Using Pseudo-Relevance Feedback to Improve Image Retrieval Results

Mouna Torjmen, Karen Pinel-Sauvagnat, and Mohand Boughanem

IRIT, 118 Route Narbonne-31062 Toulouse Cedex 4 -France
{torjmen,sauvagna,bougha}@irit.fr

Abstract. In this paper, we propose a pseudo-relevance feedback method to deal with the photographic retrieval and medical retrieval tasks of ImageCLEF 2007. The aim of our participation to ImageCLEF is to evaluate a combination method using both english textual queries and image queries to answer to topics. The approach processes image queries and merges them with textual queries in order to improve results.

A first set of expirements using only textual information does not allow to obtain good results. To process image queries, we used the *FIRE* system to sort similar images using low level features, and we then used associated textual information of the top images to construct a new textual query. Results showed the interest of low level features to process image queries, as performance increased compared to textual queries processing.

Finally, best results were obtained combining the results lists of textual queries processing and image queries processing with a linear function.

1 Introduction

In Image Retrieval, one can distinguish two main approaches [1] : (1) Context Based Image Retrieval and (2) Content Based Image Retrieval:

- The context of an image is all information about the image coming from others sources than the image itself. For the time being, only textual information is used as context. The main problem of this approach is that documents can use different words to describe the same image or can use the same words to describe different concepts. Moreover image queries can't be processed.
- Content Based Image Retrieval (CBIR) systems use low-level image features to return images similar to an example image. The main problem of this approach is that visual similarity does not always correspond to semantic similarity (for example a CBIR system can return a picture of blue sky when the example image is a blue car).

Most of the image retrieval systems combine nowadays content and context retrieval, in order to take advantages of both methods. Indeed, it has been proved that combining text- and content-based methods for images retrieval always improves performance [2].

C. Peters et al. (Eds.): CLEF 2007, LNCS 5152, pp. 665–673, 2008.

Images and textual information can be considered as independent and content and contextual information of queries can be combined in different ways:

- Image queries and textual queries can be processed separately and the two results lists are then merged using a linear function [3], [4].
- One can also use a pipeline approach: a first search is done using textual information or content information, and a filtering step is then processed using the other information type to exclude non-relevant images [5].
- Other methods use Latent Semantic Analysis (LSA) techniques to combine visual and textual information, but are not efficient [1] [6].

Some other works propose translation-based methods, in which content and context information are complementary. The main idea is to extract relations between images and text, and to use them to translate textual information to visual one and vice versa [7]:

- In [8], authors translate textual queries to visual ones.
- Authors of [9] propose to translate image queries to textual ones, and to process them using textual methods. Results are then merged with those obtained with textual queries. Authors in [10] also propose to expand the initial textual query by terms extracted thanks to an image query.

For the latter methods, the main problem to construct a new textual query or expand an initial textual query is term extraction. To do this, the main solution is *pseudo-relevance feedback*. Using pseudo-relevance feedback in context based image retrieval to process image queries is slightly different from classic pseudo-relevance feedback. The first step is to use a visual system to process image queries. Images obtained as results are considered as relevant and the associated textual information is then used to select terms in order to express a new textual query.

The work presented in this paper also proposes to combine context and content information to answer to the photographic retrieval and medical retrieval tasks. More precisely, we present a method to transform image queries to textual ones. We use *XFIRM* [11], a structured information retrieval system, to process english textual queries, and the *FIRE* system [12] to process image queries. Documents corresponding to the images returned by *FIRE* are used to extract terms that will form a new textual query.

The paper is organized as follows. In Section 2, we describe textual queries processing using the *XFIRM* system. In Section 3, we describe the image queries processing using in a first step, the *FIRE* system, and in a second step a pseudo-relevance feedback method. In Section 4, we present our combination method, which uses both results of the *XFIRM* and *FIRE* systems. Experiments and results for the two tasks (medical retrieval and photographic retrieval [13], [14]) are exposed in section 5. We discuss results in section 6 and finally we conclude in Section 7 .

2 Textual Queries Processing

Textual information of collections used for the photographic and medical retrieval tasks [14] is organized using the XML language. In the indexing phase, we decided to only use documents elements containing positive information: $\prec$ *description* $\succ$, $\prec$ *title* $\succ$, $\prec$ *notes* $\succ$ and $\prec$ *location* $\succ$.

We then used the *XFIRM* system [11] to process queries. *XFIRM* (*XML Flexible Information Retrieval Model*) uses a relevance propagation method to process textual queries in XML documents. Relevance values are first computed on *leaf nodes* (which contain textual information) and scores are then propagated along the document tree to evaluate *inner nodes* relevance values.

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

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

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, $idf_i = log(|D|/(|di| + 1)) + 1$, with $|D|$ the total number of documents in the collection, and $|di|$ the number of documents containing i, and ief_i is the inverse element frequency of term i, i.e. $log(|N|/|nf_i| + 1) + 1$, where $|nf_i|$ is the number of leaf nodes containing i and $|N|$ is the total number of leaf nodes in the collection.

idf_i allows to model the importance of term i in the collection of documents, while ief_i allows to model it in the collection of elements.

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

$$r_n = \rho * |L_n^r| . \sum_{ln_k \in L_n} \alpha^{dist(n, ln_k) - 1} * RSV(q, ln_k) + (1 - \rho) * r_{root} \qquad (2)$$

$dist(n, ln_k)$ is the distance between node n and leaf node ln_k in the document tree, i.e. the number of arcs that are necessary to join n and ln_k, and $\alpha \in]0..1]$ allows to adapt the importance of the *dist* parameter. In all the experiments presented in the paper, α is set to 0.6.

L_n is the set of leaf nodes being descendant of n, and $|L_n^r|$ is the number of leaf nodes in L_n having a non-zero relevance value (according to equation 1). $\rho \in]0..1]$, inspired from work presented in [15], allows the introduction of document relevance in inner nodes relevance evaluation, and r_{root} is the relevance score of the *root* element, i.e. the relevance score of the whole document, evaluated with equation 2 with $\rho = 1$.

Finally, documents d_j containing relevant nodes are retrieved with the following relevance score:

$$r_{XFIRM}(d_j) = max_{n \in d_j} r_n \qquad (3)$$

Images associated to the documents are lastly returned by the system to answer to the retrieval tasks.

3 Image Queries Processing

To process image queries, we used a third-steps method: (1) a first step is to process images using the *FIRE* System [12], (2) we then use pseudo-relevance feedback to construct new textual queries, (3) the new textual queries are processed with the *XFIRM* system.

We first used the *FIRE* system to get the top K similar images to the image query. We then get the N associated textual documents (with N $\leq$ K, because some images do not have associated textual information) and extracted the top L terms from them. To select the top L terms, we evaluated two formula to express the weight w_i of term t_i.

The first formula uses the frequency of term t_i in the N documents.

$$w_i = \sum_{j=1}^{N} tf_i^j \tag{4}$$

where tf_i^j is the frequency of term t_i in document d_j.

The second formula uses terms frequency in the N selected documents, the number of documents in the N selected containing the term, and a normalized *idf* of the term in the whole collection.

$$w_i = [1 + log(\sum_{j=1}^{N} tf_i^j)] * \frac{n_i}{N} * \frac{log(\frac{D}{d_i})}{log(D)} \tag{5}$$

where n_i is the number of documents in the N associated documents containing the term t_i, D is the number of all documents in the collection and d_i is the number of documents in the collection containing t_i.

The use of the $\frac{n_i}{N}$ parameter is based on the following assumption: a term occuring one time in n documents is more important and must be more relevant than a term occuring n times in one document. The log function is used on $\sum_{j=1}^{N} tf_i^j$ to emphasize the impact of the $\frac{n_i}{N}$ parameter.

We then construct a new textual query with the top L terms selected according to formula 4 or 5 and we process it using the *XFIRM* system (as explained in section 2).

In the photographic retrieval task, we obtained the following queries for topic Q48, with $K = 5$ and $L <= 5$:

Textual query using equation 4: ”south korea river”
Textual query using equation 5: ”south korea night forklift australia”

The original textual query in english was: ”vehicle in South Korea”. As we can see, the query using equation 5 is more similar to the original query than the one using equation 4.

4 Combination Function

To evaluate the interest of using both content and context information, we combined results of image queries and textual queries processing and we evaluated new relevance scores $r(d_j)$ for documents d_j:

$$r(d_j) = \lambda * (r_{XFIRM}(d_j)) + (1 - \lambda) * (r_{PRF}(d_j)) \qquad (6)$$

where $r_{XFIRM}(d_j)$ is the relevance score of document d_j according to the *XFIRM* system (equation 3) and $r_{PRF}(d_j)$ is the relevance score of d_j according to the *XFIRM* system after image queries processing (see section 3).

In order to answer to both retrieval tasks, we then return all images associated to the top ranked documents. Figure 1 illustrates our approach.

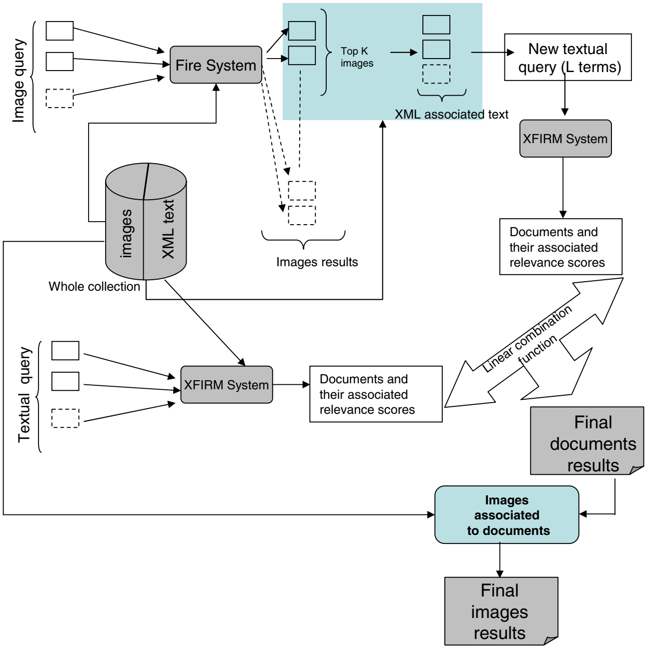
Image query
Fire System
Top K images
XML associated text
New textual query (L terms)
XFIRM System
images
XML text
Whole collection
Images results
Documents and their associated relevance scores
Linear combination function
Textual query
XFIRM System
Documents and their associated relevance scores
Final documents results
Images associated to documents
Final images results

Fig. 1. Query processing with the combination of image and textual query processing approach

5 Evaluation and Results

5.1 Photographic Retrieval Task

- **Evaluation of textual queries**
  We evaluated english textual queries using the*XFIRM* system with parameters $\rho = 0.9$ and $\rho = 1$. Results, which are almost the same, are presented in table 1.
- **Evaluation of image queries**
  Table 2 shows results using the two formula described in section 3. We notice that the use of term frequency in selected documents is not enough, and that

the importance of the term in the collection need to be used in the term weighted function (results are better with equation 5 than with equation 4). If we now compare table 1 and table 2, we see that processing image queries with the *FIRE* system and our pseudo-relevance feedback system gives better results than using only the *XFIRM* system on textual queries. It shows the importance of visual features to retrieve images.

- **Combination of textual and image queries results**

  Table 3 shows our results for the combination approach. For all these experiments, L is set to 5.

  Let us first compare runs Runcomb1 and Runcomb4, which use eq. 4 and K=6, and eq. 5 and K=15. For both, we use $\rho = 1$ and $\lambda = 0.9$ for the combination. Results show that using eq. 5 with K=15 is more efficient that eq. 4 with K=6, which confirms results obtained using only image queries.

  In order to evaluate the combination function, we then use eq. 5, and fix $\rho = 1$ and K=15. We test $\lambda = 0.5$ and $\lambda = 0.9$ (runs Runcomb3 and Runcomb4). Results are almost the same but combining equally the two sources of evidence gives slightly better results.

  Finally, we vary $\rho = 0, 9$ and $\rho = 1$, and fix equation 5, $\lambda = 0.9$ in equation 6 and K=15 (runs Runcomb4 and Runcomb2). Better results are obtained with $\rho = 1$, which means that the document relevance should not be taken into account in the evaluation of inner nodes relevance values (equation 2).

Table 1. Textual queries results using the *XFIRM* system

Run-id	ρ	MAP	P10	P20	P30	Bpref	GMAP
RunText0609	0.9	0.0634	0.1400	0.1175	0.1133	0.0719	0.0039
RunText061	1	0.0633	0.1400	0.1175	0.1128	0.0719	0.0039

Table 2. Image queries results using pseudo-relevance feedback with the *FIRE* and *XFIRM* systems

Run-id	K	L	ρ	Eq.	MAP	P10	P20	P30	Bpref	GMAP
RunPRF061tf	6	5	1	eq. 4	0.063	0.140	0.117	0.113	0.071	0.003
RunPRF061tfnNidf	6	15	1	eq. 5	0.123	0.210	0.200	0.179	0.138	0.006
RunPRF0609tfnNidf	6	15	0.9	eq. 5	0.125	0.211	0.200	0.179	0.138	0.006

Table 3. Results using the combination function

Run-id	K	λ	ρ	Eq.	MAP	P10	P20	P30	Bpref	GMAP
RunComb1	6	0.9	1	eq. 4	0.103	0.150	0.124	0.118	0.091	0.031
RunComb2	6	0.9	0.9	eq. 5	0.109	0.143	0.129	0.126	0.096	0.029
RunComb3	15	0.5	1	eq. 5	0.135	0.221	0.198	0.183	0.140	0.035
RunComb4	15	0.9	1	eq. 5	0.130	0.210	0.198	0.186	0.145	0.026

5.2 Medical Retrieval Task

For this task, we only evaluated the combination method described in section 4. RComb09 uses equation 5 with $\rho = 1$, K=15, L=10 and $\lambda = 0.9$. RComb05 , our official run, uses equation 4 with ρ=1, K=6, L=5 and $\lambda = 0.5$.

Results are significantly better for run RComb09. However, as many parameters are involved (K, L, λ and the equation used to select terms) it is difficult to conclude on which parameters impact the results. Further experiments are thus needed.

Table 4. Results of the Medical retrieval task

Run-id	Eq.	L	K	λ	MAP	R-prec	Bpref	P10	P30	P100	P500	P1000
RComb09	eq.5	10	15	0.9	0.110	0.141	0.213	0.166	0.152	0.144	0.067	0.041
RComb05	eq.4	5	6	0.5	0.048	0.070	0.168	0.05	0.075	0.058	0.058	0.038

6 Discussion

The number of textual information resources used to construct new textual queries from image queries (i.e the K number of images selected from FIRE results) has a great impact on results. Increasing K improves results by introducing relevant information. Another factor that impacts on results is the number of new query terms L. In our experiments, when K and L increase, the MAP metric also increases. Moreover, processing textual queries or images separately does not allow to obtain the best results: combining the two sources of evidence clearly improves results.

Finally, we'd like to conclude with the type of textual information used. In the Medical and Photographic Retrieval Tasks, textual information is encoded using the XML language, and as a consequence, we decided to use an XML-oriented information retrieval system to process textual queries (*XFIRM*). However, elements are not organized in a hierarchic way as in can be the case in XML documents (no ancestor-descendant relationships between nodes), and the functions used by the *XFIRM* system to evaluate nodes relevance may be not appropriate in that case. Other experiments are consequently needed with a plain-text information retrieval system. Combining the *XFIRM* system with the *FIRE* system may be however interesting with fully encoded-XML collections.

7 Conclusion and Future Work

We participated in the Photographic and Medical Retrieval Tasks of ImageCLEF 2007 in order to evaluate a method using a content- and context-based approach to answer to topics. We proposed a new pseudo-relevance feedback approach to process image queries and we tested an XML oriented system to process textual queries. Results showed the interest of combining the two sources of evidence (content and context) to answer to image retrieval.

In future work, we plan to:

- Add low level features results extracted from *FIRE* to the combination function in the Medical Retrieval Task, as visual features are very important in the medical domain.
- Sort images using concepts level features [16] instead of low level features to construct new textual queries in the Photographic Retrieval Task.
- Use specific domain ontology to expand textual queries (original textual queries and queries obtained with our pseudo-relevance feedback approach).

References

1. Westerveld, T.: Image retrieval: Content versus context. In: Content-Based Multimedia Information Access, RIAO 2000 Conference Proceedings, pp. 276–284 (2000)
2. Deselaers, T., Müller, H., Clogh, P., Ney, H., Lehmann, T.M.: The clef 2005 automatic medical image annotation task. International Journal of Computer Vision 74(1), 51–58 (2007)
3. Boll, S., Klas, W., Wandel, J.: A cross-media adaptation strategy for multimedia presentations. In: ACM Multimedia (1), pp. 37–46 (1999)
4. Jones, G.J.F., Burke, M., Judge, J., Khasin, A., Lam-Adesina, A.M., Wagner, J.: Dublin city university at clef 2004: Experiments in monolingual, bilingual and multilingual retrieval. In: CLEF, pp. 207–220 (2004)
5. Mori, Y., Takahashi, H., Oka, R.: Image-to-word transformation based on dividing and vector quantizing images with words (1999)
6. Zhao, R., Grosky, W.: Narrowing the semantic gap - improved text-based web document retrieval using visual features (2002)
7. Lin, W.C., Chang, Y.C., Chen, H.H.: Integrating textual and visual information for cross-language image retrieval: A trans-media dictionary approach. Inf. Process. Manage. 43(2), 488–502 (2007)
8. Lin, W.C., Chang, Y.C., Chen, H.H.: Integrating textual and visual information for cross-language image retrieval. In: Proceedings of the Second Asia Information Retrieval Symposium, pp. 454–466 (2005)
9. Chang, Y.C., Lin, W.C., Chen, H.H.: A corpus-based relevance feedback approach to cross-language image retrieval. In: Peters, C., Gey, F.C., Gonzalo, J., Müller, H., Jones, G.J.F., Kluck, M., Magnini, B., de Rijke, M., Giampiccolo, D. (eds.) CLEF 2005. LNCS, vol. 4022, pp. 592–601. Springer, Heidelberg (2006)
10. Maillot, N., Chevallet, J.P., Valea, V., Lim, J.H.: Ipal inter-media pseudo-relevance feedback approach to imageclef 2006 photo retrieval. In: Working Notes for the CLEF 2006 Workshop, 20-22 September, Alicante, Spain (2006)
11. Sauvagnat, K.: Modéle flexible pour la recherche d'information dans des corpus de documents semi-structurés. PhD thesis, Toulouse: Paul Sabatier University (2005)
12. Deselaers, T., Keysers, D., Ney, H.: FIRE — flexible image retrieval engine: ImageCLEF 2004 evaluation. In: CLEF Workshop (2004) (2004)
13. Müller, H., Deselaers, T., Kim, E., Kalpathy-Cramer, J., Deserno, T.M., Clough, P., Hersh, W.: Overview of the ImageCLEFmed 2007 medical retrieval and annotation tasks. In: Working Notes of the 2007 CLEF Workshop, Budapest, Hungary (2007)
14. Grubinger, M., Clough, P., Hanbury, A., Müller, H.: Overview of the ImageCLEF 2007 photographic retrieval task. In: Working Notes of the 2007 CLEF Workshop, Budapest, Hungary (2007)

15. Mass, Y., Mandelbrod, M.: Experimenting various user models for XML retrieval. In: [17] (2005)
16. Snoek, C.G.M., Worring, M., van Gemert, J.C., Geusebroek, J.M., Smeulders, A.W.M.: The challenge problem for automated detection of 101 semantic concepts in multimedia. In: MULTIMEDIA 2006: Proceedings of the 14th annual ACM international conference on Multimedia, pp. 421–430. ACM Press, New York (2006)
17. Fuhr, N., Lalmas, M., Malik, S., Kazai, G.: INEX 2005 workshop proceedings (2005)