University of Miskolc
Faculty of Mechanical Engineering
Hatvany József Informatikai Tudományok Doktori Iskola

Ontology-Based Modeling and Reuse of Technical Documentation

PhD Thesis

Author:
Tanja Sieber
MSc Electrical Engineering

Supervisor:
Dr. Kovács László

Head of the PhD School:
Prof. Dr. Tóth Tibor

Miskolc, 2008

Jury

Chairman:

- Prof. Dr. Tóth Tibor (ME)

Board members:

- Prof. Dr. Baranyi Péter (BME, MTA-SZTAKI)
- Prof. Dr. Bernhard Bauer (opponent, University of Augsburg)
- Prof. Dr. Cser László (opponent, BCE)
- Prof. Dr. Szigeti Jenö (ME)
- Dr. Kovács Emöd (EKF)
- Dr. Dudás László (ME)
- Dr. Kovács Szilveszter (secretary, ME)

Contents

1 Introduction

Manufacturers and developers of technical products are legally required to provide correct technical documentation. Technical documentation is a summarised term for all data, that a user of a product can be provided with. Technical documentation aims in transferring knowledge to the future reader of the documentation.

During a product life cycle, technical documentation can be split up into two main types: Internal and external technical documentation. *Internal technical documentation* provides necessary data to employees, reflecting the definition and the development of the product, e.g. specifications, documentation of the quality assurance measures, construction documents etc.. *External technical documentation* is directed towards the user, who has to be informed about the correct usage of the product with the help of manuals, operating or safety instructions.

In general, development processes and their related authoring activities are embedded in complex structures and accompanied by the following challenges regarding the technical documentation:

- Distributed Data and Knowledge: Problems occur very often, where dated knowledge has been produced without exactly knowledge about the future potential recipient. The 'one-size-fits-all'-approach is out of place for knowledge exchange situations.
- Diversity of Available Data: Automated processing and reuse of concerned data is therefore highly difficult to realise.
- Gap between Technical Documentation and Knowledge Management: evolved by lacking in process-driven and -supported loop between efficient knowledge management strategies and the transfer in explicited, formal knowledge into the technical documentation.
- Different Understanding and Treatment of Technical Documentation: the processing of internal technical documentation seems to rely on its own specific rules and does not utilise the methodologies widely used in external documentation. In internal technical documentation processes the act of writing the documentation is regarded as a subordinated task of the actual process, e.g. faults diagnosis, usability testing with low focus on writing understandable and re-usable documentation. This attitude of treating technical documentation as mandatory regulated acts is reflected along the product life cycle and results quite often in no or low reasonable support for technical authors working in the field of external technical documentation.
- Missing Reference Model: It seems obvious that ontologies and semantic technologies can help not only to ensure a common vocabulary but also to move towards a new documentation paradigm and to enable a new possible processing of documentation-relevant data. In fact, there is no reference model for technical documentation along a product life cycle available, that provides the evaluation of the several approaches and methodologies.

This thesis combines the concepts *technical documentation*, *ontology-based modeling* and *reuse*.

The driving force for combining the three topics is the idea for the necessary shift in documentation paradigm to understand, formalise and use knowledge in such a way, that it enables

- higher reuse of technical documentation and
- cost-saving approaches for processing technical documentation.

2 Related Work

Rockley (Rockley, 2003) as well as Ziegler (Ziegler, 2005) worked out basic concepts for the reuse of technical documentation. The definition and naming of these concepts influenced strongly the discussions in the area of content engineering. Manufacturers of content management systems have been requested by the defined concepts to enable the implementation of these desired reuse possibilities. Rockley and Ziegler base their concepts on already modularised documents. This is not the case in every kind of technical documentation, especially not in internal. The way how to achieve this desired modularisation is missing within the work of Rockley and Ziegler.

Amongst existing methodologies for structuring technical documentation ((Rockley, 2003), (Ament, 2003), (Wiegand, 1998), (Muthig&Schäflein-Armbruster, 1999 and 2001), (Lobin, 2000), (DITA, 2007a and 2007b), and (Ley, 2006)) the recent is the structured text control (Ley, 2006): Ley proposed this methodology

with application field of the development of aircraft checking procedures. For this application field Ley introduced modeling primitives for direct usage for the required declarations.

Most of the methodologies are based on the functional aspects of texts, and have their great benefit in concentrating in addition on how to write to optimise the understandibility of texts. This is according to (Ziegler, 2006) desired on the top level for linguistical and terminological consistency. Stock (Stock, 2007 focusses in his work on the development of a new document model and its embedding in the process of documentation. This document model was developed within his research project at the BMW Group within the department of AR-based checking procedures.

The aims of the to develop model have been

- modularisation,
- possibility to generate documents according to product instances, and
- 100% accuracy of the documentation.

Stock developed therefore a graph-based document structure that consists causal relations and thereout resulting possible actions. His approach is in creating a all-vehicle-context-containing documentation that can generate at runtime in dependancy of the actual vehicle context the appropriate part of the documentation. Parallel Stock defined a specific product model that splits up the product into so-called functional groups.

Ontologies in scope of this research work are understood as conceptual base for *communicating knowledge*. The objective of technical documentation is given as *transferring knowledge*. In addition the ontological basis - the semiotic triangle - is strongly reflecting the linguistic understanding of the world. Ontologies should therefore constitute a possibility to achieve the objective of technical documentation. Although there is a lot of research in using ontologies for E-learning purposes (cf. (Knight, Gašević, & Richards, 2005)), there is no research interest in using ontologies in the field of technical documentation.

Comparing the methodologies, that are used for structuring technical documentation with those used for the design of ontologies ((Noy&Guinness, 2001), (Pinto&Martens, 2004), (Gómez&Fernández, 2004), (Sure, 2003), (Uschold&Grünninger, 1996), and (Swartout et al, 1996)), there is on the one

hand a lack of common understanding for basic terms like data, knowledge, and information, and on the other hand a gap to the used ontology-relevant terms signs, concepts, and instances.

The content principle as well as ontologies seem not yet to attract interest in the field of technical documentation as well as the consideration of reuse-optimized and cost-effective processing of technical documentation along the whole product development cycle.

3 Research goals

The research work took place at the University of Miskolc in 2006 and 2007 and was supported by SAP AG, department KPS - Knowledge Productization Services in 2006. The work presented in this thesis is mainly the result of research performed within the BAYHOST-founded project ECIDISI and the supported research work at SAP AG. It is higly influenced of practical insights, that I have gained in the last 10 years during my professional experiences as service diagnostic developer at GM Europe GmbH and consultant and trainer in the field of technical documentation.

This research work faces the following research questions:

- How do the concepts ontology and technical documentation rely on?
- What kind of methodologies are qualified for enabling and optimising reuse of technical documentation along a product development cycle?
- How can economical and quality-assurance methods be applied to technical documentation?
- How can the new cognitions be applied in practice? What kind of shift in paradigm will solve what kind of existing challenges?

4 Applied methodologies

In order to close the gap between technical documentation and ontologies as foundation for the new to develop data model signs are understood according to Peirce: signs can exist on three levels. Thus, Peirce differs the following semiotic entities:

- *Indices* - Sign objects have a causal corelation to their meaning, e.g. the causal correlation between the footprints in the sand that a subject perceives and the presumed preceeded event, that a person has been walking here, whose footprints are still visible in the sand.
- *Icons* - Sign objetcs have a natural coherence to their meaning, e.g. an image of a person.
- *Symbols* - Sign objects are arbitrarily correlated to their meaning. This correlation is produced by convention, e.g. linguistic agreements, traffic signs.

This understanding of signs and the embedding into the developped so called semantic data model offered the possibility to create with this underlying basis the reference model for technical documentation - the semantic communication model. In order to evaluate the existing methodologies for technical documentation and ontology design, the Quality Function Deployment (QFD) has been used. QFD centers the concept quality as basic for every design and manufacturing decision (cf. (Akao, 1990)). The fundamental task of QFD is the systematic transfer of customer requirements into product specifications (ASI, 1990). This happens with the help of the so-called Houses-of-Quality (HoQ) as core instrument. According to Herzwurm the QFD can be adapted for the evaluation of software (cf. (Herzwurm, 1997)).

The benefit in transferring the spirit of QFD for evaluation purposes lies

- in the precise capture of requirements, that is forced by the so-called *Voice of Customer Analysis*,
- in the usage of weighting factors for the determined requirements,
- in the usage of an intervall-based proceeding for rating the degree that achieves the evaluated methodology as regards the fulfillment of the requirements, and
- the possibility to extend the developped HoQ for the design of a product that covers the determined requirements in a better way than the evaluated products.

For the development of the to be executed evaluation the method of QFD was applied with requirements weighting factors between 1 and 3 and rating degrees between 1 and 5. The customers are the enterprises (resp. their employees) that are interested in applying that methodology providing optimised reuse pre-requirements.

For the development of the new technical documentation paradigm 'intelligent content syndication', the concept of lean production has been transferred to documentation application. Lean production has been developed by Eiji Toyota and Taiichi Ohno at the Toyota Motor Company after the Second World War and was therefore titled Toyota Production System (TPS) (cf. (Ohno, 1993)). The labour costs are count fixed factor, viz. the production system saves not on employees, but tries to optimise the productivity of employees by the use of several approaches.

The lean production is not a technique by itself: the basis for this production system is a holistic approach consisting of the coaction of several concepts. According to (Traeger, 1994) the lean production can be partitioned into the following principles:

- Kaizen (process of continuos improvement),
- Jidoka (automatisation resp. autonomation),
- Muda (avoidance of waste),
- Just-in-Time (abdication of stock holding),
- Supplier-Relationship,
- Customer-Orientation, and
- Andon (visual control).

The concept of lean production is wide spread and is implemented in transferred areas of software engineering (lean programming), health care (lean healthcare) or military (lean military). Despite the fact, that the concept is implemented worldwide very successfully and according to (Traeger, 1994) easy transferable to other areas e.g. non-productive industries, there are only three publications (Abele, 1993), (Baker, 2005) and (Schott, 2006) identifying the possible mapping to lean documentation.

5 Summary of research work

The contributions of this thesis cover a broad spectrum of aspects related to practical *technical documentation engineering along a product development cycle*. The technical documentation is a relatively new (approx. 15 years) engineering subject and only since 10 years in Germany teached at universities of applied sciences in a very pragmatic way.

The same spirit is recognizable in industrial applications: technical documentation has to be written and is considered as a subordinated task. Mostly no or the next best methodology is used for structuring technical documentation regardless if the methodlogy serves reuse or cost-efficiency requirements.

The driving force for my thesis was therefore my idea for the necessary development of a documentation paradigm that

- serves understanding, formalising and using knowledge

in such a way, that it enables

- higher reuse of technical documentation and
- cost-saving approaches for processing technical documentation.

With respect to these central objectives I provided in my thesis the following new scientific results.

(1) Development of a new reference model for technical documentation

One basic approach of my research thesis has been, that ontologies and existing methodologies for their design promise for use within technical documentation to gain a higher reuse and enable more intelligent processing of technical documentation.

I compared the methodologies, that are used for structuring technical documentation with those used for the design of ontologies, and showed that there is on the one hand a lack of common understanding for basic terms like data, knowledge, and information, and on the other hand a gap to the used ontology-relevant terms sign, concept, and instance. As a consequence there exists no clearance, if and how ontologies and/or their design methodologies can serve technical documentation purposes.

I analysed terminological statements/definitions and models for the concepts sign, data, information and knowledge and developed with the gained cognitions the novel semantic data model for closing the gap between data and semiotics. Based on the underlying understanding of signs in terms of Peirce I gave definitions for data and the appropriate terms within this novel model.

I extended this model in two directions in order to achieve a clear identification and solution for the following issues:

- In the first extension – the multi-layer semantic data model - I introduced a semantic network based on a lattice of concepts. This was addressing the integration of conceptual dated knowledge, i.e. ontologies.
- In the second extension – the semantic communication model I addressed the integration of basic communication theories. I transferred therefore Austins speech-act theory into a write-act, that takes place in technical documentation. I analysed several communication theories and integrated and transferred the concepts of message and noise into my model.

This novel semantic communication model constitutes a reference model for technical documentation engineering.

Thesis 1: *I have developed a new reference model for technical documentation that provides a consistent term-system based on the understanding of sign in terms of Peirce. The model integrates herewith data and semiotics. I have created an multi-layer extension of that and proved a more flexible modelling of concepts. I have worked out the extension of this multi-layer model in integrating the model into a communicative frame, that represents technical documentation as a communicative act.*

(2) Development of a new evaluation method for methodologies

I transferred the method QFD (Quality function deployment) developped by Akao and implemented it for the creation of an evaluation matrix for methodologies for ontology-design and technical documentation with respect to the optimisation of reuse. I gained therewith a supporting methodic approach for the definition of the requirement criteria.

The criteria for reuse I evolved from my novel semantic communication model. This model reflects the communicative frame in which each kind of technical documentation is embedded.

I extended the obtained evaluation matrix in form of an guideline implementing textual explanations for the to give points. For the application of the methodologies I assumed a special use case and applied the new evaluation method for the evaluation of methodologies for structuring technical documentation and for ontology-design. I could show this way, that none of the analysed and evaluated methodologies fulfills the requirements concerning reuse in such a way, that optimised reuse possibilities can be gained. In fact, the possibility of personalisation as a specialisation of reuse is lacking in all evaluated methodologies.

Thesis 2: *I transferred the method QFD (Quality function deployment) and implemented it for the creation of my novel evaluation method for methodologies for ontology-design and technical documentation. I developed the requirement criteria for this method with respect to the optimisation of reuse.*

(3) Development of a new lean documentation model

The novel understanding of technical documentation and actual lack of reuse focussing and optimisation along the product development cycle lead to my approach to transfer the concept lean production to the domain of technical documentation.

I worked out the basic concepts of lean production and transferred the measures supporting these concepts on technical documentation. I integrated the to be taken measures into into the novel model for technical documentation and its engineering – the lean documentation.

I analysed further the transaction cost theory with the result, that transferred to technical documentation the consideration of transaction costs demand precise descriptions of the desired process chain of documentation. This is actually lacking in practice.

I developed therewith – by this novel model with the background of the transaction cost theory – the holistic approach to optimise the productivity of employees who are participating in the chain of technical documentation processes along the product development cycle.

Thesis 3: *I developed a new lean documentation model based on the concepts and measures of lean production. I analysed the applicability of transaction cost theory for technical documentation.*

(4) Application and integration of the results into the intelligent content syndication paradigm

I used the developed semantic communication model and the novel evaluation method for the analysis of reuse potentials at the SAP AG within the departement KPS (Knowledge Productization Services) and along the Product Innovation Lifecycle (PIL). I applied the holistic approach for lean documentation in the analysis as well as in the design of reuse and cost-efficiency optimiziations at the SAP AG and at GM Europe GmbH in the Service Technology Group department.

The lessons learned are (in a nutshell):

- Different processes drive technical documentation and knowledge management projects, but "human issues" might dominate other ones.
- Guidelines for involved persons in industrial context have to be pragmatical, otherwise they are unlikely to be understood and to be used at all.
- The existing boundaries between departments shouldn't be underestimated.
- The more linguistic background the involved persons have, the more difficult is standardising work.
- Improvements to be implemented should be achievable in several loops.

The developed model of lean documentation served well in this practical application as it is not a stringent framework but a conceptual holistic approach offering the possibility to work out measures for the application domain. I have done this in both cases on the one hand by introducing paradigm shifts in editorial organisations and on the other hand in operational ones. The thereout gained improvements enable horizontal as well as vertical integration and syndication of intelligent dated knowledge. The reorganisations are partly implemented and show promising effects for cost-efficient processing of technical documentation.

Thesis 4: *The new developed models provide an efficient holistic paradigm to improve analysis and implementation of existing technical documentation along product development cycle. The new paradigm of intelligent content syndication considers therewith both, internal and external technical documentation. In the practical application I worked out several new measures and splitted them up into novel paradigm shifts in editorial and operational reorganisations.*

6 Applicability of results

I finally conclude by presenting some impact that this work already has on current (research) projects combined with a future outlook on the upcoming research agenda for ontology-based processing of technical documentation.

Semantics@PIL: Based on the research results and the new developed documentation paradigm of intelligent content syndication, modelling primitives for documentation processes and their characteristics within technical documentation along the Product Innovation Lifecycle at SAP are developed. It is desired that the thereby resulting elements possess common behaviours, interfaces and interaction patterns that enable their enterprise wide management.

7 Further research directions

ORQE - Ontology- and Rule based Quality Engineering: The ORQE project focusses on the analysis of (1) ontology-based enrichements of internal technical documentation that is provided in the field of quality ensuring measures of a product development cycle and (2) the therout resulting possibilities of automised further processing of documentation modules for succeeding documentation purposes and self generating diagnosability of documented products/services.

ROMTeDoc - Development of an Re-Use Optimized Methodology for Technical Documentation: The ROMTeDoc project addresses the further development of the Re-Use Function Deployment. It aims in utilizing the Quality Function Deployment for anchoring the requirements for Re-Use optimizations within technical documentation processes. In contrast to the presented research work, where QFD has been used in an adapted manner for the evaluation of methodologies, ROMTeDoc aims in the deployment of measures and metrics which are meeting the requirements.

8 References

Akao, Y. (Ed.). (1990). Quality Function Deployment QFD: Integrating Customer Requirements into Product Design. New York: Productivity Press.

Ament, K. (2003). Single sourcing: Building modular documentation. Norwich, NY: William Andrew.

ASI (American Supplier Institute) (1990): Quality Function Deployment – Excerpts from the Implementation Manual for Three Day QFD Workshop. Version 3.4. In: QFD-Insitiute (Ed.): Transactions from the Second Symposium on Quality Function Deployment. Novi, Michigan, pp. 21-85.

Austin, J. L. (1962). How to do things with words: The William James lectures delivered at Harvard Univ. in 1955. Cambridge, Mass.: Harvard U.P.

Bicknell, B. A. (1995). The road map to repeatable success: Using QFD to implement change. Boca Raton: CRC Press.

Brassard, M. (1996). The memory jogger plus+: Featuring the seven management and planning tools (rev. ed). Salem, NH: GOAL/QPC.

DITA (2007a). DITA from A to B: helping you get started with topic-based structured writing, from http://www.ditausers.com.

DITA (2007b). DITA Infocenter: A searchable knowledge base of specifications for DITA Users, from http://www.ditainfocenter.com.

Erikkson, I.& McFadden, F. (1993): Quality Function Deployment: a tool to improve software quality. In: Information & Software Technology. Nr. 9, pp. 491-498.

Herzwurm, G. (1997). Kundenorientierte Evaluierung von Software-Tools zur Unterstützung von quality function deployment. Studien zur Systementwicklung. Köln: Univ. Lehrstuhl für Wirtschaftsinformatik Systementwicklung.

Knight, C., Gaševic, D., & Richards, G. (2005). Ontologies to integrate learning design and learning content. In Tattersall Colin, & Koper Rob (Eds.), Journal of Interactive Media in Education.

Ley, M. (2006a). Kontrollierte Textstrukturen.: Ein (linguistisches) Informationsmodell für die Technische Kommunikation. Dissertation, Justus-Liebig-Universität. Giessen.

Ley, M. (2006b). Aspekte der Informationsstrukturierung: Über Strukturierungsprinzipien, die Ebenen der Textstruktur und Dokumentgrammatiken. technische Kommunikation, 28(4/06), 51–53.

Lobin, Henning (2000): Informationsmodellierung in XML und SGML. Berlin; Heidelberg; New York; Barcelona; Budapest; Hongkong; London; Mailand; Pairs; Singapur; Tokio: Springer.

Muthig, J., & Schäflein-Armbruster, R. (1995ff). Funktionsdesign: eine universelle und flexible Standardisierungstechnik. In C. Felkner, & W. Sturz (Eds.), Technische Dokumentation – wirtschaftlich organisieren, systematisch erstellen, kundengerecht gestalten. Augsburg.

Noy, N. F., & McGuinness, D. L. (2001). Ontology Development 101: A guide to Creating your first ontology.

Ohno, T. (1993). Das Toyota-Produktionssystem. Frankfurt/Main: Campus-Verl.

Pinto, H. Sofia, & Martins, J. (2004). Ontologies: How can they be built? Knowledge and Information Systems.

Rockley, A. (2003). Managing enterprise content: A unified content strategy (1. ed.). Indianapolis, Ind.: New Riders.

Schott, G. (2006). Entwicklung eines Lean Documentation Konzepts zur Effizienzanalyse und Optimierung der Softwaredokumentation bei der SAP AG. HS Karlsruhe. Karlsruhe.

Sure, Y. (Mai 2003). Methodology, Tools & Case Studies for Ontology based Knowledge Management. Universität Friedericiana zu Karlsruhe. Karlsruhe.

Swartout, B., Patil, R., Knight, K., & Russ, T. (1996). Towards Distributed Use of Large-Scale Ontologies. KAW.

Uschold, M., & Grünninger, M. (1996). Ontologies: Principles, Methods and Applications. Knowledge Engineering.

Wiegand, H. Ernst. (1989). Aspekte der Makrostruktur im allgemeinen einsprachigen Wörterbuch: alphabetische Anordnungsformen und ihre Probleme. In F. Josef Hausmann, A. Burkhardt, G. Ungeheuer, H. Ernst Wiegand, H. Steger, & K. Brinker (Eds.), Vol. Halbbd. 1. Wörterbücher. Ein internationales Handbuch zur Lexikographie (pp. 371–409). Berlin: de Gruyter.

Ziegler, W. (2006). Content Management -Grundlagen von Redaktionssystemen: Vorlesungsmanuskript zur Veranstaltung Content Management, HS Karlsruhe. HS Karlsruhe. Karlsruhe.

9 Publications

Sieber, T. (1999). Funktion von Wissen im Kommunikationsprozess. Seminararbeit, FH Karlsruhe. Karlsruhe.

Sieber, T. (2002). SGML/XML in der Technischen Redaktion. Diplomarbeit, FH Karlsruhe. Karlsruhe.

Sieber, T. (2005). Ontologies and their representation. Doktoranduszok Fóruma. Miskolc: University of Miskolc.

Sieber, T., & Kovács, L. (2005a). Functional Design and Information Modelling as Methods to Structure Published Content. In University of Miskolc (Ed.), 5th International Conf. of PhD Students (pp. 171–175).

Sieber, T., & Kovács, L. (2005b). Technical Documentation:Terms, problems and challenges in managing data, information and knowledge. In University of Miskolc (Ed.), 5th International Conf. of PhD Students (pp. 165–170).

Sieber, T., & Kovács, L. (2005c). Mapping XML Documents into Object-Model. In University of Budapest (Ed.), 6th Internat. Symp. of Hungarian Researchers on Computational Intelligence (pp. 343–354). Budapest.

Sieber, T., & Bartz, W. (2006). Ich weiß, was Sie meinen!: Semantic Web – Kommunikation auf gleichem Niveau. technische Kommunikation, 28(2), 56–59.

Sieber, T., & Kammerer, M. (2006). Sind Metadaten bessere Daten?: Metadaten als Mittler zwischen Daten und Prozessen. technische Kommunikation, 5(05/2006), 56–58.

Sieber, T., & Kovács, L. (2006). Unbemerkt und unsichtbar: Metadaten als Triebfeder für intelligente Content-Automatisierung. technische Kommunikation, (06/2006), 42–45.

Sieber, T., & Chen, W. (2007). Intelligente Produkte finden ihre Fehler selbst.: Einsatz von Ontologien in der Produktentwicklung. technische Kommunikation, 6, 28–31.

Sieber, T., & Kammerer, M. (2007a). Daten, Wissen und Information.: Eine Grundlagenanalyse unter besonderer Berücksichtigung der technischen Dokumentation. Unpublished manuscript.

Sieber, T., & Kammerer, M. (2007b). Bridging the gap between data and semiotics: semantic data model. International Journal of Teaching and Case Studies (IJTCS).

Sieber, T., Kovács, L., & Kammerer, M. (2007). Technical Documentation – a Roundtrip from Data to Meaning. Proceedings of 1st European Semantic Technology Conference (ESTC). Vienna.

Sieber, T., & Kovács, L. (2007). Mulit-layer semantic data model. Unpublished manuscript.

Sieber, T., & Lautenbacher, F. (2007). Enterprise Content Integration: Documentation, Implementation and Syndication using Intelligent Metadata (ECIDISI): 2007 -17. Technischer Report, Universität Augsburg. Augsburg. Retrieved February 4th 2008, from http://www.informatik.uni-augsburg.de/lehrstuehle/swt/vs/publikationen/reports/2007-17/.

Ált. Inf. 2008-334 ME.