

Knowledge Management: A Careful Look

Won Kim, Cyber Database Solutions, Austin, Texas

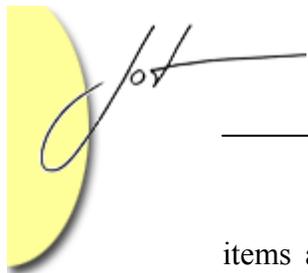
Seung Soo Park, Department of Computer Science and Engineering, Ewha Women's University, Seoul, Korea

Abstract

During the past three decades, corporations, governments, and educational institutions in industrialized nations have adopted information technology to run their daily operations. Information technology, both computer hardware and software and communications technologies, is at the core of knowledge-based economy. Knowledge-based economy is a major driving force for the economy and society of industrialized nations [www.wkforum.org]. Knowledge management has become an important subject in this context. However, the term “knowledge management” is vaguely understood. One reason is that the term “knowledge” itself has been both overused and loosely used. To understand the technical issues and challenges in knowledge management, one must first understand the term knowledge better. In this article we will take a careful look at the term “knowledge”, and on that basis shed some light on the concept of knowledge management, and then in turn point out some major issues and challenges in knowledge management.

1 TYPES OF KNOWLEDGE

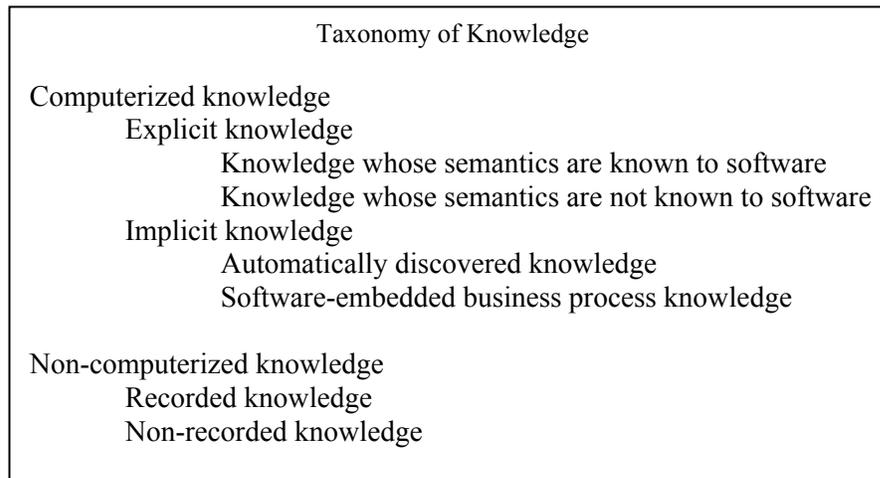
The term “knowledge” is often used interchangeably with the term “information”. Although the marketing literature on some data-processing products talk of knowledge being more useful and advanced than information, and information being more useful and advanced than raw data, we do not believe there is a compelling distinction between information and knowledge. (Knowledge is also known in some philosophical circles as true or verified belief.) The term “data”, however, can be distinguished from “knowledge”. “Data” refers to uninterpreted and unprocessed raw data, while “knowledge” refers to either data or “value-added data”. “Metadata” is also “data”. Metadata is data about data, and is found in data dictionaries or system catalogs in database systems. “Value-added data” is data obtained by querying, manipulating, analyzing, and interpreting raw data. A list of customers and items they purchased from a store would be raw data. The metadata for this data describes the data, for example, customer name as a 20-byte string, purchase date in mm/dd/yy date format, purchase



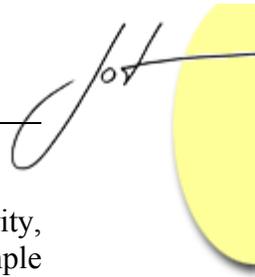
items as a list of 20-byte strings, etc. A purchase pattern discovered from the list of customers and items they purchased, such as a sudden surge in the purchases of canned goods or purchases of certain combinations of items, would be knowledge.

Sometimes one refers to “knowledge” as “actionable data” as a way of distinguishing data that is relevant to a given objective from one that is not. For example, the discovery of a sudden surge in the purchases of canned goods may be actionable knowledge to a retail chain, in that the retail chain may then order additional canned goods or raise the price on canned goods. However, such discovery as “most husbands are men” would not be useful to a rental car company. “Actionable data” may be raw data or data obtained from raw data; in other words, raw data, such as customer names and addresses, may be directly relevant to a given objective, such as a marketing campaign. This is why we believe knowledge should be regarded as encompassing both raw data and data obtained from raw data, rather than just data obtained from raw data.

To help understand knowledge in a systematic way, we provide a taxonomy of knowledge in the Figure below. Knowledge may be either computerized or non-computerized. Computerized knowledge is one that is stored in a computer system and is amenable to processing by computer software and hardware. Non-computerized knowledge is one that resides in human brains or is recorded in recordable, but not computer-processable, media, such as paper. In the context of “knowledge-based economy”, for example, knowledge means both computerized and non-computerized knowledge. “Knowledge management systems”, on the other hand, deal only with computerized knowledge.



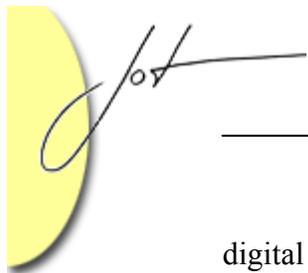
Computerized knowledge has two types. One is explicit knowledge and another is implicit knowledge. Explicit computerized knowledge is knowledge that is captured and stored in a computer system, such as databases, files, data warehouses, the Web, information portals, etc. Explicit knowledge in turn comes in two types. One is knowledge whose semantics are known to computer software such that the computer



software can automatically perform content-based searches, and enforce data integrity, etc., that is, do something beyond just storing and retrieving it in its entirety. An example is the tables of alphanumeric data stored in relational databases. The semantics of the tables are known to relational database systems, and as such relational database systems answer queries by evaluating search conditions against stored data, and enforce such semantic constraints as data type, value range, UNIQUE, null-not allowed, etc. Another is the HTML or XML documents stored in Web servers. The semantics of such documents, such as the tags, are known to the Web browser which display the documents. Another type of explicit knowledge is one whose semantics are not known to computer software, and as such computers merely store such knowledge and retrieve it for output in its entirety and human users need to interpret the contents of the knowledge. Examples are most types of multimedia data, such as photographs, images, video, voice, sound, music, broadcast, etc. (stored without any tags to give hints on the contents). These are merely stored as files in computer systems and are retrieved by their file names; human users view them and/or listen to them to understand their contents.

Implicit computerized knowledge also has two types. One is knowledge that can be automatically discovered from raw data through a process known as data mining. There are many data-mining algorithms in use today, such as neural networks, decision trees, Bayesian networks, genetic algorithms, etc. Data mining has been used to detect unusual or unforeseen patterns from raw data, such as fraudulent uses of credit cards, fraudulent insurance claims, customer churn behavior, customer purchase behavior, customer overseas calling behavior, failure patterns of automobile parts, etc. Another type of implicit computerized knowledge is the business process knowledge that is embedded in the flow and logic of computer software. Examples are the business processes embedded in such enterprise software as ERP, SCM, and CRM systems. ERP (enterprise resource planning) systems attempt to automate and thread the business processes of multiple related parts of an enterprise (e.g., from order taking, to inventory management, to shipping, to invoicing, etc.). SCM (supply chain management) systems help an enterprise to coordinate business interactions with the suppliers of a variety of supplies and services. CRM (customer relationship management) systems are used to manage a variety of related customer management activities (customer segmentation, customer reward program, marketing campaigns, etc.). There are also numerous “vertical” application software in which certain knowledge about the vertical industry is captured. Examples are patient record management systems for doctors’ offices, securities management systems for stockbrokers, airline reservation systems for airline reservation agents, etc. These vertical applications all use industry-specific jargons in their menus, and guide the users through business procedures.

Non-computerized knowledge includes all human knowledge that has not been digitized and stored in a computer system. There are two types of non-computerized knowledge. One is knowledge that has been recorded in recordable, but not computer-processable, media, such as paper, tapes, the wall, etc. Another is knowledge that has not been recorded in any recordable media. Non-computerized but recorded knowledge can often be easily computerized (into explicit knowledge whose semantics are not known to software) using such tools as OCRs (optical character recognition systems), scanners,



digital cameras, microphone and recorders, etc. The types of non-recorded knowledge that are particularly difficult to capture in a computer system include deep domain expertise, human intuition, culture, etc. One key obstacle faced by electronic learning systems, without the instructors standing by, is the difficulty in capturing certain types of deep domain expertise that are needed to answer unforeseen questions from students.

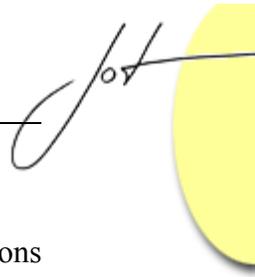
2 ADOPTING KNOWLEDGE MANAGEMENT

Knowledge management means managing knowledge at the disposal of an enterprise or individual to achieve a certain objective set forth by the enterprise or the individual. In general it refers to managing both computerized and non-computerized knowledge. From this perspective, knowledge management is not something mysterious and esoteric. In fact, everyone in the world does knowledge management every day. A person who organizes his/her income tax records by year, and keeps Java and C programming books separate from Tom Clancy novels, etc., is doing knowledge management. A library or a bookstore that keeps books and periodicals under a certain classification scheme is also doing knowledge management. Further, anyone who uses a word processor to do a job for an organization is a “knowledge worker”, and anyone who has created a computer file and retrieved it has done management of computerized explicit knowledge. Anyone and any organization that uses the Microsoft Office software on desktops and laptops, PDAs and handheld email devices, relational database systems, enterprise software, etc., are already in the thick of knowledge management.

It is important to remember that it is just not possible to capture all human knowledge in computer systems, and also today’s information technology is just not ready to do automatic intelligent processing of most of the knowledge (multimedia data, in particular) stored in computer systems.

Knowledge management involves three basic elements: making knowledge ready for use, making use and sharing of the knowledge, and protecting the knowledge. In particular, knowledge management requires capturing, organizing, storing and updating knowledge (i.e., making knowledge ready for use); encompasses querying, manipulating, analyzing, discovering, visualizing, reporting, and transmitting knowledge (i.e., making use of knowledge); and protecting knowledge (i.e., securing from unauthorized access, and protecting from computer system failures). As knowledge includes both raw data and data obtained from raw data, file management and database management, both of which manage raw data, are by default knowledge management.

Enterprises and individuals adopt knowledge management to derive certain benefits. The benefits are in general a combination of cost savings, increased revenue, enhanced customer satisfaction, shortening of the business cycle or decision-making cycle, etc. in the case of enterprises; and time saving, convenience, etc. in the case of individuals. Before adopting knowledge management, an enterprise must clearly understand and justify its objectives. As part of the objectives, it must decide at what level of the enterprise knowledge management will be applied – the entire enterprise, a particular



department, or a particular group within a particular department, etc. All considerations for the implementation of knowledge management follow from the objectives defined.

The next step is to identify all sources of knowledge the enterprise will manage for the objectives chosen. The knowledge sources will include both computerized and non-computerized knowledge. They include sources maintained within the enterprise, knowledge obtainable from knowledge providers and the Internet, knowledge obtainable from partners, etc.

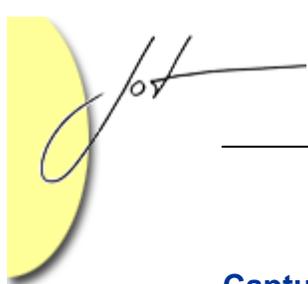
Then a “knowledge management process” needs to be defined and enforced. A knowledge management process includes all the procedures for making knowledge ready for use, making use of and protecting the knowledge by the knowledge workers and knowledge administrators in the enterprise. All the personnel to be involved in the knowledge management process must be properly trained on the procedures relevant to them.

Next all necessary computerized tools for knowledge management need to be determined, selected, installed, administered, and upgraded. Some of the tools are for use by knowledge workers, while others are for use by knowledge administrators. All the personnel involved must be properly trained on the use of all necessary tools to do their jobs.

3 TOOLS FOR KNOWLEDGE MANAGEMENT

There are a wide variety of tools for knowledge management. For non-computerized knowledge, tools would include such mundane objects as bookshelves, (physical) file folders and organizers, library indexes, telephone directories, address books, etc. In the context of knowledge management, one usually refers to tools based on information technology for managing computerized knowledge. These include computer software and hardware for supporting the three elements in knowledge management, and also communication systems infrastructure and communication devices. The tools differ depending on a large number of factors, such as the volume and types of knowledge, physical organization of the enterprise (i.e., geographic dispersion), the number and behavior of the workers (e.g., telecommuting, traveling, teleconferencing, etc.), the nature of the primary activities of the enterprise, the vertical industry of which the enterprise is a part, the financial wherewithal of the enterprise, cost of ownership of the tools, security requirements, government regulations the enterprise must observe, etc. Some tools are indispensable, while others merely make knowledge management easier.

In terms of “primitive” functions, the tools for knowledge management may be organized into nine categories. (Certain tools actually perform more than one primitive function.) These are summarized briefly below for completeness.



Capture

Tools for knowledge capture include data entry software (database system, file system, database bulk loader, order entry system, etc.) and devices (Point of Sale system, OCR, scanner, keyboard, digital camera, etc.).

Transform

Tools for knowledge transformation include span a wide range. They include natural language translation systems, voice to text converter, text to voice converter, multimedia data compression and decompression systems,. They also include ETL (extract transform and load) tools for data warehousing, multidimensional OLAP (M-OLAP) servers (that create data cubes by pre-computing all possible combinations of attributes), data preparation tools (format conversion, restructuring of tables, sampling, etc.) for data mining, data cleanser, etc.

Store

Tools for storing knowledge include database systems, file systems, data warehouses, repositories, information retrieval systems, Web servers, information portals, groupware, etc. They also include such hardware devices as disks, CDs, DVDs, tapes, etc. Further, they include such systems as SAN (storage area network) and NAS (network attached storage).

Query and Update

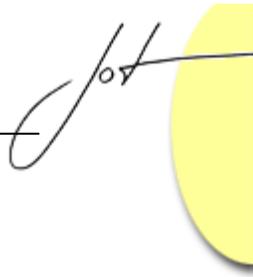
This category includes database systems, file systems, data warehouses, repositories, information retrieval systems, Web servers, information portals, groupware, OLAP (online analytical processing) servers, etc. It also includes Web search engines, natural language understanding systems, etc. Further, it includes such tools as matching fingerprints, faces, etc.

Report and Output

This category includes query tools and report writers, visualization tools, etc. This also includes hardware devices such as monitors, speakers, etc.

Discover

Tools for automatically discovering knowledge include data mining tools, text mining tools (for generating automatic summaries of texts, extracting features from an article, etc.), statistical analysis packages, business analytics, website analyzers, etc.



Share and Learn

Tools for sharing and learning knowledge include knowledge community tools (that make it possible to easily find knowledge experts to answer specific questions), electronic and distance learning systems, etc. Insofar as most tools for storing and querying knowledge allow multiple users to share a common knowledge (data)base, such tools may be regarded as tools for sharing knowledge.

Protect

This category actually consists of two sub-categories. One is for tools that protect the integrity (correctness) of knowledge. This subcategory includes tools that prevent invalid data from making its way into a database, and database and file backup and restore systems to preserve a knowledgebase from computer system failures. This also includes such tools as anti-virus software to prevent viruses from wreaking havoc on not only the knowledgebase but also the entire computer system. Another subcategory is for tools that protect a knowledgebase from unauthorized access. It includes tools that prevent unauthorized access to knowledge, encryption and decryption techniques, firewall, intrusion detection system, etc.

Transmit

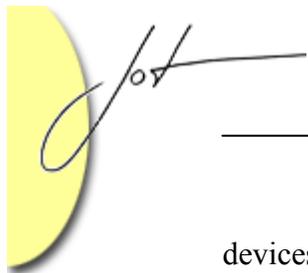
This includes networking software and hardware, such as hubs, switches, routers, etc. It also includes such system as QoS system.

Besides the above nine categories of tools for “primitive” knowledge management functions, there are applications for both knowledge workers and tools for knowledge administrators. Applications include desktop software for ordinary knowledge workers, such as word processing systems, spreadsheets, graphics packages, presentation makers, etc. They also include vertical applications for knowledge workers, such as patient record management software for doctor’s offices, securities management software for stockbrokers, business analytics, Point of Sale systems for supermarkets and department stores, etc. Applications also include enterprise applications such as ERP, SCM, CRM systems; and horizontal applications such as website analyzers, finance management systems, payroll systems, inventory control systems, etc.

4 MAJOR ISSUES IN KNOWLEDGE MANAGEMENT

There are at least four major issues in knowledge management. They include administration, integration, multimedia, indexing, and non-computerized knowledge. Let us examine these in turn. The first two are evident from the discussions thus far.

Advances in desktop applications have empowered millions of knowledge workers around the world during the past three decades. People carry laptops, PDAs, email

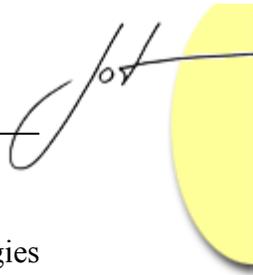


devices on their trips to stay connected to their jobs and to the world in general. It is said that in industrialized nations some 70% of the workers may be categorized as knowledge workers; that is, they make use of some information technology tools for knowledge management. Unfortunately, however, the increasing variety of desktop applications, increasing sophistication of such applications and the underlying operating systems have turned a vast number of who should be “casual” knowledge workers into in essence pseudo-system administrators and knowledge administrators at the same time. In other words, a knowledge management environment consists in general of an infrastructure for enterprise applications (operating system, networking system, database system, application server, Web server, OLAP server, groupware, etc.) as well as a variety of applications (business-specific vertical applications, word processor, spreadsheet, graphics package, Web browser, email, etc.). Who should really be casual knowledge workers now have to manage a large directory of files and email trails; learn and use a large number of menu options; be able to disable a long print job that is printing print-unfriendly files from the Web; download and update plug-ins from the Internet, and even contend with viruses.

The software and hardware infrastructures for enterprise applications and communications require professional system administrators and database administrators, as well as technical support from software and hardware vendors. System administrators need to select, install, maintain and upgrade all hardware and software necessary for knowledge management. Database administrators need to plan, monitor, and tune performance and reliability of databases.

The variety of hardware and software adopted for knowledge management lead to integration challenges. System administrators need to make different software and hardware from different vendors, and different versions of software and hardware from the same vendors, all interoperate. Further, if knowledge management must be done across multiple knowledge sources, the knowledge may have to be physically consolidated and moved into a data warehouse, with all attendant challenges in designing a data warehouse, migrating and transforming data from different sources, etc. If the knowledge in different sources is such that it cannot be moved into a single data warehouse (e.g., databases of the Motor Vehicle Departments in different states in the US), a means must be introduced to interoperate the disparate knowledge sources, that is, as a federation of independent but cooperating knowledge sources.

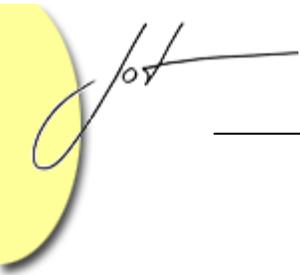
Relational database technology has been matured during the past three decades. Simply put, it is a technology for storing, querying, updating, and protecting alphanumeric data formatted in two-dimensional tables. However, managing knowledge that comes in the form of multimedia data is rather young and relatively immature. Multimedia data include text (such as articles in newspapers and magazines), images (both two and three-dimensional), diagrams and drawings, maps of all types, voice, sound, music, video, broadcast, etc. Today, technology exists for compressing and decompressing multimedia data, edit and store it as files of different formats, and transmit it over the network. Technology also exists that can extract features and rudimentary summary of textual documents, distinguish objects and background from images, identify



the gender of speakers and changes in speakers in broadcast, etc. Further, technologies for matching fingerprints, voiceprints, registered logos, etc. have been in use for some time. However, techniques and technologies need to be developed to index, search for exact match, search for similarity match, classify and categorize the full spectrum of multimedia data. Such technologies may be first developed for “vertical” data types, such as satellite images of suspected weapons factories, areas to be bombed, faces of terrorist suspects and supporters, etc., and then extended to apply to general multimedia data.

The vast amount of data on the Web, and all computerized explicit knowledge whose semantics are not known to software (e.g., multimedia data files) present a huge challenge to retrieving them. Unless they are properly (manually) indexed or tagged, much of such knowledge can go undetected. Even if such knowledge is found, it may be mixed in with a lot of irrelevant search results and may be ignored by the users. Creating and maintaining indexes into and tags for all computerized knowledge such that all (or most) of it may be found is a major challenge. Ideally, such indexes and tags should be created and maintained automatically; however, it is beyond the capabilities of today’s information technology, and a mix of automatic and manual (human) means must be used.

The problem of turning such non-recorded knowledge as deep domain expertise and human intuition into recorded knowledge will remain the ultimate challenge in knowledge management. It is in essence what the field of artificial intelligence has struggled with for at least four decades. Further, in the interest of “job security”, some experts may be reluctant to part with their expert knowledge by having it recorded. As a stopgap measure, some vendors of knowledge management solutions, such as Verity, offer “community” tools designed for customer technical support personnel to route technical questions (from customers) regarding products and services to employees with proper knowledge to answer them. Such tools are to complement the customer support databases that typically contain data such as questions asked or problems reported, when and who originated such questions or problems, answers or solutions given, who addressed them, etc. Efforts to replace knowledge engineers with software modules such as intelligent software agents have had some limited results in data gathering and automated classifications. However, knowledge management tools are not likely to replace humans to an extent much beyond what they already have done. After all, it takes humans to capture knowledge, humans to ask questions, and humans to answer them.



About the authors



Won Kim is President and CEO of Cyber Database Solutions (www.cyberdb.com) and MaxScan (www.maxscan.com) in Austin, Texas, USA. He is also Dean of Ewha Institute of Science and Technology, Ewha Women's University, Seoul, Korea. He is Editor-in-Chief of ACM Transactions on Internet Technology (www.acm.org/toit), and Chair of ACM Special Interest Group on

Knowledge Discovery and Data Mining (www.acm.org/sigkdd). He is the recipient of the ACM 2001 Distinguished Service Award.



Seung Soo Park is Dean of Engineering College at Ewha Women's University in Seoul, Korea. He is an associate professor of computer science and engineering. His research interests include artificial intelligence, data mining and bioinformatics. He received his Ph.D. in computer science from the University of Texas in Austin.