

Evaluating Word Sense Disambiguation Tools for Information Retrieval Task

Fernando Martínez-Santiago, José M. Perea-Ortega,
and Miguel A. García-Cumbreras

SINAI Research Group*, Computer Science Department, University of Jaén, Spain
{dofer,jmperea,magc}@ujaen.es

Abstract. The main interest of this paper is the characterization of queries where WSD is a useful tool. That is, which issues must be fulfilled by a query in order to apply an state-of-art WSD tool? In addition, we have evaluated several approaches in order to apply WSD. We have used several types of indices. Thus, we have generated 13 indices and we have carried out 39 different experiments, obtaining that some indices based on WSD tools even outperforms slightly the non disambiguated baseline case. After the interpretation of our experiments, we think that only queries with terms very polysemous and very high IDF value are improved by using WSD.

1 Introduction

Nowadays, the information unit managed by most IR models is the word. A theoretical good idea is the elaboration of IR systems based on concepts better than words or the lemmas of those words. We define a concept as a lexicographic-independent representation of an idea or object. Given a language, it does not care the vocabulary available in order to represent such a concept. Thus, a concept-based IR system translates words into concepts. State-of-art WSD tools obtain about 60% of precision/recall [1] [2] for “*fine-grained all words*” task¹. Is this enough to improve an IR system? For which topics an improvement can be achieved and which topics performance deteriorates? Which features might be good predictors for improvement by WSD? After the interpretation of our experiments, we think that only queries with terms very polysemous and very high IDF value are improved by using WSD.

2 Experimental Framework

In the experiments carried out in this paper we have used two disambiguated collections provided by the NUS [1] and UBC [2] teams, and the default collection

* <http://sinai.ujaen.es>

¹ Fine-grained all words is the name of a usual WSD task. In this paper, we have used WSD in a very similar way.

others do not. This requires a more detailed study based on the confidence of the WSD system, the semantic weight of the term, and other factors that we are investigating nowadays.

Finally, we have reported a set of experiments: we have created indices based on the best sense per term, two first senses per term, term+sense, NUS-best sense+UBC+best sense and the only experiment that outperforms the base line is the one based on NUS best sense index.

Acknowledgments

This work has been supported by a grant from the Spanish Government, project TIMOM (TIN2006-15265-C06-03), and the RFC/PP2006/Id.514 granted by the University of Jaén.

References

1. Cai, J.F., Lee, W.S., Teh, Y.W.: Nus-ml:improving word sense disambiguation using topic features. In: Proceedings of the Fourth International Workshop on Semantic Evaluations (SemEval 2007), Prague, Czech Republic, pp. 249–252 (2007)
2. Agirre, E., de Lacalle, O.L.: Ubc-alm: Combining k-nn with svd for wsd. In: Proceedings of the Fourth International Workshop on Semantic Evaluations (SemEval 2007), Prague, Czech Republic, pp. 342–345 (2007)
3. Agirre, E., Nunzio, G.M.D., Ferro, N., Mandl, T., Peters, C.: CLEF 2008: Ad Hoc Track Overview. In: Peters, C., et al. (eds.) CLEF 2008. LNCS, vol. 5706, pp. 15–37. Springer, Heidelberg (2008)