

Sistemi Informativi

Laboratorio 5

Catalin Copil

Mattia de Stefani

Giulio Lovisotto

April 30, 2015

1 Descrizione

Computeremo il PAGERANK con la libreria `networkx` a tempo di indexing e salveremo i risultati su un file ($id \rightarrow pagerank$). La nostra funzione di reperimento combinerà gli score di BM25 (rk) con pagerank (pr) nel seguente modo:

$$score = \alpha \cdot \ln(rk + 1) + (1 - \alpha) \cdot pr,$$

dove α e' un parametro tra 0 e 1 che determina l'importanza del ranking (primo termine) e del pagerank (secondo termine). E' stato introdotto tale parametro per evitare che il valore del pagerank influisse troppo sul punteggio finale. Sperimentalmente proveremo diversi valori per α e sceglieremo quello che garantisce la massima *map*. Abbiamo scelto di usare il logaritmo naturale per ridimensionare l'ordine di grandezza dello score di BM25 (i cui score valgono anche piu di 10), e di sommare 1 per evitare valori negativi del logaritmo.