Multimodal Docker Unified UIMA Interface: New Horizons for Distributed Microservice-Oriented Processing of Corpora using UIMA

Daniel Bundan Giuseppe Abrami Alexander Mehler

Goethe University Frankfurt | Text Technology Lab

Konvens 2025

KONVENS'25
10-12 September 2025
Hildesheim, Germany

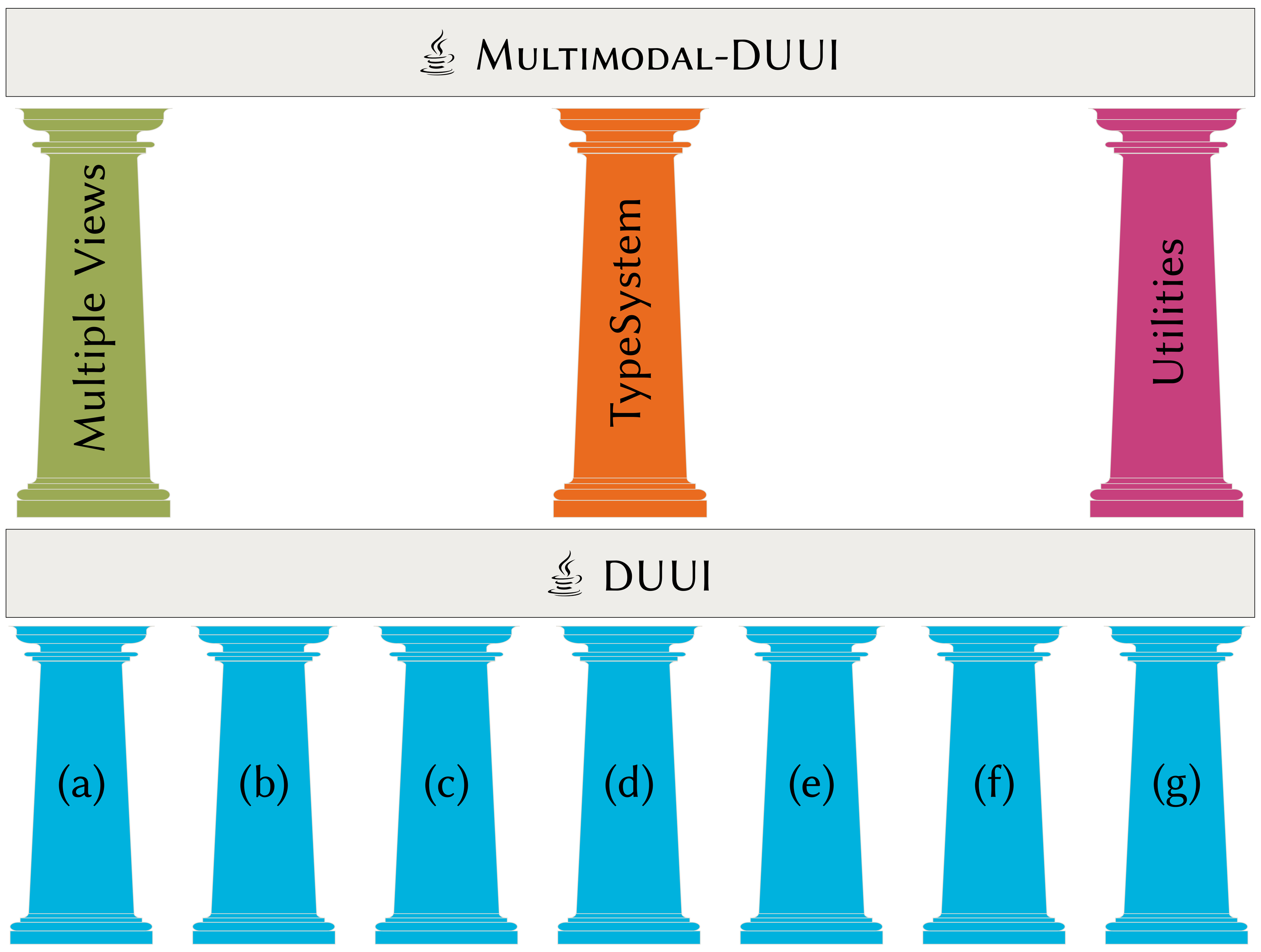
Motivation

- Collecting and processing **big corpus data** is a significant systematic challenge.
- Particularly for **multimodal corpora** including videos, audio, images and a mixture of them.
- Only a **few systems** in the field of NLP can process multimodal data automatically.
- **Individual approaches** lack integration & reusability (c.f. Abrami et al. 2025).
- Docker Unified UIMA Interface (**DUUI** – Leonhardt et al. 2023) provides parallel, scaled and unified processing of large corpora, **but only for textual ones**.

Solution

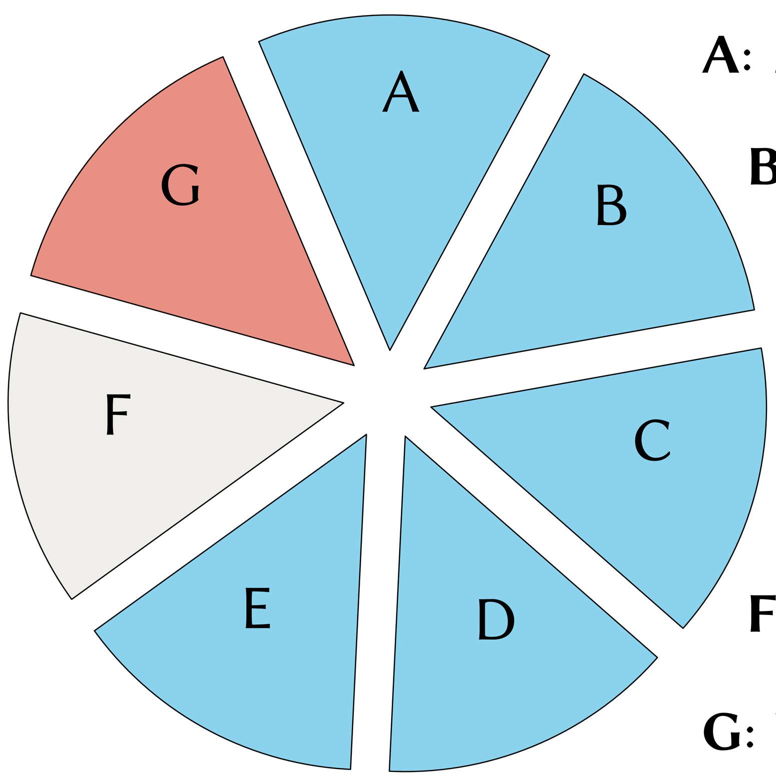
Substantial **upgrade** of DUUI to enable multimodal utilization capabilities, resulting in **Multimodal-DUUI**.

Components



- Multiple Views
  - Usage of UIMA views to serialize different modalities.
  - UIMA views can be used as a source as well as a target for DUUI-Components.
  - Allows for seamless processing of multimodal data in a single pipeline.
- TypeSystem
  - Defines annotation schema for UIMA Documents.
  - New classes added to enable multimodality and interoperability between modalities.
  - Extensible framework for multimodal annotations.
- Utilities
  - `MultimodalUtil` class for extracting multimodal content from annotated UIMA documents.
  - Bridges between annotations and operative media data.
  - Can extract video and audio segments based on annotations.
  - Handles image cropping based on annotations.

Features

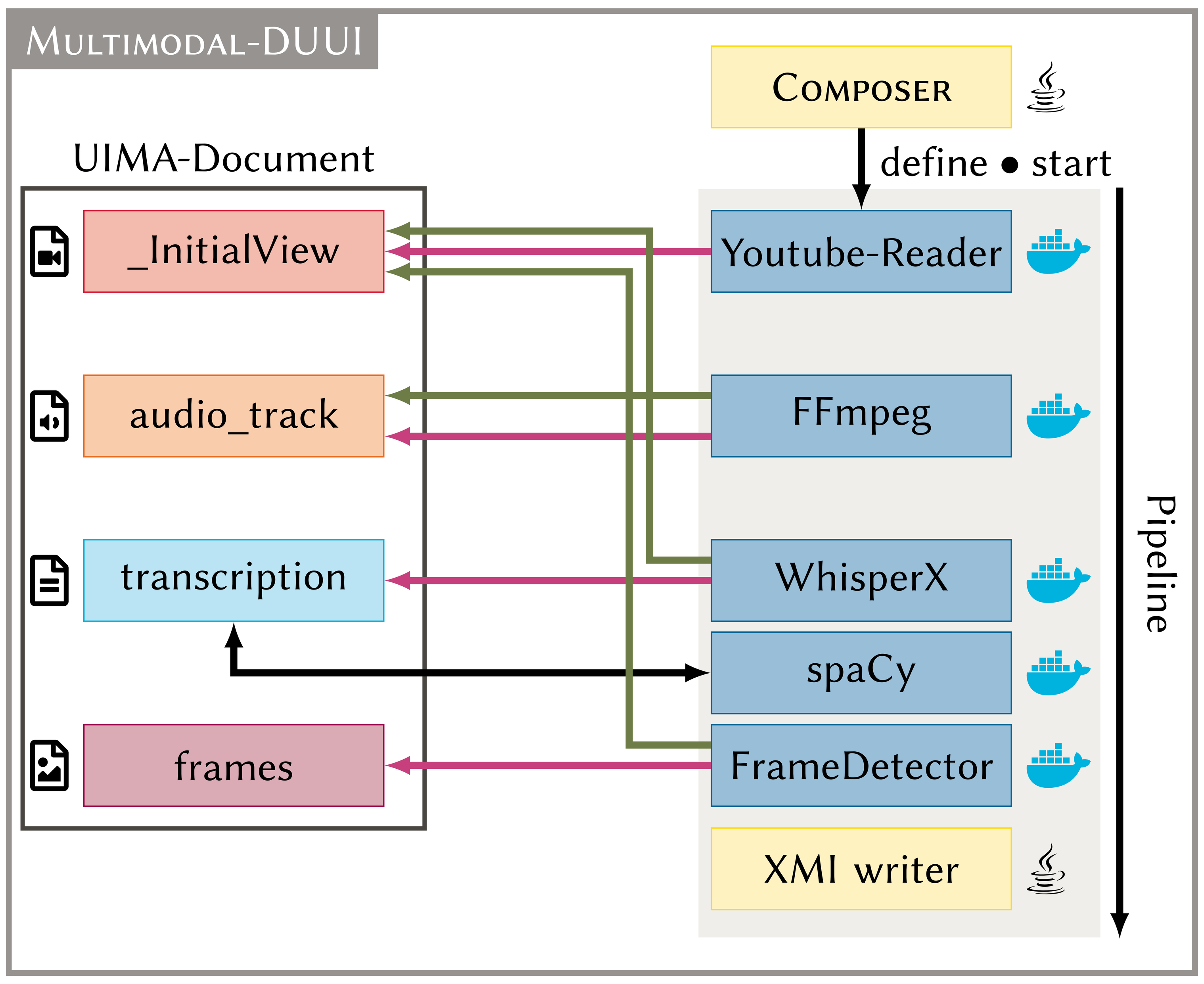


A: Multi-Layer-Annotation
B: Multimodality
C: Distributed processing
D: Schema-based
E: Reusability
F: Transposability
G: Web and API-Based

Resources

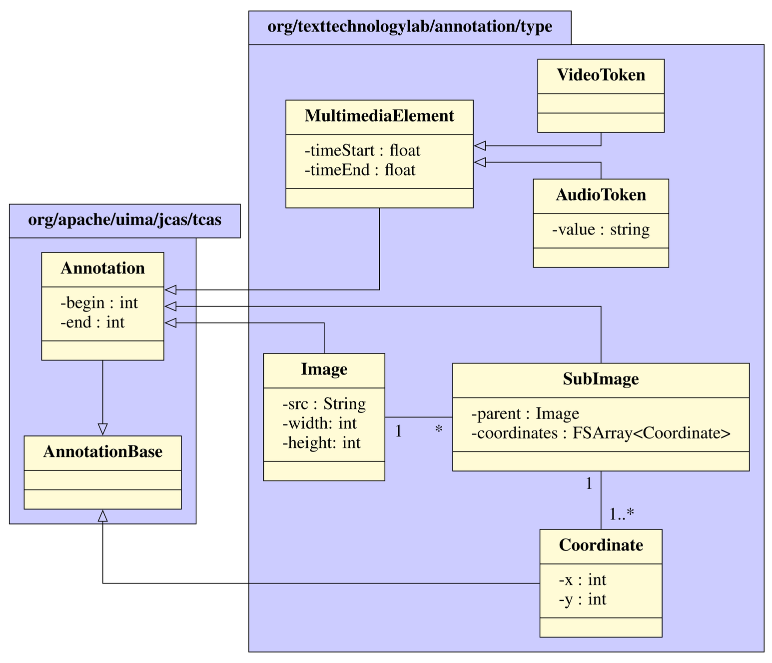
DUUI

Architecture & Pipeline



→ **Source View:** Specify from which view a Component reads from
← **Target View:** Specify from which view a Component writes to
+ **Auto-Creation:** Non-existent views are automatically created
↔ **Flexibility:** View assignments can be changed without code modification

UIMA TypeSystem



Use Cases

- **Social Media:** YouTube, TikTok, Instagram analysis
- **Surveys:** Collecting and processing multimodal interviews (e.g. Research Unit CORE 2023)
- **Parliament:** Speech transcription & analysis (e.g. Abrami, Bagci, and Mehler 2024a)
- **Multimodal AI:** Combined audio-visual processing

Performance Evaluation

- **Dataset:** 100 German YouTube videos (PleIAs 2024), avg. 12.15 minutes
- **Task:** Video transcription using WhisperX (large-v2)
- **Hardware:** Single Nvidia L40S GPU, single thread processing
- **Comparison:** SparkNLP (Kocaman and Talby 2021) vs. Multimodal-DUUI
- **Result:** SparkNLP (812,94s) • Multimodal-DUUI (1 100,18s)
- **Observation:**
  - **Overhead:** Base64 encoding adds processing time
  - **Trade-off:** Containerization overhead vs. dependency isolation
  - **Nativity:** SparkNLP is not capable of realizing multimodal processes.

Future Work

- **Component Expansion:** More multimodal models
- **Web Integration:** API-based application
- **UIMA Integration:** Propose as Apache extension

References

Abrami, Giuseppe, Mevlüt Bagci, and Alexander Mehler (May 2024a). "German Parliamentary Corpus (GerParCor) Reloaded." In: *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*. Ed. by Nicoletta Calzolari, Min-Yen Kan, Veronique Hoste, Alessandro Lenci, Sakriani Sakti, and Nianwen Xue. Torino, Italia: ELRA and ICCL, pp. 7707–7716.

Abrami, Giuseppe, Markos Genios, Filip Fitzermann, Daniel Baumartz, and Alexander Mehler (2025). "Docker Unified UIMA Interface: New perspectives for NLP on big data." In: *SoftwareX* 29, p. 102033.

Kocaman, Veysel and David Talby (2021). "Spark NLP: Natural language understanding at scale." In: *Software Impacts*, p. 100058.

Leonhardt, Alexander, Giuseppe Abrami, Daniel Baumartz, and Alexander Mehler (2023). "Unlocking the Heterogeneous Landscape of Big Data NLP with DUUI." In: *Findings of the Association for Computational Linguistics: EMNLP 2023*. Ed. by Houda Bouamor, Juan Pino, and Kalika Bali. Singapore: Association for Computational Linguistics, pp. 385–399.

PleIAs (2024). *YouTube-Commons*. dataset. Accessed: 2025-04-29.

Research Unit CORE (2023). *Critical Online Reasoning in Higher Education (CORE) (FOR 5404)*. https://core.uni-mainz.de. Accessed: 2024-07-10.

www.texttechnologylab.org

Text Technology

Gefördert durch DFG Deutsche Forschungsgemeinschaft

SPP 2431 NEW DATA SPACES

ENTAILab
539634240

CORE
5404
462702138