

Using Text and Image Retrieval Systems: Lic2m experiments at ImageCLEF 2006

Romaric Besançon, Christophe Millet
CEA-LIST/LIC2M
BP 6 92265 Fontenay-aux-Roses CEDEX - FRANCE
besanconr@zoe.cea.fr,milletc@zoe.cea.fr

Abstract

In the ImageCLEF 2006 campaign, the LIC2M participated in the imageclefphoto ad hoc task. We perform experiments on merging the results of two independent search systems: a cross-language information retrieval system exploiting the text part of the query and a content-based image retrieval system exploiting the example images given with the query. The merging is performed *a posteriori* using a weighted sum of the scores given by each system. This kind of merging can improve the results, but gain in the submitted runs remains quite small, comparatively to our experiments of last campaigns. This is due to the relatively poor results of the CBIR part. A first analysis give some hints on possible improvements: example images are often chosen to be visually different to show several aspects of possible relevant images for the chosen topic, therefore the merge of CBIR results for each example image can be irrelevant. However, using only one example image (provided we can choose the best one), or using CBIR results for each example image, but only in correspondence with text results can improve the results.

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; I.4.7 [Computing Methodologies]: Image Processing and Computer Vision—*Feature Measurement*

General Terms

Measurement, Performance, Experimentation

Keywords

Linguistic Processing, Cross-lingual Text Retrieval, Content Based Image Retrieval

1 Introduction

The CEA-LIST/LIC2M laboratory participated in ImageCLEF 2006 to perform experiments on merging strategies to integrate the results obtained from the cross-language text retrieval system and the content-based image retrieval (CBIR) system that are developed in our lab, using a simple merging strategy similar to the one used in previous ImageCLEF campaigns [1].

We use text and visual information from the queries: the title provided for the text retrieval system, and the example images for the CBIR system. Both systems are general-domain systems

and are used independently on each part of the query. Then, *a posteriori* merging strategies are applied on the results provided by each system.

We present in section 2 the retrieval systems for text and image and the merging strategies used. We present the results obtained in section 3.

2 Retrieval systems

Both text retrieval system and CBIR systems are the same that were used in previous ImageCLEF campaigns [1]. The basic principles of the systems are presented here.

2.1 Multilingual Text Retrieval System

The multilingual text retrieval system has not been specially adapted to work on the text of the ImageCLEF corpora, and has simply been used as is: no special treatment has been performed to take into account the structure of the documents (such as title, description, location,date): all fields containing some text have been taken as is. The system works as follows:

Document and query processing The documents and queries are processed through a linguistic analyzer, that extracts relevant linguistic elements such as lemmas, named entities and compounds. The elements extracted from the documents are indexed into inverted files. The elements extracted from the queries are used as query “*concepts*”. Each concept is reformulated into a set of *search terms* for each target language (in the case of imageClefPhoto, only one target language was used) either using a monolingual expansion dictionary (that introduces synonyms and related words), or using a bilingual dictionary.

Document Retrieval Each search term is searched in the index, and documents containing the term are retrieved. All retrieved documents are then associated with a *concept profile*, indicating the presence of query concepts in the document. This concept profile depends on the query concepts, and is language-independent (which allow merging results from different languages). Documents sharing the same concept profile are clustered together, and a weight is associated with each cluster according to its concept profile and to the weight of the concepts (the weight of a concept depends on the weight of each of its reformulated term in the retrieved documents). The clusters are sorted according to their weights and the first 1000 documents in this sorted list are retrieved.

2.2 Content-based Image Retrieval System

The content-based image retrieval system we used in ImageCLEF 2006 is the system PIRIA (Program for the Indexing and Research of Images by Affinity)[3], developed in our lab. The query image is submitted to the system, which returns a list of images ranked by their similarity to the query image. The similarity is obtained by a metric distance that operates on every image signatures. These indexed images are compared according to several classifiers : principally *Color*, *Texture* and *Form* if the segmentation of the images is relevant. The system takes into account geometric transformations and variations like rotation, symmetry, mirroring, etc. PIRIA is a global one-pass system, feedback or “relevant/non relevant” learning methods are not used.

Color Indexing This indexer first quantifies the image, and then, for each quantified color, it computes how much this color is connex. It can also be described as a border/interior pixel classification [5]. The distance used for the color indexing is a classical L2 norm.

Texture Indexing A global texture histogram is used for the texture analysis. The histogram is computed from the Local Edge Pattern descriptors [2]. These descriptors describe the local structure according to the edge image computed with a Sobel filtering. We obtain a 512-bins texture histogram, which is associated with a 64-bins color histogram where each plane of the RGB color space is quantized into 4 colors. Distances are computed with a L1 norm.

Form Indexing The form indexer used consists of a projection of the edge image along its horizontal and vertical axes. The image is first resized in 100x100. Then, the Sobel edge image is computed and divided into four equal sized squares (up left, up right, bottom left and bottom right). Then, each 50x50 part is projected along its vertical and horizontal axes, thus giving a 400-bins histogram. The L2 distance is used to compare two histograms.

2.3 Search and Merging Strategy

Both systems are used independently to retrieve documents from textual and visual information. For the CBIR results, since queries contain several images, a first merging has been performed to obtain a single image list from the results of each query image: the score associated to result images is set to the max of the scores obtained for each query image.

Results obtained by each system are then merged using a weighted sum of the scores obtained by each system. To make results from the different systems comparable, we tried several normalization functions, presented in Table 1, where α_i is the weight associated with the scores of the i^{th} system, RSV_{max} is the the highest score obtained for a query, RSV_{min} the lowest score, RSV_{avg} the average score and RSV_{δ} the standard deviation of the scores. These functions have for instance been tested by [4] for data fusion in the multilingual tracks of previous CLEF campaigns. The submitted runs used the normRSV merging function.

sumRSV	$\sum \alpha_i * RSV_i$
normRSVMax	$\sum \alpha_i * RSV_i / RSV_{max}$
normRSV	$\sum \alpha_i * (RSV_i - RSV_{min}) / (RSV_{max} - RSV_{min})$
Zscore	$\sum \alpha_i * [(RSV - RSV_{avg}) / RSV_{\delta} + (RSV_{avg} - RSV_{min}) / RSV_{\delta}]$

Table 1: Different weightings for score merging

Based on the results from the previous campaigns, we also considered a conservative merging strategy: we use the results obtained by one system only to reorder the results obtained by the other, the score of a document is modified using the same merging coefficient.

3 Results for the ImageClefPhoto task

We used, for the text retrieval part, textual queries in English, Spanish, French and German. We used English and German as independent target languages (as the annotations in both languages refer to the same images and form more an aligned corpus, it did not seem interesting to use both languages as a single multilingual corpus). We only submitted runs with the English target language. For the CBIR system, we tested the color and texture indexers.

We present in Table 2 the results obtained by the CBIR system alone and the text system alone.

We present in Table 3 the results obtained by the merging of the two systems, with a normRSV merging schema, and for different values of α (α being the weight associated with the text results, $1-\alpha$ to the image results). In this merging, we used the English topics with the English annotations and the color indexer. The runs submitted for merged results used $\alpha = 0.7$.

Results are given for the mean average precision (*map*) and the number of relevant documents retrieved (*relret*).

CBIR results			text results for English			text results for German		
indexer	map	relret	topics	map	relret	topics	map	relret
color	0.0468	961	eng	0.1427	1835	eng	0.0916	1445
texture	0.0363	887	spa	0.1416	1427	spa	0.127	1394
			fre	0.1031	1380	fre	0.117	1302
			ger	0.1009	1067	ger	0.145	1381

Table 2: Comparative results for the CBIR system and the image system alone and the text system alone

We see from these results that this simple a-posteriori merging of text and image results based on a weighted sum of the scores can increase the mean average precision and number of relevant documents retrieved (best value of α is around 0.9 or 0.8).

eng			fre		spa		spa	
α	map	relret	map	relret	map	relret	map	relret
1	0.1427	1835	0.1031	1380	0.1416	1427	0.1009	1067
0.9	0.146	1951	0.108	1536	0.144	1702	0.0969	1397
0.8	0.146	1927	0.112	1545	0.14	1714	0.097	1399
0.7	0.14	1864	0.109	1536	0.136	1713	0.0978	1399
0.6	0.128	1807	0.105	1533	0.129	1715	0.0948	1399
0.5	0.112	1750	0.1	1499	0.125	1707	0.0906	1400

Table 3: Comparative results for α in merging strategies

However, the gain is still small, due to the fact that CBIR results are surprisingly quite poor. On one hand, we are investigating in more details the flaws of the indexers for this new image corpus. On the other hand, a first analysis of the results show that merging the CBIR results for the example images before merging with the text results is not a good idea: when several example images are given, they can provide different aspects of what the results should look like, and therefore can be as different as possible, in the range of relevant images. The merging of results base on purely visual similarity can be irrelevant in this case. The analysis of the CBIR results show that the rate of common images in the results for the different example images of a same topic does not exceed, in average, 16 to 18%. Table 4 present the results obtained using only one example image for each topic. The best example image has been taken (according to the reference), and the gain on mean average precision in this case is more than 11%. The problem still remains to find the best image example (in our results, the average precision obtained for each image example does not correlate well with the average score given by the CBIR system).

eng		
α	map	relret
1	0.1427	1835
0.9	0.151	1974
0.8	0.156	1955
0.7	0.159	1914
0.6	0.156	1887

Table 4: Results merging text results with each image example

Another solution to this problem can be to consider each result of an example image as an independent result, and merge all results according to the same schema. In this case, the weight associated to the text result is α and the weight associated to each CBIR result is α/n where n is the number of example images. Table 5 present the results obtained with this method. The gain in this case is around 8%, but this method does not need to determine *a priori* the best example image.

eng		
α	map	relret
1	0.1427	1835
0.9	0.149	1990
0.8	0.153	1962
0.7	0.154	1915
0.6	0.149	1851

Table 5: Results merging text results with all image example

4 Conclusion

The experiments performed by the LIC2M in the ImageCLEF 2006 campaign show that merging results from different media may increase the performance of a search system: a well-tuned *a posteriori* merging of the results obtained by two general purpose systems (no particular adaptation of the systems was made for the two tasks) can improve the mean average precision. An analysis of the CBIR results show that merging the results obtained for different example images can increase the noise in global results since example images are often chosen to be visually different to show several aspects of possible relevant images. Some solutions are proposed to cope with this aspect, such as taking only one example image (the best), or using all example image, but merging each with text results (not between them). Both solutions lead to better results in term of mean average precision.

More sophisticated solutions could be considered, such as working on the image analysis to try to determinate the similarities of the example images and find similar images in the collection based on these similarities, instead of considering each example image independently.

References

- [1] Romaric Besançon and Christophe Millet. Data fusion of retrieval results from different media: experiments at ImageCLEF 2005. In Carol Peters, Fredric C. Gey, Julio Gonzalo, Gareth J.F.Jones, Michael Kluck, Bernardo Magnini, Henning Müller, and Maarten de Rijke, editors, *Accessing Multilingual Information Repositories: 6th Workshop of the Cross-Language Evaluation Forum, CLEF 2005*. Springer, 2005.
- [2] Ya-Chun Cheng and Shu-Yuan Chen. Image classification using color, texture and regions. *Image and Vision Computing*, 21(9), September 2003.
- [3] Magali Joint, Pierre-Alain Moëllic, Patrick Hède, and Pascal Adam. PIRIA : A general tool for indexing, search and retrieval of multimedia content. In *SPIE Electroning Imaging 2004*, San Jose, California USA, January 2004.
- [4] Jacques Savoy and Pierre-Yves Berger. Report on CLEF-2005 evaluation campaign: monolingual, bilingual, and GIRT information retrieval. In Carol Peters, editor, *Working Notes for the CLEF 2005 Workshop*, Vienna, Austria, 2005.
- [5] Renato. O. Stehling, Mario. A. Nascimento, and Alexandre X. Falcão. A compact and efficient image retrieval approach based on border/interior pixel classification. In *CIKM '02: Proceedings of the eleventh international conference on Information and knowledge management*, McLean, Virginia, USA, 2002.