

# IPL at ImageCLEF 2010

Alexandros Stougiannis, Anestis Gkanogiannis, and Theodore Kalamboukis

Information Processing Laboratory  
Department of Informatics  
Athens University of Economics and Business  
76 Patission street  
Athens 10434,Greece  
stougiannis@gail.com (utumno,tzk)@aueb.gr

**Abstract.** This paper describes the participation of the IPL team in the ImageCLEF-2010 campaign. Our research group has participated in the ad hoc task the case-based and the modality tasks of image retrieval based only on textual data.

To reach our aim we performed a quantitative evaluation of Lucene search engine using several state of the art similarity functions. Our runs performed moderately well using the MAP and the early precision (P10) metric.

## 1 Introduction

This paper presents the participation of the IPL at the CLEF 2010 Image retrieval task. The main goal of the medical ImageCLEF task is to improve the retrieval of medical images from heterogeneous and multilingual document collections containing images and text. Queries are formulated with sample images and some textual description explaining the search goal.

Image retrieval systems do not currently perform as well as their text counterparts [1]. Although the last decades CBIR systems have been extensively studied with significant advances they still demonstrate poor performance when applied to large databases with a wide spectrum of imaging modalities. However the retrieval performance has shown improvement by fusing the results of textual and visual techniques.

This year's image collection contains approximately 77,000 images. Some records had only one image while others had many images. This situation makes a purely text-based retrieval of images problematic however, since although only one image from a record containing many images might be relevant to a query based on the content, all the images of the record will be equally retrieved, reducing in this way the precision of the search.

Most of our effort this year was concentrated in a quantitative evaluation of a search engine under several state of the art similarity functions.

Our retrieval system is based on Lucene [2]. Lucene is a very popular open-source IR toolkit, which has been used in many search-related applications. Our evaluation was relied on the CLEF-2009 image collection. This test collection

contains a set of about 74000 images, a set of 25 topics and the relevance judgements.

To achieve this goal we incorporated within the lucene search engine three similarity functions, as will see in more details in the next section, namely the BM25, the axiomatic similarity function and the document pivoted normalization function.

Each figure in the database represented in XML format as the following simple example.

```

simple record
< record >
  < pmid > 12345 < /pmid >
  < articleurl >
    http://www.someurl.com/somearticle.html
  < /articleurl >
  < caption > somecaption < /caption >
  < figureid > fig.xyz < /figureid >
  < figureids >
    < figureidss > fig.xyz.a < /figureidss >
    < figureidss > fig.xyz.b < /figureidss >
    < figureidss > fig.xyz.c < /figureidss >
  < /figureids >
  < meshterms >
    < meshterm > meshterm1 < /meshterm >
    < meshterm > meshterm2 < /meshterm >
    < meshterm > meshterm3 < /meshterm >
    < meshterm > meshterm4 < /meshterm >
  < /meshterms >
< /record >

```

We note that several figures referred to the same image with the same title and same captions, so we decided to unify those records into one, with one header "figureid" and the rest figures as relevant ones. Thus we have two different fields for the figures of the case (figureid, figureids), and the first one plays the role of the header figure and the rest the role of relevant figures. Moreover the title and caption of the unified figure were merged into a single field named "caption" that contains all the textual information that accompanies each figure.

For the indexing with the Lucene, we used our own Analyzer which performs tokenization, removes stop words, short words (less than 2 characters), long words (more than 50 characters), transforms each word to lower case, and performs stemming using the Porter stemmer.

## 2 AdHoc Retrieval

The ad hoc task involves retrieving relevant images using the text associated to each image query. For this task we investigated four similarity functions with the lucene search engine: the default similarity function[3], the BM25 [4, 5], the

axiomatic [6, 7] and the default with document pivoted normalization [8]. The aim was to examine the behavior of each of these similarity functions on small documents as the ones of the CLEF database. The evaluation was based on the CLEF-2009 database. In the following we give briefly the definition of each function. More details can be found in [11].

1. BM25

$$score(D, Q) = \sum_{t \in Q \wedge D} IDF(t) \frac{tf(t, D) \cdot (k_1 + 1)}{tf(t, D) + k_1 \cdot (1 - b + b \cdot \frac{dl}{avgdl})} \quad (1)$$

The inverse document frequency is computed according to the classical BM25 model:

$$IDF(t) = \log \frac{N - n(t) + 0.5}{n(t) + 0.5} \quad (2)$$

where  $N$  is the number of documents in the collection and  $n(t)$  is the number of documents where the term  $t$  appears.

2. Axiomatic

It has been shown in an axiomatic approach to develop retrieval functions proposed in [6] that the following scoring function (2) is more robust than other state-of-the-art with comparable optimal performance. The following function (F2-EXP) according to [6, 7] is one of the best performing functions.

$$score(D, Q) = \sum_{t \in Q \wedge D} tf(t, Q) \cdot \left( \frac{N}{df(t)} \right)^{0.35} \frac{tf(t, D)}{tf(t, D) + 0.5 + \frac{0.5 \cdot |D|}{avgdl}} \quad (3)$$

where  $tf(t, Q)$  is the number of occurrences of term  $t$  in query  $Q$ ,  $df(t)$  is the number of documents that contain term  $t$ ,  $|D|$  is length of document  $D$ , and  $avgdl$  is the average document length in the collection.

3. Document Pivoting Normalization

This approach uses the default scoring function of lucene but  $tf(t, D)$  which is given by the square root of the term frequency is now estimated by the relation (4)

$$tf(t, D) = \frac{1 + \log(freq(t, D))}{1 + \log(avgFreq(D))} \quad (4)$$

where  $avgFreq(D) = \frac{1}{|U_D|} \sum_{t \in d} freq(t, D)$ ,  $U_D$  =the discrete terms in  $D$ ,  $lengthNorm = \frac{1}{\sqrt{(1-slope) \cdot pivot + slope \cdot |U_D|}}$  and  $pivot = \frac{1}{|D|} \sum_{d \in D} |U_D|$

The results out of these four functions from the CLEF-2009 image-collection are presented in table 1.

In table 2 are presented the Ad Hoc runs we submitted and a short description for each one.

**Table 1.** Performance of scoring functions on the CLEF-2009 image collection.

<b>CLEF-2009</b>	<b>Deafult Lucene</b>	<b>BM25</b>	<b>Axiomatic</b>	<b>Pivoted Norm</b>
numrel	2362	2362	2362	2362
numrelret	1844	1812	1847	1847
map	0.4070	0.3513	0.3835	0.4132
R-prec	0.4462	0.3827	0.4169	0.4546
bpref	0.4297	0.3614	0.3979	0.4390
P5	0.6800	0.6160	0.6480	0.6880
P10	0.6040	0.5760	0.6040	0.6520
P20	0.6060	0.5340	0.5720	0.6080
P30	0.5720	0.5107	0.5293	0.5627

**Table 2.** Ad Hoc runs on the 2010 image collection.

<b>Ad Hoc - submitted runs</b>	<b>Description</b>
ipl aueb AdHoc default TC	Default Lucene Indexing, one searchable field ("caption")
ipl aueb adhoq Pivoting TC	Pivoting normalization Lucene indexing, one searchable field ("caption")
ipl aueb adhoq default TCg	Default Lucene Indexing, one searchable field ("caption"), alternative method for duplicate figures recognition
ipl aueb adhoq Pivoting TCg	Pivoting normalization Lucene indexing, one searchable field ("caption"), alternative method for duplicate figures recognition
ipl aueb adhoq default TCM	Default Lucene Indexing, two searchable fields ("caption", "mesh")
ipl aueb adhoq Pivoting TCM	Pivoting normalization Lucene indexing, two searchable fields ("caption", "mesh")

In the AdHoc version of the Medical Retrieval, the unit of retrieval is a figure. A figure is identified by the figureid. The goal of the retrieval is to correctly retrieve the most relevant figures for each one of the 16 queries-figures. To complete this task we create indexes with the Lucene search engine using only the textual information of the figures (title + caption).

Moreover as we can see in the runs with suffix TCM the documents are expanded with the MeSH terms. In this case we have the two searchable fields. In that case we weight the two fields by ratio 0.9/0.1 in favor of "caption" over the "mesh". The scoring function is given by:

$$score(Q, D) = 0.9 * score(Q, D_C) + 0.1 * score(Q, D_{MeSH}) \quad (5)$$

where the field  $D_C$  contains the title and caption and the  $D_{MeSH}$  the MeSH terms.

**Table 3.** Ad Hoc top 10 performing participations based on MAP.

Run name	MAP
XRCE AX rerank comb,trec	0.3572
WIKI AX MOD late,trec	0.3380
ipl aueb AdHoc default TC	0.3235
ipl aueb adhoq default TCg	0.3225
ipl aueb adhoq default TCM	0.3209
XRCE CHI2 LOGIT IMG MOD late,trec	0.3167
ipl aueb AdHoc pivoting TC	0.3155
ipl aueb adhoq Pivoting TCg	0.3145
XRCE AF LGD IMG late,trec	0.3119
ipl aueb adhoq Pivoting TCM	0.3102

**Table 4.** Ad Hoc top 10 performing participations based on MAP.

Run name	P@10
WIKI AX MOD late,trec	0.5062
ipl aueb AdHoc default TC	0.4687
ipl aueb adhoq default TCM	0.4687
ipl aueb adhoq default TCg	0.4562
ipl aueb AdHoc pivoting TC	0.4500
ipl aueb adhoq Pivoting TCg	0.4500
ipl aueb adhoq Pivoting TCM	0.4375
XRCE AX rerank comb,trec	0.4375
XRCE AF LGD IMG late,trec	0.4375
OHSU pm major all mod	0.4375

In table 3 and 4 we present the top 10 performing participations in the Ad Hoc track based on the MAP metric and the early precision (P@10).

### 3 Case based Retrieval

In the Case-based version of the Medical Retrieval, the unit of the retrieval is not an image but a case. Cases are identified by a PubMed id ("pmid" tag in the xml format). Unfortunately not all cases are marked with a "pmid", so the identifier is the "ArticleURL" tag of the xml format. In other words, a case is pointed by a link in the web of an article that describes it.

A simple fusion method has been implemented to obtain a simple list of relevant figures. The total score of each relevant document is calculated by the sum of the scores for all the retrieved figures in the document:

$$score(Q, D) = \sum_{F \in D} score(Q, F) \quad (6)$$

**Table 5.** Case based runs on the 2010 image collection.

<b>Ad Hoc - submitted runs</b>	<b>Description</b>
ipl aueb casebased CT	Unified caption and title "contents" searchable field
ipl aueb casebased CTM 0.1	"contents" and "mesh" terms searchable fields, 0.1 ratio of "mesh"
ipl aueb casebased CTM 0.2	"contents" and "mesh" terms searchable fields, 0.2 ratio of "mesh"
ipl aueb casebased CTM 0.3	"contents" and "mesh" terms searchable fields, 0.3 ratio of "mesh"
ipl aueb casebased CTM 0.4	"contents" and "mesh" terms searchable fields, 0.4 ratio of "mesh"
ipl aueb casebased CTM 0.5	"contents" and "mesh" terms searchable fields, 0.5 ratio of "mesh"

**Table 6.** Case-based top-10 performing participations based on MAP.

<b>Run name</b>	<b>MAP</b>
PhybaselinefbWMR 10 0.2sub	0.3165
baselinefbWMR 10 0.2sub	0.2926
PhybaselinefbWsub	0.2699
PhybaselinefbWMD 25 0.2sub	0.2699
baselinefbWsub	0.2553
IRIT SemAnnotator-1.5.2 BM25 N34 1.res	0.2521
IRIT SemAnnotator-2.0 BM25 N34 1.res	0.2521
IRIT SemAnnotator-1.5.2 BM25 N34.res	0.2444
IRIT SemAnnotator-2.0 BM25 N28 1.res	0.2444
PhybaselineRelfbWMR 10 0.2sub	0.2435

Finally, the documents are sorted by their final fusion score.

Table 5 summarizes the runs we submitted and a short description for the Case-based task.

Similarly in this case we use two types of indexes. For the first type, only the field "contents" (title+caption) is searchable. At the retrieval phase we get the top 1000 results and iterate over them to add the score of the results with the same "articleurl". This procedure produces the run named "ipl aueb casebased CT".

In the second type of indexes, there is an additional searchable field "mesh", we performed 5 runs. The number in the run's name denotes the weight given in the "mesh" field. A value 0.1 means a weight 0.1 for the score of the "mesh" field and 0.9 for the score of the "contents" field.

Table 6 contains the top 10 performing participations in the case-based track based on the MAP metric. IPL was at the 28<sup>th</sup> position at the sequence with MAP equal to 0.1228.

**Table 7.** IPL’s Modality results based on accuracy.

Run name	Accuracy
ipl aueb rhcpp full CT	0.74
ipl aueb rhcpp full CTM	0.71
ipl aueb svm full CT	0.53
ipl aueb svm full CTM	0.49

## 4 Modality Runs

For the task of modality classification, a list of 2,390 records was given for training. The goal was to correctly classify the 2,620 records of the test set into one of the eight modalities: "CT" , "GX" , "MR" , "NM" , "PET" , "PX" , "US" , "XR".

We used two classifiers on this task: a modified Perceptron-type algorithm we have described in previous work [9] and the well known SVM classifier [10]. Again we used two types of indexes, with and without MeSH terms. Table 7 contains our results of the modality classification

## 5 Conclusions

Our baseline textual system performed quite well, with a MAP of 32%. We will continue to improve our image retrieval system by adding more image tags using automatic visual feature extraction and heuristics. There was only a minimal improvement in performance with the use of the image modality. As future research we plan to improve the pseudo-relevance feedback strategy.

## 6 References

1. Hersh, W, Muller H, et al. *Advancing biomedical image retrieval: development and analysis of a test collection*. J. Am. Med. Inform. Assoc. 13(5), 488-96, 2006.
2. Lucene. <http://lucene.apache.org/java/docs/>.
3. Lucene’s default similarity function, <http://lucene.apache.org/java/3.0.1/api/core/org/apache/lucene/search/Similarity.html>
4. K. Sparck Jones, S.Walker, and S. E. Robertson. *A probabilistic model of information retrieval: development and comparative experiments*. Information Processing Management, 36(6):779 808, 809 840, 2000.
5. [http://en.wikipedia.org/wiki/Okapi\\_BM25](http://en.wikipedia.org/wiki/Okapi_BM25)
6. H. Fang and C. Zhai. *An exploration of axiomatic approaches to information retrieval*. In Proceedings of the 2005 ACM SIGIR Conference on Research and Development in Information Retrieval, 2005.

7. *Evaluation of the Default Similarity Function in Lucene*, Hui Fang, ChengXiang Zhai, July 15, 2007
8. *Pivoted Document Length Normalization*, A. Singhal, C. Buckley, M. Mitra, Proceedings of the 19th annual international ACM SIGIR conference on Research and development in information retrieval, 21 - 29, 1996
9. Gkanogiannis A. and Kalamboukis T., *A modified and fast Perceptron learning rule and its use for Tag Recommendations in Social Bookmarking Systems*, 71-83, Folke Eisterlehner, et al., (Eds.), ECML PKDD Discovery Challenge 2009 (DC09), International Workshop at the ECML/PKDD in Bled, Slovenia, September 7th, 2009.
10. Joachims, T., *Making large-scale support vector machine learning practical* in Scholkopf, B. et al, Advances in kernel methods: support vector learning, MIT Press, 1999, 169-184.
11. <http://ipl.cs.aueb.gr/eng/stougiannis/>