UNIKASSEL
VERSITÄT

Fachgebiet
Wirtschaftsinformatik
FB 07 Wirtschaftswissenschaften

Universität Kassel · Prof. Dr. Jan Marco Leimeister · D – 34109 Kassel

Univ.-Prof. Dr.
Jan Marco Leimeister

e leimeister@uni-kassel.de
t +49 (0) 561 804-6064
f +49 (0) 561 804-6067

Pfannkuchstraße 1
34121 Kassel

Sekretariat
Mo.-Do.: 9:00-13:00 Uhr

Mechthild Häckl
e mechthild.haeckl@uni-kassel.de
t +49 (0) 561 804-6068

27.09.2021

Verbesserung künstlicher Intelligenz durch Daten: Aufbau eines Stop Word Corpus für den IT support

Hintergrund

Daten sind die Basis von vielen Algorithmen – dennoch wird hauptsächlich der Fokus auf die algorithmische Weiterentwicklung gelegt, anstatt auf die Verbesserung der zugrunde liegenden Daten. Eine Möglichkeit Datenqualität (und dadurch eine trainierte künstliche Intelligenz) zu verbessern ist durch den Aufbau eines Corpus von Stop Words. Stop Words beschreiben Wörter, die bei einer Informationsgewinnung nicht beachtet werden (z.B. wenn Such-Anfragen gestellt werden), da sie von niedriger Relevanz sind oder sehr häufig vorkommen (z.B. der, kann, soll, ich,…). Die bestehenden Listen von Stop Words bestehen häufig auf Basis frei zugänglicher Texte wie Nachrichten-Artikel.
Dadurch ergeben sich Biases in den Datensätze. Werden diese Stop Words beispielsweise in der Domäne des IT supports verwendet, könnten Wörter als unterschiedlich/ falsch wichtig gewertet werden. Durch den Aufbau eines domänen-spezifischen Corpus können alle verwendeten Techniken (insbesondere NLP-Techniken), die im IT-Support verwendet werden, qualitativ verbessert werden.

Mögliches Thema für Bachelor-/Masterarbeit

Im Rahmen einer Abschlussarbeit wird ein IT support Ticket-Datensatz zur Verfügung gestellt. Zudem können Interviews mit Domänen-Expert:innen (IT support Mitarbeitende und Machine Learning/ Natural Language Expert:innen) geführt werden. Durch eine qualitative Analyse (Interviews) und eine quantitative Analyse (mit dem zur Verfügung gestellten Datensatz) soll ein neuer Corpus für Stop Words speziell für den IT support erstellt werden.
Dieser wird dann mit State of the Art Corpora verglichen – es wird also getestet, ob eine künstliche Intelligenz mit den identischen dahinterliegenden Algorithmen allein durch die Verbesserung der Daten verbessert wurde.
Methode, Entwicklungsumgebung und Programmiersprache sind frei wählbar. Ergebnisformat kann von einer Excel-Tabelle, über ein ausführbares Skript bis zu einem fertig deployten und downloadbaren Package für die Open-Source Community reichen. (Programmier-Kenntnisse

sind keine Voraussetzung und können während der Abschlussarbeit angeeignet werden, z.B. in R oder Python).
Der Umfang der Arbeit unterscheidet sich je nach Bachelor-/Masterarbeit.

Fragen und Bewerbungen an

Mahei Li
Mahei.li@uni-kassel.de