

LOCALIZED LEARNING OBJECTS METADATA ENRICHMENT THROUGH CYRILLIC TRANSLITERATION

ZORAN ZDRAVEV

Faculty of Computer Science, Goce Delcev University in Stip, Macedonia, zoran.zdravev@ugd.edu.mk

MARGITA KON POPOVSKA

Faculty of Computer Science and Engineering, "Ss. Cyril and Methodius" University in Skopje, Macedonia,
margita@ii.edu.mk

JOVAN PEHCEVSKI

Faculty of Informatics, European University, Skopje, Macedonia, jovan.pehcevski@eurm.edu.mk

Abstract:

A myriad of digital resources that can be used in educational purposes and which can be Learning Objects has emerged on the Internet. Imperative of the scientific research community is to make these Learning Objects available to all who have an interest in education: companies, universities, schools, professors, teachers, students, pupils, lifelong learners, etc. The objective to be achieved is that Learning Objects are easily found and that they can be reused an unlimited number of times. Learning Objects Metadata contains all relevant information about digital Learning Object and therefore they are the most important element in searching and retrieving. The fact is that the discovery of Learning Objects that have "English" metadata is much easier than in the case of localized Learning Objects that have "non-English" metadata. The process of localization of the learning objects means applicability of different languages (multilingualism) and different cultural contexts (multiculturalism). The process of localization of the software, and therefore of the digital learning resources, is divided into three parts: the first part is adjusting to the "local environment" (locale), the second part is a translation and adaptation of the user interface and the third part is the translation and adaptation of the documentation. The third part includes the localization of metadata. A proposed process for improvement of discovery and exchange of the localized Learning Objects from a localized repository is given as metadata enrichment - Localized Metadata Enrichment (LME). In this paper we propose method for localized Learning Objects metadata enrichment through Cyrillic transliteration.

Keywords: Learning Objects Repository, LO Metadata Enrichment, LO Metadata Localization

1. INTRODUCTION

There are already a huge number of digital resources on the Internet that can be used for learning. The organization of these digital resources in the form of Learning Objects and they are stored in Learning Objects Repositories. With this continuous growth of the number of Learning Objects that exist online and in repositories a problem appears: how to find exactly those learning objects we need at the particular moment. To improve the availability of Learning Objects a standardization of metadata that describe Learning Objects is introduced, specifications for interoperability of repositories are adopted and they are organized as global federations of independent repositories. This works flawlessly in the case when Learning Objects and Learning Objects Repository are in English language, but when it comes to Learning Objects and Learning Objects Repository that are adapted to different languages and different cultural contexts global availability is significantly hampered.

Since most of the learning objects are non-textual (animations, images, video, audio) the discovering of

learning objects in repositories can be an impossible task without metadata. As expected, the number of learning objects in repositories will grow exponentially, and the lack of metadata will be a fundamental and critical limiting factor for the ability to find, discover, manage and use the objects.

In this paper we propose a process of Localized Learning Objects Metadata Enrichment through Cyrillic transliteration that would improve the availability of Localized Learning Objects stored in Localized Learning Objects Repository. With the application of this localized metadata enrichment globalized Learning Objects Discovery and Exchange is improved. The paper is organized in four parts. The second part briefly defines Learning Objects and explains the need for metadata, localization of Learning Objects and identifies the standards and specifications that are important for creating a Localized Learning Objects Repository. The concept of these repositories is given in the third part. The process of localized metadata enrichment through transliteration is described in the fourth section. The conclusion is given at the end

2. LEARNING OBJECTS, METADATA AND STANDARDIZATION

2.1. Learning Objects and Metadata

There is no single definition of the Learning Objects. We accept David Wiley's definition: A Learning Object is any digital resource that can be reused to support learning. He emphasizes that a learning object should be digital and reusable. The size and content of the learning object is associated with reusability, i.e. depends on reusability. Others agree that learning objects are modules or units that should be delivered through or by means of computers, which are independent and that provide a whole learning content in a planned learning. Learning Objects should be independent, i.e. it should be possible to use them independently from other objects and contents, that they should possess at least a minimum amount of information from which something can be learned and that their use is conditioned by computers.

Generally, regardless of different definitions, learning objects are digital resources, modular in nature and used in the learning process. Their size can vary, they can be applied in different areas and have different levels of granularity. Learning objects can be connected with other learning objects in order to create a greater teaching unit (Figure 1). In relation to learning objects research and development are directed towards their reusability and therefore it is obvious that they should be digital resources.

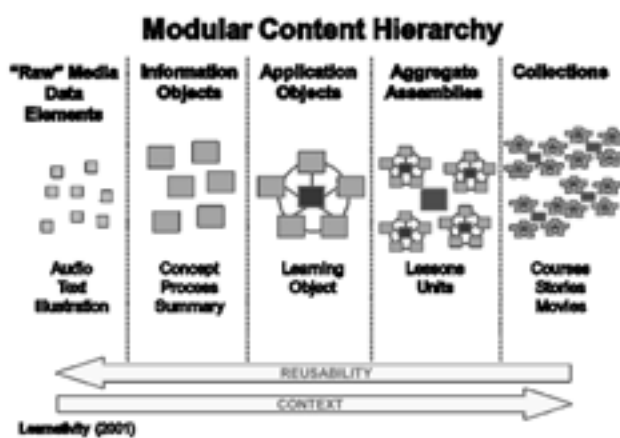


Fig.1: Modular Content Hierarchy

When it comes to learning objects as digital resources it means that they can be, but are not limited only to: texts, simulations, animations, websites, tutorials, tests, multimedia, video clips, sounds, images, illustrations, diagrams, graphs, maps or exams. All digital resources are a huge collection of data, bits and bytes of information. Digital resources are stored in repositories, and are described by metadata.

Metadata is information about an object, either physical or digital. For learning objects metadata represent data about an object. Technically it is the XML scheme used to describe learning objects. The purpose of metadata for learning objects is to support discovery learning objects, and thus facilitate their reusability. The objectives of the

metadata are to enable users to seek and use learning objects.

2.2. Localization of Learning Objects

The expansion of the Internet greatly influences the rise of the awareness and the need for localization of digital resources. In this sense the term localization is most often used to adapt the software and digital content (such as web sites) to the language and culture of certain ethnic or geographically defined groups. When speaking about Learning Objects, we mean that they are in a digital form, that they can be software, just as they may well be text documents, videos, presentations, audios, images or websites. This means that there is no essential difference between the localization of software and the localization of learning objects and therefore in further text when localization of software is mentioned it is identified with localization of learning objects.

Internationalization is a requirement for localization and implies respect and implementation of international standards and avoidance of contents or symbols that radiate strongly, or are burdened with a distinctive cultural knowledge (knowledge, not meaning).

The process of software localization is divided into three parts [1]:

- The first part is an adaptation to the "locale"
- The second part is a translation and adaptation of the user interface and
- The third part is a translation and adaptation of the documentation

Adaptation to the locale is the first and the basic task in the localization process. According to the international standard ISO/IEC 15897 (ISO/IEC 1999), the locale is a "definition of a subset of user information about the technological environment that depends on language, territory, or other cultural traditions".

Information about the locale is usually identified through language, by using a code for the language consisting of two letters (ISO 639-1) and by territory (state) using the code for the territory, which also has two letters (ISO 3166-1). This information does not depend solely on the language (e.g. they are different for UK and U.S., though these countries use the same language) or it does not depend only on the state (for example, Canada has two official languages, English and French, each of these combinations language/country has a way of showing the date, time, numbers and other elements).

POSIX (Portable Operating System Interface for Computer Environments) is the first standard that defines the basic data of the localization. The POSIX model has six main categories (ISO/IEC 9945-2) which define it:

- Classification of characters (signs) and the manner of conversion.
- Method of ordering.
- Money format (monetary).
- Numerical, non-monetary formatting.
- Formats of date and time.

- Formats of informative and diagnostic messages and interactive responses.

This is a minimal package of elements for environment localization of any software, of course including learning objects as well.

The adaptation of the user interface is the second component of localization and it comprises localization of messages (dialogues) and all menus and their associated elements (buttons, legends, tapes, etc.).

The third component is the translation of documentation, which covers the translation of texts for the license and files for user help.

In the third component, the metadata is left last, but not the least according to their importance. As already explained, the metadata are the most important element in the search. They contain all relevant information about the learning objects. If there are enough good metadata, then the probability of discovering appropriate learning objects is greater. Finding the learning objects in repositories is very similar to the general Internet browsing. Modern search engines can create a wealth of information that is universally available on the web and that could easily be found. The techniques for finding information used by these search engines are usually effective only when applied to Web collections which are written in English and Latin alphabet. However, there are many challenges to face in using search engines in "non-English" web collections.

Accordingly, the discovery of learning objects that are with "English" metadata is much easier than in the case of learning objects that are with "non-English" metadata. Here two typical cases are isolated:

- Metadata are written in non-Latin letters – it means it needs to be taken into account whether the query is written in Latin or not.

- The search should be done on an extended group of related words (e.g. work, works, working), which is often not so simple for non-English languages because of different grammatical rules of word formation.

Solving these problems is done by special algorithms for transliteration and words stemming.

2.3. Standards used for Learning Objects discovery and exchange

To enable global retrieval and exchange of learning objects accredited standards for interoperability of digital content for learning are required. With the use of accredited standards the risk in the implementation of large investments in technologies for learning are also reduced. A number of institutions and bodies work on the accreditation of these standards; here we would like to mention some of the most influential: IEEE LTSC, CEN and IMS GLC. The standards of interoperability are generally divided into: 1. standards and specifications for discovery contents and standards; 2. specifications for contents using. The standards and specifications for discovery contents we would like to emphasize as important are: OAI-PMH, IEEE LOM, IMS DRI and IMS LOD. The last specification, IMS LOD, is still a draft version. IEEE LOM facilitates sharing and exchange of learning objects by creating conditions for the development of catalogues and lists.

- OAI-PMH - Open Archives Initiative Protocol for Metadata Harvesting.
- IEEE LTSC LOM - Learning Technology Standards Committee, Learning Objects Metadata Standard.
- IMS LOD - Learning Object Discovery and Exchange specification
- IMS DRI - Digital Repositories Interoperability specification

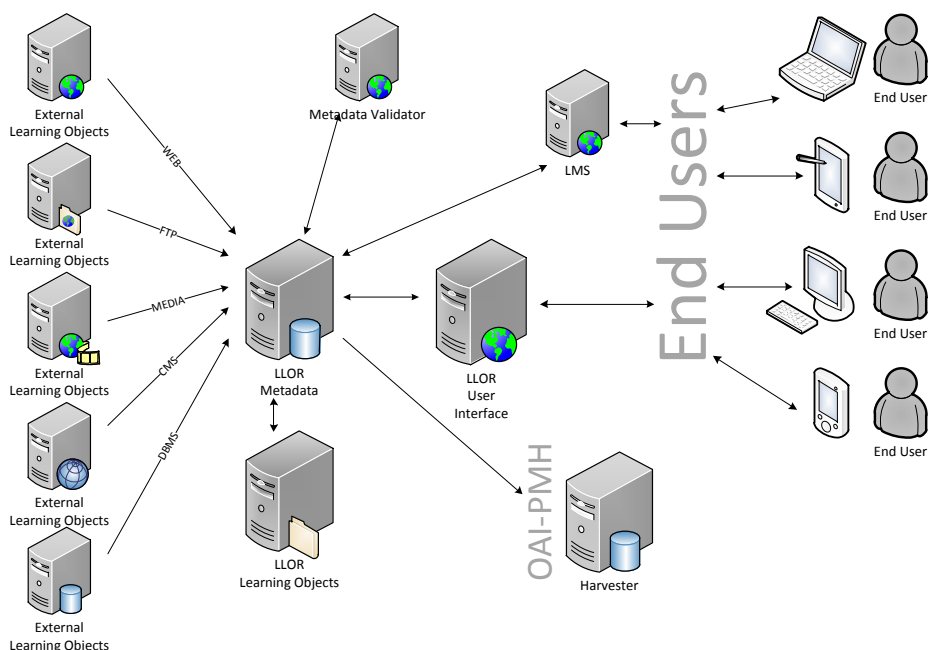


Fig.2: Localized Learning Objects Repository

3. FEDERALIZED LOCALIZED LEARNING OBJECT REPOSITORY

The process of localization of the learning objects and repositories means adaptation of the repositories and the learning objects stored in them to be used in different languages (multilingualism) and different cultural contexts (multiculturalism).

Localized learning object repositories (LLOR) can function as a standalone repository and perform all the functions to be performed by a repository. A DBMS server has the central role in the architecture of the LLOR repository, where the metadata and the locations of files that are added to the metadata are stored (see Figure 2). Files that are picked up in the repository are stored on file servers, and communication with users is through the web server. For those files that are located on another network location in the metadata a link is written and these are the so called external learning objects. End users access the repository through any LMS or LCMS or directly, as already indicated, through a web interface repository. In such a case functionalities for localized search can be built in, which will meet the main goals of discovery and exchange of learning objects. But in that way a repository

will remain isolated and learning objects will not be available to users who are not members of this repository.

The solution for such repositories is that they are associated in a federation of repositories (see Figure 3). In such a federation of repositories there is a server Harvester tasked to collect, i.e. to harvest metadata from the associated localized repositories by protocol OAI-PMH.

These metadata are then validated on the server for validation and are ultimately saved in the global repository. The global repository is available to end users through a system for management of contents and learning. Associated local repositories can also function independently

Since the federation may be accompanied by repositories with different linguistic and cultural backgrounds, the incorporation of functionalities for localized search will not be a solution. In this case a federalized repository learning objects discovery should be enabled with enrichment of metadata with localized data. So, during the harvesting of metadata the data necessary to detect learning objects will be gathered.

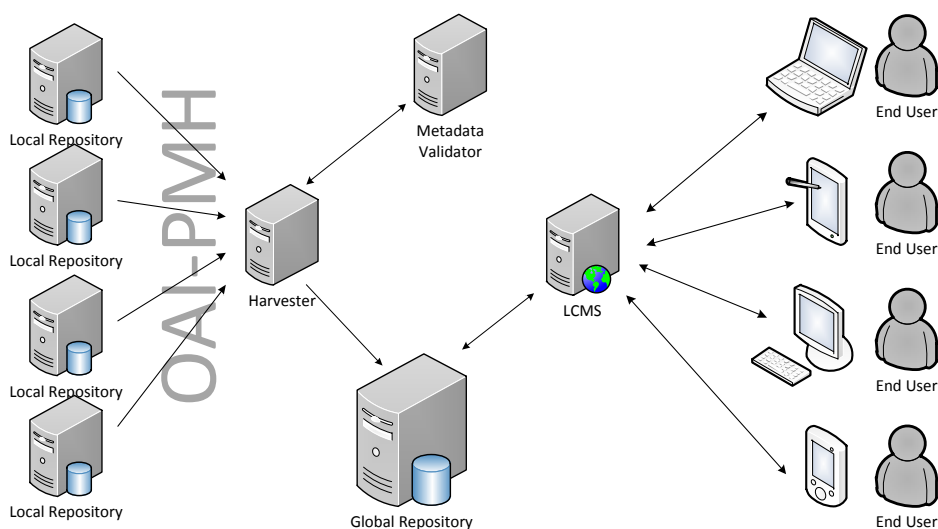


Figure 3: Federalization of repositories

4. TRANSLITERATION AND METADATA ENRICHMENT PROCESS

Enrichment of metadata is a process which, based on user-entered metadata, using certain algorithms, automatically generates additional metadata that further describe the learning object and thus facilitate its discovery in the repository.

The process of enrichment of metadata (LME – Localized Metadata Enrichment) consists of three components: Localized Metadata Transliteration (LMT), Localized Metadata Word Stemming (LMWS), Keywords and Metadata Vocabulary Bank (KwM-VB) (Fig.4). In the environments where Cyrillic keyboards are often used

LMT is necessary. LMT will solve the problem of typing, during the search, with Cyrillic or Latin letters –it doesn't matter which letters will be used, the searched results will be the same. Other components LMWS and KwM-VB are optional and they can be considered as a "useful" upgrade.

LMT-Localized Metadata Transliteration - Transliteration is the process of converting a text from one alphabet to another in a systematic manner according to specific, predetermined rules. In terms of information technology, transliteration is mapping from one system of writing into another. This is done word by word, or ideally letter by letter. The objective of transliteration is that the reader

can reconstruct the original spelling of unknown transliterated words, based on the information given.

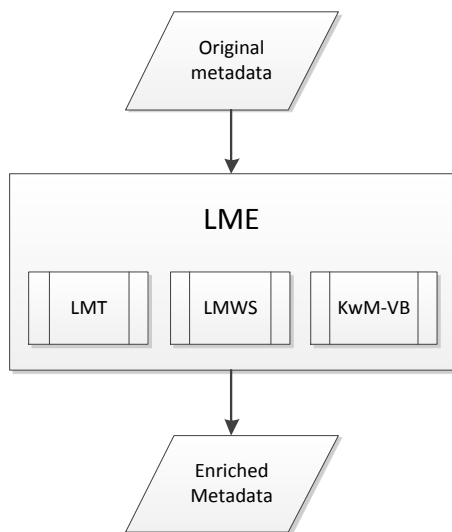


Fig. 4: Localized Metadata Enrichment

The process of metadata enriching thru transliteration runs in three scenarios (Fig. 5):

- Imported metadata are in Cyrillic. In this case, all the words are transliterated from Cyrillic into Latin.
- The entered metadata are Latin according to the recognized standards. In this case, transliteration from Latin into Cyrillic is performed.
- The entered metadata are Latin, but not standard. The difference between this and the previous case is that now the entering of the metadata is done with the so-called “Cyrillic fonts” and where real mapping is Latin, but instead Cyrillic letters the signs “ { [}] \ | ` ~ @ ^ ” appear. So now the transliteration of “non-standard” Latin into Cyrillic is performed first.

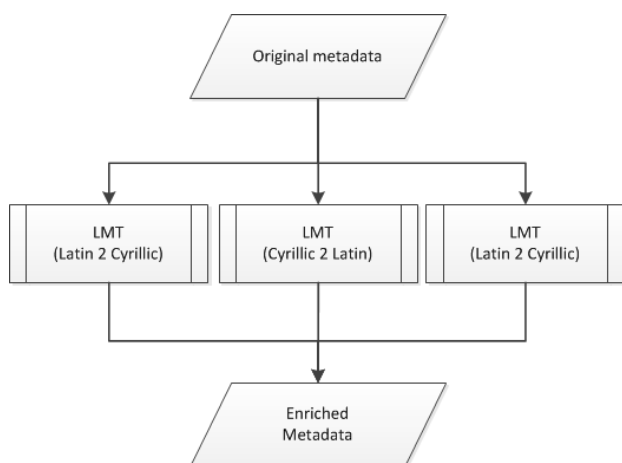


Figure 5: Three scenarios of LMT

LMWS – Localized Metadata Word Stemming - Setting the search queries in a search engine, in any human language, depends much on the grammatical rules of that language. This fact in most cases makes these rules for searches in English unsuitable.

KwM-VB – Keywords and Metadata Vocabulary Bank. This is actually about a multilingual dictionary – a thesaurus that initially gives about 2255 words and

expressions. Words and expressions that are used as standard metadata and keywords are taken from here. The dictionary and the concept of the dictionary are taken from the project LRE (Learning Resource Exchange for Schools), a member of EUN (European Schoolnet). LRET (Learning Resource Exchange Thesaurus), formerly known as ETB (European Thesaurus Browser) is published as a result and it is now managed through the project ASPECT of VBE (Vocabulary Bank for Education). Words and expressions in this vocabulary-thesaurus have been translated to desired language.

5. CONCLUSION

The number of digital resources that can be used is increasing daily. The installation of such resources on the Internet is not enough to enable their discovery and exchange. Organization and storage of these digital resources in the form of Learning Objects is made in repositories. Learning Objects are described with their metadata. Metadata are key elements through which the discovery and exchange of Learning Objects is made. If there is sufficient metadata for each Learning Object, their search will be much easier and more successful. In this paper we propose a methodology for localized metadata enrichment through Cyrillic transliteration which easier search, discovery and exchange are achieved. In the environments where Cyrillic keyboards are often used LMT is necessary. LMT will solve the problem of typing, during the search, with Cyrillic or Latin letters –it doesn’t matter which letters will be used, the searched results will be the same.

6. REFERENCES

- [1] Dagiene, V., & Zilinskiene, I. (2009). Localization of Learning Objects in Mathematics. 10th Int. Conference: Models in Developing Mathematics Education (pgs. 129–133). Dresden: The University of Applied Sciences (FH).
- [2] Hodgins, W. H. (2002). The Future of Learning Objects. In M. L. Corradini, & J. R. Lohmann (Ed.), e-Technologies in Engineering Education - Learning Outcomes Providing Future Possibilities (pgs. 76-82). Davos: Engineering Conferences International, Inc. - New York.
- [3] IEEE LTSC. (2002, July 15). IEEE 1484.12.1-2002 Draft Standard for Learning Object Metadata. New York, USA: Institute of Electrical and Electronics Engineers (IEEE).
- [4] Ip, A., Young, A., & Morrison, I. (2002). Learning Objects - Whose are they? 15th Annual Conference of the National Advisory Committee on Computing Qualifications, (pgs. 315-320). Hamilton.
- [5] Janevski, I., Takasmanov, K., & Pehcevski, J. (2008). NABU: A Macedonian Web Search Portal. International Conference on Innovations in Information Technology, 2008. IIT 2008., (pgs. 185 - 189). Al Ain.

- [6] Puustjarvi, J. (2007). Syntax and Semantics of Learning Object Metadata. In K. Harman, & A. Koohang, Learning Objects: Standards, Metadata, Repositories, and LCMS (pgs. 41-62). Santa Rosa, California, USA: Informing Science Press.
- [7] Wiley, D. A. (2000). Connecting learning objects to instructional design theory: A definition, a metaphor,. (D. A. Wiley, Editor) Downloaded January 18, 2011 from The Instructional Use of Learning Objects: Online Version: <http://reusability.org/read/chapters/wiley.doc>
- [8] Zdravev, Z. N., & Grceva, S. (2005). Developing Learning Objects Systems through implementation of Learning Object Metadata Standard. The Second International Conference on Informatics, Educational Technology and new Media in Education. 1, pgs. 273-279. Sombor: Teachers Faculty, Sombor.
- [9] Zdravev, Z. N., & Grceva, S. D. (2007). Developing Learning Content Management Systems Based On Learning Objects – Issues And Opportunities. TIO2007 - 4th International Conference on Technology, Informatics and Education. Novi Sad.
- [10] Zdravev, Z. N., Kon Popovska, M., (2011). Localization and internationalization of digital learning resources. TIO6-2011 (pgs. Submitted). Cacak-Serbia: Technical Faculty Čačak - University of Kragujevac.
- [11] Zdravev, Z. N., Kon Popovska, M., & Pehcevski, J. (2011). Localized Learning Objects Discovery and Exchange. Sixth International Conference - Computer Science 2011. 01 - 03 September 2011, Ohrid, Macedonia
- [12] Zdravev, Z. N., Warrick, W., Hathaway, D., Grceva, S. D., Sterjoska, A., & Ivanoski, V. (2006). Creating Localized Repository Of Digital Tools And Resources. The Third International Conference on Informatics, Educational Technology and New Media in Education - Proceedings (pgs. 411-417). Sombor, Serbia: Faculty of Education in Sombor Publishing Centre.