INTERNATIONAL DESIGN CONFERENCE - DESIGN 2008
Dubrovnik - Croatia, May 19 - 22, 2008.

DESIGN 2008

USER SPECIFIC SUPPLY OF DOCUMENTS FOR PRODUCT DEVELOPMENT KNOWLEDGE BY MEANS OF A COMPREHENSIVE TOPIC MAP APPLICATION

H. Weber, M. Lenhart and H. Birkhofer

Keywords: knowledge transfer, topic maps, modularization, user specific documents, information retrieval

1. Introduction

The general objective of the current ***pinn****gate*-project within the department of product development and machine elements at the Technische Universität Darmstadt is a flexible knowledge system for learning, teaching and application of product development knowledge [Birkhofer et al. 2004]. This includes support for authors in building a high quality contents basis and also addresses the need for a flexible configuration and access system for the users, with regard to the individual requirements.
The tools currently available for the transfer of product development knowledge consider the specific needs of the users only in a limited way. Because e.g. textbooks are often written for a wide spectrum of readers, there is a lack of individual and user specific addressing. This problem is mainly due to the static nature of textbooks and also many electronic documents. Although the latter provide advanced access and navigation possibilities, a topic is often described in only one way. The adaptation of available documents to different user needs and the process of creating new documents are very time and work intensive. Another problem arises if the documents are written by different authors with different points of view.
Approaches for solving these problems can be found in the domain of psychology of learning and cognition connected with realizations from the modular storage of documents and information retrieval theories. According to such sources, semantic nets play an important role for structuring the knowledge domain. Nevertheless, the respective models and methods are developed separately and have to be well attuned for efficient usage in an integrated system.

2. State-of-the-art

In the domain of knowledge transfer (in general and also with computer support) many findings and results are available on how to facilitate the learning process. The following sections outline the already developed fundamentals about user specific documents, semantic nets in general and topic maps specifically, modularization and information retrieval as bases for the current work.

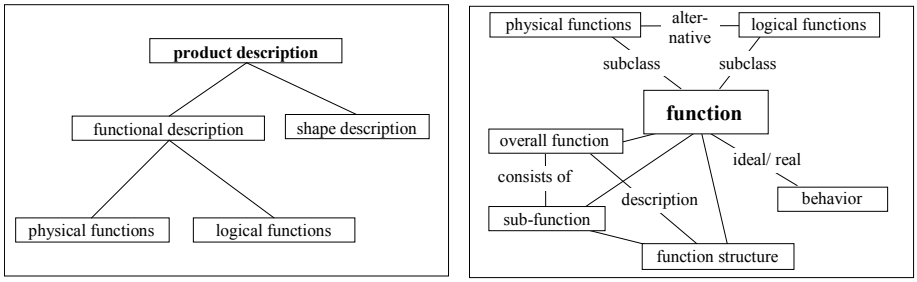
2.1 User specific documents

Documents, whether printed or electronic, have to fulfill several requirements for an efficient and sustainable transfer of product development knowledge. On the basis of cognitive theories (cognitive load theory and cognitive flexibility theory) Lenhart develops requirements for adapted documents with regard to contents, structure and layout depending on the current user, the user's expertise and the

actual situation [Lenhart and Birkhofer 2006, 2007]. One example is the required ratio of expository, instructional and additional didactic texts in the whole document. Experts in application scenarios want short instructional texts (how to perform a specific method) whereas novices in a learning scenario need more extensive descriptions of the theoretical background including explanations and definitions of unknown terms.

2.2 Semantic nets

A semantic net is a formal model of concepts and relations between them, with the aim to provide a better overview of the main concepts of the viewed knowledge domain. An advantage of the net representation over a linear structure is assumed, because it complies more with the structure of the human brain [Quillian 1968]. The cross-linking between several bits of knowledge is essential for their accessibility. Learning facts by heart is useless without the awareness of their fields of application.
Semantic nets occur in manifold forms. Depending on the complexity of the structure and the permitted types of relations, taxonomies, thesauri and ontologies are differentiated. If the terms are arranged in a strict hierarchical order, the result is a taxonomy (Figure 1a). No other information is added, e.g. information about the differences between the levels or relations of two terms at the same stage. This is the benefit of a thesaurus. Here, a standardized set of relation types can be used to label terms as (preferred) synonyms, broader/ narrower or otherwise related ("see also"). The highest level of a semantic net is reached by an ontology where any desired relation type can be used (Figure 1b).



a) *b)*

Figure 1. Sample a) taxonomy and b) ontology

The use of semantic nets in the area of product development is nothing fundamentally new. Several elements in product development literature and methodology can be seen as semantic nets. Examples are overview graphics, tables and classification systems like the catalogues for engineering by Roth [Roth 1994]. There is also increasing use of ontologies for the access of information about products and their structural elements [Pocsai 2000]. But this is not the case when regarding the transfer of product development methods and models. There are definitely several approaches for a structured transfer of product development methods, but they do not make use of the extensive possibilities of semantic nets.

2.3 Topic maps

If ontologies should be used in a computer aided system, they have to be stored in a data model like topic maps, RDF or some other description language (e.g. Ontolingua). Here, topic maps, which are internationally standardized under ISO 13250, provide a technology which is most suitable to the requirements of the current project, by focusing on information retrieval in a delimited area of knowledge as well as facilitating easy creation of relations between concepts and links to external resources (documents).
Topic maps consist mainly of three types of elements: topics, associations and occurrences [Pepper 2000]. Topics are symbols for anything the author of the topic map wants to make statements about. In the field of product development knowledge these are methods, models, theories or examples. Associations define the relations between topics. They have types designated by the topic map author

and can connect two or more topics. To explicitly mark the direction of the relations, a specific role is assigned to each topic. Occurrences constitute the connection from the topic map to any resource (document) dealing with a specific topic. These could be among others definitions, method descriptions or examples. The resources themselves are not part of the topic map but only the links to them. Principally only topics can have occurrences, but with the help of a mechanism called "reification" associations can be treated like topics and consequently have occurrences. This makes it possible to assign a description of the difference between "function" and "behavior" directly to the association between these topics.

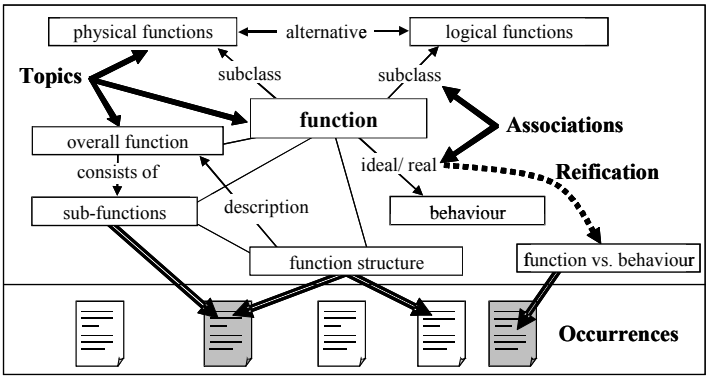


Figure 2. Sample extract from a topic map about functions

2.4 Modularization

The intentional development of modular systems is used in different disciplines. The main objectives are reuse of elements, simplification, effort reduction, standardization, individualization and flexibility. Examples are modular systems, platform strategies but also the concepts of object-oriented programming, database systems and hypertexts.

The main aim of modularization of documents is the reuse of texts, figures, tables etc. in several documents. This decreases the needed storage volume and increases the consistency by automatically changing all documents containing the same text module at once. Thus, modularization not only means dividing documents into small parts, but also the separation of contents, structure and format [Berger 2004]. Adapted documents are created out of such a modular basis in a configuration and retrieval process under consideration of the current user and his situation.

2.5 Information retrieval and navigation

Information retrieval is understood as finding and filtering relevant information for the user and his current task. The results depend on the objective requirements, which are necessary to cope with the task, the user and how he or she can articulate his individual requirements as well as the information stored in the database. The goal of modern information retrieval methods, see e.g. [Baeza-Yates und Ribeiro-Neto 1999], is an enlargement of the intersection of the demand and supply. A first step for this is user support in expressing queries. Common tools for this are full text searches, logical combination of terms with Boolean operations ("and", "or" …) and meta data searches. But these are only successful if the user roughly knows, in advance, what he or she is looking for, which is not always the case [Ahmed 2005]. Here, an additional semantic net can help the user to find unknown concepts out of their context (e.g. methods affecting a particular model).

Besides search queries there are different navigation possibilities for the users. They go from simple hierarchical structured tables of contents (site maps) over cross references (links) in (hypertext-) documents to graphical navigations with semantic nets. A good balance between widespread navigation possibilities and the resulting complexity and clear arrangement has to be found here. Too few navigation possibilities limit the accessibility of some documents, whereas too many links lead to the common "lost in hyperspace" problem.

3. Deficits and research objectives

As described, there has been much research done in the field of didactics, cognition and information technology. The fundamental problem is that the previous approaches for structuring, storing and retrieving product development knowledge were developed separately and independently from each other in most instances. Thus, the dedicated methods are based on different, incompatible and partly not formalized models. They can only be coordinated with great additional effort and with loss of quality. But a combined use of the technologies in an integrated system is necessary for a holistic approach, which regards all aspects concerning the domain of product development knowledge, the modular contents basis, the user specific requirements and the adapted information retrieval process. This is why the goal of the current research is the development of a framework for such an integrated system. Thereby, the specific requirements of imparting product development knowledge have to be continuously taken into consideration.

4. Methods

Instead of creating distributed interfaces between each pair of the existing models and methods, the new framework contains a unified description of the three fields product development knowledge domain, modular contents and users. Therefore the universally applicable topic maps standard is an appropriate means. With a suitable topic map structure for each domain the specific requirements can be fulfilled without losing the compatibility with the other fields. Thus, the different topic maps can be integrated in a superior system. By transferring search queries into topic maps the original information retrieval problem changes to a topic map matching problem which can be solved mostly by algorithms.

The fundamental approach of the comprehensive topic map application for the user specific supply with documents comprises four main parts (figure 3). These are:

1. Analysis of (modularized) documents
2. Mapping of the domain of product development knowledge
3. User modeling
4. Information retrieval with topic maps

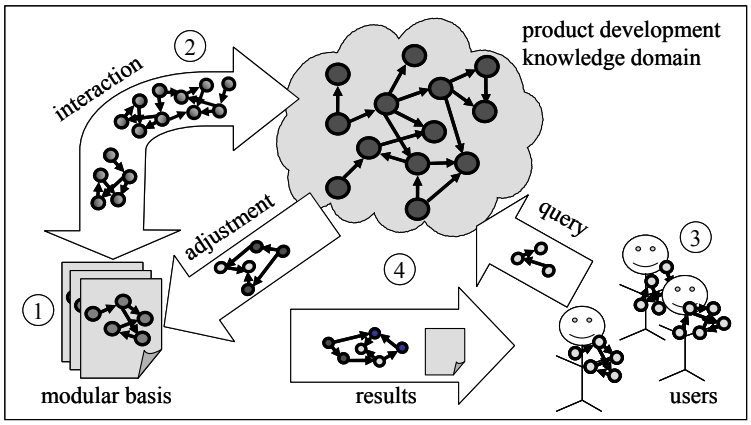


Figure 3. Functionality of the topic map-based overall system

The following sections explain the intentions of these parts, and then the characteristics of the used topic maps are summarized.

4.1 Analysis of (modularized) documents

Existing and newly written documents about product development are analyzed with regard to their contents and structure. The result is a modular basis, which is described by the module topic maps. That is, for each module the relevant terms/ concepts and additional meta data are captured and a topic

map is built in dependence of the document structure. Therefore characteristic types of contents (definitions, method descriptions, examples …) are detected. While manual methods are very work and time intensive for analyzing a great number of large documents, fully automatic methods do not provide the required quality. A compromise is a semi-automatic procedure, in which algorithmizable analyses are executed by a computer with the help of text mining methods, while the quality assurance and release of the results remains in the hands of human experts.
The analysis of the documents proceeds in an iterative interplay between modularization, text mining and topic map construction [Weber et al. 2007]. Through the modularization, the semantic contents of each module become relatively equally sized. Then each module can be used as a closed range for text mining analyses. This is advantageous because methods for detecting significant terms and associations need such ranges as references for calculating relative probabilities. In addition, clustering algorithms detect similar parts of texts and help to detect thematic borders in a document for revising the modularization.

4.2 Mapping of the knowledge domain

The knowledge domain is stored in the product development topic map. This is built using the modular basis. Additional topics and associations directly added by experts are useful, especially in the beginning of the implementation, in order to start with a basic topic map, which then is gradually extended.
The modular basis and the captured knowledge domain are coordinated by adding new significant topics found in new documents to the knowledge topic map and by assigning the modules as occurrences. Thus the product development topic map is a consolidated topic map over-reaching the module topic maps. Here the topic maps do not only serve for storing the information about each module and the product development knowledge, but also for the interchange of information in the adjustment process.

4.3 User modeling

The product development topic map and module topic maps do not depend on user specific requirements. On the basis of existing documents and expert knowledge, the semantic nets are built as universal as possible. This enables the user in some way to find suitable documents about a particular subject. Nevertheless, these documents will not fit his or her personal needs regarding the extent and structure of the texts. This is why these needs have to be mapped and taken into consideration in the selection of the results. Thereby the captured requirements act on the one hand as a filter by hiding improper documents, on the other hand as additional structuring elements. That means e.g. unknown definitions can be added to a document or the structure gets adapted.
For the implementation of a user and situation specific supply of documents, the relevant properties of the users have to be captured and processed in the system. Therefore the users are also described by topic maps. These are built upon the existing individual level of knowledge and contain a subset of topics and associations out of the knowledge map. Additional meta data about the users can be added. But these pieces of information are not easy to gather. One possibility is a self-assessment of the users for a rough classification. Here the problem arises that some users might not be able to evaluate whether his or her own level of knowledge fits the required level. Therefore, control questions are used to help detect weaknesses and strengths. An additional so called user-tracking is used; i.e. information from the usage of the system (clicked pages, search queries …) are collected. As a result, e.g. called definitions are assumed to be familiar in future uses and will then occur in a glossary rather than integrated in the following texts. Another possibility is an active rating of delivered documents by the users to detect misclassifications of documents and/or users.

4.4 Information retrieval with topic maps

The three models for the description of the knowledge domain, the modular basis and the users have to be linked for a powerful overall system. Only if all models interact in a common framework, will user queries be fulfilled successfully. This is why there have to be interfaces between the models. All

information in the different fields is stored in topic maps, so the information retrieval process also proceeds with the help of topic maps. Therefore several topic maps are generated dynamically. Queries are transferred into a topic map and sent to the system together with the user topic map no matter how the user makes an enquiry (cross-links in texts, full text searches, navigation in a graphical presentation of the knowledge topic map). This query is then matched with the product development topic map in order to determine the requirements for appropriate documents. These concern the specific subjects, types of contents and format options. Out of this a topic map for adjustment is generated and sent to the modular basis. Here modules for meeting the demands are selected and combined. Because one query seldom leads to a non-ambiguous document, the results of the topic map matching are also represented by a topic map. With its help, the user chooses one or more final documents.

4.5 Characteristics of the topic maps used

For describing the product development knowledge the seven different topic types facts, definitions, processes, models, methods, auxiliaries and examples are distinguished (table 1). The particular topics are connected by associations of types like "the functional description of a product (*model*) *is more abstract than* the describing of the physical effects (*model*)" or "a function structure (*model*) *is the output of* the function analysis (*method*)".

Table 1. Types of topics for describing the product development knowledge

Topic type	Description	Example
Facts	Technical and theoretical facts about product design	Product lifecycle phases
Definitions	Fixed meanings of (technical) terms	Definition of "function"
Processes	Knowledge about the design process and procedures	Procedure model of the VDI 2221
Models	Abstract product descriptions, their theoretical background, notations and structures	Product function, function structure
Methods	Planned procedures with concrete guidelines for supporting the product design process	Method function analysis
Auxiliaries	E.g. Forms for tools for supporting the creation of models or the performance of methods	FMEA-form
Examples	Illustrations/ applications of the types above	Results of prior method applications

In the entire system six different types of topic maps are used for storing all necessary information and for implementing the information retrieval process (table 2). Thereby quasi static topic maps, which describe an actual state, and dynamic topic maps, for processing purposes, are differentiated. A quasi static constancy means that the topic maps may change over time (e.g. growing knowledge of a user changes the user topic map), but not with every query. On the other hand the dynamic topic maps are created with every query in the information retrieval process and are discarded afterwards. Because all topic maps change during the usage of the whole system in order to fit the actual situation (contents of the modular basis, current users …), they are generally spoken of as adaptive topic maps. The module topic maps are also regarded to be quasi static, because of the bidirectional adjustment of the modular contents basis and the knowledge topic map.
As mentioned before, concept structures discovered in recently integrated documents lead to an extension of the product development topic map. Additionally, afterwards the topic maps of the existing modules are revised if new relations between concepts have now become significant in the knowledge topic map. In this case the modular basis is analyzed, if this affects the particular topic maps.

Table 2. Types of topic maps in the framework

Topic map	Description	Constancy
Product development topic map	Consolidated topic map capturing the relevant concepts of the product development knowledge domain and relations between them	quasi static
Module topic map	Description of the modular basis with regard to the contents,	quasi static

	structure and format	
User topic map	Topic map about the individual knowledge, situation and user specific requirements	quasi static
Query topic map	Represents a search query of the user to the system as input for the information retrieval process	dynamic
Adjustment topic map	Internal topic map created by the adjustment of the search query with the product development topic map for determining the requirements for the documents to be delivered	dynamic
Result topic map	Contains the results of the information retrieval process	dynamic

5. Results, conclusions and further work

First positive experiences in accessing information with the aid of topic maps have been made by implementing a prototype system called "***pinn**gate* research portal", which is based on the described approach. Thereby, the chapter about the modeling of product functions out of the current lecture notes, which are used at our department, served as a test extract from product development knowledge. This was integrated in the modular basis and analyzed with the aid of text mining methods. In combination with the modularization, the knowledge topic map was built and the modules where characterized by the contained subjects and the types of the texts. With the help of this information, a topic map based graphical navigation and a text based term search engine has been implemented for fast access to the several modules, which build the documents shown to the user (figure 4).

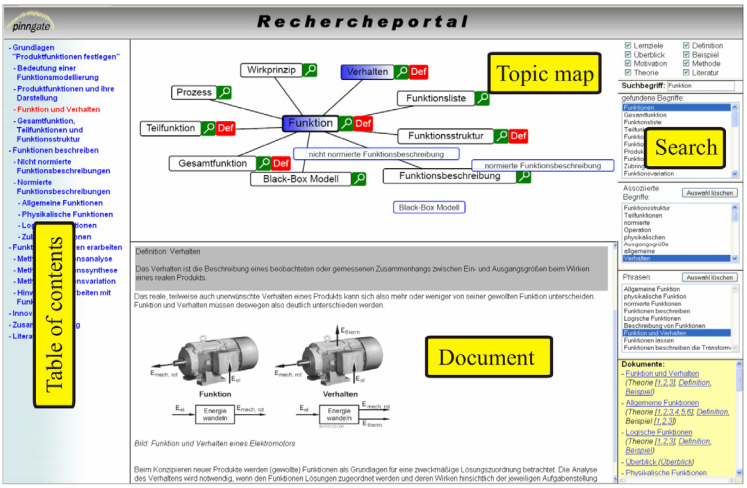


Figure 4. The *pinngate* research portal

For a first evaluation a group of 76 students and 10 research associates were asked to answer several questions with the help of the system. Thereby the main focus was on the analysis of the logged search/ navigation path and only secondary on the quality of the answers. The results were, that if modules were viewed, the test persons averagely viewed 2.5 modules before writing down an answer. In 83% of these cases the path included to the module, which contained the relevant information. In about a half of the cases the right module was even found at the first go. Even though these results may not be overvalued, because of the small size of the modular basis, these results indicate the suitability of the presented approach as a framework for an effective transfer of product development knowledge. The comprehensive use of topic maps utilizes powerful technology for this task and allows an iterative interaction between the existing and formerly incompatible approaches for knowledge management, modularization and user classification.

At the current stage not all parts of the framework are realized in the prototype. Especially the user modeling is only at its beginning. The first implemented approach is user-tracking, which captures all concepts which are dealt with in the texts shown to the current user. Using this information, a user specific topic map is built, which consists only of these concepts and their associations. This helps the user to get an overview of the relevant concepts in his current work with the system.
For further enhancement of the system, it mainly needs to be filled with more documents. On the one hand a broader spectrum of subject matter has to be covered to not be confined to a small part of the product development knowledge. On the other hand one subject has to be dealt with in several documents/ modules for better adaptability of delivered documents, because the system can only combine and arrange existing text blocks. Only a first adaptation can be realized in filtering the modules and hiding information which is currently irrelevant for the user. Additional new texts have to be written if the current modular basis is not able to fulfill some user specific queries.

References

Ahmed, S., "Encouraging Reuse of Design Knowledge: A Method to Index Knowledge", Design Studies, Vol. 26, No. 6, 2005, pp. 565-592.

Baeza-Yates, R. A., Ribeiro-Neto, B. A., "Modern Information Retrieval", ACM Press/ Addison-Wesley, 1999.

Berger, B., "Modularisierung von Wissen in der Produktentwicklung – Ein Beitrag zur einheitlichen Aufbereitung undindividuellen Nutzung in Lehre und Praxis", Dissertation, Fortschitt-Berichte VDI, Reihe 1, Nr. 376, VDI Verlag, Düsseldorf, Germany, 2004.

Lenhart, M., Birkhofer, H., "Classification of users in the context of knowledge transfer in product development", Proceedings of the 9th International Design Conference - DESIGN 2006, D. Marjanovic (Ed.), Dubrovnik, Croatia, 2006, pp. 1187-1194.

Lenhart, M., Birkhofer, H., "Levels of Expertise in Product Development – Implications for the Design of Instructional Material", Proceedings of the International Conference on Engineering Design 2007 – ICED ´07, Paris, France, 2007, CD-Rom, Paper 249.

Pepper, S., "The TAO of Topic Maps – Finding the Way in the Age of Infoglut", 2000, online: http://www.ontopia.net/topicmaps/materials/tao.html.

Pocsai, Z., "Ontologiebasiertes Wissensmanagement für die Produktentwicklung", Dissertation, Universität Karlsruhe, Shaker Verlag, Aachen, Germany, 2000.

Quillian, M. R., "Semantic memory", Minsky, M. (Ed.), "Semantic Information Processing", Cambridge, Massachusetts, MIT Press, 1968, pp. 216-270.

Roth, K., "Konstruieren mit Konstruktionskatalogen", 2nd edition, Springer, Berlin, 1994.

Weber, H., Lenhart, M., Birkhofer, H. "Iterative Semi-automatic Modularization of Documents and Topic Map Creation for Product Development Knowledge", Proceedings of the International Conference on Engineering Design 2007 – ICED ´07, Paris, France, 2007. CD-Rom, Paper 197.

Harald Weber
Technische Universität Darmstadt
Department of product development and machine elements
Magdalenenstraße 4
64289 Darmstadt
Germany
Tel.: +49 6151 16-3378
Fax.: +49 6151 16-3355
Email: weber@pmd.tu-darmstadt.de
URL: http://www.pmd.tu-darmstadt.de