Combining TEXT-MESS Systems at ImageCLEF 2008

Sergio Navarro[1], Miguel Angel García-Cumbreras[2], Fernando Llopis[1], Manuel Carlos Díaz-Galiano[2], Rafael Muñoz[1], María Teresa Martín-Valdivia[2], L. Alfonso Ureña-López[2], and Arturo Montejo-Ráez[2]

[1] Natural Language Processing and Information Systems Group, University of Alicante, Spain
{snavarro,llopis,rafael}@dlsi.ua.es

[2] SINAI Research Group, Computer Science Department, University of Jaén, Spain
{magc,mcdiaz,maite,laurena,amontejo}@ujaen.es

Abstract. This paper describes the joint work of two teams belonging to the TEXT-MESS project. The system presented at ImageCLEFPhoto task combines one module based on filtering and other based on clustering. The main objective was to study the behavior of these methods with a large number of configurations in order to increase our chances of success. The system presented at ImageCLEFmed task 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.

1 Introduction

This paper describes the joint work of two teams belonging to the TEXT-MESS project, a complete Information Retrieval (IR) system which works in a multimodal environment, with textual information and medical or general images.

1.1 ImageCLEFphoto

The goal of the photographic task is, given a query, to retrieve a diverse, yet relevant set of images at the top of a ranked list [1]. Text and visual information can be used to improve the retrieval methods, and the main evaluation point is to study how the use of clustering or filtering methods affects to the precision and diversity of the results achieved by two different IR systems.

1.2 ImageCLEFmed

The goal of the medical task is to retrieve relevant medical images from a query based on one or several medical images and a textual query [2]. The collection contains 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.

C. Peters et al. (Eds.): CLEF 2008, LNCS 5706, pp. 597–604, 2009.

We downloaded articles from the web and constructed a new textual collection including the text of the article section where the image appears. Besides, for the experiments, we worked on the combination of two IR systems, SINAI and IR-n. Both systems use adaptations to the medical domain, and their major difference is that SINAI is based on documents while IR-n is based on passages.

2 ImageCLEFphoto System Description

The complete system is composed by an IR retrieval system and two modules that work in a serial mode, a filtering module and a clustering module. The output of the IR system is the input of the first module and its output is the input of the second one. For our participation we worked with two different IR systems which participated individually in this edition of the ImageCLEFphoto task, SINAI [3] and IR-n [4].

The SINAI system combines the output of two IR systems (Lemur[1] and Jirs [5]), is automatic (without user interaction), works with English text information (not visual information), and its blind feedback algorithm is based on the Probabilistic Relevance Feedback formula (PRF) [6].

IR-n is an IR system based on passages. It allows to select between two different blind feedback algorithms, PRF and Local Context Analysis (LCA) [7]. Also it uses different mixing strategies to perform multimodal retrieval using a provided CBIR ranked list. [4].

2.1 Filtering Module

The use of the cluster term is oriented in a filtering way. After the retrieval process the documents or passages marked as relevant are filtered as follows:

1. The cluster term is expanded with its WordNet synonyms (the first sense).
2. The list of relevant documents generated by the IR system is filtered. If the relevant document contains the cluster term or a synonym its docid (the identifier of the document) is written in another list.
3. Finally, the new list with the filtered documents is combined with the original one (Lemur and Jirs) in order to improve them. A simple method to do this was to duplicate the score value of the documents in the filtered list and to add them to the original ones.

A similar filtering method is applied in the SINAI system that works with geographical information [8].

2.2 Clustering Module

The clustering strategy seeks to promote the diversity among the 20 top ranked documents returned by the system. It is based on the relevance assigned by

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

the IR system for each document and on the clusters of documents created by Carrot2 [2], an open source clustering engine. This module rises to the top of the ranking the most relevant document within each cluster and a number of documents without cluster assigned equal to the remaining number of documents until 20. The clustering can use only the image annotations or the annotations enriched with visual concepts extracted from their related images.

3 ImageCLEFmed System Description

The complete system is composed by two systems working in a parallel mode, IR-n and SINAI systems. Both use their adaptations to the IR medical domain developed for their individual participation in this edition of the task [9] [10]. The major difference between these systems is that while SINAI is a system based on documents IR-n is based on passages.

The ranking returned by each one of the systems is merged following a standard Reranking (RR) method. Finally, the result of this step is merged with the output of a CBIR system using a rerankig strategy. We tested two different multimodal RR strategies. In the following lines we describe the main modules of these systems.

3.1 Query Expansion with MeSH and UMLS Ontologies

The SINAI expansion method using MeSH ontology [3] is the same as we carried out in the past, which obtained good results [11].

Moreover, to expand the queries with UMLS [4] , SINAI group used MetaMap program [12]. In order to reduce the number of terms that could expand the query, to make it equal to that of MeSH expansion, we restricted the semantic types in the mapped terms [9] as follows: *bpoc* (Body Part, Organ, or Organ Component), *diap* (Diagnostic Procedure), *dsyn* (Disease or Syndrome) and *neop* (Neoplastic Process). Therefore, for this expansion we used the Meta Candidate terms, because these terms provide similar terms with differences in the words. For a detailed view of the process and some examples, see [9].

3.2 Negative Query Expansion Based on the Acquisition Type of the Image

IR-n system uses a query expansion method based on the acquisition type of the image to penalize those images which do not pertain to the same acquisition type found in the query. It is based on [13], but in our approach we neither used visual features nor filtering strategy but we used a modified version of the classification proposed at that work. In order to only retrieve images of the desired type in the query we used the textual query and the text annotations for the retrieval [10].

[2] http://www.carrot2.org
[3] http://www.nlm.nih.gov/mesh/
[4] http://www.nlm.nih.gov/research/umls/

3.3 TF-IDF Multimodal RR Strategy

IR-n system allows us to use an alternative RR strategy for merging a text based list and a list returned by a CBIR system.

This approach is based on two assumptions: on the one hand the textual list is more confident than the list based on images and on the other hand the TF-IDF formula is a suitable way to measure the quantity and the quality of a text. Thus, the system only uses the relevance value returned by a CBIR system for those documents which have a TF-IDF value under an established TF-IDF threshold. Further information about the reranking formula used can be found in [10].

4 Experiments Description and Results

4.1 ImageCLEFphoto

The dataset is the collection IAPR TC-12 image collection, which 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. Each image is associated with alphanumeric captions stored in a semi-structured format (title, creation date, location, name of the photographer, description and additional notes). The topics statements also have a semi-structured format which includes the query, a cluster tag and a narrative tag.

We used the SINAI system for our experiments in the following configurations:

- **LemurJirs:** This experiment combines the IR lists of relevant documents. Lemur also uses Okapi as weighting function and PRF. Before the combination of results Lemur and Jirs lists are filtered, only with the cluster term.
- **Lemur fb okapi:** The Lemur list of relevant documents is filtered with the cluster term and its WordNet synonyms. Okapi is used as weighting function, and PRF is applied automatically.
- **Lemur fb tfidf:** It is the same experiment as before, but in this case the weighting function used was TF-IDF.
- **Lemur simple okapi:** Lemur IR system uses Okapi as weighting function and without feedback. The list of relevant documents is filtered with the cluster term and its WordNet synonyms.
- **Lemur simple tfidf:** Lemur IR system is used with TF-IDF as weighting function and without feedback. The list of relevant documents is not filtered.

The following configurations were used for the IR-n system experiments:

- **IRnExp:** This experiment uses PRF as relevance feedback strategy.
- **IRnExpClust:** It uses the annotations related to IRnExp output as input for the clustering module.
- **IRnFBFIRE:** It uses a baseline experiment of the FIRE [14] system and LCA as a multimodal relevance feedback strategy.

Table 1. ImageCLEFphoto Official Textual Results

	Standalone Run				**Official Run**			
run name	**MAP**	**P20**	**CR20**	**FMea**	**MAP**	**P20**	**CR20**	**FMea**
IRnExp**Filt**	0.2699	**0.3244**	0.2816	0.3015	**0.2671**	0.3154	0.2875	0.3008
IRnExpClust**Filt**	0.2699	**0.3244**	0.2816	0.3015	0.2287	0.2090	0.3011	0.2467
LemurSimpleOkapi Filt**Clust**	0.1972	0.2795	0.2930	0.2861	0.1750	0.1987	**0.3241**	0.2464
LemurFbOkapiFilt**Clust**	0.2089	0.2808	0.2682	0.2744	0.1804	0.1897	0.2764	0.2250
LemurJirs**Clust**	0.2063	0.2769	0.2900	0.2833	0.1840	0.2051	0.2815	0.2373
LemurFbTfidfFilt**Clust**	0.2043	0.2679	0.2704	0.2691	0.1786	0.1974	0.3185	0.2437

- **IRnFBFIREClustC:** It uses the IRnFBFIRE output run to perform a clustering based on the image annotations and the visual concepts extracted from their related images.
- **IRnConcepFBFIRE:** The image annotations indexed by IR-n are previously enriched with visual concepts extracted from the image. For the retrieval phase, the system uses a baseline run of the FIRE system and LCA as a multimodal relevance feedback strategy.
- **IRnConcepFBFIREClustC:** It uses the IRnConcepFBFIREClustC run output to perform a clustering based on the image annotations and the visual concepts extracted from their related images.

Our aim is to analyze the effect of adding to the work flow of the two IR systems the SINAI filtering method and the IR-n clustering module in order to improve their performance.

Table 1 and Table 2 show the results of the textual runs and the mixed runs respectively. Furthermore, we can see the results previously obtained by the standalone runs, without adding the external filtering or clustering module, in order to observe the improvement or worsening obtained with the added module. For each run name we show a term in bold letters which identifies the external module which has been added to that base configuration, **Filt** or **Clust**.

We can observe in the Table 1 that the CR20 value has increased its value for almost all the experiments which have used the clustering module. Indeed the best CR20 value for the textual runs has been obtained using the clustering module with SINAI system.

4.2 ImageCLEFmed

In the training phase, for each IR system, we worked in the selection of those runs which better results obtained. Next, in order to figure out which are the most suitable weighting values for the reranking module, we carried out a training phase with each pair of selected runs in the previous step. The runs submitted to the competition (**TEXTMESS** runs) were the followings:

- **meshType_CT:** Standard RR strategy which fuses the output of the SINAI system with query expansion based on MeSH ontology and the output of the

Table 2. ImageCLEFphoto Official Mixed Results (Image + Text)

	Standalone Run				Official Run			
run name	**MAP**	**P20**	**CR20**	**FMea**	**MAP**	**P20**	**CR20**	**FMea**
IRnFBFIRE**Filt**	**0.3436**	**0.4564**	0.3119	0.3706	0.3354	0.4333	0.3041	0.3574
IRnFBFIREClustC**Filt**	0.3032	0.3782	0.3483	0.3626	0.3183	0.3808	0.3178	0.3465
IRnFBFIRE**Filt**ClustC	0.3032	0.3782	0.3483	0.3626	0.3097	0.3564	0.3223	0.3385
IRnConcepFBFIRE**Filt**	0.3333	0.4333	0.3316	**0.3757**	0.3272	0.4115	0.3311	0.3669
IRnConcepFBFIRE **Filt**ClustC	0.3032	0.3782	0.3483	0.3626	0.2917	0.3410	**0.3483**	0.3446
IRnConcepFBFIRE **Filt**ClustC	0.3032	0.3782	0.3483	0.3626	0.2973	0.3603	0.3446	0.3523

IR-n system with the negative expansion. Both IR systems use the image captions and article titles of the image collection.

– **umlsType_CT:** meshType_CT configuration but the SINAI system works with the query expansion based on UMLS metathesaurus using MetaMap program instead of use MeSH ontology.
– **meshType_CTS:** meshType_CT configuration but both IR systems use captions, titles and texts of the sections where the images appear.
– **umlsType_CTS:** umlsType_CT configuration but both IR systems use captions, titles and texts of the sections where the images appear.
– **meshTypeFIREidf_CT:** TF-IDF RR strategy fusing the meshType_CT output and the FIRE system output.
– **meshTypeFIRE_CT:** Standard RR strategy fusing the meshType_CT output and the FIRE system output.
– **umlsTypeFIREidf_CT:** TF-IDF RR strategy fusing the umlsType_CT output and the FIRE system output.
– **umlsTypeFIRE_CT:** Standard RR strategy fusing the umlsType_CT run output and the FIRE system output.
– **meshTypeFIRE_CTS:** Standard RR strategy fusing the meshType_CTS output and the FIRE system output.
– **umlsTypeFIRE_CTS:** Standard RR strategy fusing the umlsType_CTS output run and a CBIR system output.

Table 3. ImageCLEFMed 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

Table 4. ImageCLEFmed Official Mixed Runs Results (Image + Text)

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	0.1188	12	57
TEXTMESSumlsTypeFIRE_CTS	0.0887	14	69

Tables 3 and 4 show the 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 the best TEXT-MESS run in the textual modality is the one which uses IR-n with negative expansion based on the acquisition type of the image and SINAI with MeSH based expansion.

5 Conclusion and Future Work

On the one hand after the analysis of the ImageCLEFphoto results we can see the following conclusions: The filtering method is not useful when we use the cluster term or related words to filter retrieved documents, because some relevant documents are deleted and none of non retrieved relevant documents are included in the second step. The clustering method, without using the cluster term, can improve the results of cluster detection, although at the expense of a decrease in precision of the results that is greater than the gain obtained for the CR20.

On the other hand, we can observe in the results of the ImageCLEFmed task, that the precision values reached by our standard multimodal RR runs has gone down in the task ranking, while the TF-IDF multimodal RR runs are in the top positions of this ranking, reaching the same MAP values achieved by the runs with the same configuration but without multimodal RR (textual runs). It is explained by the low threshold used for TF-IDF RR strategy which made that the system handles all the images retrieved by the CBIR system as images with enough textual information to perform a suitable retrieval only using the relevance returned by the textual IR system. The reasons to obtain this low threshold in the tuning phase were the use of a different collection and the use of a different CBIR system from the one which was used in the test phase. These problems affected negatively to the performance of this strategy with the test collection.

As future work we plan to improve the combination of the filtering and clustering methods only applying the filter when we predict the results obtained by the IR system will be poor. Moreover, we are planning on to work on finding an alternative method to establish the TF-IDF RR threshold.

Acknowledgements

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. Arni, T., Clough, P., Sanderson, M., Grubinger, M.: Overview of the ImageCLEFphoto 2008 Photographic Retrieval Task. In: Peters, C., et al. (eds.) CLEF 2008. LNCS, vol. 5706, pp. 500–511. Springer, Heidelberg (2009)
2. Müller, H., Kalpathy-Cramer, J., Kahn, C.E., Hatt, W., Bedrick, S., Hersh, W.: Overview of the ImageCLEFmed 2008 medical image retrieval task. In: Peters, C., et al. (eds.) CLEF 2008. LNCS, vol. 5706, pp. 512–522. Springer, Heidelberg (2009)
3. García-Cumbreras, M.A., Díaz-Galiano, M.C., Martín-Valdivia, M.T., Ureña López, L.A.: SINAI at ImageCLEFphoto 2008. In: On-line Working Notes, CLEF 2008 (2008)
4. Navarro, S., Llopis, F., Muñoz, R.: Different Multimodal Approaches using IR-n in ImageCLEFphoto 2008. In: On-line Working Notes, CLEF 2008 (2008)
5. Gómez-Soriano, J.M., Montes-y-Gómez, M., Sanchis-Arnal, E., Rosso, P.: A Passage Retrieval System for Multilingual Question Answering. In: Matoušek, V., Mautner, P., Pavelka, T. (eds.) TSD 2005. LNCS (LNAI), vol. 3658, pp. 443–450. Springer, Heidelberg (2005)
6. Robertson, S.E., Sparck Jones, K.: Relevance weighting of search terms. Journal of the American Society for Information Science 27(3), 129–146 (1976)
7. Xu, J., Croft, W.B.: Improving the efectiveness of information retrieval with local context analysis. ACM Trans. Inf. Syst. 18(1), 79–112 (2000)
8. Perea-Ortega, J.M., García-Cumbreras, M.A., García-Vega, M., Montejo-Raez, A.: Filtering for Improving the Geographic Information Search. In: Peters, C., et al. (eds.) CLEF 2007. LNCS, vol. 5152, pp. 823–829. Springer, Heidelberg (2008)
9. Díaz-Galiano, M.C., García-Cumbreras, M.A., Martín-Valdivia, M.T., Ureña López, L.A., Montejo-Ráez, A.: SINAI at ImageCLEFmed 2008. In: On-line Working Notes, CLEF 2008 (2008)
10. Navarro, S., Muñoz, R., Llopis, F.: A Multimodal Approach to the Medical Retrieval Task using IR-n. In: On-line Working Notes, CLEF 2008 (2008)
11. Díaz-Galiano, M.C., García-Cumbreras, M.A., Martín-Valdivia, M.T., Montejo-Raez, A., Ureña-López, L.A.: Integrating MeSH Ontology to Improve Medical 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. 601–606. Springer, Heidelberg (2008)
12. Aronson, A.R.: Effective Mapping of Biomedical text to the UMLS Metathesaurus: the Meta Map Program. In: Proc. of the AMIA Symposium, Washington, DC, November 3-7, pp. 17–21 (2001)
13. Kalpathy-Cramera, J., Hersh, W.: Automatic Image Modality Based Classification and Annotation to Improve Medical Image Retrieval. In: MEDINFO 2007: Proceedings of the 12th World Congress on Health (Medical) Informatics; Building Sustainable Health Systems, Brisbane, Australia, pp. 1334–1338 (2007)
14. Deselaers, T., Keysers, D., Ney, H.: Features for Image Retrieval: An Experimental Comparison. Information Retrieval 11(77), 77–107 (2008)