

ReadMe

Information Retrieval Assignment 2

Submitted By:

**Himanshu Aggarwal
MT17015**

Repository of this Assignment 2 contains two python code file namely, **generator.py** and **main.py**, and three JSON files, containing processed data for the code and cache.

generator.py:

This code generates tokens from the given dataset, preprocess it, compute tf-idf values for each term and document, 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 entered by the user, providing options to generate results ranked in order of TF-IDF score or Cosine Similarity Score.

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