Enterprises face increasingly competitive environments. A s companies downsize to adapt to these environments they may be able to cut costs. But unless they have captured the knowledge of their employees, downsizing can result in a loss of critical information. Similarly, as the employee turnover rate escalates in today’s overheated job market, organizations are likely to lose access to large quantities of critical knowledge. And as companies expand internationally, geographic barriers can affect knowledge exchange and prevent easy access to information. These and other forces are pushing enterprises to explore better methods for knowledge management.

Can we create a system that will capture company-wide knowledge and make it widely available to all its members? Increasingly, organizations large and small alike are attempting to answer this question with knowledge management systems. The business world is becoming so concerned about knowledge management that, according to one report, over 40 percent of the Fortune 1000 now have a chief knowledge officer (CKO), a senior-level executive responsible for creating an infrastructure and cultural environment for knowledge sharing.1

WHAT IS KNOWLEDGE MANAGEMENT?

Enterprise knowledge management entails formally managing knowledge resources in order to facilitate access and reuse of knowledge, typically by using advanced information technology. KM is formal in that knowledge is classified and categorized according to a prespecified— but evolving— ontology into structured and semistructured data and knowledge bases. The overriding purpose of enterprise KM is to make knowledge accessible and reusable to the enterprise.

Knowledge resources vary for particular industries and applications, but they generally include manuals, letters, summaries of responses to clients, news, customer information, competitor intelligence, and knowledge derived from work processes. A wide range of technologies are being used to implement KM systems: e-mail; databases and data warehouses; group support systems; browsers and search engines; intranets and internets; expert and knowledge-based systems; and intelligent agents.

In artificial intelligence, knowledge bases are generated for consumption by so-called expert and knowledge-based systems, where computers use rule inference to answer user questions. Although knowledge acquisition for computer inferencing is still important, most recent KM developments make knowledge available for direct human consumption or develop software that processes that knowledge.

Historically, KM has been aimed at a single group— managers— through what has been generally referred to as an executive information system. An EIS contains a portfolio of tools such as drill-down access to databases, news source alerts, and other information—all aimed at supporting managerial decision making. More recently, however, KM systems are increasingly designed for entire organizations. If executives need access to information and knowledge, their employees are also likely to have an interest in and need for that information. In addition, KM technology is ideally suited for nonmanagement groups— such as customer support, where customer service requests and their solutions can be codified and entered into a database available to all customer service representatives.

IMPLEMENTING KM

As organizations store an increasing amount of information and knowledge in data and knowledge warehouses and in data and knowledge bases, they are attempting to manage that knowledge in more efficient ways. Historically, organizational knowledge has been stored on paper and in people’s minds. Unfortunately, paper has limited accessibility and is difficult to update.
And when people leave, they take most of their knowledge with them, so reuse is not always feasible. Thus, firms have moved to data and knowledge warehouses and to data and knowledge bases to improve accessibility, updatability, and archivability of data and knowledge.

**Data warehouses**

In many companies, one of the first KM tools is a data warehouse. A data warehouse acts as a central storage area—a warehouse—for an organization's transaction data. Data warehouses differ from traditional transaction databases in that they are designed to support decision making rather than simply efficiently capturing transaction data. Typically, data warehouses contain multiple years of transaction databases stored in the same database. Data warehouses are not updated on a transaction-by-transaction basis. Instead, the entire database is updated periodically.

The size of data warehouses can be substantial. Chase Manhattan Bank has a 560-Gbyte data warehouse, for example, and MasterCard Online has a 1.2-Tbyte database available to member companies for a fee. With all the data accessible in one place, relationships between data elements can be more effectively explored. Users can browse the data or establish queries, though this type of analysis generally results only in knowledge for particular individuals. An alternative approach is to use a process called knowledge discovery to determine whether there is additional knowledge hidden in the data.

**Knowledge warehouses**

Rather than the kind of quantitative data typical of data warehouses, knowledge warehouses are aimed more at qualitative data. KM systems generate knowledge from a wide range of databases including Lotus Notes databases, data warehouses, work processes, news articles, external databases, Web pages (both internal and external), and people. Thus, knowledge warehouses are likely to be virtual warehouses where the knowledge is dispersed across a number of servers.

In some cases, a Web browser can be used as an interface to a relational database. For example, Ford Research and Development uses a browseable Oracle database. The database contains manuals and design rules, specifications, and requirements. Another frequently used corporate application is a human resource knowledge base about employee capabilities and skills. Employee information can include education, specialties, previous experience, and other descriptors.

Historically, Lotus Notes has provided one of the primary tools for storing qualitative and document-based information and for facilitating virtual groups. With the recent explosion of the Internet, however, low-cost Web-based solutions within intranet environments have become the focus of KM.

**Data and knowledge bases**

Knowledge can come from top-down activity, work processes, news reports, and a wide range of other sources. Knowledge typically captured to meet top-down requirements includes manuals, directories, and newsletters. Knowledge bases capturing information generated from work processes are likely to include working papers, proposals, and other similar documents. In addition, knowledge bases can be designed to provide continuity and history in activities like customer support.

**Lessons learned.** Lessons-learned databases can be used to support operations or generate information about business in general. For example, the National Security Agency (NSA) Lessons Learned knowledge management system.

---

**Selected URLs on Data Warehousing**

- ACM SigMod—http://bunny.cs.uiuc.edu/
- Foundations of Data Warehouse Quality—http://www.dbnet.ece.ntua.gr/—dwq/
- Terminology—http://www.creda.com/
- IBM’s Page on Data Warehouses—http://direct.boulder.ibm.com/bi/tech/datamart.htm
- GOOD Group—http://loochi.bpa.arizona.edu/group.html

---

**Selected URLs on Knowledge Management**

- AAAI Spring Symposium on AI in Knowledge Management—http://ksi.cs.ucalgary.ca/AIKM97/
- American Productivity and Quality Center—http://www.apqc.org/b2/b2.htm
- Knowledge Management in Practice—http://www.apqc.org/subscribe.htm
base contains three types of lessons: informational, successful, and problem. An informational lesson might describe how an NSA employee could be moved to temporary duties in cases of emergencies. Successful lessons capture positive responses to crisis. Problem lessons provide examples of things that went wrong and potential ways to solve the problems.

Similarly, Ford Motor Co. has what the company calls TGRW—things gone right/wrong—files. TGR captures information about events that facilitate task accomplishment, while TGW captures information about events that stand in the way of task accomplishment. (Generally, TGR are easier to gather than TGW, particularly if the knowledge is archived, as few employees are anxious to be associated with things that went wrong.) TGRW knowledge bases are critical in establishing records of events that need to be addressed and monitored by project management.

**Best practices.** Best-practices knowledge bases capture knowledge of the best processes. Typically, best-practices knowledge bases are generated using benchmarking activities designed to solicity the more effective and efficient way of doing things. After an organization has knowledge of best practices, they can be incorporated.

For example, General Motors Hughes Electronics supports a “best process reengineering database.” Associated with each entry is a brief description and a contact. Typically, entries are changes in processes made throughout the organization that have led to improved processes. Making them available in a single database increases the chance that they will be seen and adopted elsewhere in the organization.

Consulting firms have been among the first to develop best-practices databases to support their consultants. Price Waterhouse was among the first with Knowledge View, which is a Lotus Notes best-practices database that allows multiple views—by industry, process, performance measure, and enabler (technology, for example). It is based on an ontology embedded in a business model that focuses on processes that lead to creation of value (for example, “Produce Products and Services”) and support process areas (for example, “Develop and Maintain Systems and Technology”).

News reports provide a means of formally integrating external information into an enterprise. For example, the professional services firm, KPMG, teamed with Story Street Partners to provide prefiltered, presorted, and presearched data on issues and companies of interest to KPMG employees.

**GENERATING KNOWLEDGE FROM DATA: KNOWLEDGE DISCOVERY**

Knowledge discovery is a new and rapidly evolving discipline that uses tools from artificial intelligence, mathematics, and statistics to tease knowledge out of data warehouses. Gregory Piatetsky-Shapiro and William Frawley define knowledge discovery as “non-trivial extraction of implicit, previously unknown, and potentially useful information from data.” Because knowledge discovery approaches can be designed to exploit characteristics and structures of the underlying application domain, knowledge discovery has found use in a wide range of applications, including fraud analysis, credit card analysis, security, customer analysis, and product analysis.

Knowledge discovery is a method that includes different tools and approaches to analyze both text and numeric data. For example, organizations have developed different ways to generate knowledge from numeric databases, such as the financial information in the US Security and Exchange Commission’s Edgar Electronic Data Gathering and Retrieval System). Price Waterhouse developed an intelligent system called EdgarScan, shown in Figure 1, to make Edgar available on the Web (http://edgarscan.tc.pw.com). EdgarScan lets users access a repository of publicly available financial information. Data is periodically extracted from the Edgar Web site (http://www.sec.gov), as shown in Figure 2, and stored in an Oracle database maintained by Price Waterhouse. User profiles are also maintained to facilitate maintenance of the database and response to users. Having access to this numeric information allows users to monitor changes in the data over time, which can facilitate comparisons between enterprises.
Price Waterhouse has also developed Odie (On-Demand Information Extractor) to scan roughly 1,000 newsletters each night to extract knowledge about management changes from text data. Odie, which has been applied to both US and European newswires, uses an understanding of the stylized language in business news articles and knowledge about syntactic patterns to understand relevant business events. In addition, Price Waterhouse is investigating the possibility of monitoring semistructured text in order to gather information to help understand other types of business events, such as acquisitions.

**REPRESENTING KNOWLEDGE**

KM systems represent knowledge in both human- and machine-readable forms. Human-readable knowledge is typically accessed using browsers or intelligent search agents. But some knowledge is accessible for machine-readable purposes, designed as an expert system’s knowledge base to support decision making. Meanwhile, ontologies are generally endemic to KM systems because they typically refer to taxonomies of the tasks that define the knowledge for systems.

**Selected URLs on Knowledge Discovery**

Knowledge Discovery Mine—http://www.kdnuggets.com/
Data Mining and Knowledge Discovery Journal—http://www.research.microsoft.com/datamine/
Third Conference on Knowledge Discovery and Data Mining—http://www-aig.jpl.nasa.gov/kdd97/
Data Mining—http://pwp.starnetinc.com/larryg/datamine.html
FAQ on Data Mining—http://www.rpi.edu/~vanepa2/faq.html
Glossary of Terms—http://www.pilotsw.com/r_and_t/whtpaper/datamine/dmglos.htm
Data Mining Paper—http://www.dbmsmag.com/9608d53.html
knowledge relating to specific modems (technical specifications, data, pictures, and so forth) and to summarize that data in a knowledge base. Whenever customer support has a question or needs to “picture” a modem, they can access it through the knowledge base. As customer support encounters particular problems, those problems are captured as specific cases that are then indexed by customer, modem, and problem type. Accordingly, whenever others have a similar problem they can find those problems documented in the database.

In other situations where the information is largely declarative knowledge (like facts and assertions), text or rules might be used to represent the information and knowledge. For example, manuals, newsletters, and other similar types of knowledge are typically provided in a document, list, or rule format (although there may be added links between the knowledge to facilitate search and understanding). Organizational rules guide promulgated behavior and would generally be of the form “If a, then b”: “If you have a baby, then you are allowed up to eight weeks of family leave.” Adaptations of these rules could potentially be used in a rule-based knowledge-based system.

If, on the other hand, information is highly filtered, then it is likely to be represented as a set of declarative statements. For example, Arthur Andersen’s knowledge base on Global Best Practices (http://www.arthurandersen.com/gbp/BPList.htm) lists five specific “Best Practices for Managing Information Resources,” including “Develop and maintain an IT strategy that is integrated and aligned with the company’s business goals.” The knowledge is declarative and is independent of particular situation information, meaning that the database lists no particular cases. Although filtering ensures that knowledge is correct and consistent, developing declarative knowledge is ultimately a political process, typically removing context and controversy. As a result, filtered knowledge can be limited in its ability to provide as deep an insight as unfiltered information.

Machine-readable knowledge

Expert systems use their knowledge bases and user responses to guide the user to recommended solutions. Expert systems can be an integral part of a KM system. For example, Deloitte & Touche’s KM system has some expert systems available to support particular processes, such as assurance activities.

Although some KM systems contain such artificial-intelligence-based systems, most KM systems use artificial intelligence primarily in the form of intelligent agents to search human-readable knowledge. We need additional research to expand the use of artificial intelligence and knowledge-based systems in KM. We need to know what forms of knowledge representation appear to work best for particular types of knowledge and how artificial intelligence can be further integrated into KM systems.

Ontologies

An ontology is an explicit specification of a conceptualization. In enterprise KM systems, ontology specifications can refer to taxonomies of the tasks that define the knowledge for the system. Ontologies define the shared vocabulary used in the KM system to facilitate communication, search, storage, and representation. Development and maintenance of an enterprise-wide ontology requires continual effort to evolve the ontology over time.

Ontologies are particularly important in ensuring that best-practices databases are able to communicate to the user the broadest range of practices and activities and allow the user to recognize when a best practice would fit in their organization. Price Waterhouse reportedly has an ontology with over 4,000 entries for its best-practices database. Since Price Waterhouse is an international firm, the ontology has been translated into other languages to broaden use and accessibility of the knowledge base. In addition, since enterprises are often involved in multiple industries, multiple ontologies may be required as part of the KM system.

Out of necessity, virtually all enterprises with a KM system have developed their own ontology. Because these firms have made this investment, ontology construction appears at this point to offer competitive advantages. However, at least one firm has expressed interest in an ontology shared across multiple organizations in order to cut development costs and to speed system development. Over time, industries are likely to form coalitions or subscribe to central services for these reasons.

Other knowledge description attributes

In addition to ontology information, additional descriptive attributes of the knowledge can prove critical to its use and maintenance. Contributor, organization, and status information are all viable descriptive attributes. Virtually all knowledge bases capture contact or contributor information, including contact or contributor names, date of contribution, the person’s role in generating the knowledge (for example, the project manager), and so on.

Many knowledge bases also include organizational information that can include the department or division in which the project was built or from which the knowledge was gathered. Status information about knowledge is also a typical kind of descriptive attribute. This kind of status information can include
whether an element of a project is planned, currently being implemented, or has been implemented. It can record whether the information is externally available or for internal purposes only.

**KNOWLEDGE FILTERING**

Unfortunately, quality and importance of the knowledge varies, depending on a number of factors, such as who is providing the knowledge to the system. In discussion groups like the Water-Cooler site, an electronic forum on best practices in acquisition management (http://www.arnet.gov/Discussions/Water-Cooler/), there is no information filtering, which typically leads to multiple and sometimes conflicting responses. Messages sent to the forum are captured in topic threads consisting of the initial message and subsequent responses.

Because knowledge quality and importance varies from source to source, systems often resort to knowledge filtering to ensure complete and correct knowledge. For example, at GM Hughes Electronics, best-process reengineering practices are captured in a database that combines human and computerized knowledge. Each entry is submitted to an editor who screens it for usefulness and relevance. At the National Security Agency, a nine-member team decides if a “lesson learned” is valid. Not all proposed lessons learned are included in the database.

Not all filtering is done by humans. Perhaps the most visible and frequent use of computer-based filters is the message filtering that categorizes and prioritizes e-mail messages. A number of products also help monitor qualitative databases. For example, grapeVine (http://www.gvt.com) monitors multiple Lotus Notes databases. The system generates “alert” messages that contain summary information with links to the document and any other discussions, based on a personal interest profile. Since it is profile-based, monitoring can be done according to individual, group, or organizational needs, cascading information up an organizational hierarchy, according to user interests.

Unfortunately, cascading as it is accomplished in collaborative systems like grapeVine has some limitations. Collaborative systems have individuals rank the importance of information coming into the company. Users might categorize certain information as “very important,” “important,” and so on. Other individuals in the enterprise then decide on what level the information must be labeled before it is delivered to them. In the case of a busy manager, we might imagine that information needs to be “very important.”

However, a limitation of using this approach is that some information ranked as “important” might turn out to be “very important.” The manager, then, would not always see the necessary or very important information. In such settings, we can imagine that there would be a tendency for the raters to rate marginal items higher than they might otherwise rate them. Meanwhile, managers might set their own level at “important” in order to assure that all ultimately “very important” announcements would be received. Ultimately, this leads not only to importance inflation, but to a deluge of information that such a system is designed to combat.

**SEARCHING FOR KNOWLEDGE**

Knowledge bases can become quite large. Ford’s initial knowledge base, for example, had the equivalent of more than 30,000 paper pages as of June 1997. Because typical knowledge bases have a great amount of information, searching them efficiently becomes an extremely critical function. The most dominant search techniques include search engines, intelligent agents, and visualization models.

**Search engines**

A wide range of well-known Internet search engines—like AltaVista, Excite, Infoseek, Lycos, WebCrawler, and Yahoo—have been used to guide users to information on the Internet. These and other search engines can be adapted to intranet environments for KM. In addition, a number of firms have developed alternative approaches to the conventional search engines. For example, Andersen Consulting has “a central repository of interfaces (‘knowledge maps’) that link to knowledge.” Users can select a map and use it to navigate directly to knowledge stored in multiple databases without needing to know which database to access.

**Intelligent agents**

Intelligent agents can be used to connect people to knowledge available on the Internet or intranets. InfoFinder, for example, learns user interests from sets of classified messages or documents, recognizing that people will tend to classify only those examples that interest them. In addition, InfoFinder uses heuristics to gather additional insights into a user’s interests. Based on message syntax, InfoFinder attempts to determine significant phrases that provide insight into user goals.

For example, one heuristic is to extract any fully capitalized word, such as ISDN, since it is likely to represent an acronym or a technical name. Another heuristic is not to extract the word if it is used for emphasis, such as “NOT.” Other syntactic heuristics include capturing bullet points, numbered lists, section headings, and diagram descriptions. These heuristics allow InfoFinder to find documents that it anticipates are of direct interest.
Visualization models

One of the dominant new trends in the search for effective enterprise KM is visualization models. Two emerging tools—Perspecta and InXight—represent different ways of visualizing knowledge space.

Perspecta (http://www.perspecta.com) creates what it calls SmartContent using metainformation derived from source documents—be it structured information in databases and tagged documents such as news feeds, or unstructured information in office documents and Web pages. For unstructured documents, Perspecta has a Document Analysis Engine that performs linguistic analysis and automatically tags documents. The SmartContent server analyzes this tagged information and identifies relationships between documents, and constructs a multidimensional information space using an Information Space Markup Language. The user “flies through” the information space, as shown in Figure 3, by manipulating the mouse. Data is downloaded to the client using Perspecta’s just-in-time Information Streaming Transport Protocol, an extension to HTTP, to conserve resources.

InXight Software (http://www.inxight.com), a spin-off from Xerox PARC, recently released its VizControl information visualization software for visualizing large hierarchies. VizControl technology offers several novel visualization formats, each of which exploit “focus + context” techniques that foreground objects of interest while preserving the overall structure of even very large data sets.

One such tool, the hyperbolic browser (or fish-eye) display shown in Figure 4, exploits hyperbolic geometry to provide exponentially more information space for hierarchies that expand exponentially with depth. Thus, a hyperbolic browser can display 1,000 nodes in a $600 \times 600$ pixel window, with those in the center displaying significant amounts of text, as opposed to the 100 or so nodes displayed in a conventional 2D browser.12

The user navigates the information space by clicking on a node or dragging the mouse over the hyperbolic plane. Current demonstration implementations map Web hierarchies identified by URLs, thus forgoing a strong semantic structure, but it is conceivable that the browser could incorporate more semantic information using technologies such as InXight’s LinguistX natural-language processing tools.

CULTURAL ISSUES

Ultimately, KM systems require a strong leadership that instills a culture of knowledge sharing. Whether KM is implemented in a centralized fashion (as Buckman Laboratories has accomplished by reorganizing its IS department into the Knowledge Transfer department4) or in a more decentralized system (of the sort that Hewlett-Packard has implemented13), KM systems require knowledge sharing. Accordingly, organizations use different incentive systems to make sure that knowledge is shared. According to Tom Davenport,4 Lotus...devotes 25 percent of the total performance evaluation of its customer support workers to knowledge sharing. Buckman Laboratories recognizes its 100 top knowledge sharers with an annual conference at a resort. ABB evaluates managers based not only on the result of their decisions, but also on the knowledge and information applied in the decision-making process.

The types of incentives and the ability to measure contributions to KM generally are contingent on the...
level or function in the organization and the particular application to which the KM system is being put. Inevitably, however, incentive systems are based on measurable activities. The Corporate Education group at Hewlett-Packard gave 2,000 free air miles for the first 50 readers and another 500 miles for anyone who posted a contribution to new knowledge bases.11

Whether such incentives actually foster a cultural framework where employees feel it is in their best interest to participate actively in KM systems remains to be seen. Clearly, however, such KM systems benefit corporations that take advantage of the technology. As enterprises are being driven toward KM systems to meet competitive pressures and create value, they are increasingly finding that these systems can facilitate reuse of existing knowledge and create new knowledge in an effort to allow better decision making.  

References

Daniel E. O‘Leary is a professor at the University of Southern California. His research interests include the impact of intelligent agents on individuals, organizations, and commerce, and the integration of AI and reengineering efforts. O’Leary is the editor-in-chief of IEEE Expert. He received a PhD from Case Western Reserve University.

Contact O’Leary at 3660 Trosdale Parkway, University of Southern California, Los Angeles, CA 90089-1421; oleary@rcf.usc.edu.