Assisted query formulation for multimodal medical case-based retrieval

André Mourão
a.mourao@campus.fct.unl.pt

Flávio Martins
flaviomartins@acm.org

João Magalhães
jm.magalhaes@fct.unl.pt

Departamento de Informática
Faculdade de Ciências e Tecnologia
Universidade Nova de Lisboa
Caparica, Portugal

ABSTRACT

Medical information retrieval systems support health care experts in diagnostic and treatment decisions through the management of large amounts of clinical data. However, the ever growing data produced in medical environments and the proficiency of non-professional users pose several challenges to a retrieval system.

In this paper, we propose a medical retrieval system, supporting semantic multimodal queries for medical case-based search. The system explores many of the commodities available in commercial search engines and provides the user with key tools to support medical information discovery: multimodal queries and semantic suggestions of medical terms. It is build upon state-of-the-art information retrieval and data fusion techniques. We also propose a new data-fusion technique, which we called Inverted Squared Rank (ISR), to better deal with the combination of ranked lists from various heterogeneous systems: similar to Reciprocal Rank Fusion approaches like RR [10] and RRF [3]. The proposed rank fusion method outperforms RR and RRF in most measures and is particularly better on the top 10 results.

The system is at `http://medical.novasearch.org/`

Keywords

assisted query, search interfaces, multimodal fusion, multimodal medical retrieval

1. INTRODUCTION

Search engines are often a key tool for both healthcare professionals and laypeople when investigating medical cases. Case-based retrieval systems can support healthcare users in many ways - suggesting new publications, exploring similar symptoms/conditions, confirming a diagnosis, etc.

We propose a search engine focused on usability and usefulness, not only for health professionals but also accessible to laypeople. At the heart of a case-base retrieval system is the support for rich queries with heterogeneous data. Textual queries can include a long descriptions of the patient condition and images often provide additional information that is difficult to convey in textual queries. For instance, medical images capture the actual exams (e.g. x-ray, MRI) providing exact visual information about the patient (e.g. position of a mass on an MRI). Thus, we designed the system in a flexible way to support multiple data-fusion techniques (e.g. CombMNZ, RR, RRF, CombSUM).

The proposed system also provides an intuitive and simplified way of accessing large medical knowledge bases. It identifies medical terms in real-time and suggests related terms based on medical ontologies. This provides a glimpse of related conditions/diagnostics which can assist users in the formulation of a more targeted query.

In general, the system combines the simplicity of web search engines (text queries, semantic autocomplete, and the general look and feel) with automatic query expansion and image query using simple drag and drop.

In this article we present the framework, focusing on the fusion techniques and the user interface for case-based search. The underlying search framework was evaluated on the case-based retrieval task of the ImageCLEF 2013 medical dataset[1].

2. RELATED WORK

Several systems designed for medical retrieval (textual or visual) are available online. The MedGIFT group [5] designed two search engine interfaces to demonstrate their work in medical retrieval: a text based case retrieval search engine[2] and a visual medical image search[3]. The visual medical image interface allows the upload of query images and searching for similar images to the ones in the articles found. Although these two systems work well in their domains, either text or images, they are independent. For instance, it is not possible to search for images using a text query or combine image and text in the query.

An example of a content based image retrieval (CBIR) system is img(Anaktisi)[4] [9]. It was created to demonstrate the CEDD and FCTH image features [2] for image retrieval in multiple datasets. It includes the IRMA medical dataset

HSD 2013 August 1, 2013, Dublin, Ireland.

[1] `http://www.imageclef.org/2013/medical`
[2] `http://fast.hevs.ch:8080/MedSearch/faces/Search.jsp`
[3] `http://fast.hevs.ch:8080/MedSearch/faces/ImageSearch.jsp`
[4] `http://orpheus.ee.duth.gr/anaktisi/`

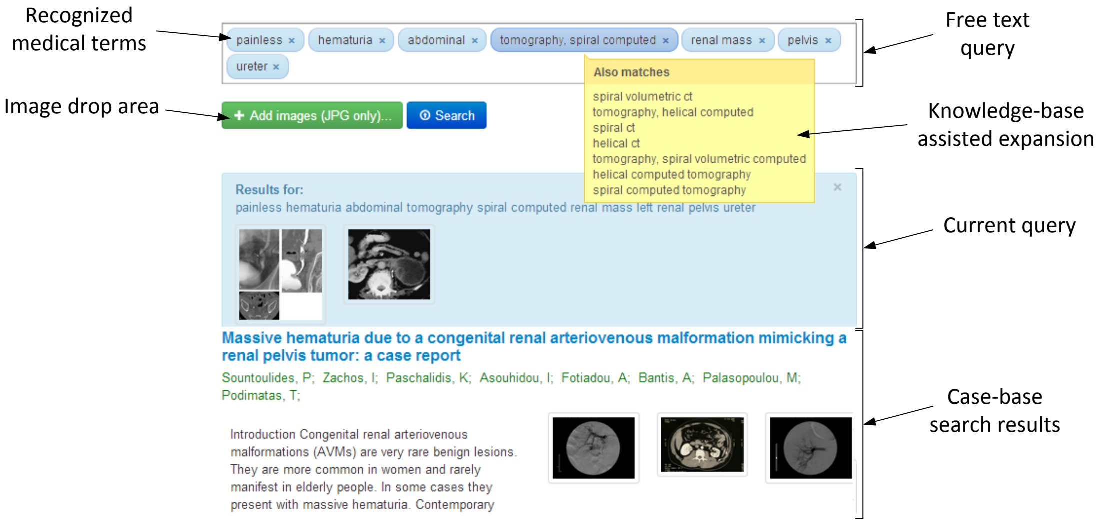

Figure 1: Search interface

and allows searching using corpus images as queries.

In our approach, images and text are completely integrated and all combinations are possible. This means, it allows searching for cases using only images, search for images using only text or use both text and images in the same query for either result type.

Outside the medical domain, we find the MMretrieval search engine[5] [8], a multimodal and multilingual search engine for retrieval of Wikipedia images. The system searches for each modality separately (including text in different languages) and applies a variety of late fusion algorithms to combine the results.

Our search engine relies on the fusion of multiple modalities (text and images). Thus, a fusion technique must be applied. Early fusion relies on combining the features from multiple modalities before searching in a common index. Late fusion relies on searching each modality index separately and combining the results in a final step. There are various data fusion methods such as Condorcet [6], CombSUM and its variants [1] (score-based), and Reciprocal Rank approaches like RR [10] and RRF [3] (rank-based) that are among the most applied but there is no clear off-the-shelf solution for all search tasks and modalities. Rank based approaches are gaining momentum, and we implemented a variant of combMNZ and RR that shows potential for multimodal combination on search engines.

We based our image retrieval on features that obtained good results on previous editions of ImageCLEF [7]. We included CEDD and FCTH (texture and color), Local Binary Pattern histograms (contours) and color histograms (color) for retrieval. Textual retrieval is based on Apache Lucene[6], with BM25L as the retrieval function.

Query expansion is useful to increase IR systems performance, making queries match more relevant documents that might not contain the exact query terms entered. Automatic query expansion (AQE) adds terms to the query without user intervention. This is already being performed on major commercial search engines. Most of the times, the expanded terms are synonyms or highly related terms and the user does not receive any feedback on the expanded terms. Interactive query expansion (IQE) gives the user the power to decide what terms are expanded however it is often an interface offered after the initial query at the cost of a more complicated interaction. Our approach is a mixture of IQE and AQE. The user can visualize what terms will be added to the query and opt-out expansion if incorrect or not desirable.

3. MEDICAL QUERY FORMULATION

Nova MedSearch is a multimodal (text and image) medical search engine that can retrieve either similar images or related medical cases. These tasks are from the medical ImageCLEF 2013 evaluation campaign. The results are displayed in an ranked list with basic information (e.g. title, keywords, images (if available)) and a link to the corresponding article details. The interface in Figure 1 takes into account both the relevancy of the images and text similarity.

3.1 Multi-part queries

Our interface aims at simplifying the inclusion of images and text data in the medical query (a screenshot of the different components of the interface is in Figure 1). For instance, we add support for drag-and-drop functionality for custom medical image queries. The free text query box allows entering a textual description of the patient, and the system automatically expands a recognized term into its related terms. In the example, we see the terms that are related to the search term "spiral computed tomography". The search results contain a visual presentation of the submitted query and the retrieved examples. In addition to general article information (title with link to full article, authors and abstract), we also display the images of the article that are most related to the query images.

3.2 Assisted query expansion

The main novelty of the search interface is the assisted query semantic-expansion. Since medical terminology is part

[5]http://www.mmretrieval.net/
[6]http://lucene.apache.org/

of natural language, terms are not unique, and multiple definitions of the same symptom/medication/disease are available. For example, our system returns "acetylsalicylic acid" and "2-(acetyloxy)benzoic acid" as terms related to "aspirin".

We implemented a guided query expansion system that interactively provides auto-complete suggestions and expansion feedback sourced from a SKOS version of the MeSH indexing terms. Medical SKOS provides domain specific expert knowledge regarding the relationships between terms. We decided to use a SKOS version of MeSH to provide two functionalities:

- word based auto-completion with terms
- automatic term expansion with semantically related terms

The process works as follows:

1. when the user starts typing a word, a dropdown box appears with the terms that match the query;
2. if the user selects a term from the list, the browser retrieves the synonyms from our framework and adds them to the query implicitly;
3. the user can then see the expanded terms by putting the mouse over the words. The user can opt-out the suggestions by clicking the ⊗ mark.

Since our system uses a SKOS representation for the terms expansion process, we can also support the SNOMED Medical ontology.

4. SEARCH-RESULTS FUSION

In this section we shall describe the search-results fusion methods that combine the rank from the multimodal information sources.

4.1 Text retrieval

The text is indexed using Lucene and the BM25L retrieval function is used. The indexed fields depend on the task. For image retrieval, we achieved good results indexing and searching only on the title, abstract and image captions. For case retrieval, we searched on the full document (title, abstract, chapters and captions).

4.2 Image retrieval

For image retrieval, we extracted a set of features that are known to be effective in medical images retrieval (CEDD, FCTH, Local Binary Pattern histograms and color histograms - see the related work section). The features of all images in the corpus are stored in a fast L_2 index. The image retrieval results are sorted by their similarity, with the score being the L_2 distances between the query image and the result images. For case-base retrieval, an additional step must be performed: the image id (IRI), must be converted into a document id (DOI) (Figure 2 (a)) and the duplicate results must be merged to have an unique document list (Figure 2 (b)). More details are present in section 4.3.

4.3 Fusion

Result fusion aims at combining ranked lists from multiple sources into a single combined ranked list. Consider these

Rank	ImgId	Score
1	a2	1
2	b1	0.93
3	b2	0.90
4	c3	0.90
5	a1	0.73
6	c2	0.48

(a)

Rank	DocId	Score
1	a	1
2	b	0.93
3	b	0.90
4	c	0.90
5	a	0.73
6	c	0.48

(b)

Rank	Doc	Score
1	a	1
2	b	0.93
3	c	0.90

Figure 2: Case based retrieval step. (a) get document id (DOI) from image id (IRI); (b) combine multiple document results into one (unique) document list. The example uses CombMAX in fusion because it is easier to visualize.

two use cases: combine the results from queries with multiple images and combine the results from text and images queries. Query images can have different modalities (e.g. x-rays, PET scans, CT scans) and represent the same concept in multiple ways (e.g. hepatolenticular degeneration images can be represented by a photo of an eye or by a light microscopy of the affected cells). Thus, we took a late fusion approach, combining the results from multiple image queries into a single image result list. We found that late fusion of the results was also useful for heterogeneous queries (e.g text only, single image, text and 3 images), as the combination of the image and text search can be ignored if the query does not contain images.

There are two main approaches for late fusion: score based and rank based. Score based approaches (CombSUM, CombMAX and CombMNZ) combine the normalized scores given by the individual searches (e.g. image search, textual search) as a basis to the create the new ranked list. The studied variant that achieves the best performance [1] is CombMNZ, but ranked based fusion is gaining momentum, and can outperform score based fusion under most conditions [3, 4]. For each document i, the score after fusion can be computed as:

$$\text{combSUM}(i) = \sum_{k=1}^{N(i)} S_k(i), \tag{1}$$

$$\text{combMAX}(i) = \max(S), \forall S \subset D_i, \tag{2}$$

$$\text{combMNZ}(i) = N(i) \times \text{combSUM}(i), \tag{3}$$

where $S_k(i)$ is the score of the i document on the k result list.

$N(i)$ refers to the number of times a document appears on a results list. A result list k does not contain all documents. Documents with a zero score or a very high rank can be safely ignored. Thus, $N(i)$ varies between 0 (the document i does not appear on any list) and the total number of results list (the document i appears on all lists). For example, in our experiments, there are two results lists: one for image search and other for textual search, limited to 1000 results each.

Rank based fusion methods consider the position of each document in each one of the individual ranks. Reciprocal Rank and Reciprocal Rank Fusion are the two methods we evaluated:

$$\text{RR}(i) = \sum_{k=1}^{N(i)} \frac{1}{R_k(i)}, \tag{4}$$

$$\mathrm{RRF}(i) = \sum_{k=1}^{N(i)} \frac{1}{l + R_k(i)}, \mathrm{with}\, l = 60. \tag{5}$$

where $R_k(i)$ is the rank of document i on the k rank.

After analyzing both score and rank based approaches, we combined elements from both to improve precision. Inverted Squared Rank (ISR) combines the inverse rank approaches of RR and RRF (using the squared rank to improve precision at top results) with the frequency component of combMNZ (results that appear on multiple lists are boosted):

$$\mathrm{ISR}(i) = N(i) \times \sum_{k=1}^{N(i)} \frac{1}{R_k(i)^2}. \tag{6}$$

5. EXPERIMENTS

To assess the proposed methods, we tested the search-results fusion on the Medical case-based search task of the ImageCLEF 2013 evaluation campaign.

5.1 Dataset

Our dataset is composed of the data released for medical ImageCLEF 2013. It is a subset of over 70,000 PubMed articles with over 300,000 images. Each article is identified with an unique identifier (DOI) and is divided into title, abstract, chapters and image captions. All images on the dataset have an unique identifier (IRI) and can be associated with the corresponding article and caption.

5.2 Results

We compared the performance of the fusion algorithms using the best textual and visual runs. Our methodology was as following: for all (36) multimodal queries present in the ImageCLEF medical 2013, we searched text and images separately and combined our image and text runs using multiple fusion algorithms. Performance was evaluated using trec_eval and the relevance judgments provided.

Table 1: Fusion comparison for the medical ImageCLEF case based queries

Comb	MAP	GM-MAP	bpref	P@10
ISR	**0.1608**	0.0779	**0.14**	**0.1800**
RRF	0.1597	**0.0787**	0.13	0.1571
RR	0.1582	0.0779	**0.14**	0.1771
combSUM	0.0804	0.0039	0.09	0.1429
combMNZ	0.0794	0.0035	0.08	0.1371
combMAX	0.0292	0.0013	0.03	0.0457

With our data, rank-based approaches outperformed score based approaches by a factor of 2. One of the reasons is the differences between the scoring of the text and images. Even though both visual and text scores have the same normalization, the interval [0...1], the distribution of the results in the score space is different. Rank based approaches can handle multi-modality better, because the scores are not used.

Regarding the differences between RR, RRF and ISR: ISR performed better in our experiments in most of the measures, with a significant performance boost on P@10. This metric is particularly important for search engines, because most users won't browse beyond the first page of results (10 first). The polynomial component promotes top ranking results to the top of the list, offering a better user experience.

6. CONCLUSIONS

Our system combines a powerful framework based on state-of-the-art image and text processing algorithms with a simple yet powerful multimodal search interface to provide a valuable tool to retrieve medical data. In addition to the interface, we introduced ISR, a variant of RR and RRF aimed at increasing relevance of the results at the top of the list. We believe that it will help users to get relevant information, reducing frustration.

The system is still a work in progress. We are planing on testing the system with health care professionals to test usability and improve it.

7. REFERENCES

[1] N. J. Belkin, P. Kantor, E. A. Fox, and J. A. Shaw. Combining the evidence of multiple query representations for information retrieval. *Inf. Process. Manage.*, 31(3):431–448, May 1995.

[2] S. A. Chatzichristofis, K. Zagoris, Y. S. Boutalis, and N. Papamarkos. Accurate Image Retrieval Based on Compact Composite Descriptors and Relevance Feedback Information. *International Journal of Pattern Recognition and Artificial Intelligence (IJPRAI)*, 24(2):207 – 244, 2010.

[3] G. V. Cormack, C. L. A. Clarke, and S. Buettcher. Reciprocal rank fusion outperforms condorcet and individual rank learning methods. In *SIGIR '09*, pages 758–759, New York, NY, USA, 2009. ACM.

[4] D. Frank Hsu and I. Taksa. Comparing rank and score combination methods for data fusion in information retrieval. *Inf. Retr.*, 8(3):449–480, May 2005.

[5] MedGIFT Group. medSearch - Medical search engine by HES-SO Valais, 2009.

[6] M. Montague and J. A. Aslam. Condorcet fusion for improved retrieval. In *Proceedings of the eleventh international conference on Information and knowledge management*, CIKM '02, pages 538–548, New York, NY, USA, 2002. ACM.

[7] 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 *CLEF 2012 working notes*, 2012.

[8] K. Zagoris, A. Arampatzis, and S. A. Chatzichristofis. www.mmretrieval.net: a multimodal search engine. In *Proceedings of the Third International Conference on SImilarity Search and APplications*, SISAP '10, pages 117–118, New York, NY, USA, 2010. ACM.

[9] K. Zagoris, S. A. Chatzichristos, N. Papamarkos, and Y. S. Boutalis. img(Anaktisi): A Web Content Based Image Retrieval System. In *2009 Second International Workshop on Similarity Search and Applications*, pages 154–155. IEEE, Aug. 2009.

[10] M. Zhang, R. Song, C. Lin, S. Ma, Z. Jiang, Y. Jin, Y. Liu, and L. Zhao. Expansion-based technologies in finding relevant and new information: Thu trec2002 novelty track experiments. In *the Proceedings of the Eleventh Text Retrieval Conference (TREC*, pages 586–590, 2002.