Computer Science and Information Technologies
Vol. 1, No. 2, July 2020, pp. 47~53
ISSN: 2722-3221, DOI: 10.11591/csit.v1i2.p47-53

Building a multilingual ontology for education domain using monto method

Merlin Florrence
Department of Computer Applications, Sacred Heart College, Tirupattur, India

Article Info

Article history:

Received Sep 24, 2019
Revised May 1, 2020
Accepted May 21, 2021

Keywords:

Methodology,
Multilingual,
Ontology

ABSTRACT

Ontologies are emerging technology in building knowledge based information retrieval systems. It is used to conceptualize the information in human understandable manner. Knowledge based information retrieval are widely used in the domain like Education, Artificial Intelligence, Healthcare and so on. It is important to provide multilingual information of those domains to facilitate multi-language users. In this paper, we propose a MOnto (Multilingual Ontology) methodology to develop multilingual ontology applications for education domain. New algorithms are proposed for merging and mapping multilingual ontologies.

Corresponding Author:

Merlin Florrence,
Department of Computer Applications,
Sacred Heart College, Tirupattur, India.
Email: merlinflorrence@gmail.com

1. INTRODUCTION

Ontology enables the natural language processing of the data in an efficient way. It retrieves the information based on the knowledge and conceptualizes the information in formal way. Enormous information is available over the internet in a specific language. It is essential to provide the information in different natural languages to benefit multi-language users. Ontologies play vital role in providing knowledge based information systems. Ontology is a "formal, explicit specification of a shared conceptualization" [1]. It is a collection of set of concepts, properties, relations, instances, axioms and rules which can be represented as, **(Ontology) O = {C, P, R, I, A}.** 'C' represents the classes or concepts of the domain. 'P' signifies the properties of the concept. 'R' denotes the binary relations between the concepts (1-1, 1-M, M-M). 'A' represents axioms and rules which are used as a basis for reasoning [2]. In ontology a set of terms for describing a domain is arranged hierarchically that can be used as a skeletal foundation for a knowledgebase [1]. This nature of ontology enables the developer to implement semantic based personalized learning applications.

The ontology developed for the educational domain contains the knowledge for developing intelligent learning system. Monolingual ontology applications for learning system are be developed by adopting the methodology [3]. Ontologies are used to represent knowledge which reflects the relevant information of the concepts and relations. There were many methodologies proposed to build ontology applications which have their own pitfalls. Modeling, evaluating and maintaining ontologies are a complex tasks in most applications such as healthcare, business, commerce and many other. There are many domains that necessitate satisfying the different language users. For example the users of government services, learning sites, education domains, healthcare domains demands to access information in their local language. In such scenario, ontology plays a vital role to provide knowledge based information. Numerous methods and tools are proposed for

building monolingual ontologies. Very few methods like Collaborative platform are proposed to build multilingual ontologies but they are limited to some languages. This chapter proposes new methodology to build multilingual ontologies. Rapid development of internet users demands on information in their natural languages which leads to the development of multilingual applications. The aim of this paper is to give an idea to develop multilingual ontologies for education domain using the proposed MOnto methodology. New algorithms are proposed for merging and mapping ontologies developed in different natural languages. The paper organized as follows: an overview of ontology based learning systems are narrated in section 2. Section 3 proposes a new methodology to build multilingual ontologies Conclusions are proposed in section 4.

2. STATE-OF-THE-ART OF ONTOLOGY-BASED LEARNING

Learning Ontologies are used in software agents, language independent applications and problem solving methods. Ontology applications are be developed using ontology development languages (OWL, RDF, TURTLE, Triple and so on) and ontology development tools(Protégé, OntoEdit, Chimaera and so on). Learning ontology application are be implemented in two different strategies: i) ontology of learning resources and ii) ontology of teaching strategy [4]. The ontology of learning resources is used for teaching knowledge modeling in e-learning system. The ontology of teaching strategies exhibits a series of macro teaching design and micro teaching activities. Ontology for learning may have personalized learning paths [5] which are used to improve the effectiveness of learning system. Personalization of e-learning process for the chosen target group will be achieved by setting up the learning path for each user according to their profile. Some models have been proposed to develop web based e-learning systems [6]. These model have been developed based on semantic web technologies and e-learning standards. These models provide two kinds of contents to the learners, they are: i) Learning content and ii) Assessment content and provides learning service and assessment service respectively. These models use the knowledge based information retrieval approach to repossess learning resources. The learning resources are described by means of metadata to implement the knowledge base.

Some ontology based learning systems have been developed to store and retrieve semantic metadata to provide better results to the learner along with personalized learning[7]. A systematic approach is proposed towards the development of semantic web services for e–learning domain. The following steps[8] are used to develop ontology for e-learning: i) determining the scope of domain, ii) reusing existing ontologies, iii) enumerating important terms in the ontology, iv) defining the classes and its hierarchy, v) defining the class properties, vi) defining the facets, vii) creating instances and viii) checking anomaly. The ontologies can be evaluated using SRIONTO (Software Risk Identification ONTOlogy) to identify the problem and risk in it [9]. The required concepts, the semantic description of the concepts and the interrelationship among the concepts along with all other ontological components have been collected from various literatures. E-learning resources can be collected using some frameworks [10]. These frameworks used to collect e-learning multimedia resources from the internet and automatically link them with topics.

Ontology-based approach can be used to develop personalized e-learning [11]. It is used to create an adaptive content based on learner's abilities, learning style, level of knowledge and preferences. In this approach, ontology is used to represent the content model, learner model and domain model. The content model describes the structure of courses and their components. The learner model describes the characteristics of learner's that are required to deliver tailored content. The domain model consists of some classes and properties to define domain topics and semantic relationships between them. It is used to assess the learner's performance by conducting the tests and the results are evaluated. The system recognizes changes in the learner's level of knowledge as they progress and the learner model is updated based on the learner's progress accordingly. However, most of the learning applications are developed either in English or in the developer language which become the hurdles of different language users to learn. Nowadays users of internet prefer to share their knowledge in their natural languages which emerges the technologies to support different natural languages. In a current scenario, enormous learning materials are available over the web which allows the user to benefit from anywhere in the world. Though the user gets large amount of information still they are longing for the information in their own languages. This motivates us to develop multilingual ontology applications to benefit different natural languages. In order to do that, MOnto methodology is proposed to build multilingual ontologies.

3. MONTO METHODOLOGY TO DEVELOP MULTILINGUAL ONTOLOGIES

A methodology is a "comprehensive, integrated series of techniques or methods creating a general systems theory of how a class of thought-intensive work ought to be performed" [12]. Methodology consists of methods and techniques where method is a process of performing some task and technique is a procedure used to achieve given objective. This research work proposes MOnto methodology to build multilingual

ontology applications. This methodology consists of five phases as given in Figure 1. viz. Input, Building MO, Ontology mediation, Retrieval and Visualization of ontology.

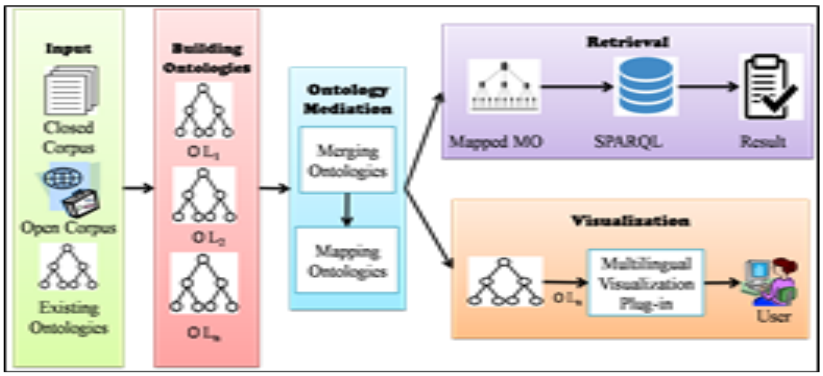


Figure 1. Monto methodology for building multilingual ontology

3.1. Phase 1: Input

This phase initializes the content to be considered for building ontologies. A set of methods and techniques are used for building ontology from distributed and heterogeneous knowledge and information sources. Information can be retrieved from different sources like, open corpus, closed corpus and existing ontologies. All the sources are under three categories: Unstructured sources, semi-structured source and structured source. Unstructured sources involve NLP techniques, morphological and syntactic analysis, etc. Semi-structured source elicits ontology from sources that have some predefined structure, such as XML Schema. Structured data extracts concepts and relations from knowledge contained in structured data, such as databases. Closed corpus is a text from the text books, study materials etc. Open corpus refers to the information available on the web. Corpus is used to represent the represents ontology by using a set of techniques to extract the knowledge from the text. In this phase, the scope and domain for building MO is identified. In order to build a new ontology for the specified domain, it is important to make sure that there is any ontology already available to the particular domain. In that case, the ontology has to be considered for reusing and re-engineering for building MO. The sources for building MO is collected as given in Table 1. The developer has to identify the domain to develop MO and has to collect the information from various sources in different natural languages. The collected resources are analyzed and classified in this initial phase.

Table 1. Document matrix for collecting resources in different natural languages

Source/Language	L1	L2	…	Ln
Open corpus	√	√		√
Closed corpus	√	√		
Existing ontology	√			√

3.2. Phase 2: Building ontologies

Once the domain is identified, the text extracted from closed corpus and open corpus in different natural languages is arranged hierarchically with the proper classifications. The terms required to build multilingual ontology are collected in different natural languages.

$L_1 = L_1t_1, L_1t_2, L_1t_3, \ldots\ldots\ldots\ldots L_1t_m$
$L_2 = L_2t_1, L_2t_2, L_1t_3, \ldots\ldots\ldots\ldots L_2t_m$
$L_n = L_nt_1, L_nt_2, L_1t_3, \ldots\ldots\ldots\ldots L_nt_m$

This can be represented as,

$$\forall\ i = 1\ to\ n \qquad L_i = L_it_1, L_it_2, L_it_3 \ldots\ldots\ldots L_it_m\ for\ m\ terms$$

where $n \neq m$

Collected terms are analyzed and irrelevant terms are filtered. The terms are classified hierarchically and the relations between the terms are established as,

$L_i t_j \overset{R_s}{\rightarrow} L_i t_k$

The relations between the terms are established and vocabularies of the terms are formulated. Using the laddering structure ontologies are developed in different natural languages (OL_1, OL_2,…, OL_n where OL is Ontology Language). 'N' Ontologies($OL_1, OL_2 \ldots OL_n$) are developed for L_i natural languages using the terms that are hierarchically structured as shown in Figure 2.

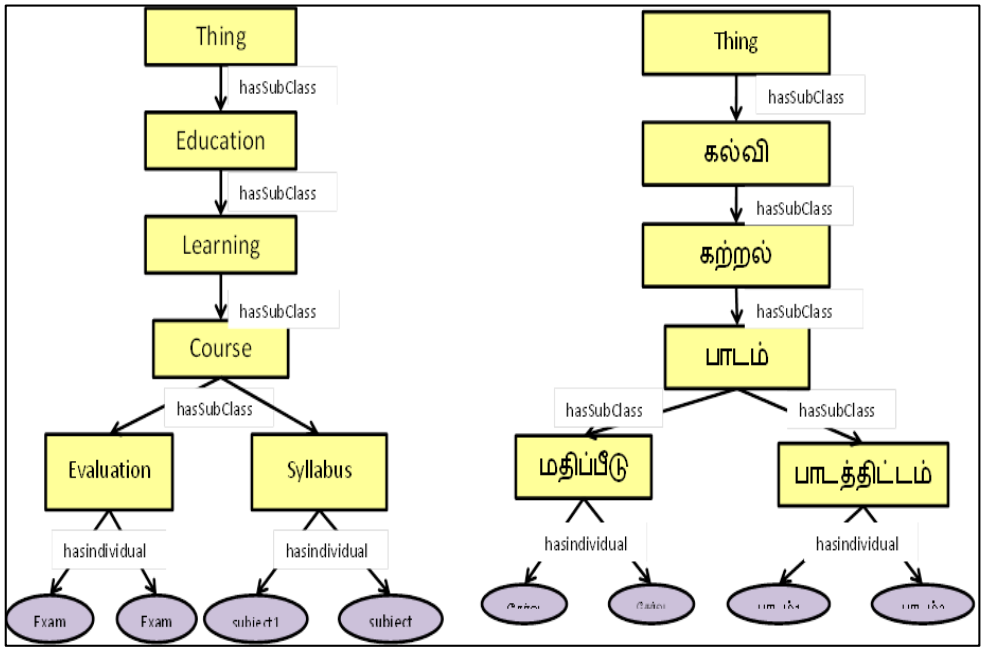


Figure 2. Illustration of building ontologies for two natural languages (Tamil and English)

3.3. Phase 3: Ontology mediation methods

Ontology mediation enables reusing of data across applications on Semantic Web, and sharing of data between heterogeneous knowledge bases. Major kinds of ontology mediation are mapping and merging. Ontology mapping is to identify the correspondence between the terms and ontology merging is creating new ontology which is the union of existing two or more ontologies. In this phase, ontologies developed in different natural languages are merged into single ontology and the correspondences between the terms of different natural languages are established. For example, OL_1, OL_2 … OL_n are the ontologies developed in different natural languages for the selected domain. where,

$OL_1 = \{ L_1t_1, L_1t_2, L_1t_3, \ldots L_1t_i \}$

$OL_2 = \{ L_2t_1, L_1t_2, L_1t_3, \ldots L_2t_i \}$

$OL_n = \{ L_nt_1, L_nt_2, L_nt_3, \ldots L_nt_i \}$

Ontologies developed in different natural languages are merged into a single ontology.

$ML = \{ OL_1 \cup OL_2 \cup \ldots \cup OL_n \}$

Correspondences between the terms in different natural languages are created

$L_nt_i \rightarrow L_nt_k$

where i and k vary from 1 to i terms in different languages.

Ontologies that are developed in different natural languages are merged into single ontology to structure multilingual ontology application. In formal, it can be represented as,

$$MO = \{X:\ O_1L_1 \cup O_2L_2 \cup \ldots..\cup O_nL_n\} \quad where\ L_i \geq 2 \wedge L_1 \neq L_n$$
$$L_1 \cap L_n\ is\ disjoint$$

Here, MO – Multilingual Ontology
L – Language
X – Set of elements

X is a collection of elements or terms which are integrated the sources of the same domain in different natural languages. Many tools like OntoClean, FCAMerge, and Observer are available to merge ontologies. The merged ontology composed of set of terms in different natural languages. Ontology merging can be done by using SMART algorithm [13]. This algorithm deals with merging and aligning of monolingual ontology of the domain. In order to overcome this, the algorithms for ontology mediation methods are proposed for merging and mapping ontology [14-21]. The research adapted those algorithms for merging and mapping multilingual ontologies.

3.4. Phase 4: Multilingual information retrieval using SPARQL

Information retrieval is the process of retrieving or extracting the information from the repository based on the user's need and query. Retrieving information in various languages can be named as multilingual information retrieval. In ontologies, SPARQL query is used to extract the knowledge from the ontology repository. RDF tags are used in SPARQL query to filter the results by means of language. This phase enables the users to extract knowledge in their own languages using SPARQL. SPARQL provides the functionality to retrieve the information in different natural languages. The sample SPARQL query is given as follows:

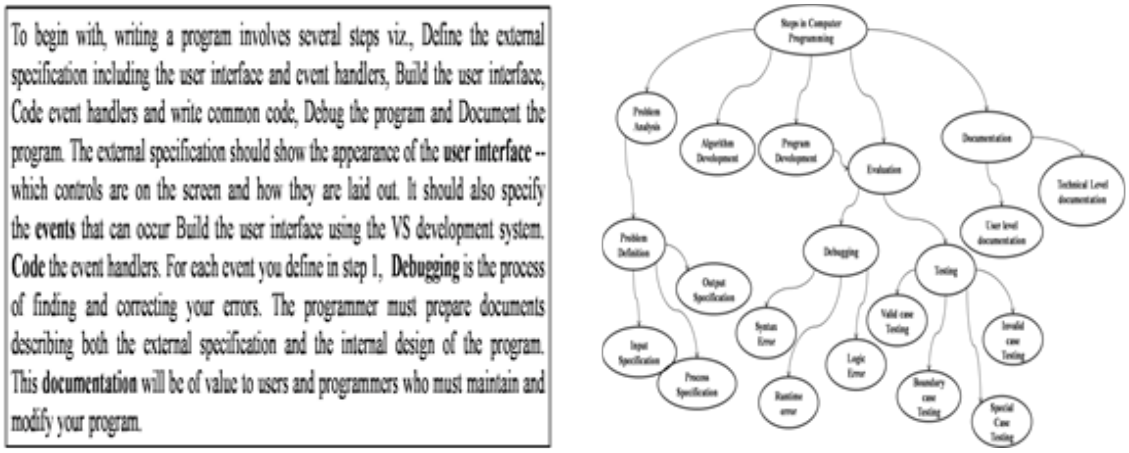
```
PREFIX scs: <http://www.shctptcs.org#>
SELECT? Subject? Object
WHERE
{
?subject scs: verse ?object.
FILTER (Lang (? object) ="ta")
}
```

The given SPARQL used 'FILTER' to sort the result and give the results of information in a specified language.

3.5. Phase 5: Visualizing multilingual ontology

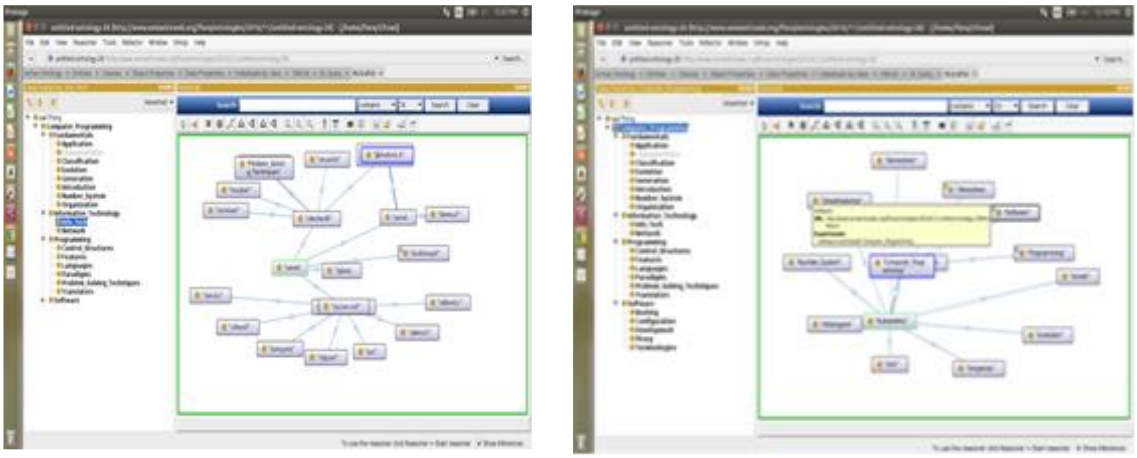
Visualization is a representation of text or object in the form of image or chart. It enables the readers to capture the knowledge effectively. Ontology is a hierarchically structured model which has numerous visualization tools (OWLGrEd, NavigOWL, IsAViz etc) and plug-ins (OntoGraf, OWLviz, CropCircles and so on). All the existing ontology visualization tools are lacking in visualizing non-English languages. Some of them require additional configuration to support different natural languages. In this phase, the new plug-in known MLGrafViz is proposed to visualize ontology in different natural languages. For example, the passage given in Figure 3 is represented diagrammatically in Figure 3 this depicts that the graphical representation of the text is clearer than the passage where the user may feel vague while reading a passage.

MLGrafViz is developed using Java and Graphviz algorithms. Initially, it allows the user to create a new ontology or to import an existing ontology into Protégé workspace. The imported ontology will be displayed in a class browser. MLGrafViz enables the user to select the language to visualize the ontology. The request is submitted to Google translate API which performs statistical machine translation and then the terms are translated into the desired natural languages. Google translate API is an open source translator used to translate text, speech, images and videos from source language to target language. It provides an API which allows the developer to build an extension and software to translate the source. Google translate uses statistical analyses instead of rule based analyses. Since ontology is hierarchically structured terms, statistical machine translator provides better result than the rule based translator. Rule based machine translation is used in translating the passage grammatically. Finally, the translated terms are displayed in MLGrafViz panel. MLGrafViz facilitates the user to visualize the ontology in different natural languages without changing the core ontology structure as depicted in Figure 4 (a), (b).



(a) (b)

Figure 3. Graphical representation, (a) Steps involved in programming – text, (b) visualization of steps involved in programming – diagrammatic representation



(a) (b)

Figure 4. MLGrafViz panel, (a) Visualization in Tamil language, (b) visualization in Zulu language

4. CONCLUSION

We have proposed MOnto (Multilingual Ontology) methodology to develop multilingual ontology application for education domain. New algorithms are proposed to perform merging and mapping of multilingual ontologies. This method allows the user to learn the subject from their own natural language which gives better understanding of the subject. This research work identifies the need of building multilingual application which plays vital role in educational domain. If the learning materials are in different natural languages, the learner will feel comfortable in learning. Learning through the natural languages is an essential thing which encourages the learner to learn many things. In future, multilingual applications can be implemented for different domain like healthcare. It is important to provide the evaluation metrics and methods to validate multilingual ontologies.

REFERENCES

[1] Asunción Gómez-Pérez, Mariano Fernández-López and Oscar Corcho. "Ontological Engineering with examples from the areas of Knowledge Management, e-Commerce and the Semantic Web." London Berlin Heidelberg, *Springer*, 2004.

[2] Naveen Kumar. "Ontology based Books Information Retrieval using SPARQL." *International Journal of Computer Applications*, vol. 67, no. 13, pp. 24-27, April 2013.

[3] Kanellopoulos, Dimitris, Sotiris B. Kotsiantis and Panayiotis E. Pintelas. "Ontology-based Learning Applications: A Development Methodology." *IASTED Conf. on Software Engineering*, 2006.
[4] Wang, Quanyu, Xingen Yu, Guilong Li and Guobin Lv. "Ontology-Based Ecological System Model of e-Learning." *International Journal of Information and Education Technology*, pp. 595-599, 2012.
[5] Bajenaru, Lidia, Ana-Maria Borozan and Ion Smeureanu. "Using Ontologies for the E-learning System in Healthcare Human Resources Management." *Informatică economică*, vol. 19, pp. 15-24, 2015.
[6] Nassim Khozooyi, Nafise seyedi, Razie Malekhoseini. "Ontology-based e-learning." *IRACST - International Journal of Computer Science and Information Technology & Security (IJCSITS)*, Vol. 2, No. 4, August 2012.
[7] Yusof, Norazah and Andi Besse Firdausiah Mansur. "Ontology Development of e-Learning Moodle for Social Learning Network Analysis." *World Academy of Science, Engineering and Technology, International Journal of Computer, Electrical, Automation, Control and Information Engineering*, vol. 7, pp. 853-858, 2013.
[8] Robin, C. R. Rene and Assistant. "Design and Development of SRIONTO: An Educational Ontology Representing Software Risk Identification Knowledge." *2nd International Conference and workshop on Emerging Trends in Technology (ICWET)*, 2011.
[9] Alhawiti, Mohammed Mfarij F. and Yasser Abdelhamid. "An Ontology-Based Framework for Collecting E-Learning Resources." *International Journal of Learning, Teaching and Educational Research*, vol. 9, 2014.
[10] Yarandi, Maryam, Hossein Jahankhani and Abdel-Rahman H. Tawil. "A personalized adaptive e-learning approach based on semantic web technology." *Webology*, vol. 10, 2013.
[11] Fernández-López, Mariano, Asun Gomez-Perez, Jérôme Euzenat, Aldo Gangemi, Yannis Kalfoglou, Domenico M. Pisanelli, Marco Schorlemmer, Geri Steve, Ljiljana Stojanović, Gerd Stumme and York Sure-Vetter. "A survey on methodologies for developing, maintaining, integrating, evaluating and reengineering ontologies." *Webology*, 2002.
[12] Fridman, Natalya and Mark A. Musen. "SMART: Automated Support for Ontology Merging and Alignment." (1999). *Proceedings of the 12th Workshop on Knowledge Acquisition, Modelling, and Management (KAW'99)*, 1999.
[13] Joseph, M.F., Lourdusamy, R. "Ontology mediation method for building multilingual ontologies." *Int. j. inf. tecnol*. Vol. 10, pp. 11–19, 2018.
[14] Ravi Lourdusamy, Merlin Florrence Joseph. "A Survey on Multilingual Ontologies." *International Journal of Modern Computer Science (IJMCS)*, pp. 71-76. 2015.
[15] Trojahn, Cássia & Quaresma, Paulo & Vieira, Renata. "An API for Multi-lingual Ontology Matching." *Proceedings of 7th conference on Language Resources and Evaluation Conference (LREC)*. pp. 3830-3835, 2010.
[16] Ganapathy, G., & Lourdusamy, R. "Ontology Merging and Matching Using Ontology Abstract Machine." *Proceedings of Knowledge Management 5th International Conference*, 2010.
[17] Dos Santos, C. T., Quaresma, P., & Vieira, R. "A Framework for Multilingual Ontology Mapping." *LREC*, 2008..
[18] Schwall, J. Creating an Ontology for a Multilingual E-Commerce Dictionary. University of Münster: s.n., 2007.
[19] Mejía, M. E., Montiel-Ponsoda, E., de Cea, G. A., & Gómez-Pérez, A. "Ontology localization." Berlin Heidelberg, Springer, *Ontology Engineering in a Networked World*, 2012.
[20] De Melo, G., & Siersdorfer, S. "Multilingual Text Classification using Ontologies." *Springer*, 2011.
[21] Dave Beckett. What does SPARQL stand for? SPARQL. [Online] October 6, 2011. [Cited: July 21, 2016.]