Chapter VIII

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 comon 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 ap-
General Strategy for Querying Web Sources in a Data Federation Environment
www.igi-global.com/chapter/general-strategy-querying-web-sources/37754?camid=4v1a