capacities and authority "foster awareness and use of OER", "encourage research on OER", "promote the understanding and use of open licensing frameworks" and "facilitate finding, retrieving and sharing of OER."

A. First interoperability norms: DC and IEEE-LOM

The idea that educational contents could be seen as "objects" to be reused in multiple contexts dates back to the late 60's but started to become a reality only by the middle of the 90s with the generalization of the Internet [6]. The aim was to insure the reuse of educational objects jeopardized by the diversity of referencing metadata schema around the world.

In 1995, the Dublin Core (DC) metadata initiative proposed a first set of standardized metadata, expressed in XML. Since then, the Dublin Core metadata schema has that was to become one of the most used vocabularies on the Web of Data. In 1996, the IEEE created the Learning Technology Standards Committee to integrate previous work on the concept of Learning Object Metadata (LOM). From then on, major resource repository initiatives bloomed rapidly: ARIADNE in Europe, MERLOT in the USA, EdNA in Australia. These and many other organizations, including ours, joined the GLOBE consortium that operates actually a large repository of nearly one million resources.

B. Potential and Limits of DC/LOM Resource Repositories

Motivations for OER repositories are the growing educational demands in all countries, the limited capacity of face to face education to fulfill the demand in a timely manner, the important effort and cost involved to build online multimedia learning materials and the new possibilities offered by the Internet.

While it is a fact that millions of documents can be found on the Internet using search engines like Google, there is no guarantee that a query will lead to trustable material on which high quality education can be built. On the contrary, OER repositories are maintained by educational institutions and professors providing a certain level of trust for the quality of referenced resources and giving precious information to users, that helps make more focused queries. These query prevent blind search based on vague keywords that leads to thousands of references that one needs to read to understand what kind of content they provide. Finally, the vast majority of these learning objects are in the public domain to be reused free of charge. These resources can be adapted or aggregated, and referenced back in a repository to extend the availability of good learning material.

After a decade of research and practice in this field a number of limitations to a larger use of OER repositories still exist. The most important ones are the multiplicity of norms and Applications profile, as well as the imprecision of most proprietary metadata schema.

III. ISO-MLR AND THE WEB OF DATA

Although the Dublin Core and the IEEE-LOM are widely used to describe learning resources, interoperability among metadata sets from multiple repositories is still challenging.

For example, instead of using ISO 8601, a DC Date element can be written in plain language making it impossible its processing by queries. Ambiguous definitions pose another challenge. For example a Date element can represent a resource creation time, a time of update or a time of publication. As mentioned above, LOM records can be based on a wide variety of Application profiles each defined in their own way by various organizations.

A. *ISO-MLR: an OER referencing standard based on RDF*

The ISO/IEC 19788 standard [5], in short ISO-MLR, is intended to provide optimal compatibility with both DC and the LOM. It insures the coherence and the non-duplication of concepts by proposing an RDF-based data model. It prevents the proliferation of non-interoperable application profiles. It supports the extension of description vocabulaires in precise ways while preserving interoperability as well as multilingual and cultural adaptability requirements from a global perspective, while integrating ressource referencing and search with other data sets in the Web of linked data.

The graph in Fig.1 shows part of the ISO-MLR RDF model. The ovals represent classes of resources, the rectangles are value types, and properties are written over the links. This graph summarizes the RDF triples in section 5 of the standard. Here are some of the triples present on Fig.1:

(Learning resource, has learning activity, Learning activity)
(Learning activity, learning method, *method value*)
(Learning resource, has contribution, Contribution)
(Contribution, has contributor, Person)
(Annotation, annotation date, *date value*) ….

Links like the second and last one represent object properties linking various kind of resources, while the others are data properties specifying values of properties.

B. *A standard for the Web of data*

ISO-MLR, uses technologies like RDF and RDF schema, integrates well to a Web of linked data. The origin of the Web of data, also termed "Semantic Web", dates back to 2001 when the actual director and founder of the Web, Tim Berners-Lee and his colleagues proposed to extend the Web to use URI's to represent not only pages of information but also people, real-world objects and also abstract concept and properties. These entities and the values of their properties can be linked together by declaring RDF triples.

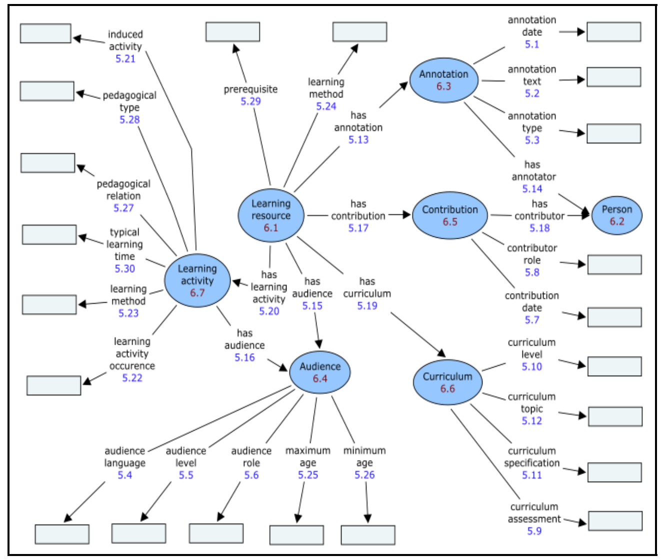

Fig. 1. Part of the ISO-MLR RDF model

It then becomes possible to describe the semantic of Web pages beyond the syntax of natural languages and their inherent ambiguity. A Web of linked data enables computer agents to follow the links and perform more intelligent operations using the knowledge behind the words.

For example, the SPARQL Query Language [7] enables queries within the huge graph of RDF triples that constitutes the Web of linked data. At the end of 2014, this graph grouped 74 billions RDF triples from 353 datasets. And it is still growing. Within this graph, the DBpedia dataset contains a large part of the information in Wikipedia, while the FOAF dataset provides information about persons having a URI on the Web. Terms in a vocabulary (concepts and properties) are linked with terms in another vocabulary. For example, "persons" in DBpedia is related to "persons" in FOAF and their geographical localization can be found in another vocabulary or dataset such as GEONAMES.

In the same way, terms in ISO-MLR are linked to terms of other vocabularies on the Web of data. For example, iso-mlr5:Person in the graph of Fig.1 has the same meaning as foaf:person or dcterms:person. This means that a computer agent that would search for an iso-mlr:learning_resource can also ask for its iso-mlr:Contributors, find these persons and retrieve their Wikipedia pages from DBpedia, their email from FOAF and their localization from GEOBASE.

IV. COMETE, A RDF-BASED RESOURCE MANAGER

COMETE is a learning resource repository manager based on the RDF approach. It allows locating, aggregating and retrieving educational resources that constitute the heritage of an organization. Basically, it is a database containing metadata about learning resources on which users can perform queries to find and discover educational material that they can reuse for their various needs.

Fig. 2 describes the technical architecture of a COMETE implantation instance. It's a 3-tiers client-server architecture developed in Java. Various web applications powered by an Apache Tomcat server provide specialized REST services that allow different types of clients to exploit the open data contained in the repository. Most of the clients use their favourite web browsers to access the system through a user-friendly web interface. A SPARQL endpoint is also available for advanced users who want to directly access the raw RDF triples to build various applications or Web services.

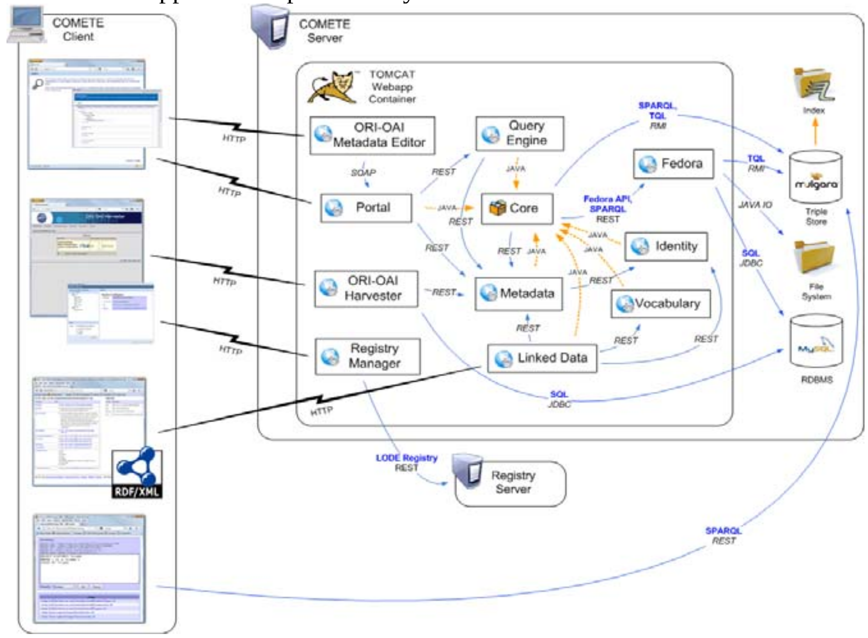

Fig. 2. COMETE architecture schema

A. *Integrating new resources in the triple store*

The integration of resources inside a COMETE repository is done by imports of their metadata records. The metadata records can be imported manually by uploading an archive file containing a collection of metadata records. Most of the time, however, either an OAI-PMH Harvester or a HTML Spider will harvest the metadata records automatically. In such a case, a Harvest Definition will declare the technical information required to access the repository to be harvested. It is also possible to program harvest schedules so that the process is executed periodically to make sure that new or updated metadata records are always imported to the system.

These records are ingested by the system and a XSL transformation extract data for generating all pertinent triples. COMETE enables data mining across multiple metadata schemas like Dublin Core, IEEE LOM and other application profiles. The result of this process is a homogeneous graph of data in accordance with COMETE's internal meta-model (partly shown on Fig. 3).

All the triples that are generated are stored into Mulgara [8], an open source RDF triple store system, where data is organized around various RDF graphs. The default graph contains all the triples about learning resources whereas some other specialized graphs manage SKOS thesaurus and other different views of the system.

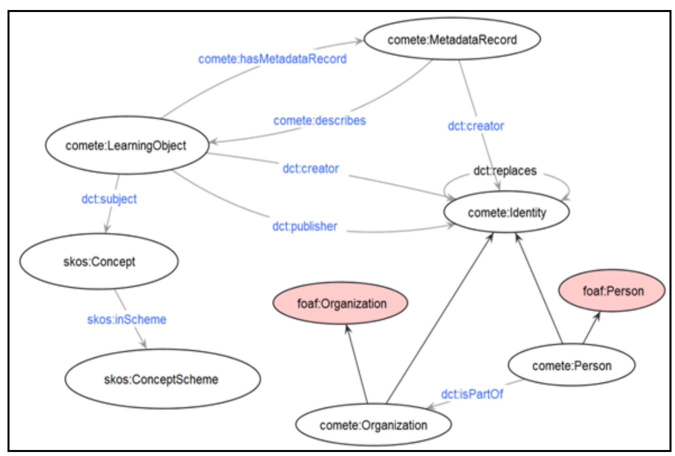

Fig. 3. COMETE metamodel (main classes)

As a semantic network, the RDF graph represents the model entities as nodes. Mains nodes are learning resources (Learning Object), persons and organizations (Identity) and element of vocabulary (SKOS Concept).

The Identity module showed on Fig. 2 implements the management of metadata about persons or organizations. This includes importation of identities, identity resolution of data that represents the same person or organization, making sure it stays unique, and completing it as new details are known. Furthermore, manual merge of identities is also provided within a set of administrative tools for a better control of data integrity.

The Vocabulary module (on Fig. 2) implements the management of vocabularies and thesauri imported from VDEX or SKOS formats, unambiguously identifying the vocabulary that a term is from, converting from one format to another, replacing a vocabulary when updates are available, publishing vocabularies automatically and providing user interface elements reusable by other modules, such as efficient vocabulary term choosers for queries to the repository.

This module manages also SKOS concept alignment between different ontologies (or vocabularies) that is taken into account by the query engine. For example the mapping between different school-level taxonomies in different countries enables one to search resources for Junior High School in United States, and the results may contain pertinent Secondary School I-III tutorials produced in Québec.

B. *Querying the triple store*

All of the previously presented modules provide rich graphs of data that allow processing more "intelligent" searches in the repository than before.

COMETE provides four search modes. The simple mode only needs a field of keywords, as in a Google search. The CERES implementation of COMETE (Fig. 4), now in use in Colleges of Québec grouping over 40 000 resources, will serve to illustrate the other modes.

Fig. 4. COMETE Advanced Search Interface

Queries like the one of Fig.4 are translated in SPARQL language by the Query Engine module (shown on Fig.2) and then run on the triple store to extract the information. Suppose now we seek the resources authored by a contributor dealing with the subject of change management ("gestion du changement" in French). Fig 4 illustrates how to fill such a query. The user could add other properties to refine the search. The right part of the figure shows the properties of a selected resource. Links to the author's and his organization's information, and also taxonomies of concepts are clickable to navigate on the global RDF graph.

A third way to search resources inside a repository is to use the Thematic Navigation mode that makes use of the Linked data module (also shown on Fig.2). It lets user discover resources directly from the categories in the available vocabularies. Results are returned and displayed to the user interface using the alignment of vocabularies integrated in the system; for example, queries may be extended with "include equivalent categories from the Library of Congress".

Finally, a fourth search mode in COMETE user interface is the Collection mode. It offers to users a list of preset complex queries to avoid having to enter them by hand. For example, one could ask the system: "give me all the resources on Algebra and non-Euclidean geometry produced last month from authors at the Université de Montréal".

C. *Linking with the Web of data*

The link with the web of data, as a global data space, is achieved by respecting the basic principles of Linked Data: representing all entities by HTTP URIs, dereferencing URIs over the HTTP protocol into a description of the identified object or concept served in different versions: HTML page for web browser clients, RDF/XML for software agent. A COMETE vocabulary details the class and property definitions in the meta-model, reusing existing vocabularies such as Dublin Core, FOAF and SKOS. The publishing of data via a SPARQL endpoint allows interaction with COMETE data by external systems.

V. Conclusion

We have presented a solution to one of the main problems in Open Educational Resources repositories, which is the multiplicity of norms, standards and application profiles that preclude efficient search for resources within multiple repositories. We have built a Linked data OER repository manager, COMETE, relying on semantic web techniques, largely compatible with the new ISO-MLR standard.

References

[1] Paquette G, Lundgren-Cayrol K, Miara A, Guérette L (2004) The Explor@2 Learning Object Manager, in R. McGreal (ed), Online education using learning objects. pp 254-268. London: Routledge/Falmer.

[2] Paquette, G. (2010). An ontology-driven System for e-learning and knowledge Management. In Paquette, G., Visual Knowledge Modeling for Semantic Web Technologies: Models and Ontologies. Hershey, PA: IGI Global, pp 302-324

[3] Allemang D. and Hendler J. (2011) Semantic Web for the Working Ontologist – Effective Modeling in RDFS and OWL. 2nd Edition. Morgan-Kaufmann/Elsevier, Amsterdam.

[4] Heath, T. et Bizer, C. (2011). Linked data: Evolving the web into a global data space. In Synthesis Lectures on the Semantic Web: Theory and Technology, 1(1), 1-136.

[5] ISO-MLR (2013) ISO-IED 19788 Information technoogy – Learning, education and training – Metatda for learning resources multipart standard. http://en.wikipedia.org/wiki/ISO/IEC_19788.

[6] Duval, E. and Robson, R. (2001) Duval, E. and Robson. R. Guest Editorial on Metadata. Interactive Learning Environments, Special issue: Metadata, Volume 9-3, December 2001, pp. 201-206

[7] SPARQL – SPARQL 1.1 Query Language, W3C Recommendation, 21 March 2013. http://www.w3.org/TR/sparql11-query/

[8] Mulgara RDF Triple Store System. http://en.wikipedia.org/wiki/Mulgara_(software)

Towards the use of Semantic Learning Object Repositories: Evaluating Queries Performance in two Different RDF Implementations

Henrique L. dos Santos, Gladys Carrillo, Cristian Cechinel and Xavier Ochoa

***Abstract*— Learning Object Repositories (LOR) are an essential component of the e-Learning ecosystem and have being normally serving the purpose of cataloging, storing, retrieving and delivering Learning Objects to be used inside e-Learning applications. Next generation of LORs needs to overcome some major shortcomings current LORs present and involve other entities that are part of the e-learning process (teachers, students, lessons, courses, activities, learning paths, etc.) in a way that they are all fully integrated and linked-up. Semantic web technologies such as RDF are the natural choice to implement these requirements into Semantic Learning Repositories. One important factor limiting the implementation of this kind of systems is the uncertainty about their performance. The present paper describes an initial study that compares the performance of two distinct RDF native database implementations (4store and Jena Apache) in the specific context of a Semantic Learning Repository. The performance tests were run to evaluate two different aspects of the databases implementation: the time to upload the RDF data to the databases, and the response time for running the queries. The results showed that 4store performed better than Apache Jena for all the scenarios we evaluated.**

***Index Terms*— performance analysis, RDF database, Semantic learning object repositories**

I. INTRODUCTION

LEARNING Object Repositories (LOR) have been the backbone for the construction of e-learning systems that provide access to a large amount of learning resources. Traditionally, these LORs have been implemented as document repositories, that is, they are centered only on one entity, in this case, the Learning Object. The information stored in a traditional LOR is the learning resource file and the metadata, in a predefined format, describing that resource. In the case of the learning resource file, some LORs store only a reference to where the file is stored and these LORs are called "Referatories". The traditional design of LOR while useful for the direct retrieval of Learning Objects, present several shortcomings when used in real-life e-learning systems. First, e-learning systems should manage much more diverse entities than just the learning objects. The learner, the teacher, the lesson (sequence) should also need to be taken into account. E-learning systems usually solve this shortcoming having several repositories for different type of entities: one for the learning objects, other of the user profile and another for the lesson structure. While this let the e-learning system to store all the needed information, it adds complexity to the system and makes very difficult to maintain the very necessary relationships (links) between entities [1]. A second major shortcoming of traditional LORs is their reliance on a single metadata format to describe the learning resources. In the best-case scenario this format will be a standard such as Learning Object Metadata (LOM) or Dublin Core (DC), otherwise, it will be an ad-hoc structure. Due to this reliance on a single metadata format, a whole area of research on Learning Object Interoperability has been developed in order to be able to interchange information between several repositories [2]. These interoperability issues, again, add complexity to the design of e-learning system, especially if it is desired that their data remain open for others to be used.

Finally, being based on predefined formats for their metadata, traditional LORs are designed to operate with a static structure. If new elements or entities are added to the e-learning system, the LOR will be unable to accommodate them and a new repository, or a major re-design, will be needed to store their information. Rapid changing and adaptable e-learning systems could only communicate with LORs as a source of information, but not as an integral part of the architecture of the system [3]. All these shortcomings demand a drastic redesign of the concept of Learning Object Repository to be the main persistence component of modern e-learning systems.

The concept of Semantic Learning Repositories solves the

This work was supported by CYTED (Ibero-American Programme for Science, Technology and Development) as part of project "RIURE - Ibero-American Network for the Usability of Learning Repositories", code 513RT0471 (www.riure.net) and by FAPERGS (Research Support Foundation of Rio Grande do Sul – Brazil) through Edital 02/2014 - PROGRAMA PESQUISADOR GAÚCHO – PqG.

H. L. dos Santos is with the Computer Engineering Course, Federal University of Pelotas, Brazil (e-mail: henriquelds94@gmail.com)

C. Cechinel is with the Faculty of Education, Federal University of Pelotas, CEP 96010-020 Brazil (phone: +55-53-39211431; e-mail: contato@cristiancechinel.pro.br).

G. Carrillo is with Faculty of Electrical and Computer Engineering at the Escuela Superior Politécnica del Litoral (ESPOL), Ecuador (e-mail: gladys.carrillo@cti.espol.edu.ec)

X. Ochoa is with Faculty of Electrical and Computer Engineering at the Escuela Superior Politécnica del Litoral (ESPOL), Ecuador (e-mail: xavier@cti.espol.edu.ec)

LOR shortcomings mentioned here. First, if all needed information can be stored in a single repository, there is no need for the e-learning system to include or connect with other types of repositories. Different types of e-learning systems could include different description for the entities and even different entities depending on the learning process they are supporting. Second, the use of Semantic Technologies leads to a format-free repository. Any metadata standard could be used to describe the existing entities. Mapping between metadata standards or ad-hoc structures is greatly facilitated by the use of RDF triplets to store information. The interoperability issues are also reduced if the data is published as Open Linked Data [4]. In this way, it can be easily consumed by any other Semantic Learning Repository or e-learning application. Finally, changes in the metadata formats and/or stored entities can be easily incorporated into the Semantic Learning Repository without need to change its functionality.

Given the internal structure of the data and the flexible nature of RDF implementation, the natural choice for the implementation of a Semantic Learning Repository is to use an RDF store. However, the perceived difference in performance between traditional Relational Database Management Systems (RDBMS) and RDF stores has been one of the main reasons why the current LORs are still implemented over RDBMS systems [5]. This perception has been formed in the early days of RDF stores, with current systems promising improved performance. However, the perception persists. Moreover, to the authors knowledge, there are no studies focused on evaluating their performance in the context of e-learning systems. The present study stress tests the performance of two most successful RDF native database implementations that are openly available (4store and Jena Apache) in the specific context of a Semantic Learning Repository. It is known that there are already benchmarks proposed for the comparison of performance among different RDF database implementations such as Berlin SPARQL Benchmark (BSBM) [6] and Lehigh University Benchmark [7], as well as other studies involving performance experiments on RDF databases [8] [9], however, as stated by [10], there is a need for testing and comparing performances for each type of application in specifics architectures, contexts and scenarios, in this case in the common queries produced by an e-learning solution.

The reminder of this paper is organized as follows. Section 2 describes the materials and the methodology of the present study, and section 3 presents the results of the tests we performed. The final remarks are presented in section 4.

II. Materials and Methods

In order to determine which RDF database performs better, a set of queries specific to the learning context was created, executed and tested on identical conditions for both implementations: 4store and Jena Apache. The tests were performed for the data model presented in Fig. 1. This data model was implemented for the APRENDE Tutoring System.

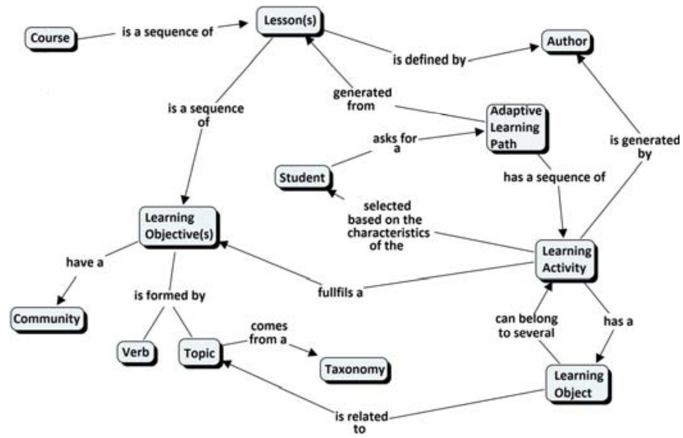

Fig. 1. Data Model

In APRENDE[1] teachers can upload learning materials (also denominated as learning activities) and associated them to learning objectives. Moreover, lessons (sequences of learning objectives) and courses (sequences of lessons) can be created and offered to the students. As the students navigate through the available lessons and courses, the system learns about their profile and recommends personalized learning materials to those lessons they are studying. A more detailed description of the APRENDE can be seen in [11]. The dataset contains a total of 1,262,954 triples, distributed over 16 graphs that were used for both evaluations. The proportional distribution of triples per graphs is shown in Table I. In order to avoid (reduce) the influence of network latency, both RDF implementations and the repository's website were running in the same machine. The hardware and software configuration are shown in Table II.

TABLE I
PERCENTAGES AND NUMBER OF TRIPLES PER GRAPHS

Graph	Percentage of total triples (%)	Number of triples
G1. Learning Activity Weight per User	80,55	1.017.339
G2. Learning Path Activity	15,29	193.140
G3. Learning Path	1,32	16.731
G4. Adaptation	0,71	9.011
G5. User	0,65	8.170
G6. Learning Style - User Profile	0,57	7.177
G7. Learning Activity	0,56	7.126
G8. Objective	0,15	1.869
G9. Learning Activity Objectives	0,06	750
G10. Lesson Objectives	0,05	663
G11. Lesson	0,04	545
G12. Course Lessons	0,02	231
G13. Course	0,008	101
G14. Taxonomy	0,005	60
G15. Sequence	0	6

TABLE II
HARDWARE AND SOFTWARE SETUP CONFIGURATION

Type	Component	Hardware/Software in use
Hardware	Processor	Intel(R) Core(TM) i5-2430M 2.40GHz 64bits (2 cores, 4 threads).
	Memory	8GB.
	Hard disk	250GB

[1] http://aprende.igualproject.org

Software	RDF Database	Apache Jena-Fuseki 1.1.1 with built-in TDB and Garlik 4store version 1.1.5.
	Operating System	Linux Fedora 19 64 bits.
	Filesystem	ext4.
	Java Version and JVM	Version 1.7.0_65, OpenJDK 64-Bit Server VM (build 24.65-b04)

The study focused on measuring two different aspects: 1) the resulting time to upload the triples in the databases; and 2) the resulting time to execute queries.

A. *Implementation of the Databases*

The implementation of the databases and the setup of the whole evaluation system consisted on the following steps: 1) Exporting RDF data that were already created to Turtle format[2]; 2) Initializing and executing each database; 3) Uploading the RDF data to the database currently under test; 4) Integrating the system website with the databases in order to allow RDF querying; and 5) Running the queries. The study focused on measuring the performance of steps 3 and 5 (upload and response time).

B. *SPARQL queries*

As the repository website was in PHP, we used Zend framework[3] to integrate it with the database in order to allow RDF querying. The Zend framework provides a HTTP Client class that allows one to send HTTP requests to an endpoint. This approach is possible because both 4store and Apache Jena provide RESTful API to the developer. In other words, there is a particular endpoint for the dataset and for each operation to be done on it (query, update, etc). Classic HTTP requests (POST, GET, PUT, HEAD, etc) can be sent to these endpoints to perform queries and updates. For the present study we tested queries and updates by using the POST method.

Apache Jena provides two alternatives for the upload of the datasets. The first one is the upload via a browser's interface at the regular address http://localhost:3030 and using a POST method to add triples in a previously created named graph. The second one is SOH (SPARQL Over HTTP), which is a set of Ruby scripts provided together with the Fuseki's distribution. In this case, the s-put script uses a PUT method to add triples in a named graph that may or not already exist. Both alternatives provided by Apache Jena were evaluated here.

The set of queries for evaluation was designed to be similar to the set that is executed when a given user access the repository. A total of 27 queries were evaluated. The queries can be divided in 4 major groups that are presented in Table IV. For the present study, the queries were executed in the same order that they are presented in Table III. This chronological sequence expressed by the execution of M1 + M2 + M3 + M4 is denominated here as Session and implies in a natural and realistic caching improvement for both databases. In other words, the Session is composed by a query mix that is naturally executed when a given user navigates on the system (the sequence of logging into the system, accessing a course, continuing a course, and completing a course). In our tests, all queries obey this chronological sequence represented by the Session. Differently from a benchmark work where a query is uninterruptedly executed numerous times, here all executions are performed in the sequence provided by the Session. Queries are complicated from M1 to M4. For instance, in M1, queries search results by using one or more graphs to authenticate users and show him their lessons and courses, whereas in M3, searches and new data insertions are executed over only one graph at a time.

During the tests, the Session was executed two times for each RDF implementation. For each Session execution, some queries of M1 and M2 groups were executed two times, in order to simulate a natural behavior of a user that visualizes a course and returns to the homepage twice. The average time of these two Session executions is then computed as the resulting time response for that given RDF implementation.

TABLE III
QUERIES

Group Mix	Queries Description	Graphs involved
M1 - Access to the site: logon process e homepage loading.	Use of FILTER, REGEX, ORDER BY, LIMIT, queries with and without FROM clause, queries with single and multiple variables, one graph search and multiple graphs search.	G1, G3, G5, G9, G10, G12, G13
M2 - Access to a course: course's content visualization, the start of a new course and creation of a new learning path.	Use of DISTINCT, ORDER BY, GROUP BY, COUNT, queries with and without FROM clause, one graph search, multiple graphs search and INSERT DATA queries	G1,G2, G3, G8, G9, G10, G12, G13, G15
M3 - Continuation of a course: learning path's information, selection and information of the next available activity, information of the current objective and activity, information of a lesson and a course's objective and completion of an activity	Use of FILTER, REGEX, LIMIT, ORDER BY, queries with single and multiple variables, one graph search, WITH DELETE INSERT queries.	G2, G3, G7, G8, G11 G13
M4 - Completion of a course: completion of the learning path.	WITH DELETE INSERT queries	G3

III. RESULTS

As mentioned earlier, the present study focused on the time response for uploading the RDF data to the databases, and for running the queries.

A. *Uploading the RDF*

Figure 2 shows the corresponding time in seconds to upload triples in the graphs in each database. In all cases 4store has better uploading time compared to the other two alternatives of Jena, but a significant difference is observed

[2] http://www.w3.org/TR/turtle/
[3] http://framework.zend.com/

when loading data in graphs with the largest number of triples (> 190000). For the graph G1 (1,017,339 triples), 4store took 8.90s, whereas Apache Jena with POST took 61.93s and Apache Jena with PUT took 245.13s.

B. Running queries

Table IV shows the total time for each mix of the Session in both implementations. As can be seen in Table IV, the time response of 4store is better than Apache Jena for all the scenarios. For instance, for the first and second mixes of the Session (M1 and M2), 4store responds 3.7 and 4.1 times faster than Apache Jena. When the user is in the middle of the course (M3), the performance of 4store is 5.7 times better than the Apache Jena. The only case where the time response for the queries in both implementations is similar is for M4 (WITH DELETE INSERT queries).

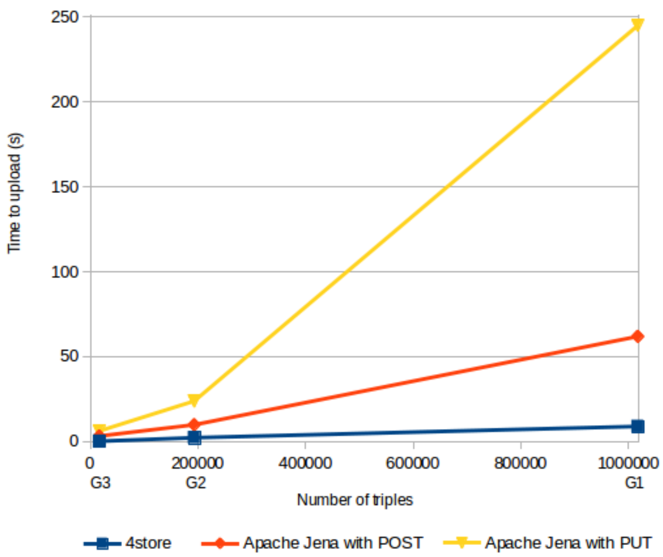

Fig. 2. Upload performance of the databases

TABLE IV
TIME RESPONSE FOR THE QUERIES

Queries Groups	Total types of queries	Total quantity of queries	Total execution time in 4store (ms)	Total execution time in Apache Jena (ms)
M1	5	16	178.66	660.20
M2	7	23	3,273.86	13,409.92
M3	10	11	253.53	1,444.36
M4	1	2	231.72	306.25

IV. FINAL REMARKS

Current Learning Object Repositories (LOR) have serious shortcomings that reduce their usefulness to implement e-learning systems that go beyond retrieving Learning Objects. The whole e-learning ecosystem demands a drastic change in the way we see, use and therefore, implement such repositories. The next LOR generation needs to consider and represent all the other entities that participate in the e-learning process (teachers, students, lessons, courses, activities, learning paths, etc.) in a way that they are all fully integrated and linked-up. We propose the use of Semantic Learning Repositories in order to provide the functionality needed by modern e-learning solutions.

This study evaluated the performance of two different RDF implementations (4store and Jena Apache) in the specific context of a Semantic Learning Repository called APRENDE. The results showed that 4store performed better than Apache Jena for all the scenarios we evaluated. More importantly, the times found for both solutions, especially for the one implemented with 4store are enough to implement an on-line system that could provide the required services to e-learning systems. Although this work is limited in scope, it is a first empirical proof that RDF stores could be successfully used to select Semantic Learning Repositories to implement larger learning solutions. Future work will focus on testing other RDF store implementations and evaluate other aspects of the system performance.

REFERENCES

[1] X. Ochoa and E. Duval, "Use of contextualized attention metadata for ranking and recommending learning objects," in *Proceedings of the 1st international workshop on Contextualized attention metadata: collecting, managing and exploiting of rich usage information*, 2006, pp. 9-16.

[2] B. Simon, D. Massart, F. Van Assche, S. Ternier, E. Duval, S. Brantner, *et al.*, "A simple query interface for interoperable learning repositories," in *Proceedings of the 1st Workshop on Interoperability of Web-based Educational Systems*, 2005, pp. 11-18.

[3] E. Bogdanov, C. Ullrich, E. Isaksson, M. Palmer, and D. Gillet, "From LMS to PLE: a step forward through opensocial apps in moodle," in *Advances in Web-Based Learning-ICWL 2012*, ed: Springer, 2012, pp. 69-78.

[4] C. Bizer, T. Heath, and T. Berners-Lee, "Linked data-the story so far," *International journal on semantic web and information systems*, vol. 5, pp. 1-22, 2009.

[5] M. Bergman, "Advantages and Myths of RDF," *AI3, April*, 2009.

[6] B. Christian and S. Andreas, "The Berlin SPARQL Benchmark," *International Journal on Semantic Web and Information Systems (IJSWIS)*, vol. 5, pp. 1-24, 2009.

[7] Y. Guo, Z. Pan, and J. Heflin, "LUBM: A benchmark for OWL knowledge base systems," *Web Semantics: Science, Services and Agents on the World Wide Web*, vol. 3, pp. 158-182, 10// 2005.

[8] M. Schmidt, T. Hornung, N. Küchlin, G. Lausen, and C. Pinkel, "An Experimental Comparison of RDF Data Management Approaches in a SPARQL Benchmark Scenario," in *The Semantic Web - ISWC 2008*. vol. 5318, A. Sheth, S. Staab, M. Dean, M. Paolucci, D. Maynard, T. Finin, *et al.*, Eds., ed: Springer Berlin Heidelberg, 2008, pp. 82-97.

[9] C. Bizer and A. Schultz, "Benchmarking the performance of storage systems that expose SPARQL endpoints," *World Wide Web Internet And Web Information Systems*, 2008.

[10] Y. Guo, A. Qasem, Z. Pan, and J. Heflin, "A requirements driven framework for benchmarking semantic web knowledge base systems," *Knowledge and Data Engineering, IEEE Transactions on*, vol. 19, pp. 297-309, 2007.

[11] X. Ochoa, G. Carrillo, and C. Cechinel, "Use of a Semantic Learning Repository to Facilitate the Creation of Modern e-Learning Systems.," in *Workshop EER: E-Learning and Educational Resources. XV International Conference on Human Computer Interaction (Interacción 2014)*, Puerto de la Cruz, Tenerife, 2014, pp. 535-542.

Leveraging Learning Analytics in a Personal Learning Environment using Linked Data

Selver Softic, Laurens De Vocht, Behnam Taraghi, Martin Ebner, Erik Mannens and Rik V. De Walle

Abstract—**We report on the reflection of learning activities and revealing hidden information based on tracking user behaviors with Linked Data. Within this work we introduce a case study on usage of semantic context modelling and creation of Linked Data from logs in educational systems like a Personal Learning Environment (PLE) with focus on reflection and prediction of trends in such systems. The case study demonstrates the application of semantic modelling of the activity context, from data collected for over two years from our own developed widget based PLE at Graz University of Technology. We model learning activities using adequate domain ontologies, and query them using semantic technologies as input for visualization which serves as reflection and prediction medium as well for potential technical and functional improvements like widget recommendations. As it will be shown, this approach offers easy interfacing and extensibility on technological level and fast insight on trends in e-learning systems like PLE.**

Index Terms—**Data Mining, Semantic Web, Electronic learning, Analytic models**

I. Introduction

Limited availability of resources along with a time efficiency focus forces the designers and decision makers of learning platforms to revise their methodologies and techniques in order to respond the challenges of time and the needs of their targeted groups. On the other hand, learners are expecting a focused and simple way to organize their learning process, without losing time on information and actions which could disturb or prolong their learning. Nowadays learning process became more individual, multi-faceted and activity driven with the tendency to ad hoc initiated collaboration and information exchange. These circumstances imply the need for a scalable, adaptive learning environment enriched with multimedia supportive materials, communication channels, personalized search and interfaces to external platforms from Social Web like e.g. Slideshare, Youtube channels etc. All these parameters increase the complexity of online learning platform design and organization. Dynamics involved in this process require shorter optimization cycles in adaptation of learning process. Also maintaining such platforms is intensively changing process demanding from maintainers to actively adapt their systems to the learner needs. Adaptation to learner needs has a strong impact on acceptance of such platforms and should be matter of continuous improvement.

Cumulated system monitoring data (e.g. logs) of such environments offer new opportunities for optimization [1]. Such data can contribute the better personalization and adaptation of the learning process but also improve the design of learning interfaces.

Main contribution of the paper is a case study done with the logs from PLE at Graz University of Technology, presenting approach using Linked Data to mine the usage trends from PLE. The idea behind this effort is aiming at gaining insights, [2] useful for optimization of PLE and adapting them to the learners by using more personalization e.g. through recommendation of interesting learning widgets.

The research activities that have been described in this paper were funded by Graz University of Technology, Ghent University, iMinds (an independent research institute founded by the Flemish government to stimulate ICT innovation), the Institute for the Promotion of Innovation by Science and Technology in Flanders (IWT), the Fund for Scientific Research-Flanders (FWO-Flanders), and the European Union.

S. Softic, B. Taraghi and M.Ebner is with Social Learning Group, Graz University of Technology, Münzgrabenstraße 35A, 8010 Graz, Austria (corresponding author e-mail: selver.softic@ tugraz.at).

L. De Vocht, E. Mannens and R. Van De Walle is with iMinds - Multimedia Lab, Ghent University, Gaston Crommelaan 8, 900 Ghent Belgium (e-mail: first.last @ ugent.be).

II. Related Work

This section reports shortly about most relevant related work regarding PLE (at Graz University of Technology) and semantic technologies used in this work.

A. Learning Analytics and Reflection of Learner Logs

Current Learning Analytics research community defines [3] Learning Analytics as the analysis of communication logs [4], [5], learning resources [6], learning management system logs as well existing learning designs [7],[8] and the activity outside of the learning management systems [9],[10]. The result of this analysis improves the creation of predictive models [11], recommendations [12],[13] and refection [14]. Learning Analytics resides on algorithms, formulas, methods, and concepts that translate data into meaningful information. Modelling, structuring and processing the collected data derived from e.g. learner behavior tracking plays a decisive role for the evaluation. Different works outlined the importance of tracking activity data in Learning Management Systems [2],[3],[4],[12],[14]. None of them addressed the issue of intelligently structuring learner data in context and processing it to provide a flexible interface that ensures maximum benefit from collected information.

B. *PLE at Graz University of Technology*

The main idea of PLE at Graz University of Technology (http://ple.tugraz.at) is to integrate existing university services and resources with services and resources from the World Wide Web in one platform and in a personalized way [15], [16]. The TU Graz PLE contains widgets [15-17] that represent the resources and services integrated from the World Wide Web. Web today provides lots of different services; each can be used as supplement for teaching and learning. The PLE has been redesigned in 2011, using metaphors such as apps and spaces for a better learner-centered application and higher attractiveness [18],[19]. In order to enhance PLE in general and improve the usability as well as usefulness of each individual widget a tracking module was implemented by prior work [20].

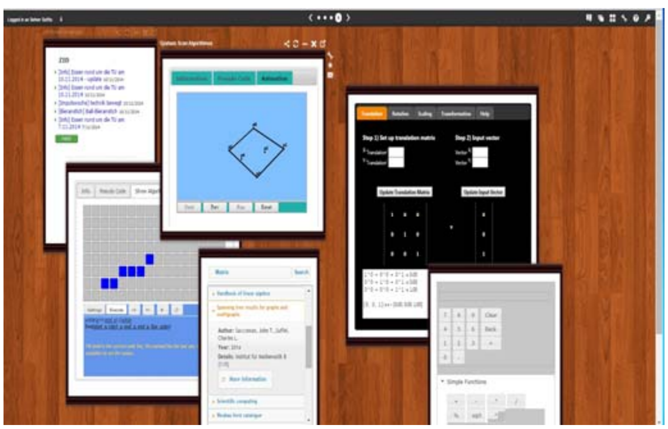

Fig. 1. PLE at Graz University of Technology.

C. *Semantics for Modeling Learners in PLE*

The Semantic Web standards like RDF (http://www.w3.org/RDF), RDFS (http://www.w3.org/TR/rdf-schema/), OWL (http://www.w3.org/2004/OWL/) and SPARQL (http://www.w3.org/TR/rdf-sparql-query/) enable data to be modeled and queried as graphs. Data schema is usually projected on specific knowledge domain using adequate ontologies. This approach has been fairly successful used to generate correct interpretation of web tables [21] to advance the learning process [1],[22] as well to support the controlled knowledge generation in E-learning environments [23]. This potential was also recognised by resent research in IntelLEO Project (http://intelleo.eu). IntelLEO delivered an ontology framework where Activities Ontology (http://www.intelleo.eu/ontologies/activities/spec/) is used to model learning activities and events related to them. The second interesting contribution from recent Ontology research work in IntellLEO project is the Learning Context Ontology which describes the context of a learning situation (http://www.intelleo.eu/ontologies/learning-context/spec/). Due to the relatedness to the problem that is addressed by this work those ontologies have been used to model the context of analytic data collected from PLE logs.

III. Approach for Mining Learner Logs

A. *Dataset*

Data used in the case study originates from Personal Learning Environment (PLE) developed for the needs of Graz University of Technology which serves currently more than 4000 users. The data was collected during two years period in order to generate analytics reports with visualization support for overall usage and process view on our environment following the research trends of previous years [3],[9].

B. *Modeling Learner Logs*

The main precondition for meaningful mining of usage trends is meaningful modelling of data. Such action assumes the choice of appropriate vocabulary or ontology. The RDF standard as such offers only the generic framework how to: organize, structure and link data.

The Activities Ontology introduced by IntelLEO research project offers a vocabulary to represent different types of activities and events related to them. Further this ontology also supports the description of environment (in our case PLE) where these activities occur. The Learning Context ontology serves as shown in figure 2. as container model to link PLE usages as event to the widget as execution environment where this event happens.

Every usage of a PLE widget creates a logging entry which produces a RDF construct represented in figure 3. Such constructs represent Linked Data which is then stored in a RDF memory store: Graph Database for Linked Data with SPARQL Endpoint, an interface where Linked Data can be queried.

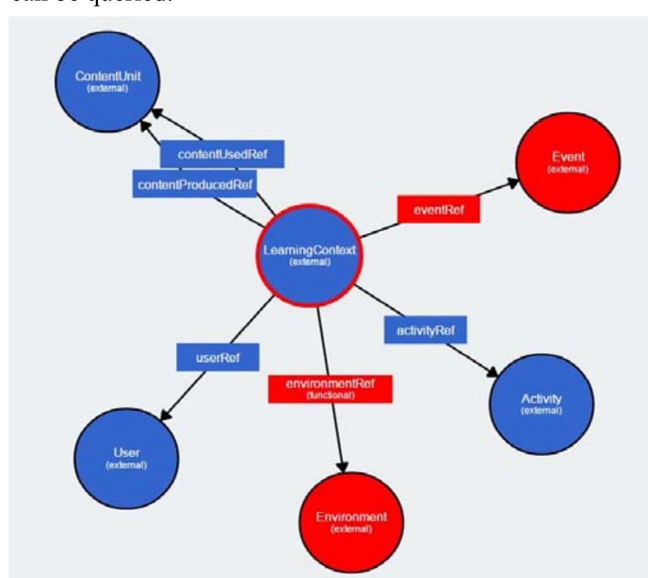

Fig. 2. Visualisation by WebVOWL beta 3.0 of a LearningContext ontology concepts and properties used to model the PLE log data.

```
@prefix ao: <http://intelleo.eu/ontologies/activities/ns/> .
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
@prefix lc: <http://www.intelleo.eu/ontologies/learning-context/ns/> .

<http://ple.tugraz.at/ns/events/log/#7912>
    rdf:type ao:Logging;
    ao: occursIn <https://ple.tugraz.at/ns/widgets/#LatexFormulaToPngWidget>;
    ao:timestamp "2012-10-04T07:52:52" .

<https://ple.tugraz.at/ns/widgets/#LatexFormulaToPngWidget>
    rdf:type ao:Enivironment;
    rdfs:label "LaTeXFormulaPNG Converter" .

<http://ple.tugraz.at/ns/learningcontext/#7912> a lc:LearningContext;
    lc:environmentRef
    <https://ple.tugraz.at/ns/widgets/#LatexFormulaToPngWidget>;
    lc:eventRef <http://ple.tugraz.at/ns/events/log/#7912>.
```

Fig. 3. Sample model of a PLE log in expressed as LearningContext.

In figure 3. a sample instance of lc:LearningContext links the usage log event denoted as instance of class ao:Logging , a subclass of an ao:Event, which occurred at certain time point inside the learning widget named LatexFormulaToPNGWidget represented through class ao:Enviroment.

What we can see with this example is that: vocabularies and ontologies which suits well to specific use case, enrich the analytic process with a high level of expressiveness in a very compact manner.

C. Querying the Models

Usage logs data presented as Linked Data graph are query able using SPARQL. In this way we are able to answer the questions like: *"Show me the top 10 used widgets?"*. Figure 4. represents exactly this question stated in the manner of SPARQL syntax. The benefit of this query is visible for instance in figure 6. where the results of this query (see figure 5.) influence the widget arrangement in the widget store. Such direct impact on system with functional operability on machine level would not be possible without standards like SPARQL and RDF.

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
PREFIX lc: <http://www.intelleo.eu/ontologies/learning-context/ns/>

SELECT DISTINCT ?widget ?ref (COUNT(?ref) AS ?count)
WHERE
{
 ?link rdf:type lc:LearningContext ;
       lc:environmentRef ?ref .

 ?ref rdfs:label ?widget .
}
GROUP BY ?ref
ORDER BY DESC(?count)
LIMIT 10
```

Fig. 4. Querying the usage of top 10 widgets in PLE.

IV. Conclusion and Outlook

Presented approach allows us mining the trends of PLE widgets usage overall time periods like presented in figure 5. This pie chart graph depicts the visual answer of the query from figure 4. The overview over distribution of widget usage can reflect the overall interest of the users within PLE for

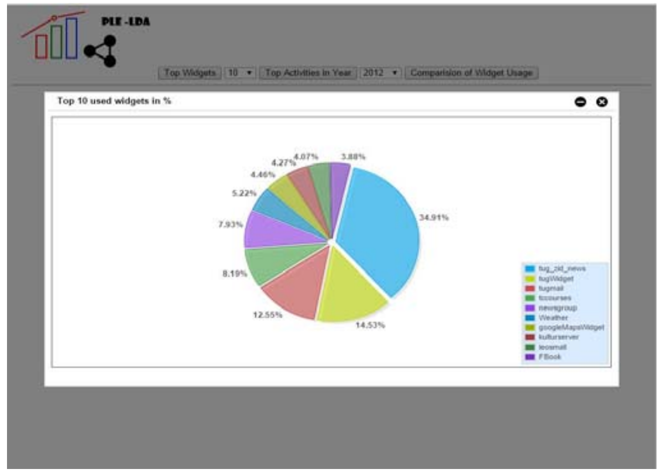

Fig. 5. Visualising top 10 used widgets in PLE.

different periods of time. Such outputs implicitly support the improvement of the quality of services for students and teachers. The same results from query in figure 4. are also used as input for ranking of widgets in widget store depicted in figure 6. This example shows the manifold application of such approach. The PLE becomes, in technical manner, extensible and well connected by standardized and intelligent interfaces and available for other web based tools and services.

Future efforts will focus on user wise statistics of learning widgets, since PLE can also provide this information. Especially the learning widget store as part of PLE could profit from this improvement. Mostly used and favored widgets by users will be ranked higher and recommended by the store itself as shown in figure 6. This process will be personalized as

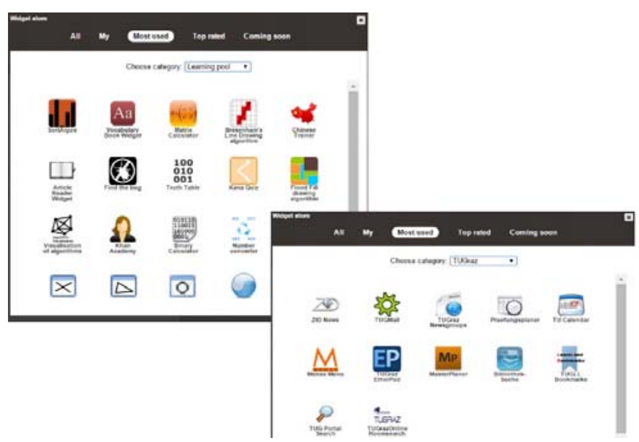

Fig. 6. Optimized widget store based on Linked Data statistics.

soon as the user information is included. The presented Learning Context ontology as such have foreseen such option already. By tracking the usages on user level the teachers will be able to draw conclusions about the popularity and quality of their learning widgets, on more granular and personal level.

Beside presented practical benefits from using Linked Data in Learning Analytics demonstrated on the example of PLE it is important to pinpoint how such approach differs from conventional methods and to which extent extends them. While conventional Learning Analytics focus more on Monitoring and drawing conclusions from analyzed and derived results Linked Data driven Learning Analytics approach delivers and derives the results on demand and in in-time. This option offers technical extensibility and claims interoperability by default, opening the interfaces toward other web platforms, sources and internet technologies. Flexibility through SPARQL standard for interactive querying allows the dynamic generation of inputs for hosting platforms, visualizations and analytic dashboards.

Such actions require data models with certain degree of expressiveness, and well-thought-out constrains. Main challenge lies in choice or construction of proper model, as well as in the decision about the granularity degree of chosen model. Sometimes, this process is limited by the quality and variety of provided data. Very important advantage of such models is their adaptability to extensions, reductions and changes of model schema. The nature of RDF, RDFS and OWL allows also to inference based on logical rules. This is especially useful for asking sophisticated questions about the context of modeled data.

Leveraging Learning Analytics with Linked Data we support standardized interfaces for information exchange, offer flexibility for visual other kinds of analytics, and also can enrich the learning system’s data with Linked Data sources from the Web as well with results from querying the graphs. The spread of applicability covers wide range of analytic methodologies like prediction, reflection and as outcome of these the intervention field.

Presented work based on case study of data from a PLE outlines the contribution of Semantic Technology Stack and Linked Data to Learning Analytics. The idea is promising and delivers great results with very low effort, what makes it especially valuable for analytical tasks targeted on improvement of learning environments like PLE.

References

[1] P. Prinsloo, S. Slade and F. Galpin, “Learning analytics: challenges, paradoxes and opportunities for mega open distance learning institutions”, Proceedings of the 2nd International Conference on Learning Analytics and Knowledge, ACM, 2012, pp. 130-13.

[2] R. Mazza and C. Milani, "Exploring Usage Analysis in Learning Systems: Gaining Insights From Visualisations Workshop on Usage Analysis in Learning Systems", Proceedings of Artificial Intelligence in Education, 2005.

[3] J. L. Santos Odriozola, K. Verbert, S. Govaerts and E. Duval, "Visualizing PLE usage", Proceedings of EFEPLE11 1st Workshop on Exploring the Fitness and Evolvability of Personal Learning Environments, CEUR WS, 2011, pp. 34-38.

[4] D. Rosen, V. Miagkikh and D. Suthers, "Social and semantic network analysis of chat logs", Proceedings of the 1st International Conference on Learning Analytics and Knowledge, ACM, 2011, pp. 134-139.

[5] A. Bakharia and S .Dawson, "SNAPP: a bird's-eye view of ter participant interaction", Proceedings of the 1st International Conference on Learning Analytics and Knowledge, ACM, 2011, pp.168-173.

[6] K. Niemann, H. C. Schmitz, M. Scheffel and M. Wolpers, "Usage contexts for object similarity: exploratory investigations", Proceedings of the 1st International Conference on Learning Analytics and Knowledge, ACM, 2011, pp.81-85.

[7] L. Lockyer and S. Dawson, "Learning designs and learning analytics", Proceedings of the 1st International Conference on Learning Analytics and Knowledge, ACM, 2011, pp.153-156.

[8] G.Richards and I. DeVries, "Revisiting formative evaluation: dynamic monitoring for the improvement of learning activity design and delivery", Proceedings of the 1st International Conference on Learning Analytics and Knowledge, ACM, 2011, pp.157-162.

[9] A. Pardo and C. D. Kloos, "Stepping out of the box: towards analytics outside the learning management system", Proceedings of the 1st International Conference on Learning Analytics and Knowledge, ACM, 2011, pp.163-167.

[10] P. Blikstein, "Using learning analytics to assess students' behavior in open-ended programming tasks", Proceedings of the 1st International Conference on Learning Analytics and Knowledge, ACM, 2011, pp.110-116.

[11] S. E. Fancsali, "Variable construction for predictive and causal modeling of online education data", Proceedings of the 1st International Conference on Learning Analytics and Knowledge, ACM, 2011.

[12] K. Verbert, H. Drachsler, N. Manouselis, M. Wolpers, R. Vuorikari, and E. Duval, "Dataset-driven research for improving recommender systems for learning", Proceedings of the 1st International Conference on Learning Analytics and Knowledge, ACM, 2011, pp. 44-53

[13] H. Drachsler, T. Bogers, R. Vuorikari, K. Verbert, E. Duval, N. Manouselis, G. Beham, S. Lindstaedt, H. Stern, M. Friedrich and M. Wolpers "Issues and considerations regarding sharable data sets for recommender systems in technology enhanced learning", Procedia Computer Science, 2010, 1, pp. 2849 - 2858.

[14] K. Verbert,N. Manouselis, H.Drachsler and E. Duval, "Dataset-Driven Research to Support Learning and Knowledge Analytics", Educational Technology & Society, 2012, 15, pp. 133-148.

[15] B. Taraghi, M. Ebner, G. Till and H. Mühlburger, "Personal Learning Environment-a Conceptual Study", iJET, International journal of emerging technologies in learning, 2010, 5, pp. 25-30.

[16] M.Ebner and B. Taraghi, "Personal Learning Environment for Higher Education--A First Prototype", World Conference on Educational Multimedia, Hypermedia and Telecommunications, 2010, pp. 1158-1166.

[17] B. Taraghi, M. Ebner and S. Schaffert "Personal learning environments for higher education: A mashup based widget concept", Proceedings of the Second International Workshop on Mashup Personal Learning Environments (MUPPLE09), Nice, France, 2009, 1613-0073.

[18] M. Ebner, N. Scerbakov, B. Taraghi, W. Nagler and I. Kamrat, "Teaching and Learning in Higher Education – An Integral Approach", Proceedings of Society for Information Technology & Teacher Education International Conference 2010, Gibson, D. & Dodge, B. (Eds.) , AACE , 2010 , pp. 428-436.

[19] B. Taraghi, M. Ebner and K. Clemens, "Personal Learning Environment -- Generation 2.0", World Conference on Educational Multimedia, Hypermedia and Telecommunications, 2012, pp. 1828-1835.

[20] B. Taraghi, C. Stickel and M. Ebner, "Survival of the Fittest -- Utilization of Natural Selection Mechanisms for Improving PLE", Proceedings of the first Workshop on Exploring the Fitness and Evolvability of Personal Learning Environments, 2011, pp. 4-9.

[21] V. Mulwad, T. Finin, Z. Syed and A. Joshi, "Using linked data to interpret tables", Proceedings of the the First International Workshop on Consuming Linked Data, 2010.

[22] Z. Jeremić, J. Jovanović and D. Gašević, "Personal learning environments on the Social Semantic Web Semantic Web", IOS Press, 2012.

[23] S. Softic, B. Taraghi and W. Halb, "Weaving Social E-Learning Platforms Into the Web of Linked Data", I-SEMANTICS, 2009, pp. 559-567.

Selver Softic obtained his master’s degree with distinction in the field of computer science. Currently he is writing his PhD on topic of application of Linked Data for learning management systems supervised by Prof. Ebner at Graz University of Technology. Mr. Softic visited in year 2013 the Multimedia Lab (Semantic Web Task Force) led by Prof. Van De Walle at Ghent University as visiting researcher . He is author of various papers on PLEs, linked Data and Semantic Web.

OpenFING: A Platform for Video Semantic Annotation as Learning Approach

Fernando Carpani, Matías Parodi, Lorena Etcheverry
Grupo de Concepción de Sistemas de Información , Instituto de Computación. Facultad de Ingeniería. Universidad de la República. Uruguay

***Abstract*— The video-recording of traditional lectures is a low-cost activity for teachers and a supplement for courses with a high number of students. Also, videos can become a helpful tutoring resource. With these ideas in mind as motivation, we developed OpenFING, a Semantic Web based collaborative platform for video annotations. In this paper we present this platform, which allows students and teachers to create video fragments, add annotations, and provides searching mechanism over structured metadata.**

***Index Terms* E-Learning, Video Annotation, Semantic Web, Linked Data.**

I. Introduction

In Uruguay, the *Instituto de Computación* is the research and teaching center in computer sciences (CS) at the *Universidad de la República* (UdelaR). It is located within the *Facultad de Ingeniería* (FING, Engineering School) and is responsible for all the undergraduate and graduate programs in CS in UdelaR. As many Latin America schools, FING is experiencing an increase in matriculation rates and scarce resources, observing low graduation and high drop-out rates. New strategies have become necessary to adapt the scholar system to this reality.

The video-recording of traditional lectures is a low-cost activity for teachers and it can be seen as a supplement for a traditional course. According to some studies, recorded lectures can become a helpful tutoring resource, mainly because videos have a slower, more step-by-step lecture style than the classroom lectures; student use of videos is voluntary and can be tailored by students to meet their learning and topic-review needs, and can occur when and where students learn most effectively [1].

In this work we present OpenFING, a Semantic Web based collaborative platform to publish and annotate videos. With this platform, teachers and students can annotate videos with topics, comments, web resources, and other kind of metadata to improve their teaching and learning activities. One of our main concerns, from the technical point of view, was to develop an architecture in which new features could be easily introduced to the platform. This lead us to the use of Semantic Web (SW) technologies [2] to develop the platform, in particular Linked Data paradigm [3] .

The rest of the paper is organized as follows. We start in Section II describing OpenFING functionalities. Then in Section III we discuss on related work. In Section IV we describe the OpenFING platform. Finally, in Section V we present conclusions and in Section VI current and future work.

II. A walk-thorugh OpenFING

In this section we present OpenFING functionalities via a set of use cases. Along the description we refer to the screen capture presented in figure 1. Italicized terms refer to screen sections.

A. *Search and find.*

A user starts the session selecting a course in the *Course Menu*. Also, the *Search Box* can be used to perform queries. Queries input may be plain text (e.x. “induction”) or contain tags to refer to specific objects in the platform (e.x. course:, lecture:). Then, the search is performed using a combination of SPARQL queries and text search on the labels and titles values. In our example, the search for “induction” returns a video lecture where the title “Inductive Set Definitions” matches the search criteria. This video contains the complete lecture about the concept he is looking for, but also other related concepts.

B. *Fragmentation and annotation.*

While the user is watching the video, he decides to mark the video fragment where the teacher defines the “Declarative View of Inductive Sets”, and annotate it with the topic “Declarative view”. To do this, he uses the *Annotation Type Selector* to declare the type of the annotation as a “Topic”, and then he writes the topic in the *Fragment Creator* text area. At this time, the fragment start time is recorded. When the user pushes the blue button, the end time is recorded and the video fragment and its annotations are saved. Both objects are associated with the user. In the system, video fragments are identified by URLs which follow the Media Fragment URI 1.0 recommendation of W3C.

C. *See annotations of other users.*

While the user watches videos, he can also see annotations created by other users in the *Annotation Viewer*. These annotations appear dynamically as the start time of related fragments is reached. When the user clicks in an annotation, the related video fragment starts in the player.

D. *Using external resources.*

OpenFING may coexist with learning platforms, such as Moodle. Users may then also annotate video fragments using URLs that refer to lecture slides, or questions in a forum. This mechanism also allows to add reference to any URL on the internet, in particular to add references to other video fragments in OpenFING, and was developed at zero cost because the use of standard dereferenceable URIs.

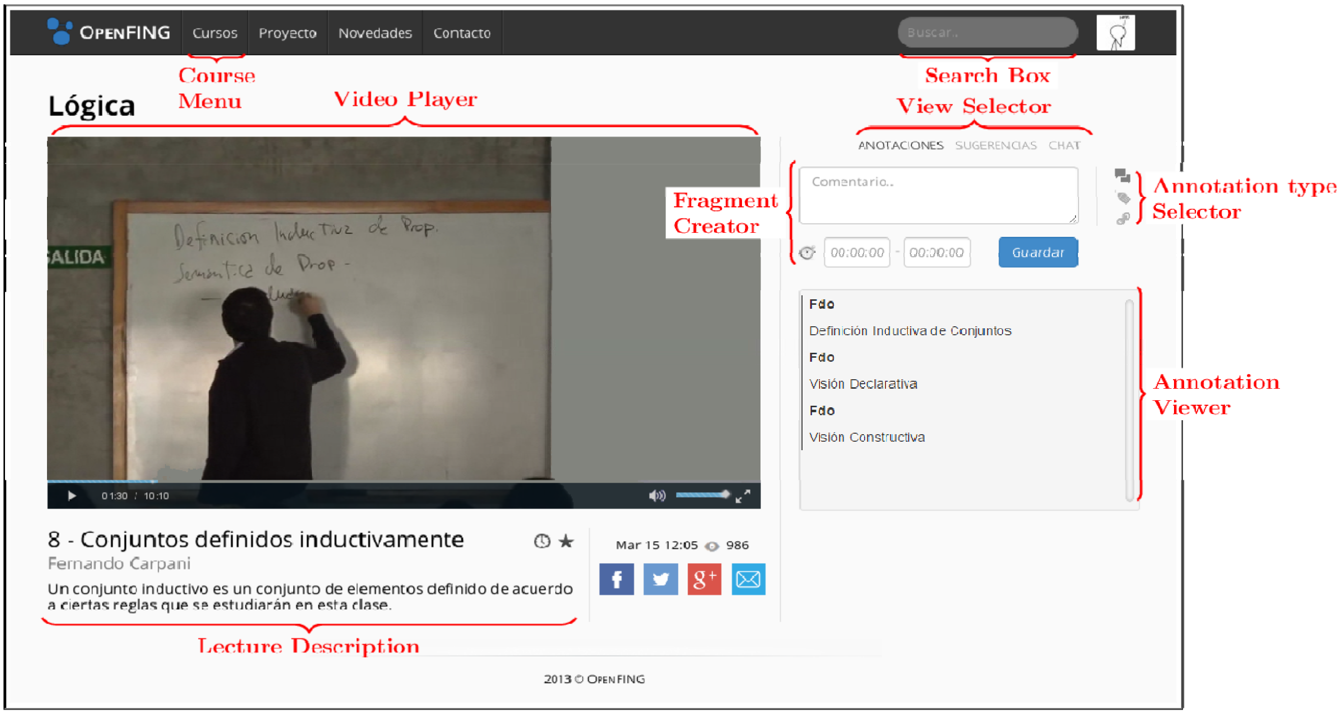

Fig 1: OpenFING client application screenshot

E. Recommended videos and resources.

While users watch videos, related videos and resources are shown in the recommendations panel (Fig. 2), which is accessible from the *View Selector*. The contents of this panel change dynamically according to the annotations found in the video. The recommendation criteria implemented so far is very simple, and retrieves video-fragments that refer to the same topic, but other criteria can be easily added to the platform.

Fig 2: Recomendations Panel

F. Teachers Activities.

Students may use OpenFING without involving the teachers, but their participation may improve the experience. For example, teachers can curate users annotations assessing its correctness, or help in the organization of topics according to some taxonomy. Also, teachers can evaluate the comprehension of a certain topic by checking the annotations created by students. Finally, teachers can also propose the creation of annotations as a learning activity, as suggested in [4].

III. Related Work

Some works deal with the use of annotations in e-Learning. In [4] the authors review a set of learning experiences that use annotations, and extract some recommendations about the use of annotations as a learning activity. In [5], an experiment about social annotation in an educational environment is presented which concludes that is a good way to promote the student engagement in the educative process. None of these works deal with video annotations. Several works treat video annotations, but only a few focus on educational videos. The work presented in [6] is close to OpenFING, but they do not use Semantic Web Technologies. About the use of Semantic Web technologies in e-Learning, some works should be taken into account. OpenCourseWare (OCW) Universia Team experience about producing and consuming Linked Data is presented in [7]. The paper introduces LOCWD, a vocabulary to describe OCW resources. In [8] a platform with some similarities to OpenFING is described where the search mechanism exploits LOD.

IV. OpenFING: A Semantic Web Based Platform

OpenFING is strongly based on SW technologies. Its data model is composed of two ontologies, and the video fragments and annotations are recorded as RDF triples in a triplestore. This allows to exploit data via SPARQL queries and reasoning strategies like Entailment Regimes. In the following sections we present the designed ontologies and the architecture of the OpenFING platform.

A. *Design decisions*

OpenFING is a collaborative platform to publish, watch, search and annotate videos. It is based on the following assumptions:

- **Each video is a set of video-fragments,** and each video-fragment is linked to a set of annotations.
- **An annotation implies a link between fragments and the annotation itself.** This links are dinamically created by human users or by software agents (e.x. Semantic Enricher in fig. 5). The set of annotations linked to a fragment are metadata tracks for the video.
- **Data can be represented as a distributed graph**. This graph is distributed because some parts might be in different servers with a decentralized management.

The aforementioned characteristics lead us to implement OpenFING using SW technologies. We now briefly describe some of the benefits of this approach with respect to traditional database technologies.

- **Our vision of data is a graph**. So, SW technologies like RDF provide a natural way to manage these data. In a relational database we need to map this vision to tables.
- **SW technologies simplify a distributed model.** These technologies allow us to distribute data among different sites. For example, the interaction between several instances of OpenFING platform can be easily implemented using SPARQL federated queries.
- **SW simplifies the extensibility of the data model.** In SW, adding new things is simpler than in a relational approach. For instance, adding a new attribute in a table may impact all the tuples in the table. Although this can be solved with a more complex design, this usually implies greater maintenance and operational costs**.** On the contrary, in the SW approach only the objects that refer to this new property are modified.
- **SW simplifies the interoperability with other systems.** External systems and users can obtain data via querying the SPARQL endpoint.

B. *The Data Model as Ontologies*

OpenFING data model is composed of two OWL2[9] ontologies: a generic vocabulary about video-fragments and its metadata, called **MMC**, and a specialization of the former to include some particularities of social components, called **OFM**. MMC is a specialization of the W3C Ontology for Media Resources[10], an it includes concepts like *Media*, *Metadata*, *Topic*, etc. OFM includes concepts like *University*, *Teacher*, *Lecture*, *Course*, etc. Figure 3 presents the main components of the OFM ontology while Figure 4 shows an example that illustrates how this ontology is used to represent data and represents the annotations built by two different users (Lucia and Carlos). These annotations refer to different courses, so, different fragments on different videos, but both annotations refer to the same topic.

OFM is aligned with standard vocabularies and ontologies like W3C Ontology for Media Resource, the Academic Institution Internal Structure Ontology (Aiiso) [11], and the

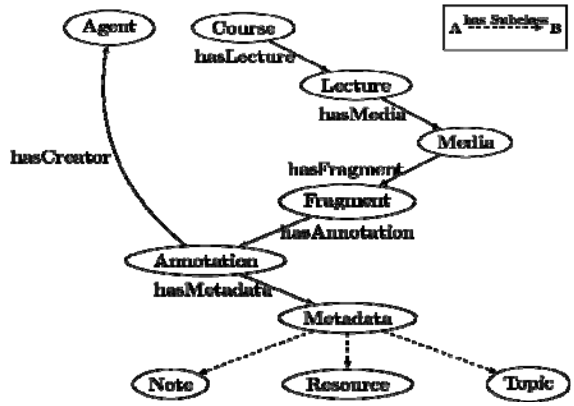

Fig 3: An extract of OFM ontology

Friend of a Friend Vocabulary (FOAF)[12]. Usually, Linked Data based systems use terms from different vocabularies. We think that this strategy delays the development because it is difficult to use all terms in a correct way. We choose to define our own terms in the design of our ontology and then map its concepts on standard vocabularies using **rdfs:subpropertyOf** and **rdfs:subclassOf** relationships. This strategy allows us to publish a SPARQL endpoint which can resolve queries using terms of standards vocabularies, while we use our ontology to define the data.

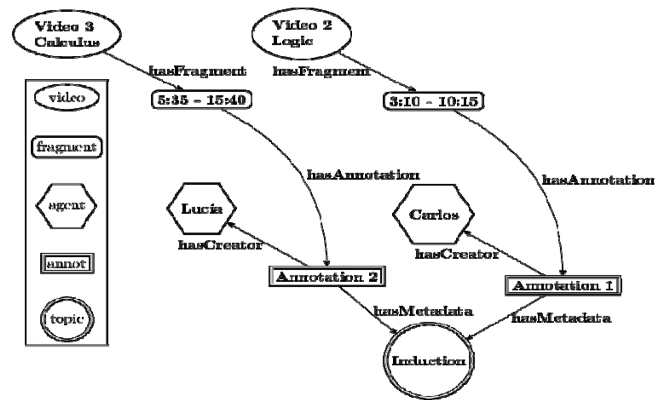

Fig 4: Structure made by two users (agents) who uses the client.

C. *Architecture*

Figure 5 presents the functional architecture of OpenFING. The main component, *OpenFING server*, acts as an application proxy and a wrapper over the triplestore by implementing the access following the datamodel. The server exchanges mostly JSON structures with the client using HTTP, and uses SPARQL protocol to access to the triplestore. The *Annotations* component is implemented as a RDF triplestore. These annotations can be either generated by users, via the client application, or semi-automatically by the *Semantic Enricher (SE)*. This component searches relevant resources in the web using techniques from areas like Natural Language Processing, Machine Learning, Data Mining. For example, it can add a reference to the DBPedia page about a certain topic.

OpenFING Client is a web interface built on top of HTML5 and Javascript, that takes advantage of HTML5 *video* and *track* tags. Finally, the V*ideos* component can be distributed over different websites, because the client only needs the video URL to use it. The platform is still under development, but we have some early prototypes.

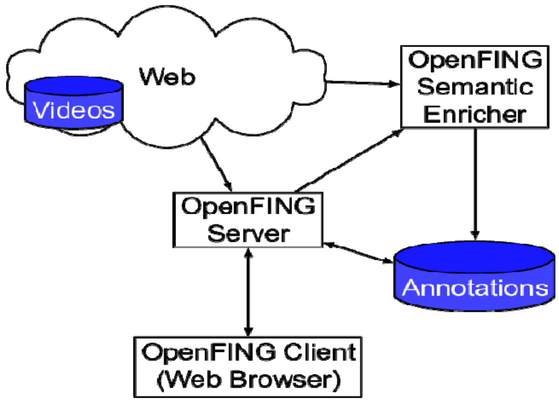

Fig 5: OpenFING Architecture

V. Conclusions

In this work we presented OpenFING, a Semantic Web based collaborative platform to publish and annotate videos. With this platform, teachers and students can annotate videos with topics, comments, web resources, and other kind of metadata to improve their teaching and learning activities. The development of video-lectures is usually considered as a high cost activity for teachers. Our low-cost approach, based on the publication of video-recorded traditional lectures, has still proven to be useful to students.

We also believe that our approach actively promotes the involvement of students in their learning process, since annotations and video-fragments are mainly added by them. In this sense, our approach is aligned with the ideas proposed in Blended Learning strategies[13].

From a technological point of view, we believe that Semantic Web technologies allowed us to develop a flexible environment, in which we can add new features in a simple way. We also think that HTML5, JS, NodeJS, SPARQL stack works as a good prototyping platform since it reduces programming and testing times.

VI. Current and Future Works

The development of OpenFING is still a work in progress. New versions of OpenFING server and clients are being developed using NodeJS and HTML5. In the near future we expect to extend the Semantic Enricher component using two approaches: querying LOD, and using Natural Language Processing of documents. On the educational dimension, we are starting to develop strategies to study the effects of using OpenFING over students and teachers.

Acknowledgment

This work will not be possible without the voluntary work of the OpenFING filming and edition team, composed by voluntary students.

References

[1] H. D. Brecht y S. M. Ogilby, «Enabling a Comprehensive Teaching Strategy: Video Lectures», *JITE*, vol. 7, pp. 71–86, 2008.

[2] P. Hitzler, M. Krötzsch, y S. Rudolph, *Foundations of Semantic Web Technologies*. Chapman & Hall/CRC, 2009.

[3] T. Heath y C. Bizer, *Linked Data: Evolving the Web into a Global Data Space*, 1st ed. Morgan & Claypool, 2011.

[4] E. Novak, R. Razzouk, y T. E. Johnson, «The educational use of social annotation tools in higher education: A literature review», *Internet High. Educ.*, vol. 15, n.º 1, pp. 39 - 49, 2012.

[5] F. Gao, «A case study of using a social annotation tool to support collaboratively learning», *Internet High. Educ.*, vol. 17, n.º 0, pp. 76 - 83, 2013.

[6] A. L. Franzoni, C. P. Ceballos, y E. Rubio, «Interactive Video Enhanced Learning-Teaching Process for Digital Native Students», en *Advanced Learning Technologies (ICALT), 2013 IEEE 13th International Conference on*, 2013, pp. 270-271.

[7] N. Piedra, E. Tovar, R. Colomo-Palacios, J. Lopez-Vargas, y J. Alexandra Chicaiza, «Consuming and producing linked open data: the case of OpenCourseWare», *Program*, vol. 48, n.º 1, pp. 16-40, 2014.

[8] H. Q. Yu, C. Pedrinaci, S. Dietze, y J. Domingue, «Using Linked Data to Annotate and Search Educational Video Resources for Supporting Distance Learning», *Learn. Technol. IEEE Trans. On*, vol. 5, n.º 2, pp. 130-142, abr. 2012.

[9] M. Krötzsch, P. F. Patel-Schneider, S. Rudolph, P. Hitzler, y B. Parsia, «OWL 2 Web Ontology Language Primer», W3C, oct. 2009.

[10] P.-A. Champin, T. Bürger, T. Michel, J. Strassner, W. Lee, W. Bailer, J. Söderberg, F. Stegmaier, J.-P. EVAIN, V. Malaisé, y F. Sasaki, «Ontology for Media Resources 1.0», W3C, W3C Recommendation, feb. 2012.

[11] N. Sahbeer y R. Styles, «Academic Institution Internal Structure Ontology». Talis Inc., 14-may-2008.

[12] D. Brickley y L. Miller, «FOAF Vocabulary Specification 0.97», Namespace document, ene. 2010.

[13] D. Ifenthaler, «Blended Learning», en *Encyclopedia of the Sciences of Learning*, N. Seel, Ed. Springer US, 2012, pp. 463-465.

Implementing “Rethinking Education”: Matching Skills Profiles with Open Courses through Linked Open Data technologies

Maria Zotou, Agis Papantoniou, Karel Kremer, Vassilios Peristeras, and Efthimios Tambouris

***Abstract*—The "Rethinking Education" EC strategy highlights the importance of learning outcomes in education. The so-called Massive Open Online Courses (MOOC) is becoming an increasingly important part of education. In this paper, we present a novel approach and proof-of-concept application to enable automatically matching MOOCs learning outcomes with learners' needs related to specific skills and competencies. For this purpose, we exploit the ESCO classification as well as Linked Open Data (LOD) technologies.**

***Index Terms*—ESCO, LOD, rethinking education, skills, competences, MOOC, Open Courses, Semantic Web**

I. Introduction

The "Rethinking Education" Strategy [1] from the European Commission (EC) aims at a significant shift in education, providing more focus on 'learning outcomes', i.e. all knowledge, skills and competencies acquired during learning which people can use as qualification evidence. An important relevant EU initiative is the EURES. The EU EURES Job Mobility Portal [2] provides information on worker mobility and access to more than two million job vacancies published by European public employment services. One of the main objectives of this portal is to enable competence-based job matching. For this purpose, a comparison between individuals competence profiles and employers needs is performed to find the best match.

From the education perspective, there is also a shift towards innovative approaches on learning technologies [3] and Open Courses (OC). The EC Working Document entitled “Analysis and mapping of innovative teaching and learning for all through new Technologies and Open Educational Resources in Europe” [4] clearly points out that the EC sees Technc and Open Educational Resources as opportunities to res education in the European Member States.

In this respect, a rapidly increasing number of Massive Open Online Courses (MOOCs) are becoming available. These MOOCs open up education by providing accessible, flexible, affordable and fast-track completion of courses for free. The European MOOCs Scoreboard [5] illustrates this propagation of MOOCs throughout Europe.

With such an ambition to focus on learning outcomes empowered by an EC agenda to enhance competence-based matching and with MOOCs flourishing, the problem of formalizing a dynamic lifelong learning process becomes apparent. The question that one may logically ask "how can I find open courses that match my profile, so that I can further enhance my skills and competences?” makes the problem much clearer.

Such matching typically imposes the key challenge of having software "systems" talking to each other in a common language supported by semantics that are understood from both sides. This is clearly a semantic interoperability challenge [6]. The European Interoperability Framework [7] provides a set of general recommendations to enable such communication. The European Skills/Competences, Qualifications and Occupations multilingual (ESCO) classification [8] provides the necessary common vocabulary to facilitate exchange of information related to skills and qualifications.

From the technology perspective, we argue that the adoption of the Linked Open Data (LOD) paradigm [9] and its related semantic technologies enables semantic interoperability.

Based on the above, this paper investigates the problem of formalizing the lifelong learning process in a dynamic way. The investigation is supported by matching skills profiles (like the ones found in a CV) with courses openly provided through the Internet. More specifically, it proposes a methodology on:

- How to transform skills profiles using a semantic interoperability hub and link them with the web of data
- How to enable candidate profiles to interact with open digital learning resources and learning outcomes and
- How to match skills profiles with open courses' learning outcomes.

The methodology is being applied for the implementation of a pilot, proof-of-concept application.

The paper is structured as follows. Section 2 provides a brief description of ESCO, while section 3 presents the proposed methodology. Section 4 describes the

M. Zotou is with the Department of Applied Informatics of the University of Macedonia, Greece; e-mail: mzotou@uom.gr.
A. Papantoniou is with TenForce BVBA e-mail: agis.papantoniou@tenforce.com.
K. Kremer is with TenForce BVBA, e-mail: karel.kremer@tenforce.com.
V. Peristeras is with the European Commission, Interoperability Solutions for European Public Administration (ISA) Unit, e-mail: Vassilios.PERISTERAS@ec.europa.eu
E. Tambouris is with the Department of Applied Informatics of the University of Macedonia, Greece; e-mail: tambouris@ uom.gr

implementation of a pilot system, while section 5 discusses our results and findings. Finally section 6 concludes and presents the next steps.

II. What is ESCO

To achieve and foster semantic interoperability in the area of job search, the European Commission is developing the multilingual European Skills/Competences, Qualifications, and Occupations (ESCO) classification. ESCO offers a reference vocabulary for the labor market and for the education and training sectors bringing together job offers supply and demand. The classification's first release, ESCO v0, which went live in October 2013, is a fully functional pilot version and can be accessed at https://ec.europa.eu/esco.

ESCO organizes available knowledge about the European labor market and the training and education sectors in three main pillars: occupations, skills/competences, and qualifications. Each pillar has its own hierarchical structure, making it easy for users to browse through the classification and find what they are looking for.

ESCO organizes terminology into *concepts* and *terms*. Each occupation, skill/competence, and qualification is a concept and is uniquely identified by a uniform resource identifier (URI). Multiple terms are linked to concepts and are used to name a concept in a human-readable way, depending on the language and the gender desired by the user. ESCO is being built on existing standards. A concept has one preferred term per language, but can have multiple non-preferred terms as well. It is based on the EURES taxonomy and also makes use of existing standardized classifications like ISCO-08, NACE [10], and FoET (Fields of Education and Training) [11].

Finally, the ESCO data model is represented in Simple Knowledge Organization System (SKOS) [12] which is the World Wide Web Consortium's (W3C) recommendation for representing any type of controlled vocabulary including taxonomies, ontologies, thesauri and classification schemes. SKOS is built on top of the Resource Description Framework (RDF; www.w3.org/TR/rdf-primer) and one of its main objectives is to enable the production, publication, and use of controlled vocabularies as linked data.

III. Methodology

Our goal has been to allow candidates to match their professional profiles with digital learning resources and receive access to available content regarding their skills and external relevant knowledge from multiple courseware providers, capitulating on the linked data approach. In order to achieve the above objective, we applied a multi-step research approach, resulting in a pilot implementation as a proof-of-concept. The steps are as follows:

Step 1: Identify a controlled vocabulary that appropriately describes skills, occupations and qualifications. As already discussed, this vocabulary is the ESCO taxonomy, which can be used as a semantic interoperability hub, linking skills with available learning resources.

Step 2: Enrich the description of open courseware metadata by linking their learning outcomes with ESCO concepts. We chose the Coursera MOOC, which provides an API [15] for accessing information about its courses. More specifically, we retrieved information on two exemplary courses. We then used the RDFazer tool [14] to tag the courses' metadata with corresponding ESCO skills and extracted the generated RDF triples.

Step 3: Set up an RDF store to host all the RDF data created during the previous steps to enable discoverability. We chose the Virtuoso triple store and installed it on a web server for storing all our data.

Step 4: Locate additional MOOC providers and present supplementary external open courseware suggestions relevant to the candidate's skills. We chose the Udacity MOOC which also provides an API [16] for accessing course information. We then communicated directly with Udacity's API to retrieve courses that foster the skills in question.

Step 5: Build a simple and intuitive interface to visualize the process of matching skills profiles with open courses through Linked Open Data technologies and the results. We developed a pilot implementation using the Eclipse IDE and languages such as Java, SPARQL, JSON, HTML and Javascript.

IV. Pilot Implementation

We have developed a pilot application based on linked data and semantic technologies, in an effort to test and validate our approach. The application aims to address and put into practice the objectives originally set, i.e. enhancing employability via learning, and following the LOD paradigm in order to bridge learning outcomes with skills, occupations and qualifications, and thus provide formal, informal and lifelong learners the opportunity to be linked with the web of data.

We examine the case scenario where a candidate wishes to create a professional CV and add match his/her profile with the corresponding ESCO skills as well as with open digital courses available. The system's architecture diagram is

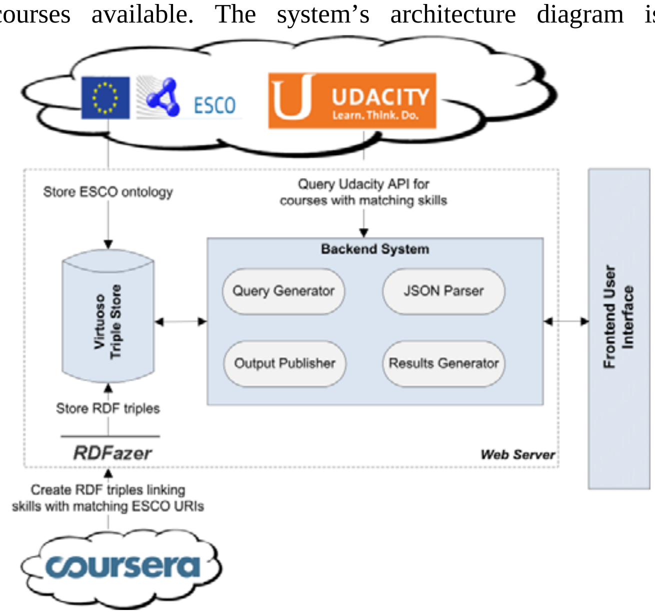

Fig. 1. Pilot software architecture diagram

presented in Fig. 1. The approach is that the candidate will have at his disposal ESCO concepts in order to do that, therefore the ESCO URIs will be injected in the CV amplifying the information regarding each skill.

The following Figure shows the user interface where the candidate can provide some initial personal information and add skills to be added to his/her CV. In this example, the user submits four skills, i.e. Data analysis, Python, Web design and Java.

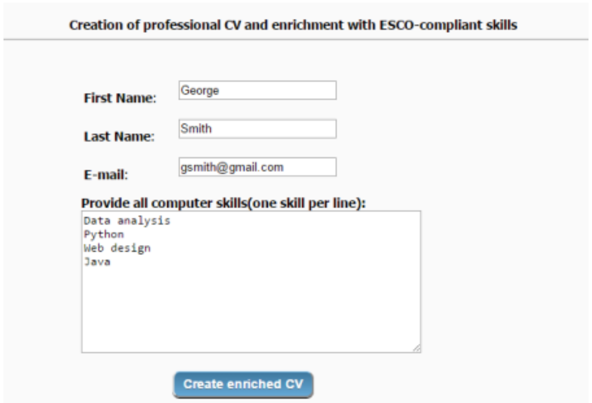

Fig. 2. Pilot's User Interface

Initially, we downloaded the ESCO ontology in its SKOS/RDF form and stored it in a Virtuoso triple store. This provided us the opportunity to query and retrieve 2,523,582 triples of information regarding skills, occupations and qualifications.

Next, we chose two courses available in Coursera as exemplary cases, namely "Programming for Everybody (Python)" and "An Introduction to Interactive Programming in Python". Using the Coursera API, we retrieved information about each course in JSON format, accessible via a web browser. More specifically, we retrieved the course's ID, title and section "About the course", which includes additional information on the curriculum, learning objectives and learning outcomes.

The next step was to enrich the courses' metadata by annotating each course's HTML snippet with the URIs of ESCO skills that matched the context.

This process was carried out using the RDFazer plugin for Google Chrome. We configured RDFazer to communicate with our Virtuoso store, and we queried the triples in order to tag the HTML page with RDFa concepts. As shown in Fig. 3, we searched for concepts in the page that indicated a specific skill, we searched for that skill in the triple store, and then we linked the matching ESCO skill URI with the HTML concept found in the course's description. Fig. 3 shows the annotation process for the first course, where we tagged the word "Python" with the ESCO skill "Python (programming language)", the phrase "data analysis" with the ESCO skill "Data analysis", the phrase "web design" with the ESCO skill "Web design" and the phrase "web browser" with the skill "Use internet browsers", while the second course was annotated with the skill "Python (programming language)". The RDF triples that were derived from this annotation process were stored in Virtuoso.

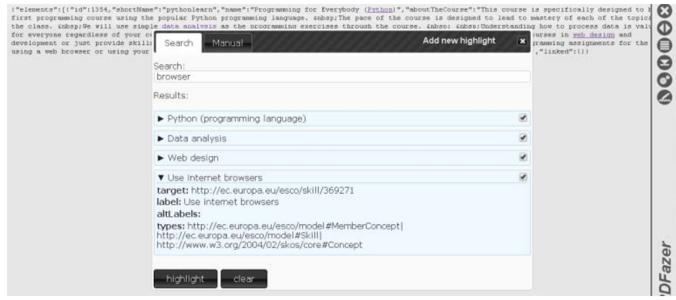

Fig. 3. Annotation of course skills with matching ESCO URIs

The backend system includes four components, as shown in Fig. 1. The Query Generator populates dynamic queries depending on the skills the candidate adds on the frontend and searches the store for the corresponding results. The queries involve searching:

1) ESCO skills URIs and labels that match or are similar to each candidate's skill. This matching is carried out automatically by the software. More specifically, the system searches in the triple store for ESCO skills that contain the given skill's name (e.g. python*). If any matches are found, the result viewed by the user in the web interface is the label of each skill linked to the ESCO website so that the user can retrieve additional information.
2) Coursera courses that have been linked with ESCO skills via RDFazer and stored in Virtuoso. The matching is again carried out by the tool, which queries the triple store for ESCO skills that have been linked as RDFa to content of the course's description. If any matches are found, the result viewed by the user is the title of each course linked to the course's page on Coursera. For example, if a user provides the skill "Python", the queries will return the courses "Programming for Everybody (Python)" and "An Introduction to Interactive Programming in Python", because both RDFa-zed courses available in the triple store include the word "python" in their descriptions. It should be noted that there is no review process included in these matches. The linking of candidate skills to ESCO skills and of ESCO skills with skills included in a Coursera course is realized through text-based matching.

The JSON parser requests information from the API provided by the Udacity MOOC on the courses available, then checks whether any courses are related with the given skills. If any matches are found, the result is the title of each course linked to the course's page on Udacity.

The Results Generator retrieves the aforementioned information per skill while the Output Publisher organizes and presents these results in an informative structure to the candidate. More specifically, a tab is dynamically created per given skill, which holds all information derived from the backend system, as shown in Fig.4.

Fig. 4. Pilot's Results page

In our example, the output in the tab for the skill "Python" includes a) the ESCO URI of the skill that matches the given attribute, b) a list of the courses available in Coursera (in this case both courses used in our example since they were linked to the skill) and c) a list of the courses available in Udacity that are related to the skill.

The pilot application can be accessed through the following URL: http://egov.it.uom.gr:8080/ESCO_impl/.

V. Discussion

The web of data provides abundant accessible and retrievable knowledge that can enhance learners' competencies in all professional fields. Learners are not always aware of the courses that exist and that they should attend in order to develop specific skills. This was the exact reason of the development of our simple pilot application; to show how the lifelong learning process can be formalized in a dynamic way

The EURES mobility portal provides the end-user with the possibility of preparing his CV online (following the Europass CV multilingual formats). Our intention was to simulate in a simplistic manner this process and show how annotation through a common and controlled vocabulary can support the process of matching skills (described in a CV) with the learning outcomes described in MOOCs that are supported by APIs. In this way the EURES portal could also be enhanced by providing information on existing MOOCs.

The overall objective is to have the user directly describe his/her skills in the CV by using ESCO URIs. If the same happens on the side of the MOOC providers, while preparing their course metadata, one can understand that there can be perfect matches.

ESCO will soon be published as Linked Open Data. Actually the whole SKOS/RDF file is already available for download from the ESCO Portal, but this isn't enough as developers using the taxonomy will want to receive metadata of specific ESCO concepts and in various formats like RDF, HTML, XML, CSV. This functionality, according to the ESCO roadmap [13] will be implemented in 2015.

What about Open Courses and Linked Data? As the pilot showed the main considerations in order to publish them as LOD are the following:

- Courses and course metadata need to be supported by an API

- ESCO should be used to annotate the content.

- Courses need to explicitly provide metadata about their learning outcomes, which currently and in most cases are only implicitly described in sections like "About the course".

VI. Conclusions

Within the context of the paper we aimed to investigate how the process of matching digital learning resources with professional profiles and competencies can be achieved. To this end, we exploited novel developments published by the European Commission such as the ESCO taxonomy as well as emerging paradigms such as Linked Open Data.

As a proof-of-concept, we developed a pilot application where candidates can create their professional profile by submitting their skills and enrich their profile by receiving supplementary content from the web of data. This pilot application demonstrates the feasibility and opportunities that derive from linking concepts from the ESCO taxonomy with skills on candidate profiles and learning outcomes of open digital resources. More specifically, the use of such an application can significantly benefit formal, informal and lifelong learners in developing appropriate competences that will increase their qualifications.

Future work will include an extension of the pilot software application to include more MOOC providers, as well as the improvement of the user interface with additional features when preparing skills profiles, a richer selection of sources, the ability to download the results in various formats etc.

References

[1] European Commission, "Rethinking Education initiative" http://ec.europa.eu/languages/policy/strategic-framework/rethinking-education_en.htm
[2] EURES, The European Jobs Network. https://ec.europa.eu/eures/
[3] European Commission, "Opening up education through new technologies", http://ec.europa.eu/education/policy/strategic-framework/education-technology_en.htm
[4] European Commission, "Opening up Education: Innovative teaching and learning for all through new Technologies and Open Educational Resources", http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex:52013DC0654
[5] Open Education Europa, "European MOOCS Scoreboard" http://www.openeducationeuropa.eu/en/european_scoreboard_moocs
[6] Wikipedia, "Semantic Interoperability", http://en.wikipedia.org/wiki/Semantic_interoperability
[7] Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of Regions 'Towards interoperability for European public services', "European Interoperability Framework (EIF) for European public services" http://ec.europa.eu/isa/documents/isa_annex_ii_eif_en.pdf
[8] European Commission, "ESCO – European Skills, Competencies, Qualifications and Occupations", https://ec.europa.eu/esco
[9] C. Bizer, T. Heath, and T. Berners-Lee, "Linked Data—The Story So Far," Int'l J. Semantic Web and Information Systems, vol. 5, no. 3, 2009, pp. 1–22
[10] NACE Rev. 2 - Statistical classification of economic activities. http://epp.eurostat.ec.europa.eu/portal/page/portal/nace_rev2/introduction
[11] UNESCO Institute for Statistics, "Global Consultation for the revised ISCED Fields of Education and Training", http://www.uis.unesco.org/Education/Documents/isced-fos-consultation-draft-2013-en.pdf
[12] A. Miles and Bechhofer, SKOS Simple Knowledge Organization System Reference, World Wide Web Consortium (W3C) recommendation, 18 Aug. 2009; www.w3.org/TR/skos-reference
[13] European Commission, "ESCOpedia", https://ec.europa.eu/esco/escopedia
[14] RDFazer, https://github.com/Rahien/RDFazer
[15] Coursera Catalog APIs, https://tech.coursera.org/app-platform/catalog/
[16] Udacity Course Catalog API, https://www.udacity.com/catalog-api