SINAI at CLEF 2005: Multi-8 Two-years-on and Multi-8 Merging-only tasks

Fernando Martínez-Santiago, Miguel A. García-Cumbreras
University of Jaén
dofer@ujaen.es, magc@ujaen.es

Abstract

This year, we have participated on *multilingual two years* on and *Multi-8 merging-only* CLEF task. Our main interest has been to test several usual CLIR tasks and investigate how they affect to the final performance of the multilingual system. Specifically, we have evaluated the information retrieval model used to obtain each monolingual result, the merging algorithm, the translation approach and the application of query expansion techniques. The obtained results show that by means of improving merging algorithms and translation resources we reach better results than improving other CLIR modules such as IR engines or the expansion of queries.

Categories and Subject Descriptors

H.3 [**Information Storage and Retrieval**]: H.3.1 Content Analysis and Indexing; H.3.3 Information Search and Retrieval; H.3.4 Systems and Software; H.3.7 Digital Libraries

General Terms

Algorithms, Languages, Performance, Experimentation

Keywords

Information Retrieval, Cross Language Information Retrieval, Collection Fusion Problem

1 Introduction

In order to evaluate the relevance of several usual CLIR modules, we have made a combination between the collection fusion algorithm 2-step RSV and several IR systems. 2-step RSV collection fusion algorithm is detailed in [4, 5]. We depict briefly this algorithm above.

1.1 The merging algorithm

In short, the basic 2-step RSV idea is straightforward: given a query term and their translations into the other languages, their document frequencies are grouped together. Therefore, the method requires recalculating the document score by changing the document frequency of each query term. Given a query term, the new document frequency will be calculated by means of the sum of the monolingual retrieved document frequency of the term and their translations. In the first step the query is translated and searched on each monolingual collection. This phase produces a T_0 vocabulary made up by "concepts". A concept consists of each term together with its corresponding translation. Moreover, we obtain a single multilingual collection D_0 of preselected documents as result of the union of the first 1000 retrieved documents for each language. The second step consists of creating a dynamic index by re-indexing the multilingual collection D_0, but considering solely the T_0 vocabulary. Finally, a new query formed by concepts in T_0 is generated and this query is carried out against this dynamic index.

Thus, the first step of 2-step RSV consists of retrieving relevant documents for each language and the alignment of the query and its translations. This year we have tested the performance of the algorithm by testing several information retrieval engines, used for retrieving relevant documents for each monolingual collection, and then applying the second step of the merging algorithm over the retrieved documents. To sum up, we have tested ZPrise with OKAPI weighting function[6], IRn passage system[2], and several relevant documents list available from Multi-8 Merging-only task.

2 Experimentation framework

The basic process is the following. In a first step each monolingual collection is preprocessed as usual (tokens extraction, stopper, stemmer). In addition, compound words are decompounded as possible to the German, Swedish, Finnish and Dutch languages. We use the decompounding algorithm depicted in [3]. The preprocessed collections have been indexed by using the passage retrieval system IRn and ZPrise. The IRn system has been modified in order to return a list of relevant documents, the documents that contain the relevant passages. Given a query and its translations into the other languages, each query is searched in the corresponding monolingual collection.

Since we have used machine translation for several languages, and 2-step RSV requires to group together the document frequency for each term and its own translations, and MT translates the whole of the phrase better than word-by-word, 2-step RSV merging algorithm is not directly feasible with MT (given a word of the original query, its translation to the rest of languages must be known). Thus, we propose in [3] an straightforward and effective algorithm in order to align the original query and its translation at term level. It aligns about 80-85% of non-empty words (Table 1).

Table 1: Percent of aligned non-empty words (CLEF2005 query set, Title+Description fields,)

Language	Translation resource	Alignment percent
Dutch	Prompt (MT)	85.4%
Finnish	FinnPlace (MDR)	100 %
French	Reverso (MT)	85.6%
German	Prompt (MT)	82.9 %
Italian	FreeTrans (MT)	83.8 %
Spanish	Reverso (MT)	81.5 %
Swedish	Babylon (MDR)	100 %

In spite of the proposed algorithm to align phrases and translations at term level works fine, it does not obtain fully aligned queries. In order to improve the system performance when some terms of the query are not aligned, we make two subqueries. The first one is made up by the aligned terms only and the other one is formed with the non-aligned terms. Thus, for each query every retrieved document obtains two scores. The first score is obtained with 2-step RSV merging algorithm over the first subquery. On the other hand, the second subquery is used in a traditional monolingual system with the respective monolingual list of documents. Therefore, we have two scores for each query, the first one is calculated by using the dynamic and global index created by 2-step RSV for all languages and the other one is calculated locally for each language. Thus we have to integrate both values. As a way to deal with partially aligned queries (i.e. queries with some terms not aligned), we have implemented several ways to combine the aligned and non-aligned score in a only score per query and retrieved document:

1. *Raw mixed 2-step RSV.* Combining the RSV value of the aligned words and not aligned words with the formula: $0.6* < RSV_aligned_doc > +0.4* < RSV_not_aligned >$

2. *Mixed 2-step RSV by using Logistic Regression.* The formula: $e^(alpha* < RSV_aligned_doc > +beta* < RSV_not_aligned >)$

3. *Mixed 2-step RSV by using Logistic Regression and local score.* The last one also uses Logistic Regression but include a new component, the ranking of the doc. It applies the formula: $e^(alpha* < RSV_aligned_doc > +beta* < RSV_not_aligned > +gamma* < ranking_doc >)$

4. *Mixed 2-step RSV by using Bayesian Logistic Regression and local score.* The last one is very similar to the previous approach, but it is based on bayesian logistic regression instead of logistic regression.

Versions two, three and four require training set (topics and their relevance assessments), and it must be available for each monolingual collection. We have used CLEF queries(140-160) relevance assessments available this year for training purposes. Thus, twenty first queries have been used as training and the other forty have been used for evaluation.

3 Expanding the queries

Some experiments based on ZPrise use pseudo-relevance feedback technique. We have adopted Robertson-Croft's approach [1] where the system expands the original query generally by 10-15 search keywords, extracted from the 10-best ranked documents. We have chosen this configuration because empirically we have obtained better results than with other configurations available at ZPrise system.

The second step of the merging method does not make use of automatic query expansion techniques such as relevance feedback (RF) or pseudo-relevance feedback (PRF) applied to monolingual queries. Since RF and PRF extend every monolingual query with collection-dependent words, the reindexing process (second step of 2-step RSV) will not take into account all of these words. Because such words are not the same for each monolingual collection, and the translation to the other languages is unknown, 2-step RSV method ignores these new terms for the second step. However, the overall performance will also improve since PRF and RF improve on monolingual experiments and usually some extended terms coincide with terms of the original query, and such terms will be aligned. Rest of expanded terms are integrated as non-aligned terms by using the approaches depicted in the section 2 for mixed 2-step RSV. Of course, the percentage of non-aligned words is increased because of the application of PRF. Table 2 shows the percentage of aligned words for expanded queries by using PRF and Machine Translation.

Table 2: Percent of aligned non-empty words (CLEF2005 query set+PRF, Title+Description fields)

Language	Alignment percent
Dutch	45.02 %
Finnish	59.97 %
French	48.11 %
German	42.23 %
Italian	44.69 %
Spanish	45.11 %
Swedish	51.2 %

4 Experiments and Results

The tables 3, 4, 5 show the official results. In order to evaluate the translation approach effect in the multilingual result, we have recovered some old experiments from CLEF 2003 for 161-200 CLEF queries(experiment ujarsv2_2003). Such experiments are based on Machine Dictionary Readable resources, and we have compared it with results of this year (experiment UJARSV2), based on Machine Translation. In order to evaluate the effect of query expansion we have developed experiments ujaprfrsv2 and UJAPRFRSV2RR. Finally, experiments UJARSV2RR, UJAUARSV2RR, UJAMENEOKRR or UJAMENEDERR use several IR systems and models to obtain the list of retrieved documents.

Table 3: Multilingual experiments (I). Experiments with capital letters are official. The "main feature" is some particularity of such experiment respect of the case base experiment. The name of the experiments: UJA[UA][PRF]RSV2[RR][_2003] means Univ. of Jaén[IRn system from Univ. of Alicante used][PRF used]2-step RSV merging algorithm[logistic regression used][CLEF 2003 results]

Experiment	Main feature	AvgP
UJARSV2	Case Base (OKAPI ZPrise IR, no PRF, MT, raw mixed 2-Step RSV)	28.78
ujaprfrsv2	UJARSV2+PRF	29.01
UJARSV2RR	different merging algorithm (see Table 4)	29.19
UJAPRFRSV2RR	UJARSV2RR+PRF	29.57
ujarsv2_2003	it uses MDR instead of MT	24.18
ujauarsv2	it uses IRn IR engine	28.81
UJAUARSV2RR	it uses IRn IR engine and a different merging algorithm	29.18

Table 4: Merging approaches. Experiments with capital letters are official.

Experiment	2-step RSV approach
UJARSV2	Raw mixed 2-step RSV
ujaprfrsv2	Raw mixed 2-step RSV
UJARSV2RR	Mixed 2-step RSV by using Logistic Regression and local score
UJAPRFRSV2RR	Mixed 2-step RSV by using Logistic Regression and local score
ujarsv2_2003	2-step RSV
ujauarsv2	Raw mixed 2-step RSV
UJAUARSV2RR	Mixed 2-step RSV by using Logistic Regression and local score

Table 5: Multi-8 merging-only experiments. Experiments with capital letters are official. "Documents" are several sets of relevant documents available for the task from Neuchatel Bilingual Runs from CLEF 2003 .

Experiment	Documents	Merging algorithm	AvgP
ujamenepr	Prosit	Raw mixed 2-step RSV	28.40
ujameprrr	Prosit	Mixed 2-step RSV by using Logistic Regression and local score	28.34
UJAMENEOK	Okapi	Raw mixed 2-step RSV	28.87
UJAMENEOKRR	Okapi	Mixed 2-step RSV by using Logistic Regression and local score	28.87
UJAMENEDF	DataFusion	Raw mixed 2-step RSV	29.42
UJAMENEDFRR	DataFusion	Mixed 2-step RSV by using Logistic Regression and local score	30.37

Table 6: Some bilingual results (except English which is a monolingual experiment).

Language	UJARSV2	ujaprfrsv2	UJAUARSV2RR	UJAMENEOKRR	UJAMENEDFRR
Dutch	30.94	38.71	34.03	35.15	44.94
English	52.06	50.73	50.96	50.29	55.71
Finnish	34.11	31.01	33.47	14.27	22.26
French	42.14	39.90	42.84	50.26	55.29
German	33.01	37.03	33.99	41.09	52.89
Italian	33.38	34.98	34.82	44.87	53.53
Spanish	37.35	40.63	39.68	43.73	51.07
Swedish	23.29	24.99	25.23	31.29	47.28

This table shows some interesting results:

- Note that the improvement for this year is considerable respect of 2003, mainly because of a best translation strategy.
- In spite of the very different performance of the bilingual experiments (Table 6), final multilingual average precision is very similar independently of the selected documents for each IR system.
- Since the simultaneous application of PRF and Machine Translation decreases dramatically the percentage of aligned words, the application of PRF improves very slightly the final result.
- Good performance of raw-mixed 2-step RSV, obtaining a result very near to the result reached by means of logistic regression and neural networks. This result is counterintuitive since the method adds two values which are not directly comparable: the score obtained by both aligned and non-aligned terms. Some of the reasons for this good result are:
  - α parameter limits the weight of the unaligned factor.
  - Not all the terms to be added to the original query are new terms since some terms obtained by means of pseudo-relevance feedback are in the initial query. Thus, these terms are aligned terms. In the same way this explains the good performance of 2-step RSV original method with expanded queries.
  - Only 20 queries available for training.
  - CLEF document collections are highly comparable (news stories from the same period). The results might be different if collections have vastly different sizes and/or topics.

Thus, 2-step RSV reaches the same precision in spite of using different IR systems. This is a drawback if the IR system used for the first step implements a IR model more sophisticated than the IR model implemented for the second step of the algorithm. In such situation, the improvement is not fully exploited by 2-step RSV merging algorithm because 2-step RSV creates a dynamic index based on classic document retrieval models (more precisely the dynamic index is created by using a document-based OKAPI weighting schema). So, what should we do to improve these results?. Since the second step is basically an OKAPI IR engine, we could improve such engine by using better IR models, and improving the translation and alignment processes.

5 Conclusions

In this work, we have tested the merging algorithm 2-step RSV in several ways. We have compared the CLEF 2003 and CLEF 2005 Multi-8 results, by using CLEF 160-200 queries. This year we have obtained better results than the 2003 edition. We think that the main reason is a better translation approach and a more refined version of the merging algorithm.

The obtained results show that the improvement of merging algorithms and translation resources are higher than the improvement obtained by expanding the query by means of pseudo-relevance feedback.

In the same way, the improvement of the monolingual IR System used to retrieve each monolingual list of documents obtains very slightly better results in the final multilingual system. In order to evaluate the impact of the monolingual IR system, we have evaluated several lists of retrieved documents by using two IR systems and some of the retrieved documents available for the Multi-8 Merging-only task, but holding the same translation approach and merging algorithm. Results show that the precision is very similar independently of the monolingual IR engine. We conclude that improvements in the selection of documents by using some monolingual IR engine is not fully exploited by 2-step RSV merging algorithm since this algorithm creates a dynamic index based on classic document retrieval models.

When pseudo-relevance feedback and machine translation is applied in the same experiment, the percentage of aligned words is too low to apply optimally some mixed variant of 2-step RSV. Thus, a more effective word alignment algorithm must be developed, especially for the new terms added to the query by means of PRF.

Finally, we think that the overall performance of the CLIR system will be overcome if we develop better translation strategies and we improve the IR model used for the creation of the dynamic index for the second step of the algorithm.

6 Acknowledgments

Thank you very much to Fernando LLopis and Elisa Noguera from University of Alicante for their support with the IRn package.

Thank you very much to Jacques Savoy and Gareth Jones to supply the results of several bilingual experiments for the Multi-8 Merging-only task and the administration of the web place.

This work has been supported by Spanish Government (MCYT) with grant TIC2003-07158-C04-04.

References

[1] D. K. Harman. Relevance feedback revisited. In Nicholas J. Belkin, Peter Ingwersen, and Annelise Mark Pejtersen, editors, *Proceedings of the 15th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR-92)*, pages 1–10. ACM, 1992.

[2] F. LLopis. University of Alicante at CLEF 2002:. In Carol Peters, Martin Braschler, Julio Gonzalo, and Michael Kluck, editors, *Advances in Cross-Language Information Retrieval, Third Workshop of the Cross-Language Evaluation Forum, CLEF 2002. Rome, Italy, September 19-20, 2002. Revised Papers*, volume 2785 of *Lecture Notes in Computer Science*, pages 103–110, 2003.

[3] F. Martínez-Santiago, Miguel García-Cumbreras, and L.A. Ureña. SINAI at CLEF 2004: Using Machine Translation Resources with Mixed 2-Step RSV Merging Algorithm. *Advances in Cross-Language Information Retrieval. Lecture Notes in Computer Science. Springer Verlag. In press.*, 2005.

[4] F. Martínez-Santiago, M. Martín, and L.A. Ureña. SINAI at CLEF 2002: Experiments with merging strategies. In Carol Peters, Martin Braschler, Julio Gonzalo, and Michael Kluck, editors, *Advances in Cross-Language Information Retrieval, Third Workshop of the Cross-Language Evaluation Forum, CLEF 2002. Rome, Italy, September 19-20, 2002. Revised Papers*, volume 2785 of *Lecture Notes in Computer Science*, pages 103–110, 2003.

[5] F. Martínez-Santiago, L.A. Ureña, and M. Martín. A merging strategy proposal: two step retrieval status value method. *Information Retrieval. In press*, 2005.

[6] S.E. Robertson and S.Walker. Okapi-Keenbow at TREC-8. In *Proceedings of the 8th Text Retrieval Conference TREC-8, NIST Special Publication 500-246*, pages 151–162, 1999.