

Basi di dati multimediali

Fase 3

Description: In questa fase del progetto approfondiremo aspetti relativi a

- Analisi di grafi,
- interrogazione, e
- relevance feedback.

Il progetto si baserà sui dati e sui risultati delle fasi precedenti.

- **Task 1: Creazione del grafo dei coautori:** Create un grafo dei coautori pesato in cui ci sia un arco tra due autori se sono stati coautori di almeno un articolo, il cui peso rappresenti la similarit' dei corrispondenti autori, misurata in termini di
 - keyword-vectors o
 - top 3 semantiche latenti degli autori (PCA e SVD).
- **Task 2: Creazione del grafo degli articoli di cui si e' stati coautori:** Create un grafo degli articoli di cui si e' stati coautori, in cui esiste un arco tra due articoli se condividono almeno un autore, e in cui i pesi rappresentano la similarità tra i corrispondenti articoli misurata in termini di
 - TF keyword-vectors, o
 - TF-IDF keyword-vectors.
- **Task 3: Partizionamento del grafo:** Dato un grafo dei "co-autori" o degli "articoli di cui si e' stati co-autori", identificare K cluster (con K dato in input), usando due algoritmi (a scelta) diversi. Visualizzare i cluster risultanti.
- **Task 4: Analisi del grafo:** Dato un grafo dei "co-autori" o degli "articoli di cui si e' stati co-autori", individuare ed elencare i K nodi maggiormente dominanti utilizzando l'algoritmo Page Rank (PR) (per un valore K dato in input).
Si veda
 - "S. Brin and L. Page. "The anatomy of a large-scale hypertextual Web search engine". Computer Networks and ISDN Systems 30: 107 -117, 1998"
- **Task 5: Ricerca all'interno del grafo (contenuto):** Dato un grafo dei "co-autori" o degli "articoli di cui si e' stati co-autori", e dato un autore, o l'identificatore di un articolo, individuare ed elencare i K nodi più simili dal punto di vista del contenuto (per un valore K dato in input).
- **Task 6: Ricerca all'interno del grafo (relazioni strutturali):** Dato un grafo dei "co-autori" o degli "articoli di cui si e' stati co-autori", e dato un autore, o l'identificatore di un articolo, individuare ed elencare i K nodi più simili dal punto di vista strutturale, utilizzando il Personalized PageRank (PPR) (per un valore K dato in input).

Si veda

- “J.-Y. Pan, H.-J. Yang, C. Faloutsos, and P. Duygulu. Automatic multimedia cross-modal correlation discovery. In KDD, pages 653 -658, 2004”

per una formulazione del personalized PageRank basata sul RandomWalks con re-start.

- **Task 7: Relevance feedback (contenuto):** Implementare un algoritmo di relevance feedback probabilistico per modificare l'importanza delle chiavi per migliorare l'accuratezza dei match rispetto al contenuto. Il sistema dovrebbe anche restituire in output le revisioni suggerite nell'importanza relativa delle chiavi nella query.

Si veda

- Gerard Salton and Chris Buckley. Improving retrieval performance by relevance feedback. Journal of the American Society for Information Science. 41, pp. 288-297, 1990.

- **Task 8: Interfaccia per le query:**

- Implementare un'interfaccia per l'interrogazione, che consenta all'utente di specificare un task tra quelli offerti dal sistema, i parametri rilevanti per quel task (tra cui il numero di risultati richiesti in output). Fare in modo che i risultati vengano presentati all'utente per ordine di matching decrescente.
- L'interfaccia deve anche consentire all'utente di esprimere il suo feedback positivo e/o negativo relativamente ai risultati ordinati restituiti dal sistema.
- Il feedback dell'utente sarà quindi tenuto in conto (rivedendo la query, oppure modificando opportunamente l'ordinamento dei risultati) e restituirà un nuovo insieme ordinato di risultati.

Consegnare:

- Il codice (opportunamente commentato) e un file README .
- L'output restituito a fronte di alcuni esempi di input.
- Una breve relazione che descriva il lavoro svolto e i risultati.