

Luca Maria Lauricella  
[https://github.com/lauriluca99/ID\\_homework2](https://github.com/lauriluca99/ID_homework2)

## **INDICIZZAZIONE E ANALYZER:**

Nell'indicizzazione di file vengono utilizzati diversi criteri quindi analyzer diversi, infatti in base a ciò che si vuole indicizzare Lucene ci mette a disposizione una serie di analyzer adatti a più situazioni. Ogni tipo di file richiede una diversa analisi dei termini (es. Case Sensitive, White Space) che varia di importanza a seconda dell'estensione. Nel nostro caso è richiesta particolare attenzione e distinzione per quanto riguarda il contenuto e il nome del file .txt. Nell'indicizzazione del nomeFile è stato utilizzato un analyzer combinato da uno StopAnalyzer e delle stopWord dove lo StopAnalyzer rimuove i segni di punteggiatura e una serie di terms poco rilevanti nella tokenizzazione invece lo StopWord ci permette di inserire una serie di parole da scartare come ad esempio l'estensione .txt che hanno i nomi dei file.

Per quanto riguarda il contenuto del file è stato utilizzato uno StandardAnalyzer offerto da Lucene che divide il testo in token utilizzando gli spazi bianchi e i segni di punteggiatura come delimitatori.

## **TEMPI DI INDICIZZAZIONE:**

all'interno di ogni cartella utilizzata ci sono 11 file di testo.

Indicizzazione 1 → 47 millisecondi  
Indicizzazione 2 → 14 millisecondi  
Indicizzazione 3 → 11 millisecondi  
Indicizzazione 4 → 11 millisecondi  
Indicizzazione 5 → 07 millisecondi

## **QUERY:**

Nelle seguenti query si noti come quando si effettua una ricerca va specificato il campo sul quale si vogliono contare gli hit: nome: / contenuto:

Di default se non viene specificato la ricerca verrà in automatico eseguita sul contenuto dei file.

Dato che nel codice viene utilizzata la QueryParser è possibile utilizzare notazioni come: -, +, \*, "".

### **Query 1:**

Catulle

Trovati 2 hits.

Riportati i primi 2 risultati

1. file testo/provaa.txt score=0.79805595

2. file testo/file1 copia 9.txt score=0.7357717

Enter the search query (q=quit):

nome: [nome del file] --> per cercare il nome file

contenuto: [contenuto del file] --> per cercare all'interno del file

**Query 2:**

contenuto: Catulle

Trovati 2 hits.

Riportati i primi 2 risultati

1. file testo/provaa.txt score=0.79805595

2. file testo/file1 copia 9.txt score=0.7357717

Enter the search query (q=quit):

nome: [nome del file] --> per cercare il nome file

contenuto: [contenuto del file] --> per cercare all'interno del file

**Query 3:**

nome: aaa

Trovati 1 hits.

Riportati i primi 1 risultati

1. file testo/aaa.txt score=1.1584632

Enter the search query (q=quit):

nome: [nome del file] --> per cercare il nome file

contenuto: [contenuto del file] --> per cercare all'interno del file

**Query 4:**

Quis nunc te adibit prova

Trovati 5 hits.

Riportati i primi 5 risultati

1. file testo/file1 copia 2.txt score=3.2328024

2. file testo/file1 copia.txt score=1.3436027

3. file testo/file1 copia 3.txt score=0.6545763

4. file testo/file1 copia 5.txt score=0.55624086

5. file testo/file1 copia 6.txt score=0.39362442

Enter the search query (q=quit):

nome: [nome del file] --> per cercare il nome file

contenuto: [contenuto del file] --> per cercare all'interno del file

**Query 5:**

"Quis nunc te adibit"

Trovati 1 hits.

Riportati i primi 1 risultati

1. file testo/file1 copia 2.txt score=3.2328024

Enter the search query (q=quit):

nome: [nome del file] --> per cercare il nome file

contenuto: [contenuto del file] --> per cercare all'interno del file

**Query 6:**

quae - puella

Trovati 5 hits.

Riportati i primi 5 risultati

1. file testo/file1 copia 8.txt score=1.0722127
2. file testo/file1 copia 3.txt score=0.6545763
3. file testo/file1 copia 5.txt score=0.55624086
4. file testo/aaa.txt score=0.48359197
5. file testo/file1 copia 6.txt score=0.39362442

Enter the search query (q=quit):

nome: [nome del file] --> per cercare il nome file

contenuto: [contenuto del file] --> per cercare all'interno del file

### **Query 7:**

no\*

Trovati 3 hits.

Riportati i primi 3 risultati

1. file testo/file1 copia 8.txt score=1.0
2. file testo/aaa.txt score=1.0
3. file testo/file1 copia 6.txt score=1.0

Enter the search query (q=quit):

nome: [nome del file] --> per cercare il nome file

contenuto: [contenuto del file] --> per cercare all'interno del file

### **Query 8:**

luca

Trovati 0 hits.

Riportati i primi 0 risultati

Enter the search query (q=quit):

nome: [nome del file] --> per cercare il nome file

contenuto: [contenuto del file] --> per cercare all'interno del file