Methods for combining content-based and textual-based approaches in medical image retrieval

Mouna Torjmen, Karen Pinel-Sauvagnat, Mohand Boughanem
SIG-IRIT-Toulouse-France
{Mouna.Torjmen , Karen.Sauvagnat , Mohand.Boughanem }@irit.fr

Abstract

This paper describes our participation in the Medical Image Retrieval task of Image CLEF 2008. Our aim was to evaluate different combination methods for purely textual and visual approaches.
Our most interesting conclusion is that combining results provided by both methods using classical combination function allows to obtain higher retrieval accuracy in terms of MAP .
MAP values than combination according to query type. Moreover, it is more reliable than using only textual retrieval or using only visual retrieval.

Categories and Subject Descriptors

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

General Terms

Measurement, Performance, Experimentation

Keywords

Contextual image retrieval, content-based image retrieval, combination, query classification

1 Introduction

In Image Retrieval, there are two main approaches [7] : (1) Context Based Image Retrieval and (2) Content Based Image Retrieval:

- The context of an image is all information about the image coming from other sources than the image itself. It is often reduced to textual information. The main problem of this approach is that documents use different words to describe the same image or can use the same words to describe different concepts.

- Content Based Image Retrieval (CBIR) systems use low-level image features to return images similar to an image used as example. The main problem of this approach is that visual similarity may not correspond to semantic similarity (for example a CBIR system can return a picture of blue sky when the example image is a blue car).

In order to take advantages of both techniques, we propose in this paper an evaluation of combination methods in the Medical retrieval task of CLEF 2008. Our aim is to compare classical combination using a linear combination function and a combination method that takes into account the query type: visual, textual, mixed. The two systems we used for content-based image retrieval and text-based image retrieval are respectively *GIFT* [3] and *XFIRM* [5].

The rest of the paper is organized as follows. Section 2 describes our approach to evaluate Medical Retrieval queries. In fact, we describe both used systems and our combination techniques. In section 3, we present an empirical evaluation of the proposed methods carried out using the Medical Retrieval Task in Image CLEF 2008. We conclude in section 4 with a discussion on our finding and suggestions for future work.

2 Retrieval approaches

2.1 The XFIRM model

We use the *XFIRM* XML search engine [5] as a base model for textual queries processing. However, as the document structure in the collection is not complex (the average document depth is low), we use a simplified version of the model.

The model is based on a relevance propagation method. During query processing, relevance scores are computed at leaf nodes level and then at inner nodes level thanks to a propagation of leaf nodes scores through the document tree. An ordered list of subtrees is then returned to the user.

Let $q = t_1, \ldots, t_n$ be a query composed of n terms. Relevance values of leaf nodes are computed using a similarity function $RSV(q, ln)$.

$$RSV(q, ln) = \sum_{i=1}^{n} w_i^q * w_i^{ln}, \quad where \quad w_i^q = tf_i^q * idf_i \quad and \quad w_i^{ln} = tf_i^{ln} * idf_i \qquad (1)$$

Where w_i^q and w_i^{ln} are the weights of term i in query q and leaf node ln respectively. tf_i^q and tf_i^{ln} are the frequency of i in q and ln respectively, $idf_i = log(|D|/(|di| + 1)) + 1$, with $|D|$ the total number of documents in the collection, and $|di|$ the number of documents containing i.

Each node in the document tree is then assigned a relevance score which is function of the relevance scores of the leaf nodes it contains.

$$r_n = |L_n^r|. \sum_{ln_k \in L_n} \alpha^{dist(n, ln_k) - 1} \qquad (2)$$

$dist(n, ln_k)$ is the distance between node n and leaf node ln_k in the document tree, i.e. the number of edges that are necessary to join n and ln_k, and $\alpha \in]0..1]$ allows to adapt the importance of the *dist* parameter. $|L_n^r|$ is the number of leaf nodes being descendant of n and having a non-zero relevance value (according to equation 1).

In the indexing phase, only two fields of documents of the medical textual collection are indexed ("caption" and "title") as they are the only ones which contain significant textual information. As the aim of the task is to return image identifier (document), we take $\alpha = 1$ in equation 2 to ensure that best ranked elements will be documents. As each document corresponds to a single image, we indexed documents using the image name and not the document name, in order to return directly the required element (image). In the rest of the paper, document relevance is thus equivalent to image relevance in the retrieval process.

2.2 The GIFT System

GIFT or GNU Image Finding Tool [3] is a free CBIR system released under GNU After license. It processes Query By Example (QBE[1]) on images, with the opportunity to improve query results by relevance feedback.

It uses common techniques from content-based information retrieval [4]. A big number of low level features can be used ($\succ$ 80000). These features are both local and global, simple color and texture... [6]. *GIFT* system has a variable length list of discrete features for every image.

In the experiments presented here, we directly used *GIFT* results kindly provided by organizers, with no further processing.

2.3 Combination of *XFIRM* and *GIFT* systems

In ImageCLEFmed 2008, there are 30 topics composed of both textual and images example queries, with 10 topics classified in the following three categories:

- **Visual** : topics where visual system alone is expected to reach good performance (a visual query example is given in figure 1)
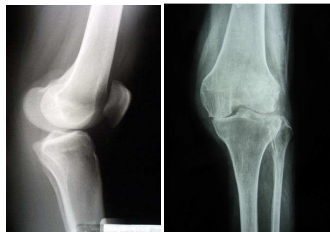
Show me images of a knee x-ray.
Zeige mir Röntgenbilder des Knies.
Montre-moi des radiographies du genou.

Figure 1: Example of a visual topic

- **Mixed** : topics with a visual influence but where text can improve results strongly (a mixed query example is showed in figure 2)

Show me all x-ray images containing one or more fractures.
Zeige mir alle Brüche auf Röntgenbildern.
Montre-moi toutes les fractures sur des radiographies.

Figure 2: Example of a mixed topic

- **Textual/semantic** : topics where the visual influence is limited as the variety with respect to visualness in the results is large (a textual query example is given in figure 3)

---

[1]http://en.wikipedia.org/wiki/QBE

microscopic giant cell
Mikroskopie Riesenzellen
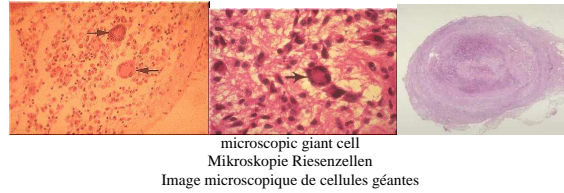Image microscopique de cellules géantes

Figure 3: Example of a textual topic

We evaluated two combination methods between the *XFIRM* and *GIFT* systems.

2.3.1 Classical combination

In this approach, we used the two aforementioned systems on the whole set of queries and merged their results to obtain a single result list.

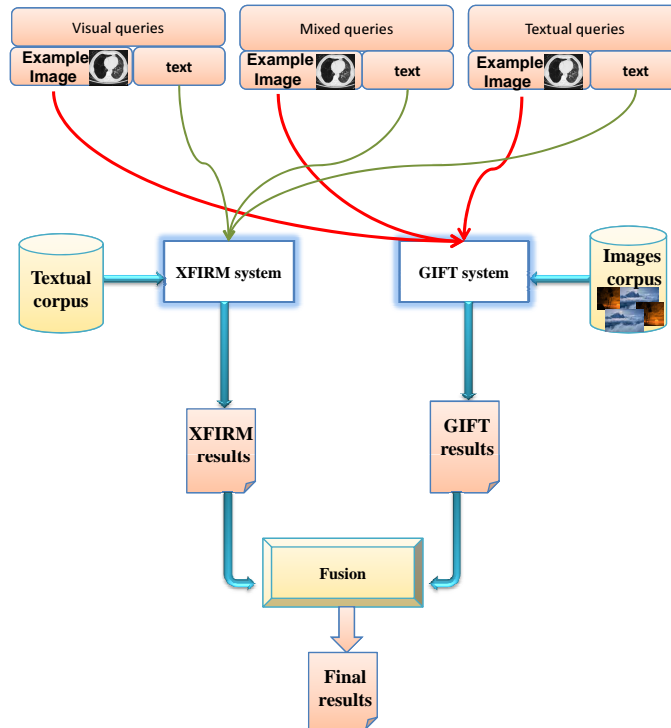The overall structure of this approach is depicted in figure 4.

Figure 4: Overview of our approach

To merge the two results lists into a single list of ranked results we first normalize scores obtained by the two systems, and then we use a simple and classic linear combination of evidences:

$$FS(image) = \alpha \cdot S_{XFIRM}(image) + (1 - \alpha) \cdot S_{GIFT}(image) \qquad (3)$$

when α is a pivot parameter $\in [0..1]$, $S_{XFIRM}(image)$ represents the image score obtained using the *XFIRM* system, and $S_{GIFT}(image)$ is the score of the same image obtained by the *GIFT* system.

When α is set to 1, only the image score from the *XFIRM* system is used. On the other hand, only the image score by the *GIFT* system is used when the value of α is set to 0. Our official runs

in Medical Image Retrieval task were submitted with α equal to 0.1, 0.5, 0.9 and 1.0. The fusion scores of the image and textual retrieval methods are then ranked in a descending order.

2.3.2 Combination according to query type

We also evaluated the processing of each category of queries with a different system. We thus used the *GIFT* system to evaluate visual topics, the *XFIRM* system to evaluate textual (semantic) topics, and a classic combination function (Equation 3) of the two systems to evaluate mixed topics. Figure 5 illustrates our approach.
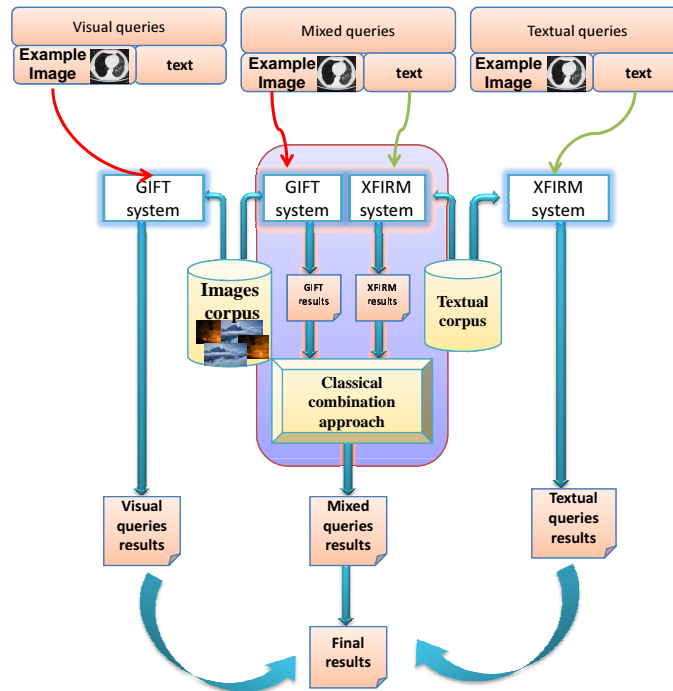

Figure 5: Overview of our approach that evaluates queries according to their types

3 Runs and Results

Results are presented in table 1. Our official runs are in grayed boxes. In these runs, the number of returned results by the XFIRM system was 250 while the number of returned results can be 1000. We corrected these runs by returning the well number of results.

XFIRMRun is the complete run of *SigRunText*, it used only textual parts of all queries and it was processed using the XFIRM system.
GiftRun is the run obtained using only example images and was processed using the Gift system. Comparing the two runs, we notice that textual based retrieval approach is much better (MAP=0.1524 versus MAP=0.0349 using the Gift system).
Our best run, *RunComb09*, used the classical combination function with $\alpha = 0.9$. This implies that the two information sources are needed to get the best performances, but textual information must be the main source of evidence and visual information must be used as a complementary evidence element to improve search results. In fact, results are improved when α increases.
RunMix09 is obtained using the approach illustrated in figure 5. For the ten mixed queries, the combination factor α used is 0.9 (the best run obtained using the classical combination).

Table 1: Results of Clef Medical Retrieval Task

Runs	α	MAP	bpref	P@10	p@30
GiftRun	0	0.0349	0.0898	0.1700	0.1511
SigRunText	1	0.1410	0.1851	0.2430	0.2244
XFIRMRun	1	0.1524	0.2454	0.2600	0.2278
SigCombAlpha01	0.1	0.0427	0.0929	0.2200	0.1600
RunComb01	0.1	0.0483	0.1021	0.2200	0.1800
RunComb02	0.2	0.0572	0.1098	0.2700	0.2111
RunComb03	0.3	0.0658	0.1170	0.3167	0.2311
RunComb04	0.4	0.0749	0.1190	0.3000	0.2452
SigCombAlpha05	0.5	0.0432	0.0948	0.2200	0.1600
RunComb05	0.5	0.0820	0.1376	**0.3733**	0.2833
RunComb06	0.6	0.0909	0.1462	**0.3733**	0.2889
RunComb07	0.7	0.1014	0.1591	0.3433	**0.2989**
RunComb08	0.8	0.1409	0.2167	0.3100	0.2933
SigCombAlpha09	0.9	0.0415	0.0947	0.2170	0.1611
RunComb09	0.9	**0.1705**	**0.2614**	0.2900	0.2611
SigMix	0.5	0.1113	0.1637	0.2870	0.2311
RunMix1	1	0.1043	0.1906	0.2067	0.1678
RunMix09	0.9	0.1101	0.1914	0.2133	0.1756

The MAP of this run showed that performance increases comparatively to textual processing and classical combination function.

More work and experimentations are needed to show the difference and the effectiveness of each approach. We plan in future work to evaluate our approaches using other systems as results are also depending of the systems used for retrieval.

Table 2 shows results for each query category using respectively the XFIRM system, the GIFT system and the best combination of both (with $\alpha = 0.9$). For all query categories, best MAP is obtained using classical combination with $\alpha = 0.9$. So the intuition that visual queries must be processed with a CBIR system, textual queries must be processed with textual based retrieval system and mixed queries must be processed with a combination of both is not validated in the results. This shows that the classification of queries into visual, mixed and textual categories is not adequate in improving results, and consequently, using classical combination remains the best way to have best results.

Table 2: Some Results of Clef Medical Retrieval Task by query category

Runs	MAP	bpref	P@10	p@30
Visual Queries (10)				
RunVisXFIRM	0.1658	**0.2256**	0.2600	0.2600
RunVisGift	0.0159	0.0615	0.1000	0.0800
RunVisComb09	**0.1667**	0.2199	**0.2800**	**0.2633**
Mixed Queries (10)				
RunMixXFIRM	0.0630	0.1609	0.1400	0.1033
RunMixGift	0.0613	0.1302	**0.3200**	**0.2800**
RunMixComb09	**0.0805**	**0.1632**	0.1600	0.1267
Textual Queries (10)				
RunTexXFIRM	0.2341	0.3496	**0.3800**	**0.3200**
RunTexGift	0.0275	0.0778	0.0900	0.0933
RunTexComb09	**0.2643**	**0.4041**	0.2800	0.2633

Our proposed approach described in figure 5 is very simple as it used queries classification

provided by organizers. We think that it can be improved and give best results by studying techniques of automatic query classification using other features. In fact, query classification is an active research field where Good [2] and Fairthorne [1] were among the first to recommend automatic query classification to improve document retrieval.

4   Conclusion and future work

In this paper, we evaluated a classical combination approach and an approach based on queries classification using two systems (GIFT and XFIRM). Best results for the MAP metric are obtained using the classical combination method, more precisely with giving more importance for textual information than visual ones. However, our proposed method of query classification processing leads to a significant performance degradation. This can be explained by the fact that classify queries into visual, textual and mixed queries is not an adequate way of classification. In addition, results are also depending of the used systems.

In future work, we plan to confirm our conclusions by evaluating this approach using systems other than GIFT and XFIRM and using other collection (as Medical Image Retrieval in Image CLEF 2007). In addition, one of the most enticing directions of future work is to study the automatic classification of queries in order to more study the impact of this technique in image retrieval.

References

[1] R.A. FAIRTHORNE. The mathematics of classification, 1961.

[2] I J Good. Speculations concerning information retrieval. Technical report, Research Report PC-78, IBM Research Centre, Yorktown Heights, New York, 1958.

[3] Henning Müller, Wolfgang Müller, David McG. Squire, Zoran Pecenovic, Stéphane Marchand-Maillet, and Thierry Pun. An open framework for distributed multimedia retrieval.

[4] Mika Rummukainen, Jorma Laaksonen, and Markus Koskela. An efficiency comparison of two content-based image retrieval systems, gift and picsom. In Erwin M. Bakker, Thomas S. Huang, Michael S. Lew, Nicu Sebe, and Xiang Sean Zhou, editors, *CIVR*, volume 2728 of *Lecture Notes in Computer Science*, pages 500–509. Springer, 2003.

[5] Karen Sauvagnat. *Modle flexible pour la recherche d'information dans des corpus de documents semi-structurs*. PhD thesis, Toulouse : Paul Sabatier University, 2005.

[6] D. Squire, W. Muller, H. Muller, and J. Raki. Content-based query of image databases, inspirations from text retrieval: inverted files, frequency-based weights and relevance feedback, 1999.

[7] Thijs Westerveld. Image retrieval: Content versus context. In *Content-Based Multimedia Information Access, RIAO 2000 Conference Proceedings*, pages 276–284, April 2000.