Sequential Page Rank Algorithm

Introduction

* Used by Google search engine to measure the relative importance of the web pages
* Assigns numerical weighting to each web page
* Depends on the number pages that are linked
* Every inbound link to page increases its PageRank value

Implementation

* Algorithm calculates the PageRank value of each page depending on the inbound links to that page
* Each inbound link is like a vote which is used to determine which pages are more important
* Page rank calculations are done iteratively
* User can either calculate page ranks until convergence or specified number of input iterations

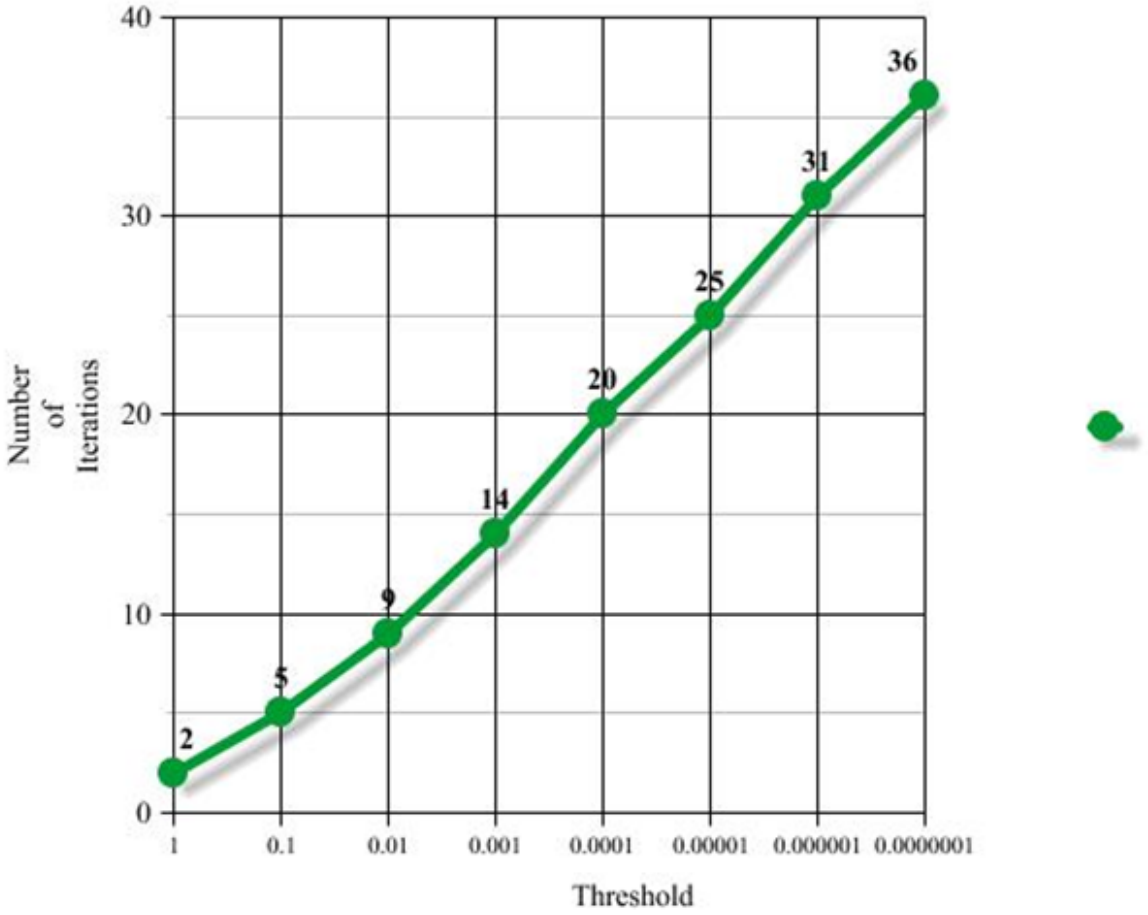
Problem Statement

Given an input file containing 1000 web pages along with the list of web pages to which they are linked. This is similar to adjacency matrix. The aim of this project is calculate PageRank value of each page by taking into account dangling nodes and determine the ten most popular web pages

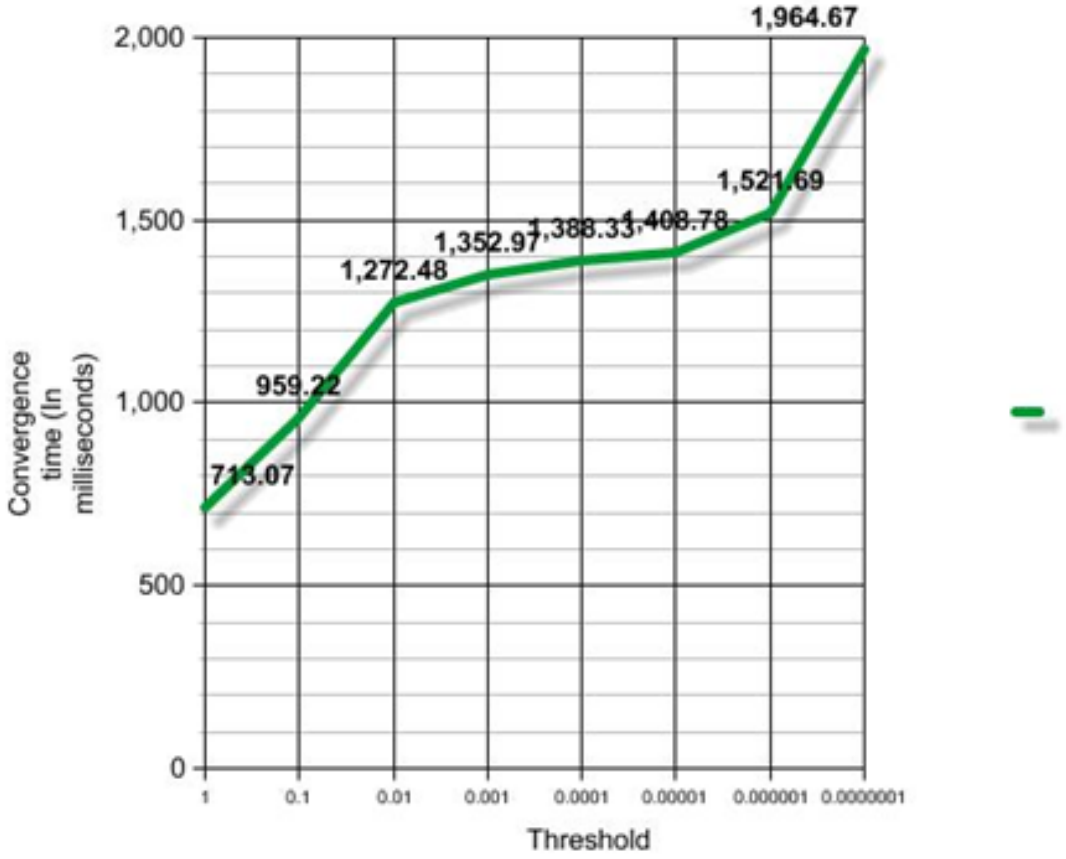
Result – Top 10 URLs with high page rank values

URL	Page Rank
4	0.13821304217473024
34	0.12302491704773691
0	0.11257935294330157
20	0.07736590523118934
146	0.05713176348278271
2	0.04792631126705502
12	0.02006643690709921
14	0.01790592635583653
16	0.01302811362009985
6	0.01295544157190792

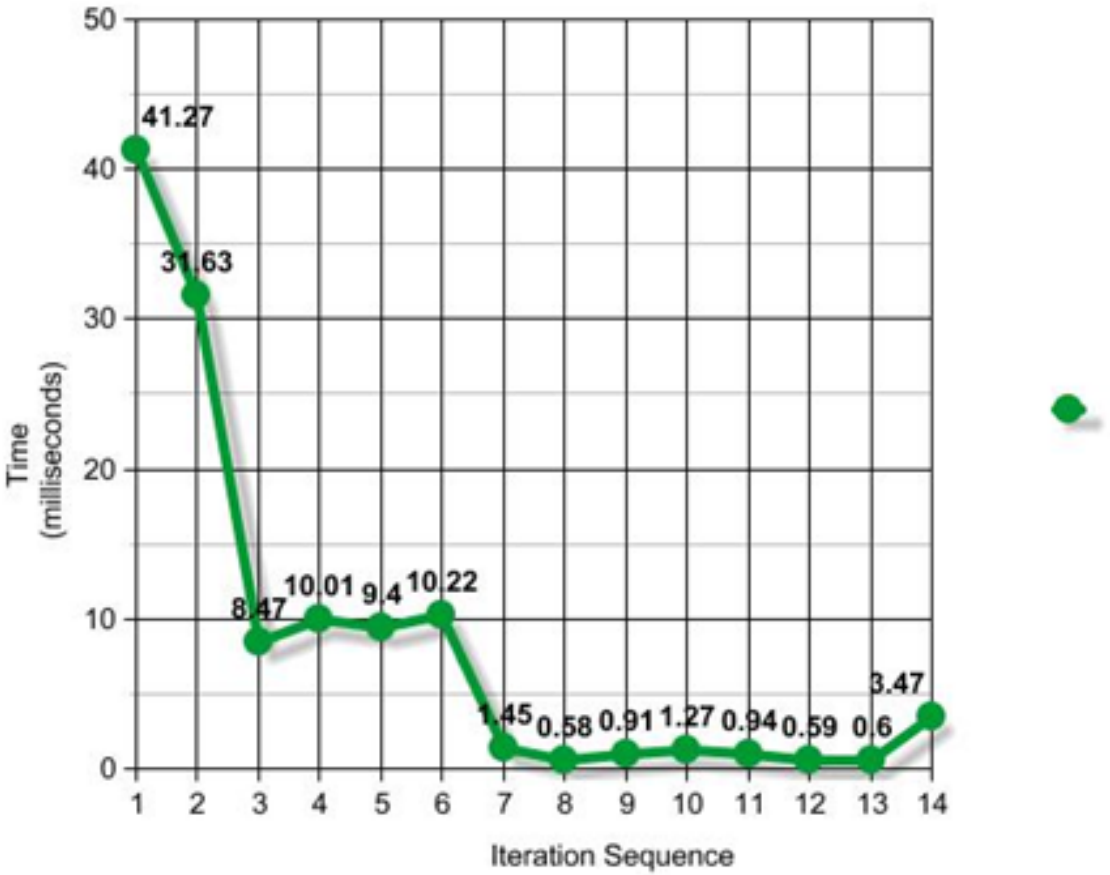
Number of iterations time v/s page rank threshold



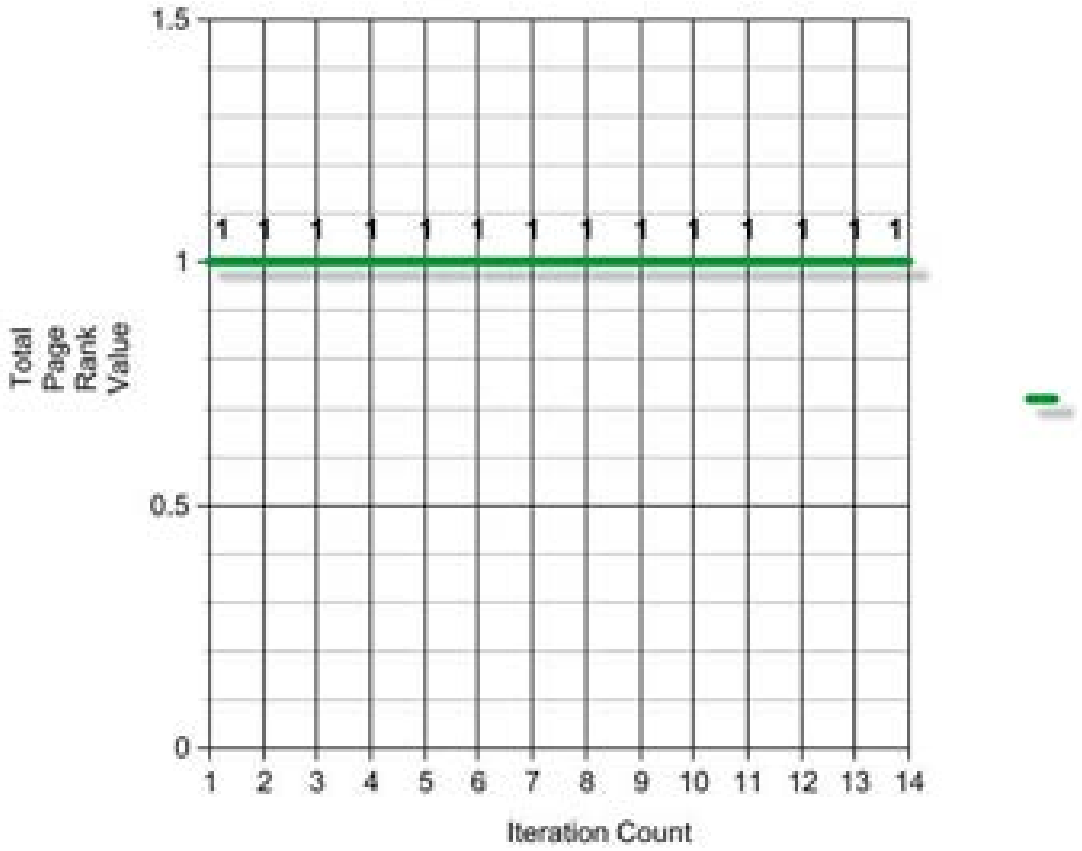
Convergence time (ms) v/s page rank threshold



Execution time(ms) v/s Sequence number



Cumulative Page rank value at each iteration



Page Rank Algorithm - Assumptions

* A specific page cannot have more than one outbound links on the same page
* A web page cannot have more than one inbound links from the same page
* Web page cannot have outbound link pointing to itself

References

* http://en.wikipedia.org/wiki/PageRank
* http://www.webworkshop.net/pagerank.html
* http://www.sirgroane.net/google-page-rank/
* http://pr.efactory.de/pagealgorithm.shtml
* http://infolab.stanford.edu/~backrub
* http://nces.ed.gov