Searching Text and Images in the Medical Domain

Allan HANBURY[a,1] and Henning MÜLLER[b]
[a] *Vienna University of Technology, Austria*
[b] *HES-SO, Sierre, Switzerland*

Abstract. This tutorial introduces the topics of medical information retrieval and medical image retrieval. After covering the basics of information retrieval, search in the medical domain including end user requirements is covered. Medical image search is also covered from the point of view of available techniques and end user requirements. The tutorial also gives concrete suggestions for improving search in the medical domain, and discusses the challenges that are currently not solved by existing approaches. The tutorial is aimed at people outside the domain of information retrieval, but who are interested in how medical search engines function and how search can solve problems in their domain.

Keywords. Medical information retrieval, medical image retrieval, challenges, information retrieval evaluation

Rationale

All parties involved in medical treatment are regularly faced with an information need that cannot be met from their own store of knowledge. For physicians, an unmet information need has been reported as occurring for 2 of every 3 patients seen [1], or for 41% of the questions they pursued [2]. This requires that they attempt to meet this information need by using available resources, which has traditionally involved searching in printed sources and asking colleagues, although searching on the Internet is of increasing importance. A recent survey done in the Khresmoi project [4] has shown that the three most common sources of online information used by physicians (in decreasing order of usage) are: general search engines (e.g. Google, Bing, Yahoo!), medical research databases (e.g. Pubmed) and Wikipedia.

Patients also have regular information needs, illustrated by the fact that 61% of American Adults seek out health advice online [3]. A Khresmoi survey of the general public revealed that the most common sources of online information used by this group are: general search engines (e.g. Google, Bing, Yahoo!), web sites providing health information (e.g. university, hospital, pharmaceutical company) and Wikipedia.

Finally, image search is particularly important in the medical domain. Internet image search applications are starting to appear (e.g. Goldminer, Yottalook). However, image search within PACS systems in hospitals is also being developed.

Given the increasing importance of search in the medical domain, it is important for end users to understand how search engines work and what their limitations are, allowing them to conduct efficient and effective online searches.

[1] Corresponding Author. hanbury@ifs.tuwien.ac.at, http://allan.hanbury.eu

1. Target audience

The target audience is people outside the information retrieval domain who wish to get a better feel for how search technology for both text and images works, and who wish to find out more about search requirements and solutions in the medical domain including their current limitations.

2. Prerequisite knowledge

Familiarity with using computers in general and a search engine is the prerequisite knowledge required. The tutorial will be at a technical level also suitable for people outside of the computer science domain.

3. Educational goal

The goal of this tutorial is to give people outside of the domain of information retrieval and search the background on how the techniques work for both text and image search, an idea for what is currently possible with state-of-the-art techniques and a brief overview of where current research and development is taking place. This should give them more confidence in using search engines in the medical domain, and beyond that provide the necessary knowledge to identify where a search-based solution could be useful in their area of work. Good and illustrative examples will be provided in order to complement the theory. A small part of the tutorial will also give a deeper background for people interested in developing their own solutions including image search as many open source components exist for this.

4. Contribution of each teacher

The outline of the tutorial topics is given below, with an indication of which teacher teaches which topic, and the approximate time required for each section.

1. Introduction to Information Retrieval (Allan Hanbury), 30'
This section covers the basics of information retrieval and how a text search engine works. This includes pre-processing (tokenization, stemming, …), indexing and retrieval. A brief overview of approaches used in web search will also be given. Finally, an introduction to how search engines can be adapted to specific domains is given.

2. Who searches for medical information and how do they search? (Allan Hanbury), 15'
Not all people searching for medical information on the Internet expect the same results. Expectations can differ in terms of the technical level (accessibly or technically written), level of specificity (overview or in-depth), language of the results, etc. This section gives an overview of the groups of people searching for medical information, the types of information they wish to access and the search processes that they use.

3. Search in the medical domain (Allan Hanbury), 20'

Approaches to adapting search engines to the medical domain are discussed and illustrated through examples of medical search engines. These approaches include the use of background information such as vocabularies, the classification of document types (e.g. primary and secondary literature) and dealing with the information quality.

4. Improving search in the medical domain (Allan Hanbury, Henning Müller), 25'

To initiate a discussion, some concrete suggestions for improving search in the medical domain will be provided. A subsequent discussion session will allow participants to describe their difficulties in medical information search. The teachers will discuss solutions to these difficulties where these exist, or will classify them as challenges not yet solved by current technology.

5. Searching for medical images (Henning Müller), 25'

Differences between text-based image search and content-based visual image search will be highlighted. The importance of images for gathering knowledge, for example of the literature will be explained. The basic structure of medical image search systems will be detailed.

6. Who searches medical images and how do they search? (Henning Müller), 20'

The part will detail the analysis of user surveys on their image search behavior performed over time and in different institutions. Beyond surveys, log files of PubMed, Goldminer and the HON media search were analyzed and will explain how people search and how image search behavior can be improved.

7. Combining text and image search (Henning Müller), 20'

Text-based search and content-based visual search have shown to be complementary in several system comparisons. Combining text and visual search well is still difficult and this part will explain these difficulties and how techniques can be combined to get optimal results with information fusion and a clear analysis of search goals.

8. Challenges for search in the medical domain (Allan Hanbury, Henning Müller), 25'

There are many challenges in the medical domain beyond Internet search for which search techniques have the potential to contribute to a solution. Examples include searching for "similar" anonymized patient records within a hospital to assist in diagnosis or treatment, or searching within a patient record to obtain an overview of the medical history. This discussion session will encourage the attendees to identify applications in their area of expertise that could take advantage of search techniques.

References

[1] Hersh WR, Hickam DH. How Well Do Physicians Use Electronic Information Retrieval Systems? A Framework for Investigation and Systematic Review. JAMA. 1998; 280(15):1347-1352

[2] Ely JW, Osheroff JA, Maviglia SM, Rosenbaum ME. Patient-Care Questions that Physicians Are Unable to Answer. JAMIA. 2007; 14:407–414

[3] Fox S, Jones S. The Social Life of Health Information. Pew Internet & American Life Project Report. 2009

[4] khresmoi.eu (accessed 10.4.2012)