IR Assignment #5 (Last)

Apache Lucene is a Java library used for full text search of documents. In this assignment, you will use Lucene to build a document index and perform search on that index.

Step 1: Download Lucene, watch Lucene videos online and refer on line examples for building index and search modules.

Step 2: You are provided a movie data file which consists of movie title, year, producer and other details. You can use the script provided to process the movie file and convert each movie into individual files on a separate directory..

Step 3: Write a lucene Indexer that will read the movie files and construct an inverted index for the movie documents. Use the standard analyzer provided by Lucene to preprocess the documents .

Step 4: Write a Search using Lucene that takes your query and the index files and displays the search results.

Sample queries are given. You should run Boolean, boosted query, phrase search, regular, fuzzy and wildcard search.

Submit the source code and the output of various queries.