



Discovering the Semantics of User Keywords



Mauricio Espinoza, Jorge Gracia, Raquel Trillo and Eduardo Mena
IIS Department, University of Zaragoza, Zaragoza, Spain

Goal

Problem:

- Although keyword-based search is a widely used technique for information retrieval, traditional techniques do not consider the specific semantics assigned by the user to keywords.
- Syntactic-based search engines are very influenced by the enormous amount of information about popular issues on the Web.

Solution:

We propose a system that takes as input a list of plain keywords provided by the user, discovers their semantics in run-time, and obtains a list of senses extracted from different ontology pools; it deals with the possible semantic overlapping among senses.

Use:

- Disambiguation process across keywords
- Access to data corresponding to the semantics of user keywords

Main steps of our approach

As motivating example, let us suppose that a user is interested in defining the concepts of the keywords "java" and "developers" (for example for semantic annotation of web pages, text disambiguation, etc.), and his intended sense is developer of the programming language java. In the following, we summarized the main steps of our system, more details can be found in [1].

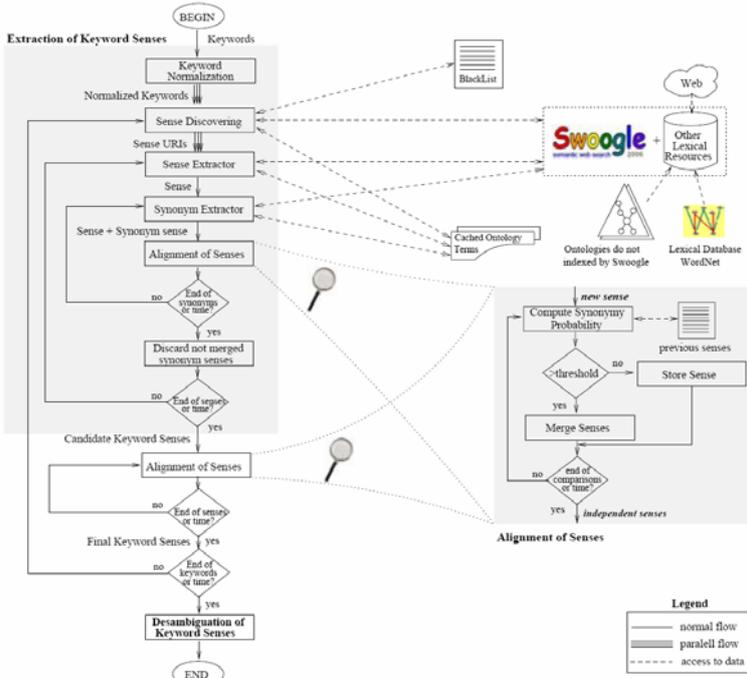


Figure 1: Main steps of our approach

References

- [1] M. Espinoza, R. Trillo, J. Gracia and E. Mena. Discovering and Merging Keyword Senses using Ontology Matching. 1st International Workshop on Ontology Matching (OM-2006) @ 5th International Semantic Web Conference ISWC-2006, Athens, Georgia (USA), CEUR-WS, November 2006.
- [2] T. Finin, L. Ding, R. Pan, A. Joshi, P. Kolari, A. Java, and Y. Peng. Swoogle: Searching for knowledge on the Semantic Web. In AAAI 05 (intelligent systems demo), July 2005.
- [3] G. Miller. WordNet: A Lexical Database for English. Communications of the ACM, 38(11), nov 1995.
- [4] J. Gracia, R. Trillo, M. Espinoza, and E. Mena. Querying the web: A multontology disambiguation method. In Sixth International Conference on Web Engineering (ICWE'06), Palo Alto (California, USA). ACM, July 2006.

Acknowledgements

This work was supported by the CICYT project TIN2004-07999-C02-02.

Contact

Dr. Eduardo Mena, (emena@unizar.es, <http://www.cps.unizar.es/~mena>)
Research group SID (<http://sid.cps.unizar.es>)

Extraction of Keyword Senses

The *sense discovering step* searches Swoogle [2], WordNet [3] and other ontology repositories to find ontological terms that match those keywords. The system builds a sense for each URI obtained with the information retrieved from matching terms in the ontology pool [1]. As matching terms could be ontology classes, properties or individuals, three lists of possible senses are associated with each keyword k : S_k^{class} , S_k^{prop} and S_k^{indv} .

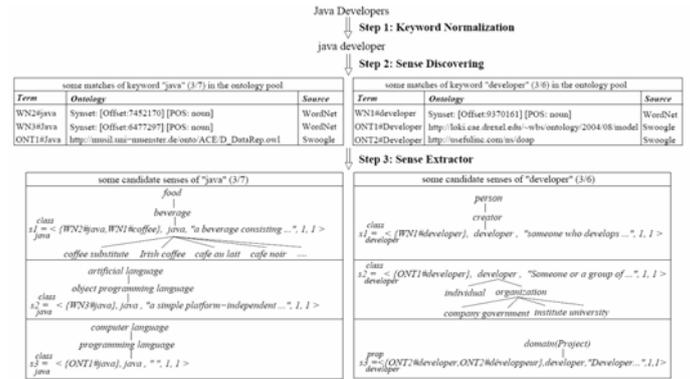


Figure 2: Extraction of senses of the user keywords "java" and "developer".

Each keyword sense is enhanced incrementally with the ontological terms of their synonyms. A sense alignment process integrates the keyword sense with those synonym senses representing the same semantics.

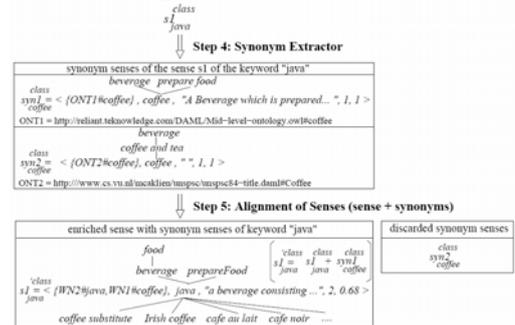


Figure 3: Semantic enrichment of the sense s1 of the keyword "java".

Alignment of Senses

This process is used in two situations by our system: 1) to check which synonym senses represent the same semantics as their keyword senses, and 2) to avoid redundancy in the list of possible senses of each user keyword. Senses are merged when the estimated *synonymy probability* among them is above a certain threshold. The synonymy measure combines a standard string distance metric with a structural similarity measure that is based on vector space techniques.

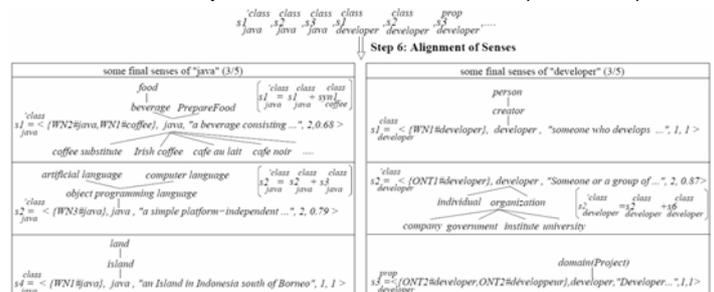


Figure 4: Sense alignment of candidate senses for "java" and "developer".

Disambiguation of Keyword Senses

A disambiguation method is used to select the most probable intended meaning of each user keyword. We propose a method to disambiguate iteratively each user keyword taking into account the possible senses of the other keywords [4]. Such a method computes a relatedness measure based on Google to compare each possible sense with respect to the senses of the other keywords.



Figure 5: Sense disambiguation of final senses for "java" and "developer".