A Query and Patient Understanding Framework for Medical Records Search

Nut Limsopatham
School of Computing Science
University of Glasgow
Glasgow, UK
nutli@dcs.gla.ac.uk

Categories and Subject Descriptors: H.3.3 [Information Search & Retrieval]: Search process

Keywords: Representation; Inference; Medical Resources

ABSTRACT

Electronic medical records (EMRs) are being increasingly used worldwide to facilitate improved healthcare services [2, 3]. They describe the clinical decision process relating to a patient, detailing the observed symptoms, the conducted diagnostic tests, the identified diagnoses and the prescribed treatments. However, medical records search is challenging, due to the implicit knowledge inherent within the medical records - such knowledge may be known by medical practitioners, but hidden to an information retrieval (IR) system [3]. For instance, the mention of a treatment such as a drug may indicate to a practitioner that a particular diagnosis has been made even if this was not explicitly mentioned in the patient's EMRs. Moreover, the fact that a symptom has not been observed by a clinician may rule out some specific diagnoses.

Our work focuses on searching EMRs to identify patients with medical histories relevant to the medical condition(s) stated in a query. The resulting system can be beneficial to healthcare providers, administrators, and researchers who may wish to analyse the effectiveness of a particular medical procedure to combat a specific disease [2, 4]. During retrieval, a healthcare provider may indicate a number of inclusion criteria to describe the type of patients of interest. For example, the used criteria may include personal profiles (e.g. age and gender) or some specific medical symptoms and tests, allowing to identify patients that have EMRs matching the criteria.

To attain effective retrieval performance, we hypothesise that, in such a medical IR system, both the information needs and patients should be modelled based on how the medical process is developed. Specifically, our thesis states that since the medical decision process typically encompasses four aspects (symptom, diagnostic test, diagnosis, and treatment), a medical search system should take into account these aspects and apply inferences to recover possible implicit knowledge. We postulate that considering these aspects and their derived implicit knowledge at different levels of the retrieval process (namely, sentence, record, and inter-record level) enhances the retrieval performance. Indeed, we propose to build a query and patient understanding framework that can gain insights from EMRs and queries, by modelling and reasoning during retrieval in terms of the four aforementioned aspects (symptom, diagnostic test, diagnosis, and treatment) at three different levels of the retrieval process.

Firstly, at the sentence level, a medical negation detection tool is used to identify the context (negative/positive) of terms, which leads to an accurate representation of the medical conditions both in the EMRs and the queries. Handling negated language is challenging in medical records search, since it is commonly used by medical practitioners to indicate that a patient does not possess a particular medical condition [1]. In particular, we improve the representation of both EMRs and queries by representing terms along with their context, in order to prevent EMRs containing terms with the opposite context to the query's intent from being retrieved, since they are unlikely to be relevant. For example, preventing the EMRs stating "no history of dementia" from being retrieved for a query searching for patients with "dementia".

Secondly, at the record level, the semantic type and relationships of medical terms extracted from knowledge-based resources (e.g. ontologies, websites) are leveraged to infer the wider medical history of the patient in terms of the four above aspects. Indeed, we propose to represent EMRs and queries within the four medical aspects, and infer the relevance of an EMR by exploiting association rules extracted from the semantic relationships of medical terms within these aspects. For example, patients with a medical history of having *CABG surgery* (treatment) can be inferred as relevant to a query searching for a patient suffering from *heart disease* (diagnosis), since a CABG surgery is a treatment of heart disease.

Thirdly, at the inter-record level, we exploit knowledge about how the four medical aspects are handled by different hospital departments to gain further understanding about the appropriateness of EMRs from different departments for a given query. Specifically, we propose to aggregate EMRs at the department level (i.e. inter-record level) to extract implicit medical knowledge (i.e. expertise of each department) and model this department's expertise, while ranking EMRs. For instance, patients having EMRs from the cardiology department are likely to be relevant to a query such as "find patients suffering from heart attack".

We evaluate our work using standard test collections provided by the TREC Medical Records track [4], which represent the scenario of practitioners searching for patients based on the relevance of their EMRs.

SIGIR'13, July 28–August 1, 2013, Dublin, Ireland.
ACM 978-1-4503-2034-4/13/07.

1. REFERENCES

[1] W. Chapman, W. Bridewell, P. Hanbury, G. Cooper, and B. Buchanan. A simple algorithm for identifying negated findings and diseases in discharge summaries. *J. Biomed. Inform.*, 34(5), 2001.

[2] W. Hersh. *Information retrieval: A health and biomedical perspective (3rd ed.).* New York : Springer, 2009.

[3] N. Limsopatham, C. Macdonald, and I. Ounis. Inferring Conceptual Relationships to Improve Medical Records Search. In *OAIR* 2013.

[4] E. Voorhees and R. Tong. Overview of the TREC 2011 Medical Records Track. In *TREC* 2011.