

Chapter XXV

Designing and Mining Web Applications: A Conceptual Modeling Approach

Rosa Meo

Università di Torino, Italy

Maristella Matera

Politecnico di Milano, Italy

ABSTRACT

In this chapter, we present the usage of a modeling language, WebML, for the design and the management of dynamic Web applications. WebML also makes easier the analysis of the usage of the application contents by the users, even if applications are dynamic. In fact, it makes use of some special-purpose logs, called conceptual logs, generated by the application runtime engine. In this chapter, we report on a case study about the analysis of conceptual logs for testifying to the effectiveness of WebML and its conceptual modeling methods. The methodology of the analysis of the Web logs is based on the data-mining paradigm of item sets and frequent patterns, and makes full use of constraints on the conceptual logs' content. As a consequence, we could obtain many interesting patterns for application management such as recurrent navigation paths, the most frequently visited page's contents, and anomalies.

INTRODUCTION

In recent years, the World Wide Web has become the preferred platform for developing Internet applications thanks to its powerful communication paradigm based on multimedia content and browsing, and to its open architectural standards that facilitate the integration of different types of content and systems (Fraternali, 1999).

Current Web applications are very complex, and the quality, as perceived by users, of highly sophisticated software products can heavily determine their success or failure. A number of methods has been proposed for evaluating their effectiveness in content delivery. Content personalization, for instance, aims at tailoring Web contents to the final recipients according to their profiles. Another approach is the adoption of Web usage-mining techniques for the analysis of the navigational behaviour of Web users by means of the discovery of patterns in the Web server log.

Traditionally, to be effective, Web usage mining requires some additional preprocessing, such as the application of methods of page annotation for the extraction of metadata about page semantics or for the construction of a Web site ontology.

In this chapter, we propose a novel approach to Web usage mining. It has the advantage of integrating Web usage mining goals directly into the Web application development process. Thanks to the adoption of a conceptual modeling method for Web application design and its supporting case tool, the generated Web applications embed a logging mechanism that, by means of a synchronization tool, is able to produce semantically enriched Web log files. This log, that we call a conceptual log (Fraternali, Matera, & Maurino, 2003), contains additional information with respect to standard (ECLF [extended common log format]) Web server logs, and some of this information is useful to the Web mining process. It refers not only to the composition of Web pages in terms of atomic units of contents and to the conceptual entities Web pages deal with, but also to the identifier of the user crawling session and to the specific data instances that are published within dynamic pages, as well as to some data concerning the topology of the hypertext. Therefore, no extra effort is needed during or after the application development to collect the data that are necessary for reconstructing and analyzing usage behaviour.

The main contribution of this chapter comes from the integration of two existing frameworks. The first one is the model-based design and development of Web applications based on the Web Modeling Language (WebML; Ceri, Fraternali, & Bongio, 2000; Ceri, Fraternali, Bongio, et al., 2002) and its supporting CASE (computer aided software engineering) tool WebRatio (Ceri et al., 2003). The second one is an evaluation of the applications based on data-mining analytics that had started by collecting the application data based both on the static (i.e., compile-time) analysis of conceptual schemas and on the dynamic (i.e., runtime) collection of usage data. The evaluation of the application is aimed at studying its suitability to respond to users' needs by observing their most frequent paths or by observing the application response in different contexts, often difficult due to the network traffic conditions, the users themselves (such as their browsers), or even security attacks.

The distinctive merit of WebML and WebRatio in this collection of application-specific data lies in the ease with which relevant data are retrieved, automatically organized, and stored. However, the illustrated results are of general validity and apply to any application that has been designed using a model-driven approach, provided that the conceptual schema is available and the application runtime architecture permits the collection of customized log data.

This chapter presents a case study on the analysis of conceptual Web log files of the Web site of a university department. Our objective is to testify to the power and versatility of the conceptual modeling of data-intensive Web applications. The aim of our study is manifold: (a) analyzing the Web logs and extracting interesting, usable, and actionable patterns, (b) evaluating the usability (in practical cases) and the expressive power of the conceptual Web logs, and (c) verifying the suitability of some KDD scenarios. In particular, KDD (knowledge discovery in databases) scenarios have been produced as a set of characteristic data-mining requests, a sort of templates, to be filled in with specific parameter values. KDD scenarios should be able to solve some frequently asked questions (mining problems) by

15 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/designing-mining-web-applications/21738

Related Content

ASCCN: Arbitrary Shaped Clustering Method with Compatible Nucleoids

Renxia Wan, Lixin Wang and Xiaoke Su (2010). *International Journal of Data Warehousing and Mining* (pp. 1-15).

www.irma-international.org/article/asccn-arbitrary-shaped-clustering-method/46940

Vertical Fragmentation in Databases Using Data-Mining Technique

Gorla Narasimhaiah and Wing Yan Betty Pang (2010). *Strategic Advancements in Utilizing Data Mining and Warehousing Technologies: New Concepts and Developments* (pp. 178-197).

www.irma-international.org/chapter/vertical-fragmentation-databases-using-data/40404

Impact of Balancing Techniques for Imbalanced Class Distribution on Twitter Data for Emotion Analysis: A Case Study

Shivani Vasantbhai Vora, Rupa G. Mehta and Shreyas Kishorkumar Patel (2022). *Research Anthology on Implementing Sentiment Analysis Across Multiple Disciplines* (pp. 412-432).

www.irma-international.org/chapter/impact-of-balancing-techniques-for-imbalanced-class-distribution-on-twitter-data-for-emotion-analysis/308500

Modeling and Querying Continuous Fields with OLAP Cubes

Leticia Irene Gómez, Silvia Alicia Gómez and Alejandro Vaisman (2013). *International Journal of Data Warehousing and Mining* (pp. 22-45).

www.irma-international.org/article/modeling-querying-continuous-fields-olap/78374

A Component-Based Framework for the Integration and Exploration of XML Sources

Pasquale De Meo, Antonino Nocera and Domenico Ursino (2012). *XML Data Mining: Models, Methods, and Applications* (pp. 343-377).

www.irma-international.org/chapter/component-based-framework-integration-exploration/60916