

Information Retrieval Assignment

DEFAULT IR - TP

Group: 10

Abhinandan De (19CS10069)

Aryan Singh (19CS30007)

Pranav Rajput (19CS30036)

Vinit Raj (19CS10065)

Approach

The problem statement was well described and we went on with our implementation as directed. The whole project was done in a pair-programming-based approach where two people would work on a subpart at a time by alternating between coding and reviewing. The major additional libraries that were used for this project were:

1. nltk
2. pandas
3. NumPy

1A. This was the main part of the problem where we were required to build the inverted index. The pandas library was used to parse CSV files, nltk was used to perform preprocessing and lemmatization. We then build the inverted index as per the regular technique. Since this was the major part of the project, all four of us (Abhinandan, Aryan, Pranav, and Vinit) contributed here. Several optimizations were also done during preprocessing to ensure faster runtimes.

1B. Here we had to parse the CSV file containing the queries and dump the results in a new file. This was mostly done with the help of pandas. Abhinandan and Pranav worked on this part.

1C. The main subroutine to be written here was to merge the posting lists after parsing the query file generated in 1B. This was done following the standard algorithm taught to us. Aryan and Vinit Worked on this.

Contributions

Name	1A	1B	1C
Abhinandan			
Aryan			
Pranav			
Vinit			