Available online at www.sciencedirect.com
ScienceDirect
Int. J. Human-Computer Studies 68 (2010) 370–385
International Journal of Human-Computer Studies
www.elsevier.com/locate/ijhcs

Easing semantically enriched information retrieval—An interactive semi-automatic annotation system for medical documents

Theresia Gschwandtner[a,*], Katharina Kaiser[a], Patrick Martini[a], Silvia Miksch[a,b]

[a]Institute of Software Technology and Interactive Systems, Vienna University of Technology, Favoritenstraße 9-11/188, 1040 Vienna, Austria
[b]Department of Information and Knowledge Engineering, Danube University Krems, Dr.-Karl-Dorrek-Straße 30, 3500 Krems, Austria

Received 20 November 2008; received in revised form 19 July 2009; accepted 14 August 2009
Available online 21 August 2009

Abstract

Mapping medical concepts from a terminology system to the concepts in the narrative text of a medical document is necessary to provide semantically accurate information for further processing steps. The MetaMap Transfer (MMTx) program is a semantic annotation system that generates a rough mapping of concepts from the Unified Medical Language System (UMLS) Metathesaurus to free medical text, but this mapping still contains erroneous and ambiguous bits of information. Since manually correcting the mapping is an extremely cumbersome and time-consuming task, we have developed the MapFace editor.

The editor provides a convenient way of navigating the annotated information gained from the MMTx output, and enables users to correct this information on both a conceptual and a syntactical level, and thus it greatly facilitates the handling of the MMTx program. Additionally, the editor provides enhanced visualization features to support the correct interpretation of medical concepts within the text. We paid special attention to ensure that the MapFace editor is an intuitive and convenient tool to work with. Therefore, we recently conducted a usability study in order to create a well founded background serving as a starting point for further improvement of the editor's usability.

Keywords: Semantic annotation; Graphical editor; UMLS

1. Introduction

One of the most crucial challenges in modern medical care is the handling of huge amounts of information. Making medical text computer-interpretable is an important step towards the effective processing of medical information. The following two examples demonstrate this need.

Analysis of patient records: Checking medical patient records against given findings, comparing parameters of different records, or statistically evaluating specific disease patterns are common tasks of medical research and clinical care. In particular, when a huge amount of information is to be captured and analyzed, the automation of these tasks is crucial for effective processing. Medical concepts within these patient records have to become easily extractable and comparable, in order to allow for queries such as how many patients above 60 years received a specific medication and how they recovered. Permitting to query concepts rather than keywords requires the previous identification, disambiguation, and tagging of medical concepts within these patient records.

Making medical knowledge computer-executable: On the other hand, the generation of computer-executable knowledge, such as contained in clinical practice guidelines (CPGs) (Field and Lohr, 1990)—systematically developed instructions and recommendations on appropriate treatment of patients in order to improve the quality of health care by providing the best available scientific evidence—is a very challenging task. For a detailed description of different approaches we refer to Peleg et al. (2003) and to Wang et al. (2002). However, the process of translating a

*Corresponding author. Tel.: +43 15880118843; fax: +43 15880118899.
E-mail addresses: gschwandtner@ifs.tuwien.ac.at (T. Gschwandtner), kaiser@ifs.tuwien.ac.at (K. Kaiser), martini@ifs.tuwien.ac.at (P. Martini), miksch@ifs.tuwien.ac.at (S. Miksch).

1071-5819/$ - see front matter © 2009 Elsevier Ltd. All rights reserved.
doi:10.1016/j.ijhcs.2009.08.002

textual CPG into a computer-executable model is an extremely time-consuming and cumbersome task, whereas medical experts and knowledge engineers have to collaborate intensively. Unfortunately, there is a great risk for knowledge engineers—who in general are not familiar with the details of medical concepts—to misinterpret the complex medical text while modeling the CPG. Thus, it appears that above anything else special attention has to be paid to ensuring the accurate interpretation of the text. Consequently, the disambiguation of the medical concepts included in the text of the document is imperative for correctly interpreting the text while further processing it into a computer-executable model.

As a first step towards the effective processing of medical text, the text needs to be translated into a computer-executable form. Thus, preprocessing the text by semantically annotating the medical concepts included is an important prerequisite to ensure an accurate interpretation of text. The semantic annotation of the narrative text of medical documents disambiguates the medical concepts included in the text and allows for semantically enriched information retrieval. Consequently, semantic annotations not only support the correct interpretation of the text, but also pave the way for more sophisticated analysis of the textual content such as statistical evaluations.

Multiple systems exist which generate a mapping from free biomedical text to medical terminologies, such as MicroMeSH (Elkin et al., 1988), Metaphrase (Tuttle et al., 1998), MetaMap (Aronson, 2001), PhraseX (Srinivasan et al., 2002), or KnowledgeMap (Denny et al., 2003). For a detailed description of annotation systems we refer to Reeve (2007).

We use the MetaMap Transfer (MMTx) program (Aronson, 2001) which maps concepts from the Unified Medical Language System (UMLS) Metathesaurus (Schyler et al., 1993)—the largest and well-accepted thesaurus in the biomedical domain—to corresponding concept chunks in the text.

1.1. The problem

It goes without saying that the complete reliability of the annotations is crucial in medical care—an extremely sensitive discipline. However, due to the ambiguity of free text, the automatic creation of an unambiguous mapping of UMLS concepts to medical concepts in the text of a document cannot be correctly accomplished by means of a semantic annotation system alone. The reliability of the MMTx results is not granted: on the one hand, MMTx cannot always determine an appropriate or distinct concept for a text chunk and, on the other hand, MMTx sometimes provides the wrong syntactical information, which causes errors in the concept assignment. This, in turn affects the reliability of the annotated semantic information.

Thus, it is of paramount importance for medical experts to control these results and, if necessary, to correct them. However, controlling and modifying the output of the MMTx program is hard to handle for physicians, since the use of MMTx is command line based, and there is no graphical user interface. It requires programming knowledge to utilize the results of the MMTx program.

1.2. An interactive editor to support these tasks

There are two significant aspects if one wants to derive the greatest possible advantage from the information gained from semantic annotations of documents:

- guaranteeing the quality of the provided information, and
- providing easy access to all available information in combination with visualizing the given information in a comprehensive form.

We have developed an interactive editor to cope with these two aspects: the MapFace editor. It greatly facilitates the handling of the MMTx program, and, being easily extendable, it has the potential to become the one tool to deal with all kinds of semantic information necessary to ensure the accurate interpretation of a medical text. In the near future, the MapFace editor is to be extended by additional components, such as a component to detect co-references in the text or a component to detect negated phrases (Gindl et al., 2008). These components take advantage of diverse information extraction methods to automatically perform these tasks, but again medical expertise is required to control the results.

Thus, the MapFace editor is designed to visualize different kinds of semantic information, to make them navigateable and modifiable, but also to advise the user of those cases, for which a sufficient result could not be computed automatically, and last but not least to visualize important coherences to support a better understanding of the medical text. Using the MMTx program and the UMLS, which is a comprehensive biomedical thesaurus, MapFace can be used to semantically annotate medical documents of any medical domain.

Usability: We have evaluated the usability of the editor (see Section 6) to identify possible shortcomings of the design. This evaluation has yielded a number of suggestions of how to further improve MapFace; we will adapt the editor accordingly, thus ensuring an intuitive and convenient way of working with it.

To start with, we give an outline of related work in Section 2. In Section 3 we describe the UMLS and the MetaMap program. We continue with a detailed description of the problems involved in the automatic semantic annotation of a document by means of the MMTx program in Section 4. In Section 5 we introduce the MapFace editor, its components, and its features. Subsequently, we describe the setting and state the results of the usability evaluation of the MapFace editor in Section 6. We outline the extensions and improvements we intend to