Ranking Semantic Similarity Association in Semantic Web

Shahdad Shariatmadari[1]
Islamic Azad Universiry, Shiraz, Iran.
shariatmadari@iaushiraz.ac.ir

Ali Mamat
University Putra Malaysia
ali@fsktm.upm.edu.my

Hamidah Ibrahim
University Putra Malaysia
hamidah@fsktm.upm.edu.my

Norwati Mustapha
University Putra Malaysia
norwati@fsktm.upm.edu.my

ABSTRACT

Discovering and ranking complex relationships in the semantic web is an important building block of semantic search applications. Although semantic web technologies define relations between objects but there are some complex (hidden) relationships that are valuable in different applications. Currently, users need to discover the relations between objects and find the level of semantic similarity between them. (I.e. find two similar papers).

This paper presents a new approach for ranking semantic similarity association in semantic web document, based on semantic association concept.

Categories and Subject Descriptors

I.2.8 [Artificial Intelligence]: Problem Solving, Control Methods, and Search

General Terms

Algorithms, Measurement, Performance, Design.

Keywords

Semantic Similarity, Semantic Association, Semantic Web.

1. INTRODUCTION

The problem of measuring “similarity” of objects arises in many applications, and many domain-specific measures have been developed. Similarity is an important and fundamental concept in AI and many other fields (One obvious example is the “find-similar-document” query on the World-Wide Web).

As semantic web[4] is going to be popular, some applications like semantic search need to investigate on relationships between objects in semantic web documents. Although populated ontologies contain a bulk of concepts and instances and relations, but there are many type of relationships that aren't defined in ontology schema.

On of the most important concepts for discovering and ranking these types of relationship, is semantic association [3]. Based on this concept, two entities are *semantically associated* if they are either *semantically connected* (there exists an alternating sequence of properties and entities connecting them) or *semantically similar* (there exists a pair of matching property emanating from them).

Although some methods have been proposed for discovering and ranking semantic connectivity association, but to our knowledge, discovering semantic similarity based on the definition in semantic association is still untouched.

[1] He is a PhD student in Faculty of Computer Science and Information Technology, UPM. Also, he is a faculty member in Department of Computer Engineering, Islamic Azad University of Shiraz-IRAN.

This paper proposes SwSim, a method for discovering and ranking semantic similarity. The remainder of the paper is organized as follows. First we review the most popular similarity techniques in Section 2, while a motivating example is defined in Section 3. Then, in Section 4, we present our method. Finally, Section 5 concludes this paper.

2. RELATED WORK

Evaluating semantic relatedness and finding semantic similarity between concepts in ontology/taxonomy is a problem with a long history. Many techniques have been proposed for evaluating semantic similarity between concepts in hierarchical ontology (more than 80% of all relations are *is-a* relationship) like WordNet and GeneOntology. These approaches can be classified into two categories: edge based and node based approaches.

The edge based approach as the simplest similarity measure, computes the distance between two concepts based on the number of edges found on the path between them [7]. In node based approach, [6] defined the similarity between two concepts as the ratio between the amount of information needed to state the commonality between these two concepts and the information needed to fully describe them. Obviously, these methods aren't applicable for ontologies with different types of relationship; however, we use the general idea of these methods in our approach.

Computing semantic similarity is an important approach in multi ontologies applications such as semantic data integration and ontology mapping.

Our approach is based on [3]. Anyanwu et al. defined the formalism of semantic association based on semantic connectivity and semantic similarity. After they defined these concepts and ρ-operator, other researchers proposed different techniques in order to discover semantic association relations in semantic web documents based on semantic connectivity [2]; however, there are few researches in semantic similarity. One of the most related work in this area is TRAKS[1].

As yet, although there are some previous approaches for discovering and ranking semantic connectivity association, to our knowledge, there has been no proposal to discover and rank semantic similarity association based on semantic association concept.

3. MOTIVATING EXAMPLE

Current methods in discovering semantic association are based on semantic connectivity. However, there are some semantic associations that can't be found by semantic connectivity.

Obviously, one path connecting two entities will be valid path for "semantic connectivity". But the same path does not necessarily is a valid result for "semantic similarity". For example, suppose entity USA and entity Tokyo.

The following two paths are valid results of semantic similarity (even though paths are quite short).

a) USA -- located_in -- North_America
b) Tokyo -- located_in – Japan

The following path is intersection (through entity 'Canada') of the entities USA and Tokyo but is not valid semantic similarity result.

USA -- neighboring_country -- Canada -- does_business_with – Tokyo

As another example, Atlanta, and USA are two entities, and there's a path that makes them connected in semantic connectivity:

Atlanta -- located_in --Georgia -- located_in -- USA

Now, in a query for semantic similarity, that takes as input two entities: USA and Atlanta, these two paths are valid results, where one of them can be found by similarity.

Atlanta --- located_in ---- Georgia ---- located_in --- USA

USA --- located_in ---- Georgia ---- located_in --- Atlanta

4. OUR APPROACH

In a semantic association search, the main knowledge-base graph is explored to discover complex relationships between a Subject and its semantically related Objects [2]. Due to the directed graph structure of the ontology (RDF/RDFS) representation, searching the semantic associations would be a graph traversing task.

Figure 1 show our model for discovering semantic similarity between two entities. Based on this model, discovering semantic similarity and ranking the results are two major building blocks.

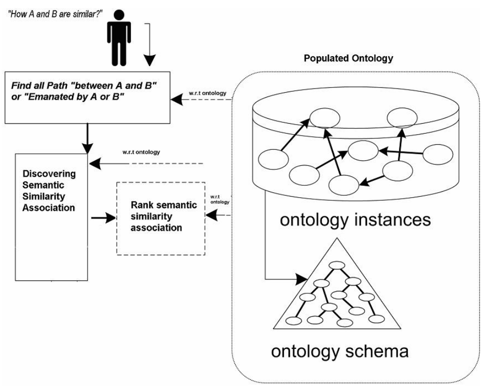

Figure 1: SwSim model

For discovering semantic similarities, we use the proposed ρ-operators [3].ρ-Iso and ρ-Join discover two related/similar entities. ρ-Iso discover two paths that have the same sequence of properties, but start from different resources. On the other hand, ρ-join discover two paths (the same properties) that intersect in a certain point. By using SPARQLeR[5], we can use two path variables and define the same path regular expression for them. It means that they will include the same properties, in the same order (moreover, in case of using hierarchy of properties, so the subsumption will work correctly).

For ranking semantic similarity, by using an approach in [8], we propose a formula. Consider two simple paths:

$s_0 \xrightarrow{a_1} s_1 \xrightarrow{a_2} s_2 \ldots s_n$

$t_0 \xrightarrow{b_1} t_1 \xrightarrow{b_2} t_2 \ldots t_n$

Similarity between these paths can be calculated based on the following formula:

$$Q(s_0 , t_0) = (1 - p) . V(s_0 , t_0) + p. E(s_1 , t_1) . Q(s_1 , t_1)$$

In this formula, V is a node (vertex) similarity function, E is edge similarity function, Q is rank function and p is a parameter.

For both functions V and E, we consider the hierarchy of classes and properties (*SubClassOf*, *SubPropertyOf*) in ontology. In this case, we use the subsumption function proposed by Aleman-Meza [2]. With parameter p ($0<p<1$), we can change the level of similarity (edge based/node based). By considering p=0, the similarity between two path will be calculated based on the degree of similarity of the nodes, however, if p=1, the similarity between two path will be calculated based on the degree of similarity of the edges.

The preliminary result show the soundness of our approach, however, we are planning to consider some metrics like context and also weighting the edges of graph, based on relation type, to improve the ranking method in near future.

5. CONCLUSION

This paper presented how to discover semantic similarity in semantic web based on semantic association concept. Although there are different approaches in discovering and ranking semantic association based on semantic connectivity, but to our knowledge this paper proposed the first approaches in discovering and ranking of semantic similarity association. Our naive algorithm and primarily examples show the soundness of our approach.

In the future, we will complete our model for discovering semantic similarity in order to rank the result in more accurate way. For this purpose we may use a semantic proximity function (use fuzzy predicates such as "close-enough") to calculate the semantic similarity between two entities.

6. REFERENCES

[1] Aleman-Meza, B. , Halaschek-Wiener1, C. , Sanket Sahoo, S. , Sheth, A. and Arpinar, B. 2005. Template Based Semantic Similarity for Security Applications , Springer.

[2] Aleman-Meza, B. 2006. Searching and Ranking Documents based on Semantic Relationships, ICDE Workshops.

[3] Anyanwu, K. and Sheth, A. 2002. The ρ Operator: Discovering and Ranking Associations on the Semantic Web, SIGMOD, ACM.

[4] Berners-Lee, T. , Hendler, J. and Lassila, O. 2001. The Semantic Web: A new form of Web content that is meaningful to computers will unleash a revolution of new possibilities", Scientific American.

[5] Kochut, J. and Janik, M. 2007. SPARQLeR: Extended Sparql for Semantic Association Discovery. Springer, 145-159.

[6]Lin, D. 1998. An information-theoretic definition of similarity. In Proceedings of the 15th Int. Conf. on Machine Learning.

[7]Resnik, P.1995. Using information content to evaluate semantic similaritity. In Proceedings of the IJCAI05, pages 448 –453.

[8]Sokolsky, O., Kannan,S., and Lee, I. 2006. Simulation-Based Graph Similarity. Proceedings of 12th International Conference on Tools and Algorithms for the Construction and Analysis of Systems(TACAS 2006), pages 426-440.