Querying Interaction Structure: Approaches to Overlap in Spoken Language Corpora

Elena Frick¹, Henrike Helmer¹ and Thomas Schmidt²

¹ IDS Mannheim (DE), ² University of Basel (CH)

BACKGROUND

Research Project „ZuMult“

ZuMult stands for German „Zu gänge zu multimodalen Korpora gesprochener Sprache“ – “Access to Multimodal Spoken Language Corpora”

Aim of the project: a backend software architecture for a unified access to spoken language resources (natural interaction) located in different repositories (incl. search functionality)

ZuMult-Blog at https://zumult.org/

MTAS-based Search Engine

Lucene indices

Search by Query

Results for searching (prolog-fragment): search for overlap between speakers

COP queries to the AGD corpora, using a Lucene index via MTAS

Search Web Services

RESTful Web Services implemented in JAX-RS

Risk Project “ZuMult”

CONTACT

Lucene

indices

More about the project:


APPROACH

Search by Query

Time intervals in which speaker “NH” is silent, but other speakers are speaking

Speaker Overlaps: segment-based vs. contribution-based approach

CONTACT

Lucene

indices

More about the project:
