Fazit

Das Ziel dieses Praktikums ist es gewesen eine Suchmaschine zu erstellen und zu implementieren, durch welche es möglich wird nach wissenschaftlichen Publikationen aus dem Bereich der digitalen Text-Forensik zu suchen.

Die Texte wurden in eine einheitliche Form gebracht und indexiert. Während der Vorverarbeitung wurde außerdem auf die Extraktion der Titel der Publikationen ein besonderer Wert gelegt, da diese aus Usability-Gründen im Suchergebnis angezeigt werden und auch für das Ranking nicht unerheblich sind. Weiterhin wurde zur Feststellung der Relevanz der bekannten BM25 Algorithmus verwendet. Des Weiteren wurden Log-Daten, die Verweildauer auf einem Dokument und die Anzahl der Erwähnungen der Paper in den anderen wissenschaftlichen Texten verwendet.

Alle Daten werden in einer Datenbank gespeichert. Die Suchmaschine ist voll funktionsfähig auf einem Webserver implementiert und kann über eine grafische Weboberfläche genutzt werden. Neben dem Titel der zur Suche relevanten Publikationen werden auch Snippets generiert, um eine angenehmere und effizientere Nutzung zu gewährleisten.

In der Evaluation der Suchmaschine hat sich gezeigt, dass gute Ergebnisse erzielt werden können, denn es befinden sich für nahezu alle Suchanfragen relevante Ergebnisse unter den ersten Suchergebnissen.

Wie bei fast allen Projekten in der Informatik handelt es sich auch bei dieser Suchmaschine mehr um ein fortlaufendes Projekt. Es können auf Grund von Nutzerfeedback immer weiter Verbesserungen vorgenommen werden. Zurzeit wird daran gearbeitet, den Upload von Texten durch den Nutzer zu gewährleisten und diese mit in die Suchmaschine einfließen zu lassen.

Zusammenfassend kann gesagt werden, dass es unserem Team im Laufe des Praktikums gelungen ist, eine Suchmaschine zu entwickeln, welche die an sie gestellten Anforderungen voll und ganz erfüllt. Es wird durch sie möglich auf einer großen Menge von wissenschaftlichen Texten eine Suche durchzuführen und relevante Ergebnisse zu erhalten.