Title
A.B. Editor et al. (Editors)

Ontology-Based Information Management in Design Processes

Sebastian C. Brandt[a], Jan Morbach[b], Michalis Miatidis[a], Manfred Theißen[b], Matthias Jarke[ac], Wolfgang Marquardt[b]

[a]*Informatik V (Inform. Systems), RWTH Aachen University, 52056 Aachen, Germany*
[b]*Process Systems Engineering, RWTH Aachen University, 52056 Aachen, Germany*
[c]*Fraunhofer FIT, Schloss Birlinghoven, 53754 St. Augustin, Germany*

Abstract

Engineering design processes are highly creative and knowledge-intensive tasks that involve extensive information exchange and communication among diverse developers. In such dynamic settings, traditional information management systems fail to provide adequate support due to their inflexible data structures and hard-wired usage procedures, as well as their restricted ability to integrate processes and product information. In this paper, we advocate the idea of Process Data Warehousing as a means to provide an information management and integration platform for such design processes. The key idea behind our approach is a flexible ontology-based schema with formally defined semantics that enables the capture and reuse of design knowledge, supported by advanced computer science methods.

Keywords: Process Data Warehousing, Ontologies, Information Management

1. Introduction

Knowledge about engineering design processes belongs to the most valuable assets of an enterprise. Typically, a vast amount of this design knowledge is manipulated by legacy tools and stored in highly heterogeneous sources, such as electronic documents and data bases. To fully exploit this intellectual capital, the knowledge must be made explicit and shared among designers and across the enterprise. Thus, the prominent concern of any successful approach is the integration of all these knowledge sources in a coherent framework that supports the mining of knowledge and its reuse on demand.

In the literature, we can identify a plethora of contributions for the support of engineering knowledge management inside manufacturing enterprises. Document management systems are widely used in industrial praxis for the storage, maintenance, and distribution of documents. A step further, Product Data Management (PDM) systems provide extended facilities for the handling of detailed product information. Regarding the process support, however, current PDM systems have largely focused on the workflow management level [14], while the fine-grained support of development activities (e.g. engineering best practices) has attracted less interest. The identified contributions adequately support information exchange, especially in the later phases of the engineering lifecycle, which are characterized by complete and well-known processes and product models. They lack essential knowledge management capabilities [5] and are less suited for the conceptual design stage [5; 14]. Conceptual engineering design processes are highly creative and dynamic processes, which are hardly predictable [12]. Any software solution has to cope with the continually changing requirements and the many degrees of freedom within these processes. Because of their hard-wired usage processes and restricted ability for interoperation, the integrated environments available today are usually unable to offer appropriate support.

This paper presents a novel approach for supporting creative, non-deterministic design processes: The *Process Data Warehouse (PDW)* allows the capture and reuse of design knowledge – data, documents, work processes, and decision making procedures, as well as their interdependencies – through ontologies. In this context, the term "ontology" denotes a conceptual data schema that represents the relevant domain entities and their interrelations by concepts and relations. Within the PDW, ontologies are used to store design knowledge in a formal, machine-interpretable way. This enables the provisioning of advanced computer science methods for managing, enriching, and searching the knowledge within the PDW. Ontology-based knowledge repositories are currently being developed in other areas than process engineering (e.g., [10; 18] in electromechanical engineering). However, these repositories are limited to the storage of product data and documents and do not record the associated work processes and decision making procedures.

2. The Process Data Warehouse

The PDW has been derived from the concept of Data Warehousing [7], where large amounts of fixedly structured data (e.g., from sales or accounting) are stored, aggregated, and then presented. To support design and other creative work processes, more flexible structures are needed, which allow the integration of complex and changing domain models [8]. This flexibility has been achieved by building the system on top of loosely connected ontology modules which are held together by a central *Core Ontology* [3]. The Core Ontology introduces top-level concepts and their relations, which are further refined within the peripheral ontology modules.

Fig. 1 displays a simplified view of the Core Ontology. Four prominent areas of conceptualization are arranged around the Object as the abstract central concept. Relations associate the concepts of one area to another.

Product area. The product area (top) contains concepts for the description of the type and version history of electronic documents and other information resources, as well as their mutual dependencies and their structural decomposition. The Product concept denotes all kind of information elements, such as data items or decision representation objects (e.g. DRL [13]). Products can be aggregated into Document Versions. The different versions of a Product can be bundled by a Version Set. A specialization of Version Set is the (logical) Document, which bundles different Document Versions.

Storage area. The storage area (right) describes at which Storage Place a particular Version Set is located, i.e. in which data base, document management system, or external tool it is stored. A Storage Place forms part of a Store, such as a document management system. This allows a tight integration with document management systems: When a user edits a document with the appropriate tool, the changes can automatically be correlated with their representation in the PDW.

Descriptive area. The descriptive area (left) contains basic concepts for describing the content or the role of Product Objects on a high semantic level. This includes Content Descriptions and Categorizations, which are grouped into Categorization Schemes. Thus, the descriptive area provides the necessary vocabulary for the content-based retrieval of data and documents.

Process area. The process area (bottom) contains the concepts needed to represent the Process Objects that create, use, or modify Product Objects. They comprise general

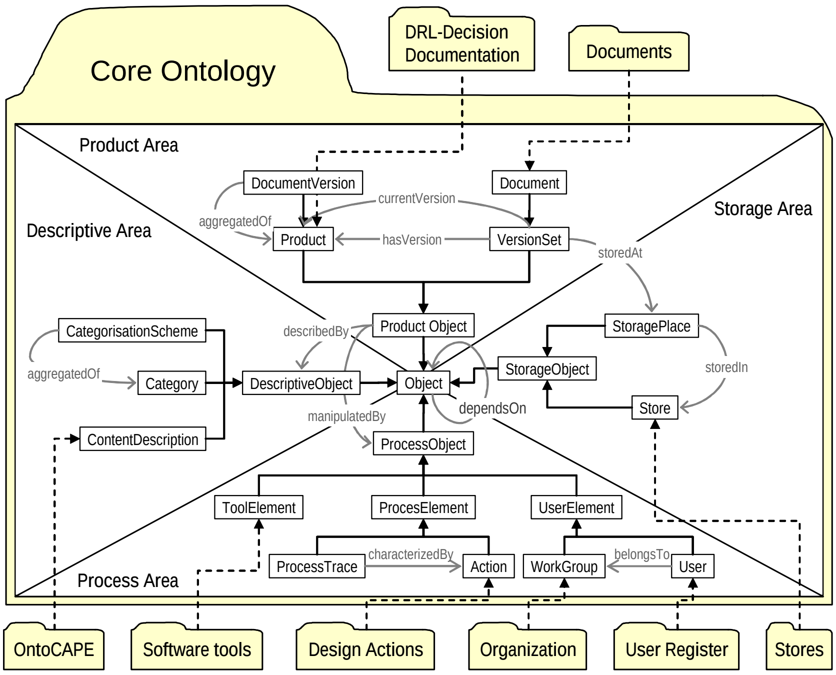

Figure 1. Simplified view of the Core Ontology and some peripheral ontology modules

process definitions (Actions) as well as Process Traces resulting from concrete executions of the Actions by Users or Software Tools.

The dependencies between elements of all four areas are explicitly modeled as an additional area orthogonal to the other four through the depends On relation. This allows the formulation of specialized relations between objects, independently of their concrete relationships and attributes.

Around these fundamental and domain-independent areas, extension points are placed that can be used to add ontology modules for specific application areas or other specializations. The most elaborate of these extensions is OntoCAPE, a large-scale ontology for the description of the process engineering domain [20], which covers fields like physicochemical properties, process equipment, and mathematical modeling. Here, it extends the descriptive area by refining the Content Description concept (refinement is indicated by dashed arrows in Fig. 1).

Unlike conventional tools that store their data in fixed schemas, the PDW uses ontologies with explicitly defined semantics to structure the data domain. The concrete data is then stored as instances of the ontological concepts. This approach allows modifications and extensions of the data structures, even during project execution. This contributes to the flexibility necessary for supporting creative design processes, as mentioned in the beginning of this section.

For reasons of interoperability, the Ontology Web Language (OWL) [2] standard from the Semantic Web approach would be the first choice for the representation of ontologies. However, since current OWL-based ontology repositories do not offer an efficiently searchable storage in relational databases, the RDFS-based KAON system

[15] is used instead. KAON enables semantic queries directly on the repository, at the cost of losing some of the expressiveness of OWL. A converter to translate from OWL (e.g., OntoCAPE) into KAON's RDFS format is currently under development.

3. Using the Process Data Warehouse

In this section, we illustrate the utilization of the PDW in engineering design processes. The chosen application scenario is part of a large case study on the design of a polyamide 6 production facility which has been described in [4]: For the design of the polyamide 6 reactor, the engineer creates a reactor model and performs simulation experiments, from which he or she derives design parameters. Based on these parameters, the reactor equipment (apparatus type, heat transfer equipment, etc.) is specified within a process flow diagram (PFD).

Model file and PFD are represented as instances of the Document Version concept, while their respective contents are annotated by concepts from the OntoCAPE ontology, which refine the Content Description concept: For example, the content of the PFD is described by instantiating the OntoCAPE concept "Pressure Vessel". For a more precise description, Pressure Vessel can be further characterized by supplementary concepts from OntoCAPE: For instance, it could be associated with the ontological concepts Blade Agitator and Heating Jacket.

The activities that the engineer performs in the scenario are represented by instances of Process Trace, which are further characterized by instantiating specializations of Action (e.g., Modeling Action to denote the creation of the reactor model). Appropriate instances of the concepts User and Tool Element describe the human and computer agents involved in the work process. As the model file and PFD mentioned above are produced during the engineering activities, the corresponding Document Versions are associated with the Process Traces. Based on the knowledge representation described above, the PDW can support the following tasks:

Design documentation. The PDW offers a unified access structure for all types of design information. Such documentation is valuable during the design process itself, for example to facilitate communication among distributed development teams. In later stages of the plant lifecycle, this information can also be used for the support of tasks like change management, plant expansion, or claims management. Information can be accessed independently of the original formats and storage locations.

Content-based retrieval of resources. Relying on the content annotations of the resources, content-specific queries can be submitted to the PDW, for instance "*Retrieve all documents that specify a reactor for the polymerization of polyamide 6*". A graphical query editor supports the composition of such queries in a "query-by-example"-based fashion, using concepts, relations, and already known instances from the ontology. Since the semantics of the query terms have been formally defined within the ontology, the computer "understands" the meaning of the query and is therefore able to retrieve the appropriate documents, even if they are represented by different (but semantically equivalent) ontological concepts within the PDW.

Navigation between resources. Content annotation and structural analysis establish semantic connections between products or documents, which can be utilized to retrieve semantically or thematically related resources. In the above scenario, for instance, model file and PFD are connected via their Content Description, as they both represent (different aspects of) the same reactor. Exploring and navigating between the available resources is supported by custom tools like the PDW front-end.

Documentation of organizational context. From the Process Objects associated with the respective resources, the user can gain information about the organizational context (i.e., the work processes and decision making procedures) the resource has been created, used, or modified in. This allows to answer questions such as "*What has this model file been utilized for?*" or "*Which decisions have been taken on the basis of this data?*". Moreover, by analyzing the Process Traces of already accomplished projects, a user can find out who already solved a certain type of problem, and contact this expert directly.

To enhance usability and user acceptance, information acquisition (i.e., the process of getting information into the PDW) must be simple and time-efficient. User interaction can be minimized by means of the following acquisition techniques and tools: Simple document metadata, as commonly stored in a document management system, can be easily derived via the storage area concepts and integrations. Annotating the content of highly-structured and formalized documents with well-known syntax and semantics (databases, xml files, model files, etc.) can be achieved by converters that extract information from the documents and map it to the ontologies of the PDW. For demonstration purposes, two prototypical converters are being developed that automatically transform and annotate model files of the process simulator *Aspen Plus* [1] and flowsheet objects of the CAE system *Comos PT* [6], respectively. Informal text documents need to be annotated by hand. Markup tools like the *OntoMat-Annotizer* [16] or *SemanticWord* [19] support the annotation of such text fragments with ontological concepts via simple drag and drop operations. The tool *TRAMP* [9] has been especially developed for structuring and annotating multimedia content as produced during three-dimensional simulations in plastics engineering.

The traces of coarse-grained work processes can be extracted from project management and other planning tools, or captured with the *Workflow Modeling System WOMS* [11]. For supplementary information like decisions taken and their arguments, specialized tools (e.g., a decision editor) are being developed.

Complementary to project coordination, environments for fine-grained process support at the engineering workplaces can store traced experiences in the PDW and reuse them when demanded by providing experience-based methodical guidance for developers. To this end, the *PRIME process-integrated modeling environment* [17], developed by the authors' research group over the last years, closely interweaves with the PDW. PRIME exploits the process area of the Core Ontology for the definition of guidance and traceability models. Based on the storage area, PRIME's process-integration mechanism offers the potential to couple the engineering tools in a coherent manner and provide high quality support for the user at the same time.

4. Summary and Outlook

The Process Data Warehouse has been developed as an ontology-based repository for the support of design processes. Its key features are (1) flexible data structures and work processes for the support of creative and dynamic work processes, (2) improved retrieval mechanisms based on the use of Semantic Web technologies, and (3) integrated representation of resources, content descriptions, and organizational context to enable experience reuse. Possible application areas have been described, and the possibilities to capture different forms of design knowledge in an integrated manner have been specified. The research described herein has been funded by the German National Science Foundation (DFG) as part of the CRC (SFB) 476 "IMPROVE".

We feel that an evaluation of our approach in an industrial setting will greatly help us to improve our concepts according to real-world requirements and needs. In the near future, we are planning to apply our approach to a large-scale industrial project dealing with continuous production processes for rubber profiles.

References

[1] Aspen Technologies Inc. Aspen Plus. http://www.aspentech.com/, accessed Feb 2006, 2006.

[2] S. Bechhofer, F. van Harmelen, J. Hendler, I. Horrocks, D. McGuiness, P. Patel-Schneider, and L. Stein. OWL Web Ontology Language Reference. http://www.w3.org/TR/owl-ref/, accessed Feb 2006, Feb. 2004.

[3] S. C. Brandt, M. Schlüter, and M. Jarke. A Process Data Warehouse for Tracing and Reuse in Engineering Design Processes. In *Proc. of the Second Internat. Conf. on Innovations in Information Technology*, Dubai, UAE, Sept. 2005.

[4] M. Eggersmann, J. Hackenberg, W. Marquardt, and I. Cameron. Applications of modelling: A case study from process design. In B. Braunschweig and R. Gani, editors, *Software Architecture and Tools for Computer Aided Process Engineering*, pages 335-372. Elsevier Science, 2002.

[5] J. X. Gao, H. Aziz, P. G. Maropoulos, and W. M. Cheung. Application of product data management technologies for enterprise integration. *Intern. Journal of Computer Integrated Manufacturing*, 16(7-8):491–500, 2003.

[6] Innotec GmbH. Comos PT. http://www.innotec.de/en/, accessed Feb 2006, 2006.

[7] M. Jarke, M. Lenzerini, Y. Vassiliou, and P. Vassiliadis. *Fundamentals of Data Warehouses*. Springer-Verlag, 2nd edition, 2003.

[8] M. Jarke, T. List, and J. Köller. The Challenge of Process Data Warehousing. In *Proceedings of the 26th International Conference on Very Large Databases – VLDB*, Cairo, Egypt, 2000.

[9] M. Jarke, M. Miatidis, M. Schlüter, and S. Brandt. Media-Assisted Product and Process Traceability in Supply Chain Engineering. In *37th Hawaii International Conference on System Sciences – HICSS*, Big Island, HI, USA, January 2004.

[10] J. Kopena, W.C. Regli. Functional Modeling of Engineering Designs for the Semantic Web. *IEEE Data Eng. Bull.* 26(4): 55-61, 2003.

[11] LPT, RWTH Aachen. Workflow Modeling System – WOMS. http://www.lpt.rwth-aachen.de/WOMS/, accessed Feb 2006, 2006.

[12] W. Marquard and M. Nagl. Workflow and information centered support of design processes – the IMPROVE perspective. *Computers and Chemical Engineering*, 29:65–82, 2004.

[13] J. Lee and K.-Y. Lai. What's in design rationale? *Human-Computer Interaction*, 6:251-280, 1991.

[14] S. Mesihovic, J. Malmqvist, and P. Pikosz. Product data management system-based support for engineering project management. *Journal of Engineering Design*, 15(4): 389–403, 2004.

[15] D. Oberle, R. Volz, B. Motik, and S. Staab. An extensible ontology software environment. In S. Staab and R. Studer, editors, *Handbook on Ontologies*, chapter III, pages 311–333. Springer, 2004.

[16] OntoMat. Annotation Portal: OntoMat Homepage. http://annotation.semanticweb.org/ontomat/index.html, accessed Feb 2006, 2006.

[17] K. Pohl, K. Weidenhaupt, R. Dömges, P. Haumer, M. Jarke, and R. Klamma. PRIME: Towards Process-Integrated Environments. *ACM Transactions on Software Engineering and Methodology*, 8(4):343–410, 1999.

[18] S. Szykman, R.D. Sriram, C. Bochenek, J.W. Racz, J. Senfaute. Design Repositories: Engineering Design's New Knowledge Base. *IEEE Intelligent Systems*, 15(3): 48-55, 2000.

[19] Teknowledge Corporation. Semantic Word. http://mr.teknowledge.com/daml/SemanticWord/SemanticWord.htm, accessed Feb 2006, 2006.

[20] A. Yang and W. Marquardt. An ontology-based approach to conceptual process modeling. In *European Symposium on Computer Aided Process Engineering – ESCAPE 14*, pages 1159–1164, 2004.