

## Homework 2

Riccardo Marasca

533971

<https://github.com/ricciric/University.git>

### Analyzer utilizzati

Come analyzer ho scelto di utilizzare lo *StandardAnalyzer* sia per il contenuto che per il titolo, ho aggiunto come parametro dell'analyzer la variabile da me creata *STOP WORDS*, contenente le seguenti stopwords, che ho ricavato direttamente dal codec:

"in", "dei", "di", "un", "una", "e", "il", "la", "in", "su", "di", "per", "con", "della", "le", "loro", "dà", "a", "che", "del", "era", "mentre".

Precedentemente avevo utilizzato un *ItalicAnalyzer*, purtroppo non si è rivelata una buona scelta, poiché mentre rileggevo il codec, ho notato che questo creava problemi con il tokenizzatore, rimuovendo a volte la vocale finale della parola ('a', 'e') contenuta nelle stop words dell'*ItalicAnalyzer*. Ho quindi riconsiderato la scelta di utilizzare l'analyzer standard. Ho scelto quest'analyzer sia per il titolo che per il contenuto, poiché nei testi che ho scelto, che riporterò successivamente, all'interno del titolo è presente una stopwords, non volevo quindi che la ricerca fosse condizionata da essa.

### File e tempi di indicizzazione

I file indicizzati sono 3:

- *Storia della Palestina*
- *Cenere e Promessa, Una storia di Rinascita*
- *Il Trono Divino, La Statua di Zeus a Olimpia*

l'indice per tutti e tre i documenti è stato creato in 589 ms.

### Query utilizzate

Riporto l'elenco delle query utilizzate:

- *Query Parser (contenuto: +storia di +israele)*
- *Term Query (contenuto: persone)*
- *Phrase Query.Builder() (contenuto: opere + contenuto: d'arte)*
- *Phrase Query\_2 (contenuto: pace, terra)*
- *Phrase Query\_3 (contenuto: fama, statua)*
- *Phrase Query.Builder\_2() (contenuto: lieto + contenuto: fine)*
- *Query Parser\_2 (contenuto: confini di +olympia)*
- *Query Parser\_3 (contenuto: giorni +antichi)*
- *Query Parser\_4 (contenuto: +arrende mai)*
- *Query Parser\_5 (contenuto: +storia statua nazione palestina)*