**Content Based Ranking**

Documents are ranked based on their contribution to the corpus using different similarity measures. It has 11 measures based on distance between document vectors, 7 based on Boolean vectors and 1 based on structure of document. For ranking documents, we are using harmonic mean as cohesion measure. We developed an efficient algorithm to calculate structured based similarities, also we modified Russell Rao’s algorithm to get similarity.

**Getting Started**

These instructions will get you a copy of the project up and running on your local machine for development and testing purposes. See deployment for notes on how to deploy the project on a live system.

**Prerequisites**

* Python3
* Scikit-learn
* SciPy
* Matplotlib

Download and install python3 from <https://www.python.org/downloads/>

Install other required packages using following commands:

*pip install -U scikit-learn*

*pip install -U scipy*

*python -mpip install -U matplotlib*

**Running the tests**

Run the program using following command:

*python ranking.py*

It prints ranked document list. It also plots two heatmaps showing similarity between documents and one histogram showing contribution of documents.

**Authors**

1. Kadam Soham Dnyaneshwar

2. Ritesh Ghodrao

3. Abhishek Jain

4. Vatsal Bhanderi

**Guidance**

Dr. Rajendra Kumar Roul