

# MIRACLE (FI) at ImageCLEFphoto 2009

R. Granados<sup>1</sup>, X. Benavent<sup>2</sup>, R. Agerrí<sup>1</sup>, A. García-Serrano<sup>3</sup>, J.M. Goñi<sup>1</sup>,  
J. Gomar<sup>2</sup>, E. de Ves<sup>2</sup>, J. Domingo<sup>2</sup>, G. Ayala<sup>2</sup>

<sup>1</sup> Universidad Politécnica de Madrid, UPM

<sup>2</sup> Universidad de Valencia

<sup>3</sup> Universidad Nacional de Educación a Distancia, UNED

{[rgranados@fi.upm.es](mailto:rgranados@fi.upm.es), [xaro.benavent@uv.es](mailto:xaro.benavent@uv.es), [rodrigo.agerri@upm.es](mailto:rodrigo.agerri@upm.es), [agarcia@lsi.uned.es](mailto:agarcia@lsi.uned.es)}

## Abstract

The Miracle-FI participation at ImageCLEF 2009 photo retrieval task main goal was to improve the merge of content-based and text-based techniques in our experiments. The global system includes our own implemented tool IDRA (InDexing and Retrieving Automatically), and the Valencia University CBIR system. Analyzing both “topics\_part1.txt” and “topics\_part2.txt” task topics files, we have built different queries files, eliminating the negative sentences with the text from title and clusterTitle or clusterDescription, one query for each cluster (or not) of each topic from 1 to 25 and one for each of the three images of each topic from 26 to 50. In the CBIR system the number of low-level features has been increased from the 68 component used at ImageCLEF 2008 up to 114 components, and in this edition only the Mahalanobis distance has been used in our experiments. Three different merging algorithms were developed in order to fuse together different results lists from visual or textual modules, different textual indexations, or cluster level results into a unique topic level results list. For the five runs submitted we observe that MirFI1, MirFI2 and MifFI3 obtain quite higher precision values than the average ones. Experiment MirFI1, our best run for precision metrics (very similar to MirFI2 and MirFI3), appears in the 16th position in R-Precision classification and in the 19th in MAP one (from a total of 84 submitted experiments). MirFI4 and MirFI5 obtain our best diversity values, appearing in position 11th (over 84) in cluster recall classification, and being the 5th best group from all the 19 participating ones.

## Categories and subject descriptors

**H.3 [Information Storage and Retrieval]:** H.3.1 Content Analysis and Indexing; H.3.2 Information Storage; H.3.3 Information Search and Retrieval; H.3.4 Systems and Software; H.3.7 Digital libraries. **H.2 [Database Management]:** H.2.5 Heterogeneous Databases; **E.2 [Data Storage Representations].**

## Keywords

Information Retrieval, Content-Based Image Retrieval, Merged result lists, Indexing, Named Entities Recognition

## 1 Introduction

The Miracle-FI participation at ImageCLEF 2009 photo retrieval task [17] main goal was to improve the merge of content-based and text-based techniques in our experiments. The global system includes our own implemented tool IDRA (InDexing and Retrieving Automatically), and the Valencia University CBIR system.

The BELGA Collection image captions were preprocessing to build a semi-structured XML description similar to the ImageCLEFphoto08 task [11]. Also the images have been preprocessed for CBIR module because some of them have some bands on the frame of the image with color pixels of the RGB and MCY system colors. As we realize that they don't follow any established format, the solution adopted was to reduce all the images to the 90% of his real size in order to eliminate the different bands and the white pixels frames.

Analyzing both “topics\_part1.txt” and “topics\_part2.txt” task topics files, we have built different queries files: [qf2] “**BELGAtopics-tctcd-(q-cl)-fQ.txt**”: one query for each cluster of each topic with the text from title, clusterTitle and clusterDescription (eliminating the negative sentences and negative clusters), [qf3] “**BELGAtopics-tct-(q-cl)-fQ.txt**”: the same as above but just with the text from title and clusterTitle, [qf4] “**BELGAtopics-topEnt(1..25)+capEnt(26..50)-(q-cl)-fQ.txt**”: one query for each cluster (except negatives ones) of each topic from 1 to 25 and one for each of the three images of each topic from 26 to 50, [qf5] “**BELGAtopics-cap(title+desc)-(26..50)-(q-cl)-fQ.txt**”: one query for each one of the three images of each topic from 26 to 50, obtained from the title and description fields of the XML captions.

In the CBIR system the number of low-level features has been increased from the 68 component used at ImageCLEF 2008 up to 114 components, mainly due to the use of local color histogram descriptors that were not use last year. This edition only the Mahalanobis distance has been used in our experiments.

Three different merging algorithms were developed in order to fuse together different results lists from visual or textual modules, different textual indexations, or cluster level results into a unique topic level results list: **MAXmerge** (the algorithm selects the results from the N lists which have a higher relevance value), **EQUmerge** (the algorithm selects the first result of each query (cluster), not selected yet), and **ENRICH** (this merging uses two results lists, a main list and a support list, and when a concrete result appears in both lists, the relevance will be increased).

The five runs submitted were: [run1] “**MirFI1\_T-CT-I\_TXT-IMG**”: launching [qf3], reordering textual results list with CBIR system, and merging both lists with the ENRICH algorithm, [run2] “**MirFI2\_T-CT-CD-I\_TXT-IMG**”: the same as above, but launching [qf2], [run3] “**MirFI3\_T-CT-CD\_TXT**”: textual experiment launching [qf2], [run4] “**MirFI4\_T-CT-CD-I\_TXT**”: topics 1 to 25 as [run3], and topics 26 to 50, launching [qf5], [run5] “**MirFI5\_T-CT-CD-I\_TXT**”: textual experiments with NER. [qf4]. All the queries files in the submitted experiments were built from different queries of different clusters of a topic, so the EQUmerge algorithm was applied to fuse the different clusters-based result lists into a unique one. In the results, we observe that experiment MirFI1, our best run for precision metrics (very similar to MirFI2 and MirFI3), appears in the 16th position in R-Precision classification and in the 19th in MAP one (from a total of 84 submitted experiments). MirFI4 and MirFI5 obtain our best diversity values, appearing in position 11th (over 84) in cluster recall classification, and being the 5th best group from all the 19 participating ones.

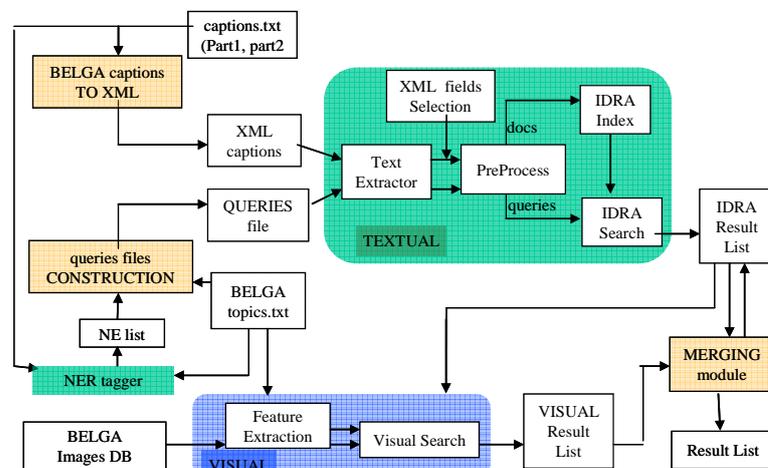


Fig. 1. System overview

## 2 System Description

The global system (shown at Fig. 1) includes our own implemented tool IDRA (InDexing and Retrieving Automatically), and the Valencia University CBIR system. The main goal, using IDRA [12] with such a large collection, was to analyze how the obtained results from the textual module could be improved using information from the content-based module. In this year, a global strategy for all experiments has been that the Content-Based module always starts working with a selected textual results list as part of his input data (different from our participation at ImageCLEF 2008 [11]).

## 2.1 Collection preprocessing

ImageCLEFphoto09 task uses the so-called “BELGA Collection” which contains 498,920 images from Belga News Agency. Each photograph is accompanied by a caption composed of English text up to a few sentences in length [3]. Image captions are provided without a specific format. Because of this, we preprocess the captions file to build a semi-structured XML description for each image, similar to the used in the ImageCLEFphoto08 task [1]. This format includes 8 tags (docno, title, description, notes, location, date, image and thumbnail), which we try to fill preprocessing the captions texts. This preprocess consists on trying to identify in each caption the appropriate part of the text to fill the XML tags. We can see an example of this transformation in Fig. 2.

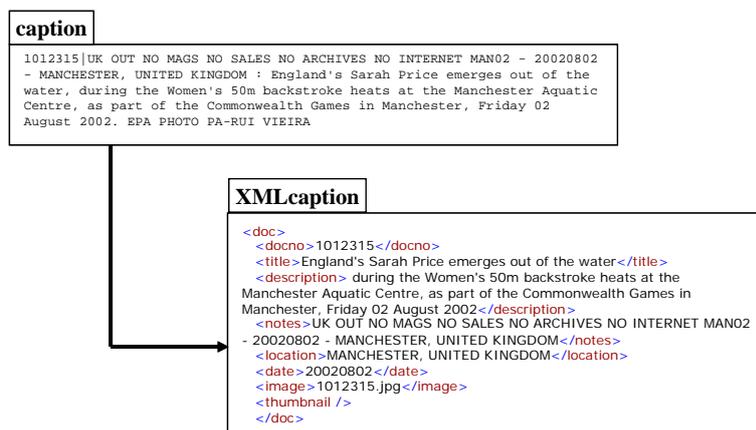


Fig. 2. “caption TO XML” example

The images of the database have been pre-processed for the Content-Based Image module because some of them have extra-information on the image itself. This extra-information consists of some bands on the frame of the image with color pixels of the RGB and MCY system colors. This kind of information is often used for color calibration. So that, the first attempt was to use this extra-information in order to calibrate the color images of the database. But, after a visual analysis of different images we realize that they don't follow an established format. At fig. 3 different images formats are shown: two vertical color bands, two horizontal color bands, only one color band, some color bands have the two color systems (RGB and MCY), others only one of the color systems, others extra white frame of different sizes. Therefore, the solution adopted was to reduce all the images to the 90% of his real size in order to eliminate the different bands and the white pixels frames.



Fig. 3. Different format images from the Belga News Agency database.

## 2.2 Queries files construction

Analyzing both “topics\_part1.txt” and “topics\_part2.txt” task topics files, we built different queries files to be launched against the IDRA indexation as the first step in the generation of our experiments. The IDRA queries

format separated by a blank. The different queries files constructed are explained in the following using the example in Fig. 4 of this year's topics (from 'Topics – part 1') shown in the official website of the task [14].

```

<top>
<num> Number: 0 </num>
<title> soccer </title>
<clusterTitle> soccer belgium </clusterTitle>
<clusterDesc> Relevant images contain photographs of the Belgium team in a soccer match. </clusterDesc>
<image> belga38/00704995.jpg </image>
<clusterTitle> spain soccer </clusterTitle>
<clusterDesc> Relevant images contain photographs of the Spain team in a soccer match. </clusterDesc>
<image> belga6/00110574.jpg </image>
<clusterTitle> beach soccer </clusterTitle>
<clusterDesc> Relevant images contain photographs of a soccer beach match. </clusterDesc>
<image> belga33/06278068.jpg </image>
<clusterTitle> italy soccer </clusterTitle>
<clusterDesc> Relevant images contain photographs of the Italy team in a soccer match. </clusterDesc>
<image> belga20/1027435.jpg </image>
<clusterTitle> soccer netherlands </clusterTitle>
<clusterDesc> Relevant images contain photographs of the Netherlands team in a soccer match or the teams
in Netherlands' league. </clusterDesc>
<image> belga10/01214810.jpg </image>
<clusterTitle> soccer -belgium -spain -beach -italy -Netherlands </clusterTitle>
<clusterDesc> Relevant images contain photographs of any aspects or subtopics of soccer which are not
related to the above clusters. </clusterDesc>
<image> belga20/01404831.jpg </image>

```

Fig. 4. Example of ImageCLEFphoto09 topics (from 'Topics – part 1')

- [qf1] “BELGAtopics-all-(q)-fQ.txt”: one query per topic containing one stream with all the text from all the clusters (not used for runs). [qf1] would contain the query shown in Fig. 5.

```

0 soccer soccer belgium Relevant images contain photographs of the Belgium team in a soccer match. spain soccer Relevant
images contain photographs of the Spain team in a soccer match. beach soccer Relevant images contain photographs of
a soccer beach match. italy soccer Relevant images contain photographs of the Italy team in a soccer match. soccer
netherlands Relevant images contain photographs of the Netherlands team in a soccer match or the teams in Netherlands'
league. soccer -belgium -spain -beach -italy -netherlands Relevant images contain photographs of any aspects or subtopics
of soccer which are not related to the above clusters.

```

Fig 5. [qf1] example

- [qf2] “BELGAtopics-tctcd-(q-cl)-fQ.txt”: one query for each cluster of each topic with the text from title, clusterTitle and clusterDescription. We eliminate the negative sentences (those containing words “not” or “irrelevant”). We do not include the negative clusters as “**soccer -belgium -spain -beach -italy -netherlands**”. [qf2] would contain the queries shown in Fig. 6 for the topic in the example of Fig. 4:

```

0-1 soccer soccer belgium Relevant images contain photographs of the Belgium team in a soccer match.
0-2 soccer spain soccer Relevant images contain photographs of the Spain team in a soccer match.
0-3 soccer beach soccer Relevant images contain photographs of a soccer beach match.
0-4 soccer italy soccer Relevant images contain photographs of the Italy team in a soccer match.
0-5 soccer soccer netherlands Relevant images contain photographs of the Netherlands team in a soccer match or the teams in
Netherlands' league.
1-1 ...

```

Fig 6. [qf2] example

- [qf3] “BELGAtopics-tct-(q-cl)-fQ.txt”: the same as above but just with the text from title and clusterTitle. Fig 7. shows the constructed queries from the example topic for [qf3].

```

0-1 soccer soccer Belgium
0-2 soccer spain soccer
0-3 soccer beach soccer
0-4 soccer italy soccer
0-5 soccer soccer Netherlands
1-1 ...

```

Fig 7. [qf3] example

- [qf4] “BELGAtopics-topEnt(1..25)+capEnt(26..50)-(q-cl)-fQ.txt”: one query for each cluster (except negatives ones) of each topic from 1 to 25 and for each one of the three images of each topic from 26 to

50. The associated text of each query is obtained extracting the named entities (with the NER tagger module) from the clusterTitle and clusterDescription fields of the corresponding topic, in the case of topics 1 to 25, and from the associated XML files for each of the three images in the case of topics 26 to 50. In the case of the topic example in Fig. 4, the corresponding constructed queries in [qf4] would be:

0-1 soccer belgium belgium belgium soccer soccer match
0-2 spain spain soccer spain spain team soccer soccer match
0-3 beach soccer soccer soccer beach match beach match
0-4 italy soccer soccer italy italy team soccer soccer match.
0-5 soccer netherlands netherlands netherlands team soccer soccer match netherlands league
1-1 ...

**Fig 8.** [qf4] example

- [qf5] “**BELGATopics-cap(title+desc)-(26..50)-(q-cl)-fQ.txt**”: one query for each one of the three images of each topic from 26 to 50. The text for each query is obtained from the concatenation of the TITLE and DESCRIPTION fields of the XML files for these captions.

## 2.3 IDRA text-based index and retrieval

IDRA textual retrieval is based on the VSM approach using weighted vectors based on the TF-IDF weight. Applying this approach, a representing vector will be calculated for each one of the image captions in the collection. The components of the vectors will be the weight values for the different words in the collection. When a query is launched, a vector for that query is also calculated and compared with all the vectors stored during the index process. This comparison will generate the ranked results list for the launched query.

The textual retrieval task architecture can be seen in the Fig. 1. Each one of the components takes care of a specific task. These tasks will be sequentially executed:

- **Text Extractor.** Is in charge of extracting the text from the different files. It uses the JDOM Java API to identify the content of each of the tags of the captions XML files. This API has problems with some special characters, so it is needed to carry out a pre-process of the text to eliminate them.
- **Preprocess.** This component process the text in two ways:
  - o special characters deletion: characters with no statistical meaning, like punctuation marks, are eliminated.
  - o stopwords detection: exclusion of semantic empty words from a new constructed list, different from last year one.
- **XML Fields Selection.** With this component, it is possible to select the desired XML tags of the captions files, which will compound the associated text describing each image. In the captions XML files there are eight different tags (DOCNO, TITLE, DESCRIPTION, NOTES, LOCATION, DATE, IMAGE and THUMBNAIL). In the index process, the selected tags from the captions XML files had been three: TITLE, DESCRIPTION, and LOCATION.
- **IDRA Index.** This module indexes the selected text associated with each image (its XML caption). The approach consists in calculate the weights vectors for each one of the images selected texts. Each vector is compounded by the TF-IDF weights values [16] of the different words in the collection. TF-IDF weight is a statistical measure used to evaluate how important a word is to a text in a concrete collection.

$$TF - IDF = t_{i,j} * \log_2 \left( \frac{N}{n_i} \right) \quad \begin{array}{l} t_{i,j}: \text{number of occurrences of the word } t_j \text{ in caption text } T_i. \\ N: \text{total number of images captions in the collection.} \\ n_i: \text{number of captions in which appears the word } t_i. \end{array} \quad (1)$$

All the weights values of each vector will be then normalized using the Euclidean distance between the elements of the vector. Therefore, the IDRA Index process update the next values for each one of the words appearing in the XML captions collection:  $n_i$ : number of captions in which appears the word  $t_i$ ,  $T_i$ : identifier of the image XML caption,  $t_{i,j}$ : number of occurrences of the word  $t_j$  in a caption text  $T_i$ ,  $idf_j$ : inverse document frequency (  $\log_2(N/n_i)$  ) in  $T_i$ ,  $E_i$ : Euclidean distance in the corresponding vector used to normalize,  $w_{j,i}$ : weight of word  $t_j$  in  $T_i$ .

- **IDRA Search.** For the query text is also calculated his weights vector in the same way as above. Now, the similarity between the query and an image caption will depend on the proximity of their associated vectors. To measure the proximity between two vectors we use the cosine.

$$sim(T_i, q) = \cos(\vec{O}) = \frac{\sum w_{j,i} * w_{j,i}}{\sqrt{\sum w_{j,i} * w_{j,i}} * \sqrt{\sum w_{q,i} * w_{q,i}}} \quad (2)$$

This value of similarity will be calculated between the query and all the images captions indexed, and the images will be ranked in descending order as the IDRA result list.

To index the collection, the system needs approximately 2 days to index each one of the 5 parts in which the collection was divided to be indexed. These 5 indexations processes can be executed concurrently. Queries file response time depends on the concrete queries file launched (on the large of the queries texts), but it takes over 10 hours to obtain a results file for 119 queries (119 queries at cluster level).

## 2.4 Named Entity Recognition (NER) functionality

The general aim of using Named Entity Recognition (NER) in our approach was to perform retrieval using the named entities extracted from both the documents (XML-structured captions as discussed in section 2.1) and the topics. Due to time constraints, only the later was used in one of the runs submitted ([run5]).

Considering the nature of the text in the topics, namely, all words were lowercase, topics in the part 2 file, did not contain enough text, etc., would not make it easy to use an off-the-self named entity tagger, we decided instead to tag (tokenize, Part-Of-Speech and NER) the captions document as released by the imageCLEF organisers. The C&C taggers [6] were used off-the-self. After that, our task was reduced to extract those unique linguistic expressions that were tagged as Location, Person or Organization.

However, we soon realized that the resulting annotation was not correctly picking up information which seems to be crucial to determine the topic of the images. More specifically, "Time" would refer to an Organization whereas the description "Time magazine correspondent" would refer to a person and, as such, the modifier correspondent seems relevant to describe the situation captured by a given image. Most of these modifiers would not be picked-up by a NER tagger because they are not in uppercase.

This turned out to be a symptom of a more general problem as presented in [9]. For tasks such as IR, it is not sufficient to have low level tools that produce high quality linguistic annotation and analysis, but it is also required that the results of the various levels of annotation be consistent with each other. Our proposal aims to exploit the interaction between the various levels of annotations (POS, NER and Chunks) provided by the C&C taggers in order to obtain a better bracketing of named entities. The general idea is to create foci consisting of those words or expressions marked-up as named entities. Whenever the C&C tagger annotates a word as a named entity, a chunk/phrase is built around it by attaching those surrounding/satellites terms that act as modifiers of the named entity according to their POS and membership to a particular chunk. We are currently able to deal with periods and abbreviations, with prepositions, adjectives and nouns. This approach allows us to extract entities such as Paris-Roubaix race, princess Mathilde, Leonardo da Vinci international airport (instead of Leonardo da Vinci), District of Columbia, Royal Palace of Brussels, etc. The following caption (id 1470132) is illustrative of our procedure:

*“American photojournalist James Nachtwey in a file photograph from May 18 2003 as he is awarded the Dan David prize in Tel Aviv for his out standing contribution to photography. It was announced by Time magazine on Thurs day, 11 December 2003 that Nachtwey was injured in Baghdad along with Time magazine senior correspondent Michael Weisskopf when a hand grenade was thrown into a Humvee they were traveling in with the US Army. Both journalists are reported in stable condition and are being evacuated to a US military hospital in Germany.”*

On the one hand, the named entities annotated by the tagger for this text are: American James Nachtwey, Dan David, Tel Aviv, Baghdad, Michael Weisskopf, US Army, US, Germany, and Nachtwey. On the other, the descriptions extracted by our system are: American photojournalist James Nachtwey, Dan David prize, Tel Aviv, Baghdad, Time magazine senior correspondent Michael Weisskopf, US Army, US military hospital, Germany and Natchtwey. It is particularly noticeable that our system was able to recognize Time magazine, and that the topic of the caption is about the Dan David prize and a US military hospital in Germany.

This approach was used to extract both the named entities and descriptions of all the captions. We then used the lowercase version of the entities found in the captions that were present in the part 1 file of the topics. This method was also applied to the captions corresponding to the images that formed the clusters in part 2 of the topics. They constituted the queries file [qf4] for the [run5] (see section 2.2 and 3). The total process of annotation, analysis, extraction of entities/descriptions took 87 machine hours (on a standard Pentium 4 PC).

## 2.5 Visual Retrieval

The VISION-Team at the Computer Science Department of the University of Valencia has its own CBIR system mainly used for relevance feedback algorithms evaluation [8, 15], and that was used for ImageCLEF 2008 for the first time. The low-level features of the CBIR system have been adapted for the images of the new image database (2009) taking into account the results of the last year.

As in most CBIR systems, a feature vector represents each image. The first step at the Visual Retrieval system is extracting these features for all the images on the database as for each of the cluster query topic images for each question. We use different low-level features describing color and texture to build a vector of features. The number of low-level features has been increased from the 68 component (ImageCLEF 2008) up to 114 components at the current edition. This increment is mainly due to the use of local color histogram descriptors that were not use last year.

- **Color information:** Color information has been extracted calculating both local and global histograms of the images using a bin of size 10x3 on a HSV color system. Local histograms have been calculated dividing the images in four fragments of the same size. For this database, only the H (hue) component has been used so that the other values were almost zero for as it happened at the IAPR database. Therefore, a feature vector of 10 components for the global histogram, and 40 components for the local histograms represent the color information of the image.
- **Texture information:** As it was done for the IAPR data base, six feature textures have been computed for this repository respectively. The first three ones use code from the implementation done by Smith and Burn in Meastex [19]; the rest have been implemented by the authors. The total of texture features builds a vector of 64 components:
  - Gabor Convolution Energies [10].
  - Gray Level Cooccurrence Matrix also known as Spatial Gray Level Dependence [7].
  - Gaussian Random Markov Fields [4].
  - The granulometric distribution function, first proposed by Dougherty [5]. We have used here not the raw distribution but the coefficients that result of fitting its plot with a B-spline basis.
  - Finally, the Spatial Size Distribution [2]. We have used two different versions of it by using as the structuring elements for the morphological operation that get size both a horizontal and a vertical segment.

The second step is to calculate the similarity distance between the features vectors from each image on the database to each the cluster images. Last edition (ImageCLEF 2008), we tested two different metrics to calculate this distance: the Euclidean and the Mahalanobis. In all experiments better results were obtained with the Mahalanobis distance due to the fact that this measure takes into account the correlations of the data set and is scale-invariant being this feature very useful because the broad differences between the different low-level feature values. Therefore, this edition only the Mahalanobis distance has been used in our experiments. These sorted lists are passed to the merging module of the global system.

## 2.6 Merging algorithms

Different merging algorithms were developed in order to fuse together different results lists from visual or textual modules, different textual indexations, or cluster level results into a unique topic level results list. All the 3 merging algorithms detailed here require trec\_eval format [19] for input results lists, which is the format required by the ImageCLEF organization to submit the definitive runs.

**MAXmerge.** This algorithm is used to fuse together the N (configurable value) results lists obtained when a concrete queries file is launched against the N indexations corresponding to the N parts in which the collection was divided to be indexed. For each query, the algorithm selects the results from the N lists which have a higher

relevance value for the corresponding query, independently of the list the results appears in. The maximum number of results per query in the resulting list is set up to 1000 ('max' is a configurable parameter).

**EQUImerge.** When a results list contains results for queries which references clusters (not topics), this algorithm is used to select the formers represents of each cluster to build a unique results list with results for queries which references topics. The algorithm selects the first result of each query-cluster, not selected yet, to build the results list for the corresponding query. A preliminary step is carried out to separate the results into several lists depending on the number of cluster (inside each topic) the results belong to. The relevance value will be decremented (configurable value 'decr') for each result starting with the original relevance value of the first selected result. The maximum number of results per query is set up to 1000 ('max' is a configurable parameter).

**ENRICH.** This merging uses two results lists, a main list and a support list. The merged results list will have a maximum of 1000 results per query (configurable). If a concrete result appears in both lists for the same query, the relevance of this result in the merged list will be increased in the following way:

$$newRel = mainRel + \frac{supRel}{(posRel + 1)}$$

newRel: new relevance v. in the merged list  
mainRel: relevance value in the main list  
supRel: relevance value in the support list  
posRel: position in the support list

(6)

Relevance values will be then normalized from 0 to 1. Every result appearing in the support list but not in the main one (for each query), will be added at the end of the corresponding list. In this case, relevance values will be normalized according with the lower value in the main list. In the last year implementation of this algorithm, this addition didn't work correctly, so this year it has been modified in a proper way. In this experiment, main and support lists are compound by a maximum of 1000 results for each query. Also the merged lists resulting will be limited to the same number of results per query.

### 3 Experiments (submitted runs)

This campaign 5 different runs were submitted to the ImageCLEFphoto2009. All of them were based on the 5-parts indexation carried out using the TITLE, DESCRIPTION and LOCATION XML tags. Runs start from the launch of some of the queries files described in section 2.2. When one of these files is launched, it is against these 5 indexations and later, the 5 obtained results lists over each indexation are merged using the MAXmerge algorithm.

- [run1] **"MirFI1\_T-CT-I\_TXT-IMG"**: mixed (textual/visual) experiment launching [qf3], reordering this textual results list with content-based results, and merging both lists with the ENRICH algorithm.
- [run2] **"MirFI2\_T-CT-CD-I\_TXT-IMG"**: the same as above, but launching [qf2].
- [run3] **"MirFI3\_T-CT-CD\_TXT"**: textual experiment launching [qf2].
- [run4] **"MirFI4\_T-CT-CD-I\_TXT"**: the first part (topics 1 to 25) doing exactly the same as [run3], and the second one (26 to 50) launching [qf5].
- [run5] **"MirFI5\_T-CT-CD-I\_TXT"**: textual experiments treating with NER. [qf4] is launched to obtain the results list.

As it can be observed, all the queries files used in the submitted experiments were built at cluster level, that is, with different queries for the different clusters of a topic, as explained in the queries files construction (section 2.2). The obtained results lists will have results for each one of the clusters of each topic, and EQUImerge algorithm will be applied to fuse together the different clusters-based lists into a unique query-based one.

### 4 Results and concluding remarks

After the evaluation by the task organizers, obtained results for each of the submitted experiments are presented in Table 1. The table shows for each run: the identifier, the mean average precision (MAP), the R-Precision, the precision at 10 and 20 first results, the number of relevant images retrieved (out of a total of 34887 relevant images in the collection), and the cluster recall at 10 and 20. Average values from all the experiments presented

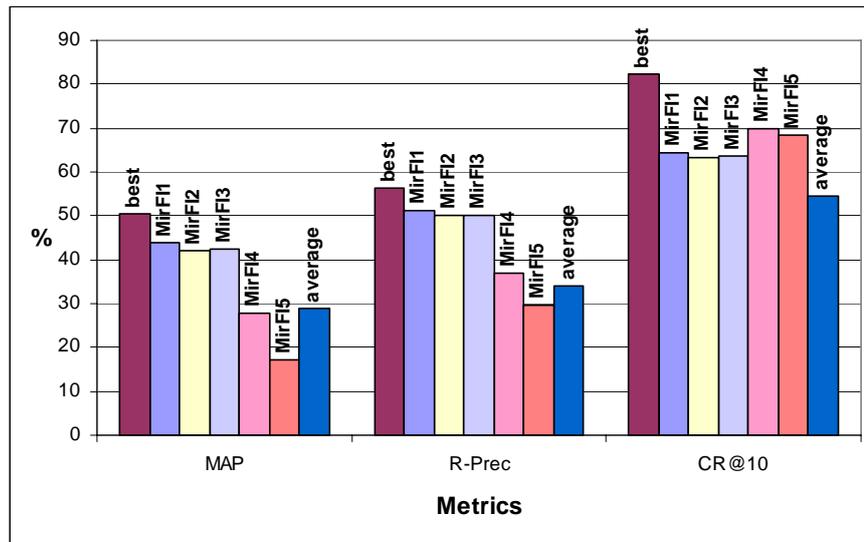
to the task for these metrics are also shown in the table, as well as the best value obtained for each of the metrics. MAP, R-Precision and CR@10 values are all shown in Fig. 9 in order to be visually compared.

**Table 1.** Results for the submitted experiments.

Run Identifier	MAP	R-Prec	Prec@10	Prec@20	RelRet	CR@10	CR@20
<b>MirFI1_T-CT-I_TXT-IMG</b>	43.78	51.39	82.00	81.80	16547	64.51	72.00
<b>MirFI2_T-CT-CD-I_TXT-IMG</b>	42.25	50.11	80.00	81.00	16301	63.24	73.41
<b>MirFI3_T-CT-CD_TXT</b>	42.33	50.12	80.80	81.40	16301	63.51	73.31
<b>MirFI4_T-CT-CD-I_TXT</b>	27.84	36.96	47.40	48.90	13627	69.83	76.76
<b>MirFI5_T-CT-CD-I_TXT</b>	17.33	29.53	23.60	26.30	13498	68.49	72.82
<b>average</b>	29.08	34.09	--	--	--	54.67	62.35
<b>best</b>	50.64	56.43	--	--	--	82.39	86.07

At first sight, we can observe that MirFI1, is our best run for precision metrics (very similar to MirFI2 and MirFI3), and appears in the 16th position in R-Precision classification and in the 19th in MAP one (from a total of 84 submitted experiments). 19 groups participate in the task and only 6 of them obtain better precision results than our best experiment.

Regarding the diversity metrics (cluster recall at 10 an 20, CR@10 and CR@20), MirFI4 and MirFI5 obtain our best diversity values, appearing in position 11th (over 84) in cluster recall classification, and being the 5th best group from all the 19 participating ones.



**Fig. 9.** Comparison of own experiments with best and average values

Comparing obtained results from experiments MirFI1 and MirFI2, we can see that not using CD (cluster description) tag from the topics is a quite better for precision results and very similar in diversity ones. So we can say that the addition of this field in the queries construction step was not very useful. Obtained results for experiments MirFI2 and MirFI3 are almost the same. So we can conclude that the use of the ENRICH merging algorithm with the visual re-ranked results list, does not affect the results in a significant way.

MirFI3 and MirFI4 are differentiated in the way of constructing the second half of the queries (topics from 26 to 50). As explained in sections 2.2 and 3, MirFI4 extracts the text from the captions corresponding to the example images included in the second part of the topics. The evaluation of the results shows that MirFI3 obtains better precision results than MirFI4, but worse diversity ones. One reason is that the use of the captions text adds more information to the queries, which is useful for the diversity aim, is noise for the precision one.

The goal of experiment MirFI5 was to analyze if results could be improved with the use of NER techniques for the construction of the queries. The obtained precision results for this experiment was our worse ones (seems

as a lot of noise was introduced) but the addition of entities information to the queries, improves the diversity results. Experiments MirFI4 and MirFI5, also show how this additional information improves the diversity results, but makes the precision ones worse.

## Acknowledgements

This work has been partially supported by the Spanish R+D National Plan, by means of the project BRAVO (Multilingual and Multimodal Answers Advanced Search – Information Retrieval), TIN2007-67407-C03-03; by the Madrid's R+D Regional Plan, by means of the project MAVIR (Enhancing the Access and the Visibility of Networked Multilingual Information for the Community of Madrid), S-0505/TIC/000267; and by the Spanish Ministry of Education and Science, by means of the project MCYT, TIC2002-03494.

## References

1. Arni, T., Clough, P., Sanderson, M., Grubinger, M.: Overview of the ImageCLEFphoto 2008 Photographic Retrieval Task. In: Working Notes for the CLEF 2008 Workshop, 17-19 September, Aarhus, Denmark (2008)
2. Ayala, G., Domingo, J.: Spatial Size Distributions. Applications to Shape and Texture Analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, n. 4, pp. 1430--1442. (2001)
3. Belga News Agency, <http://www.belga.be>.
4. Chellapa, R., Chatterjee, S.: Classification of Textures using Gaussian Markov Random Fields. *IEEE Transactions on Acoustics Speech and Signal Processing*, vol. 33, pp. 959--963. (1985)
5. Chen, Y., Dougherty, E.R.: Gray-scale morphological granulometric texture classification. *Optical Engineering*, vol. 33, n. 8, pp. 2713--2722. (1994)
6. Clark, S., Curran, J.: Wide-coverage efficient statistical parsing with CCG and log-linear models. *Computational Linguistics*, vol. 33, n. 4, pp. 493--553. (2007)
7. Connors, R.W., Trivedi, M.M., Harlow, C. A.: Segmentation of high-resolution urban scene using texture operators. *Computer Vision, Graphics, and Image Processing*, vol. 25, n. 3, pp 273--310. (1984)
8. de Ves, E., Domingo, J., Ayala G., Zuccarello, P.: A novel Bayesian framework for relevance feedback in image content-based retrieval systems. *Pattern Recognition*, vol. 39, pp. 1622--1632. (2006)
9. Finkel, J.R., Manning, C.D.: Joint parsing and named entity recognition. In: Proceedings of Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics, pp: 326--334, Boulder, Colorado. Association for Computational Linguistics. (2009)
10. Fogel, I., Sagi, D.: Gabor filters as texture discriminator. *Biological Cybernetics*, vol. 61, n. 2, pp. 103--113. (1989)
11. Granados, R., Benavent, X., Garcia-Serrano, A., Goñi, J.M.: MIRACLE-FI at ImageCLEFphoto 2008: Experiences in merging text-based and content-based retrievals. In: Working Notes for the CLEF 2008 Workshop, 17-19 September, Aarhus, Denmark (2008)
12. Granados, R., García-Serrano, Ana., Goñi, J.M.: La herramienta IDRA (Indexing and Retrieving Automatically). Demostración en la XXV edición del Congreso Anual de la Sociedad Española para el Procesamiento del Lenguaje Natural 2009 (SEPLN'09). (2009)
13. Grubinger, M., Clough, P., Müller, H., Deselaers, T.: The IAPR TC-12 Benchmark: A New Evaluation Resource for Visual Information Systems. In: Proceedings of International Workshop OntoImage'2006 Language Resources for Content-Based Image Retrieval, held in conjunction with LREC'06, pp. 13--23, Genoa, Italy, 22 May 2006 (2006)
14. ImageCLEF 2009 Photo Retrieval Task, <http://www.imageclef.org/2009/photo>
15. Leon, T., Zuccarello, P., Ayala, G., de Ves, E., Domingo, J.: Applying logistic regression to relevance feedback in image retrieval systems, *Pattern Recognition*, vol. 40, pp. 2621--2632. (2007)
16. Manning, C.D., Raghavan, P., Schtze, H.: Introduction to information retrieval. Cambridge Univ Press New York, NY, USA. (2008)
17. Paramita, M., Sanderson, M., Clough, P.: Diversity in photo retrieval: overview of the ImageCLEFPhoto task 2009. CLEF working notes 2009, Corfu, Greece. (2009)
18. Smith, G., Burns, I.: Measuring texture classification algorithms. *Pattern Recognition Letters*, vol. 18, n. 14, pp. 1495--1501. (1997)
19. Text REtrieval Conference (TREC), <http://trec.nist.gov/>