Customer Information Quality and Knowledge Management - A Case Study Using Knowledge Cockpit

Jessica Hu, Kuan-Tsae Huang, Kazashi Kuse, Geng-Wen Su, Ko-Yang Wang[1]

Knowledge Management and Asset Reuse, IBM Global Services Consulting Group

Abstract

Information quality has become the number one issue a knowledge worker faces in today's environment due to the high noise to signal ratio of nonessential content in knowledge repositories and the lack of quality measurement for the information. This paper discusses information organization and retrieval technologies for improving the quality of knowledge by reducing the noise to signal ratio, and focuses on techniques to abstract data from search results into comprehensible knowledge for users. A case study concerning a customer engagement is used to illustrate the ideas.

1. Introduction

1.1 Quality Information and Organizational Knowledge

The intellectual capital of a company is the combination of the personal knowledge of its employees and the organizational knowledge that is not affected by personnel changes. Organizational knowledge is the retrievable, reusable, and shared intellectual assets of the corporation such as processes, methodologies, best practices, patents, models, software, business know-how, etc. It represents a significant portion of the non-tangible assets of the corporation. Most companies recognize the strategic importance of their intellectual capital and spend a considerable amount of resources collecting, storing, and protecting these assets. However, only a fraction of the collected information and knowledge are utilized or reused effectively. On average, American corporations analyze only seven percent of the information they collect or generate.

There are many road blocks in transforming personnel knowledge into organizational knowledge and in utilizing organizational knowledge effectively. The following lists provides examples of some of these road blocks.

1. Culture barriers for sharing knowledge - Examples include the "not invented here" syndrome, insecurity about the quality of other people's work, the unwillingness to invest in understanding the work of others, and the insecurity of sharing knowledge with others.

[1] Contact: Ko-Yang Wang, (914) 332-3466, e-mail: kyw@us.ibm.com, IBM GS Consulting Group, Rt. 9 & Rockwood Road, Town of Mt. Pleasant, Sleep Hollow, NY 10591

2. The information quality barrier - This problem is created by the high noise to signal ratio of nonessential content in knowledge repositories and the lack of quality measurement for the information. This problem is particularly serious when the knowledge base of a corporation is growing. It may have a significantly negative impact on the sharing or growth of organizational knowledge.
3. The technology barrier. It is usually difficulty to find or retrieve the appropriate knowledge for reuse. Most employees don't know where, when, or how to find and reuse the intellectual capital of the corporation. Technology can ease this process. It is critical for lowering barriers and encouraging knowledge reuse.

Cultural barriers need to be overcome by changing the measurement and incentive system for knowledge reuse, by transforming the culture of sharing, and by generating executive support. Enabling technology that can minimize the barriers or help to overcome them needs to be develop. Organizational knowledge must be structured so that its meaning can be optimized. A key step for doing this is to associate the domain specific information of documents, also known as meta-data, with knowledge so that it can be used to categorize, organize, search, locate, compare, retrieve, reuse, and share the knowledge. Identifying the meta-data requires nontrivial efforts, time, and investment. Automatic or semiautomatic meta-data recognition is highly desired. The analysis and utilization of knowledge depends on good measurement of the knowledge based on usability, usefulness, and relevance instead of just the accuracy and quantity of the knowledge. Techniques for measuring knowledge need to be developed.

1.2 Transforming Information into Knowledge to Meet Business Challenges

1.2.1 Business Challenges

Today's business challenges include understanding and satisfying customer needs; monitoring and staying ahead of the competition; determining industry trends and adapting to the challenges; increasing market share; and entering new markets. To meet these challenges, companies need information such as:

- Industry Analysis
- Company Profiles
- Product/Service Assessments
- Competitive Analysis
- Trend Monitoring
- Key Issue Identification

1.2.2 Transforming Information into Knowledge/Intelligence

Where can the information required for addressing today's business challenges be found? Information sources can be classified as being internal or external. Internal information sources are available inside a company's fire walls. They represent the information that a company has collected and stored in its legacy systems,

DBMS, and data warehouses accessible through the company's intranet. These data stores include sales reports, human resources, financial reports, news, Web pages, competitor analysis and market intelligence. On the other hand, the Internet has proven to be the most popular external data source. It is an incredibly deep business research tool and repository of information.

Interestingly, the biggest problem people encounter in finding information on the internet or even the intranet is not that they are unable to find the information they are looking for, but that they are unable to digest the over-abundance of data/information that they uncover. This is due to the lack of quality measurements for information to evaluate and filter out noise and abstract critical data into comprehensible knowledge. This problem impinges on the quality of information searching and may even deter users from utilizing information warehouses. Technology must be utilized to help to filter out unrelated data and extract useful knowledge for decision support at every level of business.

The knowledge mining technology we proposed in IQ96 manages information warehouses and solve the problem of information overloading by:

- collecting related data and information using a technology called Web Crawler
- abstracting key information about the data into meta files which can then serve as the basis for information quality measurement
- processing meta-data to classify the data into an organized structure and to abstract the information into new higher level knowledge
- providing visualization support to present the knowledge in a comprehensible way.

1.3 An Overview of the Paper

This paper discusses information organization and retrieval technologies for improving the quality of knowledge by reducing the noise to signal ratio, and focuses on techniques to abstract data from search results into comprehensible knowledge for users. A case study concerning a customer engagement is used to illustrate the ideas.

In section two the knowledge mining technology and the knowledge cockpit framework are discussed. Scenarios for applying the technology is also briefly described. In sections three through five, we discuss the underlying technology for applying knowledge mining to actual problems. In section three, we focus on the information retrieval, knowledge organization, and issue sharing. In section four, we discuss document management issues and the technology to resolve these issues. In section five, we present a case study and the application of a customer knowledge management process in solving a client's business problems. Some lessons that we learned in the case study are presented in the last section.

2. The Knowledge Mining Technology and the Knowledge Cockpit Framework

2.1 The Knowledge Mining Technology

Data Mining is an emerging technology to extract comprehensible and valid information from large relational databases to support business decisions. The technique has been applied successfully to help corporations to analyze the market information and target their sales resources on a focused group of customers. For example, Yahoo analyzes the interests of its users based on their access and search data and sells the information to companies. Master Card analyzes the patterns of credit card transactions to zero in on potential customers. However, the problem with data mining is that only data in the relational databases can be mined. This significantly limits the utility of data mining since large amounts of critical information never get installed into relational databases. Moreover, the long time it takes to input information into relational databases frequently causes it to lose its strategic significance for mission critical decisions. Knowledge mining solves these problems by allowing free form text, current events, and other kinds of documents to be searched in a time efficient fashion.

Knowledge mining [Fayyad, 1995] is a process for abstracting valid, previously unknown, ultimately comprehensible knowledge from a variety of information sources. Knowledge mining dramatically enhances the power of information searching by expanding the diversity of information sources. Timely and comprehensive knowledge discovery is made possible with the integration of current events and transaction data into the sources. Information that was excluded from relational databases can now be harvested.

The knowledge mining techniques utilize traditional data mining techniques such as clustering, classification, value prediction, association discovery, sequential pattern discovery, and similar time sequence discovery. It also incorporates new technologies, such as compilation, tokenization, keyword abstraction, semantic analysis, and concept/feature recognition, as well as meta data analysis techniques such as context and semantics analysis.

A typical knowledge mining application incorporates four steps: 1) information collection, to harvest and filter information; 2) features recognition, to generate features/meta data; 3) meta data analysis to abstract knowledge; and 4) visualization, to present knowledge to the user.

The knowledge mining technique helps business users tap into vast amounts of information in the data warehouse, internet, and corporate intranet. It helps them to transform raw information and data into knowledge, and alert them of critical events or information. This process can change the way businesses make decisions and help them to capitalize on their investment in data like never before.

2.2 The Knowledge Cockpit Framework

The knowledge cockpit incorporates pluggable components into its framework. These components can be classified into three different categories: information warehouse, knowledge mining, and decision support.

Information warehouse components typically serve as the primary information sources for companies where information is made available by agents that actively collect, filter, store, and update information.

Information sources include documents, files, news, logs, transactions, and various kinds of databases. Information warehouse components contain relational databases and management tools, analytical tools, and information retrieval tools.

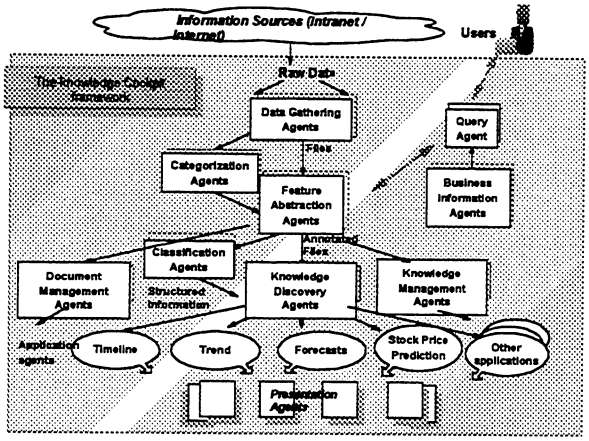

Figure 1. The Knowledge Cockpit framework

Knowledge mining components contain different agents which utilize advanced knowledge mining techniques to process, discover and synthesize knowledge. These agents transform divergent information from different information sources into a consistent network of knowledge. In turn, the decision support components utilizes this produced network of knowledge to assesses the retrieved information's quality, and to translate the knowledge into a compact, comprehensible format. The decision support components interact with the users to make critical decisions based on the interpretation and assessment of the discovered knowledge. More detail of the Knowledge Cockpit Framework can be found in *IQ96*.

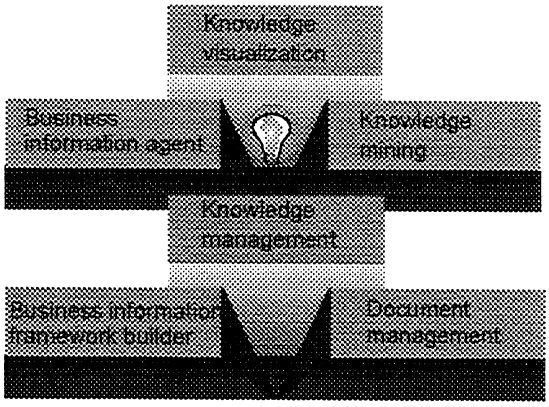

Figure 2. The Knowledge Cockpit Framework.

2.3 Applying Knowledge Mining and Knowledge Cockpit to Business Problems

2.3.1 Customer Knowledge Management

By linking current events into the mining framework, mining technology can be applied into many new areas. One potential application for knowledge mining technology is the customer knowledge management system. The customer knowledge management process creates values for customers by combining and analyzing operational data (such as sales, ordering, billing, customer relationship information) and business intelligence information (such as competitor information, market analysis, current business events) in the corporate data warehouse. Using the knowledge mining and visualization technology, advanced information can be abstracted and passed into the customer value management system. The customer value management system can then use the information to support decision making in customer care and relations management, customer portfolio management, and customer marketing programs. This technology allows enterprises to analyze operations and business intelligence information in real time.

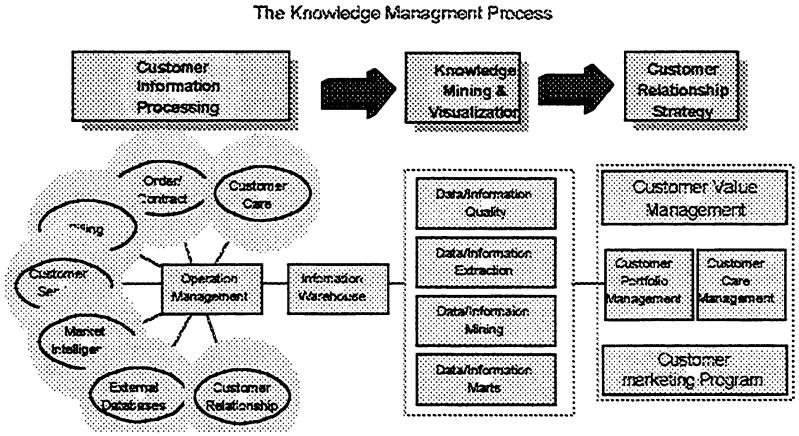

Figure3. The customer knowledge management process.

2.3.2 Information Quality

The value that a customer knowledge management process creates is only as good as the in-flow data quality, i.e. garbage in garbage out. In the real world, data may have been collected in an ad hoc fashion; unfilled fields in records will invariably be found, mistakes in data entry are quite common, data from different databases becomes out of sync over time... As a result, the customer knowledge management process cannot succeed without serious effort to ensure the quality of data.

In the knowledge mining part of the customer knowledge management process, standard data "cleansing" techniques [Fayyad, 1996] need to be applied judiciously to the in-flow data. Every piece of data comes in need to go through uniformity checks to ensure data is within limits, conformity and completeness checks to validate data

with other sources, and its accuracy with the underlying data. Trace need to be designed into every step of the system to instill the credibility of data.

In addition to the data quality, every step of the information extraction and information fusing need to be accounted for with reasons so that a human can trace back the whole process if needed. Since the information or knowledge generated by the customer knowledge management process will be highly summarized, credibility will not be established without such tracing facility. Furthermore, these reasoning can be compared with the human reasoning during real field testing and feed back to the designer for next iteration.

3. Managing Documents with Knowledge Cockpit

3.1. Document management in the information age

The explosion of new information poses a significant challenge to both individuals and organizations. The rate of new information created was estimated to be doubling every twenty years for the last century and is accelerating in a rate approaching doubling every 10 years. The majority of the information are carried in form of documents. We spend a large share of our time creating new documents and revising or recreating old ones. Organizations spend a significant portion of their revenues on handling documents. As a result, management of documents becomes increasingly important to the success of companies in modern information age.

Despite the increasing importance of document management, the traditional document management systems leave a lot to be desired. Most organizations manage their own documents in an Ad Hoc fashion in spite of spending exorbitant amount in maintaining libraries to manage books. Valuable information may be undocumented or inaccessible, carried in people's heads or hidden away on an individual's hard drive and perished with the leaving or retirement of key persons. External information that organizations pay handsomely may languish in someone's drawer while others wondering how to find the competitive information that is so critical to their functions. Documented, accessible knowledge is critical to your organization and can improve decision making, customer service and work flow efficiency.

The traditional document management systems usually stores documents in some kind of database. They require manual indexing of each document before loading into the database. As a result, most new documents will not be immediately available due to the long laborious indexing process. The retrieval of the documents also poses quite a challenge to the users. Somehow, they need to know the exact key of the document to be retrieved. In order to help users, these systems generally provide some sort of catalogue with abstract to overcome this major hurdle, which, in term, takes quite a bit of time and further delay the availability of the new documents. The emphasis of the traditional systems are on the area of efficient storage of documents rather than on the retrieval. Consequently, the benefit of the traditional document management system is not as pronounced as one would expect.

For the last five years, various search engines utilizing all sort of techniques based on Information Retrieval (IR) research have found their way into major information services and the World Wide Web. Many of the features once considered too esoteric have become common in most of the search engines available today. It is

high time for a new generation of document management systems that emphasize efficient retrieval for the users rather than efficient storage for the computers.

In general, there are various IR. techniques [Frankes, 1994, Lewis, 1996] that the new generation of the document management systems can use to help users effectively retrieval relevant documents that they cares. Currently, we understand very well how to do "natural language" queries, therefore relieve the need to know the exact database key for the documents, which are generally an alphanumeric string that bears little relation to the content of the document and are difficult to remember. We also understand very well about ranked retrieval results, term weighting, "query-by-example", and query formulation assistance. Combined with the traditional design of a high power database with different level of archiving storage (memory, hard disk, and tape/cdrom), this new generation of the document management system will provide much more benefit to companies than the traditional ones.

Such a system will provide information at users' finger tip when they need it. However, an efficient document retrieval system can only solve part of an organization's information management problems. It is rarely enough to address the whole range of the organization's information needs. Typically, a complete solution requires other tools such as routing and extraction, handling multimedia, handling OCR documents, integrate with database management systems for structured data, and work flow for managing documents and their use in the organization. Currently, these systems must be integrated using customized software, and even then the integration is often very superficial.

To summarize, there is a need for a new generation of document management system that integrates IR. techniques with traditional database systems. Such a system can be further integrated with current data mining tools to discover patterns, associations, and exceptions among documents. It will provide synergy between various disciplines and tools and therefore, the maximum benefit for the users.

3.2 Managing Documents with Knowledge Cockpit

3.2.1 Organization of documents in Knowledge Cockpit

There are two distinct types of organizing documents in Knowledge Cockpit. The first type are those documents with hypertext links. Most of the web pages we collect over the World Wide Web(WWW) fall into this type. The main issue of storing this type of documents lies in the difficulty of maintaining the links. The dynamics of WWW is such that pages come and go without notifying anyone who links to them. As a consequence, users often see the disappointing message of "page does not exist" when they finally get something they feel some interest. Such a chaotic, unmanaged network works out fine with an anarchical, democratic, loosely coupled community but will not be acceptable in business environment with any degree of seriousness.

In Knowledge Cockpit, we choose to take advantage of the automatic link generation feature that is implemented in the Lotus Notes. Basically, we restructure the web pages into hierarchy of Notes databases that somewhat preserve the hierarchy of their hypertext links. We, then, point to the databases from the Cockpit

navigator instead of point to the individual pages. Lotus Notes will automatically generate the links to each documents in the database. Since databases that serve as the container of documents are much more stable than individual pages. We do not have to spend a lot of effort to add or delete the links between them. The Lotus Notes will automatically manage the links within the databases when add/delete pages occur. Otherwise, each time an add/deletion of a page occurs, we have to manually update all the pages that point to that particular page. It is very time consuming and labor intensive.

The second type of organizing documents are organizing documents according to their intended retrieval pattern. Generally speaking, human retrieve information by their content. The basic assumption of keyword searches lies on the observation that keywords often coincide with concepts. Especially using a sizable number of keywords combined together to form a "model query", the retrieval process can be very accurate as well as efficient. However, users often have difficulty coming up with multiple representative keywords to form an effective query. 95% of the query consists of single keyword only. From linguistic studies, it is well known that single word is "ambiguous", meaning it's meaning changes under different context. That is why the search engines often retrieve way too many unrelated pages along with the intended pages.

To rectify the over recall, search engines increasingly classify the documents into many categories according to their intended use. Following the trend, Knowledge Cockpit also classify documents into broad categories according to their usage. However, instead of asking the users to navigating through the category structure, we attach these categories as meta data to the documents. In doing so, we are able to combine keyword searches with category structure. For example, instead of asking users to examine all the documents in the Restaurant->French->Vegetarian sub category to find what they want, users can submit a query that basically says that "retrieve all the documents in the French Restaurant category that contain vegetarian in it."

4.2.2 Categorization of Documents

Categorization of documents is a labor intensive process. The effectiveness of categorization is also depending upon the categories that one chooses. It is much easier to classify a document in a well thought out and rather complete set of categories than a set that is less complete and also has many overlaps. However, categories do change with time. It is a rather daunting task that one has to reclassify all the documents from time to time when the number of documents is fairly large.

In the prototype, the number of documents is not big. We plan to classify documents by hand. In the future, we are planning to build an automatic classifier to address the reclassification problem. There are several techniques in building automatic classifier. At this point of time, we are still in the evaluation stage. We haven't decided on which technique or a combination of techniques will best suited for our purpose.

4. A Case Study

4.1 The Customer Environment and Issues

The following real case demonstrates the results achieved by applying the technologies mentioned in the previous sections. In this case, both a sales information management tool and a business intelligence prototype were developed to help the client corporation better integrate its internal and external sales information. In the future the intent is to eventually integrate these two tools into one system.

Business Environments

The customer is an electronics product manufacturing company that sells about 20 types of products including office equipment to over 10,000 clients through three major sales channels; retail, whole, and government. In the US, the business is focused on marketing since the majority of their products are manufactured outside US. Therefore, gaining a better understanding of their product sales is important in identifying opportunities to increase revenue and market shares. The company was seeking "hard numbers" to provide their sales managers with for strategic planning.

The industry itself is not growing rapidly, so one of the most important business strategies is to increase their market share. The company has top shares (up to 50%) for several products in the industry. Their major business approaches are effective advertisement by customer analysis, sales force by incentives, and cost reduction by sales process and logistics optimization.

Some of their products are household commodities, so mass-advertisement is essential to increase the shares of these products. Since this can be costly, they are paying close attention to cost effectiveness. To do that, they need to carefully and timely analyze their customers, products, competitors, industry, and external environments, from various aspects. For instance, two days after the cloning sheep news was published, the company made an advertisement using the sheep. This type of association (their products and crone sheep) in advertisements is one of the targets of knowledge mining. For this purpose, all kinds of related news need to be examined and customer impact is also evaluated with reasonable processing speed.

This company has adopted a sales incentive system using a sales target quota. There are marketing managers for product lines in each branch, and each marketing manager has several sales managers, or direct sales people. Higher levels of managers such as product line managers or branch marketing manager have annual and semiannual quotas, and sales people have monthly quota as well as annual and semiannual ones. The quota is assigned from the top level, namely product line managers, according to their performance in the last two years. The quota is then distributed downwards through the chain. The incentive is a bonus salary which is calculated by an open formula which counts both monthly and two-year performance.

Under this situation, it is very important for each manger and sales person to know not only their current sales figures but also their potential ones. The former is in their existing legacy system, and they can get it as fixed-format reports. But, the latter is very difficult to systemize because this information is only in the brain of each person. Each sales person has to consider a large amount of information, such as client characteristics and risks, to estimate the magnitude of the potential sales. Each marketing manager also needs to understand the

reliability of their sales people and his own client background information to grasp the potential sales under his group. They currently do not have any systems and tools to support these processes.

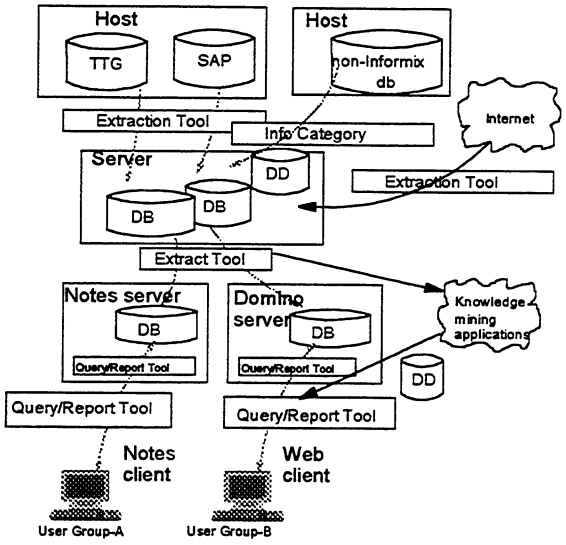

Figure 5. Customer Sales Information Management Tool.

4.2 Legacy System Integration

Transaction systems

In this case study, Lotus Notes is identified as a front end to host-based transaction processing systems since it provides a solution that effectively integrates transaction systems with new client/server systems needs to preserve and leverage the strengths of each platform with minimal tradeoffs in functionality and ease of use.

We have performed research on the technology called MQSeries link for Lotus Notes and CICS link, that integrates Notes with transaction processing systems. The technology removes the constraints of partial solutions, providing an IS organization with a robust, transaction-oriented solution for integrating these systems with Notes. These links leverage existing transaction processing systems and allows Notes to act as a common front end to otherwise standalone and incompatible systems. It also includes support for remote users.

MQSeries and CICS links for Lotus Notes are Notes server tasks that translate data between Notes and a set of APIs (MQI in the case of MQSeries and ECI for CICS). This link technology also controls the flow of data between the Notes application and the system transaction. The actual translation, connecting, delivery and reply from the target system are all under the control of the link technology. CICS link for Lotus Notes works directly with CICS, which runs natively on OS/2, MVS and VSE. MQSeries link provides messaging and queuing support, routing the message to the appropriate target system in the network so that it can be accessed by programs (e.g. a Lotus Notes server) servicing these queues. It avoids the use of private, logical connections between application

programs, so that all communication occurs through queues only. A Notes application consists of a variety of objects, one of which is the Notes form, which in turn contains several types of fields. MQSeries link for Lotus Notes (or CICS link for Lotus Notes) provides an extension to a field definition describing how the data can be obtained or stored.

The heart of MQSeries link is a Notes Database named MQSeries link Application Transaction Map (MATM). This is a normal Notes database that is managed and accessed as any other Notes application. This database controls what MQSeries link does as a Notes server task. The MQSeries API includes transactional features that enable it to be used in high-performance, time-critical applications. The request handling of MQSeries and MQSeries link was designed to optimize performance. When the information is received by MQSeries link from a Notes application, that Notes application is released to continue on with other functions. It does not need to wait for a reply. In addition, after issuing the request, MQSeries link is also free to handle other incoming requests. It does not wait for that specific reply.

4.3 Business Data Management Systems

We have researched on a variety of integration techniques and products that allow application developers to leverage the power of both Notes and DBMS. There are three different ways in Lotus Notes to allow the application developer to retrieve values stored in DBMS tables or stored values into DBMS tables. They are @DB functions, Lotus Script Data Object, and NotesPump.

- @DB functions are launched automatically as field formulas, macros, or manually by selecting a button on a Lotus Notes form. Access to DBMSs from Notes gives developers the opportunity to do a variety of tasks like generate keyword lists and launch stored procedures and queries stored in external databases.
- LotusScript Data Object. The LotusScript Data Object (LS:DO) is a set of LotusScript classes, methods, properties and events that allow access to relational and legacy databases using ODBC. The LS:DO allows developers to combine Notes with external databases to link Notes applications with operational systems. Since the LS:DO provides both read and write access to back-end databases, Notes workflow and replication capabilities are available to developers.
- NotesPump™. NotesPump is server-based software that allows administrators to centrally schedule and execute the high bandwidth exchange of data between Lotus Notes and relational database systems. NotesPump's DB2 link (via ODBC) coupled with IBM's existing suite of replication products gives NotesPump users access to IMS, VSAM and DB2 data on a wide range of platforms.

In this case study, we have developed Notes applications based on @Db functions and Lotus Script Data Object technology. By applying these two technologies, we was able to make the data stored in DBMS available from Lotus Notes and Web interfaces.

4.4 Business Intelligence Prototype

A competitor information system prototype has been developed with the company using the Knowledge Cockpit Framework. A Notes application has been built to contain the competitor information in an organized fashion. The navigator of the Notes application use the door and room metaphor to guide the users to different type of competitor information. There are four doors that lead to Products and Services, News, Financial Reviews, and Market Shares. Clicking on one of the doors will lead the user into the corresponding room with a large file cabinet that contains all the competitor's information in that particular category labeled with each competitor's name on each drawer. Clinking on any of the draw with open that competitor's information in the particular category that user is interested.

The competitor's information in the Notes application is updated daily for the News category using a Java Web Crawler application.[Maes, 1994] The Web Crawler will visit various news sources including that companys' home page to collect company news for that particular day and import them into a Notes Database. Since Notes will maintain the links within each database automatically, the management of the links toward company news will be upto date all the time as mentioned in the section 4.1.

The Product and Services category will be updated less frequently since they generally do not change day to day. However, the structure of the product pages in the company's home page varies from company to company. The Java application we develop will pull all the pages underneath the company's home page and examine the link structure for the whole collection of pages. From the link structure, we did some heuristic pruning and processing to identify the top of each product page. The Web Crawler will then collect all the pages underneath these product pages and import them into a Notes database for that particular company.

For Financial Review and Market Share information, we are planning to process 10K/10Q information and also acquire competitor's information from other vendor. The purpose of the prototype is not so much to undercut the business of the competitor information provider but rather than the integration and easy sharing of the competitor information.

Case Study - Business intelligence

Text Mining technology would be used to capture/extract "useful" information from raw data which has non-structure or semi-structure. This is the first step of our "extracting" procedure.

Data Mining, Ontology approach would be involved in second step to identify relationship among "useful" information. these results are expected to be the knowledge since human being not only recognize those features, conceptualize keyword, but also understand association or correlation for each of them.

A company-oriented Ontology application are in development in which there are a set of conceptualized keywords or terminologies and their certain relationship are represented in a standard form to be represented. This ontology will be used to extract financial and market shared information from 10K/10Q.

Business Processes

Sales process optimization is one of the most important ways to gain cost reduction and client satisfaction. In the company, a sales order is processed as follows. One sales order consists of quantities and sales amounts (K$)

of multiple product items. For example, 100 digital cameras at 20K$, 200 still cameras at 80K$, and 50 optional lens covers at 0.2K$ are in one sales order.

Sales Process:

1. Check the stocks of each item in the order.
2. Make a new internal order (backorder or future order) for unstocked items.
3. Credit check the client.
4. Management approval for the credit check
5. Item shipment request to a warehouse
6. Shipment from the warehouse
7. Invoice issue

The average length of the process is 2 days. Each sales manager and sales person checks the status of their orders every day using the existing legacy CICS reporting system. At the end of each month, each sales person will check the monthly sales amount against the target quota. If the sales doesn't not reach the target, the sales person will push related divisions to speed up the process.

The existing system only provides the total amounts and quantities for each sales person, but does not give them more detailed information, such as customers or product models. For executives, CEO and product-line mangers, there is an executive sales information system, which was built from scratch with PowerBuilder. The system provides very abstract sales figures by product lines, by branches, and by channels. Any further details can not be acquired from the system.

The optimization of logistics is also useful in reducing the sales process cycle. Some products have many optional parts that need to be stored in warehouses in a well-organized fashion. The process for returned products and parts is also important.

Key Issues

There are three major issues from the knowledge management point of view.

a) The company does not have a way of capturing individual/group knowledge in the system.

Each sales person has sales knowledge about his/her customers; ordering patterns, decision makers, key persons, constraints, competitors' approaches, local events, and so on. He/she also understands their products; characteristics, comparisons with competitors' products, CM, etc. Most of this knowledge is in spread sheets, notepads, memos, or the brain of the sales person, not in the system. Each sales manager has knowledge about his/her sales staff as well as customers and products. If it is close to the end of a sales period, sales managers need to simulate outlook sales volume by using the system data and their individual information (out of the system). Sales managers keep this information in a lot of personal ways.

In order to store knowledge in a system they need to have a business process model. The number of report types in the current system, more than 2,000, is also a big issue. The number is continuing to increase to meet user

(sales mangers and sales people) requirements. These report systems are being maintained by some full-time IT people. First of all, business styles are different among the product lines. For example, some products are personal consumer oriented while others are in business use. This implies each product line has its own business process, but the existing legacy system is trying to manage all products in one way, so the current system is overflowed to meet huge varieties of users' personal, group, and division unique requirements. This is also one of the reasons why irregular data is in the current system

b) There is no way to extract knowledge from data and information.

The current system can make a huge amount of reports, but can not analyze and summarize them. Each sales person or manager carefully examines many reports to detect trends, problems, and so on. Data mining and text mining are potential technologies to solve the issue, but they target professional users, such as analysts. For a general use, much more research is needed.

Web pages are also good data and information sources for business, but some level of analysis and summarization are essential for practical use. They are not very useful to general users, sales people and managers, without having been pre-analyzed or summarized information.

c) There is no way to share knowledge that is captured and extracted.

The users are currently using very personal ways to keep this kind of knowledge. A small-scale sharing between two persons could be realized, but large-scale sharing is very difficult. For example, useful wen pages should be shared as a knowledge by analyzing and summarizing them, and by attaching its usage.

The sales information stored in the company's IDMS system has been identified as the major data source for understanding company performance. In order to provide a better interface to sales managers and sales representatives, Lotus Notes has been utilized to present information in a more accessible and friendly way. The technologies introduced in previous sections for the integration of legacy systems are applied here.

d) Other issues

Here we describe data quality issues in the customer's current system, which was built with IMS and CICS, several years ago. There are several types of irregular data in the current system mainly due to human input and nonstandadized business processes. After sales people acquire sales orders from their clients, they fill-in sales order forms by hand. The forms are input into the system by hand, too. In order to absorb these nonstandard processes, several logic and data structures of the existing legacy systems were built in very ad-hoc ways. The followings are examples of irregular data.

a) The total amount and quantity of a sales order is not equal to the total order times in the order.

e.g.. 264 of 98,064 sales records in June.

b) The invoice month of sales recrods is not the same as the month the records are processed.

e.g.. 268 of 145,340 sales records of a product-line for five months.

We also found that the data in the sales databases are not consistent with the data in the human resouces databases. For example, in the sales databases, each sales representive has a month quota required. When matching the sale representive names against the names in human resouces databases, we found some of the names have no

match. We need to take those invalid data out from the sales databases when we present the sales result in Lotus Notes databases.

Customer Expectations

The company is willing to develop a flexible knowledge management system that is highly adaptable to different people, different product lines, different sections, different branches, and even different personnel. Currently each division, group, and person's requirements and working processes have many varieties, so it is impossible to develop business processed for each variety. They need an infrastructure that will allow users to customize detail processes on the standard one, according to their unique requirements. They want to make the data sources very open; existing IMS, WS RDB, Internet, and so on.

They started from solving issue b). The new system allows users to customize ways of extracting and browsing data and information. Each individual, group, and division can have its unique and flexible way. As a first step we provided a flexible method to browse the daily and monthly sales information. Then they will have a mechanism for capturing user's personal knowledge (estimation, expectation, forecast, value-add) into the as-is daily and monthly sales information. By analyzing the daily sales information, remarkable data or information will be emphasized by colors or font size. They are also planning to introduce a competitor information system, which automatically gathers and analyzes competitor information, and summarize them into predefined templates.

5. Lessons Learned: From Quality Information to Customer Intelligence

The information warehouse from various databases can be considered a gold mine of information for some people, and at the same time a junk yard to others. Team selling requires communication and collaboration across diverse distribution channels and locations. Our Knowledge Cockpit is a customer knowledge management system that unites various operational databases from field sales, telesales, resellers, partners to customers support and services to create multiple views of business intelligence for customer development and retention from different perspectives. Our experiences from working with customers in various industries such as consumer packaged goods, transportation, financial services, securities brokerage, manufacturing, computer software and hardware, communications, and chemical processing also suggest us that flexibility of providing different views and interpretations of business intelligence is one of critical success factors. The following three points are key to our lesson learned.

5.1 Value Quality Information

The information can become useless if not managed through a quality life cycle and fully organized for corporate usage. Maintaining high-quality content and frequently refreshing it with new information is critical for business intelligence. Likewise, companies need to manage both tacit and structured information and knowledge as dynamically as possible to make it attractive and maximize it value.

5.2 Linkage to Intranet and Extranet

Having access to quality information alone is not sufficient for succeeding at customer knowledge management. Once customer data and information is captured, companies must established practices to compel its knowledge mining and dissemination throughout the firm to increase productivity, share knowledge and foster collaboration for customer care. Many companies still operate under the mistaken assumption that employees automatically gain their knowledge. Moreover, as companies expand to become more global, there are opportunities for sharing of multiple sources of tacit knowledge from both customers and business partners. These factors provide compelling reasons for the need to create frameworks for unleashing what is learned at the individual contact level onto the organizational level. The corporate infrastructure must support the continuous interaction through both Intranet to its employees as well as Extranets to its customers and business partners.

5.3 Active participation from various communities

Finally, the effectiveness and value of customer knowledge management depends on the active participation of many professionals. People need to make it a habit to contribute their information and knowledge for reuse, and to attentively refine existing business intelligence. Every business unit should promote and encourage others to use and submit intelligence as well. Each party's willingness to take the time to share their knowledge is critical. The time that one spends contributing will be more than compensated for when one taps the reservoirs business intelligence to provide efficient, quality service to customers. Teamwork among business units and customers is key to the success of customer relationship management.

Knowledge Cockpit provides a way for company to effectively leverage their well of information. Constant and systematic approaches of tracking quality information and deriving business intelligence to create the real impact to business operations is critical to the success of the program. One can expect from the customer knowledge management to increase client satisfaction, optimize marketing and sales resource and maximize business opportunity.

Appendix

Technologies to support Knowledge Management

Business executives make on-the-fly data queries from huge data warehouses hundreds of miles away using a laptop computer. Data analysts tap into huge external databases on the Internet that, combined with their company's data warehouse, reveal their customers' buying patterns in rich detail. Salespersons plug into their department's data mart, see they're overstocked on light bulbs, and then cut a deal to sell them -- all on the customer's sales floor.

Data Mining

Every day, huge amounts of data pour into your business -- sales transactions, inventory figures, billing information, marketing campaign responses and much more. More data, in fact, than anyone can be expected to

look at, which is why the average business uses only 7% of the data it collects. But data mining is changing all that.

What is data mining? It's a process that allows businesses to go into their data -- all of their data -- and extract the important pieces of information. Using a set of end user tools that makes data mining accessible to everyone -- not just data analysts but also marketing managers and salespeople -- this process changes the way businesses make decisions and capitalizes on their investment in data like never before.

Text Mining

If we suffer from information overload, much of the blame rests on text-based documents. Articles, opinions, letters and other forms of free text constitute 80% of the world's stored information. How do we sort through these documents to get the information we need, without getting information overload? Text mining.

Text mining solutions are designed to cut through mountains of text-based data and discover the hidden nuggets of information that you need to know. Applying text mining techniques to customer complaint letters, for example, can discover the major reasons for customer complaints and present them to you in a quick, intelligible manner. By searching and mining business news wire documents that mention your company, text mining can reveal a corporate image of your company in vivid detail.

Text mining solutions include:

- InfoDetector, which mine text-based documents to find out what kind of customers you have, and what they're buying. The main functions provided by InfoDetector are: automatic data segmentation (clustering), comparison of two segmentation, interactive visualization of results, generation of reports and data export.
- Technology Watch, which spot technology trends, identify the key players in technology areas and watch your competition. Technology Watch is intended to support strategic decision-making in companies as well as research and defense organizations. It can analyze patents or other data downloaded from international data banks which relate to chemistry, pharmacy, aeronautic, physics, engineering, patents and more. Technology Watch analyzes patent-related and other text documents, and automatically classifies them by content into a number of groups. In each group, the patents or documents are as similar as possible, while the different groups are as distinct as possible from each other.
- Text Navigator, which sift through large numbers of documents and separate the wheat from the chaff in meaningful clusters. Text Navigator does just this, by analyzing the content of a set of text documents and grouping similar items together. Text Navigator can summarize several hundred or several thousand newspaper articles. It completely and automatically clusters together documents according to their contents, in order to display a large collection of documents in an organized fashion and identify their basic components.

Intelligent Miner

Intelligent Miner is a heavy-duty data miner that enables users to identify hidden correlation in their data by performing predictive modeling, database segmentation, link analysis and deviation detection using a variety of

data mining techniques. Yet Intelligent Miner is easy to use because it's designed for decision makers as well as data analysts. Don't let vital, hidden, actionable data continue to slip through your company's fingers

Organizations generate and collect large volumes of data which they use in daily operations. The challenge is to fully capitalize on the value of this data, to discern information implicit in the data.
Data Mining is the process of extracting valid, previously unknown, and ultimately comprehensible information from large databases and using it to make crucial business decisions. It is quickly being recognized as an essential business intelligence tool ... a necessary ingredient to discovering the information necessary to improve a company's market presence and differentiate their products and services in today's global marketplace.

Discovery-driven Data Mining

The Intelligent Miner enables users to identify hidden correlation in their data by performing predictive modeling, database segmentation, link analysis and deviation detection using a variety of data mining techniques.

It extends the analytical capabilities available to data analysts, escaping the confines of verification-driven systems and entering the realm of data-driven discovery. Analysts are no longer limited by those hypotheses which they are able to imagine. Its data-driven discovery allows users to increasingly leverage the data warehouse and more than justify that investment, efficiently analyzing vast amounts of data.

How to Discover that Hidden Information

The Intelligent Miner provides a variety of proven data mining techniques:
Clustering. Partition a database so that records that have similar characteristics are grouped together. Two techniques of this type are provided: Demographic clustering and Neural clustering For example, group a set of individuals based on demographics and insurance policies held..
Classification. Given examples of objects belonging to different groups, develop a profile of each group in terms of attributes of the objects. This profile is then used to predict the group of a new object. Two techniques of this type are supported: Tree induction and Neural Induction For example, "Buyers of expensive sport cars are typically young suburban professionals whereas luxury sedans are bought by elderly wealthy persons."
Value Prediction. Given a database of records, discover the dependency of one attribute's value upon the values of other attributes in the same record, and automatically generate a model that can predict that specific attribute's value for a new record. For example, predict likelihood of fraud for a credit card.
Association Discovery. Given a database of transactions, where each transaction consists of a set of items, discover all associations such that the presence of one set of items in a transaction implies the presence of another set of items. For example, "of the shoppers who purchased milk, 55% also purchased some other type of dairy product, and 42% also purchased bread."
Sequential Pattern Discovery. Given a database of transactions over a period of time, find inter-transaction patterns such that the presence of a set of items is followed by another set of items. For example, "42% of new checking account customers who apply for an ATM card in the same transaction, will also apply for our charge account within 90 days".

Similar Time Sequence Discovery. Given a database of time sequences, find sequences similar to a given one, or find all occurrences of similar sequences. For example, given NorthSouth Airlines daily closing price for 1995, find all other stocks with similar behavior.

Preparing the Data for Mining. Automation of some of the most typical data preparation tasks is aimed at improving the analyst's productivity, by reducing or even eliminating the need for programming specialized routines. Depending on the data mining technique, analysts may select, sample, aggregate, filter, cleanse and/or transform data in preparation for mining or results presentation. Examples of Data Preparation Functions supported: Extract data from an Oracle or Sybase database and load into a DB2 database for mining (Intelligent Miner for AIX only). Merge pharmaceutical transaction data with purchased demographic data regarding physicians. Categorize vehicle speed as high, medium or low. Discard customer demographic data with invalid state code. Calculate frequency count of transactions by customer.

Understanding the Results. Specialized visualization functions bring out unusual features that might otherwise be "drowned out". The role of the data analyst in result interpretation is not only to visualize (graphically or logically) the output of the data mining operation, but also to present the most promising results to the decision maker. The Intelligent Miner supports a range of presentation graphics, specialized to the type of data mining results visualized.

Decision Makers can also use popular Decision Support tools, such as IBM's QMF*, Intelligent Decision Server* (IDS), Lotus 1-2-3, or AS, to access mining results stored in DB2 tables. Or they can use custom tool extensions or applications which invoke Intelligent Miner function via its published Application Program Interface (API). For example, using IBM's IDS, analysts will be able to quickly construct and execute applications which perform data mining operations. And, data mining applications address cross industry business problems with Customer Segmentation and Item Set Analysis using this same API.

References

Choo, Chun Wei and Ethel Auster. "Scanning the Business Environment: Acquisition and Use of Information by Managers." Annual Review of Information Science and Technology. Editor: M. Williams. Medford, NJ, Learned Information, Inc. For the American Society for Information Science, 1993.

Fayyad, U. and G. Piatetsky-Shapiro, P. Smyth. "From Data Mining to Knowledge Discovery: An Overview." Advances in Knowledge Discovery and Data Mining. AAAI Press/the MIT Press. 1996.

Fayyad, U. Advance in Knowledge Discovery and Data Mining AAAI Press. Menlo Park, CA. 1995.

Frankes, W. and R. Baeza-Yates editors. Information Retrieval: Data Structures & Algorithms Prentice Hall, NJ, 1992, pp.1-4

Lewis, D. and K. Sparck Jones. "Natural Language Processing for Information Retrieval," CACM, Vol.39 No.1, Jan. 1996. pp.92-101

Maes, P. "Agents That Reduce Work and Information Overload," CACM., Vol. 37, No. 7, July 1994. pp. 31-40.

Nonaka, I. and H. Takeuchi. The Knowledge-Creating Company. Oxford University Press. London. 1995.

"Programming Mobile Agents in Java." Aglets Workbench. (http://www.ibm.co.jp/trl/aglets.)

Simoudis, E., J. Han, and U. Fayyad. Editors. Proceedings of the Second International Conference on Knowledge Discovery & Data Mining. AAAI Press. Menlo Park, CA. 1996.

Uschold, M. and M. Gruninger. "Ontologies: Principles, Methods, and Apllications." The Knowledge Engineering Review. Vol.11:2, 1996. pp. 93-136

"White Paper -- Notes and DBMSs." Lotus Web Page. (http://www.lotus.com)

"White Paper -- Notes and Transaction Systems - MQSeries and CICS Link for Lotus Notes." Lotus Web Page. (http://www.lotus.com)

"White Paper -- Lotus Notes : An Overview." Lotus Web Page. (http://www.lotus.com)