HOMEWORK 2

Davide Moliterno

Link github: <https://github.com/chila99/IDDHomework2>

Il Progetto ha lo scopo di indicizzare i documenti relativi ai testi di canzoni di differenti artisti selezionati e quindi di poter successivamente effettuare delle ricerche attraverso specifiche query per effettuare il retrieval di tali file. Nello specifico è composto da due classi diverse chiamate Indexing e Searching che permettono di effettuare le due operazioni sopracitate. Nello specifico la classe di Indexing, a partire dal path della directory in cui sono salvati i vari file, si avvale di due tipi di Analyzer diversi: uno WhiteSpaceAnalyzer e uno StandardAnalyzer con una serie di stopwords delle lingue italiano, inglese e spagnolo per il campo “contenuto”. Questo per far in modo che il titolo possa essere cercato anche attraverso phrase query molto specifiche mentre per motivi spazio-temporali si è scelto di non indirizzare determinate parole molto ricorrenti all’interno dei vari testi e far in modo che le query fossero case-insensitive. Il tempo impiegato dal sistema per indicizzare l’intera directory contenente circa 500 file testuali di testi di canzoni è mediamente pari a 1.3s. Dall’altra parte la classe di searching permette in loop(fino a che non si inserisce il valore 0) di effettuare query di diverso tipo ottenendo vari risultati. Nello specifico verranno ritornati sempre i primi 10 risultati nel ranking. Le possibili query sono le seguenti: “campo: query” e “query”. La prima cercherà esclusivamente nel campo indicato mentre la seconda effettuerà una ricerca in entrambi i campi. Infine è anche possibili fornire al sistema una phrasequery inserendo tra apici(“”) la frase da voler cercare. Nota: quando si cerca per il titolo è stato usato un analyzer speciale che rende maiuscole tutte le iniziali delle parole della query.

Esempi di query con risultati: