

# SINAI at ImageCLEF 2007

M.C. Díaz-Galiano, M.A. García-Cumbreras, M.T. Martín-Valdivia,  
A. Montejo-Raez, L.A. Ureña-López  
University of Jaén. Departamento de Informática  
Grupo Sistemas Inteligentes de Acceso a la Información  
Campus Las Lagunillas, Ed. A3, E-23071, Jaén, Spain  
{mcdiaz,magc,maite,amontejo,laurena}@ujaen.es

## Abstract

This paper describes the SINAI team participation in the ImageCLEF campaign. The SINAI research group has participated in both the ad hoc task and the medical task. The experiments accomplished in both cases result from very different approaches.

For the ad hoc task the main Information Retrieval (IR) system used combines the document lists retrieved by two IR systems, and uses online translators for the bilingual experiments.

For the medical task, we have used the MeSH ontology to expand the queries. The expansion consists in searching terms of the query in the MeSH ontology in order to add similar terms. We have processed the set of collections using Information Gain (IG) in the same way as in ImageCLEFmed 2006.

## Categories and Subject Descriptors

H.3 [Information Storage and Retrieval]: H.3.1 Content Analysis and Indexing; H.3.3 Information Search and Retrieval; H.3.4 Systems and Software

## Keywords

Visual and textual retrieval, Information Gain, Indexing, Machine Translators, Ontologies, MeSH

## 1 Introduction

This is the third participation of the SINAI research group at the ImageCLEF campaign. We have participated in the ad hoc task [6] and the medical task [7].

The ad hoc task involves retrieving relevant images using the text associated to each image query. As a cross-language retrieval task, multilingual image retrieval based on query translation can achieve a higher performance than monolingual retrieval.

This year, a new IR module has been tested. This module works with two different IR systems and the final relevant list is the result of the combination of both IR lists. The Machine Translation Module developed last year has been updated and used for the bilingual task. English, Spanish, French, Italian and Portuguese are the languages used this year.

The goal of the medical task is to retrieve relevant images based on an image query. This year, two new collections have been introduced. We have filtered all the collections using IG to select the best tags of each one [3]. Moreover, we have expanded the queries using MeSH ontology: we have selected similar terms to the query in the ontology and have added them to the query itself.

The following section describes the ad hoc experiments. In Section 3, we explain the experiments for the medical task. Finally, conclusions and further work are presented in Section 4.

## 2 The Ad Hoc Task

Given a multilingual query, the goal of the ad hoc task is to find as many relevant images as possible from an image collection.

The proposal of the ad hoc task is to compare results with and without pseudo-relevant feedback (PRF), with or without query expansion, using different methods of query translation or using different retrieval models and weighing functions.

### 2.1 Experiments Description

In our experiments we have used the five following languages: English, French, Italian, Portuguese and Spanish.

This year we have combined lists of relevant documents returned by two different IR systems: Lemur<sup>1</sup> and Jirs [4].

As translation module we have used SINTRAM (SINai TRANslation Module), our Meta Machine Translation system that uses some online Machine Translators for each language pair and implements some heuristics to combine the different translations [5]. After a complete research we have found that the best translators were:

- Systran for French, Italian and Portuguese
- Prompt for Spanish

The dataset is the collection IAPR. The IAPR TC-12 image collection consists of 20,000 images taken from different locations around the world and comprises a varying cross-section of still natural images. It includes pictures of a range of sports and actions, photographs of people, animals, cities, landscapes and many others of contemporary life.

The collections have been preprocessed using stopwords removal and the Porter's stemmer.

The dataset of the collection has been indexed using both IR systems, namely, Lemur and Jirs. One parameter for each experiment is the weighing function, such as Okapi or TFIDF. Another is the use or not of PRF.

A simple fusion method has been implemented to obtain a simple list of relevant documents. In the first step, both lists are normalized between 0 and 1. Then, some heuristics are implemented:

- Weighing each list. Some experiments are based on a weighing function that gives a percentage of importance to the Lemur list and a different one to the Jirs list. The final score of each relevant document is calculated by the sum of each score multiplied by its weight. Finally, the documents are sorted by their final fusion score.
- Using a threshold. Filtering relevant documents by a threshold value is another heuristic. If the score of a document is worse than this parameter, it is not included in the final list. Finally, the resultant list is sorted by the score of the included documents.

With the ad hoc 2006 framework (using the same collection, queries and relevance judgements) [1] and the heuristics already described we have evaluated several configurations in order to obtain the best ones.

1. The basic case, with Lemur, uses English queries, Lemur as IR system and Okapi with PRF as weighing function. It obtains a MAP value of 0.1672
2. The basic case, with Jirs, uses English queries, Jirs as IR system and Okapi with PRF as weighing function. It obtains a MAP value of 0.1513
3. For the first heuristic the weight for Lemur and Jirs changes between 1 and 0.1. For instance, the experiment that weighs both lists in the same way uses 0.5 as the weight value for both Lemur and Jirs. The best result was 0.1678, using a weight of 0.6 for Lemur and 0.4 for Jirs.

---

<sup>1</sup><http://www.lemurproject.org>

Language	Experiment	IR	Expansion	Weight	MAP	Best MAP
English	EN-EN-Exp2	Lemur	without	Okapi	0.1591	0.2075
English	EN-EN-Exp1	Jirs	without	Okapi	0.1473	0.2075
Spanish	ES-EN-Exp9	Jirs	without	Okapi	0.1555	0.1558
Spanish	ES-EN-Exp10	Lemur	without	Okapi	0.1498	0.1558
Portuguese	PO-EN-Exp8	Lemur	without	Okapi	0.1490	0.1490
Portuguese	PO-EN-Exp7	Jirs	without	Okapi	0.1350	0.1490
French	FR-EN-Exp4	Lemur	without	Okapi	0.1264	0.1362
French	FR-EN-Exp3	Jirs	without	Okapi	0.1195	0.1362
Italian	IT-EN-Exp5	Jirs	without	Okapi	0.1231	0.1341
Italian	IT-EN-Exp6	Lemur	without	Okapi	0.1198	0.1341

Table 1: Summary of results for the photo task: Monolingual and bilingual runs with Lemur and Jirs IR systems

Language	Experiment	IR	Expansion	Weight	MAP	Best MAP
English	EN-EN-Exp11	Fusion	without	Okapi	0.0786	0.2075
Spanish	ES-EN-Exp15	Fusion	without	Okapi	0.0559	0.1558
Portuguese	PO-EN-Exp14	Fusion	without	Okapi	0.0423	0.1490
French	FR-EN-Exp12	Fusion	without	Okapi	0.0323	0.1362
Italian	IT-EN-Exp13	Fusion	without	Okapi	0.0492	0.1341

Table 2: Summary of results for the photo task: Monolingual and bilingual runs with lists fusion.

- In the case of the second heuristic, different values, from 0.1 to 0.9, are tested as threshold. The best result was 0.1524, obtained with 0.1 value as threshold.

Finally, all the Lemur weighing functions have been tested and the best one was again Okapi with feedback, with a result of 0.1672.

## 2.2 Results and Discussion

With the results obtained with the 2006 framework, the new 2007 queries were run. We sent 15 runs: five runs using Lemur, five using Jirs and five with the fusion of both lists.

The results obtained with each IR system (using only text) and the best MAP for each language is shown in Table 1.

Good results have been obtained this year with both IR systems. Only the English runs have obtained a loss of MAP of around 25%. Our best Spanish result is similar to the best one obtained. For Portuguese we have obtained the best one, and for French and Italian our results are a bit worse: only a loss of MAP of around 8%. The MAP values are low because only the title is used as query.

From these results we can conclude that Lemur IR system works better than Jirs, but the difference is not very significant.

The results obtained by applying the fusion method and the best MAP for each language is shown in Table 2.

The main conclusion for these fusion results is that there may be a problem, because it is not logical to obtain so poor results.

Experiments with the 2006 framework gave us similar results between the ones obtained with each simple IR system and the fusion ones.

## 3 The Medical Task

The main goal of the medical ImageCLEF task is to improve the retrieval of medical images from heterogeneous and multilingual document collections containing images and text. Queries are

formulated with sample images and some textual description explaining the research goal. For the medical task we have used the list of retrieved images by FIRE<sup>2</sup> [2], which was supplied by the organizers of this track.

Last year, our efforts concentrated on manipulating the text descriptions associated with these images and mixing the results partial lists with the GIFT lists [3]. We also focused on preprocessing the collection using Information Gain (IG) in order to improve the quality of results and to automatize the tag selection process. However, this year we have concentrated on improving the queries using MeSH ontology.

### 3.1 Preprocessing the Collection

In order to generate the textual collection we have used the ImageCLEFmed.xml file that links collections with their images and annotations. It has external links to the images and to the associated annotations in XML files. It contains relative paths from the root directory to all the related files.

The entire collection consists of six datasets (CASImage, Pathopic, Peir, MIR, endoscopic and MyPACS) containing about 66,600 images (16,600 more than the previous year). Each subcollection is organized into cases that represent a group of related images and annotations. In every case a group of images and an optional annotation is given. Each image is part of a case and has optional associated annotations, which enclose metadata and/or a textual annotation. All the images and annotations are stored into separated files. ImageCLEFmed.xml only contains the connections between collections, cases, images and annotations.

The collection annotations are in XML format and most of them are in English. We have preprocessed the collections to generate a textual document per image [3].

We have used the IG measure to select the best XML tags in the collection. Once the document collection was generated, experiments were conducted with the Lemur retrieval information system by applying the KL-divergence weighing scheme.

### 3.2 Expanding Queries with MeSH Ontology

The Medical Subject Headings (MeSH) is a thesaurus developed by the National Library of Medicine<sup>3</sup>. MeSH contains two organization files, an alphabetic list with bags of synonymous and related terms, and a hierarchical organization of descriptors associated to the terms. A term is composed by one or more words.

We have used the bags of terms to expand the queries. If all the words of a term are in the query, we generate a new expanded query by adding all its bag of terms. To compare the words of a particular term and those of the query, we first put all the words in lowercase and we do not remove stopwords. In order to reduce the number of terms that could expand the query, we have only used those that are in A, C or E categories of MeSH (A: Anatomy, C: Diseases, E: Analytical, Diagnostic and Therapeutic Techniques and Equipment) [8].

### 3.3 Experiment Description

Our main objective is to investigate the effectiveness of the query expansion together with filtering tags using IG in the text collection. We have carried out these experiments using a corpus with 20%, 30%, 40%, 50% and 60% of tags for the 2007 collection, because these settings led to the best results on the 2006 corpus.

Finally, the expanded textual list and the FIRE list are merged in order to obtain one final list (FL) with relevant images ranked by relevance. The merging process was done by giving different importance to the visual (VL) and textual lists (TL):

$$FL = TL * \alpha + VL * (1 - \alpha) \quad (1)$$

---

<sup>2</sup><http://www-i6.informatik.rwth-aachen.de/~deselaers/fire.html>

<sup>3</sup><http://www.nlm.nih.gov/mesh/>

Experiment	Precision
LIG-MRIM-LIG_MU_A.eval	0.3962
<b>SINAI-SinaiC100T80.eval</b>	<b>0.3716</b>
miracleTxtENN.txt.eval	0.3518
OHSU-oshu_as_is_1000.eval	0.3453
UB-NLM-UBTI_1.eval	0.3230
IPAL-IPAL1_TXT_BAY_ISA0.1.eval	0.3057
RWTH-FIRE-ME-tr0506.eval	0.3044
GE_EN.treceval.eval	0.2714
iclefmed2007_text2_out.txt.eval	0.1976
UNALCO-nni_FeatComb.eval	0.0082
DEU_CS-DEU_R2.eval	0.0028

Table 3: Performance of official runs in Medical Image Retrieval

Experiment	Precision
SinaiC20T100	0.2950
SinaiC30T100	0.3340
SinaiC40T100	0.3507
SinaiC50T100	0.3487
SinaiC60T100	0.3452
SinaiC100T100	0.3668

Table 4: Performance of only textual information runs in Medical Image Retrieval

We have submitted runs with  $\alpha$  set to 0.5, 0.6, 0.7 and 0.8.

The baseline experiments contain the 100% of the tags. To compare the mixed results we have accomplished experiments with an  $\alpha = 1$ .

### 3.4 Results and Discussion

The total runs submitted to ImageCLEFmed2007 for textual retrieval and mixed retrieval were more than 100.

Table 3 shows the best results for the groups participating in ImageCLEFmed2007. In this table we can observe that the best result obtained by our system is the experiment with 100% of the tags and  $\alpha = 0.8$  (80% of textual information).

In Table 4 we can see the results of only textual experiments. The best result obtained is 0.3668 of precision value, using 100% of tags in the collection. Surprisingly, the IG selection experiments do not improve the results. However, by using 40% of tags the loss of precision is lower than 5%.

## 4 Conclusions and Further Work

This year the ad hoc task obtained good results, but always with a low MAP, because only the title was used. It could be interesting to develop a new fusion module and to apply a query expansion module based on Google, tasks on which we are already working.

This year two new collections have been included in the ImageCLEFmed2007. Adding this new information in the collection improves the results obtained with the Lemur IR system. We want to investigate what type of query (textual, visual or mixed) is influenced the most by these new collections. Moreover, our next step will focus on using the UMLS ontology<sup>4</sup>, in order to include the multilingual features of the collections.

<sup>4</sup><http://www.nlm.nih.gov/pubs/factsheets/umls.html>

## 5 Acknowledgements

This project has been partially supported by a grant from the Spanish Government, project TIMOM (TIN2006-15265-C06-03).

## References

- [1] Clough, P., Grubinger, M., Deselaers, T., Hanbury, A., and Müller, H.: Overview of the ImageCLEF 2006 Photographic Retrieval and Object Annotation Tasks. In Proceedings of the Cross Language Evaluation Forum (CLEF 2006), 2006.
- [2] Deselaers, T., Weyand, T., Keysers, D., Macherey, W., and Ney, H.: FIRE in ImageCLEF 2005: Combining Content-based Image Retrieval with Textual Information Retrieval. Working Notes of the CLEF Workshop, Vienna, Austria, September 2005.
- [3] Díaz-Galiano, M.C., García-Cumbreras, M.A., Martín-Valdivia, M.T., Montejo-Raez, A., and Ureña-López, L.A.: SINAI at ImageCLEF 2006. In Proceedings of the Cross Language Evaluation Forum (CLEF 2006), 2006.
- [4] Gómez-Soriano, J.M., Montes-y-Gómez, M., Sanchis-Arnal, E., and Rosso, P.: A Passage Retrieval System for Multilingual Question Answering. 8th International Conference of Text, Speech and Dialogue 2005 (TSD'05). Lecture Notes in Artificial Intelligence (LNCS/LNAI 3658). pp. 443-450. Karlovy Vary, Czech Republic. 2005.
- [5] García-Cumbreras, M.A., Ureña-López, L.A., Martínez-Santiago, F., and Perea-Ortega, J.M.: BRUJA System. The University of Jaén at the Spanish task of QA@CLEF 2006. In Proceedings of the Cross Language Evaluation Forum (CLEF 2006), 2006.
- [6] Grubinger, M., Clough, P., Hanbury, A., and Müller, H.: Overview of the ImageCLEF 2007 Photographic Retrieval Task. Working Notes of the 2007 CLEF Workshop. Sep, 2007. Budapest, Hungary.
- [7] Müller, H., Deselaers, T., Kim, E., Kalpathy-Cramer, J., Deserno, T.M., Clough, P., and Hersh, W.: Overview of the ImageCLEFmed 2007 Medical Retrieval and Annotation Tasks. Working Notes of the 2007 CLEF Workshop. Sep, 2007. Budapest, Hungary.
- [8] Chevallet, J.P., Lim, J.H., and Radhouani, S.: Using Ontology Dimensions and Negative Expansion to solve Precise Queries in CLEF Medical Task. Working Notes of the 2005 CLEF Workshop. Sep, 2005. Vienna, Austria.