CSC 575 Final Project

Introduction

This project is an implementation of an Information Retrieval system which makes use of different components such as inverted index, TF-IDF weighted term-document matrices, query expansion, and more to present the user with the most relevant documents to one of their queries.

Dataset

This is a text-based dataset used for information retrieval, publicly available from the University of Glasgow's Information Retrieval Group. The data was compiled by the Centre for Inventions and Scientific Information (CISI). (Data: https://www.gla.ac.uk/schools/computing/research/researchsections/ida-section/informationretrieval/)).

Running

Virtual environment instructions

Optional, but suggested to avoid package management clashes. Skip to "Running the system".

Installing & starting

NOTE: If you have virtualenv installed skip to 3.

1. Navigate to the project folder in your terminal
2. pip3 install virtualenv
3. python3 -m venv venv
4. source env/bin/activate

Use 'deactivate' to exit the virtual environment.

Running the system

Make sure to run these commands from the project folder in your terminal.

1. pip3 install -r requirements.txt
2. python3 app.py