Multilingual Domain-sensitive Temporal Tagging with HeidelTime

Jannik Strötgen

Max Planck Institute for Informatics, Saarland Informatics Campus

jannik.stroetgen@mpi-inf.mpg.de

Due to the prevalence of temporal expressions in diverse types of documents and the importance of temporal information in any information space, the detection of temporal expressions and the normalization of their semantics to some standard format is an important task in NLP. This task is known as temporal tagging, and many types of applications, e.g., in information retrieval, question answering, and digital humanities, can benefit from the output of temporal taggers to provide more meaningful and useful results.

Research on temporal tagging has focused on processing English news articles for quite a long time. Only more recently, challenges of texts of other domains have also been studied (e.g., clinical documents, literary narratives, and colloquial text), and further languages have been addressed. A general overview on temporal tagging is given by Strötgen & Gertz (2016).

In this system demonstration, we present the temporal tagger Heidel-Time. It uses Timex3 tags, which are defined in the temporal markup language TimeML (Pustejovsky et al., 2005), to annotate dates (`July 2017`), times (`9 pm`), durations (`two days`), and set expressions (`twice a week`). HeidelTime is a rule-based system initially developed for English news articles, but later extended to process more languages and to tackle the challenges of different domains. It is publicly available (`https://github.com/HeidelTime/heideltime/`) and constantly maintained, e.g., it was recently extended with automatically created language resources for more than 200 languages in addition to the 13 languages for which resources have been manually developed by several researchers.

The main goals of this demonstration are (i) to show how to use Heidel-Time for multilingual, domain-sensitive temporal tagging, (ii) to explain how HeidelTime's language resources can be adapted without even touching HeidelTime's source code, and (iii) to discuss further application scenarios that can benefit of a temporal tagger's output.

References: ● Pustejovsky et al. (2005): Temporal and Event Information in Natural Language Text. *Language Resources and Evaluation*, 39(2-3):123–164. ● Strötgen & Gertz (2016): Domain-Sensitive Temporal Tagging. *Synthesis Lectures on Human Language Technologies, Morgan & Claypool Publishers*.