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, giovanni.fulantelli}@itd.cnr.it

Stefan Dietze, Besnik Fetahu
L3S Research Center
Hannover, Germany
{dietze, 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, 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 paper, we discuss the strategies to evaluate the educational relevance of OER. 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 objects, 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.

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”;

1 http://linkedup-project.eu/
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 a 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 OpenLearn2 and the mEducator Educational Resources3) as well as implicitly educationally relevant datasets (such as BBC Programmes4 and Europeana5) 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 API6. 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

---

1. OpenLearn: http://openlearn.open.ac.uk/
2. mEducator: http://datahub.io/it/dataset/meducator
3. BBC Programmes: http://backstage.bbc.co.uk/
4. Europeana: http://data.europeana.eu/
5. DBpedia Spotlight API: http://spotlight.dbpedia.org/
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