

# Vorlesung Information Retrieval

## Übungsblatt 4 (Aufgaben 11-12)

**Ausgabe:** 18.05.2009 (in der Vorlesung)  
**Abgabe:** 25.05.2009 (in der Vorlesung)  
**Besprechung:** 08.06.2009 (in der Übungsstunde)



Sommersemester 2009

Wir betrachten den Textkorpora aus Aufgabe 4.

**d1** : *Marcus tried to assassinate Caesar.*

**d2** : *Marcus was a Roman.*

**d3** : *Caesar was a ruler. All Romans were either loyal to Caesar or hated him.*

**d4** : *Everyone is loyal to someone. People only try to assassinate rulers they are not loyal to.*

**t1** : *Marcus*

**t2** : *try*

**t3** : *assain*

**t4** : *Caesar*

**t5** : *Rome*

**t6** : *rule*

**t7** : *loyal*

**t8** : *hate*

**t9** : *people*

### Aufgabe 11:

Das Relevance-Feedback ergab, dass Dokument d2 und Dokument d3 relevant sind.

- Berechne  $p_i$  und  $q_i$  für die Terme t1 bis t9 mit Smoothingfaktor  $\lambda = 0,5$ .
- Berechne die Ähnlichkeit zwischen der Anfrage  $q = (0 \ 1 \ 1 \ 0 \ 0 \ 0 \ 0 \ 1)$  und Dokument d5 = „Marcus tried to rule the people of Rome“

### Aufgabe 12:

Es ergibt sich folgende Term-Dokument-Matrix.

$$M = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 2 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 2 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

Mit SVD erhalten wir:

$$U = \begin{pmatrix} 0.1261 & -0.1237 & 0.6739 & 0.4013 & -0.5901 & -0.0104 & -0.0509 & 0.0302 & -0.0406 \\ 0.2856 & 0.2940 & 0.3922 & -0.2046 & 0.3772 & -0.3165 & -0.5705 & -0.0624 & -0.2541 \\ 0.2856 & 0.2940 & 0.3922 & -0.2046 & 0.2692 & 0.4262 & 0.5056 & 0.3468 & 0.0794 \\ 0.4818 & -0.5959 & 0.0745 & -0.4853 & -0.0562 & -0.0994 & 0.1159 & -0.3146 & 0.2152 \\ 0.2208 & -0.3834 & 0.0132 & 0.6712 & 0.5901 & 0.0104 & 0.0509 & -0.0302 & 0.0406 \\ 0.3803 & 0.0343 & -0.2686 & 0.0653 & -0.1592 & 0.7294 & -0.4276 & -0.1136 & -0.1570 \\ 0.5684 & 0.3538 & -0.3418 & 0.2024 & -0.2155 & -0.3699 & 0.3550 & -0.0948 & -0.2751 \\ 0.1922 & -0.2852 & -0.1954 & -0.0718 & -0.1030 & -0.1712 & -0.2101 & 0.8677 & -0.0389 \\ 0.1881 & 0.3195 & -0.0732 & 0.1371 & -0.0562 & -0.0994 & -0.2175 & 0.0188 & 0.8819 \end{pmatrix}$$

$$S = \begin{pmatrix} 3.4785 & 0 & 0 & 0 \\ 0 & 2.2721 & 0 & 0 \\ 0 & 0 & 1.8149 & 0 \\ 0 & 0 & 0 & 1.2014 \end{pmatrix}$$

$$V = \begin{pmatrix} 0.3390 & -0.0579 & 0.8445 & -0.4105 \\ 0.0997 & -0.2232 & 0.3786 & 0.8927 \\ 0.6685 & -0.6480 & -0.3547 & -0.0863 \\ 0.6544 & 0.7259 & -0.1328 & 0.1647 \end{pmatrix}$$

Beantworte die folgenden Anfragen mit Hilfe der zweiwertigen Singulärwertzerlegung

**q1:** *Who assassinated Caesar*

**q2:** *Loyalty and assassination*

Berechnen Sie die sortierten Trefferlisten für die beiden Anfragen gemäß der Cosinus-Ähnlichkeit.