

Some Experiments in Question Answering with a Disambiguated Document Collection

Davide Buscaldi and Paolo Rosso

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

Abstract. This paper describes our approach to the Question Answering - Word Sense Disambiguation task. This task consists in carrying out Question Answering over a disambiguated document collection. In our approach, disambiguated documents are used to improve the accuracy of the retrieval phase. In order to do this, we added a WordNet-expanded index to the document collection. The expanded index contains synonyms, hypernyms and holonyms of the words already in the documents. Question words are searched for in both the expanded WordNet index and the default index. The obtained results show that the system that exploited disambiguation obtained better precision than the non-WSD one.

1 Introduction

The evaluation of the impact of Word Sense Disambiguation (WSD) on Information Retrieval (IR) has been the object of many research efforts in the last decade [4,7,6]. One of the objectives of the QA-WSD and CLIR-WSD tasks at CLEF¹ 2008 was to attempt to find new evidence in favor of the utility of WSD in IR or not, by providing participants with disambiguated collections to perform tests on. The QA-WSD task put its focus on the Question Answering task, that can be seen as a specialized kind of IR.

The available collections were the one disambiguated using the method of the University of Basque Country (UBC) [1], and the one disambiguated by the method of the National University of Singapore (NUS) [8]. This is the first time that disambiguated collections of this size have been developed and released for a large-scale evaluation.

Our system is constituted by a modified version of the QUASAR system described in [2]. For this task the search engine (JIRS) has been replaced by Lucene², which can work with multiple indices. This capability was needed in order to put in different indices the terms extracted from the documents and the terms derived from WordNet [5]. The method we developed to take advantage of the disambiguated collection is similar to the “Index Term Expansion” method

¹ <http://www.clef-campaign.org>

² <http://lucene.apache.org>

described in [3], in which the geographical terms in documents were expanded with their WordNet holonyms. In this case we added to the WordNet index all the hypernyms, holonyms, and synonyms of the disambiguated words in the document collection.

In the following section, we describe the system. In Section 3 we discuss the experiments carried out and the obtained results. Finally, in Section 4 we draw some conclusions.

2 WordNet-based Index Expansion

Previous to the indexing phase, all documents are split into sentences. These are used later to form the passages. In the indexing phase, we create two indices: the first one (*text*) contains all the terms of the sentence; the second one (expanded index, or *wn* index) contains all the synonyms of the disambiguated words; in the case of nouns and verbs, it contains also their hypernyms. In the case of nouns, their holonyms (if present) are also added to the index. For instance, let us consider the following sentence from document GH951115-000080-03:

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 collection. For these words we can find their synonyms and related concepts in WordNet, as listed in Table 1.

Table 1. Expansion of 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, logical argument, line of reasoning, line.*

During the search phase, the *text* and *wn* indices are both searched for question terms. The top 20 sentences are returned for each question. The passages are built from these sentences, by appending them the previous and following sentences in the collection. For instance, if the above example was a retrieved sentence, the resulting passage would be composed by the sentences:

- GH951115-000080-2 : “The real question is how these policies are best defeated and how the great mass of Labour voters can be won to see the need for a socialist alternative.”
- GH951115-000080-3 : “Splitting the left from the Labour Party would weaken the battle for progressive policies inside the Labour Party.”
- GH951115-000080-4 : “It would also make it easier for Tony Blair to cut the crucial links that remain with the trade-union movement.”

Figure 1 shows the first 5 sentences returned for the question “What is the political party of Tony Blair?” using only the *text* index; in Figure 2 we show the first 5 sentences returned using the *wn* index.

The Labour Party , under Tony Blair , is poised to achieve political power for the first time in 16 years	GH950821-000120-5
No Headline Present Political peace : A truce is agreed on the Westminster front as party leaders John Major , Paddy Ashdown , and Tony Blair celebrate VJ-Day	GH950821-000164-0
Blair puts the family centre stage LABOUR leader Tony Blair , in a further move to occupy the political centre ground , yesterday staked his claim for Labour to be the ' ' party of the family	GH950330-000184-0
Blair should beware I AM mystified by your political editor ' s comment about Tony Blair ' ' dispatching the Bennite Marxist left ' ' and rescuing the Labour Party from the ' ' false historical perspective ' ' of a ' ' Marxist intellectual analysis ' ' (March 23	GH950327-000057-0
Blair wrestles with age-old links TRADE union leaders of all political persuasions have confronted Labour leader Tony Blair with a demand that the unions retain a 50 % voice in the party they created	GH951002-000214-1

Fig. 1. Top 5 sentences retrieved with the standard Lucene search engine.

The Labour Party , under Tony Blair , is poised to achieve political power for the first time in 16 years	GH950821-000120-5
The Labour Party has been set a simple test by Tony Blair	GH950310-000026-16
No Headline Present Political peace : A truce is agreed on the Westminster front as party leaders John Major , Paddy Ashdown , and Tony Blair celebrate VJ-Day	GH950821-000164-0
'The investigation is understood to have the full support of Labour Party leader Tony Blair	GH950227-000161-5
On the eve of the Labour party conference , there were also sharp words for Tony Blair	GH951002-000228-9

Fig. 2. Top 5 sentences retrieved with the WordNet extended index.

It can be noted that sentences retrieved with the expanded WordNet index are shorter, because the keyword *political* was found only in the expanded index and not in the text.

Our system had some limitations on the type of questions it could answer. The reason is that the base system was developed for the 2006 edition of CLEF QA, which included guidelines that were different from the ones adopted in CLEF 2008. In 2006, questions did not include questions with references to a previous question (anaphora). Therefore, our system cannot solve anaphoras. We refer the reader to the description in [2] for a detailed description of the base system.

3 Experiments

The participation at CLEF 2008 consisted in submitting two mandatory runs, one with the basic system (labeled as “*no WSD*” in Table 2) that does not use semantic information, and one with the system described above (*WSD-NUS* in Table 2), using as collection the NUS-disambiguated collection. Of the 200 questions in the test set, only 49 had an answer in the disambiguated collection (the other questions had their answers in Wikipedia, which was not featured for the QA-WSD track), according to the organisers. However, we manually checked the data and found that it was possible to find an answer to 25 of the Wikipedia questions, bringing the number of questions with an answer in the collection to 74.

In Table 2 we show the results obtained by the two mandatory runs and another run that used the UBC-disambiguated collection (*WSD-UBC*). The results of this last run are not official (i.e. we evaluated the run ourselves instead of the track organizers).

Table 2. Results obtained with the three runs over the 49 questions that had (officially) an answer in the collection and all questions.

run ID	49 Questions			All Questions				
	R	X	U	Accuracy	R	X	U	Accuracy
no WSD	8	0	0	16.32%	10	0	0	5.00%
WSD-NUS	7	0	0	14.29%	8	0	1	4.00%
WSD-UBC	6	0	0	12.24%	7	0	1	3.50%

The table shows that, apart from the fact that the complete question set was not suitable for the evaluation, the runs that were carried out on the disambiguated collections obtained worse results. In order to understand the reason of this, we carried out an analysis of the average number of passages that contained the answer for each of the questions. Of the 49 questions, only three answers were present in more than nine passages. The average number of passages containing the answer for each question in the remaining 46 questions is 2.04. This number justifies the small differences between the WSD based system and the base one (the systems retrieve the same sets of relevant passages, independently from the method used).

Therefore, we carried out some additional experiments with the sets of questions from CLEF 2005 and 2006, in order to check what would be the results with questions that better fit the used collections. The questions were the same of the English-Spanish bilingual test sets, but in this case we employed them in a monolingual environment, with an English target collection. In Table 3 we show the results obtained with these questions. The evaluation was carried out taking into account the 2005 and 2006 guidelines.

Table 3. Results obtained with the CLEF QA 2005 and 2006 question sets, with the base system, the WSD-based system and the UBC collection, the WSD based system and the NUS collection.

run ID	CLEF 2005 questions			CLEF 2006 questions				
	R	X	U	Accuracy	R	X	U	Accuracy
no WSD	30	6	0	15.00%	28	2	1	14.00%
WSD-NUS	36	5	0	18.00%	29	1	2	14.50%
WSD-UBC	37	5	1	18.50%	31	1	3	15.50%

From the results shown in Table 3 it can be observed that the average results are comparable to the ones obtained with the 49 questions of the CLEF 2008 test set: this confirms the fact that the whole question set included too many questions that the system could not answer. The results on the 2005 and 2006 test sets show also that with questions that present a higher redundancy of the answer the use of the disambiguated collection allowed to obtain a higher accuracy of the WSD system with respect to the non-WSD one. The answer redundancy for the 2005 collection was 29.28 answers per question, while in the 2006 collection was of 25.71 answers per question. There was no significant difference between the use of the NUS and the UBC document collections. Note that NIL questions were excluded from the computation of the results, since they were not taken into account in the evaluation at CLEF QA-WSD 2008.

4 Conclusions

The obtained results do not provide a decisive argument in favour of the utility of Word Sense Disambiguation in Information Retrieval. However, it is noteworthy that the WSD-based QA system performed better than the non-WSD one under two conditions: higher answer redundancy and the use of the disambiguated collection. We did not observe any significant difference on the smaller question set between the WordNet enhanced method and the base system. We believe that most errors were due to the poor performance of the QA system and not to the retrieval process. In the future, we will attempt to evaluate the impact of the use of the disambiguated collections only in passage retrieval, independently from the rest of the QA system.

Acknowledgements

We would like to thank the TIN2006-15265-C06-04 and the AECI-PCI A010317/07 research projects for partially supporting this work.

References

1. Eneko Agirre and Oier Lopez de Lacalle. UBC-ALM: Combining k-NN with SVD for WSD. In *Proceedings of the 4th International Workshop on Semantic Evaluations (SemEval 2007)*, pages 341–345, Prague, Czech republic, 2007.
2. Davide Buscaldi, José Manuel Gómez, Paolo Rosso, and Emilio Sanchis. N-gram vs. keyword-based passage retrieval for question answering. In Carol Peters, Paul Clough, Fredric C. Gey, Jussi Karlgren, Bernardo Magnini, Douglas W. Oard, Maarten de Rijke, and Maximilian Stempfhuber, editors, *Evaluation of Multilingual and Multi-modal Information Retrieval*, volume 4730 of *Lecture Notes in Computer Science*, pages 377–384. Springer, 2007.
3. Davide Buscaldi, Paolo Rosso, and Emilio Sanchis. A wordnet-based indexing technique for geographical information retrieval. In Carol Peters, Fredric C. Gey, Julio Gonzalo, Henning Mller, Gareth J.F. Jones, Michael Kluck, Bernardo Magnini, Maarten de Rijke, and Danilo Giampiccolo, editors, *Lecture Notes in Computer Sciences*, volume 4730 of *Lecture Notes in Computer Science*, pages 954–957. Springer, Berlin, 2007.
4. Julio Gonzalo, Felisa Verdejo, Irin Chugur, and José Cigarrán. Indexing with wordnet synsets can improve text retrieval. In *COLING/ACL '98 workshop on the Usage of WordNet for NLP*, pages 38–44, Montreal, Canada, 1998.
5. George A. Miller. Wordnet: A lexical database for english. In *Communications of the ACM*, volume 38, pages 39–41, 1995.
6. Paolo Rosso, Edgardo Ferretti, and Vicente Vidal. Text categorization and information retrieval using wordnet senses. In *2nd Global WordNet Conference (GWC 2004)*, pages 299–304, Brno, Czech Rep., 2004.
7. Mark Sanderson. *Word Sense Disambiguation and Information Retrieval*. PhD thesis, University of Glasgow, Glasgow, Scotland, UK, 1996.
8. Yee Sen Chan, Hwee Tou Ng, and Zhi Zhong. US-PT: Exploiting Parallel Texts for Word Sense Disambiguation in the English All-Words Tasks. In *Proceedings of the 4th International Workshop on Semantic Evaluations (SemEval 2007)*, pages 253–256, Prague, Czech republic, 2007.