Using Translation Heuristics to Improve a Multimodal and Multilingual Information Retrieval System

Miguel Ángel García-Cumbreras, Maria Teresa Martín-Valdivia, Luis Alfonso Ureña-López, Manuel Carlos Díaz-Galiano, and Arturo Montejo-Ráez

University of Jaén. Departamento de Informática
Grupo Sistemas Inteligentes de Acceso a la Información
Campus Las Lagunillas, Ed. A3, E-23071, Jaén, Spain
{magc,maite,laurena,mcdiaz,amontejo}@ujaen.es

Abstract. Nowadays, the multimodal nature of the World Wide Web is an evidence. Web sites which include video files, pictures, music and text have become widespread. Furthermore, multimodal collections in several languages demand to apply multilingual information retrieval strategies. This paper describes a new retrieval technique applied on a multimodal and multilingual system that have been tested on two different multilingual image collections. The system applies several machine translators and implements some novel heuristics. These heuristics explore a variety of ways to combine the translations obtained from the given set of translators, and the configuration of the retrieval model by using different weighting functions, and also studying the effect of *pseudo-relevance feedback* (PRF) on this domain. Our results show interesting effects by these variations, allowing the determination of the parameters for the best retrieval model on this data and reporting the loss in performance on each language.

1 Introduction

Human communication is intrinsically multimodal, exchanging information on several media: text, audio, image or video. With the advances of technology, current communication systems are become increasingly multimodal. Content-Based Multimedia Information Retrieval (CBMIR) provides new paradigms and methods for searching on this multimodal information. Furthermore, multimodal collections that include several languages are creating the necessity of applying multilingual information retrieval strategies. The retrieval of information on these systems involves the use of the text associated to any piece of information, no matter its format, in order to provide a response to a textual query.

As a cross-language retrieval task, a multilingual image retrieval based on query translation can achieve high performance. The ad-hoc task involves the

F. Masulli, S. Mitra, and G. Pasi (Eds.): WILF 2007, LNAI 4578, pp. 438–446, 2007.

retrieval of relevant images using the text associated to each image query. This paper describes a new retrieval technique applied on a multimodal and multilingual system that works with several translators combined using different heuristics.

Given a multilingual query, the main goal is to find as many relevant images as possible from a given image collection. In our experiments, we have used the collections supplied by the CLEF[1] organization in order to accomplish the ImageCLEF[2] task. The CLEF (Cross-Language Evaluation Forum) campaign is an international meeting whose purpose is to organize a competition to evaluate different multilingual systems. Furthermore, the ImageCLEF task (the cross-language image retrieval track) includes a multimodal collection and runs as part of CLEF. Specifically, we have used two different image collections: the St. Andrews collection of historic photographs and the IAPR TC-12 image collection.

Documents in the collection are in English but the textual queries include several languages: nine different languages for the St. Andrews collection (English, Dutch, Italian, Spanish, French, German, Danish, Swedish and Russian) and seven languages for the IARP collection (Dutch, English, French, German, Italian, Portuguese and Spanish). The collections have been preprocessed using stop-words removal and the Porter's stemmer algorithm for suffix stripping. Certain tags have been selected as relevant to this task. The collections have been indexed using the LEMUR[3] Information Retrieval (IR) system. We have translated all queries into English before passing them to the IR system.

We have also developed a new translation module which combines a set of Machine Translators following some heuristics. These heuristics are, for instance, the use of the translation made by the translator by default, a combination with the translations of every translator, or a combination of the words with a higher punctuation (i.e. those words appearing in all translations get one point, and two points for those appearing in the default translation).

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

The paper is organized as follows. First, we present our framework and we introduce briefly the image collections used. Then, we describe the experiments accomplished and we show the results obtained. Finally, we present our conclusions and plans for future research.

2 Experimentation Framework

Our system, applied on a multimodal and multilingual environment, works with several translators combined using different heuristics.

[1] http://www.clef-campaign.org/
[2] http://ir.shef.ac.uk/imageclef/
[3] Available at `http://www.lemurproject.org/`

2.1 Collections Description

We have used two different image collections to accomplish the experiment. The first one is the St. Andrews collection of historic photographs. The St. Andrews collection contains 28,133 historic photographs provided by St. Andrews University Library[4][1], which holds one of the most important collections of historic photography in Scotland. Most of the photographs are in black and white and were taken by Scottish photographers or Scottish photographer companies. All of images in St. Andrews collection have associated textual captions written in British English. The captions consist of 8 fields including title, photographer, location, date and one or more pre-defined categories (all manually assigned by domain experts). Figure 1a shows a sample image with its corresponding annotation.

Short title: Rev Willian Swan.
Long title: Rev William Swan.
Location: fife, Scotland
Description: Seated, 3/4 face studio portrait of a man.
Date: ca.1850
Photographer: Thomas Rodger
Categories: [ministers][identified male][dress - clerical]
Notes: ALB6-85-2 jf/ pcBIOG: Rev William Swan ( ) ADD: Former owners of album: A Govan then J J? Lowson. Individuals and other subjects indicative of St Andrews provenance. By T. R. as identified by Karen A. Johnstone " Thomas Rodger 1832-1883. A biography and catalogue of selected works".

(a) Sample image caption

(b) Sample topic with title: "Woman in white dress"

Fig. 1. Samples from St. Andrews collection

We have used 28 topics with this collection. Each topic consists of a short sentence or phrase describing the search request in a few words (title), and a description of what constitutes a relevant or non-relevant image for that search request (narrative). Both topic title and narratives have been translated into the following languages: German, French, Italian, Spanish (European), Spanish (Latin American), Chinese (Simplified), Chinese (Traditional) and Japanese. Translations have also been produced for the titles only and these are available in 25 languages including Russian, Croatian, Bulgarian, Hebrew and Norwegian. All translations have been provided by native speakers and verified by at least another native speaker. Figure 1b shows a sample of a topic.

Colour information, which typically plays an important role in CBIR, is ineffective due to the nature of the St. Andrews collection (historic photographs). Also unlike typical CBIR benchmarks, the images in the St. Andrews collection are very complex containing both objects in the foreground and background which prove indistinguishable to CBIR methods. There are obvious limitations

[4] http://www-library.st-andrews.ac.uk/

with the existing collection: mainly black and white images, domain-specific vocabulary used in associated captions, restricted retrieval scenario (i.e. searches for historic photographs) and experiments with limited target language (English) are only possible (i.e. we cannot test further bilingual pairs). To address these issues and widen the image collections available to ImageCLEF participants, the organizers replaced the database in 2006 by a new collection of images from a personal photographic collection with associated textual descriptions in other languages, in addition to English. This new collection is the IAPR TC-12 image collection.

The IAPR TC-12 Benchmark was created under Technical Committee 12 of the International Association of Pattern Recognition (IAPR[5]). This image collection consists of 20,000 images taken from locations around the world and comprising a varying cross-section of still natural images. It includes pictures of a range of sports, actions, photographs of people, animals, cities, landscapes and many other aspects of contemporary life. The collection contains many different images of similar visual content, but varying illumination, viewing angle and background. This makes it a challenge for the successful application of visual analysis techniques. Each image in the collection has a corresponding semi-structured caption consisting of the following seven fields: a unique *identifier*, a *title*, a free-text *description* of the semantic and visual contents of the image (what we called here *narrative*), *notes* for additional information, the *provider* of the photo, *where* the photo was taken, and *when* the photo was taken. These fields are given in English and German. Figure 2a shows a sample image with its corresponding English annotation.

In order to increase the reliability of results, a total of 60 topics was provided to participants of ImageCLEF 2006. Each original topic comprised a title (a short sentence or phrase describing the search request in a few words), and a narrative (a description of what constitutes a relevant or non-relevant image for each request). In addition, three sample images were provided with each topic in order to test relevance feedback (both manual and automatic) and query-by-example searches. The topic titles were then translated into 15 languages including German, French, Spanish, Italian, Portuguese, Dutch, Russian, Japanese, and Simplified and Traditional Chinese. All translations were provided by at least one native speaker and verified by at least one another native speaker. Figure 2b shows a sample of a topic.

2.2 Preprocessing and Translation Heuristics

For both collections, a pre-processing of documents and queries was performed using stop-words removal and the Porter's stemmer algorithm for suffix stripping [4]. Resulting documents (not queries) were indexed with the LEMUR IR System. Our main objective during these two years was to test the translation module, as detailed in experiment below.

[5] http://www.iapr.org/

(a) Sample image caption from the IAPR TC-12 collection

(b) Sample topic with three images

Fig. 2. Samples from IAPR TC-12 collection

In the ImageCLEF 2005 evaluation forum, using the St. Andrews collection, nine languages were studied. We developed a Machine Translation Module that worked with several online machine translators to translate the queries from any language to English.

We have used some Machine Translators (in brackets the translator by default for each language):

- Epals (German and Portuguese)
- Prompt (Spanish)
- Reverso (French)
- Systran (Dutch and Italian)

The merging of translators' results has already been studied in previous research works by other authors [9,7]. In our experiments, in order to combine the translations produced by former applications, we explored the following heuristics:

- The first one is the use of the translation made by the translator by default
- The second one is a combination with the translations of every translator
- The third one is a combination of the words with a higher punctuation, where this punctuation is built using some features such as the word frequency in the translations, the named entities recognized and the translator used.

After some testing, the best heuristic was the second one and the most suitable translators were found to be Systran (for Dutch, French, German, Italian, Russian and Swedish) and Prompt (for both European and Latinoamerican Spanish).

The search space for the best set of parameters covers the use of the query title or title+narrative as final query, the weighting function to be used, such as Okapi or TFIDF, and the convenience of use of pseudo-relevance feedback (PRF). Table 1 shows a summary of experiments submitted and results obtained for all these languages [5]. The results obtained showed that in general the use of query expansion improves the results, as is displayed in Figure 3. Only one Italian experiment without query expansion got a better result. In the case of the use of only title or title + narrative, the results are not conclusive, but the use of only title seems to get better results (see figure 4 for a graphical view of this fact).

Table 1. Summary of results for the English monolingual adhoc runs and the best ones for other bilingual runs (with the Okapi weighting scheme) on the St. Andrews collection

Language	Initial Query	Expansion	MAP	% over mono.	Rank
Monolingual En(English)	title + narr	with	**0.3727**	n/a	31/70
Monolingual En(English)	title	without	0.3207	n/a	44/70
Monolingual En(English)	title	with	0.3168	n/a	45/70
Monolingual En(English)	title + narr	without	0.3135	n/a	46/70
Bilingual DeEn(German)	title	with	0.3004	58.8%	4/29
Bilingual DuEn(Dutch)	title	with	0.3397	66.5%	2/15
Bilingual FrEn(French)	title + narr	with	0.2864	56.1%	1/17
Bilingual ItEn(Italian)	title	without	0.1805	35.3%	12/19
Bilingual RuEn(Russian)	title	with	0.2229	43.6%	11/15
Bilingual SpEn(Spanish Eur)	title	with	0.2416	47.3%	5/33
Bilingual SpEn(Spanish Lat)	title	with	0.2967	58.1%	8/31
Bilingual SwEn(Swedish)	title	without	0.2074	40.6%	2/7

In 2006 the collection used was the IAPR TC-12 collection, as pointed out previously. This time a new translation module was developed, which combines the following Machine Translators and heuristics: Epals (German and Portuguese), Prompt (Spanish), Reverso (French) and Systran (Dutch and Italian). Heuristics applied are, for instance, the use of the translation made by the translator by default, a combination with the translations of every translator, or a combination of the words with a higher punctuation (scoring two points if it appears in the default translation, and one point if it appears in all of the other translations). To evaluate our framework we can summarize the results obtained at each participation of our group in the CLEF campaign.

The experiments performed over the IAPR TC-12 collection for the ImageCLEF 2006 evaluation forum show also interesting results. This time seven languages were considered in our experiments: Dutch, English, French, German, Italian, Portuguese and Spanish. In table 2, we can see the global results, for the English monolingual run and the other bilingual runs [6]. The best machine translation heuristic was the use of the translation made by the default translator (for each language pair) plus the words that appear in two or more of the other translations, obtaining the best results with pseudo-relevance feedback and Okapi a weighting function. The results show that there is a loss of MAP between the best monolingual experiment and this bilingual experiment, namely

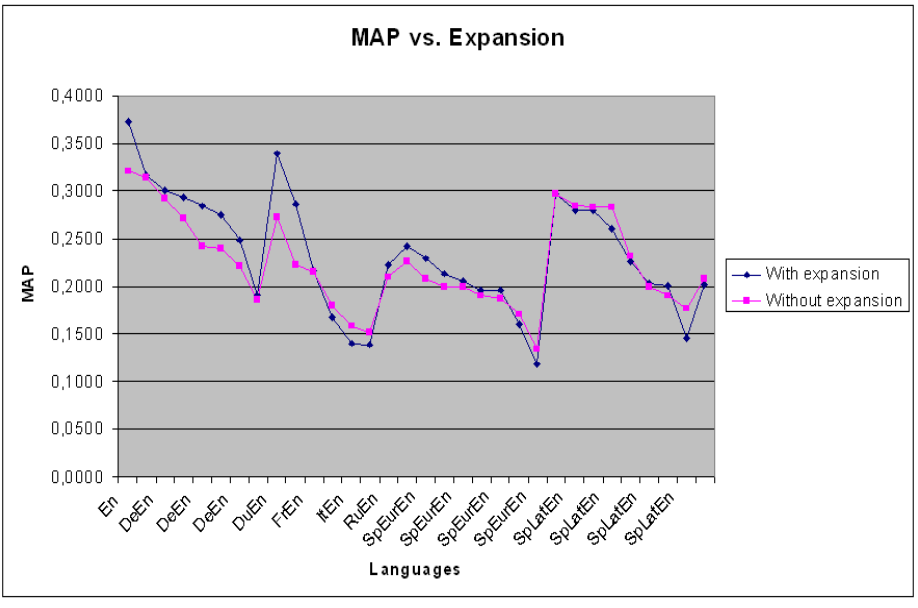

Fig. 3. Graphical view of results for different biligual experiments using the Okapi weighting scheme on the St. Andrews collection. MAP versus expansion.

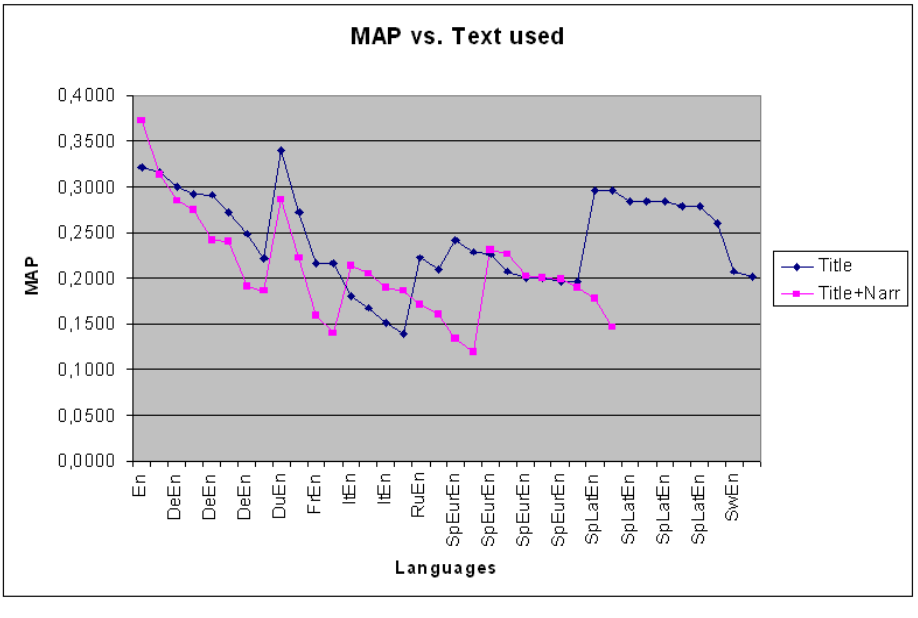

Fig. 4. Graphical view of results for different biligual experiments using the Okapi weighting scheme on the St. Andrews collection. MAP versus text used in query (title or title+narrative).

aproximately 28%. Yet, the other results in the English monolingual task are less acceptable compared to the German bilingual ones, which were found to be more robust to variations in the parametrization of the system. In general, there is a loss of precision compared to the English monolingual results. The Spanish result is the best over all bilingual experiments.

Table 2. Summary of results for the English monolingual adhoc runs and the other bilingual runs on the IAPR TC-12 corpus

Language	Initial Query	Expansion	Weight	MAP	Rank
Monolingual En(English)	title + narr	with	Okapi	**0.2234**	9/49
Monolingual En(English)	title + narr	without	Okapi	0.0845	38/49
Monolingual En(English)	title + narr	with	Tfidf	0.0846	37/49
Monolingual En(English)	title + narr	without	Tfidf	0.0823	39/49
Bilingual DeEn(German)	title + narr	with	Okapi	**0.1602**	4/8
Bilingual DeEn(German)	title + narr	without	Okapi	0.1359	7/8
Bilingual DeEn(German)	title + narr	with	Tfidf	0.1489	5/8
Bilingual DeEn(German)	title + narr	without	Tfidf	0.1369	6/8
Bilingual DuEn(Dutch)	title + narr	with	Okapi	0.1261	4/4
Bilingual FrEn(French)	title + narr	with	Okapi	0.1617	5/8
Bilingual ItEn(Italian)	title + narr	with	Okapi	0.1216	13/15
Bilingual PtEn(Portuguese)	title + narr	with	Okapi	0.0728	7/7
Bilingual SpEn(Spanish)	title + narr	with	Okapi	**0.1849**	4/7

3 Conclusions and Future Work

We have reported on our experimentation in a multimodal and multilingual environment. We have used image collections where each image has an associated text.

We have tried a new Machine Translation module. The application of some heuristics improves the bilingual results, but we consider that further study on the queries with the poorest results is needed in order to improves them. The results are quite good, but in some languages the translation is not that relevant.

Another main conclusion is that the pseudo-relevance feedback increases significantly the results, recommending its use in any case.

Our next work will be the pursuit of improvement on the results focusing on the retrieval phase, applying new techniques of query expansion (using Thesaurus or web information), investigating other heuristics for the Machine Translation module and the combination of results of the text retrieval with another module that works with image information.

Acknowledgements

This work has been partially supported by a grant from the Spanish Government, project TIMOM (TIN2006-15265-C06-03). We would like to thank the Cross-Lingual Evaluation Forum in general, and Carol Peters in particular.

References

1. Reid, N.: The photographic collections in St. Andrews University Library. Scottish Archives 5, 83–90 (1999)
2. Clough, P., Mller, H., Deselaers, T., Grubinger, M., Thomas, M.: The CLEF 2005 Cross-Language Image Retrieval Track. In: Peters, C., Gey, F.C., Gonzalo, J., Müller, H., Jones, G.J.F., Kluck, M., Magnini, B., de Rijke, M., Giampiccolo, D. (eds.) CLEF 2005. LNCS, vol. 4022, Springer, Heidelberg (2006)
3. Müller, H., Geissbhler, A., Ruch, P.: Report on the CLEF experiment: Combining image and multi-lingual search medical image retrieval. In: Peters, C., Clough, P.D., Gonzalo, J., Jones, G.J.F., Kluck, M., Magnini, B. (eds.) CLEF 2004. LNCS, vol. 3491, Springer, Heidelberg (2005)
4. Porter, M.F.: An algorithm for suffix stripping. In: Readings in information retrieval, pp. 313–316. Morgan Kaufmann Publishers, San Francisco (1997) ISBN 1-55860-454-5
5. Martín-Valdivia, M.T., García-Cumbreras, M.T., Díaz-Galiano, M.Á., Ureña-Löpez, M.C., Montejo-Ráez, A.: A.: SINAI at ImageCLEF 2005. In: Peters, C., Gey, F.C., Gonzalo, J., Müller, H., Jones, G.J.F., Kluck, M., Magnini, B., de Rijke, M., Giampiccolo, D. (eds.) CLEF 2005. LNCS, vol. 4022, Springer, Heidelberg (2006)
6. Díaz-Galiano, M.C., García-Cumbreras, M.Á., Martín-Valdivia, M.T., Montejo-Ráez, A.: Ureña-López, L.A.: SINAI at ImageCLEF 2006. In: Proceedings of the Cross Language Evaluation Forum, CLEF 2006 (2006)
7. Callison-Burch, C., Flournoy, R.: A Program for Automatically Selecting the Best Output from Multiple Machine Translation Engines. In: Proceedings of the Machine Translation Summit VIII, Santiago de Compostela, Spain (2001)
8. Gollins, T., Sanderson, M.: Improving Cross Language Retrieval with Triangulated Translation. In: SIGIR (2001)
9. Aceves-Prez, R.M., Montes-y-Gmez, M., Villaseor-Pineda, L.: Enhancing Cross-Language Question Answering by Combining Multiple Question Translations. In: Gelbukh, A. (ed.) CICLing 2007. LNCS, vol. 4394, Springer, Heidelberg (2007)
10. Dominique, L., Patrick, S., Sophie, N.: Cross lingual question answering using QRISTAL for CLEF 2005. In: Peters, C., Gey, F.C., Gonzalo, J., Müller, H., Jones, G.J.F., Kluck, M., Magnini, B., de Rijke, M., Giampiccolo, D. (eds.) CLEF 2005. LNCS, vol. 4022, Springer, Heidelberg (2006)