

Aufgabe 1

Wort	d_1	d_2
der	0	1
die	2	1
flieht	0	1
jagt	1	0
katze	1	1
maus	1	1
vor	0	1

$$\text{dist}(d_1, d_2) = \arccos \frac{d_1 \cdot d_2}{|d_1| \cdot |d_2|} = \arccos \frac{4}{\sqrt{7} \cdot \sqrt{6}} = 51.89^\circ$$

Aufgabe 2

Wort	d_1	d_2
distanz	1	1
dokumente	1	0
er	0	1
hatte	0	1
keine	0	1
seinem	0	1
text	1	1
zu	0	1

$$\text{dist}(d_1, d_2) = \arccos \frac{d_1 \cdot d_2}{|d_1| \cdot |d_2|} = \arccos \frac{2}{\sqrt{3} \cdot \sqrt{7}} = 64.12^\circ$$

Wort	d_1	d_3
den	0	1
distanz	1	0
dokumente	1	1
er	0	1
für	0	1
im	0	1
ordner	0	1
speichert	0	1
text	1	1

$$\text{dist}(d_1, d_3) = \arccos \frac{d_1 \cdot d_3}{|d_1| \cdot |d_3|} = \arccos \frac{2}{\sqrt{3} \cdot \sqrt{8}} = 65.91^\circ$$

Dokument 1 liegt näher bei den Suchbegriffen als Dokument 2.

Aufgabe 3

Sind $\vec{a}$, $\vec{b}$ die ursprünglichen Dokumentvektoren und $\vec{a'}$ der Vektor des Dokuments mit den verdoppelten Wörtern, so gilt $\vec{a'} = 2\vec{a}$ und damit

$$\frac{\vec{a'} \cdot \vec{b}}{|\vec{a'}| \cdot |\vec{b}|} = \frac{(2\vec{a}) \cdot \vec{b}}{|2\vec{a}| \cdot |\vec{b}|} = \frac{2 \cdot \vec{a} \cdot \vec{b}}{2 \cdot |\vec{a}| \cdot |\vec{b}|} = \frac{\vec{a} \cdot \vec{b}}{|\vec{a}| \cdot |\vec{b}|}$$

Die Dokumentdistanz bleibt gleich.

Aufgabe 4

Wort	d_1	d_2	$d_1 \cdot d_2$	d_1^2	d_2^2
and	1	1	1	1	1
as	0	2	0	0	4
by	0	1	0	0	1
chance	0	1	0	0	1
choose	0	2	0	0	4
faire	2	1	2	4	1
foule	2	0	0	4	0
is	2	0	0	4	0
not	0	1	0	0	1
that	0	1	0	0	1
the	0	1	0	0	1
true	0	1	0	0	1
view	0	1	0	0	1
you	0	1	0	0	1
Summe			3	13	18

$$\text{dist}(d_1, d_2) = \arccos \frac{d_1 \cdot d_2}{|d_1| \cdot |d_2|} = \arccos \frac{3}{\sqrt{13} \cdot \sqrt{18}} = 78.69^\circ$$