

Closing the gap between SNOMED CT and unstructured plain medical texts: Reflections on a first experience in French

Brice de Behault, EarlyTracks
Thibault Bughin, EarlyTracks
Céline Andris, EarlyTracks

Audience

This presentation is targeted to implementers from healthcare organisations.

Objectives

First, to show that it is possible to reconcile structured and unstructured views in electronic patient record (EPR). Second, to share a first-hand experience trying to close the gap between a standard terminology and the reality observed in EPR.

Abstract

A significant part of a patient's clinical information is stored in unstructured format in EPR. Free texts have always been a very natural way of expression for medical practitioners, but this format represents an important limitation to the access, retrieval and sharing of information. Because of the lack of machine readability, IT and medical managers from different hospitals have each developed their own independent system to try to introduce structured data into EPR. Typically, this means asking practitioners to structure content at input time. This implies, among other things, to force a change in the practitioners' habits, to sacrifice the expressiveness of the medical information and to reduce the ergonomics of the EPR encoding system, which in turn seriously endangers end-user adoption.

Thanks to recent developments in the area of natural language processing, it is now possible to automatically and highly effectively extract specific pieces of information from a text, and consequently resolve the dilemma between structured and unstructured data. This means that medical concepts such as disorders, symptoms and procedures can be automatically recognized and extracted from plain text with great accuracy.

Based on that high quality level of extraction, additional processing can be used to semi-automatically generate local extensions of terminologies like SNOMED CT. Eventually, when quality meets the acceptable requirements of end users, this whole process ends up bridging the gap between structured and unstructured content as both of them become interchangeable, leading to ergonomic, natural and expressive inputs.

In this presentation, we will introduce the high-level architecture and methodology that we have built (first as a research project) as well as the results obtained based on partnerships with hospitals and EPR providers.

These results will then be used to highlight the observed high importance of proper local (hospital-level) SNOMED CT extensions on top of national and international editions in order to achieve a satisfactory adoption of SNOMED CT by practitioners.