DAEDALUS at ImageCLEF Medical Retrieval 2011: Textual, Visual and Multimodal Experiments

Sara Lana-Serrano[1,3], Julio Villena-Román[2,3], José Carlos González-Cristóbal[1,3]

[1] Universidad Politécnica de Madrid
[2] Universidad Carlos III de Madrid
[3] DAEDALUS - Data, Decisions and Language, S.A.
slana@diatel.upm.es, jvillena@it.uc3m.es, josecarlos.gonzalez@upm.es

Abstract. This paper describes the participation of DAEDALUS at ImageCLEF 2011 Medical Retrieval task. We have focused on multimodal (or mixed) experiments that combine textual and visual retrieval. The main objective of our research has been to evaluate the effect on the medical retrieval process of the existence of an extended corpus that is annotated with the image type, associated to both the image itself and also to its textual description. For this purpose, an image classifier has been developed to tag each document with its class (1st level of the hierarchy: Radiology, Microscopy, Photograph, Graphic, Other) and subclass (2nd level: AN, CT, MR, etc.). For the textual-based experiments, several runs using different semantic expansion techniques have been performed. For the visual-based retrieval, different runs are defined by the corpus used in the retrieval process and the strategy for obtaining the class and/or subclass. The best results are achieved in runs that make use of the image subclass based on the classification of the sample images. Although different multimodal strategies have been submitted, none of them has shown to be able to provide results that are at least comparable to the ones achieved by the textual retrieval alone. We believe that we have been unable to find a metric for the assessment of the relevance of the results provided by the visual and textual processes.

Keywords: Image retrieval, domain-specific vocabulary, ontology, semantic expansion, indexing, context, image classification, multimodal, visual, textual.

1 Introduction

This paper describes the participation of DAEDALUS research team at ImageCLEF Medical Retrieval task [1] of ImageCLEF 2011. Last campaign, our research goal was to compare among different query expansion techniques using different approaches: methods based on linguistic information such as thesauri or knowledge bases, and statistical techniques based on term frequency [2]. Those experiments, in turn, were continuing the research line that was opened in previous campaigns [3] [4]. However, in spite of all our efforts, our best run was the baseline experiment.

This year we have focused on mixed experiments that combine textual and visual retrieval. The main objective of our research has been to evaluate the effect on the medical retrieval process of the existence of an extended corpus that is annotated with the image type, associated to both the image itself and also to its textual description. For this purpose, we have developed an image classifier to obtain the class (or classification label) for each image in the test corpus.

In the following sections we will describe our approach, the experiments that we submitted, the results achieved, and some preliminary conclusions.

2 Description of the System

The architecture of our system is composed of five different modules:

- the **expander module**, which performs the expansion of the content of textual documents and/or topics with related terms using textual algorithms;
- the **textual** (text-based) **retrieval module**, which indexes descriptions in order to search and find the most relevant ones to the text of the topic;
- the **visual** (image-based) **retrieval module**, in charge of the indexing and retrieval of images;
- the **visual classifier**, used to determine the class that corresponds to a given image;
- the **result combination module**, which uses different operators to combine, if necessary, the result lists provided by the previous retrieval subsystems**.**

For the textual retrieval process, several semantic expansion techniques have been applied: image descriptions and topics are parsed and tagged using the UMLS-based terminological dictionary [5] to identify and disambiguate medical terms, and semantic expansion with MeSH concept hierarchy [6] using the UMLS entities detected in document and topics as basic root elements to expand with their hyponyms (i.e., other entities whose semantic range is included within that of the root entity).

Lucene [7] has been used as the information retrieval engine for the whole textual indexing and retrieval task.

A specific adhoc engine has been developed for the visual retrieval module. This engine determines the most relevant images given a sample image and, optionally, a set of classes and/or subclasses that filter out the type of images to retrieve. The image classifier and the visual retrieval module have been implemented based on LIRE [8].

All documents contained in the corpus for both the image retrieval and the textual retrieval processes have been tagged along with its class (first level of the classification hierarchy: Radiology, Microscopy, Photograph, Graphic, Other) and subclass (second level of the hierarchy: AN, CT, MR, etc.).

Several experiments have been carried out using two different types of tagging, for comparison. In the first type, just the information provided by the organizers as part of the task has been used. In the second type, the tagging has been done using the output provided by our own image classification module.

3 Textual-based Runs and Results

Several experiments using different semantic expansion techniques have been performed. In all of them, the input for the retrieval process was both the topic, expanded depending on the experiment, and also the image class to retrieve, obtained by means of an analysis of the textual topic.

Table 1 shows a description of the submitted experiments.

Table 1. Description of textual runs.

Run Id	*Description*
BasTxtC	baseline (lowercase + stemming + stopword) and corpus tagged using provided classes.
BatTxtC_MC	baseline (lowercase + stemming + stopword) and corpus tagged using computed classes.
SemAC	semantic annotation with UML and corpus tagged using provided classes.
SemAC_MC	semantic annotation with UML and corpus tagged using computed classes.
SemEC	semantic annotation with UML and MeSH and corpus tagged using provided classes.
SemEC_MC	semantic annotation with UML and MeSH and corpus tagged using computed classes.

Table 2 shows the values of MAP, R-Precision, P_5, P_10 and P_15 for each of the submitted experiments. It can be noticed that the inclusion of semantic expansion tends to improve the overall results.

Table 2. Results of textual runs.

Run Id	*MAP*	*Rprec*	*P_5*	*P_10*	*P_15*
BasTxtC	**0.1966**	0.2668	0.4200	0.3900	0.3778
BasTxtC_MC	0.1918	0.2607	0.4067	0.3867	**0.3800**
SemAC	0.1818	0.2637	0.4267	0.3767	0.3667
SemAC_MC	0.1859	0.2569	0.4133	0.3833	0.3511
SemEC	0.1906	**0.2868**	0,4400	0.3867	0.3756
SemEC_MC	**0.1955**	**0.2795**	**0.4267**	**0.4000**	0.3644

The following figures (Figures 1, 2 and 3) present a detailed analysis of the results, grouped according to the type of the topic (visual, mixed or semantic). In general, results associated to semantic topics are better for all experiments. The worst results are achieved in mixed topics: this is because no relevant document was retrieved for topics 12 and 19 in any of the experiments. This issue has to be further studied.

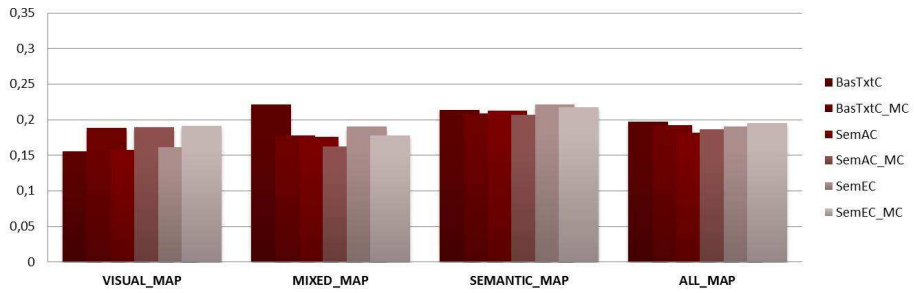

Figure 1. MAP values, by topic type.

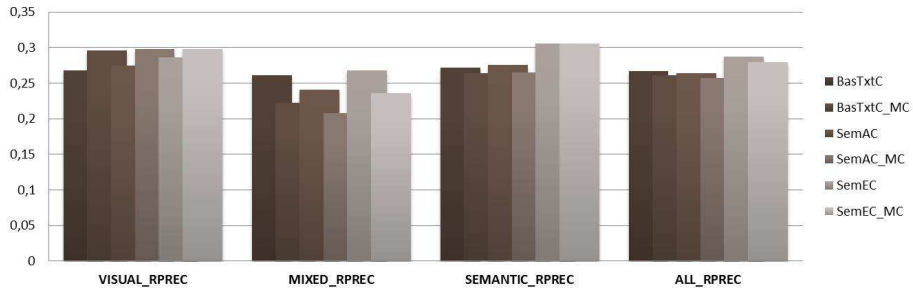

Figure 2. R-Precision values, by topic type.

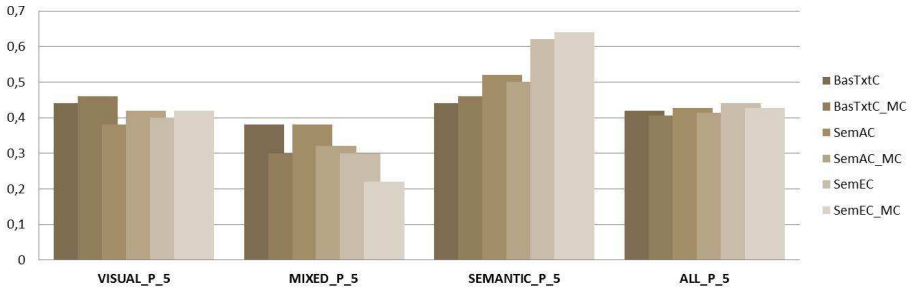

Figure 3. P_5 values, by topic type.

4 Image-based Runs and Results

Two different strategies have been adopted in our visual experiments to determine the class and/or subclass of the retrieved documents: in the first one, this information is extracted from an adhoc analysis of the textual topic, whereas in the second strategy, this information is inferred from the classification of the topic images.

The different experiments are dependent on the corpus used in the retrieval process and the strategy for obtaining the class and/or subclass. Table 3 shows a description of the finally submitted runs.

Table 3. Description of visual runs.

Run Id	*Description*
BasIm	No information about class/subclass is used
Img_C	class extracted from textual topic and corpus tagged using provided classes.
Img_SC	subclass extracted from textual topic and corpus tagged using provided classes.
Img_C_MC	class extracted from textual topic and corpus tagged using computed classes
ImgSC_MC	subclass extracted from textual topic and corpus tagged using computed subclass.
Img_C_MCMI	class computed from topic's images and corpus tagged using computed classes.
Img_SC_MCMI	subclass computed from topic's images and corpus tagged using computed subclass.

The following table shows the values of MAP, R-Precision, P_5, P_10 and P_15 for each of the submitted experiments. It can be noticed that the best results are achieved in runs that make use of the image subclass based on the classification of the example images (the so-called “computed subclass”).

Table 4. Results of visual runs.

Run Id	*MAP*	*Rprec*	*P_5*	*P_10*	*P_15*
BasImg	0.0125	0.0397	0.0867	0.0733	0.0667
ImgC	0.0139	0.042	0.0867	0.0733	0.0778
ImgC_MC	0.0147	0.0471	0.0867	**0.0967**	0.0889
ImgC_MCCI	0.0147	0.0471	0.0867	**0.0967**	0.0889
ImgSC	0.014	0.038	**0.1133**	**0.0967**	0.0867
ImgSC_MC	**0.017**	**0.0473**	0.100	0.0933	**0.0933**
SC_MCCI	**0.017**	**0.0473**	0.100	0.0933	**0.0933**

As shown in both our own and also in other participants’ experiments, the performance of the runs based only in visual retrieval is really poor as compared to the results achieved by the textual retrieval.

5 Mixed-based Runs and Results

After the release of results by the organizers of the task, we realised that our combination process in the mixed-based runs had a severe bug that invalidated our results. Thus we fixed that bug, repeated the same experiments and evaluated them.

Table 5 shows the main statistics associated to experiments that achieve the best performance. Those results correspond to the combination of the best textual experiments along with the best visual experiments. The run identifier for the mixed experiments has the following pattern: <TextualRunId>#<VisualRunId>.

Table 5. Results of mixed runs.

Run Id	*MAP*	*Rprec*	*P_5*	*P_10*	*P_15*
SemEC#SC	0.1281	0.2101	0.3467	0.2900	0.2622
SemEC#SC_MC	0.1384	0.2167	0.3733	0.3100	0.2867
SemEC#SC_MCCI	0.1384	0.2167	0.3733	0.3100	0.2867
SemEC_MC#SC	0.124	0.1883	0.3333	0.2800	0.2622
SemEC_MC#SC_MC	0.1336	0.1973	0.3533	0.3100	0.2822
SemEC_MC#SC_MCCI	0.1336	0.1973	0.3533	0.3100	02822

Even though a large series of different combinations have been performed, none of them has shown to be able to provide results that are at least comparable to the ones achieved by the textual retrieval alone. We believe that this is because of the fact that we have been unable to find a metric for the assessment of the relevance of the results provided by the visual and the textual retrieval processes. In other words, we are unable to sort out if a result image (whether visual or textual) is appropriate or not for its inclusion in the final result list. The effect is that the combination of both approaches, instead of improving the overall performance, decreases it.

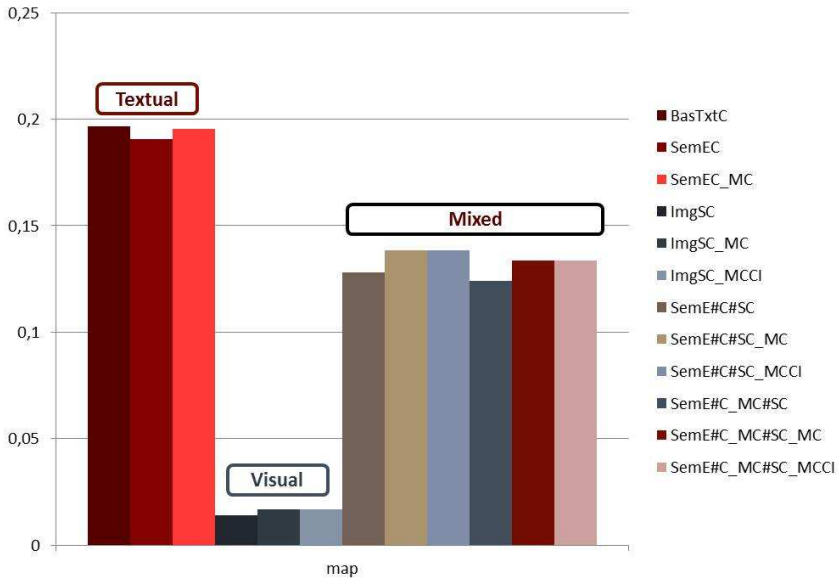

Figure 4. MAP- Mixed retrieval vs. Textual and Visual retrieval

6 Conclusions and Future Work

Considering the results achieved by our own experiments and also by other participants in the task, it can be concluded that, for this scenario, the application of retrieval techniques based exclusively on visual content provides poor results in comparison to textual techniques. We think that the best strategy to improve this kind of engines is to incorporate advanced techniques for the extraction and characterization of semantic information within visual resources. However, to be able to deal with this kind of solutions, it is necessary to have access to large database of semantically annotated resources whose cost (both economic and computational) is not affordable.

Another issue that must be tackled is to find metrics that allow to compare the relevance of a document resulting from the textual retrieval process along with another document returned by the visual retrieval process. Once we find this metric, we will be able to define mixed retrieval processes that actually improve the results over the results achieved by textual and visual retrieval independently.

Acknowledgements

This work has been partially supported by several Spanish research projects: MA2VICMR: Improving the access, analysis and visibility of the multilingual and multimedia information in web for the Region of Madrid (S2009/TIC-1542), MULTIMEDICA: Multilingual Information Extraction in Health domain and application to scientific and informative documents (TIN2010-20644-C03-01) and BUSCAMEDIA: Towards a semantic adaptation of multi-network-multiterminal digital media (CEN-20091026). Authors would like to thank all partners for their knowledge and support.

References

1. Jayashree Kalpathy-Cramer, Henning Müller, Steven Bedrick, Ivan Eggel, Alba Garcia Seco de Herrera, Theodora Tsikrika, The CLEF 2011 Medical Image Retrieval and Classification Tasks, *CLEF 2011 working notes*, Amsterdam, The Netherlands, 2011.
2. Lana-Serrano, Sara; Villena-Román, Julio; González-Cristóbal, José Carlos. MIRACLE at ImageCLEFmed 2008: Semantic vs. Statistical Strategies for Topic Expansion. *Evaluating Systems for Multilingual and Multimodal Information Access 9th Workshop of the Cross-Language Evaluation Forum*, CLEF 2008, Aarhus, Denmark, September 17-19, 2008, Revised Selected Papers. Peters, Carol et al (Eds.). Lecture Notes in Computer Science, 2008 (printed in 2009).
3. Villena-Román, Julio; Lana-Serrano, Sara; González-Cristóbal, José Carlos. MIRACLE at ImageCLEFmed 2007: Merging Textual and Visual Strategies to Improve Medical Image Retrieval. *Advances in Multilingual and Multimodal Information Retrieval. 8th Workshop of the Cross-Language Evaluation Forum*, CLEF 2007, Budapest, Hungary, Revised Selected Papers. Carol Peters et al

(Eds.). Lecture Notes in Computer Science, Vol. 5152, 2008. ISSN: 0302-9743/1611-3349.
4. Martínez-Fernández, José Luis; García-Serrano, Ana M.; Villena-Román, Julio; Martínez-Fernández, Paloma. Expanding Queries Through Word Sense Disambiguation. *Evaluation of Multilingual and Multi-modal Information Retrieval: 7th Workshop of the Cross-Language Evaluation Forum*, CLEF 2006, Alicante, Spain, September 20-22, 2006, Revised Selected Papers. Carol Peters et al. (Eds.). Lecture Notes in Computer Science , Vol. 4730, 2007. ISSN: 0302-9743
5. U.S. National Library of Medicine. National Institutes of Health. Unified Medical Language System (UMLS). http://www.nlm.nih.gov/research/umls/.
6. U.S. National Library of Medicine. National Institutes of Health. Medical Subject Headings (MeSH). http://www.nlm.nih.gov/mesh/.
7. Apache Lucene project. http://lucene.apache.org/.
8. LIRE. Lucene Image Retrieval. http://www.semanticmetadata.net/lire/.