

Gross- und Kleinschreibung sowie Satzzeichen sind nicht zu berücksichtigen.

### 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

Die Dokumentdistanz bleibt gleich.

#### **Aufgabe 4**

- Auffinden von Dokumenten (Datenbanken, Internet)
- Erkennung von Plagiaten
- automatische Klassifikation von Dokumenten