

Themenbeschreibung

Projekt: Angewandte Informationswissenschaft

1. **Welche Themen im Bereich der Informationswissenschaft und Sprachtechnologie interessieren mich? (z.B. Informetrie, Social Media, Gamification, Information Retrieval, Computerlinguistik, ...) Welche(r) Aspekt(e) genau?**
 - ⇒ Besonderes Interesse an Computerlinguistik
 - ⇒ Durch mein Praktikum bin ich viel in Kontakt mit dem Thema Artificial Intelligence und Machine Learning gekommen, in diesem Zusammenhang ist mir die Bedeutung von vielen linguistischen Problemen noch mal deutlicher geworden
 - ⇒ Daher Themenbereich Computerlinguistik, z.B. Autokorrektur
2. **Welche Programmiersprachen und Tools liegen mir bzw. welche würde ich gerne ausprobieren? (z.B. Python, Java, Prolog, Flask, D3, Django, ...)**
 - ⇒ Programmierkenntnisse: Java, Python, Prolog
 - ⇒ Bevorzugen würde ich Python
3. **Wie könnte ich mich in Form eines Projektes mit dem gewählten Thema auseinandersetzen? (z.B. Datenanalyse, Visualisierung, App, Text Mining, Suchmaschine, Wissensordnung, ...)**
 - ⇒ Datenanalyse: Was sind die häufigsten Schreibfehler die entstehen, wie können diese bearbeitet werden → dazu eine passende Visualisierung
 - ⇒ Oder: eigene Autokorrektur, hierbei muss natürlich der Umfang deutlich eingeschränkt werden, nur bestimmte Fehler könnten ausgewählt werden
4. **Ist dies in der Bearbeitungszeit (ca. 4-8 Wochen) realisierbar? Wenn nicht, was könnte ich eingrenzen/erweitern?**
 - ⇒ S.o.: nur bestimmte Fehler korrigieren, z.B. Buchstabendreher, weggelassene Buchstaben, versehentliche Leerzeichen, Groß- und Kleinschreibung usw.
5. **Material**
 - ⇒ Linguistische Methoden (z.B. aus dem Bereich Information Retrieval):
 - N-Grams (z.B. Henrichs's Pentagram Index)
 - Acquaintance: N-Grams in the Vector Space
 - HAIRCUT: n -Grams Processed Probabilistically
 - Dependencies within Character Sequences
 - ⇒ Linguistisches Identitiy Matching
 - ⇒ NLP Methoden
 - ⇒ Machine Learning