Automatic Template Filling in a Biomedical Domain

Abstract for the Expert Session in **Trends in Bioinformatics** on February 3rd 2016
by Janos Brauer

The amount of information in the biomedical literature available for researchers is growing exponentially[1]. For tasks such as understanding the state-of-the-art in a given area and generating new hypotheses, researchers need to navigate through a vast amount of information from biomedical literature.

Information Extraction (IE) from biomedical literature has proven to be a very effective tool to cope with the abundance of papers. "*The goal of IE research is to build systems that find and link relevant information while ignoring extraneous and irrelevant information*"[2].

Researchers of the Deutsche Krebsgesellschaft have been manually extracting information out of oncology papers' abstracts into tabular schemas. This task is both time-consuming and error-prone, therefore this project investigates the automation of the process by using the Information Extraction methodology Template Filling.

Template Filling is a promising method for helping researchers seeking information for a particular need to efficiently extract precise answers from one or multiple documents. In Template Filling a template, which is an abstract schema, contains slots to be filled with specific information of interest. Fill Rules serve as extraction guidelines to fill the slots of the template.

Existing software tools for Template Filling are either domain-specific or publicly unavailable, for which reason the creation of a new tool for the field of oncology research is conceptualized in this project. The integration of existing tools for Natural Language Processing (NLP) in a biomedical context into a Template Filling system is investigated and prototyped. Additionally, solutions for challenges of IE in the biomedical domain such as synonyms of terms or language translations are researched.

Text	Template
Sentence 1 One hundred ten Patients were randomly assigned	**Study Count**: 110
Sentence 2 Patients received treatment by laser	**Treatment**: Laser therapy

Figure 1: Text on the left is used to fill slots of the template on the right

In my expert session I will talk about Template Filling more thoroughly. I will present how and which tools to use to fill the schema of a template. Finally, I will address future work and pending challenges in the field of Template Filling in a biomedical domain.

[1] R. Smith. Strategies for coping with information overload, 2010.
[2] J.Cowie, W. Lehnert. Information extraction. Communications of the ACM 39.1 (1996): 80-91.