International Journal of Information Management 35 (2015) 662–671

Contents lists available at ScienceDirect

International Journal of Information Management

journal homepage: www.elsevier.com/locate/ijinfomgt

On the relevance of reports—Integrating an automated archiving component into a business intelligence system

Michael Schulz*, Patrick Winter, Sang-Kyu Thomas Choi

School of Business Administration & Economics, Philipps University at Marburg, 35037 Marburg, Germany

ARTICLE INFO

Article history:
Received 10 February 2015
Received in revised form 26 June 2015
Accepted 21 July 2015
Available online 11 August 2015

Keywords:
Archiving
Business intelligence (BI)
Operational BI
Information storage

ABSTRACT

In the last years, the scope of business intelligence (BI) systems has been extended from strategic to operational decision support (operational BI). This has led to an increase in the number of information needs and, at the same time, to a decrease in the "efficiency" of reports in terms of how many information needs they address. As a consequence, the number of reports has exploded. This slows down knowledge workers' manual or automated search for information, resulting in high search costs to companies. However, it can be observed that in many cases only a small subset of all reports is (still) relevant to knowledge workers. The remainder is an unnecessary burden that could be sorted out without obstructing the access to information that still is needed. In this paper, we develop a framework to identify such reports and archive them automatically. The relevance of reports is concluded from users' information retrieval behavior as recorded in the log files of the BI system, particularly of its search component. We evaluate the proposed framework through a simulation study. The results indicate that the integration of an automated archiving component into a BI system can significantly reduce search effort and, hence, search costs.

1. Introduction

It is long understood that a company's competitiveness largely depends on how *effectively* it can make use of information (e.g., Menon & Varadarajan, 1992). Due to an increasing volume and variety of data available for analysis on one hand and changes in the audience of information systems (IS) on the other hand, however, the number of information resources stored in many IS has significantly increased in the last years. As a consequence, it has become difficult for knowledge workers to locate relevant information in reasonable time (that is, *efficiently*) because distinguishing between relevant and irrelevant information resources takes (too) long (e.g., Davenport & Beck, 2000). Enabling them to do so, therefore, is a major challenge to modern information management (IM).

Not meeting this challenge can have severe negative consequences for companies. In the extreme case, commonly referred to as information overload (see (Edmunds & Morris, 2000) for a review), it may prevent the effective use of information and, hence, significantly weaken competitiveness. Putting the danger of this happening aside, increasing search effort may cause knowledge workers to base their working process on only a subset of all available information resources that they can look through in a given amount of time (e.g., to the end of a deadline), leading to results of lower quality (Chewning & Harrell, 1990; Hwang & Lin, 1999; O'Reilly, 1982). In order to avoid this, they may also try to still find all relevant information, thereby spending a lot of time that they could have spent for other tasks, causing opportunity costs to the company (Cleverley & Burnett, 2015; Haas & Hansen, 2007). The same applies if they recreate existing information resources that they do not find. The impact this has on business, commonly referred to as search costs, has been quantified by the International Data Corporation (IDC), a market research firm specialized in IT (Feldman & Sherman, 2003): "*Using the scenarios outlined above, IDC estimates that an enterprise employing* 1000 *knowledge workers wastes at least* $2.5 *to* $3.5 *million per year searching for nonexistent information, failing to find existing information, or recreating information that can't be found. The opportunity cost to the enterprise is even greater, with potential additional revenue exceeding* $15 *million*

* Corresponding author. Fax: +49 64212826554.
E-mail addresses: michael.schulz@wiwi.uni-marburg.de (M. Schulz), patrick.winter@wiwi.uni-marburg.de (P. Winter), thomas.choi@wiwi.uni-marburg.de (S.-K.T. Choi).

http://dx.doi.org/10.1016/j.ijinfomgt.2015.07.005
0268-4012/

annually" (p. 9). This was over ten years ago; more recent studies (e.g., Schubmehl & Vesset, 2014) give even higher cost estimates.

When considering the general case, approaching the challenge of reducing search effort is difficult. This is because across different IS, usually both, information resources *and* user groups, are heterogeneous, even within one company. Therefore, it is hard to find general patterns of the formers' relevance to the latter. For this reason, we restrict ourselves to a special case in this paper, the case of business intelligence (BI) systems. With these, the company's data are analyzed to derive information that can be used for decision support later on. The results of these analyzes are saved in reports, which are the only and, thus, comparatively homogeneous information resources that exist in BI systems (e.g., Golfarelli, Rizzi, & Cella, 2004). Reports after their creation remain in the system, so that they can be accessed later on by all knowledge workers (who have the necessary rights). The major drawback associated with this process is that while new reports are steadily created, such that are not relevant anymore usually do not get deleted. Therefore, the total number of reports increases over time.

In the past, this has not been a major problem because BI systems back then were almost exclusively used for strategic purposes (Herring, 1988), for which the total number of reports required is relatively low. This has changed in the last years since numerous companies have started to employ BI systems also for operational purposes (operational BI) (White, 2005). As we will elaborate on in more detail later, this has increased the total number of reports stored dramatically – e.g., in a case reported in (Eckerson, 2008) from 1,400 to 4,000 within only one year. This can lead to the aforementioned consequences if no techniques to reduce search effort are employed. Therefore, there is an acute need in BI to develop and introduce such techniques, which is why we focus particularly on this field in this paper.

When aiming to reduce the total number of (active) reports, care must be taken to not obstruct the access to information that still is needed. A natural approach to do so is to sort out reports that have become irrelevant. The difficulty in this, however, lies in discerning these from the remainder. In this paper, we develop a technique to do so automatically. More concretely, we investigate the integration of an archiving component into a BI system that identifies and archives reports based on the information retrieval (IR) behavior of its users (as recorded in its log files). By this, we transfer the concept of archiving from the level of data (Inmon, 2010) to the level of information resources, constituting an information storage (IST)-based approach to IM. We propose a framework for archiving that consists of four parts: which elements an archiving component should have, which types of relevance patterns reports can exhibit, which indicators can be used to infer their relevance patterns, and how the archiving component needs to interact with the BI system's other components. We evaluate our framework through a simulation study.

We structure this paper as follows: in Section 2, we elaborate in more detail on the historical development of BI and how it has affected the report portfolio. We further briefly review and discuss some alternative approaches to reduce search effort. In Sections 3 and 4, we present our archiving framework and the simulation study to evaluate it, respectively. Section 5 concludes this paper with an outlook for further research.

2. Background

2.1. Historical development of BI and consequences

Companies have employed IS to support their business processes for many years now. While the data stored in these systems are recorded for operational use, it soon has been recognized that they also provide a valuable basis for decision support (Sprague, 1980). For this purpose, they are extracted from the operational IS, transformed, and loaded into analytical IS (Moore & Chang, 1980). The latter often are tailored to certain user groups or certain purposes, which is why they exhibit various functionalities and appear under various labels (e.g., "management IS", "expert systems", etc.). In the 1980s, the more general term "business intelligence" became popular (e.g., Gilad & Gilad, 1988). We use this term in this paper to emphasize the goal of deriving information from data, regardless of what happens with this information later on. Nevertheless, the common understanding of BI was still such that its primary application was strategic decision making and its primary audience, therefore, the top and middle management (Hannula & Pirttimäki, 2003; Herring, 1988).

Enabled and, as some may argue, driven by technological progress, the scope of BI systems has been extended in the last decade. The possibility to store and analyze large amounts of data in reasonable time (e.g., through in-memory databases) has motivated companies to base no longer only strategic but also operational decisions on data. While this in the beginning has promised competitive advantages (Marjanovic, 2007), it today has become a necessity to avoid competitive disadvantages (Nadj, Morana, & Maedche, 2015). Doing so within operational IS is difficult, however, because these cannot simply be put on hold for analysis and, further, usually lack the necessary analytical functionalities (such as, e.g., historization of data). As BI systems are separate from operational business and provide these functionalities, it is not surprising that they soon were employed for this purpose (e.g., Marjanovic, 2007), constituting operational BI. As a consequence, the number of reports stored in these systems has increased dramatically, as mentioned earlier. This is essentially due to the following two reasons:

First, an increase in the number of information needs (INs) to be fulfilled with the aid of BI systems (Böhringer, Gluchowski, Kurze, & Schieder, 2009), which on one hand simply results from a lot more knowledge workers being concerned with operational decisions than with strategic decisions. On the other hand, a lot more and more heterogeneous data have to be stored for operational decision support (in particular, disaggregated data). This is amplified by the availability of new data sources (such as, e.g., sensor networks). Once these data are stored in the system, it is likely that they will be analyzed sometime out of curiosity, bringing new INs into being. Because a report can address only a few INs (often just one), many new reports have to be created to fulfill all of them.

Second, a decrease in the "efficiency" of reports, that is, in the ratio between their number and the number of INs they are suited to fulfill. This is caused by a new practice of granting all knowledge workers access to BI systems, so that they can create reports by themselves (self-service BI, SSBI) (Imhoff & White, 2011) instead of having to wait for experts to create them. By introducing this practice, companies have reacted to the observation that the traditional way of supplying knowledge workers with information through the IT is too time-consuming to be efficient and too slow to be effective in supporting heterogeneous decisions as they occur in operational BI (Böhringer et al., 2009). This is particularly true because for supporting operational decisions, INs often have to be fulfilled in (near) real-time (Işik, Jones, & Sidorova, 2013). The problem associated with SSBI is that the new audience of BI systems contains a lot of users with a low expertise in BI. These may not be aware of or fully comprehend the existing reports and, therefore, create new reports to fulfill INs that also could have been fulfilled using the existing ones. Furthermore, they foremostly create so-