

ReadMe

Information Retrieval Assignment 1

Submitted By:

Himanshu Aggarwal
MT17015

Repository of this Assignment 1 contains two python code file namely, **generator.py** and **main.py**, and three JSON files, containing processed data for the code.

generator.py:

This code generates tokens from the given dataset, preprocess it, construct an inverted index using default dictionary, and save it as JSON file for ease of usage in future.

main.py

This is the main file for running and evaluating this assignment. It contains code for evaluating queries in the form CLI of the following forms:

x AND y
x OR y
x AND NOT y
x OR NOT y
x ANDS y
x ANDN y

First four commands evaluate the query using merge postings algorithm, implemented using set operations for better performance.

x ANDS y, evaluates AND query using Skip pointers algorithm.

x ANDN y, evaluates AND query using iteration, just to compare with the set operations.

On running this code, a user friendly, menu-driven CLI appears.