Improving Searching and Browsing Capabilities of Learning Object Repositories

Julià Minguillón, M. Elena Rodríguez, and Jordi Conesa

Estudis d'Informàtica, Multimèdia i Telecomunicació,
Universitat Oberta de Catalunya,
Rambla Poblenou 156, Barcelona, Spain
{jminguillona,mrodriguezgo,jconesac}@uoc.edu
http://www.uoc.edu

Abstract. Learning object repositories are a basic piece of virtual learning environments used for content management. Nevertheless, learning objects have special characteristics that make traditional solutions for content management ineffective. In particular, browsing and searching for learning objects cannot be based on the typical authoritative metadata used for describing content, such as author, title or publication date, among others. We propose to build a social layer on top of a learning object repository, providing final users with additional services for describing, rating and curating learning objects from a teaching perspective. All these interactions among users, services and resources can be captured and further analyzed, so both browsing and searching can be personalized according to user profile and the educational context, helping users to find the most valuable resources for their learning process. In this paper we propose to use reputation schemes and collaborative filtering techniques for improving the user interface of a DSpace based learning object repository.

Keywords: Learning Object Repositories, Browsing, Searching, Recommendation Systems, Collaborative Filtering, DSpace, Metadata, Paradata

1 Introduction

Since the introduction of Information and Communication Technologies in the field of education, almost every educational institution has adopted an e-learning solution, although with different approaches. Virtual learning environments are one of the most common tools used to implement an e-learning platform, typically including a module for content management. Usually, content is understood as complete courses, but the reality is that learning resources can be very different according to their type, format and granularity [2].

In order to encourage its usage, the most important issue for a learning object repository integrated in a virtual learning environment is being able to build a true social learning network around it, promoting the creation, sharing

and reuse of learning resources among the members of the learning community, mainly both learners and teachers. This can be only done if the learning object repository, regardless of its technology, provides its users with a virtual learning environment and a true learning experience. These learning experiences can be used in order to get knowledge about the needs of learners, their preferences in the use of learning objects and to identify learning objects in the repository that are less attractive to them. All this information can be used in order to improve the repository by updating the descriptions of learning objects according to what we learnt (making the appropriate learning object easier to find through additional metadata [3]), improving the services of the repository itself (making it easier to use) and improving the quality of the institutional repository by deleting or improving the irrelevant resources and promoting the more useful ones. The learning object repository is an important element of the virtual learning environment but it is not the only one and, of course, learners may search for resources outside the institutional "walled garden", mainly through Google and other search engines. Nevertheless, the main problem for learners is to filter among the thousands of results returned by a general purpose search engine. We encourage our learners to use the institutional repository as part of their learning process.

This paper is organized as follows: Section 2 describes the use of learning object repositories as an important tool for supporting both learners and teachers. Section 3 describes the functionalities of an ideal learning object repository which uses a social layer on top of it to improve searching and browsing capabilities. Finally, Section 4 outlines the main advantages of the proposed system, the implementation issues and the current and future research topics of this project.

2 Learning Object Repositories

Learning objects are stored in learning object repositories, which can be considered a specific kind of content management system for educational resources but much more versatile [7]. Although, as stated before, traditional CMS tools can be used to store, describe and share learning objects (such as Drupal or OpenCMS, among many other open source software tools), these tools are usually oriented towards web content. According to [5], repositories are differentiated from other digital collections because the content is deposited in the repository together with its metadata; and such content is accessible through a basic set of services (i.e. put, get, search, etc.). Depending on the specific needs of the community using the repository, this will provide additional tailored services, but all repositories should at least provide two basic ones: content preservation and content reusing [1].

Obviously, digital repositories are a way to organize learning objects (and their parts that can be processed separately) in collections, although there are several specific issues that must be firstly addressed. For example, an exercise (which is basically a text defining a problem and, optionally, its solution, another text which may include references to the use of software or tables with data, for

instance) is a typical learning object. But, differently to classical items in a collection of a digital repository, exercises may have neither a title nor even an author, the two main fields used for finding a book. Other typical learning objects can be data sets, mathematical proofs, equations, simulations, and so. Usually, learners search through these kinds of resources not by title or author, but by keyword or, even better, using a hierarchical taxonomy specially designed. Therefore, it becomes necessary to rethink the traditional way of describing learning resources, using criteria related to the learning process but maintaining a minimum description for archiving purposes. Learning object repositories should be designed for final users, that is, learners and teachers [4], promoting content reutilization rather than preservation. Both concepts (preservation and reutilization) are somehow contradictory (institutional, top-down vs social, bottom-up) but a tradeoff can be achieved by combining digital repositories with web 2.0 services.

2.1 DSpace as a learning object repository

DSpace[1] is an open source platform developed by MIT and Hewlett-Packard in 2002 for creating digital repositories, as initially outlined in [8]. DSpace preserves and enables easy and open access to all types of digital content including text, images, moving images, mpegs and data sets. It is used by more than one thousand institutions and it has a large community of developers, becoming a *de facto* standard for building open repositories.

DSpace organizes resources in a hierarchical structure based on communities (and, recursively, subcommunities) and, finally, collections. These contain the items (i.e. the resources) which are described using a metadata profile, usually non qualified Dublin Core. With the default DSpace user interface, users can search and browse by author, title, publication date and keywords, as well as through the hierarchical structure of communities and collections. But, according to their nature, some learning objects may have or not title, author, creation date, etc., so they cannot be accessed by classical retrieval mechanisms used in digital libraries or repositories. In fact, DSpace has to be customized to change the basic fields used for searching and browsing, as well as all the workflows related to the process of adding new resources to the repository. From a teaching perspective, learners should retrieve resources not from a list of search results but within a specific educational context. DSpace (as any other large collection of resources) suffers from the "Google effect", that is, a search based on a simple keyword such as "Statistics" may return thousands of resources, which is not a good result from a teaching perspective[2]. It is well known that most users click on the first three results (up to 62.53%, see [3]), so it is very important to determine a proper ranking of learning resources, providing learners with the most appropriate content according to their profile and context.

[1] http://www.dspace.org/
[2] See http://dspace-dev.dsi.uminho.pt as an example of recommender system.
[3] http://www.webuildpages.com/jim/click-rate-for-top-10-search-results/

3 Improving browsing and searching

As described in [6], our goal is to provide a layer of web 2.0 services on top of each learning object stored in the repository. These services include adding comments to a learning object, rating it, starring it as a favorite resource, tagging it, sharing it through different social networks and, finally, subscribing to it in order to be aware of all the activity generated around such learning object. All the information generated during the interaction between users, services and resources is stored in each learning object and/or user profile, if available. All this information is known as paradata[4], and it can be used for our adaptation purposes in two different ways: supporting users when browsing and searching (i.e. filtering before finding) and sorting results (i.e. filtering after finding). In our case, paradata is composed of 5-tuples $\{U, S, R, X, T\}$ meaning that user U used service S on resource R with result X in moment T. By means of data mining techniques, these data can be analyzed and reintroduced into the system to enhance browsing and searching capabilities.

3.1 Possible uses of paradata

For each resource, we know the following: the number of times it has been accessed, the number of times it has been downloaded, the number of comments placed on it, the number and average of ratings, the number of times it has been favorited, the number of times it has been shared and the number of users subscribed to it.

On the other hand, for each user, we know the following according to its role (manager, teacher, learner or anonymous visitor). For teachers, we are interested in knowing the list of subjects she is in charge of, as her activity on resources related to those subjects will have an important weight. For learners, we know the list of subjects she is/has been enrolled in, the languages she is competent (and her preferred one), previous and current professional experience (if available) as well as all the repository services used in a period of time. Although it is out of the scope of this paper, all this information could be available by means of specifications like IMS Learner Information Profile, in order to promote interoperability with other e-learning systems.

Once the layer of services is available to all users and the proposed system has been gathering interaction data during an adequate period of time (i.e. an academic semester), it is possible to use such data for computing the heuristics that will be used by the reputation scheme to rank both users and resources, providing useful information about:

- The most popular resource: popularity can be both rank based or activity based, or any combination of both. Not only the most popular resources are interesting, the worst ranked ones need to be analyzed by teachers in order to detect potential problems. On the other hand, it might be useful to identify

[4] http://nsdlnetwork.org/stemexchange/paradata

unused resources. This may lead to detecting wrong metadata descriptions which may cause a resource to be non findable.
- The most active users: analogously, users can be ranked according to their level of activity. In a virtual learning environment, where peer-to-peer learning is promoted as part of the underlying pedagogical model, learners can build a reputation by creating, sharing and answering other peers' questions, all these actions rated by the other users. On the other hand, teachers can act as peer experts in one or more subjects.
- The most common tags used for describing resources: although resources have been already described by both librarians and teachers according to several taxonomies (domain specific keywords, resource type, etc.), users can add their own tags for describing resources as in delicious. These tags can be analyzed and further incorporated into metadata as new keywords, for example.
- Relationships between resources: like Amazon, collaborative filtering can be used to detect which resources can be potentially more interesting according to the implicit navigational behavior of users with similar profiles. This information will be used to determine the most adequate resources related to a given one.

3.2 Improving the user experience

When a user reaches the DSpace repository, there is usually a list of the most recently added resources. We propose to replace the items in this list with the most adequate ones according to her profile (i.e. the most relevant ones with respect to the subjects she is enrolled in). A second list with the resources previously used and/or the most popular is also desirable.

On the other hand, when the same user performs a search and obtains a list of results, we propose to modify two different aspects. Firstly, only a few results are shown (i.e. five to ten), the most important ones according to the underlying reputation scheme which uses both user profile and the available paradata. Secondly, the resources most related to these search results are also shown, promoting a browsing strategy through a local network of resources, updating it according to user's navigational behavior. From a teacher's perspective, this is better than providing learners with hundreds of resources without any contextual support.

4 Discussion

Learning object repositories are a very important piece of any e-learning platform, although they are currently being underused by final users, specially learners. In order to promote repository usage, we propose to add a layer of web 2.0 services on top of the repository, bridging resources to the learning process. The interactions between users, services and resources generate a lot of paradata that

may be captured and further analyzed for personalization purposes. Both searching and browsing can be adapted to user's profile, improving her experience when using the repository.

Nevertheless, there are some well known issues that must be faced regarding to the way learning resources must be described. It is not easy to establish a taxonomy for describing all the resources in an institutional repository. Due to the variety of contents, such a taxonomy would have probably too many levels, making it too complex for most users. On the other hand, learning activities that promote the use of the repository must be also designed, in order to encourage learners to adopt a more active role in their learning process.

This work is part of a three year research project (2011-2013) on analyzing usage of repositories and social networks in virtual learning environments. Currently now we are modifying our DSpace institutional repository[5] in order to include the layer of services as well as the mechanisms for paradata gathering. We expect that in Fall 2011 we will be able to deploy a first version of the repository and start capturing users' interactions. Current and future research lines around this topic include the creation of reputation schemes for both users and resources, using explicit and implicit paradata. Making repository services available from social networks (where students are) is also an interesting possibility.

Acknowledgments. This work has been partially supported by Spanish Government funded project MAVSEL ref. TIN2010-21715-C02-02 and by the IN3.

References

1. Akeroyd, J.: Information management and e-learning. some perspectives. Aslib Proceedings: New Information Perspectives 57(2), 157–167 (2005)
2. Conway, P.: Modeling the digital content landscape in universities. Library Hi Tech 26(3), 342–354 (2008)
3. Ferran, N., Casadesús, J., Krakowska, M., Minguillón, J.: Enriching e-learning metadata through digital library usage analysis. The Electronic Library 25(2), 148–165 (2007)
4. Ferran, N., Guerrero-Roldán, A.E., Mor, E., Minguillón, J.: User centered design of a learning object repository. In: Kurosu, M. (ed.) Proceedings of the 1st International Conference on Human Centered Design, HCI International. Lecture Notes in Computer Science, vol. 5619, pp. 679–688. San Diego, CA (July 2009)
5. Heery, R., Anderson, S.: Digital repositories review. Tech. rep., JISC (2005)
6. Minguillón, J., Rodríguez, M.E., Conesa, J.: Extending learning objects by means of social networking. In: Advances in Web-Based Learning - Proceedings of the 9th International Conference on Web-based Learning. pp. 220–229 (2010)
7. Minguillón, J., Sicilia, M.A., Lamb, B.: From Content Management to E-Learning Content Repositories, chap. Content Management for E-Learning, pp. 27–41. Springer (2010)
8. Smith, M., Barton, M.R., Branschofsky, M., McClellan, G., Walker, J.H., Bass, M.J., Stuve, D., Tansley, R.: Dspace: An open source dynamic digital repository. D-Lib Magazine 9(1) (2003)

[5] http://openaccess.uoc.edu/webapps/o2/?locale=en