INTERNATIONAL DESIGN CONFERENCE - DESIGN 2006
Dubrovnik - Croatia, May 15 - 18, 2006.

DESIGN 2006

INTELLIGENT SEARCH FOR PRODUCT DEVELOPMENT INFORMATION – AN ONTOLOGY-BASED APPROACH

J. Ponn, F. Deubzer and U. Lindemann

Keywords: information and knowledge management, information retrieval, error prevention, ontology-based search

1. Introduction

The active management of information and knowledge is nowadays considered an important means to achieve an enterprise's effectiveness and competitiveness [Studer et al. 1998]. Product development is conceived as an information processing activity respectively an information intensive activity. Thus, adequate strategies, methods and tools for information and knowledge management in the product development process become a vital factor for a company's success. A survey in industry, conducted by the authors of this paper, has shown that personal communication still represents the most favorable source of knowledge and information. However, there are situations where a personal exchange of knowledge is not possible. The knowledge bearer might be temporarily (e.g. business trip, vacation, sickness) or permanently (e.g. retirement, company leaving) not available or simply unknown to the searcher. Here, digital databases within the enterprise or publicly available sources (such as the internet) represent potential alternatives. When searching for information within digital repositories, however, certain problems arise. Finding relevant contents is often a difficult process, consuming time and nerves of the involved actors.

In this paper, two industrial case studies are presented that address these general issues on a concrete level. Although the scenarios differ concerning a variety of aspects (industrial branch, processes in focus, type of information needs etc.) the common general goal was to improve the involved actors' access to relevant information and knowledge via digital databases in specific situations. In both cases, ontologies were considered a possible solution approach. The concrete task in both projects consisted in the investigation of the potential of an ontology-based search. Chapter 2 contains a description of the different scenarios and the analysis of the particular problems that were identified in both case studies. In chapter 3, the motivation for an ontology-based approach is explained and related work is depicted. In chapter 4, the procedure of putting the approach into practice and corresponding results are described (development and evaluation of the ontologies). Chapter 5 discusses major implications of the presented work and concludes the paper.

2. Motivation and Problem

The two case studies were conducted by researchers of the Institute for Product Development at the Technische Universität München (TUM) in the context of industrial projects. The project motivation resulted from the fact, that current access mechanisms to relevant contents in digital databases proved to be inadequate. The development of a concept for a targeted search for relevant contents was therefore articulated as the general goal. As a first step in both projects, a detailed clarification of the task was carried out. Therefor, a general survey in industry (in project 1) and interviews among involved company employees (in both projects) were conducted. Questions of interest concerned: typical search situations, expectations concerning results of the search in these situations, recurring problems concerning the search in a digital database and requirements for a search engine. In addition, a particular use case was developed in each project in order to enable exemplary search processes and obtain concrete results. These measures helped to determine the main problems, to evaluate the given search systems and their weaknesses and to clarify the requirements for a new approach. In the following, details of both projects (scenarios) are presented.

2.1 Scenario 1 – matching customer needs and technical solutions

Project 1 involved several companies within the German drive and automation sector (providers of system solutions and component providers) as well as a provider of an e-market platform. The investigation in the project concentrated on processes in sales departments and product development, where the interconnectivity between customer needs and corresponding high quality technical solutions has to be provided. The focus within the supply chain were client-provider-constellations between OEM producers and system solution providers on the one hand, and system solution providers and component providers on the other hand. From the general survey (41 participants) it became clear, that the most typical search situations included the search for experts, for companies and for reference products. Expected results of a search for directly product related information were (functional) product descriptions, technical data and existing or potential future fields of application. In contrast, detailed drawings or CAD models were mentioned by a much smaller percentage of the respondents, which can be explained by the high number of sales employees (28) as opposed to product development engineers (7) participating in the survey. The rest of the respondents held positions in project management, product management and general management.

The task in the created use case was formulated as follows: "The detection of an object's presence is a sub-function of a system with the overall function of automatic screwing of mobile telephone cases. Before parts of the cases can be screwed together, their presence has to be assured. For the realization of the sub-function, adequate solutions (components) have to be determined." The focus was placed on the search for existing products (solutions that have already been developed in the past) that can potentially be adopted to the new case. Two sub-scenarios were defined, formulated queries and obtained search results were documented and evaluated. Sub-scenario (1a) deals with an internal search and addresses the fact that within a company a lot of information and knowledge on technical solutions to customer needs as well as practical experience concerning applications exist. However, because of the distribution across departments, databases etc., the access to relevant contents in specific situations is often limited. These circumstances turn out to be a problem already, when just considering information and knowledge management within one single enterprise. Sub-scenario (1b) covers the even more complex situation, when information and knowledge on applications as well as on technical solutions are shared and exchanged on platforms, addressing a high number of companies as providers and clients of contents (for instance e-market applications).

Sales employees of the partner companies who had access to company internal sources worked on the task for sub-scenario (1a). They stated that there was a variety of different systems available (e.g. intranet, specific databases, file server, digital catalog on CD). The difficulty in choosing an adequate source, data redundancy and lacking up-to-dateness of the contents represented major problems. In addition, the actual use of some systems within daily work was described as negligible. One reason was e.g. the fact, that login processes represented certain barriers making the systems unpopular. As a major requirement for a search engine, the possibility to combine search topics (such as product category, e.g. "sensors", and a particular product property, e.g. "durability") in order to put a focus on the search was articulated. Sub-scenario (1b) was worked on by a mechanical engineering student using specific publicly available sources (websites of the involved partners in the project) and general internet search engines. The student reported that the search functionalities of the different websites were generally limited. The search by keywords was mainly unsatisfactory, a full text search and search filters were hardly implemented. The structure of the regarded company and e-market websites was characterized as clearly arranged. However, in the search by categories, information on functions and applications was scarce. Instead, information on many concrete products was available. Their suitability for the given task was often hard to determine, though. A long searching procedure arose also due to the fact, that promising products were found in different categories. These were established according to the product portfolio, and therefore differed on each website.

2.2 Scenario 2 – preventing errors in product development

The second scenario originates from a project that was carried out in the automotive industry (German premium car manufacturer). Here, the error prevention within car body design based on operating experience in the product development process was regarded. In this particular case, a database was accessible through the company intranet, containing design knowledge documenting past problems and best practices on certain parts. Relevant information concerned e.g. a particular component failure that has occurred before in another project and should be avoided in the current product development. The structure of the database was organized according to the physical structure (assemblies and parts) of the automobile. The editing process allowed only a few editors to enter data into the system. Entries were made e.g. after design reviews. The overall intention was to gain high quality non-redundant data, which resulted in the establishment of a strict editing process. The database was not a primary working medium but represented an additional source of information to support the designers within the development process.

A major problem in this case study was the deficient search functionality with mainly unsatisfactory results, leading to the fact that only a couple of people were actually using the system (mainly the editors). The wish for a fast access to relevant contents in specific situations was articulated. The goal of the project was to improve the search mechanisms, thereby encouraging the feeding of more information into the system and increasing the number of system users. The conducted interviews highlighted the fact that personal communication represented the most important source for information, which could never be replaced by a digital system such as the one in focus of the project. Still, the system's potential was seen for situations, in which personal communication is not possible or in which detailed information is sought after (e.g. drawings of components). Although the company employees involved in the project were quite familiar with the contents of the system (since they were all editors), a lack of relevant results was observed in most of the search situations, leading to a decreased satisfaction and reduced use of the database. Therefore, the goal was to enhance these search mechanisms. A question of particular interest was: What functionalities does a search engine have to feature in order to offer similar advantages as personal communication?

2.3 Comparison of the scenarios

Figure 1 gives an overview over important project characteristics. Both scenarios were located within the field of product development. Scenario 1 had a focus on sales processes, in particular finding the right solutions for articulated customer needs, preferably by reusing or adapting existing solutions (within the company or from an external source). In contrast, the focus in scenario 2 lay on component design. Therefore, the information sought after was quite different.

aspect	scenario 1	scenario 2
industrial branch	drive and automation	automotive
company (-ies)	OEM producers, providers of system solutions, component providers	OEM producer
processes/ departments	sales, product development	car body design
sources of information (in focus of project)	a) internal sources: databases, intranet etc. b) company websites, e-market platform	specific knowledge database on the company intranet
expected results of search processes (in focus of project)	information on existing products: (functional) product descriptions, technical data, fields of application	results of design reviews, component failures, corresponding solutions, best practices

Figure 1. Comparison of both scenarios

In both cases, digital databases were regarded that contained valuable information and knowledge. However, the accessibility to the contents and the quality of results in search processes showed high potential for optimization. The scenarios have a major problem in common: the discrepancy between concepts in the mind of the searching actors and the way the contents can be retrieved using the available mechanisms. The general core of the problem is a lack of common understandig and the existence of too many different notions and terms between providers and clients of information.
This matter shall be illustrated by giving examples originating from scenario 1: first, there are varying vocabularies concerning the same object within a company's departments. R&D engineers think in terms of part specifications (materials, geometry, pneumatic pressure etc.) whereas sales employees deal with customer relevant vocabulary (aesthetics, performance, features etc.). Second, customers express their needs (e.g. the function a system has to fulfill), whereas manufacturing companies offer technical components (e.g. pneumatic cylinders), so that a match between the two is often a problem. As a third aspect, clients want to compare different products on the market in order to determine the optimum solution for their needs, bearing the difficulty that different companies use different vocabularies (product classification etc.). An additional barrier is the difference in spoken languages (German, English etc.) when considering the global market. In scenario 2, there was no such multitude of viewpoints. However, similar problems concerning the lack of common understanding were identified (e.g. existence of a variety of terms for the same object in different departments).
In addition, the existing mechanisms to access contents often just supported a single view or category (e.g. product component structure), leading to inadequate search results concerning the keyword search as well as the search by categories. The possibility to switch between different views or to combine several categories and therefore allow for a more focussed search, was expressed as a major requirement. In both cases, an ontology-based approach was chosen to address the mentioned issues, the motivation therefor is explained in the next chapter. The focus was put on searching and retrieving information, not on the preparation and storage of contents.

3. Solution Approach – Ontologies

Many definitions of the term ontology exist. One of the newest that is widely used is given by [Studer et al. 1998]: "An ontology is as a formal, explicit specification of a shared conceptualization. A 'conceptualization' refers to an abstract model of some phenomenon in the world by having identified the relevant concepts of that phenomenon. 'Explicit' means that the type of concepts used, and the constraints on their use are explicitly defined. 'Formal' refers to the fact that the ontology should be machine readable. 'Shared' reflects the notion that an ontology captures consensual knowledge, that is, it is not private to some individual, but accepted by a group."
[Uschold & Jasper 1999] adopt the following characterization: "An ontology may take a variety of forms, but necessarily it will include a vocabulary of terms, and some specification of their meaning. This includes definitions and an indication of how concepts are inter-related which collectively impose a structure on the domain and constrain the possible interpretations of terms." The intention of this rather broad interpretation was to facilitate a comparison of goals and technical approaches in different research communities (databases, software engineering, knowledge engineering etc.).
Nowadays, ontologies are used for a variety of different purposes (e.g. natural language processing, knowledge management, e-commerce, etc.). There is also a growing interest in developing and applying ontologies within the domain of product development. [Eris et al. 1999] explain the need for an ontological approach with the apparent lack of cumulative quality in empirical studies of product development projects. In order to allow for an improved collaboration within the design research community, for instance by providing a tool for conference paper classification, an initial ontology of product development processes was created. In subsequent papers, the framework was evaluated and further developed (e.g. [Mabogunje et al. 2002]).
[Storga et al. 2005] present the framework of a 'Design Ontology' as a formal design language that is meant to facilitate the description, explanation, understanding and reuse of product development knowledge. Therefor, informal definitions of the concepts from the product development domain were converted into a formal design model. The first proposal for the 'Design Ontology' consists of a core of about more than 150 different terms, which were extracted from the Genetic Design Model System

(GDMS). The classification and definition of terms was done according to the SUMO (Suggested Upper Merged Ontology).

[Li et al. 2005] concentrate on the representation and retrieval of design knowledge. Their approach comprises a systematic and intuitive way of design knowledge representation by using ontologies, the extraction of design knowledge embedded implicitely in 3D CAD models, an ontology-based personalized retrieval system and the transformation between user-specific semantics and system semantic representations. A layered FBSO model (Function-Behavior-Structure Ontology) was used for design knowledge representation, including ontologies for functional concepts, behavior concepts, features, components and mating relations. Furthermore, eight kinds of relationships were defined.

[Shin et al. 2005] deal with error inducing systems and design induced error (DiE). The aim of their work is to provide theoretical explanations of DiE and to develop an ontology to support the automated or semi-automated processing of accident and failure reports. They describe two strands of ontology development: first, the development of a metatheory (as a means of synthesizing several different theories that focus on similar or different aspects of the regarded domain), and second, the study of documentation of accidents and failures to identify evidence of DiE and to explore how an ontological approach may assist in the (semi-)automated processing of the documentation.

When reflecting the common problems identified in the two specific scenarios in focus of this paper (see chapter 2), and comparing them with the possibilities offered by an ontology-based approach (as discussed within this chapter), the conclusion is drawn, that ontologies represent a promising instrument for our case. [Uschold & Jasper 1999] developed a classification of ontology applications that contains two cases with relevance for our approach: (1) the ontology as enabler for a common access to information, in particular as a promoter of common understanding among humans, and (2) the ontology-based search, i.e. the use of an ontology for searching an information repository for desired resources (e.g. documents, names of experts etc.). In the following, we will describe how an ontology-based approach was put into practice within the two scenarios.

4. Procedure and Results

After the problem clarification phase, the focus in scenario 1 was placed on developing adequate classes (concepts on a superordinate level) for the regarded domain and their general relations. In additition, taxonomies for each class should be developed by explicitely considering established and widely accepted classifications as far as they were available. The project's goal was to evaluate the potential of the ontology-based approach, therefore building the basis for further activities. The development of a software system (linking the available heterogeneous databases and providing an ontology-based search engine) was planned for a follow-up project. In scenario 2, being less generic, the project was performed in close collaboration to a provider of ontology software. The represented domain and its entities were more specific. Therefore, the focus was placed on optimizing the search within an existing particular database with support of ontology techniques. The implementation and concrete validation of the ontology were the major goals. Despite these differences concerning project details, procedure and methods in both cases were rather similar and can roughly be divided into two phases: ontology development and ontology evaluation.

4.1 Ontology development

The development of the ontology in scenario 1 took place in a series of workshops at the Institute for Product Development amongst the participation of several employees of the involved companies. First, eight general classes were defined top-down (see figure 2, upper left part). Classifications were generated for these classes with the aim of considering existing and widely accepted definitions to the most possible degree. Therefore, an industrial classification (SIC code) was adopted for the class "industrial sector". For "product" and "product property" a classification established by the German Engineering Federation was used. For "product requirement", the same classification as for "product property" was adopted, since product requirements represent desired/requested product properties, and hence the concepts are very related. For "function" and "principle", existing classifications within product development methodology were investigated and compared (e.g. [Pahl & Beitz 1996]). A first taxonomy of functions was established, which has to be refined in further project activities. Parallel to the top-down classification, contents for the ontology were developed bottom-up with the help of the defined use case. Therefore, the relevent terms were reduced to a limited number that allowed the derivation of conclusions with a justifiable effort. For each of the top level classes, the relevant terms with respect to the use case (i.e. concrete instances) were gathered (see figure 2, upper right part). This concretization allowed the derivation of important relations between the classes, which were then generalized and assigned to the top level classes (see figure 2, lower part). Importance was placed on defining just the most important relations, not all relationships between classes that are theoretically possible. In figure 2, relationships are displayed within the cells of a matrix that has to be read in the following order: element in row is related to element in column by relationship given in the cell of the matrix (e.g. product fulfills function).

class / sub-class		description	instances (use case) e.g.
industrial sector		classification according to SIC-Code	screw machine products, communications equipment
product	end user product	product subjected to a transformation	mobile telephone
	OEM plant	system that carries out the transformation, consists of at least one station	screwdriving center
	OEM station	part of an OEM plant	screwdriving system, process control unit
	OEM component	part of an OEM station	laser, camera, photo sensor, push button, spindle
product requirement		basis for the evaluation of a product	package, availability, power consumption, EMC
function		purpose of a product	detect presence, detect position, generate signal
principle		abstract description of function fulfillment	electric, magnetic, optical, hydraulic, pneumatic
product property		specification parameter of a product, which is compared to the requirements	package, availability, power consumption, EMC
company		provider of components / system solutions	
person		actor in a company in a certain role	

	industrial sector	product	product requirement	function	principle	product property	company
industrial sector		requires	determines				
product	is deployed in		fulfills	fulfills	bases on	possesses	is produced by
product requirement		is fulfilled by					
function		is fulfilled by			is realized by		
principle		is realized in					
product property		is realized in	fulfills		specifies		
company		produces					

Figure 2. Specific ontology in scenario 1 (excerpt), definition of classes and relations

In scenario 2 the raw data for the initial ontology was collected by documenting search processes. As in scenario 1, the focus was placed on a specific topic (here: fasteners in interior design of car doors) in order to reduce the complexity to a manageable degree. The initial ontology was developed by three researchers from the Institute for Product Development by analyzing search protocols, which had been filled out by designers in the area of car interior processing queries in the system. Important terms were extracted and clustered into categories (classes) and general relations between the classes were derived. Additional terms and synonyms (that were not found in the protocols) were added in order to make the ontology large enough for an evaluation. The initial ontology was then implemented in an ontology software (Onto-Edit®) and presented to the involved company employess in a workshop. Hereby, functions of the ontology-based search from the technical point of view and first potential improvements of the system could be demonstrated and discussed. In addition, the contents of the ontology were discussed and revised. The class "function", which had been defined for the first version of the ontology, was e.g. dropped in that phase, since the discussions revealed slight confusions concerning the meaning of the term. It was rather used in the sense of "requirement" and represented no additional benefit. The second version of the ontology consisted of 10 top level classes (see figure 3, upper part) and 15 relations (see figure 3, lower part).

class	instances (use case) e.g.
requirement	assembling force, stiffness, vibration behavior
material	PP, ABS, POM, PA, metal
component	retainer, clip, screw, fastener, adhesive joint
module	car interior, attachment parts, trunk
design specification	radius, stiffness, material thickness

class	instances (use case) e.g.
assembly	blind assembly, repeated assembly
production	bonding, welding, plugging
experience	
failure	rattling, squealing, breaking, noise
critical area	material thickness, material type

	requirement	material	component	module	design spec.	assembly	production	experience	failure	critical area
requirement			is set up for							
material			is employed in							
component	has to fulfill	consists of		is employed in		is processed in	is produced by		is affected by	
module			includes							
design spec.			specifies					results from		
assembly										
production										
experience				refers to	is required for					
failure								can be avoided by		results from
critical area										

Figure 3. Specific ontology in scenario 2 (excerpt), definition of classes and relations

4.2 Ontology evaluation

In scenario 1, only a qualitative evaluation was possible, since the implementation of the ontology in selected databases was not part of the project. However, the relevance of the derived classes and relationships for industrial practice could be assured. Company employees participating in the project manually combined terms of the ontology in keyword searches related to the defined use case. This "manual simulation" of the ontology led to more relevant search results.

In scenario 2, the ontology was implemented in Onto-Edit® and integrated in a search engine provided by another software company involved in the project. The regarded knowledge base was indexed and search queries for the application case "fasteners" were run. Search processes with and without ontology support were compared. The results were promising: in situations where the conventional search delivered not enough or no results at all, the ontology-supported search widened the searching space and delivered about 10 times more results. The reason was that synonyms and terms closely related to the keyword in the ontology were also considered in the search (e.g. for "retainer" there existed the synonyms "clip", "dome" and "clipdome"). Of course, 100 or more search results cannot all be examined for relevance. Therefore combinations of keywords, suggested by the ontology software, were applied to confine the search space again (e.g. components in combination with failures, modules and materials). Applying 2 keywords reduced the number of search results from over 100 down to 20-30. A combination of 4 keywords lead to 2-4 search results, most of them of high relevance for the search topic. Other aspects that represented an improvement in comparison with the existing system were the consideration of typing errors, the search within indexed documents and the explicit exclusion of terms from the search. Aspects criticized by the involved employees were the new layout of the program (GUI) they were not used to and a generic thesaurus that was also implemented (as an example in this respect, the term "dome" was related to the term "church" by the thesaurus, which brought no benefit in the considered use case).

5. Discussion and Conclusions

In summary, two scenarios were described dealing with processes in sales and product development. Deficits in knowledge management via digital databases were addressed as core problems. The potential of ontologies as a common framework for communication and a tool for improving search engines was discussed. In the projects, initial ontologies have been developed for clearly defined use cases, (partially) implemented and evaluated. In scenario 1, there are some special implications concerning the domain of sales processes. One of the biggest problems here is to match the client/customer view with the engineer's view, i.e. needs and requirements with technical solutions and product specifications. Two classes that potentially close the gap are "functions" and "principles".

[Kitamura & Mizoguchi 2003] present an approach towards organizing knowledge by functional decompositions. They distinguish explicitely between functions ("what goal is intended to be achieved") and ways of function achievement ("how the goal is achieved") which includes the (physical) principle of the achievement. In scenario 2, the consideration of failures (defects, errors) is considered an important aspect. One of the major goals for setting up the system in the first place was to store the knowledge on failures in order to avoid them in the future. The ontology-based approach enables the users to link failures to critical areas in which they occur, components that are affected by the failure, but also to best practices or lessons learned (class "experience"). Therefore, knowledge concerning error prevention can be transferred from one area to another with the help of the system (e.g. experience made by car interior designers concerning problems with particular components can be transferred to designers dealing with similar components but in other modules, such as car boot). In both projects it became clear that there exist typical "standard paths" through the ontologies. The classes are connected in a logical order that is processed in the minds of the involved actors. In the ontology these patterns are explicitely documented (e.g. component is affected by failure, failure results from critical area, failure can be avoided by experience, experience refers to module etc.). A navigation through the ontology (enabled by an ontology browser application) is therefore closely related to the processes in the minds of the involved actors. Moreover, the ontology provides valuable terms (concepts) that may not be on the user's mind, but are still relevant for the search process.
The work described in this paper represents a first step towards the solution the discussed issues. The ontologies have yet to be expanded and tested within a broader field of application. Future investigations have to address the topic of ontology development effort and the possibility to include approaches towards multiple viewpoints in design information retrieval (e.g. [French et al. 2002]).

References

Eris, Ö., Hansen, P., Mabogunje A. and Leifer, L., "Toward a pragmatic ontology for product development projects in small teams", Proceedings of the ICED 1999, Munich, 1999.

French, J.C., Chapin, A.C. and Martin, W.N., "Multiple Viewpoints as an Approach to Digital Library Interfaces," Workshop on Document Search Interface Design and Intelligent Access in Large-scale Collections, Portland (OR), 2002.

Kitamura, Y. and Mizoguchi, R., "Organizing knowledge about functional decomposition", Proceedings of the ICED 2003, Stockholm, 2003.

Li, Z., Anderson D.C. and Ramani, K., "Ontology-based design knowledge modeling for product retrieval", Proceedings of the ICED 2005, Melbourne, 2005.

Mabogunje, A., Hansen, P. K., Eris, Ö. and Leifer, L., "Product development process ontology", Proceedings of the International Design Conference Design 2002, Dubrovnik, 2002.

Pahl, G., Beitz, W., "Engineering Design", London: Springer 1996.

Shin, I.J., Busby, J.S., Hibberd, R.E. and McMahon, C.A., "A theory-based ontology of design induced error, Proceedings of the ICED 2005, Melbourne, 2005.

Storga, M., Andreasen, M. M. and Marjanovic, D., "Towards a formal design model based on a genetic design model system", Proceedings of the ICED 2005, Melbourne, 2005.

Studer, R., Benjamins, V. R. and Fensel, D., "Knowledge engineering: principles and methods", Data and Knowledge Engineering, 1998, pp. 161-197.

Uschold, M. and Jasper, R., "A framework for understanding and classifying ontology application", Proceedings IJCAI99 Workshop on Ontologies and Problem-Solving Methods, Stockholm, 1999.

Josef Ponn, Dipl.-Ing.
Technische Universität München (TUM), Institute for Product Development
Boltzmannstr. 15, 85748 Garching, Germany
Tel.: +49 (0) 89 289 15141
Fax.: +49 (0) 89 289 15144
Email: ponn@pe.mw.tum.de
URL: http://www.pe.mw.tum.de