Short title of the focus theme: Health Information Search
Long title of the focus theme: Health information search –free text, visual information, knowledge bases and combinations

Editorial: Health information search to deal with the exploding amount of health information produced

Henning Müller, Allan Hanbury, Najeeb Al Shorbaji

Summary

Objectives: This focus theme deals with the various aspects of health information search that are necessary to cope with the challenges of an increasing amount and complexity of medical information currently produced.
Methods: This editorial reviews the main challenges of health information search and summarizes the five papers of this focus theme.
Results and conclusions: The five papers of the focus theme cover a large part of the current challenges in health information search such as coding standards, information extraction from complex data, user requirements analysis, multimedia data analysis and the access to big data. Several future challenges are identified such as the combination of visual and textual data for information search and the difficulty to scale when analyzing big data.

Keywords: Health information search, information retrieval, knowledge bases, terminologies

Introduction

Medicine is one of the most information intensive professions and the amount of information available for decision making on single patients has been increasing steadily. It will most likely continue doing so in the foreseeable future. Communication of health information and related technologies are foreseen to be the next revolution in healthcare [1]. Medical images are even estimated to occupy 30% of world storage capacities, meaning that information compression and efficient access methods are required [10].
The goal of most hospital information systems or more general health information systems is to deliver the right information to the right people at the right place and time and in the right format [2].
Information retrieval or information search allows for effective information access on local document repositories and on the Internet [3], the largest linked information infrastructure in the world. Information retrieval has been extensively used in the medical domain. A very large variety of approaches exist, from Boolean text retrieval to ranked retrieval, natural language processing and mapping of text to terminologies.
Currently, there are several big challenges in health information search that researchers need to tackle to be able to influence care processes:

- The data representation needs to be interoperable between the various actors of the health system so that data quality can be high and ambiguity can be reduced. Both semantic interoperability (meaning of words and concepts) and syntactic interoperability (exchange format).
- The exact use cases and usage scenarios need to be known— why do people search and in which situations is search required? Only with this information can systems be integrated into the clinical workflow and help physicians to take into account all available data. It is important to know what is needed when and how the information is being used.
- Natural language processing and the analysis of free text that includes word synonyms, potential spelling mistakes and unusual abbreviations will remain a challenge even if many parts can be coded using a standardised terminology. Most likely a completely correct coding will not be realistic and tools for automatic free text analysis and error correction will remain necessary.
- Multimedia data have been used increasingly, from signals to images and now 3D volumes of tomographic slices. Including these data into retrieval processes is far from trivial and much research is underway. With combined modalities such as Positron Emission Tomography/Computed Tomography (PET/CT) or dual energy, the amount of data produced and requiring interpretation will also increase significantly. The increasing use of 4D data (3D images with a time component) requires innovative visualisation techniques to obtain the most benefit from the data. Furthermore, usually only extremely small parts of the data are of interest for analysis, and locating and identifying these specific parts is difficult.
- Many other complex data are available such as time series, for example for chronic diseases such as diabetes. These are available both on a population level and on an individual level. Detecting events, especially if these allow pre-emptive action to be taken, can be of tremendous benefit.
- Very large scale computing now becomes, in principle, possible, allowing treatment planning to be based on an entire patient record placed into the context of huge amounts of stored medical knowledge (big data). Still, legal barriers prevail [4] and also computational aspects need to be dealt with before full advantage can be taken of the opportunities available.

Influence of these techniques in developed countries can be important but in developing countries the potential for improvement using information and communication technologies is even much bigger. Organizations such as the WHO (World Heath Organization) can clearly help to distribute knowledge and particularly knowledge accessibility in the entire world to increase health and reduce risks[1]. Again, information retrieval can help in that the information is made accessible for many user groups and that the right information reaches the right people.

This focus theme can obviously only cover a small part of the vast research domain of health information search. Of the ten initially received submissions, five finally passed the review process and are presented in this section.

[1] http://www.who.int/governance/eb/constitution/en/index.html

Papers in this focus theme

The papers in this focus theme cover several of the challenging domains in information retrieval ranging from the use of basic terminologies for information coding and reuse, towards a survey to find out more about visual information retrieval requirements. The last two articles describe concrete applications, one on the diversity of information presented to a user and another one on the use of visual image retrieval and the identification of organs in images to help radiology information access.

In [5], Nitsuwat et al. describe the representation of information using standard terminologies. An adaptation of the International Classification of Diseases (ICD) terminology for a national situation in Thailand is described. This International Classification of Diseases-Thailand Modification (ICD-10-TM) is then used as a knowledge base in a system for semi-automatic coding. The results show the successful implementation of the system (95% accuracy) and further changes to it to avoid coding mistakes. User satisfaction was equally high.

In [6], Rasmussen et al. deal with the consistent coding of medical data using SNOMED (Systemized Nomenclature in Medicine) in Denmark. Using such an expressive terminology can be extremely helpful for interoperability between the various components in the health system. The work described defines guidelines to keep the coding across persons consistent despite redundancy and incompleteness. The study analyses 14 different templates from Denmark and Sweden.

Markonis et al. describe in [7] a survey that was executed among radiologists to get more insights into information needs concerning visual information. 34 radiologists were surveyed. A problem is that text search for images may be common but visual search tools are not known by most radiologists. Online image search is in any case a common task to obtain information in addition to searching for images in books or asking colleagues. Many search functionalities that are currently not available were requested, such as search for similar images or similar cases. Trust in the data found is also important, so showing evidence for a given diagnosis such as biopsy results or comments of other colleagues underlining the statements (as in social networks) helps to accept the information found.

In [8], Denecke describes an approach for diversity-aware information retrieval for medical web content. Diversity in search results has received much attention as user needs can vary and diversity in the results can help users to find the information required faster. The system described is evaluated using blog posts looking to identify whether information is factual or affective. The proposed approach obtains improved results at a statistically significant level.

Cavallaro et al. describe in [9] a system to identify structures in images and thus allow retrieving only partial volumes, which can help in making radiology work more efficient by saving loading time. Such a system is made possible through visually identifying structures in 3D and then allowing radiologists to navigate by anatomic region or by marking a similar volume of interest and finding similar volumes in other data sets.

Conclusions

The papers in this focus theme correspond well to current work and challenges in the field of health information search but cannot fully cover the wide spectrum of work in this field. Facing an ever-increasing production of data and an increase in the number of facts to be combined for medical decision-making, information overload is frequent and new information needs regarding search applications appear in many fields.
The papers deal with several of the current challenges in health information search and present results and solutions for these challenges.

Acknowledgements
This focus theme on medical information retrieval is partially funded by the EU through the FP7 project Khresmoi (FP7-257528).

References
[1] Chase D, Healthcare's medical instrument of the future: communication, Forbes Journal, 2011.
[2] Haux R. Health information systems - past, present, future. Int J Med Inform. 75(3-4):268-81, 2006.
[3] Hersh W, Information retrieval: a health and biomedical perspective, Springer, 2009
[4] Elger B, Iavindrasana J, Lo Iacono L, Müller H, Roduit N, Summers P, Wright J, Health Data Depersonalisation for Prospective research in the life sciences, Comput Methods Programs Biomed, 99(3):230-251, 2010.
[5] Nitsuwat S, Paoin W, Development of ICD-10-TM Ontology for a Semi-automated Morbidity Coding System in Thailand, Methods Inf Med, 2012.
[6] Rasmussen A, Rosenbeck Gøeg K, SNOMED CT implementation: Mapping guidelines facilitating reuse of data, Methods Inf Med, 2012.
[7] Markonis D, Holzer M, Dungs S, Vargas A, Langs G, Kriewel S, Müller H, A survey on visual information search behavior and requirements of radiologists, Methods Inf Med, 2012.
[8] Denecke K, An Architecture for Diversity-aware Search for Medical Web Content, Methods Inf Med, 2012.
[9] Cavallaro A, Kriegel HP, Petri M, Schubert M, Semantic Localization driven Partial Image Retrieval in CT Series, Methods Inf Med, 2012.
[10] EU Commission, Riding the wave - How Europe can gain from the rising tide of scientific data, European Commission, October 2010.

Contact details:
Henning Müller
HES-SO
TechnoArk 3
3960 Sierre, Switzerland
tel +4 127 607 9036
henning.mueller@hevs.ch