Overview

UTEMPL is a flexible tool for information extraction and text mining:

- based on UIMA
- easy use of several input formats:
  - Medline abstracts
  - Journal articles (HTML or XML)
  - Patents
  - ...

- standardized interfaces: integration of
  - NLP tools from different developers
  - new analysis engines for new tasks
  - high exchangeability
  - combination and selection of various internal and external applications

Advantages

- easily adaptable to new tasks
- allows to compare and evaluate different tools or methods
- efficient adaptation of existing workflows

UTEMPL workflow

Processing tools

- Text Zoning
- Sentence Splitter
- several Tokenization Tools
- PoS Tagger
- various CRF based Named Entity Recognizer
- ProMiner containing various dictionaries and regular expression
- Pattern based relation extraction
- Dependency Parser based relation extraction

References


Contact

Juliane Fluck
Fraunhofer SCAI
Schloss Birlinghoven
53754, Sankt Augustin,
Juliane.Fluck@scai.fraunhofer.de
http://www.scai.fraunhofer.de