

Analyse von Widersprüchen in natürlich- sprachigen Texten mit Hilfe von Natural Language Processing

Bachelor-/Masterarbeit

Konsistenz und Widerspruchsfreiheit innerhalb des eigenen Datenhaushaltes ist für Unternehmen ein wichtiger Qualitätsanspruch. Die Realisierung dieses Anspruchs hängt jedoch stark von der Art der jeweiligen Daten ab. Während strukturierte Daten (z.B. csv-, xls-Dateien) aufgrund ihrer fest definierten und normalisierten Struktur direkt maschinell verarbeitet werden können, weisen unstrukturierte Daten (z.B. Rechnungen, Verträge oder Berichte) diese Eigenschaften nicht von vornherein auf und können daher nicht von herkömmlichen Revisions- und Controlling-Tools verarbeitet werden. Aus diesem Grund erfolgt die Prüfung von Konsistenz- und Widerspruchsfreiheit solcher Dokumente in der Praxis typischerweise manuell und unter Nutzung von Stichproben.

Eine Möglichkeit, diesen Prozess zu automatisieren und somit gleichzeitig die Nutzung von Stichproben zu überwinden, wird durch den Einsatz von Natural Language Processings (NLP) bereitgestellt. NLP ist ein Forschungs- und Anwendungsbereich, in dem untersucht wird, wie Computer eingesetzt werden können, um natürlich-sprachige Texte zu verstehen und zu bearbeiten, um dadurch verschiedene Aufgaben zu erfüllen. Bevor NLP für den konkreten Einsatz einer Widerspruchserkennung eingesetzt werden kann, ist jedoch entscheidend, die unterschiedlichen Eigenschaften und syntaktischen Formen von Widersprüchen innerhalb der deutschen Sprache zu ermitteln. Letzteres spiegelt das eigentliche Ziel der geplanten Bachelor-/Masterarbeit wider, welche wiederum in die folgenden Kernthemen unterteilt ist:

- Literaturrecherche zur Identifikation relevanter Merkmale
- Konzeption, Entwicklung und Test eines Prototyps, welcher die zuvor ermittelten Merkmale im Rahmen einer Textanalyse adaptiert
- Evaluation und Interpretation der Ergebnisse
- Dokumentation der Arbeit

Für die Bearbeitung dieser Arbeit werden im Idealfall folgende Eckdaten erfüllt:

- Du hast bereits erste Erfahrungen mit der Programmiersprache Python gesammelt (idealerweise im Bereich NLP) oder bist bereit, diese durch intensive Einarbeitung aufzubauen
- Du hast gute Deutschkenntnisse (erforderlich für die Ergebnisbeurteilung der NLP-Module)

Die Bachelor-/Masterarbeit bettet sich in das Forschungsprojekt *Difa* (Data Intelligence for Audit,) ein – einer Forschungs Kooperation zwischen der Abteilung Wirtschaftsinformatik / Very Large Business Applications (VLBA) der Universität Oldenburg und der Volkswagen AG ein. Das Projekt beschäftigt sich u.a. damit, datenbasierte Techniken für die interne Revision zu entwickeln.

Bei Interesse an einer Bearbeitung des ausgeschriebenen Themas, sende eine Nachricht an:

M.Sc. Gerrit Schumann
University of Oldenburg
Department of Computing Science
Mail: gerrit.schumann@uol.de

DEPARTMENT FÜR INFORMATIK

ABTEILUNG
WIRTSCHAFTSINFORMATIK I
VERY LARGE BUSINESS APPLICATIONS

PROF. DR. JORGE MARX GÓMEZ

TELEFONDURCHWAHL
(0441) 7 98 – 4470
Sekretariat – 4478

FAX
(0441) 7 98 – 4472

EMAIL
Jorge.Marx.Gomez@uni-oldenburg.de

GEBÄUDE A4
Uhlhornsweg 84 – Raum A4 3-315

OLDENBURG
03.02.2022



VERY LARGE
BUSINESS
APPLICATIONS
Carl von Ossietzky
Universität Oldenburg

POSTANSCHRIFT
D-26111 Oldenburg

PAKETANSCHRIFT
Ammerländer Heerstraße 114 - 118
D-26129 Oldenburg

TELEFONZENTRALE
(0441) 7 98 – 0

BANKVERBINDUNG
Landessparkasse zu Oldenburg

IBAN: DE46 2805 0100 0001 9881 12
BIC: SLZODE22