Multilingual Information Delivery Based on a Domain Ontology

Nauris Paulins, Irina Arhipova, Signe Balina

Abstract: As internationalisation becomes more and more intensive, organizations need to work with information in different languages in their work. Interlingua content delivery in e-learning courses could be effective learning content adaption method for delivering information for much wider audience. This paper suggests high-level framework for multilingual learning content delivery. This approach is based on a domain ontology, which allows to annotate and retrieve learning resources in a language independent way. Such method can improve accessibility, sharing and reuse of multilingual learning content.

Key words: Adaptive learning, multilingual ontology, semantic annotation, domain adaption.

INTRODUCTION

The e-learning market has exploded over the last years. At the same time globalisation impact requests for multilingual content, where the same content must be produced in multiple languages. According to IEEE Learning Technology Standards Committee, a learning object (LO) is any entity, digital or non-digital, that may be used for learning [1]. E-learning courses are built for specific knowledge domain, but resources in these courses are divided into small chunks, that can be reused in various learning courses. But in different domains the meaning of a specialized term can differ, which is crucial for correct learning content translation. A specialized course content requires specialized knowledge on the domain, specific subject fields can always be approached from different perspectives, also they can change or develop in other directions, and translators need constant training in new subjects. That means that content should be adapted before it is presented to the user. Such process requires a great effort and huge amount of time, and as stated by Aguado and Montiel [2] translators need to repeat this training in all their working languages.

As mentioned by several authors [3]–[5] the implementation of a semantic layer can improve the management of learning resources. Semantic layer basically is connected to the concept of ontologies. Ontology is metadata, which includes extra description about concepts and relationships of some specific field in course, therefore ontology is described as a specification of conceptualization [6]. An essential requirement for the metadata is their description in a machine readable way, thereby allowing interoperability on the Semantic Web [7]. Ontologies allow systems process data semantically, because there is a common knowledge base, made up of terms and the relationships between these terms.

The analysis of literature has proved that the most acceptable form of ontologies for e-learning purpose is domain ontologies, which can be taxonomies [8] or more complex structure with several relations and rules [9]. The main factors to be considered for the usage of ontologies in e-learning systems are as follows: formalization level of ontologies and conceptualization of ontology.

However, the multilingual information delivery does not include only content translation and annotation, but also such steps like information retrieval, storage and exchange. The translator can easily check the correspondence between terms in different language or find out variations between different conceptualizations. Researchers mostly analyse specific steps of this process, but do not include overall framework for multilingual information delivery. This paper proposes a framework for multilingual ontological knowledge delivery from content providers to students.
In the course of the research project „Competence Centre of Information and Communication Technologies” run by IT competence centre in Latvia, co-financed by European Regional Development Funds, the framework has been developed for dynamic and adaptive information delivery in multilingual form, also focusing on “small” languages, which will improve possibility to share and reuse learning resources across different communities.

This framework includes main components and steps for content delivery in multilingual form, but also includes several steps connected by intelligent tutoring systems and learning path development with ontology support, like it is done by Mizoguchi [10]. The primary objective of the proposed framework is to develop ontology-based e-learning framework, which allows for building adaptive and multilingual learning course in order to promote learning process for students.

THE PROPOSED FRAMEWORK

The proposed high-level framework is composed of several components and includes main steps for semantically based tutoring process (Figure 1).

![Diagram of Multilingual learning content delivery framework](image)

Figure 1: Multilingual learning content delivery framework: 1) LO annotation; 2) LO localization; 3) LO storage; 4) Course/Material development; 5) Content publication.

Basically, all components are separate independent systems, where Learning Management system is the main executive component. Main steps included in the proposed framework for multilingual information delivery are as follows:
• **Annotation of Learning Objects** – the main target of this step is to improve LO management by means of machine readable metadata and ontologies (i.e. taxonomies).

• **Learning Object localization** – this is very important step to prepare LO for multilingual use. There are included such steps like natural language processing and generation, machine translation and also cultural audit for better adaption.

• **Multilingual Learning Object storage** – multilingual content must be stored in such a way that LO has semantic relation between original and translated objects.

• **Learning content development** – main steps for course content retrieval from LO storage and control processes for interaction with learners; these steps are, for example, pedagogic rules, theme hierarchy.

• **Learning content publication** – basically, this is a learning interface for content publication and visualisation. This step works also like the main entry point for information input from learners.

The following sections of the research describe the framework’s parts in detail.

**MULTILINGUAL LEARNING OBJECT ANNOTATION**

As already mentioned, semantic annotation and linking is a time consuming and complex process. According to Knoth [7], the time complexity is a quadratic with respect to the number of LOs stored in repository.

Metadata and domain ontologies are mostly defined at different levels, where LO metadata is an instance of a learning object metadata schema. At the same time metadata is enriched with keywords in the ontology. An example of such system is the SELeNet project that defines the ACM Computer Classification System (http://www.acm.org/class) for the computer science domain as a Resource Description Language Schema [11]. Such method, which employs classes to represent concepts, can also bring some disadvantages, like rigidity of the schema or complication at reasoning level. The solution for such problems could be specialized ontologies with rich set of properties such as Simple Knowledge Organization System (SKOS) [12].

Each domain in the framework ontology is represented by an instance of the *framework: domain* class, which is a subclass of *skos:collection class*, while each *isPartOf* property is a sub-property of the *skos:member property*. Basically, it means that domain knowledge is a collection of different subdomains and each subdomain is a part of some parent domain. These competences are assigned together with SKOS properties *skos:prefLabel* and *skos:altLabel*. Ontology mapping plays an important role in dealing with multiple ontologies, which expresses relation between ontology concepts and properties. SKOS Mapping contains a set of properties for specifying mapping relations between concepts from different domain ontologies (*broadMatch*, *narrowMatch*, *ExactMatch*, etc.).

The multilinguality of ontology is solved by means of translation of concepts during the localization phase. However, to guarantee quality and effective pedagogical reusability, it is necessary to provide the content provider with the ability to modify annotation of LO, according to the context and pedagogical needs.

**LEARNING OBJECT LOCALIZATION FOR MULTILINGUAL USE**

Multilingual content management requests also for multilingual ontologies, therefore it is necessary to guarantee that the same information will be recognizable in different languages. Organizations with huge amount of multilingual resources express the need for multilingual ontologies. Therefore it is necessary to develop an extra module for ontology localization to a specific language and culture community. The translation process is not simple due to the ambiguous nature of natural language processing. Therefore this
process has been divided in several steps: 1) translation process for ontology labels, 2) disambiguating label sense, and 3) ranking translations. These results are given to an expert which can evaluate results and make necessary corrections.

Translation service is mainly based on multilingual resources from where possible translations are obtained and consequently compared against labels in original language and its domain context by means of label translation system. If obtained translation is rated as correct, the system updates ontology model with translations and linguistic information, which is associated with ontology. An important part of this process is a semantically represented sense of each label. This approach is adapted from [13] which uses a method where discovered semantics is extracted from different ontology pools.

STORAGE OF MULTILINGUAL LEARNING OBJECTS

Most of learning objects’ repositories such as MERLOT or CARE store metadata records, which describe LO. It is expected that these repositories play an important role – both humans and software agents are to be capable of searching and retrieving the information stored in them [14]. LO are identified with their metadata. Due to metadata, LO can be searched by search engines, but there are difficulties:

- Domain specific resources can be hard to find,
- Learning resources found for specific learning objective can be difficult to filter;
- The resources lack the educational properties;
- Quality assurance mechanisms are in insufficient amount.

Due to ontologies, LO repositories are content storage and retrieval systems, which highly facilitate searching and filtering of the learning resources. Ontologies enable better organization of LO, which can be done by means of taxonomies [15].

As in Figure 1, there are several parts of repositories – LO repository, LO template repository, ontology and semantic relation repositories. These parts are stored separately, but they are linked. That means that repositories must follow a series of specifications and standards, which enable interoperability with other repositories. In this project IMS Digital Repository interoperability specification (IMS DRI) is mainly used [16]. Though, there are additional standards, which are used from interoperability perspective. Tortosa et all. [17] have summarised main standards and specifications for repository interoperability and LO exchange, as in Table 1.

<table>
<thead>
<tr>
<th>Category</th>
<th>Sub-category</th>
<th>Standard</th>
</tr>
</thead>
<tbody>
<tr>
<td>LO</td>
<td>Metadata</td>
<td>IEEE LOM</td>
</tr>
<tr>
<td></td>
<td>Packaging</td>
<td>IMS CP</td>
</tr>
<tr>
<td></td>
<td>Resources</td>
<td>IMS RLI</td>
</tr>
<tr>
<td></td>
<td>Vocabulary</td>
<td>IMS VDEX</td>
</tr>
<tr>
<td>Searches</td>
<td>Interoperability Repositories</td>
<td>IMS DRI</td>
</tr>
<tr>
<td></td>
<td>Query interoperability</td>
<td>CEN SQI</td>
</tr>
<tr>
<td></td>
<td>Publication interoperability</td>
<td>CEN SPI</td>
</tr>
<tr>
<td>Services</td>
<td>Data exchange</td>
<td>IMS LIS</td>
</tr>
</tbody>
</table>

Main functions included in IMS DRI specification are – LO submission and storage, search and retrieval, information collection form other repositories. For resource organization is better to use IMS RLI (Resource List Interoperability) standard, this norm determines the optimal way to organize, describe and exchange list of repository resources, which are transferred using IMS CP (Content Packaging) standard.
One of the main functions and pillars of interoperability is searching and querying through these repositories. Multilingual content and domain-specific LO collections make this task more challenging. Within the proposed framework one can concentrate on corpus-based approach, utilizing a range of information retrieval specific statistical measures. When compared to dictionary based, corpus based translation gives better performance [18].

LEARNING COURSE CONTENT DEVELOPMENT

As in all intelligent systems, there are three main parts – domain knowledge which contains the teaching content and meta information about learning course subject, pedagogical aspects which include knowledge regarding various pedagogical decisions and user models recording information related to user’s activities in the course.

The course content is based on smaller chunks, which are as training cases or learning elements. At the beginning level the learning content is generated accordingly to the information about the user and the course. The main task of the system is to decide which components are shown to the learner and which are not. The system also must determine which actions are possible in each learning component. These decisions are based on the learner’s profile and their previous activities in the course. The adaption of the course requires the following sequence of steps to be included in tutoring process:

1. Each component in the course has to be checked against state transition function, which must determine the next state of training case if component is selected in a certain state.
2. When the component is selected, the system selects a set of components, which will provide the course continuation.
3. The selected components can be shown or hidden in order to determine which components will be actual display of learner;
4. Learner’s model information should be updated periodically in order to determine changes in component display.
5. During the component selection the system should determine actions within this component.

These are main components, which provide adaption of the content when learners navigate in the course. The initial sequence is mainly defined by the content provider, theme hierarchy, and teaching rules.

LEARNING CONTENT PUBLICATION

Learning content publication is provided within learning environments, like Learning Management System (LMS), Learning Course Management System (LCMs), Virtual Learning environment (VLEs). However, such a system must have intelligence capabilities to be able to automatically adapt the content towards the realisation of the personalised learning concept. One of the instructional designer’s tasks is to make learning process more effective, involving the use of new media and visualisation techniques. The explanation of the positive effect of visualisation is provided by cognitive load theory [19] and cognitive theory of multimedia learning which is presented by Mayer [20]. CISCO researchers [21] has proved that multimodal learning is more effective than traditional unimodal learning. The adaption of content is carried out by manipulation of LO which is embedded in learning material. Such manipulations can be done for different reasons: providing explanation variants, which depend on the user model, and providing additional or comparative explanation. It is possible that a system, according to the user model, can hide or change some fragments of material, for example, it can change some pictures with text if the user is mostly tended on verbal perception. During the content visualisation and publication the students’ cognitive load and characteristics of their learning style should be considered.
CONCLUSIONS AND FUTURE WORK
In this paper the ability to deliver multilingual learning content was analysed. Ontology usage for educational content annotation can be used for reasoning across multiple domains. E-learning systems should support annotation LO in multilingual settings, if they are meant to be interoperable. It is important in regions of “small languages”, like in Europe, where there are many countries with small population. It will improve possibility to share and reuse learning resources across different international groups. Due to an increasing amount of resources, there should be provided interoperability between different systems, which can be done with wide range of standards. Information access and retrieval also should be based on semantic features, which will improve quality and accuracy of searching results. However, personalised information delivery is strictly based on student’s model and their interest domain, which can be determined during learning process.

Further research should include framework implementation in real systems. This should include data on how to implement and evaluate the proposed framework. The multimedia content annotation and adaption for multilingual use should be evaluated more deeply as well.

ACKNOWLEDGEMENTS
This research is part of a project „Competence Centre of Information and Communication Technologies“ run by IT competence centre, contract No. L-KC-11-0003, co-financed by European Regional Development Fund.

REFERENCES

ABOUT THE AUTHOR
PhD student, Nauris Paulins, Depament of Computer Systems, Faculty of Information Technologies, Latvia University of Agriculture – Jelgava, Latvia, Phone: 0037163005703, E-mail: nauris.paulins@llu.lv

Professor Irina Arhipova, Department of Control System, Faculty of Information Technologies, Latvia University of Agriculture – Jelgava, Latvia, Phone: 0037163005707, E-mail: irina.arhipova@llu.lv

Professor Signe Balina, University of Latvia, Faculty of Economics and Management, Aspazijas Blvd. 5, Riga, Latvia, Phone: 0037167034646, E-mail: signe.balina@lu.lv