SINAI System: Combining IR Systems at ImageCLEFPhoto 2007

M.Á. García-Cumbreras, M.C. Díaz-Galiano, M.T. Martín-Valdivia, A. Montejo-Ráez, and L.A. Ureña-López

SINAI Research Group, Computer Science Department, University of Jaén, Spain
{magc,mcdiaz,maite,amontejo,laurena}@ujaen.es

Abstract. This paper describes the SINAI team participation in the ImageCLEFPhoto 2007 campaign. This year we have developed a system that combines the document lists retrieved by two Information Retrieval systems (Lemur and JIRS). Online machine translators have been used for the bilingual experiments. The results obtained show that if we only use title text our system works bad. Because of the low MAP values fusion method does not improve the results.

1 Introduction

This is the third participation of the SINAI research group at the ImageCLEF campaign [1]. We have participated in the AdHoc task [2] and int the medical task.

The AdHoc task involves retrieving relevant images using the text associated to each image query. As a cross-language retrieval task, multilingual image retrieval based on query translation can achieve higher performance than monolingual retrieval.

This year, a new Information Retrieval (IR) module has been tested. This module works with two different IR systems and the final relevant list is the result of the combination of both IR lists (voting system). The Machine Translation Module developed last year has been updated and used for the bilingual task. English, Spanish, French, Italian and Portuguese are the languages used this year.

Given a multilingual query, the goal of the Image CLEF Photographic task is to find as many relevant images as possible from an image collection.

The proposal is to compare results with and without pseudo-relevant feedback (PRF), with or without query expansion, using different methods of query translation or using different retrieval models and weighting functions.

The following sections describe the SINAI system, and our experiments are detailed. Finally, conclusions and further work are presented.

2 System Description

2.1 Collection Preprocessing

The dataset used is the IAPR collection. The IAPR TC-12 image collection consists of 20,000 images taken from different locations around the world and

C. Peters et al. (Eds.): CLEF 2007, LNCS 5152, pp. 512–517, 2008.

comprises a varying cross-section of still natural images. It includes pictures of a range of sports and actions, photographs of people, animals, cities, landscapes and many others of contemporary life.

The collections have been preprocessed using stopwords removal and the Porter's stemmer.

The dataset has been indexed using both IR systems, namely, Lemur[1] (used past years) and JIRS [3]. Java Information Retrieval System (JIRS) is a Passage Retrieval system oriented to Question Answering tasks although it can be applied as IR system.

One parameter for each experiment is the weighting function, such as Okapi or $TF \cdot IDF$. Another is the use or not of pseudo-relevance feedback (PRF).

2.2 Queries Processing

Given a set of multilingual queries, in the bilingual subtasks, the first step is its translation into English.

As translation module we have used SINTRAM (SINai TRAnslation Module), our Meta Machine Translation system that uses some online Machine Translators for each language pair, and implements some heuristics to combine the different translations [4].

We have made previous experiments using the same translation module. The best result for each language is obtained by the following translators:

- Systran for French, Italian and Portuguese queries
- Prompt for Spanish queries

Then, the original and translated English queries have been preprocessed, as usual (stopper and stemmer), and run against the IR index.

2.3 Experiments Description

In our experiments we have used English queries (monolingual) and the four following bilingual: French, Italian, Portuguese and Spanish.

Our system combines lists of relevant documents returned by Lemur and JIRS IR systems.

A simple fusion method has been implemented to obtain a simple list of relevant documents. In a first step, both lists are normalized between 0 and 1. Then, some heuristics are applied:

- *Weighting each list.* Some experiments are based on a weighting function that gives a percentage of relevance to the Lemur list another and to the JIRS list. The final score of each relevant document is calculated by the sum of each score multiplied by its weight. Finally, the documents are sorted by their final fusion score.

[1] http://www.lemurproject.org

– *Using a threshold*. Another heuristic filters relevant documents by a threshold value. If the score of a document is worse than this parameter then it is not included in the final list. This final list is rankid again by the score of the documents.

Figure 1 describes the architecture of our system. Each query is translated and run against the Lemur and the JIRS Information Retrieval systems. Then, several fusion methods are applied to combine both relevant documents lists.

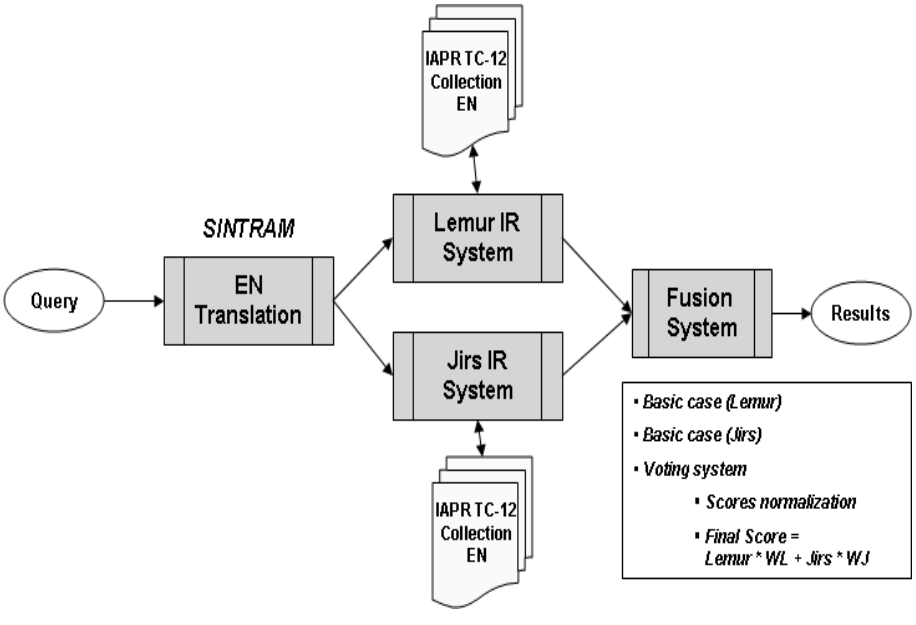

Fig. 1. Architecture of the SINAI system applied to ImageCLEFPhoto 2007

Using the AdHoc framework of ImageCLEFPhoto 2006 [5], all the described heuristics have been evaluated, in order to obtain the best configuration of parameters.

1. The Lemur baseline uses English queries and Lemur as IR system. Several weighting functions and the use or not of pseudo-relevance feedback (PRF) have been tested. The best result was obtained using Okapi as weighting function with PRF. It obtains a MAP value of 0.1672
2. The JIRS baseline uses English queries, JIRS as IR system and Okapi with PRF as weighting function. It obtains a MAP value of 0.1513
3. In the other experiments, the score of the Lemur subsystem and the JIRS one are weighted, between 0.0 and 1.0. For instance, the experiment that weights both lists in the same percentage applies the formula: $0.5 \cdot W_{lemur} + 0.5 \cdot W_{jirs}$ The best result was 0.1678 (MAP), using a weight of 0.6 for Lemur and 0.4 for JIRS.

4. To apply the second heuristic, different values, from 0.1 to 0.9, are tested as threshold. The best result was 0.1524 (MAP), obtained with a threshold=0.1.

3 Results and Discussion

We have accomplished 15 experiments: five experiments using Lemur, five using JIRS and five with the fusion of both lists.

The results obtained with each IR system (using only text, Okapi as weighting method and without expansion) and the best MAP achieved by CLEF participants for each language is shown in Table 1.

Table 1. Summary of results for the photo task: Monolingual and bilingual runs with Lemur and JIRS IR systems

Language	Experiment	IR	MAP	Best MAP
English	EN-EN-Exp2	Lemur	0.1591	0.2075
English	EN-EN-Exp1	JIRS	0.1473	0.2075
Spanish	ES-EN-Exp9	JIRS	0.1555	0.1558
Spanish	ES-EN-Exp10	Lemur	0.1498	0.1558
Portuguese	PO-EN-Exp8	Lemur	0.1490	0.1490
Portuguese	PO-EN-Exp7	JIRS	0.1350	0.1490
French	FR-EN-Exp4	Lemur	0.1264	0.1362
French	FR-EN-Exp3	JIRS	0.1195	0.1362
Italian	IT-EN-Exp5	JIRS	0.1231	0.1341
Italian	IT-EN-Exp6	Lemur	0.1198	0.1341

The results obtained with both IR systems, compared with other participants with the same configuration, are good. Only the English runs have obtained a loss of MAP of around 25%. Our best Spanish result is similar to the best one obtained. For Portuguese we have obtained the best one, and for French and Italian our results are a bit worse: only a loss of MAP of around 8%.

From these results we can conclude that the Lemur IR system works better than JIRS, but the difference is not very significant.

The results in terms of MAP are low. The experiments accomplished last year, with the same collection and same queries, gave us better results. The Table 2 show, for each language, the best result obtained in 2006, the best one obtained in 2007 and the loss of MAP obtained in 2007 (in percentage).

In 2006 title and narrative were used. In 2007 only title. All results are obtained applying query expansion and the Okapi weighting function.

After a complete analysis, the first conclusion is that if we only use the title of the query (very few words) instead of title and description, MAP results are decreased notably.

English monolingual queries obtain a loss of precision of 28%. Italian and Portuguese queries obtained better results with the new model. Spanish and French queries obtained a loss of precision around 15%.

Table 2. Comparison of results for the monolingual and bilingual runs obtained in 2006 and 2007. Last column shows the loss of MAP.

Language	MAP-2006	MAP-2007	Loss of MAP(%)
Monolingual En	0.2234	0.1591	28%
Bilingual FrEn(French)	0.1617	0.1362	15.76%
Bilingual ItEn(Italian)	0.1216	0.1341	+10.27%
Bilingual PtEn(Portuguese)	0.0728	0.1490	+104.67%
Bilingual EsEn(Spanish)	0.1849	0.1558	15.73%

The new model has worked well with the bilingual runs, but the monolingual one has decreased its results. The main reason is that the new version used of the Lemur IR system works bad than the previous one, using the same configuration.

Finally, the results obtained by applying the fusion method and the best MAP for each language is shown in Table 3.

Table 3. Summary of results for the photo task: Monolingual and bilingual runs with lists fusion

Language	Experiment	IR	MAP	Best MAP
English	EN-EN-Exp11	Fusion	0.0786	0.2075
Spanish	ES-EN-Exp15	Fusion	0.0559	0.1558
Portuguese	PO-EN-Exp14	Fusion	0.0423	0.1490
French	FR-EN-Exp12	Fusion	0.0323	0.1362
Italian	IT-EN-Exp13	Fusion	0.0492	0.1341

Fusion results have not improved the single ones. Lower MAP values decreased when we combine relevant lists. Other techniques must be used when the queries have few words.

4 Conclusions and Further Work

We have presented a system that combines document lists retrieved by two IR systems (Lemur and JIRS), and uses online translators for the bilingual experiments.

The results obtained have a low MAP, because only the title was used. Because of the low MAP values fusion method obtained poor results.

As future work, it could be interesting to develop a new robust fusion module in order to improve MAP values, and to apply a query expansion module based on Google[6], tasks on which we are already working.

Acknowledgements

This work has been partially 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] Di Nunzio, G.M., Ferro, N., Mandl, T., Peters, C.: CLEF 2007 Ad Hoc Track Overview. In: Peters, C., et al. (eds.) CLEF 2007. LNCS, vol. 5152. Springer, Heidelberg (2007)

[2] Grubinger, M., Clough, P., Hanbury, A., Müller, H.: Overview of the ImageCLEF 2007 Photographic Retrieval Task. In: Working Notes of the 2007 CLEF Workshop, Budapest, Hungary (2007)

[3] Gómez-Soriano, J.M., Montes-y-Gómez, M., Sanchis-Arnal, E., Rosso, P.: A Passage Retrieval System for Multilingual Question Answering. In: Matoušek, V., Mautner, P., Pavelka, T. (eds.) TSD 2005. LNCS (LNAI), vol. 3658, pp. 443–450. Springer, Heidelberg (2005)

[4] García-Cumbreras, M.A., Ureña-López, L.A., Martínez-Santiago, F., Perea-Ortega, J.M.: BRUJA System. The University of Jaén at the Spanish task of QA@CLEF 2006. In: Working Notes of the 2006 CLEF Workshop (2006)

[5] Clough, P., Grubinger, M., Deselaers, T., Hanbury, A., Müller, H.: Overview of the ImageCLEF 2006 Photographic Retrieval and Object Annotation Tasks. In: Working Notes of the 2006 CLEF Workshop (2006)

[6] Martínez-Santiago, F., Montejo-Ráez, A., García-Cumbreras, M.A.: SINAI at CLEF Ad-Hoc Robust Track 2007: applying Google search engine for robust cross-lingual retrieval. In: Working Notes of the 2007 CLEF Workshop (2007)