Knowledge Processes and Ontologies

Steffen Staab[1,2], Hans-Peter Schnurr[2], Rudi Studer[1,2], York Sure[1]
[1]Institute AIFB, University of Karlsruhe, D-76128 Karlsruhe, Germany
http://www.aifb.uni-karlsruhe.de/WBS
[2]Ontoprise GmbH, Haid-und-Neu Straße 7, 76131 Karlsruhe, Germany
http://www.ontoprise.de
mailto:{staab,studer,sure}@aifb.uni-karlsruhe.de
mailto:{staab,studer,schnurr}@ontoprise.de
tel. +49-721-6084751, fax +49-721-6086580

November 29, 2000

Abstract

Technology for knowledge management has so far focused on the management of knowledge containers. We present an approach that is oriented towards managing knowledge contents instead by identifying knowledge items at various levels of formality. This is done by providing various types of meta data that are tied to ontologies for conceptual interlinkage. Knowledge items are embedded into knowledge processes, which are supported by a suite of ontology-based tools. In order to handle this sort of rich knowledge process, we introduce a meta process that puts special emphasis on constructing and maintaining the ontology when introducing knowledge management systems. In order to elucidate our approach, we describe a case study about the building of CHAR, the Corporate History AnalyzeR.

Keywords: Ontology, Knowledge Mangement, Knowledge Process

1 Introduction

In recent years Knowledge Management (KM) has become an important success factor for enterprises. Increasing product complexity, globalization, virtual organizations or customer orientation are developments that ask for a more thorough and systematic management of knowledge — within an enterprise and between several cooperating enterprises. Obviously, KM is a major issue for human resource management, enterprise organization and enterprise culture — nevertheless, information technology (IT) plays the crucial enabler for many aspects of KM. As a consequence, KM is an inherently interdisciplinary subject.

IT-supported KM solutions are built around some kind of organizational memory that integrates informal, semi-formal and formal knowledge in order to facilitate its access, sharing and reuse by members of the organization(s) for solving their individual or collective tasks [2]. In such a context, knowledge has to be modeled, appropriately structured and interlinked for supporting its flexible integration and its personalized presentation to the consumer. Ontologies have shown to be the right answer to these structuring and modeling problems by providing a formal conceptualization of a particular domain that is shared by a group of people in an organization [5].

There exist various proposals for methodologies that support the systematic introduction of KM solutions into enterprises. One of the most prominent methodologies is CommonKADS that puts emphasis on an early feasibility study as well as on constructing several models that capture different kinds of knowledge needed for realizing a KM solution [8]. Typically, these methodologies conflate two processes that should be kept separate in order to achieve a clear identification of issues: whereas the first process addresses aspects of introducing a new KM solution into an enterprise as well as maintaining it (the so-called ”Knowledge Meta Process”), the second process addresses the handling of the already set-up KM solution (the so-called ”Knowledge Process”) (see Figure 1). E.g. in the approach described in [7], one can see the mixture of aspects from the different roles that, e.g., “knowledge identification” and “knowledge creation” play. The knowledge meta process would certainly have its focus on knowledge identification and the knowledge process would rather stress knowledge creation.

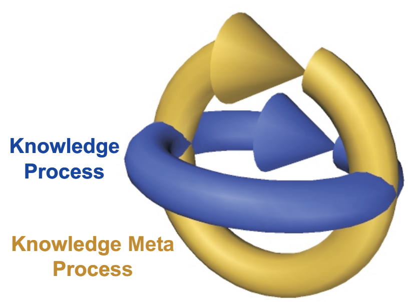


Figure 1: Two orthogonal processes with feedback loops

We start with discussing the interaction between the two types of processes in Section 2 of the paper. Here, we contrast the more document-oriented view with the more knowledge item-oriented view and indicate the implications for the involved knowledge (meta) processes. The notion of knowledge items is further elaborated in Section 3. Special emphasis is put on the various formality levels of these knowledge items and the role meta data play in this context. In Section 4, we elaborate in detail the second type of process, the Knowledge Process. We set up an ontology-based approach and identify five subprocesses that constitute the Knowledge Process: knowledge creation, knowledge import, knowledge capture, knowledge retrieval and access, and knowledge use. As we will see, ontologies constitute the glue to tie together all these knowledge subprocesses. Furthermore, ontologies open the way to move on from a document-oriented view on KM to a contents-oriented view of KM, where knowledge items are interlinked, combined and used within inferencing processes.

In section 5 we define a methodology for introducing an ontology-based KM solution into enterprises, i.e. we address the first type of process mentioned above, the Knowledge Meta Process. The methodology extends and improves the CommonKADS methodology [8] by introducing — among others — specific guidelines for developing and maintaining the respective ontology. Special emphasis is put on a stepwise construction and evaluation of the ontology.

We illustrate our concepts and methods by outlining the development and usage process of a concrete KM solution, i.e. the Corporate History AnalyzeR (CHAR) (cf. Section 6). CHAR is a valuable example for the gain in functionality and usability that results from our ontology-based KM approach. Flexible view generation, integrated results and knowledge inferred as part of KM queries constitute functionalities that can be achieved by such an integrated approach. Our approach shows that one can achieve a clear identification and handling of the different subprocesses that drive the development and usage of KM applications. All these subprocesses are supported by appropriate tools that are tied together by the ontology infrastructure (see for more details [9]).

Knowledge Management is a process which is not only governed by IT. Hence, one needs to keep the balance between human problem solving and automated IT solutions. This balancing distinguishes KM from traditional knowledge-based systems. Nevertheless, the extensive knowledge modeling tasks that are inherent in ontology-based KM approaches support Alun Preece's saying "Every KM project needs a knowledge engineer".

2 Knowledge Items in Knowledge Processes and Meta Processes

The core concern of IT-supported knowledge management is the computer-assisted capitalization of knowledge [1]. Because information technology may only deal with digital, preferably highly-structured, knowledge the typical KM approach distinguishes between computer-based encoding in an organizational memory and direct transfer that is done by humans. Sticking to what is almost readily available, KM systems have tended to serve *either* the needs of easy access to documents

Table 1: Approaching the Knowledge Process — two extreme positions

	Document focus	Knowledge item focus
1.	Find out what the core knowledge needs are	
2.	Find out which *business documents* and *databases* deal with these knowledge needs	Find out which *knowledge items* deal with these knowledge needs
3.	Build an *Infrastructure* for your organizational memory system	Organize the *knowledge processes* to allow for creation, handling, and process support of and around knowledge items
4.	Re-organize *processes* to deal with creation and distribution of knowledge	Build an *Infrastructure* for your organizational memory system

(*e.g.*, building on groupware, etc.) *or* the encoding of knowledge that facilitates the direct transfer of knowledge by humans (*e.g.*, by people yellow pages, skill databases, etc.).

Introducing KM to a company (*i.e.*, moving along the Knowledge Meta Process in "the yellow circle" in Figure 1), a very simple, pragmatic approach has typically been pursued, which however meant that only the low hanging fruits were picked. This approach is summarized in the left column of Table 1. What appears preeminent in this approach is the focus on the handling of documents (steps 2 and 3) and the existing, but minor role of the appendix "process". In spite of its immediate successes, this approach shows several disadvantages. In particular, it often leads to the consequence that the Knowledge Process steps ("the blue circle") of creation, import, capturing, retrieving/accessing, and using are only very loosely connected, if at all (*cf.* Figure 2). The underlying reason is that for each of these steps different types of business documents play a major role, which makes "knowledge re-use" — and not only knowledge re-finding — extremely difficult.

Focus on the Knowledge Process. In fact, there are several commercially sucessful solutions that focus on the process thought of handling knowledge, but due to the evolving nature of business solutions they have not quite reached the level that allows for the immediate connection of knowledge with process steps and the overarching combination of knowledge items across several process steps.

From knowledge containers to knowledge items. Subsequently, we show how *domain ontologies* may act as the glue between knowledge items, bridging between different Knowledge Process steps. Thereby, we argue for a refocus on the Knowledge Process and its core items, which need not be documents! This shift becomes visible in the second column of Table 1, which positions *knowledge items* and *Knowledge Processes* in the center of consideration.

The reader may note that we contrast two rather extreme positions in Table 1. As becomes obvious in recent research papers, current knowledge management research tends to move away from the document focus to a focus on knowledge items and processes [1, 10]. While for a multitude of settings we still see the necessity for the document-oriented view, we argue for a more constructivist view of the Knowledge Processes. In particular, we believe that the exuberant exchange and trading of knowledge within and across organizations still has to begin — and that it needs a knowledge item-oriented view such as we plead for.

3 Knowledge Items

Relevant knowledge items appear in a multitude of different document formats: text documents, spreadsheets, presentation slides, database entries, web pages, construction drawings, or e-mail, to name but a few. The challenge that one must cope with lies in the appropriate digestion of the knowledge, *e.g.*, by "simple" reuse, or by aggregation, combination, condensation, abstraction, and by derivation of new knowledge from aggregations. Following only the lines of traditional document management, IT support for knowledge management cannot take advantage of the contents of the business documents, but only of its explicit or implicit classification. At the other extreme of this spectrum, there are expert systems that structure and codify all the knowledge that is in the system. Though such an approach may sometimes be appropriate, it is certainly not the way to follow in the typical knowledge management scenario, where not everything can be codified, a lot of knowledge is created sporadically, and the worth of knowledge re-use is only shown over time and not necessarily obvious from the very beginning.

Hence, one must search for the adequate balance between reuse, level of formality, and costs to codify knowledge. For instance, certain helpdesk scenarios imply long term use of extremely well-defined knowledge items [4]. Then it may be worth to codify extensively and to spend some considerable amount of time and money on coding. On the other hand, a sporadic discussion is typically not worth coding at all, since it lives on the spur of the moment and often is negligible and, hence, not reusable after some short time.

As a way to balance these conflicting needs and to flexibly manage various degrees of encoded knowledge, we advertise the use of various notions of *meta data*. The different notions of the term “meta data”, *i.e.* data about data, may be classified at least into the following categories:

1. Data describing other data. We may again divide this category into two orthogonal dimensions.

   (a) The one dimension concerns the formality of this data. Meta data may range from very informal descriptions of documents, *e.g.* free text summaries of books, up to very formal descriptions, such as ontology-based annotation of document contents.

   (b) The second dimension concerns the containment of the meta data. Parts of meta data may be internal to the data that is described, *e.g.* the author tag inside of HTML documents, while others may be stored completely independently from the document they describe, such as a bibliography database that classifies the documents it refers to, but does not contain them.

2. The second major connotation of meta data is data that describes the structure of data. For our purpose, one might refer to this notion by the term “meta meta data”, because we describe the structure of meta data. Also, in our context this notion boils down to an ontology that formally describes the domain of the KM application, possibly including parts of the organization and the information structures (*cf.* [1]). The ontology allows to combine meta data from different parts of the Knowledge Process and data proper that adhere to the ontology description.

Meta data in its first connotation fulfills a double purpose. It condenses and codifies knowledge for reuse in other steps of the KM process by being connected through mutual relationships and the ontology (the meta meta data). Furthermore, it may link knowledge items of various degrees of formality together, thus allowing a sliding balance between depth of coding and costs.

4 Knowledge Process

Once a KM system is fully implemented in an organization, knowledge processes essentially circle around the following steps (*cf.* Figure 2).

- *Knowledge creation* and/or *import*, *i.e.* contents need to be created or converted such that they fit the conventions of the company, *e.g.* to the the knowledge management infrastructure of the organization;
- then knowledge items have to be *captured* in order to elucidate importance or interlinkage, *e.g.* the linkage to conventionalized vocabulary of the company;
- *retrieval of* and *access to knowledge* satisfies the "simple" requests for knowledge by the knowledge worker;
- typically, however, the knowledge worker will not only recall knowledge items, but she will process it for further *use* in her context.

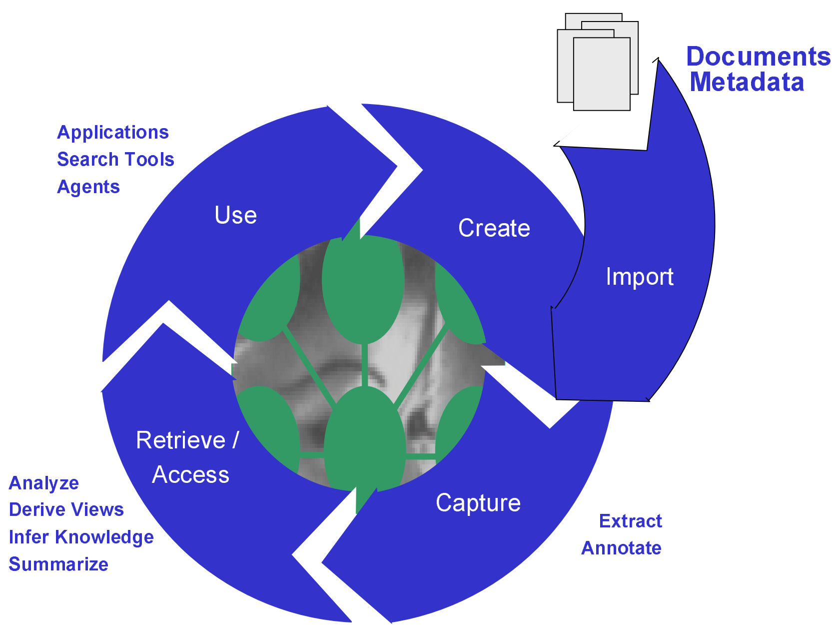


Figure 2: The Knowledge Process

Knowledge Creation. Creation of computer-accessible knowledge proceeds between the two extremes of very formal and very informal knowledge. What is often overlooked is that comparatively deep coding can often be done without requiring any extra efforts. Business documents, in general, are not arbitrarily changing knowledge containers, but reasonably often come with some inherent structure, part of which is often required by quality management and, *e.g.*, engineering requirements. Thus, in a contribution to [10] we have proposed to embed the structure of knowledge items into *document templates*, which are then filled on the fly by doing daily work. The granularity of this knowledge then lies in the middle between the extremes of coarse representations of business documents only and an — for the purpose of KM — overly fine one, such as found in expert systems. Thus, one finds several degrees of formality between the two extremes of very formal and very informal knowledge. We compare some of them in Table 2.

Table 2: Degrees of formal and informal knowledge

Degree	Model	Interface	Example
Thoroughly Formal	Relational	Form Interface	Database interface
Formal	Content-structured document	Tight XML Structure	XML-EDI
Partially Formal	Document template	Loose XML Structure	Investment recommendation template (cf. Table 3)
Informal	Free text	No predefined structure	ASCII text file

In this comparison, we use the term “content-structured documents” to refer to XML structures that are tightly (sometimes explicitly, sometimes implicitly) linked to a domain model. For instance, XML-EDI documents come with a predefined structure alluding to a standard framework for exchanging data, such as invoices, healthcare claims, or project status reports. By document templates we refer to similar structures, which however come with a larger degree of freedom, including large chunks of informal knowledge items (*e.g.*, *cf.* Table 3). One may note that these different degrees of formality are often combined, *e.g.* unstructured documents may have attached a form for adding Dublin Core meta data.

Careful analysis of the knowledge items in use allows for the possibility to add formal knowledge parts into the process of creating these documents, thus pushing the degree of formality slightly upwards without endangering the overall usage of the system, which could be incurred by an expert systems-like approach to Knowledge Management.

Table 3: Filled Investment Recommendation Template.

<investmentrecommendation>			
<author>		Henrik Oppermann	</author>
<plandate>		October 18th, 2003	</plandate>
<interviewpartners>			
	<name>	York Sure	</name>
	<name>	Hans-Peter Schnurr	</name>
	<name>	Steffen Staab	</name>
</interviewpartners>			
<recommend>		strong buy	</recommend>
<details>			
	<peergroup>	···	</peergroup>
	< ··· >	···	< ··· >
</details>			
</investmentrecommendation>			

Knowledge Import. For many KM purposes the import of knowledge items into the KM system of the organization has the same or more importance than their creation within the organization. The overall situation is akin to data warehousing — only that the input structures are more varying and the target structures are much richer and more complex than this is the case for the standard data warehouse.

For imported knowledge accurate access to relevant items plays an even more important role than for “home-made” knowledge. The reason is that for home-made knowledge items, people may act as a backup index. This is not the case for recently imported knowledge that no one has seen yet. In fact, access studies to KM systems have shown that OM parts that cover imported knowledge are less heavily exploited than those covering home-grown ones [6] — though it seems implausible that they would contain less useful contents.

Knowledge Capture. Once that knowledge items have been created, but not yet, or only incompletely, captured apart from their context, *e.g.* from their database entries or their business document containers, the next process step is the capturing of their essential contents. Besides of common indexing and abstracting techniques known from library science, we provide means to capture document excerpts as well as interlinkage between excerpts by our tool, OntoAnnotate (*cf.* Figure 3).

OntoAnnotate allows to create objects and describe them with their attributes and relations exploiting knowledge that is found on web pages, in spreadsheets, or in text documents. Thus, in the example we capture the facts that the company M.A. Hanna sells to GE its Shape Distribution Business, link them to the excerpts from which they origin and note who has done the annotation.

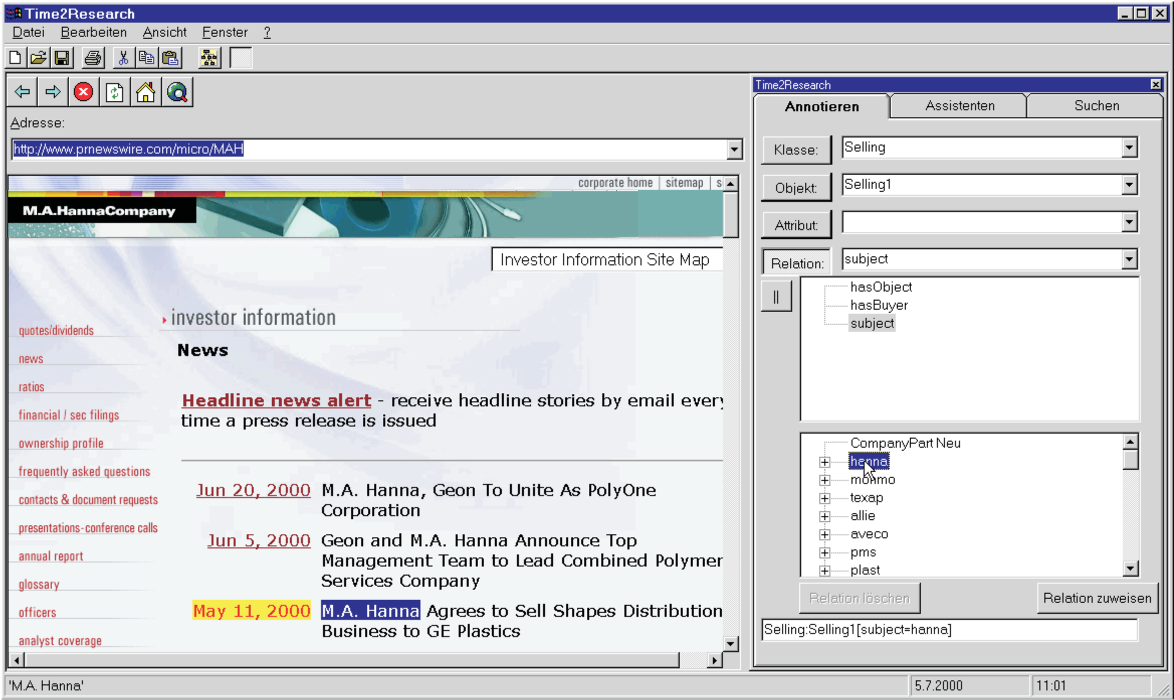


Figure 3: Capturing Knowledge about a Selling

By this annotation process meta data are created that conform to the ontology and, hence, can be aligned with related information to yield analyses and derivations such as described in our case study in Section 6. The origins of the meta data may be used to validate the overall information.

Knowledge Retrieval and Access. Large parts of knowledge retrieval and access from an ontology-based organizational memory are performed through conventional GUI, exploiting means like information retrieval or taxonomy-enhanced database views. In addition, one may use the ontology to derive further views. In particular, we exploit the ontology for navigation purposes (like Yahoo). Thus, knowledge workers may explore what is in the organizational memory without being required to ask a particular question — which is often a hard task for newbies. Also, the ontology allows to derive additional links and descriptions, *e.g.* the ontology allows to derive state descriptions for points in time for which no explicit data exists, or it provides new hyperlinks that are not given explicitly. Thus, we may complete views without requiring that all information is given. An actual example for the latter is given in detail in Section 6.

Knowledge Use. Knowledge use deals with the most intricate points of knowledge management. It is the part that is most often neglected, because many KM systems assume that once some relevant document is found everything is done. Eventually, however, the way that knowledge from the organizational memory is used is quite involved. Therefore topics like proactive access, personalization, and, in particular, tight integration with subsequent applications play a crucial role for the effective re-use of knowledge. Very often it is not even the knowledge itself which is of most interest, but the derivations that can be made from the knowledge. For instance, in our case study we will see that no single knowledge items about a company may be relevant to a market analyst, but the overall picture presented by analysis.

In addition, usage data tells a lot about the organizational memory *and* about the organization. For instance, one may analyze which processes, customers and techniques are tied to core figures of the organization.

5 Knowledge Meta Process: Methodology for Ontology-based Knowledge Management

Ontologies aim at capturing domain knowledge in a generic way and provide a commonly agreed understanding of a domain, which may be reused and shared across applications and groups. Ontologies typically consist of definitions of concepts, relations and axioms. Until a few years ago the building of ontologies was done in a rather *ad hoc* fashion. Meanwhile there have been some few, but seminal proposals for guiding the ontology development process (*e.g.* [11]). For instance Guarino & Welty [3] give formal guidelines for constructing a consistent and reusable ontology.

In contrast to these methodologies, which mostly restrict their attention within the ontology itself, our approach focuses on the application-driven development of ontologies. We cover aspects from the early stages of setting up a KM project to the final roll out of the ontology-based KM application. Here, we focus our attention on the development of the ontology. Thereby, we integrate some lessons learned of our practical experiences into the steps to be taken to perform each ontology development activity. The ontology development process is sketched in Figure 4, we will now describe its steps in detail.

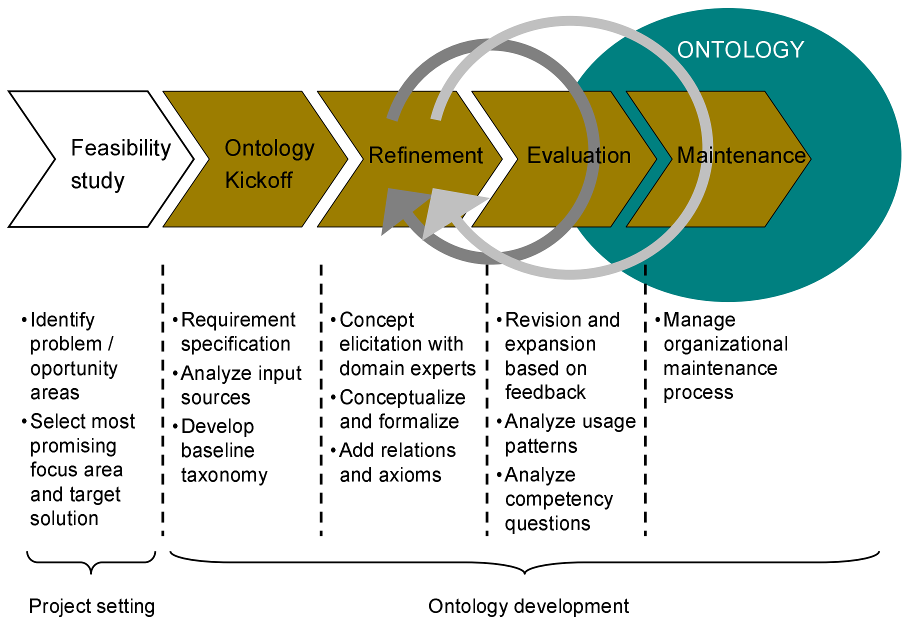


Figure 4: The Knowledge Meta Process

Feasibility study. Any knowledge management system may only function satisfactorily if it is properly integrated into the organization in which it is operational. Many factors other than technology determine success or failure of such a system. To analyze these factors, we must initially perform a feasibility study to first identify problem/opportunity areas and potential solutions, and second, to put them into a wider organizational perspective. The feasibility study serves as a decision support for economical and technical project feasibility, in order to select the most promising focus area and target solution. An approach for carrying out a feasibility study is described by the CommonKADS methodology (*cf.* [8]). It should be carried out before actually developing ontologies and serves as a basis for the kickoff phase.

Kickoff phase for Ontology Development. The output product of the kickoff phase is an ontology requirements specification document (*cf.* Table 4) describing what an ontology should support and sketching the planned area of the ontology application. It should also guide an ontology engineer to decide about inclusion, exclusion and the hierarchical structure of concepts in the ontology. In this early stage one should look for already developed and potentially reusable ontologies. In summary, it should clearly describe the following information:

Table 4: Ontology Requirements Specification Document (ORSD)

Domain	**:**	**Business strategy in the chemical industry**
Date	**:**	2000/11/26
Ontology Engineer	**:**	T. Model

Goal of the ontology:

- Tracking and analyzing corporate business histories

Domain and Scope:

- Merger & acquisition, restructurings, management changes and other strategic activities in the chemical industry

Supported Applications:

- Web-based Corporate History Analyzer

Knowledge Sources:

- Research analysts (domain experts)
- Related websites (company homepages, chemical industry networks)
- Newspaper articles
- Ad hoc news tickers

Users and Use Cases:

- Users: Research analysts, strategic consultants
- Use Case 1: Track strategies of specific companies
- Use Case 2: Analyze strategic moves of competitors

Competency Questions:

- Attached Competency Questionnaire

Potentially reusable ontologies:

- not known

1. Goal of the ontology
2. Domain and scope
3. Applications supported by the ontology
4. Knowledge sources (*e.g.* domain experts, organization charts, business plans, dictionairies, index lists, db-schemas etc.)
5. Potential users and usage scenarios

6. Competency questionnaire (*i.e.* an overview of possible queries to the system, indicating the scope and content of the domain ontology, *cf.* Table 5)

7. Potentially reusable ontologies

Table 5: Competency Questionnaire

Domain	**:**	**Business strategy in the chemical industry**
Date	**:**	2000/11/26
Ontology Engineer	**:**	T. Model

CQ Nr.	Competency Question	Concepts	Relation
CQ1	What are the subsidiaries, divisions and locations of company X?	company, subsidiary, division, location	company *has* subsidiary company *has* division company *has* location
CQ2	Which companies acquired company X?	company, acquisition	company *makes* acquisition acqusition *has* buyer acqusition *has* seller
CQ4	Which companies merged in 1990 in the rubber industry?	company, merger, year, industry	company *makes* merger company *isPartOf* industry merger *happensIn* year
CQ5	Who is CEO of company X?	CEO, company,	company *has* CEO
CQ6	Which activity of company X leads to operation in region Y?	activity, company, operation, region	company *performs* activity activity *leadsTo* operation operation *takesPlaceIn* region
CQ7	Is there any regional expansion of company X due to the acquisition of company Y?	expansion, company, region, acquisition	company makes expansion company makes acquisition expansion takesPlaceIn region
CQ8	...		

Refinement phase. The goal of the refinement phase is to produce a mature and application-oriented target ontology according to the specification given by the kickoff phase. This phase is divided into different subphases:

1. The gathering of an informal "baseline taxonomy" containing relevant concepts given during the kickoff phase.

2. A knowledge elicitation process with domain experts based on the initial input from the baseline taxonomy to develop a "seed ontology" containing relevant concepts, relations between them and axioms on top. The seed ontology is usually expressed at an epistemological level.

3. A conceptualization and formalization phase to transfer the seed ontology into the "target ontology" expressed in formal representation languages like F-Logic, OIL or DAML-ONT.

The usage of potentially reusable ontologies (identified during the kickoff phase) may improve the speed and quality of the development during the whole refinement phase. These ontologies might *e.g.* give useful hints for modeling decisions.

Evaluation phase. The evaluation phase serves as a proof for the usefulness of developed ontologies and their associated software environment. In a first step, the ontology engineer checks, whether the target ontology suffices the ontology requirements specification document and whether the ontology supports or "answers" the competency questions analyzed in the kickoff phase of the project. In a second step, the ontology is tested in the target application environment. Feedback from beta users may be a valuable input for further refinement of the ontology. A valuable input may be as well the usage patterns of the ontology. The prototype system has to track the ways users navigate or search for concepts and relations. With such an "ontology log file analysis" we may trace what areas of the ontology are often "used" and others which were not navigated. This phase is closely linked to the refinement phase and an ontology engineer may need to perform several cycles until the target ontology reaches the envisaged level — the roll out of the target ontology finishes the evaluation phase.

Maintenance phase. In the real world things are changing — and so do the specifications for ontologies. To reflect these changes ontologies have to be maintained frequently like other parts of software, too. We stress that the maintenance of ontologies is primarily an organizational process. There must be strict rules for the update-delete-insert processes within ontologies. We recommend that the ontology engineer gathers changes to the ontology and initiates the switch-over to a new version of the ontology after thoroughly testing possible effects to the application, *viz.* performing additional cyclic refinement and evaluation phases. Similar to the refinement phase, feedback from users may be a valuable input for identifying the changes needed. Maintenance should accompany ontologies as long as they are on duty.

In the following section we bring this general methodology to life in our case study CHAR.

6 Case Study: the Corporate History AnalyseR — CHAR

6.1 Knowledge Meta Process instantiated: The development of CHAR

Feasibility study. The active tracking and management of knowledge relevant to one's business is a major task for knowledge-intensive companies. While the correct analysis of market situations and competitors are critical requirements for success, the failure to provide adequate knowledge about one's business environment may incur large losses. The trouble is that management and professionals have a hard time gathering information, analyzing it, and performing their operational work. Here comes corporate research into play. The task of the corporate researcher is the tracking of relevant knowledge in the outside business setting and the communication of important knowledge to stakeholders within the company. Nowadays, the market analysts, consultants and inhouse market research departments try to track the activities of their industry with traditional methods. Newspaper articles, online databases and annual company reports are analyzed and competitors web pages are thoroughly tracked. The results are presented to the management, published in reports and distributed. There exist several problems in this research process:

- Information archives are document-based. For a collective gathering of facts, this view is too coarse.
- Typical document management systems rely almost exclusively on information retrieval techniques, which are too inaccurate given the task at hand.
- Implications may only be made transparent if background knowledge is used. *E.g.*, if one company sells one of its units, the implication is that this unit may no longer be the sole supplier for the company as it was before. Nowadays, systems are typically not supporting such background knowledge.
- Different people may contribute similar knowledge, but they may need different views onto the same basic piece of information.

A KM system that covers such knowledge about the outside world should, (i), support the collective gathering of information on the level of facts rather than documents, (ii), integrate the gathering task smoothly into the common research process, (iii), allow to intelligently combine facts, (iv), check new facts against the available background knowledge, (v), allow a multiple view access to the knowledge via a single entry portal, and, (vi), allow to route derived facts back into the common workplace environment. With these aims in mind, we have developed the ontology-based application CHAR, the Corporate History AnalyzerR.

Kickoff phase for Ontology development. Starting from these requirements, the question was how to bring the required conceptual structures and reasoning capabilities into action following the further steps of our Knowledge Meta Process shown in Figure 4. User requirements were collected in the requirements specification phase during structured interviews with corporate research analysts. The Ontology Requirements Specifications Document shown in Table 4 has been an output of this first phase. There, the ontology had to be specified as described in Section 5. In a next step, the ontology had to be developed in detail. Therefore, the domain experts were asked, what questions they would expect to be answered by a system supporting their corporate research work. These competency questions allowed to find the most important concepts and relations between concepts. One result of analyzing the questionnaires has been that the system should deliver answers about the acquisitions, mergers and restructuring of companies over specific time periods. The example Competency Questionaire in Table 5 lists competency questions CQ1 to CQ5 that were found during the ontology kickoff. In addition, the need for a clarification of organizational "wording" was recognized, *e.g.* "Is a business unit the same as a division or a department?"

Refinement phase. First in the refinement phase, the concepts were brought into a taxonomic hierarchy by the knowledge engineer. Figure 5 shows some part of the taxonomic structure.

Second, the taxonomy was shown to the domain experts to gather additional concepts and relevant attributes and relations between the concepts. This work was supported by our ontology engineering tool OntoEdit (*cf.* [9]), which allows the epistemological modeling of the ontology and the formalization in different representation languages, like *e.g.* F-Logic, OIL and DAML-ONT. One particularity of the corporate history domain was that time had to be modeled. There are many objects which have a life time. For instance a company has a start time when it is founded and there is an end time when this company is acquired by another company, is merged with another company, or goes bankrupt. So duration had to be modeled. Likewise, an "Acquisition" is an "Event" which occurs at a specific point in time, requiring a model for time points. One of the key characteristics of CHAR is that the

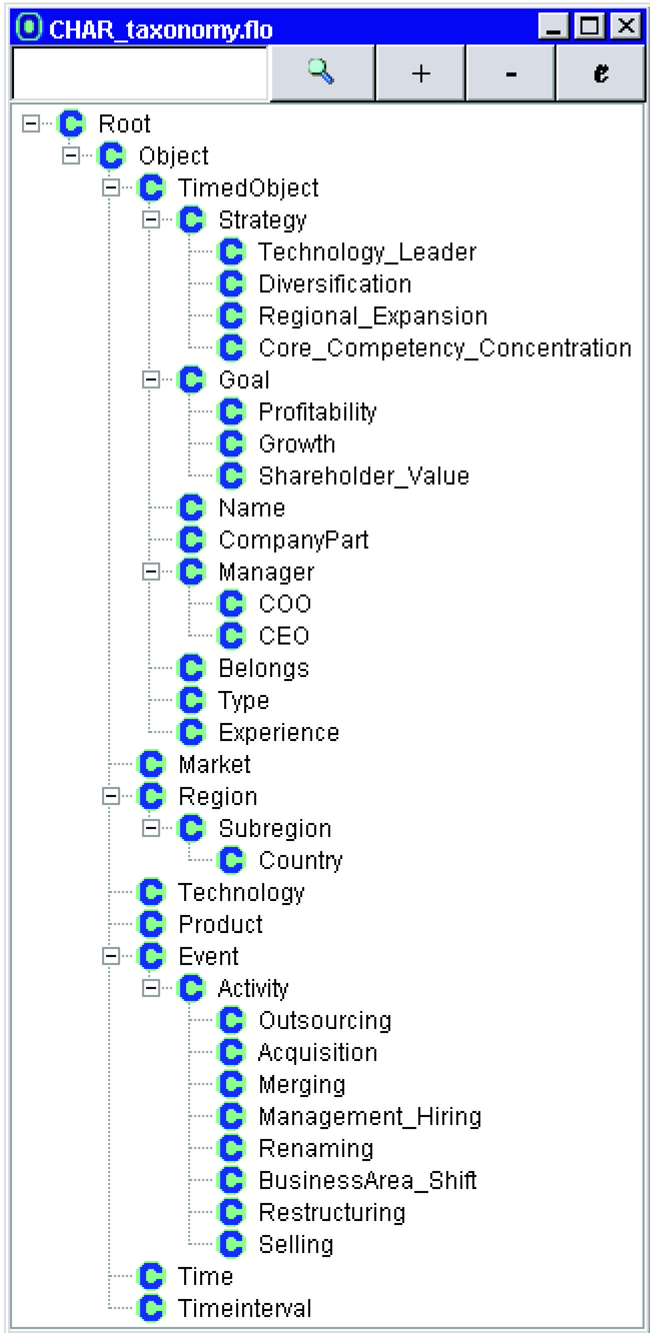


Figure 5: CHAR taxonomy

user provides facts to the system about actions like acquisitions and mergers and the system inferes the consequences. For this purpose rules had to be modeled for all possible activities which incur such consequences. For instance, “If two companies are merged, a new company with a new name is created, the old companies ceases to exist, they are from now on subsidiaries of the new company” or

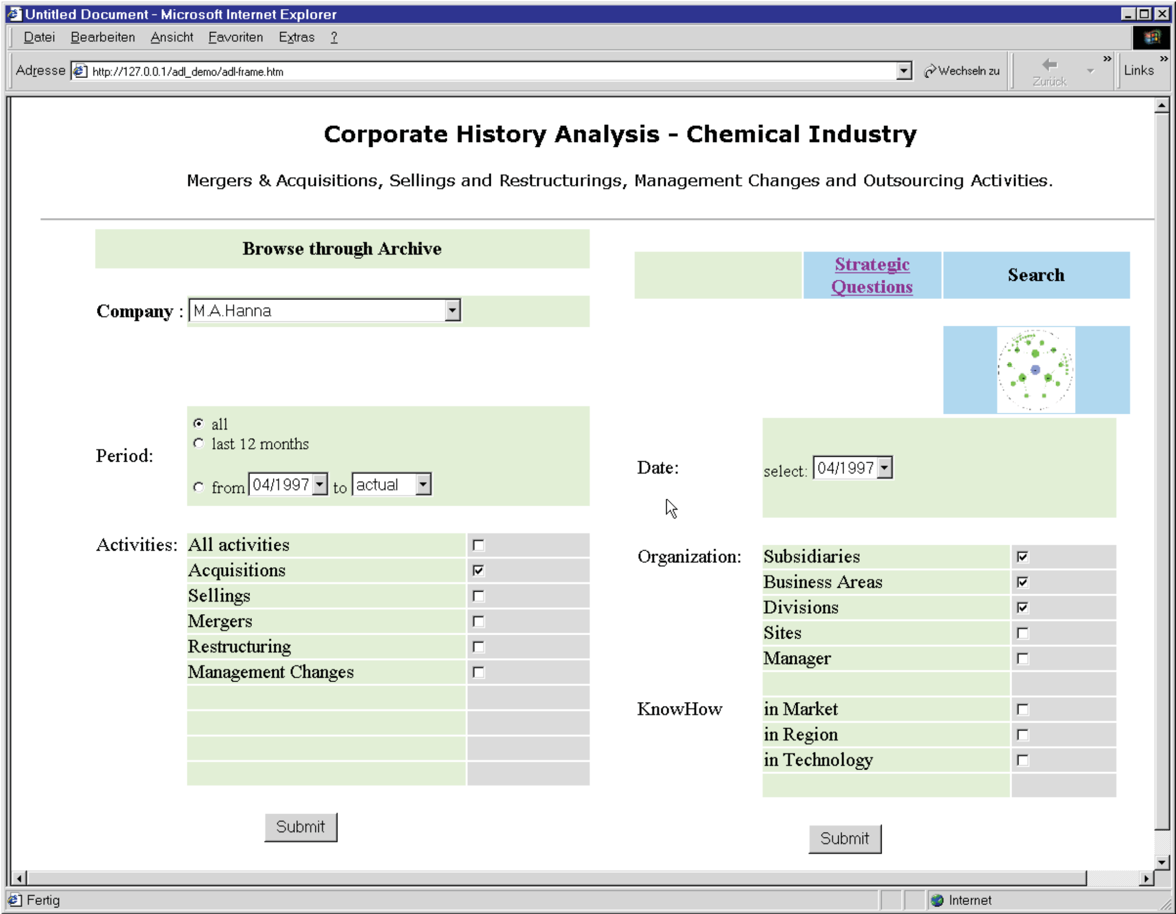


Figure 6: CHAR

"If a division is outsourced it becomes part of a new company or it becomes a company on its own."

For our concrete example, CHAR, a Web interface had to be developed that used the ontology for querying and augmenting the knowledge repository (*cf.* Figure 6). There, we needed to perform a query development step to formalize the views and competency questions described earlier. This development step depended mostly on the actual application setting. It was independent from the ontology construction, on which the paper reports.

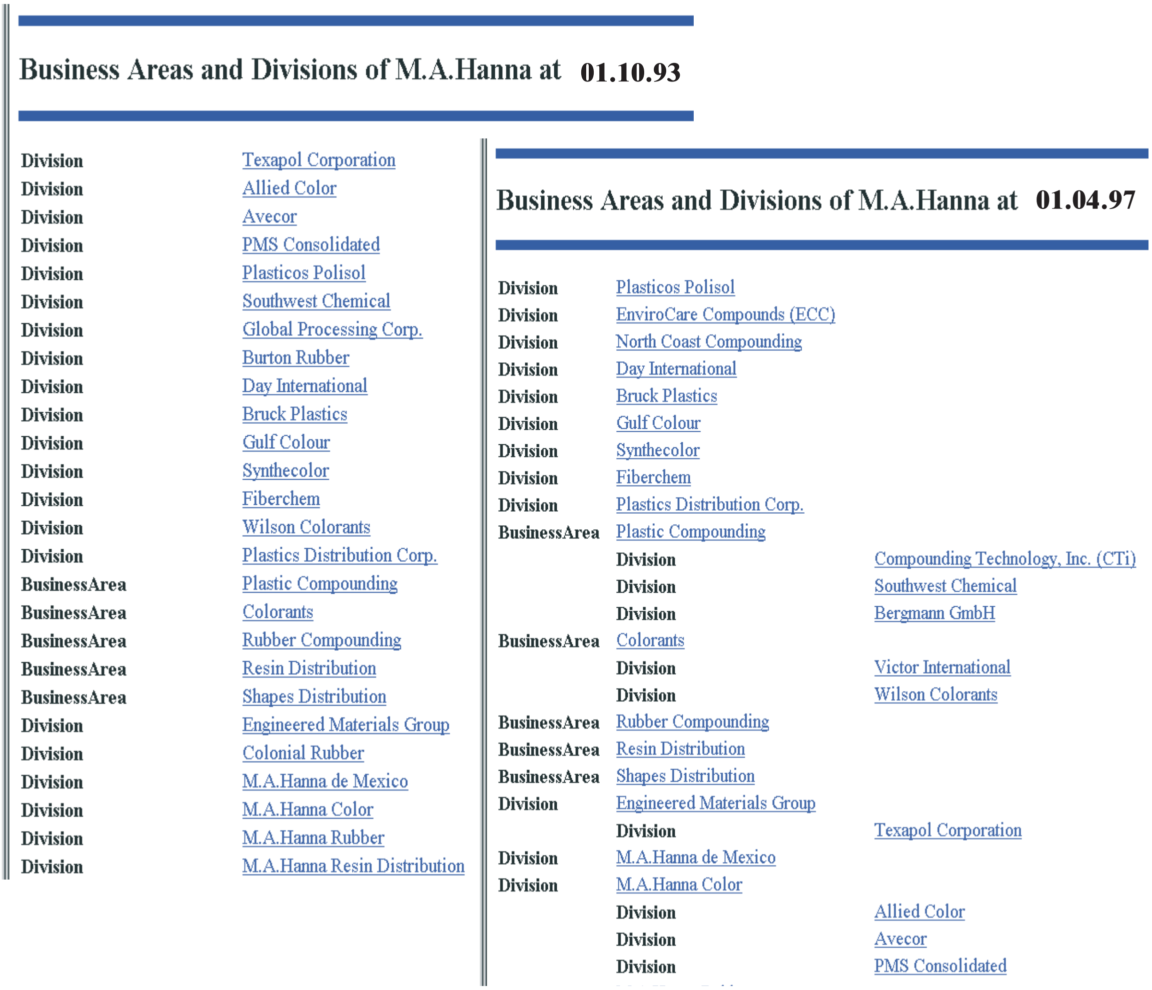
Evaluation phase. In the evaluation and testing phase the usability of the ontology was investigated. With regard to the CHAR ontology we found that the competency questions CQ2 to CQ5 (Table 5) could be handled successfully by the ontology. However, it had been envisaged that complex knowledge content could be queried and depicted on one screen, such as company purchases, regions of previous activities and regions of activities of the purchased company. Beta testing the system the research analyst found that singleton knowledge items were spread over different screens of the application. Thus, they were very hard to combine by the research analyst in order to answer strategic questions. Hence, he wanted better support from the ontology. Through the discussion we came up with additional competency questions (CQ6 and following) that should be handled by the ontology. Hence, corresponding axioms were added to the ontology and a new query window for strategic questions was introduced to the application.

Maintenance phase. We are currently in a maintenance phase, where through a shift in the goal orientation of the application, we face requirements for considerable extension of the ontology and the application. Besides of corporate histories, the system is now required to support the extended tracking of market circumstances, including, *e.g.*, a more detailed view of available technology and better comparisons with peer groups. Major problems that we now must face are, first, all the documentation and versioning problems that have so far been neglected in practically all ontology development environments. Also, there are ontology parts for which our development methodology had not yet been followed. There, developers who were new to the project often need to reverse-engineer some ontology structures in order to find answers about their actual usage. For the future, we expect that comprehensive use of the proposed methodology will mostly prevent these problems.

6.2 Knowledge Process instantiated: Usage of CHAR

CHAR, the Corporate History AnalyzeR should allow many people to contribute factual knowledge in a way that is embedded into their common work process and that is organized around a semantic background. On the other hand, to deliver answers, it should provide multiple views onto the same knowledge base for different time frames, for different regional foci, for varying intra-organizational structures and for different strategic questions, to name but a few.

Providing Knowledge. The process of providing new facts into the knowledge warehouse should be as easy and as smoothly integrated into the common working tasks as possible. For this reason we offer various modes of contributing knowledge. First, one may enter information through a form-based interface. Second, when the information that is to be provided is produced during the writing of documentations or reports one may use a template approach in order to generate knowledge by writing these documents. Third, one may use wrapper mechanisms in order to provide data from



Business Areas and Divisions of M.A.Hanna at 01.10.93

Division	Texapol Corporation
Division	Allied Color
Division	Avecor
Division	PMS Consolidated
Division	Plasticos Polisol
Division	Southwest Chemical
Division	Global Processing Corp.
Division	Burton Rubber
Division	Day International
Division	Bruck Plastics
Division	Gulf Colour
Division	Synthecolor
Division	Fiberchem
Division	Wilson Colorants
Division	Plastics Distribution Corp.
BusinessArea	Plastic Compounding
BusinessArea	Colorants
BusinessArea	Rubber Compounding
BusinessArea	Resin Distribution
BusinessArea	Shapes Distribution
Division	Engineered Materials Group
Division	Colonial Rubber
Division	M.A.Hanna de Mexico
Division	M.A.Hanna Color
Division	M.A.Hanna Rubber
Division	M.A.Hanna Resin Distribution

Business Areas and Divisions of M.A.Hanna at 01.04.97

Division	Plasticos Polisol	
Division	EnviroCare Compounds (ECC)	
Division	North Coast Compounding	
Division	Day International	
Division	Bruck Plastics	
Division	Gulf Colour	
Division	Synthecolor	
Division	Fiberchem	
Division	Plastics Distribution Corp.	
BusinessArea	Plastic Compounding	
	Division	Compounding Technology, Inc. (CTi)
	Division	Southwest Chemical
	Division	Bergmann GmbH
BusinessArea	Colorants	
	Division	Victor International
	Division	Wilson Colorants
BusinessArea	Rubber Compounding	
BusinessArea	Resin Distribution	
BusinessArea	Shapes Distribution	
Division	Engineered Materials Group	
	Division	Texapol Corporation
Division	M.A.Hanna de Mexico	
Division	M.A.Hanna Color	
	Division	Allied Color
	Division	Avecor
	Division	PMS Consolidated

Figure 7: Comparing two Organizational Structures

tables and lists on the web. Fourth – and most important for CHAR, one may use our annotation tool in order to add meta data to data given in documents. A snapshot of the annotation tool noting some action about M.A.Hanna is shown in Figure 3: The user reads or works with documents using a text or spreadsheet processing tool or an internet browser. When he detects some relevant change being described in the document and this change might become relevant, he highlights the word or phrase and uses the annotation tool to select the type of the phrase (*e.g.* "M.A.Hanna is a company") and its relation to other material (*e.g.* "Hanna sells Shapes Distribution Business to GE Plastics on May 11, 2000"). The document, these facts and meta data about the annotater, the time of annotation, etc. is all stored in the back-end knowledge repository.

Querying for Knowledge. The query interface of the Corporate History Analyzer has been developed in order to deal with organizational and strategic questions that depend on spatio-temporal constraints. It renders views that may be seen on a common web browser. Actually, they look just like common web sites with some choices for selection. These selections are also controlled by the knowledge of the backend system, *e.g.* one may select companies that are known to exist in the knowledge repository. Figure 6 depicts the main views that are offered by CHAR, *viz.* Activities, Organization, Know How, Strategic Questions and General Query Possibilities (indicated by "Search"). The first major category of queries relevant to the corporate history is about organizational structures and the activities that change organizational structures. For instance, the view of "Acquisitions of M.A.Hanna" returns all its purchases and corresponding views are offered for Selling, Mergers, Restructuring and Management Changes (Figure 7). What is interesting to note at this point is that it is rather difficult to get a clear picture of what is really happening with M.A.Hanna. It is difficult and time-consuming for the human analyst to detect some trend in lists of single knowledge items. Observations become much easier, when different types of knowledge items may be related and contrasted. For instance, Figure 7 depicts two snapshots of the organization of M.A.Hanna that are automatically derived from single activities, like acquisitions and restructurings, and that give the analyst a neat picture of how formerly isolated purchases that M.A.Hanna made before 1994 were more tightly integrated in the company in 1997 (*e.g.* "Compounding Technology" having been reorganized into the Business Area "Plastic Compounding").

Thus, besides sophisticated support that is based on concrete facts and figures and that does not aim at interpretation thereof, CHAR supports strategic questions indicating possible answers to questions about business competitors that cannot be answered definitely, but that rely on some conjectures. For instance, the purchase of a company from abroad may lead to a gain of market share in that area, and thus to a regional diversification.

7 Conclusion

The importance of knowledge processes has been rapidly recognized lately. We here have shown that there are (at least) two dimensions on which these Knowledge Processes need to be analyzed: whereas the first dimension considers the Knowledge Meta Process that is needed to introduce a knowledge management solution into an enterprise, the second dimension addresses the Knowledge

Processes that are performed while running a knowledge management solution. During the Knowledge Meta Process an ontology is developed that acts as the conceptual glue between different steps of the Knowledge Process proper. The development of the Corporate History Analyzer is an intriguing example that highlights the assignments of the different tasks to steps in our overall framework.

Of course, the conceptualization of domains and the related meta data will have to reflect the changing environment of the enterprise. *I.e.* evolving ontologies and, thus, evolving metadata will become an important aspect for maintaining a knowledge management solution. For the future, we therefore envision that the (semi-)automatic analysis of the Knowledge Processes feeds back into the Knowledge Meta Process cycle. For instance, new concepts that come up during the use of the KM solution may be introduced into the ontology — which is maintained following the methodology for KM system development and maintenance. The challenges then lie, *(i)*, in ontology learning and evolution for analysing KM processes and, *(ii)*, in methodological issues for dynamically updating the KM solution. Our framework may then leverage these evolving systems, because it offers a clear distinction of relevant processes and responsibilities.

Acknowledgements. We thank our colleagues and students at AIFB, University of Karlsruhe, and ontoprise GmbH, J. Angele, S. Decker, M. Erdmann, A. Hotho, A. Maedche, M. Maier-Collin, D. Oberle, D. Wenke, without whom the research would not have been possible. The research presented in this paper has been partially funded by EU in the IST-1999-10132 project On-To-Knowledge and by BMBF in the project GETESS (01IN901C0).

References

[1] A. Abecker, A. Bernardi, K. Hinkelmann, O. Kühn, and M. Sintek. Toward a technology for organizational memories. *IEEE Intelligent Systems*, pages 40–48, May / June 1998.

[2] R. Dieng, O. Corby, A. Giboin, and M. Ribiere. Methods and tools for corporate knowledge management. *Int. Journal of Human-Computer Studies*, 51(3):567–598, 1999.

[3] N. Guarino and C. Welty. Identity, unity, and individuality: Towards a formal toolkit for ontological analysis. *Proceedings of ECAI-2000*, August 2000. available from http://www.ladseb.pd.cnr.it/infor/ontology/Papers/OntologyPapers.html.

[4] L. Morgenstern. Inheritance comes of age: Applying nonmonotonic techniques to problems in industry. *Artificial Intelligence*, 103:1–34, 1998.

[5] D. O'Leary. Using AI in knowledge management: Knowledge bases and ontologies. *IEEE Intelligent Systems*, 13(3):34–39, May/June 1998.

[6] D. O'Leary. Knowledge management: An analysis based of knowledge use and reuse. *IEEE Intelligent Systems*, 2001.

[7] G. Probst, K. Romhardt, and S. Raub. *Managing Knowledge*. J. Wiley and Sons, 1999.

[8] G. Schreiber, H. Akkermans, A. Anjewierden, R. de Hoog, N. Shadbolt, W. Van de Velde, and B. Wielinga. *Knowledge Engineering and Management — The CommonKADS Methodology*. The MIT Press, Cambridge, Massachusetts; London, England, 1999.

[9] S. Staab, J. Angele, S. Decker, M. Erdmann, A. Hotho, A. Maedche, H.-P. Schnurr, R. Studer, and Y. Sure. Semantic community web portals. In *WWW9 — Proceedings of the 9th International World Wide Web Conference, Amsterdam, The Netherlands, May, 15-19, 2000*. Elsevier, 2000.

[10] S. Staab and D. O'Leary, editors. *Bringing Knowledge to Business Processes. Papers from 2000 AAAI Spring Symposium*, Technical Report SS-00-03, Menlo Park, CA, 2000. AAAI Press.

[11] M. Uschold and M. Gruninger. Ontologies: Principles, methods and applications. *Knowledge Sharing and Review*, 11(2), June 1996.