

**Einführung in IR - Winter 2006/07**

Dipl.-Inform. Sascha Kriewel, LF 137

sascha.kriewel@uni-due.de

**Übungsblatt 7      Nicht-Probabilistische Retrievalmodelle  
keine Abgabe**

---

**Aufgabe 14: Vektorraum-Modell**

(a) Es seien folgende Dokumenten-Vektoren gegeben:

$$d_1 = (0, 0, 0, 1, 2, 3)$$

$$d_2 = (0, 0, 0, 0, 4, 0)$$

$$d_3 = (1, 3, 0, 1, 4, 1)$$

$$d_4 = (0, 0, 4, 1, 5, 1)$$

Und dazu der Fragevektor  $q = (0, 1, 0, 1, 1, 2)$ . Berechne die Retrievalgewichte der Dokumente  $d_1$  bis  $d_4$  bezüglich der Anfrage einmal mit Hilfe des Skalarprodukts und einmal über das Kosinusmaß. Zur Erinnerung: das Kosinusmaß ist gegeben durch

$$\cos(x, y) = \frac{\langle x, y \rangle}{\|x\| \cdot \|y\|}$$

Finde heraus, welche anderen Ähnlichkeitsmaße für Vektoren es gibt.

(b) Das (bescheidene) Indexierungsvokabular sei folgende Termmenge  $T = \{\text{computer, system, information, data, user, interface}\}$ . Die Dokumentmenge enthalte vier Dokumente mit den Repräsentationen:

$$d_1 = \{\text{computer, system, information, system, user}\}$$

$$d_2 = \{\text{computer, interface, user, interface, user}\}$$

$$d_3 = \{\text{information, data, user}\}$$

$$d_4 = \{\text{system, data, information, system, user}\}$$

Bestimme die  $tf \cdot idf$  Gewichte nach der in der Vorlesung beschriebenen Methode. Wähle für den Logarithmus die Basis 2 (jede Basis ist im Prinzip zulässig).

Die Anfrage bestehe nun aus den Termen 'computer', 'user' und 'system'. Wie sehen Dokumentenvektoren und Anfragevektor aus? Bestimme die Retrievalwerte.