

SINAI at ImageCLEF 2009 medical task

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

Abstract

This paper describes the SINAI team participation in the ImageCLEF 2009 medical task. We explain the experiments accomplished in the medical retrieval task (ImageCLEFmed). We have experimented with query and collection expansion. For expansion, we have carried out experiments using MeSH ontology. With respect to text collection, we have used different collections, one with only caption, and the other with caption and title. Moreover, we have experimented only with textual search, using the LEMUR toolkit as Information Retrieval system.

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

General Terms

Algorithms, Experimentation, Languages, Performance

Keywords

Query expansion, Document expansion, MeSH ontology, Information Retrieval

1 Introduction

This paper presents the fifth participation of the SINAI research group at the ImageCLEF medical retrieval task. In previous years we have experimented with the expansion of the queries with medical ontologies [1, 2]. We have obtained good results expanding queries with MeSH ontology and UMLS thesaurus. This year our main goal is to study if expand collection and queries with the same ontology obtains good results.

The goal of the medical task is to retrieve relevant images based on an image query[3]. This year, the same collection as 2008 is used but with a larger number of images. The data set used contains all images from articles published in Radiology and Radiographics, more than 70,000 images.

Previous years we developed a system that tests different aspects, such as the application of Information Gain in order to improve the results[4], the expansion of the topics with the MeSH¹ ontology[2], and the expansion of the topics again with the UMLS² metathesaurus and minor textual information but more specific[1].

¹<http://www.nlm.nih.gov/mesh/>

²<http://www.nlm.nih.gov/research/umls/>

The following section describes the collections used and the expansion of collection and queries using MeSH ontology. In Section 3, we explain the experiments and obtained results. Finally, conclusions are presented in Section 4.

2 The textual expanded collection

In 2008 a new collection was introduced in this task. We have created different textual collections using information of this collection in the web [1]. This year, the task has been separated in two subtask, one based in image retrieval (*ad hoc*) and the other based in medical case retrieval. For these reason, we have created three different textual collections, two collections to image based retrieval and one to case based retrieval. The collections are:

- **C**: Contains *caption* of image to use in image based retrieval.
- **CT**: Contains *caption* of image and *title* of the article to use in image based retrieval.
- **TA**: Contains *title* and text of the full *article* to use in medical case based retrieval.

This year we have experimented with expansion of the textual collection. Our initial aim was to expand the textual collection in the same way that the query expansion in experiment of last year [1]. Nevertheless, the time request needed for expansion of minimal textual collection with UMLS and MetaMap software has been excessive. For this reason, we only have used the MeSH ontology to expand collections and topics.

From the collections used in image based retrieval, we have created other two expanding the text using MeSH ontology. These new collections have been named **CM** and **CTM** respectively. This year we have not used the collection with full article to image based retrieval for two reasons. The first one is that we want to experiment the expansion of collection with minimal textual information. The second reason is that the expansion of a big textual collection requires a large amount of time.

3 Experiment Description and Results

This year we have two types of topics sets: *ad hoc* and case based topics. We have expanded this topics set and we have obtained the next topics sets:

- **t**: Original topics set to use in *ad hoc* retrieval.
- **tM**: Topics set to *ad hoc* retrieval expanded with MeSH ontology.
- **cbt**: Original topics set to use in case based retrieval.
- **cbtM**: Topics set to case based retrieval expanded with MeSH ontology.

The name of the experiment includes the name of the group, the name of the collection (described in section 2) and the name of the topic set. The data set of the collection has been indexed using Lemur³ IR system, by applying KL-divergence weighing function and using Pseudo-Relevance Feedback (PRF).

Table 1 shows the main average precision (MAP) of image based retrieval experiments. Different from what we expected, these results show that query expansion does not improve the results. The expansion using only the collection with more textual information (CTM collection) obtains the best results.

Table 2 shows the results of medical case based experiments. As we can see in this table the query expansion does not improve the results.

³<http://www.lemurproject.org/>

| Experiments | MAP |
|-------------|---------------|
| sinaiCt | 0.3289 |
| sinaiCTt | 0.3569 |
| sinaiCMt | 0.3124 |
| sinaiCTMt | 0.3795 |
| sinaiCtM | 0.2754 |
| sinaiCTtM | 0.3077 |
| sinaiCMtM | 0.2838 |
| sinaiCTMtM | 0.3286 |

Table 1: Results of image based experiments

| Experiments | MAP |
|-------------|--------|
| sinaiTAcbt | 0.2626 |
| sinaiTAcbtM | 0.2605 |

Table 2: Results of case based experiments

4 Conclusions

This year we have used topic and the expansion of the collection. The topic expansion has been carried out in the same way as previous year. The expansion of the collections improves the results when the topics are expanded, too. However, the obtained results are not successful. At the moment, we are investigating the reasons of these unexpected results.

Acknowledgements

This work has been supported by the Regional Government of Andalucía (Spain) under excellence project GeOasis (P08-41999), the Spanish Government under project Text-Mess TIMOM (TIN2006-15265-C06-03) and the local project RFC/PP2008/UJA-08-16-14.

References

- [1] Díaz-Galiano, M.C. and M.A. García-Cumbreras and M.T. Martín-Valdivia and L.A. Urea-López and A. Montejo-Ráez. Query Expansion on Medical Image Retrieval: MeSH vs. UMLS. Lecture Notes in Computer Science. Springer. 2009.
- [2] Díaz-Galiano, M.C., García-Cumbreras, M.A., Martín-Valdivia, M.T., Montejo-Raez,A., and Ureña-López, L.A.: SINAI at ImageCLEF 2007. In Proceedings of the Cross Language Evaluation Forum (CLEF 2007), 2007.
- [3] Müller, H. and Kalpathy-Cramer, J. and Eggel, I. and Bedrick, S. and Rahouani, S. and Bakke, B. and Kahn, C. and Hersh, W. Overview of the CLEF 2009 medical image retrieval track. CLEF working notes. Corfu, Greece. 2009.
- [4] 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.