

Repetitorium der Computerlinguistik

Marina Sedinkina

Ludwig Maximilian University of Munich
Center for Information and Language Processing

29.April 2019

- 1 Übersicht
- 2 Themen
- 3 Voraussetzung für Bestehen des Kurses

- vor 10.06: Themen, die für die Anfertigung der Bachelorarbeit nützlich sein sollen.
- nach 10.06: Nützliches für nach der Bachelorarbeit
- Dozent: M.Sc. Marina Sedinkina (sedinkina@cis.uni-muenchen.de)
- <http://cis.lmu.de/~sedinkina/>

- Wie kann ich eine Forschungsfrage formulieren?
- Wie kann ich relevante Literatur finden?
- Welche Tipps gibt es für den Schreibprozess?
- Wie kann meine Arbeit gegliedert werden?
- Welche Bewertungsmaßstäbe werden normalerweise an eine Bachelorarbeit angelegt?

Schreibzentrum: <https://www.schreibzentrum.fak13.uni-muenchen.de/schreibberatung/index.html>

- *Offene Sprechstunde:* Montag und Mittwoch, 14 bis 15 Uhr: Theresienstr. 39, Raum B037.
- *Individuelle Termine*
- *Freitagsworkshops:* <https://www.schreibzentrum.fak13.uni-muenchen.de/freitagsworkshops/index.html>
- *Schreibtag:* 11. Mai 2019: Bachelor-, Master- und Zulassungsarbeiten, aber auch Tipps rund um Motivation und Schreibflow, individuelle Schreibberatung. Anmeldung auf der Website.
- Im Verlauf des Semesters veröffentlicht das Schreibzentrum auch *neue Termine* für Workshops zum journalistischen und kreativen Schreiben.

Thema: Aufbau von Machine Learning Experimenten

- Was ist eine geeignete Repräsentation meiner Daten für das maschinelle Lernen?
- Welche Metriken gibt es, um die Performanz eines Verfahrens zu beurteilen?
- Warum ist es notwendig, die Daten in Trainings-, Entwicklungs- und Evaluations-Sets einzuteilen?
- Wie kann ich herausfinden, ob mein Modell
 - die Trainingsdaten zu genau modelliert und nicht generalisiert? (overfitting)
 - nicht ausdrucksstark genug ist, um die wesentlichen Merkmale zu modellieren? (underfitting)

- Wie kann ich optisch ansprechend und konsistent
- ...formatieren?
- ...untergliedern?
- ...Literaturreferenzen verwalten?
- ...Abbildungen und Tabellen einfügen?

- Welche Prinzipien sollte man bei der Visualisierung von Information (Daten, Auswertung von Experimenten) beachten?
- Erstellung von Diagrammen mit matplotlib
- Linien-, Balken-, Kreisdiagramm etc.
- Modifizierungen: Achsenbeschriftung, Farbe, Dicke etc
- Visualisierung von Wortvektoren

- Wie kann ich mit Beamer
- ...Folien erstellen?
- ...Abbildungen und Tabellen einfügen?
- ...Design der Präsentation anpassen?
- ...Animationen verwenden?

- die Firmen einladen, die die Arbeit von Computerlinguisten in Unternehmen kurz vorstellen könnten.
- eine gute Möglichkeit:
 - die Firmen kennenzulernen, die sich mit Sprachtechnologien bzw. Computerlinguistik beschäftigen
 - einen Überblick verschaffen, an welchen Aufgaben und Projekten der Computerlinguist arbeiten kann.

- Mehrere Firmen haben sich bereit erklärt, die Arbeit von Computerlinguisten in Unternehmen kurz vorzustellen.
 - **HolidayCheck/Burda**
 - **Capgemini**
 - **Siemens**
 - **Sulzer GmbH**
 - vielleicht: **Volkswagen AG**
- letztes Jahr:
<https://marinapollo.github.io/rep18.github.io/>

- Für eines der 5 Themen/Sitzungen muss **eine** Kurzzusammenfassung geschrieben werden
- 1 Thema = 1 Sitzung
- Umfang ca. 1-2 Seiten
- Bitte an **sedinkina@cis.uni-muenchen.de** schicken
- sich für die Prüfung anmelden!!!