Research Article

International Journal of Advanced Computer Research, Vol 9(45)
ISSN (Print): 2249-7277   ISSN (Online): 2277-7970
http://dx.doi.org/10.19101/IJACR.2019.940101

E-Learning resource reuse, based on bilingual ontology annotation and ontology mapping

Tatyana Ivanova*
Associate Professor of Computer Systems and Technologies at Technical University, Sofia, Bulgaria

Received: 10-August-2019; Revised: 07-October-2019; Accepted: 11-October-2019

Abstract
Semantic annotation of e-Learning resources is very important for successful finding or recommendation of the most suitable ones for specific learning goals or learners. Significant research has done recently on usage of ontologies to improve learning, but most of developed ontologies are the only English language labelled and describe learning only from a specific point of view. Usage of bilingual and multilingual ontologies for resource annotation could make interlingual content delivery and reuse in e-learning more effective. It also can make learning content adaptable for a much wider audience. In this paper, we present an approach for the annotation of e-Learning resources, based on a mapped system of bilingual ontologies. We propose a knowledge-based flexible and easily extensible knowledge model and discuss how knowledge-based system, implemented this model can be used for comparison of resources, using ontology mapping. As e-Learning is complex domain that mixes pedagogy, psychology, scientific and presentation subdomains, modelling this domain is very difficult tack. We believe that relatively independent modelling of all the subdomains and specifying relations between them is the most promising approach. Our ontological model aims to ensure strict separation of different type knowledge, used in the learning process (pedagogical from domain-specific, general from domain-specific, linguistic from semantically–rich). This can simplify the ontology building process, ontology reuse, ontology evaluation, and also comparison of e-Learning systems, annotated by ontologies, following this model.

Keywords
Document metadata, Ontology, Bilingual ontology, Knowledge-based model, Ontology-based e-learning resource annotation.

1.Introduction
Explicit semantic modelling of the knowledge has an important role in e-Learning. Explicit semantic relations may help users of educational content to understand the domain or to navigate overall e-Learning system. Links to similar content are important for finding additional learning resources. Explicit links between concepts are useful for easy comprehension and understanding the subject. Today international languages (as English and Russian) will propose the highest quality learning content, but others prefer to use learning resources, information and knowledge, presented in his native language. Ontologies are more and more frequently used in e-Learning systems, to support almost all e-Learning tasks and resource reuse. e-Learning standards, learners, learning content, learning process, and more, are modeled, using ontologies in many projects and in such a way personalization, flexibility, and resource reuse are achieved [1]. But almost all ontologies developed and tested in e-Learning projects use lexicalization only in English language. These ontologies are not usable for describing and annotation of non-English learning content. The semantic representation of metadata in almost all-natural languages are needed to support search and delivery of the most suitable resources for every learner and learning tack. Recently, the increased usefulness of ontologies for practical applications has led to building some ontologies that use other language as a base human language (for representing labels, comments, textual definitions in ontologies) [2−4]. But most ontologies in e-learning projects continue to use English-based lexicalization.

*Author for correspondence

Many ontologies have been developed to support e-Learning tacks, including resource annotation, but almost all of them are rarely used in practical e-Learning systems. There are several reasons that make them difficult to use: Almost all ontologies use English language terminology, but learners prefer his native language; most of ontologies describe resources only from a specific point of view; ontologies are not mapped and mapping process is complex and error-prone. The goal of our research is to show how ontologies lexicalized in two languages (bilingual ontologies) can benefit e-Learning. We present an approach for the annotation of e-Learning resources, based on a mapped system of bilingual ontologies. Our aim is to present an ontological model that ensures strict separation of different type knowledge (pedagogical from domain-specific, general from domain-specific, linguistic from semantically–rich). This can simplify the ontology building process, ontology reuse, ontology evaluation, and also comparison of e-Learning systems, annotated by ontologies, following this model.

We propose a mapped ontology-based flexible and easily extensible knowledge model and discuss how knowledge-based system, based on this model can be used for comparison of resources, using ontology mapping. The structure of this paper is as follows: In the next section short discussion of related works is made. Section 3 presents our approach for the annotation of e-Learning resources, based on a mapped system of bilingual ontologies. Section 4 discusses how knowledge-based system, implemented this model can be used for comparison of resources, using ontology mapping. Section 5 concludes the article and discusses some aspects of future research.

2.Analysis of used materials and related methods

The main related area is ontology-based e-Learning resource description and annotation. As our model uses mapped ontologies, and mappings between ontological multilingual systems, the mapping of bilingual and multilingual ontologies is also important area, related to our research.

2.1Ontology-based e-Learning resource description and annotation

In E-learning systems and repositories, learning objects (LOs) are described (and annotated) using metadata. Several standards are developed and used for describing and annotating, e-Learning resources: IEEE LOM (Learning Object Metadata: http://ltsc.ieee.org/wg12), Dublin Core Metadata Initiative, or DCMI (dublincore.org), IMS Learning Object Metadata (IMS LOM) (http://www. imsglobal.org/metadata), Canadian Core Learning Resource Metadata Protocol CanCor (http://www.cancore.ca) and UK Learning Object Metadata Core (http://www.cetis.ac.uk/profiles/ uklomcore/ uklomcore_v0p2_may04.doc). The Dublin Core Standard is not mainly for educational metadata, but it defines some relationships with different standards. All these standards are mainly related to resource structure, presentation and pedagogical points of view. They do not include structured representation of the learning domain terminology.

Many attempts are made to represent e-Learning standards in a semantic way using ontologies. In [5], the authors introduce a mapping of the standard IEEE LOM in the ontology language web service modelling language (WSML) to provide a basis for translating existing IEEE LOM metadata records to WSML and to serve as a basic learning object ontology. The ontology web language (OWL) OBAA metadata ontology developed in [6] represents all the metadata from IEEE LOM standard. OWL LOM ontology, developed in [7] offers a shared vocabulary, giving a common semantics for any application which uses LOM.

In the resent semantic-based e-Learning metadata description approaches several extensible standards are combined. For example, in [8] a VideoAula, OBAA (extensible e-Learning metadata standard), IEEE-LOM, and IMS AccessForAll ontologies are used to get the metadata’s semantic. All these ontologies are developed using English language terminology and are mainly for educational and structural purposes (they are not intended for a comprehensive description of the e-Learning domain).

Attempts to translate the educational standards, conceptual data schema into a variety of European languages is presented in [9]. Most of these translations have not been made publicly available in finalizing form. The idea that the original English versions of these names and values should be regarded as linguistically neutral and equivalents in alternative languages should be provided through the user interface or any other mechanisms is proposed in [9] and widely accepted. We have found just a few papers, describing research on multilingual ontologies in e-learning. Lightweight domain ontologies are used in [10] for LO annotation. Ontologies are represented in a simple knowledge organization system (SKOS), and contain at about 1700 concepts. These concepts were translated semi-automatically by providers of educational content with the help of machine translation into French, Spanish, German, Italian Lithuanian languages. The system BONy [11] is a cognitive mobile e-Learning management system (LMS) that supports a multilingual access to information by the ontological representation of knowledge and an interconnection among learning objects accordingly to Semantic WEB methodology, best practices and standards. Italian, English, Spanish, Greek, German, Polish, Hungarian, Slovakian Czech and Catalan languages are involved.

Multilingual ontology for e-Learning is developed in [3] by translation of English monolingual domain ontology semi-automatically by providers of educational content with the help of machine translation into 5 languages (French, Spanish, German, Italian and Lithuanian). This multilingual ontology is used to annotate textual content of LOs. In such a way the multilingual ontology is used for ensuring sharing and reuse of learning resources.

The LT4eL project [2] improves the accessibility of e-Learning resources by integrating semantic knowledge to enhance the management, distribution and retrieval of the learning material. In this project two-stage approach for semi-automatic extending of multilingual domain ontology is proposed. On the first stage the developed ontology extends to new domains. Then the mapping between the ontology and various lexicalizations has been carried out to enhance the ontology with the other language terminology.

Semantic annotation of e-Learning resources, based on pedagogical or domain ontology is still underdeveloped for Bulgarian. Moreover, it is difficult to find good domain ontology, including terminology in Bulgarian. Language tools as morpho-syntactic tagger for Bulgarian (working with more than 95 % accuracy), a dependency parser (with more than 84% accuracy), a general clunker and a named entity grammar are available [4]. Ontology-based lexicon for Bulgarian also is presented in [4].

Lack of widely-used standards in the ontological representation of e-Learning metadata and difficulties in ontology development, maintenance and evolution are the main problems in ontology-based e-Learning resource-annotation area. Developed ontologies in e-Learning are related to different standards and are labelled mainly in English, and rare-in some other international languages. To bridge the standardization gap, mappings between ontologies, developed on different standards, should be performed. And labelling in national languages (as Bulgarian) of some ontologies in e-Learning should be performed to make them usable for annotation of learning content in different national languages.

It is clear, that ontology mapping is crucial both when we develop complex ontological systems and when use resources, described by different ontological systems. There are grand numbers of researches, related to ontology mapping. We will make a brief survey of the mapping approaches, related to e-Learning resource annotation and take special attention to the effects of usage of ontologies, having labels in two or more languages in the ontology mapping process.

2.2Multilingual ontology mapping

Ontology mapping is a process of creation and maintenance of alignments between elements of two ontologies covering overlapping areas of knowledge. Formally, an alignment (or mapping) is a quintuple $m=(id, t_i, t_j, alType, sim)$, where *id* is the alignment identifier, t_i and t_j are homogeneous ontology terms from different ontologies, *s* is the similarity degree of *m* (sim∈[0,1]), alType is a type of alignment: the semantics associated to an alignment. The alignment type can be subsumption, equivalence, part-of or any other semantic relationship. There is some confusion about the usage of the term's ontology alignment, ontology mapping and Ontology mapping. The terms ontology alignment and ontology mapping can be used both for referring to the process and results of the process execution.

Language aware or multilingual ontology mapping is defined in [12] as a type of ontology mapping where the matcher is capable of dealing with ontologies expressed (or labelled) in multiple languages. Can usage of labels in two or more languages make mapping more easily, or will add new problems? We have made a short analysis of the recent research on multilingual ontology mapping trying to answer this question.

There are only a few researches on multilingual ontology mapping (e.g. [13−16]). Multilingual ontology mapping usually uses traditional ontology mapping algorithms and multilingual textual or thesaurus-based linguistic resources. So, multilingual ontology mapping requires revision of some of the mapping algorithms to make them well-working in a multilingual environment. Most of the approaches to map ontologies, having labels in different languages relies on general-purpose machine translation services as Bing or Google [14, 15]. The main problem of this translation is that it is statistically and machine-learning-based and works well only on corpora similar to those on which the system was trained. So, translation accuracy of specific domain classifications and ontologies that contain specialized terminology is low. If ontologies use short labels or non-standard orthography and syntax, this makes the translation task more error-prone.

Some other approaches are based on semantic mapping. Ontology labels are parsed by multilingual natural language processing and then matched using language-independent and domain-aware background knowledge acting as an Interlingua [12].

Multilingual ontology mapping has several different aspects:

- Mapping the ontology, having labels in one language to ontology, having labels in other language (cross lingual ontology mapping);
- Mapping the ontology, having labels in several languages to ontology, having labels in other different languages;
- Mapping the ontology, having labels in several languages to ontology, having labels in other languages in the case that the two ontologies use at least one common language;

The first case is about mapping monolingual ontologies, having different natural language labels and it is closely related to label translation. Ontology-based cross-language mapping is a process of establishing correspondences (find relations) among the ontological resources from two independent ontologies where each one is lexicalized in a different natural language [16]. This is the Cross-Lingual Ontology Mapping (CLOM).

According to [14], CLOM strategies can be grouped into several categories:

- Instance-based approach;
- Schema-based;
- Manual processing;
- Corpus-based approach;
- Based on linguistic enrichment of ontologies.

The *manual* CLOM process has higher precision and recall, but it is difficult, expensive, and is infeasible for large, complex or frequently changed ontologies.

The *schema-based approach* techniques consider the location of the concept in the ontology structure (e.g., tree, graph) and how the mappings of concepts can contribute to the mappings of adjacent concepts.

The *instance-based* approach requires rich sets of instances embedded in ontologies, and it is not satisfied for mapping terminological ontologies.

C*orpus-based* approach is used in several researches, using various natural languages and textual corpuses. In [17] for example, it is intended to align the English thesaurus WordNet and the Chinese thesaurus HowNet. It relies on bilingual corpora and domain-specific ontologies. Architecture, based on the use of Information Retrieval techniques for suggesting alignments between two or more multilingual ontologies is proposed in [18].

A *linguistically motivated* mapping method is proposed in [19], advocating a linguistic-driven approach in the ontology development process that generates enriched ontologies with human-readable linguistic resources. Such enrichment is difficult to apply automatically because it is not standardized. A minimal API for multi-lingual mapping is presented in [20]. Two strategies are implemented: direct (translation-based) and indirect. If two languages derive from the same root language, they have a similar vocabulary. In such cases, a direct mapping can be performed. The direct alignment strategy of multilingual ontology mapping is usually based on machine translation. Trojahn et al. [13] presented a translation-based multilingual ontology mapping framework, where ontology labels are first represented by collections of phrases in the target natural language. Matches are then generated using specialized monolingual mapping agents that use various techniques (i.e. structured-based mapping algorithms, lexicon-based mapping algorithms and so on). The basic idea of indirect alignment between multilingual ontologies is to compose alignments which already exist [20]. The indirect alignment strategy is based on composition of alignments. Indirect alignment in [21] is made as an alignment between French and Portuguese ontologies by using intermediary alignments in English, i.e., French – English and English-Portuguese alignments. According to [22], there is no integrated solution in automated ontology mapping that is a clear success.

In many mapping projects, various techniques are combined in order to generate high quality mapping results [23]. Used techniques and the strategy of its usage highly depend on the ontologies, mapping goals and available external resources. When ontologies have no common language lexical elements, translation also is one of the important steps.

2.3Mapping ontologies that use different natural language labeling, but having at least one common language

In this case, common language can be used as a basis for monolingual mapping, and used lexical elements of other languages allow reducing the problems raised when two different concepts have the same label [18]. Multilingual resources in some cases provide term translation mechanisms that have already been adapted to the represented domain and can help in solving some ambiguity problems [24]. Machine learning approach for mapping two ontologies using a small set of manually aligned concepts is proposed in [25] and the authors show that multilingual information can improve the mapping quality. The quality of machine translation systems is limited and depends greatly on the pair of languages considered. As a consequence, a pure translation-based approach is not sufficient to find a significant number of mappings [16]. As semantic resources and all the research, related to ontology mapping are performed mainly using English natural language, using ontologies, labelled both in Bulgarian and English will ensure both semantic knowledge representation in Bulgarian and usage of all the results from English-based ontology engineering and mapping.

Using one or other mapping strategy depends on the available resources (textual resources quality, alignments, dictionaries and translators) and features of the languages the ontologies are written. Statistics from 2010 show that the number of non-English speaking Internet users is almost three times the number of English-speaking ones [26]. On the other hand, very small number of ontologies use lexicalization of language, different from English. So, if we wish to make easier and efficient mapping as a result of usage of bilingual ontology labelling, we also must think about how to ensure the development or evolution of such type ontologies in the interested languages. It is easy to find good ontological resources written in English in some domains, but most of them are tack-specific, and its usage for other tracks require further development. There are some other languages that are objects of some such a research in recent years. MultiFarm dataset has been designed as a benchmark for multilingual ontology mapping. The MultiFarm dataset is composed of a set of ontologies translated in eight different languages – Chinese, Czech, Dutch, French, German, Portuguese, Russian, and Spanish and the corresponding alignments between these ontologies. But such dataset is not proposed in Bulgarian. So, using semi-automatic mapping tool to make easy for human experts to recommend candidate alignments between multilingual ontologies or evaluate automatically proposed ones is important for ensuring the high-quality mapping and to support the evaluation of the mapping quality. In the next section we propose a flexible and extensible model for annotation and comparison of resources for learning, based on Bulgarian-English bilingual ontologies.

3.Results of our work-ontological model for describing learning resources

3.1Description of the model

There are many researches about using ontologies for semantic description of e-Learning resources, but there is no widely accepted standard. Describing e-Learning content is difficult because of its complexity. e-Learning is a complex domain, including some high-quality domain knowledge, pedagogical, psychological, linguitstical knowledge, as well as some knowledge about computers, internet, e-Learning standards, etc. Every subdomains of e-Learning are important, but various projects take in mind only some of these subdomains, which are the most important for he's project objectives. Every learning resource is about the specific learning domain and is organized, following some pedagogical notations. It is important for resource comparison to describe semantically its specific content, pedagogy, and show how they are related to general scientific knowledge, and pedagogical theories. So, the needed semantic description should contain all learning domains-related details and links to more general knowledge, as well as pedagogical details and links to general pedagogical theory.

We propose a flexible ontological model that can ensure comprehensive semantic descriptions of e-Learning resources, sufficient for finding the needed resources, ensuring needed information for automated resource comparison and recommendation that can reuse previously developed domain or pedagogical ontologies. Our leading principles are:

- Deep and comprehensive knowledge modeling;

- Ensure usage of multilingual learning resources, including Bulgarian language texts;
- Ensure modular usage only of the needed semantic descriptions for every resource;
- Maximal possible automation of ontology and mappings generation;
- Ontology reuse.

Our aim is to represent ontologically both deep semantic of e-Learning domain and it's pedagogical, lexical and multilingual specifics. Our research is related to the bilingual domain of Computer science (in Bulgaria it is bilingual, as using English language resources is a must for every expert in this area), but many other learning domains (as Bioinformatics, Medicine…) are bilingual in many countries. So, we propose an ontology-based model that describes specific relations between learning concepts, as well as general interdomain relations and pedagogical specifics, in the context of English and Bulgarian language lexicalization.

As computer science has many loosely related branches, we believe that a single ontology in the computer programming area will contain a lot of unnecessary information. A system of mapped ontologies can describe clearly all the elements of the specific programming subdomain and its relation to other computer science domains. We propose the system of bilingual ontologies for different programming languages mapped to global computer science ontology.

Our model includes the following ontologies (*Figure 1*):

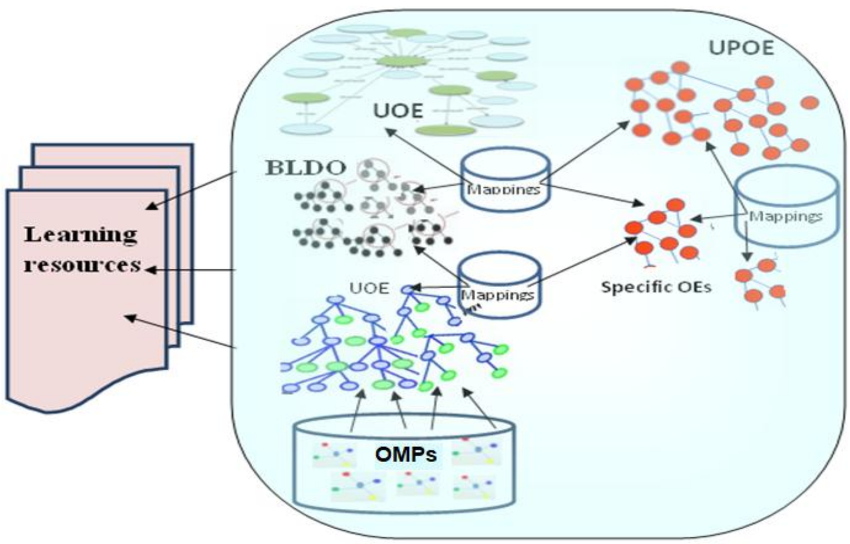


Figure 1 The ontology-based model for annotation of learning resources

For representing pedagogical knowledge:
- Upper pedagogical ontology for e-Learning (UPOE)
- Specific pedagogical ontologies (specific OEs)
- Upper ontology for e-Learning (UOE)
- Bilingual linguistic domain ontology (BLDO)
- Comprehensive domain ontology (CDO)
- Ontology Mapping patterns (OMPs)

Pedagogical ontologies include all general and specific pedagogical terminology, important for the learning. It may include information for the e-Learning content, pedagogical structure (which prior knowledge is necessary to understand or what type of interaction with the learner is used), information about learning goals, target groups of learners, etc. UPOE includes only general terminology and specific learning components or theories are described by specific OEs.

We use *UPOE* of mapping e-Learning standards and approaches. It contains general terms, included in Instructional ontologies (representing different instructional models, learning theories, approaches), and Authoring ontologies (modeling authors' activities, the logical structure of the learning content). We believe that it is a good idea to standardize this ontology and use it in description of many courses or other learning content. Translation of English labels to Bulgarian should be performed (or evaluated) by experts and teachers.

One and more specific *OEs* should be mapped to UPOE and used for a detailed description of important pedagogical issues of describing resources according to its specific pedagogical properties (as personal preferences, learning styles, goals, the concept of prior knowledge, etc.). These pedagogical ontologies represent clear definitions of important for the learning resource pedagogical issues in semantic machine-processable way. Specific pedagogical ontologies can represent terminology of specific learning approaches and standards. Different specific pedagogical ontologies can be used for describing resources from different pedagogical points of view. For example, learner profile ontology, curriculum ontology, or some IMS Learning Design (LD) ontology is used to describe corresponding standards or participants of the learning process.

We can map to UPOE one or more pedagogical ontologies containing terms, related to the concrete course instructional theories (instructional model, learning approach, goals, assessment, e.g., tests and performance tracking collaboration, e.g., group formation, peer help, educational adaptive hypermedia learners and learners' activities models, etc). It can represent in details whole pedagogical model of the course. Pedagogical ontologies described clear pedagogical theories should be standardized.

Translation of English labels to Bulgarian should be performed (or evaluated) by experts and teachers. The UPOE ensure terminology for mapping between several specific ontologies, describing comprehensively specific pedagogical theories. For example, if our learning content is intended for programmers, dyslexia learning disability terminology are not needed to include in the semantic description of the content, as well as social learning concepts are not so important in e-Learning course for artists. So, modular architecture of pedagogical ontology will ensure links between various pedagogical theories and comprehensive representation of the used theory specifics in a relatively small semantic pedagogical model.

All possible mappings between specific pedagogical ontologies and UPOE have to be done to ensure explicit specification of the relationships between the described course pedagogy, and other pedagogical issues. This will support pedagogical comparison of different resources and is the most useful when learner have specific personal preferences, learning style, or disability (for example, dyslexia).

Upper ontology for e-Learning (UOE) contains general domain terms, related to the e-Learning and semantics of the learning domain that can be found in OntoWordNet. Ontology development steps, discussed in [2] can be used to develop this ontology. In this ontology many general domain terminologies, related to various domains may be added. We develop our upper ontology by managing and mapping elements from general domain ontologies in E-learning and computer science domain (as an ontology of software design domain [27], ACM classification, IMS LD ontology [28] and competency-based ontology [29]). Topic hierarchy of computer sciences is important to be bilingual, as Bulgarian learners know well only Bulgarian terminology, but search engines work better when English language terminology is used. Web translation services as Google, Bing, or WebTrance (http://webtrance.skycode.com/) can be used to translate English labels to Bulgarian. The quality of such translation is good. Evaluation can be done by usage of learning textual resources (if they contain terms both in English and Bulgarian), or by involving teachers or learners.

The main role of the *UOE* is to support broad semantic interoperability among a large number of domain-specific ontologies, and e-Learning ontologies (including pedagogical, psychological, about e-Learning standards, etc.), used for semantic description of e-Learning courses. It contains terms, useful in mapping general terms in various domain ontologies and general terms related to e-Learning.

Bilingual linguistic domain ontology (BLDO)

A linguistic ontology can be seen both as a lexicon and as ontology [30]. It usually contains the structured presentation of a domain terminology. Our BLDO is lightweight domain ontology as it is constructed for a specific domain of the learning content. It covers used and defined concepts and relations in a learning content expressed by using labels in Bulgarian and English natural language. This ontology presents the structure of the course knowledge (or learning content) as lexical items related to their meanings without comprehensive modeling of deep semantic of the learned knowledge (it fir example does not contain concept definitions, or all the relations between the concepts). In the general case, small changes in the concept's properties, relations, or order of learning will not lead to changes in this ontology. In this way, course evolution can be supported without any needs of frequent changes in big ontologies. A simple knowledge organization system (SKOS) language is one of the best languages for the development of multilingual lightweight ontologies. SKOS allows easily labeling with lexical strings in several natural languages. SKOS define properties skos:prefLabel to specify a preferred string label (in English, for example) and skos:altLabel to specify an alternative string label for a concept. SKOS can easily connect different representations of the same concept in multiple languages. SKOS proposes semantic relations between concepts, as skos:broader, skos:narrower, and skos:related. This ontology concepts and relations which are neither part of the upper layer, nor of the specific domain ontology, are important for the alignment between them. For example, 'loop for' is at the domain layer for C++ programming ontology and 'programming code' is in the upper layer, but the concept of "repeating the code" which is more specific than 'repeat' and more general than "looping" is in the Linguistic domain ontology. In our view, this ontology is a useful instrument for mapping the specific course domain to the upper ontology. Alignment of the domain ontologies to an upper ontology is important as it will ensure inheritance of the knowledge represented in the upper ontology. In such a way more general constraints will reflect the structure of the domain.

Our BLDO allows abstraction, a Natural Language (NL) independent representation of knowledge and connects different syntactic representations of concepts and relations (in two natural languages). Other NL presentation of ontology elements can appear: in URIs, in labels (using rdfs:label, skos:preflabel, etc.), in NL definitions, comments, and at the same time, English language labeled ontological elements can be seen as a natural language independent knowledge representation (as English is a widely used language).

Comprehensive domain ontology contains domain terms that cannot be found in WordNet, including multiword concepts. This ontology represents not only the hierarchy of learning concepts. It also includes properties and relations in the way that they are presented in the learning content. So, this ontology represents a comprehensive semantic model of the learning content. Mappings between these strictly specific terms and related to them general domain terms in the Upper ontology for e-Learning is important for semantic annotation, comparison and search of the e-Learning content. Parallel labeling in Bulgarian and English is useful in resolving of natural language ambiguity problems. Currently, most of the developed domain ontologies include proper lexical terms and expressions only in English. Manual enrichment with Bulgarian language elements will be difficult, time-consuming and expensive tack. Automatic linguistic enrichment of ontologies is proposed in [19]. It uses linguistic resources and human intervention for supervising the process of the enrichment of the ontological content with proper lexical expressions in natural language.

As in Bulgarian texts for e-Learning bilingual terminological expressions frequently coexist (for example, after many terms in Bulgarian, its English variant is presented in brackets), and definitions are clearly formulated, e-Learning resources are very useful source in the ontology enrichment process. So, ontology enrichment will beneficiate of a clarity and greater linguistic expressivity of the e-Learning textual sources to obtain proper translations in Bulgarian for ontology concepts and roles.

To develop bilingual English-Bulgarian domain ontology in the computer science domain, we use enrichment strategy, very similar to the proposed in [19]. We first perform mapping process from English terminological ontology to the used English terms in e-Learning texts, and then enrich ontology, using closely-related to these terms' Bulgarian equivalents and definitions. Involving students in this process is very useful both for understanding logical links between learning concepts and for ontology enrichment. We also use linguistic tools (as English-Bulgarian dictionary http://www.ectaco.co.uk/English-Bulgarian-Dictionary/) and semi structured sources as Wikipedia in the enrichment process. We will present conclusions about the usability of Wikipedia in ontology mapping and enrichment process later.

Other multilingual language sources as Bulgarian WordNet also can be useful in enrichment or mapping process, but they are not publicly available. Ontologies, enriched by translation of terms (including idioms and multiword terms), natural language descriptions of concepts, presence of synonyms, will offer useful resources for supporting an ontology mapping process. *Ontology design patterns (ODPs)* are used for modeling deep semantics of learning concepts. If two learning objects contain one and the same concept, but its definitions are different, this may be important, or not. For example, differences between defining the square as rectangle, having equal sides, or as rhombus, having right angle, are important for the strict structure of the mathematical theory, but not so important for calculating its area. But ODPs can be used for explicit semantic representation of these differences. The quality and the specifics of ODPs are important for ensuring deep semantic comparison of learning content. Mapping ODPs, describing complex concepts, must contain information about similarity and difference in its semantic (definitions, relationships, pedagogical aspect). Comparison and mapping of the ODPs, used in two similar courses are important for comparing the deep semantic of e-Learning content. As there is no unique widely accepted strict definitions for many concepts in the computer technology area, these ODPs are very useful in the process of comparison of similar or closely related courses.

In the learning context, it is important not only to know to what concepts, learning is targeted the course, but what are the exact definitions of these concepts. Various courses can use different definition approaches, or definitions may have different granularity or clarity. Mappings between ODPs, used for explicit representation of basic course concept definitions are very useful in deep semantic comparison of the learning content. The proposed ontological system is intended for e-Learning resource annotation, searching and comparison. In the next chapters we will discuss annotation and mapping specifics.

3.2The e-Learning resource annotation

Every Learning Objects Repository (LOR) can use its own metadata standard. And our resource also cannot be from some LOR, it can only be some text from a web site, or tutorials. So, metadata, needed for annotating different resources is very different. Our model can support annotation of different resources by importing of its describing domain and pedagogical ontologies (BLDO, CDO, OMPs, specific OEs if exist) and mapping learning content elements to classes or relations in these ontologies. One and the same Upper ontologies (domain and pedagogical) can be used for describing various resources, and its specific ontologies should be mapped to the upper ontologies. When e-Learning resources are developed by different teachers or experts, or when they contain significant knowledge differences, some of the ontologies, included in its semantic description models may be different.

Placing labels in both Bulgarian and English languages have following benefits:

- This will support easy comparison of our Bulgarian language course with English language courses in related domains. The best learning and scientific resources in the computer programming area are written in English.
- This will support search both by the usage of Bulgarian and English language terminology. The most suitable language for searching in internet is English.
- This will encourage learners to memorize and use both Bulgarian and English language terminology. As we know, the usage of the English language is a must for Bulgarian experts in computer science.
- Parallel labeling is useful in resolving of natural language ambiguity problems in the mapping or ontology management process. For example, the Bulgarian word ”функция” may be translated as function, procedure, role, method, etc., and all these translations have different meaning, but when the label “method” is specified as a parallel English language label in Java programming ontology, the meaning will be only one, and the term ambiguity will be removed.

For comparison of two learning resources, described using different systems of mapped ontologies, we must perform mappings between the used ontological systems.

3.3Mappings, mapping strategies and comparison of resources described by different ontological systems

Ontology mapping is crucial both for development of the ontological system for describing e-Learning resources (internal mappings), and for comparing of resources, described by different ontological systems
Internal mappings between ontologies, included in the model are described above. Every learning course or single resource comes with annotations and uses his own ontologies. Here we will discuss mappings between ontological systems, describing different resources, needed for resource comparison.

We present a suggestion-based mapping approach that uses lexicons, thesauruses and web-based sources (as Wikipedia, DBpedia) as external knowledge. Bilingual linguistic resources-based methods for generating mappings and validating them by experts or teachers are used for mappings between ontologies, describing different learning resources. We also believe that proposing easy-to-use graphical interface for involving learners, teachers and other users in the mapping process will ensure correct mappings.

The proposed mapping process consists of six stages:

1. Mappings between Upper ontologies, if different upper ontologies are used**.** In many cases, resources use one and the same upper ontology.
2. Mapping between Bilingual linguistic domain ontologies uses Corresponding English-based glossaries.
3. Mapping between precise domain ontologies.
4. Mapping between pedagogical ontologies.
5. Mapping between ontology design patterns. ODPs encapsulate the complexity of the concept definitions and mappings between ODPs describe comprehensively relations between similar concepts, closely-related concepts, or make possible comparisons of different definitions of the same concept. Score computation similarity term and relations with ODPs.
6. Ontology mapping evaluation.

Mapping between Bilingual linguistic domain ontologies is important to find general thematic similarity between resources. To find similarity between linguistically different labels, structural comparison, synonyms, Bulgarian language labels and English-based thesauruses as WordNet are used. In many cases, machine translation services as Bing or Google are very useful in the cases when bilingualism is only partial (some of the labels have no equivalent in other language). We use top-down mapping strategy. Upper-level differences in many cases indicate a significant thematic difference.

Mapping between precise domain ontologies is important in a comprehensive comparison of proposed learning content. Problems may be occurred in parsing labels using syntactic NLP techniques, as labels can be short or can use specific way of coding of natural language terms. So, NLP processing, adapted to each natural language should be supported, and domain and context aware label disambiguation in multilingual environment will be useful.

Statistical or fuzzy mapping also can be used when one-to-one translations or mappings are not possible (for example, when the meaning of closely related terms in different courses is slightly different). Example: procedures and functions in Pascal both are functions (but clearly distinguished type functions) in C++ and methods in Java. Data type system and type conversion specifics are very different in C++ and Visual basic, for example, but there are global concepts as "**weakly typed" and "strongly typed"** Usually different Part-of-speech I s different ontological element (nouns represent classes, verbs represent relations) according to [31], Part-of-speech (POS) information has been shown to solve 87% of all word ambiguities in English-language text. So, using only class labels when mapping between classes is performed, and only relation labels when napping between relations is done, leads to significantly reduced of ambiguities. Comparing equivalent word(s) (labels) in Bulgarian is another source of mapping information.

Mapping between pedagogical ontologies is important for personalization and recommendation of the content for concrete learners. It is a good idea to use one and the same upper ontology for e-Learning in all the ontological resource descriptions. Other used pedagogical ontologies can be very different for different resources. For example, in one system learner profile ontology may be used, in other – curriculum ontology, or some IMS LD ontology. Ontologies also can be based on different e-Learning standards. So, mapping, and conclusions, followed from it should be case-based, but it is good to develop some standards in this area. Conclusions about usability of resources on the base of pedagogy-related description may by very important. Resources for children or dyslexics for example, would not be useful for professional programmers, and documentation-based professional texts would not be understandable (and recommendable) for beginners.

Mapping between ontology design patterns can give two types of information:

- About defined concepts in the learning content;
- About similarities in definition of defined concepts in the learning content.

This information can be used to score computation similarity term and relations using ODPs. Ontology mapping evaluation will depend on the type of mappings, but it is a good idea to involve specialists (experts, teachers, learners) as a valuable part of the evaluation process and checking conclusions, derived from performing mappings.

Our approach can be implemented by two main components:

- A back-end module implementing an ontology mapping and linguistic resources-based techniques for suggesting sets of candidate mappings;

- A set of user interface facilities that have for supporting learners, teachers or experts to perform mapping activities as a part of his work or learning.

We can use learning resources in the mapping process to disambiguate unclear labels of concepts and relations. Mapping statistics will provide valuable information about similarities and differences between learning courses.

4.Use case and discussion

We will describe how we have used our model and the corresponding ontological system, to describe learning resources in the software development domain.

To evaluate our model, we have developed ontologies that describe two learning curses, one for programming in C++, and other for programming in Java. The two courses are targeted for training junior programmers. So, pedagogy is the same, and we only develop two different BLDO, CDO, some OMPs and mappings between them. As there is no unique widely accepted strict definitions for many concepts in the computer technology area, these OMPs are very useful in the process of comparison of similar or closely related courses. Mechanisms for dynamic updates of domain ontologies or (semiautomatic) enrichment are needed because of computer science domains are rapidly changing areas. A static ontology will soon be outdated.

Bilingual domain ontologies and OMPs, used for annotation are also useful in the learning process. They can be used by teachers for supporting test generation and directly by the learners as a visual representation of learning concepts and its relationships. We develop lightweight bilingual linguistic ontologies using ontologies in our previous research and enriching them with additional Bulgarian language labels. Linguistic ontologies contain at about 80 terms and 24 relations each, and precise domain ontologies - at about 60 terms and relations.

We're currently experimenting different techniques for performing semiautomated mapping and ontology enrichment. Approaches, using and multilingual lexical resources are complementary. We investigate the idea of combining machine translation and terminology extraction from e-Learning resources, containing some English language terms. WordNet and other language (linguistic semantic and syntactic) resources can be used, but only general or frequently used programming terminology can be found in WorldNet. We also have done some evaluation of the usefulness of Wikipedia for automated extraction of additional knowledge, related to learning programming in C++ and Java. We search and browse materials for at about of 100 concepts, used in our ontologies, distributed as follows:

6 concepts from Upper Domain Ontology (UDO) used only for Java programming as "servlet", "virtual machine", "applet", "run time environment".

6 concepts from UDO, used only for C++ programming, as "header file", "resource file", "dl library".

20 concepts from UDO that are common Java and C++, as "algorithm", "program", "programming approach".

8 concepts from Linguistic Domain Ontology only for Java, as "bytecode", "reference", "applets life cycle", "referent datatype".

8 concepts from Linguistic Domain Ontology only for C++, as "pointer", "address arithmetic", "structure".

20 concepts from Linguistic Domain Ontology, common for Java and C++, as "class", "object", "datatype", "array", "statement", "operator", "loop", "condition".

8 concepts from Specific Domain Ontology only for Java, as "overloading", "overriding", "default access mode", "boxing".

8 concepts from Specific Domain Ontology only for C++, as "Incrementing a Pointer", "decrementing a Pointer", "short int", "long int".

20 concepts from Specific Domain Ontology, common for Java and C++, as "public access", "private", "constructor", "loop for", "iterator expression", "call statement", "return type".

We are interested in finding good definitions or explanations for these concepts in Bulgarian and English. Our results are shown in *Figure 2* (for Java), *Figure 3* (for C++), and *Figure 4* (common Java and C++).

Wikipedia's content is linked and a well-working search engine is proposed. Our results show that general and frequently-used programming concepts are well-defined and explained in English. There are also good resources; describing general concepts in Bulgarian, but Bulgarian language content for specific programming concepts are insufficient.

Our conclusions are that Wikipedia is a good external knowledge source for mapping English language labels, but not sufficient for mappings, based on Bulgarian language.

We currently work on usage of bilingual English-Bulgarian ontologies, learning content and Wikipedia as external resource for mappings between Bulgarian language terminologies in computer science area.

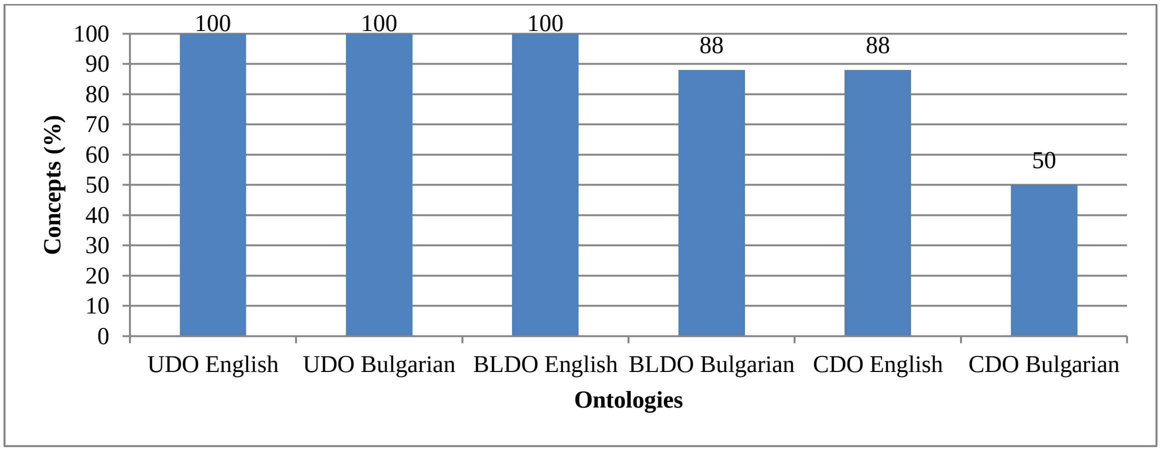


Figure 2 Percentage of explained Java only concepts

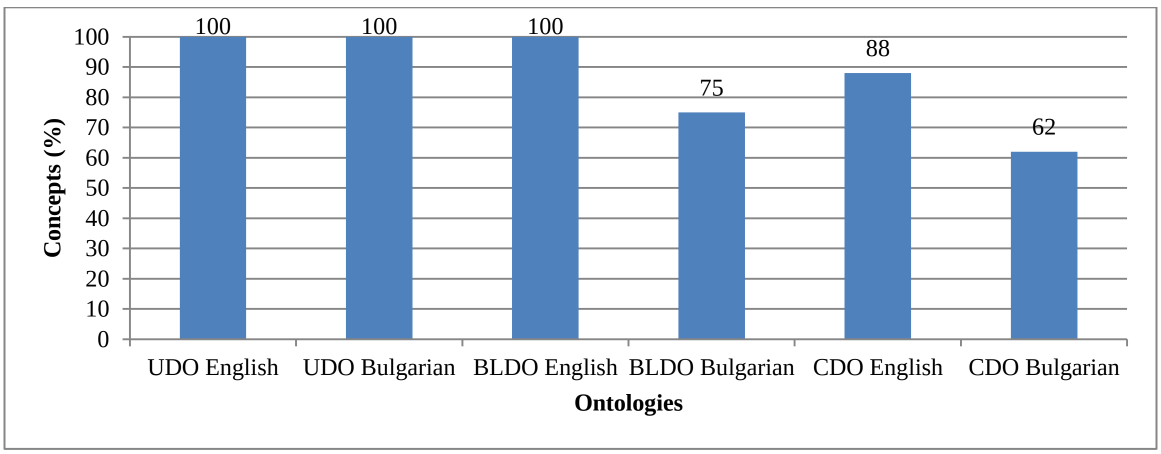


Figure 3 Percentage of explained C++ only concepts

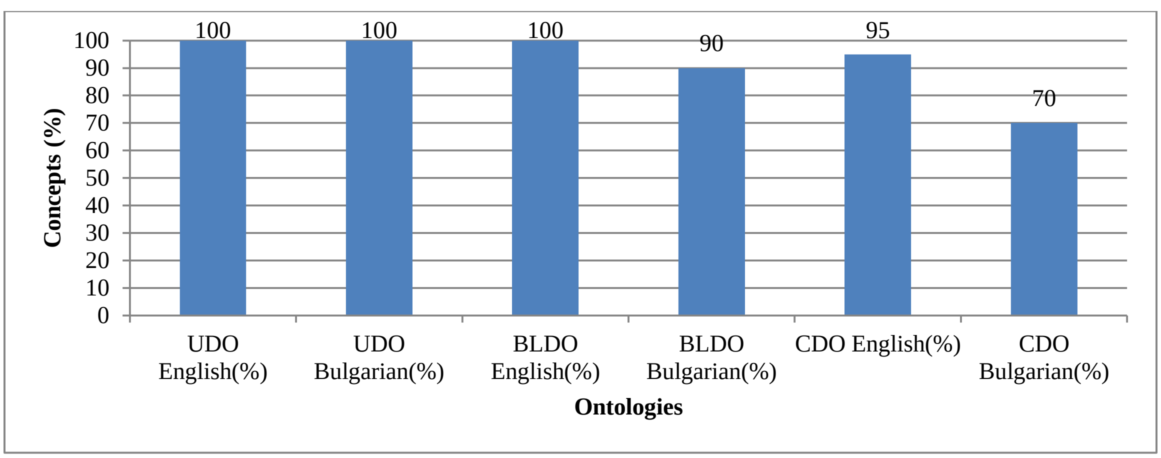


Figure 4 Percentage of explained common Java and C++ programming concepts

5.Conclusion

In this paper, we propose an ontology-based model for annotation, searching and comparison of e-Learning content. Our knowledge-based model can be used both for content in English and Bulgarian languages. It is intended for learning in computer science and programming domain, but it can be used in many other learning domains. The proposed knowledge model is generally complex and it seems to be difficult to use them in real applications. Its main strength is strict separation of different type knowledge (pedagogical from domain-specific, general from domain-specific, linguistic from semantically–rich). This can simplify the ontology building process, as global, linguistic and pedagogical knowledge is already modeled ontologically in many scientific projects. So, ontology reuse, ontology mapping and mappings reuse will simplify building of this complex ontological system. Standardization, reuse and automated evolution of previously developed ontologies will make the ontology development process more efficient.

We also try to establish the link between conceptual knowledge and its associated linguistic representations in Bulgarian and English languages by labeling ontological elements both in Bulgarian and English, use some descriptions, synonyms, related words. We show that using bilingual ontologies, e-Learning texts and Wikipedia as an external knowledge source can give some benefits for ontology mapping.

Bilingual lexicalization gives some new ideas for sense disambiguation. Storing ontology patterns and mappings in libraries for reuse is another way to make easier development of such complex ontology – based knowledge models.

Acknowledgment

The author would like to thank the Research and Development Sector at the Technical University of Sofia for the financial support.

Conflicts of interest

The author has no conflicts of interest to declare.

References

[1] Al-Yahya M, George R, Alfaries A. Ontologies in E-learning: review of the literature. International Journal of Software Engineering and Its Applications. 2015; 9(2):67-84.

[2] Lemnitzer L, Mossel E, Simov K, Osenova P, Monachesi P. Using a domain-ontology and semantic search in an e-learning environment. In innovative techniques in instruction technology, e-learning, e-assessment, and education 2008 (pp. 279-84). Springer, Dordrecht.

[3] Knoth P. Semantic annotation of multilingual learning objects based on a domain ontology. In: doctoral consortium workshop at the fourth European conference on technology enhanced learning. 2009.

[4] Simov KI. Ontology-based lexicon of Bulgarian. JLCL. 2009; 24(2):40-55.

[5] Sánchez-Alonso S, Sicilia MA, Pareja M. Mapping LOM to WSML: an ontology of learning objects. In proceedings of the 2007 (pp. 92-101).

[6] Gluz JC, Vicari RM. An OWL ontology for IEEE-LOM and OBAA metadata. In international conference on intelligent tutoring systems 2012 (pp. 691-3). Springer, Berlin, Heidelberg.

[7] Casali A, Deco C, Romano A, Tomé G. An assistant for loading learning object metadata: an ontology-based approach. In proceedings of the informing science and information technology education conference 2013 (pp. 77-87). Informing Science Institute.

[8] Behr A, Primo TT, Vicari R. Obaa-leme: a learning object metadata content editor supported by application profiles and educational metadata ontologies. Proceedings of the workshops of the Brazilian congress of informatics in education 2014 (p. 455).

[9] IMS Global Learning Consortium. MS Metadata Best Practice Guide for IEEE 1484. Standard for Learning Object Metadata.

[10] Paulins N, Arhipova I, Balina S. Multilingual information delivery based on a domain ontology. In proceedings of the international conference on computer systems and technologies 2014 (pp. 430-6). ACM.

[11] Manente M. Babylon and ontology: multilingual and cognitive e-learning management system via PDA phone. Education, Audiovisual and Culture Executive Agency (EACEA), Brussels. 2008.

[12] Bella G, Giunchiglia F, McNeill F. Language and domain aware lightweight ontology matching. Journal of Web Semantics. 2017; 43:1-7.

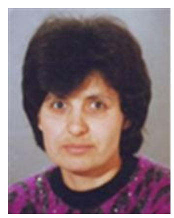
[13] Trojahn C, Quaresma P, Vieira R. A framework for multilingual ontology mapping.2008.

[14] Fu B, Brennan R, O'Sullivan D. Cross-lingual ontology mapping–an investigation of the impact of machine translation. In Asian semantic web conference 2009 (pp. 1-15). Springer, Berlin, Heidelberg.

[15] Wang S, Isaac A, Schopman B, Schlobach S, Van Der Meij L. Matching multi-lingual subject vocabularies. In international conference on theory and practice of digital libraries 2009 (pp. 125-37). Springer, Berlin, Heidelberg.

[16] Spohr D, Cimiano P, Hollink L. Multilingual and cross-lingual ontology matching and its application to financial accounting standards. In proceedings of the international semantic web conference 2011.

[17] Ngai G, Carpuat M, Fung P. Identifying concepts across languages: a first step towards a corpus-based approach to automatic ontology alignment. In COLING: the 19th international conference on computational linguistics 2002.
[18] Dragoni M, Petrucci G. Supporting multilingual ontology matching with MoKi. In international semantic web conference (Posters & Demos) 2015.
[19] Pazienza MT, Stellato A. Linguistically motivated ontology mapping for the semantic web. In Italian workshop on semantic web applications and perspectives 2005(pp.1-16).
[20] Jung JJ, Håkansson A, Hartung R. Indirect alignment between multilingual ontologies: a case study of Korean and Swedish ontologies. In KES international symposium on agent and multi-agent systems: technologies and applications 2009 (pp. 233-41). Springer, Berlin, Heidelberg.
[21] Dos Santos CT, Quaresma P, Vieira R. An API for multilingual ontology matching.2010.
[22] Shvaiko P, Euzenat J. Ten challenges for ontology matching. In OTM confederated international conferences on the move to meaningful internet systems 2008 (pp. 1164-82). Springer, Berlin, Heidelberg.
[23] Li J, Tang J, Li Y, Luo Q. Rimom: a dynamic multistrategy ontology alignment framework. IEEE Transactions on Knowledge and data Engineering. 2008; 21(8):1218-32.
[24] Lin F, Krizhanovsky A. Multilingual ontology matching based on Wiktionary data accessible via SPARQL endpoint. arXiv preprint arXiv:1109.0732. 2011.
[25] Spohr D, Hollink L, Cimiano P. A machine learning approach to multilingual and cross-lingual ontology matching. In international semantic web conference 2011 (pp. 665-80). Springer, Berlin, Heidelberg.
[26] Meilicke C, GarcíA-Castro R, Freitas F, Van Hage WR, Montiel-Ponsoda E, De Azevedo RR, et al. MultiFarm: a benchmark for multilingual ontology matching. Web Semantics: Science, Services and Agents on the World Wide Web. 2012; 15:62-8.
[27] Damaševičius R. Ontology of domain analysis concepts in software system design domain. In information systems development 2009 (pp. 319-27). Springer, Boston, MA.
[28] Amorim RR, Lama M, Sánchez E, Riera A, Vila XA. A learning design ontology based on the IMS specification. Journal of Educational Technology & Society. 2006; 9(1):38-57.
[29] Paquette G. A competency-based ontology for learning design repositories. International Journal of Advanced Computer Science and Applications. 2014;5(1):55-62.
[30] Hirst G. Ontology and the lexicon. In Handbook on ontologies 2009 (pp. 269-2). Springer, Berlin, Heidelberg.
[31] Wilks Y, Stevenson M. Sense tagging: semantic tagging with a lexicon. arXiv preprint cmp-lg/9705016. 1997.



Tatyana Ivanova received her M.S. in Mathematics and Informatics from Sofia University and Ph.D. degree in Artifical Intelligence Systems from Technical University of Sofia, in 1987 and 2009, respectively. She is currently an Associate Professor of Computer Systems and Technologies at Technical University, Sofia, Bulgaria. Her research interests are Semantic Web, E-learning, Ontological Engineering, Databases.
Email: tiv72@abv.bg