

User-Oriented Filtering of Qualitative Data

Carsten Felden, Peter Chamoni

Universität Duisburg-Essen, Lotharstraße 65, 47048 Duisburg,

{c.felden/chamoni}@uni-duisburg.de

Abstract. Several business intelligence concepts show that it is possible to build a data warehouse based on heterogeneous quantitative and qualitative data. But the danger of flooding the management with information still remains. Business information should be made available according to the personal need of a manager by a self-defined push mechanism. The concept of active data warehousing is going to be expanded in this way, first. Messages about new relevant information should be created for quantitative and qualitative data as well. Due to this, there is a need for an enterprise specific ontology. This ontology works as an intermediate between current information and user profiles. Just interesting information are passed to the user. The second approach clusters qualitative data in favour of visualization. The usage of increasing cell structures provides a self organizing map. Decision makers will get the opportunity to search in document volumes, which are clustered.

1 Introduction

The information to decision makers is nowadays afflicted with the problem of information overload [14]. The World Wide Web (WWW) offers the opportunity, to integrate textual information (qualitative data) from enterprise external sources. This leads back to the problem of information overload.

In this paper we propose the enhancement of an existing market data information system (MAIS) with internal and external data by two methods in favour of text documents analysis. The first approach realizes the automation of individual information filtering in the sense of an active data warehouse to reduce the information overload. Due to this, the user profile plays an important role, because an individual filtering of information for decision makers can be made according to his profile. This enables decision makers to fulfil their tasks faster, because of the selection and individual representation of decision relevant information. The second approach supports the more efficient information search of a user by clustering text documents, to enable a more powerful retrieval. The results of our findings are summarized in a conclusion.

2 Conceptual Frameworks

There is an implicit need of fast decision-relevant information support. This leads to information systems which can recognize situations automatically and enforce user actions. Thus the data warehouse nowadays changes into an active data warehouse which provides a high availability of business-critical information.

2.1 Basic concept

There already exist approaches to integrate quantitative and qualitative data in a data warehouse. This has to be understood as a basis for the further development, whereas an active data warehouse should also refer to qualitative data [4]. The underlying information system was realized in a project of the VEW Energie GmbH and the University of Duisburg.

There are several approaches for the distinction of interesting and uninteresting internet documents to be mapped into a data warehouse. For this reason, non-overlapping classes are used, in order to make a clear separation. The examination of the documents and the setting up of classification criteria has to be differentiated. Alternative automatic procedures are statistic methods, e.g. the Rocchio algorithm, Support Vector Machine (SVM), or artificial neural networks [18]. In context of MAIS development, a neural network was used for relevance determination. The integration process is illustrated in the following figure 1.

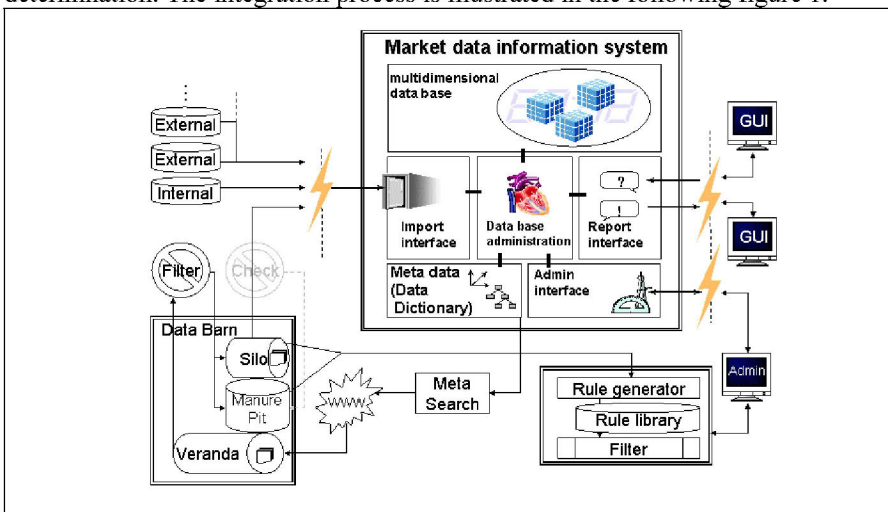


Fig. 1. Market data information system [7]

The set of dimensions of the implemented multidimensional data model are the basis of the retrieval process to find external data. If a coded retrieval query refers to a dimension itself, the result has to be linked to this dimension. If the coded