Computers in Human Behavior xxx (2015) xxx–xxx

Contents lists available at ScienceDirect

Computers in Human Behavior

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

Adopting the metadata approach to improve the search and analysis of educational resources for online learning

Maria-Cruz Valiente, Miguel-Angel Sicilia *, Elena Garcia-Barriocanal, Enayat Rajabi

Computer Science Department, University of Alcala, Polytechnic Building, Ctra. Barcelona km. 33.6, 28871 Alcala de Henares, Madrid, Spain

ARTICLE INFO

Article history:
Available online xxxx

Keywords:
Collaborative learning
e-learning innovations
Knowledge Society
Metadata-supported learning
Teaching/learning strategies

ABSTRACT

The emergence of large collections of learning resources created through the harvesting and aggregation of metadata raises important concerns on the suitability of educational resource descriptions as provided in metadata schemas. For learning purposes, both teachers and students usually seek information on their own, and the vast majority of the search that they do in search engines like Google is driven by multiple keywords or classifications. Therefore this type of metadata-based learning resources could help them obtain better results related to the educational resources they are looking for and provide the basis for collaborative learning environments which enable knowledge sharing and reuse in terms of web-based search systems. This paper reports an exploratory study based on the availability and suitability of keywords and classifications in metadata-based educational resources to improve collaborative learning between teachers and students through the search and analysis of learning resources from a large sample obtained from the Global Learning Objects Brokered Exchange (GLOBE).

1. Introduction

The adoption and implementation of e-learning innovations in the context of the Knowledge Society have become more demanding in recent years (García, Colomo, & Lytras, 2012a; Lytras, 2010; Lytras & Ordóñez de Pablos, 2011) as well as the empowerment of teachers to work with their students and other teachers in order to share and reuse educational resources (Damiani, Lytras, & Cudré-Mauroux, 2010; García, Colomo, & Lytras, 2012b). In this context, the proliferation of educational resource repositories of different kinds has raised the need to aggregate the descriptions of resources into larger collections, thereby providing a critical mass for users, especially for learners, with educational needs that may not be confined to a single thematic repository. Since metadata defines the set of properties that educational resources should include for their retrieval, the use, for example, of the *Learning Object Metadata* (LOM) standard (IEEE 1484.12.1, 2002), combined with harvesting protocols (i.e. protocols for the collection of metadata from repositories) such as the *Open Archives Initiative Protocol for Metadata Harvesting* (OAI-PMH) (Open Archives, 2008), has facilitated the deployment of such collections for collaborative learning purposes. LOM defines the structure of a metadata instance that a learning object should conform (IEEE 1484.12.1, 2002). In this context, a learning object can be defined as any digital resource that can be reused to support the learning process (Wiley, 2000). From here on, we use the term 'learning object' or 'learning resource' to refer to the LOM-based educational resource and the term 'educational resource' to refer to the resource itself. Fig. 1 shows a general overview of the learning object model defined by the LOM standard. OAI-PMH is a low-barrier mechanism for repository interoperability, that is, the standard used to harvest the metadata from other repositories where the data providers are the repositories that expose that structured metadata via OAI-PMH. Then, service providers will make OAI-PMH service requests to harvest the metadata involved in the process (Open Archives, 2008). Repositories in this situation represent network accessible servers that can process the different OAI-PMH requests, enabling the sharing and reuse of Web educational resources targeted toward the professional growth of teachers and improved learning for students.

However, the aggregation of heterogeneous collections provided by different learning communities for its application in a common collaborative learning environment (in our particular case, teachers and students) is a complex task to tackle. The variety of learning resources in granularity (i.e. how big a learning object is) and the different technical formats such as video, application and text, makes also difficult to use full-text indexing as commonly used in search systems on scholarly literature collections. In consequence, search engines tend to rely on indexing metadata instead

* Corresponding author.
E-mail addresses: maricruz.valiente@uah.es (M.-C. Valiente), msicilia@uah.es (M.-A. Sicilia), elena.garciab@uah.es (E. Garcia-Barriocanal), enayat.rajabi@uah.es (E. Rajabi).

http://dx.doi.org/10.1016/j.chb.2014.12.059
0747-5632/

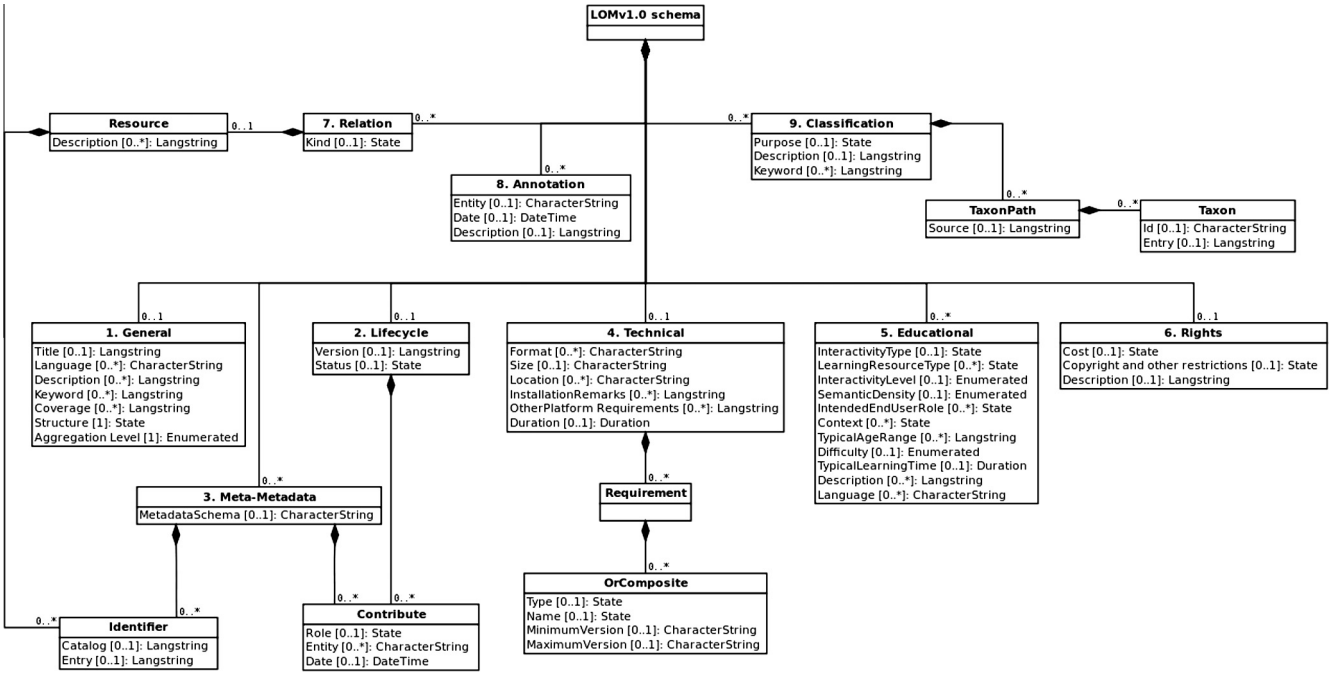

Fig. 1. UML class diagram representing an overview of the learning object model defined by the LOM standard. *Source:* http://en.wikipedia.org/wiki/Learning_object_metadata.

of the contents themselves. This raises the question on how effectively different metadata elements properly describe and categorize the educational resource space. Therefore, we focus on gathering empirical evidence on the keyword and classification spaces of large aggregated collections, how that spaces compare to other description mechanisms and to which extent they could be effectively used to interlink with other Web resources in the context of collaborative learning environments.

In doing so, a large portion of the Global Learning Objects Brokered Exchange (GLOBE[1]) collection was subject to analysis using different methods. GLOBE enables the sharing and reuse of learning objects between different learning resource repositories worldwide, and it is nowadays the most diverse and large collection available openly (GLOBE, 2011). LOM has the problem that it does not establish the elements that should be present in each learning object (i.e. fields are not marked as required, recommended or optional). Therefore, an empty LOM record is a valid LOM record. This characteristic creates several problems when the metadata is shared among different repositories. For this reason, GLOBE defines an application profile that enhances the LOM standard by defining a set of required fields (GLOBE, 2011). An application profile specifies a set of metadata elements selected from one or several metadata schemas which are combined for the definition of a new domain-specific schema (GLOBE, 2011). In this scenario, the LOM-based general keyword field that describes the learning content is established as a recommended field, and the LOM-based classification space is basically optional apart from the textual label of the classification taxon which is defined as a mandatory field. Both, general keywords and the classification space are the two key items that we want to analyze in order to determine the availability and suitability of keywords and classifications as a searching tool for collaborative learning domains where teachers and students are involved.

There have been some empirical studies on the actual use of LOM metadata. For example, Friesen (Friesen, 2004) collected samples from several international repositories to study the real use of LOM remarking the fact that the potential value was not being realized. After that one, Ochoa, Klerkx, Vandeputte, and Duval (2011) became the study with a largest empirical base, up to 630,317 metadata instances from GLOBE. In their study, the most comprehensive study to date on the use of LOM for heterogeneous resource collections, the *Keyword* element was found to be used in more than 55% of the metadata instances, and with the *Taxon* element around 60%.

However, none of the existing studies on the use of LOM metadata analyzes the availability and suitability of metadata for educational purposes. This paper represents an extension of the work done by Sicilia, Sanchez-Alonso, Garcia-Barriocanal, Minguillón, and Rajabi (2013). Our intention is to carry out an enhanced and deeper study aimed at the improvement of collaborative learning in terms of metadata-based educational resources.

1.1. Metadata

Metadata is generally defined as "data about data" although is better understood as "any statement about an information resource", regardless of any specific domain (Garshol, 2004). In computer science, the concept of metadata is usually understood as the description of information regarding objects on the network. In our context, these objects will be learning objects that will support collaborative learning through web-based search systems.

One of the most well-known vocabularies for metadata is Dublin Core (Weibel, Kunze, Lagoze, & Wolf, 1998), which is composed by 13 properties that can be used to describe information resources. Examples of these properties are "title", "creator", "subject", "description", "publisher", "date" and "language" as shown in Table 1.

Besides, metadata could be categorized into five types depending on their functionality (Baca, 2008): *administrative*, *descriptive*, *preservation*, *technical* and *use*. Metadata categories with the corresponding descriptions and some examples are given in Table 2.

2. Materials and methods

As mentioned earlier, Dublin Core (Weibel et al., 1998) is one of the most well-known ways to describe information resources. The

[1] http://www.globe-info.org.