

Stop Stupidity - Automatische Überprüfung vermeintlicher Fakten mit Information Retrieval

Fake News sind Meldungen, welche dem Leser durch absichtlich falschen Inhalt oder unseriöse Berichterstattung ein falsches Bild der Wirklichkeit vermitteln. Solche Meldungen sind nicht einfach zu erkennen und können oft nur durch eine manuelle, meist zeitintensive Recherche überprüft werden. Das Ziel dieser Arbeit ist, ein System zu entwickeln, welches diese Überprüfung automatisiert. Anhand von Behauptungen, sogenannten Claims, soll eine Klassifizierung vorgenommen werden. Damit sollen Claims als False, Half-True oder True eingestuft werden können. Dazu wird ein System implementiert, welches auf Ansätzen und Ideen basiert, welche in verschiedenen Arbeiten im Rahmen des CLEF Fact Checking Lab in den Jahren 2018 und 2019 angewendet wurden. Für jeden Claim wird eine Suche mit Google durchgeführt. Sätze aus Webseiten, welche in Bezug zum Claim eine hohe inhaltliche Relevanz vorweisen, werden extrahiert. Ausgehend von diesen Sätzen werden verschiedene Features generiert. Diese dienen als Input für ein Modell, welches mit einem XGBoost-Algorithmus trainiert wird. Die Ergebnisse zeigen, dass im Vergleich zu den besten Arbeiten des Fact Checking Lab 2018 ähnlich gute Resultate erreicht werden und auch die Baseline klar übertroffen wird. Die besten Ergebnisse wurden durch verschiedene Kombinationen von Features sowie mit einem grösseren Datensatz erreicht. Erste Analysen zeigen, dass durch eine ungenügende Normalisierung und Unterteilung der Sätze zu viele falsche Klassifizierungen auftreten. Um bessere Resultate zu erzielen, könnten sich weiterführende Arbeiten vermehrt mit der korrekten Identifikation eines Satzes auseinandersetzen.

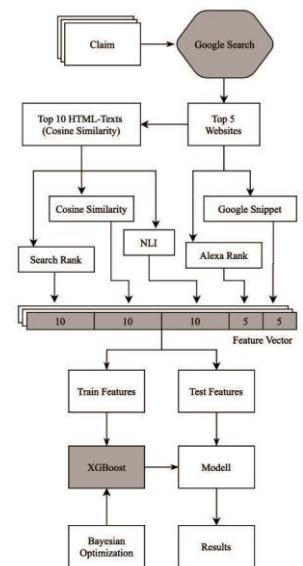


Diplomierende

Jonas Huber
Martin Wädensweiler

Dozent

Martin Braschler



Übersicht über das implementierte System zur Überprüfung vermeintlicher Fakten: Mit dem zu überprüfenden Claim wird mit Google nach Webseiten gesucht. Aus den ersten fünf der gefundenen Webseiten werden die zehn Sätze mit der höchsten Cosine Similarity zum Claim extrahiert. Daraus werden die Features Search Rank, Cosine Similarity und NLI (Inferenz zwischen Claim und extrahiertem Satz) generiert. Weitere Features werden aus dem Alexa Rank und dem Google Snippet der Top 5 Webseiten gebildet. Diese Features dienen als Input für einen XGBoost Classifier, welcher den Claims die Labels True, Half-True und False zuweist.