

Sistemi Informativi

Laboratorio 7

Catalin Copil

Mattia de Stefani

Giulio Lovisotto

May 14, 2015

1 Descrizione

La nostra funzione di reperimento utilizzerà i primi N documenti prodotti da BM25 per costruire il grafo delle citazioni R_q , tale grafo verrà espanso per ottenere un grafo allargato B_q , e su quest'ultimo verrà calcolato HITS. La funzione combinerà poi gli score di BM25 (sc) con i punteggi di authority ($auth$) e hubbiness (hub) per i primi N documenti, e li riordinerà per $hits_{score}$ decrescente:

$$hits_{score} = \alpha \cdot sc + \beta \cdot auth + \gamma \cdot hub,$$

α varia in $[0, 1]$, mentre β, γ sono parametri che possono variare tra $[-1, 1]$, in quanto vogliamo cogliere possibili influenze negative sul di tali valori. Nella precedente funzione i valori di $sc, auth, hub$ vengono normalizzati in $[0, 1]$ prima del calcolo.

Vogliamo provare diverse combinazioni di tali valori per ottenere la massima precisione.