

Sprachtechnologie

7. Übung

Prof. Tobias Scheffer
Uwe Dick

Sommer 2015

Ausgabe am: 08.06.15
Besprechung am: 15.06.15

Aufgabe 1

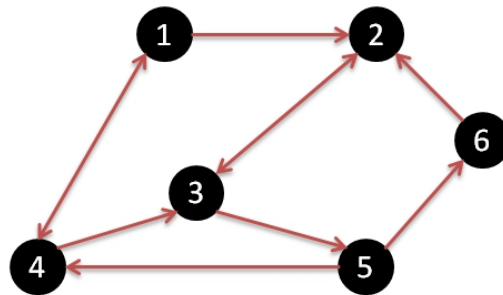
Edit-Distanz

Bestimmen Sie mit dem in der Vorlesung vorgestellten Algorithmus die Edit-Distanz der Wörter *abloratu* und *abitur* (Vorlesung Indexieren und Suchen, S.42). Geben Sie die Transformationsregeln an, um ausgehend von dem ersten Wort, das zweite Wort mit möglichst wenigen Transformationen zu erzeugen.

Aufgabe 2

HITS

Angenommen, Sie haben eine Linkstruktur zwischen Dokumenten (1 bis 6) bestimmt (siehe Abbildung), wobei alle Kantengewichte auf 1 festgelegt sind. Bestimmen Sie mit Hilfe des HITS-Algorithmus aus der Vorlesung das relevanteste Dokument. Dabei enthalten nur die Dokumente 1, 3, 4, 5 relevante Schlüsselwörter. Führen sie dafür die Authority- und Hub-Update-Regeln jeweils $k = 5$ mal aus.



Aufgabe 3

PageRank

Bestimmen Sie mit dem PageRank Algorithmus die Scores aller Dokumente für den Graphen aus Aufgabe 2. Welches Dokument erhält den höchsten Ranking-Score? Hinweis: Am leichtesten geht es mit einem kleinen Skript für Matlab oder Octave oder mit der numpy-Library für Python, z.B. mit der Anaconda-Distribution (<https://store.continuum.io/cshop/anaconda/>).