

GRK 1994: Adaptive Preparation of Information from Heterogeneous Sources (AIPHES)

Funded by
DFG Deutsche
 Forschungsgemeinschaft

Technische Universität Darmstadt
 Universität Heidelberg
 Heidelberger Institut für Theoretische Studien

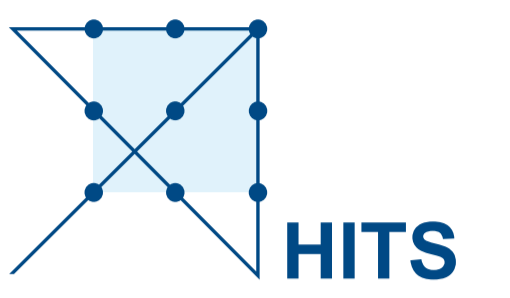


TECHNISCHE
 UNIVERSITÄT
 DARMSTADT



RUPRECHT-KARLS-
 UNIVERSITÄT
 HEIDELBERG

Heidelberger Institut für
 Theoretische Studien



Research Agenda

Computational Linguistics

Area A: Graph-based discourse-semantic processing for heterogeneous document sources

- A1 – Entity Linking/Cross-document Coreference Resolution
- A2 – Identification of complex event structures in discourse
- A3 – Opinion and Sentiment - extrapositional aspects of discourse

Area C: Representation and analysis of text-based structures

- C1 – Motif Analysis of Text-Based Graphs
- C2 – Methods for contextual and constraint-based ranking
- C3 – Deep Learning embeddings for adaptive language processing

Network Analysis and Machine Learning

Language Technology

Area B: Natural Language Processing for multi-document summarization

- B1 – Structured summaries of complex contents
- B2 – Content selection based on linked lexical resources
- B3 – Data-driven paraphrasing and harmonization of language style

Area D: Criteria and methods for quality assessment of heterogeneous sources and dossiers

- D1 – Multi-level models of information quality in online scenarios
- D2 – Manual and Automatic Quality Assessment of Summaries from Heterogeneous Sources

Information Management

AIPHES: Executive Summary

Motivation

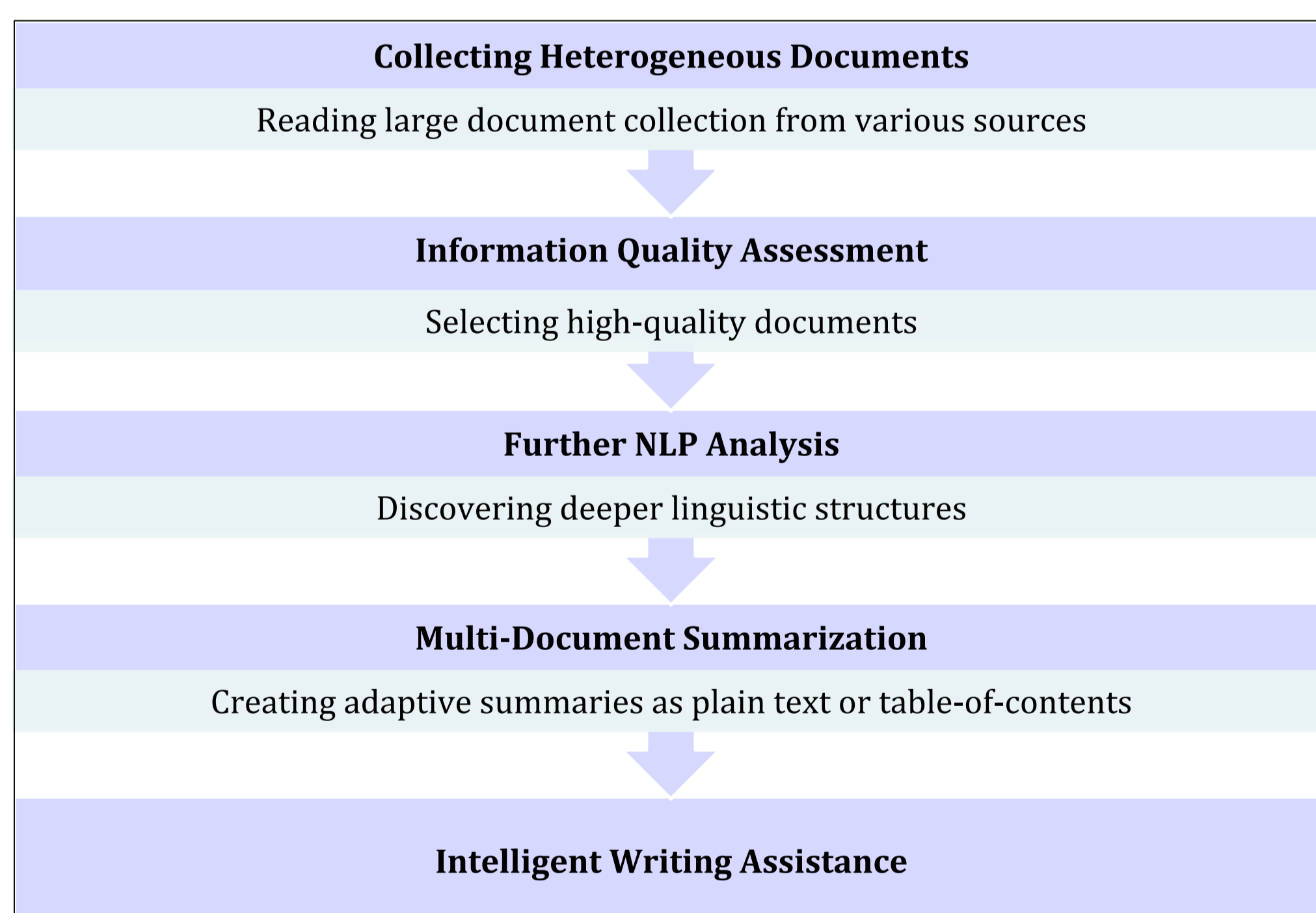
- Currently, the preparation and updating of multi-document summaries is done manually
- E.g. editorial departments e.g. Zeit Online (online version of the German weekly newspaper)

Topic

- Extracting structured knowledge from heterogeneous sources using automated methods in order to create an informative and stylistically homogeneous dossier
- Focus on multi-document summarization for German
- Adapting to different text types, thematic domains, user groups, languages

- Research Training Group, 04.2015 – 2019
- 11 PhD students, 8 PI's and 13 associated researchers
- Funding: about five million Euros, DFG

Intelligent Writing Assistance



Knowledge Transfer between Research and Online Editorial Journalists

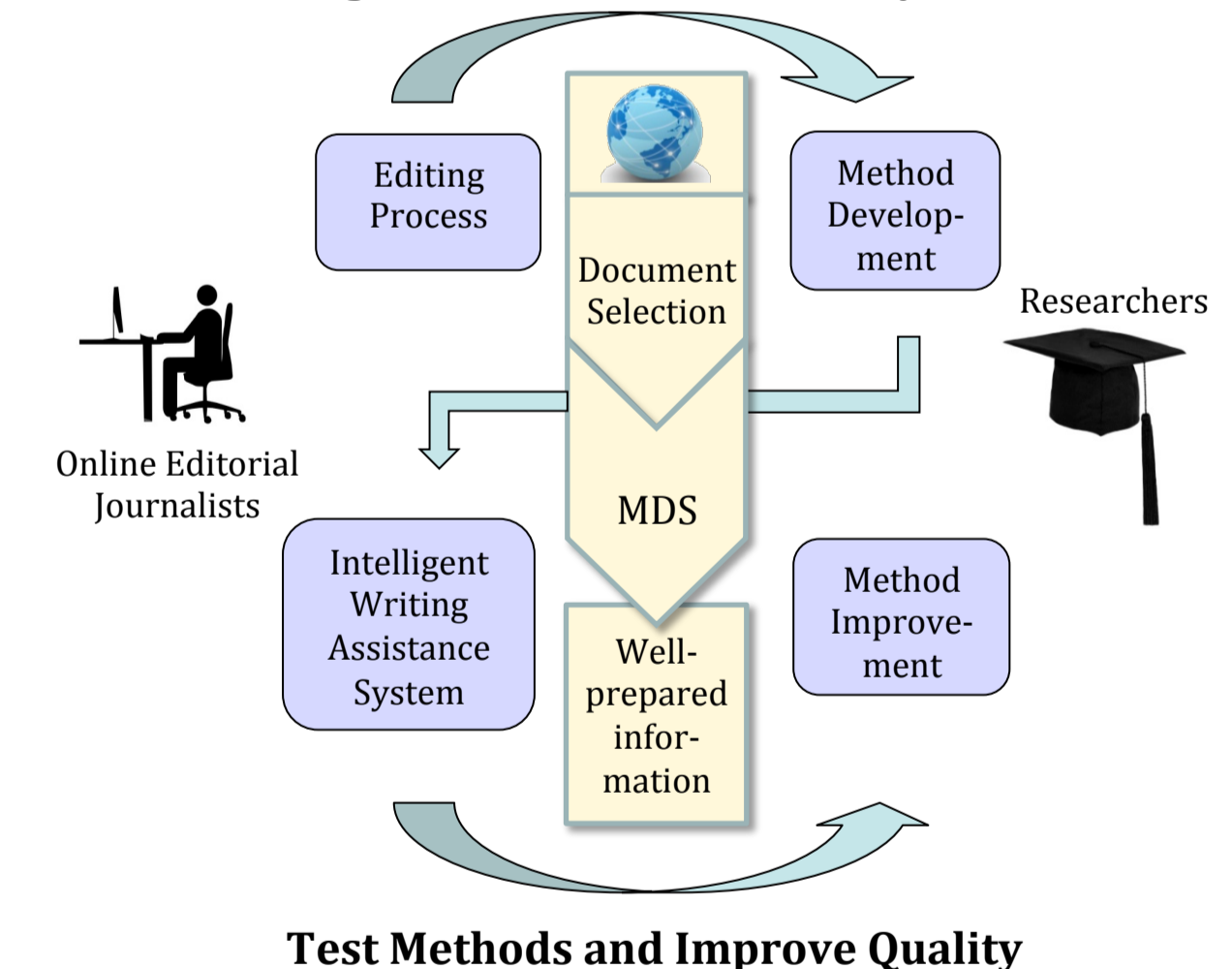
Research → Online Editorial Journalists

- Develop innovative methods for intelligent writing assistance in automatically summarized texts
- Automatically analyze manual interventions in intelligent writing assistance systems to improve methods
- Get feedback about the tools and methods from end users who work with them in practice

Online Editorial Journalists → Research

- Transfer knowledge about the editorial process
- Specify quality criteria and demands for source documents and multi-document summaries
- Test the developed methods and tools in practice
- Give feedback about and manually evaluate the automatic document summaries

Transfer Knowledge about Processes, Quality Criteria, Demands



Test Methods and Improve Quality

AIPHES for eHumanities: Expected Advancements

Discourse Analysis in Large Text Collections

Analysis of complex structures across large text collections, to identify relevant entities, events and extrapositional aspects in heterogeneous and potentially noisy texts.

Benefits the fields:

- Linguistics
- Education

Discovering Topics in Texts

Discovery of prominent topics in large and heterogeneous document collections with the help of graph- and ranking-based algorithms, and deeper linguistic structures.

Benefits the fields:

- History
- Psychology

Creating Textual Summaries

Creating summaries from large and heterogeneous document collections, based on an analysis of inherent structures and document relevancy, and data-driven standardization.

Benefits the fields:

- History
- Literature

Quality Assessment in Texts

Assessment of document quality in large and heterogeneous text collections with the help of multiple dimensions of quality, external resources and user interaction.

Benefits the fields:

- Education
- Literature

Partners

Future Management Group
 (Walluf)

Holtzbrinck Digital
 (München)

Hochschule Darmstadt,
 Fachbereich Media

Principle Investigators:

Prof. Dr. Chris Biemann (TU Darmstadt)
 Dr. Judith Eckle-Kohler (TU Darmstadt)
 Prof. Dr. Anette Frank (Universität Heidelberg)
 Prof. Johannes Fürnkranz (TU Darmstadt)

Prof. Dr. Iryna Gurevych (TU Darmstadt), Spokesperson
 Dr. Christoph Rensing (TU Darmstadt)
 Prof. Dr. Michael Strube (Heidelberger Institut für Theoretische Studien)
 Prof. Dr. Karsten Weihe (TU Darmstadt)

<http://www.aiphes.tu-darmstadt.de>