

Gender Gap Tracker für die Schweizer Medien

Studiengang: BSc in Informatik | Vertiefung: Data Engineering
Betreuerin: Prof. Dr. Mascha Kurpicz-Briki
Experte: Alain Joray

Die Bachelor Thesis „Gender Gap Tracker für die Schweizer Medien“ untersucht den Gender Gap in textbasierten Deutschschweizer Online-Medien anhand der Anzahl Zitate von Männern und Frauen.

Einleitung

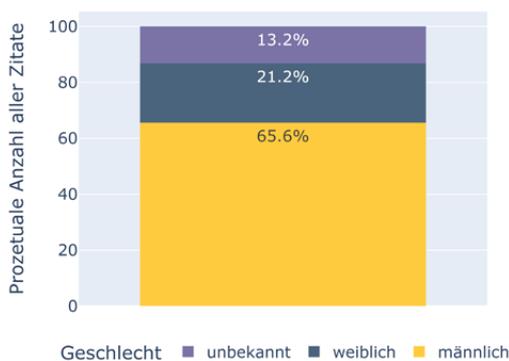
Die Datengrundlage bildet eine Datenbank mit über 350'000 Artikeln aus vier Nachrichtenportalen, die in einem separaten Vorprojekt (Project 2) mittels Web-crawling aufgebaut wurde. Die Inspiration für dieses Projekt und das gewählte Vorgehen kam von der kanadischen Studie „The Gender Gap Tracker: Using Natural Language Processing to measure gender bias in media“.

Vorgehen

Zum Ermitteln des Gender Gaps setzt diese Arbeit - in Anlehnung an die Vorbildsstudie - auf Machine Learning basierte Werkzeuge aus dem Bereich Natural Language Processing. Konkret verwendet die geschriebene Software ein Spacy Modell um mithilfe von Named Entity Recognition, Part of Speech Tagging, Coreference Resolution und Dependency Parsing Zitate und Personen aus den Texten zu extrahieren. Zum Überprüfen der Qualität haben wir manuell Test Sets erstellt, anhand derer wir die Performance der Software messen können (vgl. Diagramm Testresultat).

Resultate

Die Resultate zeigen, dass im Durchschnitt 65.6% der Zitate von Männern sind und 21.2% von Frauen (vgl. Diagramm Anzahl Zitate prozentual pro Geschlecht).



Anzahl Zitate prozentual pro Geschlecht

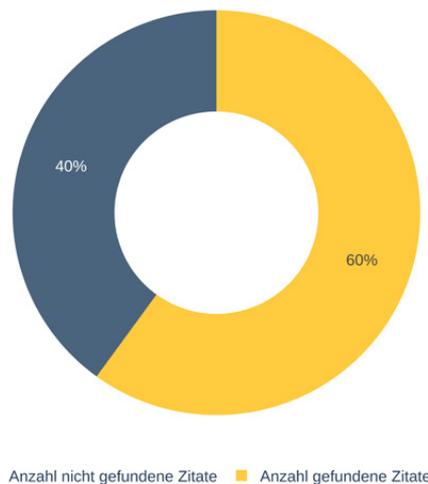
Der kleinste Unterschied aller Portale liegt bei 61.6% und 25.3%.

Limitationen

Aufgrund des zeitlichen Rahmens konnte in dieser Arbeit nur die Extraktion der gängigsten Formen von Syntaktischen Zitaten implementiert werden. Das Programm erfasst laut unseren Tests etwa 60% der Syntaktischen Zitate (vgl. Diagramm Testresultat). Ohne genauere Angaben zu der Verteilung der Anzahl Zitate auf deren unterschiedlichen Kategorien zu haben, schätzen wir diesen Anteil auf etwa 50%.

Interpretation

Unter Berücksichtigung der Limitationen bedeuten diese Ergebnisse, dass in der Deutschschweiz ein signifikanter Gender Gap in den textbasierten Online-Medien vorliegt und Frauen deutlich unterrepräsentiert sind, gemessen an der Anzahl der gefundenen Zitate. Der Unterschied ist damit grösser, als in den meisten anderen Studien, die für diese Arbeit zum Vergleich herbeigezogen wurden.



Testresultat



Anina Aeschbacher



Felix Michael Neidhart