

A Bilingual Ontology Mapping and Enrichment Method for Domain Ontologies in E-learning

Tatyana Ivanova
Technical University of Sofia
Bulgaria
{tiv72@abv.bg}



ABSTRACT: English language is accepted as international language for publishing research in e-learning and Semantic Web area, but people are needed from representing data and knowledge in his native language. This paper proposes an approach for mapping and enrichment of domain ontologies, labeled in two natural languages (bilingual ontologies). Our approach combines syntactical, structural and semantic mapping. It also can reuse existing mappings and use learner's or expert's feedback to improve mapping. We discuss the application of our approach for mapping ontologies describing e-learning content.

Keywords: Ontology, Multilingualism, Ontology Mapping, Ontology Alignment, Bilingual Ontology, E-learning

Received: 18 September 2019, Revised 12 December 2019, Accepted 19 December 2019

© 2020 DLINE. All Rights Reserved

DOI: 10.6025/jcl/2020/11/1/1-11

1. Introduction

As the English language content in every scientific domain can be found and it has highest quality, possibilities to find such content is very important for most web users. But many users prefer or can use only content in his native language. There are also users, that use both resources in native language and in international languages. So, mappings of international language content to the content, written in other languages have become more and more important. Many ontologies, describing almost all scientific domains have been built recently. Most of them are labeled in international languages, as English, and only small number have labels, written in non-international natural languages.

The easiest way for automated comparing textual content in two different languages is when this content is annotated by usage of bilingual thesauruses or ontologies, labeled in these languages. But most of developed ontologies are labeled only in English. Mapping techniques are also developed mainly for English language – labeled ontologies. Good general bilingual thesauruses do not include important domain-specific terminology. To find, compare, or recommend ontologically described resources, written in different natural languages, multilingual or cross-lingual ontology mapping should be performed. Ontology mapping is a process of finding correspondences between elements of two or more ontologies, describing overlapping areas of knowledge. There is a lot of ontology mapping research, but almost all of the research results are related to monolingual (English – language labeled) ontology mapping. Main mapping approaches can be classified as lexical (syntactic), structural, instance-based and semantic. All

the lexical approaches highly depend on the used natural language. Mapping systems usually use combination of mapping techniques, and the mapping quality highly depends on the mapped ontologies, including its natural language labeling. So, ontologies, labeled in more than one language (bilingual or multilingual ontologies), and mapping techniques for such type of ontologies are needed to support information retrieval, recommendation and integration in multilingual environment.

There are two main ways of development of multilingual ontologies – by linguistic enrichment of ontology labels and by aligning two or more ontologies monolingual ontologies developed (labeled) in different natural languages. The conceptual structure and logic of the parts, expressed in different languages in multilingual ontology, developed by linguistic enrichment, are identical. Conceptualization mismatches should be considered while developing multilingual ontology by ontology matching.

Our literature review related to multilingual ontologies shows that most of the available or developed in research projects ontologies are labeled in English and there is only small number of bilingual or multilingual ontologies. There are no theoretical barriers to label ontology in several languages, but in practice this is expensive, labor-intensive tack as it can be done only by team of experts in the scientific domain, knowledge engineering and linguistics in two or more natural languages. So, labeling ontologies in two or more languages is a practical problem, related to ontology development.

In this paper we propose combined approach for mapping and enrichment of domain ontologies and taxonomies, labeled both in two natural languages. Such ontologies can enable the users whose native languages are only national, to search and retrieve high quality content both in its native and some international language (usually the scientific quality and actuality of the content in international languages are higher). For example, Bulgarian learners, speaking both Bulgarian and English could use one and the same bilingual ontology for personalized access to knowledge in Bulgarian and in English. The language-independent knowledge is usually represented using such ontologies, and then labels, comments, and other language-specific elements are added. As comprehensive manual translation of ontology labels is labor-intensive, we assume that only some of labels can be translated or only some of synonyms are added. Our approach is intended for mapping such incomplete bilingual ontologies, and use translations, found in the mapping process to enrich linguistical elements of the mapped ontologies. We will discuss use case of usage such mappings in e-learning.

2. Ontology Mapping Approaches

2.1. Types of Ontology-Mapping Approaches

There are some good surveys on ontology mapping. A classification of recent schema and ontology matching techniques is proposed in [16], [15]. According to the type of sources of mapping information the approaches can be classified into four main types (figure 1):

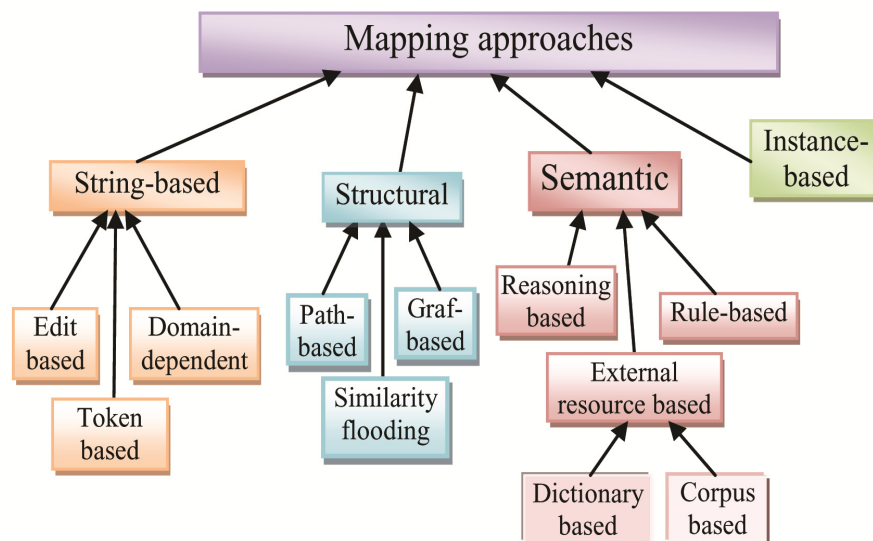


Figure 1. Classification of ontology mapping approaches and techniques

- Terminological, or lexical
- Instance based, extensional
- Semantic based
- Structural, or graph-based, intentional
- Combined approaches.

Lexical approaches use string-similarity based metrics to detect similarities between labels of ontology entities (usually between concepts). Pure lexical methods work well only when there are many lexical overlap between the labels of the ontology entities. In other cases, structural, semantic mapping, or mapping, based on external sources are used. In instance-based mapping semantic relations between concepts of two ontologies are determined based on the overlap of their instance sets. Instance based mapping is based on statistical or set-based measures are used to calculate similarity of ontology instances. Semantic approaches are usually based on background knowledge, or reasoning algorithms. Semantic mapping ensures also finding natural language independent semantic relations between concepts. Semantic approaches try to find not only equivalence relations between entities in different ontologies, but also subsumption, paronymy, or other semantic relations. In semantic approaches some formal semantics also can be used to interpret the input and justify the results [15]. Example semantic mapping systems, tested on OAEI benchmark are Aroma, ASMOV, Taxomap, STROMA [3]. Reasoning – based systems are AROMA, YAM++, LogMap, S-Match, MaasMatch.

Linguistic information about lemma’s semantic can be extracted from various sources: Dictionaries/thesauri, e.g. WordNet[9]; Textual sources, using information retrieval techniques ; Reference domain ontologies as a semantic bridge between two ontologies; Wikipedia [2], LOD, or other semi-structured web sources.

The structure level mapping 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. Three structure-based ontology mapping techniques are discussed in [17]: Similarity-flooding, Anchor-PROMPT and Anchor-Flood.

2.2. Ontology Mapping Strategies

To achieve high precision and recall in automated ontology mapping, a single ontology mapping approach (for example the string-based syntactic mapping) is not sufficient. Combination of linguistic, structural and semantic mapping techniques usually leads to good results as they can complement each other. Linguistic matchers calculate context-independent similarity coefficients between entity labels in some language, while the structural matchers measure the similarity between contexts in which labels occur, and semantic approaches can clarify the context. Mapping strategies (or multiple strategies) are combinations of mapping metrics, algorithms or approaches. There are two kinds of multistrategy mapping: hybrid and composite (fig. 2). The approach is hybrid when different mapping criteria or properties are used within a single algorithm. Example is the usage of combined metric that includes lexical and structural components. By contrast, a composite mapping approach combines the results of several independently executed mapping algorithms, and every one of them can be simple or hybrid [21]. COMA, RiMOM [11] and Cupid matching tools for example are based on composition of several matchers (syntactical and structural). COMA can construct new matchers and combines them flexibly testing various matching strategies.

An interactive ontology matching strategy uses expert knowledge towards improving the quality of the final alignment. These strategies use lexical, structural and/or semantic approaches to find candidate mappings and the expert feedback to validate the

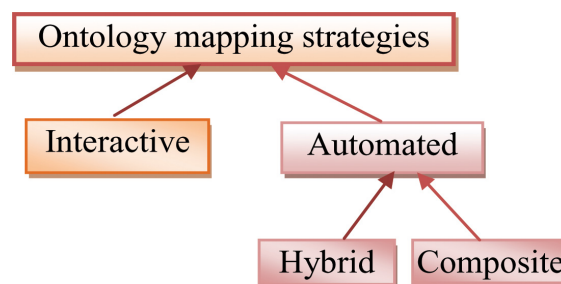


Figure 2. Ontology mapping strategies

proposed correspondences [18]. The main aim of this strategy is to decrease the number of interactions without decreasing proportionally the quality of the generated alignment. Future generating of the candidate correspondences depends on the previous feedbacks given by the expert.

3. Related Work

As the Bilingual ontology alignment problem can be seen as a specialization of the Multilingual ontology mapping problem, and Multilingual ontology mapping is a generalization of Bilingual ontology mapping, most of the mapping techniques can be used both in the multilingual and bilingual context.

3.1. Multilingual Ontology Mapping

The ontologies that are developed to access different natural languages are called Multilingual Ontologies (MO) [12]. Language-aware or multilingual ontology matching is defined in [5] as a type of ontology matching where the matcher is capable of dealing with ontologies expressed in multiple languages. A multilingual ontology provides additional information that can be exploited in the matching process. On the other hand, widely usage of lexical comparison between names of entities during the mapping process make the most of mapping algorithms inapplicable for the ontologies, labeled in different languages. So multilingual ontology matching requires revision of most of the matching algorithms to make them applicable or more suitable for the multilingual context. At this time there are only a few researches on multilingual ontology matching (e.g. [20]).

Multilingual ontology mapping can use traditional ontology matching algorithms, machine translation and multilingual textual or thesaurus-based linguistic resources. Most of the approaches rely on general-purpose machine translation services (as Bing or Google) to reduce the problem to monolingual English-only matching. The main problem of machine translation is that it is statistics and machine-learning-based and works well only on corpora similar to those on which the translation 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 requires preprocessing and makes the translation task more error-prone.

Some other approaches use also semantic matching. Ontology labels are parsed by multilingual natural language processing (including translation) and then matched using language-independent and domain-aware background knowledge acting as an Interlingua [5], or using monolingual, English language-based semantic mapping.

Multilingual ontology mapping has two different aspects:

1. Mapping the ontology, having labels in one language to ontology, having labels in other language (cross lingual ontology mapping);
2. Mapping the ontology, having labels in several languages to ontology, having labels in one or more languages;

The first case is about mapping monolingual ontologies, having different natural language labels (Cross-Lingual Ontology Matching, CLOM). CLOM is the process of establishing correspondences (find relations) among the ontological resources from two independent ontologies where each one is lexicalized in a different natural language [20].

Mapping ontologies, having labels or other lexical elements, written in two or more different natural language usually requires translation. As ontologies have no common language lexical elements, translation also is one of the important steps. A minimal API for multi-lingual matching is presented in [10]. Two strategies are defined and implemented: direct (translation-based) and indirect.

The direct alignment strategy of multilingual ontology mapping is usually based on machine translation. Automatic translation before alignment is used by most of cross-lingual alignment systems: CLONA, XMap, AML, YAM++, LogMap [18]. When ontology labels are written in two languages derived from the same root language they have a similar vocabulary. In such cases, a direct matching is recommended. Direct matching is also recommended when there is no appropriate stored alignments, or when one of the mapped ontologies is labelled in international language, as English.

The basic idea of indirect alignment strategy between multilingual ontologies is to compose alignments which already exist [10].

The indirect alignment strategy is based on composition of alignments. Some translation-based methods for matching two specific multilingual ontologies are based on translation of mapped ontologies to a common broadly accepted natural language (as English) [1]. Indirect alignment approach is tested in [6] as an alignment between French and Portuguese ontologies by using intermediary alignments in English, i.e., French – English and English – Portuguese alignments. Many sophisticated mapping systems as LogMap and Cross-Lingual Ontology Matching System CroLOM [18] apply first NLP on each natural language before translation. Then, it uses the public translation tools as Yandex, Bing, or Google Translate in order to translate all entities in English as pivot language and then compute the similarity between translated entities using WordNet for disambiguation. The Uby project aims at aligning concepts from different sources such as WordNet, GermaNet, FrameNet, Wiktionary and Wikipedia. It comprises more than 4.2 million lexical entries and 0.75 million links that were both manually and automatically generated (using mapping algorithms) [5].

According to [8], main CLOM techniques can be grouped into several approaches (like monolingual mapping):

- Manual processing
- Corpus-based approach [14]
- Instance-based approach [7]
- Based on linguistic enrichment of ontologies[16]
- Using machine learning[19]
- Combined approaches.

Our classification of multilingual ontology mapping approaches is presented on figure 3.

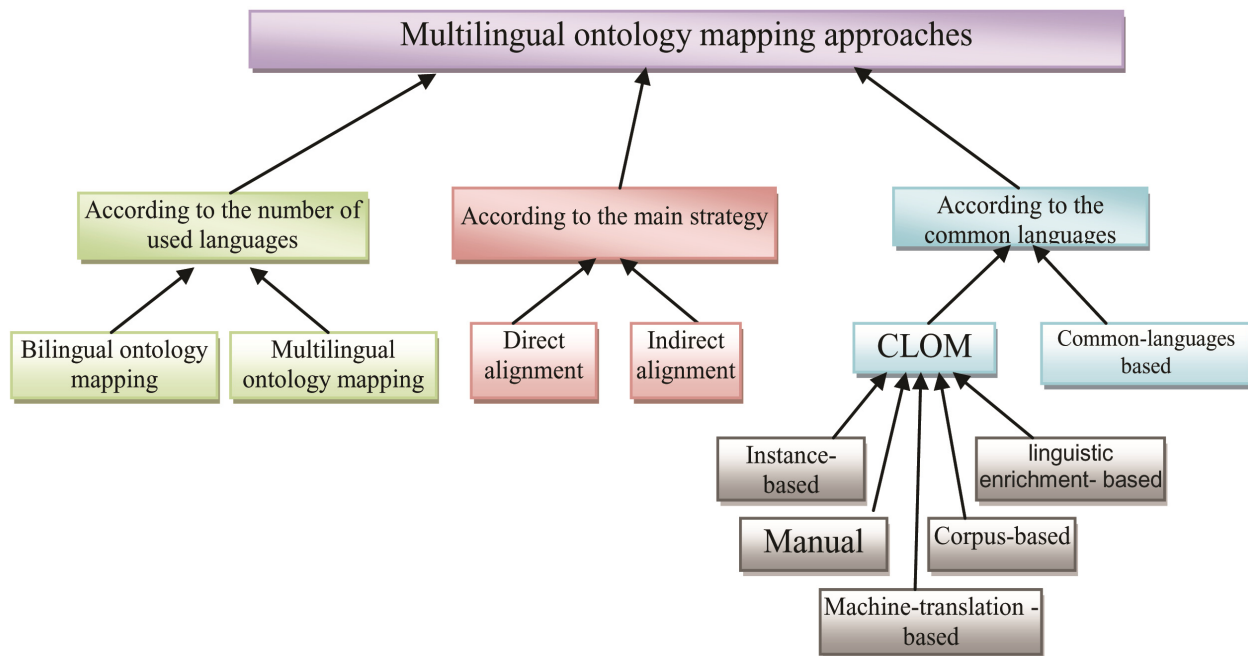


Figure 3. Classification of multilingual ontology mapping approaches

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 labeled. MultiFarm dataset [13] has been designed as a benchmark for multilingual ontology matching. 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.

3.2. Bilingual Ontology Alignment

We could not find information about mapping approaches, developed for direct or indirect mappings of bilingual ontologies. Research on mapping bilingual ontologies is closely related to ontology development and mapping supported by usage of semi-structured web resources (as Wikipedia), thesauruses and machine-readable dictionaries. An approach for aligning WordNet and HowNet to construct a bilingual ontology based on the co-occurrence of the words in the sentence pairs of a parallel corpus is presented in [22]. Authors used aligned ontologies to develop initial version of bilingual domain specific ontology by extraction of domain-specific terminologies and then enriching it by adding additional terms and some axioms.

4. Our Approach for Mapping Bilingual Ontologies

We present an approach for mapping bilingual or multilingual ontologies. It is the most useful when mapping ontologies, containing terminology in one international (e.g. English) and two or more non-international languages. We believe that such type of ontologies are most useful for information integration in multilingual environments (as Web, e-learning, E-commerce, etc.) when they include terminology written in international language. The presence of international language terminology will make possible usage of well working terminology-based mapping approaches, tested in the international languages (as English language) and can act as a bridge between rare and international languages. Many open and free linguistic resources, available only in small number of languages can be used to support mapping if ontologies contain labels in international language and labels in other (non international, rare) languages can be used to translate these mappings in the non-international languages, or to disambiguate ambiguous mapping proposals. Incomplete labeled ontologies also can be used (when not all of the terms are labeled in every of the used languages, but every term is labelled in the some language). As ontologies used in many applications need from constant evolution, and manual development is expensive, our aim is also to include in our approach possibilities for ontology enrichment. For example, when we map two classes on the base of labels in one language, and in one of the ontologies the mapped class has label in the other language, but in the other ontology the mapped class is labelled only in one language, our approach enriches the second ontology by taking the label in the other language from the first ontology and adding it in the second ontology.

So, our mapping approach is intended for mapping incomplete bilingual ontologies. The two ontologies can contain entities, labelled in the two used languages, as well as entities, labelled in only one of the languages. Our approach is based on mapping strategies that combine monolingual and translation – based cross-lingual mapping techniques. We use independent lexical similarity metrics in two used languages, and combined metric, which calculates similarity, using labels in the two languages, if labels in the two languages are specified. Our approach also involves humans (learners, teachers, or experts) in the mapping process. We will explain our mapping strategies and used techniques, and then show how they works, discussing a use case of mapping English-Bulgarian bilingual ontologies in e-learning.

4.1 Mapping Strategies

Our mapping process has two main phases: establishing some anchor mappings and use them iteratively for finding others by calculating structural similarities or by reasoning with mappings. Mapping establishment has also two phases: finding candidate correspondences, and verification (automated or interactive, by users or experts). After establishing mappings, they are used as anchors in the iterative mapping process. The two main mapping steps are repeated iteratively. During the iteration all previously established mappings are used, until no new mappings are found in the latest iteration.

More comprehensively, our mapping strategy is as follows (fig. 4):

1) Establish initial anchor mappings based on lexical and syntactic metrics. This can be done by:

- Finding monolingual candidate mappings, if ontologies have a common language. Monolingual candidate mappings are searched first in international language and then in non-international language. Cross lingual links in Wikipedia or Google translation can be used to find candidate mappings in the case when ontologies have no common language. Using labels or synonyms in two or more languages we calculate more precise values of lexical similarity measures and in such a way we can achieve higher precision and recall of the lexical mapping techniques.

- Evaluation of the proposed candidate mappings to select the best ones as anchors. We evaluate candidate mappings first by comparing proposed ones by different techniques and mark the mapping proposals that are different, but contain the same element as possibly contradictory. Then use external resources, as translation services, Wikipedia, learning resources, or other

texts to solve contradictions when translation or disambiguation is needed. Previously stored mappings from libraries also can be used in the evaluation process. Asking user feedback about unresolved contradictions is also important step in establishment of anchor mappings.

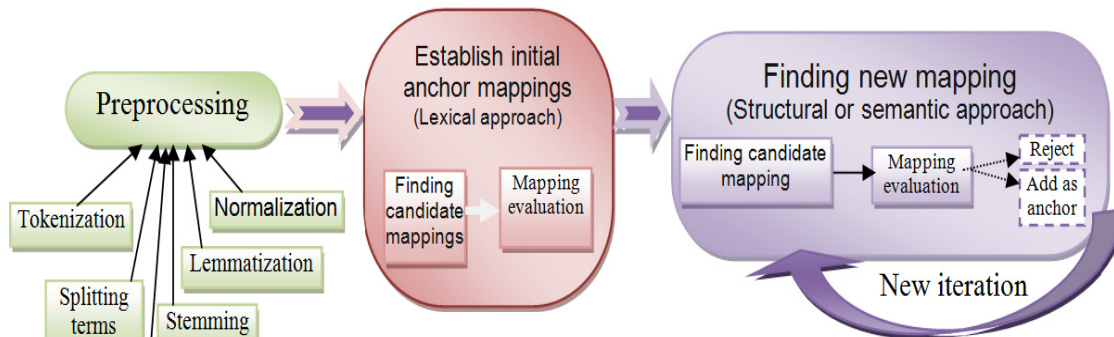


Figure 4. Multilingual ontology mapping strategy

2) Consider established aligned pairs as anchors and use them to perform iterative mapping process. We use all the established mappings as anchors in searching new mappings by applying structural and semantic approaches, including reasoning.

We store created mappings and metadata, related to its context in the alignment library for future reuse. The mapping context includes information about language, ontology domains, and tacks, in which this mapping is accepted.

We also can use accepted mappings to enrich one or the two ontologies by using labels, or its synonymy, found in the other ontology or Wikipedia during the mapping process. For example, if we establish mapping between classes, named “case” and “case statement” in two ontologies, we can add “case statement” as a synonymy label in the first ontology, “case” as a synonymy label in the second ontology, and also “условен оператор кейс” as Bulgarian labels in the two mapped ontologies, incompletely labeled in English and Bulgarian.

As mapping of different type ontology entities (e. g. classes, properties, relations) has some specifics, we will discuss them in the next subsection.

4.2. Mapped Entities

Mapping classes and properties is important for the alignment of bilingual ontologies, used in many application areas. Mapping relations is also important, as relationships in some cases are part of class definitions, and we can use them in future class mapping process. So, in the phase of generating the candidate correspondences mappings between all important entities and reasoning about them is performed iteratively.

4.2.1. Class Mapping Specifics

We use two variants of equivalents between concepts: **Equivalence relationship** and **Semantic equivalence relationship**.

Equivalence Relationships are the most important in our mapping. We map two classes as equivalent, when they have sufficient both structural and syntactical or linguistic similarity. Sufficient structural similarity is when all properties, relations and upper-classes are mapped as equivalent. Syntactical or linguistic similarity means that labels of entities after the preprocessing must be the same or some of them may be synonymous.

We use **Semantic equivalence relationship** when mapped concepts have the same semantic as domain entities, but have some differences in definitions, properties, or other characteristics in the mapped ontologies. For example, we can use the multilingual labels in the ontology mapping as additional linguistic information (some context). For example, the concept “operator” in Bulgarian incorporates “operators” and “statements” in English. So, different English language labels in this case will help to

disambiguate the sense of the label “operator” in two different ontologies to perform correct semantic mappings. We use also numeric evaluation metric of semantic equivalence (including weights to describe similarities, and thresholds for accepting or rejecting mapping proposals). The inclusion of synonyms (e.g., ‘startirane’ and “puskane” are both synonyms of English verb “run” for program running) or terminological variants (as abbreviations and acronyms) also can make clear the context of ontology labels.

We also define a context of the concept and a context similarity measure. A context of the concept includes all the entities, closely related to the concept (direct hypernyms, hyponyms, synonyms, and other directly related terms in all the used natural languages. Context similarity measure calculates the percentage of mapped terms, included in the context of the discussed term. We store all the Equivalence and Semantic equivalence mappings and its contexts and context similarity measures in the mapping library for the future use.

Our mapping approach also can establish subsumption mappings. This is done only as a lazy mapping procedure (when needed, for example to answer the question “are the concepts operator and conditional operator related or not?”). We do not store subsumption mappings in the library.

4.2.2. Mapping Properties and Relations

Equivalent properties can be lexically similar, but also should have the same (or mapped as equivalent) Domains, Ranges, Sub Properties or Super Properties. Differences in domains, ranges, or property hierarchies are variants of semantic heterogeneity and should be resolved during reasoning or validation by users or experts.

Main steps for establishing initial property mappings are:

- Grouping the ontology properties in categories: Data properties and object properties (having several subcategories: FunctionalProperty, InverseFunctional Property, SymmetricProperty, TransitiveProperty).

- Lexical mapping of properties, belonging to one and the same category. It is possible to map properties from different groups, but such mappings cannot be equivalences and its specifics should be explained in every case. For example, we can have the functional property has type (variable, datatype) in Java programming ontology, and the non-functional property has type (variable, datatype) in JavaScript ontology. It is possible to map these properties, but this mapping cannot be “equivalence” as there are important differences in Java and JavaScript typization systems, and it is important to have in mind or explain these differences.

- Using external resources for disambiguation or synonymy detection;

- Using domain and range built-in properties and already mapped classes to check if lexically-similar properties have the same semantic. Equivalent properties must have the same domains and ranges, so to map the two properties, its domains and ranges must be previously mapped as equivalent.

- Interactive validation by users or experts. Automatic mapping of relations and properties is more difficult and error-prone, so validation by humans is the most important. And in some domains, as e-learning the validation is useful for humans (learners) as it can help the understanding of important relationships between concepts.

Accepted mappings are further used in the iterative mapping process.

4.3. Used Measures, its Importance and Combination

Our approach uses lexical metrics, based on included multilingual terminology in the ontologies. We use labels in the both languages and its synonyms or abbreviations (if presented) in the mapping process as elements of one language-based or combined lexical and syntactic metrics. If the mapped entity does not include label in some of used languages, we use translation service or Wikipedia to find needed translation (including disambiguation techniques) and enrich label in the ontology with this translation if it is correct. We also store approved (at the second stage) mappings and its context for future use. Mappings between specific domain terminologies are valuable, as some of these terms can not be found in linguistic sources and are translated incorrectly by translation tools. Our approach rely on sequential application of three main type mapping approaches: lexical (string-based) , structure-based, and semantic-based to match bilingual ontologies.

4.3.1. Lexical Similarity Measures (Cross-lingual based Similarity)

We use pre-processing before comparing entity names. The preprocessing can include translation and is language and ontology specific - dependent. For example, in one ontology, names can have specific prefixes, and in the other – not. Compound terms also can be written in different ways (for example, using “_”, or not). And different languages can use different terminological order in compound terms (for example, in “ontology mapping” in English the head noun is the second token, and in Bulgarian variant “svurzvane na ontologii” the head noun is the first token. All these naming specifics are detected during lexical analysis of the ontology labels as a preprocessing step. Specific rules for detection of naming specifics or head nouns are used for different languages and ontologies in this step to cope with language and naming specifics.

We calculate combined lexical similarity between two entities by using their labels, acronyms and synonyms in all the languages, used in the mapped ontologies (for example, both in Bulgarian and English for Bulgarian-English bilingual ontology). To calculate combined (multilingual) lexical similarity measure for two mapping entities, we first calculate monolingual lexical similarity measures for every used language, using entity labels written in this language. Then we calculate the average of the monolingual similarity measures for different languages. The values of such multilingual lexical similarity measure describe entities more precise than analogous similarity measures in one language. For example, when only English – language ontology labels “data” and “date” are used to calculate lexical similarity, the similarity measure will be higher than in the case when Bulgarian language labels “Danni” and “Vreme” are used in the similarity-calculation formula. In the first case, higher similarity measure can lead to the wrong mapping proposal to map different concepts “data” and “date”. Usage of Bulgarian language labels in the lexical similarity metric will eliminate this erroneous mapping. We use a string-based match between the pairs of labels, synonyms, acronyms, and an aggregation on the resulting string similarities.

4.3.2. Structural Similarity

The idea of the structural similarity is based on the assumption that, when two entities are similar, its surrounded terms are also similar. We use two types of structural similarities: Context-based similarity and Path – based similarity.

Context-based similarity is a structural similarity, based on the similarity of the entity neighbors (direct hypernyms, direct hyponyms, entity properties, synonyms). Using Context-based similarity we can estimate for example the closeness in definition of the terms in different ontologies.

Path – based similarity is a kind of semantic similarity, and comparing paths in bilingual or monolingual environment can help for correct translation or disambiguation. As in hierarchy every class inherits all the features of all super classes in the path, two classes (or relations) are considered similar if the classes in their relative paths, in their class hierarchies, are already similar. This similarity is important when comparing two learning courses (close concepts should be related to similar paths) and we first detect it, using mainly the classes and relation’s labels in international language.

Semantic similarities, based on additional information from Wikipedia, SemRep [4] or learning resources (as a corpus) provide well-defined semantics for the entity of ontologies and are useful for automated refining candidate alignments, or finding non-syntactic, synonymy-based alignments. Usability of external resources highly depends from the natural languages, used for labeling in ontologies, and these resources are usable mainly when ontologies have labels in some international language.

5. Use Case and Discussion – Mapping English-Bulgarian Bilingual Ontologies

To perform some initial evaluation of our approach, we mapped two domain ontologies, comprehensively describing learning courses for programming in C++ and Java. Learning courses are in Bulgarian, but contain also English terms for valuable programming entities. In scientific and learning domains a precise, strict and clearly defined terminology is used. So, ontologies, describing such type of content include well-structured natural language terminology for classes, properties, relations and represent definitions, classifications, important relationships between terms. Our ontologies are mainly terminological and does not contain valuable amount of instances. They describe closely related domains with valuable overlapping. Most of the entities are labeled in English and almost all entities are labeled in Bulgarian. Strict definitions of concepts are presented by ontological structures, so structural similarity of described ontologies is closely related to course similarity. We use learning content and Wikipedia as external semantic sources for disambiguation and finding synonymy for enrichment of ontologies.

Our aim is to perform maximal equivalence mappings as they are the most important for comparing e-learning content. We start mapping by string-based comparison – common prefix, suffix, common Bulgarian or English labels, or compound label compo-

nents, to obtain candidate alignments, which then will be enriched or refined. Finding clear differences between mapped entities in Semantic equivalence mappings is important when comparing e-learning courses.

Involving students in the mapping process ensure some type of evaluation, enrichment, and is also useful part of the learning process. After initial lexical comparison of entity labels we use reasoning-based strategies to establish correct mappings (If some direct super or sub-concepts are similar, then the two compared concepts may be similar; If the relation and related concepts are similar, then the two compared concepts may be similar; If some definition paths are similar, then defined concepts may be similar). Derive and explain all subsumption mappings (if some equivalence class is missing, what is the situation) is both important in learning and for improving mapping quality. It is also important to distinguish between metrological and taxonomical hierarchies when mapping ontologies, related to e-learning. In metrological hierarchies entities are connected through part-of-whole relations. In taxonomical hierarchies entities are connected through superclass-subclass relations (parent-child relationships).

Other structural information (domain and range properties, class-subclass relations, sibling concepts) is also important for calculating similarity between course terminologies.

“Positive integer” and “unsigned integer” for example mainly refer to the same concept, but they cannot be considered as absolute synonyms. The concept “Integer” has slightly different meaning in different languages or language versions have some differences.

We also made experiment to estimate the role of usage of two language labels in the mapping. We first perform mapping our Bulgarian-English domain ontologies by usage only Bulgarian language labels, then by usage only English-language labels. In the first case, precision was 0.09 less and recall was 0.12 less than precision and recall in the second case. In the case of usage of only the English language labels, precision was 0.04 less and recall was 0.05 less than precision and recall in case when bilingual mapping was performed. This experiment confirms the expectations that the use of two-language labels can improve ontology mapping. The improvement is higher when ontologies are labelled in Bulgarian (non-international language).

6. Conclusion

Labeling ontology entities in two natural languages (international and non-international) is useful for some tasks, including ontology mapping, information integration, e-learning. We propose ontology mapping approach that combines lexical, semantic and structural ontology mapping strategies for mapping bilingual ontologies. Our approach also can use external sources and user feedback for improving mapping quality. Multilingual context contains additional information for supporting ontology mapping and our approach can benefit from this information to increase mapping precision and recall. Our approach also uses various techniques, and our future research is to find flexible strategy of its combination in dependence with the mapped ontology specifics, and the context of its usage.

References

- [1] Al-Feel, H. et al. (2013). An inter-lingual reference approach for multi-lingual ontology matching. *arXiv preprint arXiv:1309.6650*, 2013.
- [2] Arnold, P., Rahm, E. (2014). Extracting Semantic Concept Relations from Wikipedia. *In: Proceedings 4th Int. Conf. Web Intelligence, Mining and Semantics (WIMS)*, 2014.
- [3] Arnold, P., Rahm, E. (2014). Enriching ontology mappings with semantic relations. *Data & Knowledge Engineering*, 93 (1) 2014.
- [4] Arnold, P., Rahm, E. (2015). SemRep: A repository for semantic mapping. *Datenbanksysteme für Business, Technologie und Web*, 2015.
- [5] Bella, G., et al. (2017). Language and domain aware lightweight ontology matching. *Web Semantics: Science, Services and Agents on the World Wide Web*, 43, p. 1-17, 2017.
- [6] Dos, S. et al. (2010). An API for multilingual ontology matching. *In: Proceedings 7th conference on Language Resources and Evaluation Conference (LREC)*, p. 3830-3835, 2010.
- [7] Dragoni, M. and Petrucci, G. (2015). Supporting Multilingual Ontology Matching With MoKi. *In: International Semantic Web*

Conference (Posters & Demos), 2015.

- [8] Fu, B. et al. (2009). Cross-lingual ontology mapping - an investigation of the impact of machine translation. *In: Proceedings of the 4th Annual Asian Semantic Web Conference (ASWC 2009)*, Vol. 5926 of LNCS, p 1-15, 2009. http://dx.doi.org/10.1007/978-3-642-10871-6_1
- [9] Hertling, S., Paulheim, H. (2012). WikiMatch—using Wikipedia for ontology matching. *Ontology Matching*, 2012.
- [10] Jung, J. J. et al. (2009). Indirect alignment between multilingual ontologies: A case study of Korean and Swedish ontologies. *Agent and Multi-Agent Systems: Technologies and Applications, Lecture Notes in Artificial Intelligence*, p. 233-241, 2009.
- [11] Li, J. et al. (2009). RiMOM: A dynamic multistrategy ontology alignment framework. *IEEE Transactions on Knowledge and Data Engineering*, 21 (8) 1218-1232, 2009.
- [12] Lourdusamy, R. and Florence J. M. (2016). Methods, approaches, principles, guidelines and applications on multilingual ontologies: a survey *ICTACT Journal on Soft Computing*, 7 (1) 2016.
- [13] Meilicke, C. et al. (2012). MultiFarm: A benchmark for multilingual ontology matching. *Web Semantics: Science, Services and Agents on the World Wide Web*, vol. 15, p. 62-68, 2012.
- [14] Ngai, G. et al. (2002). Identifying concepts across languages: a first step towards a corpus-based approach to automatic ontology alignment. *In: Proceedings of the 19th International Conference on Computational Linguistics*, vol.1, p 1-7, 2002.
- [15] Otero-Cerdeira, L. et al. (2015). Ontology matching: A literature review. *Expert Systems with Applications*, 42 (2) 949-971.
- [16] Pazienta, M., Stellato, A. (2005). Linguistically motivated ontology mapping for the Semantic Web. *In: Proceedings of the 2nd Italian Semantic Web Workshop*, p. 14-16, 2005.
- [17] Ramar, K., Gurunathan, G. (2016). Technical review on ontology mapping techniques. *Asian Journal of Information Technology*, 15 (4) 676-688.
- [18] Shvaiko, P. et al. (2017). OM-2017: Proceedings of the Twelfth International Workshop on Ontology Matching. *In: OM 2017-12th International Workshop on Ontology Matching*, p. 1-233, 2017.
- [19] Spohr, D. et al. (2011). A machine learning approach to multilingual and cross-lingual ontology matching. *The Semantic Web—ISWC*, p. 665-680, 2011. http://dx.doi.org/10.1007/978-3-642-25073-6_42
- [20] Spohr, D. et al. (2011). Multilingual and cross-lingual ontology matching and its application to financial accounting standards. *In: Proceedings of the 10th International Semantic Web Conference*, 2011.
- [21] Tang, J., Liang, B. Y., Li, J. Z. (2005). Toward detecting mapping strategies for ontology interoperability. *In: The 14th International World Wide Web Conference (WWW 2005)*.
- [22] Yeh, J. F. et al. (2004). Automated alignment and extraction of bilingual ontology for cross-language domain-specific applications. *In: Proceedings of the 20th International Conference on Computational Linguistics*, 2004.