Computers & Education 117 (2018) 31–49

Contents lists available at ScienceDirect

Computers & Education

journal homepage: www.elsevier.com/locate/compedu

Cross-repository aggregation of educational resources

Marcos Mouriño-García[a], Roberto Pérez-Rodríguez[a], Luis Anido-Rifón[a], Manuel J. Fernández-Iglesias[a,*], Víctor M. Darriba-Bilbao[b]

[a] Dept. of Telematics Engineering, University of Vigo, Campus Universitario, 36310 Vigo, Spain
[b] Dept. of Informatics, University of Vigo, Campus As Lagoas, 32004 Ourense, Spain

ARTICLE INFO

Keywords:
Architectures for educational technology systems
Cooperative/collaborative learning
Distributed learning environments
Media in education

ABSTRACT

The proliferation of educational resource repositories promoted the development of aggregators to facilitate interoperability, that is, a unified access that would allow users to fetch a given resource independently of its origin. The CROERA system is a repository aggregator that provides access to educational resources independently of the classification taxonomy utilized in the hosting repository. For that, an automated classification algorithm is trained using the information extracted from the metadata of a collection of educational resources hosted in different repositories, which in turn depends on the classification taxonomy used in each case. Then, every resource will be automatically classified on demand independently of the original classification scheme. As a consequence, resources can be retrieved independently of the original taxonomy utilized using any taxonomy supported by the aggregator, and exploratory searches can be made without a previous taxonomy mapping. This approach overcomes one of the recurring problems in taxonomy mapping, namely the one-to-none matching situation. To evaluate the performance of this proposal two methods were applied. Resource classification in categories existing in all repositories was automatically evaluated, obtaining maximum performance values of 84% (F_1 score), 87.8% (area under the receiver operator characteristic curve), 86% (area under the precision-recall curve) and 75.1% (Cohen's κ). In the case of resources not belonging to one of the common categories, human inspection was used as a reference to compute classification performance. In this case, maximum performance values obtained were respectively 69.8%, 73.8%, 75% and 54.3%. These results demonstrate the potential of this approach as a tool to facilitate resource classification, for example to provide a preliminary classification that would require just minor corrections from human classifiers.

1. Introduction

Open educational resources (OER) are educational materials in digital form that are freely available to educators, students and self-learners to be used and re-used in learning, teaching and research (UNESCO, 2002). One of the biggest challenges about OER is access (D'antoni, 2006), that is, how to make potential users aware of the existence of open educational resources to be utilized in their educational projects. Instruments to promote the use of open educational resources and to facilitate access to them were developed, including virtual learning environments, thematic portals, virtual communities, wikis, open magazines, social networks and repositories. Educational resource repositories are the most widespread platform because they offer benefits such as the preservation and reuse of content, permanent access, visibility, and ease of search and retrieval using metadata (Gibbons, 2009). There is

* Corresponding author.
E-mail addresses: marcos@gist.uvigo.es (M. Mouriño-García), roberto.perez@gist.uvigo.es (R. Pérez-Rodríguez), lanido@det.uvigo.es (L. Anido-Rifón), manolo@det.uvigo.es (M.J. Fernández-Iglesias), darriba@uvigo.es (V.M. Darriba-Bilbao).

http://dx.doi.org/10.1016/j.compedu.2017.09.014
Received 15 December 2016; Received in revised form 27 September 2017; Accepted 29 September 2017
Available online 05 October 2017

a great number and variety of online repositories (Roy, Sarkar, & Ghose, 2010) such as OERCommons (Open Educational Resources, 2013), MERLOT (Cafolla, 2006), Open Stax CNX, Edna or Lornet among others (Ternier et al., 2009).

Metadata are the key elements for repositories to represent and organize educational resources. Due to the huge amount of learning objects available, manual generation of metadata is not feasible as a general solution. Note that this would be a resource-consuming process, and metadata created by humans are bound to include errors (Meire, Ochoa, & Duval, 2007). As a consequence, automatic metadata generation techniques were proposed. Although some metadata is relatively straightforward to obtain (e.g., date of creation, original source of the resource) other metadata elements are more complicated to create. Thus, significant efforts have been made in order to automatically generate high quality metadata (Broisin, Vidal, Meire, & Duval, 2005; Meire et al., 2007; Rodriguez, Bollen, & Sompel, 2009).

Each of the repositories above collects resources according to a particular metadata schema. Examples of these schemas (Anido-Rifón et al., 2014) are IEEE LOM, ISO/IEC MLR, or DublinCore Metadata Initiative, as well as several application profiles such as LRE Metadata Application Profile (Massart, Shulman, & Van Assche, 2011), Open Discovery Space Application Profile (Niemann et al., 2013) or UK LOM Core (Campbell, 2011) among others.

Due to the large number of existing repositories and providers, several repository alliances, networks or aggregators were deployed to promote the sharing and reuse of educational materials, such as ARIADNE (Ternier et al., 2009), MACE (Boeykens, Santana Quintero, & Neuckermans, 2009), MELT (Kurilovas & Dagiene, 2009), Edutella (Nejdl et al., 2002), GLOBE, ELENA (Dolog, Henze, Nejdl, & Sintek, 2004), LRE, Open Discovery Space (Nikolas, Sotiriou, Zervas, & Sampson, 2014) or PROLEARN (Wolpers & Grohmann, 2005). The main challenge in repository aggregation is interoperability, that is, a unified and integrated access to the collected resources independently of the underlying metadata scheme or application profile (Stefaner et al., 2007).

Indeed, the heterogeneity of the classification approaches in existing repositories (Dietze et al., 2012) led to the emergence of several methods to overcome it, including the most commonly used metadata mapping techniques. One of the recurring problems is related to the different degrees of equivalence encountered when mapping individual elements, namely one-to-one, many-to-one and one-to-none. As a consequence, in many cases there do not exist exact equivalents between elements, and meaning and scope superpositions occur in some cases (Chan & Zeng, 2006), while in others an existing equivalence will not be found by the mapping engine (Hillmann & Westbrooks, 2004).

Aggregators allow users to search and retrieve resources in different ways. Simple search is used to fetch resources according to the keywords provided, while advanced search supports resource filtering according to specific values of metadata elements. Finally, browsing or exploratory search enables users to navigate a category tree to access specific elements (Neven & Duval, 2002; Roy et al., 2010).

Despite being a desirable feature, not all aggregators implement exploratory search or browsing, as it would require all resources being classified according to a common set of categories or taxonomy. This becomes an issue when trying to aggregate repositories implementing different taxonomies (e.g., OERCommons and MERLOT). Approaches exist that try to address this situation, such us the definition of a new common taxonomy (Kawase et al., 2013), or the application of ontology mapping, ontology matching or ontology alignment techniques (Doan, Madhavan, Domingos, & Halevy, 2004, pp. 385–403), which in turn have several drawbacks. For example, they use to be performed manually, a laborious and error-prone process (Doan et al., 2004, pp. 385–403). To try to overcome this, several authors have proposed the introduction of machine learning (ML) techniques (Sebastiani, 2002) to automatically generate mappings between ontologies (Doan et al., 2004, pp. 385–403; Nezhadi, Shadgar, & Osareh, 2011; Shvaiko, 2013). However, in a similar way to metadata mapping, ontology mapping techniques perform mappings between individual elements, so the different degrees of equivalence mentioned above (i.e., one-to-one, many-to-one, and one-to-none) are not addressed.

The proposal discussed in this paper follows a different approach. While ontology mapping techniques act at the taxonomy level by computing equivalences between different nodes (cf. Fig. 1a), the proposed solution processes educational resource metadata elements in each of the repositories (e.g., title, description, keywords) to classify the target resource again according to each of the taxonomies in existing repositories (Fig. 1b).

Thus, CROERA (Cross-Repository Open Educational Resources Aggregation) is an approach to the aggregation of repositories that has the features below:

- It provides access to educational resources independently of the taxonomy used by each of the integrated repositories. The aggregated resources are classified automatically using ML techniques according to each of the taxonomies of the repositories integrated. This allows users to access resources in several repositories no matter their taxonomies using a taxonomy with which the user is most familiar, or that is considered as the most appropriate or useful in that situation.
- It addresses the heterogeneity of categories, thus enabling browsing (i.e., exploratory searching) through any of taxonomies included in the aggregator.
- It does not require any mapping or matching technique for metadata or ontologies, thus avoiding the drawbacks of such techniques, in particular the one-to-none matching situation.

To evaluate the performance of the resource classifier, two approaches were followed. On the one hand, the quality of the classification of the resources classified under common categories is addressed automatically. On the other hand, in order to assess the quality of the classification of those resources not belonging to the set of common categories, human experts were used.