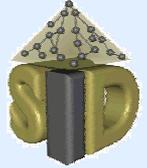




Overview of a Semantic Disambiguation Method for Unstructured Web Contexts



Jorge Gracia and Eduardo Mena
IIS Department, University of Zaragoza, Spain

Goals

- Discovering the intended meaning of words in unstructured web contexts, by applying word sense disambiguation techniques.
- Representing such meanings as terms of dynamically accessed ontologies.

Our proposal

1. **Context selection.** We select the context words that are most highly related to the ambiguous keyword, by computing their web-based relatedness [1].
2. **Obtaining candidate senses.** Online and local resources are accessed to provide a set of candidate senses for the keyword. Each sense corresponds to an ontology term or to the integration of various ontology terms.
3. **Disambiguation algorithm.** It is performed in three steps:
 - a. We first explore the semantic *relatedness* among the keyword senses and the words in the context [1,2].
 - b. We measure the *overlap* between the words in the context, and the words that appear in the semantic definition of the sense.
 - c. If the right sense is not clear yet, we consider the *frequency* of usage of senses as an additional factor for disambiguation.

Example

KEYWORD: bank

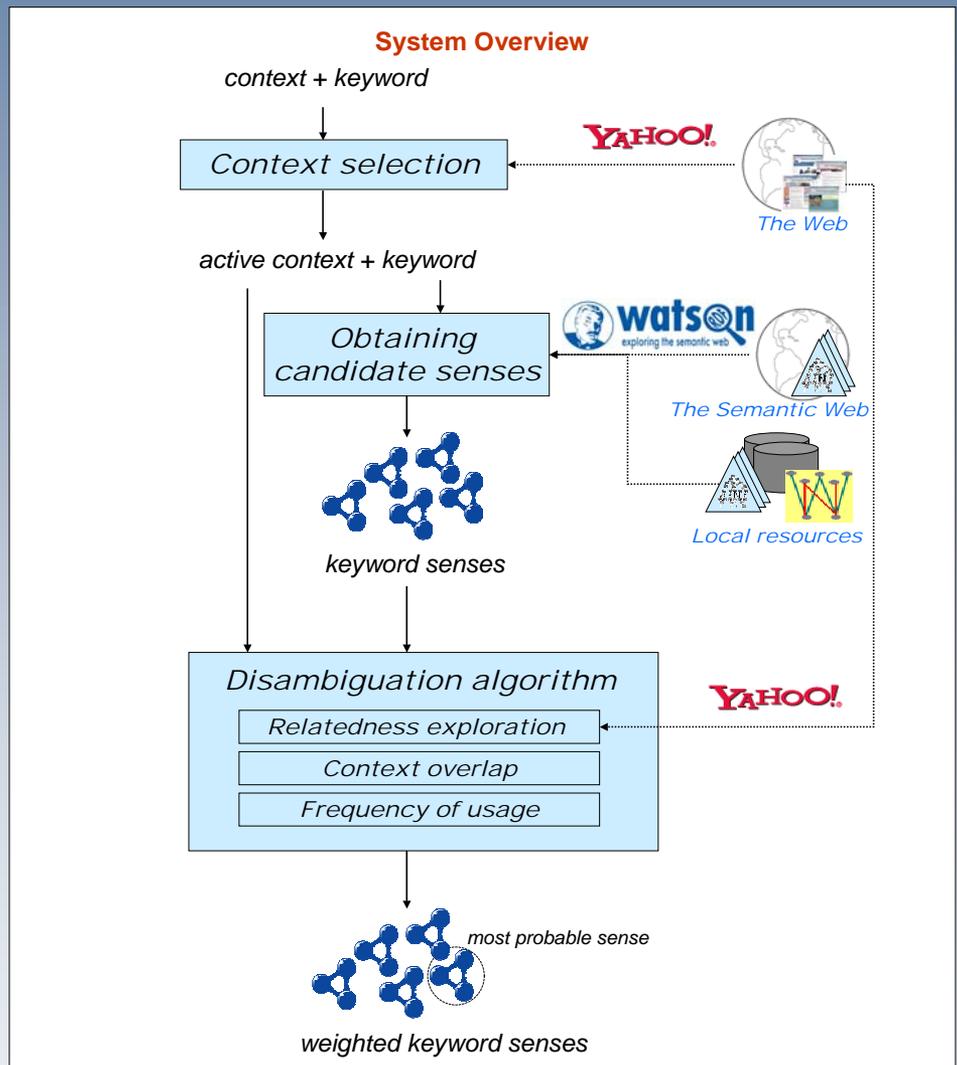
CONTEXT (tags in Flickr):
bank building skyscraper finance city curtain wall modern glass Birmingham alabama 35205 2005 10-25-fav most favorited geo:lat=33.5184 geo:lon=-86.8065 geotagged

Context Selection

ACTIVE CONTEXT: city finance building wall

Disambiguation algorithm

SELECTED SENSE: [WordNet#4] "a building in wich Commercial Banking is transacted..."



Contact

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

References

- [1] J. Gracia and E. Mena. Web-based measure of semantic relatedness. In Proc. of WISE'08, Auckland, New Zealand, pages 136-150. LNCS, September 2008.
- [2] J. Gracia, R. Trillo, M. Espinoza, and E. Mena. Querying the web: A multiontology disambiguation method. In Proc. of ICWE'06, Palo Alto, California, USA. ACM, July 2006.

Acknowledgements

This work was supported by the CICYT project TIN2007-68091-C02-02.

Experimental Results

Disambiguation of 350 polysemic keywords found in Flickr pictures, using their tags as context.

	Disambiguation	MFS baseline	Random baseline
Accuracy	58%	43%	20%