Medical case–based retrieval: integrating query MeSH terms for query–adaptive multi–modal fusion

Alba G. Seco de Herrera, Antonio Foncubierta–Rodríguez and Henning Müller

University of Applied Sciences Western Switzerland (HES–SO), Sierre, Switzerland

ABSTRACT

Advances in medical knowledge give clinicians more objective information for a diagnosis. Therefore, there is an increasing need for bibliographic search engines that can provide services helping to facilitate faster information search.

The ImageCLEFmed benchmark proposes a medical case–based retrieval task. This task aims at retrieving articles from the biomedical literature that are relevant for differential diagnosis of query cases including a textual description and several images. In the context of this campaign many approaches have been investigated showing that the fusion of visual and text information can improve the precision of the retrieval. However, fusion does not always lead to better results.

In this paper, a new query–adaptive fusion criterion to decide when to use multi–modal (text and visual) or only text approaches is presented. The proposed method integrates text information contained in MeSH (Medical Subject Headings) terms extracted and visual features of the images to find synonym relations between them. Given a text query, the query–adaptive fusion criterion decides when it is suitable to also use visual information for the retrieval.

Results show that this approach can decide if a text or multi–modal approach should be used with 77.15% of accuracy.

Keywords: Query–adaptive fusion, MeSH, ImageCLEFmed, multi–modal fusion.

1. INTRODUCTION

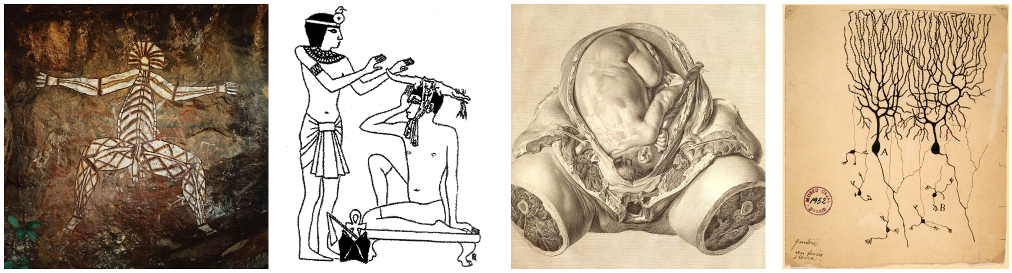
Medical topics have been represented in images since prehistoric times with early illustrations leaning towards symbolic representations. Illustration have been developing from symbolism to greater realism (see Figure 1). Advances in medical technologies have changed the physicians' vision and understanding of the human body. Different modalities of medical images, such as x–ray or light microscopy, sometimes show objective evidence of diseases and decrease the dependence on the patient's sometimes subjective descriptions. Figure 2 shows examples of findings in medical images that help physicians in their work on patient cases. The rapid development of medical knowledge[1] forces clinicians to increasingly use bibliographic search engines to support diagnosis because of the difficulty in keeping updated in even a specific field.[2,3] Evidence–based medicine is another important reason to search for positive or negative evidences for cases. Therefore, there is a need for solutions regarding biomedical information search. The biomedical open access literature of PubMed Central* is a resource very extensively used.[3] Indeed PubMed Central alone contained almost 2 million images in 2014. However, clinicians still fail regularly when searching for the information they need.[4]

The ImageCLEFmed† benchmark proposed a case–based retrieval task based on a subset of 70,000 redistributable articles of PubMed Central. The campaign aims at evaluating and comparing algorithms that retrieve articles from the biomedical literature that are relevant for differential diagnosis of the query cases. This work gives a new approach to solve this task. Many approaches have been explored over the years for searching in the biomedical literature.[5–7] Moreover, previous studies[8] have shown that the combination of visual and text

Further author information: (Send correspondence to A.G.S.H.)
A.G.S.H., Email: alba.garcia@hevs.ch

*http://www.ncbi.nlm.nih.gov/pmc/

†http://imageclef.org/



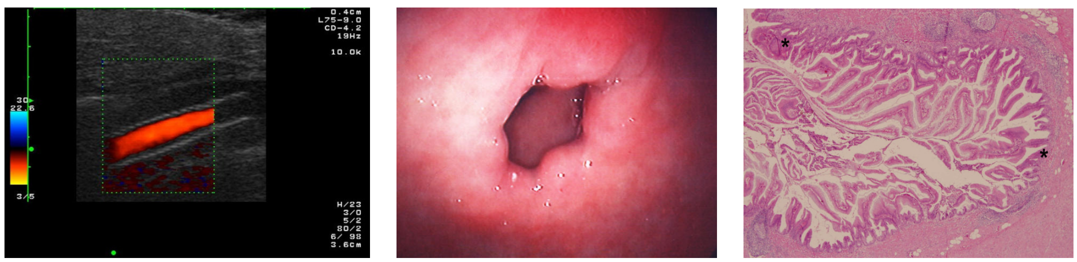
(a) Rock painting, 6000 B.C. Aboriginal 'X–ray style' figure. Kakadu National Park, Northern Territory, Australia.

(b) The Ebers Papyrus, 1200 B.C. Egyptian papyrus which describes therapy of migraine.

(c) Copperplate engraving of a woman who died near the end of term by William Hunter, 1774. National Library of Medicine.

(d) Drawing of Purkinje cells and granule cells from pigeon cerebellum by Santiago Ramón y Cajal, 1899. Instituto Santiago Ramón y Cajal.

Figure 1. Examples of historical medical illustrations.



(a) Findings using color Doppler after endovascular treatment (stenting) in a 52–year–old woman suffering from recurrent transient ischemic attacks.

(b) A complete healing at the polypectomy site on an endoscopy after a 12–week course of proton pomp inhibitor therapy.

(c) Hematoxylin and eosin stain on the appendix tissue reveals villous adenoma with moderate to severe dysplasia located suppurative appendicitis.

Figure 2. Examples of medical images that help in the diagnosis and treatment planning of cases.

information can improve the quality of the retrieval. However, visual information is not always useful for this task and even can decrease the performance of the retrieval.[9–11]

Text retrieval techniques commonly use terminologies for query expansion.[12,13] The queries can be expanded automatically with synonyms from such a terminology, for example. Díaz Galiano et al.[14] considered terms associated with MeSH (Medical Subject Headings) descriptors as synonyms and used these to expand queries. More recently Dramé et al.[15] explored the use of term synonyms to expand queries. However, visual retrieval techniques cannot apply these methods directly for synonym extraction because visual information cannot be directly represented as words. Nevertheless, language modelling techniques can be extended easily to visual techniques.

Some efforts have been made to find a relation between images and text. Recently, Simpson et al.[16] review the techniques applied to limit the semantic gap between images and its meaning in terms of natural language. A method based on global feature mapping is presented. However, most of the approaches use joint probabilistic models to deal with this problem.[17–20]Additionally, some approaches are based on image region categorization.[21,22]

In this paper, we propose a new method for query–adaptive multi–modal fusion. The goal is to change the formulation of the retrieval algorithm based on the user query. Kennedy[23] reviews the methods proposed for adapting retrieval strategies according to the intentions of the user. Most of the techniques are based on query classification using natural language analysis of the query. Although other strategies have been proposed, such as the prediction of the quality of each available tool based on statistical measures of the returned results or the adaptation strategies based on the user context.

This work suggests a criterion to decide when to use multi–modal (text and visual) or only text approaches for medical case–based retrieval. To that end, a method to find 'synonyms' between text information contained in MeSH (Medical Subject Headings) terms extracted and visual features contained in visual descriptors is proposed. This approach is based on probabilistic latent semantic analysis[24] to find the synonym relations. The query–adaptive fusion criterion allows to know when a given a text query is suitable to also use visual information for the retrieval.

The rest of this paper is organized as follows. Section 2 describes the dataset and approach used in this work. Section 3 presents the experimental results of the proposed method for medical case–based retrieval. Conclusions and future directions are discussed in Section 4.

2. METHODS

This section describes the details concerning the dataset and the techniques employed to carry out the experiments.

2.1. Dataset

In this paper the data and evaluation scenario provided by the ImageCLEFmed 2013 benchmark are used. The data used are a subset of PubMed Central containing in total over 1.5 million images and being updated with new data very regularly. The distributed subset of ImageCLEF contains only articles allowing redistribution. The case–based task of ImageCLEFmed is used for the experiments. The 2013 collection provided for this task consists of over 300,000 images of 75,000 articles of the biomedical open access literature. 35 query topics were distributed as part of the benchmark. Each topic consists of a narrative case description with patient demographics, symptoms and test results including imaging studies but not the final diagnosis. An example topic is shown in Figure 3.

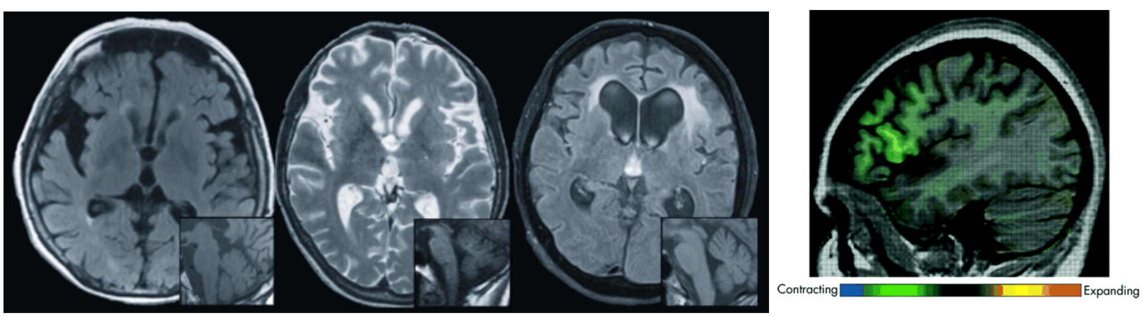


Figure 3. Images from one of the topics in the case–based retrieval ImageCLEFmed task. These images correspond to the following text query: 'A 55–year–old man with progressive behavioural and personality changes. MRI shows frontal lobe atrophy with preservation of posterior brain structures.'.

The goal of the task is to retrieve articles that might best suit to the provided case in terms of usefulness for differential diagnosis. For more details on the task see.[25]

2.2. Retrieval baseline

The Apache Lucene[‡] framework was used for text retrieval. The Lucene configuration used applies tokenization, stemming, stop word removal and term frequency–inverse document frequency (tf/idf) weighting.[26]

For the visual content of the images, multiple features are used, as this was a successfully used technique in the past.[25,27] A combination of the following four visual descriptors selected from the Parallel Distributed Image Search Engine (ParaDISE)[28] are applied:

[‡]`http://lucene.apache.org/`

- *Grid BoC* – A $n \times n$ spatial grid representation of the Bag of Color (BoC);[29]
- *BoVW–SPM* – A spatial pyramid matching[30] of the Bag–of–Visual–Word representation of the Scale Invariant Feature Transform SIFT;[31]
- *CEDD* – Color and Edge Directivity Descriptor;[32]
- *Tamura* – Tamura texture description.[33]

The selection of the descriptors is based on previous work with good performance on the ImageCLEFmed 2012 tasks.[9]

2.3. MeSH Term Extraction

Most of MEDLINE publication records are manually annotated with MeSH terms, which can be retrieved using the Entrez search system API[§].[34] In this work it was possible to retrieve MeSH terms for 73,584 documents (98.6%) of the ImageCLEFmed dataset and to construct two binary sparse document – MeSH term matrices: one covering all 18,299 MeSH terms referenced by the document corpus and a second matrix covering only 5,583 MeSH terms marked as *major topic* for documents. Each image belonging to a document is represented as a binary histogram which characterized the annotated MeSH terms contained in the document. Each binary histogram is a binary vector–form representation of MeSH terms occurrence in the document.

Queries were mapped to MeSH terms by a score–based phrase matching algorithm favouring MeSH terms with words occurring rarely in the document corpus.[35] Matching synonyms were replaced by their primary MeSH terms. Only MeSH terms occurring in the document–MeSH term matrices were considered for query mapping. Hence, textual queries are also represented as a binary histogram of the extracted MeSH terms.

2.4. Visual and Text Word Synonymy

Collins dictionary[36] defines a '*synonym*' as '*a word that means the same or nearly the same as another word*'. Furthermore Foncubierta–Rodríguez[37] extends the definition of synonyms to visual words based on criteria derived from Probabilistic Latent Semantic Analysis (PLSA).

Definition 2.1 (Synonyms). *A pair of visual words w_n, w_m can be considered* synonyms *if the following three conditions are met:*

1. *There is at least one visual topic z_j to which both w_n and w_m belong;*
2. *w_n and w_m have a complementary distribution in the collection;*
3. *w_n and w_m have a similar contextual distribution with the rest of the words.*

were a visual topic *z is defined as the representation of a generalized version of the visual appearance modelled by various visual words. It corresponds to an intermediate level between visual words and the complete understanding of visual information. A set of visual topics $\mathcal{Z} = \{z_1, \ldots, z_{N_Z}\}$ can be defined in a way that every visual word can belong to none, one or several visual topics. In this case, visual topics correspond to each of the topics or aspects derived from a PLSA analysis.* According to this definition of visual synonymy, Foncubierta–Rodríguez[37] defines a synonymy matrix as:

Definition 2.2 (Synonymy visual word space). **S** *is a symmetric* synonymy matrix *if:*

$$\mathbf{S} = \begin{pmatrix} 1 & s_{12} & \cdots & s_{1N_W} \\ s_{21} & 1 & \cdots & s_{2N_W} \\ \vdots & \vdots & \ddots & \vdots \\ s_{N_W1} & s_{N_W2} & \cdots & 1 \end{pmatrix} \quad (1)$$

[§]https://www.ncbi.nlm.nih.gov/books/NBK21081/

where s_{ij} measures the synonymy of the visual words w_i and w_j.

$$s_{ij} = s_{ji} = \begin{cases} 1 \text{ if } i = j \\ \sigma_{ij} \text{ if } w_i, w_j \text{ are synonyms} \\ 0 \text{ otherwise} \end{cases} \tag{2}$$

and σ_{ij} is the synonymy value of the words w_i, w_j. The synonymy value *of two words w_n, w_m is defined as the maximum significance value for which both words are significant for the same visual topic.*

$$\sigma_{nm} = \sigma_{mn} = \max_j \left\{ \min_{n,m} \{v_{n,j}, v_{m,j}\} \right\} \tag{3}$$

where $v_{i,j}$ is the normalized value of the probability $P(w_i|z_j)$ obtained from PLSA.

Medical text can be represent as an histogram of MeSH terms (see Section 2.3). Images can also be represented as a histogram of visual features that is built using descriptors as the descriptors mentioned in Section 2.2. Therefore it is possible to consider both text and visual features to create a common vocabulary. Definition 2.2 is extended from language modelling techniques, therefore it can also be used for the synonym relation between text and visual information keeping mathematical sense of synonyms.

The synonymy matrix from a set of MeSH terms and visual descriptors is obtained considering the relative properties of visual words based on their behaviour on training data. For each of the images in the training set, the histogram of MeSH terms and the histogram with the visual features are concatenated. As a result the following symmetric synonymy matrix is obtained:

$$\mathbf{S_{tv}} = \begin{pmatrix} 1 & t_{12} & \cdots & \cdots & t_{1M} & tv_{1M+1} & \cdots & tv_{1M+N} \\ t_{21} & 1 & \cdots & \cdots & \cdots & \cdots & \cdots & tv_{2M+N} \\ \vdots & \vdots & \ddots & \cdots & \vdots & \vdots & \vdots & \vdots \\ t_{M1} & \cdots & \cdots & \cdots & t_{MM} & tv_{MM+1} & \cdots & tv_{MM+N} \\ vt_{M+11} & \cdots & \cdots & \cdots & vt_{M+1M} & v_{M+1M+1} & \cdots & v_{M+1M+N} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ vt_{M+N1} & \cdots & \cdots & \cdots & vt_{M+NM} & v_{M+NM+1} & \cdots & 1 \end{pmatrix} \tag{4}$$

where t_{ij} is the synonymy value of two MeSH terms, v_{ij} is the synonymy value of two visual features and $tv_{ij} = vt_{jj}$ is the synonymy value of a MeSH term and a visual feature. M is the dimension of the textual histogram (the number or MeSH terms in the set) and N the dimension of the visual histogram.

2.5. Query–adaptive fusion criterion

Not all medical case text descriptions need query images to find relevant articles. Often the relevant articles for a topic do not contain images or contain only general biomedical illustration (such as statistical figures or graphs). In ImageCLEFmed 2013[25] best results for case–based retrieval were actually achieved by pure text runs. Participants usually decreased their results when using multi–modal approaches. However, we believe that visual information can improve the precision of the retrieval.

The basic hypothesis of this work is defined as follows:

Hypothesis 2.3. *If the extracted MeSH terms of a text query have synonym relations with the visual features, then visual information can improve retrieval.*

Similar to the use of text synonyms, using multi–modal retrieval (text and visual information) only when there is a synonym relation between the text query and the visual features can made the retrieval more consistent because only articles that are really related to the topic will be retrieved.[38]

This work focuses on the synonym relation between text and visual features, i.e., on the submatrix of the matrix S_{tv}:

$$\mathbf{A} = \mathbf{S_{tv}}(i,j), \quad \forall i \in [M, M+N] \quad \textit{and} \quad \forall j \in [1, M] \tag{5}$$

The following criterion is proposed to predict when it is suitable to use visual information in addition to text based on the query:

DEFINITION 2.4 (QUERY–ADAPTIVE FUSION CRITERION). *Let* $\mathbf{q} \in [0,1]^M$ *be the binary histogram of MeSH terms occurrence in the textual query. If* $\exists i/\mathbf{q}(i) \neq 0$ *and* $\exists j/\mathbf{A}(i,j) \neq 0$ *then the textual query is* suitable *to be fused with a visual query.*

3. EXPERIMENTAL RESULTS

To demonstrate the effectiveness of the proposed technique, the data distributed by ImageCLEFmed case–based task in 2013 was used in the implemented experiments.

The synonymy matrix of a set of MeSH terms and each visual descriptor is calculated based on a training set of 5,000 random images. To study the effect of the latent variable z the synonym matrices are calculated for $N_Z = 50, 100, 200, 300$. Minimum significance percentiles $p = 0th, 50th, 75th, 99th$ are also considered in the study, removing all words with a maximum significance $m_i = \max_j t_{i,j}$ below the given percentile.

The two sets of MeSH terms described in Section 2.3 (*major* and *all*) are analysed in this article. When using the set of *all* MeSH terms, the calculation of the synonymy matrix was restricted to 50,000 synonyms due to computational limitations. All synonyms were calculated when using the *major* set of MeSH terms. The choice of the latent value and the percentile does not affect to the performance when using *all* the MeSH terms.

The result of the Average Precision per topic is summarized in Table 1. This table shows a comparison between the runs. In general, the *text* approach has a higher Average Precision than the *visual* approach. Fusion of text and visual approaches (*mix*) can improve the Average Precision although for several topics is better to use the *text* approach. The query–adaptive criterion presented in Section 2.5 allows the automatic selection of the *text* or *mixed* approach for each of the topics. Table 1 shows the Average Precision per topic for the approaches using *all* and *major* MeSH terms. For the *major* approach, Table 1 shows the results for the latent values and percentiles corresponding to the approach with accuracy 77.15%. Results are compare with the best mix run submitted to ImageCLEFmed 2013.

Table 2 shows the accuracy of correct decision obtained when applying the proposed approach with various parameters and only *major* MeSH terms. These results are not presented for *all* MeSH terms because there is no difference between the parameters, showing the stability of the method. Indeed, using *major* MeSH terms the accuracy of the query–adaptive criterion is always the same except in two cases.

Table 3 summarizes best results achieved with the proposed query–adaptive fusion criterion. This result shows an accuracy of 77.15% when using *major* MeSH terms for most of the parameters values. Accuracy using *all* MeSH terms is lower with62.86%, probably due to the restriction in the number of synonyms.

4. CONCLUSIONS AND FUTURE WORK

A query–adaptive fusion criterion for the use of multi–modal techniques in medical case–based retrieval is presented. The proposed method integrates the textual information of MeSH terms with the visual descriptors creating a matrix of synonym relations between both kinds of features (text and visual). The synonym matrix is then used to decide if a text query is suitable for a multi–modal approach or if text alone would lead to best results.

The performance of the experiments is assessed on the very challenging dataset of the case–based retrieval task of ImageCLEFmed 2013. Experimental results indicate that it is indeed effective, showing that correct decisions are taken in 77.15% of the cases. The results are also very stable regarding parameter choices. Therefore, the current work opens an area of research on multi–modal decision for medical case–based retrieval.

Future work includes hierarchical relationships between MeSH terms as well as a study of synonym relation between visual descriptors and terms of the Unified Medical Language System (UMLS). Visual query reweighting based on synonym relations between text and visual features is also an interesting field. Finally, the presented work can be explored for automatic visual descriptor selection.

Table 1. Average Precision per topic using various approaches. Correct decisions taken by the proposed approaches are shown in bold type.

#Topic	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
Best mix ImageCLEF	0.1117	0.0598	0.9167	0.0005	0.2658	0.2066	0.0630	0.0871	0.3548
Visual	0.0010	0	0.3333	0	0.0034	0.0383	0.0011	0.0033	0
Text	0.1055	0.0310	0.6789	0.0081	0.4491	0.2207	0.1432	0.0864	0.0434
Mix	0.1049	0.0306	0.6782	0.0074	0.4492	0.2261	0.1421	0.0799	0.0434
All	**0.1055**	0.0306	**0.6789**	0.0074	**0.4492**	**0.2261**	**0.1432**	0.0799	**0.0434**
Major	**0.1055**	**0.0310**	**0.6789**	**0.0081**	**0.4492**	0.2207	**0.1432**	**0.0864**	**0.0434**
#Topic	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**
Best mix ImageCLEF	0.0025	0.0059	0.0119	0.3326	0.1659	0.138	0.047	0.0111	0.1374
Visual	0	0	0	0.0058	0.0363	0	0.0022	0	0.1000
Text	0.0357	0.0038	0.0482	0.2915	0.3044	0.2003	0.0367	0.2000	0.0242
Mix	0.0357	0.0037	0.0481	0.3049	0.3121	0.1893	0.0344	0.2000	0.0669
All	**0.0357**	**0.0038**	0.0481	**0.3049**	**0.3121**	0.1893	0.0344	**0.2000**	**0.0669**
Major	**0.0357**	**0.0038**	**0.0482**	**0.3049**	0.3044	0.1893	0.0344	**0.2**	**0.0669**
#Topic	**19**	**20**	**21**	**22**	**23**	**24**	**25**	**26**	**27**
Best mix ImageCLEF	0.2097	0.0754	0.0720	0.1985	0.2081	0.0589	0.0085	0.2202	0.2317
Visual	0.0057	0.0118	0.001	0.0038	0.0006	0.0033	0.0005	0.0035	0.2572
Text	0.1896	0.1063	0.1118	0.2419	0.3514	0.1217	0.0106	0.2780	0.0793
Mix	0.1934	0.1115	0.1098	0.2455	0.3506	0.1228	0.0102	0.2704	0.3157
All	**0.1934**	**0.1115**	0.1098	**0.2455**	0.3506	**0.1228**	**0.0106**	0.2704	**0.3157**
Major	**0.1934**	**0.1115**	**0.1118**	0.2419	**0.3514**	0.1217	**0.0106**	**0.278**	**0.3157**
#Topic	**28**	**29**	**30**	**31**	**32**	**33**	**34**	**35**	**Mean**
Best mix ImageCLEF	0.0212	0.1325	0.2894	0.5048	0.2435	0.1045	0.0823	0.0503	0.1608
Visual	0.1699	0.0081	0.0574	0	0.001	0.1265	0.0002	0.0013	0.0336
Text	0.0918	0.1642	0.2419	0.5069	0.1381	0.2876	0.2820	0.1536	0.1791
Mix	0.1278	0.1686	0.2783	0.5063	0.1372	0.2868	0.2786	0.1404	0.1889
All	**0.1278**	**0.1686**	**0.2783**	0.5063	0.1372	**0.2876**	0.2786	0.1404	0.1890
Major	0.0918	**0.1686**	**0.2783**	0.5063	**0.1381**	**0.2876**	**0.282**	**0.1536**	0.1885

Table 2. Accuracy (%) of correct decisions obtained by the proposed approached when using *major* MeSH terms. The results are shown for several latent values (z) and percentiles p.

z \ p	**0**	**50**	**75**	**99**
50	45.72	77.15	77.15	62.86
100	77.15	77.15	77.15	77.15
200	77.15	77.15	77.15	77.15
300	77.15	77.15	77.15	77.15

Table 3. Accuracy (%) of correct decisions obtained by the proposed approaches when using *all* and *major* MeSH terms.

Run	**Accuracy**
Major	77.15
All	62.86

ACKNOWLEDGMENTS

The authors thank Mario Taschwer for assistance with the MeSH term extraction and for helpful discussions. This work was partly supported by the EU 7^{th} Framework Program in the context of the Khresmoi project (FP7–257528).

REFERENCES

1. L. Hunter and B. K. Cohen, "Biomedical language processing: What's beyond PubMed?," *Molecular Cell* **21**, pp. 589–594, Mar 2006.
2. G. Luo, C. Tang, H. Yang, and X. Wei, "MedSearch: A specialized search engine for medical information retrieval," in *Proceedings of the 17th ACM Conference on Information and Knowledge Management*, *CIKM '08*, pp. 143–152, ACM, (New York, NY, USA), 2008.
3. S. Z. Shariff1, S. A. Bejaimal, J. M. Sontrop, A. V. Iansavichus, R. B. Haynes, M. A. Weir, and A. X. Garg, "Retrieving clinical evidence: A comparison of PubMed and Google Scholar for quick clinical searches," *Journal of Medical Internet Research* **15**(8), 2013.
4. D. Markonis, M. Holzer, S. Dungs, A. Vargas, G. Langs, S. Kriewel, and H. Müller, "A survey on visual information search behavior and requirements of radiologists," *Methods of Information in Medicine* **51**(6), pp. 539–548, 2012.
5. J. Kalpathy-Cramer, A. García Seco de Herrera, D. Demner-Fushman, S. Antani, S. Bedrick, and H. Müller, "Evaluating performance of biomedical image retrieval systems an overview of the medical image retrieval task at ImageCLEF 2004–2014," *Computerized Medical Imaging and Graphics* **39**(0), pp. 55 – 61, 2015.
6. H. Müller, N. Michoux, D. Bandon, and A. Geissbuhler, "A review of content–based image retrieval systems in medicine–clinical benefits and future directions," *International Journal of Medical Informatics* **73**(1), pp. 1–23, 2004.
7. W. Hersh, *Information Retrieval — A health and Biomedical Perspective*, Springer, 2003 – second edition.
8. A. García Seco de Herrera and H. Müller, *Fusion Techniques in Biomedical Information Retrieval*, pp. 209–228. Springer, 2014.
9. A. García Seco de Herrera, D. Markonis, R. Schaer, I. Eggel, and H. Müller, "The medGIFT group in ImageCLEFmed 2013," in *Working Notes of CLEF 2013 (Cross Language Evaluation Forum)*, September 2013.
10. M. S. Simpson, D. You, M. M. Rahman, D. Demner-Fushman, S. Antani, and G. Thoma, "ITI's participation in the 2013 medical track of ImageCLEF," in *Working Notes of CLEF 2013 (Cross Language Evaluation Forum)*, September 2013.
11. A. Mourão, F. Martins, and J. a. Magalhães, "NovaSearch on medical ImageCLEF 2013," in *Working Notes of CLEF 2013 (Cross Language Evaluation Forum)*, September 2013.
12. C. E. Crangle, A. Zbyslaw, J. M. Cherry, and E. L. Hong, "Concept extraction and synonymy management for biomedical information retrieval," in *The thirteenth Text REtrieval Conference (TREC 2004)*, 2004.
13. Z. Shi, B. Gu, F. Popowich, and A. Sarkar, "Synonym–based query expansion and boosting–based re–ranking: A two–phase approach for genomic information retrieval," in *The Fourteenth Text REtrieval Conference (TREC 2005)*, 2005.
14. M. C. Díaz-Galiano, M. T. Martín-Valdivia, and L. A. Ureña López, "Query expansion with a medical ontology to improve a multimodal information retrieval system," *Computers in Biology and Medicine* **39**(4), pp. 396–403, 2009.
15. K. Dramé, F. Mougin, and G. Diallo, "Query expansion using external resources for improving information retrieval in the biomedical domain," in *Proceedings of the ShARe/CLEF eHealth Evaluation Lab*, 2014.

16. M. S. Simpson, D. Demner-Fushman, S. K. Antani, and G. R. Thoma, "Multimodal biomedical image indexing and retrieval using descriptive text and global feature mapping," *Information Retrieval* **17**(3), pp. 229–264, 2014.
17. V. Lavrenko, R. Manmatha, and J. Jeon, "A model for learning the semantics of pictures," in *Proceedings of the Seventeenth Annual Conference on Neuronal Information Processing Systems*, **16**, pp. 553–560, 2003.
18. K. Barnard, P. Duygulu, D. Forsyth, N. de Freitas, D. M. Blei, and M. I. Jordan, "Matching words and pictures," *Journal of Machine Learning Research* **3**, pp. 1107–1135, 2003.
19. P. Duygulu, K. Barnard, J. F. de Freitas, and D. A. Forsyth, "Object recognition as machine translation: Learning a lexicon for a fixed image vocabulary," in *Computer Vision–ECCV 2002*, pp. 97–112, Springer, 2002.
20. N. Rasiwasia, J. Costa Pereira, E. Coviello, G. Doyle, G. R. Lanckriet, R. Levy, and N. Vasconcelos, "A new approach to cross-modal multimedia retrieval," in *Proceedings of the international conference on Multimedia*, pp. 251–260, ACM, 2010.
21. R. Datta, W. Ge, J. Li, and J. Z. Wang, "Toward bridging the annotation-retrieval gap in image search," *Advances in Multimedia Computing* **14**(3), pp. 24–35, 2007.
22. C. Lacoste, J.-H. Lim, J.-P. Chevallet, and D. T. H. Le, "Medical–image retrieval based on knowledge-assisted text and image indexing," *IEEE Transactions on Circuits and Systems for Video Technology* **17**(7), pp. 889–900, 2007.
23. L. Kennedy, S.-F. Chang, and A. Natsev, "Query–adaptive fusion for multimodal search," *Proceedings of the IEEE* **96**(4), pp. 567–588, 2008.
24. T. Hofmann, "Unsupervised learning by probabilistic latent semantic analysis," *Machine learning* **42**(1-2), pp. 177–196, 2001.
25. A. García Seco de Herrera, J. Kalpathy-Cramer, D. Demner Fushman, S. Antani, and H. Müller, "Overview of the ImageCLEF 2013 medical tasks," in *Working Notes of CLEF 2013 (Cross Language Evaluation Forum)*, September 2013.
26. A. García Seco de Herrera, R. Schaer, D. Markonis, and H. Müller, "Comparing fusion techniques for the ImageCLEF 2013 medical case retrieval task," *Computerized Medical Imaging and Graphics* **39**, pp. 46–54, 2015.
27. H. Müller, A. García Seco de Herrera, J. Kalpathy-Cramer, D. Demner Fushman, S. Antani, and I. Eggel, "Overview of the ImageCLEF 2012 medical image retrieval and classification tasks," in *Working Notes of CLEF 2012 (Cross Language Evaluation Forum)*, September 2012.
28. R. Schaer, D. Markonis, and H. Müller, "Architecture and applications of the parallel distributed image search engine (ParaDISE)," in *FoRESEE 2014, 1st International Workshop on Future Search Engines at INFORMATIK 2014*, 2014.
29. A. García Seco de Herrera, D. Markonis, and H. Müller, "Bag of colors for biomedical document image classification," in *Medical Content–based Retrieval for Clinical Decision Support*, H. Greenspan and H. Müller, eds., *MCBR–CDS 2012*, pp. 110–121, Lecture Notes in Computer Sciences (LNCS), October 2013.
30. S. Lazebnik, C. Schmid, and J. Ponce, "Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories," in *Proceedings of the 2006 IEEE Conference on Computer Vision and Pattern Recognition, CVPR*, pp. 2169–2178, IEEE Computer Society, (Washington, DC, USA), 2006.
31. D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision* **60**(2), pp. 91–110, 2004.
32. S. A. Chatzichristofis and Y. S. Boutalis, "CEDD: Color and edge directivity descriptor: A compact descriptor for image indexing and retrieval," in *Lecture notes in Computer Sciences*, **5008**, pp. 312–322, 2008.
33. H. Tamura, S. Mori, and T. Yamawaki, "Textural features corresponding to visual perception," *IEEE Transactions on Systems, Man and Cybernetics* **8**, pp. 460–473, June 1978.
34. N. R. Coordinators, "Database resources of the National Center for Biotechnology Information," *Nucleic Acids Research* **41**(D1), pp. D8–D20, 2013.
35. M. Taschwer, "Textual methods for medical case retrieval," tech. rep., Institute of Information Technology (ITEC), Alpen-Adria-Universität Klagenfurt, Austria, May 2014.

36. unknown, “Collins: English dictionary.” `http://www.collinsdictionary.com/`. Accessed: 2014-12-14.
37. A. Foncubierta-Rodríguez, *Description and Retrieval of Medical Visual Information based on Language Modelling*. PhD thesis, University of Geneva, 2014.
38. unknown, “WHSL medical subject headings for PubMed searching: Medical subject headings (MeSH).” `http://libguides.wits.ac.za/whsl-mesh`. Accessed: 2014-12-14.