Toward a Technology for Organizational Memories

Andreas Abecker, Ansgar Bernardi, Knut Hinkelmann, Otto Kühn, and Michael Sintek
German Research Center for Artificial Intelligence

TO MEET THE GROWING NEED FOR ENTERPRISEWIDE KNOWLEDGE MANAGEMENT, THE AUTHORS HAVE DEVELOPED AND FIELDED A THREE-LAYERED MODEL FOR PROCESSING KNOWLEDGE. THIS ARTICLE SHOWS HOW THEIR ORGANIZATIONAL MEMORY SERVES AS AN INTELLIGENT ASSISTANT AND DEALS WITH BOTH FORMAL AND NONFORMAL KNOWLEDGE ELEMENTS IN A TASK-ORIENTED FASHION.

THE RECOGNITION THAT KNOWLEDGE is one of an enterprise's most important assets, decisively influencing its competitiveness, has fueled interest in comprehensive approaches to the basic activities of *knowledge management*: the identification, acquisition, development, dissemination, use, and preservation of the enterprise's knowledge. Traditionally, enterprises have addressed knowledge management from either a management or a technological point of view. Managers understand that the knowledge their employees possess is one of their company's most valuable assets. They are concerned with the effective use of personal knowledge and the qualitative and quantitative adaptation of this knowledge toward a changing environment. The *technological approach*, by contrast, deals with questions about what information technology should be provided to support knowledge management.[1]

We find that effective knowledge management requires a hybrid solution, one that involves both people and technology.[2] As this article shows, our long-term vision is a *corporate* or *organizational memory* at the core of a learning organization, supporting sharing and reuse of individual and corporate knowledge and lessons learned. Arranged around such an OM, intelligent knowledge-management services actively provide the user working on a knowledge-intensive operational task with all the information necessary and useful for fulfilling this task (see Figure 1).

Our view of an organizational memory grew out of our practical experiences and also conforms well with definitions suggested in the literature: an OM's main function is to enhance the organization's competitiveness by improving the way it manages its knowledge. To achieve this goal, short-term efforts should concentrate on *knowledge preservation*,[3] which is based largely on explication of tacit knowledge and which is supported by expert systems, issue-based information systems, best-practice databases, and lessons-learned archives. Gaële Simon introduces the term *knowledge capitalization* for nearly the same process, which "allows [us] to reuse, in a relevant way, the knowledge of a given domain previously stored and modeled, in order to perform new tasks."[4] However, Simon emphasizes the exploitation of existing documents, which are based primarily on (maybe structured) natural language. In the long run, we feel an OM should also support *knowledge creation* and *organizational learning*.[5] As this article shows, an OM must be more than an information system but must also help to transform information into action.

Practical requirements

In recent years, we have performed several case studies, prototype developments, and evaluations concerning knowledge-based systems for supporting complex tasks in technical domains, such as motor design or the configuration of production facilities. Discussions with industrial customers and our growing understanding of their particular needs turned us away from the expert-system approach centered around the idea of an autonomous problem-solver. We have moved instead to the idea of an OM, which emphasizes the support of the human user by providing, maintaining, and distributing relevant information and knowledge. We have also concluded that an OM cannot simply be a passive information system, but must act as an intelligent assistant to the user.

Our experiences showed that the following requirements are crucial for an OM's success in industrial practice:[6]

- *Collection and systematic organization of information from various sources.* Knowledge needed in work processes is currently scattered among various sources, such as paper and electronic documents, databases, e-mails, CAD drawings, and the heads and private notes of individuals. The primary requirement for an OM is to prevent the loss and enhance the accessibility of all kinds of corporate knowledge by providing a centralized, well-structured information depository.
- *Minimization of up-front knowledge engineering.* Even though the advantages of having an OM are generally recognized, organizations are reluctant to invest time and money into a novel technology whose benefits are distant and uncertain. Furthermore, prospective users have little or no time to spare for requirements and knowledge acquisition. An OM thus must exploit readily available information (mostly databases and electronic or paper documents), provide benefits quickly, and be adaptable to newly arising requirements.
- *Exploiting user feedback for maintenance and evolution.* As with up-front knowledge engineering, OM maintenance efforts must be minimized. An OM also must deal with incomplete, potentially incorrect, and frequently changing information. To keep an OM up to date and gradually improve its knowledge, is is important to collect feedback from its users, who must be enabled to point out deficiencies and suggest improvements without significantly disrupting the usual workflow.
- *Integration into existing work environment.* To gain user acceptance, an OM must tap into an organization's existing flow of information.[3] At a technical level, the OM thus must directly interface the tools currently used to do the work, including word processors, spreadsheets, CAD systems, simulators, and workflow-management systems.
- *Active presentation of relevant information.* In industrial practice, costly errors are often repeated due to an insufficient flow of information. A passive information system cannot avoid this situation, because workers are often too busy to look for information or don't even know that pertinent information exists. An OM therefore should actively remind workers of helpful information and be a competent partner for cooperative problem-solving.

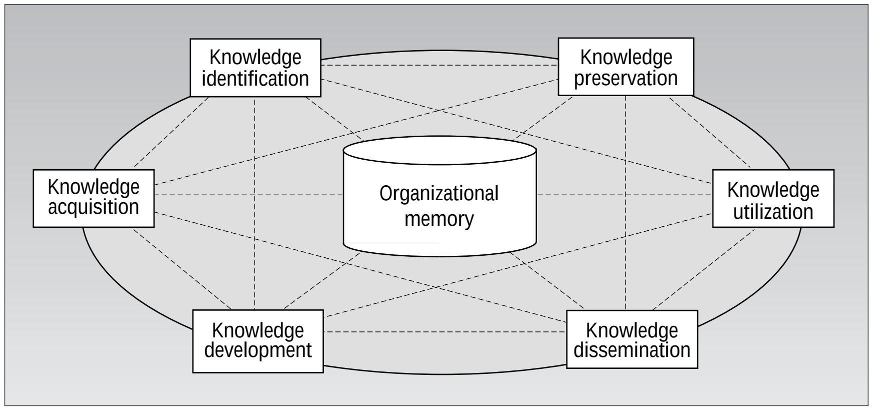

Figure 1. The organizational memory assists in the basic knowledge-management activities.

Organizational memory assistant systems

Our case studies underscore the need for computer-assisted knowledge capitalization. This corresponds well with Thomas Davenport's demand for a hybrid solution for knowledge management: information technology can easily capture, transform, and distribute large amounts of highly structured knowledge. But, for tacit, hard-to-formalize knowledge that must be interpreted in a broader context and combined with other types of information, humans are the "recommended tool."[2] If the effort for formalizing knowledge is too high, it should be left informal and processed by humans. However, information technology can increase the quality of a person's decision-making and problem-solving, such as by providing relevant informal knowledge in the actual work context. Thus, the machine amplifies human knowledge.

The hybrid approach for knowledge management corresponds well with the shift of focus in artificial intelligence. While an important AI goal has been to build knowledge-based systems that solve challenging problems on their own, an intelligent-assistant system cooperates with a human user in solving a problem. It contributes to the solution, for instance, by solving subproblems, performing calculations, or verifying or criticizing the user's solutions.

This approach has drawn considerable interest for a variety of reasons. Well-known difficulties of conventional knowledge-based systems such as brittleness or limited user acceptance called for adequate solutions. To solve important problems, it is often better to let the computer work out what might be done but to let the human user decide, thus distinguishing workload versus decision competence and responsibility. The combination of the assistant system and the user improves both problem-solving capabilities and user acceptance. Fredrick Brooks illustrated this cooperation by the formula[7]

$$IA > AI$$

meaning that an intelligence-amplification system—machine *and* a mind—can beat an AI system—a mind-imitating machine working by itself.

Providing knowledge as a central OM service. In the following, we will concentrate on one central service of the OM: providing the necessary knowledge whenever it is needed. For this, the OM realizes an active knowledge dissemination and usage approach that does not rely on a user's queries but automatically provides knowledge useful for solv-

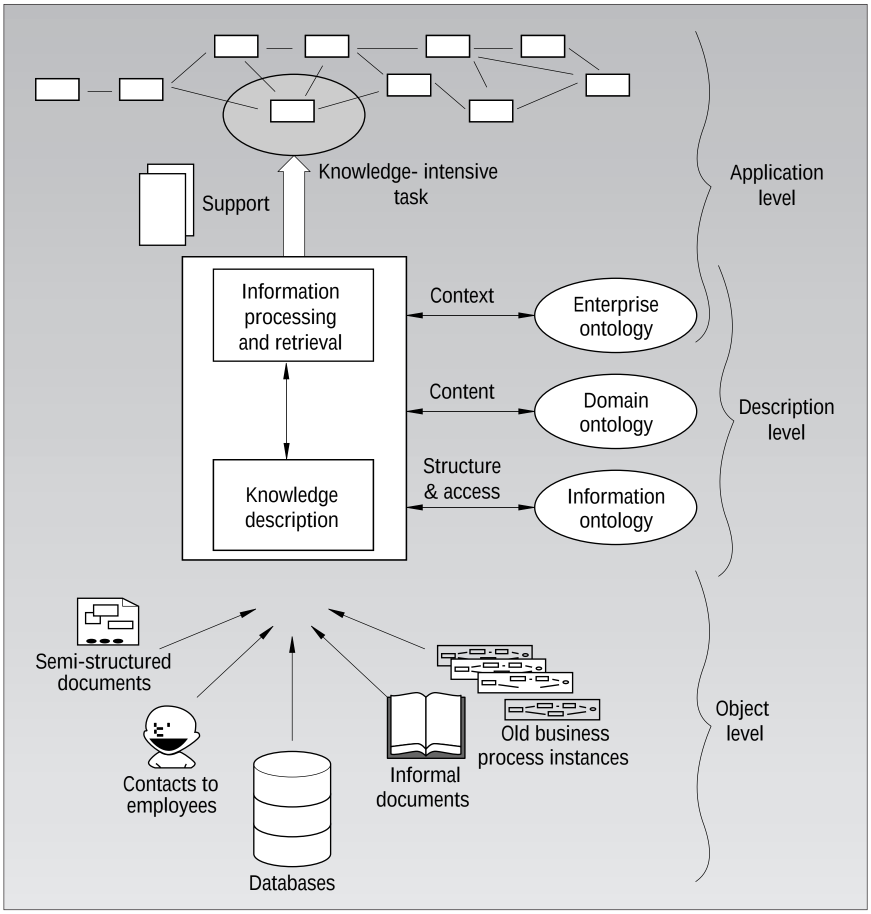

Figure 2. Intelligent support by context-sensitive knowledge supply.

ing the task at hand. The resulting system acts as an intelligent assistant that

- accompanies the execution of tasks, and
- presents relevant information that helps workers do their jobs better and more effectively.

For an OM to be effective, users must receive relevant information at the right time without being overwhelmed with a flood of irrelevant data. Information is relevant only if users can perform their task better with this information than without it. Thus, relevance of information is always defined with respect to its use. Consequently, actively providing information in an OM—in contrast to conventional information filtering—is primarily oriented according to a task model in addition to a user model. Until now, knowledge on task-specific relevance has been only implicitly represented in application programs, encoded in database queries, or not represented at all, but hidden in assumptions underlying the active by-hand navigation in hypertext information systems. Instead, we propose to explicitly represent the relationship between task, application situation, and knowledge context in a declarative way. Explicit modeling facilitates application development and maintenance, makes automatic analyses possible, and allows for systematic evolution of the OM content and behavior over time.

A functional view

The biggest profit from support by an OM will likely come in tasks that are complex, difficult, and important by nature. To perform these tasks, the human experts need considerable skill and knowledge. Such *knowledge tasks* deal with the acquisition, creation, packaging, and application of knowledge, and can be increasingly identified inside the core competencies of modern enterprises.[8] Given their characteristics, a complete automation—or even a very detailed partition into subtasks—is usually not feasible because there is no predetermined task sequence that, if executed, guarantees the desired outcome. In fact, what we call *knowledge tasks*, or *knowledge-intensive tasks*, essentially amount to the notion of *wicked problems* introduced by Rittel in the 1970s and extensively discussed by E. Jeffrey Conklin and William Weil, in the context of organizational memories.[9]

An OM helps a user perform knowledge tasks by actively providing useful information and knowledge. Knowledge tasks are often embedded into more "tame" work processes, which are linked to them by exchange of information, decisions, and documents. Thus, the embedding business processes naturally provide the context for performing, analyzing, and supporting knowledge tasks. We propose a three-layered model as sketched in Figure 2, which points out the main issues to be addressed when building a system for realizing context-sensitive, active knowledge supply.

Our approach models and executes processes and tasks on the *application level*. When a knowledge worker recognizes an information need within the actual flow of work, a query to the OM must be derived. This query is instantiated and constrained as specifically as possible on the basis of the actual work context. In the opposite way, the OM can also store new information created within a given working situation in a contextually enriched form such that subsequent retrieval processes might compare the query situation with the creation situation for estimating context-specific relevance.

As one of many possibilities for realizing the application level, we include conventional business-process models and workflow-management systems. Doing so lets us rely on a body of well-understood knowledge already formalized in enterprises and used to guide and support work processes.

Because an OM relies substantially on existing information sources, the *object level* is characterized by a variety of sources, heterogeneous with respect to several dimensions concerning form and content properties. The OM performs the mapping from the application-specific information needs to these heterogeneous object-level sources via a uniform access and utilization method on the basis of a logic-based, knowledge-rich *knowledge description level*.

Object level. This level comprises manifold information and knowledge sources, ranging from machine-readable formal representations to human-readable informal representations. Crucial parts of corporate knowledge to be processed by the computer must be formalized, whereas other parts that need only

```
recommendation: c1
iff   the machine-type of crankshaft is (farmer or hobby)
then  the weight of crankshaft should be [150 ... 180]
 "Somewhat heavier crankshafts (weighing 150-180 grams) are used in low-end machines,
since a low weight is less important for occasional users."
 (a)
constraint: c5
if    the machine-type of crankshaft is professional
then  the bearing-type of main-end-bearing must be full-needle
and   (the width of main-end-bearing) = (the width of low-end-bearing)
 "Only a full-needle bearing is durable enough for heavy use."
 (b)
```

Figure 3. Example of (a) a recommendation and (b) an implicational design constraint.

be understood by humans might be left informal. The decision whether to formalize or not rests on cost-benefit analyses, stability of knowledge, and the question whether some portion of knowledge *can* reasonably be formalized at all.

On the one hand, an OM will have to deal primarily with less formalized knowledge contained in more or less structured, preferably electronic documents. These informal or semistructured knowledge representations are well tailored to human needs. Readily producible and easily understandable natural language, graphics, or images can express and exchange various kinds of knowledge. Lessons-learned databases—for example, those stored in an intranet or a Lotus Notes environment—are a typical example of informal knowledge. Other frequently found sources of informal knowledge are human-resource knowledge bases about employee capabilities and skills. Existing informal documents such as manuals of technical systems are an important knowledge source that should be made accessible through an OM.

On the other hand, such informal knowledge can neither be operationalized for automatic problem-solving (such as expert-system rules) nor processed by complex query-answering mechanisms (such as databases). Hence, its usefulness for supporting human problem-solving is limited. Elsewhere, we presented an OM prototype for crankshaft design for a manufacturer of motor-powered tools and vehicles.[6] In this system, parts of the company's design knowledge are formalized as rules that are employed for critiquing and suggesting viable solutions. We distinguished two types of design rules: hard constraints, which should always be satisfied, and soft recommendations, which allow occasional exceptions. The reasons for the design rules, which are important for assessing the relative merits of several viable design alternatives, are attached as natural-language annotations taken directly from the expert interview protocols. Figure 3 shows a recommendation and an implicational design constraint.

The costs and effort required for formalizing major amounts of knowledge are usually prohibitive. Therefore, an OM primarily contains informal knowledge sources, organized and made accessible by more formal notions. In this specific example, it made sense to formalize the described rules because they influenced highly important design decisions. Moreover, because the system only generated advice and critique and was never intended for fully automatic operation, the requirements on consistency and completeness of the knowledge base were much easier to achieve than for a conventional knowledge-based system. Also, the underlying knowledge-evolution approach aims at a continuous knowledge capture on the job and does not depend on a successfully completed knowledge-acquisition phase before system deployment.

Knowledge-description level. This level enables a uniform, intelligent access to a diversity of object-level sources. Because legacy information systems must be incorporated without modification, we propose a separate, knowledge-rich information-modeling level. Essentially, its purpose is to ease

- precise selection and efficient access to information and knowledge recognized as relevant in a given task context and application situation, and
- better comprehension and interpretation by the user and the system in a given task and application context.

The "Information modeling" sidebar describes the operations of this level in detail.

Application level. As already mentioned, the OM's application level links the information model and the concrete application situation. Parameters of the actual work context map onto expressions of the OM repository and result in the appropriate queries and assertions. Using this basic functionality, we can realize the OM's various services in different ways, ranging from dedicated programs, which perform a particular, well-defined task, to flexible and universal query interfaces, which allow the user to access the information contained in the OM. As usual, dedicated approaches result in tight support of specific activities, while more flexible approaches cover larger areas but offer less specific support, thus requiring more user interaction.

Concentrating on providing knowledge items relevant to assist the human expert in solving knowledge-intensive tasks, we want to discuss our approach for realizing a flexible but concise coupling to an enterprise's activities. This approach relies on an explicit modeling of the relation between task, application situation, and knowledge context.

Business-process-oriented knowledge management. The first prerequisite for representing task-specific relevance of knowledge items is a suitable representation of the tasks in question. In general, any particular task in an enterprise is part of some comprehensive process—for example, product development. To develop a general approach for context-specific relevance assessment, we selected business-process modeling as a promising starting point.[10] Business-process models have proved valuable both in business-process reengineering and as a basis for dynamic enactment by workflow-management systems. Beyond the plurality of the various systems and tools, the Workflow Management Coalition defined a standard model.[11] This widely accepted methodology

Information modeling

Every information and knowledge item is described by a number of attributes representing the information metamodel, the information content, and the creation and application context. The concepts for the knowledge descriptions are specified in ontologies (see Figure A).

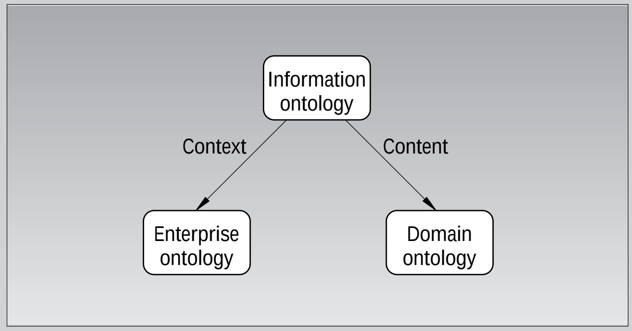

Figure A. Three ontologies span the basic dimensions of information modeling.

Information metamodeling

The *information metamodel* describes the different kinds of information sources with their respective structure, access, and format properties. The vocabulary for the information-source metamodels comes from the *information ontology,* which also contains generic concepts and attributes that apply to all kinds of information—such as the timeliness, the author, the reliability of information, or the type of statements an information source makes. For instance, it might express *descriptive* knowledge about products and processes or *prescriptive* knowledge stating how to do certain things. The information ontology also introduces concepts and attributes specific for certain kinds of information sources. For instance, access to an external commercial database involves costs and time delays, whereas personal competencies must be accompanied by the level of expertise of an employee and her availability.

Essentially, the information ontology comprises all aspects of information and knowledge sources that are not content-specific. It also provides links into the domain ontology used for content description, and it provides links into the enterprise ontology used to describe the creation context and the intended utilization context of knowledge items. The simplified example shown in Figure B gives an impression of how the several ontologies interact.

If an enterprise information ontology is accomplished, it should be reusable with only small adaptations for most enterprises. Its concrete design is still subject to our ongoing research. There are already contributions for specific kinds of information sources. For instance, it is usual to model logical and layout structure of printed documents. Embedding documents and document flow into organizational context is investigated in Office Automation and Document Analysis and Understanding.[1]

Besides the more syntactic and contextual issues, information metaproperties are of particular importance for realistic applications. Concerning those, hints emerge in the areas of lessons-learned archives (for distinguishing form, content, and availability as the basic knowledge-modeling dimensions[2]) and business-knowledge navigation (for identifying form, quality, and resource constraints as crucial retrieval factors besides the content[3]). The retrieval system can derive just such retrieval constraints from the formalized query context, thus supporting the precise selection of useful information.

Content modeling

For modeling the content of information sources, we use terms from a *domain ontology*. For the design of this ontology, we can build upon metamodeling mechanisms for databases, formal knowledge, and text documents:

- *Ontologies* and *data models* are used in knowledge-based and database systems, respectively, to specify the basic assumptions that

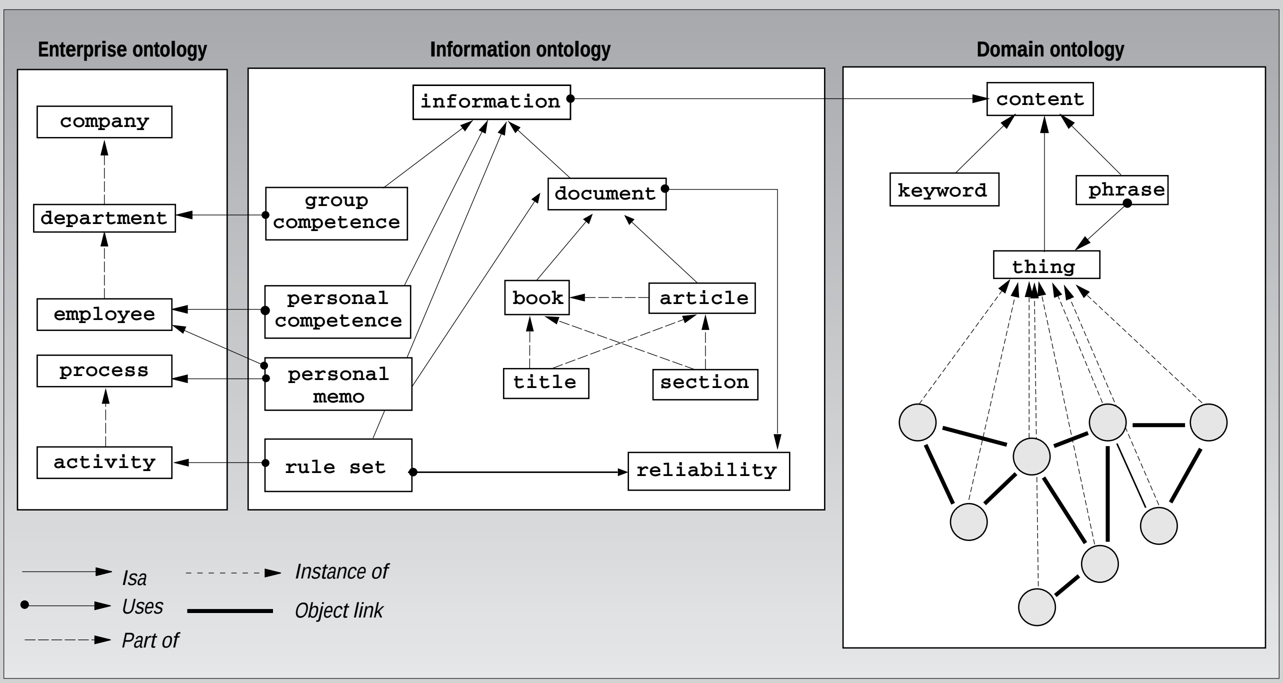

Figure B. Simple knowledge-description ontologies in some more detail.

went into the system's conceptualization.[4] Semantic data models are simple kinds of ontologies.

Using formal ontologies for content description lets an inference component make formal inferences over kinds of the represented knowledge and lets us formulate retrieval heuristics that exploit the domain structure. While most such *conceptual information retrieval* approaches extend the keyword-based content characterization of classical information retrieval by embedding the content identifiers into a formal structure, they nevertheless rely on quite a *sparse* content characterization. Recent approaches have proposed more detailed content models that allow precoordination of concepts, for example.[5]

- *Classification systems* are used in digital libraries and document-management systems. Classification systems are pragmatically designed for optimum human ease of access. They do not aim at a semantically clear formal model.
- *Thesauri* are repositories of lexical semantics. Highly sophisticated, hand-crafted thesauri (such as WordNet[6]) capture more semantics than most known formal ontologies. In contrast, similarity thesauri, which are usually automatically built up from collections of available documents, represent weighted-term associations for use in query expansion in information retrieval or for document classification.

The integration and conjoint use of these kinds of metalevel descriptions poses an interesting research question, because they differ considerably in their depth of modeling the world, the methods used, and their typical use. Because an OM contains both formal and informal knowledge, we propose an integrated use of ontology and thesaurus for domain modeling. The concepts of the domain ontology are the basic primitives for the formal knowledge representation. In addition, the integrated use of ontology and thesaurus contains lexical information for use in classifying and accessing informal knowledge. For instance, the concepts of the ontology are extended with equivalent linguistic terms that occur in textual documents.

Context modeling

In addition to the usual modeling dimensions of information retrieval, we focus on context as highly relevant for retrieval within an organization. Context modeling concerns two issues:

- the intended application context of a knowledge item, and
- the context a knowledge item was created in.

For instance, if a notice about some customer or supplier has been created within a certain business context—such as price negotiations—this context information can be very valuable in determining the relevance of this notice in a new application context. For particular kinds of information—such as best-practice reports, lessons-learned or formal design rules—the application task can be specified in advance. We suggest expressing information context in terms of the organizational structure and the process models. These in turn are expressed in terms of the *enterprise ontology*. The design of the enterprise ontology builds on insights and developments from enterprise modeling, from business-process modeling, and organizational modeling in knowledge-based systems.[7,8]

Table A. Sample knowledge-item description.

Metaproperties	Name	"How to achieve good payment conditions From Borg Inc."
	Author	James T. Kirk
	Nature	Activity-related advice
	Type	Heuristic, experience-based
	Form	English text, MS Word source, version 28.0
	Source	File E:\home\experiences\ds9-12-99pn.doc
	Availability	Always
	Costs	None
Context	Creation process	Project ds9 for Starfleet Corp. in Dec. '99
	Creation activity	Price negotiation with hardware suppliers
	Creation department	Purchasing dept.
Content	Product	20 SUN Ultra
	Supplier	Borg Inc.
	Contact person	Dr. Darth Vader

Knowledge description in lessons-learned archives

Table A shows a sample knowledge-item description in kind of a frame representation that was inspired by knowledge-item descriptions for organizing lessons-learned or best-practice databases.[2]

Implementation aspects

For representing and infering knowledge descriptions, most research approaches rely on description logics or relational databases and Datalog; some newer projects build on Frame-Logic. From the performance and integration points of view, we prefer database-oriented systems with deductive, object-oriented facilities, extended by mechanisms to handle the various forms of uncertainty naturally coming into play in information retrieval.

References

1. A. Celentano, M.G. Fugini, and S. Pozzi. "Knowledge-Based Document Retrieval in Office Environments: The Kabiria System," *ACM Trans. Information Systems*, Vol. 13, No. 3, 1995, pp. 237–268.
2. G. van Heijst, R. van der Spek, and E. Kruizinga, "Organizing Corporate Memories," *Proc. KAW '96*, Univ. of Calgary, Knowledge Science Inst., 1996; http://ksi.cpsc.ucalgary.ca/KAW/KAW96/KAW96Proc.html.
3. D. Steier, S.B. Huffman, and W.C. Hamscher, *Meta-Information for Knowledge Navigation and Retrieval: What's in There*, white paper, Price Waterhouse Technology Center, Menlo Park, Calif., 1995; http://www.pw.com/tc/2176.htm.
4. T. Guber, "A Translational Approach to Portable Ontologies," *Knowledge Acquisition,* Vol. 5, No. 2, 1993, pp. 199–220.
5. P.E. van der Vet and N.J.I. Mars, "Coordination Recovered," *Informatiewetenschap 1996*, Delft, The Netherlands, 1996, pp. 129–151.
6. G.A. Miller, "WordNet: A Lexical Database for English," *Comm. ACM*, Vol. 38, No. 11, 1995, pp. 39–41.
7. M. Uschold et al., *The Enterprise Ontology,* tech. report, AIAI, Edinburgh, Scotland, Oct. 1995.
8. W. Post et al., "Organizational Modeling in CommonKADS: The Emergency Medical Service," *IEEE Intelligent Systems*, Vol. 12, No. 6, Nov./Dec. 1997, pp. 46–52.

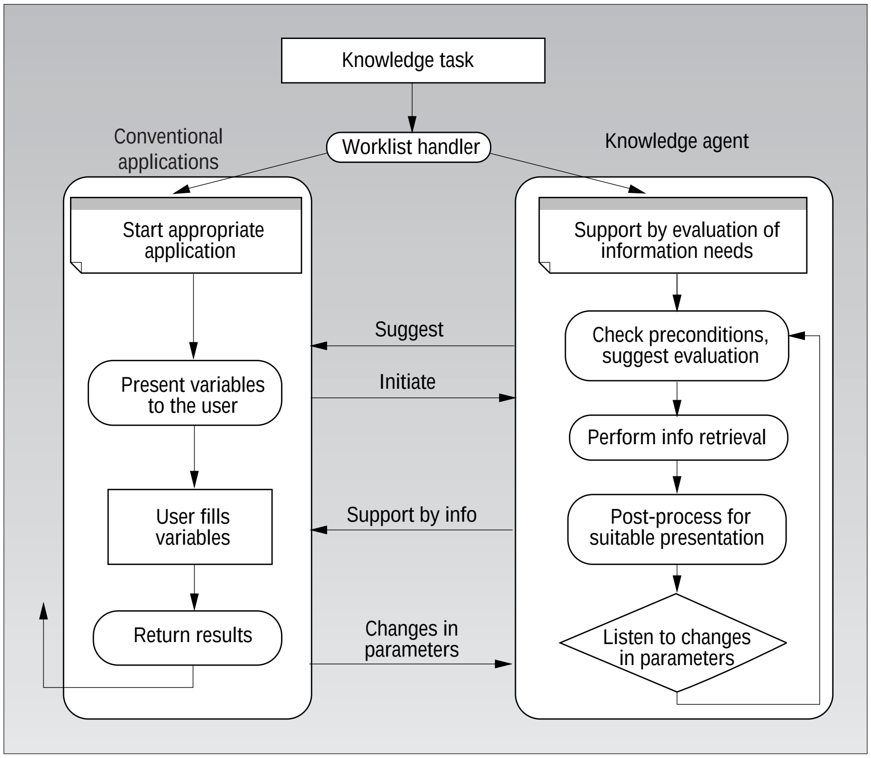

Figure 4. Principal steps for knowledge-task support.

offers a declarative task representation that can be extended to reflect the OM-specific needs. The dynamic execution of a process model by some workflow engine offers the necessary hooks to realize the active support by the OM.

Descriptions of tasks in a business process comprise the predecessors and successors in the process, the particular activity with tools and resources, and the variables that are accessible from the workflow system and the tools. Knowledge tasks as parts of a business process are not deeply modeled. Consider, for example, the credit process in a bank. Making the decision whether a customer is creditworthy is surely a knowledge task. The process model, however, represents it as a simple task of filling out a particular form.

To enable the OM to provide optimal support for tasks of this category, we developed a representation frame for knowledge tasks. This representation extends the formalisms adapted from the business-modeling approach: it characterizes a knowledge task by the specification of supporting information that helps the human achieve the goals of the task. To fulfill an information need, actions of varying complexity can be imagined, ranging from database queries using well-defined selections, to arbitrary deductions on content descriptions of a document base or calculations by some expert system.

We interpret the complete process model as the *global context* of the knowledge task, as the process ultimately describes the objective of the task sequence. During enactment by the workflow system, the instances of the variables form the *local context* of the particular task instance, providing the necessary information about the environment for the actual activity. This is completed by a variety of *workflow control data* that the workflow system offers and that inform about the state of the particular process instance.

This integration of workflow-related information into the knowledge-handling mechanisms provides answers to three guiding questions that play a crucial role in effective support:

- What is the overall goal of a particular activity, and which support is needed? (this is represented in the process model),
- What contextual information is already known at this particular instance? (this is answered by the workflow activity instance), and
- When is the support appropriate? (when the activity is started during workflow execution).

Extended representation of knowledge tasks. To realize the concepts just discussed, we extend the representation of a knowledge task beyond what is usually represented in a business process model by

- A set of information-need descriptions. Each information need bears a name and is characterized by a set of preconditions (governing when the information need has to be considered), a parametrized information retrieval query, and a declaration of the local goals to which it contributes.
- A set of postprocessing rules that influence the presentation of nonformalized information or guide the processing of formalized data. Results of these processing steps might trigger preconditions of some information needs.

This modeling of some knowledge task occurs during process definition. The human expert creating the extended process model formulates information needs that result in relevant information at runtime.

Processing a knowledge-task model. The evaluation of the extended representation of a knowledge task at runtime lets the OM realize its role as an assistant system. Controlled by the workflow enactment service, the activation of a knowledge task starts both the conventional application (for example, an editor with a form needing some key data and a decision whether the credit should be granted) and an additional knowledge agent (see Figure 4). The latter evaluates the preconditions and offers the available information as possible support to the user. On the user's demand, the system instantiates the current parameters and performs the information retrieval. The postprocessing rules determine the presentation of the result, which finally enables the user to proceed in the knowledge task. If the user decides that the task has been completed, control transfers back to the workflow system.

Knowledge acquisition and maintenance

Knowledge acquisition and maintenance—the main reasons why knowledge-based systems so often failed in industrial practice—also pose a serious challenge for OMs. As our industrial case studies showed, minimizing costs for up-front knowledge engineering is critically important. Furthermore, an OM resides in a dynamically changing environment and is thus subject to frequent changes and adaptations.

Enterprises can successfully develop and maintain an OM by adhering to the following principles:

- Exploit easily available information sources.
- Forgo a complete formalization of knowledge.
- Use automatic knowledge-acquisition tools.
- Encourage user feedback and suggestions for improvements.
- Check the consistency of newly suggested knowledge.

Numerous experiences have shown that for reasons of cost-efficiency and practicability, knowledge acquisition and maintenance should rely as little as possible on human experts. Our approach thus relies mostly on documents and databases, as a cheap, plentiful, and easily available source of information. More expensive user feedback should be used sparingly to detect missing, invalid, or outdated knowledge.

In the following description, we distinguish between techniques for extending the object-level knowledge and the ontologies.

Updating the object level. As we've discussed, for every document added to the OM, a knowledge description must be added. Generating knowledge descriptions for textual documents can be supported by document-classification and information-extraction techniques. The description of the document *content* in part can be extracted automatically. In our department, we developed document-analysis techniques that extract relevant information, learn appropriate classes of documents, or classify documents according to a given set of classes.[12] Some of the techniques are specifically qualified to extract information from printed documents and thus allow the reuse of already existing documents in an OM.

Databases are a second source of useful knowledge. Knowledge discovery from databases or data-mining techniques use tools from AI, mathematics, statistics, and visualization to extract knowledge from operational databases or data warehouses. For example, classification techniques let users extract customer profiles from sales databases.

Exploiting automatic thesaurus generation for ontology construction. Agreed-upon domain ontologies do not exist for many domains. Therefore, the development of domain ontologies and data models takes place prior to filling in the object-level information (with possible iterations). It consumes a considerable portion of the total development effort. Figure 5 shows a procedure for knowledge acquisition and maintenance that is based on the stated principles.

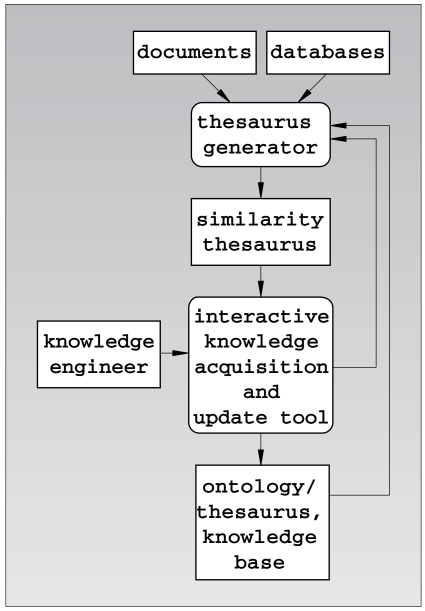

Figure 5. Thesaurus generation as the basis for conjoint ontology/thesaurus evolution.

For exploiting the information from document collections, we developed an automatic thesaurus-generation tool that combines several state-of-the-art thesaurus-generation methods. To focus the thesaurus-generation process, a list of keywords might be supplied that are known to play an important role in the respective application domain. Such keywords are usually available from databases or, in the case of an update, from an existing OM.

The thesaurus generator extracts relevant terms and relations from the given set of documents. Even though current techniques only allow the automatic extraction of similarity relations (or, more precisely, "has-to-do-with" relations), the so-obtained similarity thesaurus shows in a condensed form many interesting relationships between important domain terms. By adjusting generation parameters, we can extract different kinds of terms and relations, without having to read the respective documents.

As compared to a manual knowledge-acquisition-from-text approach, in which a domain expert or a knowledge engineer reads selected texts to identify relevant pieces of knowledge, automatic thesaurus generation is particularly effective when large amounts of text are available. These texts might even be of rather poor quality and contain a mixture of relevant and irrelevant information. This is often the case with documents that are routinely created during work processes. But, high-quality documents summarizing the essentials that might be used for a manual knowledge acquisition are often not available or are not detailed enough to be really useful.

The knowledge engineer then integrates the information obtained from the thesaurus generator into the OM semiautomatically, scanning the similarity thesaurus and deciding which relations should be formalized and added to the knowledge base or ontology, which should be included in the thesaurus integrated with the ontology, and which should be ignored.

Because a manual inspection and classification can only be performed for a limited number of similarity relations, a tool supports the knowledge engineer that highlights the most prominent relations not yet known in the current organizational memory. This tool also lets him specify actions based on various criteria, which are then automatically applied to the other terms and relations in the similarity thesaurus. The described knowledge-acquisition and update procedure can be reiterated at regular intervals when a sufficient number of new documents or database entries have accrued.

WHILE THIS VIEW ON ORGANIZATIONAL memory covers the necessary steps to realize an active assistant system for providing and managing context-sensitive information, it is by no means a closed or complete presentation of the topic. In general, organizational memory cannot be understood as a closed research area of its own; it merely grows out of a pragmatic integration of manifold AI—and other—techniques driven by an ambitious application goal. Figure 6 shows a number of important research areas that contribute to OM technology, organized according to our three-layered view on OM realization. Dan O'Leary presents a further idea of possible protagonists in this domain.[1]

Our view of OMs grew out of industrial experiences and is constantly checked against reality by application projects that use and criticize our basic research results. Important aspects of the principles we've discussed have been realized, for example,

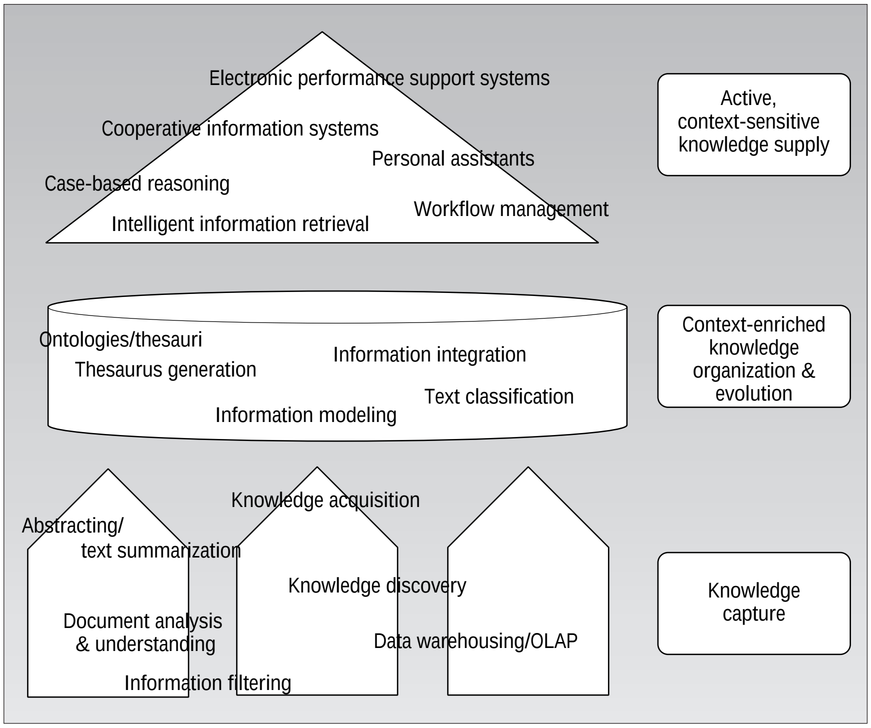

Figure 6. Several research areas contributing to an OM technology.

in a recently developed system for intelligent fault recording and maintenance support in a mechanical-engineering domain.[13] This application exploits vast amounts of well-structured but nonformal technical documentation together with a concise domain ontology and a crisp task model and has proved already to successfully contribute to effective knowledge conservation and dissemination. ■

References

1. D. O'Leary, "Enterprise Knowledge Management," *Computer*, Vol. 31, No. 3, Mar. 1998, pp. 54–61.

2. T.H. Davenport, "Some Principles of Knowledge Management," Apr. 1996, http://www.bus.utexas.edu/kman.

3. E.J. Conklin, *Designing Organizational Memory: Preserving Intellectual Assets in a Knowledge Economy*, white paper, Group Decision Support Systems, Washington, D.C., 1996; http://www.gdss.com/DOM.htm.

4. G. Simon, "Knowledge Acquisition and Modeling for Corporate Memory: Lessons Learnt from Experience," *Proc. KAW '96*, Univ. of Calgary, Knowledge Sciences Inst., 1996; http://ksi.cpsc.ucalgary.ca/KAW/KAW96/KAW96Proc.html.

5. I. Nonaka and H. Takeuchi, *The Knowledge-Creating Company*, Oxford Univ. Press, Cambridge, UK, 1995.

6. O. Kühn and A. Abecker, "Corporate Memories for Knowledge Management in Industrial Practice: Prospects and Challenges," *J. Universal Computer Science, Special Issue on Information Technology for Knowledge Management*, Vol. 3, No. 8, 1997, pp. 923–954; http://www.iicm.edu/jucs_3_8.

7. F.P. Brooks, "The Computer Scientist as Toolsmith II," *Comm. ACM*, Vol. 39, No. 3, Mar. 1996, pp. 61–68.

8. T.H. Davenport, S.L. Jarvenpaa, and M.C. Beers, "Improving Knowledge Work Processes," *Sloan Management Rev.*, Vol. 37, No. 4, Summer 1997, pp. 53–65.

9. E.J. Conklin and W. Weil, *Wicked Problems: Naming the Pain in Organizations*, white paper, Group Decision Support Systems, 1997; http://www.gdss.com/wicked.htm.

10. D. Karagiannis, S. Junginger, and R. Strobl, "Introduction to Business Process Management Systems Concepts," *Business Process Modeling*, B. Scholz-Reiter and E. Stickel, eds., Springer-Verlag, Berlin, 1996.

11. *Workflow Management Coalition Terminology & Glossary*, document WFMC-TC-1011, The Workflow Management Coalition, Hampshire, UK, 1996; http://www.wfmc.org.

12. A. Dengel and K. Hinkelmann, "The Specialist Board—A Technology Workbench for Document Analysis and Understanding," *Integrated Design and Process Technology, Proc. Second World Congress*, Soc. for Design and Process Science, Austin, Texas, 1996, pp. 36–47.

13. A. Bernardi, "Electronic Fault Recording: A Corporate Memory for Maintenance Support of Complex Machines," *Proc. Int'l Symp. Management of Industrial and Corporate Knowledge*, IIIA, Compiègne, France, 1997, pp. 129–136; http://www.dfki.uni-kl.de/~bernardi/papers/ismick97.ps.

Andreas Abecker is a research scientist at DFKI—the German Research Center for Artificial Intelligence—in Kaiserslautern. Within the DFKI Knowledge Management Group, he investigates methods and tools for building, maintaining and using OMs for learning organizations. His research interests also include knowledge representation and reasoning as well as machine learning and data mining. He received a masters in computer science from Kaiserslautern University.

Ansgar Bernardi is a research scientist at DFKI, where he has participated in research on knowledge management and leads an R&D project that resulted in the deployment of a corporate memory for maintenance support. His research interests include the representation of technical knowledge, AI support for production planning and scheduling, and decision support for comprehensive product design. He received a masters in computer science from Kaiserslautern University.

Knut Hinkelmann is head of DFKI's Knowledge Management Research Group. His research interests are knowledge management, knowledge representation and reasoning, deductive databases, and business-process management. He received a masters in computer science and a PhD in knowledge processing and deductive databases from the University of Kaiserslautern. He is member of the IEEE and AAAI. Contact him at DFKI, Postfach 2080, 67608 Kaiserslautern; knut.hinkelmann@ dfki.de; http://www.dfki.uni-kl.de/~hinkelma.

Otto Kühn is a senior researcher in the DFKI Knowledge Management Group, where his work focuses on knowledge acquisition, ontology construction, and automatic thesaurus generation. He performed several case studies with major European companies in which requirements for OM systems in industrial practice were investigated and some prototypical implementations were realized. He received a masters in psychology from the University of Heidelberg.

Michael Sintek is a research scientist with DFKI's Knowledge Management Group, where he investigates and implements methods and tools for object-oriented and logics-based knowledge representation and reasoning for OMs. He received a masters in computer science from the Kaiserslautern University.