

Semantic Search and SemSime

Subjects: Computer Science, Information Systems

Contributor: Anna Formica

This paper presents *SemSim^e*, a semantic similarity method for searching over a set of digital resources previously annotated by means of concepts from a weighted reference ontology.

It is based on a frequency approach for weighting the ontology, and refines both the user request and the annotations of the digital resources with rating scores. Such scores are High, Medium, and Low and, in the user request, indicate the preferences assigned by the user to each of the concepts representing the searching criteria whereas, in the annotations of the digital resources, they represent the levels of quality associated with each concept in describing the resources. The experiment we have performed shows that *SemSim^e* outperforms the previous semantic search method SemSim.

Keywords: similarity reasoning ; semantic search ; reference ontology ; semantic annotation

1. Introduction

The most significant improvement within the Semantic Web research area pertains to reasoning and searching abilities. In this perspective, semantic similarity reasoning, which relies on the knowledge coded in a reference ontology ^[1], is a different technique with respect to the well-known deductive reasoning used in expert systems. In ^[2], we proposed SemSim, a semantic search method based on a Weighted Reference Ontology (WRO). In SemSim, both the resources in the search space and the requests of users are represented by means of an Ontology Feature Vector (OFV), which is a set of concepts from the WRO. We distinguish the user request, also denoted as Request Vector (RV), from the description of a resource, also referred to as Annotation Vector, indicated by AV. In the search process, SemSim contrasts the RV against each AV, and the result is a ranking of the resources that exhibit the highest similarity degree with respect to the request defined by the user.

In ^[2], we analyzed two different approaches in order to weigh the reference ontology, namely the frequency-based and the uniform probabilistic approaches. In the experiment described in that paper, we show that SemSim by the frequency-based approach outperforms the SemSim by the uniform probabilistic approach, as well as the most representative similarity methods from the literature.

2. SemSime

In this work, we present a new method, referred to as *SemSim^e*. It relies on the frequency-based approach and revises SemSim along two directions. According to the first direction, in contrasting the RV with the AV, *SemSim^e* takes into consideration the cardinality of the set of the concepts (features) in the user request rather than the maximal cardinality of the compared OFV. This choice allows us to give more relevance to the features which are requested by the user rather than the extra features contained in the annotation vectors available in the search space. Along the second direction, SemSim has been enhanced with the rating scores High (H), Medium (M), and Low (L) in the OFV, with regard to both the request and the search space resources. Within the request, rating scores denote the preferences given by the user to the concepts of the WRO used to specify the query whereas, within the annotation vectors, rating scores represent the levels of quality associated with the concepts when they describe the resources. Consider an example rooted in the tourism domain, where the user is searching for a vacation package by specifying the following features: InternationalHotel (H), LocalTransportation (M), CulturalActivity (H), and Entertainment (L). On the basis of the given rating scores, he/she gives a high preference to resorts which are international hotels offering cultural activities, and less priority to the remaining features, in particular to the entertainments. Analogously, a holiday package annotated with HorseRiding (H), Museum (M), and ThaiMeal (L) is characterized by a high quality level with regard to the horse riding service, rather than the facilities in visiting museums or having Thai meals at lunch or dinner. Note that, in ^[3], a proposal concerning rating scores was given, where the concepts of the WRO are weighted according to the uniform probabilistic approach ^[4], rather than the frequency-based one. Furthermore, in our approach we assumed that, given a

facility (for instance, HorseRiding) included in a tourist package, the higher the user's priority about that facility, the higher the expectancy about the quality of the same facility and, therefore, the greater the availability of the user for considering more expensive solutions.

In this paper, we have experimented *SemSim^e* in the domain of tourism and we have compared it to the SemSim method defined in [2] and a further evolution of SemSim, referred to as SemSimRV. Essentially, SemSimRV is the original SemSim method where, in line with the first direction adopted in *SemSim^e* illustrated above, more priority has been given to the features indicated by the user in his/her request. The results of the experiment show that *SemSim^e* outperforms both these methods.

References

1. Thomas R. Gruber; A translation approach to portable ontology specifications. *Knowledge Acquisition* **1993**, 5, 199-220, [10.1006/knac.1993.1008](https://doi.org/10.1006/knac.1993.1008).
2. Anna Formica; Michele Missikoff; Elaheh Pourabbas; Francesco Taglino; Semantic search for matching user requests with profiled enterprises. *Computers in Industry* **2013**, 64, 191-202, [10.1016/j.compind.2012.09.007](https://doi.org/10.1016/j.compind.2012.09.007).
3. Michele Missikoff; Anna Formica; Elaheh Pourabbas; Francesco Taglino; Enriching semantic search with preference and quality scores. *Semantic Computing* **2017**, 1, 141-148, [10.1142/9789813227927_0016](https://doi.org/10.1142/9789813227927_0016).
4. Formica, A.; Missikoff, M.; Pourabbas, E.; Taglino, F. *Weighted Ontology for Semantic Search*; Lecture Notes in Computer Science; Springer: Berlin/Heidelberg, Germany, 2008; Volume 5335, pp. 1289–1303.

Retrieved from <https://encyclopedia.pub/entry/history/show/7870>