

Praktikum 4

Hinweise

Es gelten weiterhin alle Hinweise und Anforderungen von Praktikumsblatt 2!

Aufgabe 1 – Signaturverfahren

1. Implementieren Sie ein Signaturverfahren für das boolesche Retrieval-Modell. Wählen Sie eine geeignete Hash-Funktion sowie Blocksignaturen und verwenden Sie eine geeignete Datenstruktur für die Signaturen! Nutzen Sie zur Erzeugung der Signaturen die Parameter $F = 64$, $D = 4$ und ermitteln Sie den dazu optimalen Wert für m .
2. Optional: Implementieren Sie das Signaturverfahren so, dass die Parameter F und D veränderbar sind.
3. Ermöglichen Sie dem Nutzer eine Suche mittels Konjunktion und Diskjunktion von zwei Suchbegriffen.

Aufgabe 2 – Vektorraummodell

1. Implementieren Sie die Suche mit dem Vektorraummodell mithilfe von invertierten Listen. Beachten Sie dabei die folgenden Hinweise:
 - (a) Nutzen Sie $tf.idf$ für die Generierung der Termgewichte!
 - (b) Die Anfragevektorbildung realisieren Sie bitte nach Salton/Buckley (1988).
 - (c) Verwenden Sie den Basisalgorithmus mit invertierten Listen, wie er bei Henrich vorgestellt wird.