An End-To-End Integration of Automatic Annotations into CATMA

Thomas Bögel* and Marco Petris† and Jannik Strötgen* and Michael Gertz*

* Database Systems Research Group, Heidelberg University, Im Neuenheimer Feld 348, 69120 Heidelberg, Germany
† Institute for German Studies, University of Hamburg, Von-Melle-Park 6, 20146 Hamburg, Germany

Automatic annotations in CATMA

Goal: reducing manual annotation effort by integrating automatic annotations into CATMA [1].

Context of heureCLEA

• BMBF-funded Digital Humanities project
• Investigating temporal phenomena in literary texts

→ UIMA: flexible & extendible pipeline architecture

Results of the annotation process

Integration of UIMA annotations into CATMA

Information about CATMA

• Sharing texts, annotations and annotation schemes allows for collaborative annotation and analysis of texts
• Unlimited and free creation of hierarchical annotation schemes tailored to your research question
• Powerful query language for exploration and analysis of text and annotations
• An easy to use graphical user interface keeps entry barriers low

UIMA components & Machine Learning [2,4]

• Surface phenomena for predicting complex annotations
• Challenges due to literary domain
• Flexible pipeline with multiple systems for the same task
• Voting/expert schema for specific tasks
• Hybrid approaches (heuristics & machine learning) yield good performance
  – Correct ML errors by explicit user feedback
  – Tailor ML models towards individual users

References


Project Information:

http://heureclea.de/

This work was presented at 2. Jahrestagung der Digital Humanities im deutschsprachigen Raum (D Hd), February 25–28, 2015, Graz, Austria.