Text-mess in the Medical Retrieval ImageCLEF08 Task

S. Navarro[1], M.C. Díaz[2], R. Muñoz[3], M.A. García[4], F. Llopis[5],
M.T. Martín[6], L.A. Ureña[7], A. Montejo[8]
[1,3,5]Natural Language Processing and Information Systems Group. University of Alicante, Spain
[2,4,6,7,8]Sistemas Inteligentes de Acceso a la Información, SINAI Group. University of Jaén, Spain
{snavarro[1],rafael[3],llopis[5]}@dlsi.ua.es
{mcdiaz[2],magc[4],maite[6],laurena[7],amontejo[8]}@ujaen.es

Abstract

This paper describes our participation in the Medical Retrieval task at ImageCLEF 2008. We present the joint work of two teams belonging to the TEXT-MESS project using a new system that combines the 2 individual systems of these teams. The aim of the experiments performed is to figure out if there are techniques used in one of the two systems which can complement the other system in order to improve their performance. The best results obtained in the training phase and in the competition has been reached with a configuration which uses the IR-n system with a negative query expansion based on the acquisition type of the image mixed with the SINAI system with a MeSH based query expansion. We have obtained a MAP of 0.2777 for our best run, obtaining the 5th place in the ranking of textual participant runs submitted, and the 6th place in the global classification.

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 Managment**]: H.2.5 Heterogenous Databases

General Terms

Measurement, Performance, Experimentation

Keywords

Information Retrieval, Image Retrieval, Multimodal Reranking, Late Fusion, Multimodal Relevance Feedback, PRF, LCA, MeSH, Automatic query expansion, Negative query expansion

1 Introduction

This paper describes our participation in the Medical Retrieval task at ImageCLEF 2008. We present the joint work of two teams belonging to the TEXT-MESS project using a new system that combines the 2 individual systems of these teams. The aim of the experiments performed is to figure out if there are techniques used in one of the two systems which can complement the other system in order to improve their performance.

Our experiments has been focused on to use a reranking method combining different configurations of the individual systems. Moreover, we have made experiments with two different reranking methods - the standard one and TF-IDF multimodal reranking - in order to perform a multimodal participation with our joint systems.

This paper is structured as follows: Firstly, it presents the main characteristics of the SINAI and IR-n system, then it moves on to explain the experiments we have made to evaluate the join system, and finally it describes the results and conclusions.

2 The System

The complete system is composed by two systems adapted to the IR medical domain - SINAI and IR-n -. They work in a parallel mode. The ranking returned by each one of the systems is merged following a standard reranking method. Furthermore, for our participation in the mixed modality - runs that mix a visual and a textual approach - the system uses two modalities for multimodal reranking. The standard reranking method and the IR-n TF-IDF multimodal rerankig described in [6].

2.1 The SINAI System

SINAI system use Lemur software[1] to textual information retrieval. One of the purposes of the improvements added to this system is to compare the performance of query expansion using two different ontologies: MeSH and UMLS. Experiments with the MeSH ontology have been carried out in the past [1] obtaining good results. The expansion method using MeSH is the same as presented last year.

On the other hand, the UMLS metathesaurus is a repository of biomedical ontologies and associated software tools developed by the US National Library of Medicine(NLM)[2]. It is built from different *source vocabularies*. One of the source vocabularies is MeSH ontology.

To expand the queries we have used MetaMap program [5] that was originally developed for use in information retrieval. In order to reduce the number of terms that could expand the query, to make it equal to that of MeSH expansion, we have used MetaMap, restricting the semantic types in the mapped terms [2].

2.2 The IR-n System

IR-n is an information retrieval system based on passages. Those type of IR systems, unlike document-based systems, can consider the proximity of words with each other, that appear in a document in order to evaluate their relevance [4].

This system has added for its current participation in this task a common approach to the multimodal issue. It allows two working modes. The first one is the standard one for merging two lists, based on set values to the weighting factor of each list in order to create a joined list. The second one is the TF-IDF multimodal reranking, it is a variation of the standard one. It bases the calculus of the relevance of an image on the quantity and the quality of its annotations in order to decide whether the relevance value returned by the textual IR system is enough to rank a document or it is needed to add the relevance returned by a CBIR system [6].

Furthermore, in order to adapt the system to this restricted domain, it has used two automatic query expansion techniques related with the medical domain. On one hand it adds expanded terms to the queries based on MeSH ontology. And on the other hand when in the query there are terms related to the type of the images that have to be retrieved, the system uses a negative term expansion based on the terms related to other types of a taxonomy of types of images. In order to move away from the top positions of the ranking, those documents which do not belong to the type/s requested in the query [6].

Finally in this CLEF edition IR-n has added Local Context Analisy (LCA) [7] as alternative strategy to PRF, in order to compare its behaviour within this restricted domain.

[1]http://www.lemurproject.org

[2]http://www.nlm.nih.gov/

3 Training

This section describes the training process that was carried out in order to obtain the best possible features for improving the performance of the whole system. The collections and resources are described first, and the next section describes specific experiments.

3.1 Data Collections

For this year task, a new collection has been used in order to evaluate and to compare the participant systems. This collection is the Goldminer collection [3]. The subset used contains all images from articles published in Radiology and Radiographics including the text of the captions and a link to the html of the full text articles. We have used two versions of this textual collection for generate our submitted runs - which were obtained by SINAI group following the method described in [2] -:

- **CT**: It contains image captions and article titles.
- **CTS**: It contains captions, titles and texts of the sections where the images appear.

The training for our participation have been done with the Consolidated Collection. This collection consists the image collection and topics used in ImageCLEFmed 2005-2007 merged into one single new collection, with relevance judgments done for all topics based on all collections. We have used a preprocessed version of this collection which is compounded by a textual document per image, which has been translated to English when it has been need it [3] -.

3.2 Experiments

In the training phase, initially for each IR system, we have worked in the selection of those runs which better results obtained on its training phase. Next, in order to figure out which are the most suitable weighting values for the re-reanking module, we have carried out a training phase with each pair of selected runs in the previous step.

The runs selected using SINAI system have been the following:

- **OnlyText**: Baseline experiment.
- **OnlytextMeshSimAceSinrepe**: Query expanded with MeSH ontology.
- **OnlytextUmlsFiltered**: Query expanded with UMLS metathesaurus using MetaMap program.

The runs selected using IR-n system have been the following:

- **EXP_NEGATIVA**: It uses negative expansion based on the acquisition type of the image
- **EXP_NEGATIVA_MeSH**: It uses negative expansion based on the acquisition type of the image and query expansion based on MeSH ontology.
- **EXP_PRF_NEGATIVA_MESH**: It uses negative expansion based on the acquisition type of the image, query expansion based on MeSH ontology and PRF as relevance feedback strategy.

The Table 1, show us the MAP obtained for each standalone run and the MAP obtained with the best standard reranking run with the weights used.

We can see that the best results have been obtained combining the IR-n run which have used the negative expansion based on the acquisition type of the image, with the SINAI runs which have used MeSH or UMLS for the query expansion. We select these two configurations for our textual submissions of this year, - we have submitted 2 runs using the CT collection and 2 more using

[3] *http://goldminer.arrs.org*

Table 1: Best Textual Reranking Results

SINAI Run	IR-n Run	SINAI map	IR-n map	SINAI/IR-n weights	RR map
OnlyText	EXP_ NEGATIVA	0.2042	0.2174	0.3/0.7	0.2227
OnlyText	EXP_ NEGATIVA_MeSH	0.2042	0.2051	0.3/0.7	0.2135
OnlyText	EXP_ PRF_NEGATIVA_MESH	0.2042	0.2029	0.3/0.7	0.2144
Onlytext **SimAceSinrepe**	**EXP_ NEGATIVA**	**0.2204**	**0.2174**	**0.6/0.4**	**0.2299**
Onlytext MeshSimAceSinrepe	EXP_ NEGATIVA_MeSH	0.2204	0.2050	0.6/0.4	0.2051
Onlytext MeshSimAceSinrepe	EXP_ PRF_NEGATIVA_MESH	0.2204	0.2029	0.8/0.2	0.2170
Onlytext **UmlsFiltered**	**EXP_ NEGATIVA**	**0.1989**	**0.2175**	**0.3/0.7**	**0.2247**
Onlytext UmlsFiltered	EXP_ NEGATIVA_MeSH	0.1989	0.2051	0.3/0.7	0.2174
Onlytext UmlsFiltered	EXP_ PRF_NEGATIVA_MESH	0.1989	0.2029	0.3/0.7	0.2158

the CTS collection - . Furthermore we this two configurations to figure out the best parameters to perform a multimodal reranking in order to merge their output with the ranking list returned by a CBIR.

For this training phase we have used the University of Geneva CBIR baseline run for the 2007 query set - in the moment of the experiments we did not have available a CBIR baseline that answer to the whole consolidated query set -. Since that the reranking results have been obtained with runs that for the 2007 queries - 30 queries of a total of 85 queries - have used the reranking technique and for the other queries they have used a purely textual retrieval. Bearing that in mind we have evaluated the results of the reranking experiments as if its improvement or worsening was minimized by its partial using with the training query set.

The Table 2, shows us the resulting MAP obtained for the previous reranked runs, the MAP of the CBIR baseline run, the multimodal reranking strategy used to combine those runs, the best configuration of weighting values - for the standard reranking strategy - or the best threshold -for the TF-IDF reranking - and its MAP result.

Its important to stand out the extremely low threshold percentage value used in the best TF-IDF reranking run - specially if we compare it with the 60

For our participation in the task, we have submitted these 4 multimodal runs within these mixed modality - mixing a visual and a textual approach - using CT collection and 2 runs more with the best configuration using the collection CTS.

4 Results in ImageCLEF Medical Retrieval 2008

The Tables 3 and 4 shows us the official MAP results, the ranking position within the textual and the mixed modality respectively and the ranking position within all the participant runs.

In the Table 3 we can see that our best TEXT-MESS run in the text modality has been ranked in the 5th place. It is the same configuration that obtained the best results in our training phase. It uses IR-n with negative expansion based on the acquisition type of the image - Type - and

Table 2: Best Multimodal Reranking Results

MMRR **Strategy**	**RR Run**	**CBIR Run**	**RR map**	**GIFT map**	**RR/GIFT weights**	**MMRR map**
Standard	OnlytextUmlsFiltered + EXP_NEGATIVA0.3_0.7	GE_ GIFT4	0.2247	0.0474	0.6/0.4	0.2250
TF-IDF	OnlytextUmlsFiltered + EXP_NEGATIVA0.3_0.7	GE_ GIFT4	0.2247	0.0474	Th:0.00002	0.2251
Standard	**OnlytextMeshSimAceSinrepe + EXP_NEGATIVA0.6_0.4**	**GE_ GIFT4**	**0.2299**	**0.0397**	**0.9/0.1**	**0.2310**
TF-IDF	textMeshSimAceSinrepe + EXP_NEGATIVA0.6_0.4	GE_ GIFT4	0.2299	0.0397	Th:0.00002	0.2303

Table 3: Official Textual Runs Results

run name	**map**	**txt rk**	**CLEF rk**
TEXTMESSmeshType_CT	0.2777	5	6
TEXTMESSumlsType_CT	0.1413	37	49
TEXTMESSmeshType_CTS	0.1026	49	65
TEXTMESSumlsType_CTS	0.0858	52	70

SINAI with MeSH based expansion - mesh -.

Table 4: Official Mixed Runs Results

run name	**map**	**txt rk**	**CLEF rk**
TEXTMESSmeshTypeFIREidf_CT	**0.2777**	**2**	**6**
TEXTMESSmeshTypeFIRE_CT	0.2223	7	29
TEXTMESSumlsTypeFIREidf_CT	0.1412	10	50
TEXTMESSumlsTypeFIRE_CT	0.1325	11	56
TEXTMESSmeshTypeFIRE_CTS	1188	12	57
TEXTMESSumlsTypeFIRE_CTS	0.0887	14	69

We can observe that our standard reranking runs has gone down in the ranking, while the TF-IDF runs are in the top positions of the ranking, but they obtain the same MAP as the MAP obtained by the same configuration but without multimodal reranking - textual submissions -. It has happened because the threshold that we have used for TF-IDF reranking strategy is too much low for the competition collection. It makes that the system treat all the documents retrieved by the CBIR as if they have enough textual information to perform a suitable retrieval - skipping their CBIR relevance value-. We also have had a tuning problem with the parameters of the standard reranking strategy. Since that we only trained it using only a subset of the whole training query set and that we have used a different collection and a different CBIR from the ones we used in the competition. All of this has affected negatively to the performance of this strategy with the test collection.

5 Conclusion and Future Work

The major finding in this results has been to check that the performance of the global system can be improved joining two standalone systems which use complementary and successful methods for improving the retrieval.

Furthermore, the fact that the negative query expansion based on the acquisition type of the image using a non visual approach has had a good behaviour complementing the work done by the SINAI system using MeSH, is a non expected good result, that we should study and exploit in the future.

Finally, in order to avoid the problems experienced with the TF-IDF reranking strategy, we are planning on to work on finding an alternative method to establish the TF-IDF reranking threshold. Which instead of use the documents retrieved, use the whole collection to work out this value.

6 Acknowledgement

This research has been partially funded by the Spanish Government within the framework of the TEXT-MESS (TIN-2006-15265-C06-01 and TIN-2006-15265-C06-03) project.

References

[1] M.C. Díaz-Galiano, M.A. García-Cumbreras, M.T. Martín-Valdivia, A. Montejo-Raez, and L.A. Ureña López. SINAI at ImageCLEFmed 2007. In *In on-line Working Notes, CLEF 2007*, 2007.

[2] M.C. Díaz-Galiano, M.A. García-Cumbreras, M.T. Martín-Valdivia, L.A. Ureña López, and A Montejo-Ráez. SINAI at ImageCLEFmed 2008. In *In on-line Working Notes, CLEF 2008*, 2008.

[3] M.C. Díaz-Galiano, M.A. García-Cumbreras, L.A. Ureña López, M.T. Martín-Valdivia, and A. Montejo-Raez. SINAI at ImageCLEF 2006. In *Working Notes*, 2006.

[4] Fernando Llopis. *IR-n: Un Sistema de Recuperacin de Informacin Basado en Pasajes.* PhD thesis, University of Alicante, 2003.

[5] Sergio Navarro, Fernando Llopis, and Rafael Muñoz. A.R. Effective Mapping of Biomedical text to the UMLS Metathesaurus: the MetaMap Program. In *Proc.of the AMIA Symposium*, pages Nov.3–717–21, 2001.

[6] Sergio Navarro, Fernando Llopis, and Rafael Muñoz. Different Multimodal Approaches using IR-n in ImageCLEFphoto 2008. In *In on-line Working Notes, CLEF 2008*, 2008.

[7] Jinxi Xu and W. Bruce Croft. Improving the effectiveness of information retrieval with local context analysis. *ACM Trans. Inf. Syst.*, 18(1):79–112, 2000.