

Acquisition of a German Fan Fiction Corpus and Analysis in the Context of Gender Representation

Thema:

Acquisition of a German Fan Fiction Corpus and Analysis in the Context of Gender Representation

Art:

[MA](#)

BetreuerIn:

[Thomas Schmidt](#)

BearbeiterIn:

Jonathan Sasse

ErstgutachterIn:

[Christian Wolff](#)

Status:

[abgeschlossen](#)

Stichworte:

[Fan Fictions](#), [NLP](#), [Computational Literary Studies](#), [Digital Humanities](#)

angelegt:

2022-02-11

Antrittsvortrag:

2022-07-04

Abschlussvortrag:

2023-01-23

Hintergrund

Fanfiktion stellt eine gute Datenquelle für Forschungsfragen aus dem Bereich des Natural-Language-Processing dar. Für statistische Analysen als auch das Training von NLP-Modellen, steht hier eine große Anzahl an frei zugänglichen Texten von verschiedensten Autoren und Schreibstilen zur Verfügung.

Zielsetzung der Arbeit

Es soll ein Korpus für deutsche Fanfiktion geschaffen werden, der für weiterführende Arbeiten verwendet werden kann. Zudem sollen schon erste statistische Analysen mit Fokus auf die Verteilung der Geschlechterrollen in den erlangten Texten durchgeführt werden.

Konkrete Aufgaben

- Websites für deutsche Fanfiktion sammeln und gemäß Nutzbarkeit evaluieren

- Fanfiktions-Texte mit sämtlichen nützlichen Metadaten mit einem Web-Scraper crawlen
- Daten statistisch auswerten
- Zusätzliche Personen-Geschlechter-Daten mit Hilfe eines pre-trained Named-Entity-Recognition Models erlangen und auswerten

Erwartete Vorkenntnisse

- gute Kenntnisse in Python, Web-Scrapern und Datenbanken
- Grundkenntnisse für die Verwendung von neuronalen Netzwerken (TensorFlow, Transformers, NER)

Weiterführende Quellen

- [Beyond Canonical Texts: A Computational Analysis of Fanfiction](#) (Milli & Bamman, EMNLP 2016)
- [FanfictionNLP: A Text Processing Pipeline for Fanfiction](#) (Yoder et al., NUSE 2021)
- [GottBERT: a pure German Language Model](#) (Scheible et al., 2020)

From:

<https://wiki.mi.ur.de/> - **MI Wiki**

Permanent link:

https://wiki.mi.ur.de/arbeiten/german_fan_fictions

Last update: **11.01.2024 14:01**

