

# A Linked Data approach to evaluate Open Education Resources

Davide Taibi, Giovanni Fulantelli  
Institute for Educational Technology  
National Research Council of Italy  
Palermo, Italy

[davide.taibi@itd.cnr.it](mailto:davide.taibi@itd.cnr.it), [giovanni.fulantelli@itd.cnr.it](mailto:giovanni.fulantelli@itd.cnr.it)

Stefan Dietze, Besnik Fetahu  
L3S Research Center  
Hannover, Germany  
[dietze@l3s.de](mailto:dietze@l3s.de), [fetahu@l3s.de](mailto:fetahu@l3s.de)

**Abstract**— The social environments based on the Web 2.0 paradigm have modified the way people behave on the Web. One of the direct consequences of this change is that the amount of online resources produced and shared by users has increased considerably. Amongst them, it is possible to find materials that can be exploited for educational purposes. For example, YouTube, Flickr, Slideshare, more and more often collect resources that can be used in educational contexts. In this scenario, finding methods to support the evaluation of the educational relevance of online resources is becoming one of the greatest challenges faced by the educational technologists today. In this paper we propose an approach for the evaluation of the relevance of educational resources based on recent advancements of Linked Open Data.

**Keywords**— *Linked Open Data, OER, Educational resource relevance*

## I. INTRODUCTION

The use of Open Educational Resources (OER) in learning settings, at different educational levels, has become more pervasive in the last years. This trend is mainly due to the availability of a really huge number of OER repositories as well as to the technological maturity reached by standards used to describe OER metadata.

However, in the last few years the flexibility yielded by Web 2.0 technologies has changed the way repositories are considered: from static collectors of resources to dynamic and social environments able to support exchange of knowledge [1]. The development of the Social Web has emphasized this condition [2] [3] [4], thus supporting social activities within OER repositories: mechanisms for commenting, voting, explicitly encouraging the sharing of resources are more and more often implemented in OER repositories. At the same time, the barriers between specialized OER repositories and generalist Web 2.0 repositories, such as YouTube, Flickr, Slideshare, are starting to fade away, so that the social environments have included specific areas to store educational resources, where academic institutions are publishing their OER (in addition to their official repositories).

Educational resources are therefore not only hosted by repositories specifically designed for promoting sharing of

learning materials (usually supported by universities or research institutions), but even by general-purpose Web 2.0 sharing environments which represent a meaningful source of educational resources.

In this scenario, new needs have raised up. In fact, it is necessary to embrace an integrated methodology to provide a channel to access learning resources distributed in a plethora of digital sources on the Web. At the same time, it is necessary to develop automated ways to analyze and evaluate the huge amount of learning resources in order to support teachers activities in selecting the resources suitable for their course [5][6].

Concerning the first aspect, the technological advancements of the Web of Data, facilitated particularly by the Linked Data (LD) approach, provides the technological background to support reuse and sharing of knowledge across the Web.

Recent efforts on using Linked Data approaches for fostering educational data sharing have been widely documented in [7]. Furthermore, the huge amount of data released from the Linked Universities movement (e.g., from The Open University UK or Oxford University) demonstrates the increasing interest of the academic world towards the Linked Data paradigm. However, these approaches so far have mainly focuses on resources explicitly designed for educational purposes., while the exploitation of the vast body of knowledge not directly related to learning has not yet been considered widely.

The Social Web offers new opportunities to tackle the second issue raised before: social activities such as rating, liking and commenting, can activate mechanisms to infer relevant clues about the educational relevance of resources on the Web. This issue is more acute in emerging countries in which the digital divide between students and teachers especially in higher education is more evident. In fact, students have good English skills and are able to use social media environments not only for leisure but also for learning activities, while teachers are not so experienced with the use of these tools for learning [8]. In this scenario students' social activities could be exploited to support teacher in selecting educational contents.

In the framework of the LinkedUp project<sup>1</sup>, several data curation activities have been carried out, aimed at assessing, cataloging, annotating and exposing all sorts of Web data of educational relevance. In this project a new RDF schema has been developed aiming at aligning disparate schemas used by different datasets. In addition, to enable an initial classification of different resource types, a vocabulary of educational resource types has been introduced to describe the properties of each dataset. Finally, a preliminary dataset catalog (named Linked Education catalog), adopting the VoID vocabulary for the description, cataloging and annotation of educationally relevant datasets, has been established.

The Linked Education catalog includes datasets of potential educational relevance. The existing dataset annotations often do not facilitate a comprehensive understanding of the underlying data, and the information provided by public registries or endpoints, such as the DataHub, are often insufficient to deduce the actual nature of the underlying distributed data. Additional knowledge is required about aspects such as:

- representation of resource types
- representation of topics in the datasets
- evaluation of educational relevance of learning resources
- assessment of the data quality

With respect to the first point, the representation of resource types, d'Aquin et al. in [9] analyzes and collects datasets explicitly referenced to education in which resources are exposed as linked data. The main aim of this work has been to analyze how the datasets are connected to each other, considering not only direct links between them but also the reuse of common vocabularies and types.

The representation of the topics covered by the datasets has been conducted by considering an approach which exploits the data enrichment capability of DBpedia spotlight tool, and the relations between concepts and categories in DBpedia. The evaluation of the educational relevance of the learning resources and the assessment of the data quality are in a preliminary stage.

In this paper, we discuss how the strategies to evaluate the educational relevance of OER are modified by the social web dynamics and by the linked data approach to the management of knowledge. In the next section, we introduce the potentials of social web and linked data to create the models that can be used to evaluate educational resources. Then, we illustrate an example of application of the enriched models. Finally, we conclude by pointing out the future work to be done in order to fully exploit the remarkable potentials of Social Semantic Web for OER.

## II. EDUCATIONAL RESOURCE RELEVANCE

Since 2005, Duval [10] has highlighted that “quality is not a binary property of learning material but it is multi-faceted”;

in particular, he has stressed the importance to consider the learning context in the definition of a “quality” measure for learning objects. He has proposed the LearnRank function to assign a relevance rank to learning object, this function is based on the similar elaborations of Pagerank, a well-known algorithm used by search engines to evaluate the relevance of web pages. The definition of learning context involves aspects related to: learning motivation, learning setting, location, educational level, and so on. Ochoa [11][12] has studied the relevance ranking metrics for Learning objects; his research has mainly focused on Learning Objects and how they are managed in Learning Objects repositories. Ochoa, following Duval, considers relevance of LOs as a multi-dimensional problem, and he has elaborated a model to transform the relevance concept into a numerical value. The evaluation of educational resources in the mentioned studies has been based on three main aspects: Human review, Text similarity, User profile.

In the last few years, social media environments have spread and evolved, so that they have been growingly considered valuable sources for learning materials. This “social” evolution of the Web has favored the diffusion of non-institutional sources of resources that can be used in educational settings, thus increasing the number of available resources

At the same time, the Linked Open Data movement has pointed out the potentials of connecting data. Consequently, also the institutional repositories have added functionalities to open up their data.

Finally, the semantic web and the Linked Open Data have changed the way knowledge is represented and managed, by providing datasets connected according to semantic relationships at an extraordinary pace.

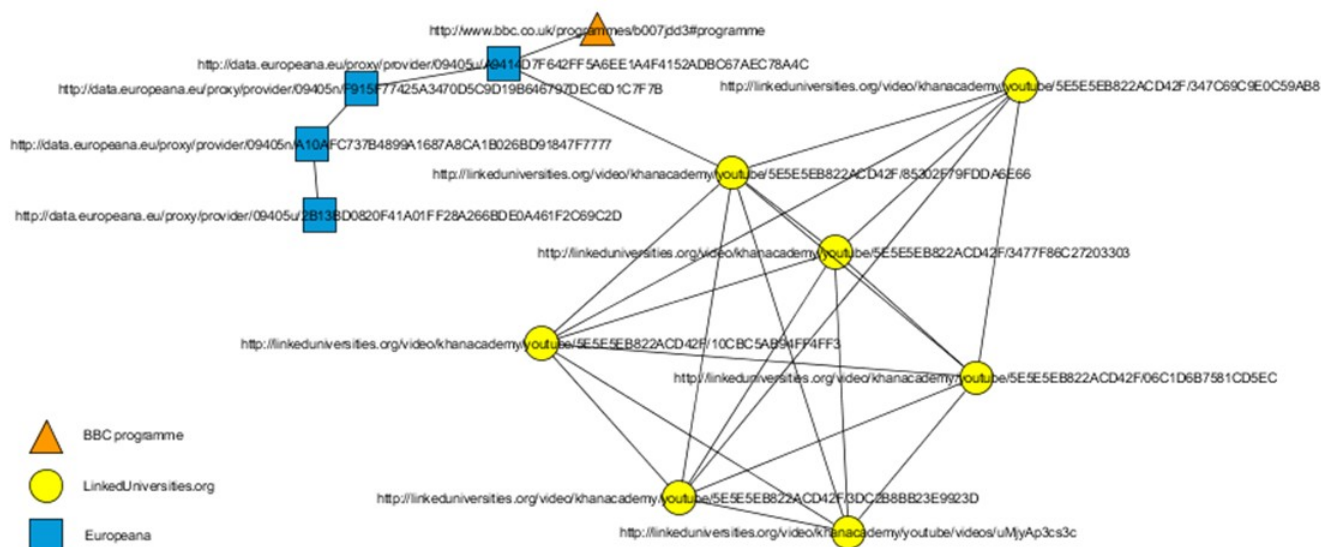
The popularity of social media environments in the web 2.0 context, and the increasing publication of datasets in the background of the semantic web, have led to an evolution and improvement of the previous models to evaluate the relevance of Educational Resources. In particular, the three main factors these models were based on - Human review, Text similarity, User profile – have been deeply influenced by the Social Semantic Web.

Specifically to the Human review aspect, studies have highlighted how peer-review mechanisms in institutional repositories have some drawbacks mainly due to:

- Only a few resources can be evaluated for economic reasons;
- Evaluation is based on the judgment of very few people, so it is high the risk of having personal opinions influencing the final ranking;
- The process is not scalable.

Peer review mechanism in Merlot is exemplary of the difficulties met by a very important institution to perform the human review strategy.

<sup>1</sup> <http://linkedup-project.eu/>



**Figure 1:** A network of learning resources

By changing the perspective and considering the social environments instead of the institutional repositories, the workload for the human review mechanism is not on a limited group of experts, but it is shared amongst all the members of the community. Despite a potential loss of quality in the evaluation process, the scalability and subjectivity issues are solved by this approach. Furthermore, the Wikipedia model has proved that if the community is large enough, also the quality can be improved [13].

Concerning the second mechanism used to evaluate educational resources, the text similarity, the repositories usually implement vector space representation for metadata of learning resources to perform an evaluation of the relevance of resources respect to the learning context in which the resource is used. This type of representation is traditionally adopted in information retrieval activities, and has been successfully used in the early web search engines. Nevertheless, the evolution of the semantic web has permitted the evolution of techniques based on concept similarity rather than text similarity.

Finally, the user profile strategy is traditionally developed by making users to select explicitly the topics of their interest from a taxonomy, which must be the same adopted to classify the resources. Recent work [14] has focused on the definition of the learner portfolio using semantic web approaches, in order to represent students' learning objectives, learning outcomes, learning curriculum. The semantic representation can be used to connect students' profile to the educational resources needed to reach the educational objectives.

### III. EVALUATING EDUCATIONAL RELEVANCE: A LINKED DATA APPROACH

In the framework of the LinkedUp project both explicitly educational datasets (such as OpenLearn<sup>2</sup> and the mEducator

Educational Resources<sup>3</sup>) as well as implicitly educationally relevant datasets (such as BBC Programmes<sup>4</sup> and Europeana<sup>5</sup>) have been selected in order to create the Linked Education catalog, as stated in the introduction. The selection took into consideration the heterogeneity of the data with the aim of creating an integrated dataset which combines a wide variety of educational as well as educationally related resources. Additionally, given the diversity of available resources, a *classification* of the available datasets, indicating their *main purpose*, *nature* and *educational relevance* was required. In order to address these issues, in [15] we presented an *approach*, composed by a set of data processing techniques to enable cross-dataset analytics. Our approach takes advantage of established datasets such as DBpedia and Freebase by exploiting the DBpedia Spotlight API<sup>6</sup>. Initially, our current implementation enriches resource titles and descriptions, as these are the most frequently used properties. This approach has been exploited to create associations between the unstructured title and description of each resource to DBpedia concepts, so that the unstructured text has been enriched through DBpedia entities.

These generated enrichments have been used to create a weighted graph, in which resources are the nodes and the edges are weighted taking into consideration the shared enrichments between resources. Figure 1 shows a part of the resulting network including example resources from BBC, Linked Universities and Europeana datasets.

In the proposed approach the properties of the network (such as degree, centrality measures) can be evaluated to provide a relevance measure of the learning resources involved in the network. Moreover, votes, comments, number of visualizations, and other social activities related to each

<sup>3</sup> <http://datahub.io/it/dataset/meducator>

<sup>4</sup> <http://backstage.bbc.co.uk/>

<sup>5</sup> <http://data.europeana.eu/>

<sup>6</sup> <http://spotlight.dbpedia.org/>

<sup>2</sup> <http://openlearn.open.ac.uk/>

resource can also be used to weight the relationships between resources affording a relevance index that take into consideration both social and semantic aspects.

#### IV. CONCLUSIONS

The educational relevance of Open Educational Resources have been widely debated in the literature, and in the last few years the interest in educational data mining and learning analytics approaches has rapidly increased. Consequently, metrics to measure different aspects of educational processes, and in particular relevance ranking metrics for educational resources, have been proposed. The evolution of the Social Semantic Web has led to reconsider such metrics by introducing further dimensions to capture the new dynamics of education. In this paper we have proposed an approach based on the recent advancements of the Linked Open Data to highlight the “conceptual learning context” related to educational resources. This is a preliminary work, and we are going to test the concepts presented in this paper in the near future, in order to validate them.

#### ACKNOWLEDGMENT

This work has been developed in the framework of the LinkedUp – Linking Web Data for Education Project, partly funded by the European Union under FP7 Grant Agreement No 317620.

#### REFERENCES

- [1]. Iiyoshi T. and Vijay Kumar M.S. (2008). Introduction in Opening Up Education: The Collective Advancement of Education through Open Technology, Open Content, and Open Knowledge. Ed. Iiyoshi & Vijay Kumar (The Carnegie Foundation for the Advancement of Teaching / The MIT Press, 2008).
- [2]. Breslin J. and Decker, S. (2007). The Future of Social Networks on the Internet: The Need for Semantics. *Internet Computing*, 11, (6), 2007, 86-90.
- [3]. Zaidatun Tasir, Yahya Mohammed Hashem Al-Dheleai, Jamalludin Harun, & Nurbiha A. Shukor. 2011. Students Perception towards the Use of Social Networking as an elearning Platform. *Proceeding of The 10th WSEAS International Conference on EDUCATION and EDUCATIONAL TECHNOLOGY (EDU'11)*, Penang, 2 – 6 October 2011.
- [4]. Mchichi Tarik and Afdel Karim. 2011. The use of web 2.0 innovations on education and training. In *Proceedings of the 2011 international conference on Computers and computing (ICCC'11)*, Vladimir Vasek, Yuriy Shmaliy, Denis Treck, Nobuhiko P. Kobayashi, and Ryszard S. Choras (Eds.). World Scientific and Engineering Academy and Society (WSEAS), Stevens Point, Wisconsin, USA, 140-145.
- [5]. Miguel Vargas, Manuel Ortega. 2006. A Framework for Learning Object Evaluation (FLOE). *Proceedings of the 5th WSEAS International Conference on E-Activities*, p.90-97, November 20-22, 2006, Venice, Italy.

- [6]. Juhaida Abdul Aziz , Rosseni Din , Parilah M. Shah , Rashidah Rahamat , Siti Zuraida Manaf , Mohamed Amin Embi. 2010. Collaborative filtering for teaching in a learning 3.0 environment. *Proceedings of the 10th WSEAS international conference on E-Activities*, p.83-88, December 01-03, 2011, Jakarta, Indonesia.
- [7]. Dietze, S., Sanchez-Alonso, S., Ebner, H., Yu, H., Giordano, D., Marenzi, I., Pereira Nunes, B (2013). *Interlinking educational Resources and the Web of Data – a Survey of Challenges and Approaches*, Emerald Program: electronic Library and Information Systems, 47, (1), 2013.
- [8]. Glennie, J., Harley, K., Butcher, N., and van Wyk, T, (eds.). *Perspectives on Open and Distance Learning - Open Educational Resources and Change in Higher Education: Reflections from Practice*. Commonwealth of Learning - UNESCO, 2012.
- [9]. d'Aquin, M., Adamou, A., Dietze, S., *Assessing the Educational Linked Data Landscape*, *ACM Web Science 2013 (WebSci2013)*, Paris, France, May 2013.
- [10]. Duval, E. (2005). *Policy and Innovation in Education – Quality Criteria*, chapter *LearnRank: the Real Quality Measure for Learning Materials*, pages 457–463. European Schoolnet.
- [11]. Ochoa X. (2011). *Learnometrics: metrics for learning objects*. In *Proceedings of the 1st International Conference on Learning Analytics and Knowledge (LAK '11)*. ACM, New York, NY, USA, 1-8. DOI=10.1145/2090116.2090117 <http://doi.acm.org/10.1145/2090116.2090117>
- [12]. Ochoa, X. (2008), 'Learnometrics: Metrics for Learning Objects', PhD thesis, University of Louvain.
- [13]. Giles, Jim. (2006). *Internet Encyclopedias Go Head to Head*. *Nature*, 438.15. (15 Dec. 2005): 900–901. Web. 28 April 2006
- [14]. Taibi D., Gentile M., Fulantelli G., Allegra M., (2010). *An Ontology to Model e-portfolio and Social Relationship in Web 2.0 Informal Learning Environments*, *International Journal of Computers Communications & Control*, ISSN 1841-9836, 5(4):578-585, 2010.
- [15]. Taibi D., Fetahu B., Dietze S. (2013). *Towards Integration of Web Data into a coherent Educational Data Graph*, *ACM Proceedings of Linked Learning 2013: the 3rd International Workshop on eLearning Approaches for the Linked Data Age*.

### **Creative Commons Attribution License 4.0 (Attribution 4.0 International, CC BY 4.0)**

This article is published under the terms of the Creative Commons Attribution License 4.0

[https://creativecommons.org/licenses/by/4.0/deed.en\\_US](https://creativecommons.org/licenses/by/4.0/deed.en_US)