MEDINFO 2007
HISA Ltd Australia
© 2007 The authors. All rights reserved.

A Framework to Extract, Interpret and Structure Clinical Information From Free-Texts

Gisele Azambuja [a], Tatjana Zrimec [a], Andrew Hopkins[b]

[a] *Centre for Health Informatics, University of New South Wales*
[b] *Cardiology Department, The Liverpool Hospital*

Abstract

The major barrier in existing hospital information systems is that most information is stored in free-text medical narratives and is unavailable for decision support and research. Our aim is to develop a general medical text processing system to extract clinical information from different types of free-text clinical records. We propose a framework to encode and represent clinical information relying on UMLS Knowledge Resources.

Keywords:

natural language processing, medical records

Introduction

For more than three decades physicians, computer scientists and linguists have addressed Natural Language Processing (NLP) techniques in medicine. Applications include extracting selected types of information from free-text clinical notes and automatic coding of patient information in the terms of a controlled medical vocabulary [1, 2]. When referring to natural language commonly used in medicine, NLP systems are also called Medical Language Processing (MLP) systems [3]. When adequate processing of texts is achieved, MLP systems can make clinical information available for a variety of applications and still preserve the knowledge integrity while offering freedom of expression to physicians.

Conventional medical narratives are commonly stored in existing electronic medical records as free-text clinical notes. We propose a method and provide a framework to automatically process clinical free-text reports from different medical domains in order to extract and represent relevant clinical information. This information can then be easily accessed for other computer applications. In this paper we focus particularly on creating a structure to represent information from free-text clinical notes.

Materials and methods

We based the development of our system on the assumption that, when reading free-text clinical notes, physicians want to know what, when, where, and why health care activities were performed as well as their outcomes. These questions are related with general categories of information as follows: “Event”, “Time”, “Place”, “Reason” and “Outcome”. In our schema, designed to represent information from the free-text, we used the translation of above mentioned categories into semantic types from Unified Medical Language System (UMLS). In this way, “Event” was mapped to “Health Care Activity”, “Time”, to “Temporal Concept” and “Place” became “Health Care Related Organization”. The category “Reason” was mapped to both, “Disease or Syndrome” and “Sign or Symptom”. Since, in our schema, most classes and sub-classes are based on UMLS semantic types they can be related with each other by predicates existing in the UMLS Semantic Network. Thus, assertions such as, “Health Care Related Organization *carries_out* Health Care Activity” and “Diagnostic Procedure *is_a* Health Care Activity” are represented in our structure. However, the UMLS semantic types do not cover all classes of information necessary to represent and structure free-text clinical notes. Therefore, we added other classes to the schema, like “Outcome”. Similarly, we also added some relationships between classes that are well known as existing in free-text medical narratives, even though they do not exist in the UMLS Semantic Network. An example is the relation *occurs_in* between the classes “Health Care Activity” and “Temporal Concept”.

Results

We have created a schema to represent information from clinical notes. The idea here is to represent sentences from a clinical text using an expandable structure which consists of classes of information, relationships among them and a set of attributes for each class of information. At this stage we are evaluating the schema and applying automated techniques to discover rules for the processing task. To date, we have defined some hypotheses which have the potential to develop into these rules.

References

[1] Haug PJ, Ranum DL, Frederick PR. Computerized extraction of coded findings from free-text radiologic reports. Work in progress. Radiology 174 (2): 1990, 543-8.
[2] Friedman C, Alderson PO, Austin JH, Cimino JJ, Johnson SB. A general natural-language text processor for clinical radiology. J Am Med Inform Assoc. 1 (2): 1994, 161-74.

[3] Friedman C, Hripcsak G. Natural language processing and its future in medicine. Acad Med 74 (8): 1999, 890-5.

Address for correspondence

gsele@student.unsw.edu.au