

IR CS F469 Assignment Design Document

By

Pavan Kumar Reddy Yannam 2019A7PS0038H B.V.S. Ruthvik 2019A7PS0017H P.V. Sri Harsha 2019A2PS1521H

Application Architecture

Query Input	Process Input Query and Retrieve functions.
Preprocessing	Removing stop words Removing non alphanumeric characters Stemming or Lemmatization Get closest words using edit distance.
Creating Inverted Index	•Generating words present in each documents •Generating an inverted index containing information on the documents in which each word appears.
Boolean Operations	•AND •OR •NOT

Data Structure

Inverted Index - Implemented using Dictionary, Sets and Lists

Running time

Preprocessing = 21.34 sec Building index = 0.88 sec Search/retrieval = 0.43 sec