Computers in Biology and Medicine 85 (2017) 98–105

Contents lists available at ScienceDirect

Computers in Biology and Medicine

journal homepage: www.elsevier.com/locate/compbiomed

A novel tool for the identification of correlations in medical data by faceted search

Danilo Schmidt[a,*], Klemens Budde[a], Daniel Sonntag[b], Hans-Jürgen Profitlich[b], Matthias Ihle[c], Oliver Staeck[a]

[a] *Medizinische Klinik für Nephrologie, Charité – Universitätsmedizin Berlin, 10117 Berlin, Germany*
[b] *German Research Center for Artificial Intelligence, DFKI, 66123 Saarbrücken, Germany*
[c] *Averbis GmbH, 79106 Freiburg, Germany*

ARTICLE INFO

Keywords:
Faceted search
Knowledge based systems
Medical domain
Decision support

ABSTRACT

This work focuses on the integration of multifaceted extensive data sets (e.g. laboratory values, vital data, medications) and partly unstructured medical data such as discharge letters, diagnostic reports, clinical notes etc. in a research database. Our main application is an integrated faceted search in nephrology based on information extraction results. We describe the details of the application of transplant medicine and the resulting technical architecture of the faceted search application.

1. Introduction

Due to new diagnostic technologies, such as medical imaging, detailed laboratory tests, computerized data collections, and electronic data storage, new possibilities have arisen. As medical records may cover a very long history of diseases (years to decades) and include a vast number of diagnoses, symptoms, results, medications, and laboratory values, there is a clear need for advanced search capabilities in information systems for an easy retrieval of relevant data. For example, physicians treating patients with complex chronic diseases could benefit from an easy search system to quickly examine a long disease history or other complex patient data. However, most medical information systems lack good search capabilities. In particular, for data with a large amount of unstructured text, it is difficult to perform a search for relevant informations in an appropriate amount of time. For example, identifying individuals or groups with certain attributes within a large cohort of patients is challenging and very time-consuming.

In this paper, we propose a three stage process: (1) offline textual information extraction from medical records for the use of transplant medicine; (2) the generation of interesting faceted search capabilities on the results of the previous stage; (3) the combination of the information extraction results with structured laboratory values (future work). Such a faceted search application (see chapter V System Architecture) uses a technique for accessing information organized according to a faceted medical classification system, allowing users to explore a collection of diagnoses, symptoms, results, medications, and laboratory values by applying multiple filters.

Thus, in the medical domain a user-centric faceted search system holds major benefits: first, the physician himself should have the ability to determine which parameters (laboratory values, diagnosis, drug, etc.) are integrated in the process of data handling and he or she can identify the impact of different variables (e.g. high urinary protein, urinary tract infection, tacrolimus drug level). Second, the addition or elimination of variables should affect the results in a traceable way. These results are provided by the faceted search. The faceted search should however follow well trained and established patterns of conventional medical decision making and differential diagnosis. Thus, the faceted search allows the analysis of complex data sets along a cognitive chain of decision-making [1,2].

We focus on the physician for evaluating predictors, influencing factors and diagnostic findings that allow him to find associated diagnoses by inclusion and exclusion of parameters. Finally, the user should have control of his actions and retrace the path to the solution. Furthermore, the faceted search allows physicians to identify groups of patients with similar attributes. This can provide valuable decision support, where physicians are confronted with situations where rare or complex diseases require a high degree of specialist knowledge. Evaluation of comparable cases simplifies the differential diagnosis and finding of correct therapeutic concepts.

In the following sections we describe the background of our motivation to bring the faceted search into the field of medicine. We

* Corresponding author.
E-mail addresses: danilo.schmidt@charite.de (D. Schmidt), klemes.budde@charite.de (K. Budde), daniel.sonntag@dfki.de (D. Sonntag), profitlich@dfki.de (H.-J. Profitlich), matthias.ihle@averbis.com (M. Ihle), oliver.staeck@charite.de (O. Staeck).

http://dx.doi.org/10.1016/j.compbiomed.2017.04.011
Received 9 December 2016; Received in revised form 5 April 2017; Accepted 13 April 2017
0010-4825/

give a short overview about the complex problems of kidney transplant medicine. Furthermore, we describe the underlying database and explain the potential of the faceted search. In the next sections we place our project in the context of related work. We describe the repository, the architecture and the user interface of our faceted search tool. In the conclusion and the outlook we explain the advantage of our tool for this special application and propose next steps to extend the facilities of the faceted search.

2. Background

2.1. Kidney transplant medicine

Kidney diseases have become a global health burden. Annually more than 3 billion Euros in Germany are estimated to account for the treatment of renal replacement therapies in patients with end-stage renal disease. The probability of occurrence of dialysis-dependent renal insufficiency is increasing by 3–5% per year [3]. Kidney transplantation is the most common treatment for patients with end stage renal disease, as it offers the lowest morbidity, lowest cost, a significant survival benefit, and the highest quality of life in comparison to maintenance dialysis. While over 8000 patients are on the waiting list for a kidney transplant, only 2195 kidney transplantations were performed in Germany in 2015 [4]. The main goals after transplantation are to increase the long-term graft survival and the reduction of complications. These goals are especially important due to the current organ shortage.

After years of chronic disease and dialysis treatment, kidney transplant recipients are largely affected by complex comorbidities. Additionally, they are at high risk of severe complications like rejections, infections, drug toxicity, malignancies, and cardiovascular diseases. The majority of patients have to take 5–10 different medications every day during their entire life. In order to archive optimal post-transplant outcome, it is necessary to identify and influence relevant risk factors and predictors for complications and poor outcome. Due to complexities of post-transplant management, kidney transplant recipients should remain in life-long specialized care. As a result of this complexity, numerous medical data are available, and decision-making is highly complex in clinical practice. Medical data are largely filed in multiple poorly structured or unstructured medical reports. An integrated faceted search application should help to automatically analyze and visualize this data.

2.2. Data description

The faceted search application is based on the nephrology database TBase. The web-based electronic patient record TBase was implemented in a German kidney transplantation program as a cooperation between the Nephrology of Charité- Universitätsmedizin Berlin and the AI Lab of the Institute of Computer Sciences of the Humboldt University of Berlin [5,6].

Since the year 2000 the relevant medical data of patients after kidney transplantation of the Charité were prospectively collected in the TBase database. The database is implemented as the electronic patient record in the daily routine for patient-centered care after kidney transplantation. The nephrology department of Charité to date contains records of more than 6500 transplant recipients and patients on the waiting list. Currently, TBase automatically integrates essential laboratory data (9.9 million values), clinical pharmacology (237.821 prescribed medications), diagnostic findings from radiology, pathology and virology (146.851 findings), and administrative data from the SAP-system of the Charité (70.591 diagnoses, 25.520 hospitalizations). Furthermore, all follow-up visits (131.584) are documented in TBase. In addition, the nephrology department of the Charité treats another 6000–8000 medical cases per year. This collection of detailed medical data opens new possibilities for the detection of rare disease patterns and quality management.

2.3. Potentials of faceted search

Two groups of applications of faceted search in the medical field can be distinguished: first, the use in clinical research, and second, the implementation in the individual treatment as a decision support system in the clinical routine. Applied to the clinical research the faceted search can help identify subgroups of patients with special characteristics (e.g. the search for certain medications, laboratory values in a defined range, distinct symptoms, or diagnoses matching the inclusion criteria of clinical trials). The faceted search could also be used to systematically detect side effects or drug interactions by identifying abnormal laboratory values, symptoms or diagnoses associated with certain agents or combinations of agents.

Another aspect is the application of the faceted search in the clinical routine and treatment of individual patients. For instance, it can be valuable for the identification of patients with special attributes in case of new therapeutic options, identification of medication side effects or contraindications.

Furthermore, it can be used to review the historical and actual medical reports of a certain patient to provide clinical decision support. It could identify possible drug interactions, link symptoms or alterations of laboratory values to a ranked list of potential unidentified diagnoses and thus increasing patient safety and help the clinician finding diagnoses and optimal treatment. Faceted search can be applied in the development of a decision support system that analyses both structured and unstructured medical reports of big cohorts of patients to link symptoms to a ranked list of diagnoses.

2.4. Text-mining in the clinical setting

Clinical text mining is aimed at extracting relevant information from unstructured clinical data and mapping it to a normalized content [7–9]. Since clinical texts differ significantly from other domains, special systems are developed for this purpose [10]. Well-known English-language systems are cTAKES [11], HITEx [12,13] and IBM's MedKAT [14].

With sufficient large amounts of training data, good results can be achieved with statistical methods, based on the presence of large sets of associations between texts and the normalized content [15,16]. If the number of associations is small, more knowledge-intensive, training-independent methods are necessary [17]. Here, rule-based approaches often deliver better results [18,19].

3. Related work

In [1] Sacco describes an approach of a guided interactive diagnostic system based on dynamic taxonomies. He uses a set-oriented search technique to support the user on free (un-ordered) browsing and exploring the information database. We extend this approach by a special multi-facet functionality. Biron et al. [20] describe an information retrieval system for computerized patient records implemented at Laon Barard Cancer Center in Lyon. The goal was to develop an application for full-text search on any unstructured medical information available in different repositories at CLB using the Solr search engine. Its most common use was the search on patient records at the bedside. Facets in their systems only included non-medical facts like age, sex or housing department. They came to the conclusion that negation detection is essential to avoid the numerous false-positive failures and that the consideration of synonyms and time labelling will lead to valuable additional results. In our application, both negation handling (via advanced text mining) and synonyms (by using medical ontologies) are covered and extend previous work. In particular, our approach shows the following main advantages: