Indexing with WordNet Synonyms
May Improve Retrieval Results

Davide Buscaldi and Paolo Rosso

Natural Language Engineering Lab.,
ELiRF Research Group,
Dpto. de Sistemas Informáticos y Computación (DSIC),
Universidad Politécnica de Valencia, Spain
{dbuscaldi,prosso}@dsic.upv.es

Abstract. This paper describes a method developed for the Robust
- Word Sense Disambiguation task at CLEF 2009. In our approach, a
WordNet expanded index is generated from the disambiguated document
collection. This index contains synonyms, hypernyms and holonyms of
the disambiguated words contained in documents. Query words are in-
tegrated by terms extracted by means of a pseudo relevance feedback
technique. The set of terms made of query words and terms resulting
from pseudo relevance feedback are searched for in both the expanded
WordNet index and the default index. The results show that the use of
the extended index did not prove useful, obtaining $14-16\%$ less in MAP
with respect to the base system. However, for some queries, expanding
index terms with synonyms resulted particularly useful.

1   Introduction

The use of WordNet senses to improve the precision of Information Retrieval (IR)
systems is probably one of the holy grails of modern research in IR. Since 1993,
starting with the work of Ellen Voorhees [8], many researchers attempted to use
effectively WordNet in IR, sometimes with good results [7], in other cases with-
out success [3]. The Robust-WSD task introduced in CLEF 2009 represents an
interesting attempt to foster further investigation in this field. In 2008, we partic-
ipated in the QA-WSD task using an index expansion method based on WordNet
hypernyms, synonyms and holonyms, which exploited the disambiguated collec-
tion [1]. The results did not show any relevant difference between the use of
disambiguation or not, although we observed that passages returned using the
disambiguated collection and our method tended to be shorter with respect to
the base system. We took the opportunity presented by the Robust WSD Task
at CLEF 2009 to test the same method in this task. A novelty for this partici-
pation was the introduction of a naïve Pseudo Relevance Feedback[5,9] method,
consisting in the expansion of the query with the top 5 terms (according to their
tf.idf weights) resulting from the unexpanded query.

In the following section, we describe the retrieval system. In section 3 we
describe the characteristics of our submissions and discuss the obtained results.

2   Description of Retrieval System

The core of the system is a standard Lucene[1] search engine (version 2.4.1).
During the indexing phase, we create two indices: the first one (*text*) containing
all the terms of the sentence; the second one (expanded index, or *wn* index)
containing all the synonyms of the disambiguated words (we consider the sense
with the highest score to be the "right" sense). In the case of nouns and verbs,
it contains also their hypernyms. For nouns, the holonyms (if available) are also
added to the index, in a similar way to the GeoWorSE system that participated in
the 2008 GeoCLEF track [2]. For instance, let us consider the following sentence
from document `GH951115-000080`:

> Splitting the left from the Labour Party would weaken the battle for
> progressive policies inside the Labour Party.

The underlined words are those that have been disambiguated in the col-
lection. For these words we can found their synonyms and related concepts in
WordNet, as listed in Table 1.

Table 1. Expansion of the index terms of the example sentence. NA : not available
(the relationship is not defined for the Part-Of-Speech of the related word).

lemma	ass. sense	synonyms	hypernyms	holonyms
split	4	separate part	move	NA
left	1	–	position place	–
Labour Party	2	labor party	political party party	–
weaken	1	–	change alter	NA
battle	1	conflict fight engagement	military action action	war warfare
progressive	2	reformist	NA	NA
policy	2	–	argumentation logical argument line of reasoning line	–

Therefore, the *wn* index will contain the following terms: *separate, part, move,
position, place, labor party, political party, party, change, alter, conflict, fight,
engagement, war, warfare, military action, action, reformist, argumentation, log-
ical argument, line of reasoning, line.*

---

[1] `http://lucene.apache.org`

Previously to the search phase, in the default configuration, the original query is expanded with a naïve Blind Relevance Feedback (BRF) method. The *text* index is searched for question terms. The top 5 resulting documents are analysed to extract up to 5 keywords that are used to expand the original query. The keywords are selected according to their *tf.idf* weight (idf is calculated over the entire document collection). The expanded query is then submitted to the search engine to produce the final list of relevant documents.

In the WSD configuration, search is carried out in a similar way, with the difference that all nouns and adjectives are also searched for in the *wn* index.

In Table 2 we show the expansion terms obtained for topic `147-AH` : "*Oil accidents and birds*", using the two different configurations.

Table 2. Terms extracted for pseudo relevance feedback, topic 147-AH. Original query: "Oil accidents birds"

mode	term	tf.idf weight
No-WSD	gero	52.07
	pigeon	31.68
	fli	29.21
	spill	28.66
	wildlife	24.24
WSD	spill	200.60
	pipeline	174.10
	river	64.05
	arco	63.93
	fish	61.82

3   Experiments

We submitted four runs with the WSD system, two using the NUS labeled collection and two with the UBC labeled collection. For each collection, we submitted one run using only the topic title and another one using both the title and the description. As baseline, we submitted two non-WSD runs, one in the configuration "title only" and one in the configuration "title and description".

In Table 3 we show the results obtained by the two non-WSD runs and the four WSD runs.

The results show that the use of the disambiguated collection in general did not prove useful, since the base system obtained a better average MAP in all configurations. There are differences of $\sim 16\%$ in MAP between the normal and WSD runs in the title only configuration, and up to 14.21% between in TD configuration. The difference ($\sim 1\%$ in TD configuration) between the use of the NUS disambiguated collection and the UBC disambiguated collection is tiny, demonstrating that the disambiguation method is not relevant.

Table 3. Results obtained by our system at the CLEF 2009 Robust WSD track. TD: Title and Description. TO: Title Only. NUS: NUS labelled collection. UBC: UBC labelled collection. gm_AP: Geometric Mean Average Precision.

run ID	WSD	type	avg. MAP	avg. R-Prec	gm_AP
NLEL0901	n	TD	40.26%	38.72%	17.50%
NLEL0906	n	TO	33.42%	32.98%	8.75%
NLEL0902	y	TD NUS	27.14%	26.57%	6.87%
NLEL0904	y	TD UBC	26.05%	25.59%	6.42%
NLEL0903	y	TO NUS	17.48%	17.63%	1.14%
NLEL0905	y	TO UBC	17.53%	18.67%	1.24%

We analyzed some of the queries in which the standard system performed considerably better than the one which used the disambiguated collection. We find that disambiguation errors were the reason of the bad results. For instance, let us examine the results for topic $AH - 141$, "Letter Bomb for Kiesbauer": the base system obtained 100% MAP, placing the only relevant document ($GH950610 - 000164$) at the top of the list of retrieved documents, while the WordNet-based system obtained 0.4%, placing the relevant document only in the 255th position. The best matching document, according to the WordNet-based system, was $LA010894 - 0146$, titled "Viacom, Blockbuster to merge in Paramount bid". The reason of this result is that all references to "*Blockbuster*" in the document were assigned to the first sense of "blockbuster" in WordNet: "a large bomb used to demolish extensive areas". The effect, given the indexing method, was to add the hypernym "general purpose bomb" to the *wn* index, and to obtain a high relevance for this document for any query containing "bomb", since "blockbuster" is very frequent in the document (and so its hypernym).

Although the overall results and the study of this cases demonstrate that, *in general*, expanding the index with WordNet is more harmful than useful, we observed that in 49 topics (30.6% of the total) the outcome of using WordNet for indexing was to obtain better MAP than in the case of not using it. In Figure 1 we show the difference in MAP for those topics.

We focused on the topics for which the MAP increment was more important - topic $AH - 180$ "Bankruptcy of Barings" and topic $AH - 200$ "Flooding in Holland and Germany". In the first case, in the WSD run, one of the extracted expansion terms was "Leeson", the surname of the person responsible for the bankruptcy. This term is very specific to the topic and was key to obtain the improvement for the WordNet-based system, since it appears in every relevant document. Looking for terms in the *wn* index allowed to find this expansion term that the base system could not find, mainly because of the term "bankruptcy", that did not appear in any of the relevant documents.

In topic $AH - 200$, *Holland* was not used in the relevant documents. One of its synonyms, *Netherlands*, was used instead. Since this term is listed in WordNet as one of the synonyms of "Holland", the WordNet-based system was able to found it in the expanded index.
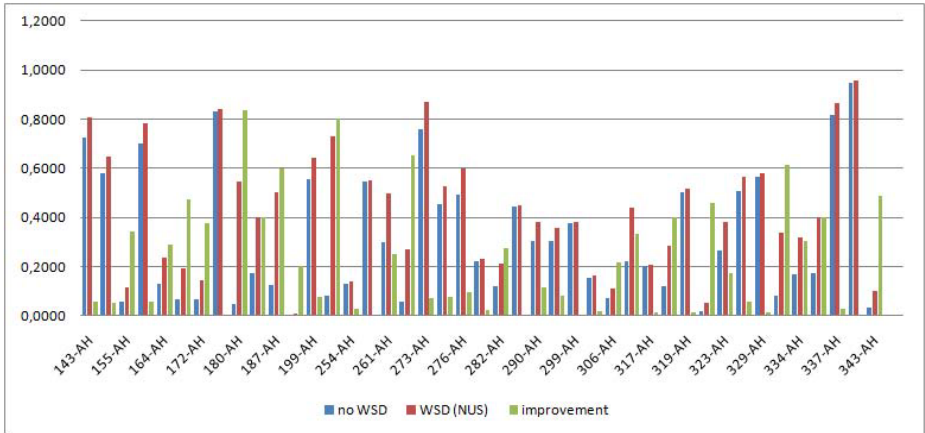
Fig. 1. Comparison of MAP values obtained for topics in which the use of the WordNet-based indexing method allowed to obtain better results than the base system

In order to understand which improvements were due to the relevance feedback method and which depended from the adopted WordNet-based expansion, we carried out some experiments with a system that did not include the BRF phase. The results of these experiments can be seen in Table 4.

Table 4. Results obtained by the system without blind relevance feedback

run ID	WSD	type	avg. MAP	avg. R-Prec	gm_AP
noBRF01	n	TD	30.56%	30.71%	10.72%
noBRF02	y	TD NUS	17.18%	18.15%	2.46%

These results shows a significant drop in precision with respect to the results obtained using BRF. In Figure 2 we show the precision increments obtained over the base system without relevance feedback (noBRF01 in Table 3) with the use of the WordNet expansion method (noBRF02 run) and the use of BRF without WordNet (this configuration corresponds to the NLEL0901 run).

As can be seen in Figure 2, in some topics the use of the WordNet-based method allowed to obtain a great MAP increase $(30 - 60\%)$ over the base system with no contribution from BRF. We examined the results obtained in such topics in order to understand how they were produced. Topic `AH-183`, "Asian dinosaur remains" resulted in a $\sim 55\%$ increase over the baseline. The reason was that some relevant documents contained references to China and Mongolia, places where some dinosaur remains were found, but no reference to Asia. Therefore, the expansion method allowed to retrieve and rank better these documents because "Asia" was added to the wn index, as an holonym of China and Mongolia. The same reason is due to the increase obtained for topic `AH-266`,
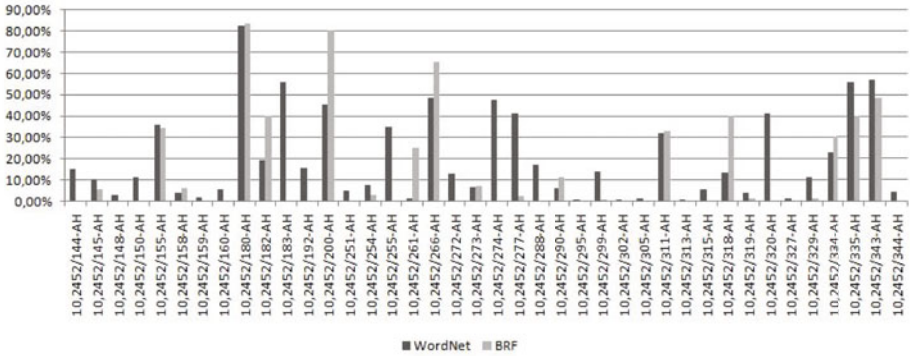
Fig. 2. MAP increases obtained with the WordNet-based indexing method and BRF, for topics in which WordNet resulted in an improvement over the base system MAP

"Discrimination against European Gypsies", where "Europe" was not included in the relevant documents, but only in their expansion. In topics `AH-255` ("Internet Junkies"), `AH-274` ("Unexploded World War II bombs") and `AH-277` ("Euthanasia by medics") the reason for the improvements was different, since the topic included both a term and one of its synonym (one in the title and the other in the description), but documents included only one of them. The synonyms were "addict" for "junky" in `AH-255`, "Second World War" for "World War II" in `AH-274` and "mercy killing" for "euthanasia".

4   Conclusions

The obtained results did not show any particular improvement depending on disambiguation accuracy. Sanderson's work [6] suggested that only high precision (more than 90%) in the disambiguation process may produce an improvement of the results in Information Retrieval. However, the collections used in the task were disambiguated with methods that, although being ones of the best systems available, were able to obtain $\sim 60\%$ in precision for the all-words task at Semeval 2007 [4]. Disambiguation errors proved to cause drops in retrieval accuracy, such in the case of topic $AH - 141$, and errors were also propagated by the retrieval method selected: adding hypernyms and synonyms from errors introduced even more errors.

However, the analysis of the results obtained for some topics showed that a synonym of a non-ambiguous term was key to obtain an improvement, proving that at least in some cases, when terms are not ambiguous, adding synonyms to the index results useful. More research should be carried out on the cases in which disambiguation proved useful, in order to discover how the information extracted from WordNet affected the retrieval process. We should also check whether adding information only for not ambiguous terms could be better than expanding automatically disambiguated terms. Finally, we will need to determine

the cases in which the improvement was indirectly derived from terms extracted using the relevance feedback and those in which the use of WordNet related information affected directly the results.

Acknowledgements

References

1. Buscaldi, D., Rosso, P.: Some experiments in question answering with a disambiguated document collection. In: Peters, C., Deselaers, T., Ferro, N., Gonzalo, J., Jones, G.J.F., Kurimo, M., Mandl, T., Peñas, A., Petras, V. (eds.) CLEF 2008. LNCS, vol. 5706, pp. 442–447. Springer, Heidelberg (2009)
2. Buscaldi, D., Rosso, P.: Using geowordnet for geographical information retrieval. In: Peters, C., Deselaers, T., Ferro, N., Gonzalo, J., Jones, G.J.F., Kurimo, M., Mandl, T., Peñas, A., Petras, V. (eds.) CLEF 2008. LNCS, vol. 5706, pp. 863–866. Springer, Heidelberg (2009)
3. Mihalcea, R., Moldovan, D.: Semantic indexing using wordnet senses. In: Proceedings of the ACL-2000 Workshop on Recent Advances in Natural Language Processing and Information retrieval, Morristown, NJ, USA, pp. 35–45. Association for Computational Linguistics (2000)
4. Pradhan, S.S., Loper, E., Dligach, D., Palmer, M.: Semeval-2007 task 17: English lexical sample, srl and all words. In: SemEval 2007: Proceedings of the 4th International Workshop on Semantic Evaluations, Morristown, NJ, USA, pp. 87–92. Association for Computational Linguistics (2007)
5. Robertson, S.E.: On term selection for query expansion. J. Doc. 46(4), 359–364 (1990)
6. Sanderson, M.: Word sense disambiguation and information retrieval. In: SIGIR 1994: Proceedings of the 17th annual international ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 142–151. Springer, New York (1994)
7. Schütze, H., Pedersen, J.O.: Information retrieval based on word senses. In: Proceedings of the 4th Annual Symposium on Document Analysis and Information Retrieval, pp. 161–175 (1995)
8. Voorhees, E.M.: Using wordnet to disambiguate word senses for text retrieval. In: SIGIR 1993: Proceedings of the 16th annual international ACM SIGIR conference on Research and development in information retrieval, pp. 171–180. ACM, New York (1993)
9. Xu, J., Bruce Croft, W.: Query expansion using local and global document analysis. In: SIGIR 1996: Proceedings of the 19th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 4–11. ACM, New York (1996)