Metadata for Electronic Learning Resources

Metapodatki pri elektronskih učnih virih

Jernej Južna, Alenka Kavčič, Saša Divjak

University of Ljubljana, Faculty of Computer and Information Science
jernej.juzna@fri.uni-lj.si, alenka.kavcic@fri.uni-lj.si, sasa.divjak@fri.uni-lj.si

Abstract

Over the last decades, extensive catalogues and large repositories of educational content have been made available to schools, teachers and students. As the number of resources in these repositories continues to grow, the demand for better identification of resources is increasing: more precise and detailed tagging of resources is needed in order to enable teachers and learners quickly and easily search, find and access the needed learning materials.

The problem of identification of resources and consequently more accurate search can be solved by using better and richer metadata. Since creation of needed metadata is long-lasting and complex procedure, approaches how to deal with this issue are presented in this article. Social tagging and automated metadata generation are examples of such approaches.

The number of repositories that offer these resources is increasing as well, so extra attention is given to metadata exchange that provides a way to perform federated search. Since exchange between different systems can be made only by using standard formats, an insight of common standards is given as well.

Keywords: *metadata, creating metadata, social tagging, metadata standards, metadata exchange*

Povzetek

Šolam, učiteljem in učencem so na spletu na voljo obsežni katalogi in zbirke elektronskih učnih virov. Medtem ko se število virov v teh zbirkah neprestano veča, narašča tudi potreba po natančnejši identifikaciji teh virov, da lahko omogočimo uporabnikom hitro in enostavno preiskovanje, odkrivanje in dostop do potrebnih učnih materialov.

Problem identifikacije in natančnejšega iskanja rešujejo boljši in kvalitetnejši metapodatki učnih virov. Pridobivanje metapodatkov ni enostavno, saj so običajni postopki kompleksni in dolgotrajni. V članku je tako predstavljenih nekaj novih načinov, kako se omenjenim problemom izogniti, kot sta na primer: družbeno označevanje in avtomatsko pridobivanje metapodatkov.

Hkrati s številom učnih virov raste tudi število ponudnikov teh virov. Problem dostopnosti istega vira pri večih ponudnikih rešuje izmenjava metapodatkov, ki omogoča združeno iskanje. Predstavili smo tudi najpogostejše načine izmenjave metapodatkov in standardne formate zapisov, ki tako izmenjavo omogočajo.

Ključne besede: *metapodatki, pridobivanje metapodatkov, družbeno označevanje, standardi metapodatkov, izmenjava metapodatkov*

1. Introduction and motivation

In Slovenia, extensive collections of learning resources have been created during the recent years, mostly supported by Ministry of Education and Sport. All this content is freely available to teachers and students through the many educational Web portals. At the same time links to similar collections from all over Europe are being established. Rapid growth of quantity of learning materials and increasing number of repositories raise a new dilemma: how to find a learning resource that suits our needs, and where to search for it? The answer is given in well defined, quality metadata that are descriptive enough to cover wide variety of all possible learning resources, and in exchange of these metadata between different portals.

2. Metadata

Metadata are data about data, or in other words, information about the content. Metadata of a learning resource are therefore the data about its title, author, content, license, technical details and location. Since they are well defined, we can say that they provide a way for the computer to "understand" the resource. Therefore, we can achieve better and faster search since we know where to look for. The importance of quality metadata is most obvious in search, as only with high quality metadata we can guarantee that users will be able to find learning resource they are looking for.

3. Creating metadata

The many methods of metadata creation can be divided into two groups depending on who creates the metadata and when they are created. The first group consists of so called *a priori* metadata (before the usage) that are added by authors and professional indexers, whereas the second group is called *a posteriori* metadata (after the usage) that are created by users themselves or by means of automatic metadata generation.

3.1. Authors and professional indexers

Authors are usually most familiar with the resources and may naturally add the basic descriptive metadata, such as the title, author, keywords, description, user manual, etc., preferably already during the material creation. However, it has been shown [15] that the authors are not best suited for classifying a given resource, since the classification systems are usually very complex. In order to correctly assign these advanced metadata, expert indexers have to be employed. They are familiar with classification systems and usually with the portal itself, as in many cases they also publish the resources on the Web. They can also double-check the metadata and make the appropriate corrections.

3.2. Automatic metadata generation

The fastest and cheapest way to generate lots of metadata is by using machines. Unfortunately, only some specific metadata can be automatically created [11]. They are mostly technical data about the size, type, location, time of creation, etc. of the learning resource. Additionally, computers are suitable for monitoring resource usage and its accesses and these data can be used as a good indicator of popularity/quality of the material in question. Regrettably, this approach is not suitable for creating the title, short description, keywords of a given content (or any kind of data that needs understanding of that content), since there are no known algorithms that are precise enough to be fully autonomous.

Nonetheless, automatic metadata generation can be efficiently used as computer assistance when manually filling in the metadata, since it can be very successful in correctly predicting user input. For example, above mentioned algorithms are not suitable for autonomous work, but can still produce a very good result in suggestions. Important conclusions about missing metadata can also be drawn directly from the user profile and current resource metadata. Example of such a computer system is "Automatic Metadata Generation" [1].

3.3. Folksonomy and social tagging

Folksonomy is taxonomy, created by users and used to categorize and retrieve Web content through open-ended labels called tags [9]. Folksonomy is therefore a user created classification system, which is the result of social tagging, where tags are open-ended and shared between users in order to ease the search, discover, and navigate over time. A well-developed folksonomy is ideally accessible as a shared vocabulary that is both originated by, and familiar to, its primary users [16]. Two widely cited examples of Web sites using folksonomic tagging are Flickr [5] and del.icio.us [2].

The main advantage of this system is the freedom of labelling a resource the way that the user wants in order to easily identify and search for it. At the same time, such system creates a social network that can be used to find similar resources or people with the same interests. If, for example, one user discovers that several resources he is interested in were labelled by one person, there is a great chance that the same person has labelled some other resources that are related to what the user is searching for.

However, folksonomy cannot be used as a substitute for taxonomy, but only as an enrichment of current taxonomy [9]. The main problem is that different tags can have the same meaning, or even worse that the same tag can have several different meanings. Additionally, in order for the system to work, we need many users that are willing to add the tags instead of only searching or examining the resources.

4. Metadata formats

The metadata of a given resource can be stored internally, together with the learning resource, or externally, separated from it. In the first case, metadata can be easily transferred along with the resource without a fear of being lost or swapped with others and are always accessible to the user. On the other hand, the second form can be easily integrated into databases which enable fast searches without the need of resources themselves.

There are many standard formats for metadata description, one of the most common in learning environment being the IEEE LOM (Learning Object Metadata) [6]. LOM standard is a data model, usually encoded in XML file, which describes learning resources. The whole format consists of nine main parts, each of which includes different fields. The five parts are more important: general data (with fields: title, keywords, description, language, etc.), environment (technical type, size, location, etc.), educational (learning type, context, difficulty, etc.), rights (license, price, etc.) and classification (taxonomy, etc.).

The main weakness of this standard is its complexity, which is due to the numbers of different fields. Its practical use is also hindered due to the limited vocabularies and the fact that all the fields are optional. Complexity can be dealt with using specialized tools that format metadata in required way, and the practical usability can be greatly increased by expanding the standard. One of such standard extension is LRE LOM (Learning Resource Exchange LOM) [7], which expands most of the vocabularies but also defines that title, author, language, location, license and classification fields are mandatory.

There are many more standards for metadata description, but their use is not so common in learning environments. Their main disadvantage is lack of explicit support for the educational aspect of the resource. One of the widely used in Internet environment for all kinds of Web resources is Dublin core [3].

5. Exchanging metadata

Learning resources are generated by different content providers and many of them use their own version of repository to publish these resources. With the growing number of content providers all over Europe, users are puzzled over which portal to use in order to find the resources they need. By storing resource metadata in some standard format outside the resource itself, it has become possible

to effectively exchange the metadata between different providers. This has two main advantages: the content itself remains at the original location where it can be maintained by its authors/provider, and search thought some portal's content is available to users of other portals, since metadata are all that is needed for effective search. This kind of search is called federated search and instead of individual repository we can talk about network of repositories.

European projects CALIBRATE [12] and MELT [13] both use metadata exchange in order to connect repositories all over Europe. As a result, Learning Resource Exchange network, lead by European Schoolnet [4], was established, with one of its partner being SIO (Slovensko izobraževalno omrežje - Slovenian Educational Network) [14]. Metadata are stored in LRE LOM format, and SQI (Simple Query Interface) [8] and OAI-PMH (Open Archive Initiative – Protocol for Metadata Harvesting) [10] protocols are used in order to exchange metadata.

5.1. Simple Query Interface (SQI)

SQI protocol is used for exchanging only specific metadata, which satisfy given user-defined filters. User issues a query that searches only for the content with desired properties. That query is executed in local repository and at the same time is also sent to all other repositories in the network that supports SQI. Results of all searches are returned to the user. That way, all the searches are done on the fly, so the returned metadata are always up to date, but the search can take longer due to the delays on the network or can even fail if distant repository is not working at the moment.

5.2. Open Archive Initiative – Protocol for Metadata Harvesting (OAI – PMH)

OAI-PHM is a protocol for harvesting metadata, which is physical exchange of all metadata. By using OAI-PHM protocol, all metadata from one provider are transferred to all other providers in the network. That way, the search can be performed locally and still include results form other repositories without network delays or fear of broken connections. Since data are transferred from their original location, regular updates are needed in order to keep the metadata up to date.

6. Conclusion

The usage of quality metadata stored in standard formats enables users of learning resources easier and faster searching for content, which suits their needs, language or preferred way of teaching or learning regardless of the portals, where they are performing the search. Creation of needed metadata is long-lasting and complex procedure, but with the usage of advanced computer algorithms and social tagging that could easily change in the near future.

References

[1] Automatic Metadata Generation, url: http://ariadne.cs.kuleuven.be/amg/

[2] del.icio.us, url: http://del.icio.us

[3] Dublin Core Metadata Initiative, url: http://dublincore.org/

[4] European Schoolnet, url: http://www.eun.org/portal/index.htm

[5] Flickr, url: http://www.flickr.com

[6] IEEE LOM Standard, url: http://ltsc.ieee.org/wg12

[7] Learning Resource Exchange, url: http://lre.eun.org

[8] LOR interoperability, url: http://www.prolearn-project.org/lori

[9] A. Mathes, "Folksonomies – Cooperative Classification and Communication Through Shared Metadata", University of Illinois Urbana-Champaign, 2004.

[10] Open Archives Initiative, url: http://www.openarchives.org/

[11] Project AMeGA, url: http://ils.unc.edu/mrc/amega

[12] Project CALIBRATE (IST-28025), url: http://calibrate.eun.org

[13] Project MELT (ECP-2005-EDU-038103), url: http://info.melt-project.eu

[14] SIO - Slovensko izobraževalno omrežje, url: http://sio.edus.si/

[15] Understanding Metadata, NISO Press, 2004, url: http://www.niso.org/standards/resources/UnderstandingMetadata.pdf

[16] Wikipedia - Folksonomy, url: http://en.wikipedia.org/wiki/Folksonomy

Acknowledgment

European project "MELT: A Metadata Ecology for Learning and Teaching" is co-founded by European Commission as part of eContentPlus program (Grant agreement number ECP-2005-EDU-038103).

Authors

Jernej Južna *is a postgraduate student at the Faculty of Computer and Information Science, University of Ljubljana, where he is employed as a researcher on the MELT project.*

Alenka Kavčič *is an assistant at the Faculty of Computer and Information Science, University of Ljubljana. Her main research interest is in the field of hypermedia and computer based learning, especially adaptive hypermedia systems and user modelling.*

Saša Divjak *is a professor at the Faculty of Computer and Information Science, University of Ljubljana. He is actively involved in e-learning and use of ICT in schools.*