



UIMA als Plattform für die nachhaltige Software-Entwicklung in den Digital Humanities



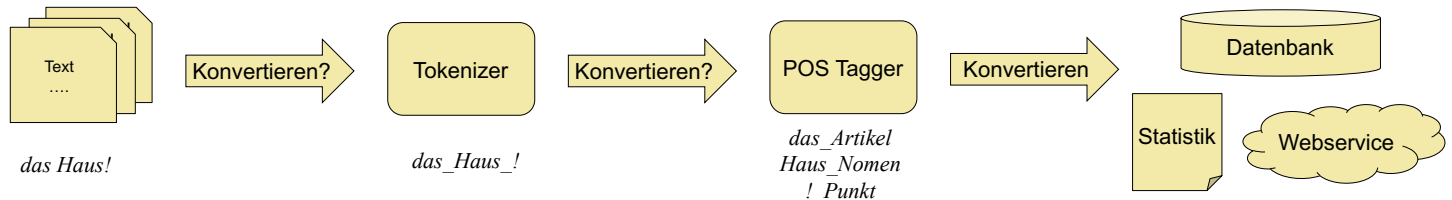
Johannes Hellrich^{1,2} – Franz Matthies² – Udo Hahn²

¹ Graduiertenkolleg „Modell Romantik“

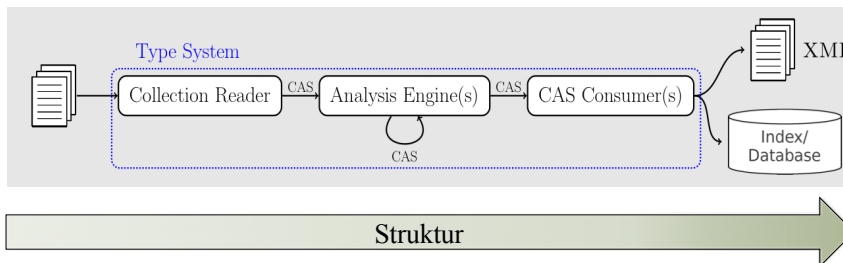
² Jena University Language & Information Engineering (JULIE) Lab
Friedrich-Schiller-Universität Jena, Jena, Deutschland



Computerlinguistische Pipelines

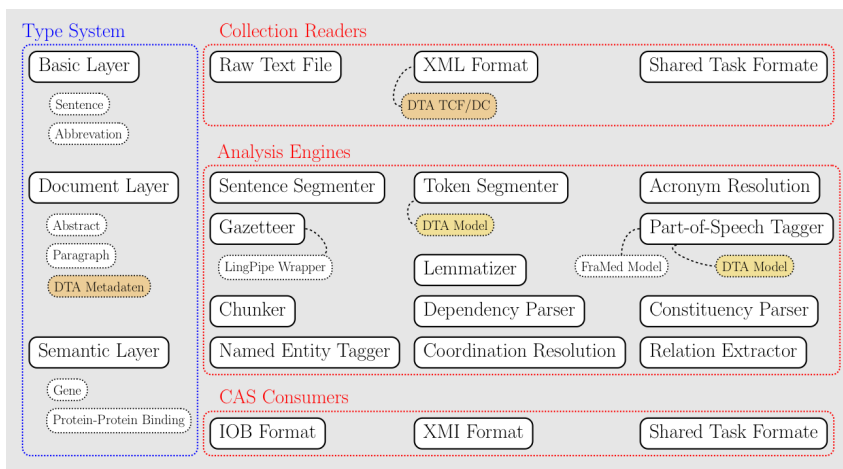


Unstructured Information Management Architecture (UIMA)



- **Type System:** Definiert mögliche Annotationen
- **CAS (Common Analysis Structure):** Speichert Texte und Annotationen
- **Pipeline:**
 - *Collection Reader:* Liest Daten in die CAS
 - *Analysis Engine:* Fügt Annotationen hinzu
 - *CAS Consumer:* Speicherung der Daten in der CAS

JULIE Component Repository (JCoRe)



Open Source

- GitHub
- Größtenteils Apache Lizenz

Zugänglich

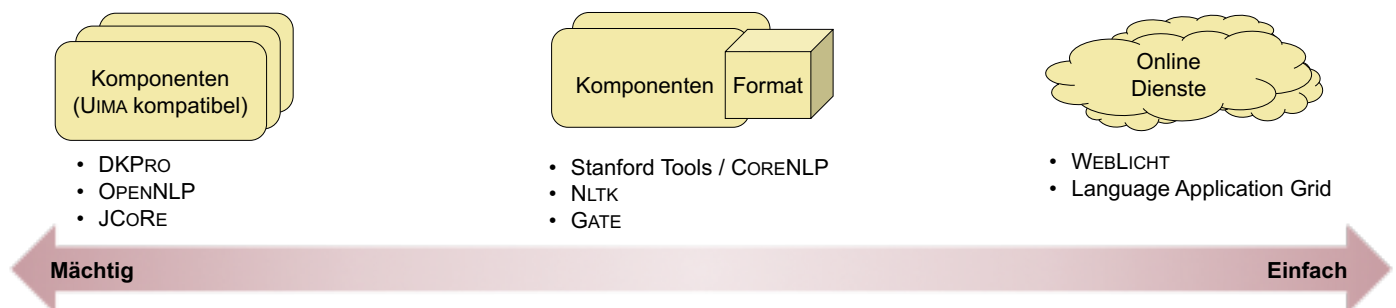
- Pipeline Builder
- Maven Central

Langjährige Erfahrung

- UIMA Workshops auf GSCL 2009 & LREC 2008
- UIMA Tutorial auf SMBM 2008
- IBM UIMA Innovation Award 2007 & 2008
- Publikation zum Typensystem 2007
- Publikation zur Modernisierung 2016

<http://julielab.github.io>

Computerlinguistische Repositorien & Frameworks



Literatur

- Hahn, Udo / Buyko, Ekaterina / Tomanek, Katrin / Piao, Scott / McNaught, John / Tsuruoka, Yoshimasa / Ananiadou, Sophia (2007): "An annotation type system for a data-driven NLP pipeline", in: *LAW 2007 – Proceedings of the Linguistic Annotation Workshop @ ACL 2007* 33-40.
- Hahn, Udo / Matthies, Franz / Faessler, Erik / Hellrich, Johannes (2016): "UIMA-based JCoRe 2.0 goes GitHub and Maven Central: State-of-the-art software resource engineering and distribution of NLP pipelines", in: *LREC 2016 – Proceedings of the 10th International Conference on Language Resources and Evaluation* 2502-2509.

Danksagung

Johannes Hellrich wird im Rahmen des Graduiertenkollegs „Modell Romantik. Variation – Reichweite – Aktualität“ (GRK 2041/1) von der Deutschen Forschungsgemeinschaft (DFG) gefördert.

Friedrich-Schiller-Universität Jena  Deutsche Forschungsgemeinschaft