Chapter 12
Website Topology Modification with Hotlinks Using Mined Webusage Knowledge

Thendral Puyalnithi
VIT University, India

Madhu Viswanatham V
VIT University, India

ABSTRACT

The hotlinks are the special links introduced in the website to reduce the time to access certain webpages in a webpage that is present in the deeper levels of the topology. Hotlinks selection mechanism plays a vital role in quick access of webpages. The problem is to decide which webpage should be having hotlinks and where the hotlinks should be placed in the website tree topology. We have proposed a methodology which starts by finding the frequent webpage access pattern of visitors of the website. The frequent pattern is found using Associative mining, Apriori algorithm or Frequent Pattern Tree algorithm. Then the frequent patterns are passed through page ranking mechanism. We find the pattern which is having the highest priority. Then the hotlinks are created for the members (webpages hyperlinks) of the pattern. Thus, the work is about assigning hotlinks for a set of pages which are frequently visited. Thus, by updating the topology by introducing hotlinks we can reduce the time to access the web pages.

INTRODUCTION

Website

A website is a collection of related web pages, including multimedia content, typically identified with a common domain name, and published on at least one web server. A web site may be accessible via a public Internet Protocol (IP) network, such as the Internet, or a private local area network (LAN), by referencing a uniform resource locator (URL) that identifies the site. All publicly accessible websites collectively constitute the World Wide Web, while private websites are typically a part of an intranet.

Web pages, which are the building blocks of websites, are documents, typically composed in plain text interspersed with formatting instructions of Hypertext Markup Language (HTML, XHTML). They may incorporate elements from other websites with suitable mark-up anchors. Web pages are accessed and transported with the Hypertext Transfer Protocol (HTTP), which may optionally employ encryption (HTTP Secure, HTTPS) to provide security and privacy for the user. The user's application, often a web browser, renders the page content according to its HTML mark-up instructions onto a display terminal. Hyperlinking between web pages conveys to the reader the site structure and guides the navigation of the site, which often starts with a home page containing a directory of the site web content. Some websites require user registration or subscription to access content. Examples of subscription websites include many business sites, parts of news websites, academic journal websites, gaming websites, file-sharing websites, message boards, web-based email, social networking websites, websites providing real-time stock market data, as well as sites providing various other services. There are three categories in web mining-web structure mining, web content mining, web usage mining.

Web Structure Mining

This is the process of analysing the nodes and connection structure of a website through the use of graph theory. There are two things that can be obtained from this: the structure of a website in terms of how it is connected to other sites and the document structure of the website itself, as to how each page is connected. Web structure mining is the process of using graph theory to analyse the node and connection structure of a web site. According to the type of web structural data, web structure mining can be divided into two kinds:

Extracting patterns from hyperlinks in the web: a hyperlink is a structural component that connects the web page to a different location. Mining the document structure: analysis of the tree-like structure of page structures to describe HTML or XML tag usage. PageRank algorithm is used by Google to rank search results. The name of this algorithm is given by Google-founder Larry Page. The rank of a page is decided by the number of links pointing to the target node.

Web Usage Mining

This is the process of extracting patterns and information from server logs to gain insight on user activity including where the users are from, how many clicked what item on the site and the types of activities being done on the site.

Challenges in Web Mining are:

1. **The Web is Too Huge:** The size of the web is very huge and rapidly increasing. This seems that the web is too huge for data warehousing and data mining.
2. **Complexity of Web Pages:** The web pages do not have unifying structure. They are very complex as compared to traditional text document. There are huge amount of documents in digital library of web. These libraries are not arranged according to any particular sorted order.
3. **Web is Dynamic Information Source:** The information on the web is rapidly updated. The data such as news, stock markets, weather, sports, shopping, etc., are regularly updated.

Related Content

Optimal Ordering Strategy of a Replenishment Policy for Deteriorating Items Under Retailer's Partial Trade Credit Policy

Gour Chandra Mahataand Puspita Mahata (2011). *International Journal of Strategic Decision Sciences (pp. 78-94).*

www.irma-international.org/article/optimal-ordering-strategy-replenishment-policy/54743

A Multi-Criteria Decision Aiding System to Support Monitoring in a Public Administration

Maria Franca Norese (2009). *International Journal of Decision Support System Technology (pp. 59-71).*

www.irma-international.org/article/multi-criteria-decision-aiding-system/37433

Evaluating the Performance of Decision Making Units in the Food Production Industry

Emmanouil Stiakakisand Angelo Sifaleras (2010). *Decision Support Systems in Agriculture, Food and the Environment: Trends, Applications and Advances (pp. 173-192).*

www.irma-international.org/chapter/evaluating-performance-decision-making-units/44761

Integrated Traffic Management Using Data from Traffic, Asset Conditions, Energy and Emissions

Thomas Böhm, Christoph Lackhoveand Michael Meyer zu Hörste (2017). *Decision Management: Concepts, Methodologies, Tools, and Applications (pp. 1524-1538).*

www.irma-international.org/chapter/integrated-traffic-management-using-data-from-traffic-asset-conditions-energy-and-emissions/176817

Factors Influencing Potential Users' Behavior Towards Usage of Cloud Computing

Shruti Bansal, Ajay Jainand Karuna Gupta (2023). *Data-Driven Approaches for Effective Managerial Decision Making (pp. 82-101).*

www.irma-international.org/chapter/factors-influencing-potential-users-behavior-towards-usage-of-cloud-computing/323314