MIRACLE Approach to ImageCLEF 2004: Merging Textual and Content-Based Image Retrieval

J.L. Martínez-Fernández[1,4], A. García-Serrano[2], J. Villena[3,4], and V. Méndez-Sáenz[2]

[1] Advaced Databases Group, Computer Science Department, Universidad Carlos III de Madrid, Avda. Universidad 30, 28911 Leganés, Madrid, Spain
[2] Artificial Intelligence Department, Universidad Politécnica de Madrid, Campus de Montegancedo s/n, Boadilla del Monte 28660, Spain
{agarcia, vmendez }@isys.dia.fi.upm.es
[3] Department of Telematic Engineering, Universidad Carlos III de Madrid, Avda. Universidad 30, 28911 Leganés, Madrid, Spain
[4] DAEDALUS – Data, Decisiond and Language, S.A. Centro de Empresas "La Arboleda", Ctra. N-III km. 7,300 Madrid 28031, Spain
{jmartinez, jvillena}@daedalus.es

Abstract. This paper presents the image retrieval techniques tested by the MIRACLE (Multilingual Information RetrievAl for the CLEf campaign) research group as part of the ImageCLEF 2004 initiative. Two main lines of research continuing the past year's experiments were considered: the application of linguistic techniques to improve retrieval performance and the combination of textual and content-based image retrieval.

1 Introduction

The Multilingual Information RetrievAl for the CLEf campaign (MIRACLE) research group participation in ImageCLEF 2004 centred on two main goals:

- The application of lexical linguistic knowledge in the image retrieval task (based on morphological, syntactic and semantic features of lexical entries).
- To make a first attempt at the use of content-based image retrieval techniques and the combination of these techniques with text-based ones.

The task defined for the ImageCLEF track is centred on the retrieval of images according to a user query based on two different sources of information: the textual descriptions of the pictures and the content of the image file.

Taking into account the textual source of data, linguistic knowledge is introduced using well-known tools for the English language such as the Brill tagger 2, WordNet 3, EuroWordNet 5 and previously developed modules for entity recognition. The availability of EuroWordNet for German, French, Spanish and Italian allowed the definition of experiments where semantic information for these languages is applied. The success of applying this lexical linguistic knowledge depends highly on the quality of the resources for the different languages under consideration. For these languages, EuroWordNet was used for translation and query expansion.

C. Peters et al. (Eds.): CLEF 2004, LNCS 3491, pp. 699 – 708, 2005.
© Springer-Verlag Berlin Heidelberg 2005

Regarding the content-based analysis of images files, the tool GIFT/Viper 0.1.9 6, a public package devoted to image processing, was used. This software provides an implementation of an index and retrieval engine for images. An adapted client is also provided, which has been integrated with the retrieval system used by the MIRACLE team.

The organization of ImageCLEF 2004 proposed three different tasks:

1. an ad hoc bilingual retrieval task, where images are accompanied by English captions,
2. a medical retrieval task, where a set of scan, x-ray, pictures and short textual descriptions of the medical diagnosis are provided, and
3. a user-centred search task, where the main goal is to take user interaction into account in the retrieval process.

In 1 an in-depth description of the different tasks can be found. The MIRACLE team took part in the first two tasks, the first one paying more attention to textual descriptions and the second one to testing the aforementioned content-based image indexing and searching tool. As a result, 45 runs were submitted for both tasks, and a great human effort was made for this CLEF track. The following sections include a detailed description of the experiments, evaluation and analysis of the results.

2 Text-Based Image Retrieval

A flexible system was built to process the text captions provided for each image. The figure in the appendix shows a graphic representation of the different processes followed in the retrieval process according to the languages considered. As previously mentioned, different tools were used to process English queries. A *tagger*, based on Brill's work 2, can be used to attach a morphosyntactic tag to each word. A *proper-names* detection module can be applied at the output of the Brill tagger. A *shallow parser* which, in a final step, was in charge of dividing the text into sentences, whose constituent phrases could be recognized and extracted. Finally, a *semantic component* was used to implement query expansion based on semantic information contained in the WordNet database. Optionally, the linguistic category of a given word was used when the semantic expansion was carried out. For example, if a word acting as a name is going to be expanded, only synonyms of the given word that could act as a name are considered.

For languages other than English, EuroWordNet was used, where available. For languages not covered by EuroWordNet, web translation tools, like Systran[1] or Translation Experts[2] were applied. The main objectives of these experiments were to test EuroWordNet when used in translation tasks and as a synonym expansion tool. For translation purposes, the inter-lingual index (ILI) supplied with EuroWordNet was applied. Again, it was possible to consider the linguistic category of the word when asking for its translations. So, if a name was going to be translated, only words that could act as a name in the target language were taken into account.

[1] "*Altavista's Babel Fish Translation Service*", http://babelfish.altavista.com/
[2] "*Translation Experts*", http://www.transexp.com

2.1 Monolingual English Experiments

The monolingual English experiments were defined by applying different combinations of the modules described in the figure included in the appendix. This allows for the definition of experiments where the linguistic category of the word could be used to filter out possible synonyms. Detected proper names can be treated as special words and used to drive the retrieval process. All tested combinations of these modules are detailed in Table 1, where five different sets of experiments can be distinguished, depending on the kind of linguistic knowledge applied. Of course, image captions must also be indexed according to the process followed in the query. This means that, when proper names were used in the query, image captions were indexed using the same proper-names recognition module. The same applies when common nouns were recognized in the query.

Table 1. Run definitions for the ad hoc retrieval task

	Monolingual English Experiments	
	Query Process	**Run Name**
Baseline	Topic Words	**mirobaseen**
	Topic Words + Synonyms	**mirosbaseen**
Only Nouns	Nouns	**mironounen**
	Nouns + Synonyms without category	**mirosnounen**
	Nouns + Synonyms with category	**miroscnounen**
Baseline + Proper Names	Topic Words + Proper Names	**miroppbaseen**
	Topic Words +Synonyms + Proper Names	**mirosppbaseen**
Nouns + Proper Names	Nouns + Proper Names	**miroppnounen**
	Nouns + Synonyms without category + Proper Names	**mirosppnounen**
	Nouns + Synonyms with category + Proper Names	**miroscppnounen**
Shallow Parsing	Topic and Narration Words	**mirorppbaseen**
	Topic and Narration Words + Synonyms with category	**mirorscppbaseen**

In Table 1, 'Topic Words' means that all recognized words (excluding stopwords) were used to search for the corresponding index database. 'Synonyms' means that all synonyms for a word found in WordNet were used to expand the query, without any refinement, i.e., no disambiguation process is carried out to select the right synonym. 'Nouns' stands for the situation where the query text was tagged and only words acting as nouns were selected as part of the final query. 'Proper Names' is used to mark that only recognized proper names in the text were used as part of the query. 'Synonyms with category' is used to distinguish the process in which not all the synonyms of a word were taken into account, but only those synonyms that could act with the same category as the initial word were included in the query. Finally, in the last two experiments, the narrative of the query (only available for the English queries) was used as the input to the Shallow Parsing module. This module is used to parse the text and get a more precise category for the word.

Average precision results obtained for monolingual experiments are shown in Table 2. The position obtained for each defined run in the absolute ranking produced by ImageCLEF organizers is shown in the third column. This ranking is an ordered list of average precision numbers obtained for every experiment submitted. This simplifies the comparison between different experiments and systems.

Taking these results into account, it is important to highlight some points: first of all, the basic experiment (taken as the baseline) produced the best results. In Table 2, there are two jumps in average precision scores: after run 4 and 8. These differences in precision show that, when all words are used in the characterization of the textual captions, the results are better and the inclusion of more linguistic information (such as proper nouns or synonyms) does not lead to an improvement. On the other hand, if only common or proper nouns are used to represent the documents there is a loss of precision, perhaps due to the fewer number of words used for document characterization. These initial results must be analyzed by taking into account the

Table 2. Average precision results for monolingual English experiments

Run Name	Average Precision	Rank
mirobaseen	0.5865	1
mirosbaseen	0.5623	4
miroppbaseen	0.5609	6
mirosppbaseen	0.5388	8
miroppnounen	0.3384	87
mirosnounen	0.3383	88
mirorppbaseen	0.3366	90
mirosppnounen	0.3337	92
mirorscppbaseen	0.2703	112
miroscppnounen	0.2568	116
mironounen	0.2525	119
miroscnounen	0.2461	120

features of the image captions, i.e., titles are too short for linguistic tools to carry out good parsing. So, further experiments regarding the length of captions should be made.

2.2 Bilingual Experiments

For the bilingual experiments two different approaches, depending on available resources, were considered. These two approaches were:

- A EuroWordNet-based approach, where information contained in the ILI was used to translate the original query. This approach was used for Spanish, German, French and Italian languages.
- A translator-based approach, where online translation tools, in particular Systran and Translation Experts tools, were used to translate queries from the source (or query) language to the target language (English in Image CLEF tasks).

All bilingual experiments used a base indexing process, where all words (excluding stopwords) were included in the index. In Table 3, the last two letters of the run name shown in the first column indicate the query language for the corresponding experiment. Run names where a 'w' appears after the 'miro' part are those where EuroWordNet is used. In every experiment, the target language is English.

Table 3 shows average precision figures obtained for the multilingual experiments defined in the previous section.

Table 3. Average precision for multilingual ad hoc retrieval experiments

Run Name	MAP	%Monolingual	Rank
mirobaseru	0.3866	65.93	73
mirobasedu	0.3807	64.91	76
mirobasesw	0.3043	51.89	99
mirowbaseit	0.2857	48.72	106
mirobaseda	0.2799	47.72	107
mirowbasees	0.2687	45.82	113
mirowbaseesc	0.2615	44.59	114
mirowbasege	0.2455	41.87	122
mirobaseja	0.2358	40.21	124
mirowbasefr	0.2188	37.31	127
mirobasezh	0.1777	30.30	135
mirobasefi	0.1700	28.99	141

According to these results, one important fact worth mentioning is the loss of precision. Taking into account the best monolingual experiment (%Monolingual column), a decrease of 34% in precision is obtained, again highlighting the importance of the quality of the translators used in multilingual environments. Situations in which EuroWordNet have been used as a translation tool can be

compared with the results obtained from CLEF 2003 4 and an important decrease in precision can be noticed. This fact could mean that EuroWordNet is not a good tool for translation purposes. Last year bilingual experiments with French, German, Italian and Spanish achieved around 40% average precision, while this year average precision for these languages is around 30%. It is also worth mentioning that other participants, according to official results, have obtained only a decrease of 10% in precision for some bilingual tasks (but not using EuroWordNet as a translation tool), so, in our situation, there is room for improvement.

3 Content-Based Image Retrieval

In 2004, ImageCLEF organizers defined a new task where the main focus is image content-based retrieval. For this purpose a set of medical images, including scans, x-ray images and photographs of different illnesses were made available to ImageCLEF participants. A more detailed description of the image collection used can be found in 1.

The CBIR system used was GIFT/Viper 0.1.9 6 developed under the GNU licence which enables query by example (using an image as the starting point for the search process) and implements relevance feedback methods. This software was developed by the Vision Group at the Computer Science Center of the University of Geneva. Although different search algorithms can be added, the provided *separate normalisation* algorithm has been used in these experiments.

The first step in the search process for this task must involve an image, but textual descriptions of the medical cases were used to try to improve the retrieval results. The search process can be divided into the following steps:

1. The initial query, made up of one image, is introduced into the CBIR system to obtain a set of images related to the query.
2. The CBIR system returns a list of images along with the corresponding relevance values. Relevance feedback is applied to try to improve results and the number of images used in this refinement process is a configuration parameter of the system.

The MIRACLE research group has not submitted runs where only image content is used in the retrieval process. Algorithms and methods to characterize images based on their content is not the main focus of this research group. Nevertheless, there was a great deal of interest in testing whether the analysis of the content of the image could improve text-based image retrieval. The next section is devoted to the description of the approach followed to mix both kinds of technique.

4 Merging Text-Based and Content-Based Image Retrieval

The first step of the MIRACLE team in content-based image retrieval led to the definition of experiments where content-based image retrieval was applied in combination with text retrieval. This was the case of the ad hoc retrieval task, where some runs mixing results obtained using textual search and CBIR search were submitted. The text retrieval subsystem was the one used in text-based experiments,

although for the initial test and tuning of the overall system, last year's data and text search systems was used.

The process of mixing textual and image results begins by taking the first N elements of the list with the images returned by the text search subsystem and their relevance figures and building a query for the CBIR subsystem. The content search is carried out followed by a new search considering the 5 (RF_IMG in Fig. 1) first elements returned. Finally, results obtained with this last relevance feedback approach are combined with the original results list returned by the textual search subsystem. Fig. 1 shows the search process followed.

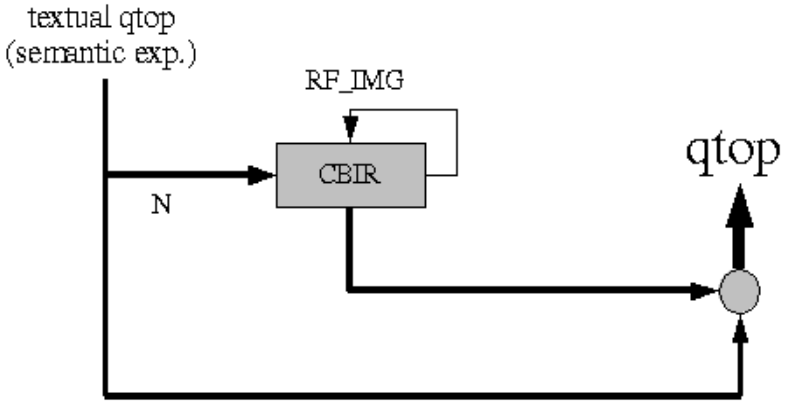

Fig. 1. Text and CBIR subsystem combination model

The combination of the partial results lists provided by each system to provide a unique results list with a global relevance value is:

$$\begin{cases} \sqrt[k]{REL\_VIS^{\,weight\_vis} \times REL\_TXT^{\,weight\_txt}}, & \text{for elements in both lists and } k = weight\_vis + weight\_txt \\ factor\_vis, & \text{for elements appearing only in the list obtained with the CBIR subsystem} \\ factor\_txt, & \text{for elements appearing only in the list obtained with the textual search subsystem} \end{cases}$$

In this expression, *REL_VIS* and *REL_TXT* are the relevance value returned by the CBIR subsystem and the text search subsystem respectively. *factor_vis*, *factor_txt*, *weight_vis* and *weight_txt* are parameters to be empirically established and can be used to adjust the overall system according to the results obtained, for example, giving more importance to textual results or CBIR results.

Several sets of experiments were carried out applying this system. One of them was built using the results of the textual runs described in Section 2. The goal was to evaluate the effect of the image content analysis subsystem on the retrieval process. Table 4 shows average precision obtained for these runs. In these experiments, the values for the parameters defined for the image and textual results combination expression were: N=10, *factor_vis* = 0.5, *factor_txt* = 0.75, *weight_vis* = 1 and *weight_txt* = 2.

Table 4. Average precision values for text and CBIR mixing experiments

Run Name	Average Precision	Rank	Initial Text search Experiment
enenrunexp1	0.5838	2	mirobaseen
enenrunexp7	0.5339	9	mirosppbaseen
enenrunexp4	0.3373	89	mirosnounen
enenrunexp10	0.2533	118	miroscppnounen

Compared to results in Table 2, these results are very close to (and always below) the ones where only a textual search was applied. This could be due to the selected values of the configuration parameters defined in the combination algorithms. These results show, at least, that textual retrieval performance is not compromised by the inclusion of the CBIR system, but more tests should be made to extract a more valid conclusion.

5 Conclusions

The basic objective to be fulfilled this year was to take another step forward in finding a right combination of linguistic and statistical methods to improve the Information Retrieval process. The MIRACLE group is also very interested in the field of multimedia retrieval so, the content-based image retrieval task defined this year, as part of the ImageCLEF track, was a great opportunity to take a first step in this direction. From our point of view, the results obtained for the ad hoc retrieval task were good enough as we explain in the following.

The average precision values for the monolingual English task were a little bit better than those obtained last year, highlighting that it is difficult to improve the results for this task. For monolingual experiments, it seems that the best performance figures that can be obtained with current technology have been reached, so new techniques and methods must be included in the retrieval process. On the other hand, bilingual tasks, in the way we have developed them, have to be improved until the same level of precision as that of monolingual environments is reached.

Some more refined ways of managing semantic information in the retrieval process could be investigated given that the improvement of the retrieval performance, when semantic lexical knowledge is applied, highly depends on the quality of the resources for the different languages to be considered. Besides, different semantic domains have

different degrees of development of the lexical resources, so we also want to prove the influence of that aspect in the results obtained for the different tasks.

In the CLEF workshop held this year, participants had a great deal of interest in content-based image retrieval, but the results obtained were not as good as those of the textual task. This fact compels us to increase our efforts devoted to this kind of retrieval for the following campaigns.

Acknowledgements

This work has been partially supported by the projects OmniPaper (European Union, 5th Framework Programme for Research and Technological Development, IST-2001-32174) and MIRACLE (Regional Government of Madrid, Regional Plan for Research, 07T/0055/2003).

Special mention must be done to our colleagues at other members of the MIRACLE group: Santiago González Tortosa, Michelangelo Castagnone, José Carlos González, José Miguel Goñi, Javier Alonso, Paloma Martínez, César de Pablo y Francisco Bueno.

References

1. Clough, P., Müller, H. and Sanderson, M.: The CLEF Cross Language Image Retrieval Track (ImageCLEF) 2004, in Fifth Workshop of the Cross--Language Evaluation Forum (CLEF 2004), Lecture Notes in Computer Science (LNCS), Springer, Heidelberg, Germany (in print), 2005.
2. Brill E.: Some Advances in Transformation Based Part of Speech Tagging, Proceedings of the Twelfth National Conference on Artificial Intelligence, 1994
3. Miller, G.A.: WordNet: A lexical database for English. Communications of the ACM, 38(11):39—41, 1995
4. Villena J., Martinez-Fernandez J.L., Fombella J., García-Serrano A., Ruíz-Cristina A., Martínez P., Goñi J.M., González-Cristóbal J.C.: Image Retrieval: the MIRACLE Approach, 4th Workshop of the Cross-Language Evaluation Forum, CLEF 2003, Trondheim, Norway, August 21-22, 2003, Revised Selected Papers, Series : LNCS , Vol. 3237, Peters, C.; Gonzalo, J.; Braschler, M.; Kluck, M. (Eds.), 2005
5. Eurowordnet: Building a Multilingual Database with Wordnets for several European Languages", http://www.illc.uva.nl/EuroWordNet/, March, 1996
6. The GNU Image-Finding Tool GIFT 0.1.9, http://www.gnu.org/software/gift/, last accessed 12.08.2004

Appendix. Query Processing Applied for the Ad Hoc Retrieval Task