Association for Information Systems
AIS Electronic Library (AISeL)

AMCIS 2000 Proceedings

Americas Conference on Information Systems (AMCIS)

2000

Vertical Integration of Business News from the Internet within the Scope of Strategic Enterprise Management (SAP SEM)

Marco Meier
University of Erlangen - Nuremberg, marco.meier@wiso.uni-erlangen.de

Alexander Fuelleborn
SAP, alexander.fuelleborn@asp.com

Peter Mertens
University of Erlangen - Nuremberg, mertens@wiso.uni-erlangen.de

Vertical Integration of Business News from the Internet within the Scope of SAP Strategic Enterprise Management (SAP SEM)™

Marco Meier, Department of Information Systems I, University of Erlangen-Nuremberg, marco.meier@wiso.uni-erlangen.de
Alexander Fülleborn, SAP AG, Development SEM Business Information Collection, alexander.fuelleborn@sap.com
Peter Mertens, Department of Information Systems I, University of Erlangen-Nuremberg mertens@wiso.uni-erlangen.de

Abstract

In order to improve decision support through knowledge management it is becoming increasingly important to link Enterprise Resource Planning (ERP) systems with data warehouses, business intelligence and knowledge management applications. In the context of efficient definition and execution of strategies it is furthermore a key issue to combine internal and external as well as quantitative and qualitative information. The Internet is already one of the most important media for accessing external data and it might continue to grow in significance. In the research project MINT (**M**anagement Information from the **INT**ernet) supported by SAP AG and the Bavarian Research Center for Knowledge-based Systems the prototype of an Editorial Workbench has been developed. This system helps to manage knowledge spread in internal and external sources in order to distribute the right information to the responsible manager in time. SAP decided to use the prototype for the development of their new product Strategic Enterprise Management (SEM)™.

Introduction

In the last years a lot of companies successfully managed to support their operative activities with traditional Enterprise Resource Planning (ERP) systems. Because of the fact that value creation has become one of the most important topics in enterprise management a new focal point is on strategic issues. However, the key is not only to define the right targets. A characteristic of the top performers is their ability to execute strategies faster than their competitors. For that reason there is an increased need of vertical integration. That means to link ERP systems with data warehouses, business intelligence and knowledge management applications.

Almost all big ERP suppliers offer new products in order to meet this challenge. One of the first and most comprehensive approaches is SAP Strategic Enterprise Management™ (SAP SEM), which represents an additional layer beyond the underlying ERP systems and the data warehouse (SAP Business Information Warehouse (SAP BW)). So it ends up in a vertical integration solution. The system consists of five modules for business information collection (SEM-BIC), planning and simulation (SEM-BPS), consolidation (SEM-BCS), corporate performance monitoring (SEM-CPM) and stakeholder relationship management (SEM-SRM). The focus of the research described in this paper is on the collection of external business news from the Internet for decision support, which is covered by SEM-BIC. Additional facts about the other components of SAP SEM can be found at (SAP 1999).

Several studies prove that today less than half of the information required for strategic decisions comes from data that is at hand within the company (Bauer 1996, p. 46). However, the external data available is for the most part of a qualitative nature. An analysis by HERGET and HENSLER showed, for example, that publicly accessible databases contain approximately 72 % textual, 19 % numeric and 9 % other information (Herget et al. 1995). Various organizational units in the company already collect unstructured information from numerous data sources for environmental scanning. This is of vital importance to the company and includes news on competitors as well as technological developments. This information is only partly – if at all – integrated into Management Information Systems (MIS). The principal reasons for this situation are the high costs for integrating information from different media sources and the absence of effective information logistics in large companies. Furthermore, neither the function range nor the data models in the current systems are suitable for linking qualitative and quantitative data from internal and external sources adequately (Meier et al. 1999).

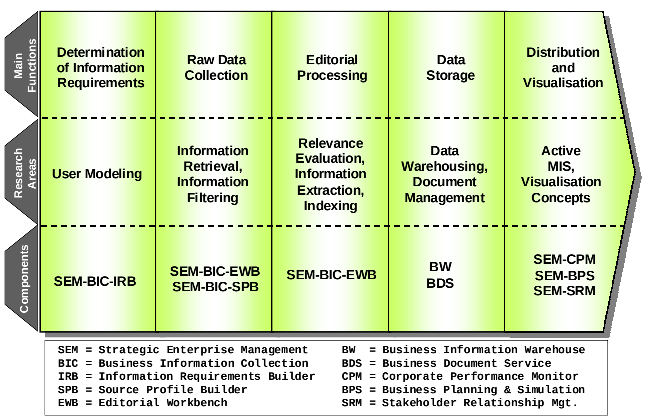
Present technological developments offer promising solution approaches to tackle these problems. The most significant ones referring to the vertical enhancement of ERP systems are the rapid expansion of business news on the Internet and recent developments in natural language processing, especially text mining. The Internet allows direct access to most of the popular, professional information providers, such as market research institutes, online databases or newspaper archives. Almost every company maintains a web site containing press releases, financial and product information, etc. Since the quality and quantity of data from the Internet still varies depending on the subject area, the retrieval results need to be edited to enable a systematic integration into MIS (Mertens 1999).

To bridge the gap between the described business needs and technology pressure was the trigger for the project MINT (**M**anagement Information from the **Int**ernet). The objective was to gain insights in approaches for computer-aided procurement of external management information from the Internet as well as their integration into decision support systems. Apart from the combination and advancement of available theories in different areas, like user modeling, information retrieval or text mining, the work pursued also a very pragmatic research target. This aim was to design and implement the prototype of an "Editorial Workbench". Based on information requirements of managers, this system was to support the collection and filtering of external facts from the Internet as well as linking them with internal management information.

Methodology

The development followed the exploratory model (Sametinger et al. 1992). Based on a pre-study about the suitability of the Internet for MIS in autumn 1996 a rough function and data model as well as a first prototyp version have been designed within six months. It procured a first impression about the look and feel of the system. Thereafter, influenced by permanent feedback of potential users, the models and the prototype have been particularized and enhanced. Partners were customers of SAP Enterprise Controlling (SAP EC) software from different industries, controllers, management consultants, press departments and information brokers.
The basic framework for the development of the Editorial Workbench was the value chain of an information supply process, as shown in Figure 1.

Figure 1: Information Supply Process



It starts with identifying the information requirements of the management. In the next step raw data, consisting of documents that probably contain relevant facts, are collected from the Internet. Before this is stored and distributed to the decision makers, it has to be processed editorially by filtering important content and linking it to internal management information
The way that we handled the problem was first to investigate the state of the art for research areas that could contribute to a solution. In the second row of Figure1 these areas are assigned to the main functions. The result was that in computer science some sophisticated concepts already exist. But often they were simply not practical and the best solutions concentrated only on parts of the information supply process. So the approach was to combine and adapt the methods by evolutionary prototyping.
Finally, SAP AG decided to use the prototype for the development of the new SAP product Strategic Enterprise Management (SEM), especially the component Business Information Collection (SEM-BIC). The last row in Figure1 shows the modules within SAP SEM which will cover the main functions. This paper refers to the components only as far as the MINT project is concerned. More detailed information about SAP SEM is available at (SAP 1999).
The development of the commercial product module SEM-BIC started after two years. Half a year later the first pilot installation at a pharmaceutical company was initiated. SAP itself uses the system for market monitoring. The first experiences have been evaluated by semi-structured interviews (see evaluation chapter below).
One member of the R&D-team beard the main responsibility of the project which lasted in total three and a half years. He has been supported by four students and several developers from the industry partner.
The model proved to be useful, because the requirements were not exactly known from the beginning and available technologies in the Internet environment changed rapidly. As there was no complete specification, the system could not be evaluated by comparing it to its specification. Instead, adequacy of the system had to be evaluated by subjective judgments.
Some interface problems occurred during the prototyping phase because of the rapid parallel development of the several modules. At last the prototype met the requirements, since the target was not a stable productive system, but a concrete practical blueprint for the development of the commercial system which will fulfill all requirements for software quality.

Editorial Workbench

Determination of Information Requirements

As a first step, information requirements of managers have to be determined in order to identify the suitable keywords for searches in the Internet. This kind of user modeling is also a prerequisite for an active, individualized distribution of the results. It has to be specified who (receiver) needs which information (content) at what time (trigger).

Receivers could be single decision makers or groups of managers. This solution refers to the concept of stereotypes in user modeling by RICH (Rich 1979). The component Information Requirements Builder offers role-based profile templates, for example, for CIOs or product managers. A content specification is created by combining business objects and keywords. The Information Requirements Builder prototype offers various selection menus for the characteristics "company", "industry sector" and "region". The content description is completed by keywords that refer to the selected characteristics. The details of interest concerning customers might be their incoming orders, stock price or human resources policy. Relevant facts for a country would be demographic statistics such as population structure or labour market data. It is also possible to define own keywords. Further elements of information requirements are triggers that determine at what time the retrieval process starts.

Raw Data Collection

The next step is to support the collection of data from the Internet. The objective is to relieve professional in-house information brokers of routine work so that they can concentrate on more demanding editorial tasks. Specific activities in raw data collection include selecting suitable sources, preparing and executing queries, monitoring contents for changes as well as initial data check-in.

The process of finding the sources that fit an information requirement may be compared to the selection of suppliers in an industrial company. The proposals are determined by metadata regarding sources and selection rules. The component uses information categories, frequency of update and regional references as criteria. This type of selection is an attempt to restrict the flood of retrieval results in advance and to increase the probability of finding relevant data.

Following the selection of appropriate sources, the Editorial Workbench automatically generates source-specific queries, for example CGI-strings, insofar as information providers allow direct access.

A further relief of the strain on in-house information brokers is achieved by the integration of tools that monitor changes in the content of web pages, for example of customers or competitors. A detailed description of monitoring functions can be found at (NetMind 1999; Daily Diffs 1999).

Editorial Processing

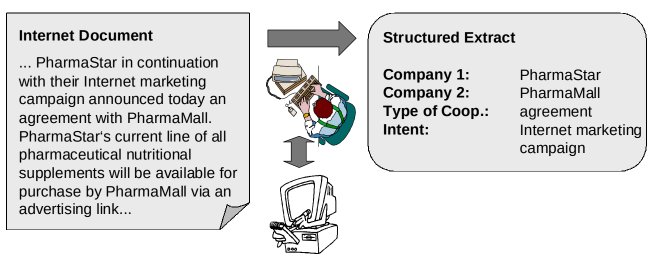
The objective of the editorial processing is to extract relevant facts from the documents and to link them with internal management data. Total automation of this phase is not possible even with state-of-the-art systems. However, there are areas where the system can do preparatory work and support the editor by text mining methods. In the general framework of knowledge discovery in databases, data mining techniques are usually dedicated to information extraction from structured databases. Text mining approaches, on the other hand, support information extraction from unstructured textual data.

The indexing of new messages is split into two processes: the raw indexing and the linking of business objects. The first step requires no user interaction and runs immediately after the transformation operations. Single words are identified by punctuation marks or white spaces, such as blank characters, new lines or tabulators. Terms are reduced to principal forms. In order to determine the infinitive of irregular verbs, such as "sell" from "they sold", an external dictionary called "WordNet" which was developed at Princeton University (Princeton 1997) is used. The Editorial Workbench eliminates plural endings and translates comparatives and superlatives to the corresponding adjectives (Porter 1980).

The results of this procedure are used as a basis for the systems's proposal of a business object to which the message should be linked to. It has to be emphasized that the suggestions are not restricted to terms that are in the original text. By building document vectors, the Editorial Workbench calculates similarity measures of documents and recommends index terms which have been assigned to messages with similar content in the past.

Information Extraction is also divided into two phases: identification of potential extracts and the actual registration (Appelt et al. 1997). The groundwork is the semantic analysis of the document done by the external tool "Link Grammar Parser" (Temperley et al. 1998).

Figure 2: Adaptive Information Extraction



Based on a collection of rules, the Editorial Workbench recognizes predefined events, for example mergers or strategic alliances and the corresponding characteristics such as type of the cooperation and names of the companies involved. The system learns new rules by monitoring the editor. First, he or she can define the structure of new events. Then the computer records how the user completes the required fields using context-sensitive menus (see Figure 2) and tries to deduce principles. A detailed explanation of the basic structure of such rules, which had to be adapted for dynamic document collections, is given by SODERLAND (Soderland 1997). Again the user can accept or reject the recommendations of the system. Every modification contributes to the learning process of the system.

Data Storage

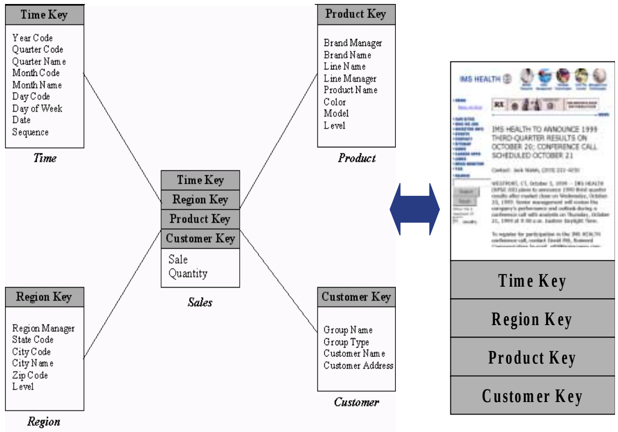
It would be logical to organize external information according to the pattern of internal data storage. For data warehouses, often the star schema is used (Inmon 1996). It is yet to be verified whether this concept is also suitable for qualitative data.

While a key figure such as revenue is assigned to one specific business object (one region, one customer group, one product, etc.), a distinctive allocation like this is not always suitable for textual data. For example, if two competitors merge, this message has to be linked to both companies. However, for the star schema, unique key values are a prerequisite and they allow only one instance for one dimension. As a result, two data records (one for each customer) would have to be built. If further aspects of the news item are taken into account, for example, one company is from Brazil and the other one from the USA, the number of keys rises to four. That is why changing and deleting the data becomes more difficult and could interfere with consistency.

A second argument is that the star schema is widespread because of its advantages in aggregating data for top-down as well as slice-and-dice navigation. For qualitative data this is often not advisable. The two statements: “Biohealth founded a new plant in Brazil.” and “PharmaStar sold $10 million in Brazil.” have to be assigned to the characteristic Brazil, but there is no suitable way to reduce them to one message.

To conclude, the star schema does not offer any advantages for organizing external qualitative data compared with relational comment and indexing tables (see Figure 3). The connection of internal and external data is achieved by using the same vocabulary as in the dimension tables.

Figure 3: Enhancement of the star schema



Distribution and Visualisation within a Management Portal

Finally, external information units have to be distributed together with internal data according to the information requirements. Therefore the prototype of a Management Portal has been designed which enables managers to unlock internally and externally stored information and provide them a single gateway to personalized information needed to make business decisions.

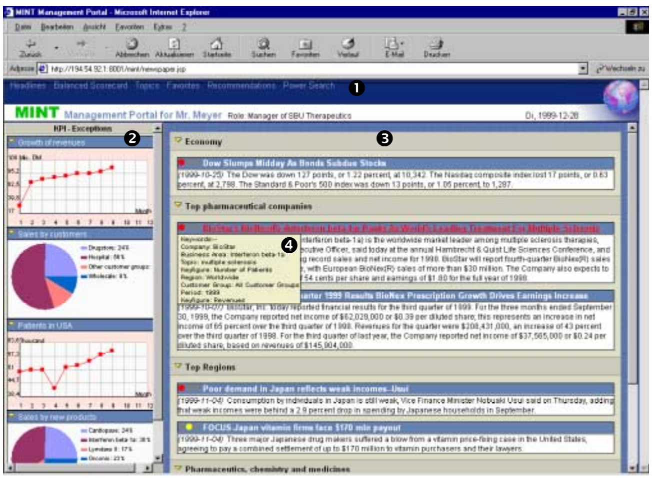
Basically, one can distinguish between active (push) and passive (pull). Both play an important role in modern distribution of information to managers (Mertens et al. 1996; Vandenbosch 1997). Critical news, for example a significant drop in incoming orders, should immediately be delivered by the system to the responsible manager. In order not to aggravate the problem of information overload, one must carefully define what is critical. For quantitative data, this can be achieved by threshold values. Problems arise with qualitative data. Experiences with push services on the Internet show that it is not sufficient to restrict the delivery to some information categories. The amount of messages is still too large, in particular with regard to the limited time of managers (Vandenbosch et al. 1997).

An interesting approach is used at BASF AG. Experts in the various company departments evaluate new messages in their activity fields. If more than 30 %, for example, of these people label the news as “very important”, it is automatically delivered to the next hierarchical level (Mertens et al. 2000). If the decision maker requires more detailed data, navigation by the pull principle is supported. Distribution is closely linked with visualisation methods. The following requirements are important in this context:

1. Integrated representation of internal, external, quantitative and qualitative data
2. Similarity with generally accepted news media for managers
3. Hierarchical and flexible navigation between overview and details
4. Individual preselection of critical news
5. Simple full-text search
6. Intuitive relevance feedback

A promising idea in this context is the portal approach, which is similar to electronic newspapers. In the taskbar ❶ the manager can select different views and functions: headlines, balanced scorecard, topics, favorites, recommendations, full text search. The left frame ❷, in the sense of exception reporting, contains critical developments in key performance indicators (KPI) of the balanced scorecard. In the right frame ❸ current top news from internal and external sources are listed and sorted by relevance. If the mouse pointer touches a headline, a structured extract ❹, containing the assigned keywords, is displayed.

Figure 4: Management Portal

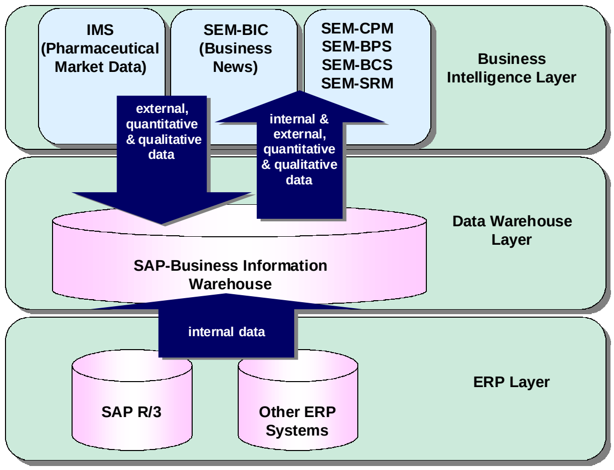


By selecting a headline, the system displays related internal data, e.g. sales reports, and shows the full textual information. Through different menus it is possible to drill down in the internal data. In this case new appropriate external information will be presented in the left frame dynamically.

Evaluation

A popular approach in the context of using data from the Internet for decision support is “Web Farming” by HACKATHORN (Hackathorn 1999). His focus is more on organizational issues. Indeed he lists some single technologies and tools that might be helpful for some parts of the information supply process, but no comprehensive software solution like the Editorial Workbench is described. In that Web Farming and the Editorial Workbench are complementary. The Web Farming approach may be used to set up the organizational framework and the Editorial Workbench as the vertically integrating software solution.

Figure 5: SEM architecture at pilot customer from pharmaceutical industry

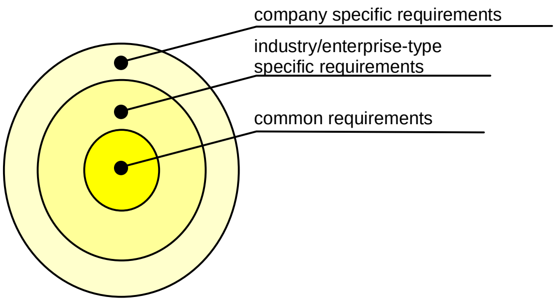


The first market version of SEM-BIC was intensively tested by a pilot customer from the pharmaceutical industry. Figure 5 shows the vertical integration of ERP and business intelligence within the scope SAP SEM at that company. The SAP BW contains internal and external data from ERP systems, the Institute for Medical Statistics (IMS) and Business News from SEM-BIC. The other SEM modules (see also Figure 1) process these data. Although there are still starting points for enhancement and improvement, the first feedback was surprisingly positive. More details about this project can be found at (Meier et al. 2000).

Future Work

The prototype is currently used for the development of SAP SEM. A further objective is not only to deliver software functionality, but also business content; concerning information requirements and sources. In typical cases users should be able to start immediately with these standardized contents or will only have to modify them slightly. Another important issue for future research is to show how information requirements of decision making bodies could be structured within a kernel-shell architecture (see Figure 6). The kernel contains the information requirements which all companies have in common. The elements in the shells refer to specific industry and company problems.

Figure 6: Kernel-shell architecture



Acknowledgements

This research was supporetd by the Bavarian research Center for Knowledge-based Systems, Erlangen-Tennenlohe, Germany, and SAP AG, Walldorf, Germany.

References

Appelt, D.E., and Israel, D. "Building Information Extraction Systems," *ANLP-97 Tutorial*. http://www.ai.sri.com/~appelt/ie-tutorial/, Date: 1999-08-14.

Bauer, M. "Altbekanntes in neuer Verpackung?" *Business Computing* (4) 1996, p. 46.

Daily Diffs (ed.) "Daily Diffs Free Web Page Change Monitoring," http://www.dailydiffs.com, Date 1999-05-20.

Hackathorn, R.D. *Web Farming for the Data Warehouse – Exploiting Business Intelligence and Knowledge Management*, San Francisco 1999.

Herget, J., and Hensler, S. "Online-Datenbanken in Wirtschaft und Wissenschaft: aktuelle Nachfragestrukturen und Nutzungstendenzen," *WIRTSCHAFTSINFORMATIK* (37:2) 1995, pp. 129-138.

Inmon, W.H. "*Building the Data Warehouse*," 2nd ed., New York, 1996.

Meier, M., and Fülleborn, A., "Integration externer Führungsinformationen aus dem Internet in SAP Strategic Enterprise Management (SEM)," *WIRTSCHAFTSINFORMATIK* (41:5) 1999, pp. 449-457.

Meier, M., and Schröder, J. "Integration interner und externer Führungsinformationen in einem Pharmaunternehmen Probleme und Lösungsansätze," *WIRTSCHAFTSINFORMATIK* (42:2) 2000, 137-145.

Mertens, P. "Integration interner, externer, qualitativer und quantitativer Daten auf dem Weg zum Aktiven MIS" *WIRTSCHAFTSINFORMATIK* (41:5) 1999, pp. 405-415.

Mertens, P., and Griese, J. *Integrierte Informationsverarbeitung 2: Planungs- und Kontrollsysteme in der Industrie*, 8th ed., Wiesbaden 2000.

Mertens, P., Hagedorn, J., Fischer, M., Bissantz, N., and Haase, M. "Towards Active Management Information Systems," Humphreys, P., Bannon, L. et al. (eds.), *Implementing Systems for Supporting Management Decision-Concepts, Methods and Experiences*, London 1996, pp. 305-325.

Mertens, P., and Schrammel, D. *Betriebliche Dokumentation und Information*, 2nd ed., Meisenheim am Glan 1977.

NetMind (ed.) "NetMind – Products." http://www.netmind.com/html/products.html, Date: 1999-04-30.

Porter, M. F. "An Algorithm for Suffix Stripping," Program (14:3) 1980, pp. 130–137.

Princeton University Cognitive Science Laboratory (ed.) "WordNet." http://www.cogsci.princeton.edu/~wn/, Date: 1999-04-01.

Rich, E., "User Modeling via Stereotypes," Cognitive Science (3) 1997, pp. 329-354.

Sametinger, J., and Stritzinger, A., "Exploratory software development with class libraries," *Shifting Paradigms in Software Engineering*, Klagenfurt, Austria, 1992, pp. 82-97.

SAP AG (ed.) "SAP Strategic Enterprise Management – The Functions – A Closer Look," Walldorf 1999.

Soderland, S.G. *Learning Text Analysis Rules for Domain-Specific Natural Language Processing*, Amherst 1997.

Temperley, D., Sleator, and D., Lafferty, J., *Link Grammar*, http://www.link.cs.cmu.edu/link, Date: 1999-03-30.

Vandenbosch, B., and Huff, S.L. "Searching and Scanning: How Executives Obtain Information from Executive Information Systems," *MIS Quarterly* (21:1) 1997, pp. 81-107.