Knowledge Harvesting, Articulation, and Delivery

Kemal A. Delic

Dominique Lahaix

Harnessing expert knowledge and automating this knowledge to help solve problems have been the goals of researchers and software practitioners since the early days of artificial intelligence. A tool is described that offers a semiautomated way for software support personnel to use the vast knowledge and experience of experts to provide support to customers.

A consequence of the global shift toward networked desktops is visible in customer technical support centers. Support personnel are overwhelmed with telephone calls from customers who are experiencing a steady increase in the number of problems with intricate software products on various platforms.

Support centers are staffed with less knowledgeable (and less experienced) first-line agents answering the simple questions and solving common problems. Expert (and more expensive) technicians resolve more complex problems and execute troubleshooting procedures. The work of both (the first-line agents and the technicians) is supported by various technical tools, but they always have to use their brains and experience to handle effectively the stream of problems they encounter. This knowledge is seen as the key ingredient for the efficient functioning of support centers.[1]

Kemal A. Delic

Kemal Delic is a technical consultant at HP's Software Services Division in Grenoble, France. He is responsible for knowledge technologies. He received a Dipl.El.Ing. degree from the Faculty of Electrical Engineering at the University of Sarajevo in Bosnia. Before joining HP in 1996, he was a senior scientific consultant with CPR Consortium in Pisa, Italy. He is married and has two children. In his free time he enjoys reading medieval history.

Dominique Lahaix

Dominique Lahaix is the knowledge and electronic services manager at HP's Software Services Division. He came to HP at the Grenoble Division in 1988. He received an engineering degree in computer science in 1985 from the Institut National des Sciences Appliquées de Lyon. Born in Burgundy, he is married and has three children. In his spare time he enjoys playing the saxophone, reading philosophy, and outdoor activities such as skiing, soccer, and running.

The number of calls and their complexity have both increased. At the same time, support solution efficiency has decreased as the cost for providing those solutions has increased. As a result, there is a need for a knowledge sharing solution in which the first-line agents will be able to solve the majority of problems and escalate to the technicians only the complex problems. To enable such a solution, we have to:

- Find efficient knowledge extraction methods
- Create compact, efficient knowledge representation models
- Use extracted knowledge directly in the customer support operations.

This article describes the HP approach to providing customer support in the Windows®-Intel business segment. This segment includes networked desktop environments known for their high total cost of ownership. Help-desk services for this segment are supposed to solve the majority of problems with software applications, local area networks, and interconnections.

The system described here, called WiseWare,* is a knowledge harvesting and delivery system specifically designed to provide partially automated help for HP customer support centers in their problem solving chores.

Partial automation of help-desk support is seen as a suitable, cost-effective solution that will:

- Shorten the time spent per call
- Decrease the number of incoming calls (because of proactive mechanisms)
- Decrease the number of calls forwarded to the next support level
- Decrease the overall labor costs.

The objective is to reduce dramatically the support costs per seat per year.

Where Is Knowledge?

To find the most efficient knowledge extraction methods, we must first answer the question, "Where is the knowledge?" Books, technical articles, journals, technical notes, reports, and product documentation are all classical resources that rely on a human being's ability to extract, evaluate, and apply knowledge. Mechanized efforts still can't replace these human attributes.

Current support solutions usually are based on electronic collections in a free-text format, in which the important concepts are expressed using natural human language. The latest release of WiseWare uses technical notes, frequently asked questions, help files, call log extracts, and user submissions as the primary raw material. According to the knowledge resource, different knowledge representations and extraction methods are used.

Extensive research in the field of artificial intelligence has created several knowledge representation and extraction paradigms in which the final use for knowledge determines the characteristics of the representation scheme. The earliest knowledge extraction efforts, known as *information retrieval*, initially had small industrial impact. However, recent interest in the Internet and in electronic book collections has revived the business interest in information retrieval. Some of the hottest products on the market today are search engines. Different search methods (by keywords or by concepts) are being used and other search methods (by examples and by natural language phrases) are being investigated. Recent synergy with artificial intelligence methods has created a promising subfield known as intelligent information retrieval.[2] The majority of today's customer support solutions can be classified as enriched information retrieval systems.

Electronic Document Libraries

Developments in the information retrieval field have transformed free-text collections into more refined collections known as electronic document libraries. Electronic document libraries have an articulated structure (author, subject, abstract, and keywords), enabling efficient searches and classification. They combine advanced technological methods (such as hypertext and multimedia) to fit users' information retrieval needs. Some of the best support solutions today are in a digital library class and represent sophisticated document management systems.

Case-Based Retrieval

Early hardware support documentation contained troubleshooting diagrams that made it possible for service technicians to troubleshoot equipment consistently by following the diagrams and performing the appropriate tests and measurements. The recent revival of these diagrams is

* WiseWare is an internal tool and not an HP product.

Glossary

Cluster. Natural association of similar concepts, words, and things.

Concept. Group of words conveying semantic content. It can be described graphically as relationships between words having different attributes (and in some cases as numerical measurements of strength).

Data Mining. Collective name for the field of research dealing with data analysis in large data depositories. It includes statistics, machine learning, clustering, classification, visualization, inductive learning, rule discovery, neural networks, Bayesian statistics, and Bayesian belief networks.

Information Retrieval. Identification of documents or information from the collection that is relevant for the particular information need.

Keyword. Characteristic word that may enable efficient retrieval of relevant documents. Two criteria used to assess the value of a keyword are the number of documents retrieved and the number of useful documents (recall and precision)

Knowledge. Group of interrelated concepts used to describe a certain domain of interest. Complex structures formed by emulating human behavior in certain activities (for example, assessment, problem solving, diagnosing, reasoning, and inducing). Different schemes are used to enable knowledge representation such as rules, conceptual graphs, probability networks, and decision trees. Knowledge is found in large text collections and is biologically resident in human brains.

Knowledge Map. Graphical display of interrelated concepts.

Knowledge Base. Complex entity typically containing a database, application programs, search and retrieval engines, multimedia tools, expert system knowledge, question and answer systems, decision trees, case databases, probability models, causal models, and other resources.

Metrics. Group of measurement methods and techniques introduced to enable quantification of processes, tools, and products

Natural Language Processing. Activity related to concept extraction from, formalization of, and methods deployment in a problem area.

Paradigm. A theoretical framework of a discipline within which theories, generalizations, and supporting experiments are formulated.

Problem Domain. Area of interest defined by terminology, concepts, and related knowledge.

Search. Activity guided by a find and match cycle in which a search space is usually explored with an appropriate choice of search words (keywords). Advanced search is done by concepts.

seen in interactive troubleshooting systems that enable PC hardware technicians to solve hardware problems. So far, such systems are implemented as *case-based* retrieval (or reasoning) systems. The majority of these systems provide only retrieval; just a few include the reasoning component. The case-based retrieval paradigm is based on the human ability to solve problems by remembering previously solved problems. The support system plays the role of an electronic case database in which the knowledge consists of documented experience (cases). Creation and maintenance of the cases is an expensive and nontrivial process. Currently, these activities are performed by humans and are used mainly for hardware support. Such systems cannot deal efficiently with large, complex, and dynamic problem areas.

Rule-Based Systems

Some support centers have tried to use expert systems based on rules, but they have discovered that the rule-based systems are difficult to create, maintain, and keep consistent. Crafting a collection of rules is a complex chore. It is not clear if this technology will have a role in future knowledge representation and extraction development.

Model-Based Systems

A model-based paradigm in which various statistical, causal, probability, and behavioral models are used is another example of knowledge representation for customer support. The knowledge here is expressed by the fault/failure model that contains quantified relationships between causes, symptoms, and consequences. Basic decision making is enabled with such models. Although some limited experiments with this highly sophisticated knowledge representation paradigm have been done, no system is in operational use in support centers.

New Research

The newest research in the field of data mining and knowledge discovery[3] may offer some potentially effective knowledge representation methods for deployment in customer support centers. This research aims at the extraction of previously unknown patterns (insights) from the existing data repositories. Research in artificial intelligence has identified the initial assembly of a low-cost knowledge base as a potential "engineering bottleneck." The knowledge authoring environment discussion later in this article addresses that issue. Because most of the knowledge for WiseWare comes from text sources, we will focus our attention here on the knowledge extraction process.

WiseWare and Knowledge Refinement

Knowledge is a fluid, hard-to-define but essential ingredient for all human intellectual activities. It is difficult to extract, articulate, and deploy. The prevailing quantity of knowledge is encoded in the form of text (90 percent) expressed in natural language and is articulated as a web of interrelated concepts. A goal of research in natural language is to enable automatic and semiautomatic extraction of knowledge. Content analysis must be automated to efficiently provide suggestions and solutions for users. As we have already seen, several knowledge representation paradigms are being invented and investigated (for example, semantic nets, rules, cases, and decision trees). Additionally, we can deploy various techniques to extract concepts (symbolic knowledge) and numerical quantities (numerical and statistical knowledge).

Refinement Process

Human experts use spreadsheets, outline processors, and some vendor-specific tools to refine source text, but have not yet developed systematic, efficient processing methods. In the future, we would like to automate some phases of this process, leading toward more efficient and effective deployment.

Knowledge refinement is seen as a process for converting raw text into coherent, compact, and effective knowledge forms suitable for software problem solving and assistance (for example, decision trees, rules, probability models, and semantic nets). The basic raw material (the knowledge in its primary form) remains accessible. This preserves previous investments in knowledge and enables integration into future, more sophisticated solutions.

We can describe the knowledge refinement process as efforts made to transform raw text to a compact representation and then to operational knowledge. Associated costs increase as raw text moves through the refinement process to become operational knowledge.

Currently WiseWare content is partitioned into three conceptual categories: fixes, step notes, and technical notes. The first two contain shallow, specific knowledge and the third contains complex technical concepts. A fix is a simple, short document that describes with fewer than 100 words a known and recurring problem with a known solution, the fix (see **Figure 1a**). A fix often helps the customer out of the immediate problem but does not provide a long-term solution. It is essentially a "quick fix."

A step note usually walks the user through a procedure that prevents the problem from occurring in the future (see **Figure 1b**). The step note requires more of the user's time to solve the immediate problem than the fix does, but it saves time in the future.

Both fixes and step notes offer additional references. Those references contain keywords providing links to technical notes that explain the most relevant related subjects in depth. Technical notes require deep technical knowledge to be properly understood and applied.

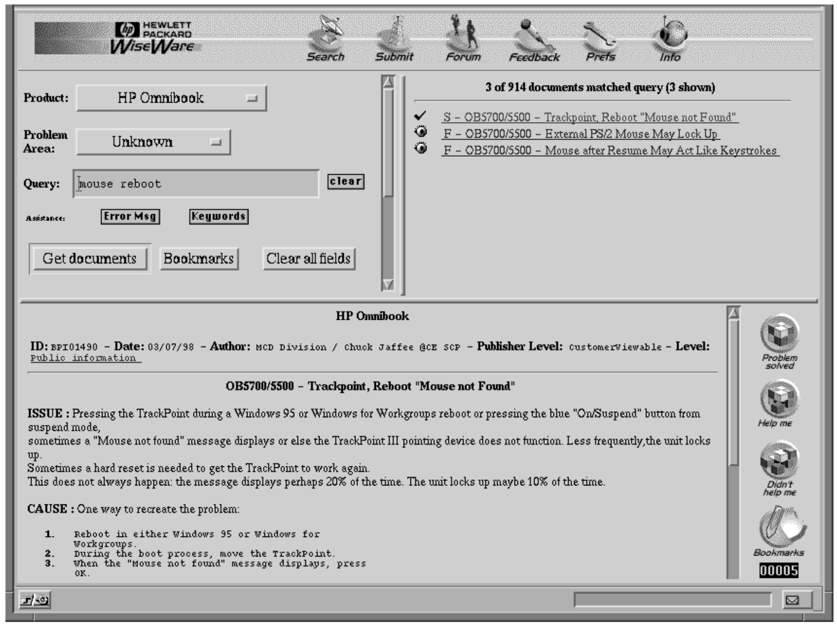
The whole collection of fixes, step notes, and technical notes is tagged to associate the content of each document with the proper problem classes. Consequently, WiseWare content is perceived by the user as a repository of advice and solutions for given problems (quick fixes, step-by-step procedures, and technical theory).

Some generic activities in the refinement process can be denoted as:

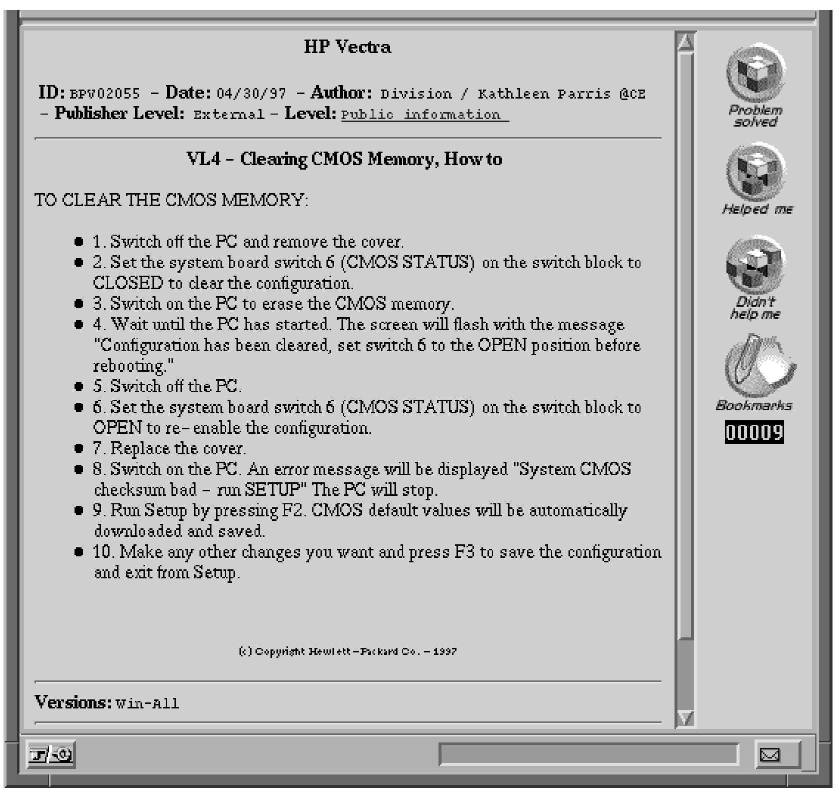
- Assessment
- Extraction
- Filtering
- Summarization
- Clustering
- Classification.

Figure 1

Two WiseWare screens: (a) WiseWare fix screen, (b) WiseWare step note screen.



(a)



(b)

We can describe the evolution of WiseWare as going from answering questions to giving advice and finally to problem solving and troubleshooting. The support costs in this evolution have escalated as the problems have become more complex.

Knowledge Authoring Environment

Since a critical mass of knowledge can be reached only if multiple authors contribute to the knowledge base, the knowledge authoring environment must be able to deal with multiauthor issues effectively. Additionally, because the knowledge authoring environment is deployed on a worldwide basis, the issue of different languages is relevant as well. Finally, deployment in different time zones requires very high reliability and availability of the knowledge authoring environment.

The quality of the knowledge is constantly monitored and refined. Areas for improvement are pinpointed by analyzing results reported on the knowledge base logs. As weak points are identified and strengthened, better system performance will help to optimize return on investment figures. Even user satisfaction can be assessed from the various logs and usage traces that will reflect a combined measure of system quality and usefulness.

Future worldwide cooperation among support centers to share knowledge is our objective. Ideally, each center will deploy and create the necessary knowledge locally. Centers operate in different time zones, have different cultural and social contexts, and have the ability to manipulate huge amounts of data, information, and knowledge. Coordinating the knowledge bases for all support centers pose several challenging problems. The complexity of these problems is reduced by careful engineering and incremental deployment. The result is a low-cost, knowledge-based support, adding new value to the support business.

In a very advanced situation, and from a long-term perspective, extracted knowledge will become the crucial ingredient for the next development phase. In this phase, human mediation in problem solving could be removed. Support could be delivered electronically without human intervention. For example, imagine intelligent agents traveling over the network to the troubled system to fix a problem.[4] Current viruses on the Internet are doing exactly the opposite task. What if the trend were reversed? Support knowledge could be adapted so that healing viruses could travel through a system, delivering remote fixes. To understand how this could become a reality, let's review the history of WiseWare.

WiseWare Architecture

In November of 1995, the first challenge was posed to the WiseWare team when the French call center decided to outsource low-end software support services. Their support personnel were without computer technology background and demonstrated poor English language skills. The knowledge department in HP's Software Services Division in Europe responded to the challenge and delivered the first operational WiseWare solution in April of 1996. Since then, new releases are issued every two months with steady improvements.

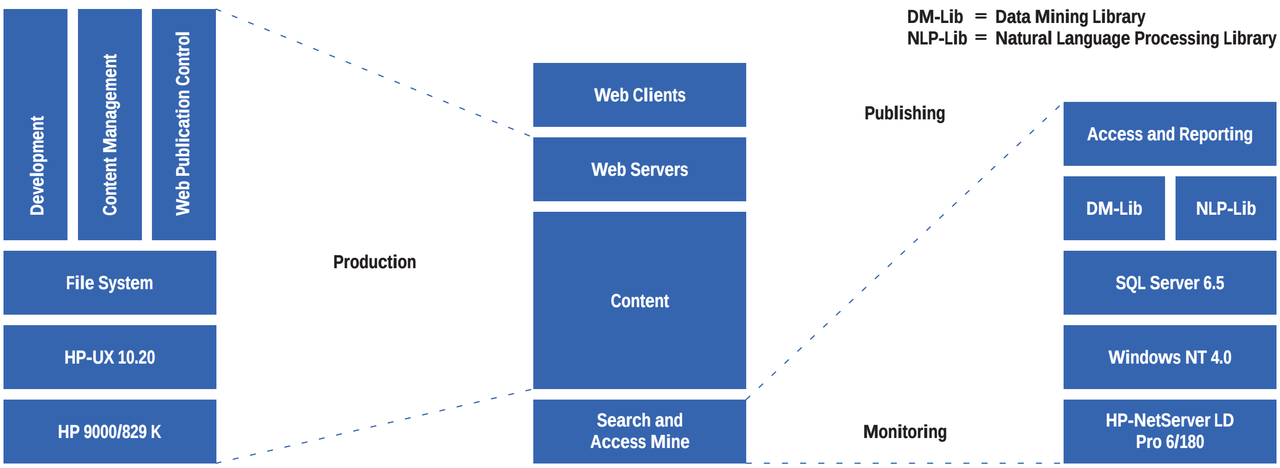
In the WiseWare release 4.1, mirroring intranet servers (Europe and the United States) cover three super regions. The number and quality of accessible documents is constantly improved, while use of the system is closely monitored from access and search logs. We have established close links with software vendors who allow us privileged access to their documents. (The legal framework for cooperation and alliances is defined as well.) All activities and services undergo quality assurance scrutiny in preparation for ISO-9000 certification.

WiseWare provides approximately 80,000 documents to 13 call centers worldwide. The average problem resolution assistance rate is over 30 percent. More than 40 products are covered in the various types of documents offering quick fixes for agents and in-depth technical knowledge for advanced WiseWare users.

WiseWare is a distributed system with three major parts: production, publishing, and monitoring (see **Figure 2**). They are implemented on UNIX® and Windows NT platforms, with intranet technology providing the necessary glue for client/server solutions. It is a nonstop, highly available system. The key advantage of the WiseWare system lies in the tight loop between the monitoring and production areas in which the principal objective is to provide users with highly adaptable documents for everyday problem-solving chores. Data mining and natural language processing modules dynamically create user, problem, and document profiles that will then drive the production side, enabling technical and business insights

Figure 2

WiseWare system architecture.



to be gleaned from large and extensive access and search logs.

At this time, customers call the express hubs and explain their problems to support personnel, using natural language constructs that sometimes blur the real nature of the problem. According to their understanding, support personnel create and launch a search phrase. It is a Boolean construct containing relevant keywords or free-text phrases that roughly represent the problem. Different search, hit, and presentation strategies are currently used, but formation of the effective search query and reduction of the number of relevant replies are largely still unresolved. A mixture of artificial intelligence techniques and traditional information retrieval and database methods is being offered as potential solutions.

Table I shows how one, two, and three words in a typical search phrase can influence the number of relevant documents returned with current version of WiseWare. A well-formed phrase helps to quickly pinpoint relevant documents while retaining necessary coverage of the problem area. Notice the quick decrease in the number of relevant documents returned as the phrase becomes longer.

Support center personnel work under time-pressured, stressful circumstances. As a result, the whole human-computer interaction issue must be carefully considered. Efficiently delivering advice and problem-solving assistance can depend on the smallest detail. Besides the quality of the material in the supporting knowledge base, questions regarding query formulation and presentation of the retrieved information will influence final acceptance from the users. Support activities can be treated as symbiotic human-machine problem solving in a bidirectional learning paradigm. The user learns how to manipulate the system (facilitated by language features such as localization and query wizards). At the same time, the system adapts to the user's methods of accessing the knowledge base. The WiseWare system learns user behavior from access and language patterns. Interaction with the system customizes the environment to suit the specific user's profile. The reasoning activity is still done by humans and is supported by refined electronic collections. Good synergy and efficient functioning of such human-computer systems are the current objectives.

Because the support centers are located in different geographical, cultural, and language areas, the natural language layer is seen as crucial for search and presentation.

Table I
Search phrases and number of documents returned

	Search Phrases/Documents Returned					
Application/Documents	**Change**	**Change+User**	**Change+User+ Password**	**Create**	**Create+New**	**Create+New+ Message**
WinNT/6046	1371	9	3	–	–	–
Win3.x/4397	1224	1	0	–	–	–
CC:Mail/2600	–	–	–	727	12	2
MS:Mail/3963	–	–	–	878	21	2

Technological advances in visual search and delivery combined with audio and video techniques may improve the quality and efficiency of the system. Better architecture combined with object-oriented (multimedia) databases will add another dimension to the delivery phase. These improvements will be made over time and will be accelerated by technological developments in related fields.

Conclusion

Accessible knowledge is the essential ingredient for successfully dealing with the rising quantity and complexity of customer support calls. A semiautomated system with refined knowledge in reusable forms can enable users to share knowledge among different, geographically dispersed customer support centers. The overall objective of HP's WiseWare server is to provide low-cost, effective customer support. This is a simple objective but one that is difficult to achieve, especially when significant effort and investment are required to achieve technological breakthroughs in the problem-solving field.[5]

In the short term, incremental deployment of advanced methods such as data mining and natural language processing techniques will improve system quality and usage. In the long run, it is very likely that most of the client-hub telephone voice communication will be gradually replaced by computer-computer communication. Several layers of the present problem- solving architecture will disappear or will be replaced by some new elements. The problem-solving knowledge along with search and access log collections being developed now will serve as the fundamental basis for future electronic support.

Acknowledgments

We would like to thank Markus Baer, Markus Brandes, and Jean-Claude Foray from HP's support labs for their role in WiseWare development. Its development would not have been possible without the understanding and support of Jim Schrempp and Alain Moreau. Finally, special thanks to all WiseWare team members.

References

1. http://www.HelpDeskInst.com/
2. http://ciir.cs.umass.edu/
3. http://www.kdnuggets.com/
4. http://retriever.cs.umbc.edu/agents/
5. http://www.gartner.com/

▶ Go to Next Article

▶ Go to Journal Home Page