CEDIFOR

Centre for the Digital Foundation of Research in the Humanities, Social, and Educational Sciences

CEDIFOR Board Members at TU Darmstadt

Prof. Dr. Petra Gehring (Institut für Philosophie)
Prof. Dr. Iryna Gurevych (UKP Lab)
Prof. Dr. Andrea Rapp (Institut für Sprach- und Literaturwissenschaft)
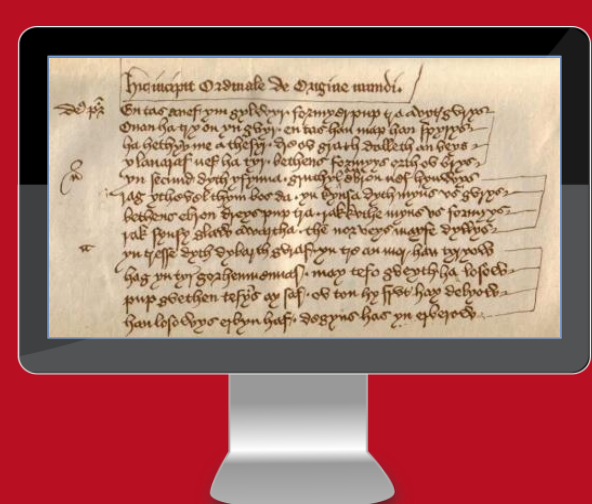Dr. Johannes Daxenberger (UKP Lab)

UBIQUITOUS KNOWLEDGE PROCESSING

TECHNISCHE UNIVERSITÄT DARMSTADT

CEDIFOR: Vision

- We serve as a **primary contact point for eHumanities projects** in universities, research institutes, archives, libraries, and museums in the Rhein-Main area and beyond
- We cover research projects in the **humanities, social sciences, and educational science**
- Our results are **sustainable** and our services are **long-term**
- We work with **text, image, video, and audio**
- We work and publish **nationally and internationally**
- We base our work on the results of the **LOEWE Research Cluster "Digital Humanities"** (2011-2014)

Research Innovations

Epistemological Innovation

Consulting and Supervising <u>Novel Research Concepts</u>

- Integration of humanities with computer science for mutual benefit
- Discovery, elaboration, and realization of research projects

Developing <u>Tools</u> for Multimedia Research Subjects

- Integration of multi-code, multimodal, and multimedia research subjects
- Creation of highly structured corpora: internally linked and stratified across time

Technological Innovation

Infrastructural Innovation

Creating a Research Infrastructure for the <u>Common Usage of Tools and Data</u>

- Enable shared usage of research subjects, data, and tools
- Create sustainable results, workflows, and best practices

eHumanities: The Current State

Research Environments
- **Annotation**: <u>CSniper</u>, <u>WebAnno</u>, <u>DiGAT</u>
- **Processing**: NLTK, GATE, UIMA, <u>DKPro</u>, eHumanities Desktop
- **Discovery**: Semantic CorA, <u>Corpus Explorer</u>

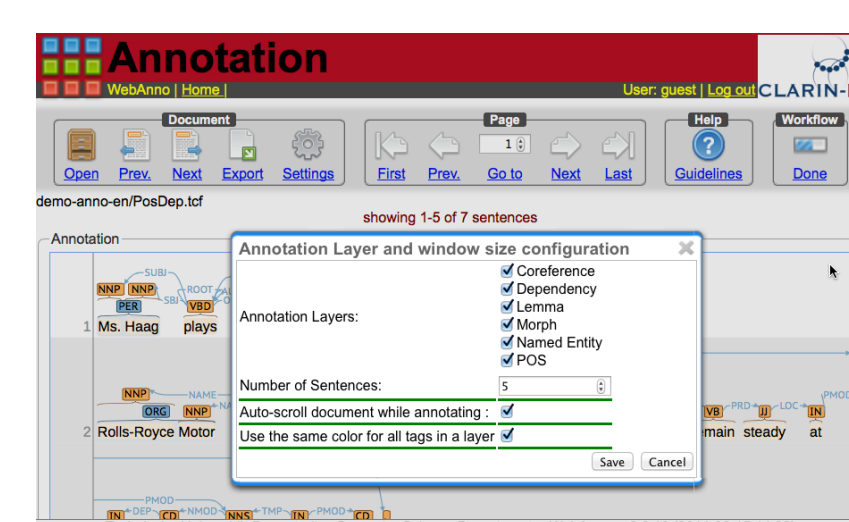**Assignment and Linking of (Meta-)data**
- Based on the **Lexical Markup Framework**: <u>UBY</u>
- Based on the **Semantic Web** standard: OWL, RDF

Development and Analysis of Multimodal/Multimedia Data
- **Image**: ImageDB
- **Video/Audio**: ELAN

Infrastructure
- DARIAH-DE
- CLARIN-D

CLARIN-D
DARIAH-DE

Challenges

- Researchers in humanities have difficulties to manage the coordination, interpretation, and linking of **growing amounts of metadata**
- Lack of accessible tools for the analysis of **multimedia data** (image, video and audio)
- **Error rates** in applied Natural Language Processing methods are too high in practice

Tools developed by or in collaboration with UKP Lab are <u>highlighted</u>.

Pilot Projects Carried Out by TU Darmstadt

2015

Inferences Between Linguistic Attributes and Higher Level Features

- Visualization of correlations between linguistic attributes and **personality traits**
- Example traits: performance motivation, emotional stability, locus of control
- Use case: **analysis of student texts**

This is a cooperation project with Dr. Sonja Rohrmann, GU Frankfurt

Computer-Supported Discovery of Metaphors

- Context-based discovery of metaphors in **large corpora**
- Computer-supported **visualization and processing** of the data
- Theory and content-based **classification of metaphors**

This is a cooperation project with Dr. Andrea Rapp and Dr. Petra Gehring, TU Darmstadt

Automatically Analyzing and Extending the Structure of Links in EGOlink

http://ieg-ego.eu

Goal: Reduce human effort to enrich EGO articles with internal and external links by developing (semi-)automatic methods for linkification
1. Automatic identification and ranking of potential anchors in EGO articles
2. Linking the identified anchors to their counterparts in other EGO articles and/or external resources
3. User study to assess the success of the project

This is a cooperation project with Dr. Michael Piotrowski, IEG Mainz

2016

Automatically Extracting Audiovisual Features to Analyze Videos in Educational Research
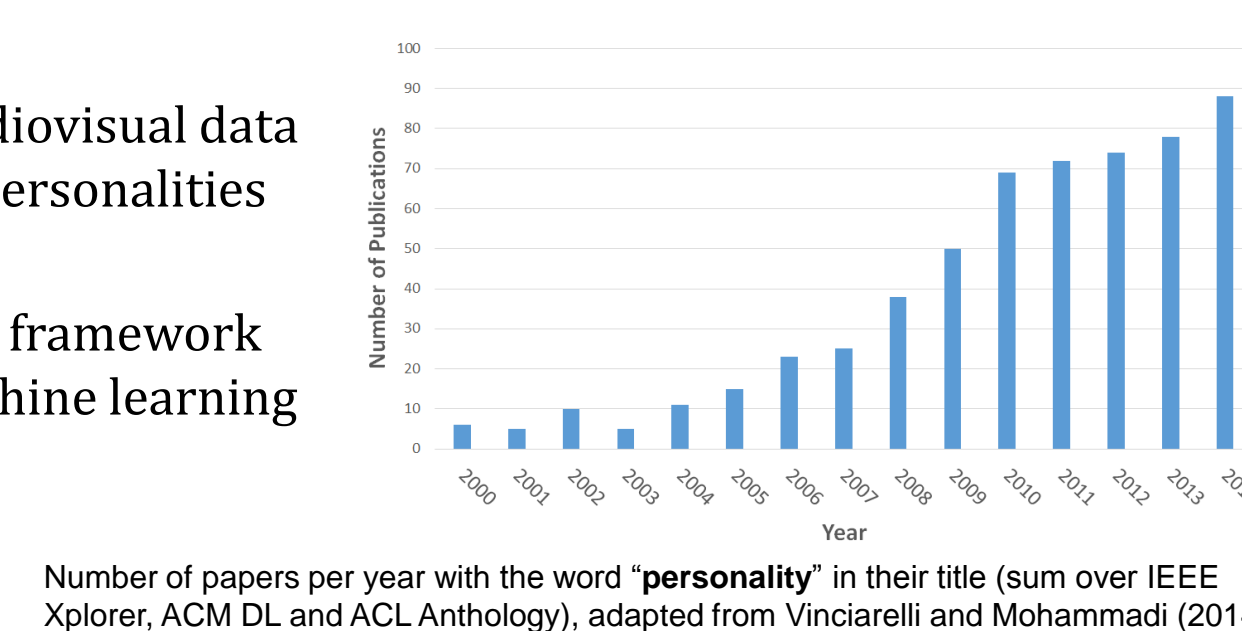
Motivation & Goals
- Reduce manual pre-processing effort for qualitative analysis of audiovisual data
- Reduce the bias introduced by human judgements about student personalities

Methods
- Implementing feature extractors for audiovisual data in the DKPro framework
- Automatic personality perception of students with the help of machine learning

Data
- Classroom video recordings of students in a physics class
- Expert annotations about the personality of students

Number of papers per year with the word "**personality**" in their title (sum over IEEE Xplorer, ACM DL and ACL Anthology), adapted from Vinciarelli and Mohammadi (2014)

Collaboration with:
Dr. Susanne Görke, *Altorientalische Philologie, Johannes Gutenberg Universität Mainz*

Building a Framework for Semantic Search on Translated Cuneiform Scripts
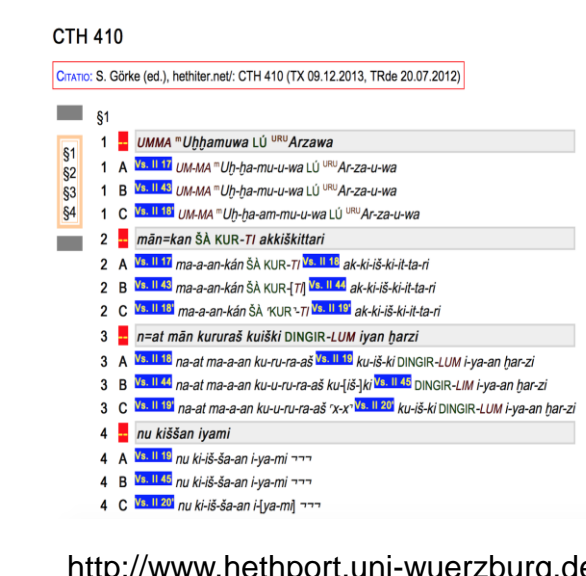
Motivation & Goals
- Enable semantic search in translations of cuneiform documents
- Make the translations more accessible by adding another layer of information

Methods
- Sophisticated lexical-semantic methods based on modern NLP technology
- Text mining algorithms for a deeper exploration of the documents

Data
- Originals and German and English translations of Hittite documents
- Akkadian cuneiform documents and their translations

http://www.hethport.uni-wuerzburg.de

Collaboration with:
Caroline Verena Wahle, Graduiertenkolleg Unterrichtsprozesse, Universität Koblenz-Landau

CEDIFOR: Collaboration Strategy

Subject Areas

CEDIFOR aims to establish sustainable partnerships with a wide range of disciplines.
Existing collaborations include:
- *Learning Sciences*
- *Education and Educational Psychology*
- *Psychology and Psychotherapy*
- *Ancient Eastern Philology*
- *German Philology*
- *European history*

CEDIFOR has a special focus on **Digital Libraries projects**. Our expertise covers:
- Enriching digital sources with the help of automatic annotation and conversion
- Applying state-of-the-art Natural Language Processing techniques to make digital collection more accessible
- Ensuring long-term availability of digital collections by means of sustainable workflows

Digital Libraries

Open Content

We **emphasize open content,** both for the resources we are working with, as well as for the resources we are producing. This is inspired by the following observations:
- The amount of public data, extracted e.g. from social media or online learning platforms, is constantly growing
- Open Content creates higher visibility
- Open Content ensures reusability and long-term availability

Partner Institutions

Rhein-Main
- Leibniz-Institut für Europäische Geschichte Mainz
- Universität Gießen
- Hochschule für Musik und Darstellende Kunst Frankfurt
- Universitätsbibliothek Frankfurt
- Deutsche Nationalbibliothek Frankfurt
- Städel Museum Frankfurt
- Freies Deutsches Hochstift / Frankfurter Goethe-Museum
- MPI für Europäische Rechtsgeschichte Frankfurt
- ULB Darmstadt

National and International
- Monumenta Germaniae Historica München
- Bibliothek für Bildungsgeschichtliche Forschung Berlin
- Universität Hildesheim
- Universität Bamberg
- Max Planck Institut Nijmegen
- Harvard University
- Centre for the Study of Manuscript Cultures Hamburg
- Europeana Den Haag

http://www.cedifor.de

References
- Vinciarelli, A., & Mohammadi, G. (2014). A Survey of Personality Computing. IEEE Transaction on Affective Computing, 5(3), 273 – 291.
- http://commons.wikimedia.org/wiki/File:Origo_Mundi_kynsa_gwersow.jpg

Contact
Dr. Johannes Daxenberger, daxenberger@ukp.informatik.tu-darmstadt.de