



Bachelor- oder Masterarbeit

Entwicklung eines Tools zum Überprüfen von Publikationen auf Plagiate

Stand der Technik

Das an der Leibniz Universität Hannover verwendete Tool zur Überprüfung von Plagiaten Docoloc bietet eine Basisfunktionalität zum Überprüfen von Plagiaten. Jedoch hat es einige Nachteile und erkennt nicht zusammenhängende Sätze bei einer Quelle. Das Tool erkennt ebenfalls nicht, wenn Texte aus dem Englischen übersetzt wurden und dieser Text dann eingereicht wird. Viele Dozenten nutzen dieses Tool auch nicht, sondern geben einzelne Sätze bei Suchmaschinen ein und überprüfen dann händisch die Ergebnisse. Ein weiteres Problem ist, dass viele Institute auch große Sammlungen an vergangenen Abschlussarbeiten haben, welche nicht öffentlich abrufbar sind, jedoch ein hohes Potential haben die Quelle für ein Plagiat zu sein. Diese werden allerdings häufig bei Plagiatsprüfungen nicht genutzt.

Arbeitsauftrag

Es soll eine Webseite in Python entwickelt werden, welche es dem Nutzer erlaubt beliebige PDF Dokumente hochzuladen und auf Plagiate überprüfen zu lassen. Dabei soll es möglich sein je nach Installation verschiedene Suchmaschinen zu nutzen.

Dazu muss das PDF zunächst aufgeteilt werden in:

- Beiwerk, wie z.B. Titelseite, Inhaltsverzeichnis, Literaturverzeichnis, Kopf und Fußzeilen, etc.
- Bilder, Tabellen, Quelltexte, etc und deren Beschreibungen/Titel
- Fußnoten
- Der eigentliche textuelle Inhalt

Anschließend soll der textuelle Inhalt in Sätze überführt werden und diese mithilfe von Online-Services übersetzt und diese Übersetzungen mit Hilfe von Suchmaschinen auf Plagiate überprüft werden.

Ein Focus wird hierbei gelegt auf der Ähnlichkeit von Umgebungstexten. Sollte also ein Absatz komplett aus einer Quelle übernommen sein, sollte das Tool dieses erkennen und entsprechend berichten.

Die Ergebnisse sollen anschließend auf einer Webseite angezeigt werden.

Anforderungen

Die Arbeit kann sowohl als Bachelor, als auch als Masterarbeit erledigt werden. Bei einer Bachelorarbeit wird der Arbeitsauftrag zum Aufteilen der PDF verringert und geringere Anforderungen an die abschließende Ausgabe der Ergebnisse gestellt.

Wir erwarten grundlegende Kenntnisse in der Programmierung mit Python und eine Bereitschaft sich in neue Themen einzuarbeiten.