

Text Mining for Historical Documents /  
Historische Quellen mit (computer-)linguistischen  
Methoden analysieren  
Group Work

Caroline Sporleder and Martin Schreiber

Computational Linguistics &  
Kultur- und Mediengeschichte  
Universität des Saarlandes

Wintersemester 2008/09

16.02.2009

## Aim:

- develop an outline of a concrete interdisciplinary project
- project doesn't have to be put into practice but the sketch of the project should be detailed and draw on expertise from CoLi and History

## What to do?

- form a group (2-3 history students, 2-3 coli students)
- choose a collection of data
- have a close look at the data
- come up with interesting historical questions the collection might answer
- think about how computational linguistics could help
  - ⇒ how could **access** to the collection be improved to help with answering historical research questions (e.g. automatic creation of meta-data, intelligent search methods, automatic creation of domain ontologies or social network graphs)
- develop a clear outline of how you would carry out the project
- present your project at the end of next week

## How should you do it?

- talk a lot with to each other
- have a very close look at the data
- be creative and use your expertise
- do some additional (internet) research (e.g., find related collections which might be useful for your research question, find useful external resources, read up on a specific NLP technique you might want to use)

## What to address

- what is the collection you chose about?
- what are the historical research questions you want to answer?
- why are these questions relevant/interesting etc?
- are there any subquestion you have to answer first?
- what kind of NLP processing do you have to do to help with access to the collection?
- inhowfar would this help the historians?
- what are the exact processing steps?
- which problems need to be solved?
- what tools are you going to use? / Will you use additional resources (which, how, why) or annotate data etc. ?
- what are your fallback options if something doesn't work?

# Practical Project for Coli Students

- Coli students have to do a small practical project (e.g., solve a sub-problem of the project)
- this can be done alone or together with the other coli student(s) in your group
- you then have to write a short report about this