PhD Program Knowledge Discovery in Scientific Literature

Presenter: Erik-Lân Do Dinh, Spokesperson: Prof. Dr. Iryna Gurevych

http://www.kdsl.tu-darmstadt.de



label classification, argumentation mining, focused crawling, relation extraction, temporal network analysis and figurative language analysis. KDSL applies these methods to answer research questions related to the educational research. Thus, we enable new forms of intelligent information access for humanities researchers. We are part of the Research Initiative "Knowledge Discovery in the Web" at the TUDA. We also work with a LOEWE Research Center "Digital Humanities" (Univ. of Frankfurt, TU Darmstadt), the Institute for Computational Linguistics (Univ. of Heidelberg), and the BMBF-funded research center for Digital Humanities CEDIFOR (Univ. of Frankfurt, TU Darmstadt, German Institute for International Educational Research).

A researcher is interested in a research question that might be answered on basis of a data set X.

There are around 1000 publications containing work on data set X.

The system searches the Web for relevant publications, structures the results by research questions, variables, and methods, personalizes the results and visualizes research gaps.

Motivation & Goals

Future expert information for scientific communities relies on the ubiquitous availability of literature on the web. Intelligent services have their foundations in modern **computer science methods**, which need to be developed and extended to be valuable to the experts using them. Basic research on this is therefore a central goal of the program, laying foundations for digital libraries of the future.

Research Areas

- Crawling and Semantic Structuring of Scientific Publications in the Web
- Temporally Dynamic Networks of Topics and Authors in Scientific Publications • Scalable Multilabel Classification for Educational Research

Data & Corpora

TECHNISCHE

UNIVERSITÄT

DARMSTADT

Knowledge

Discovery in

Scientific

iterature

Our data collections include modern corpora in the field of educational research, such as FIS Bildung Literaturdatenbank and **peDOCS**, containing references and full text articles. We also employ historical educational corpora from scientific literature, like Natur und Staat (1903 - 1911).

Research Program

The main topic of the PhD program is knowledge discovery in the vast amount of scientific literature ubiquitously available on the Web. This research employs methods of intelligent identification and analysis of structures in scientific texts on all scales, enabling completely new and unforeseen forms of access to scientific information.

The underlying data is represented by full texts of publications together with their manifold connections via citations, authors and the linguistic and semantic knowledge extracted from the texts.

Our goals require intensive research on dataund text-mining methods and their application to unstructured scientific information and historical corpora. The methods are for to querying, indexing, example applied discovering and preparing knowledge on the web. Implicit knowledge contained in scientific literature is thus tapped and rendered usable.

- Knowledge Extraction and Consolidation for Scientific Publications in the **Educational Domain**
- Domain-Adaptive Text Mining to Support Knowledge Discovery in Scientific **Historical Literature**
- When "Evolution" leads the Way: Argumentation, Constructions and Metaphors in German Popular Biopolitics on the Eve of the 20th Century.

Semantic

Analysis

Personalized

Processing

• Metaphor in Use: Analysing Figurative Speech in the Literary and Scientific Discourse of the 20th Century

Common Methods

- Corpus based work and annotation (Núñez, Gerloff)
- Multilabel classification (Nam)
- Focused crawling and relation classification (Remus)
- Co-occurrence analysis (Ma)
- Metaphor detection (Do Dinh)
- Argumentation analysis (Kirschner)

Supervisors and Associated Researchers

Dr. Sabine Bartsch (Literary Studies, English Linguistics) **Prof. Dr. Chris Biemann** (Computer Science, Language Technology) **Dr. Judith Eckle-Kohler** (Computer Science, Ubiquitous Knowledge Processing) **Prof. Dr. Johannes Fürnkranz** (Computer Science, Knowledge Engineering) **Prof. Dr. Petra Gehring** (Philosophy, Language and Technical Philosophy) **Prof. Dr. Iryna Gurevych** (Computer Science, Ubiquitous Knowledge Processing) **Prof. Dr. Nina Janich** (Literary Studies, German Linguistics) **Prof. Dr. Andrea Rapp** (Literary Studies, Digital Philology) **Prof. Dr. Karsten Weihe** (Computer Science, Algorithmics)





UBIQUITOUS **KNOWLEDGE** PROCESSING

The research program "Knowledge Discovery in Scientific Literature" focuses on the educational research as the target domain. To this end, we use the data collected at the German Institute for Educational Research (DIPF) and TUDA. We closely cooperate with users of the developed innovative the technology and humanities researchers.

PhD Program

The supervision of young researchers in the PhD program strongly relies on close contacts between members of the program, regular joint meetings, co-supervision from multiple disciplines and a lively exchange in the research and qualification program "Language and Knowledge Engineering".

The program strives to publish at leading scientific conferences and provides its software as open source products on the basis of the DKPro framework.







Language and **Technical Philosophy Digital Philology**

This work has been supported by the Deutsches Institut für Internationale Pädagogische Forschung Technische Universität Darmstadt, Fachbereich Informatik.

Project duration: 2013 until 2016

Funding volume: 352 TEUR

http://www.kdsl.informatik.tu-darmstadt.de





und Bildungsinformatio