Mutua Fadhla Mohamed

SM3201434

The folder includes the pre-calculated dataset openalex_papers.csv of 50 papers and openalex_papers2.csv of 5000 papers; however, if required, it can be reconstructed using Data_construct.ipynb. The primary script for execution is final.ipynb, and a complete list of the libraries utilized is provided in the accompanying report.

!NOTE!: In use was ChatGpt which helped with: data manipulation, idea for sparse matrix and Vectorized BM25 for speed up, help on which metrics to represent and filling queries. Finally an essay-style report was provided to it and it returned a tex file for the report.