

Aufgabe 9 Positionelle Indexes

Nachfolgend ein Ausschnitt eines positionellen Indexes der Form „Term: doc1: position1, position2, ...; doc2: position1, position2, ... ; etc.“:

angels: 2: (36,174,252,651); 4: (12,22,102,432); 7: (17);
fools: 2: (1,17,74,222); 4: (8,78,108,458); 7: (3,13,23,193);
fear: 2: (87,704,722,901); 4: (13,43,113,433); 7: (18,328,528);
in: 2: (3,37,76,444,851); 4: (10,20,110,470,500); 7: (5,15,25,195);
rush: 2: (2,66,194,321,702); 4: (9,69,149,429,569); 7: (4,14,404);
to: 2: (47,86,234,999); 4: (14,24,774,944); 7: (199,319,599,709);
tread: 2: (57,94,333); 4: (15,35,155); 7: (20,320);
where: 2: (67,124,393,1001); 4: (11,41,101,421,431); 7: (16,36,736);

1. Welche Dokumente, falls überhaupt, passen auf folgende Suchanfragen? (Ausdrücke in Anführungszeichen sind Phrasensuchen)
 - a. „fools rush in“
 - b. „fools rush in“ AND „angels fear to tread“
2. Rekonstruieren Sie den Inhalt des Dokumentes 2 anhand des Indexes.
3. Erstellen Sie ein Teil des Biword Indexes, das für die Beantwortung der ersten Anfrage notwendig ist.
4. Nehmen Sie an, die Terme „in“ und „to“ sind als Stop-Words aus den o.g. Suchanfragen ausgefiltert worden. Wie kann ein positioneller Index in einem Suchsystem mit Stop-Word Filterung kombiniert werden?

Aufgabe 10

Skip-Lists

Gegeben sei eine Suchanfrage bestehend aus zwei Wörtern.

Für den einen Term besteht die Posting-Liste aus folgenden 16 Einträgen:

(4, 6, 10, 12, 14, 16, 18, 20, 22, 32, 47, 81, 120, 122, 157, 180)

Für den anderen Term besteht die Posting-Liste nur aus einem Eintrag:

(47)

Bestimmen Sie die Anzahl notwendiger Vergleichsoperationen für

1. Herkömmliche Posting-Listen
2. Posting-Listen mit Skip-Pointers und einer Skip-Länge von Wurzel P. Begründen Sie Ihre Antwort.

Aufgabe 11

Wildcard Query

Gegeben sei die Wildcard Suchanfrage Sh*sp*re (Shakespeare).

- a) Erstellen Sie für diese Query, die Anfragen an einen Bigram Index.
- b) Geben Sie die Anfrage bei Google ein. Wie erklären wir uns das Ergebnis?

Aufgabe 12

Levenshtein-Distanz

Erklären Sie was die Levenshtein-Distanz ist und zu welchem Zweck man sie berechnet?

Berechnen Sie die Levenshtein-Distanz zwischen den Wörtern "glue" und "bleu". Stelle dazu die Matrix auf und nimm folgende Kosten an: substitute: 1 insert, delete: 2

Aufgabe 13
Matrizenprodukt

Schauen Sie Definition für die Multiplikation von Matrizen nach und berechnen Sie folgende Produkte:

$$\begin{pmatrix} 1 & 2 & 1 \\ 2 & 1 & 2 \\ 3 & 0 & 2 \end{pmatrix} \quad \begin{pmatrix} 1 & 1 & 2 \\ 2 & 1 & 2 \\ 1 & 3 & 3 \end{pmatrix}$$

$$\begin{pmatrix} 1 & 2 & 1 & 3 \\ 2 & 1 & 2 & 0 \\ 3 & 0 & 2 & 1 \end{pmatrix} \quad \begin{pmatrix} 1 & 1 \\ 2 & 1 \\ 2 & 1 \end{pmatrix}$$