12 May 2006

By: Vlad Tarko, Senior Editor, Sci-Tech News

Google Measures the "Impact" of Scientific Papers

And proves to be better than the traditional citation index

Google PageRank algorithm seems to offer a better way of measuring the "impact" of scientific papers than the traditional citation indices, which only measures how many times a certain paper has been cited by other papers, and might even eventually replace the citation index.Researchers have found that the Google PageRank algorithm, which measures the relative importance of Web pages, provides a systematic way of finding important papers. The Google index also proved better at uncovering scientific "gems" that escape conventional rankings. Sidney Redner and Pu Chen of Boston University and Huafeng Xie and Sergei Maslov at the Brookhaven National Laboratory found that various papers that proved to be highly influential, like the 1933 paper by Wigner and Seitz, "On the Constitution of Metallic Sodium", which now is textbook material, or Feynman and Gell-Mann's 1958 paper "Theory of the Fermi Interaction", which introduced a new theory that subsequently became the "standard model" of weak interactions, or the 1963 paper by Glauber, "Photon Correlations", which won the last year's Nobel Prize for physics, were not highly rated according to the citation index but Google PageRank algorithm nevertheless highlighted them. The reason why Google algorithm is better than the citation index is that it goes beyond the surface, launching many random "walkers" on the network of citations. While the citation index only looks at how many citations each paper has, Google looks further at how many citations the paper that cites a certain paper (and so on) has. Thus, a paper may turn out to be highly influential although it isn't cited directly by many. In their study, researchers simply applied Google PageRank algorithm to the entire network of citations for all articles in the Physical Review family of journals published between 1893 to June 2003. The network in the experiment consists of 353,268 "nodes", which represent all articles published during this time, and 3,110,839 "links" that represent all citations to Phys. Rev. articles from other Phys. Rev. articles. The team also found that papers having high citation indices were also rated highly by the Google PageRank technique. The papers that had anomalously high Google rank numbers compared with their citation rank turned out in the statistical analysis in the form of so-called "outliers" - papers that didn't seem to fit into the general correlation scheme (the outliers are usually left out of the statistical analyses because the are thought to mess up the data) - but these proved to be precisely the exceptional papers. "I imagine using Google PageRank to help organize scientific literature searches," says Redner. "The technique might also emerge as a more useful measure of scientific impact than merely the number of citations alone."

Copyright (c) 2001-2009 Softpedia. All rights reserved. Softpedia™ and Softpedia™ logo are registered trademarks of SoftNews NET SRL.