Making Efficient Use of Open Educational Resources Using a Multi-Layer Metadata Approach

Birgit Zens & Peter Baumgartner

Abstract: Repositories of educational content enable a worldwide exchange of learning resources and support an economic use of existing materials. Metadata provide information describing objects and are essential for the exchange and re-use of content. Due to a rapidly growing amount of content, the scalability of metadata creation is becoming an important issue. The creation of structured metadata by experts is cost-intensive and does not scale. In contrast, automatically generated metadata are cheap to obtain, but provide only limited information on the information content of objects. A third way of enriching objects with metadata is social tagging, resulting in folksonomies, i.e., collaboratively user-generated, unstructured vocabularies. Combining these three approaches can provide for a cost-effective way of enriching content with useful types of metadata allowing users to find resources that fit their needs. This approach is pursued in the MELT project, funded by the European Commission. Its (cost-)effectiveness is currently being evaluated.

Learning Resource Exchange for Schools

The use of digital educational resources in schools has been growing in recent years. Large repositories of educational content (e.g. the MERLOT initiative)[1] make it possible to exchange digital learning resources worldwide and support an economic use of existing resources. The *Learning Resource Exchange* initiative (LRE)[2] uses a brokerage system that supports federated searching across a network of linked content repositories, including public and private content providers across Europe. The primary goal is to make existing learning resources more widely available for schools across national borders.

A major issue emerging from utilizing such large networks of content repositories is how to enhance the recall and precisely uncover content that meets the users' needs. The solution to this problem is the enrichment of resources with relevant and high quality metadata.

Metadata and the Problem of Scalability

Metadata provide information that describe objects and are defined as "attributes required to fully/adequately describe a Learning Object" (IEEE Learning Technology Standards Committee, 2005)[3]. This information is important for the recall of a given learning object, its evaluation regarding specific tasks, and its eventual use (Ochoa & Duval, 2006). Without appropriate metadata, learning resources are not really reusable.[4]

Metadata can be classified into two groups: *a priori* and *a posteriori* metadata. A priori metadata are created before usage of a given object by professional indexers and authors of learning objects. Posteriori metadata, in contrast, are created after usage by the users themselves or by means of automatic metadata generation (Juzna, Kavcic, & Divjak, 2007).

A priori metadata use several standard formats. The most common learning technology standard is the IEEE LOM (Learning Object Metadata)[5]. The major disadvantage of this standard lies in its complexity, its limited vocabularies, and the lack of mandatory fields (Juzna et al., 2007). In order to increase the practical usability of the IEEE LOM

---

1 http://www.merlot.org/
2 http://lre.eun.org
3 http://ltsc.ieee.org/wg12/
4 http://ariadne.cs.kuleuven.be/amg
5 http://ieeeltsc.org/

data model, the LRE LOM was developed (EUN, 2007)[6], providing an application profile for learning resource exchange that defines mandatory, recommended, and optional elements of LOM and extends the vocabularies for the LRE.

Since these classification systems are very complex, experienced indexers are needed. Thus, the enrichment of content with a priori metadata is very cost-intensive and does not scale. As the amount of learning objects continues to grow rapidly, it becomes less and less feasible to enrich all available learning resources with appropriate metadata. One solution to this problem is automatic metadata generation (Meire, Ochoa, & Duval, 2007).

Automatic Metadata Generation

Automatically generated metadata are created from the content of the learning object and from the context that the learning object is deployed in. These data refer to, e.g., file size, location, time of creation of the learning resource, or language of the object. Automatic metadata generation is a cheap way of mass metadata creation. However, the information of automatically generated metadata with respect to the information content of a given learning object and its practical use is limited.

In contrast, metadata that are added by the users of learning resources provide information about the content of given objects in terms of keywords regarded as salient by the users. These metadata are called *social tags* or *folksonomies*.

Social Tagging and Folksonomies

Social Tagging is regarded as "democratic metadata generation" (Vuorikari, 2007, p. 4) and describes the process of collaboratively adding open-ended labels or keywords, so-called *tags*, to a digital learning resource. Social tagging allows users to add tags to given objects and thus reflects the view of multiple users. It results in many accumulative metadata records related to a given resource. This collaborative metadata generation stands in contrast to expert indexing, which is controlled by one individual or institution (Vuorikari, 2007).

Social tagging results in *folksonomies*. Folksonomies are user-generated taxonomies, i.e., community-generated, unstructured, bottom-up vocabularies that are familiar to the users. The term *folksonomy* is a combination of the terms *taxonomy* (a classification technique) and *folk*.

Folksonomies facilitate the sharing of content within a social network of users and potentially promote an efficient discovery of learning resources that meet the user's needs. Two prominent examples of services using a social tagging approach are Flickr[7] and del.icio.us[8].

MELT: A Multi-Layer Metadata Approach

Combining the three metadata enrichment approaches previously described in this paper, i.e., expert indexing, automatic metadata generation, and collaborative social tagging appears to provide a cost-effective way of enriching learning resources with useful types of metadata that allow users to find resources that fit their needs, language, culture, and preferred ways of teaching and learning.

MELT (A Metadata Ecology for Learning and Teaching)[9] is a Content Enrichment project supported by the European Commission that bridges together 17 public and private sector content partners with the goal of promoting the exchange of learning resources across Europe. MELT uses an existing brokerage system that supports federated

---

6 http://insight.eun.org/intern/shared/data/insight/lre/AppProfilev3p0.pdf
7 http://www.flickr.com
8 http://www.del.icio.us
9 http://info.melt-project.eu

searching across a network of linked content repositories. This technical infrastructure will make MELT content freely available to schools all over Europe[10].

MELT pursues a multi-layer metadata enrichment approach that includes expert indexing, automatic metadata generation, and social tagging. First, expert indexers enrich the MELT content using a LOM-based application profile, the LRE LOM[11], and translate the metadata to at least one other language (mostly English). Second, a new framework for automatic metadata generation is deployed (Meire et al., 2007). Third, teachers are provided with social tagging tools so that they can add their own metadata to MELT content they have used.

To assess the effectiveness of this multi-layer metadata approach, MELT is currently being evaluated.

Evaluation of MELT

Objectives

The effectiveness of the MELT approach will be assessed by using the following main success indicators:

- Effectiveness and efficiency of the search and retrieval process
- Utility of metadata enriched by experts for finding relevant content
- Utility of folksonomies for finding relevant content
- Effectiveness of automatically generated metadata regarding discovery of content
- User satisfaction
- Use of the retrieved content
- Use of content across languages and across countries

Participants

The participants of the evaluation are teachers across Europe[12], mainly from the four pilot countries Belgium, Hungary, Finland, and Austria. In the run-up to the evaluation, a kick-off questionnaire assessed the characteristics of the participants.

Out of a total of 40 individuals, 67.5 percent are male and 32.5 percent are female. Most of the participants are between the ages of 41 to 50 years (53.8%), one third of the participants are below 40 years and 12.9 percent are older than 50 years. The majority of the participants teach at secondary schools (63.2%), 23.7 percent teach at primary schools, and another 23.7 percent teach adults (upper secondary school, teacher training, university level). The taught subjects cover a broad range.

The participants' experience with using the Internet is quite high (M=4.6; SD=0.56)[13]. Experience with using search engines like Google or Yahoo is also high (M=4.5; SD=0.64). Out of a total of 40 participants, 90 percent (36 individuals) indicated that they are familiar with social bookmarking or social tagging. Out of a total of 40 teachers, 95 percent indicated that they use digital learning objects in the classroom, 76 percent thereof often. The teachers obtain the content from regional educational repositories, search engines, school book publishers, and they create the learning objects by themselves. On a five point rating scale ranging from 1 "not at all" to 5 "absolutely", the participants indicated that applying educational content in the classroom is beneficial (M=4.5; SD=0.51).

The motivation of the teachers for participating in the project is quite good. The most frequently indicated reasons for participation were "own interest" and wanting "to be an innovate teacher". Most of the participants expected the project to be fun (M=4.2; SD=0.7)[14].

---

[10] This brokerage system was established by the earlier CALIBRATE project funded by the European Commission.

[11] http://insight.eun.org/intern/shared/data/insight/lre/AppProfilev3p0.pdf

[12] Hungary, Belgium, Finland, Austria, Italy, Spain, Estonia, Germany, Slovenia, and Sweden

[13] Five point rating scale 5 being the highest value.

[14] 5 point rating scale ranging from 1 "not at all" to 5 "absolutely".

Evaluation Methods

The evaluation methodology[15] combines quantitative and qualitative methods. The data sources consist of both explicit feedback of the MELT users, i.e. teachers across Europe, and implicit feedback capturing the users' behaviour by means of logging data. Explicit feedback is assessed using questionnaires and interviews.

Questionnaire on Searching & Finding Content

The *Search & Find Questionnaire* investigates the search and discovery process in a naturalistic environment and on a longitudinal basis. Therefore, each participant is prompted to complete the *Search & Find Questionnaire* several times directly after conducting a relevant content search. The questionnaire contains standardized closed questions and open fields. The Search & Find Questionnaire primarily assesses the effectiveness and efficiency of the search and retrieval process, the utility of metadata enriched by experts, and the utility of folksonomies for finding relevant content, and user satisfaction.

Questionnaire on Using Content

The *Use of Content Questionnaire* examines the use of the discovered learning resources in class. An important question is whether the used content originated from a different language or nationality different from the users' own nationality and language. The *Use of Content Questionnaire* is completed by the teachers for each learning object used in class. The questionnaire contains standardized closed questions and open fields.

Interviews

Additionally, interviews are conducted at two points in time: in the beginning of the tagging activities of the teachers and towards the end of the project. The interviews focus on positive and negative experiences made in the project, i.e., experiences with tagging, problems, challenges, and benefits. The interviews at the second point in time also focus on the use of content in class.

Log Data

The questionnaires and interviews are complemented by the use of log data that provide indicators for discovery and use of objects. The log data will be related to the metadata information of objects that were viewed and saved. Moreover, the log data will be linked to the data derived from the questionnaires.

Outlook

The evaluation is currently in progress. The results are expected to provide information on the effectiveness of the new metadata approach of MELT. A cost-effectiveness analysis will complement the study and take into account the relation of costs and effectiveness for content providers.

References

Juzna, J., Kavcic, A., & Divjak, S. (2007). Metadata for Electronic Learning Resources. *Information Society IS 2007*. Lubljana, Slovenia, October 2007.

Meire, M., Ochoa, X., & Duval, E. (2007). SamgI : Automatic Metadata Generation v2.0. In C. Montgomerie & J. Seale (Eds.), *Proceedings of World Conference on Educational Multimedia, Hypermedia and Telecommunications 2007* (pp. 1195-1204). Chesapeake, VA: AACE.

Ochoa, X. & Duval, E. (2006). Quality Metrics for Learning Object Metadata. In E. Pearson & P. Bohman (Eds.), *Proceedings of World Conference on Educational Multimedia, Hypermedia and Telecommunications 2006* (pp. 1004-1011). Chesapeake, VA: AACE.

Vuorikari, R. (2007). Folksonomies, Social Bookmarking and Tagging: State-of-the-Art. *Insight. Observatory for new technologies and education.* May 2007.

[15] http://eunbrux09.eun.org/shared/data/melt/MELT_D7P1_final.pdf