

Gross- und Kleinschreibung sowie Satzzeichen sind nicht zu berücksichtigen.

### Aufgabe 1

Bestimme die Dokumentdistanz der Texte:

- Die Katze jagt die Maus
- Die Maus flieht vor der Katze

### Aufgabe 2

Jemand sucht nach einem Dokument, das möglichst gut mit den Begriffen

Dokumente, Distanz, Text

übereinstimmt. Welches der folgenden Dokumente müsste eine Suchmaschine zuerst präsentieren, wenn sie den Dokumentwinkel als Ähnlichkeitsmass verwendet?

- (a) Er hatte keine Distanz zu seinem Text.
- (b) Den Text speichert er im Ordner für Dokumente.

### Aufgabe 3

Wie verändert sich die Dokumentdistanz zwischen zwei Dokumenten, wenn man in einem der Dokumente willkürlich jedes Wort verdoppelt?

- (a) Die Dokumentdistanz wird grösser.
- (b) Die Dokumentdistanz wird kleiner.
- (c) Die Dokumentdistanz bleibt gleich.

### Aufgabe 4

Bestimme die Dokumentdistanz der beiden Shakespeare-Zitate:

- Faire is foule, and foule is faire  
– *Macbeth*, 1. Akt, 1. Szene / *Hexen*
- You that choose not by the view Chance as faire, and choose as true.  
– *Der Kaufmann von Venedig*, 3. Akt, 2. Szene / *Bassanio*

Wie üblich, sind Gross- und Kleinschreibung sowie Interpunktion nicht zu berücksichtigen.

Gross- und Kleinschreibung sowie Satzzeichen sind nicht zu berücksichtigen.

**Aufgabe 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**

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

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

Dokument 1 liegt näher bei den Suchbegriffen als Dokument 2.

**Aufgabe 3**

Die Dokumentdistanz bleibt gleich.

**Aufgabe 4**

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