NLP04 Information Extraction from Research Papers for DIGIS

Background

The objective is to devise approaches for the automated extraction of geochemical data and metadata from research papers and implement them prototypically pipeline for the geochemical data infrastructure DIGIS.

We extract specific mentions of methods from papers. This information can be part of the paper or included in tables or figures. The structure depends on the journal.

DIGIS

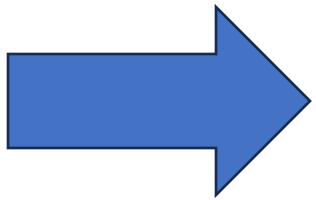
Digital Geochemistry Infrastructure for GEOROC 2.0

Goal

- A protoype for extraction specific information from research papers

Tasks

- Compare existing approaches for information extraction for a given set of papers
- Implement a prototype
- Draft a data pipeline



Daniel Kurzawe

kurzawe@sub.uni-goettignen.de

Mathias Göbel

goebel@sub.uni-goettingen.de

NLP23 Textual Criticism and Plagiarism

Background

“The identification of textual variants, or different versions, of either manuscripts or of printed books” (Wikipedia) is a major task in philology entitled “textual criticism”. The analysis of a single text in different variants starting from the very first sketch up to the latest authorized version is provided with a historical-critical edition. Before the digital age, these editions used an obnoxious amount of signs marking and categorizing these differences, like the Leiden convention has standardized. However, newer visualization technologies provide more and more interactive views to these editions.

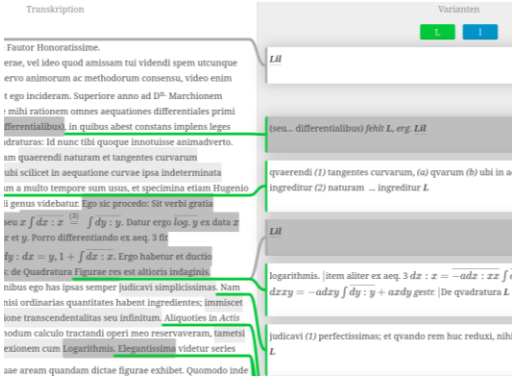
What are the shared approaches of plagiarism detection and textual criticism? Can they benefit from each other?

Goal

- Investigate if/how a software for plagiarism detection can be utilized to deal with a set of documents that represent the same text. Comparison of both methods.

Tasks

- Input material selection (assisted)
- Data conversion
- Usage of PD software/visualization, publish result using web technology
- Workflow to automatize main steps and to scale up (deal with as many editions as possible)



Daniel Kurzawe

kurzawe@sub.uni-goettingen.de

Mathias Göbel

goebel@sub.uni-goettingen.de