

# Projektseminar Angewandte Informationswissenschaft – Projektidee

## Sentiment-Analyse mit Twitterdaten

Name: Sabrina Wirkner

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?

- Mich interessiert im Bereich der Informationswissenschaft und Sprachtechnologie vor allem die Computerlinguistik. Die Analyse von Sprache, also Spracherkennung, Automatisches Übersetzen, Schreibkorrektur, finde ich spannend, besonders aber interessiere ich mich für die Sentiment-Analyse. Ich habe bereits einmal beim Testen einer emotionalen Suchmaschine mitgemacht, wodurch ich auf das Thema der Sentiment-Analyse mit schriftlichen Daten aufmerksam wurde. Wie gut eine Maschine Gefühle in Texten verstehen kann, aber auch was für Probleme auftauchen, würde ich gerne selbst herausfinden.

2. Welche Programmiersprachen und Tools liegen mir bzw. welche würde ich gerne ausprobieren? (z.B. Python, Java, Prolog, Flask, D3, Django, ...)

- Ich würde gerne mit Python arbeiten, da ich mit dieser Programmiersprache am besten zurechtkomme. Außerdem ist Python bekannt dafür, gut geeignet für die Verarbeitung von Sprache zu sein.

3. Wie könnte ich mich in Form eines Projektes mit dem gewählten Thema auseinandersetzen? (z.B. Datenanalyse, Visualisierung, App, Text Mining, Suchmaschine, Wissensordnung, ...)

- Meine Idee für ein Projekt (und auch für die Bachelorarbeit) ist die Sentiment-Analyse mit Twitterdaten, also mit Texten, die aus maximal 140 Zeichen bestehen. Zu diesem Thema würde ich gerne eine Klassifikation aufbauen, indem ich eine oder mehrere Methoden des Machine Learnings anwende (z.B. Naive Bayes / Bag of Words, Support Vector Machines, Pointwise Mutual Information o.Ä.) und diese mit manuell klassifizierten Twitterdaten teste. Dafür müsste ich zudem ein Sentiment-Lexikon (mit Gewichten) nutzen.

4. Ist dies in der Bearbeitungszeit (ca. 4-8 Wochen) realisierbar? Wenn nicht, was könnte ich eingrenzen/erweitern?

- Ich denke, dass das Projekt in der Bearbeitungszeit realisierbar ist. Den Umfang kann man leicht anpassen, indem man weniger oder mehr Methoden testet und Ansätze hinzu- oder wegnimmt. Wichtig für mich ist, dass ich am Ende eine Basis für eine Klassifikation habe, die funktioniert und die sich nach Bedarf erweitern lässt. Da das Aufbereiten und Klassifizieren von Daten aufwändig ist, ist hier meine **Frage**, ob ich einen Teil der bereits aufbereiteten und klassifizierten Twitterdaten für das Projekt nutzen darf.