REINA at CLEF 2009 Robust-WSD Task: Partial Use of WSD Information for Retrieval

Angel Zazo, Carlos G. Figuerola, José L. Alonso Berrocal and Raquel Gómez
REINA Research Group - University of Salamanca
C/ Francisco Vitoria 6-16, 37008 Salamanca, Spain
http://reina.usal.es

Abstract

This paper describes the participation of the REINA research group at CLEF 2009 Robust-WSD Task. We have participated in both monolingual and bilingual subtasks. In past editions of the robust task our research group obtained very good results for non-WSD experiments applying local query expansion using co-occurrence based thesauri constructed using windows of terms. We applied it again. For WSD experiments, our intention was to use the WSD information and WordNet for expansion, but we did not have time to use them. We only used the lemma proposed by the POS tagger of the WSD collection as a stemmer. In bilingual retrieval experiments, two on-line machine translation programs were used to translate topics, and translations were merged before performing a monolingual retrieval. We also applied the same local expansion technique.

Our non-WSD runs obtained the top rank considering the GMAP measure (monolingual and bilingual subtasks). However, regarding expansion we viewed that the settings tuned for a system not always produces retrieval improvement when the conditions change: number of query terms, query language, document or query subject, linguistic approach, etc.

Our WSD runs also obtained very good positions in the rankings even thought we only used partial WSD information. However, in comparison with non-WSD runs the retrieval performance made worse. We detected some homonym errors in the POS tagger and probably these errors are worse than the errors carried out by the Porter stemmer used in the non-WSD experiments.

Categories and Subject Descriptors

H.3.1 [**Content Analysis and Indexing**]: *Indexing methods, Thesauruses*; H.3.3 [**Information Search and Retrieval**]: *Query formulation*; H.3.4 [**Systems and Software**]: *Performance evaluation*; I.2.7 [**Natural Language Processing**]: *Machine Translation, Language parsing and understanding*

General Terms

Measurement, Performance, Experimentation

Keywords

Robust Retrieval, Word Sense Disambiguation, Query Expansion, CLIR

1 Introduction

Robust retrieval tries to obtain stable performance over all topics by focusing on poorly performing topics. Word sense disambiguation (WSD) is the identification process of sense of a word used in a given sentence. The goal of the CLEF 2009 robust task was to test whether WSD can be used beneficially for retrieval systems. For this, the organizers provided document collections annotated with WSD from previous CLEF campaigns. Our research group has participated in the monolingual (English) and the bilingual (Spanish to English) subtasks with non-WSD and WSD experiments. For non-WSD ones, this year we used the same approach of our group at past CLEF robust tasks: an IR system based on the vector space model and applying a query expansion technique that uses co-occurrence based thesauri built with windows of terms. For WSD experiments, our primary approach was to use the WSD information and WordNet for retrieval, but we had no time to use them. Nevertheless, we have used the information of the part-of-speech (POS) tagger as a stemmer, instead of the Porter stemmer for English used in our non-WSD experiments.

Our main focus was monolingual retrieval. The steps followed are explained below. For bilingual retrieval experiments we used machine translation (MT) programs to translate topics into document language, and then a monolingual retrieval was implemented.

2 Non-WSD Experiments

At past CLEF robust campaigns our non-WSD runs got very good rankings [1, 2], therefore we decided to use this year the same information retrieval system and the same settings for our monolingual experiments. We used the well-known vector space model, using the **dnu-ntc** term weighting scheme. For documents, letter *u* stands for the *pivoted document normalization* [3]: we adjusted *pivot* to the average document length and *slope* set to 0.1. We decided to remove the most frequent terms in each collection, those which had a document frequency of at least a quarter of the number of documents in the collection. We use the Porter stemmer for English. It should be noted that we automatically removed certain phrases from the descriptions and narratives of the topics, such as "Find documents that ..." or "Encontrar documentos sobre ..."

The last step was to apply local query expansion using windows of terms. This technique uses co-occurrence relations in windows of terms from the first retrieved documents to build a thesaurus to expand the original query. Terms close to query terms must have higher value of relation than other terms in the document. In this case, it is important to define the distance value between two terms. If distance is zero, both terms are adjacent. If distance is one, then there exists one term between the two terms, and so on. To compute the distance, stop words are removed and sentence or paragraph limits are not taken into account.

To expand the original query, terms with a high co-occurrence value with all terms of the query must be selected. We use the measurement of scalar product with all query terms to obtain the terms with highest potential to be added to the original query. A description of this procedure can be found in [4].

Taking into account that the geometric average (GMAP), rather than the mean of the average precision (MAP), turned out to be the most stable evaluation method for robustness, several tests were carried out to obtain the best performance using the training topics. The highest improvement achieved with this expansion technique was by using a distance value of 2, taking the first 5 retrieved documents, and adding 10 terms to the original query. For the rest of the experiments we used the same settings when this expansion was applied. Runs without expansion were also submitted at CLEF task.

For the bilingual experiments, the CLIR system was the same as that used in monolingual retrieval. A previous step was carried out before searching, to translate Spanish topics into English. We used two on-line machine translation (MT) programs: Systran[1] and Reverso[2]. For each topic

[1] `http://www.systransoft.com`
[2] `http://www.reverso.net`

Table 1: Results of the runs submitted at CLEF 2009 Robust-WSD Task.

RunId	Subtask	Expansion	Topic Fields	MAP	GMAP
ROB1	MONO-EN	Yes	TD	41,94	19,16
ROB2	MONO-EN	Yes	TDN	**44,52**	**21,18**
ROB3	MONO-EN	Yes	T	37,09	13,49
ROB4	MONO-EN	No	TDN	43,50	21,05
ROB5	MONO-EN	No	TD	40,66	18,69
ROBWSD1	WSD-MONO-EN	Yes	TD	37,70	15,56
ROBWSD2	WSD-MONO-EN	Yes	TDN	**41,23**	**18,38**
ROBWSD3	WSD-MONO-EN	Yes	T	34,63	10,49
ROBWSD4	WSD-MONO-EN	No	TDN	40,42	18,35
ROBWSD5	WSD-MONO-EN	No	TD	38,10	16,34
BILI1	BILI-X2EN	Yes	TD	34,37	12,22
BILI2	BILI-X2EN	Yes	TDN	**38,42**	**15,11**
BILI3	BILI-X2EN	Yes	T	28,72	5,41
BILI4	BILI-X2EN	No	TDN	37,31	14,76
BILI5	BILI-X2EN	No	TD	34,52	12,99
BILIWSD1	WSD-BILI-X2EN	Yes	TD	28,60	7,78
BILIWSD2	WSD-BILI-X2EN	Yes	TDN	**30,32**	9,38
BILIWSD3	WSD-BILI-X2EN	Yes	T	23,33	2,64
BILIWSD4	WSD-BILI-X2EN	No	TDN	29,75	**9,57**
BILIWSD5	WSD-BILI-X2EN	No	TD	28,75	7,57

we combined the terms of the translations in a single topic: this is another expansion process, although in most cases the two translations were identical. Finally, a monolingual retrieval was performed. The local query expansion using co-occurrence based thesauri built with terms windows was also applied.

3 WSD Experiments

We had no time to use all the WSD information and WordNet for retrieval. Nevertheless, we used a piece of the WSD information as a stemmer. The POS tagger proposes a "lema" for each word it analyzes. We use this lemma for indexing instead of the stem returned by the Porter stemmer used in the non-WSD experiments. We also applied the same expansion technique with the same settings used in non-WSD experiments.

4 Results

Table 1 shows MAP and GMAP measures of all runs we submitted. For the monolingual non-WSD experiments, i.e. our base experiments, the improvement obtained applying local query expansion was about 3% in MAP and 2% in GMAP respects no expansion. Our run ROB2 obtained the top rank in GMAP measure in the subtask. However, for the rest of the experiments this expansion not always produces retrieval improvement.

For the bilingual non-WSD experiments, our run BILI2 obtained the top ranks in GMAP and MAP measures in the subtask. For bilingual retrieval evaluation, a common method is to compare results against monolingual baselines. Our bilingual runs achieved values about 92% MAP and 85% GMAP of the monolingual ones. This shows that the use of on-line MT programs to translate topics is a good approach for cross-language information retrieval.

For the WSD experiments our runs ROBWSD2 (monolingual) and BILIWSD2 (bilingual) also obtained good positions in the rankings. In these cases, we only made a partial use of the WSD information from the topics and documents collections.

5 Conclusions

For monolingual retrieval we used a simple document retrieval system based on the vector space model and we applied a local query expansion technique as a basis of our runs. The use of query expansion can be used to improve retrieval, in fact this is the approach used at TREC and CLEF robust tasks, but we have verified that the settings for a system not always produces retrieval improvement when the conditions change (number of query terms, language, document or query subject, linguistic approach, etc.). The problem is that poorly performing topics behaved differently when changing the retrieval conditions. We think that regarding robustness the objective must be to make good information retrieval systems, rather than to tune some query expansion techniques.

For the bilingual retrieval, the use of on-line MT programs to translate topics is a good approach for CLIR. Collecting terms from some translations of a topic is a technique that also improves the systems performance.

For the WSD experiments we only made a partial use of the WSD information of the topics and documents collection, in spite of this our runs obtained a good positions in the subtasks. In all cases our non-WSD experiments obtained better results than WSD experiments. We think that the reason is the information we used form the POS tagger, owing on the fact that we detected some errors in it, primarily homonym errors, both in Spanish and in English. In Spanish this kind of error is sometimes introduced by the elimination of accent sings in the process. For example, in the Spanish topic "*Pesticidas en alimentos para bebés*" (Pesticides in baby food), the word "*bebés*" (babies, a noun) was tagged as verb, and the lemma proposed was "*beber*" (to drink). The explication is here: in Spanish the word "*bebes*" (note the missing accent) is the second-person form of the present tense of the verb "*beber*" (*tú bebes agua*, you drink water.)

We think that errors based on homographs are the most important ones to deteriorate the retrieval performance. It is very important that the POS tagger works fine, otherwise any process that depends on it will increase the error. Probably these errors are worse than the errors carried out by the Porter stemmer used in the non-WSD experiments.

References

[1] G. M. Di Nunzio, N. Ferro, T. Mandl, and C. Peters. CLEF 2006: Ad-hoc track overview. *Lecture Notes in Computer Science*, 4730:21–34, 2007.

[2] G. M. Di Nunzio, N. Ferro, T. Mandl, and C. Peters. CLEF 2007: Ad-hoc track overview. *Lecture Notes in Computer Science*, 5152:23–32, 2008.

[3] A. Singhal, C. Buckley, and M. Mitra. Pivoted document length normalization. In *Proceedings of the 19th Annual International ACM SIGIR Conference*, pages 21–29, 1996.

[4] A. F. Zazo, C. G. Figuerola, J. L. A. Berrocal, E. Rodríguez, and R. Gómez. Experiments in term expansion using thesauri in Spanish. *CLEF 2002, Lecture Notes in Computer Science*, 2785:301–310, 2003.