2002

Knowledge management for the analysis of complex experimentation

Maule, R.

MCB UP Limited

Internet Research: Electronic Networking Applications and Policy Volume 12 · Number 5 · 2002 · pp. 427-435

Knowledge management for the analysis of complex experimentation

R. Maule
G. Schacher and
S. Gallup

The authors

R. Maule and **G. Schacher** are Professors and **S. Gallup** is Associate Professor, all at the Naval Postgraduate School, CSUMB, Monterey, California, USA.

Keywords

Knowledge management, Artificial intelligence, Knowledge-based systems

Abstract

Government agencies carry out many events each year designed to determine future requirements and capabilities. These events include field experiments, surveys, interviews, simulations and workshops. Similar themes are evident across many of these events. Unfortunately, mechanisms for passing information from one event to the next, or for developing bodies of knowledge in the topical areas they address, have yet to be fully developed. The task is difficult on two fronts. In response to this need a knowledge management capability was developed to help provide structure for dynamic and static data and thereby, aid in the analysis of complex experimentation. The system warehouses qualitative and quantitative data and supports mining operations through a number of traditional and artificial intelligence-based techniques. Described are the information architecture of the system, the knowledge processing methodologies, and the structure of the thematic data sets that form the knowledge ontologies.

Electronic access

The research register for this journal is available at
http://www.emeraldinsight.com/researchregisters

The current issue and full text archive of this journal is available at
http://www.emeraldinsight.com/1066-2243.htm

Internet Research: Electronic Networking Applications and Policy
Volume 12 · Number 5 · 2002 · pp. 427–435
 · ISSN 1066-2243
DOI 10.1108/10662240210447173

Introduction

A prominent function of knowledge systems is to help address information overload and access problems through tools for data synthesis and extraction. As the underlying technologies have evolved so has the discipline. Knowledge engineers, in the project under evaluation, utilize Web- and artificial intelligence (AI-)based content management tools to find, extract and disseminate knowledge. Many of the practitioners are without the traditional ties to AI, linguistics or cognition common to their predecessors. Rather, they use the tools of AI to build knowledge applications that convey understanding for top-level decision makers. They report to chief knowledge officers who are the focal point for enterprise-wide knowledge, communications and information.

The problem of information overload is particularly acute in governmental operations given the large volumes of information that must be processed. Needed are systems to enable the effective application of knowledge, whether stored or dynamic, to achieve understanding. As in previous eras, knowledge technologies remain relatively unfamiliar to those outside of the discipline (Srikantaiah and Koenig, 2000) and this often complicates implementation for practitioners. In addition, resources requiring synthesis are expanding at an exponential rate. Not helping matters is that decision makers external to the discipline often expect the systems to provide knowledge at the touch of a button while those in the profession realize knowledge management (KM) as an extension to traditional information processing. The system developed herein focuses on the processing of information and the generation of knowledge to aid in understanding.

This paper thereby presents both a theoretical base and an operational KM system, based on the implementation variables and limitations described above. The KM system was designed to aid in the manipulation of information to support knowledge generation for the analysis of complex experimentation and serves as a decision aid for experimentation analysis. There are capacities for simulated intelligence at various point(s) in the processing architecture. Qualitative and quantitative data, and various forms of multimedia, are structured to aid in

Internet Research: Electronic Networking Applications and Policy
Volume 12 · Number 5 · 2002 · 427–435

analysis and the generation of correlations between and across variables and components. Mining operations are supported through a number of traditional and AI-based techniques. The system is both an evolutionary advancement of previous technologies, and due to the complexity of the subject matter, has broken new ground in the development of methodologies to process dynamic information and provide contextual analysis to aid understanding.

Research in knowledge management

Many find the field of KM worthy of research if for no other reason than the complexity of the discipline and the high cost of the tools. The interdisciplinary nature of the field is also intriguing – for perhaps the first time in the history of computing very diverse disciplines are coalescing to address issues unique to the concept of KM. Each of the fields brings its unique perspective and some history to the field of KM.

The information technology disciplines are addressing processing technologies, both hardware and software. The cognitive science and AI communities are advancing algorithms to address information synthesis and human capacities for understanding. The networking and telecommunications fields, being somewhat responsible for interconnecting the machines that collectively produced the information glut, are aiding with the solution through the internetworking of the various processing technologies. On the networks and information grids are shared meanings. Knowledge ontologies, in turn, support semantic webs of understanding (Berners-Lee *et al.*, 2001).

Research in the above areas has both basic and applied components. And, the complexity of the problem extends applied research methodologies well beyond the traditional systems modeling, case study, or organizational analysis to address knowledge engineering and knowledge architecture. Basic research is currently occurring in areas related to cognition and intelligence, both human and machine. Also to be addressed are mechanisms to integrate humans and machines in virtual knowledge settings – with this latter issue introducing the disciplines of 3D visualization, simulation, and virtual reality. The need for these latter areas becomes readily apparent in large knowledge spaces as 2D representations become inadequate for representations of complex scenarios. 3D and virtual technologies can help structure knowledge in complex and distributed environments.

Leading-edge research organizations, tasked with inventing or guiding future technologies, provide some assistance in categorizing the underlying themes. The Defense Advanced Research Projects Agency (DARPA) is following the lead of the World Wide Web Consortia (W3C) with the concept of a "semantic Web" as the next evolution of the Web (Berners-Lee *et al.*, 2001). This involves descriptions that are mapped across Internet sites and knowledge systems that interact with these descriptions to create shared knowledge spaces. The descriptions are terms that have their meanings or "ontologies" (representations of words and their meanings) indexed, with the indexes shared and synchronized across participating sites. While the benefits of such ubiquitous knowledge access are obvious, the implementation of this concept is far more difficult and has led to an evolution of the traditional knowledge engineering discipline from AI-based acquisition and assimilation to include Web-based information and communications management in support of the dynamic knowledge requirements of chief decision makers. The systems and methodologies developed in this paper are in support of such an environment.

The concepts identified herein address both systems and ontology design and implementation, and the means to assign levels of synthesis [often encapsulated as metadata] to a wide variety of media to include concepts, themes, presentations, ideas, external systems, etc. The challenge is to use ontologies to internetwork information resources, distributed databases, and the heterogenous technologies representing current and historical knowledge. This will support the vision of the semantic Web, wherein knowledge can be located and communicated based on common understanding as enabled through shared, discipline-specific terminology. A supporting focus of this project was to establish

Internet Research: Electronic Networking Applications and Policy
Volume 12 · Number 5 · 2002 · 427–435

information tagging schemes and resource structures, and metadata from these structures, to serve as knowledge ontologies on which shared meaning and understanding would occur. Once operable, these systems help enable network-centric operations and shared understanding that is based on both dynamic and historical, contextual knowledge.

Knowledge technologies

Internet technologies are enablers of distributed knowledge management and provide the means to implement global access (Tschaitschian *et al.*, 2000). Architectural issues address the design of navigation, labeling and search mechanisms to optimize system efficiency and user satisfaction (Sherman, 2000). Knowledge is thereby structured to proactively achieve a goal or objective (Saward, 2000). In these contexts the knowledge can be an abstraction of an information source to provide application relevance, or a manipulation of information to answer a question or support a concept.

System design is the key to usability. Design processes begin with detail diagrams that outline information flows, network diagrams that provide integration points to other systems, search and navigational systems to draw knowledge from the information, and some manner of content weighting (how much content and in what mixtures). Collectively, these provide a means to calculate knowledge within a schema – with the schema represented as a concept hierarchy, and as a set of knowledge units or frames (Gao and Sterling, 2000).

Consolidating data from various sources into a consistent knowledge base is one of the greatest challenges of KM. Once integrated, this enables business intelligence and strategic options to be visualized and cause and effect relationships to be mapped (Primix Solutions, 2000). Better decision-making, faster turnaround times, improved organizational communications, and higher levels of cooperation and interaction among personnel are results of a properly implemented and maintained KM system (Schwartz *et al.*, 2000).

Schwartz *et al.* (2000) have identified acquisition, organization and distribution as basic tenets of Internet-based KM. Acquisition includes information gathering, inquiry, validation and encoding. The organization stage addresses the ranking, classification, and association of information. Linkages and search mechanisms are addressed at this stage, including keywords, indexes, and cataloging. Information distribution involves various delivery techniques, the creation of product awareness, end-user "identification" with the product, and technical matters in the utilization of the system that may range from Web push/pull to querying techniques in databases.

Knowledge portals

Knowledge portals are a relatively new but highly significant innovation that is causing a fundamental change in traditional information practices (Ennov, 2001; Hoffman, 2001). Basically, the amount of electronic information has reached a stage such that users need mechanisms to find relevance among the abundance (Silver, 2001). Portals are a means to personalize, sort and filter information (Moore, 2001). They are Web sites that provide gateways to other Web sites and various Web services (Wesker, 2001). One of the more popular services increasingly integrated into portals is realtime chat and messaging, meaning that the portal and its Web sites and services is empowered with collaborative abilities (Loria, 2001).

With the ability to aggregate sites and services "portals" are often considered the "killer application" of KM (*Knowledge Management*, 2001). Still, to keep things in perspective, portals are but the latest evolution of a series of Web-based systems designed to provide enhanced business intelligence and decision support (Ruber, 2000). A simultaneous occurrence was the evolution of Web searching systems with qualitative database technologies to integrate classification and indexing systems for unstructured information into the portals. Many of the original document-based artificial intelligence systems have been integrated into the portal technologies. Supported information services include Web sites, file systems, notes and collaborative tools, relational and object-oriented databases, and other forms of both structured and unstructured information.

One industry in which the concept is evolving at a very rapid pace is the US Government, where portals are being designed, developed

Internet Research: Electronic Networking Applications and Policy
Volume 12 · Number 5 · 2002 · 427–435

and implemented as a means to organize information producing entities, and to categorize and filter vast quantities of quantitative and qualitative data. KM "communities of practice" have arisen as a means to support those responsible for the design, development and maintenance of knowledge systems. These communities tend to address portals as the best available means to integrate work being done in various locales and by various groups – and to ensure active participation by key decision-makers.

Given the complexity of systems to integrate structured and unstructured information, along with the various techniques and services to aid in the use of information processing technologies, it has become apparent that there is a need for knowledge frameworks. These frameworks would need to address the processes through which data is translated into information and knowledge, and the systems to help structure the knowledge for end-user consumption. Sub-frameworks would include information sharing mechanisms within the client organization and knowledge processes between organizations. The system described herein develops such a framework and applies it in the development of an operational knowledge system and portal.

Experimentation KM system

In the context of the KM system, "knowledge management" is used in the broadest sense and includes the management of numerical values obtained from automated collection systems, qualitative data from human subjective opinions, synthetic results from both human and machine synopsis, and result sets tailored to address specific long-range plans and objectives. The challenge was to create a KM system that enabled ready access and retrieval of new and archival information by analysts and subject-area specialists. The developed system supports the analysts and their need to ascertain the effectiveness of various processes, systems and technological capabilities within large-scale experimental frameworks.

The most important step in creating the KM system was to insure that it would support the major strategic, operational and tactical questions addressed currently and in the future. The system was established in a very broad configuration, to support a wide variety of issues, and this has made it relatively easy to model knowledge and problem needs. Qualitative and quantitative data provide detailed operational assessments of the technologies. Concept analysis helps determine the effectiveness of the systems and processes.

Client application

The system itself is distributed, using servers running Internet Information Server and Apache on Intel and Sun platforms. The servers communicate through portals and ULR linkages using a one-time or single sign-on procedure to validate users across the system. The multi-tier, multi-vendor, multi-application architecture has proven useful for load-balancing as well as system and media debugging. There are both production and development systems, each with multiple versions of the software applications. This has led to a situation in which there may be significant differences in the application of specific software functions but this is unavoidable given the speed with which the software and field is evolving.

A key objective of the system was to provide an easy-to-use interface for the end-users so every attempt was made to keep the system and interfaces as simple as possible. The portal technologies have been quite helpful in this regard. The underlying framework for the applications is Java 2 Enterprise Edition (J2EE). To date there has been three years in development of the system and resources. In addition, numerous specialized systems are integrated into the architecture and provide output particular to specific needs. The basic storage parameters and functionality of the system are documented in Table I. At the core

Table I Media types, storage parameters, and system functionality

Functionality	Resource level 1	Resource level 2	Resource level 3
Warehousing	Qualitative data	Quantitative data	Multimedia
Mining	Text mining	Data mining	Categorization
Search	Full text	Full text	Full media
Presentation	Documents	Tables	Streams
Correlation	Text-level, AI	Data-level, AI	Media-level
Analysis	AI synthesis	AI synthesis	Reference

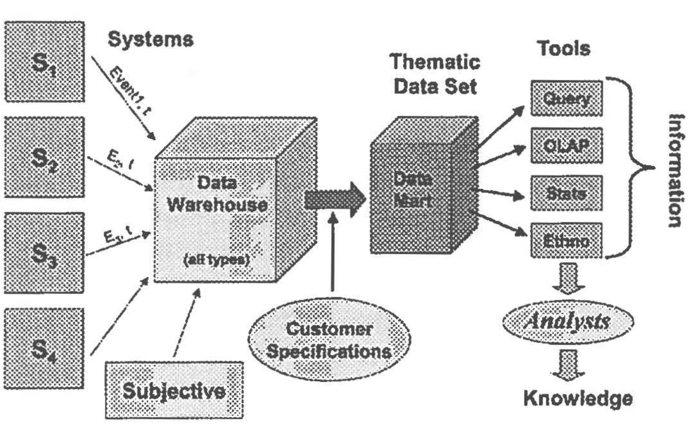
is the warehousing and various methods for knowledge and information location, ranging from mining to search, to presentation and correlation. Efforts are being made to provide similar capabilities across qualitative and quantitative data. Correlation and analysis functionality are designed to support the human analysts and provide various levels of AI-based synthesis and extraction to help summarize and draw insight from extremely large repositories and data sets.

Once the data have been deposited in the KM system they are immediately available for research and analysis. Two of the most common analysis methods utilized are linear analysis and quasi multi-dimensional analysis. Results may be in the form of a text document explaining the linear progression of an event over time, a graphical interpretation of an event, a chart detailing processes within an event, or a simulation. Results can aid in the evaluation of new systems and assist in the design of new initiatives (Figure 1).

Qualitative and quantitative analysis

Generally speaking, quantitative research methods use structured data to draw statistical conclusions. Results might include validity measures, reliability assessments, sampling results, measurement data and surveys. Qualitative research involves unstructured data, such as interviews, documents and participant observation data to understand and explain events and processes. Qualitative research excels at "telling the story," often from the participant's viewpoint, providing the rich descriptive detail that sets quantitative results into the correct context.

Figure 1 KMS collaborative architecture for client topic and concept analysis



The KM system collects and catalogs or "warehouses" qualitative and summarized quantitative data and organizes it in a way that makes it easier for researchers and analysts to develop knowledge – or answers to questions from research sponsors. In turn, this knowledge is fed back into the knowledge bases and repositories to provide higher-level knowledge. Several levels of systems have been developed, to provide a range of knowledge analysis services.

The first level system is basically the data warehouse that holds the qualitative and quantitative information. The repository contains reference reports specific to a technology being tested; discussions of the operations of the systems from the users; observation reports assessing the technologies and their usage in conjunction with other systems; and final analysis of the overall effectiveness of a system or process. Analysis is supported by quantitative data from active input and processing technologies. The knowledge system accentuates base information with fulltext search, conceptual analysis, and AI processing algorithms for qualitative knowledge mining. There are also separate knowledge processing systems that process data for inclusion in the base repository, such as the ethnographic qualitative knowledge management system (EQKMS).

The EQKMS provides detailed ethnographic assessments of specific events. In a typical application, analysts will assess a pattern of events and appropriately code the data using a specific knowledge ontology. Then, the analyst will look for identified events that exist across functional or operational areas. This information is extracted to produce knowledge that provides a precise assessment of the details of an event or rationale underlying an event. Such matters as the functioning of a technology in a specific situation, human-in-the-loop operations at specific nodes, and the effectiveness of operations that integrate data from various systems are assessed.

The system can support online collaboration among analysts and researchers, with users able to load, manipulate and process information remotely. Users are able to determine who will

Internet Research: Electronic Networking Applications and Policy
Volume 12 · Number 5 · 2002 · 427–435

see their information and when they will see it – which is useful for analysis during development and prior to display to the client. Passwords and encryption secure the service.

Qualitative mining methodologies

A number of methodologies are available for information access and the processing of information into knowledge. Information may be browsed through broad categorizations to provide an overview of the information available or by topic to narrow resource searches to a specific area of interest. In one option the resources are provided in a hierarchical representation, by topic and theme. A fulltext search may use either keywords or phrases. Titles of documents, keywords associated with particular documents, author searches, and full text search for all pertinent documents in the repository are available. Advanced search techniques use multiple phrases and boolean operators to narrow or expand from the basic search with filtering selected by author, date, category or topic. Search results can be browsed as text or HMTL with the search terms highlighted. From this level the analysts can begin their work to translate the information into knowledge.

Qualitative data mining utilities are supported at the document level, as a layer above the initial search and filtering. A concept analysis AI provides an assessment of hits by the frequency of important words in the selection and is used to generate metadata and build the knowledge ontologies. Next, a metadata term can be selected to produce a synthetic document that yields text excerpts extracted from the source. This can be a highly effective technique to generate key findings since at this level the complex search and filtering techniques have been overlaid with the metadata and the knowledge extracted from the synthetic document is within the appropriate situational context. A second level of AI synthesis is used to derive document summaries based on the searches and filters. While these AI algorithms may not extract knowledge in the same manner as a human, given the gigabytes of qualitative information requiring processing these tools can be critical in the knowledge generation process.

Qualitative analysis for complex system relationships

In studying highly complex systems we are often dealing with "systems of systems" that are intricately coupled and interrelated. This makes research on specific issues very difficult and is an underlying challenge of modern KM. The blurring of boundaries between independent systems makes it difficult to draw useful distinctions between the various data types that emerge. Comparison and explanation often require both qualitative and quantitative analysis.

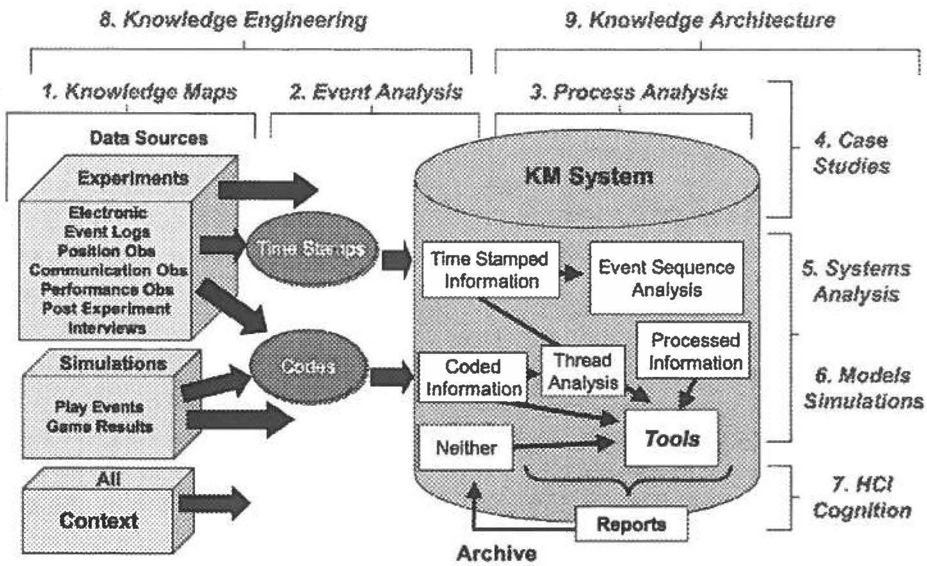
Qualitative research is broadly defined as any kind of research that produces findings not arrived at by means of statistical procedures or other means of quantification. Where quantitative researchers seek causal determination, prediction, and generalization of findings, qualitative researchers seek illumination, understanding and extrapolation to similar situations, prediction and answers. Qualitative analysis can provide perspective, insight, and wisdom. The term "structured" and "unstructured" are also used to refer to quantitative and qualitative data, respectively. In this context, structured (quantitative) data addresses the records, systems output, and mechanical processing that reflect what has occurred. Unstructured (qualitative) data refers to the notes, memoranda and insight that provide guidance for the future. It helps predict what will or may occur.

Experimentation and knowledge generation

Planning documents, technical specifications, e-mail memoranda, observations, formal reports, and interviews are all means to generate qualitative data. The process through which such information is codified and placed within the system is dependent on the knowledge methodology. In the system herein, the data are collected, categorized based on their origin or theme, warehoused in one or more repositories, and analyzed for current and near-term client applications (Figure 2). Information to be archived in the system comes from differing sources and is characteristically of high variability. There may be neither a common structure nor a common core of assumptions.

Internet Research: Electronic Networking Applications and Policy
Volume 12 · Number 5 · 2002 · 427–435

Figure 2 Data processing continuum highlighting extracting processes for knowledge management generation and KM research methodologies



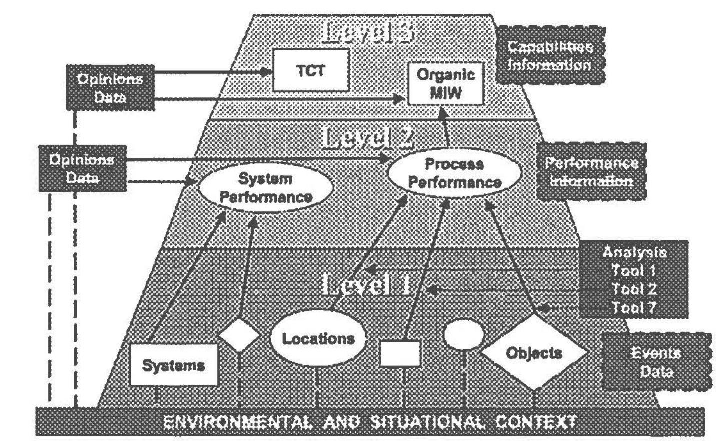
There is often an overt desire to test a range of operational structures and situations.

Even a given type of event may have some of its conditions change over time, and even from event to event. The experimentation the system was designed to support is highly dynamic. On the other hand, there is a desire to synthesize results from many events to obtain conclusions on current capabilities and project future potential. This means that the KM system must be robust enough to accept change in the configurations under which information is obtained. To support this complexity, the warehousing process consists of examining documents in their original formats and converting them into files in the appropriate structure. Documents are first categorized according to environmental contexts and then sorted according to the topics they are addressing. Report-length documents are broken into component documents to better represent specific topic areas and to facilitate searches and analysis.

Transformation of data into knowledge

When data are entered into the KM system there effectively begins a transformation from an unstructured, loosely related collection of data into semi-structured information (Figure 3). Before information can become knowledge, pertinent characteristics must be extracted

Figure 3 Data and information levels for knowledge synthesis and extraction



through the application of various analysis models. Examination, analysis, and abstraction models are applied against the warehouse data and information to derive knowledge. The analysis models group and filter data and information, commonly related as:

(1) Time marked events, including objective and subjective data that directly address events, such as:
   - systems event data;
   - notes from observations, including human performance at a particular time or place; or
   - simulated events addressing a series of actions over time.

(2) Opinions or derived results, concerning the performance of a particular system or process in a specific situation, or cumulative results across time.
(3) Context and meta-data that provide the environment for the data or information, including:
   - scenarios;
   - operational and tactical assessments;
   - specific environmental variables unique to the situation or context.

The last step in the transformation of data into knowledge is the most arduous and can take many forms. Research analysts, working alone or in groups, access the information to answer a specific question or derive results on a particular experiment. Information included can come from myriad systems, platforms, records and observations – all contributing uniquely to the final product. The end result, or knowledge, may summarize an event through a written analysis summary; provide a graphical interpretation of the interactions between players in an event over time (such as a cluster analysis); or visually tell a story, such as an event recreation or simulation. In this manner the KM system helps integrate quantitative data and qualitative information to produce knowledge that can be carried forward and used for future planning.

Conclusion

The KM system is a data and information warehouse, and an information and knowledge processing system, for the analysis of complex experiments within various situational and environmental contexts. The system supports decision-making based on accumulated experiences, including operational processes in critical systems and process efficiency measurements in selected environments. Analysts use the system to generate and correlate knowledge that is in turn fed back into the system to provide higher-level knowledge, often referred to as "understanding." Various automated systems are used to generate synthetic knowledge and the results exist as a layer in the KM system architecture. Cumulatively, the human and machine-assisted knowledge generation aid clients in situational and environmental understanding through scenarios and outcomes-based analysis. Future projections can then be derived, based on the analysis methodologies and related events in similar environments.

The KM system methodologies also help to define measurable relationships between systems and processes, including human-in-the-loop operations at specific nodes. A multi-level analysis with drill down capability aids in reach-back or thread analysis from high-level conclusions to supporting information. This includes synthesis from results to higher-level decision-making. The integration of qualitative expert opinions with quantitative data provide a more accurate assessment of the status of a concept or initiative at the macro level, and component analysis of hardware, software, systems and processes at the micro level.

References

Berners-Lee, T., Hendler, J., and Lassila, O. (2001), "The semantic Web", *Scientific American*, May, available at: www.sciam.com/2001/0501issue/0501berners-lee.html
Ennov, B. (2001), "Corporate portals require complete KM strategies", *KMWorld*, July/August, p. S22.
Gao, X. and Sterling, L. (2000), "Semi-structured data extraction from heterogeneous sources", in Schwartz, D., Divitini, M. and Brasethvik, T. (Eds), *Internet-Based Organizational Memory and Knowledge Management*, Idea Group Publishing, Hershey, PA, pp. 83-102.
Hoffman, T. (2001), "Making personalized retrieval a reality in knowledge enterprises", *KMWorld*, July/August, pp. S14-S15.
Knowledge Management (2001), "Through the looking glass: a progress report on the enterprise portal", *Knowledge Management*, February, pp. 17-23.
Loria, M. (2001), "Knowledge management and collaboration", *KMWorld*, July/August, pp. S8-S9.
Moore, A. (2001), "Enterprise information portals: the power and the peril", *KMWorld*, July/August, pp. S2-S3.
Primix Solutions (2000), "Customer knowledge management", *Knowledge Management*, October, pp. 90-1.
Ruber, P. (2000), "Opening doors to structured data", *Knowledge Management*, October, p. 81.
Saward, G. (2000), "The challenge for customer service: managing heterogeneous knowledge", in Schwartz, D., Divitini, M., and Brasethvik, T. (Eds), *Internet-based Organizational Memory and Knowledge Management*, Idea Group Publishing, Hershey, PA, pp. 57-81.

Internet Research: Electronic Networking Applications and Policy
Volume 12 · Number 5 · 2002 · 427–435

Schwartz, D., Divitini, M. and Brasethvik, T. (2000), "On knowledge management in the Internet age", in Schwartz, D., Divitini, M. and Brasethvik, T. (Eds), *Internet-Based Organizational Memory and Knowledge Management*, Idea Group Publishing, Hershey, PA, pp. 1-19.

Sherman, L. (2000), "Creating useful knowledge structures", *Knowledge Management*, October, pp. 82-3.

Silver, B. (2001), "The face of knowledge management: a status report", *KMWorld*, January, pp. 14-15.

Srikantaiah, T. and Koenig, M. (2000), *Knowledge Management for the Information Professional*, American Society for Information Science, Medford, NJ.

Tschaitschian, B., Abecker, A. Hackstein, J. and Zakraoui (2000), "Internet-enabled corporate knowledge sharing and utilization", in Schwartz, D., Divitini, M. and Brasethvik, T. (Eds), *Internet-Based Organizational Memory and Knowledge Management*, Idea Group Publishing, Hershey, PA, pp. 25-41.

Wesker, M. (2001), "Using a portal to solve business problems", *KMWorld*, July/August, pp. S4-S5.