



## Chapter 11

# Complementing the Data Warehouse with Information Filtered from the Web

Witold Abramowicz, Pawel Jan Kalczynski and Krzysztof Wecel  
The Poznan University of Economics, Poland

*The data warehouse is considered to be the best way to organize transactional data. However, as many researches claim data warehouse should be augmented with external textual information. The objective of this chapter is to examine the requirements for profiling in the data warehouse environment. Profiles created in the data warehouse are then utilized to filter information. The goal of the sketched system is to support users in his situated actions. We explore many issues concerning personalization, such as information overflow, user models, and situatedness. We also analyze the factors that influence the filtering process. Finally, we draw some conclusions that should be considered during extension of the evaluated system.*

## INTRODUCTION

One of the problems, technology has to cope with, is the overflow of information. Users have access to a great deal of information resources through such means as commercial online services, business portals, financial wires, mail messages, electronic bulletin boards, and news articles (Abramowicz, Kalczynski & Wecel, 2001b). Although users of electronic sources have access to a rich body of information, only a small part of information is actually relevant to their interests. Thus, the problem of picking and presenting to the user only documents containing interesting information arises. One solution is to use filters that selectively eliminate the irrelevant information based on user preference (Foltz, 1990).

Another solution to dealing with the increasing flow of data is data warehousing. Data warehousing has grown during the last 10 years into a technology proven in

Previously Published in *Managing Information Technology in a Global Economy*, edited by Mehdi Khosrow-Pour, Copyright © 2001, Idea Group Publishing.

This chapter appears in the book, *Data Warehousing and Web Engineering* by Shirley Becker. Copyright © 2002, Idea Group Publishing.



many successful organizations as the best way to organize, store, analyze and report data collected through the course of doing business. The data warehouse stores and manages historical data, and can be reshaped and reused to meet the business requirements as they evolve. Many data warehousing tools exist to extract, transform, store, aggregate, access, present, mine and extrapolate data.

The chances of success in a data warehousing project can be greatly increased by using one of the newest business trends: knowledge management. Knowledge management allows businessmen to make better use of resources, find innovative solutions to business problems, and improve quality of management (Abramowicz, 2001). Data warehousing is, in essence, a form of knowledge management. It promises to improve data quality and integrate internal and external data (Allison, 2000).

This chapter presents the incorporation of information filtering into the data warehouse environment through appropriately defined profiles.

The specific contributions include: (1) exploiting information filtering techniques to supplement the data warehouse with external information, (2) requirements concerning the profiles, (3) suggestion to explore the data warehouse to build profiles.

The rest of this chapter is organized as follows: the next section presents the background of our project. The following section discusses some personalization issues in the data warehouse environment and reviews related work. The section entitled, *The Profiles in a DW Environment*, presents requirements for profiles in a DW. Finally, the last section provides conclusions.

## BACKGROUND

### Data Warehousing

We assume that the idea of data warehousing is commonly known (Inmon, Hackathorn, 1994; Kimball, 1996). For the purposes of our project we use the SAS/Warehouse Administrator™ (SAS Institute, Inc., 1997b) from the SAS Institute (<http://www.sas.com>).

To be more precise, we briefly describe this specific DW environment. According to the data warehousing paradigm, the repository is organized into a number of *subjects*. A warehouse *subject* can be defined as a collection of data and information concerning a particular issue (e.g. client, product). Each subject must contain a single *detail logical table*. The detail logical table consists of a single multidimensional table or a number of detail tables organized into a star schema.

It is very beneficial that each subject may also contain an *information mart* – a logical grouping of information items. The information mart may contain any information generated from detail data or summary data in the warehouse, and additionally text documents, HTML pages, or spreadsheets.



11 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/complementing-data-warehouse-information-filtered/7869](http://www.igi-global.com/chapter/complementing-data-warehouse-information-filtered/7869)

## Related Content

---

### Acquiring Semantic Sibling Associations from Web Documents

Marko Brunzeland Myra Spiliopoulou (2008). *Data Warehousing and Mining: Concepts, Methodologies, Tools, and Applications* (pp. 1987-2003).

[www.irma-international.org/chapter/acquiring-semantic-sibling-associations-web/7744](http://www.irma-international.org/chapter/acquiring-semantic-sibling-associations-web/7744)

### World Wide Web Personalization

Olfa Nasraoui (2005). *Encyclopedia of Data Warehousing and Mining* (pp. 1235-1241).

[www.irma-international.org/chapter/world-wide-web-personalization/10787](http://www.irma-international.org/chapter/world-wide-web-personalization/10787)

### Data Warehousing and the Organization of Governmental Databases

Franklin Maxwell Haper (2008). *Data Warehousing and Mining: Concepts, Methodologies, Tools, and Applications* (pp. 2427-2437).

[www.irma-international.org/chapter/data-warehousing-organization-governmental-databases/7772](http://www.irma-international.org/chapter/data-warehousing-organization-governmental-databases/7772)

### Vertical Data Mining

William Perrizo, Qiang Ding, Qin Dingand Taufik Abidin (2005). *Encyclopedia of Data Warehousing and Mining* (pp. 1181-1184).

[www.irma-international.org/chapter/vertical-data-mining/10776](http://www.irma-international.org/chapter/vertical-data-mining/10776)

### A Literature Overview of Fuzzy Database Modeling

Z.. M. Ma (2008). *Data Warehousing and Mining: Concepts, Methodologies, Tools, and Applications* (pp. 187-207).

[www.irma-international.org/chapter/literature-overview-fuzzy-database-modeling/7641](http://www.irma-international.org/chapter/literature-overview-fuzzy-database-modeling/7641)