mEx - An Information Extraction Platform for German Medical Text

Roland Roller, Christoph Alt, Laura Seiffe and He Wang
Language Technology Lab, DFKI, Berlin, Germany

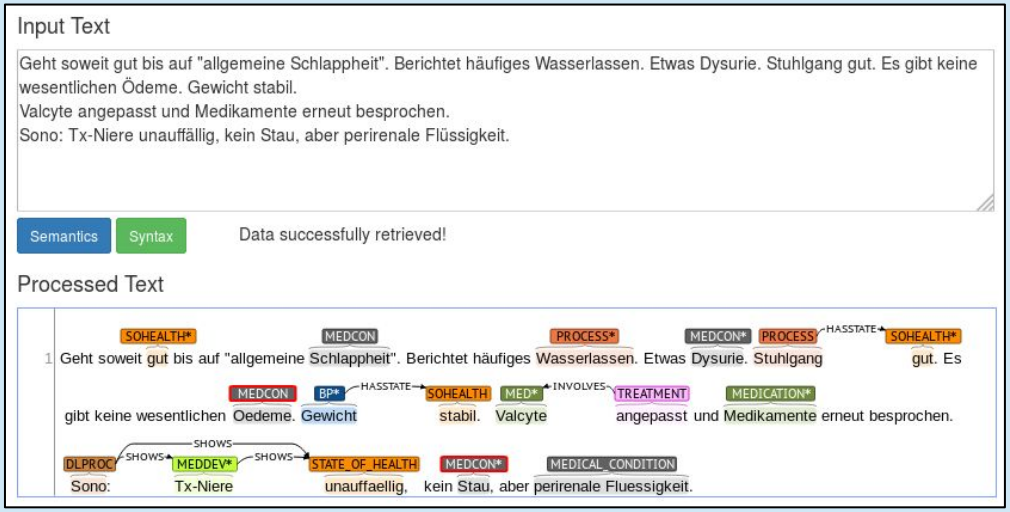
Problem/Motivation

Within the clinical routine, text data usually contains additional information beside the structured data (e.g. lab values). This information can be very useful for clinical decision support and end-point prediction. However accessing this information is often difficult as most NLP tools are not specialized on this kind of text. For non-English clinical text, the situation is worse. Only a few tools exist which makes the development cumbersome.

Solution

We present mEx, an Information Extraction platform for medical text. Some characteristics: ***Language:*** *German* - ***Domain:*** *Nephrologie* - ***Technologie:*** *Semantic & Syntactic components (see right side)* - ***Access:*** *via web & REST interface*

Interface



Information Extraction Components

Part-of-Speech Tagger A POS tagger assigns each token a part-of-speech, currently we integrate the Jena Part-of-Speech Tagger (jPOS).

Dependency Tree Parser A dependency tree parser infers the syntactic structure of a sentence. mEx integrates a re-trained dependency parser optimized for German clinical text.

Named Entity Recognition A NER component detects mentions of pre-defined entities in text, such as drug mentions, body parts or diseases. Our NER component bases on a Bi-LSTM with CRF.

Factuality Detection Negations and vague descriptions are a vital part of clinical documentation, as doctors often speculate on the presence of diseases. mEx integrates a modified version of NegEx.

Relation Detection A relation describes a particular relationship between concepts or entities, such as a medical condition occurs in a particular body part. Our relation extraction component bases on a CNN.

Concept Normalization In medical documentation, different entity mentions can refer to the same concept, e.g. ‘pain in the head’ and ‘cephalgia’ both refer to the concept of ‘headache’. mEx employs a two-step concept normalization (candidate search and disambiguation) in which mentions are linked against an identical concept within a biomedical ontology.

This research was supported by the German Federal Ministry of Economics and Energy (BMWi) through the project MACSS (01MD16011F).