Abstract

The aim of my work was to find out the specific features of Czech medical reports in terms of the possibility of extracting specific information from them.

For my work, I had a total of 268 anonymized narrative medical reports from two outpatient departments. I have studied standards for preserving electronic health records and for transferring clinical information between healthcare information systems. I have also participated in the process of implementing electronic medical record in the field of dentistry.

First of all, I tried to process narrative medical reports using natural language processing (NLP) tools. I came to the conclusion that narrative medical reports in the Czech language are very different than a typical Czech text, especially because it mostly contains short telegraphic phrases and the texts lack typical Czech sentence structure. It also contains many misspellings, acronyms and abbreviations. Another problem was the absence of existence of the Czech translation of the main international classification systems. Therefore I decided to continue the research by developing the method for pro-processing the input text for translation and its semantic annotation.

The main objective of this part of the research was to propose a method and support software for interactive correction and semantic annotation of narrative medical reports that would allow their easier use with fewer errors outside of their original context.

I have developed a three-phase pre-processing method to encourage secondary use of medical reports. The method of three-phase pre-processing of narrative clinical reports was verified on 49 anonymous Czech medical reports in the field of cardiology.

Keywords

Clinical reports, natural language processing, information extraction, electronic health record