CPS842 Project

Xiangyu Lu 500679658
Marc McCombe 500786634

Scheme we followed to implement the pagerank score:

- 2d array to create matrix. To store the page to page relationship
    - Based off of the .X section in cacm.all
- Normalize the matrix (be aware of rows with all zeros)
- a = 0.85 damping value
- The matrix * (1-a)
- The matrix * (a/N) to get final P matrix
- Do the iteration, 15- 20 iterations for our set
- Normalized the matrix for matching the similarity score: by multiplying 10000 to each unit in the matrix.
- Use the linear combination to combine the previously calculated cosine similarity score with the PageRank score as follows: score(d, q) = w1*cos-score(d, q) + w2*pagerank(d) where w1+w2=1. Note that w1 and w2 should be set as input parameters.

How to run the project

- Open the terminal window under linux system, then type the following;
- % javac Invert.java
- % java Invert
- Following the instruction
- Then posting.txt and dictionary.txt will be generated
- %javac Search.java
- %java Search
- User enters weight values for cosine similarity score and pagerank score
- User enters desired query term, then hit return; ranked documents with the final score will be printed out;
- %javac Eval.java
- %java Eval
- User enters the weight values for cosine similarity score and pagerank score (w1 and w2).
- The program will give AP values of each query found in query.txt as well as the final MAP value