

Extraction of temporal expressions and clinical events from clinical notes

Marcia Barros¹, André Lamúrias¹, Marta Antunes¹, Joana Teixeira¹,
Gonçalo Figueiró¹, Alexandre Pinheiro¹, Francisco M. Couto¹

¹ LaSIGE, Departamento de Informática, Faculdade de Ciências, Universidade de Lisboa

The amount of clinical notes is increasing and there is a need to implement methods to automatically extract useful information from them, like temporal expressions or clinical events. For example, in SemEval 2016 (Semantic Evaluation) competition there is a Clinical TempEval task that provides clinical notes and pathology reports for cancer patients from the Mayo Clinic, and asks participants to extract temporal information (the spans of time expressions, the spans of event expressions, the attributes of time expressions, the attributes of event expressions, the relation between an event and the document creation time, narrative container relations). Thus, we are developing a system for addressing these challenges, more precisely to identify the span of temporal expressions and clinical events. The system is based on a machine learning framework to identify biomedical entities, named IBEnt, which applies the StanfordNER tool to build a recognition model. We are using a particular library, the SUTime, available as part of the Stanford CoreNLP pipeline, which allows the recognizing and normalizing of time expressions. We also created a list of rules that to complement that tool. The list of rules was manually created by a team of curating by analyzing a set of texts annotated as well as expressions in the Unified Medical Language System (UMLS) which were considered as temporal concepts. We also plan to use IBEnt relation extraction capability to identify the relation between an event and the document creation time, and narrative container relations. As preliminary results for

identify the span of time expressions and the spans of event expressions we achieved a medium precision of 67%, a recall of 68% and a F-Score of 67% and a medium precision of 74%, a recall of 76% and a F-Score of 75%, respectively. We intend to use ontology (such as SNOMED-CT) mapping techniques to improve the precision of the results.

Preference for presentation: Oral

Location: University of Minho

Author for Correspondence: marcia.c.a.barros@gmail.com