CPS 842 Project Report – Harley Guanlao 500695678

My project consists of a combination of assignment 1, 2, and new programs. Included in order to run the programs should be invert.py, cacm.all, search.py, eval.py, stopwords.txt, qrels.text, query.text, ui.py, pagerank.py.

* Running invert.py should give you dictionary, postings, and links. (all X -> Y citations in list format)
* Running pagerank.py shall give you prScores. (PageRank Scores for all documents in dict format)
* Ui.py is called to format the information and return the ranked list.
* Eval.py requires two inputs now (w1 and w2)
* Running Search.py will still implement the pagerank Score however the weights are hard coded to 0.5, 0.5.

Everything in this project (post assignment 1 and 2) I wrote myself, including adjustments made to search.py, eval.py, ui.py, invert.py.

Adjustments Made:

* Invert.py was made to include all the citations from .X and format them into an easily readable file.
* Search.py was made to involve the use of weights.
* Ui.py was made to involve the use of weights and include prScores
* Eval.py was made to allow the users to input weights of their choosing between 0 and 1

New:

* Pagerank.py was created to parse through the list of links or citations and create a dict consisting of documents and their weights, with multiple iterations easily possible; pushing a streamlined version onto the file prScores for use elsewhere.

References:

The only references I used were the quick tutorial references just to keep the basics and possibilities of python readily available.

<https://www.w3schools.com/python/python_for_loops.asp>

<https://www.w3schools.com/python/python_dictionaries.asp>