CALIFORNIA STATE SCIENCE FAIR
2003 PROJECT SUMMARY

Name(s)
Tony Wu

Project Number
S1229

Project Title
Developing a More Effective Ranking Algorithm with Query Analysis than Google's PageRank

Abstract

Objectives/Goals
The Researcher developed a new ranking algorithm of Web pages. This algorithm ranks Web pages by counting the number of query keywords and analyzing the distribution of the query keywords within a page. The objective was to see if the new algorithm is more effective than Google's PageRank in determining the relevance of a page.

Methods/Materials
Wrote a VB program that retrieves Google's top 500 -1000 Web pages with a query of 3 - 4 keywords and ranks them using the new algorithm. Randomly mixed the top 10 pages ranked by the researcher's algorithm with Google's top 10 pages and let 11 people evaluate these 20 web pages regarding their relevance to the query. Also, repeated above steps using other 10 queries in different subjects and let three people determine their relevance to the queries.

Results
The users' average evaluation scores showed that the researcher's algorithm retrieved 70.77% of relevant and partial relevant web pages while Google only 28.79% within top 10 pages.

Conclusions/Discussion
The researcher's algorithm is more effective than Google's PageRank in terms of the relevance of Web pages to search query for a certain group of people represented by the testing users and queries provided by these users, which was statistically significant at the .01 confidence level.

Summary Statement
For this project, the researcher created a new ranking algorithm of Web pages in Visual Basic code and compared it to Google's PageRank.

Help Received
Mr. Robert Ferazzi in University High School, Irvine, CA, Prof. Jeffrey D. Ullman, Prof. Gio Wiederhold, Dr. Jan Jannink and Ph.D. Student Glen Jeh in Computer Science Dept., Stanford University advised the experimental design and verified the research methodology.