

Ingegneria dei dati 2025/2026

Homework 2

(da svolgere individualmente)

Paolo Merialdo

Homework 2

1. scrivere un programma che indicizza i file .txt contenuti in una directory del proprio laptop. In particolare, si devono considerare due campi (e quindi creare due indici): il nome del file, il contenuto del file. Per ciascun campo utilizzare un analyzer appropriato
2. scrivere un programma che legge una query da console, interroga l'indice e stampa il risultato. Usare una semplice sintassi per la query (ad esempio, una query inizia con la parola chiave nome o contenuto seguita da una sequenza di termini (eventualmente racchiusi tra virgolette per esprimere una phrase query))
3. testare il sistema con una decina di query diverse

Scrivere una relazione che, oltre a riportare l'url del proprio progetto su Github (o analogo) descriva:

- gli analyzer che si è scelto di utilizzare (motivando le scelte)
- il numero di file indicizzati e i tempi di indicizzazione
- le query usate per testare il sistema

Homework 2

Termini di consegna: inviare la relazione entro le ore 18:00 del 3 novembre 2025 attraverso il seguente modulo online:

<https://forms.office.com/e/EbqbR7dvK4>