Extending the Data Warehouse with Company External Data from Competitors' Websites: A Case Study in the Banking Sector

Dipl.-Ing. Nevena Stolba[1]

Technische Universität Wien

Institut für Softwaretechnik und Interaktive Systeme

Wissenschafterinnenkolleg Internettechnologien

Favoritenstrasse 9-11 / E 188

A-1040 Wien

E-Mail: stolba@wit.tuwien.ac.at

http://wit.tuwien.ac.at

Dr. Beate List[1]

Technische Universität Wien

Institut für Softwaretechnik und Interaktive Systeme

Wissenschafterinnenkolleg Internettechnologien

Favoritenstrasse 9-11 / E 188

A-1040 Wien

E-Mail: list@wit.tuwien.ac.at

http://wit.tuwien.ac.at

[1] This research has been funded by the Austrian Federal Ministry for Education, Science, and Culture, and the European Social Fund (ESF) under grant 31.963/46-VII/9/2002.

Extending the Data Warehouse with Company External Data from Competitors' Websites: A Case Study in the Banking Sector

Organisations require a data warehouse that facilitates modern performance measurement in order to support managements' decision making process. Current data warehouse systems represent traditional performance measurement and lack a balanced set of indicators. Furthermore, the performance information is based on a closed system that retrieves its data from company internal sources only. Changes in the organisation's environment are ignored, although significant changes even require an adjustment of business objectives or eventually business strategy. Consequently, the performance of an organisation must not be measured in isolation. A data warehouse that facilitates modern performance measurement is supposed to provide a single source of information on the performance of the company and its environment. This paper illustrates how a data warehouse facilitates modern performance measurement by the integration of company external data into the data warehouse. As a proof of concept in a business environment, a feasibility study has been implemented in the banking sector.

1 Motivation

Modern organisations utilise a data warehouse for performance measurement. But building a data warehouse is still very much driven by technology and well-established methodologies for the development process are not available. State-of-the-art performance measurement theories are not associated with data warehouse development at all. Data warehouses of large companies support mostly financial analyses (e.g. turnover, cost, margin, etc.) and to a much lower degree customer relationship management (e.g. customer satisfaction, customer retention, new customer acquisition, customer profitability, market and account share, etc.) (Chamoni and Gluchowski, 2004). Present data warehouses retrieve the data from company internal data sources only, e.g. operational systems, legacy databases or other data that is created inside the company. Unfortunately, today data warehouses represent mainly the traditional way of performance measurement.

Traditional performance measurement is primarily focused on financial reporting and represents an inward focus on the business. In a global environment, the markets are getting more and more dynamic and hard to predict, and change

occurs from outside. Significant changes in an organisation's environment often require an adjustment of business objectives or eventually business strategy. Consequently, internal data is insufficient for successful decision making. To obtain a non-isolated view on the performance of a business, external data (such as data gathered from current and potential customers, competitors or suppliers) are required. Drucker (1998) recommends the development of information systems that collect external data in order to facilitate non-traditional performance measurement.

The utilisation of external data is fully accepted in the performance measurement community (Kueng, Wettstein and Meier, 2001). This understanding has to be gained by the data warehouse community as well. Combining internal data and external data in a data warehouse provides decision makers with the ability to analyse business activities not in isolation but rather in relation with the organisational environment.

In order to integrate the external environment, Hackathorn (1999) argues for the extensive usage of Web content as well as the integration into a data warehouse. From his perspective "The Web is the mother of all data warehouses!". He points out that "Information derived from the Web has terrific potential if only we can make business sense of it all" but "the immense resources of the Web are largely untapped". Hackathorn has promoted the integration of Web sources into data warehouses extensively. He created the term Web farming and defined it "as the systematic refining of information resources on the Web for business intelligence".

As a further step towards a data warehouse that facilitates modern performance measurement, this work integrates company-external data from Web sources. We utilise a wrapper technology to extract data from competitors' websites and to render the results into the data warehouse. A feasibility study has been realised in the banking sector, in order to demonstrate firstly, that the integration of company-external data extends the analysis capabilities of the data warehouse heavily, secondly, that competitors' websites do provide data appropriate for performance measurement and thirdly, that existing wrapper technology is well-suited for data extraction from company-external Web sources in order to integrate the selected data into a data warehouse.

The paper is structured as followed: In section 2 the position of company external data in organisations is discussed. Section 3 presents an overview of the utilised wrapping tool LIXTO. In section 4 the organisational setting of the bank, the architecture of the bank's data warehouse and the integration process of external data into the data warehouse is presented. Some analysis examples are given in section 5. Section 6 provides an overview of related work. The paper is concluded in section 7.

2 Company External Data

Company external data is becoming more and more important. Today, an enterprise must know more about its customers, its suppliers, its competitors and other external factors. What matters is, how a company is doing compared with its current competitors stated Eccles more than a decade ago (Eccles, 1991). Peter Drucker admonishes IT executives to look outside their enterprises for information. (Drucker, 1998). He remarked that the single biggest challenge for the next 10 to 15 years is to collect outside data because change occurs from the outside. He predicted that the obsession with internal data would lead to organizations being blindsided by external forces. He stated further, to focus inward on costs and efforts, rather than outward on opportunities, changes, and threats is becoming an increasingly dangerous tendency considering the globalization of economies and industries, the rapid changes in markets and in consumer behaviour, the crisscrossing of technologies across traditional industry lines, and the increasing instability of currencies.

Meanwhile, monitoring external changes can be characterised as an aspect that generated a degree of consensus within the performance measurement community (Kueng, Wettstein and Meier, 2001): “Performance Measurement Systems (PMS) should monitor changes in the organization’s environment. If the changes in the environment are significant, the business objectives and eventually the business strategy must be changed. Consequently, changes in the environment may determine the performance indicators to be measured. Some authors, (see Simons, 1999) suggest that a PMS should include an external monitor component.”

Data warehouse maturity models (see Hatcher and Prentice, 2004; Chamoni and Gluchowski, 2004) have been established in order to enable an organisation to objectively evaluate their use of existing information resources, and to rank themselves on one of those levels. Most organisations can be found at a low level, because beside other factors, the majority of data warehousing efforts take place in enterprises that focus on company-internal data (Hackathorn, 1999). In order to reach a high level, e. g. the Information Evolution Model (Hatcher and Prentice, 2004) requires that “An enterprise has to extend information boundaries beyond its own borders and maximize every opportunity to leverage experience and knowledge to stay ahead of its competitors. Achieving this requires that a company shifts focus from enterprise information management to building the ability to monitor the marketplace and quickly realign the company to meet market demands. To accomplish this, an organisation must establish systems to scan information from external sources.”.

3 LIXTO

Integrated information access utilising various Web sources is a difficult problem using the Web infrastructure of today. Wrapper technology can help to solve the problem by extracting relevant data from Web sources and preparing it for integration. *Lixto Suite* (www.lixto.com) is a product based on technology to access, transform, and syndicate information from Web sources. *Lixto* has been developed by Lixto Software GmbH, a privately owned company located in Vienna, Austria. The latter is a spin-off of the Database and Artificial Intelligence Group at the Vienna University of Technology and the EC3 Electronic Commerce Competence Centre. *Lixto* is comprised of two tools: The *Lixto Transformation Server* (Herzog and Gottlob, 2001) and the *Lixto Visual Wrapper* (Baumgartner, Flesca and Gottlob, 2001).

Wrapper technology is used to extract the relevant information from HTML documents and translate it into XML which can be easily queried and further processed. Based on a new method of identifying and extracting relevant content of HTML documents and translating it to XML format, the wrapper generation tool *Lixto Visual Wrapper* has been designed and implemented. Once a wrapper is built, it can be applied automatically to continually extract relevant information from a permanently changing Web page.

The process of wrapping comprises two steps: First, the identification phase, where relevant fragments of web pages are identified and extracted based on given instance examples. So-called extraction rules are specified semi-automatically by a wrapper designer, utilising the visual point-and-click interface of *Lixto Visual Wrapper*. The first phase is succeeded by the structuring phase, where the extracted data is mapped to some destination format, i.e., enriched with XML tags.

The *Lixto Transformation Server* enables application designers to format, transform, merge and deliver XML data to various devices. It specifies a set of predefined components that can be used to create XML data flow applications. Each component features input and output channels which pass XML documents to subsequent components. The *Lixto Transformation Server* provides a mediation service to integrate various applications and XML data stores. From this integrated data, personalised views can be formatted according to the needs of any delivery channel such as HTML, XML, SMS, databases, data warehouses and email.

Currently, Lixto is conceived by many the most innovative and advanced wrapper generation tool, as pointed out e.g., (iX, 2004) and in the mediation approach of SEWASIE (Beneventano et al., 2003) where the authors describe that Lixto is the most suitable tool for their approach. Moreover, it has been proven that Lixto *Lixto Visual Wrapper* is more expressive than other known wrapper languages and tools (Gottlob and Koch, 2004). Since about two years, Lixto is successfully applied in various domains such as automotive, the energy and the mobile market.

4 Case Study: Extracting Competitors Web Data

4.1 The Organisation: An Online Bank

The described organisation is an online bank, which is primarily operating in Germany. Deregulations of the banking sector, the single European currency, the global European market and numerous mergers as well as the Internet as the enabling technology for online banking have been turning this sector into a highly competitive market. Additionally, online banks face world wide competition and deal with highly flexible customers. In order to cope with future challenges in the marketplace and to keep up with competitors, the banking group's main objectives are: firstly, to deliver high customer value through a reasonable product price and to deliver high quality to clients and secondly, to reduce administrative cost.

The online bank operates a data warehouse, which does not yet contain any external data. We developed a prototype for integrating external data available on the competitors' Website. In order to keep the study concise, we decided to focus on the most important products of the online bank and its competitors:

- Current accounts
- EC cards
- Credit cards

In a future version of the prototype, we are going to integrate other products like loans, stocks, bonds, and savings accounts into the data warehouse as well. The goal of the feasibility study is to integrate competitors' product data, which are available on the Web, in order to analyse it and to compare it with the bank's internal data. Identifying trends and dependences between competitors' pricing or product development policy and the online bank's business provides valuable information to decision makers. Banks' Web sites offer mainly prices, fees, costs or interest rates on their products. We know that the integration of financial measures is insufficient for performance measurement. Nevertheless, we start off with financial measures, because they are easy to gather, well understood by everyone and present to an adequate extent the added value of external data.

According to the Wal-Mart data warehouse development methodology (Westerman, 2001), we derived business questions from the organisation's business goals. We prioritised and selected those business questions that represent the online bank's position in the market in terms of price. The most important business questions we intended to answer by means of integrated external data are:

- How are we positioned on the market with our prices of account costs?
- Are there any dependencies between the pricing policy of the competition and our business?
- What is our position in the credit-card sector?

4.2 Architecture

The architecture of the online bank's data warehouse is presented in Figure 1. This feasibility study extends the data warehouse with external data sources. In order to extract external data from competitors' Websites, we utilised the LIXTO suite. The Web data is extracted and transferred into XML documents. Then, these XML documents are transferred into a relational database (MySQL). After that, external data is ready to be integrated into the bank's data warehouse. Internal data sources are loaded the common way.

To load external data on a regular basis into the data warehouse we have divided the process into the following four steps:

- **Data Discovery:** Identification of competitors and selection of Web sites that contain relevant information to answer the business questions.
- **Data Acquisition**: Set up LIXTO for data extraction on a regular basis.
- **Data Structuring:** Filtering, validating and transforming Web data into a format adequate for integrating into the data warehouse.
- **Data Integration**: Feeding the data warehouse with external data.

4.3 Data Discovery

We identified two main competitors that have an impact on the online bank's strategy and maintain internet sites of significance for performance measurement. We describe the online bank we are dealing with as ***Bank A***, the first competitor (also an online bank) as ***Bank B*** and the second competitor (a traditional bank) as ***Bank C***. The competitors have been chosen manually, since this is a task that cannot be done automatically. Bank B has been selected, because it is the market leader, operates also in Germany and targets a similar customer segment. It is one of the eldest online banks in Germany and is well-known for its competitive prices and numerous customers. Bank A is quite young. It was established 2 years ago. Bank B is a role model for online banks and the direct competitor to Bank A. We have chosen Bank C, because it is a typical traditional bank that has those customers we want to acquire in the future. Consequently, we need to know the conditions of bank C and attempt to be much better.

Generally, there are a lot of other Web pages available that provide daily data on banking products e.g. the Yahoo Banking Center or financial advisors.

In order to simplify the comparison of the products that these three banking institutes offer (Table 1), we unified the product names:

1. Business Account
2. Enhanced Account
3. Basic Account
4. Youth and Student Account

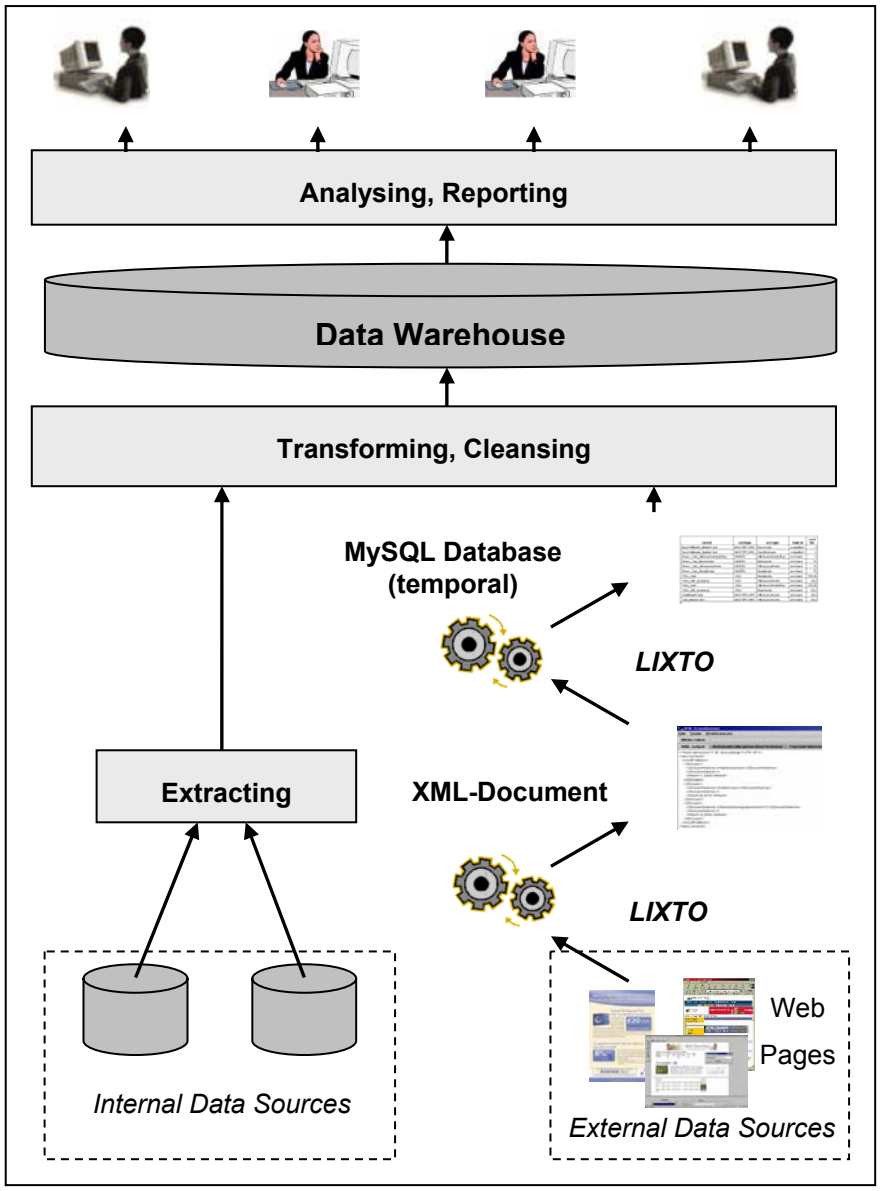

Figure 1: Architecture of the Online Bank's Data Warehouse

	Bank A	Bank B	Bank C
Business Account	X	X	X
Enhanced Account	X	-	X
Basic Account	X	X	X
Youth and Student Account	-	X	X

Table 1 Account Types of Bank A, Bank B and Bank C

The *Business Account* includes an EC card and a credit card, whereby the owner of the account can choose between different credit cards.

The *Enhanced Account* is customised for customers who are interested in "all inclusive charging". It can be extended anytime. This account includes a free EC card and charges the credit card. Bank B offers no Enhanced Account.

The *Basic Account* is designed for selective users with individual demands. Every single activity is charged separately. EC card and credit card are available against extra payment.

The *Youth and Student Account* is customised for teenagers and students. It is much cheaper than the other accounts and offers an EC card and a credit card with a very low limit. Bank A offers no Youth and Student Account.

4.4 Data Acquisition

One of the main goals of this study is to show how much more knowledge on competitors we can gain by analysing their data available on the Web. Therefore, we need to set up a data extraction mechanism that operates on a regular basis. To do so, we have selected the LIXTO Tool Suite. In the following subsection, we select data from the Web and turn it into an XML document. This functionality is provided by the LIXTO Visual Wrapper.

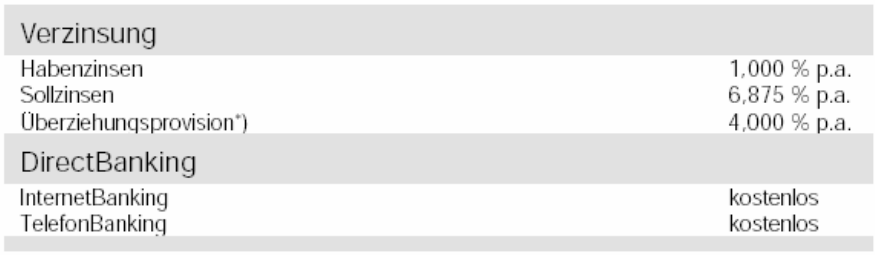

Figure 2: An Example of Web Data to be loaded into the Data Warehouse

Before a wrapper is generated, Web data answering the organisation's business questions have to be selected. Figure 2 presents a Web page that shows an example of data available on the Web. This page is used as an external source for interest rates on accounts. After receiving an internal representation of the example Web page in LIXTO, a wrapper is generated and the required values for extraction are specified. The output of the extractor is translated into XML and the user can choose the HTML attributes that appear in the XML output. Figure 3 shows the XML output of the wrapper program based on the example Web page (Figure 2). Generally, the structure of a Web page does not change very often. However, if a change occurs, the wrapper has to be readjusted.

```
<?xml version="1.0" encoding="UTF-8"?>
<document>
 <rootPattern>
  <Zinsen>
   <ZinsenName>Habenzinsen</ZinsenName>
   <ZinsenName />
   <Wert>1,000</Wert>
  </Zinsen>
  <Zinsen>
   <ZinsenName>Sollzinsen</ZinsenName>
   <ZinsenName />
   <Wert>6,875</Wert>
  </Zinsen>
  <Zinsen>
   <ZinsenName>Überziehungsprovision*)</ZinsenName>
   <ZinsenName />
   <Wert>4,000</Wert>
  </Zinsen>
 </rootPattern>
</document>
```

Figure 3: XML Output of LIXTO Wrapper

4.5 Data Structuring and Data Integration

The XML document has to be transformed into the data warehouse. For this purpose LIXTO offers the Transformation Server, which transforms the XML document into a desired output format (e.g. Mail, SMS, database, HTML). We have converted the XML document into a MySQL database before integrating into the data warehouse, because the LIXTO Transformation Server does not support our selected data warehouse, the Teradata database. We loaded the data from the MySQL database into the data warehouse with the Teradata load utility *FastLoad* and performed some additional transformations and format changes.

5 Case Study: Data Analysis

5.1 Data Model

In this subsection we describe the data model (Figure 4) and some analysis examples. The online bank's data warehouse is organised as 'Hub and Spoke' architecture. The 'Hub' represents a normalised relational database.

The entity ***Account*** stores data on account types of the online bank's own accounts as well as the competitors' accounts. The entity *Account* gets the data from internal systems and from the Web. It includes data about account type, bank id, different costs e.g. internet banking, interest due, debit interest, money transfer costs, etc.

The entity ***Account_Fee*** contains data on account costs. The entity ***Process_Instance*** holds data of the account activation business process. A Workflow Management System is utilised to automate the process. Each process instance relates to a customer account that has been activated. Workflow related control data (e. g. processing time) are stored in the entity *Process_Instance*. The entity is fed by an internal data source and is already an integral part of the data warehouse. ***EC Card*** and ***Credit Card*** are entities that describe the characteristics of different types of EC and credit cards as well as their fees e.g. card fee, additional card fee, card reorder fee, blocking fine, etc.

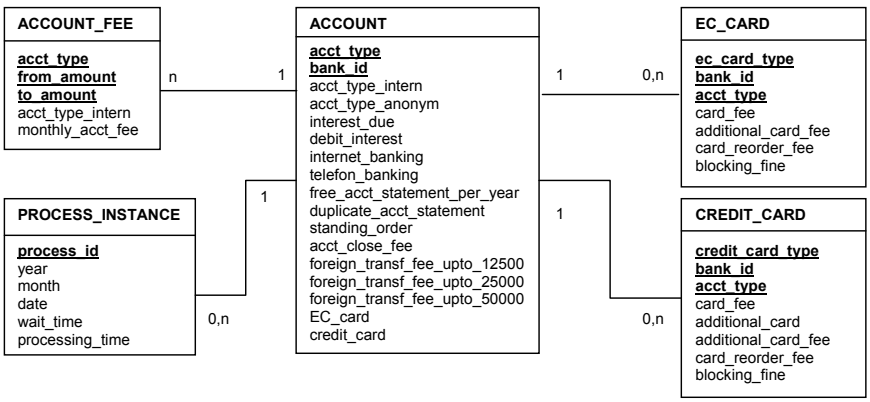

Figure 4: Data Model with Company Internal and External Data

5.2 Business Questions

In this subsection, we answer the three business questions described in subsection 4.1. Again, we would like to stress that financial measures are insufficient for performance measurement. But, the focus of this work is mainly to demonstrate the added value of external data. We use financial measures, because they are easy to gather, easy to process and well-understood by everyone.

Business Question 1: How are we positioned on the market with our prices of account costs?

For this reason, we compared all account costs of Bank A with the account costs of the competitor banks (see Figure 5). The entity *Account_Fee* is queried to answer this business question. The result shows that Bank A is positioned in the middle of the price range. Compared to Bank C, a traditional bank, Bank A offers much better conditions and provides highly competitive products. But the position is worse compared to the competitor Bank B, the other online bank.

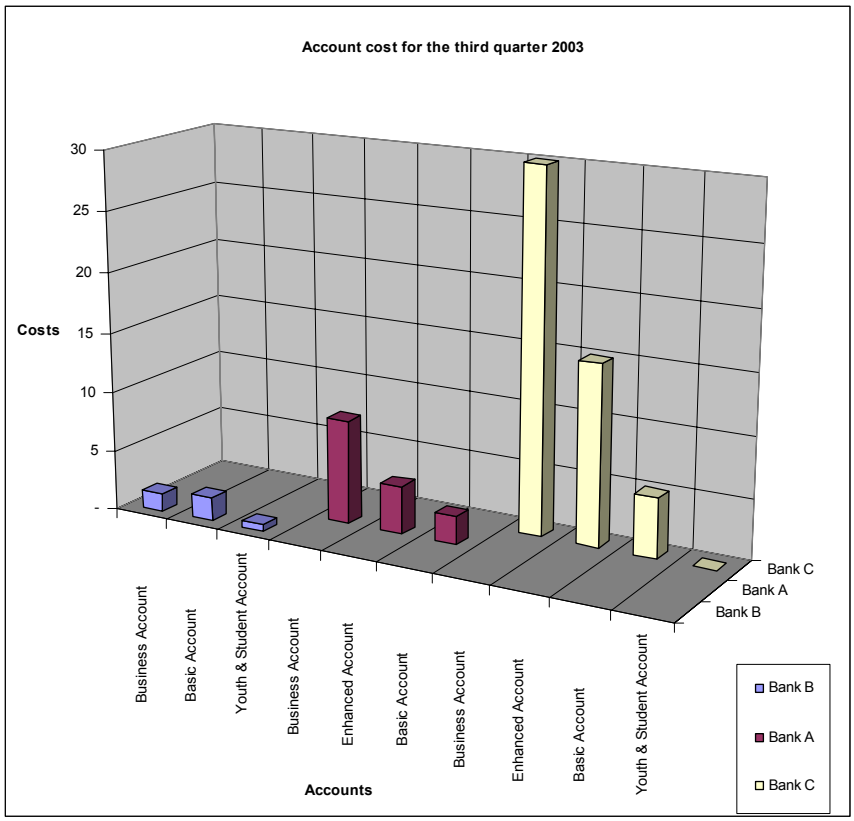

Figure 5: Comparison of Account Cost for the Third Quarter 2003

Business Question 2: Are there any dependencies between the pricing policy of the competition and our business?

Analyses about account activations at Bank A show that there are some periods, where less accounts than usual have been activated. The consideration is to find out, if there were some promotions and special offers on the competitors' side during that time. The entities *Process_Instance* and *Account* are queried to answer this business question. The result of this analysis shows (Figure 6), that most business accounts were activated in July 2003. In August, some of the competitors changed their prices for business accounts and the number of activations dropped massively. As a reaction to this event, the business managers of Bank A corrected the price of the business account and offered some special bonuses for new activations. Afterwards, Bank A started winning new customers in this segment. A similar situation has been observed in November 2003, for resident accounts (Account_Basic). Here, we found through the analysis of the competitors' Web data, that one of our competitors had decreased the price of account costs for new customers by 25%.

Although it is widely known that the price is not the only factor in the race to gain and keep customers, Bank A has made the experience that special offers have a huge impact on the number of activations. The reason for this could be that the bank is part of a financial service group and potential customers are addressed personally.

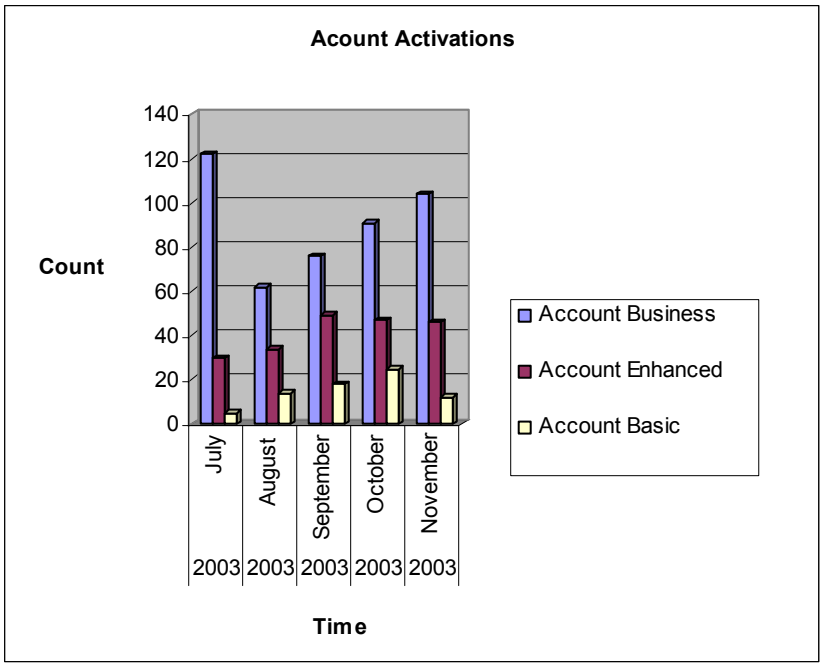

Figure 6: Number of Account Activations at the Online Bank

Business Question 3: What is our position in credit-card sector?

The last emphasis in this study refers to the credit card market. This is an important product for banking institutes and there is a wide variety of different product combinations offered with or without banking accounts (Table 2). The entity *Credit_Card* is queried to answer this business question. Often, additional offers like special credit card packages have an impact on the decision for or against a bank. In this case, one can see that the card fees of Bank A are rather high compared to the competitors. But an advantage is that Bank A cooperates with all three main credit card providers. Both competitors cooperate with one credit card provider only.

	Card Type	Account Type	Card Fee	Blocking Fine
Bank B	MASTERCARD	Business Account	0	36,34
	MASTERCARD	Basic Account	0	36,34
Bank A	DINERS	Business Account	35	35
	MASTERCARD	Business Account	36,33	14,53
	VISA	Business Account	63,95	36,34
	DINERS	Enhanced Account	70	35
	MASTERCARD	Enhanced Account	43,66	14,53
	VISA	Enhanced Account	79,94	36,34
	DINERS	Basic Account	70	35
	MASTERCARD	Basic Account	43,66	14,53
	VISA	Basic Account	79,94	36,34
Bank C	MASTERCARD	Business Account	24,22	36,34
	MASTERCARD	Enhanced Account	36,33	36,34
	MASTERCARD	Basic Account	36,33	36,34

Table 2: Credit Card Cost

6 Related Work

The integration of external data into the data warehouse has received little coverage in the literature. The main focus of related literature is on data extraction and transformation. A semi-automatic transformation approach for the integration of Web data into a data warehouse has been proposed by Zhu, Bornhoevd and Buchmann (2001). In this approach Web data is integrated based on an object model, called MIX (Metadata based Integration model for data X-change). The MIX model represents data together with a description of their underlying interpretation context, and uses domain-specific ontologies to enable an interpretation of the available data and metadata. The authors compare the structure of MIX objects and data warehouse tables, designed as star schema to identify semantic correspondences between them. These correspondences are explicitly described as mapping rules, based on which the transformation can be accomplished automatically via tree restructuring. The focus of the work conducted by Zhu, Bornhoevd and Buchmann is on the extraction and transformation of Web data. Our focal point is complementary. In our work data extraction and transformation is covered by a commercial tool, called LIXTO. We integrate company-external data in order to present innovative ways to balance performance measures and extend the performance measurement capabilities of a data warehouse. A case study in a commercial environment has not been conducted by Zhu, Bornhoevd and Buchmann (2001).

Criteria for evaluating Web sites to be used as input in Web warehousing (integrating data extracted from the web into data warehouses) have been presented by Zhu and Buchmann (2002). The authors identified 12 criteria and grouped them into three categories: source stability, data quality, and application-specific or contextual requirements. Based on these criteria, four Multi-Criteria Decision Making (MCDM) methods have been applied to evaluate and screen Web sources. These selection criteria could be integrated into our research work, in order to raise the quality of performance measures.

Vrdoljak, Banek and Rizzi (2003) propose a semi-automated methodology for designing Web warehouses from XML sources modelled by XML Schemas. In the proposed methodology, a design is carried out by first creating a schema graph, then navigating its arcs in order to derive a correct multidimensional representation. Relevance is given to the problem of detecting shared hierarchies and convergence of dependencies, and of modelling many-to-many relationships. The approach is implemented in a prototype that reads an XML Schema and produces as output the logical schema of the warehouse. In a previous work Golfarelli, Rizzi, and Vrdoljak (2001) proposed a semiautomatic approach for building the conceptual schema for a data mart starting from the DTDs describing the XML sources. Again, our focus is complementary. We do not address data extraction or transformation; we focus on modern performance measurement.

7 Conclusion

A data warehouse that facilitates modern performance measurement is supposed to provide information on the performance of the company and its environment. In order to move towards these requirements, we have extended a data warehouse by the integration of company-external data from competitors' Web sources.

A feasibility study has been implemented in the banking sector and can be seen as a proof of concept of this approach in a commercial environment. The analysis examples of the extended data warehouse provide an added value to the bank and show that company-external data extend the analysis capabilities of a data warehouse in general. The focus on financial measures and on a narrow set of business questions is a limitation of the feasibility study. However, the need for integrating company-external data into the data warehouse has been demonstrated. Further work will address this limitation and extend the data warehouse with non-traditional performance measures.

Lixto, a commercial tool has been utilised to access and integrate Web data into the data warehouse of the online bank. Lixto offers a user-friendly interface that facilitates to conveniently generate wrappers and transformations. An adjustment of a generated wrapper is only required when the layout of the Web page accessed changes dramatically. In the banking sector, we have experienced stable Web pages that are updated regularly. This is a basic requirement for setting up wrapper technology. Nowadays, there is plenty of data available on the Web that can be extracted for performance measurement purposes. All banks have Web pages that provide detailed data on their products. Additionally, there are Websites e.g. the Yahoo Banking Center that compare products of different banks on a daily basis. Our feasibility study has provided evidence that wrapper technology in general, and Lixto in particular is well-suited for extracting data from company-external Web sources in order to integrate the data into the data warehouse. Furthermore, the approach and technical realisation of this study are ready to be applied in an operational data warehouse environment.

8 Acknowledgement

We would like to thank Lixto Software GmbH (www.lixto.com) for providing a free software licence and for their active support during our case study.

This research has been funded by the Austrian Federal Ministry for Education, Science, and Culture, and the European Social Fund (ESF) under grant 31.963/46-VII/9/2002.

9 References

Baumgartner, R., Flesca, S., Gottlob G.: Visual Web Information Extraction with Lixto, 27th International Conference on Very Large Data Bases, Morgan Kaufmann 2001.

Beneventano, D., Bergamaschi,S., Guerra, F., Vincini, M.: Building an integrated ontology within SEWASIE system, Workshop Semantic Web and Databases at the 29th International Conference on Very Large Data Bases, 2003.

Chamoni, P., Gluchowski, P.: Integrationstrends bei Business-Intelligence-Systemen Empirische Untersuchung auf Basis des Business Intelligence Maturity Model, Wirtschaftsinformatik 46 / 2, Vieweg Verlag, 2004.

Drucker, P.: The Next Information Revolution, Forbes ASAP, 1998.

Eccles, R.: The Performance Measurement Manifesto, Harward Business Review, 1991.

Golfarelli, M., Rizzi, S., Vrdoljak, B.: Integrating xml sources into a data warehouse environment. Proceedings IEEE International Conference on Software, Telecommunications and Computer Networks (SoftCOM 2001), 2001.

Gottlob, G., Koch, C.: Monadic Datalog and the Expressive Power of Languages for Web Information Extraction, Journal of the ACM 2004.

Hackathorn, R.: Web Farming for the Data Warehouse, Morgan Kaufmann 1999.

Hatcher, D., Prentice, B.: The Evolution of Information Management, Business Intelligence Journal, TDWI, Spring 2004.

Herzog, M., Gottlob, G.: InfoPipes: A Flexible Framework for M-Commerce Applications, TES Workshop at the 27th International Conference on Very Large Data Bases, 2001.

iX, Gekelterte Semantik, Magazin für professionelle Informationstechnik 7/2004, Heise Verlag 2004.

Kueng, P., Wettstein, Th., Meier, A.: Performance Measurement Systems Must Be Engineered; Communications of AIS, Volume 7, Article 3, 2001.

Simons, R.: Performance Measurement Control Systems for Implementing Strategy, Upper Saddle River NJ: Prentice Hall 1999.

Vrdoljak, B., Banek, M., Rizzi, S.: Designing Web Warehouses from XML Schemas, Proceedings 5th International Conference on Data Warehousing and Knowledge Discovery (DaWaK 2003), Springer-Verlag, Lecture Notes in Computer Science, 2003.

Westerman, P.: Data Warehousing using the Wal-Mart Model, Morgan Kaufmann 2001.

Zhu,Y., Bornhoevd, Ch., Buchmann, A.: Data Transformation for Warehousing Web Data, Proceedings of the 3rd International Workshop on Advanced Issues of E-Commerce and Web-Based Information Systems, 2001.

Zhu,Y., Buchmann, A.: Evaluating and Selecting Web Sources as External Information Resources of a Data Warehouse, Proceedings of the 3rd International Conference on Web Information Systems Engineering (WISE02), 2002.