

Multimedia Retrieval by Means of Merge of Results from Textual and Content Based Retrieval Subsystems

Ana García-Serrano³, Xaro Benavent², Ruben Granados¹,
Esther de Ves², and José Miguel Goñi¹

¹ Universidad Politécnica de Madrid, UPM

² Universidad de Valencia

³ Universidad Nacional de Educación a Distancia, UNED
agarcia@lsi.uned.es, xaro.benavent@uv.es

Abstract. The main goal of this paper it is to present our experiments in ImageCLEF 2009 Campaign (photo retrieval task). In 2008 we proved empirically that the Text-based Image Retrieval (TBIR) methods defeats the Content-based Image Retrieval CBIR “quality” of results, so this time we developed several experiments in which the CBIR helps the TBIR. The TBIR System [6] main improvement is the named-entity sub-module. In case of the CBIR system [3] the number of low-level features has been increased from the 68 component used at ImageCLEF 2008 up to 114 components, and only the Mahalanobis distance has been used. We propose an ad-hoc management of the topics delivered, and the generation of XML structures for 0.5 million captions of the photographs (corpus) delivered. Two different merging algorithms were developed and the third one tries to improve our previous cluster level results promoting the diversity. Our best run for precision metrics appeared in position 16th, in the 19th for MAP score, and for diversity value in position 11th, for a total of 84 submitted experiments. Our best and “only textual” experiment was the 6th one over 41.

Keywords: Information Retrieval, Textual-based Retrieval, Content-Based Image Retrieval, Merge Results Lists, Fusion, Indexing.

1 Introduction

The main goal of this paper it is to present our experiments in ImageCLEF 2009 Campaign (photo retrieval task) [9] with different merging algorithms applied to the results obtained from the two subsystems developed: the Textual and the Content based Image Retrieval ones (TBIR and CBIR). The global system includes our own implemented tool IDRA (InDexing and Retrieving Automatically) [6], and the Valencia University CBIR system [3, 4, 7].

Given that the paradigms are intrinsically different, and also that we proved empirically last year that the TBIR defeats the CBIR “quality” of results [5, 10], this time we were interested in experiments in which the CBIR helps the TBIR.

The IDRA or TBIR System main improvement of this year was the named-entity sub-module. In case of 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 campaign only the Mahalanobis distance has been used in our experiments.

Our participation at ImageCLEF 2009 propose an ad-hoc management of the topics delivered, and the generation of XML structures for 0.5 million captions of the photographs (corpus) included in the so-called Belga Collection and delivered for the ImageClef 2009 International Competition.

Two different merging algorithms were developed in order to fuse different results lists from visual or textual modules, from different textual indexations, and other: *MAXmerge* (the algorithm selects the results from the N lists which have a higher relevance value) 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 is increased). Finally in order to improve our previous cluster level results looking forward the diversity we implemented the *EQUImerge* (that selects the first result of each query for a different cluster, not selected yet).

A more detailed presentation of the experiments is included in the following.

2 TBIR, CBIR and MERGE Subsystems Descriptions

The global system includes our own implemented tool IDRA (InDexing and Retrieving Automatically), and the Valencia University CBIR system. 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 its input data (decided from our participation at ImageCLEF 2008 [5]).

2.1 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.

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 (as explained in the following), but it takes over 10 hours to obtain a results file for 119 queries (119 queries at cluster level).

The textual retrieval tasks (sequentially executed) are the following.

Text Extractor. Uses the JDOM Java API to identify the content of each of the tags of the captions XML files.

Preprocess. This component processes the text in two ways, characters with no statistical meaning like punctuation marks, are eliminated and stop-words detected, with a new constructed list.

XML Fields Selection. With this component, it is possible to select the desired XML tags of the captions files (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. We preprocessed the Belga collection in order to have the same format as in the CLEF08 campaign.

IDRA Index. This module indexes the caption by calculating the weights vectors for each one. Each vector is compounded by the TF-IDF weights values [8] of the different words in the collection. All the weights values of each vector are normalized using the Euclidean distance between the elements of the vector. Therefore, the IDRA Index process update the values for each one of the words appearing in the XML captions collection.

IDRA Search. For the query text, its weights vector is also calculated. 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. 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.

2.2 Content-Based Information

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 [4,7], 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, the first step at the Visual Retrieval system is extracting the visual features for all the images on the database and for each of the cluster query topic images. 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 used 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 because the rest of values were almost zero, 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 database, six texture features have been computed for this repository respectively. The first three ones use code from the implementation done by Smith and Burn in Meastex (trec.nist.gov); the rest have been implemented by the authors. The total of texture features builds a vector of 64 components: Gabor Convolution Energies, Gray Level Cooccurrence Matrix also known as Spatial Gray Level Dependence, and Gaussian Random Markov Fields. The granulometric distribution function that we have used here is not the raw distribution but the coefficients that result of fitting its plot with a B-spline basis. Finally, for the Spatial Size Distribution we 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 of the cluster images. In 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. This is very useful because of the broad differences among the low-level values. Therefore, only the Mahalanobis distance has been used in our experiments. These sorted lists are passed to the merging module of the global system.

2.3 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 three require trec_eval format for input results lists, which is the format required by the ImageCLEF organization to submit the definitive runs.

Since N different indexes are created for the collection and, consequently, the IDF is computed per index fraction rather than globally, we use the MAXMERGE algorithm to fuse together the N (configurable value) results lists obtained when a concrete queries file is launched against the N indexes 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).

As the topic is divided into different queries when textual description of clusters are given, it is necessary to post process the different result lists in order to produce a unique result list which aims to provide diversity as expected from a topic with described clusters. It is the case in which the EQUIMERGE algorithm selects the top results from each cluster result list, and builds a unique list, by selecting in each step a result from a different cluster, if it has not been yet selected. 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 selected results per topic is set up to 1000 ('max' is a configurable parameter).

The ENRICH merge algorithm 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:

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

where, *newRel* is the new relevance value in the merged list, *mainRel* is the relevance value in the main list, *supRel* is the relevance value in the support list and *posRel* is the position in the support list. 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 Preprocessing the Corpus

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. Image captions are provided without a specific format and we preprocess them in order to build a semi-structured XML description for each image, similar to the used in the ImageCLEFphoto08 task [2]. This format includes 8 tags (docno, title, description, notes, location, date, image and thumbnail). The Named Entities (NE) of the captions were tagged using a pipeline of taggers by C&C tools: tokenizer, Part Of Speech (POS) and NER taggers [1]. We focused on NEs referring to Locations, Persons and Organization.

The images of the database have been pre-processed for the Content-Based Image module because some of them have extra-information consisting on 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. There are different images formats, e.g. with two vertical color bands, or 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, some 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.

Analyzing both “topics_part1.txt” and “topics_part2.txt” task topics files, we built different queries files to be launched against the IDRA indexation. The different queries files constructed are explained in the following:

- [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).
- [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”.
- [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 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 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.

- [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.

4 Results

Five runs were submitted to the ImageCLEFphoto2009. Runs start from the launch of some of the queries files described in section 3 to five parts of the corpus and the five obtained results lists 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*” 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 named entities. [qf4] is launched to obtain the results list.

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), the cluster recall at 10 and 20, and the F-Measure. 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.

Table 1. Results for the submitted experiments

Run	MAP	R-Prec	Prec@10	Prec@20	RelRet	CR@10	CR@20	F-Measure
[run1]	43.78	51.39	82.00	81.80	16547	64.51	72.00	72.21
[run2]	42.25	50.11	80.00	81.00	16301	63.24	73.41	70.64
[run3]	42.33	50.12	80.80	81.40	16301	63.51	73.31	71.12
[run4]	27.84	36.96	47.40	48.90	13627	69.83	76.76	56.47
[run5]	17.33	29.53	23.60	26.30	13498	68.49	72.82	35.10
average	29.08	34.09	65.50	64.38	10940.5	54.67	62.35	58.48
best	50.64	56.43	84.80	83.20	19066	82.39	86.07	80.87

We can observe that [run1] is our best run for precision metrics (very similar to [run2] and [run3]), and appears in the 16th position in R-Precision classification and in the 19th in MAP one (from a total of 84 submitted experiments). Regarding the

diversity metrics (cluster recall at 10 and 20, CR@10 and CR@20), [run4] and [run5] obtain our best diversity values, appearing in position 11th in cluster recall classification. The last column, F-Measure metric, which combines both precision and diversity values, situates our best experiment in position 11th. Another of the classifications provided by the organizers, shows [run3] as the 6th best run for all the experiments (41) based just in textual information.

Comparing obtained results from experiments [run1] and [run2], we can see that not using CD (cluster description) tag from the topics is slightly 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 [run2] and [run3] are very similar, 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.

Experiments [run3] and [run4] are different in the way of constructing the second half of the queries (from topic 26 to 50). The evaluation of the results shows that [run3] obtains better precision results than [run4], 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, but adds noise for the precision one. Experiments [run4] and [run5] show how named entity additional information improves the diversity results, but make the precision ones worse.

5 Concluding Remarks

Our experiments in ImageCLEF 2009 Campaign (photo retrieval task) were developed in order to make the CBIR module “help” the TBIR. The TBIR System main improvement was the named-entity sub-module, but the experiments with this new feature seems not to improve our previous results. We should to analyze in deep this unexpected performance.

In case of 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 only the Mahalanobis distance has been used. We propose an ad-hoc management of the topics delivered, and the generation of XML structures for 0.5 million captions of the photographs (corpus) delivered.

Two different merging algorithms were developed and the third one tries to improve the diversity of our previous cluster level results.

For the five runs submitted we obtained middle results comparing with the other groups in the campaign: our best run for precision metrics appears in the 16th position in R-Precision classification (0.5139), near the best (0.5643), and in the 19th in MAP one (0.4378), quite better than average and not very far from the best (0.5064), from a total of 84 submitted experiments. Our best diversity values, appeared in position 11th in cluster recall classification. F-Measure based classification, combining both precision and diversity values, situates our best experiment in position 11th. Having into account just textual experiments, our best was the 6th one over 41. Table 1 shows that our best runs values are always over the average and not so far from best ones. Analysing results per topic set, the most important thing we note is that using as query

text the one from the relevant images captions ([qf5]) introduces noise and results get much worse (rank 14 to 75 in provided “25 queries-part 2” classification).

Acknowledgments. Work partially supported by Spanish R+D Plan TIN2007-67407-C03-03 and TIC2002-03494; by Madrid’s R+D Plan MAVIR S-0505/TIC/000267.

References

1. Agerri, R., Granados, R., García-Serrano, A.: Extracting descriptions for Image Photo Retrieval. In: Working Notes 7th International Workshop in Adaptive Multimedia Retrieval (AMR 2009), UNED, Madrid (2009), <http://nlp.uned.es/amr2009/24-25/09/>
2. Arni, T., Clough, P., Sanderson, M., Grubinger, M.: Overview of the ImageCLEFphoto 2008 Photographic Retrieval Task. In: Peters, C., Deselaers, T., Ferro, N., Gonzalo, J., Jones, G.J.F., Kurimo, M., Mandl, T., Peñas, A., Petras, V. (eds.) CLEF 2008. LNCS, vol. 5706, pp. 500–511. Springer, Heidelberg (2009)
3. Ayala, G., Domingo, J.: Spatial Size Distributions. Applications to Shape and Texture Analysis. *IEEE Trans. on Pattern Anal. and Machine Intelligence* 23(4), 1430–1442 (2001)
4. de Ves, E., Domingo, J., Ayala, G., Zuccarello, P.: A novel Bayesian framework for relevance feedback in image content-based retrieval systems. *Pattern Recognition* 39, 1622–1632 (2006)
5. Granados, R., Benavent, X., Garcia-Serrano, A., Goñi, J.M.: MIRACLE-FI at ImageCLEFphoto 2008: Some Results Using Different Approaches to Merge Visual and Text-Based Features. In: Peters, C., Deselaers, T., Ferro, N., Gonzalo, J., Jones, G.J.F., Kurimo, M., Mandl, T., Peñas, A., Petras, V. (eds.) CLEF 2008. LNCS, vol. 5706, pp. 568–571. Springer, Heidelberg (2009)
6. Granados, R., García-Serrano, A., Goñi, J.M.: La herramienta IDRA (Indexing and Retrieving Automatically). In: Congreso SEPLN 2009 (2009)
7. Leon, T., Zuccarello, P., Ayala, G., de Ves, E., Domingo, J.: Applying logistic regression to relevance feedback in image retrieval systems. *Pattern Recognition* 40, 2621–2632 (2007)
8. Manning, C.D., Raghavan, P., Schtze, H.: Introduction to information retrieval. Cambridge Univ. Press, New York (2008)
9. Lestari Paramita, M., Sanderson, M., Clough, P.: Diversity in photo retrieval: overview of the ImageCLEFPhoto task 2009. In: CLEF 2009 Workshop, Part II. LNCS, vol. 6242, Springer, Heidelberg (2010)
10. Villena-Roman, J., Lana-Serrano, S., Martínez-Fernandez, J.L., Gonzalez-Cristobal, J.C.: MIRACLE at ImageCLEFphoto 2007: Evaluation of Merging Strategies for Multilingual and Multimedia Information Retrieval. In: Peters, C., Jijkoun, V., Mandl, T., Müller, H., Oard, D.W., Peñas, A., Petras, V., Santos, D. (eds.) CLEF 2007. LNCS, vol. 5152, pp. 500–503. Springer, Heidelberg (2008)