

Thesis  
B.Sc.

Thesis  
M.Sc.

# Analysis of and Mining in Mailing Lists

## Motivation

In previous work, we created a database that contains the operation data of the Internet Engineering Task Force (IETF) – like mailinglist posts, document histories, and meeting attendees. The IETF is standardizing Internet technology. The data contain the mailing lists that are used to discuss and achieve consensus, but the data is not limited to that. Until now, we have looked into some time series aspects and the problem of identifying individuals and their many representations in the data. Continuation may be concerned with refinement of existing work or also utilization of methods that we have not yet used. Examples are Sentiment Analysis, Data Mining, Statistics and Time Series Analysis. A clear focus should be defined on individual basis.



Goal of this research is to gain better understanding of communication used in the standardization process in the IETF. We want to better understand who is important and who ends discussions about certain topics (e.g. identify lead users and further stereotypes). We also want to characterize the different areas of Internet standardization. E.g. we would expect security standardization to be have a more aggressive tone than routing standardization. We would also like to improve the tools for better data exploration. Finding the right questions is an important part of thesis and multiple questions are to be tackled.

## Your Task

- Survey related work and tools that you might want to use
- Familiarize with the existing database
- Define (in exchange with advisor) research hypotheses
- Consider adding new data
- Extract inputs and outputs as well as to generate factors
- Write scripts and code to analyze the data
- Proof/disproof your hypothesis with the data and tools
- Thesis writeup

## Contact

Dr. Heiko Niedermayer [niedermayer@net.in.tum.de](mailto:niedermayer@net.in.tum.de)  
Daniel Raumer [raumer@net.in.tum.de](mailto:raumer@net.in.tum.de)

