

Chapter 1

Development of Efficient Decision Support System Using Web Data Mining

G. Sreedhar

Rashtriya Sanskrit Vidyapeetha (Deemed University), India

A. Anandaraja Chari

Rayalaseema University, India

ABSTRACT

The management of web sites imposes a constant demand for new information and timely updates due to the increase of services and content that site owners wish to make available to their users, which in turn is motivated by the complexity and diversity of needs and behaviours of the users. Such constant labour intensive effort implies very high financial and personnel costs. The growth of World Wide Web and technologies has made business functions to be executed fast and easier. E-commerce has provided a cost efficient and effective way of doing business. Web mining is usually defined as the use of data mining techniques to automatically discover and extract information from web documents and services. Also, web data mining is commonly categorized into three areas: web content mining that describes the discovery of useful information from content, web structure mining that analyses the topology of web sites, and web usage mining that tries to make sense of the data generated by the navigation behaviour and user profile.

INTRODUCTION

Over the last few years there has been a remarkable increase in use of the World Wide Web (WWW) for a wide and variety of purposes. There was also a fast growth in its applications. This led the Internet users to realize the importance and the benefits gained from a globally interconnected hypermedia system. On the other hand, it causes a larger number of useless, meaningless and badly designed websites on the Internet world causing unwanted additional traffic; this is all because of an unorganized, non-planned websites development processes. Due to the unceasing growth of web sites and applications, developers and evaluators have interesting challenges not only from the development but also from the quality assurance

DOI: 10.4018/978-1-5225-1877-8.ch001

point of view. Today, web is not only an information resource but also it is becoming an automated tool in various applications. The management of web sites imposes a constant demand for new information and timely updates due to the increase of services and content that site owners wish to make available to their users, which in turn is motivated by the complexity and diversity of needs and behaviours of the users. Such constant labour intensive effort implies very high financial and personnel costs.

BACKGROUND: WEB MINING

The data on World Wide Web are available in three different formats: web content, web structure and web usage. Web mining is usually defined as the use of data mining techniques to automatically discover and extract information from web documents and services. The authors of O.Etzioni (1996) and R. Cooley (1997) discuss in their research that web data mining can be defined in two distinct forms: first, it is defined as chain of order tasks and second, it is defined considering type of web data used in web data mining process. Web mining is the application of data mining techniques to extract knowledge from web data, i.e. web content, web structure, and web usage data. Web Data Mining extracts and analyses useful information from huge amounts of data stored in different data sources to provide additional information and knowledge that are crucial of decision making process. A decision support system is a computer-based information system that supports business and organizational decision-making activities. According to J. Srivastava et al. (2002) also web data mining is commonly categorized into three areas. They are:

1. Web content mining
2. Web structure mining
3. Web usage mining

Web Content Mining

Web content data is web pages content availed to users to satisfy their needs of information. This can be in the form of text, HTML pages, images, audio, video etc. In this category, the HTML pages are common and more familiar form of web content data. The author S.Nestorov et al. (1998) discuss in their research that HTML documents are often considered as semi structured as different elements of documents are not designed according to specific schema. In HTML, elements are tagged in a way to enable designing layout of document. Generally, HTML elements are of two types: first concerns with way of displaying documents in browser and second concerns with information about document itself like title and other document relationship. XML document is another known form of web content data which enables storing and transporting information. It is having structured information and includes contents and information about contents. Each XML document has specific structure and XML is a mark-up language which allows identifying document structure and adding the information. In XML, there are no predefined tags and it is language to describe and add mark-up to documents using XML specification. The applications processing XML document or the style sheets decide the semantics of XML document. Another type of web content data is dynamic server pages which are processed by the web server and generated result is sent to web browser. In contrast, without any change, the static contents are sent to browser. Some of familiar dynamic server page contents are like JSP (Java Server Page), ASP (Active Server Page) and PHP (Pre-Hypertext Processor). The author Manoj Pandia et al. (2011) said that Web Content Mining

9 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/development-of-efficient-decision-support-system-using-web-data-mining/173819

Related Content

Internet Banking Service Quality, Customer Satisfaction and Customer Loyalty: The Case of Vietnam

Pham Long and Phan Dien Vy (2016). *International Journal of Strategic Decision Sciences* (pp. 1-17).

www.irma-international.org/article/internet-banking-service-quality-customer-satisfaction-and-customer-loyalty/149659

Thinking in Terms of Design Decisions When Developing Maturity Models

Tobias Mettler (2012). *Decision Making Theories and Practices from Analysis to Strategy* (pp. 170-183).

www.irma-international.org/chapter/thinking-terms-design-decisions-when/65962

Debt Strategy Trends of Emerging Market Firms: Evidence from India

B. Rajesh Kumar and K.S. Sujit (2017). *International Journal of Strategic Decision Sciences* (pp. 86-101).

www.irma-international.org/article/debt-strategy-trends-of-emerging-market-firms/189236

Efficient Implementation of Hadoop MapReduce based Business Process Dataflow

Ishak H.A. Meddah, Khaled Belkadi and Mohamed Amine Boudia (2017). *International Journal of Decision Support System Technology* (pp. 49-60).

www.irma-international.org/article/efficient-implementation-of-hadoop-mapreduce-based-business-process-dataflow/173477

An Integrated Risk Management Framework: Measuring the Success of Organizational Knowledge Protection

Stefan Thalmann, Markus Manhart, Paolo Ceravolo and Antonia Azzini (2017). *Decision Management: Concepts, Methodologies, Tools, and Applications* (pp. 470-486).

www.irma-international.org/chapter/an-integrated-risk-management-framework/176767