

# FCSE at Medical Tasks of ImageCLEF 2013

Ivan Kitanovski, Ivica Dimitrovski, and Suzana Loskovska

Faculty of Computer Science and Engineering, University of Ss Cyril and Methodius  
Rugjer Boshkovikj 16, 1000 Skopje, Macedonia  
{ivan.kitanovski, ivica.dimitrovski, suzana.loshkovska}@finki.ukim.mk

**Abstract.** This paper presents the details of the participation of FCSE (Faculty of Computer Science and Engineering) research team in ImageCLEF 2013 medical tasks (modality classification, ad-hoc image retrieval and case-based retrieval). For the modality classification task we used SIFT descriptors and *tf - idf* weights of the surrounding text (image caption and paper title) as features. SVMs with  $\chi^2$  kernel and one-vs-all strategy were used as classifiers. For the ad-hoc image retrieval task and case-based retrieval we adopted a strategy which uses a combination of word-space and concept-space approaches. The word-space approach uses the Terrier IR search engine to index and retrieve the text associated with the images/cases. The concept-space approach uses Metamap to map the text data into a set of UMLS (Unified Medical Language System) concepts, which are later indexed and retrieved by the Terrier IR search engine. The results from the word-space and concept-space retrieval are fused using linear combination. For the compound figure separation task, we used unsupervised algorithm based on breadth-first search strategy using only visual information from the medical images. The selected algorithms were tuned and tested on the data from ImageCLEF 2012 medical task and based on the selected parameters we submitted the new experiments for ImageCLEF 2013 medical task. We achieved very good overall performance: the best run for the modality classification ranked 2nd in the overall score, the best run for the ad-hoc image retrieval ranked 3rd.

**Keywords:** information retrieval, medical imaging, medical image retrieval, modality classification, compound figure separation

## 1 Introduction

In this paper we present the experiments performed by the Faculty of Computer Science and Engineering (FSCE) team for the medical tasks at ImageCLEF 2013. Our group participated in all medical subtasks. To acquire the optimal parameters we evaluated our approaches on the ImageCLEF 2012 dataset and then based on those parameters we submitted the runs for ImageCLEF 2013.

The paper is organized as follows: Section 2 describes our approach for the modality classification task, section 3 shows the algorithm for the compound separation task, section 4 presents the ad-hoc image retrieval task, section 5 contains the details for the case-based retrieval task.

## 2 Modality classification task

### 2.1 Introduction

Imaging modality is an important information on the image for medical retrieval. In user studies, clinicians have indicated that modality is one of the most important filters that they would like to be able to limit their search by. Using the modality information, the retrieval results can often be improved significantly. The ImageCLEF 2013 medical modality classification task is a standardized benchmark for systems to automatically classify medical image modality from PubMed journal articles [1]. The 2013 dataset has 31 classes (the same number of classes and the same classification hierarchy as in 2012) but larger number of compound figures are present making the task significantly harder but corresponding much more to the reality of biomedical journals [1].

Our approach uses visual features with combination of textual features extracted from the surrounding text content of the images. SVMs with  $\chi^2$  kernel were used as a classifiers. The algorithms are explained in details in the remainder of this section.

### 2.2 Visual features

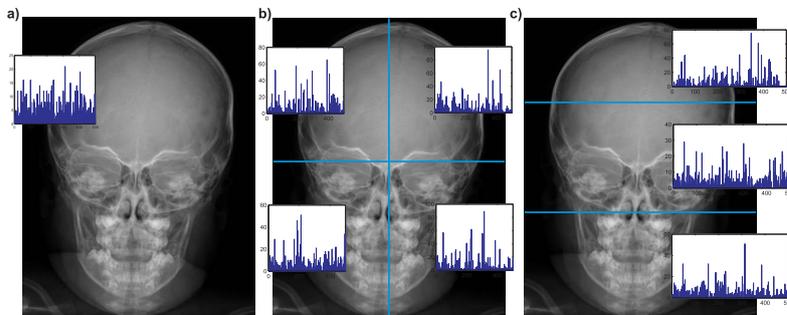
Collections of medical images can contain various images obtained using different imaging techniques. Different feature extraction techniques are able to capture different aspects of an image (e.g., texture, shapes, color distribution...) [2]. Texture is especially important, because it is difficult to classify medical images using shape or gray level information. Effective representation of texture is needed to distinguish between images with equal modality and layout. Local image characteristics are fundamental for image interpretation: while global features retain information on the whole image, the local features capture the details. They are thus more discriminative concerning the problem of inter and intra-class variability [3].

The bag-of-visual-words approach is commonly used in many state of the art algorithms for image classification [4]. The basic idea of this approach is to sample a set of local image patches using some method (densely, randomly or using a key-point detector) and calculate a visual descriptor on each patch (SIFT descriptor, normalized pixel values). The resulting distribution of descriptors is then quantified against a pre-specified visual codebook which converts it to a histogram. The main issues that need to be considered when applying this approach are: sampling of the patches, selection of the visual patch descriptor and building the visual codebook.

We use dense sampling of the patches, which samples an image grid in a uniform fashion using a fixed pixel interval between patches. We use an interval distance of 6 pixels and sample at multiple scales ( $\sigma = 1.2$  and  $\sigma = 2.0$ ). Due to the low contrast of some of the medical images (for example, radiographs), it would be difficult to use any detector for points of interest. Also, it has been pointed by Zhang et al. [4], that a dense sampling is always superior to any

strategy based on detectors for points of interest. We calculate an opponentSIFT descriptor for each image patch [5], [6]. OpponentSIFT describes all the channels in the opponent color space using SIFT descriptors. The information in the  $O_3$  channel is equal to the intensity information, while the other channels describe the color information in the image. These other channels do contain some intensity information, but due to the normalization of the SIFT descriptor they are invariant to changes in light intensity [6].

The crucial aspect of the bag-of-visual-words approach is the codebook construction. An extensive comparison of codebook construction variables is given by van Gemert et al. [7]. We employ  $k$ -means clustering on 250K randomly chosen descriptors from the set of images available for training.  $k$ -means partitions the visual feature space by minimizing the variance between a predefined number of  $k$  clusters. Here, we set  $k$  to 500 and thus define a codebook with 500 codewords [3].



**Fig. 1.** Three different spatial pyramids used in our experiments, a) 1x1, b) 2x2 and c) 3x1. The spatial pyramid constructs feature vectors for each of the specific part of the image.

Dense sampling gives an equal weight to all key-points, irrespective of their spatial location in the image. To overcome this limitation, we follow the spatial pyramid approach [8]. We used a spatial pyramid of 1x1, 2x2, and 1x3 regions. Since every region is an image in itself, the spatial pyramid can easily be used in combination with dense sampling. The resulting vector with 4000 bins  $((1 \times 1 + 2 \times 2 + 1 \times 3) \times 500)$  was obtained by concatenation of the eight histograms (each histogram is  $L1$  normalized). Fig. 1 shows an example of the histograms extracted from an image for the spatial pyramids of 1x1, 2x2 and 3x1.

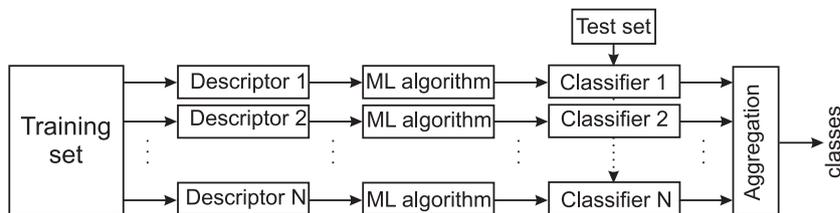
### 2.3 Textual features

Images in the collection belong to a medical article, so they can be indexed using the surrounding text content. The text representation adopted in this work included information from the title of the paper and the image caption,

which can be found in the XML file corresponding to each image in the data set. With that, a text corpus for the image collection was built, and standard text processing operations were applied, including tokenization, stemming, and stop-word removal using Terrier IR [9]. We calculate the weight for each term in each document using  $TF - IDF$  weighting model. The calculated weights were adopted as textual features.

## 2.4 Feature fusion schemes

Different features (in our case visual and textual) bringing different information about the content of the images clearly outperform single feature approaches [10], [3]. Following these findings, we combine the two different features described above using high level feature fusion scheme. The fusion schemes is depicted in Fig. 2.



**Fig. 2.** High level fusion scheme for the different descriptors.

The high level fusion scheme averages the predictions from the individual classifiers trained on the separate descriptors.

## 2.5 Classifier setup

We used the libSVM implementation of SVMs (Support Vector Machines) [11] with probabilistic output [12] as classifiers. To solve the multi-class classification problems, we employ the *one-vs-all* approach. Each of the SVMs was trained with a  $\chi^2$  kernel. Namely, we build a binary classifier for each modality/class: the examples associated with that class are labeled positive and the remaining examples are labeled negative. This results in an imbalanced ratio of positive versus negative training examples. We resolve this issue by adjusting the weights of the positive and negative class [6]. In particular, we set the weight of the positive class to  $\frac{\#pos + \#neg}{\#pos}$  and the weight of the negative class to  $\frac{\#pos + \#neg}{\#neg}$ , with  $\#pos$  the number of positive instances in the train set and  $\#neg$  the number of negative instances. We also optimize the cost parameter  $C$  of the SVMs using an automated parameter search procedure [6]. For the parameter optimization, we used the dataset from 2012. After finding the optimal  $C$  value, the SVM is trained on the 2013 set of training images.

## 2.6 Results and discussion

In this section, we present and discuss the results obtained from the experimental evaluation of the proposed method. First, we compare and evaluate the performance of the proposed method for the ImageCLEF 2012 dataset. Next, we present the results obtained for this year, ImageCLEF 2013 dataset.

The first three rows in Table 1 show the results of our method applied on the ImageCLEF 2012 dataset. These results include visual, textual and mixed runs. From the presented results, we can note that the better predictive performance of the visual run compared to the textual run. The high level feature fusion scheme helps in increasing the predictive performance. Furthermore, from the presented results, we can also note that our method has a very high accuracy/performance. Compared with the results from the groups that participate in the ImageCLEF 2012 medical task [13] our visual run is second best, the textual and mixed runs are ranked first. The mixed run with accuracy of 77.0 will be ranked first in the overall ranking if we have submitted this run in the last years modality classification task.

**Table 1.** Results of the runs of modality classification task for ImageCLEF 2012 and 2013.

Dataset	Run Type	Accuracy
ImageCLEF 2012	visual	66.10
	textual	62.90
	mixed	77.00
ImageCLEF 2013	visual	77.14
	textual	63.71
	mixed	78.04

The second three rows in Table 1 shows the results of our method applied on this year modality classification task. These results include also visual, textual and mixed runs. The accuracy of 78.04 obtained with the mixed run is second best in the overall ranking. The high level feature fusion scheme increases the predictive performance for this year dataset also.

## 3 Compound figure separation

Compound figures contain figures of several types, they cannot be classified into unique classes and need to be separated before a detailed classification into the figure types can be performed. In this work, a unsupervised technique of compound figure separation is proposed and implemented based on breadth-first search strategy using only visual information from the medical figures. All pixel values in the figure are examined/traversed searching for enclosed region separated with white border/pixels. The sensitivity of the border is controlled by threshold parameter. The regions smaller than predefined value are discarded.

In some of the figures the separating borders between the contained subfigures are in black color, therefore before applying our algorithm we invert the output figure. For the given test dataset our algorithm correctly classified 68.59% of the figures.

## 4 Ad-hoc image retrieval

In this section, we give an overview of the application of our methods to ad-hoc medical image retrieval and present the results of our submitted runs. We participated only in the textual retrieval.

### 4.1 Proposed approach

The approach uses the image caption and the title of medical article in which it is referenced i.e. surrounding text. The approach seeks to combine word-space and concept-space approaches with the goal to achieve better overall retrieval performance.

The word-space component indexes and retrieves the surrounding text of the medical images in a traditional way. The surrounding text of the medical images is first preprocessed performing stop words removal and stemming, and creating a standard inverted index. In the retrieval phase, the system pre-processes the query and applies stop words removal and stemming to the query as well. Weighting models are applied to calculate the score for the relevancy of every medical article in respect to the given query. Once the score is calculated the documents are sorted and returned.

The concept-space component works by analyzing the text by the presented medical concepts. The first step is to map the surrounding text of the medical images to medical concepts. The mapping can be done using a variety of toolkits, services or libraries such as [14], Meshup [15] etc. The problem in this approach arises in the way documents will be indexed and then evaluated in the retrieval phase with respect to queries. Classical information retrieval models, directly or indirectly, depend on the number of terms which the document and query share to compute the relevance score [9]. But, the number of terms which a query and document share in the word-space could be very different in the concept-space. For example, if a query and the document share one term "x-ray" in word-space, they can share up to six terms in concept-space [16]. On the other hand, if they share a phrase of two terms "lung x-ray" in word-space, then they will share only one term in concept-space.

The results from both components are then normalized and passed to a fusion component (the diagram is depicted on Figure 3). It can use any of the known strategies for late fusion [17]. In this study, we used a simple linear combination of the normalized results.

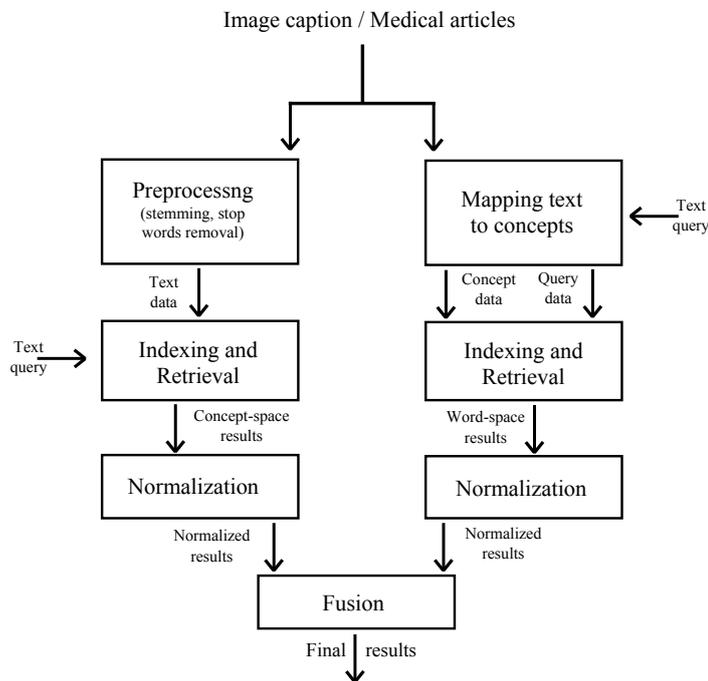


Fig. 3. Diagram of the process flow

## 4.2 Retrieval framework

For the word-space approach Terrier IR [9] is used as a search engine. For the preprocessing stage, Porter stemmer [18] and stop words are applied. In the retrieval phase, several weighting models were evaluated: PL2 [19], BM25 [19], BB2 [19], DFR-BM25 [19], TF-IDF [20], DirichletLM [21]. Additional experiment was performed with query expansion on the best performing model to test its maximum output.

The concept-space approach requires a mapping mechanism to match the text data to medical concepts. In this approach, Metamap is used as mapping tool and the extracted medical concepts are UMLS [14] concepts. The mapping is performed only on the surrounding text of the medical images. After the concepts are extracted, new artificial text is generated containing only the UMLS concepts. The same process is repeated for the queries. Once the artificial text is constructed it is passed to the search engine for indexing. Terrier IR indexes the artificial text, with no additional preprocessing (no stemming and stop words removal). The retrieval is performed by passing the artificial queries to the search engine. In this phase, the same weighting models are applied as in the word-space

approach. Basically, the concept-space approach can be viewed as a word-space approach with more complex preprocessing.

Before the fusion phase, the results from the word-space and concept-space are normalized using min-max normalization [22]. The normalized results are then passed to the fusion component which applies linear combination. This kind fusion provides modularity and control over the extent in which components influence the final result.

### 4.3 Evaluating on ImageCLEF 2012

The proposed framework was first evaluated on the ImageCLEF 2012 dataset. This phase is used to find the optimal weighting models and appropriate parameters. The results of the word-space assessment are depicted on Table 2. The results show that the BM25 model provides the best performance for the word-space retrieval. An additional experiment was performed with the best model by assigning weights to key words in the queries using Terri query language (For example. words such as "MRI", "CT" etc. are given 1.5 weight). The results for the experiment with the word weights (BM25-ww) show an increase in performance.

**Table 2.** Comparison of weighting models for word-space ad-hoc retrieval

<i>Model</i>	<i>MAP</i>	<i>P10</i>	<i>P20</i>	<i>Rprec</i>	<i># of rel. docs</i>
BB2	0.2056	0.3429	0.2714	0.2411	473
BM25	0.2266	0.3381	0.3000	0.2559	494
DFR-BM25	0.2091	0.3476	0.2738	0.2236	474
PL2	0.2055	0.3429	0.2643	0.2353	472
TF-IDF	0.2085	0.3524	0.2714	0.2194	471
DirichletLM	0.1601	0.2619	0.2024	0.1614	434
BM25-ww	0.2407	0.3619	0.2929	0.2620	490

The results of the concept-space assessment are depicted on Table 3. In this case the best results are provided with the DirichletLM model.

**Table 3.** Comparison of weighting models for concept-space ad-hoc retrieval

<i>Model</i>	<i>MAP</i>	<i>P10</i>	<i>P20</i>	<i>Rprec</i>	<i># of rel. docs</i>
BB2	0.1257	0.1700	0.1025	0.1433	173
BM25	0.1230	0.1550	0.1025	0.1441	172
DFR-BM25	0.1227	0.1550	0.1000	0.1441	173
PL2	0.1065	0.1500	0.0950	0.1137	168
TF-IDF	0.1226	0.1550	0.1025	0.1402	172
DirichletLM	0.1568	0.2450	0.1475	0.1888	232

The results of the mixed assessment are depicted on Table 4. The mixed assessment is consisted of two experiments. The first one is by combining the

best word-space and concept-space approaches. The second experiment is done by combining the word-space with word weights and concept-space approaches.

**Table 4.** Results for the mixed run of the ad-hoc retrieval on ImageCLEF 2012

<i>Type</i>	<i>MAP</i>	<i>P10</i>	<i>P20</i>	<i>Rprec</i>	<i># of rel. docs</i>
Mixed	0.2385	0.3762	0.2738	0.2496	492
Mixed-ww	0.2528	0.3857	0.2690	0.2600	488

#### 4.4 Results and discussion

Based on the results obtained from the experiments over the ImageCLEF 2012 dataset, the runs for the ImageCLEF 2013 ad-hoc retrieval task was submitted. Another experiment was made, only now using ImageCLEF 2013 data and submitted the results only from the best performing techniques. For word-space text-based retrieval we submitted the run using BM25 weighting model word weights and for the concept-space text-based retrieval we submitted the run using DirchletLM weighting model. Finally, for the mixed retrieval we submitted the linear combination of the two previous spaces. The results from our runs on ImgeCLEF 2013 are presented on Table 5.

**Table 5.** Submitted runs for ad-hoc retrieval

<i>Type</i>	<i>MAP</i>	<i>GM-MAP</i>	<i>bpref</i>	<i>P10</i>	<i>P30</i>
word-space	0.2435	0.0430	0.2424	0.3314	0.2248
word-space-ww	0.2507	0.0443	0.2497	0.3200	0.2238
concept-space	0.1456	0.0244	0.1480	0.2000	0.1286
mixed	0.2464	0.0508	0.2338	0.3114	0.2200
mixed-ww	0.2479	0.0515	0.2336	0.3057	0.2181

## 5 Case-based retrieval

In this section, we give an overview of the application of our methods to case-based retrieval and present the results of our submitted runs. We participated only in the textual retrieval of the cases.

### 5.1 Proposed approach

The proposed approach for this task is similar to the ad-hoc retrieval task, with the difference that in this case the retrieval unit is a medical article, not an image.

Two approach combines the word-space and concept-space, just as with the ad-hoc retrieval. For the word-space component, we index the entire text of the medical articles, which includes the title, abstract, article text and captions of the images in the article (we refer to this as "fulltext"). The indexing and retrieval is done using Terrier IR and several weighing models are applied to analyze their performance for this type of task. For the concept-space component, only the title and abstract of the medical article are used for extraction of medical concepts. The tool for medical concept extraction is Metamap and the extracted results are UMLS concepts. The rest of the process for the concept-space approach is identical to the concept-space ad-hoc retrieval. The final result is provided with the late fusion of both components using linear combination.

## 5.2 Evaluating on ImageCLEF 2012

The proposed framework was again evaluated on the ImageCLEF 2012 dataset. The results of the word-space assessment are depicted on Table 6. The results show that the BM25 model provides the best performance for the word-space case-based retrieval. An additional experiment was performed with the best model by adding query expansion. The results for the experiment with the query expansion (BM25-qe) show that the query expansion increase retrieval performance by roughly 4%.

**Table 6.** Comparison of weighting models for word-space case-based retrieval

<i>Model</i>	<i>MAP</i>	<i>P10</i>	<i>P20</i>	<i>Rprec</i>	<i># of rel. docs</i>
BB2	0.1598	0.1435	0.1326	0.1604	217
BM25	0.1818	0.1522	0.1391	0.1757	222
DFR-BM25	0.1816	0.1522	0.1413	0.1767	222
PL2	0.1780	0.1478	0.1370	0.1861	227
TF-IDF	0.1805	0.1522	0.1326	0.1662	221
DirichletLM	0.1811	0.1652	0.1283	0.1744	225
BM25-qe	0.1994	0.1957	0.1522	0.2198	232

The results of the concept-space assessment are depicted on Table 7. In this case the best results are provided with the DirichletLM model. An additional experiment was performed using query expansion on the best performing model, which provides an improvement of roughly 2%.

The results of the mixed assessment are depicted on Table 8. The mixed assessment is consisted of two experiments. The first one is by combining the best word-space and concept-space approaches. The second experiment is done by combining the word-space and concept-space approaches, both with added query expansion.

**Table 7.** Comparison of weighting models for concept-space case-based retrieval

<i>Model</i>	<i>MAP</i>	<i>P10</i>	<i>P20</i>	<i>Rprec</i>	<i># of rel. docs</i>
BB2	0.0691	0.1000	0.0630	0.0874	132
BM25	0.0705	0.1000	0.0652	0.0815	127
DFR-BM25	0.0706	0.1000	0.0652	0.0888	127
PL2	0.0686	0.0826	0.0674	0.0835	129
TF-IDF	0.0699	0.0957	0.0609	0.0815	131
DirichletLM	0.0841	0.0957	0.0565	0.0988	134
DirichletLM-qe	0.1073	0.1261	0.0870	0.1069	158

**Table 8.** Results for the mixed run of the case-based retrieval on ImageCLEF 2012

<i>Type</i>	<i>MAP</i>	<i>P10</i>	<i>P20</i>	<i>Rprec</i>	<i># of rel. docs</i>
mixed	0.1758	0.1565	0.1370	0.1915	222
mixed-qe	0.2186	0.2043	0.1804	0.2337	235

### 5.3 Results and discussion

Using the models and optimal parameters learned with the experiments over the ImageCLEF 2012 dataset, the experiments over the ImageCLEF 2013 dataset were performed. The best results were provided with in the case of the mixed experiment using query expansion.

**Table 9.** Results for the ImageCLEF 2013 dataset

<i>Type</i>	<i>MAP</i>	<i>P10</i>	<i>P20</i>	<i>Rprec</i>	<i># of rel. docs</i>
word-space	0.2026	0.2057	0.1743	0.2115	549
word-space-qe	0.2019	0.2314	0.1957	0.2011	596
concept-space	0.0438	0.0829	0.0671	0.0730	300
concept-space-qe	0.0632	0.0857	0.0771	0.0850	334
mixed	0.1832	0.1857	0.1586	0.2036	550
mixed-qe	0.2059	0.2229	0.1957	0.2235	604

## References

- de Herrera, A.G.S., Kalpathy-Cramer, J., Fushman, D.D., Antani, S., Müller, H.: Overview of the imageclef 2013 medical tasks. In: Working notes of CLEF 2013. (2013)
- Dimitrovski, I., Loskovska, S.: Content-based retrieval system for X-ray images. In: International Congress on Image and Signal Processing. (2009) 2236–2240
- Tommasi, T., Orabona, F., Caputo, B.: Discriminative cue integration for medical image annotation. Pattern Recognition Letters **29**(15) (2008) 1996–2002
- Zhang, J., Marszalek, M., Lazebnik, S., Schmid, C.: Local features and kernels for classification of texture and object categories: A comprehensive study. International Journal of Computer Vision **73**(2) (2007) 213–238

5. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision* **60**(2) (2004) 91–110
6. van de Sande, K., Gevers, T., Snoek, C.: Evaluating color descriptors for object and scene recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **32**(9) (2010) 1582–1596
7. van Gemert, J.C., Veenman, C.J., Smeulders, A.W.M., Geusebroek, J.M.: Visual word ambiguity. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **99**(1) (2010)
8. Lazebnik, S., Schmid, C., Ponce, J.: Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories. In: *IEEE conference on Computer Vision and Pattern Recognition*. (2006) 2169–2178
9. Ounis, I., Amati, G., Plachouras, V., He, B., Macdonald, C., Johnson, D.: Terrier information retrieval platform. In: *Advances in Information Retrieval*, Springer (2005) 517–519
10. Tommasi, T., Caputo, B., Welter, P., Güld, M., Deserno, T.: Overview of the clef 2009 medical image annotation track. In: *Multilingual Information Access Evaluation II. Multimedia Experiments – LNCS 6242*, Springer Berlin/Heidelberg (2010) 85–93
11. Chang, C.C., Lin, C.J.: LIBSVM: a library for support vector machines. (2001) Software available at <http://www.csie.ntu.edu.tw/~cjlin/libsvm>.
12. Lin, H.T., Lin, C.J., Weng, R.C.: A note on Platt’s probabilistic outputs for support vector machines. *Machine Learning* **68** (2007) 267–276
13. Müller, H., de Herrera, A.G.S., Kalpathy-Cramer, J., Demner-Fushman, D., Antani, S., Eggel, I.: Overview of the imageclef 2012 medical image retrieval and classification tasks. In: *CLEF (Online Working Notes/Labs/Workshop)*. (2012)
14. Aronson, A.R.: Effective mapping of biomedical text to the umls metathesaurus: the metamap program. In: *Proceedings of the AMIA Symposium*, American Medical Informatics Association (2001) 17
15. Trieschnigg, D., Pezik, P., Lee, V., De Jong, F., Kraaij, W., Rebholz-Schuhmann, D.: Mesh up: effective mesh text classification for improved document retrieval. *Bioinformatics* **25**(11) (2009) 1412–1418
16. Abdulahhad, K., Chevallet, J.P., Berrut, C., et al.: Mrim at imageclef2012. from words to concepts: A new counting approach. In: *Notebook Papers of Labs and Workshop (CLEF)*. (2012)
17. Müller, H., de Herrera, A.G.S., Kalpathy-Cramer, J., Fushman, D.D., Antani, S., Eggel, I.: Overview of the imageclef 2012 medical image retrieval and classification tasks. *Working Notes of CLEF* (2012)
18. Macdonald, C., Plachouras, V., He, B., Lioma, C., Ounis, I.: University of glasgow at webclef 2005: Experiments in per-field normalisation and language specific stemming. In: *Accessing Multilingual Information Repositories*. Springer (2006) 898–907
19. Amati, G., Van Rijsbergen, C.J.: Probabilistic models of information retrieval based on measuring the divergence from randomness. *ACM Transactions on Information Systems (TOIS)* **20**(4) (2002) 357–389
20. Hiemstra, D.: A probabilistic justification for using  $tf \times idf$  term weighting in information retrieval. *International Journal on Digital Libraries* **3**(2) (2000) 131–139
21. Zhai, C., Lafferty, J.: A study of smoothing methods for language models applied to information retrieval. *ACM Transactions on Information Systems (TOIS)* **22**(2) (2004) 179–214

22. Jain, A., Nandakumar, K., Ross, A.: Score normalization in multimodal biometric systems. *Pattern recognition* **38**(12) (2005) 2270–2285