A Multimodal Approach to the Medical Retrieval Task using IR-n

Sergio Navarro, Rafael Muñoz, Fernando Llopis

Natural Language Processing and Information Systems Group

University of Alicante, Spain

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

Abstract

In our participation in the Medical Retrieval task we wanted to figure out if good results can be achieved with IR-n - our IR passage based system - for this restricted domain. We have focused on comparing the behaviour of two relevance feedback methods in this task - LCA and PRF -. Furthermore, in order to adapt our system to this task we have used two automatic query expansion techniques related with the medical domain. On one hand we have added to our system an automatic query expansion method based on MeSH ontology and on the other hand we have added a negative query expansion based on the acquisition type of the image. Finally we have added a multimodal re-ranking module - late fusion -. We have used 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 as to which system is more confident for that image - the system based on text or the one based on images -. A major finding of the results is that our passage based system fits very well to this task. Within the textual runs submitted by all the participants we have reached the 6th place for our baseline and the 1st place for a run using PRF and query expansion adapted to the medical domain. Our results for multimodal re-ranking have not been successful due to problems with the parameters tuning for the test collection of this year.

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, Multimodal Relevance Feedback, PRF, LCA, MeSH, Automatic query expansion, Negative query expansion

1   Introduction

This is the first time we are participating in the Medical Retrieval task. We had experience of participation in the photo retrieval task of 2007 . Our participation in ImageCLEFphoto [1] involved the use of an information retrieval system based on passages. We analysed the suitability of our system for the short text annotations related to the images in the collection [11].

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. In our participation in the Medical Retrieval task we wanted to figure out if good results can be also achieved with our passage based system for this restricted domain.

However, we noticed that in spite of the improvements in the general results brought by the relevance feedback - we used PRF [12] relevance feedback strategy -, this process also adds wrong terms for the expansion in some of the cases. Therefore we decided to focus part of our efforts on finding an alternative strategy for the relevance feedback, Thus, we are comparing in this CLEF edition PRF with Local Context Analisy (LCA) [13], as alternate strategy.

Furthermore, in order to adapt our system to this task we have used two automatic query expansion techniques related with the medical domain. On one hand our system adds expanded terms to the queries based on MeSH ontology. And on the other hand when in the query there are terms related to the type of the images that have to be retrieved, the system uses a negative term expansion based on the terms related to other types of a taxonomy of types of images. In order to move away from the top positions of the ranking, those documents which do not belong to the type/s requested in the query.

Another important conclusion in our ImageCLEFphoto participation was that a multimedia approximation along with a suitable mixing procedure would constitute the key to successful participation in this task. Thus, we have added to our system a new operation mode that led to the establishment of 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.

This paper is structured as follows: Firstly, it presents the main characteristics of the IR-n system focusing focusing on the relevance feedback strategies, the automatic query expansion and the multimodal re-ranking strategy, 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 [9].

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 [9].

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.

---

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

2.1 Relevance Feedback

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 [13]. In fact, almost all the systems that participated at ImageCLEF 2007 used Probabilistic Relevance Feedback (PRF) [6] [8] [5] [3].

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. In our experiments we have compared the performance of these two methods within the Medical Domain.

2.2 Automatic Query Expansion

2.2.1 Expanding Queries with MeSH

The Medical Subject Headings (MeSH) is a thesaurus developed by the National Library of Medicine. We have used MeSH terms and its synonyms to expand the queries. If all the words of a term are in the query - a term is composed of one or more words - we expand the query with the synonymous of this term. To compare the words of a particular term and those of the query, we first put all the words in lowercase and we do not remove stopwords. This method was implemented by Sinai group in their 2007 participation [3]. It leads to an improvement of the results, thus we have added it to our system in order to study its behaviour using a passage based system with LCA as relevance feedback method.

2.2.2 Negative Expansion Based on the Acquisition Type of the Image

We have added a module for a negative query expansion. It is based on the acquisition type of the images. In this way, there was a previous work that used a filtering strategy according to the acquisition modality of the images [7]. That work used a supervised machine learning system witch used visual features to classify and to annotate images according to their acquisition modality.

In our approach we neither have used visual features nor filtering strategy but we have used a modified version of the classification proposed at this work. In order to analyse whether the negative expansion based on the acquisition modality as well can improve the precision without the use of visual features.

Our approach use the textual query and the original text annotations for the retrieval. In order to only retrieve images of the desired type in the query, our system looks for the terms that represent each acquisition type - Table 1 - in the query. And if there are terms regarding to one or more acquisition modality types, then the query is expanded with a negative weight for the terms related to the other types witch are not in the query.

2.3 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

Table 1: Taxonomy of Image Acquisition Types

Type	Terms
Angiogram	angiogram, angiograms, angiography, angiographies
CT Scan	ct, cts, tomography, tomographies
MRI:	mri, mris, resonance, resonances
Ultrasound	ultrasound, ultrasounds
Scintigraphy	scintigraphy, scintigraphies
X-Ray	x ray, x rays, xray, xrays
Microscope	microscope
Gross Path	gross path
Endoscopy	endoscopy, endoscopies

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 text related to an image.

3 Training

IR-n is a parameterizable 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 MeSH for the automatic query expansion, the use of negative expansion based on the acquisition type of the image and the use of multimodal re-ranking,.

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 the medical retrieval task the training for our participation has been done with the Consolidated Collection. This collection consists of the image collection and topics used in ImageCLEFmed 2005-2007 merged into one single new collection, with relevance judgments made for all topics based on all collections. This consolidated collection related files consists of six datasets (CASImage, Pathopic, Peir, MIR, endoscopic and MyPACS) containing 66,662 images. Each subcollection is organized into cases that represent a group of related images and annotations. In every case a group of images and an optional annotation is given. Each image is part of a case and has optional associated annotations, which enclose metadata and/or a textual annotation. All the images and annotations are stored into separated files. There are an index file which contains the connections between collections, cases, images and annotations. The collection annotations are in XML format and most of them are in English.

For this year task, a new collection has been used in order to evaluate and to compare the participant systems. This collection is the Goldminer collection [2] The subset used contains all 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. The database distributed include an xml file with the image id, the captions of the images, the titles of the journal articles in which the image had appeared and the PubMed ID of the journal article.

The SINAI group of the University of Jaen kindly has provided us with the preprocessed consolidated collection for this participation- which they used last year for their participation -, and the preprocessed Goldminer collection of this year - only with the captions per article, non full text articles -. They have preprocessed the collections in order to generate a textual document per image, which has been translated to English when it need [4].

Table 2 shows the characteristics extracted from the textual annotations in the preprocessed collections using IR-n splitter and the number of queries related to each collection.

Table 2: Data Collection

Collection	NoImgs	WDAvg	SentAvg	Language	NrQ
Consolidated	66,662	93,87	4,29	English	85
Goldminer	67,115	55,37	3,82	English	30

Bellow is a descriptions of Table 2 columns:

- **NoImgs**: is the number of images.

- **WDAvg**: is the average of words by image annotations.

- **SentAvg**: is the average of sentences by image annotations.

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

---

[2] http://goldminer.arrs.org

- **NrQ**: Number queries related to the collection.

3.2 Experiments

The experiment phase aims to establish the optimum values for the configuration of the system for the collection.

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)**: Indicates which relevance feedback uses the system - PRF, LCA -.

- **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, $term$ indicates the k terms extracted from the best ranked passages or documents from the original query.

- **Automatic query expansion based on MeSH (mesh)**: Indicates if the system has to use this expansion or not.

- **Negative expansion based on the acquisition type of the image (NType)**: Indicates if the systems has to use this negative expansion or not.

- **Multimodal Re-ranking Strategy (rr)**: Indicates if the system has to use the multimodal re-ranking strategy or not, and which one in particular. 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 maximum TF_IDF value found in the image list which the system uses to calculate the threshold in order to perform the TF-IDF multimodal re-ranking process.

For the experiments we have worked with DFR as the weighting schema. We have taken this decision based on the training results obtained with participation in the English monolingual task within the ImageCLEFphoto edition of last year.

Furthermore, is is important to take in account that for the training phase of the re-ranking strategies we have used the University of Geneva CBIR baseline run for the 2007 query set - we have not found a CBIR baseline that answer to the whole consolidated query set -. Since that the re-ranking results have been obtained with runs that for the 2007 queries - 30 queries of a total of 85 queries - use the re-ranking technique and for the other queries they use a purely textual retrieval. Bearing that in mind we have evaluated the results of the re-ranking experiments as if its improvement or worsening was minimized by its partial use with the training query set.

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) and the Recall.

In the Table 3 we show the best configurations obtained for each combination of $NType$, $mesh$ and $relFB$, in order to compare the performance of the different relevance feedback and query expansion methods. It's data is presented in the increasing order of MAP value.

On one hand we can observe that the only run which is able to improve the baseline is the one that only uses the negative expansion based on the image type. On the other hand we have to highlight that the passage size for all the best runs is lower than the average of sentences per image of the collection. Furthermore we can see that PRF always obtains better results than LCA for the same combination of parameters.

Table 3: Best Textual Runs

NType	mesh	ps	c	avgld	relFB	exp	num	k	map	recall
yes	no	3	3.5	120	no	0	0	0	**0.2174**	0.5095
no	**no**	**3**	**3.5**	**120**	**no**	**0**	**0**	**0**	**0.2170**	**0.5048**
no	no	3	7.5	40	prf	1	5	10	0.2142	0.4980
yes	no	2	6	40	prf	1	10	5	0.2132	0.5103
no	no	3	6.5	40	lca	1	20	5	0.2102	0.4993
no	yes	2	6	40	prf	1	10	5	0.2065	0.5052
yes	no	2	5.5	40	lca	1	10	5	0.2053	0.4971
no	yes	2	5.5	40	lca	1	10	5	0.2052	0.4870
yes	yes	3	3	40	no	0	0	0	0.2050	**0.5189**
no	yes	3	3	40	no	0	0	0	0.2043	0.5111
yes	yes	3	7.5	40	prf	1	5	5	0.2029	0.5159
yes	yes	2	3.5	40	lca	1	5	5	0.2020	0.5049

The total number of submissions that we are allowed to send for the official results is 10. Thus, in order to have at least an official result per technique for our participation in the task within the textual modality, we have selected the best three runs in MAP terms, and furthermore, the best run only using LCA in order to compare it with PRF run, and finally, two runs using the two query expansion modules - NType and mesh - with and without relevance feedback - Table 3-.

Table 4: Textual Runs Submitted 2008 task

NType	mesh	relFB	map	recall
no	no	no	0.2170	0.5048
no	no	prf	0.2142	0.4980
no	no	lca	0.2102	0.4993
yes	**no**	**no**	**0.2174**	**0.5095**
yes	yes	no	0.2050	0.5189
yes	yes	prf	0.2029	0.5159

For the training of the two re-ranking techniques, we have worked with the best run of the textual training process and the baseline CBIR run of the University of Geneva for the 2007 query set. Furthermore, in order to complete the number of runs that we are allowed to send we have also trained the re-ranking strategies with the best run in Recall terms, in order to study how it can influence this measure in the final MAP obtained with a re-ranking strategy. The Table 5, shows us the best results obtained in that training process.

Table 5: Best Multimodal Runs

NType/mesh/relFB	map	recall	wTxt/ wImg	rr map	rr recall	% tImg	rrTfIdf map	rrTfIdf recall
yes/no/no	0.2174	0.5095	0.9/0.1	**0.2182**	0.5095	0.00002	0.2176	**0.5101**
yes/yes/no	0.2050	0.5189	0.9/0.1	**0.2065**	0.5189	0.00002	0.2052	**0.5197**

We can see that the two techniques improve the MAP results sightly. We believe that this improvement would have had a greater impact in the overall results of MAP if we had used a CBIR run for all the training query set. Furthermore it is important to highlight that despite the fact that standard re-ranking obtains better MAP results than the TF-IDF re-ranking we can

observe that TF-IDF improves the Recall measure, while the first one does not.

We have sent these four latter runs for our participation. We have mixed our textual runs with a CBIR baseline obtained with FIRE [2] and distributed by the organization. It is important to stand out the extremely low threshold percentage value used in the best TF-IDF re-ranking runs - specially if we compare it with the 60% value obtained for the training phase of the TF-IDF re-ranking technique in our ImageCLEFphoto participation of this year [10] - paper pendent of publication -. This is a risk for the success of the TF-IDF re-ranking strategy in the participation of this year, due to wich we think that this value is very dependent on the characteristics of the collection and for this year edition the competition uses a different collection for testing the participant systems.

4 Results in ImageCLEF Medical Retrieval 2008

Table 6 shows us the official results obtained by our runs in the ImageCLEF Medical Retrieval 2008 task for the textual modality. We can see that our best run has been ranked in the 1st place for the textual systems classification - txt rk - from a total of 69 submitted runs within this classification. And in the 2nd place within the general classification of the task - CLEF rk - from a total number of 119 submitted runs.

Table 6: Official Textual Runs Results

run name	NType	mesh	relFB	map	txt rk	CLEF rk
IRn2ExpPRFNegMesh	**yes**	**yes**	**prf**	**0.2881**	**1**	**2**
IRn2baseline	**no**	**no**	**no**	**0.2768**	**6**	**8**
IRn2ExpNeg	yes	no	no	0.2755	7	9
IRn2ExpNegMesh	yes	yes	no	0.2666	11	14
IRn2Explca	no	no	lca	0.2616	13	17
IRn2ExpPrf	no	no	prf	0.2552	15	19

With the new collection, our best run in the training phase - the run which only uses negative expansion - has obtained worse results than the baseline run. However, the best ranked run in the task is one that in the training phase was not between the best runs. In fact baseline run is the run that showed better results have been shown in the training phase and in the competition. Furthermore, we can see that the LCA submitted run shows better results than the runs which only use PRF - contrary to what we experience in the training phase -.

Table 7 shows our official results within the mixed runs submisions.

Table 7: Official Mixed Runs Results

run name	NType	mesh	relFB	RR	map	mix rk	CLEF rk
EXPNegativaRERANKING-IDF	yes	no	no	rrTfIdf	0.2755	3	9
EXPNegativaMeshRERANKING-IDF	yes	yes	no	rrTfIdf	0.2666	4	14
EXPNegativaRERANKING	yes	no	no	rr	0.0507	20	86
EXPNegativaMeshRERANKING	yes	yes	no	rr	0.0300	31	106

With respect to this results, despite the fact that we have obtained the 3rd position from a total of 36 submitted runs in this modality. However, this results does not represent the ones that can be achieved with these strategies. The reason is that if we observe the MAP values obtained by the TF-IDF re-ranking techniques, we can see that these are the same results obtained by the

textual run with the same parameters. It has happened because the threshold that we have used for TF-IDF re-ranking strategy is too much low for the competition collection. It makes that the system treat all the documents retrieved by the CBIR as if they have enough textual information to perform a suitable retrieval - skipping their CBIR relevance value-. We also have had a tuning problem with the parameters of the standard re-ranking strategy. Since that we only trained it using a subset of the whole training query set and that we have used a different collection and a different CBIR from the ones we used in the competition. All of this has affected negatively to the performance of this strategy with the test collection.

5 Conclusion and Future Work

A major finding of these results is that our passage based systems fits very well to this task, demonstrating at another instance its usefulness in narrow texts domain. Indeed, our baseline run only uses a passage size of 3 sentences and with DFR as weighting schema, has obtained the 6th best result of the textual participant systems. And moreover, it has showed a constant behaviour in the training and in the competition tests.

With respect to the results obtained with the relevance feedback strategies, it is difficult to have a conclusion, because the results in the training and in the participation are opposite. We would like to work further in the future with LCA in order to filter the terms of the query that it uses in order to measure which documents have greater probabilities of be relevant in order to perform the expansion. The idea is to filter those terms of the query which do not pertain to the medical domain.

An interesting result has been to check that the negative query expansion based on the acquisition type of the image always was being used in our best runs - in the training phase and in the competition -, although it have not improved meaningfully the results. It demonstrates that this technique can improve the results even when it is based only on the captions of the image - there are not any visual system tagging the images with its acquisition type -. In future works we would like to compare its behaviour with a visual system for tagging the images.

Furthermore we have to work to analyse the reasons why our system does not improve always its performance when it uses the MeSH query expansion, Especially when this techniques has given good results in other systems [3]. Finally, in order to avoid the problems experienced with the TF-IDF re-ranking strategy, we are planning on to work on finding an alternative method to establish the TF-IDF re-ranking threshold. Which instead of using the documents retrieved, uses the whole collection to work out this value.

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] M.C. Díaz-Galiano, M.A. García-Cumbreras, L.A. Ureña López, M.T. Martín-Valdivia, and A. Montejo-Raez.

[4] M.C. Daz-Galiano, M.A. Garca-Cumbreras, M.T. Martn-Valdivia, A. Montejo-Raez, and L.A. Urea-Lpez. Sinai at imageclef 2007. In *Working Notes of the 2007 CLEF Workshop*, Budapest, Hungary, September 2007.

[5] H. Jair Escalante, Carlos A. Hernndez, Aurelio Lpez, 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.

[6] 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.

[7] William Hersh Jayashree Kalpathy-Cramera. Automatic Image Modality Based Classification and Annotation to Improve. In *MEDINFO 2007*, 2007.

[8] 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.

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

[10] Sergio Navarro, Fernando Llopis, and Rafael Muñoz. Different Multimodal Approaches using IR-n in ImageCLEFphoto 2008. In *In on-line Working Notes, CLEF 2008*, 2008.

[11] 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.

[12] 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.

[13] 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.