



# Maschinelle Sprachverarbeitung

## Übung

Einführung

Mario Sängner

[mario.saenger@informatik.hu-berlin.de](mailto:mario.saenger@informatik.hu-berlin.de)

# Überblick

---

- Organisatorisches
- 1. Übungsaufgabe
- Gruppenbildung

# Organisation

---

- Über das Semester verteilt gibt es sechs Übungsaufgaben
  - Aufgaben sind sehr implementierungsintensiv
  - Erstellung von Programmen "*from scratch*" oder Nutzung von bestehenden Bibliotheken
  - Umsetzung mit Hilfe von Java (oder zum Teil Python)
- Lösungen sind in zweier Teams anzufertigen
  - Jedes Team muss 5 der 6 Aufgaben lösen
  - Fester Abgabetermin je Übungsaufgabe
  - Abgabe der Lösungen erfolgt über HU-Box
- Zur Lösung der Aufgaben sind 2-3 Wochen Zeit

# Ablauf einer Übung

---

- Vorstellung der Lösungen mittels Kurzvorträgen (5-10 Min)
  - Vortragende Teams werden „zufällig“ ausgewählt
  - Freiwillige Präsentationen sind immer möglich
  - Jedes Team muss einmal im Semester vortragen!
- Ziel: Teilen von gesammelten Erfahrungen und Erkenntnissen!
  - Auch Fehlritte sind interessant!
- Ausgabe und Erläuterung der nächsten Übungsaufgabe

# Übungsaufgaben

---

- Analyse eines PubMed-Korpus
- Implementierung einer Volltextsuche
- Part-of-speech tagging mit Hidden-Markov-Modellen
- Klassifikation von Textdokumenten
- Regelbasiertes Named Entity Recognition (NER)
- NER mit Maschinellern Lernen

# Wettbewerb

---

- Alle Übungsaufgaben sind als Wettbewerb konzipiert
  - Die 5 Teams mit der besten / schnellsten Lösung erhalten Punkte
  - Am Ende des Semesters bekommt das beste Team ein kleines Präsent
- Teilnahme am Wettbewerb ist nicht verpflichtend!

# Allgemeine Informationen

---

- Webseite der Vorlesung
  - Vorlesungsmaterialien + weiterführende Referenzen
  - [https://hu.berlin/vl\\_masprach1819](https://hu.berlin/vl_masprach1819)
- Webseite der Übung
  - Übungsaufgaben + Materialien
  - [https://hu.berlin/ue\\_masprach1819](https://hu.berlin/ue_masprach1819)
- Rückfragen zur Übung per Mail an mich:
  - [mario.saenger@informatik.hu-berlin.de](mailto:mario.saenger@informatik.hu-berlin.de)
  - RUD 25, 4.404

# Gruppenbildung

---

- Findet Euch bis spätestens zum 22.10.2018 in Zweiergruppen zusammen
  - Jede Gruppe kann einen „fancy“ Gruppenname wählen 😊
  - Gruppenname + Name der Gruppenteilnehmer per Email an mich
- Google-Doc-Sheet zur Gruppenbildung
  - <https://bit.ly/2HOaEpa>
  - Kontaktiert Euch bitte selbstständig!



---

# Fragen?