

Chapter 2.45

Enhanced Knowledge Warehouse

Krzysztof Wecel

The Poznan University of Economics, Poland

Witold Abramowicz

The Poznan University of Economics, Poland

Pawel Jan Kalczynski

University of Toledo, USA

INTRODUCTION

Enhanced knowledge warehouse (eKW) is an extension of the enhanced data warehouse (eDW) system (Abramowicz, 2002). eKW is a Web services-based system that allows the automatic filtering of information from the Web to the data warehouse and automatic retrieval through the data warehouse. Web services technology extends eKW beyond the organization. It makes the system open and allows utilization of external software components, thus enabling the creation of distributed applications.

The enhanced knowledge warehouse is an add-in to the existing data warehouse solutions that offers a possibility of extending legacy data-based information resources with processed information acquired from Web sources of business information.

eKW is characterized by transparent filtering and transparent retrieval. The system automatically acquires interesting documents from Web and internal sources and enables business users, who view data warehouse reports, to access these documents without the necessity of formulating any document retrieval queries. Instead, the system takes advantages of meta-information stored in business metadata and data warehouse reports, and builds and runs appropriate queries itself. The system also has an event-alerting capability.

eKW, by design, is a part of the semantic Web. The main role of eKW in the semantic Web is mediation between the world (external information) and internal information systems utilized within the organization. One of the most popular ways of conforming to semantics is to use ontologies, which provide a means for defining the concepts of the exchanged data. eKW is a kind of data

mediator that employs ontologies as a conceptualization layer (W'cel, 2003).

BACKGROUND

Document management systems may be divided into two major categories: information retrieval (IR) and information filtering (IF) systems.

The classical models of information retrieval systems were defined by Salton (1983). As distinct from structured data management, information retrieval is imprecise and incomplete. This is due to the inaccurate representations of document contents and user information needs. The distinctive feature of information retrieval systems is a relatively constant document collection. The collection stores documents and their representations (indices). When a user submits a query to the information retrieval system, the query is compared against all indices in the collection. Documents whose indices match the query are returned as the resulting subset of the collection. Most search engines on the Web and digital libraries available on the market can serve as examples of information retrieval systems (Baeza-Yates, 1999).

The main objective of information filtering systems is to block unwanted information from the incoming stream of documents. As distinct from IR systems, filters do not have any fixed collection of documents. Instead, they hold a collection of standing queries (profiles). The profiles represent long-term information interests of their owners. When a new document arrives at the filter its representation is created and compared against all profiles in the collection. Some libraries maintain systems that inform users about new volumes included in the library. Such systems are usually based on the server-side filtering architectures and their profiles usually consist of semi-structured elements (e.g., author, subject, title). This idea is referred to as selective dissemination of information (SDI).

JUSTIFICATION

Possessing and maintaining large information resources does not itself provide any guarantees that users will manage to find the piece of information (document) they need. First of all, users must be aware that the particular piece of information already exists in their resources. Yet, even if they are aware, searching may be very labor consuming. Such a situation is highly undesirable for organizations, as they do not exploit their potentials to increase the effects. That is why knowledge management is nowadays considered the capability of re-use of information and is becoming an increasingly vital issue.

Usually, users are not eager to learn how to operate in several different information systems. Integration of information resources enables the exploitation of the capabilities of many systems through a single user interface (UI). Data warehouse users should be capable of finding interesting documents through a single UI. Not only should the users be aware that the desired information does exist, but also relevant documents should be disseminated to them mechanically. Such a mechanical distribution of the relevant information gives more time for other tasks.

Business users have certain problems with finding information sources, building correct retrieval queries and formulating proper filtering profiles. Thus, introducing the system capable of relieving the users from the necessity of formulating queries and solving the most commonly reported problems with accessing information on the Web would increase productivity of those seeking relevant external information.

The constantly growing number of content providers and the exploding volume of business information are not accompanied by the corresponding growth of capabilities of exploitation of the resources by the contemporary organizations. Therefore, automatic acquisition, organization and presentation of information became not only

4 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/enhanced-knowledge-warehouse/25155

Related Content

The Effect of Conflict and Knowledge Sharing on the Information Technology Project Team Performance

Barbara Hewitt, Diane B. Walz and Alexander McLeod (2020). *International Journal of Knowledge Management* (pp. 1-20).

www.irma-international.org/article/the-effect-of-conflict-and-knowledge-sharing-on-the-information-technology-project-team-performance/243636

The Role and Significance of Traditional Leadership in South African Local Governance

Joshua Mawere, Pfarelo Eve Matshidze, Stewart Lee Kugara and Thanyani Madzivhandila (2022). *Handbook of Research on Protecting and Managing Global Indigenous Knowledge Systems* (pp. 249-273).

www.irma-international.org/chapter/the-role-and-significance-of-traditional-leadership-in-south-african-local-governance/289300

Modeling and Automated Examination of Communication Processes in Integrated Health Information Systems: A Systematic Approach

Samrend Saboor, Alexander Hörband and Elske Ammenwerth (2013). *International Journal of Knowledge-Based Organizations* (pp. 19-36).

www.irma-international.org/article/modeling-automated-examination-communication-processes/76323

A Framework for Synthesizing Arbitrary Boolean Queries Induced by Frequent Itemsets

Animesh Adhikari (2013). *International Journal of Knowledge-Based Organizations* (pp. 56-75).

www.irma-international.org/article/framework-synthesizing-arbitrary-boolean-queries/77884

Social Media Driven Management Education

Owen P. Hall Jr. (2017). *International Journal of Knowledge-Based Organizations* (pp. 43-59).

www.irma-international.org/article/social-media-driven-management-education/176917