



Übung zur Vorlesung
Einsatz und Realisierung von Datenbanksystemen im SoSe16

Moritz Kaufmann (moritz.kaufmann@tum.de)
<http://db.in.tum.de/teaching/ss16/impldb/>

Blatt Nr. 11

Hausaufgabe 1

Berechnen Sie für folgende drei Dokumente die TF-IDF-Werte:

1. „Beim Fußball dauert ein Spiel neunzig Minuten – und am Ende gewinnen die Deutschen“
2. „Beim Fußball muss das Runde (der Ball) in das Eckige (das Tor)“
3. „Nie war ein Tor so wertvoll wie jetzt“

Welches Ranking ergibt sich gemäß der Relevanzwerte für die Anfrage: „Fußball“ \wedge „Tor“. Zur Ermittlung des TF Wertes gehen sie davon aus, dass alle Wörter eines Dokuments *interessant* sind?

Hausaufgabe 2

Ihr System ist durch eine DDoS Attacke blockiert. Schreiben sie eine Streaming SQL Anfrage, die IP's identifiziert über die das System angegriffen wird. Als Metrik soll dabei das Verhältnis von neu geöffneten Verbindungen im Vergleich zur Anzahl der Datenpakete pro IP in einem 5 Minuten Fenster nicht über 20% liegen. Das System empfängt dabei folgenden Eventtyp:

Packet [IP,TYPE,DATE]

TYPE kann dabei einer der folgenden Werte sein:

NEW_CONNECTION, DATA, CLOSE_CONNECTION.

Hausaufgabe 3 In Abbildung 1 gezeigte Netzwerk von Web-Seiten wird ein kleines Beispiel für einen Webgraphen gezeigt. Lösen sie folgende Aufgaben.

1. Berechnen Sie, für das in Abbildung, den PageRank, sowie die HITS-Werte nach 2 Iterationen. Nutzen Sie $1/|V|$ als Anfangswert für den PageRank und 1 für HITS. $a = 0.1$
2. Formulieren sie eine Iteration des Pagerank Algorithmus in SQL. Der Graph ist dabei in der Tabelle *edges(From,To)* gespeichert, die aktuelle PageRank Gewichtung in der Tabelle *pagerank(Vertex,Weight)*. Sie können die Anzahl der Knoten als Konstante annehmen, z.B. 1000.
3. Formulieren sie die SQL Anfrage nun als rekursive SQL Anfrage um.

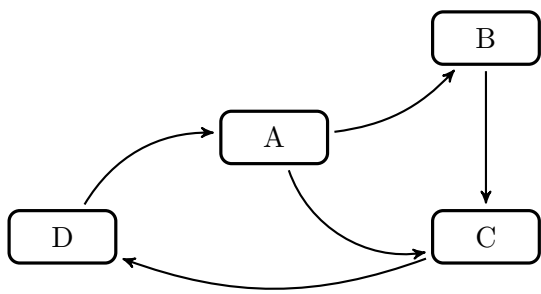


Abbildung 1: Ein kleiner Webgraph.