

## Evaluierung von Ranking-Algorithmen für natürlichsprachige Datenbankabfragen

Die vorliegende Arbeit befasst sich mit der Frage nach effizienten Suchmethoden innerhalb einer umfangreichen Datensammlung. Die Suchanfrage ist dabei natürlichsprachig formuliert, da von Anwendern keine spezifischen Syntax-Kenntnisse wie beispielsweise SQL erwartet werden. Für die Repräsentation relevanter Antworten werden mittels unterschiedlicher Algorithmen Rankings erstellt, welche die Ergebnisse geordnet darstellen.

Zur Klärung der Fragestellung werden zwei unterschiedliche Ansätze zum Durchsuchen der Datenmenge verglichen. Zum einen handelt es sich dabei um die Suche innerhalb einer relationalen Datenbank und zum anderen um den Einsatz von Information Retrieval-Methoden. Für die Versuche wurden dazu das SODA-System sowie das Terrier IR-System genutzt, deren Performances anschliessend aufgrund der berechneten Precision- und Recall-Werte verglichen wurden. Als Vergleichswert wurde ein vorhandenes Relevance Assessment als Goldstandard miteinbezogen, um qualitative Auswertungen zu ermöglichen. Dieses Relevance Assessment wurde im Rahmen der INEX-Konferenz erzeugt und ist Teil eines Datenkorpus, das ebenfalls die Datengrundlage sowie die Anfragen für die Experimente dieser Arbeit enthält.

Die Auswertung der insgesamt 45 Anfragen zeigte, dass SODA in jenen Fällen eine bessere Performance erreichte, in denen eine präzisionsorientierte Antwort erwartet wurde. Über alle Queries betrachtet, wurden so die höheren Präzisionswerte als bei Terrier beobachtet. Terrier hingegen erzielte die besseren Werte unter Berücksichtigung der Ausbeute, indem es für mehr Anfragen Ergebnisse erzielte als SODA.

Performance-Einbussen bei SODA waren in jenen Fällen festzustellen, in denen sehr lange Anfragen formuliert wurden oder dann, wenn das System SQL-Statements erzeugte, welche das gesamte Datenset einer Tabelle abfragten. Terrier hingegen lieferte keine Ergebnisse zurück für die Fälle, in denen das Relevance Assessment genau ein einziges relevantes Dokument auswies.



Diplomierende  
Stefanie Nadig  
Jan Scheuermeier

Dozierende  
Martin Braschler  
Kurt Stockinger

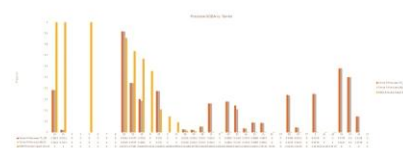


Bild klein 1.