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.

![Modular Content Hierarchy](image)

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 LODE. The last specification, IMS LODE, 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.
• IMS LODE - 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.

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.

![Diagram of Localized Metadata Enrichment](image)

**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.

![Diagram of Three scenarios of LMT](image)

**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) s published as a result and 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**


Keywords: E-Learning, LMS, education design, AI, education, auto-poiesis

1. INTRODUCTION

The fields of the semantic web and of the software agents technologies are mature [1], [2], but they have not found till today a widespread application specifically in the field of Education. Even if the development of the research in the field of Computer Science and the potentialities that they could propose in different field of application, they are not always fully used and so far there’s no significant step forward in the researches related to the introduction of technologies in the field of Education. Such situation is tied to various factors:

- in the field of Education so far the amount of investment and of research is inferior to the amount spent in other fields;
- in the field of Human Sciences and mostly of education there exists prejudices towards technologies in general, and towards Artificial Intelligence specifically.

That second aspect is due to a scarce knowledge, among researchers in Education, about approaches and researches that today the Cognitive Sciences and the research on software agents propose. In the school context and in the research in the educational field the skepticism is present because the prevalent model in education properly relies on the complexity of the learning process and, since technology is wrongly seen as deterministic, it is meant to undermine the role of the subject’s autonomy and, in the field of education and training, the flexibility that should characterize the learning processes.

Such vision comes from an atavic rejection towards technologies and from the idea created by the projects on Artificial Intelligence of the last century Sixties and Seventies. Those projects were aimed at simulating the behaviour of the human mind and at supporting deterministic models of learning.

The mistrust is caused by past experiences such as the programmed instruction in which the introduction of technologies had fostered linear and reductionist processes and had induced the conviction that the use of AI could be connected only to comportamentistic approaches.

The wrong information is producing a series of problems that can be summarized into two extreme situations that come from the same reason:

- the rejection for technologies by the education operators;
- the introduction of new technologies in a technicistic way. In other words the expert in technology is the one delegated to provide the artifact as if the solution was neutral and didn’t need to involve a planning decision.

The lack of a co-design process among experts of different fields has promoted solutions in which instructivist models are present, much more simple and widespread.

The new technologies are, in this way, seen as tools to get the content acquisition. Today, instead, the teching/learning process seems more and more finalized to the acquisition of competences such as the learning to learn, the self-regulated learning, the construction of a professional habitus, the problem posing and the problem solving.

The only way to foster the development of the research is surely a better communication among communities of the two fields. Very often researches on similar topics have been developed in autonomous ways with different languages and approaches; very often the comparison and the realization of joint projects have been hard. The result is having lame projects, valid in one of the two fields, but weak and lacking in the other.

The gap between the research’s results in some sciences, specifically in the cognitive sciences, today fundamental in the knowledge society, and their application in other areas, specifically in Education, seems to be at the basis of the conference organized by MIT, entitled “2011 MIT SDM Conference on Systems Thinking for Contemporary
Challenges. Addressing Complexity and Innovation in Healthcare, Education, and Product Development” that will take place on October 24-25, 2011. 

Here follows the conference’s introduction:

"Across industries and government organizations around the globe, a systems-based approach is increasingly seen as critical to addressing the urgent and complex problems we face today. For many organizations, the question is not whether to employ systems thinking, but how to apply it before it is too late.

The Massachusetts Institute of Technology’s annual Conference on Systems Thinking for Contemporary Challenges, sponsored by the System Design and Management Program, will focus on addressing complexity and innovation in healthcare, education, and product development. MIT has carefully chosen speakers not only for their expertise in addressing complex systems challenges, but also for their role in leading the implementation of the day-to-day tasks that produce results” [3].

In strict relation to the researches here described the direction to take today seems the one which suggests a co-disciplinarity from the very first phase of the project, during which experts in education can approach the potentialities of new technologies and, specifically, of software agents and the experts in computer engineering can acquire consciousness of the problems today present in the fields of education, as Koper underline [4].

2. ITS AND SUBJECT MATTER ORIENTED ARTIFACTS

The state-of-the-art of the research shows in the last 10 years the collaboration between researchers in the computer science field and the cognitive psychologists, but not the same collaboration between computer science experts and experts in didactics and pedagogy.

Models related to Intelligent Tutoring System (ITS) were created, that is, web applications that support the student in disciplinary learning path. In the majority of cases, the activities imply the dialogue between the student and the computer directly during the interaction without the presence of other professionals such as “human” tutor or peers. Examples of ITS are:

- Baghera, an environment that supports students and teachers in teaching/learning Geometry [5];
- MyClass, an environment for the managers’ training which includes an intelligent tutor. That tutor can replan the teaching strategies and adapt them to the student or the group of students, in every phase of the teaching/learning process;
- Andes, an environment created for learning Newtonian Physics, used in many colleges and high schools in the USA [6];
- GRAMY, an environment to experiment the demonstrations of geometry theorems;
- Advanced Geometry Tutor: an environment to support students in the creation of the demonstrations of geometry theorems and to test their validity.

The above mentioned lists shows the all the proposal are subject-matter oriented, that is, tied to a specific topic and built for specific disciplinary goals; they are based on a wide and articulated disciplinary knowledge base (KB), whose implementation requires much time and deep knowledge.

Their goal is to guide the student in the acquisition of disciplinary contents. The chatbots have a central role, they guide the student, also with maieutic modalities, step by step, towards the solution of the problematic situation, providing feedback to the proposed solutions.

As many researches have underlined [7], the pertinence of the feedback is correlated to the accuracy of the questions and two kinds of problems can emerge: on one side the generic answers generated by vague questions, provided to students who would need a punctual orientation and, on the other side, obvious replies provided to students who, instead, would need a deepen the topic. To solve those problems solutions were adopted such as giving to the student the chance to activate dialogues about topics that they consider already acquired, or offering different levels of support, leaving the student free to decide.

Johnson found some problems often related to the dynamic student-agent; he believes that those problems come from a poor “social intelligence” of the synthetic agent:

- Criticizing the same errors several times;
- Interrupting the student’s activities also after insignificant errors;
- Transmitting negative emotional inputs to the student’s actions;
- Not respecting the student’s work;
- Not being able to encourage the student when he needs;
- Not being able to provide support when the student is confused and frustrated [8].

To solve those problems he proposes the presence of a team not only constituted by the student and the synthetic tutor, but also by human tutors and a work group to distribute in a dynamic way role and responsibilities and the use of multiple strategies, since the experimentations done highlightes, for example, that the intervention of the tutor is sometimes welcome, while other subjects, in similar situations, prefer to proceed in autonomy. On the other way the tutor has to be able to balance a presence to encourage with an absence to give a sense of responsibility.

Also as an effect to those critics, the research has been addressed, in the last years, towards solutions that would take into account the emotional aspects and has deeply studied the role played also by the physical appearance of the avatar of the tutor, his facial expression [9].

The main problem of that approach is in any case the connection of the ITS to a specific path, while the need of the school and mainly University, is to have open LMS that are addressed to a specific path of the teacher during the design process. There is the need of a learning environment that include also the tools for the design process and is supported by software agents.