Statistical Natural Language Processing

Organizational Information

Henning Wachsmuth

`https://ai.uni-hannover.de`

Organizational Information

Course

- Lectures. Henning Wachsmuth
- Tutorials. Gabriella Skitalinska
- Language. English

Information

- Web. https://www.ai.uni-hannover.de/en/teaching/courses/snlp
- Stud.IP. https://studip.uni-hannover.de/dispatch.php/course/details?sem_id=649188504c46a121538d68ebca9f78cd

Time and location

- Lectures. Thursday 11:00–12:30, as of October 12, Appelstr. 11, A145
- Tutorials. Wednesday 13:15–14:45, as of October 18, Appelstr. 9A, MZ2
  Notice: The lab "Human Language Technology" is mostly aligned with this course.

Consultation?

- Set up appointment via e-mail: h.wachsmuth@ai.uni-hannover.de

Organizational Information
This Course

Overall goals

- Learn major skills needed to tackle typical natural language processing (NLP) tasks.
- Get to know the main NLP techniques used nowadays.

Contents

- Some basics of data science, linguistics, and foundations of NLP
- Several statistical NLP techniques, mainly based on machine learning
- Several NLP tasks and applications

Competences

- Understanding of theory and practice of NLP
- Development of NLP methods for given tasks
- Scientific evaluation of NLP methods

Organizational Information
Basics the Course Builds upon

Required basics

- Models and algorithms. Concepts and methods from bachelor studies
- Languages. Understanding of natural and formal languages
- Math. Basic probability theory and linear algebra
- Programming (lab only). Some experience with software development

Covered basics

- Linguistics. Fundamental language concepts and phenomena
- Statistics. Selected concepts and methods from data science
- Machine learning. Fundamental learning concepts and methods
- Programming (lab only). Implementation in Python

Organizational Information
Outline of the Course

Introduction

1. Overview
2. Basics of Data Science
3. Basics of Natural Language Processing

Techniques

4. Representation learning
5. NLP using Clustering
6. NLP using Classification and Regression
7. NLP using Sequence Labeling
8. NLP using Neural Networks
9. NLP using Transfomers

Application

10. Practical Issues

Organizational Information
Course Elements

Teaching

- Lectures. Presentation of course content and organizational info
- Tutorials. Presentation of assignments and solutions, Q&A

Assignment sheets (details in first tutorial)

- Amount. 6 in total, bi-weekly (all pencil-and-paper)
  First sheet published on Oct 16; to be submitted by Oct 30, 23:59 (UTC+1)
- Group work. You need to submit with 3–4 people
- Bonus. (a) Min. 60% of all points: exam grade + 1/3, (b) Min. 80%: + 2/3
  Example for (b): grade of 2.7 is changed to 2.0; only grades < 5.0 can be improved.

Exams (details in some weeks)

- Oral, $\sim$30 minutes, questions on all lecture parts, in English
- Registration. November 15–30, 2023
- First exam dates tentatively in first half of February 2024
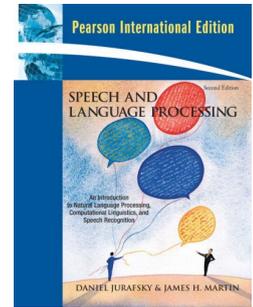 A list of example questions will be provided early enough.

Organizational Information
Textbooks (Not Obligatory)

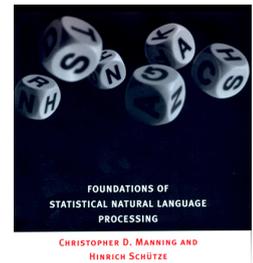**Speech and Language Processing**

(Jurafsky and Martin, 2009)

- Oriented towards computational linguistics
- Comprehensive
- Draft of upcoming third edition freely available

Foundations of Statistical Natural Language Processing
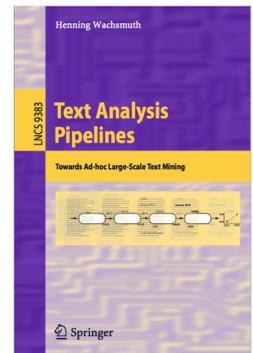
(Manning and Schütze, 1999)

- More oriented towards computer science
- Comprehensive, a bit outdated

Text Analysis Pipelines

(Wachsmuth, 2015)

- Rather oriented towards computer science
- Focused on NLP processes
- Book preprint freely available

References

- **Jurafsky and Martin (2009).** Daniel Jurafsky and James H. Martin. Speech and Language Processing: An Introduction to Natural Language Processing, Speech Recognition, and Computational Linguistics. 2nd edition, Prentice-Hall, 2009.
  Free draft of 3rd edition: https://web.stanford.edu/ jurafsky/slp3/

- **Manning and Schütze (2009).** Christopher D. Manning and Hinrich Schütze. Foundations of Statistical Natural Language Processing. MIT Press, 1999.

- **Wachsmuth (2015).** Henning Wachsmuth: Text Analysis Pipelines — Towards Ad-hoc Large-scale Text Mining. LNCS 9383, Springer, 2015.
  Free preprint: https://webis.de/downloads/publications/papers/wachsmuth_2015b.pdf