

# Basi di Dati Multimediali

## Fase 2 della sperimentazione di approfondimento

**Descrizione:** Questa fase della sperimentazione e' dedicata ad aspetti relativi a

- riduzione della dimensionalita' dello spazio ed
- grafi sociali.

Ci baseremo sui risultati prodotti nella fase 1, utilizzando gli stessi dati.

- **Task 1: Operazioni di ricerca**

- **Task 1a:** Scrivere un programma che, dato un utente, identifichi e restituisca le top-5 semantiche latenti (topics) utilizzando,

- \* PCA (nello spazio TF-IDF), e
- \* SVD (nello spazio TF-IDF)

Si consiglia di utilizzare i packages di Matlab per svolgere questo Task.

- **Task 1b:** Scrivere un programma che, dato un utente, identifichi (e restituisca in ordine di similarita') i 10 autori piu' simili, confrontando i rispettivi

- \* keyword vectors,
- \* vettori di differenziazione (TF-IDF2 o PF), e
- \* le top-5 semantiche latenti (PCA, SVD).

- **Task 1c:** Scrivere un programma che, data un autore, trovi e restituisca in ordine di rilevanza gli articoli per lui piu' rilevanti (di cui lui non e' autore), basandosi sulla similarita' tra i seguenti vettori relativi all'utente dato

- \* keyword vector,
- \* differentiation vector (TF-IDF2 or PF), e
- \* top-5 semantiche latenti (PCA, o SVD)

ed i document vector degli articoli (TF-IDF).

- **Task 2: Analisi della rete degli autori**

- **Task 2a:** Scrivere un programma che

1. crei una matrix di similarita' *author-author* (utilizzando i keyword vectors),
2. effettui la trasformazione SVD su questa matrice di similarita' *author-author* , e
3. restituisca le top-3 semantiche latenti (in termini di appartenenza dei autori) sottostanti questa matrice di similarita' *author-author*.

- **Task 2b:** Scrivere un programma che operi sulla rete dei coautori, e

1. crei una matrice di similarita' *coauthor-coauthor* (utilizzando i keyword vectors),
2. effettui la trasformazione SVD su questa matrice di similarita' *coauthor-coauthor* , e

3. estituisca le top-3 semantiche latenti (in termini di appartenenza dei autori) sottostanti questa matrice di similarita' *coauthor-coauthor*.

- **Task 3: Analisi delle comunita'**

- **Task 3a:** Scrivere un programma che consideri le top-3 semantiche latenti (gruppi) della matrice di similarita' *coauthor-coauthor* (o *author-author*) e suddivida gli autori in 3 gruppo sulla base del loro grado di appartenenza alle 3 semantiche
- **Task 3b:** Scrivere un programma che consideri le top-3 semantiche latenti (gruppi) della matrice di similarita' *coauthor-coauthor* (o *author-author*) e associ un keyword vector a ciascuna semantica/gruppo sulla base dei keyword-vectors degli autori e del loro grado di appartenenza alla corrispondente semantica latente/gruppo.