

**RUHR-UNIVERSITÄT BOCHUM**

# **COMPUTERLINGUISTISCHE HAUSARBEITEN**

# Was gehört in eine Projektskizze?

## Ziel der Arbeit

- Wie lautet das (übergeordnete) Ziel/ die **Forschungsfrage**?
- Warum ist das eine relevante Frage? Wo ist die Forschungslücke (Bezug zu bestehender Literatur)?
- Wie wird dieses Ziel im Projekt konkret **operationalisiert**?  
Was ist die Begründung für diese Vorgehensweise?

# Was gehört in eine Projektskizze?

## Daten

- Woher bekommt man die nötigen Daten?
- ggf. Wie müssen diese zunächst aufbereitet werden? Welche Tools gibt es dafür?
- Wie viele Daten werden angestrebt? Wie viele davon als Trainings- und Testdaten?

# Was gehört in eine Projektskizze?

## Methoden

- Welche Methoden werden angewandt?  
Mit kurzer Begründung, warum diese zur Forschungsfrage passen (ggf. mit Bezug auf bestehende Literatur)
- Konkret: Welche NLP-Tools benötigen Sie dafür? (Wo) sind diese (für die jeweilige Sprache!) verfügbar?
- Wie beantworten Sie damit die Forschungsfrage?

# Was gehört in eine Projektskizze?

## Evaluation

- Wie evaluieren Sie am Ende Ihr Programm?  
Mit welcher Baseline vergleichen Sie das Ergebnis?

# Computerlinguistische Projektskizzen

## Was gehört *nicht* in eine Projektskizze?

- Details zur Aufteilung des eigenen Codes
  - d.h. man sollte den geplanten Ablauf abstrakt beschreiben, aber nicht die konkrete Aufteilung in Funktionen etc. nennen
- Details zu genutzten (built-in) Funktionen
  - aber: welche externen Libraries potenziell genutzt werden können ist relevant!
- Rechtschreib- und Grammatikfehler
  - Eine Projektskizze sollte den gleichen sprachlichen Anforderungen genügen, die auch an eine Hausarbeit gestellt werden, auch wenn es nur ausformulierte Stichpunkte sind

# Wozu verfasst man Projektskizzen?

- Nicht für den/die Dozent:in sondern in erster Linie für sich selbst
  - Ist das, was ich vorhabe, schlüssig und praktisch durchführbar?
  - Gibt es die Daten und Tools, die ich benötige, oder müsste ich dazu selbst noch Code schreiben, was dann ggf. zu zeitaufwendig ist?

# Computerlinguistische Hausarbeit

- **Ziel:** Ein Paper, wie es bei einer computerlinguistischen Konferenz eingereicht werden könnte
- Die Hausarbeit umfasst im Wesentlichen die gleichen Punkte, wie Ihre Projektskizze, nur jetzt mit mehr Inhalt gefüllt!
  - Einleitung (Problemstellung/Forschungsfrage/Hypothese, Motivation)
  - Bisherige Arbeiten zum Thema
  - Methodisches Vorgehen (inkl. Beschreibung der Daten)
  - Ergebnisse (Beschreibung + Evaluation)
  - Diskussion der Ergebnisse (inkl. Analyse der Fehler)
  - Limitationen und Ausblick auf mögliche weitere Forschung (knapp!)



# Computerlinguistische Hausarbeit

## Was gehört sonst noch rein?

- Beispiele!
  - Ganz wichtig: Wann immer möglich, konkrete Beispiele zur Illustration angeben
- Tabellen/Visualisierungen
  - Ergebnisse/Details zu den Daten möglichst in Tabellen statt im Fließtext darstellen
  - Häufig noch besser: Ergebnisse mit Diagrammen visualisieren (z.B. mit matplotlib)
- Ein roter Faden
- Am Ende die Beantwortung der Forschungsfrage
- Präzise Ausdrucksweise, korrekte Terminologie

**Ziel: leser:innenfreundlich gestalten!**

# Computerlinguistische Hausarbeit

## Was gehört *nicht* in eine CL-Hausarbeit?

- Auch hier gilt das gleiche wie für die Projektskizze! Vor allem:
  - Keine zu konkreten Details zur Aufteilung des eigenen Codes
    - D.h. kein Bericht, was man in separate Funktionen ausgelagert hat oder wie viele Funktionen man nutzt, sondern Code eher abstrakt beschreiben
  - Keine Details zu genutzten (built-in) Funktionen
  - Keine Screenshots von Code!
- Nicht: „The history of my research“
  - d.h. keine Wertung, z.B. was leicht/schwierig war, sondern ein sachlicher Bericht des finalen Projekts
- Kein sprachliches Geschwurbel, vage Ausdrucksweise („man weiß doch was gemeint ist“) sondern präzise und terminologisch korrekte Beschreibungen

# Computerlinguistische Hausarbeit

## Formale Anforderungen

- 7-9 Seiten (Referenzen und ggf. Appendix zählen extra) im vorgegebenen Format:
  - <https://github.com/acl-org/acl-style-files>  
(Das ist das echte Template für ACL-Konferenzen)
- In Word bitte **APA** Stil zum Zitieren benutzen! In LaTeX kann einfach das voreingestellte Zitationsformat verwendet werden

# Computerlinguistische Hausarbeit

## Umfang

- BA Vertiefung 1: 2 CP
- BA Vertiefung 2: 5 CP
- Mastermodul 1: 4 CP (ggf. + 3 CP wenn große Studienleistung erwünscht)
- Mastermodul 2: 6 CP (ggf. + 3 CP wenn große Studienleistung erwünscht)

## Fristen

- Themenabsprache bis zum 19.02.  
(vom 21.02.-11.03. bin ich im Urlaub und nicht erreichbar)
- Abgabefrist: zunächst 31.03. – individuelle Verlängerung nach vorheriger Absprache möglich

# Themenideen

## Adversarials

- Vergleich mit dem Original-Code auf den deutschen Daten
- Vergleich desselben Ansatzes für mehrere Sprachen
- Adversarials aus der Aufgabenstellung bzw. zufälligen Ausschnitten aus der Aufgabenstellung erzeugen

## Essay Scoring

- Vergleich des Einflusses bestimmter Features

## Gap-Fill Bundles

- Andere Sprachmodelle zugrunde legen (z.B. S-BERT)