

Aufgabe 14

Jaccard Koeffizient:

Die Mengen sind die k-gram der Terme: A für Frageterm „bord“, B für Term „border“, „lord“, „morbid“ or „sordid“.

Hier nehmen wir an: $k = 2$, d.h. Die Mengen sind hier jeweils die Bi-gram der Terme.

Jaccard Koeffizient: $J(A,B) = \frac{|A \cap B|}{|A \cup B|}$

a) bord und border

A: {bo, or, rd}

B: {bo, or, rd, de, er}

$J(A,B) = \frac{|\{bo, or, rd\}|}{|\{bo, or, rd, de, er\}|} = 3/5$

b) bord und Lord

A: {bo, or, rd}

B: {lo, or, rd}

$J(A,B) = \frac{|\{or, rd\}|}{|\{bo, lo, or, rd\}|} = 2/4 = 1/2$

c) bord und morbid

A: {bo, or, rd}

B: {mo, or, rb, bi, id}

$J(A,B) = \frac{|\{or\}|}{|\{bo, mo, or, rd, rb, bi, id\}|} = 1/7$

d) bord und sordid

A: {bo, or, rd}

B: {so, or, rd, di, id}

$J(A,B) = \frac{|\{or, rd\}|}{|\{bo, or, rd, so, di, id\}|} = 2/6 = 1/3$

Aufgabe 15

Zuerst spaltet eine Menge von Dokumente in den Blöcke

Block 1		Block 2		Block 3	
Term ID	Doc ID	Term ID	Doc ID	Term ID	Doc ID
1	1	7	2	7	3
2	1	2	2	3	4
3	1	3	2	2	4
4	1	1	3	1	4
5	1	3	3	7	4
6	2	5	3		

=> nach den Term ID innerhalb der Blöcke sortieren und wenn die Term ID gleich sind, setzen alle Doc ID in einer Posting-Liste zusammen.

Block 1		Block 2		Block 3	
Term ID	Doc ID	Term ID	Doc ID	Term ID	Doc ID
1	1	1	3	1	4
2	1	2	2	2	4
3	1	3	2 → 3	3	4

4	1	5	3	7	3 → 4
5	1	7	2		
6	2				

=> Merge Blöcke

Term ID	Doc ID
1	1 → 3 → 4
2	1 → 2 → 4
3	1 → 2 → 3 → 4
4	1
5	1 → 3
6	2
7	2 → 3 → 4

=> schreibt das Ergebnis in Database ein

Aufgabe 16

Gegeben sind Dokumente:

d1: A C F E J

d2: A C J H I K

d3: J K L M N

d4: A F J L

d5: C H M N

Zuerst setzen die ersten 10 Postings in einen Block zusammen, dann zweite und bis alle Dokumente in den Speicher gehalten werden.

Block 1		Block 2		Block 3	
Dictionary	Postings	Dictionary	Postings	Dictionary	Postings
A	D1 → D2	K	D2 → D3	C	D5
C	D1 → D2	J	D3 → D4	H	D5
F	D1	L	D3 → D4	M	D5
E	D1	M	D3	N	D5
J	D1 → D2	N	D3		
H	D2	A	D4		
I	D2	F	D4		
10 Postings		10 Postings			

=> sortieren die Dictionaries innerhalb der Blöcke

Block 1		Block 2		Block 3	
Dictionary	Postings	Dictionary	Postings	Dictionary	Postings
A	D1 → D2	A	D4	C	D5
C	D1 → D2	F	D4	H	D5
E	D1	J	D3 → D4	M	D5
F	D1	K	D2 → D3	N	D5
H	D2	L	D3 → D4		
I	D2	M	D3		
J	D1 → D2	N	D3		
10 Postings		10 Postings			

=> Merge die Blöcke

Dictionary	Postings
A	D1 → D2 → D4
C	D1 → D2 → D5
E	D1
F	D1 → D4
H	D2 → D5
I	D2
J	D1 → D2 → D3 → D4
K	D2 → D3
L	D3 → D4
M	D3 → D5
N	D3 → D5

Aufgabe 17

1.

$$\text{a) } A * B = \begin{pmatrix} 5 & 3 & 2 \\ 1 & 0 & -1 \\ 4 & 2 & 0 \end{pmatrix} * \begin{pmatrix} 7 & 3 & -2 \\ -1 & -5 & 3 \\ -8 & 0 & 1 \end{pmatrix}$$

$$= \begin{pmatrix} 5*7+3*(-1)+2*(-8) & 5*3+3*(-5)+2*0 & 5*(-2)+3*3+2*1 \\ 1*7+0*(-1)+(-1)*(-8) & 1*2+0*(-5)+(-1)*0 & 1*(-2)+0*3+(-1)*1 \\ 4*7+2*(-1)+0*(-8) & 4*3+2*(-5)+0*0 & 4*(-2)+2*3+0*1 \end{pmatrix}$$

$$= \begin{pmatrix} 16 & 0 & 1 \\ 15 & 3 & -3 \\ 26 & 2 & -2 \end{pmatrix}$$

$$b) A \cdot A^T = \begin{pmatrix} 5 & 3 & 2 \\ 1 & 0 & -1 \\ 4 & 2 & 0 \end{pmatrix} \cdot \begin{pmatrix} 5 & 1 & 4 \\ 3 & 0 & 2 \\ 2 & -1 & 0 \end{pmatrix} = \begin{pmatrix} 38 & 3 & 26 \\ 3 & 2 & 4 \\ 26 & 4 & 20 \end{pmatrix}$$

$$c) A^T \cdot A = \begin{pmatrix} 5 & 1 & 4 \\ 3 & 0 & 2 \\ 2 & -1 & 0 \end{pmatrix} \cdot \begin{pmatrix} 5 & 3 & 2 \\ 1 & 0 & -1 \\ 4 & 2 & 0 \end{pmatrix} = \begin{pmatrix} 42 & 23 & 9 \\ 23 & 13 & 6 \\ 9 & 6 & 5 \end{pmatrix}$$

2.

a)

$$A \cdot a$$

$$B \cdot a$$

$$C \cdot b, C \cdot d, C \cdot e, C \cdot f$$

$$D \cdot b, D \cdot d, D \cdot e, D \cdot f$$

$$E \cdot b, E \cdot d, E \cdot e, E \cdot f$$

$$F \cdot c$$

b)

$$A \cdot a = \begin{pmatrix} 1 \cdot 1 + (-1) \cdot 2 \\ 2 \cdot 1 + 0 \cdot 2 \end{pmatrix} = \begin{pmatrix} -1 \\ 2 \end{pmatrix}$$

$$B \cdot a = \begin{pmatrix} 0 \cdot 1 + 1 \cdot 2 \\ 2 \cdot 1 + 3 \cdot 2 \\ -4 \cdot 1 + 2 \cdot 2 \end{pmatrix} = \begin{pmatrix} 2 \\ 8 \\ 0 \end{pmatrix}$$

$$C \cdot b = \begin{pmatrix} 1 \cdot 1 + (-2) \cdot (-1) + 1 \cdot 3 \\ 2 \cdot 1 + 0 \cdot (-1) + 3 \cdot 3 \end{pmatrix} = \begin{pmatrix} 6 \\ 11 \end{pmatrix}$$

$$D \cdot b = \begin{pmatrix} 3 \cdot 1 + 0 \cdot -1 \cdot (-1) \cdot 3 \\ 2 \cdot 1 + 4 \cdot -1 + (-2) \cdot 3 \\ -1 \cdot 1 + 0 \cdot -1 + 2 \cdot 3 \end{pmatrix} = \begin{pmatrix} 0 \\ -8 \\ 5 \end{pmatrix}$$

$$E \cdot b = b$$

c)

$$A = \begin{pmatrix} 2 & 1 \\ 4 & -1 \end{pmatrix}$$

1. Schritt: Berechnung der Determinante der Matrix A

$$\det(A) = (2 \cdot -1) - (4 \cdot 1) = -6$$

$$\text{inverse Matrize: } \begin{pmatrix} -1/6 & -1 \\ -4 & 2 \end{pmatrix} = \begin{pmatrix} 1/6 & 1/6 \\ 2/3 & -1/3 \end{pmatrix}$$

$$B = \begin{pmatrix} 1 & 2 & 3 \\ 2 & -4 & -2 \\ 3 & -2 & 1 \end{pmatrix}$$

Diese Matrix ist nicht invertierbar, weil:

- Die Determinante von Matrix B ist gleich 0

- Die Zeilenvektoren sind linear abhängig. Der dritte Zeilenvektor (3,-2,1) ist die Summe der

beiden ersten $(1,2,3)$ und $(2,-4,-2)$, d.h. Die Summe der beiden ersten minus den dritten Vektor ergibt den Nullvektor.