Using Ontologies for Developing a Search Mechanism in Social Networks

Marzieh Doroudian, Reza Akbari, Marzieh Ahmadzadeh
Shiraz University of Technology
Iran
{M.Doroudian, akbari, ahmadzadehm}@sutech.ac.ir

ABSTRACT: *Online social networks are a particular type of virtual community and in recent years they have attracted many people. These networks provide various features for users and allow them to construct a profile and share their information. Hence searching in users' profiles and their information is one of the most common operations in social networks. In this paper, we present a search method based on ontology in which for each profile in social network an ontology is created and then the same is done for user's query. Using ontology matching techniques these ontologies are matched with each other and the similarities between them are computed. Finally according to similarities, results related to user's query are found and represented. The experiments show that the presented method is efficient and provide satisfactory results.*

Keywords: Ontology, Ontology Matching, Social Networks

Received: 25 May 2014, Revised 1 July 2014, Accepted 5 July 2014

1. Introduction

Online social networks are a new generation of websites that currently have considerable popularity among the most famous and popular websites. The advancements in technology make it possible to develop new applications. Social networking play an important role in online communications. Also, they affect the distribution and organization of the data, information and knowledge. The efficient search in social networks is necessary. Hence, major web search companies are deploying services that leverage social networks. Yahoo and Google are examples of such search engines [1].

These networks provide features such as chat, messaging, email, video and audio transfer, etc. for users to enable them to interact with each other. Searching in social networks is one of the most common operations for which it is important to provide relevant results to users. In this area, there are many technologies but we cannot be sure which one provides links related to the user query correctly with fewer errors. So, in this paper, we present a method which tries to provide search results that are more accurate and desirable.

The rest of this paper is organized as follows: the definitions of social networks and an overview of some search methods are presented in Section 2. The proposed method is described in Section 3 and in Section 4, the experiments and their results will be examined. Finally, the paper concludes in Section 5.

2. Related Works

Expansion of the Internet has increased significantly in recent decades and led to enhancing interactions among individuals. So individuals, organizations and institutions can share their information through the Internet and have access to large amounts of information on the internet easily. Hence searching in the internet is considered as one of the most important applications of this global network.

People by searching in the internet can find a significant percentage of what they are looking for. They usually express what they are looking for in the form of a query request and according to this, a set of data with different formats are presented to them such as text, images and other items. To provide this information as an output, it is necessary to perform operations about which numerous studies have been carried out and varied technologies and methods have been proposed.

Stemming: In this method, first a list of keywords from the user query and the content of web pages or other documents is extracted and then these words are compared with each other according to their structural and morphological features. The internal structure of words is considered in morphology as an area in linguistics. Based on morphology, the various forms of a word may not affect the search results. For example, if the word is used in the plural or singular form, it will not affect the meaning of the word and both can be displayed as search results relevant to that word [2]. Although stemming is commonly used in information retrieval systems and significantly improves matching coverage, it also introduces noise. For example, as stated in [3], "*using the stemmer, both* "*marine vegetation*" and "*marinated vegetables*" stem to "*marin veget*", *which is undesirable*".

WordNet based methods: These methods (e.g. [4], [5]) for obtaining the similarity between two concepts use of WordNet. WordNet is a database of English words in which each word (noun, verb, adjective, and adverb) with a set of other words, constitute a group and describe a specific and distinct concept, as long as they are synonymous to each other. For each concept there can be several groups that each one has a specific relationship with that concept. Those includes synonymous to that concept, groups that are expressed in more general sense which are referred to as Hypernym and groups that are expressed in more specific sense and are called Hyponym [6].

Although WordNet has collected a relatively comprehensive set of words, but it does not provide any information about some compound words.

Translation models: These models are based on statistics and probability. Their goal is to determine which candidate document is more likely a translation (or transformation) of the user's query [7]. Although translation models have been used for different purposes such as document retrieval [7], question answering [8] but it seems that these models are less likely to be effective on short texts such as queries.

Query Expansion: In this technique, a set of related terms is added to the query and thus the query is changed from its initial form into a richer form which can cover more concepts. Therefore more relevant content will be extracted. These terms are extracted from documents, web pages or data sets that are retrieved in response to the initial query [9]. So the way to do an initial search is also very important because it is possible that extracted terms are not correct. In addition adding words to the initial form of the query may cause undesirable changes in its meaning.

In this article, the issue is finding users' required information in a social network. Online social networks are a particular type of virtual community and there are several definitions for them. Including:

Online social network: "*OSNs form online communities among people with common interests, activities, backgrounds, and/or friendships. Most OSNs are Web-based and allow users to upload profiles* (*text, images, and videos*) *and interact with others in numerous ways*" [10].

Social network site: "*We define social network sites as web-based services that allow individuals to* (1) *construct a public or semi-public profile within a bounded system,* (2) *articulate a list of other users with whom they share a connection, and* (3) *view and traverse their list of connections and those made by others within the system*" [11].

Unlike the web, which is organized around content, online social networks are formed around the users. In other words, social networks are comprised of user accounts and connections between users.

There is a variety of information in social networks that usually organized in the form of user profiles. One of the major search issues is quick and reliable access to this information.

3. Proposed Method

In this section, we present a method for searching in which the notion of ontology and its matching process is used because the ontology plays an important role in the search. Therefore, brief definitions of these two concepts will be presented:

Ontology: Ontology in a general sense refers to the branch of philosophy that studies the nature of existence or being as such. "*Ontology is derived from the two Greek words namely ontos which means* "*to be*" *and logos meaning* "*word*" " [12]. There are several definitions of ontologies. Here, one of the most cited definition aboutn the otology which is proposed by Gruber [13] is given:" ontology is a formal, explicit specification of a shared conceptualization."

Ontology Matching: as stated in [14], "*Ontology matching is the process of discovering similarities between two source ontologies. The result of a matching operation is a specification of similarities between two ontologies. The input of the operator is a number of ontology and the output is a specification of the correspondences between the ontologies*".

The search method is proposed based on the presented concepts. This method has several steps which are shown in Figure 1. The method starts by creating the ontology of profiles of members of the social network. Next the ontology of the user's query is created. After that the matching process between ontologies obtained from the two previous steps is executed. Finally, the scores are assigned to each result and they are sorted.

3.1 Creating the ontology of members profiles

Profiles on social networks include a variety of parts which usually contain personal information (e.g. name and photo), interests, job information, educational background and contacts. In this paper, job information is used in order to describe the performance of the proposed method. Thus, an ontology is created for each user's job title. For each of these titles that are stored in the form of a list of words, a set of related words (semantically) is extracted. These words are extracted from WordNet comprised of three categories: hypernyms, synonyms and hyponyms. After extracting words, ontology building process begins. These words will be classes of ontology which will be related to each other based on "*Super-class, Sub-class*" relation. In Figure 2 and 3, two ontologies are shown which are obtained by the proposed method.

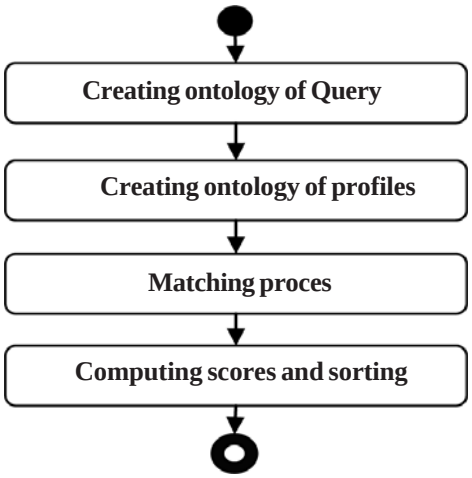

Figure 1. The main steps of the proposed search method

As shown, there is the same node in both ontologies called "*Thing*". This is a root to these ontologies because OWL language is used and ontologies that are created by using this language have this node as root node by default [15]. The number of Thing's children is equal to the number of synonyms for a word. Each word can have several different meaning and for each of these meanings, there is a set of synonyms and for each set of synonyms there is a set of hypernyms and usually a set of hyponyms. Each set often consists of several terms. Therefore, a term from each set will be selected randomly. Given that some sets have common terms, it should be checked that the selected terms are not repeated because these terms are used as ontology

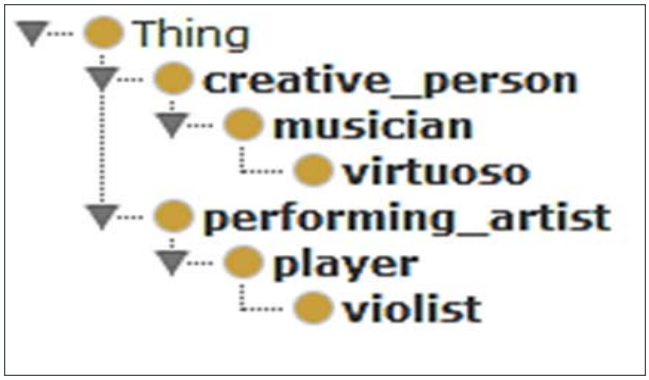

Figure 2. Musician's ontology

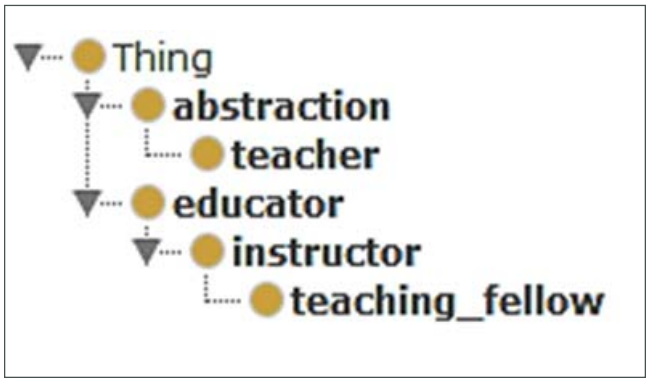

Figure 3. Teacher's ontology

classes and ontology classes cannot have the same name. The words are stored in a matrix of 3 * *n* where n is the number of synsets. In the second row of the matrix, each cell contains a term extracted from one of n synsets and for each term, one term of its hypernyms and also one term of its hyponyms are extracted and placed in upper cell and lower cell respectively. If a word does not have hyponym, null value is placed in the lower cell.

After extraction of the matrix terms, the process of constructing begins. If we consider ontology as a graph, the first row of the matrix can be considered as first-level children, second row will be intermediate nodes and third row are considered as leaves. In other words, each column of the matrix forms a branch of the ontology graph and thus each cell is child or subclass of its upper cell.

3.2 Creating the ontology of query

Query is received in the form of a string then it is decomposed into its tokens by a tokenizer. Among these tokens, prepositions, etc. are removed and only keywords remain. If necessary, the main forms of keywords are extracted. Then, for each keyword, ontology is created. Finally, these ontologies are given to ontology matching process as input.

3.3 Ontology matching and measuring similarity

The outputs of the first and second step are used as the input of this step. In this section, created ontology or ontologies for query are matched with each ontology for profiles. To compute the similarity between ontologies, the similarity between their entities must be earned. For this purpose, there are many techniques, including:

- String-based techniques
- Language-based techniques
- Constraint-based techniques
- Linguistic resources
- Graph-based techniques
- Statistics and Data analysis techniques [16]

In the proposed method, String-based techniques and Linguistic resources are used. To computing similarity measure, classes of each ontology are extracted then these techniques applied on them. The similarity between classes is calculated and stored in the similarity matrix.

String-based techniques) First the edit distance between classes in the query's ontology and profile's ontology classes (two by two) are calculated using the Levenshtein Distance and the results are stored in a similarity matrix. Edit distance, computes the number of required operations to transform one string (or word) to another. To compute the final edit distance, we use the following formula, because scoring range should be between 0 and 1

$$Levenstein = 1.0 - \frac{op}{max\ (source.\ Length,\ Target.Length)} \qquad (1)$$

where Op is the number of required operations to transform one string to another. Source and Target are two input strings to calculate the edit distance and Length computes the number of characters in the input string.

Linguistic resources) To apply these techniques, the query's ontology classes and profile's ontology classes (two by two) are compared with each other, and then similarity measures are computed according to the following conditions:

1. If both class names are absolutely identical, similarity will be equal to 1.
2. If profile's ontology class belongs to synsets of query's ontology class, the similarity will be equal to 0.8.
3. If profile's ontology class belongs to hyponyms of query's ontology class, the similarity will be equal to 0.5.
4. If profile's ontology class belongs to hypernyms of query's ontology class, the similarity will be equal to 0.3.

These techniques are also applied to the other profiles' ontology and the similarity matrix will be generated. Thus for each profile's ontology two similarity matrix will be generated.

Assigning scores to each user) This step takes the similarity matrixes and calculates the average of the values Sim in each matrix. Thus, for each user two values are obtained. Since the results of the second technique are more important than first technique, for each of them is considered a weight. Finally, for each user a similarity value will be obtained using the following formula:

$$sim_{user_i} = \frac{(w_1 \times a) + (w_2 \times b)}{w_1 + w_2}, \; w_1 > w_2, \; 1 \le i \le k \tag{2}$$

$$a = \frac{\sum_{i=0}^{n} \sum_{j=0}^{m} matrix\_lingua\,[i]\,[j]}{n \times m} \tag{3}$$

$$b = \frac{\sum_{i=0}^{n} \sum_{j=0}^{m} matrix\_ED\,[i]\,[j]}{n \times m} \tag{4}$$

Where K is number of users. matrix_ED is first technique's similarity matrix and Matrix_Lingua is second technique's similarity matrix. After performing the above steps, the values *Sim* are sorted in descending order.

4. Experiment

It seems that the proposed method can provide an acceptable and reliable response. Therefore, the results of the experiments will be presented to evaluate this approach. In this experiment several Google+ members' job title have been used. Like other social network Google+ allows users to create profile and share their information. Each social network usually defines a particular structure for its profiles and therefore their titles, topics and types of information are different. Figure 3 shows the structure of Google+ profiles. As shown in Figure 4, Google+ profiles are composed of about 9 parts and each part has a subset.

In the first experiment the proposed method receive a single word as the input query. Since this method is only used information contained in "*Work*" just this section is depicted in detail. One of the fields in "*Employment*" is "*Job title*", which displays the title of the members and its content. According to the previous sections, for each job title, an ontology is created and then the same has been done for user's query. Finally, the matching process and the scoring have been done. Figure 5 shows the results of the first experiment in which the user is looking for "*Teacher*". Only the first 30 results are shown in this figure. The results show tht the ontology based mathod has effeciency and present results in a meaningful approach.

In the second experiment the proposed method receives a query with more than one word as input. If the number of keywords extracted from the user query is more than a word or in other words, the number of query's ontologies is more than one ontology, each of them separately with each of the profiles' ontology are matched. Thus, to each user is assigned several scores for which an average is taken to form the user's final score. Finally the resulted scores are sorted in a descending order. Figure 6 shows the results of the second experiment in which the user is looking for "*Expert Programmer*". Only the first 30 results are shown in this figure.

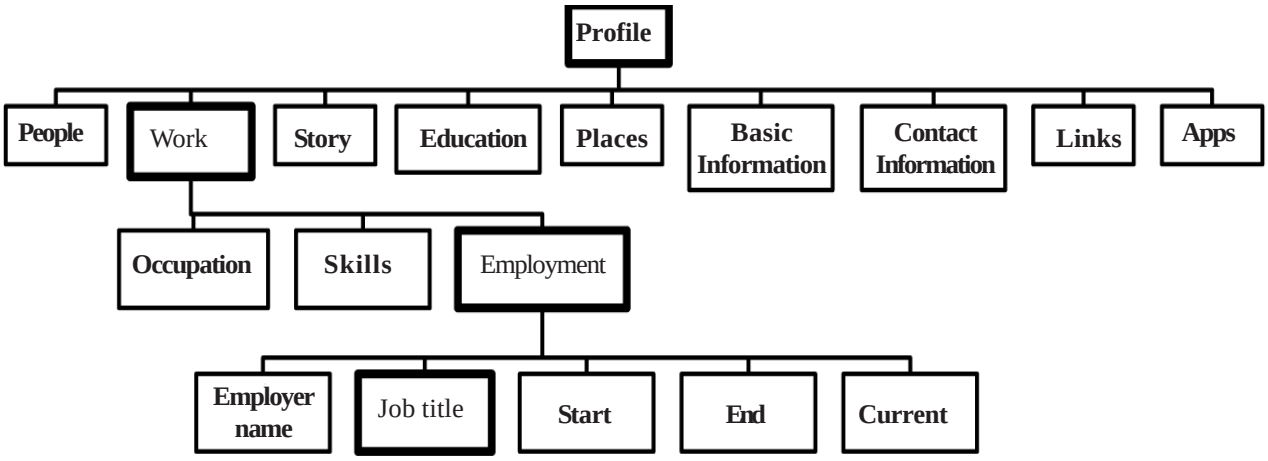

Figure 4. Structure of Google + profile

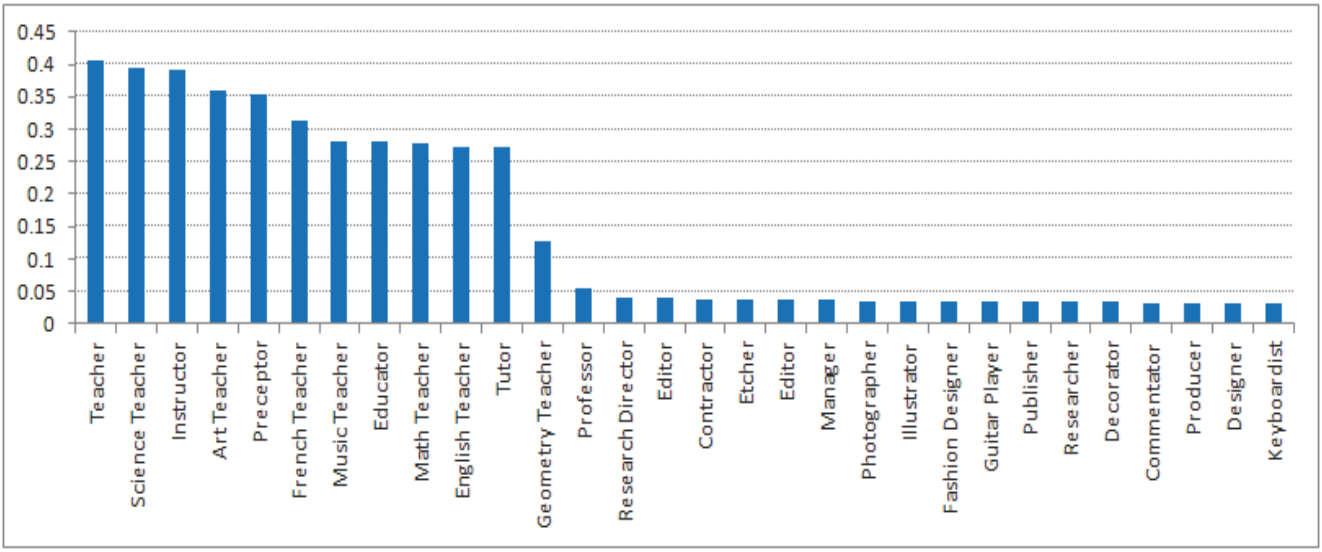

Figure 5. Results of first experiment in which the user is looking for "*Teacher*"

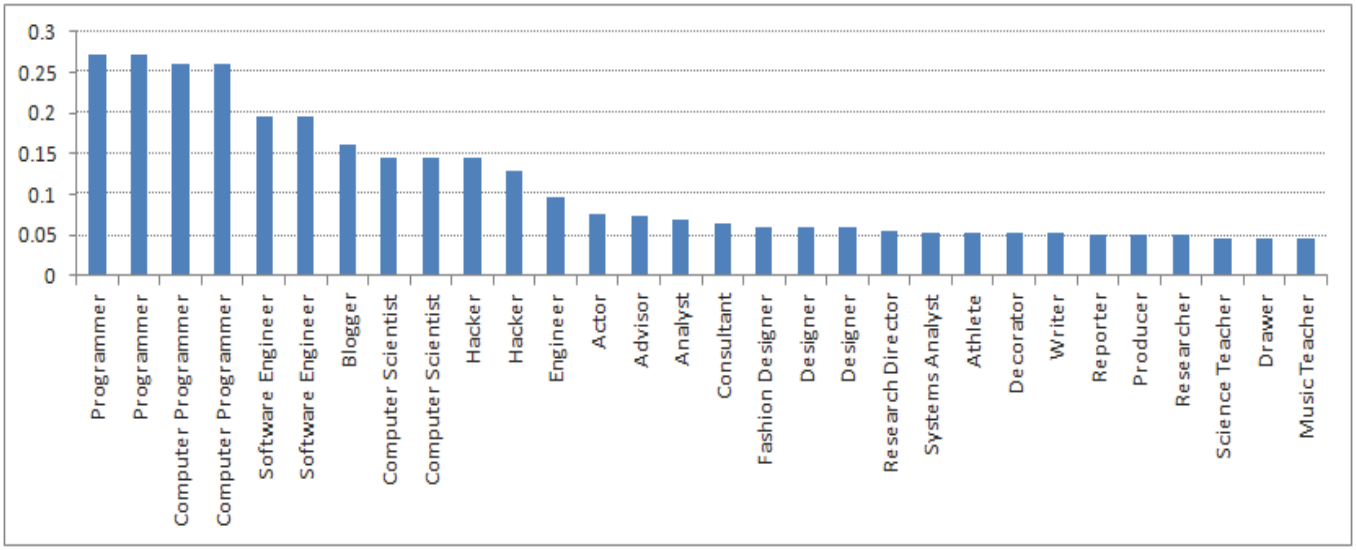

Figure 6. Results of second experiment in which the user is looking for "*Expert Programmer*"

As can be seen in these tables, members are sorted by scores earned. In other words, they are organized according to similarity of their job titles with the specified job title. The results show that the proposed method can successfully find the desired information and it is able to provide information that is relevant to a user query.

5. Conclusion

Searching and finding relevant information and providing it to users is known as one of the key problems in computer science. The effecient search in social nework is important for the users. In this paper, a method for ontogy-based search in social networks was presented. Therefore, first, the definitions of these networks were presented as well as a review on existing h for matching text. Then the proposed method was presented together with the results of the experiments, which were performed using this method. The presented results showed that the proposed method works effectively and is able to provide reliable results.

References

[1] Mislove, A., Marcon, M., Gummadi, K. P., Druschel, P., Bhattacharjee, B. (2007). Measurement and Analysis of Online Social Networks, *In*: Internet Measurement Conference (IMC), San Diego, California, USA, p. 29–42.

[2] Krovetz, R. (2000). Viewing morphology as an inference process, Artificial Intelligence, 118, p. 277-294.

[3] Metzler, D., Dumais, S., Meek, C. (2007). Similarity Measures for Short Segments of Text, *In*: *Advances in Information Retrieval*, ed: Springer Berlin Heidelberg, p. 16-27.

[4] Agirre, E., Alfonseca, E., Hall, K., Kravalova, J., Pasca, M., Soroa, A. (2009). A study on similarity and relatedness using distributional and WordNet-based approaches, *In*: Proceedings of Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics, ed: Association for Computational Linguistics, , p. 19-27.

[5] Saric, F., Glavas, G., Karan, M., Snajder, J., Basic, B. D. (2012). TakeLab: systems for measuring semantic text similarity, *In*: *First* Joint Conference on Lexical and Computational Semantics (*SEM), Montre´al, Canada, p. 441-448.

[6] Miller, G. A., Beckwith, R., Fellbaum, C., Gross, D., Miller, K. (1990). Introduction to WordNet: An On-line Lexical Database, International Journal of Lexicography, 3, p. 235-244.

[7] Berger, A., Lafferty, J. (1999). Information retrieval as statistical translation, *In*: 22nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, Berkeley, California, USA, p. 222-229.

[8] Murdock, V., Croft, W. B. (2005). A translation model for sentence retrieval, *In*: Human Language Technology and Empirical Methods in Natural Language Processing, Vancouver, British Columbia, Canada, p. 684-691.

[9] Sahami, M., Heilman, T. D. (2006). A web-based kernel function for measuring the similarity of short text snippets, *in* 15th *international conference on World Wide Web*, Edinburgh, Scotland, p. 377-386.

[10] Schneider, F., Feldmann, A., Krishnamurthy, B., Willinger, W. (2009). Understanding Online Social Network Usage from a Network Perspective, *In*: Proceedings of the 9th ACM SIGCOMM conference on Internet measurement conference, Chicago, Illinois, USA, p. 35-48.

[11] Boyd, D. M, Ellison, N. B. (2008). Social Network Sites: Definition, History, and Scholarship, *Journal of Computer-Mediated Communication*, 13, p. 210-230.

[12] Breitman, K. K., Casanova, M. A., Truszkowski, W. (2007). Semantic Web: Concepts, Technologies and Applications. Springer Verlag.

[13] Gruber, T. R. (1993). *A Translation Approach to Portable Ontology Specifications*, Knowledge Acquisition, 5, p. 199-220.

[14] J. Wiley, Semantic Web Technologies: Trends and Research in Ontology-based Systems.

[15] Pulido, J. R. G., Ruiz, M. A. G., Herrera, R., Cabello, E., Legrand, S., Elliman, D. (2006). *Ontology languages for the semantic web: A never completely updated review*, Knowledge-Based Systems, 19, p. 489-497.

[16] Euzenat, J., Shvaiko, P. (2007). Ontology Matching: Springer-Verlag Berlin Heidelberg.