The Third International Conference on e-Learning (eLearning-2012), 27-28 September 2012, Belgrade, Serbia

BUILDING TERMINOLOGICAL RESOURCES IN AN E-LEARNING ENVIRONMENT

RANKA STANKOVIĆ, IVAN OBRADOVIĆ, OLIVERA KITANOVIĆ, LJILJANA KOLONJA
University of Belgrade, Faculty of Mining and Geology, {ranka|ivano|olja|lily}@rgf.bg.ac.rs

Abstract: *At the University of Belgrade Faculty of Mining and Geology a terminological resource dubbed RudOnto is being developed, with the aim of becoming the reference resource for mining terminology in Serbian in e-format. As a central terminological resource it is also invaluable in the learning process, and it is included in the blended learning approach at the Faculty of Mining and Geology through an export of its subsets to the Moodle e-learning environment. This paper gives an outline of RudOnto, details about Moodle implementation on the Faculty of Mining and Geology, and describes the mechanism for generating Moodle glossaries from RudOnto.*

Keywords: *E-Learning, Terminological resources, Glossary, Moodle*

1. INTRODUCTION

The importance of terminological resources for specific domains in electronic format is growing with the rapidly expanding availability of various texts on the web. First and foremost, they are indispensable in information an document retrieval systems. In addition to monolingual resources, machine translation systems and cross-language information retrieval emphasize the need for development of bilingual and multilingual terminological resources as well.

However, terminological resources also have a strong educational impact. It is of paramount importance that students involved in the study of a specific domain get acquainted with the proper terminology related to that domain both in their mother tongue, and in other languages. Hence the need to integrate terminological resources in e-format into the e-learning environment.

The importance of developing both Serbian terminological resources and multilingual resources involving Serbian as one of the languages in e-format for mining engineering terms has been recognized several years ago. Namely, various applications developed in this area followed by the development of an information system for planning, monitoring and management of mine exploitation, indicated that such resources would greatly contribute to their functionality [5].

The only available multilingual terminological resource in printed form involving Serbian is a Dictionary of mining in five languages (Serbian, English, French, German and Russian), published by the Mining Institute [9]. This dictionary has been used for almost three decades as the main reference, especially for translators of works published in this field. A newer edition was not printed, despite the fact that the development of mining engineering has generated many new terms and has also made some terms obsolete.

Another multilingual resource is a thesaurus of geological terms with more than 3000 dictionary entries with their English equivalents, developed within the GeolISS project and available for search on the web (http://geoliss.ekoplan.gov.rs/term).

Due to this scarcity of terminological resources, a team of researches at the University of Belgrade Faculty of Mining and Geology (FMG) embarked on the task of developing RudOnto, a complex terminological resource aimed at gradually becoming the reference resource for mining terminology in Serbian in e-format. RudOnto is presently implemented within a mining engineering information system, which has previously also been developed at FMG. Although it has reached a considerable size to date, it is still being intensively enlarged and refined. However, it is also already being used, among other things for the production of controlled dictionaries related to planning and management of exploitation, to mine safety protection systems, mining equipement management, human resources management systems, as well as in a business intelligence module of the information system for defining OLAP cubes.

Given the fact that RudOnto is targeted to become the reference resource for mining terminology in Serbian, it was only natural to introduce it in the teaching process at FMG. Taken into account its e-format and the fact that for several years now FMG has been developing and using its Moodle e-learning platform, the decision has been made to develop an interface between RudOnto and Moodle. In this paper we describe how we have tackled this problem[1].

In the next section we give a brief outline of terminological resources in general and RudOnto in particular. Section 3 offers an overview of state of the art of Moodle implementation and blended learning at FMG. In Section 4 we describe how RudOnto is used for generating Moodle glossaries, followed by concluding remarks, mostly about future work planned.

[1] The research outlined in this paper was supported by the Serbian Ministry of Education and Science grant #III47003.

2. THE RUDONTO TERMINOLOGICAL RESOURCE

Presently, RudOnto is a terminological resource organized basically as a taxonomy of terms, in which each term is followed by a definition, its synonyms, and bibliographical reference to their source, as well as equivalent terms in other languages (presently only the English equivalent has been implemented). Namely, terminological resources can be organized in various forms: as indexes, glossaries, taxonomies, thesauruses and ontologies. Basically, each of them is an extension of the one that precedes it, thus forming the so called ontological spectrum [3], as depicted in Figure 1. We will give a brief outline of these forms, but it should be stressed that the boundaries between them are not always easily determined.

Index is the simplest form, basically just a list of terms, usually arranged in alphabetical order. Glossaries, which are on the next level of the spectrum, are lists of terms with definitions, and they can be monolingual, bilingual or multilingual. In the case of a bilingual or multilingual terminological resources, corresponding terms are usually linked by appropriate mechanisms.

In taxonomies semantic relations between terms are introduced, or more precisely, between concepts represented by specific terms. The elementary semantic relationship is the hypernym/hyponym relation between the broader and the narrower concept, respectively. Hence, a taxonomy in essence offers a hierarchical classification of concepts, in addition to the terms that describe them, and the definitions that explain them in more detail.

Thesauruses offer and even more complex semantic structure, and due to this complexity, they are very often confined to a specific domain. Besides displaying a richer semantic structure through more elaborate semantic relations such as holonymy/meronymy (part of) and the like, thesauruses are primarily aimed at facilitating document retrieval and achieving consistency in indexing documents stored in a database. Hence, they provide assistance to persons who associate terms or indexes to documents in a database, one the one hand, and persons who want to retrieve documents with the help of the terms used for indexing, on the other. Although developed within the IT environment, thesauruses are thus primarily intended for human use.

The most complex semantic structure is offered by ontologies, although boundaries set between thesauruses and ontologies in literature differ. Some authors even tend to consider ontologies and thesauruses as one and the same, the difference being only in the purpose they are built for. For some other authors, a terminological resource qualifies for an ontology only if new knowledge can be derived from segments of knowledge already existing in the resource. According to this view. an ontology is a formal representation of knowledge, which includes a vocabulary with a set of concepts, semantic relationships between these concepts, and a simple reasoning mechanism related to a specific domain.

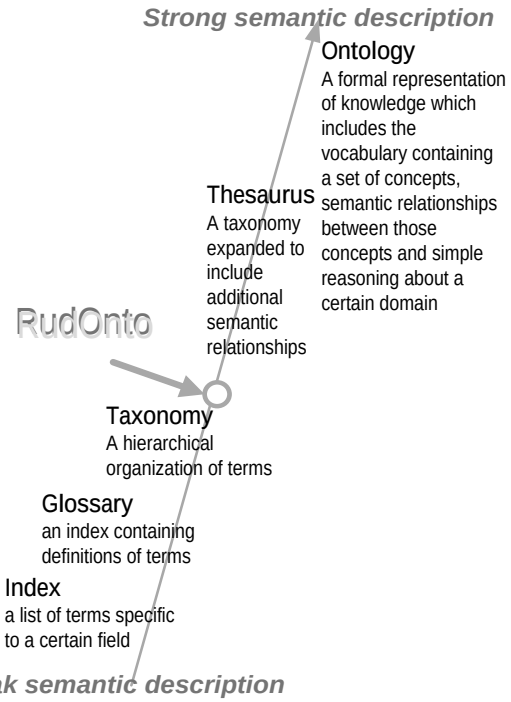

Figure 1: Semantic scale of terminological resources

Although currently realized essentially as a taxonomy, RudOnto includes some semantic relations besides hypernymy/hyponymy, thus displaying some characteristics of a thesaurus. Its further development will be focused on enriching the set of semantic relations between concepts, in order to fully promote it to the level of a thesaurus, and possibly an ontology, as anticipated by its name.

The current version of RudOnto contains close to 7000 general concepts that are most frequently used in mining engineering practice, with approximately 1600 English equivalents, and occasional translations to other languages.

The existing semantic relations between concepts are mainly those of hypernymy/hyponymy, that is, between broader (more general) and narrower (more specific) concepts. Another common semantic relation implemented is synonymy, which relates terms describing the same concept. However, as we have already mentioned, some specific relations are implemented as well, such as "has/is constructive characteristic (of)" or "has/is technological parameter (of)". By the introduction of such relations the semantic structure of RudOnto acquired features of semantic network. As RudOnto is a multilingual resource, another type of relations is introduced, namely those between equivalent terms in different languages, or the so called translational equivalents. However, although a term in one language can have several equivalents in another, for practical reasons one of them is chosen as the basic translational equivalent, whereas all others are represented as synonyms.

As we have already mentioned, RudOnto is used, among other things, for production of controlled dictionaries. A controlled dictionary is a consistent collection of terms selected for a specific purpose, chosen by its author. For

example, a controlled dictionary from RudOnto is used for labeling and annotation of cartographic content both in Serbian and English. Such annotation is enabled by the existing relations between Serbian and English translational equivalents, and by the strict relation between the dictionary and the content of the database. Some other controlled dictionaries that can be derived from RudOnto are the Geostatistics dictionary, Mine safety protection dictionary, Mineral resource exploitation dictionary, Petroleum exploitation dictionary, but also dictionaries of general terms, namely those not strictly related to mining engineering, but indispensable for information systems in this area (e.g. measurement units), as well as meta-classifications (file formats, data types, and the like).

Controlled dictionaries derived from RudOnto also serve for validation, classification and specification of the content stored in its database. In addition to that, as the importance of efficient and flexible database search and information extraction on the web is growing each day, performance of the search related to mining engineering data can be greatly improved by the use of RudOnto in query expansion [10].

Given the variety of applications of RudOnto and in view of the need to secure its functionality within the information system, an UML (Unified Modeling Language) engineering model with a special structure has been developed, whose main features are depicted in Figure 2. Assuming basic familiarity with this language we will briefly comment this model. The class *Rečnik* in the model is the lexicographic superclass whose instances are inherited. It is implemented as an abstract class, whereas concepts, represented by one basic term and several possible synonymous terms, are registered using the class *Koncept*. Concepts can be both general, common to all subfields of mining engineering, or specific mining engineering concepts.

The hypernym/hyponym relation in RudOnto is enabled by involution, so that each hyponym can appear only once in the resource hierarchy, and have only one hypernym above it. All other semantic relations between concepts are realized by the *RelacijeTermina* class. Translational equivalents of a term in one or more languages in the UML model are realized by the *MultijezičkiLeks* class. As for the source from which a concept or a term has been taken (together with their meaning), it is registered by the *Bibliografija* class, whereas the corresponding author who made the entry to the resource is registered by the *Metapodatak* class. Finally, there is a *Multimedija* class used for implementing illustrations: pictures, formulas in the form pictures, or any other relevant multimedia content. Multimedia documents proper are not entered into the resource database. Instead, they are represented by their locations on the sever (URIs) or some other metadata.

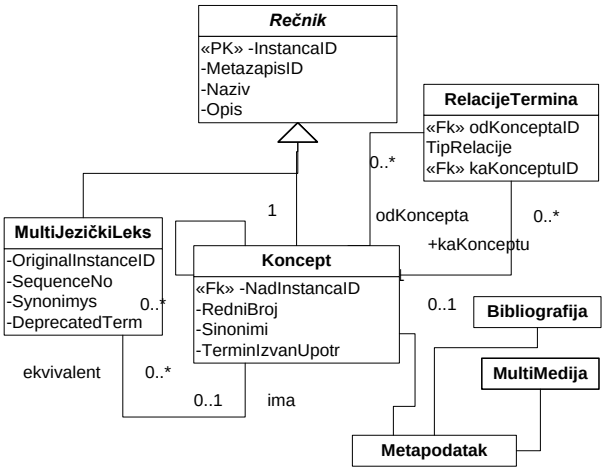

Figure 2: UML model of RudOnto structure

Figure 3 features panels from the information system module dedicated to management of RudOnto; the left hand side of the larger panel shows the hypernym/hyponym hierarchy of concepts, while the right hand side offers the full entry for the selected concept in the hierarchy. The entry consists of the basic term used for this concept, its synonyms (none in this case), and its definition. The smaller panel illustrates yet another feature of RudOnto. Namely, for each term, such a panel can be invoked, containing all available translational equivalents of a term in other languages. Thus the example in Figure 3 shows a panel with the translations of the term *Variogram* in French, German, Russian, Greek, Italian, Chinese and English.

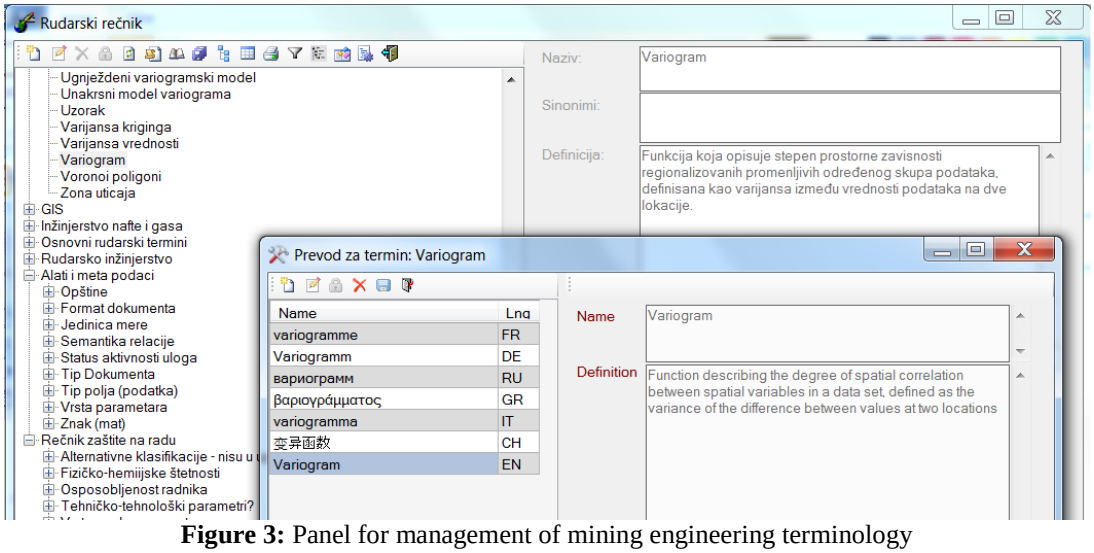

Figure 3: Panel for management of mining engineering terminology

3. MOODLE @ FMG

For several years now, blended learning is being implemented within a growing number of subjects at the Faculty of Mining and Geology [11]. The Faculty started with this implementation by developing its own FMG CMS (Course Management System) as the core of the blended learning system at FMG. In 2009, following an initiative of a number of teachers and teaching assistants, the introduction of Moodle FMG CMS was initiated.

The initial version of Moodle was 1.9.9, but migration to Moodle 2.3.1 is now underway and is expected to be completed by the beginning of the next academic year. The current installation represents a slightly customized default installation, where customization pertains basically to the appearance of the environment. In the course of development of e-learning courses on Moodle during the past three years additional modules have been added on an "as needed" basis, such as the module for mathematical formulas. By the end of the academic year 2011/12 the number of courses created on Moodle has doubled in comparison with the previous academic year, with more than 40 courses and about 830 students enrolled.

In blended learning at FMG Moodle has so far been extensively used for sharing classroom material, such as lecture notes or exercises, as well as additional material, such as external links, audio and video files, animations, etc. Activities in the form of assignments, quizzes and forums have also been introduced. Most of the courses feature course material organized by course creators in blocks, either in the weekly format or the topics format. In addition to that, Moodle is being used as a grading tool, for assigning scores to activities, and as a tool to monitor student activity and performance. Finally, Moodle also turned to be a handy tool in exchanging personal messages between students and teachers.

Terminology is an important issue in the learning process as students need to be acquainted with the proper use and meaning of terms in the field of their study. In addition to that, especially in the case of the so-called "minor" languages, such as Serbian, it is of paramount importance to make students familiar with the appropriate terminology in at least one of the "major" languages. When blended learning is implemented, where e-learning is an important part of the learning process, then the ever expanding number of available texts in electronic form on the web makes this issue even more critical.

Bearing this in mind, we have recognized the necessity of developing electronic dictionaries for Moodle in the form of Moodle glossaries. In the first, experimental phase, we decided to explore these development possibilities independently from RudOnto. Thus, we started with a glossary of basic concepts related to Geographical information systems (GIS) within the course Information technologies. For each concept separate Serbian and English entries were created. In line with the standard requirements for glossaries, besides the basic Serbian and English terms, each entry contained a short definition of the term in the respective language. However, no synonyms were taken into consideration, nor have relations between equivalent terms in the two languages been introduced. An example of an entry in English, followed by the corresponding entry in Serbian, is:

geodatabase (GDB), A collection of geographic datasets of various types held in a common file system folder, a Microsoft Access database file, or in a multiuser relational database ...
геобаза података (ГДБ), Колекција географских скупова података различитих типова који се смештају у заједничком каталогу датотека, Microsoft Access бази, или у вишекорисничкој релационој бази података ...

In this experimental phase we developed yet another glossary of statistical terms within the subject Informatics 2. When this glossary is concerned, we found it more practical to proceeded in a somewhat different manner, placing both the Serbian and English equivalents within the same entry. An example of such an entry is:

population, A population is the set of all individuals of interest in a particular study
популација, Популација је скуп скуп свих индивидуа од интереса у неком истраживању.

In the course of our experimenting we concluded that Moodle glossaries are not proper lexical resources, since the format offered by the standard distribution of Moodle does not provide for all the necessary features of a glossary. There is, hence, a need to investigate possibilities of transforming Moodle dictionaries into forms closer to proper multilingual language resources. It should start with transforming Moodle glossaries into proper (multilingual) glossaries, with the final goal of enriching Moodle with ontologies, as the importance of such resources in e-learning is growing rapidly. Namely, ontologies are in the core of the Web of knowledge – the Semantic Web, which can serve as an invaluable e-learning tool [4]. In the Semantic Web ontologies are integrated into repositories of learning objects, with the aim of organizing different concepts stored within these resources in what is known as “knowledge domain ontology” [2].

A well known implementation of Semantic Web in education was realized within the AquaRing project [1]. Namely, this project turned to ontologies in order to support annotation and retrieval of educational content, as well as enhanced education and informal learning, and to improve access to large digital repositories. To that end an educational ontology has been developed and an infrastructure based on the Semantic Web implemented. The M-OBLIGE model aimed at developing multitutor ontology-based learning environments offers another example of the use of semantic resources in education [8]. Local ontologies, which describe domains of each individual tutor, and external ontologies, which describe more general concepts, form the basis of this model. Ontology processors use these ontologies to select the

most suitable tutor for a student depending on the set of new concepts the student needs to learn. Domain expertise can be shared within this model, and it can also serve as a framework for web based integration of multiple tutors.

However, realizing that the transformation of Moodle glossaries to proper language resources is a long term goal, we concluded that at least the development of Moodle glossaries in parallel with RudOnto should be terminated. We decided to continue developing RudOnto as the main terminological resource and derive Moodle glossaries from it. Thus we settled for the current format of Moodle glossaries, and looked into the possibilities of exporting subsets of RudOnto concepts into this format. To that end we have developed a mechanism which we describe in the following section.

4. RUDONTO TO MOODLE GLOSSARY

Transformation of subsets of concepts from RudOnto as the central resource to a Moodle glossary is simple and speedy due to a wizard integrated in the information system. Figure 4 depicts a panel from this wizard, which serves for export into Moodle, but also to various other formats, namely TBX, OWL, RDF and LMF.

As hypernymy/hyponymy relations between concepts in RudOnto form a hierarchical tree, subsets of concepts form sub-trees of this tree. The export of a subset of concepts in the form of a sub-tree develops as follows. The users selects a node (concept) in the hierarchy depicted in Figure 3. This node represents the root of the sub-tree that is going to be exported. Positioned on this node, the user invokes the export wizard and selects his/her export options. The options allow the user to generate both a monolingual or a bilingual glossary, within several available types.

Without going into details of all export parameters, we will explain just the most important. First of all, the user defines the scope of the subset to be exported. The options are the export of immediate hyponyms of the selected node (its "children" - concepts at the first lower level only), or the export of all subordinate concepts (its "descendants" - concepts at all lower levels). In both cases the user can also decide either to include the selected root node in the export or to exclude it.

As we have already mentioned, the standard format of Moodle glossaries does not offer all features needed for a proper lexical resource. Hence the export from RudOnto had to be realized in line with what was available in Moodle [12]. The export of the basic term used for a concept and its definition from RudOnto to Moodle is pretty straightforward. The basic Serbian term is mapped into the Moodle glossary element CONCEPT, whereas its definition is mapped into the Moodle glossary DEFINITION field. The user can opt for either the Serbian definition only or for the definition and its translation as well. As for other elements to be exported, only the KEYWORDS element in the Moodle glossary remained. This element, represented in the corresponding XML schema by the ALIASES tag, is hence used as a repository for different types of terms: Serbian synonyms of the basic term, its available translational equivalent in the chosen language, and the inflectional forms of the Serbian term and its synonyms. Namely, as Serbian is a morphologically very rich language, there was a need to provide for all inflectional forms of terms, as they can be of importance for annotation of lessons. This morphological expansion is realized by the use of lexical resources [6] and the Vebranka web service [7]. The user can select which of the aforementioned term types will be mapped into the KEYWORDS field.

Moodle glossaries also feature the CATEGORIES element which defines to which category from a predefined set of categories the dictionary belongs. In the case of export from RudOnto the user can select from several options: no category, the category of a dictionary is the root of the sub-tree, or the category the concept is an ancestor of the root concept.

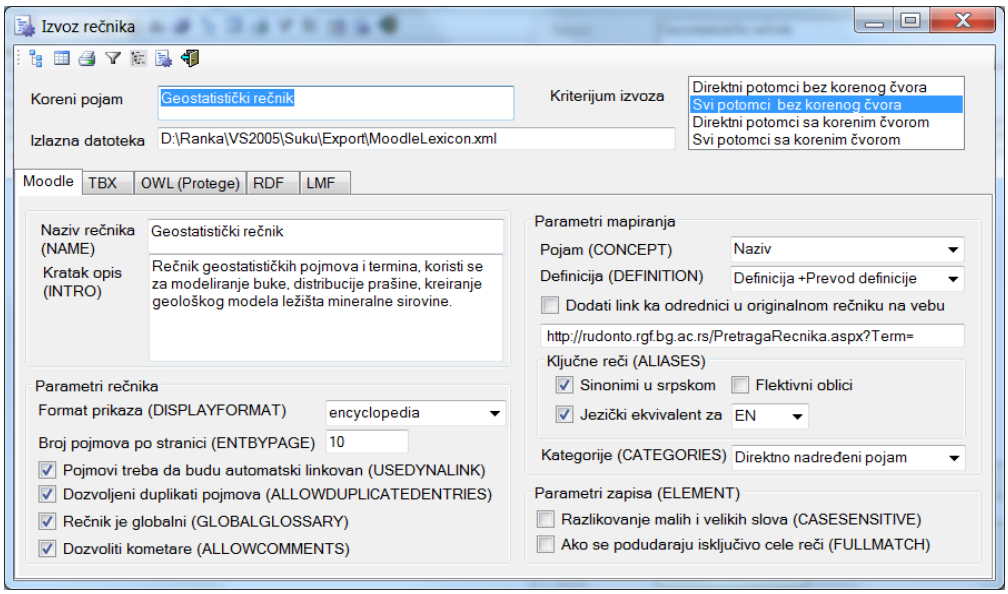

Figure 4: Panel for data export from RudOnto

Finally, the user can opt to generate for each concept exported to Moodle a link to the corresponding concept in RudOnto, as well as an URI to the concept in the original terminological resource on the web, if existent.

An example of export from RudOnto to Moodle is:

<GLOSSARY>
<INFO>
<NAME>Geostatistički rečnik</NAME>
<INTRO>Rečnik geostatističkih pojmova i termina, koristi se za modeliranje buke, distribucije prašine, kreiranje geološkog modela ležišta mineralne sirovine.</INTRO>
<ALLOWDUPLICATEDENTRIES>1</ALLOWDUPLICATEDENTRIES>
<DISPLAYFORMAT>encyclopedia</DISPLAYFORMAT>
<SHOWALPHABET>1</SHOWALPHABET>
<SHOWALL>1</SHOWALL>
<ALLOWCOMMENTS>1</ALLOWCOMMENTS>
<USEDYNALINK>1</USEDYNALINK>
<DEFAULTAPPROVAL>1</DEFAULTAPPROVAL>
<GLOBALGLOSSARY>1</GLOBALGLOSSARY>
<ENTBYPAGE>10</ENTBYPAGE>
<ENTRIES>
<ENTRY>
<CONCEPT>Geostatistika</CONCEPT>
<DEFINITION>Grana primenjene statistike koja se bavi analizom skupa podataka koji u sebi sadrže prostornu komponentu.
A branch of applied statistics that focuses on analysis of data sets containing spatial component.</DEFINITION>
<FORMAT>1</FORMAT>
<USEDYNALINK>0</USEDYNALINK>
<CASESENSITIVE>0</CASESENSITIVE>
<FULLMATCH>0</FULLMATCH>
<TEACHERENTRY>1</TEACHERENTRY>
<ALIASES>
<ALIAS>Geostatistics</ALIAS></ALIASES>
<CATEGORIES><CATEGORY>Geostatistički rečnik</CATEGORY></CATEGORIES>
</ENTRY>

5. CONCLUSION

The terminological resource features offered by Moodle in the form of its glossaries can serve as an initial form for development of such resources within this e-learning environment. However, as they are presently configured, within the standard Moodle distribution, their format is inadequate, especially when bilingual or multilingual terminological resources are concerned. Hence, there is a need to transform Moodle glossaries by adding more features to them that will make them proper terminological resources. Until this is achieved, transformation of subsets of concepts from external terminological resources to Moodle glossaries, despite its imperfection, can still be realized in a moderately successful way, as we have demonstrated in this paper. In our future work we plan to proceed along two lines: refinement of the export mechanism for generating Moodle glossaries that we have described here, and development of a Moodle module for proper terminological resources.

LITERATURE

[1] Bianchi, S., Mastrodonato, C., Vercelli, G., and Vivanet, G., *Use of Ontologies to Annotate and Retrieve Educational Contents: the AquaRing Approach*, JE-LKS, Journal of e-Learning and Knowledge Society — Vol.5, No.1, pp.211-220, 2009.

[2] Bouzeghoub, A. Elbyed, A., *Ontology Mapping for Web-Based Educational Systems Interoperability*, IBIS – International Journal of Interoperability in Business Information Systems, No.1, pp.73-84, 2006.

[3] Daconta, M. O. et al., 2003. *The Semantic Web: A Guide to the Future of XML, Web Services, and Knowledge Management*, John Wiley & Sons.

[4] Devedžić, V., *Semantic Web and Education*, Springer, 2006.

[5] Kolonja, B., Stanković, R., Vuković, F., Obradović, I., *An Approach to Management Decision Support: The Coal Mine Pljevlja Management Information System*, Proceedings of the Fifteenth International Symposium on Mine Planning and Equipment Selection, Torino, Italy, Vol. 1, pp. 436-441, 2006.

[6] Krstev, C., Stankovic, R., Obradovic, I., Vitas, D., Utvic, M., *Automatic Construction of a Morphological Dictionary of Multi-Word Units*, in IceTAL. Reykavik, Iceland: Springer, pp. 226–237, 2010.

[7] Krstev, C., Stankovic, R., Vitas, D., Obradovic, I., *The Usage of Various Lexical Resources and Tools to Improve the Performance of Web Search Engines* in 6th LREC, Marrakech, Marocco, 2008.

[8] Mitrović, A., Devedžić, V., *A model of multitutor ontology-based learning environments*, International Journal of Continuing Engineering Education and Lifelong Learning, Vol.14, No.3, pp.229–245, 2004.

[9] Nešić, G. (ed), *Rudarski rečnik. Srpsko-hrvatski. English. Français. Deutsch. Russkiĭ.*, Beograd, Rudarski institut, pp. 1292., 1970

[10] Stanković, R. Obradović, I., Kitanović, O., *GIS Application Improvement with Multilingual Lexical and Terminological Resources*, Proceedings of the 5th International Conference on Language Resources and Evaluation, LREC 2010, Valetta, Malta, May 2010, pp.2283-2287, 2010.

[11] Stanković, R., Obradović, I., Kitanović, O., Prodanović, J., Ilić, V., *An aproach to implementation of blended learning in a university setting*, Proceedings of the Second International Conference on e-Learning, eLearning 2011, September 2011, Belgrade, Serbia, D. Milošević (ed.), Metropolitan University, pp.61-66, 2011.

[12] XSD schema for moodle glossary http://moodle.org/pluginfile.php/159/mod_forum/attachment/262740/moodle_glossary.xsd