Different Multimodal Approaches using IR-n in ImageCLEFphoto 2008

Sergio Navarro, Fernando Llopis, Rafael Muñoz

Natural Language Processing and Information Systems Group

University of Alicante, Spain

`snavarro,llopis,rafael@dlsi.ua.es`

Abstract

This paper describes the approach of the university of Alicante to the problem of finding a suitable handling of multimodal sources within the ImageCLEF ad-hoc competition. We have worked on to add modifications to the most common multimodal techniques used in the image retrieval area in order to improve their performance. Moreover, we have added a clustering module in order to increase the number of different clusters that can be found within the top 20 images returned. Finally, we have studied the effect of using visual concepts in the retrieval phase and in the clustering phase. We can see in the results that with these multimodal techniques we have improved up to a 27% our results in a MAP way, respect the ones obtained using our last year configuration - a textual run using PRF -. Furthermore, we have seen that the use of LCA in a multimodal way outperforms clearly the MAP and P20 results obtained with other common methods used - it has obtained 0.3436 MAP, 4th place in the published task results, and 0.4564 P20, 5th place in the published task results -. Finally our TF-IDF re-ranking run method has showed the best behaviour for the top 20 documents returned from our submissions, obtaining a F-measure value of 0.4051 - based on P20 and CR20 measures -. It makes us to conclude that the combination of these two mutimodal techniques will be the key for improving the performance in our system in future works.

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 Re-ranking, Late Fusion, Intermedia Pseudo Relevance Feedback, Multimodal Relevance Feedback, PRF, LCA, Visual Concepts

1   Introduction

This is the second time we are participating in the ImageCLEFphoto[1] task. Our participation in ImageCLEFphoto of last year involved the use of an information retrieval system based on passages. We analyzed the suitability of our system for the short text annotations related to the images in the collection. We concluded that our system improved the results more in comparison to other systems, which were similar to our system except the fact that they did not use passages. The experiments also showed that relevance feedback is a good tool for improving results [7].

However, we noticed that in spite of the improvements in the general results brought by the relevance feedback - we used PRF [9] relevance feedback strategy -, this process also adds wrong terms for the expansion in some of the cases. Thus, we believe that solving the problem of using non relevant documents for the expansion is an important issue, mainly because in our participation of this year we have planned to add to our system a content based information retrieval (CBIR) system - such systems use only visual information (image characteristic) for IR and nowadays this type of systems have low precision -. Thus, in order to avoid the great number of non relevant documents existent in the image based list, we need an accurate relevance feedback method. Such a solution could help our system to obtain more suitable terms for the expansion of the query.

So far, the most common relevance feedback strategy between the participants of ImageCLEF-photo task on last years edition was to use PRF [9] [4] [3] [5]. Thus, we are comparing in this CLEF edition it with the Local Context Analisy (LCA) [10] as an alternate strategy for a multimodal approach. This strategy has been used in other works within the image retrieval dominion, but only for textual retrieval not in a multimodal way [11].

Furthermore we have added to our system a new operation mode that let to establish a multimodal re-ranking strategy based on mixing the ranking that IR-n returns with the ranking that a CBIR returns. We have experimented with two operation modes, one merges the two list in a classical re-ranking way, and the other mode bases the calculus of the relevance of an image on the quantity and the quality of the text related to the image in order to take the decision of which system is more confident for that image.

Moreover, in order to attempt to bridge the existent semantic gap between the image and the natural language we have taken the opportunity of experiment with the output of one of the participant systems in the visual detection task, which has been made available by the organization.

Finally, in order to improve the recall of the different image cluster detected within the 20 top ranked documents, we have added to our system a clustering module based on the image annotations and the visual concepts related to each image - returned by the visual concept detection system -.

This paper is structured as follows: Firstly, it presents the main characteristics of the IR-n system focusing on the relevance feedback strategies, the multimodal re-ranking strategy, the handling of the image concept annotations and the clustering module, then it moves on to explain the experiments we have made to evaluate the system, and finally it describes the results and conclusions.

2   The IR-n System

In our approach, we used IR-n - an information retrieval system based on passages -. Passage-based IR systems treat each document as a set of passages, with each passage defining a portion of text or contiguous block of text. Unlike document-based systems, these systems can consider the proximity of words with each other, that appear in a document in order to evaluate their relevance [6].

The IR-n passage-based system differs from other systems of the same category with regard to the method proposed for defining the passage - that is - using sentences as unit. Thus, passages are defined by a number of consecutive sentences in a document [6].

---

[1] *http://www.imageclef.org*

IR-n uses stemmer and stopword lists to determine which information in a document will be used for retrieval. For a list of stemmers and stopwords used by IR-n, see www.unine.ch/infor/clef.

IR-n uses several weighting models. Weighting models allow the quantification of the similarity between a text - a complete document or a passage in a document - and a query. Values are based on the terms that are shared by the text and query and on the discriminatory importance of each term.

2.1 Multimodal Relevance Feedback

Under a textual viewpoint, most IR systems use relevance feedback techniques [1]. These systems usually employ local feedback. The local feedback assumes that top-ranked documents are relevant. The added terms are, therefore, common terms from the top-ranked documents. Local feedback has become a widely used relevance feedback technique. Although, it can deter retrieval, in case most of the top-ranked documents are not relevant, results in TREC an CLEF conferences show that is an effective technique [10].

In the selection of terms, PRF gives more importance to those terms which have a higher frequency in the top relevant documents than in the whole collection. An alternative query expansion method relies on the Local Context Analysis (LCA), based on the hypothesis that a common term from the top-ranked relevant documents will tend to co-occur with all query terms within the top-ranked documents. That is an attempt to avoid including terms from top-ranked, non-relevant documents in the expansion. Furthermore, in the case of polysemus words, this method will help to retrieve documents more related to the sense of the query, since it is logical to think that the user will use words from the domain associated with this sense to complete the query.

The IR-n architecture allows us to use query expansion based on either the most relevant passages or the most relevant documents.

Under a multimodal viewpoint, relevance feedback involves the use of two systems, each one threats with a type of media - text or image -. The n-top documents returned by one system are used to enrich the query before use it as input of the other system.

For our multimodal approach we have used the n-top ranked documents from the baseline run of the FIRE system [2] - it was supplied by the task organization -. In order to select the terms to expand the query processed by IR-n, the system lets us to select between PRF and LCA term expansion strategies.

2.2 Multimodal Re-ranking Strategy

This strategy involve the merging of the list returned by the text based IR system and the list returned by the CBIR system. That is done giving a different weight to the normalized relevance value or ranking position for a document in each list. We have included the classical re-ranking strategy and also a variation of it in order to try to improve it and to compare its behaviour.

In the re-ranking strategy, the IR-n list and the CBIR list are merged in order to obtain one final list with documents ranked by relevance - the final relevance (FR) -. The merging process was done by giving different importance to the visual relevance (VR) given for a document in the visual list and the textual relevance (TR) given by the textual IR system:

$$FR(d) = TR(d) * wText + VR(d) * wImg \qquad (1)$$

- where d is a document.

- where VR is a normalized value of the relevance value returned by the CBIR for a document.

- where TR is a normalized value of the relevance value returned by the textual IR system for a document.

Despite this strategy usually improves the results, it usually adds a great number of non relevant images in the ranking. That is due to the low precision that CBIR systems usually obtains. It makes that when we use an image of the CBIR list there are a high level of probability to be selecting a non relevant image.

In an effort for overcome this circumstance, we have modified the re-ranking strategy. We have based our approach on two assumptions. On one hand that the textual list is more confident than the list based on images and on the other hand we assume that the TF-IDF formula is a suitable way to measure the quantity and the quality of a text.

In order to reduce to the minimum the number or non relevant images used from the image based list, we have established a TF-IDF threshold (2). The images which annotations have a TF-IDF value over this threshold are skipped.

$$threshold = (MaxTFIDF * tImg)/100 \qquad (2)$$

- $MaxTFIDF$ is the maximum TF-IDF value found in the image list.

- $tImg$ is a user defined parameter which indicates the percentage value respect the $MaxTTIDF$ in order to work out the threshold.

Thus, in the ranking formula (3) the system uses the *threshold* in order to avoid the risk of use the CBIR relevance values for those images which annotations have enough quantity and quality of text to perform a suitable textual retrieval - without the use of the image -.

$$FR(d) = \begin{cases} TR(d) + VR(d), & \text{if threshold } > TFIDF(d) \\ TR(d), & \text{else} \end{cases} \qquad (3)$$

- where $TFIDF$ is the TF-IDF value of the the text related to an image.

2.3   Clustering Module

Usually, when the users performs a query in a IR system that works with images, they find that there are several similar images between top 20 results. Thus, if they want to find different relevant images they have to navigate through the rest of the returned list.

Our approach is a naive attempt to solve this problem using Carrot2 [2] an open source clustering engine for text. Our clustering module pass the query and the documents of the ranking list - which optionally can be enriched with its related visual concepts - to Caroot2. It uses the texts passed to perform the clustering. For each cluster returned, our module selects in the ranking list the images with the best relevance within its cluster. If there are a number of clusters lower than twenty, the images without a cluster assigned are selected until complete the selection of twenty images. Afterwards, the module adds to the relevance value of the selected images the maximum relevance value in the whole list, in order to take up this images to the top 20 positions in the ranking.

We have worked with Lingo [8], the default clustering algorithm of Carrot2.

Lingo has 2 parameters that influence the number and contents of clusters is creates:

- **Cluster Assignment Threshold**: determines how precise the assignment of documents to clusters should be. For low values of this threshold, Lingo will assign more documents to clusters, which will result in less documents ending up in "Other topics", but also some irrelevant documents making its way to the clusters. For high values of this parameter, Lingo will assign less documents to clusters, which will result in better assignment precision, but also more documents in "Other topics" and less clusters being created.

---

[2] *http://www.carrot2.org*

- **Candidate Cluster Threshold**: determines how many clusters Lingo will try to create, higher values of the parameter will give more clusters. However, it is not possible to exactly predict the number of clusters based on the value of this parameter before clustering is actually run.

3 Training

IR-n is a parametrizable system, which means that it can be adapted in line with the concrete characteristics of the task at hand. The parameters for this configuration are the number of sentences that form a passage, the weighting model to use, the type of expansion, the number of documents/passages on which the expansion is based, the average number of words per document, the use of multimodal relevance feedback, the use of multimodal re-ranking, the use of clustering and the use of visual concepts for the retrieval phase or the clustering phase.

This section describes the training process that was carried out in order to obtain the best possible features for improving the performance of the system.

The collections and resources are described first, and the next section describes specific experiments.

3.1 Data Collection

In this edition of ImageCLEFphoto the IAPR TC-12 collection has been used for the evaluation of the participant systems - it was used for the 2006 and 2007 editions too -.

Table 1 shows the characteristics extracted from the textual annotations in the collection using IR-n splitter.

Table 1: Data Collection

NoDocs	WDAvg	WD Max	SentAvg	Sent Max	Language
20.000	38,2	118	2,6	6	English

Bellow is a descriptions of Table 1 columns:

- **NoDocs**: is the number of documents.

- **WDAvg**: is the average of words by document.

- **WD Max**: is the maximum number of words in a document.

- **SentAvg**: is the average of sentences by document.

- **Sent Max**: is the maximum number of sentences in a document.

- **Language**: is the language of the collection.

IAPR TC-12 corpus uses a semi-structured format based on XML. Thus, as in our last participation we have used for the text retrieval the five text fields corresponding to the title (TITLE), description (DESCRIPTION), notes (NOTES), place (LOCATION), and date of the photo (DATES) as input for IR-n.

The queries also have a semi-structured format, this year the organization has employed the same queries than the last two years, but with two additional tag, the cluster tag, which defines how the clustering of images should take place, and the narrative field, which obtain a longer description about the topic. We can see an example of a query in the Figure 1, there we saw how the text within the narrative tag contains relevant terms for the query, as 'church', 'cathedral' or 'mosque'. Despite that we have the belief that narrative field contains relevant information that could be very useful for a successful query expansion, in order to work with a more realistic environment, where the user does not give this kind of long descriptions, we only have used the title field for our participation.

Figure 1: Metadata File Example

```
<top>
<num> Number: 2 </num>
<title> church with more than two towers </title> <cluster> city </cluster>
<narr> Relvant images will show a church, cathedral or a mosque with three or
more towers. Churches with only one or two towers are not relevant. Buildings
that are not churches, cathedrals or mosques are not relevant even if they
have more than two towers. </narr>
<image> SampleImages/02/16432.jpg </image>
<image> SampleImages/02/37395.jpg </image>
<image> SampleImages/02/40498.jpg </image>
</top>
```

3.2 Experiments

The experiment phase aims to establish the optimum values for the configuration of the system for the collection. We have worked in the training phase with the query subset of this year jointly with the extra data distributed by the organization. A 2007 GIFT baseline run and the output of the visual concept detection system developed by Thomas Desealers from RWTH Aachen University.

Below is a description of the input parameters of the system:

- **The Passage size (ps)**: Number of sentences in a passage.

- **Weight model (wm)**: We used DFR weighting model.

- **Relevance Feedback (relFB)**: Indicating which relevance feedback uses the system - PRF, LCA - and if its the multimodal version - MM -.

- **Relevance Feedback parameters**: If exp has value 1, this denotes we use relevance feedback based on passages. But, if exp has value 2, the relevance feedback is based on documents. Moreover, num denotes the number of passages or documents that the relevance feedback will use from the textual ranking, $ncbir$ denotes the number of documents that the multimodal relevance feedback will use from an image based list and finally, $term$ indicates the k terms extracted from the best ranked passages or documents from the original query.

- **Visual Concept (vcon)**: Indicate if the system has to use the visual concept detection module for the retrieval.

- **Multimodal Re-ranking Strategy (rr)**: Indicate if the system has to use the multimodal re-ranking strategy, and which one. The standard one - RR - or TF-IDF version - RRidf -

- **Multimodal Re-ranking parameters**: $wTxt$ and $wImg$, are the weight of the textual list and the image based list respectively in the standard re-ranking formula. And $tImg$ is the percentage respect the $MaxFIDF$ value found in the image list which the system use to calculate the threshold in order to perform the TF-IDF multimodal re-ranking process.

- **Clustering (clust)**: Indicate if the system has to use the clustering module and whether for the input it has to use the visual concepts annotation jointly with the textual annotations - $clustC$ value - or it only has to use the textual annotations - $clust$ value -.

For the experiments we have worked with a passage size of 3, and with DFR as the weighting schema. We have taken this decision based on the training results obtained on last edition.

In respect the training of the clustering module, we have not done any experiment due to we do not have any training query set to help us with the best clustering parameters selection in

order to increase the number of cluster selected between the 20 top ranked images returned. We have set for all the clustering runs a value of 0.5 for the cluster assignment parameter that is the default value used by Carrot and a value of 20 clusters for the candidate cluster parameter - in order to set within the top 20 ranked images an image pertaining to each cluster -.

In the next tables we show the results obtained in the training phase. In order to evaluate the experiments, we use as evaluation measure the Mean Average Precision (MAP).

In the Table 2 we show the best configurations obtained for each combination of *vcon* and *relFB*, in order to compare the performance of the different relevance feedback techniques used depending of the use of visual concept or not .

Table 2: Best Relevance Feedback Configurations

relFB	vcon	c	avgld	exp	num	ncbir	k	map
NO	NO	8.5	120	0	0	0	0	0.2373
NO	YES	6	120	0	0	0	0	0.2296
PRF	NO	8	80	2	5	0	5	0.2710
LCA	NO	8.5	120	1	5	0	5	0.2667
PRF	YES	9	40	2	5	0	10	0.2756
LCA	YES	8.5	40	2	15	0	5	0.2853
PRF_MM	NO	6	120	2	5	5	5	0.3206
LCA_MM	**NO**	**6**	**40**	**2**	**0**	**20**	**5**	**0.3430**
PRF_MM	YES	8	40	1	5	5	20	0.3192
LCA_MM	YES	8	40	2	0	20	20	0.3337

On one hand we can see that while for the textual relevance feedback runs, the system has obtained better results when it has used visual concepts. On the other hand the use of visual concepts has decreased the performance of the multimodal and baseline runs. It is important to highlight that LCA usually uses a greater number of document for the expansion than PRF, And it is important to highlight too, that for the multimodal runs, PRF best run uses a low number of documents of the top ranked documents of each list - textual list and image based list - while the best LCA run only uses a greater amount of documents of the image based list. It could be explained by the good performance of the LCA method discriminating the non relevant documents of the image based list and by the more enriching terms added by the relevant annotations related to the image based list than the annotations related to the textual list. Since that the relevant documents of the image based list must give access to annotations with terms which a textual IR system difficulty could access -.

Table 3 show us in a synthesized way the relevance feedback results - the data are presented in decreasing order of MAP - in order to make easier the comparison between LCA and PRF. We can see how LCA obtains better results for all the runs that use multimodal information - multimodal relevance feedback and/or visual concepts -.

Table 3: Best Relevance Feedback Results

vcon	map_0	MM	PRF map	LCA map
NO	0.2373	**YES**	0.3206	**0.3430**
YES	0.2296	**YES**	0.3192	**0.3337**
YES	0.2296	NO	0.2756	**0.2853**
NO	0.2373	NO	**0.2710**	0.2667

In order to compare the two basic methods for mixing multimodal results - multimodal re-ranking and multimodal relevance feedback -. For the re-ranking experiments, we only have

mixed with the FIRE baseline list the output returned by the best baselines and the best textual relevance feedback runs. The following table - Table 4 - shows us the re-ranking configurations which best results returned - the data are presented in decreasing order of re-ranking map -.

Table 4: Best Re-ranking Results

relFB	vcon	NoRR map	wTxt	wImg	RR map	% tImg	RRidf map
LCA	YES	0.2853	0.6	0.4	**0.3120**	60	0.3095
PRF	NO	0.2710	0.6	0.4	**0.2943**	60	0.2938
NO	YES	0.2296	0.5	0.5	0.2780	60	**0.2791**
NO	NO	0.2373	0.6	0.4	0.2785	60	0.2785

We can see how the re-ranking always improves the results, and how small is the difference between the results of the two strategies - RR and RRidf -.

The optimal parameters for almost all RR runs is to use a weight of 0.6 for the textual list and 0.4 for the image based list. And for all the RRidf runs 60% is the best value in order to obtain the TF-IDF threshold which decide for each image whether its annotation is enough to evaluate its relevance or is need it the normalized relevance value of the CBIR system for that image in order to add it to the textual relevance.

Finally, the MAP results obtained using the clustering module with each one of the previous runs - the baseline runs, the best multimodal relevance feedback runs and the best multimodal re-ranking runs - are showed in the Table 5. In this table the data are presented in decreasing order of clustering MAP -.

Table 5: Clustering Results

strategy	vcon	map_0	map clust	map clustC
LCAMM	NO	0.3430	0.3070	**0.3260**
LCAMM	YES	0.3337	0.2777	**0.3038**
LCA_RR	YES	0.3120	0.2738	**0.2917**
LCA_RRIDF	YES	0.3095	02688	**0.2837**
PRF_RRIDF	NO	0.2938	0.2523	**0.2656**
LCA	YES	0.2853	0.2473	**0.2648**
PRF_RR	NO	0.2943	0.2498	**0.2645**
PRF	NO	0.2710	0.2314	**0.2437**
NO	NO	0.2373	0.2050	**0.2157**
NO	YES	0.2296	0.1966	**0.2056**

We can see definitely that despite clustering get worse results in MAP way - we expected it happen - the use of visual concepts in the clustering phase improves the performance of the clustering module in comparision. Moreover, we can see in this global results that LCA has been used in all the top ranked runs.

For our participation in ImageCLEFphoto08 we have sent 10 runs corresponding with the runs showed in the Table 6. We have selected for the TEXT modality the best run of the training process with and without clustering for which has not been used any image resource, neither visual concepts nor CBIR list. For the MIXED modality we have sent the best run for each combination of $VisualConcept$, MM, RR and $RRIDF$ techniques with and without clustering using visual concepts .

4 Results in ImageCLEFphoto08

Table 6 shows us the official results obtained by our mixed runs and textual runs respectively in the imageCLEFphoto08 task, we have found a not meaningful difference in our MAP values with the official ones. It has been due to we were using an older version of the trec_eval tool for the experiments than the one used by the organization of the task.

Table 6: Mixed Results in ImageCLEFphoto08

run name	strategy	vcon	clust	Of. MAP	Of. P20	Of. CR20	F-Mea
IRnConcepExpFIREidf	LCA_RRIDF	yes	no	0.3044	0.4359	0.3783	**0.4051**
IRnConcepExpFIRE	LCA_RR	yes	no	0.3114	0.4282	0.3641	0.3936
IRnExpFIREidf	PRF_RRIDF	no	no	0.2847	0.4103	0.3820	0.3956
IRnConcepExpFIREClustC	LCA_RR	yes	clustC	0.2789	0.3628	**0.3989**	0.3800
IRnConcepFBFIRE	LCAMM	yes	no	0.3333	0.4333	0.3316	0.3757
IRnConcepExpFIREidfClustC	LCA_RRIDF	yes	clustC	0.2834	0.3654	0.3771	0.3712
IRnExpFIRE	PRF_RR	no	no	0.2937	0.4013	0.3446	0.3708
IRnFBFIRE	LCAMM	no	no	**0.3436**	**0.4564**	0.3119	0.3706
IRnConcepFBFIREClustC	LCAMM	yes	clustC	0.3032	0.3782	0.3483	0.3626
IRnFBFIREClustC	LCAMM	no	clustC	0.3267	0.3936	0.3353	0.3621
IRnExpFIREidfClustC	PRF_RRIDF	no	clustC	0.2656	0.3397	0.3804	0.3589
IRnExpFIREClustC	PRF_RR	no	clustC	0.2439	0.3205	0.3330	0.3266
IRnExpClustC	PRF	no	clustC	0.2423	0.2603	0.2921	0.2753

Table 7: Textual Results in ImageCLEFphoto08

run name	strategy	vcon	clust	Of. MAP	Of. P20	Of. CR20	F-Mea
IRnExp	PRF	no	no	0.2699	**0.3244**	0.2816	**0.3015**
IRnExpClust	PRF	no	clust	0.2298	0.2192	**0.2967**	0.2521

The data are presented in decreasing order of F-measure. We have worked out this value basing in the official P20 and CR20 measures published by the organization. We use F-measure in order to evaluate which run is the most balanced for P20 and CR20 measures, according with the objective of the task for this year.

Observing the results we can see that despite LCAMM is the best run in MAP and P20 measures - LCAMM has obtained 0.3436 MAP, 4th place in the published task results, and 0.4564 P20, 5th place in the published task results -. The low result in CR20 shades its good performance. Furthermore, we see that TF-IDF re-ranking obtains our best F-measure result, and it is very close to the 10 best ones published by the organization, in F-measure terms - 1st task run 0.4650, 10th task run 0.4148 and our run 0.4051-. In deed all the TF-IDF runs which not use clustering always have better F-measure than the standard re-ranking runs.

Moreover, we see that clustering always decreases the f-meassure, due to this reason, it affects always very negatively the P20 precision, however it does not improve too many the - CR20 - measure. Indeed, it decreases the CR20 measure too for all the TF-IDF re-ranking runs.

Finally we would like to highlight that our best textual run has reached the 6th place in terms of F-measure within the 10 best results published by the organizations in this modality - 1st task run 0.4008, 10th task run 0.2861 and our run 0.3015 -.

5 Conclusion and Future Work

As we expected a major finding of these results is that in MAP terms LCA has obtained a better performance than PRF in a multimodal environment. In fact, it has been the key that has allowed to select more accurately the relevant documents within the image baseline list, avoiding the great number of non relevant documents that a CBIR system usually ranks between the top documents returned.

With respect to the multimodal TF-IDF re-ranking, it has showed that for the top 20 documents returned it accomplishes its objective of incrementing the precision and the variety of images using the CBIR output only when the textual information is not enough. Indeed the worse results obtained with the TF-IDF re-ranking than with the standard re-ranking when they used clustering, are a probe of the good performance of the TF-IDF re-ranking method. Since that the textual clustering drops from the top ranked result the images without meaningful annotations. The fact that TF-IDF re-ranking has had more penalty with the clustering, means that it is returning a greater number of relevant images with annotations without clustering..

In future works we would like to extend the experiments with LCAMM to other query sets and even to other collections in order to confirm what seems by this results a more confident method for the multimodal relevance feedback expansion. Furthermore, in an attempt of improving the global MAP results of the TF-IDF re-ranking, we will study its modification, in order to limit the images used for the re-ranking. Since an image with poor annotations and a medium CBIR relevance value, has low probabilities of being a relevant image. And finally, we will work in experimenting with the joining of LCAMM and TF-IDF re-ranking methods in order to improve the performance of our system..

6 Acknowledgement

References

[1] Aitao Chen and Fredric C. Gey. Combining Query Translation and Document Translation in Cross-Language Retrieval. In Carol Peters, Julio Gonzalo, Martin Braschler, and et al., editors, *4th Workshop of the Cross-Language Evaluation Forum, CLEF 2003, Lecture notes in Computer Science*, Lecture notes in Computer Science, Trondheim, Norway, 2003. Springer-Verlag.

[2] Ney H. Deselaers T., Keysers D. Features for image retrieval: An experimental comparison. *Information Retrieval*, 11(2):77–107, 2008.

[3] Sheng Gao, Jean-Pierre Chevallet, Thi Hoang Diem Le, Trong Ton Pham, and Joo Hwee Lim. Ipal at imageclef 2007 mixing features, models and knowledge. In *Working Notes of the 2007 CLEF Workshop*, Budapest, Hungary, September 2007.

[4] Aurelio Lpez H. Jair Escalante, Carlos A. Hernndez, Heidy M. Marn, Manuel Montes, Eduardo Morales, Luis E. Sucar, and Luis Villaseor. Tia-inaoes participation at imageclef 2007. In *Working Notes of the 2007 CLEF Workshop*, Budapest, Hungary, September 2007.

[5] Anni Jrvelin, Peter Wilkins, Tomasz Adamek, Eija Airio, Gareth J. F. Jones, Alan F. Smeaton, and Eero Sormunen. Dcu and uta at imageclefphoto 2007. In *Working Notes of the 2007 CLEF Workshop*, Budapest, Hungary, September 2007.

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

[7] Sergio Navarro, Fernando Llopis, Rafael Muñoz, and Elisa Noguera. Information Retrieval of Visual Descriptions with IR-n System based on Passages. In *In on-line Working Notes, CLEF 2007*, 2007.

[8] Stanislaw Osinski, Jerzy Stefanowski, and Dawid Weiss. Lingo: Search results clustering algorithm based on singular value decomposition. In *Intelligent Information Systems*, pages 359–368, 2004.

[9] S. E. Robertson and K. Sparck Jones. Relevance weighting of search terms. *Journal of the American Society for Information Science*, 27(3):129–146, 1976.

[10] 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.

[11] Rong Yan and G. Hauptmann. A review of text and image retrieval approaches for broadcast news video. *Information Retrieval*, 10(4-5):445–484, 2007.