

Chapter 7.11

Hyperlink Structure Inspired by Web Usage

Pawan Lingras

Saint Mary's University, Canada

Rucha Lingras

Saint Mary's University, Canada

ABSTRACT

This chapter describes how Web usage patterns can be used to improve the navigational structure of a Web site. The discussion begins with an illustration of visualization tools that study aggregate and individual link traversals. The use of data mining techniques such as classification, association, and sequence analysis to discover knowledge about Web usage, such as navigational patterns, is also discussed. Finally, a graph theoretic algorithm to create an optimal navigational hyperlink structure, based on known navigation patterns, is presented. The discussion is supported by analysis of real-world datasets.

DOI: 10.4018/978-1-59904-990-8.ch024

INTRODUCTION

The structure of a Web site is usually based on how the designer envisions the site will be used. However once the Web site is put into use, the designer's theoretical approach may turn out to be not so practical. It is only the actual use of the Web site that will give the designer clues about how the users navigate through the site and in what content they are most interested. For example, if the users are flocking towards a particular type of content, then the designer could think about establishing the site as an authority on that type of content by providing more information on the topic. Conversely, if a type of content is not generating much attention, it may be because the users are missing the information. The designer may look into changing its location on the Web site.

The users' navigational patterns help reveal the users' interests, but they can also be used to adjust the hyperlink structure of the Web site for optimal navigation. This chapter describes how Web usage can be used to construct a hyperlink structure that is easier to navigate. The discussion includes data visualization, uncovering navigational patterns using conventional data mining techniques, as well as a graph theoretical algorithm to construct an improved navigational structure.

The second section of this paper provides a background on various tools to analyze Web navigation. The next section describes an application of data mining techniques to discover Web navigation patterns. An algorithm to create an optimal hyperlink structure is presented in afterwards. The chapter concludes by summarizing the content and identifying areas for future research and development.

BACKGROUND

Web usage mining applies data mining techniques to discover usage patterns from Web data, in order to understand and better serve the needs of Web-based applications. While Web content mining and Web structure mining utilize the information found in Web documents, Web usage mining uses secondary data generated by the users' interaction with the Web. Web access logs available on most servers are good examples of the datasets used in Web usage mining. Other Web usage data may include browser logs, user profiles, registration files, user sessions or transactions, user queries, bookmark folders, as well as mouse clicks and scrolls (Kosala and Blockeel, 2000). Web usage mining includes the creation of user profiles, as well as analysing user access patterns and navigation paths.

Prior to applying data mining techniques, it is essential to understand the dataset. This is typically done by creating multiple summary reports and, if possible, using visual representations.

Before writing programs for analyzing Web access logs, one may want to consider one of the analysis tools already available. These analysis tools may provide answers to most questions regarding the usage of Web sites. The list below provides the freeware and open source Web access analysis tools listed on an Open Directory Web site (<http://dmoz.org/>). In addition to freeware and open source tools, the listing of commercial tools can also be found on the Open Directory site. This section provides a discussion on how to obtain summary reports, visualization of aggregate clickstream, as well as individual user sessions from Web access logs.

- Analog www.analog.cx
- AWStats awstats.sourceforge.net
- BBClone bbclone.de
- The Big Brother Log Analyzer bbla.sourceforge.net
- BlibbleBibble LogAnalyser www.blibbleblobble.co.uk/Downloads/LogAnalyser
- Dailystats www.perlfect.com/freescripts/dailystats
- GeoIP www.maxmind.com/geoip
- High Speed Merging ww.whurst.net/programming/hHSM/index.php
- HitsLog Script www.iris.net/soft/hitslog
- Http-Analyze www.http-analyze.org
- Kraken Reports www.krakenreports.com
- Logfile www.ratrobot.com/programming/shell
- LogFile Analyse www.jan-winkler.de/dev
- LogReport Foundation logreport.org
- MagicStats www.nondot.org/MagicStats
- Modlogan www.modlogan.org
- NedStat www.nedstat.com
- Pathalyzer pathalyzer.bzzt.net
- phpOpenTracker www.phpopentracker.de
- PowerPhlogger pphlogger.phpee.com
- RCounter rcounter.noonet.ru
- Realtracker Web site Statistics free.real-tracker.com
- Relax ktmatu.com/software/relax

12 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:
www.igi-global.com/chapter/hyperlink-structure-inspired-web-usage/37728

Related Content

Exploiting Captions for Web Data Mining

Neil C. Rowe (2005). *Web Mining: Applications and Techniques* (pp. 119-144).

www.irma-international.org/chapter/exploiting-captions-web-data-mining/31136

Semantic Clustering of Web Documents: An Ontology based Approach Using Swarm Intelligence

J. Avanijaand K. Ramar (2012). *International Journal of Information Technology and Web Engineering* (pp. 20-33).

www.irma-international.org/article/semantic-clustering-web-documents/75122

Design of an Embedded Solar Tracking System Based on GPS and Astronomical Equations

Fawzi M. Al-Naima, Ramzy S. Aliand Ahmed J. Abid (2014). *International Journal of Information Technology and Web Engineering* (pp. 12-30).

www.irma-international.org/article/design-of-an-embedded-solar-tracking-system-based-on-gps-and-astronomical-equations/113318

A Deterministic Approach to XML Query Processing with Efficient Support for Pure and Negated Containments

Dunren Che (2009). *Agent Technologies and Web Engineering: Applications and Systems* (pp. 175-194).

www.irma-international.org/chapter/deterministic-approach-xml-query-processing/5033

A Deterministic Approach to XML Query Processing with Efficient Support for Pure and Negated Containments

Dunren Che (2006). *International Journal of Information Technology and Web Engineering* (pp. 49-67).

www.irma-international.org/article/deterministic-approach-xml-query-processing/2618