Text Mining for Historical Documents

Caroline Sporleder

Computational Linguistics
Universität des Saarlandes

Wintersemester 2009/10

05.03.2010

Information about the Report

Introduction

- what is your project about?
- why is it useful (historic context)?
- what sub-steps did you have to do? what problems to solve (e.g. error-correction, NER etc.)?
- how does it relate to previous research (make a connection to the theoretical part of the seminar)?

Main Part

Detailed description of the subtasks that your system performs:

- how does it tackle these subtasks?
- why did you choose this approach?
- did you encounter specific problems that you had to solve?
- how did you solve them?

Content (3)

Evaluation

of sub-components and system overall (insofar as it makes sense)

- either quantitative (e.g., accuracy for information extraction task)
- and/or qualitative (e.g., give examples where your system improves IR over keyword-based search)
- also discuss cases where your system doesn't perform well and describe how it could be improved (error analysis)

Conclusion

- summary: in how far where the objectives for the system achieved (i.e., how well does it work and in how far might it be useful for Historians)
- outlook: what else could be done?

Formalities

- report length 20-50 pages
- include at least some references
- provide examples!
- include appendix where you indicate who did what in the group project
- submit code, should be documented and executable
- Due date: April 1, 2010