Web-Based Data Collection for the Analysis of Hidden Relationships (Web Mining of Hypertext Links)

Edna Reid
Nanyang Technological University

Follow this and additional works at: http://aisel.aisnet.org/icis2000

Recommended Citation
http://aisel.aisnet.org/icis2000/87

This material is brought to you by the International Conference on Information Systems (ICIS) at AIS Electronic Library (AISeL). It has been accepted for inclusion in ICIS 2000 Proceedings by an authorized administrator of AIS Electronic Library (AISeL). For more information, please contact elibrary@aisnet.org.
WEB-BASED DATA COLLECTION FOR THE ANALYSIS OF HIDDEN RELATIONSHIPS (WEB MINING OF HYPERTEXT LINKS)

Edna Reid
Nanyang Business School
Nanyang Technological University
Singapore

What segments of the global virtual communities are interested in your enterprise’s or competitor’s website? Are they threats and/or opportunities? What are they saying?

With the proliferation of e-businesses, our current array of research tools need to harness and exploit the rich content provided in virtual community resources such as web hypertext linkage data, online text, online advertisements, and consumer-generated information. Online texts, including computer-mediated communication (CMC), such as those found in a myriad of web sites, are being analyzed for patterns among the sites and between the words or phrases on those sites.

According to some scholars, CMC data can be used to generate cognitive maps, identify networks of people who share similar conceptual orientations, and diffuse new ideas or innovations. The techniques of web mining hypertext links and online sociometric analyses are also being used to explore trends, social networking, diffusion, and competitive analyses.

Except for a few studies, web linkage analysis has not been fully exploited by the business community. Fuld and Company, a well-known competitive intelligence enterprise, and AstraZeneca International, a leading pharmaceutical enterprise, are using web hypertext link data for competitive analysis.

Kessler (1999), a manager at Fuld, suggested the use of hypertext links as indicators of official and unofficial relationships. In conducting a business intelligence project, she needed information about a small subsidiary of a large service sector enterprise. Although limited information was available from the subsidiary’s web site, she found incoming hypertext links from a federal and a state government web site. The hypertext links (aka “reverse links”) revealed a signal that the subsidiary participates in government discount programs.

To perform “reverse link,” for example, one can go to the HotBot site (www.hotbot.com), change the drop-down box default from “all the words” to “links to this URL” and type in the specific URL. The result will indicate what web sites have hypertext links to the specified URL, which can then be used in the analysis of who is citing whom via the web.

Rousseau (1997), a bibliometric scholar, used the term “citation” to designate this relation between sites on the Internet. The interest is not the number of links that are found on a given web page, but in the number of times to which a given web page is referred.

This tutorial will present an analytical framework and five-step process for exploiting hypertext linkage data. The process facilitates a variety of e-business research as well as provides the capability of discovering hidden relationships. In addition to explaining the process, the tutorial will demonstrate how the technique was applied to a case study of Microstrategy, Inc., a provider of business intelligence software (www.microstrategy.com).

For Microstrategy, there were 1,154 links to their web site. This presentation will summarize the results of analyzing 1,154 links and outline a framework for facilitating virtual community research that anchors on technological trends, industry mapping, online citations, Porter’s Competitive Forces, and competitive intelligence (CI). The application of the framework can generate results similar to the CI services provided by online ad enterprises that allow e-business managers to see where their competitors place online ads, who they reach, and the content of the ads.
References