

# Chapter 1

## Mastering Web Mining and Information Retrieval in the Digital Age

**Kijpokin Kasemsap**

*Suan Sunandha Rajabhat University, Thailand*

### **ABSTRACT**

*This chapter aims to master web mining and Information Retrieval (IR) in the digital age, thus describing the overviews of web mining and web usage mining; the significance of web mining in the digital age; the overview of IR; the concept of Collaborative Information Retrieval (CIR); the evaluation of IR systems; and the significance of IR in the digital age. Web mining can contribute to the increase in profits by selling more products and by minimizing costs. Web mining is the application of data mining techniques to discover the interesting patterns from web data in order to better serve the needs of web-based multifaceted applications. Mining web data can improve the personalization, create the selling opportunities, and lead to more profitable relationships with customers in global business. Web mining techniques can be applied with the effective analysis of the clearly understood business needs and requirements. Web mining builds the detailed customer profiles based on the transactional data. Web mining is used to create the personalized search engines which can recognize the individuals' search queries by analyzing and profiling the web user's search behavior. IR is the process of obtaining relevant information from a collection of informational resources. IR has considerably changed with the expansion of the Internet and the advent of modern and inexpensive graphical user interfaces and mass storage devices. The effective IR system, including an active indexing system, not only decreases the chances that information will be misfiled but also expedites the retrieval of information. Regarding IR utilization, the resulting time-saving benefit increases office efficiency and productivity while decreasing stress and anxiety. Most IR systems provide the advanced searching capabilities that allow users to create the sophisticated queries. The chapter argues that applying web mining and IR has the potential to enhance organizational performance and reach strategic goals in the digital age.*

DOI: 10.4018/978-1-5225-0613-3.ch001

## **INTRODUCTION**

The development of the World Wide Web has created the successful applications, such as search engines, electronic commerce (e-commerce), weblogs, and social network communications (Yin & Guo, 2013). The analysis of web usage has mostly focused on sites composed of conventional static pages (Berendt & Spiliopoulou, 2000). However, huge amounts of information available in the web derive from databases and are presented to the users in the pattern of the dynamically generated pages (Berendt & Spiliopoulou, 2000). As enterprises expand the increasing information about their business activities on their websites, website data promises as the meaningful source for exploring innovation (Gök, Waterworth, & Shapira, 2015).

With the advent of cost-effective storage systems and high-speed network connectivity, the amount of data gathered by various transactional systems has rapidly increased (Krishna, Jose, & Suri, 2014). Devi et al. (2012) stated that the rising popularity of e-commerce makes data mining a vital technology for several applications, especially online business competitiveness. Web mining is defined as the research focusing on the application of data mining techniques to web data (Borges & Levene, 2007). Markov and Larose (2007) indicated that web mining can be categorized into three domains regarding the nature of data (i.e., web structure mining, web content mining, and web usage mining).

Information retrieval (IR) systems aim to retrieve data that satisfies certain requirements and constitute an important service in many types of networks (Feng & Chin, 2015). IR is a fundamental component of human information behavior (Ruthven, 2008). There is a need for organizing the available information in the meaningful perspective in order to guide and improve the document indexing for the IR applications taking into account more complex data (Codocedo, Lykourantzou, & Napoli, 2014). The key driver of IR system becomes the degree to which a user's search is adapted to the individual user properties and the contexts of use (Steichen, Ashman, & Wade, 2012). The relevant documents from the large data sets are retrieved with the support of ranking function in IR system (Gupta, Saini, & Saxena, 2015).

Increasing amounts of data volume applied on the web and their heterogeneous character make the search for information a challenging task (Besbes & Baazaoui-Zghal, 2015). The design of IR systems must respond to the goals, intentionality, and the domain knowledge of the users (Benoît & Agarwal, 2012). Traditional ranking models for IR lack the ability to make a clear distinction between relevant and non-relevant documents at top ranks if both have similar representations concerning a user's query (Lee, Seo, Jeon, & Rim, 2011). Information requirements recognize the state for individuals seeking information, which includes information search using the IR system (Cole, 2011). Cognitive constructivism takes individual information searchers and their information interaction with IR systems regarding the primary contexts for information behavior in modern organizations (Kasemsap, 2015a).

This chapter aims to bridge the gap in the literature on the thorough literature consolidation of web mining and IR. The extant literatures of web mining and IR provide a contribution to practitioners and researchers by describing the multifaceted applications of web mining and IR to appeal to the different segments of web mining and IR in order to maximize the business impact of web mining and IR in the digital age.

## **BACKGROUND**

The rapid development of the Internet has led to the diverse applications of accessing various web resources, such as web pages, extensible markup language (XML) documents, pictures, and audio files

26 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/mastering-web-mining-and-information-retrieval-in-the-digital-age/162887](http://www.igi-global.com/chapter/mastering-web-mining-and-information-retrieval-in-the-digital-age/162887)

## Related Content

---

### Performance Analysis of a Web Server

Jijun Luand Swapna S. Gokhale (2008). *International Journal of Information Technology and Web Engineering* (pp. 50-65).

[www.irma-international.org/article/performance-analysis-web-server/2653](http://www.irma-international.org/article/performance-analysis-web-server/2653)

### Engineering Wireless Mobile Applications

Qusay H. Mahmoudand Zakaria Maamar (2006). *International Journal of Information Technology and Web Engineering* (pp. 59-75).

[www.irma-international.org/article/engineering-wireless-mobile-applications/2603](http://www.irma-international.org/article/engineering-wireless-mobile-applications/2603)

### Disambiguating the Twitter Stream Entities and Enhancing the Search Operation Using DBpedia Ontology: Named Entity Disambiguation for Twitter Streams

N. Senthil Kumarand Dinakaran Muruganantham (2016). *International Journal of Information Technology and Web Engineering* (pp. 51-62).

[www.irma-international.org/article/disambiguating-the-twitter-stream-entities-and-enhancing-the-search-operation-using-dbpedia-ontology/159158](http://www.irma-international.org/article/disambiguating-the-twitter-stream-entities-and-enhancing-the-search-operation-using-dbpedia-ontology/159158)

### Technological Choices

Roberto Paiano, Anna Lisa Guidoand Andrea Pandurino (2009). *Designing Complex Web Information Systems: Integrating Evolutionary Process Engineering* (pp. 233-252).

[www.irma-international.org/chapter/technological-choices/8175](http://www.irma-international.org/chapter/technological-choices/8175)

### Usability and Accessibility of E-Health Websites: Enabling Nutrition and Physical Activity Education for Blind and Low Vision Internet Users

Constantinos K. Coursaris, Sarah J. Swierengaand Pamela Whitten (2014). *Evaluating Websites and Web Services: Interdisciplinary Perspectives on User Satisfaction* (pp. 204-223).

[www.irma-international.org/chapter/usability-and-accessibility-of-e-health-websites/97034](http://www.irma-international.org/chapter/usability-and-accessibility-of-e-health-websites/97034)