Information Retrieval

Assignment-2 : PageRank

DOCUMENTATION

Date : 10/11/2017

Members : Rahul Polisetti - 2015AAPS0240H

Sai Aasrith - 2015A7PS0016H

Hitesh V. Bhagchandani - 2015A7PS0023H

## power\_iteration.py

### PageRank(M, articles, teleport\_set = None, topic\_specific = False, epsilon = 1.0e-8, d = 0.85):

## power iteration for PageRank

## load\_corpus.py

* read the links files
* read the tsv file and delimit on tabs
* delete the first ten lines of comments in the tsv file
* unquote to parse the url formating of the links

## initialize.py

* load the links from load\_corpus.py
* dictionary of articles(nodes)
* each article is mapped to a list of its outlinks
* map every article to a number in alphabetical order
* create web matrix M
* save M and articles

## topic\_specific\_load\_corpus.py

* load catagories from the tsv file

## topic\_specific\_main.py

* create a topic set
* create a dictionary of topics
* add articles to the topics
* create a teleport set based on the user choice
* run the Page Rank algorithm