**Indexierung von D und Ausführung von Queries q:**

Als Index zu verwalten ist die Dokument−Thema−Ähnlichkeitsmatrix U sowie − quasi als Domänenwissen − die Feature−Thema−Ähnlichkeitsmatrix V. Letztere kann man weitgehend statisch berechnen, sobald man eine repräsentative, große Dokumentenkollektion hat. Ein neu eingefügtes Dokument d (d.h. ein 1xm−Zeilenvektor) muß dann zur Indexierung in den Themen−Vektorraum abgebildet werden, und zwar durch die Transformation d' = d x V, und der resultierende 1xr−Zeilenvektor d' wird zum Index hinzugefügt (d.h. konzeptionell wird U um eine Zeile erweitert).

Eine Query q, die als mx1−Spaltenvektor aufgefaßt werden kann, wird dann zunächst in eine Query q' in den Themen−Vektorraum transformiert, und zwar durch $q' = V^T \times q$ . Dann wird q' aufgrund des Index U evaluiert (mit der verwendeten Ähnlichkeitsfunktion des r−dimensionalen Themen−Vektorraums), und die dadurch erzeugte Rangliste von Dokumenten ist das Anfrageresultat. Im einfachsten Fall würde z.B. das Cosinus−Maß oder die Euklidische Distanz zwischen q' und den Zeilen von U (also den Dokumenten) verwendet.

**Beispiel:**

Die Dokument−Feature−Ähnlichkeitsmatrix A sei wie folgt gegeben (der Einfachheit halber seien hier auch Gewichte größer als 1 möglich):

| Features / Dokumente | Lafontaine | Schröder | Euro | Uefa | Beckenbauer |
|---|---|---|---|---|---|
| d1 | 1 | 1 | 1 | 0 | 0 |
| d2 | 2 | 2 | 2 | 0 | 0 |
| d3 | 1 | 1 | 1 | 0 | 0 |
| d4 | 5 | 5 | 5 | 0 | 0 |
| d5 | 0 | 0 | 0 | 2 | 2 |
| d6 | 0 | 0 | 0 | 3 | 3 |
| d7 | 0 | 0 | 0 | 1 | 1 |

Die SVD ermittelt daraus:

$$U = \begin{pmatrix} 0.18 & 0.00 \\ 0.36 & 0.00 \\ 0.18 & 0.00 \\ 0.90 & 0.00 \\ 0.00 & 0.53 \\ 0.00 & 0.80 \\ 0.00 & 0.27 \end{pmatrix} \qquad \Delta = \begin{pmatrix} 9.64 & 0.00 \\ 0.00 & 5.29 \end{pmatrix} \qquad V^T = \begin{pmatrix} 0.58 & 0.58 & 0.58 & 0.00 & 0.00 \\ 0.00 & 0.00 & 0.00 & 0.71 & 0.71 \end{pmatrix}$$

Intuitiv enthält die Dokumentenkollektion also zwei Themen, nämlich Politik und Sport. Ein weiteres Dokument d8 = ( 1 1 0 0 0 ) mit dem einzigen Feature "Lafontaine" würde dann auf den Themenvektor d8 x V= (0.54 0).
Eine Query $q = ( 0\,0\,1\,0\,0 )^T$, die nach dem Feature "Euro" sucht, würde auf den Themenvektor $V^T \times q = (0.58 \;\; 0)^T$ abgebildet. Unabhängig vom wirklich verwendeten Ähnlichkeitsmaß im Themenvektorraum wäre also d8 vermutlich ein sehr guter Treffer für q.