



Clinical Natural Language Processing tools for SNOMED CT

Presenter: Richard Gain, Clinithink

Audience

This presentation will be aimed at clinical users, systems developers, data analysts and anyone interested in tools that generate full, post-coordinated SNOMED CT expressions from clinical narrative.

Objectives

Learn how a Clinical Natural Language Processing engine can be integrated with SNOMED CT to harness the power of this rich clinical terminology in the way that it was designed to be used. Data entry becomes simpler, data analysis more intuitive, and clinical decision support tools functionality enhanced. (1)

Abstract

SNOMED CT is a compositional terminology which is at its most powerful when the full richness of a clinical record is expressed using post-coordinated SNOMED CT expressions. However, these expressions can be difficult and time-consuming to construct, even for SNOMED CT experts.

Many systems providers have sought to get around this limitation by constructing large numbers of data entry templates, for the many different common clinical scenarios, that offer a selection of attributes which can be used to add qualifying and contextual detail to the basic coded concepts.

There are several problems with this approach, including the difficulty of developing and maintaining the templates, the degree of constraint imposed by predefined data entry forms, and their general acceptability to clinical users as a data entry mechanism. The simple fact is that most clinicians would rather just write notes than have to fill in forms.

Using Natural Language Processing (NLP) to parse the clinical narrative is not a simple task. Records often contain complex sentence constructions including technical terms, copious abbreviations and acronyms, comma-separated lists of items which may be present or absent, and implied contexts based on section headings within the record. Speed and accuracy of processing are paramount, but there is always a trade-off between the two. For a data entry use case, the system needs to be capable of encoding typed information in near-real time. Batch processing of high data volumes will require systems capable of encoding thousands of documents per hour. Clinical NLP will never be 100% accurate because of the ambiguities of language. Word sense disambiguation is one of the great challenges for this technology, but with access to large volumes of data the statistical models steadily improve, as demonstrated by Google search.

Our approach has been to create a unique blend of natural language processing and ontology-driven relationships provided by SNOMED CT, which allows us to draw on the most useful aspects of each.

One of the most exciting new areas of work is the use of queries for data abstraction. Having amassed a large database of SNOMED CT encoded clinical records, the next challenge is to simplify the process of analysing the data according to user-specified criteria. These may be predefined criteria such as coding schemes like ICD, CPT or OCPS, or completely customer specific information requirements.

References

1. Link to [Mount Sinai Hospital press release](#)