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

Acknowledgements

C’est une grande folie de vouloir être sage tout seul. – It is a great folly to wish to be wise alone.

Francois de La Rochefoucauld, 1613-1680

This thesis is the result of my work as a researcher within the KPS (Knowledge Productization and Services) department at SAP AG in Walldorf, Germany and my academical time as a PhD student at the Hatvany József Informatikai Tudományok Doktori Iskola of the Faculty of Mechanical Engineering at the University of Miskolc, Hungary. Many people supported me and made the successful completion of this thesis possible.

First I would like to express my gratitude to my advisors, Dr. Matthias Kammerer (SAP AG) and Dr. Kovács László (Institute of Information Sciences, University of Miskolc). Their goal-oriented management allowed freedom in my research activities and at the same time gave me a strong support. Their trust in my abilities greatly enhanced my performance in my research projects I worked in. The discussions with them were always founded in an exchange of critical opinions above thesis-relevant terms and I felt very happy to have the possibility to work within this research work in a multidisciplinary area. This multidisciplinarity was reflected in the supervising representants – on the one hand computer linguist Dr. Matthias Kammerer and on the other hand Dr. Kovács László, mathematic and physician.

I would like to thank Dr. Thilo Buchholz (SAP AG) who was responsible for giving me the possibility to be sponsored for one year as a PhD student within SAP AG.

With Florian Lautenbacher (University of Augsburg, Germany) I shared the workload in ECIDISI (BAYHOST-founded project). While I typically covered the linguistic and documentation-oriented aspects, he actually gave input and ideas from the software-engineering point of view.

I would also like to thank all professors from the University of Miskolc, who were enabling me to take part at their lectures holding them in English, especially Prof. Tóth Tibor – the leader of the PhD school.

My ex-colleague Anke Sonemann (GM Europe GmbH, Germany) allowed me to use GM documentation as use cases for this thesis.

My husband, Gustav Sieber (BOSCH Powertools Kft., Hungary) provided one important stone on which this thesis was built, viz. the baseline understanding of QFD as a quality-assurance method along product development cycle in the field of mechanical engineering.

Bernd Schultheiss (TransSmart Consulting, USA) was always open to explain me details of Transaction Cost Theory. He deserves special gratitude for teaching me an irreplaceable "coolness in tricky business situations" and understanding the economic aspects of systems and processes. Especially during the last days of the thesis he supported me in proof reading the English parts of this thesis.

My brother Michael Köder (gematik, Germany) was always a great help and especially during the last weeks of this thesis he strongly supported me in proof reading the English parts of this thesis and in reunion with Dr. Sascha Schröder (POW AG, Germany) in keeping my personal writing motivation on the required level.

My parents, Manfred and Monika, my brothers, Michael and Manfred, my husband Gustav and my two sons, Tim and Marco receive my deepest gratitude and love for the many years of support, love and understanding that encouraged me and, in the end, made this thesis possible.

May 2008, Miskolc

Tanja Sieber

Contents

Part I Foundations

If we would understand history in a much better way, we could find at the beginning of every improvement a huge knowledge.

Émile Mâle, 1862-1954

1 Introduction & Overview

1.1 Introduction

Idea of this thesis

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.

The topics technical documentation and actual challenges in re-using technical documentation are now briefly introduced to motivate the idea and area of interest of this thesis.

Technical Documentation

Manufacturers and developers of technical products are legally required to provide correct technical documentation.

Definition *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.

Definition *Internal 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.

Definition *External Technical Documentation*

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:** The concerned knowledge has to be shared between many people working in different departments, different production locations, different countries and different applications. Because of this, in such a process nearly every-one has two roles: one of owning and dating knowledge (*sender*) and one of searching and needing dated knowledge (*recipient*). Problems occur very often, where dated knowledge has been produced without exact knowledge about the future potential recipient.
- **Diversity of Available Data:** There exists a great diversity of different documentations. Variants exist not only in relation with different products and their releases but also in relation with different intentional groups (e.g. detailed developer documentation versus easy-readable user documentation). Analysing the data, which is produced from the several different sources, its granularity, structure and standardisation is different depending on several influencing parameters such as: the authoring system, the modelling methodology, that has been used for structuring the data, the possibility of automated document generation (e.g.

JavaDoc, DoxyGen in software engineering), and the inconsistent syntactically and semantically usage of data and metadata. Automated processing and reuse of concerned data is therefore highly difficult to realise.

- **Gap between Technical Documentation and Knowledge Management:** Experience and knowledge of several departments are not automatically combined: If experience is directly reabsorbed into the process area and is available to all those involved, this means a quantum leap for quality and efficiency. Actually, there is a lack of a process-driven and -supported loop between efficient knowledge management strategies and the transfer in explicated, formal knowledge into the technical documentation.
- **Different Understanding and Treatment of Technical Documentation:** To make the communication between sender and recipient of success, in the field of external technical documentation some linguistic methodologies for structuring the documents are used and embedded in technical authoring systems. It is astonishing to find that 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 external technical documentation processes, Content Management Systems and modularisation are widespread, a lot of internal technical documentation is still written as a whole document. 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 act 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. The interest of manufacturers in ontology-based enabled content reuse is high, but there are questions concerning the cost and efficiency of using new (and also existing) modelling methodologies. The challenges concerning the application of semantic technologies in the area of technical documentation in general include, if and how it does fit to existing used modelling methodologies, how it can be integrated to optimise reuse, what exactly are the benefits of ontologies and how they can be created or transformed out of existing data respectively metadata. 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.

1.2 Object of Investigation of This Thesis

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 influenced of practical insights, which the author has gained in the last 10 years during her professional experience as service diagnostic developer at GM Europe GmbH and consultant and trainer in the field of technical documentation.

Central Questions

This research work faces the following research questions:

1. **How do the concepts ontology and technical documentation rely on?**
   1.1 What are the relevant terms and concepts and how do we have to understand them in scope of the research work?
   1.2 What kind of model does represent this understanding of the basic terms and concepts the best?
   1.3 How can this model be integrated, adopted and extended to represent the communicational situation of technical documentation?
2. **What kind of methodology is qualified for enabling and optimising reuse of technical documentation along a product development cycle?**
   2.1 What methodologies do already exist and how qualified are they?
   2.2 How can we evaluate the methodologies?
3. **How can economical and quality-assurance methods be applied to technical documentation?**
   3.1 What kind of problem is tackled by those methods?
   3.2 How can these methods be transferred to the field of technical documentation?
4. **How can the new cognitions be applied in practice?**
   4.1 What kind of shift in paradigm will solve what kind of existing challenges?

1.3 Reader's Guide

As follows, the four main parts of the thesis are outlined.

Part I

In Part I of the thesis those basic concepts are introduced and analysed, that are necessary as foundations for the research work and the thesis:

- *Internal and external technical documentation*, their embedding in a usual product life cycle and basic methods, which are used for modeling and structuring technical documents.
- *Reuse* and its application in internal/external technical documentation.
- *Ontologies* and methodologies for their design.

Part II

In Part II the development of the semantic communication model is introduced. Therefore a summarised terminological analysis of *signs* (semiotics), *data*, *information* and *knowledge* is given. Based on the analytical work a differing understanding of data is developed and illustrated.

This is done by

- presenting the *semantic data model* (Sieber/Kammerer), that bridges the existing gap between semiotics and data,
- deriving a more precise understanding of the relativity of metadata,
- concretising the representation of semantic data using the *multi-layer semantic data model* (Sieber/Kovács), that aims in defining clear mappings of concepts, and
- showing the *semantic communication model* (Sieber/Lautenbacher), that is used to introduce the term *semantic noise* in the context of understanding documentation as a write-act.

Part III

Part III aims in introducing one basic method applied in the field of quality- and requirements engineering, the so-called *Quality Function Deployment* (QFD) as method for defining and setting up customer requirements into measurable qualities of the product.

This analysed method is transferred and adopted in order to create a matrix for the evaluation of modeling methodologies regarding the desired optimisation of reuse.

Part IV

In Part IV of the thesis a new paradigm of technical documentation is presented by showing the worked out transfer of already existing concepts of other application areas (mechanical engineering, controlling) into the technical documentation area.

The analysed concepts are *Lean Production* and *Transaction Cost Theory*:

- In the thesis is shown, that and how the spirit of Lean Production can be transferred for technical documentation.
- Using the ideas and concepts of Transaction Cost Theory, technical documentation has to be regarded as a process.

Based on the transferred ideas a conceptual framework for *Intelligent Content Syndication* is proposed in order to optimise the reuse and cost-efficiency within technical documentation processes.

2 Foundations: Technical Documentation

This chapter starts by approaching the concept *product development cycle* and introduces the Product Innovation Lifecycle as realised at SAP. The hereby followed Quality Gates strategy demands several documents on the points of transition between processes. Section 2.2 illustrates how diverse the resultant sorts of *technical documentation* are. To provide an insight into this diversity of technical documentation, it is referred to Appendix D, where some examples of internal and external technical documentation have been collected, which have been made available by SAP AG and GM Europe GmbH. In Section 2.3, it is argued for a shift from linear oriented writing to modular oriented writing having regard to more effective reuse strategies. Section 2.4 depicts the principles of modeling methodologies and Section 2.5 summarises the research analysis work of several widespread methodologies.

2.1 Product Development Cycle

Definition *Product Life Cycle*

In an economical sense, a product life cycle is understood as the period from launching a product on the market until taking it off the market again: A product "lives" as long as it gains economical turnover on the market.

For research work, the field of product development is much more interesting.

Definition *Product Development Cycle*

The product development cycle means the process a good has to pass thorough from the idea until the saleable product.

The phase of product development is named in different companies with different terms, for example at SAP AG (SAP AG, 2006e) the phase is named *Product Innovation Lifecycle (PIL).* The phase contains - according to the complexity of the realisation of a product out of an idea - several processes. The process owners and the process actors may be spread all over the company (and in an outsourced process even outside of the company).

2.2 Product Innovation Lifecycle (SAP AG)

At SAP AG (cf. (SAP AG, 2006e)), the Product Innovation Lifecycle (cp. Figure 1) is divided up into the phases:

- Invent,
- Define,

- Develop,
- Deploy and
- Optimize

.

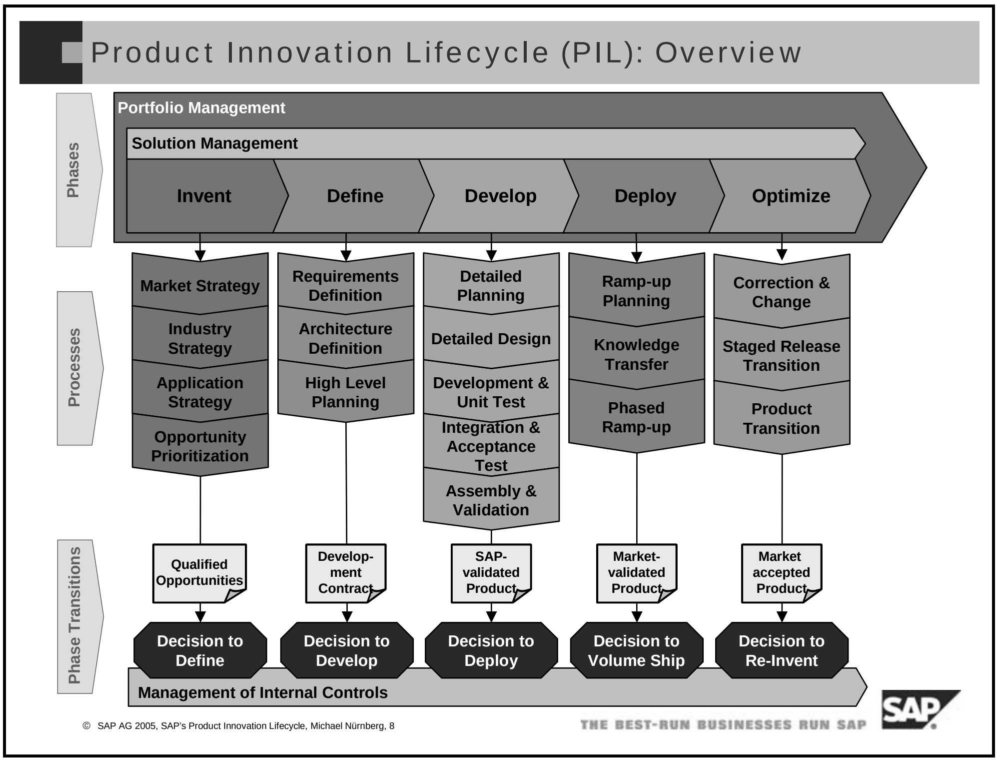


Figure 1 SAP AG: Product Innovation Lifecycle - Overview

The several phases contain diverse processes; different departments of the company, external service providers and other companies share their fulfilment.

Quality Gates

So-called *Quality Gates* host the transits in between two phases .

Definition *Quality Gates*

„Quality Gates are formal meetings, which facilitate decisions to advance to a milestone, based on evaluation of deliverables and assessment of readiness. Readiness is proven with defined deliverables on product and process compliance. The Quality Gate meeting focuses on decision proposals for critical issues and must come to an explicit decision. A decision to advance is final. Quality Gates are used to manage phase transitions as well as some important transitions“

(Müller (SAP AG), 2006)

According to the above, Quality Gates (SAP AG, 2006e & 2006h) are defined as internal control points of a standardised Product Innovation Lifecycle. They should guarantee and pinpoint a match with the process standards.

Role of Technical Documentation

Based on Quality Gates, decisions are taken if the process can be passed into the next phase of the PIL.

Definition *Deliverables*

The effected work in the phases respectively in processes and sub-processes is reflected in documented form in so-called deliverables, which need to exist at the different Quality Gates. In most cases, it will be technical documentations, which will be held at the disposal for the following processes resp. their owners and/or they must be designed as documentation for the running process.

In the context of the development phase, it will be desktop text or scenario description for example (cf. (Müller (SAP AG), 2006), (SAP AG, 2006b, 2006e & 2006f)). The therefore to develop documents/data focus on different target groups (SAP AG, 2006b) and represent - according to the involved persons, who work on the engineering of the deliverables - different views on the future product.

Reuse of Deliverables

One deliverable of a sub-process is hereby used en bloc as input for another sub-process (cp. (SAP AG, 2005d, 2006b, 2006e, 2006f & 2006h)). Because of the often very differencing formatting and objective target of the deliverables and the various granularity the use is not in that kind, that a real 1:1 reuse can take place. It is a "setting to proposal" of documents, which can be prepared further on depending on the process or tool support. The same knowledge will be found redundant in several parts within several documents, without being sourced via one common basic source (the optimum from the original knowledge-source). More often, they are changed within the process of development (and therewith possibly being falsified).

2.3 Principles for Writing Technical Documentation

The above given practical insight from the creation of internal technical documentation and missing reuse strategies are not only to be found at SAP AG: Approaches to grip knowledge-contents in a more granular way and to build up documentations more modular have been increasingly implemented in the last few years by the quick development and distribution of so called Content Management Systems.

Linear vs. Modular Writing

Definition *Linear Writing*

Linear writing means the sequential writing of a document. This can be illustrated the best as writing a book.

Linear development of documents is characterized by a hierarchic and sequential approach. Content management and the inventing of a fitting system brings up changes in the development process of documentations: With a Content Management System, documents are no longer hold as enclosed units. The content itself is administrated disconnected from layout and structure. This separation allows a more flexible reuse of the contents, for example for different documentation types, target groups or media (cf. (Rockley, 2003), (Ziegler, 2005&2006) & (Sieber&Lautenbacher, 2007)).

Central methods of the content management are:

- Single Sourcing and Cross Media Publishing,
- modular writing and
- separation of layout, content and structure.

Definition *Modular Writing*

Modular writing means the establishment of standing alone content modules, which can be used and make sense for different document layouts and reading sequences.

Therefore, all the efforts in context of the modular writing aim in increasing the reuse of those documentation-bricks. For enabling this intentioned reuse, the documentation-bricks have to be embedded into standardisation activities.

Reuse by Standardisation

For the standardisation of documentation-bricks, a rule-based acquisition of data is indispensable. Ziegler (Ziegler, 2006) sees the following possibilities of a XML-based content standardisation on five layers (see Figure 2).

.

Guideline, Term- and Textcontrol	Writing Rules
Methodology + Classification	Modeling Rules
Methodology + DTD/XSD	Information Model
Editors + XML	Data Format
Editors + Unicode	Character Format

Figure 2 Standardisation according to (Ziegler, 2006)

While the both bottom layers create syntactic requirements for the reuse, the other continuative layers are necessary for the right and correct reuse of the documentation-bricks:

- The *information model* is building therefore the basis for a structured and modular acquisition of information. It is composed of the chosen structure methodology and their illustration for example in XML with a DTD or a XML-schema.
- Building up on the chosen information model, the rules for the modularisation can be defined. Through the *classification of the modules* (e.g. via annotations) the relocating and therewith the possibility of an efficient reuse is guaranteed.
- The *writing rules* assure linguistical and terminological consistency.

2.4 Principles for Structuring Technical Documentation

According to Figure 2 (Ziegler, 2006) a central role for flexible reuse of document-bricks based on single-source-publishing are therefore the modularisation and standardisation of documentations. Depending on the editorial and technological concept as well as the chosen information model, the definition of the module dimension may range enormous. Concerning the choice of the module granularity, it must be balanced in between the acceptable redundancy on the one hand and the desired flexibility in the reuse and the therewith-combined administrative effort on the other hand.

The modularisation of information products can ground on the results of a bottom-up-analysis or been provided via top-down-methodologies (Ziegler, 2006). As two possible principles for structuring information Ley describes in (Ley, 2006) the *functional principle* and the *content principle*.

Functional and Content Principle

Definition *Functional Principle*

The functional principle assumes that bricks of documentation serve specific functions within a document. These functions can be context-dependent.

Definition *Content Principle*

The content principle means that bricks of documentation may have a relation to objects in the world that they rely on.

Single parts of a document do not always have easy identifiable functions like action-instructive characters. E.g., explaining or describing parts do have the function, to describe or explain *something*.

One method for describing these objects of a certain domain in a systematic way is the description using ontologies. According to (Ley, 2006) the structuring of a domain regarding the content can start with general concepts like abstract objects, concrete objects, qualities, quantities or situations. With increasing descriptive depth, the concepts become more specific until the domain is completely (or for the intended purpose sufficiently) described. The 'something' that has to be explained or described in the information product can be identified that way systematically.

In the following section, methodologies for structuring technical documentation are introduced and analysed, which of them use what kind of structuring principle.

2.5 Methodologies for Structuring Technical Documentation

In (Sieber&Lautenbacher, 2007) the research analysis of methodologies for modularised document structuring is presented. Summarised the most important cognitions of this analysis work are presented regarding the reuse orientation of the analysed methodologies. Some of these methodologies will be further focused in the evaluation in Part III.

(1) Rockley: Content Audit

> *„During a content audit, you look at your organization's content analytically and critically, so that you can identify opportunities for reuse and the type of reuse.“*
>
> (Rockley, 2003)

According to (Rockley, 2003), it is possible within the content audit to identify reusable contents by performing a bottom-up-analysis and comparing different documents. The relations between the contents have to be displayed in a so-called *Reuse Map*. The content audit results in this Reuse Map and gives no further instructions how to create reuse-optimised modules.

(2) Ament: Single Sourcing

> *„Single Sourcing is a documentation method that enables you to re-use the information you develop.[…] Single Sourcing is a methodology, not a technology.“*
>
> (Ament, 2003)

According to (Ament, 2003) modular writing afflicts in three actions:

- *Chunking,*
- *labeling* and
- *linking.*

Chunking means thereby the partition of information into standing alone modules based on several so-called information types. These modules have to be annotated by labeling in a unified, standardised and context-independent way. The last action - linking - combines the annotated modules by references. Ament emphasis more on the style how to write in a modular way than on giving a formal methodology how to construct the modules in a reuse-efficient way.

(3) Wiegand: Function-Positional Spacing

Wiegand describes in (Wiegand, 1998) how a structural analysis of dictionary articles can be performed. Therefore, he developed a segmentation method, which consists of the method of functional segmentation and the method of function-positional segmentation. The functional segmentation of dictionary articles includes the identification of functional text elements, which is also interesting for modeling all kinds of documents and henceforth for technical documents.

The functional segmentation determines typographical and non-typographical structure pointers, which assist the user in perceiving the structure of the article and of parts of an article that belong together. Typical typographical structure pointers are the font type and the font styles. Examples for non-typographical structure pointers are punctuation marks, brackets and arrows. The exhaustive functional text-segmentation is described as the incremental segmentation of a dictionary article considering statements and structure pointers together with a presentation as whole-part-relationships in a formally defined description language. This leads to a segmentation of the text where all elements are defined and can be associated to be part of a bigger context. On the other hand, the functional-positional segmentation does not only force the segmentation of the text, but also includes the position of all text segments in a linear order. The determined hierarchical structure of articles can then be presented as a tree-like partitive structure graph. This methodology is suitable for a bottom-up analysis of already existing documents. Its drawback is that it instructs not how to model documents or modules in an optimised way for gaining higher reuse effects.

(4) Schäflein-Armbruster/Muthig: Function design™

The function design™ (cf. (Muthig&Schäflein-Armbruster, 1999 and 2001)) method is a universal and flexible technology for structuring technical documentation. It has been developed based on theoretical thoughts upon the speech-act: Each spoken sentence contains information and serves a communicative function. Nevertheless, there is no clear 1:1 correlation between a sentence and its communicative function. Most of the times the meaning of a sentence is transported using meta-spoken ways such as pronunciation. This must be achieved differently in printed documents.

Based on that, the goal of the function design™ theory is that each sentence of a text needs to have a unique identifiable function. The function design™ method can be divided in macro and micro level: on the macro level, the kind of document is classified and on the micro level, this document is further segmented in sequence patterns, functional entities and tags. The developer must put himself into the position of the receiver of a text. The intentions and existing knowledge of the target group must be kept in mind during the whole specification process. Therefore, the data in a document can be captured in so-called functional entities that must answer the questions that a potential reader might be thinking of. Each functional entity is described with the following declarations:

- Name and (optional) abbreviation,
- purpose,
- sequential order,
- inner structure and formulation, and
- layout and design.

Basic functional entities can be grouped into structural patterns, which have a common purpose or function. This fosters already reuse of data in the function design™.

(5) Horn: Information Mapping™

The information mapping™ technique is based on knowledge from the cognitive-, medial- and learning- psychology. The focus thereby is the transformation of information into perception-optimised pieces and structures of information. The methodology could not further be analysed as there is nearly no material exempt from charges available. Horn developed based on the psychological cognitions different kind of information types with special characteristics. This shows promise regarding reuse benefits.

(6) Lobin: Textual Information Modeling

Lobin describes in (Lobin, 2000), how SGML and XML can be utilised for modeling information along textual structures. The idea of structured information is achieved by the following observations:

- Each text contains concrete (e.g. the sequence of letters) and abstract (e.g. categories like heading) entities. The classification of the several differing abstract entities happens with the help of specific presentation patterns, e.g. typographical accentuations.
- The composition of these abstract and concrete entities is not arbitrary, but follows fixed rules, which are defining the hierarchical relation between abstract entities and subordered abstract and concrete entities and the linear relation of coequal entities. These rules can be integrated into a so-called grammar.

Lobin focuses on the presentation of the grammar and the hereon-resulting texts in form of SGML/XML-DTD/Schema and instances. There is some impact on the optimised modeling of documentation.

(7) OASIS: DITA

The Darwin Information Typing Architecture (DITA) v1.0 has been approved as an OASIS Standard in 2005. DITA (DITA, 2007a and 2007b) is an XML-based specification for modular and extensible topic-based information. DITA provides a model for defining and processing new information types as specialisations of existing types. DITA populates the model with an extensible hierarchy of standard types and encourages reuse by reference either of topics or of fragments of topics.

Through use of a common specification, DITA content owners can benefit from industry support, interoperability, and reuse of community contributions. At the same time, through specialisation, content owners can address the specific requirements of their business or industry.

(8) Ley: Structured Text Control

Ley developed this structuring methodology with application field of the development of aircraft checking procedures (cf. (Ley, 2006)).

The methodology of structured text control is characterized by

- the definition of text constituents,
- the declaration for each of these text constituents of
  - the illocutive basic type,
  - the semantic structure,
  - the syntactic structure and
- the declaration of relations between the constituents.

For the above-mentioned application field Ley introduced modeling primitives for direct usage for the required declarations.

Ley’s methodology is based on speech-act based theories and linguistic understanding of documentation as communicative act. The motivation for Ley for the development of this methodology was the existence of the area of conflict between customer usability on the one side and economic effectiveness of the other side (Ley, 2006). According to Ley, an optimised compliance of both can be gained only by a reasonable information structure. The therefore developed text control raises the claim on enabling complete and consistent modeling of a clearly outlined quantity of information. There exists no evaluation of the benefit of this methodology in comparison to others. According to the focus of this research work on enabling an optimised reuse of modules, the described way of the consistent and complete declaration of modules shows promise.

.

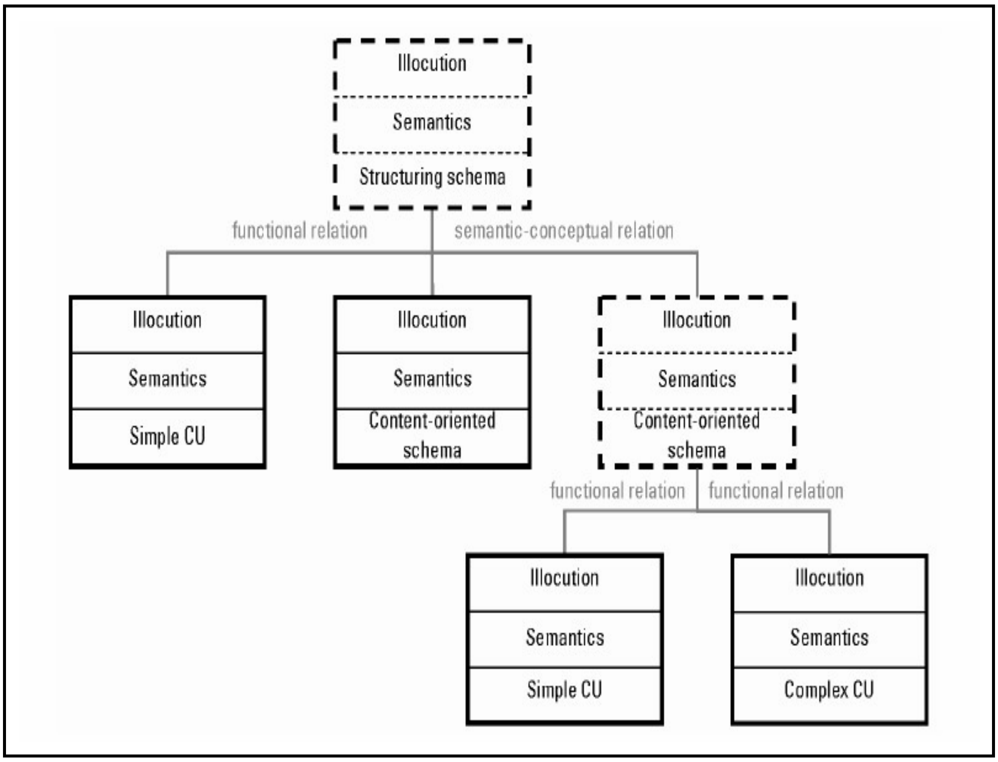


Figure 3 Structured Text Control (Ley, 2006)

2.6 Conclusion

Analysing the methodologies, new challenges occurred by comparing the used terminology within the analysed material: There was no exact defined boundary between the terms data, information, information product and document resp. documentation. The author of this thesis used the terminology within the description of the methodologies above as it was utilised in the analysed material. In Chapter „Developing a Semantic Data Model“ on page 29 exact and consistent term definitions for *data* and lemmata for *information* and *knowledge* will be proposed, which enable the exact annotation of the terminological foundation in the research-relevant context of technical documentation and therefore used modeling.

Most of the analysed methodologies ((3), (4) and (5)) are based on the functional aspects of texts, whereupon (4) and (5) have their great benefit in concentrating in addition on how to write to optimise the understandibility of texts. This is according to Figure 2 (Ziegler, 2006) desired on the top level for linguistical and terminological consistency. The methodologies (6) and (7) concentrate on XML-based authoring, whereupon (7) shows promise for higher possible reuse effects by offering inheritance and specialisation mechanisms.

As regards the desired standardisation levels (see Figure 2) according to (Ziegler, 2006), there is no methodology considering all the required levels.

Regarding the introduced principles for structuring technical documentation in Section 2.4, (8) constitutes the only methodology integrating the functional and the content principle for structuring information and provides consistent modeling primitives for a special domain (aircraft checking procedures).

The content principle as well as ontologies seems not yet to attract interest in the field of technical documentation. In the following chapter, the author provides the basic of ontologies and analyses methodologies for their design regarding the mentioned content principle of structuring documentation.

3 Foundations: Ontologies

This chapter introduces ontologies and their design. It starts with Section 3.1 demonstrating the typical communicational situations along a product development cycle as agent-based knowledge transfers. The history and foundations of the concept *ontology* are briefly described in Section 3.2. The current trend of understanding ontologies in computer science is approached in Section 3.3. Following that understanding, the core usage of ontologies is to support communication. Section 3.5 summarises the research analysis work of several widespread methodologies for the design of ontologies.

References: The first, second and third section of this chapter are mainly based on (Sieber&Chen, 2007). The last section is based on (Sieber&Lautenbacher, 2007).

3.1 The Human Agent

As defined in Chapter 2 the product development cycle contains various processes, in which everybody is permanently in charge of communicating knowledge and the data exchange is fulfilled via documents. At the same time, everybody is playing both parts: the part of the knowledge carrier and the part of the knowledge researcher.

- The knowledge carrier faces the challenge to let the knowledge enter the documentation, overall being stored appropriate in data.
- The knowledge researcher is forced to extract the relevant parts of the documentation.

In Figure 4, those roles are demonstrated (cf. (Sieber&Chen, 2007)): the human being as a "black box agent", being able to receive data via sensor from the outside world and to create data via acting in the outside world.

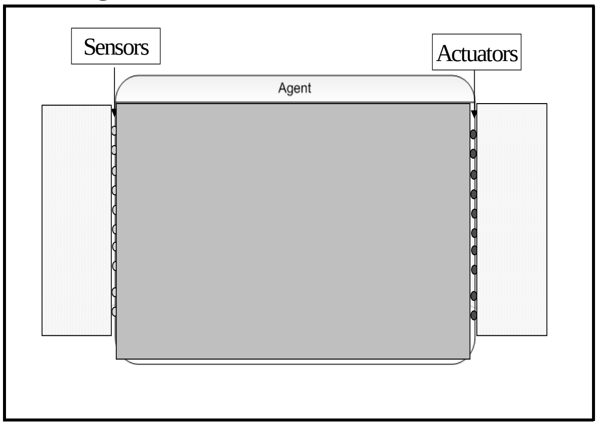


Figure 4 Black Box: Agent

Ontologies offer an approach how the intra - personal knowledge of such an agent - the black box - can be understood and captured in a formal way.

3.2 Philosophical Meaning

> *„The term "Ontology" (Greek, on = being, logos = to reason) in its original sense is a philosophical discipline, a branch of philosophy that deals with the nature and the organization of being."*
>
> (Sure, 2003)

Originally, the terminus ontology appeared in the 17th century as a synonym to Metaphysics or the First Philosophy, defined by Aristotle (cf. (Kunzmann, Burkard&Wiedmann, 1991)). Understanding ontology like this, ontology is therefore a part of philosophy, asking which things and phenomena in the world belong to the "being" and which of them are only existing in the imagination of human beings or only existing under certain circumstances.

3.3 Ontologies in Computer Science

The 20th century is characterized of linguistic movements and improvement in natural science and technology based on the methods and awareness of knowledge in the field of formal logic (cp. (Kunzmann, Burkard&Wiedmann, 1991)).

The Semiotic Triangle as Ontological Basis

The term of ontology is subordinated to a changing understanding of the interaction between the real objects of the world, symbols and thoughts. The basic discussion concerning these contexts are not a novelty at all, they can be traced back up to Platon and Aristotle. He himself for example discovered in his "Peri hermeneias", that

> *„Spoken words are the symbols of mental experience and written words are the symbols of spoken words. Just as all men have not the same writing, so all men have not the same speech sounds, but the mental experiences, which these directly symbolize, are the same for all, as also are those things of which our experiences are the images."*
>
> (Aristotle, 2007)

Aristotle's basic approach was multiple developed within the history and reputed. Ogden and Richards illustrated a schematic for the first time in the year 1923 as a semiotic triangle (illustrated in Figure 5). In principal, it means that words – or better to say in the way of Peirce signs – cannot be assigned unambiguously to objects in the world or a thereto-belonging concept.

.

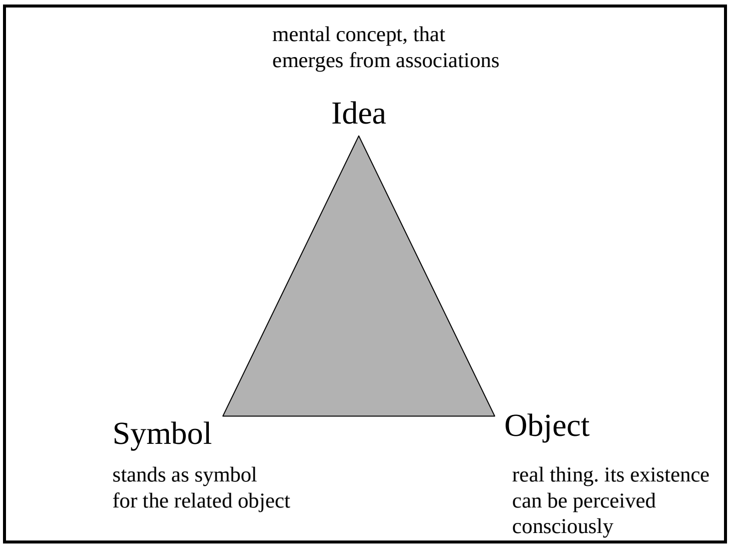


Figure 5 Ogden/Richards: Semiotic Triangle

Ontologies as Conceptual Basis for Communication Between Agents

The basic idea of ontology is, that a certain group of agents can agree upon which concepts are important within their scope of application and how they can combine them together. The term of ontology in an information technical sense means in this baseline a formal defined system of terms (concepts) and the combination of them. Tom Gruber published the most prominent definition nowadays (cf. (Gruber, 1993; 1995)):

Definition *Ontology (in terms of Gruber)*

„An ontology is an explicit specification of a conceptualization.“

(Gruber, 1993)

As given in the definition of (Gruber, 1993), an important aspect of ontologies is their consensual character. Users of ontology first need to agree on a shared conceptualisation to enable knowledge sharing on a conceptual level. Ontologies reduce thereby the possibility of interpretation of the used symbols and their according concepts.

Ontologies and Technical Documentation

According to this, ontologies can be used as a kind of conceptual basis to communicate knowledge (Sieber, 2005).

Definition *Ontology (in scope of this work)*

An ontology describes a conceptual base to communicate knowledge.

Formalisms of Ontologies

In the context of ontologies the following abstract terms are important (cf. (Sure, 2003)):

- *Classes* (often also called *concepts*),
- *relations* (often, especially in the RDF community, also called *properties*),
- *axioms* and
- *instances*.

Concepts are normally organised in taxonomies through which inheritance mechanisms can be applied. Relations represent a type of interaction between concepts of the domain. Axioms are used to model sentences that are always true. They are also used to represent knowledge that cannot be formally defined by other components. Instances can be instances of concepts and of relations.

The range of everything being titled ontology may reach therefore from

- very simple and easy kept glossary over
- taxonomy, containing concepts and relations, up to
- very complex ontologies that contain beside concepts and their relations also rules.

Application Example: Automised Validation of Product

The last ones may allow a conclusion over the data amount, which is described by the ontology. Such a complex application may be used for example to validate a product against a created specification of requirements and according to this discover at an early stage if the product really fulfils what the requirements determinate of the product.

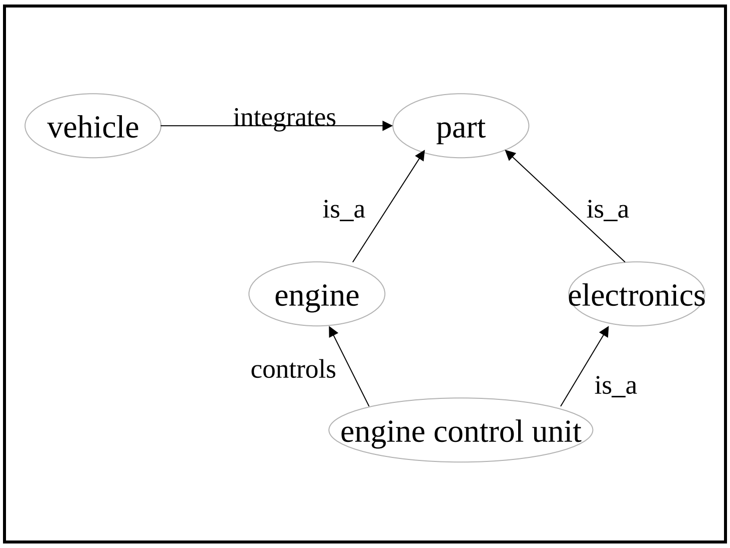


Figure 6 Example of Simple Ontology for Fault Finding Purposes

This offers promising implementations of ontologies in the field of fault diagnosis and automated testing. In (Sieber&Chen, 2007) it is described, how ontologies can be used in that context for automated fault finding (see Figure 6). Implementing axioms faults can be logically derived within automised checking procedures of the according instances.

Applications for integrating ontologies are easy to find. Thus, central question for application fields like the shown above is: Do methodologies exist for guiding the modeling of ontologies in order to gain the most possible benefits of the resulting ontology?

In the next section, existing methodologies for the design of ontologies are provided.

3.4 Design of Ontologies

In general, there are three different ways existing in order to create an ontology (Sieber&Lautenbacher, 2007):

1. Either there are no basic documents existing, which can be used for the creation, or the ontology needs to be created completely new.
2. Alternatively, knowledge is existing in more or less structured form (for example in a database, ontology of components etc.).
3. The last possibility is to create ontologies out of complete unstructured text documents under using techniques such as information extraction as well as natural language processing (NLP).

As follows, the summarised analytical research work on methodologies for ontology design belonging to the first group is provided with special regard to the description of the conceptualisation of a domain.

3.5 Methodologies for Ontology-Design

(1) Uschold & Grünninger

(Uschold&King, 1995) resp. (Uschold&Grünninger, 1996) propose a methodology consisting of four main phases:

- Definition of the purpose of ontology,
- construction of ontology with
  - collection (top-down, bottom-up or middle-out),
  - coding, and
  - integration,

- evaluation, and
- documentation.

(Grüninger&Fox, 1995) resp. (Uschold&Grüninger, 1996) extended the proposed methodology - based on the above listed steps and their collected experiences in the TOVE Ontology Project.

A drawback of this methodology is that there is no detailed description of the transition between the phases and the phases itself. It would be very interesting for example to see the influence of the defined purpose of the ontology for the following steps.

(2) Swartout et al.: SENSUS

(Swartout et al, 1996) developed a methodology for the design of taxonomies consisting on the following steps:

- Identification of so-called seed-concepts,
- manual connection of these concepts,
- addition of concepts on the path to the root,
- addition of new domain concepts,
- addition of whole subtrees,
- removal of not used concepts.

The methodology is mainly based on the implemented possibilities within a tool and not on presenting a methodology for a well-guided conceptual design of ontologies.

(3) Sure, Staab et al.: OTKM

(Staab et al., 2001) and (Sure, 2003) developed the On-To-Knowledge Methodology (OTKM) and proposed the following proceeding:

- Feasibility study:
  - Problem identification,
  - selection of possible existing solution.
- Ontology-Kick-off:
  - Requirements analysis,
  - analysis of existing sources,
  - construction of taxonomy.
- Refinement:
  - Hearings with domain experts,
  - extension of taxonomy,
  - formalisation of taxonomy,

  - addition of relations.
- Evaluation:
  - Discussion of problems and possible solving,
  - evolution.

(4) Gómez&Fernández: METHONTOLOGY

In 1997, Fernandez-Lopez et al. established the basis for the METHONTOLOGY (cf. (Gómez&Fernández, 2004)). This methodology describes the design of ontology as a sum of activities, which can be grouped into three categories: Management,- development and support-activities. The conceptualisation is involved hereby in the development activity group and has to consider the following steps:

- Construct glossary,
- construct taxonomy,
- construct binary relation diagrams,
- construct concept-vocabulary,
- describe binary relations/instance attributes, class attributes and constants
- define formal axioms and rules
- construct instances.

The methodology resembles in the activity list for conceptualisation kind of iterative brainstorming and let still miss detailed answers on the question, what exactly should be kept in concepts.

(5) Pinto et al.

(Pinto&Martens, 2004) propose a methodology similar to METHONTOLOGY consisting of specification, conceptualisation, formalisation, implementation and evolution. In each level, several activities can be executed and the design of ontology is executed analogous to the development of a prototype - in each phase a return to any previous phase is possible. The prototype of the ontology is refined in cycles.

(6) Noy&Guinness: Ontology Development Guide 101

In the Ontology Development Guide 1.0.1 (Noy&Guinness, 2001) the proceeding for the finding of concepts and relations is described. The design of ontology should therewith be an iterative process and the identified concepts should be close to the logical or physical objects of the real world. The description of the found concepts and relations should be formulated using substantives (concepts) and verbs (relations), which are - formulated in sentences - describing the domain. For the design of ontology Noy&Guinness recommend to query competency question that are domain-specific and can be used later on for ensuring that the designed ontology is complete.

3.6 Conclusion

Concerning the description of the conceptualisation, the methodologies (1)-(5) remain vague. (3) gives a well-founded process overview, but without detailed instructions how to act on. (4) goes deeper in the several phases and covers all steps within a life cycle of the ontology that has to be designed. Similarity exists to the Rational Unified Process in which the several steps concerning software development are described in a similar way. A detailed explanation how to find the concepts of ontology is neither given here. The only methodology that proposes exact action instructions for finding concepts, attributes and instances is (6).

The pragmatic aspect of ontologies - to communicate knowledge using a conceptual basis - is only rudimentary captured in (1), where the analysis of the purpose of the ontology is instructed.

As explained in Section 3.3, ontologies reflect the linguistic understanding of signs. Thus, none of the methodologies takes any linguistic background as foundation for the elements of the ontologies.

Ontologies vs. Technical Documentation

The provided definition for ontologies in scope of this research work on page 21 introduces ontologies as conceptual base for *communicating knowledge*. The given definition for technical documentation on page 2 points out the objective of technical documentation 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. Astonishing that although there is a lot of research in using ontologies for E-learning purposes (cf. (Knight, Gaševic, & Richards, 2005)), there is no research interest in using ontologies in the field of technical documentation.

The basic idea of this research thesis is that ontologies and existing methodologies for their design promise for use within technical documentation. Comparing the methodologies, that are used for structuring technical documentation with those used for the design of ontologies, 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*.

As a consequence there exists no clearance, if and how ontologies and/or their design methodologies can serve technical documentation purposes. In the following part, the author presents the consistent terminological foundations for data, knowledge and information, that have been worked out with special regard to the field of technical documentation. Based on these terminological foundations the development of the so-called semantic communication model is provided. The purpose for its development is to obtain a reference model for technical documentation, which can be used to evaluate the qualification of any of the presented methodologies (for structuring documentation as well as for design of ontologies) for gaining a higher reuse of technical documentation.

The development of such a model is necessary for closing the above shown existing gap between the understanding of basic concepts with special regard to the same objective of ontologies and technical documentation, viz. the transfer/communication of knowledge.

Part II Semantic Communication Model

The person discovering a new scientific truth had to destroy almost everything before, what he ever had learnt.

José Ortega y Gasset, 1883-1955

4 Developing a Semantic Data Model

This chapter begins by presenting the existing gap between data and semiotics. This is done by a summarised critical terminological analysis (cf. Appendix E.1), the presentation of the semiotic understanding of signs in terms of Peirce, and existing models for relationships between data, information and knowledge in Section 4.1. In Section 4.2, the semantic data model (Sieber/Kammerer) is presented with its definitions for data items and related terms.

Based on this semantic data model the challenges of technical documentation are achieved and the terms knowledge and information are defined in their decompositions in Section 4.3.

References: The chapter is based on (Sieber&Kammerer, 2007a) and (Sieber, Kammerer&Kovács, 2007).

4.1 Terminological Foundations

In the context of technical documentation and the so-called knowledge engineering the terms data, knowledge and information are used partly in a reflected way and partly in a not reflected way. In spite of multiple efforts to define the terms or to describe the underlying concepts semantically, the terms remain vague, the given definitions or statements are contradictory, sometimes contra intuitive and leading to unacceptable consequences. As mentioned in Section 2.5, the usage of the terms even in the same domain is inconsistent. Thus, misunderstandings are preassigned if in the particular discourse is not clarified in advance, in what sense a term should be understood.

Data

Following the analysed definitions and statements in (Sieber&Kammerer, 2007a) data is regarded as being

- *„observed differences“* (Willke, 2004),
- *„outputs from devices, explicated on a medium“* (Burkhard&Eppler, 2005),
- *„signs that are stored in a medium“* (Back, Bendel&Stoller, 2001),
- *„something that exists without having any meaning“* (Keller&Tergan, 2005),
- *„some correlated signs that make a sense in specific context“* (Seeger, 1997) , and/or
- *„materialised knowledge that was syntactically correct composed for answering a question“* (Kuhlen, 2004).

Due to this analysis, information scientists let *data* appear as signs, symbols, numbers, characters, images or "something" - always with the restriction that it is per se *without any meaning*.

Signs

By contrast, people with a strong linguistic background, e.g. technical writers or computer linguists refer to the *semiotic triangle* (cf. page 20) in order to explain their understanding of the world with the existing relationship between mental concepts, signs (in terms of Peirce) and referred objects.

According to Peirce, signs can exist on three levels. Thus, Peirce differs the following semiotic entities:

- *Indices* - Sign objects have a causal correlation to their meaning.
- *Icons* - Sign objects have a natural coherence to their meaning.
- *Symbols* - Sign objects are arbitrarily correlated to their meaning. This correlation is produced by convention.

Data-Knowledge-Information

In addition to the analysis of the concepts data, knowledge and information, existing models for the relation between data, knowledge and information (cf. (Werms, 2006), (North, 2004) & (Ackoff, 1989)) have been analysed.

.

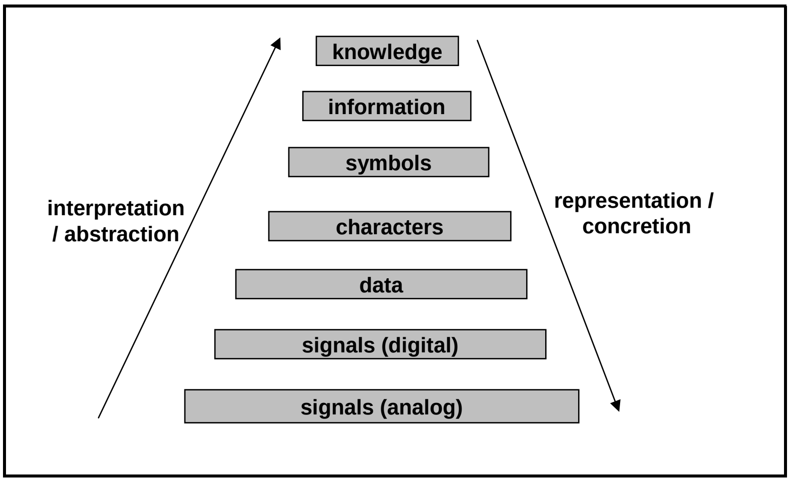


Figure 7 Approaches: Data-Knowledge-Information according to (Werms, 2006)

According to Werms, symbols are abstracted characters and those are abstracted data. This contradicts to the understanding of signs (and therefore symbols) in terms of Peirce.

.

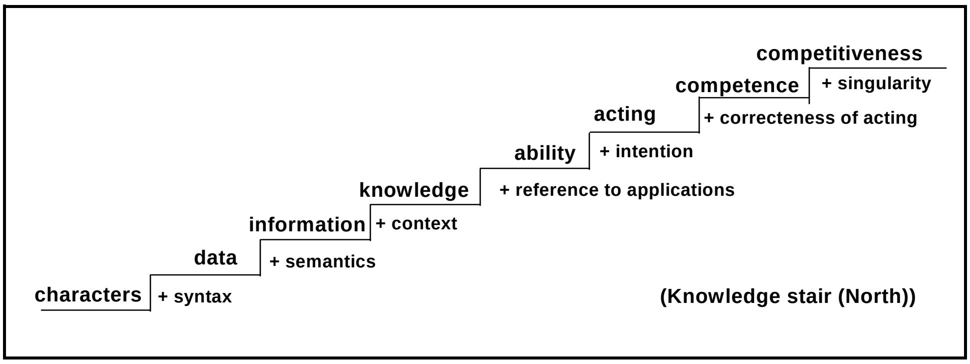


Figure 8 Approaches: Data-Knowledge-Information according to (North, 2004)

North proposes in his model, that characters have to be added up with syntax in order to achieve the data level and data needs semantics in order to achieve the information level. Again, there is a contradiction to the understanding of signs in terms of Peirce, where even a character is a sign and has therefore a meaning.

.

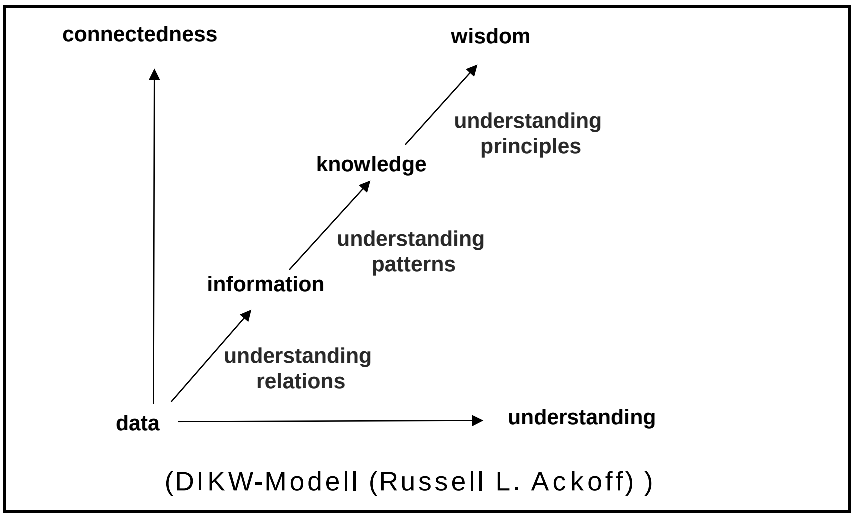


Figure 9 Approaches: Data-Knowledge-Information according to (Ackoff, 1989)

Ackoff represents data, information, knowledge and wisdom in relation to their connectedness and the necessary understanding degree. For the abstraction from data to knowledge is an understanding of relations necessary. But how can this kind of understanding exist, when the underlying base for an understanding - the knowledge - is not yet achieved at this stage according to this model?

4.2 Development of Semantic Data Model

The several analysed approaches and terminological definitions and statements about data have not reflected the research understanding of data in relation to the concept of signs (in terms of Peirce). Therefore, a data model was developed that aims in bridging the existing gap between semiotics and information sciences and unifies perspectives and cognitions, which contribute out of these scientific areas to the character and usage of *data*.

Different Levels of Data

According to the developed semantic data model (see Figure 10), data penetrates through various levels:

- At the level of *form*, data appear concretely as so-called data instances.
- At the level of *representatives*, data appear in an abstract form, as the so-called data representatives, which can be assigned to particular data representation systems.
- At the level of *meaning*, data appear again as abstractions (data items), which imply that a semantic system is allocated to the data.

Furthermore, atomic data items can be differentiated from complex data items. Atomic data items have the property that they cannot be split any further into smaller meaningful components.

.

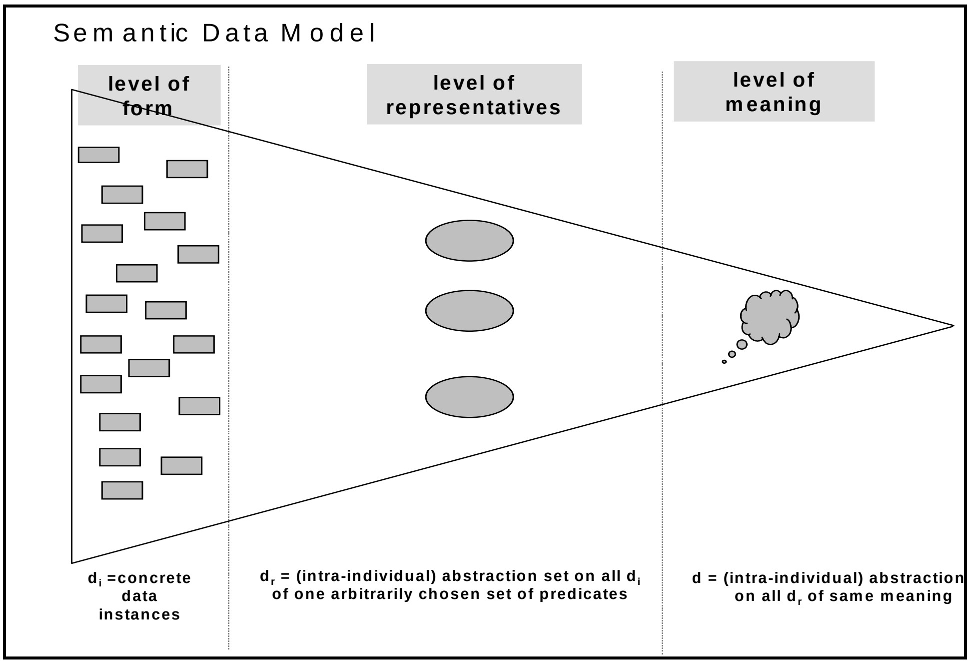


Figure 10 Semantic Data Model (Sieber/Kammerer)

Definitions: Data Item, Data Instance and Data Representative

Furthermore, the following terms, which appear in the semantic data model, are defined. The basic definition is the following for the concept *data instance:*

Definition *Data Instance*

A data instance is the concrete (extra-individual) occurrence of a semiotic entity. This means simultaneously: Every semiotic entity is a data instance.

With this definition, the gap between data and semiotics is closed. In terms of Peirce, a data instance can therefore appear on all three levels: As icon, index, or symbol.

Definition *Data Representation System and Data Representative*

A data representation system is a set of predicates on the powerset of data instances. A data representative is the (intra-individual) abstraction set on all data instances of one arbitrarily chosen data representation system.

Therewith a data representative and - a fortiori - a data instance always belong to a specific data representation system, e.g. the (abstract) data representative "11" belongs to the data representation system "decimal system, Arabic numerals, value11" and can be substantiated for example in the (concrete) data instances 11, **11**, **11**, `11` etc.

Definition *Data Item*

A data item is the (intra-individual) abstraction set on all data representatives of the same meaning.

This definition for data item implies that data (as a collective term of data items) following the semantic data model bear a meaning, otherwise no set constitution would be possible. It means also, that it depends on the meaning, what kind of abstraction set has to be accumulated.

Definition *Atomic Data Item*

An atomic data item cannot be analysed in further data, which contribute to the meaning of the data item.

An atomic data item requires therefore mapping rules in the form of functions assigning the chosen (intra-individual) data item to the chosen (intra-individual) data representation system and the chosen (intra-individual) data representation system to concrete (extra-individual) data instances.

Definition *Complex Data Item*

A complex data item is a data item, which can be analyzed in further complex and/ or atomic data items.

A complex data item requires therewith in addition to the mapping rules syntactical rules in the form of grammar for the correct syntactical composition of the atomic data items on the level of form and meaning.

Concretion Process: From Data Item to Data Instance

In the semantic data model, the data representation system plays a key role, besides the meaning of the data. This can be seen in an example showing the concretising of a data item taking place: For instance, for substantiating the atomic data item ELEVEN, a person will have various possibilities for concretising this.

- The person can decide to use the decimal system and Arabian numerals, and to represent the values between 1 and 15 in this system (for example, "11").
- Or, to use Roman numerals and map the value range mentioned (for example, "XI").
- Or, in the hexadecimal system, the Latin alphabet together with Arabian numbers and map the value range similarly (for example, "B").
- Of course, a person could just as well spell out the corresponding numbers in English, using the Latin alphabet (for example, "eleven").
- This can be extended as desired.

One thing which takes place almost unconsciously while concretising a data item is the choice of a particular data representation system, and thereby a data representative for the data item. In the following (and similarly, mostly unconscious) step, a person will select a certain formal representation for the selected data representative, and create a concrete data instance, such as typing and formatting "11".

Abstraction process: From Data Instance to Data Item

Now, data items that have been concretised in this manner confronts another person in the form of a data instance. This person in turn, and again unconsciously, tries to understand the data. This is done by abstraction.

.

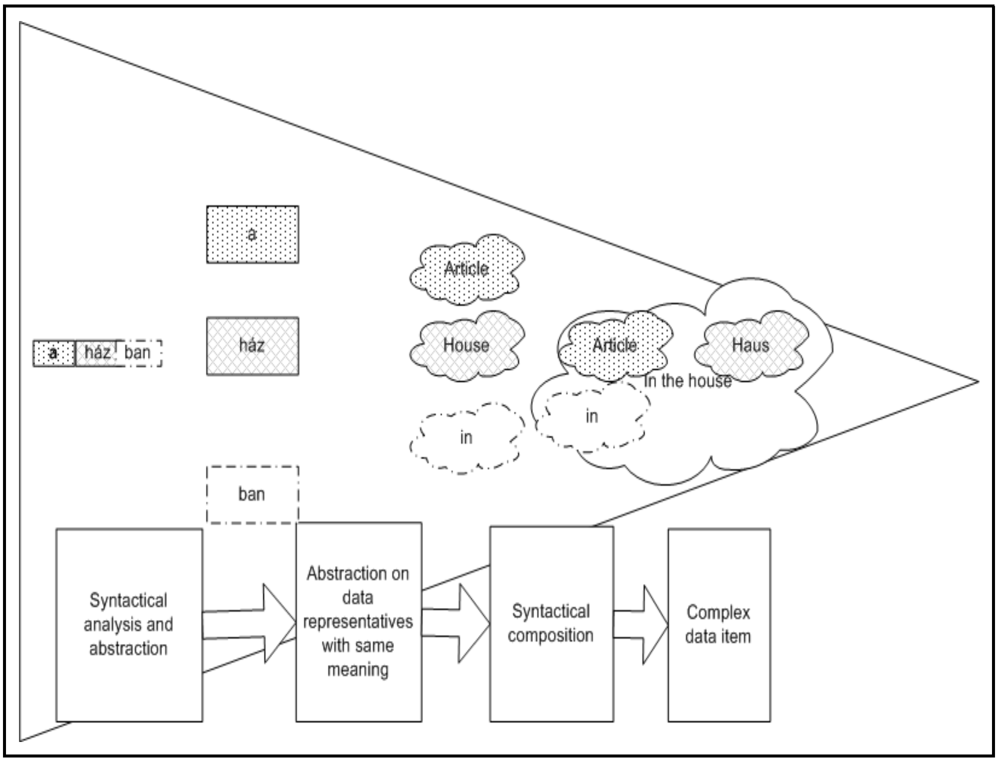


Figure 11 Example - abstraction process from concrete data instance 'a házban'

A person perceiving the concrete data instance 'a házban' of the example shown in Figure 11 is confronted by several questions during the process of abstraction:

- What is the syntactical decomposition of the given complex data instance?
- What is the variation used in concretising the chosen data representatives?
- What is the underlying data representation system?
- What is the meaning of the atomic data items?
- What is the meaning and the syntactical composition of the complex data item that was originally the basis for the concretisation?

Abstraction towards the "correct" (in terms of meaning) complex data item can happen only if the corresponding *mapping rules* are defined or can be derived without ambiguity. In the case of complex data items, *syntactical rules* have to be given in addition to the *data representation systems*. In the example above these syntactical rules are applied for the decomposition of the complex data instance (grammar of hungarian language) and for the composition to the complex data item 'in the house'.

4.3 Application in Technical Documentation

The daily challenge of technical documentation is the transfer of technical *knowledge*. As follows, the development of this concept with the background of the developed semantic data model is provided.

Related Work: Terminological Foundations

According to (Kuntz, 2005), knowledge exist as a kind of a cognitive structure and is localised implicit intra-personal and explicit in documents. (Probst, 2000) and (Sure, 2003) describe knowledge as conscious basis for actions and problem solving (cf. Appendix E.2).

Following the analysed definitions/statements of the term information (cp. Appendix E.3), the understanding of this term changed over the years. It may be assigned therewith by:

- The definition of the verb *informare* in (Stowasser et al., 1979) as assimilating the form of the object of cognition without its material substance,
- the meaning noted in (Oxford, 2007) as the act of informing used in the phrase *„Men so wise should go and inform their kings.“ (1330),*
- the measurable information like defined in (Hartley, 1928) reduced on the syntactical component of information,
- the understanding of telecommunication engineering following (Shannon&Weaver, 1972) that is also denying any semantic components of communication and focusing only on syntactical transfers and their possible noises,
- the subjectivity of information in dependence of the perceiving subject (Kovács, 2004), (Hansen&Neumann, 1978), (Keller&Tergan, 2005),
- the fact to be a kind of a meta category, that is fulfilled if the transferred message answers in a communicational frame a question (Capurro, 2000),
- a part of knowledge that influences the already existing knowledge recipient-sided (Kuhlen, 1999), (Ingwersen, 1992) and/or
- again similar to the original understanding (cf. (Stowasser et al., 1979)) a construction, that is resulted in the recipients’ activity: information as the image of an object that is constructed by the recipient (Fuchs-Kitkowski, 1976).

The basic underlying understanding of data/signs for all analysed definitions/statements differs from the understanding and definitions of the semantic data model. As follows, the understanding of the terms knowledge and information is given as perceived in the context of technical documentation with special regard to the developed semantic data model.

Knowledge and Information in Technical Documentation

As constated in the definition for technical documentation (cf. page 2), the basic action of technical documentation is the transfer resp. communication of knowledge. In general, communication between a sender and a recipient does not happen at the "same level of knowledge" (cf. (Sieber&Bartz, 2006) and (Sieber&Kovács, 2006)). Exchange of "information" (see Figure 12) often fails due to incompatibilities in the way data is processed. This is applicable largely to the communication between human beings and machines, because the underlying basis of knowledge that enables us human beings to interpret data is unknown to the machine as such.

.

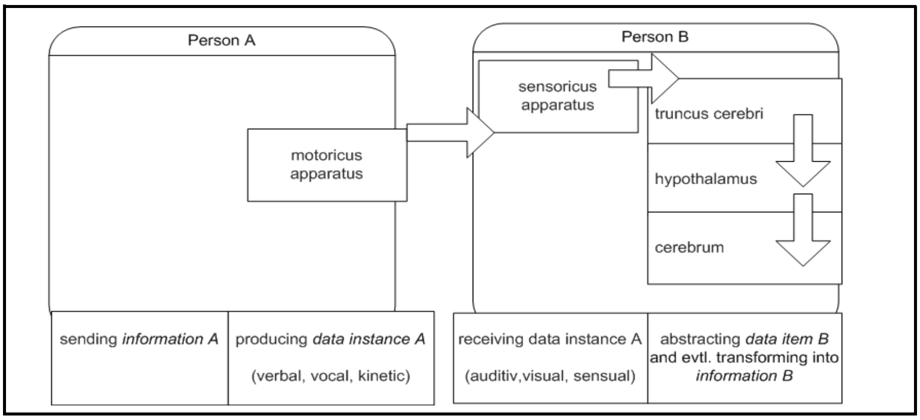


Figure 12 Information Exchange according to (Sieber, 2006)

Sender Recipient Model

The "information" exchange can be visualised using the developed semantic data model (see Figure 13), that is extended therefore to the so-called *sender recipient model* (Sieber&Kammerer, 2007a).

.

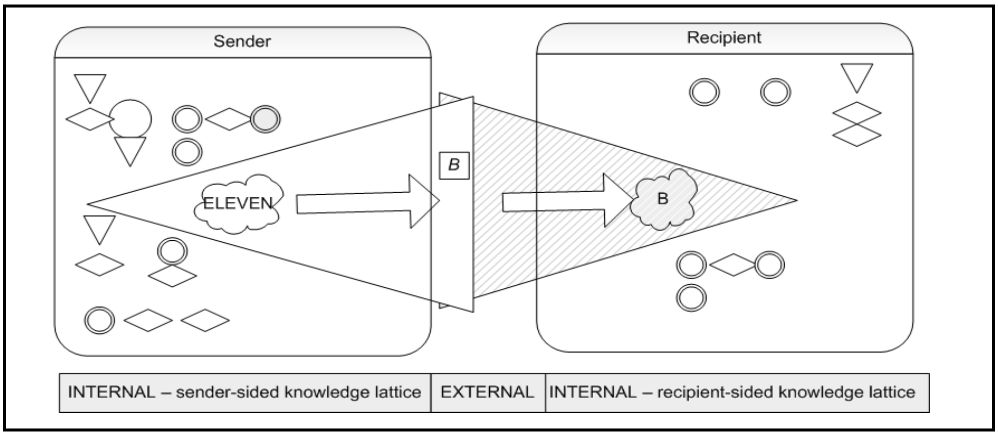


Figure 13 Sender Recipient Model (Sieber/Kammerer)

From the *point of view of the sender,* three steps in a communication are necessary:

- Decision about what sort of "information" the sender needs,
- decision about how that "information" can be "put to data items", and
- a concretising process of data items happens.

From the *point of view of the recipient* an abstraction process takes place and it is not even sure, if the abstracted data item is that data item concretised from the sender, e.g. in Figure 13 the "wrong" (in terms of meaning) data item is abstracted. According to the developed semantic data model and the sender recipient model, the following conclusion is visible: everything outside of a human being is concrete data instance.

Information as Piece of Knowledge

According to the above presented sender recipient model, the only objects that can be transmitted without interpretation are data instances. If it is possible to derive the correct abstractions out of the used data instances, the correct data is abstracted. Only in a next supplementary process that data is localised into the existing knowledge lattice and a "piece of knowledge" is created: information.

Technical documentation aims in optimising this desired transfer of knowledge from a sender to a recipient. The following lemmata for the terms knowledge and information are therefore defined as follows in the context of this thesis with the background of the semantic data model:

Lemma *Knowledge*

Knowledge is intra-personal.

Knowledge constitutes the basic for problem solving and acting.

Knowledge has to be dated for communicating purposes.

Lemma *Information*

Information is new knowledge created by intellectual re-organisation of existing knowledge.

Information is system-relative and hereby subjective.

The effect of knowledge generation out of the perceived data instance(s) is thereby emphasized: in the context of technical documentation, the recipient should be able to interpret the document viz. the data instance(s) and gain new knowledge out of it.

According to the developed semantic data model, this is only possible by the usage of data instances. In order to enable the desired "correct" (in terms of meaning) abstraction process it is recommended to put oneself in the position of the future recipient and to obey the following measures:

- Consistent usage of data representation systems that are known to the reader.
- Consistent usage of data representation systems that are suitable for the context and the conventions (e.g. usage of Arabic instead of Roman numerals in equations).
- Enough redundant data for ensuring that the intended interpretation of meaning can take place.

Metadata

Considering the last proclaimed measure to be taken and to give explicitly details about the data itself, it can be put on an object-level and so-called *metadata* can be introduced on a meta-level to describe the underlying data.

Definition *Metadata*

Metadata are data, metadata are relative and metadata describe data.

Metadata constitute therewith a basis for bringing together data that is related in terms of content, and for processing them further. They can be understood as a pre-requisite for an intelligent and efficient administration and processing, and not least as a focused, formal means of providing relevant data.

Example: XML and Metadata

The usage of metadata in XML (cf. (Sieber&Kammerer, 2007a)) has been analysed. With the help of the developed semantic data model it becomes obvious, that in XML metadata itself is put on an object-level and described by another meta-level. That way several levels can occur and the made annotations can be regarded as metadata to already existing metadata. In fact, the XML together with its specifying files DTD/XSD covers that way three levels of data. The object data instance itself (in the example: a zongarista), the data about this object (in example: név = metadata instance), the data about this metadata instance (in example: ELEMENT) and description of the data representation system, that has to be used (in example: PCDATA).

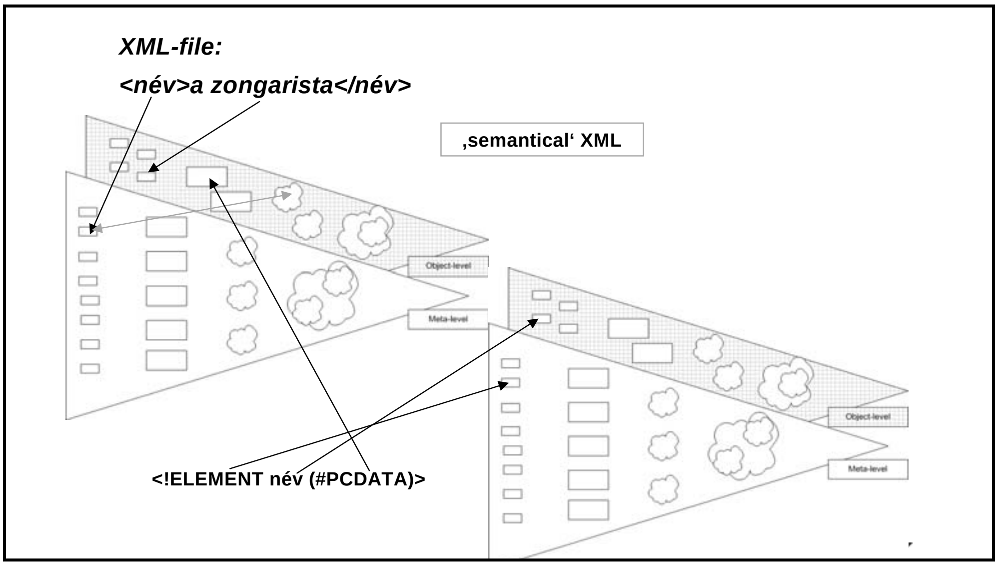


Figure 14 Analysis: Data and Metadata in XML

Using the semantic data model and the analysis it can be argued that in fact there is no "semantic" or "semantical" XML through the natural naming of XML-elements (cf. Figure 14): The metadata instance név should cover in the shown example the "semantic" bridge between future data instances on object level and their meaning, but név itself is a data instance and its meaning is only intra-personal cleared.

Aporia of data

This is founded in the *aporia of data:*

> *„Die bisherige Analyse führte zu der Erkenntnis, dass Daten nicht wirklich greifbar sind, sondern nur deren Dateninstanzen. Dateninstanzen sind semiotische Entitäten, die außerhalb des menschlichen Kognitionsapparats existieren und über die Sinneswahrnehmungen apperzipiert werden. Sowohl die Datenrepräsentanten als auch - a fortiori - die Daten selbst sind Abstraktionen und aufgrund dessen prinzipiell und ausschließlich im Kognitionsapparat des Menschen verfügbar. Dies hat zur Folge, dass Daten - obwohl nur intrapersonal vorgestellt - dennoch als extrapersonal gedacht werden müssen (denn sonst sind sie keine Daten mehr, sondern intrapersonales Wissen und/oder Information), was hier die Aporie des Datums genannt werden soll.“*
>
> (Sieber&Kammerer, 2007a)

[The hitherto analysis lead to the cognition, that data are not tangible, only data instances are. Data instances are semiotic entities that exist outside the human cognition system and are perceived by sensors. The data representatives and - a fortiori - the data items are abstract and only available in the human cognition system. This results in the fact that data – even though imagined only intrapersonal – have to be figured extrapersonal (as otherwise they are no longer data, but intrapersonal knowledge and/or information). This is called the aporia of data.]

4.4 Conclusion

Challenges for the Technical Author

In terms of the developed semantic data model, the challenge of a technical author in the field of technical documentation can be regarded as the selection and design of the data instances in such a way that the desired abstraction process takes place on the recipient's side and guides the recipient to be able to solve occurring problems or to act properly.

Central questions for the author in defining that "piece of knowledge" that the recipient will need are the following:

- What kind of actions have to be guided?
- Which solutions to what kind of problems have to be illustrated in usage with the product?
- What knowledge needs the user maximally for being able to execute the desired actions and/or solve occurring problems?
- What knowledge can be regarded as existing pre-conditions?
- How can the knowledge, that needs to be transferred - i.e. the delta between supposed pre-conditioned and maximum knowledge- be verbalized in such a way, that it contributes to a certain desired action or problem solving?

Requirements for Modeling Methodologies

On an abstract level, these questions have to be considered by the methodology that will be applied for defining, what complex part of the world is transformed into an easy to understand describing pattern. The methodology needs to face up to the above listed questions.

According to these questions and the therefrom to derive requirements for possible evaluations for methodologies, an extension has to be provided within the proposed semantic data model: The following Chapter 6 aims in giving a precise view in the insight of the developed data model by introducing several layers within the semantic-data model that reflect the existence of concepts in terms of the understanding within the ontological area and herewith covers the content principle (cf. Section 2.4).

5 Integrating a Semantic Network

The semantic data model (Sieber/Kammerer) closes the existing gap between data and semiotics. This model was extended using a semantic network based upon a lattice of concepts. The result is the multi-layer semantic data model (Sieber/Kovács) that can be used to visualize more general decoding and encoding processing between signals in the environment and (semantic) concepts.

References: The whole chapter is mainly based on (Sieber&Kovács, 2007).

5.1 Signals and their Processing

Considering the presented sender recipient model (cf. Figure 13), data instances are those objects that exist extra-personal.

Assumed that *signals* exist extra-personal, and that these signals can be produced on the sender-side using so-called actuators and received on the other side using so-called sensors, the sender recipient model can be illustrated the following way in a signal processing model (see Figure 15).

.

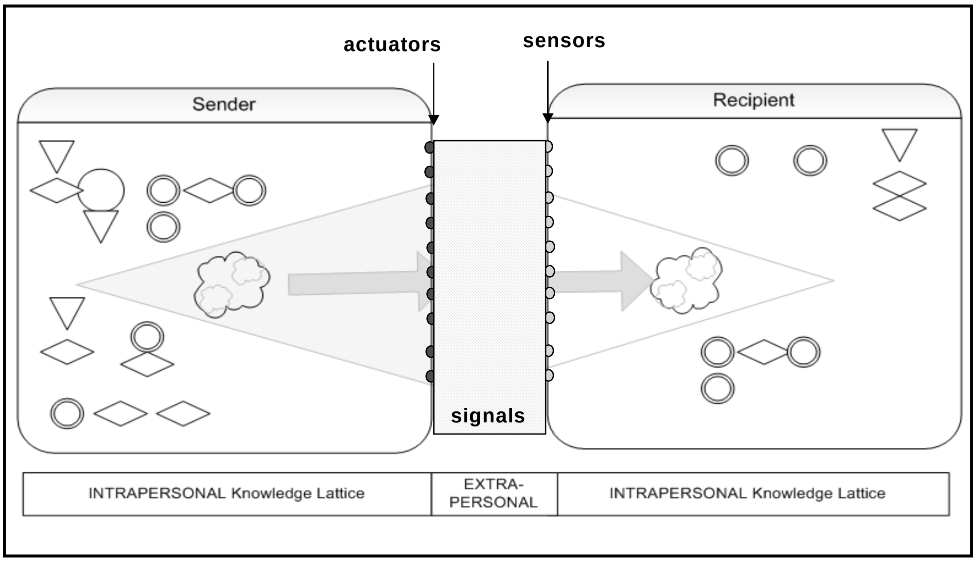


Figure 15 Signal Processing Model

According to this signal processing model, an *agent* is a unit in the world that can act as a sender or/and as recipient (see Figure 16). For the introduced "Black Box Agent" (see Figure 4) ontologies have been introduced as offering an approach how the intra - personal knowledge of such an agent - the black box - can be understood and captured in a formal way using concepts and relations.

.

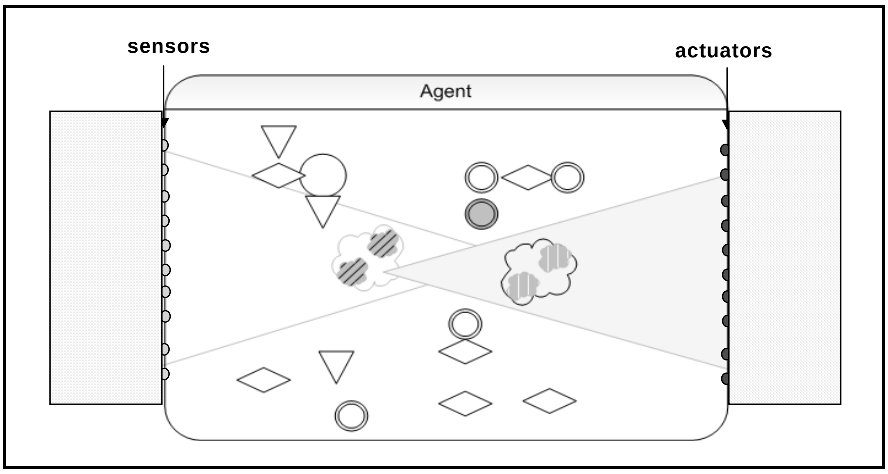


Figure 16 Agent (Sieber/Kovács)

In (Sieber, Kovács, 2007). it is assumed that an agent has a kind of an internal model of the world, which is introduced as a *semantic network*. Furtheron an *encoding engine* is that part of an agent, that has actuators and transforms data items (in terms of Sieber/ Kammerer) into concrete data instances (in terms of Sieber/Kammerer) by producing according signals. A *decoding engine* is that part of an agent that has sensors and creates an updated model of the world based on received signals.

Semantic Network

Definition *Semantic Network*

A semantic network SN is a model description form, wherein concepts and their (specialisation) relationships are described. A concept c corresponds to some symptom of the world.

$$SN = \{C, \leq\}$$

$$C = \{c\}$$

This means, that the model created by the decoding engine can be described by a semantic network as well as the internal model of the agent that constitutes the basis for the work of the encoding engine.

Degrees of Concept Level

The multi-layer semantic data model reflects this understanding of the semantic network by introduction of a *lattice of concepts* regarding the specialisation relationship. This lattice exists of several concept levels depending on the *degree* of the according concepts (cp. Figure 17).

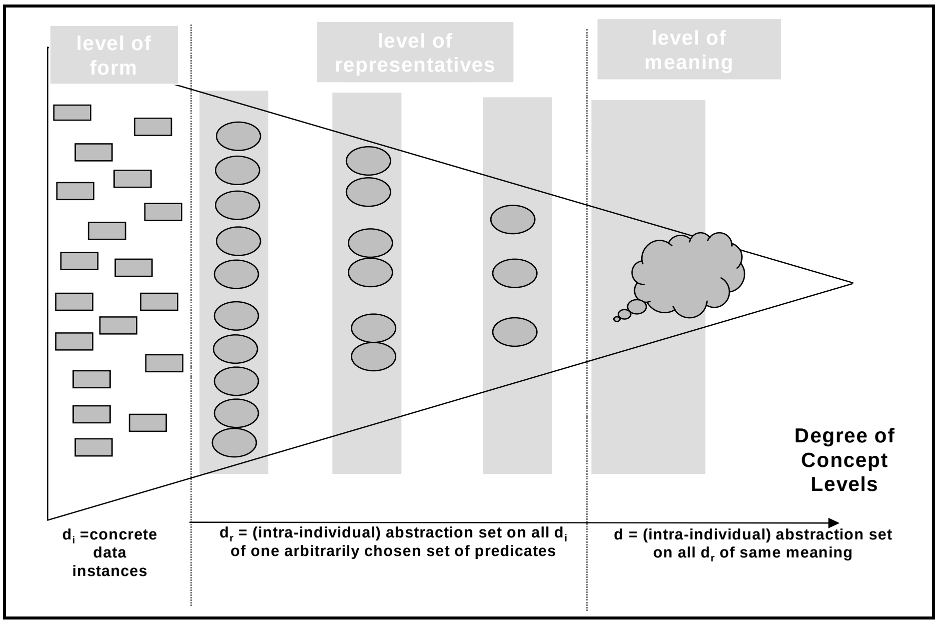


Figure 17 Degrees of Concept Level

All the elements of the world, that are connected to the agent (cf. Figure 16), form the *environment*. The sender recipient model (cf. „Sender Recipient Model (Sieber/Kammerer)“ on page 38) can thereby be modified using the introduced terms (environment, concept lattice, semantic network) and visualising the knowledge lattices by semantic networks (see Figure 18).

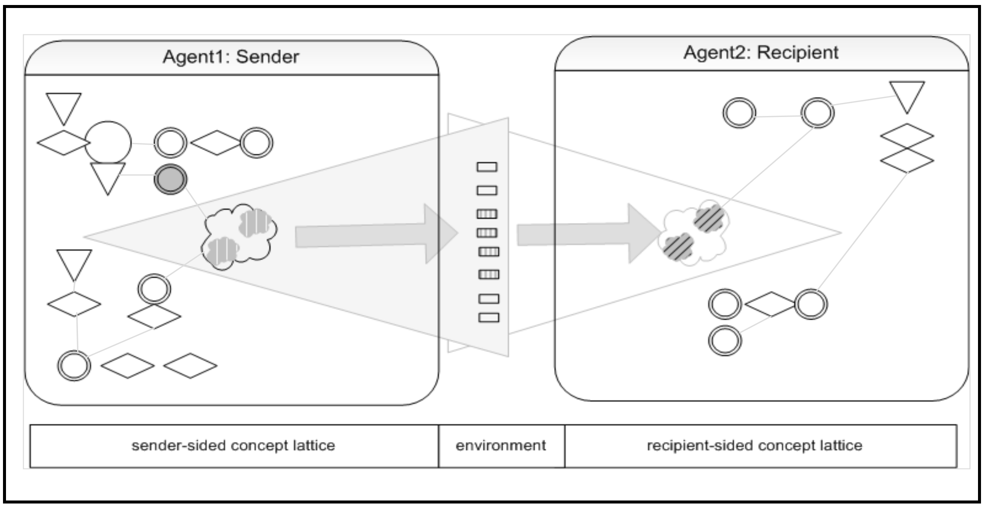


Figure 18 Concept Lattice and Environment

5.2 Transformation From Signals to Concepts

The transformation of signals to data instances to data items is executed in several layers within the multi-layer semantic data model.

Concepts used in this developed multi-layer semantic data model appear on several possible layers:

- The environmental layer,
- the signal-level concept layer
- and the higher level concept layer.

Environmental Layer

Assumed that signals have a nested structure and appear as atomic or compound signals, an atomic signal has a neglected structure and a compound one has an internal structure of atomic signals. Every signal s has a tuple of appearance attribute values and can be mapped onto a signal-space, the dimensions of which can be for example format-specific (e.g. sound, image, temperature,..) or form-specific (e.g. shape, size, colour,…) etc..

Definition *Signal Space*

A={a}: set of attribute values belonging to n different attributes, where the attribute values are given with real numbers

B: signal space, i.e. a vector space, where the dimensions are the attributes and the coordinate values are the attribute values

$$B \subseteq A^n$$

S={s}: set of signals

$$S \subseteq B$$

The recognition and possible distinction between several attributes is dependent of a recipients' sensors.

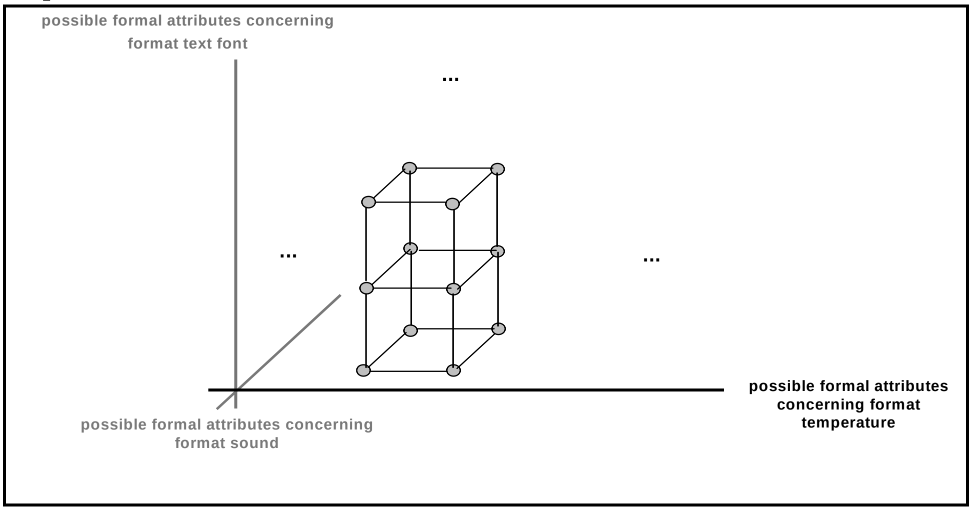


Figure 19 Signal Space

With this definition of signals according to the multi-layer semantic data model, the definition for data instance (cp. page 33) can be extended:

Definition *Data Instance*

*A signal with a corresponding concept in the model is called data instance d_i. A pair of a signal and a concept is introduced as d_i**

$$D_i^* \subseteq (S \times C)$$

$$D_i = \{s |\ s \in S,\quad \hat{\text{I}}c \in C: (s,c) \in D_i^*\ \}$$

A data instance is therewith a point in the signal-space i.e. a signal, for that a corresponding concept in a semantic network can be found. This reflects once more the understanding of the semantic data model, that a data instance is a semiotic entity in terms of Peirce and therefore exists a concept within a semantic network, i.e. a meaning can be assigned to a data instance.

Signal-Level Concept Layer

A concept on the signal-level layer corresponds to data instances in the signal-space and is associated with signal attributes of the selected point. In the defined semantic network concepts and their relations are connected.

If concepts are understood that way, that it is possible to find relations, that are specialising concepts from a higher-level to concepts to a lower level, such a level of concepts exist at the signal-level layer, for that no specialisation relation exists. Despite that concept level is reached, that enables the concretisation into data instances d_i using signal attributes. That concept level is assumed to be the signal level layer and the according concepts the instance-concepts c_i.

Definition *Specialisation Relation*

A specialisation relation is a relation that specialises a concept:

$\leq_s$: specialisation relation

$$\leq_s \subseteq \leq \subseteq C \times C$$

A *specialisation of a concept spec (c)* is:

$$spec(c)=\{c' |\ c' \in C : c' \leq_s c\}$$

Definition *Instance-concept*

A concept is called instance-concept c_i, if spec (c_i) is empty.

$$c_i=\{c \mid c \in C: spec(c) =\{\ \}\ \}$$

Definition *signal level layer*

A signal level layer is a set of instance-concepts.

Higher-Level Concept Layer

Considering the lattice of concepts, a supremum and infimum is defined for every pair of concept. The lattice is called complete lattice, if the infimum and supremum exist for every subsets of concepts within C. The c is called higher concept than c' if c' is related with c by a specialisation relationship. Based on this lattice, the concepts are divided into two groups: atomic concepts having no more specialised subconcepts and higher level concepts. The atomic concepts are the instance-level concepts.

The multi-layer semantic data model therewith shows that the higher level concepts are based on the lower level concepts. That means that the meaning of a higher level concept depends on the underlying concepts.

Mappings

Based on the lattice, a membership, instantiation, representation and generalisation mapping can be defined on the set of concepts (see (Sieber&Kovács, 2007)):

Definition *Membership Mapping*

The membership mapping $m_m(d_i)$ is a mapping in the signal space that maps a data instance to an instance level concept

Definition *Instantiation Mapping*

The instantiation mapping $m_i(c)$ maps an instance level concept to a data instance.

Several membership resp. instantiation mappings may overlap each others: a data instance may belong to several instance-concepts. A quality criterion for the correctness of instance-concepts is the existence of appropirate instantiation and membership mapping, which map the concept to the attributes of the signal space in such a way, that resulting data instances can be mapped to the instance-concept.

Definition *Representation Mapping*

The representation mapping $m_r(c)$ maps a higher level concept to a lower level concept.

Definition *Generalisation Mapping*

The generalisation mapping $m_g(c)$ maps a lower level concept to a higher level concept.

Definition *Compound Membership Mapping*

A compound membership mapping m_{cm} is a mapping that maps a data instance to a high-level concept (see Figure 20).

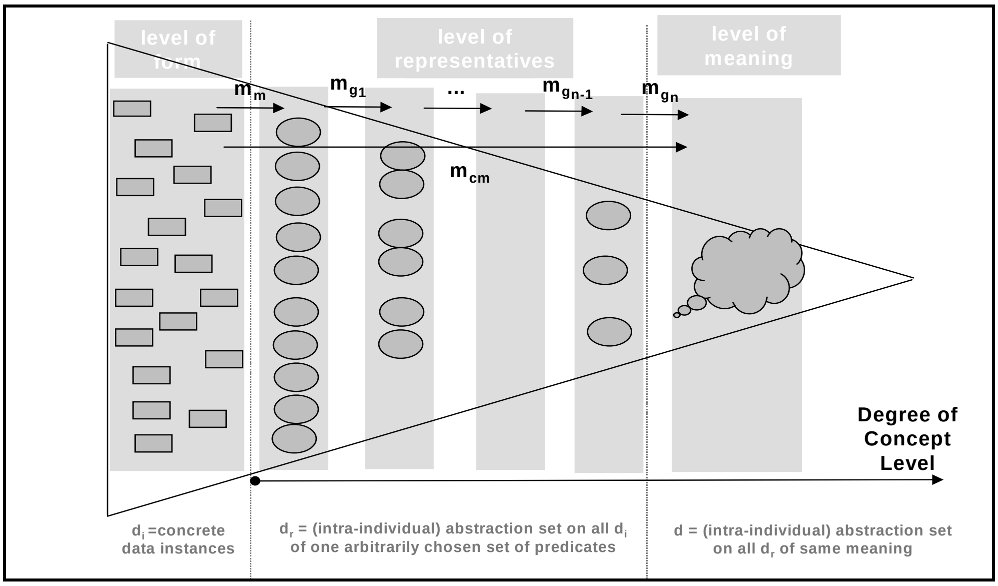


Figure 20 Compound Membership Mapping

- A *compound instantiation mapping* m_{ci} is a mapping, that maps a high-level concept to a data instance (see Figure 21).

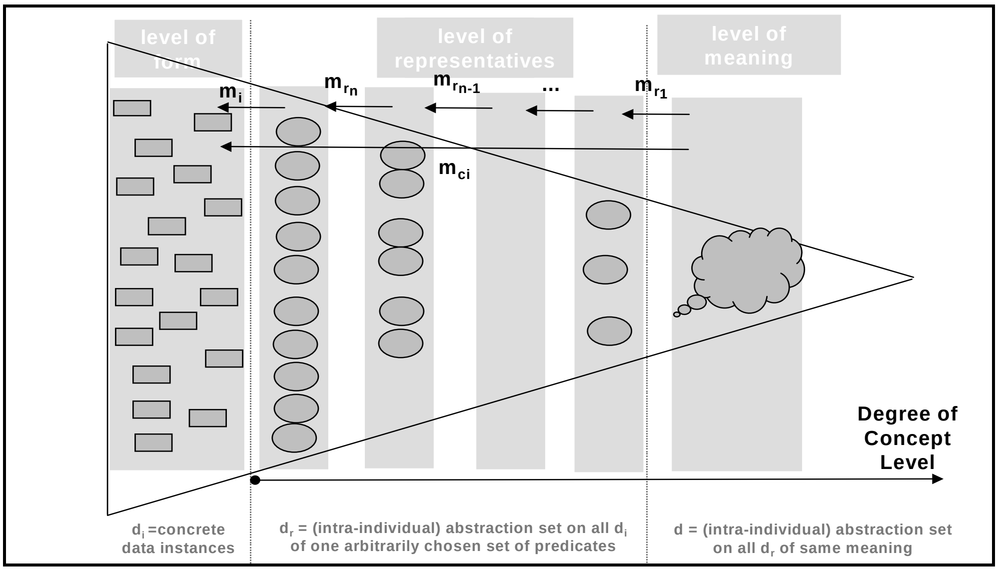


Figure 21 Compound Instantiation Mapping

Predicates

Definition *Predicate*

A predicate p(d_i,c) is a logical function of a data instance d_i and a concept c with

$$p: D_i \times C \rightarrow \{TRUE, FALSE\}$$
$$P = \{p\}$$

$$p(d_i,c)=\{TRUE, \text{ if } \exists f \in M_{cm}, f(d_i)=c) \text{ OR } (f \in M_{ci}, f(c)=d_i)\}$$

Definition *Instance Domain*

A instance domain d_{ins}(p,c) signifies for an arbitrarily chosen concept c the set of data instances d_i, for which the predicate p(d_i,c) takes the value TRUE:

$$d_{ins}(p,c) = \{x | x \in D_i, p(x,c) = TRUE\}$$

On page 33 the data representation system has been defined as a set of predicates on the powerset of data instances. The data item "ELEVEN" has been substantiated there using the data representation system "decimal system, Arabic numerals, value11". This can be analysed using the above introduced definitions for concepts, predicate and instance domain:

- "ELEVEN", "decimal system" ,"Arabic numerals" ,"value11" are concepts.
- 11, **11**, **11**, `11` are concrete data instances.
- Predicates can check, if 11 and "ELEVEN","decimal system", 11 and "Arabic numerals", 11 and "value 11" are TRUE, if according compound membership or compound instantiation mappings exist.
- The data representation system is the set of the predicates "decimal system, Arabic numerals, value11".
- All the (concrete) data instances 11, **11**, **11**, `11` etc. constitute the instance domain of "11" - the data representative of the chosen data representation system.
- The data representative "11" constitutes therewith a concept itself.

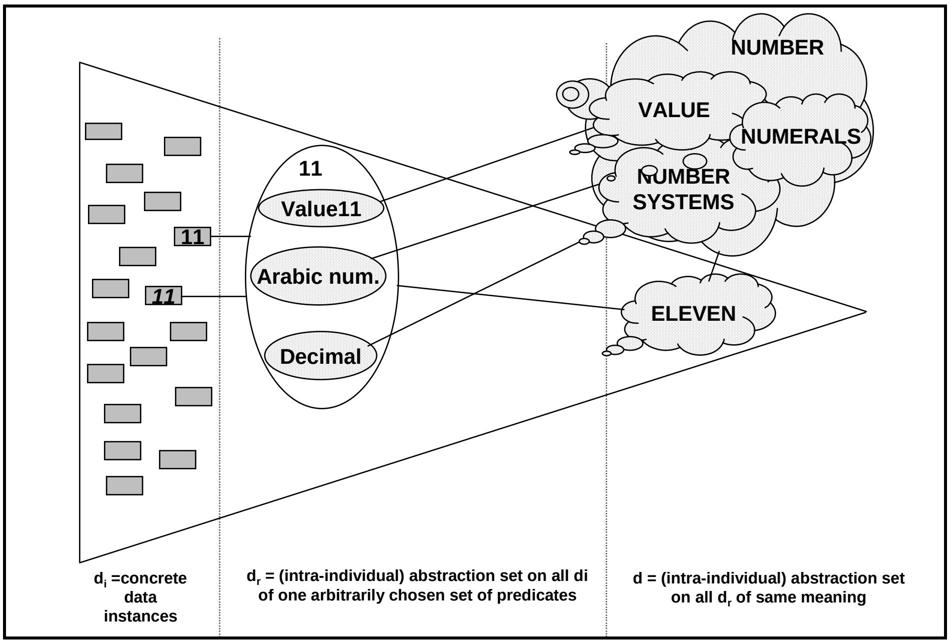


Figure 22 Example Semantic Network - Data Item ELEVEN

5.3 Conclusion

Challenges for Encoding and Decoding Processes

In the extended model, the decoding and encoding processes have been illustrated in detail and a lattice of concepts was therefore introduced. The extended model reflects the nature of signals, introduces the signal space and the so-called signal-level concept layer.

With the integration of the semantic network in the semantic data model, it was illustrated, that the level of representatives (in terms of Sieber/Kammerer) and the level of meaning (in terms of Sieber/Kammerer) are relative: both, data items and data representatives occur in the extended multi-layer semantic data model as concepts.

The abstraction from data instance to data item is executed in several steps depending on the degree of the chosen concept level - if the concretised data item can be localized on the signal-level-layer, i.e. the data item is already the instance-concept, there is only one abstraction process step to be carried out. The semantic network reflects the inner model of an agent, the concepts of whom play a decisive role for the abstraction process. This approach offers new insights in reusing and repurposing data instances: the same data instance can lack the possible abstraction to data item, because the semantic network of the recipient misses therefore needed concepts. On the other hand, the knowledge about the semantic network of the recipients' semantic network enables re-annotating and therewith repurposing of data instances in such a way, that it fits to the inner model of the recipient.

Considering the future recipients’ sensor characteristics, this offers possibilities for reuse and repurposing by mapping the same instance-concept to different data instances. This can be done by defining suitable signal attributes and attribute values that are optimised for the recipients’ sensors.

Requirements for Modeling Methodologies

Regarding possible reuse optimisations within the field of technical documentation, the extended model can be used for evaluating the quality of modeling methodologies. Quality requirements for the "correctness" of descriptions for the inner model of an agent can be derived from the point of view, that technical documentation aims in transferring knowledge: For reuse purposes the inner model should therefore be described in a consistent and formalised way, that allows semi-automised processing of technical document-bricks, i.e. data instances according to their underlying meaning, chosen representatives and signal attribute values. For doing so, the appropriate concepts, mapping functions and therewith predicates have to be consistent declared and used.

The developed multi-layer semantic data model has focussed on the development of a model serving the content principle of structuring/modeling technical documentation. In the following chapter, the developed data model is integrated into a communicational situation and serves therewith the function principle (cf. Section 2.4). This is done with the background of Austin‘s speech act and the integration of cognitions of existing communication models. It results in the so-called semantic communication model.

6 Developing a Semantic Communication Model

This chapter aims in integrating the extended form of the semantic data model (Sieber/ Kammerer/Kovács) into a communicational situation. Hence, in the first section Austin's Speech-Act Theory is transferred to the typical communication situation in the area of technical documentation. The second section summarises analysed communication models and Section 6.3 illustrates the extended semantic communication model (Sieber/Lautenbacher)

References: The first section is based on (Lautenbacher, Sieber, Bauer&Cabral, 2007), the second and third section on (Sieber&Lautenbacher, 2007).

6.1 Technical Documentation as Write-Act

Technical documentation aims in transferring knowledge to people, i.e. the basic for technical documentation is successful communication. Communication in the first instance means speaking. This is why Austin describes explicit his understanding of communicating in his *Speech-Act Theory.*

Speech Act Theory (Austin)

John Langshaw Austin (cf. (Austin, 1962)) introduced an informal description of the idea of an illocutionary act that can be captured by emphasing that when we use language as more than a mere way to state things as true/false, we actually do the action being pronounced or denoted. To further explain this theory, Austin declared three types of speech acts:

- *Locutionary acts*: saying something (the locution) with a certain meaning but not necessarily building a speech. It may be a word, sentence or sound. There are three different kinds of these acts:
  - The locutionary act is at one level the production of certain noises and as such, it is called the *phonetic act*.
  - Through the production of those noises, the speaker produces words in syntactic arrangements and this act is therefore called *phatic act*.
  - Finally, through the production of words in syntactic arrangements, with certain intentions and in certain contexts, it conveys certain messages and is in this respect dubbed *rhetic act*.
- *Illocutionary acts*: the performance of an act in saying something, or the speaker's intent. John Searle developed further this category and identified five illocutionary points (cp. (Searle, 1969): assertives (true or false statements),

directives (statements with a certain intent), commisives (statements which commit the speaker to a course of action), expressives (express the sincerity of the speech act) and declaratives (statements that connote a change of the world referred by representing as already changed).

- *Perlocutionary acts*: These acts have a direct effect on sensitive perception, feelings or actions of both the speaker and the receiver. These acts seek to change a state of mind, an idea or feeling towards a representation.

These categories created by Austin and afterwards developed by Searle apply to all languages as we see them: understanding that by language we imply a system of communication consisting of sounds, words or characters used by two sources/ destinations to exchange information.

Transfer to Technical Documentation: Write-Act-Theory

In case of written exchange, it is possible to speak about a scribal act instead of a phonetic act, but the spirit behind the Speech-Act Theory remains the same, which makes it quite interesting for documentation engineering purposes.

Illocutionary acts are performed with intentions. They are communicatively successful if the hearer/reader recognizes the speaker´s/writer´s illocutionary intention. Illocutionary acts are intentional and are generally performed with the primary intention of achieving some perlocutionary effect.

By sending certain data in the form of words or written commands, we issue not just an order or perlocutionary act. The content of the statement is imbued with a certain meaning. In spoken languages, this meaning can be implied by the tone/intonation used to communicate, or by different signs, the speaker can send. In written documents, this is much more difficult since the future reader is mostly not known and therefore the meaning must be made explicit in other ways.

How to Ensure Meaning Transfer

The challenges for technical documentation have been already explained on page 41 by using the semantic data model resp. the sender recipient model: Every *communication* based on data instances has to consider that each person has a different background and knowledge and will derive from a data instance maybe different meanings. Therefore, the data instances have to be specified in a way the recipient can understand what they are meant for.

6.2 Communication Models

With respect to a further development of the semantic data model (Sieber/Kammerer) based on the question "How takes communication place in the context of technical documentation?" the following communication models has been analysed: Aristotle (cf. (Underwood, 2003) and (Honeycut, 2007)), Lasswell (Underwood, 2003), Gerbner (Underwood, 2003), Osgood&Schramm (Underwood, 2003), Berlo, Bühler (Bühler, 1969), Per Flensburg (Flensburg, 2007), Dance&Larson (Dance&Larson, 1976), Shannon&Weaver (Underwood, 2003), Becker (Becker, 1968) and Ruesch&Bateson (Ruesch&Bateson, 1994). The analysis is published in (Sieber&Lautenbacher, 2007).

The communication models from Shannon&Weaver, Berlo, and Osgood&Schramm come out on most suitable for the context of technical documentation and are summarised as follows.

(1) Osgood& Schramm

> *„The Osgood and Schramm circular model is an attempt to remedy that deficiency: The model emphasizes the circular nature of communication. The participants swap between the roles of source/encoder and receiver/decoder.“*
>
> (Underwood, 2003)

Osgood and Schramm (1954) describe communication as a circular process. Start and end are hereby not to define. As a further central element of this communication model, the interpretation of the communicated message comes to the fore. In that context the preferably effective reduction of semantic noise is the pre-requirement for any successful communication.

(2) Shannon& Weaver

Claude E. Shannon (1949) developed and used a communication model consisting of five elements:

1. Information source.
2. Transmitter or sender --> encodes the message into signal.
3. Channel that transports the signal from sender to recipient and can be influenced by noise.
4. Receiver --> decodes the signal into message.
5. The destination.

Shannon understood this model in a technical sense and not as explanation for human communication. Weaver extends the application area of the model:

„The word communication will be used here in a very broad sense to include all of the procedures by which one mind may affect another. ´[...]. In some connections it may be desirable to use a still broader definition of communication, namely, one which would include the procedures by means of which one mechanism [...] affects another mechanism [...] In oral speech, the information source is the brain, the transmitter is the voice mechanism producing the varying sound pressure (the signal) which is transmitted through the air (the channel). [...] When I talk to you, my brain is the information source, yours the destination;“

Shannon, Weaver 1949, cited from (Underwood, 2003)

According to this, communication can be regarded as a linear process in centre of which stands the signal. The principle of this model is therewith that each human communication has an information source acting as sender that is sending its message in form of a code using a channel.

(3)Berlo

„Berlo's approach is rather different from what seems to be suggested by the more straightforward transmission models in that he places great emphasis on dyadic communication, therefore stressing the role of the relationship between the source and the receiver as an important variable in the communication process.“

(Underwood, 2003)

Berlo developed a dyadic communication model with focus on the central role of the relation between sender and receiver in a communicational situation. He depicted five elements, that are influencing the fidelity within the source/coding and the receiver/decoding: the communicational abilities, the knowledge, the social system, the culture and the attitudes of the corresponding persons.

6.3 Transfer: Semantic Communication Model

Message

According to Shannon&Weaver the concept *message* is introduced in relation to the concept data instance (in terms of Sieber/Kammerer). Thereby, the technical documentation can be regarded as a *transfer of messages* with the intention of communicating knowledge between a sender and a recipient. This means also, that the technical author resp. the modeling methodology has to define exactly, what kind the message is:

- Is it only an atomic data instance?
- It it a complex data instance?

- Is it a data instance on object level that is described by a data instance on meta level?
- Etc.

In the further-development of the semantic data model, the concept of message is therefore introduced as "package" of connected and structured data instances on object- and meta level (see Figure 23).

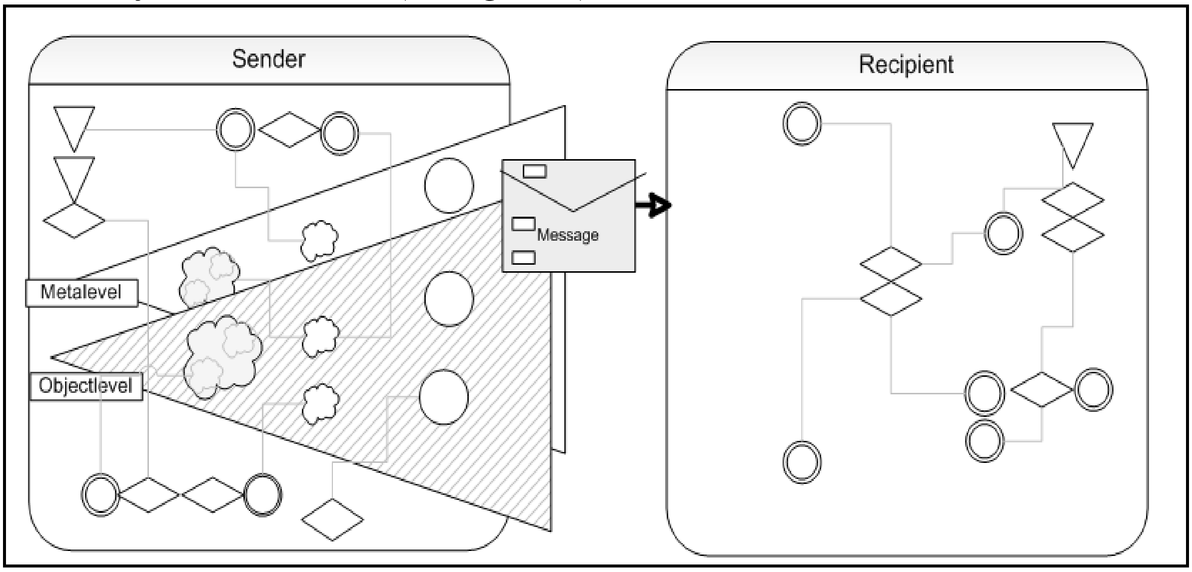


Figure 23 Semantic Communication Model: Message

Noise and Interpretation

In Section 4.3, the communication of knowledge is illustrated in relation to the developed semantic data model. In terms of Shannon&Weaver the problem in transferring messages are the existences of noise. According to the understanding of noise of Shannon&Weaver, this results in syntactical interferences. Following the proposed concept for noise from Berlo, the *semantic noise* plays a decisive role in the semantic communication model: The reduction of that semantic noise during message transfer enables the preferably fault-free interpretation (cf. to Osgood&Schramm) of the message.

6.4 Conclusion

Challenges for Technical Documentation

In the proposed semantic communication model, the concepts message and semantic noise have been introduced. The challenge for the technical author is therewith to chose and structure the documentation, i.e. the messages in such a way, that the semantic noise is reduced during the message transfer. This means that all kind of semantic-ensuring measures have to be taken, e.g. choosing appropriate writing styles, following an appropriate suitable structure within the documentation, ensuring the correctness of transferred data instances etc.

Requirements for Modeling Methodologies

Following the understanding of communication in the field of technical documentation (see Section 4.3), the process of "knowledge" transfer can be visualized as defacto the transfer of message. The complexity of the messages to be transferred reflects the thereby involved different data-levels (object and meta-level(s)) and different complexity degrees within the data-levels (atomic vs. complex data instances) used in this process.

The desired reduction of the semantic noise can be best achieved when the hereby-illustrated communicative frame is known and considered for the message transfer. For reuse purposes, which should be processed in a semi-automised manner, it will be therefore necessary, that this communicative frame (sender, recipient, their knowledge lattices, mappings for the data instances, attributes of the signals, structure of the message) is described in a consistent formal way.

The semantic communication model with the background of the multi-layer semantic data model can be used herewith to serve as reference model for technical documentation. It combines as well semiotic as information science cognitions and reflects the objective of technical documentation by understanding documentation as a write-act that can be illustrated by a communication model. It offers thereby beside the function of a reference model a base model for any kind of conceptual modeling of technical documentation.

As the described model is based on the idea of agents, the model is flexible for being used in human-human communication as well as in human-machine or machine-machine communication. It can therefore be used not only with the background of technical documentation and the exchange of messages between technical authors and future readers, but it can also serve as foundation for communication between any kind of agents, e.g. within service oriented architectures.

The developed model is founded on precise and consistent terminological definitions that constitute the most important concepts in the field of technical documentation – signs, data, knowledge, information, concepts and message. These terminological definitions can be used for any kind of discussion and further research work in the field of technical documentation as well as knowledge engineering.

The next chapter achieves the application of this semantic communication model as a reference model for the evaluation of methodologies for structuring technical documentation and ontology design with special regard to an optimisation of reuse within technical documentation.

Part III Reuse Function Deployment

Falsities exist not only because one does not know certain things, but also as one judges before knowing everything, that is thereto relevant.

Immanuel Kant, 1724 - 1804

7 Reuse within Documentation Engineering

This chapter introduces and gives an overview about the current trend in technical documentation – the industrialisation of writing. The following Section 7.2 illustrates concepts of reuse and introduces knowledge reuse as one of the existing challenges of knowledge management. This chapter ends with Section 7.3 pointing to the missing evaluation possibilities for methodologies that are evoked by the missing definition of requirements in the context of discussions about reuse and its embedding in documentation engineering.

References: The chapter is mainly based on (Sieber&Lautenbacher, 2007).

7.1 Industrialisation of Writing

As depicted in Chapter 2 the technical documentation consists of internal and external technical documentation. Both are strongly coined by existing bodies of rules and regulations, which have been developed according to product liability and quality ensuring measures. In the field of external documentation - influenced by the development and usage of new technologies and editorial systems - approaches for modularising the documentation have been evolved.

Single Sourcing

Content Management Systems are increasingly employed that do not charge with whole documents but manage the content separated from layout and structure. This separation allows a more flexible reuse. In the context of Content Management Systems exist the same terminological uncertainties and vague definitions as discussed in Chapter 4 for the terms data, information and knowledge. According to the semantic data model (Sieber/Kammerer), content is perceived in this thesis as defined in (Sieber&Lautenbacher, 2007):

Definition *Content*

Content ist als ein Sammelbegriff zu verstehen, der digitalisiert vorliegende Dateninstanzen eines bestimmten beliebigen Datums (i.S.v. Sieber/Kammerer) auf einer abstrakten Ebene inhaltlich zusammenfasst.

[Content is a collective term that subsumes digitally hold data instances of one specific arbitrary data item (in terms of Sieber/Kammerer) on an abstract level.]

This definition can be illustrated with the following example: In Figure 24 the data instances 1, **1** and *1* can be subsumed regarding their content. This is not the same like an abstraction regarding the meaning: all data instances can be subsumed regarding their meaning AGE.

.

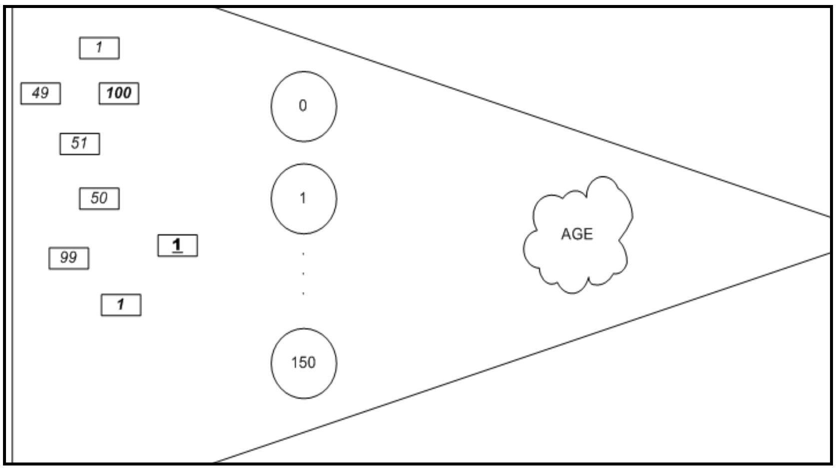


Figure 24 Example Content - Age

It is relevant for content, that only those data instances can be subsumed regarding their content, that belong to the same specific arbitrary data item: In Figure 24 and Figure 25 two different data items are given. It contradicts to the understanding of content and to the above given definition, that a subsumption regarding the content over all data instances from both data items can be executed.

.

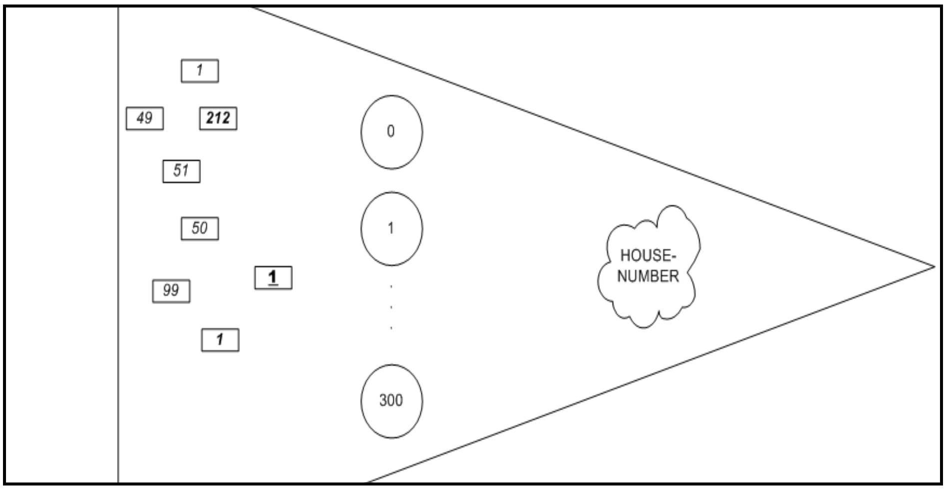


Figure 25 Example Content - House Number

In the Anglo-Saxon language, the central methods *Single Source Publishing* (SSP) and *Cross Media Publishing* (CMP) are subsumed under the term Single Sourcing.

> *"Single Sourcing implies that there is a single source for content; content is written once, stored in a single source location, and reused many times."*
>
> (Rockley, 2003)

For this thesis, we regard Single Sourcing as the subsuming term for both:.

Definition *Single Sourcing*

Based on a digitally hold data instance (in terms of Sieber/Kammerer) Single Sourcing supports that another data instance is created bounded to a desired medium. This new data instance concretises the same (as regards meaning and content) data item in the same or other representation.

Product Data Integration

This trend follows from the request to extend the spirit of Single Sourcing to the product life and development cycle. Concepts like *product data integration* and *management* are to close the gaps between product data and documentation. These challenges are easily solvable with the background of product databases and content management based editorial systems.

7.2 Reuse Strategies

Regarding the objective to optimize the reuse and thereby to enhance the cost effectiveness within documentation engineering Rockley and Ziegler investigated in defining possible reuse strategies.

Reuse (Rockley)

According to Rockley (Rockley, 2003) a digitally kept data instance can be reused either automised (*systematic reuse*) or manual (*opportunistic reuse*), for both of which exist three application scenarios:

- *Locked*: The reused data instance cannot be changed.
- *Nested*: The data instance is collected in one module and annotated with validities.
- *Derivative*: The reused data instance can be changed, derived data instances remain related with the source by a parent-child-relationship.

These reuse-methods may be regarded as basic concepts for reuse. Regarding a more effective management of granular content brickets, more precise reuse possibilities should be allocated. Different product variants are differing sometimes only in singular functionalities or components what provides for high reusability of according modules. Supplementary modules can exist that differ as regards content only in single sentences. For managing these modularized variants granular reuse-methods have to been provided in order to minimize the redundant administration of a large extent same modules.

Reuse (Ziegler)

A more detailed presentation of reuse-methods for that kind of modularised variants' management provides Ziegler in (Ziegler, 2005).

.

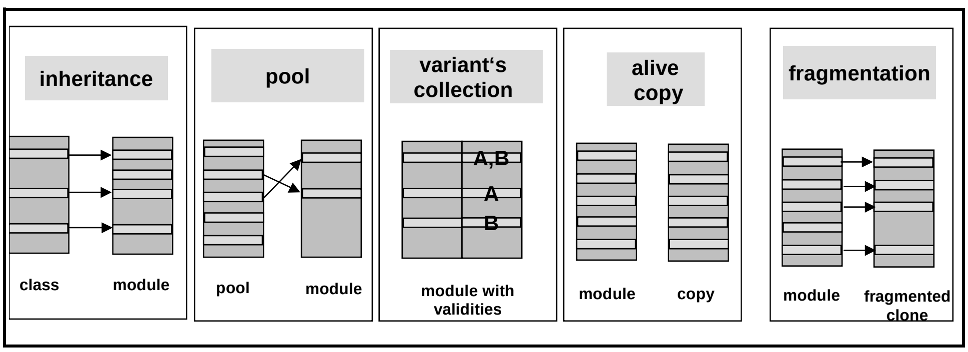


Figure 26 Reuse according to Ziegler

The several approaches for reuse according to Ziegler are (cf. Figure 26):

- *Inheritance*: For modules that show constant intersective elements, a module class can be defined. Each from this class derived module inherits automatically these basic and not changeable elements (systematic and locked reuse in terms of Rockley). Derived modules can be extended by further elements.
- *Reuse pool*: This method is based upon the opportunistic reuse (cf. to Rockley) of elements. The reuse pool contains all those elements that are provided for reuse in other modules. The required elements have to be manually selected and inserted as locked reuse into the new module.
- *Variants' collection*: Variants are administrated within one module (nested reuse). The automised evaluation of the validities happens only at publication time.
- *Alive copies*: This method is not regarded as reuse, the copied module is tagged with a copy evidence of the copy source. Changing in this source can be adapted manually in the copied module.
- *Fragmented reuse*: The combination of flexible intervention possibilities and controlled reuse is feasible by this method. The reused module contains several elements. These can be deleted or adapted within the fragmented clone. Only unchanged elements are further treated as reuses.

With the background of understanding technical documentation as write-act and the objective of it as the communication of knowledge (cf. Chapter 4 and 7), the presented reuse strategies lack the consideration of further-going reuse reflections. The reuse strategies cover only the reuse of data instances but not the challenges with regard to the integration of knowledge reuse.

Knowledge Reuse

In general, Sleeman identified six challenges for the management of knowledge (cf. (Sleeman, 2006) and (Rekowski (SAP AG), 2006)), that are summarised as follows formulating the challenges with special regard to the derived understanding of technical documentation as the communication of knowledge:

1. Knowledge *acquisition*: Suffer from a surfeit of data instances, still situations are faced, where the problem is insufficient or poorly specified knowledge. How can an appropriate authoring environment with reuse resources improve this?
2. Knowledge *modelling*: The real requirement of knowledge modelling is to understand the types of template and structure within which knowledge can be most usefully held, and reasoned with. The presented ontologies (see Chapter 3) offer such structures.
3. Knowledge *reuse*: One of the most serious impediments to cost-effective knowledge transfer is that dated knowledge is repeatedly built afresh. There is little reuse of existing domain content or of previous attempts to meet a similar knowledge transfer demand. How to enable and establish knowledge component reuse as a routine part of knowledge management and technical documentation?

4. Knowledge *retrieval and extraction*: In any large repository retrieval of knowledge is an issue. How is a subset of content relevant to a problem or task recovered? In some cases, the ideal process is not one of retrieval but dynamic extraction. Automated methods to support retrieval and extraction are as vital as architectures to integrate such capabilities.
5. Knowledge *publishing*: Assuming large repositories of well-structured, well-indexed knowledge exit, how best to disseminate this content in the right form to the right person at the right time?
6. Knowledge *maintenance*: The challenge is to know when and how to update, and to be able to predict the consequences of any changes.

According to (Markus, 2001) the synthesis of evidence from a wide variety of sources suggests four distinct types of *knowledge reuse situations* according to the knowledge reuser and the purpose of knowledge reuse. The types involve:

- Shared work producers, who produce knowledge they later reuse,
- shared work practitioners, who reuse each other's knowledge contributions,
- expertise-seeking novices, and
- secondary knowledge miners.

Each type of knowledge reuser has different requirements for knowledge repositories. Owing to how repositories are created, reusers' requirements often remain unmet. Repositories often require considerable rework to be useful for new reusers, but knowledge producers rarely have the resources and incentives to do a good job of repurposing knowledge. Solutions include careful use of incentives and human and technical intermediaries.

7.3 Conclusion

The reuse definitions according to Rockley and Ziegler assume a modularised modality how documentation is permuted. Along a product life cycle (cp. Chapter 2) this is not the case and is only permuted in consistent manner in the field of external technical documentation, if Content Management Systems have been implemented as editorial systems. Regarding the reuse of any sort of technical documents (e.g. requirement specification) for any kind of temporally down streamed engineering processes along the product life cycle modularisations of these documents and thereby enabled reuse in terms of Rockley/Ziegler constitute an important approach. Though neither Rockley nor Ziegler describe the requirements for the methodologies that should be applied for obtaining the desired reuse-optimised modularisations.

Solutions for the presented challenges within knowledge management are approached by the usage of ontologies (cp. (Sure, 2003) and Chapter „Foundations: Ontologies“ on page 19): Knowledge acquisition e.g. can be carried out through ontology modelling and a hereunto following semantic annotation. An ontology forms the conceptual structure of the knowledge base, and the "semantic" annotation populates the knowledge base with "semantic" instances. This kind of knowledge modelling and knowledge holding in suitable repositories should also heighten the possible reuse of knowledge. Neither Seelman nor Markus describe what are the requirements for ontology design methodologies that have to be applied for gaining the most effective reuse effect.

For both methodologies and their to fulfil requirements regarding reuse optimisations exist neither a reference model, that can be used for defining the requirements nor any other evaluation criteria methods.

In the next chapter the method *Quality Function Deployment* is introduced, which supports to capture requirements in a systematic way. The semantic communication model (Sieber/Lautenbacher) will be used as the developed reference model for discussing and visualising the concept *reuse* in order to derive thereout resulting requirements. The evaluation of methodologies for ontology-design and structuring documentation is presented in the chapter „Reuse Evaluation of Methodologies“ on page 73.

8 Developing Evaluation Criteria

This chapter depicts in the first section the concept *evaluation* and *Quality Function Deployment*. According to (Herzwurm, 1997), an adaptation of QFD is presented as evaluation method.

Section 8.2 illustrates the main groups for the requirements criteria that have been developed in alignment to the proceeding of applying the evaluated methodologies. This section ends in illustrating the usage of the semantic communication model as reference model for the definition of requirements criteria as regards reuse and personalisation.

References: The chapter is mainly based on (Sieber&Lautenbacher, 2007).

8.1 Evaluation Strategies

Design Science

According to House (House, 1993) the following definition for *evaluation* can be given:

Definition *Evaluation*

Evaluation means the specific benchmark of material and immaterial matters with the aid of criteria and methods of which the adequacy can be explained.

Since several years, research on evaluation mainly exists in the area of education and health care (cf. (Frank, 2000)). Also in the area of information management, the evaluation is increasingly focussed. Although it is described as the important element of each kind of research in the context of the so-called *design science* (Hever et Al., 2004), in the field of evaluating methodologies exist only few evaluation approaches like in (Moodey&Shanks, 1994) for the evaluation of the quality of Entity-Relationship-Diagrams or in (Hong, Gong&Brinkkemper, 1993) for the evaluation of object-oriented analysis and design methodologies.

The research specific area of evaluating methodologies for technical documentation is only tackled in (Ried, 2006): Ried himself titles the evaluation sheet (with seven developed criteria) as a trial of valuation and revised his valuations during his presentation several times influenced by the evoked discussions. The evaluation lacks of methodic approach for evaluating methodologies and of a reference model that reflects the technical documentation.

The developed semantic communication model serves as such a reference model. As follows, a methodic approach for capturing requirements criteria for evaluation purposes is proposed.

Quality Function Deployment

The Quality Function Deployment (QFD) centers the concept *quality* as basic for every design and manufacturing decision (cf. (Akao, 1990)). Quality is thereby directed to improve the value for the customers. 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. Besides other quantitative and qualitative methods like affinity and/or tree diagram are applied (q.v. (Brassard, 1989), (Bicknell&Bicknell, 1995) and (Erikkson&McFadden, 1993)). QFD enables as well the precise specification of customer requirements and needs as the systematic analysis of design criteria of products and services. To this, the customer requirements are correlated to the product criteria in order to achieve those product specifications that solve the problems (i.e. the customer requirements) best.

> *„Copy the spirit and not the form“*
>
> (Akao, 1990)

QFD was applied originally in manufacturing. As the method is not constituted on a theoretical framework with fixed and consistent definitions, the usage of QFD is not limited to support the product development. It can be regarded as a generic applicable analysis, planning and communication method. Increasingly the application of QFD is remarkable in other areas e.g. the software engineering.

According to Herzwurm, the QFD can be adapted for the evaluation of software (cf. (Herzwurm, 1997)). In Figure 27, the adapted House of Quality is illustrated.

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 interval-based proceeding for rating the degree that achieves the evaluated methodology as regards the fulfilment of the requirements, and
- the possibility to extend the developed HoQ for the design of a product that covers the determined requirements in a better way than the evaluated products.

.

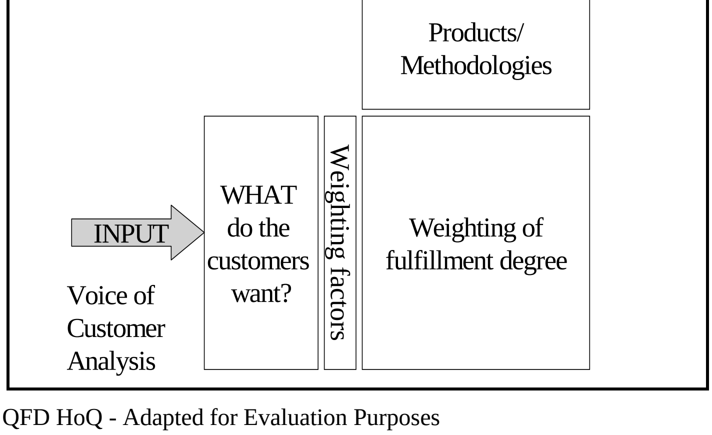


Figure 27 QFD HoQ - Adapted for Evaluation Purposes

For the development of the 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.

8.2 Development of Evaluation Matrix

The structure of applying methodologies (cf. (Sieber&Lautenbacher, 2007)) served to define the requirements criteria that should be analyzed in the evaluation. That way the criteria could be subsumed under the following main structure:

1. Pre-requirements for the application of the methodology.
2. Application of the methodology.
3. Quality of the resulted data model.
4. Quality of the future thereon based and created data instances.

The QFD-oriented evaluation matrix consists of five parts following this structure - as the 4th point is splitted into 4a personalisation and 4b reuse. As we worked with two persons on evaluating the methodologies, an evaluation guideline (see (Sieber&Lautenbacher, 2007)) was developed, which should formulate in an easy to understand and retrieve way, how to fill out the matrix and should serve in future for reflecting the different meanings of the degrees of to given points.

For the development of criteria in the first three parts the sources ((Hever et Al., 2004), (Moodey&Shanks, 1994) and (Hong, Gong&Brinkkemper, 1993)) has been of use. For the development of criteria for the forth part, the semantic communication model has been used as reference model.

Reuse within Semantic Communication Model

The methodology has to instruct the development of modules (the according future instances of those represent the future messages) by instructing definitions of their local and temporal anchoring, standardised terminology and inner structure for the future instances.

The elements, which should be formal described for enabling reuse, are grey high lightened in Figure 28:

.

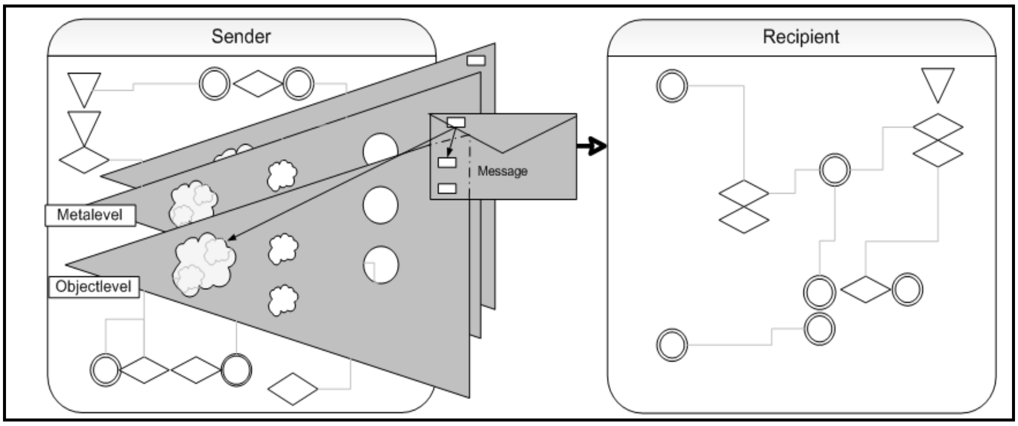


Figure 28 Criteria for Optimised Reuse

The result of the applied methodology - the resulting data model - has therefore to define the inner structure of each complex/atomic data item (on meta and object level) that is used for module creation.

As regards complex data item, it has to be defined

- of what kind of complex/atomic data items it consists,
- in which order those have to be arranged and
- how often they can occur.

As regards atomic data item, it has to be defined

- how this should be represented and
- how the inner lingual structure of the future thereon based data instance should be (writing rules, terminological rules).

The methodology has to provide explicit deriving possibilities as regards the structure of the message that will be sent to the future recipient. The resulting data model should be declarative as regards the dependency of complex/atomic data items from the communicative frame.

For the temporal anchoring, the methodology should guide the user of the methodology to considerations concerning the temporal dependency of future data instances. This can happen by introducing a special meta level, which reflects the usage of variables that can generate automised the future instances on demand by importing the according data instances.

Personalisation within Semantic Communication Model

For future preferably automised possibilities for personalisation (as a special kind of reuse), the communicative frame has to be described in those parts that allow logical conclusions concerning the facts, what kind of instances have been originated in what communicative frame and can be used for what kind of other communicative frame.

For that kind of conclusions, the knowledge lattices of the sender and the recipient are relevant as well as the existence and definition of a role concept. This means - applied on the semantic communication model - the therefore needed existence of a meta data that is provided for system internally controlling- and deriving purposes and is not sent to the recipient within the intrinsic message

.

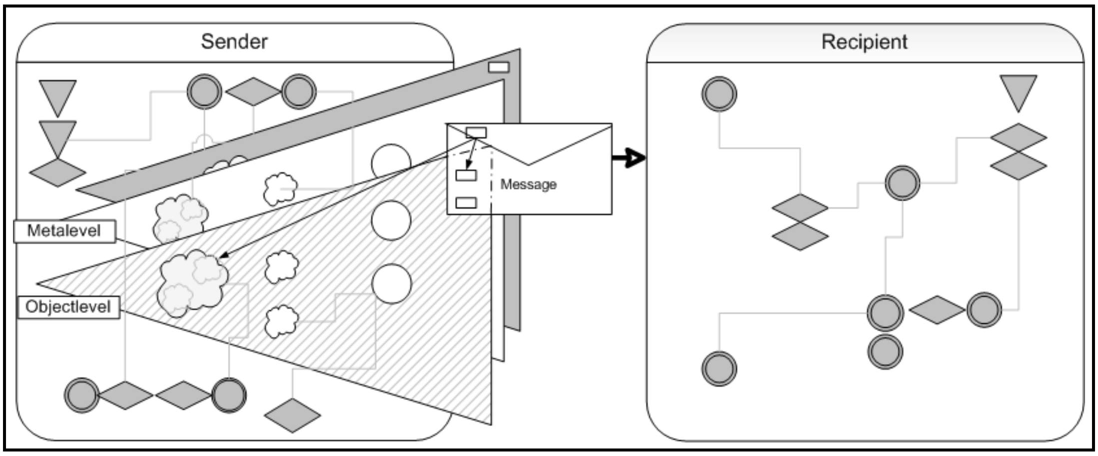


Figure 29 Criteria for Optimised Personalisation

The elements, which should be formal described for enabling personalisation, are grey high lightened in Figure 29:

- The assumed knowledge lattice of the recipient,
- the knowledge lattice of the sender, and
- the inclusion of a role concept on a metadata level.

Developed Materials

The according *evaluation matrix* and the developed *evaluation guideline* have been published and are online available in (Sieber&Lautenbacher, 2007).

In Table 1, an example of the detailed description of requirement criteria in the developed evaluation guideline is illustrated: The criteria are formulated as questions, the priority is already inserted and for an easier completion of the evaluation matrix the extremum meanings of 1 resp. 5 points are formulated in a textual way.

The evaluation matrix contained only the criteria and the priorities and had to be filled out for each to evaluate methodology in correspondence to the developed guideline.

.

Criteria	Prio (1=low, 2=mid, 3=high)					
1. Pre-requirements for the application of the methodology		**1**	**2**	**3**	**4**	**5**
1.1 Are special precognition prerequired?	2	Comprehensiveness and specifity of prerequired cognitions lies far beyond of what can be expected from the addressed intentional group				no specific cognitions prerequired that lie beyond the to be expected cognitions of the addressed intentional group

Table 1 Example: Requirement Criteria Description in Evaluation Guideline

8.3 Conclusion

The developed guideline can be used for any evaluation of methodologies. The developed structure allows easy and flexible adaptations in accordance to the focus of the evaluation. Whereas the first three parts contain general requirements for any kind of methodology, the last two parts have been developed with special regard of reuse optimisations in technical documentation. The semantic communication model serves therewith as the reference model that allows the development of the requirement criteria. As any kind of evaluation matrix, the developed one reflects subjective criteria. By the methodic approach guided by the spirit of QFD, it was tried to minimise them as well as by the usage of the developed reference model. The evaluation matrix together with the semantic communication model was presented at several meetings within the SAP AG and it served as a very good foundation for any kind of discussion about data modeling in general and reuse in particular. In the next chapter, the evaluation and the results are illustrated.

9 Reuse Evaluation of Methodologies

Within the BAYHOST-founded project ECIDISI, the evaluation of methodologies for structuring technical documentation and methodologies for ontology-design has been executed. This chapter reflects in Section 9.1 how the evaluation has been performed and summarises in Section 9.2 the evaluation results.

References: The chapter is mainly based on (Sieber&Lautenbacher, 2007), the technical report of the ECIDISI-project.

9.1 Performance of Evaluation

The following concrete use case was defined before carrying out the evaluation (cf. (Sieber&Lautenbacher, 2007)):

Use Case ‚cancel a credit card'

Sender1: requirements engineer.

Intentional group / recipient 1a: software developer and 1b: technical writer.

Sender2: technical writer.

Intentional group/ recipient 2a: bank employee and 2b: customer.

Pre-requirements for customer: loss or theft of credit card.

Pre-requirements for bank employee: call by customer.

- customer calls bank employee
- customer authentification (own use case)
- bank employee cancels credit card
- Either:
  - customer demands new card. bank employee sends customer the demanded card, customer activates credit card (own use case).
- Or:
  - bank employee waits 3 weeks. bank employee inquires again the customer.
  - Either: customer countermands the cancellation of credit card. bank employee removes blicking.
  - Or: customer demands new card. bank employee sends customer the demanded card. customer activates credit card (own use case).

Selection of Methodologies

From the analysed methodologies for structuring technical documentation (cf. Section 2.5), the following have been evaluated:

- Muthig/Schäflein-Armbruster: Function Design™ as representant of a methodology based on the functional principle (q.v. „Principles for Structuring Technical Documentation“ on page 11).
- Horn: Information Mapping™ representing a methodology based on psychological cognitions.
- OASIS: DITA as the only standard amongst the methodologies.
- Ley: Structured text control as the newest developed methodology integrating both principles - the content and the functional.

From the analysed methodologies for ontology-design (cf. Section 3.5), the following three have been selected for the evaluation:

- Sure, Staab et Al.: OTKM as representant of a well-described process-overview.
- Gómez-Fernández: METHONTOLOGY as promising methodology covering all aspects of ontology lifecycle.
- Noy-McGuinness: Ontology Development Guide 101 as guideline for the development of concepts.

All selected methodologies have been applied for the illustrated use case. For the implementation of the thereby developed ontologies, Protégé has been used.

The results of the application as well as the filled out evaluation matrix have been published in the technical report of the ECIDISI-project (Sieber & Lautenbacher, 2007). In Table 2 an example of the filled out evaluation matrix for the methodology Function Design™ is illustrated.

For each criterion, points have been given according to the evaluation guideline. For each part of the evaluation matrix, the according sum was counted. That way the filled out evaluation matrix for the several methodologies can be compared for each single criteria, or for the several desired parts or for the whole matrix.

All the filled out evaluation matrices are published and online available in (Sieber&Lautenbacher, 2007), where the detailed experiences in applying the methodologies are also described. As regards the fulfilment of criteria concerning reuse and personalisation, in the next section the collected experiences and evaluation results are summarised.

.

Methodology:	**Prio (1=low, 2=mid, 3=high)**	**Function Design™**	
Criteria		**Points (cp. evaluation guide)**	**Total**
1. Pre-requirements for the application of the methodology			**37**
1.1 Are special precog-nitions pre-required?	2	5	10
Who is the addressed intentional group?		technical writers, product engi-neers with main focus documenta-tion/communication	
What kind of precogniti-ons is prerequired?		basic interest on structuring methods is advantageous	

Table 2 Example: Filled Out Evaluation Matrix

9.2 Results: Evaluation of Methodologies

Evaluation Methodologies for Structuring Technical Documentation

As all methodologies aim in structuring technical documentation in a modularised way, it is assumed that they fulfil largely the criteria for personalisation and reuse. Concerning the personalisation criteria, all methodologies lie with 14 of 40 reachable points behind these expectations. Concerning the reuse criteria, DITA leads with 66 from 70 reachable points and covers thereby 94% of the criteria. In comparison to the other methodologies, DITA achieves this by the high degree of standardisation and the embedding of the modularisation into a complete framework. This framework assigns standardised blocks to specific characteristics and behaviours. This benefits compared to the pure structuring methodologies Function Design™ and Information Mapping™, which are bottom of the evaluated methodologies with 44 resp. 45 reached points.

Remarkable in the application of the linguistic methodologies has been that – even though the intentional group(s) and their needs have to be analyzed – this developed and worked out knowledge needs not to be documented or formal described. Such the bad results of all methodologies concerning the personalisation criteria can be explai-ned.

Degrees of freedom concerning naming and design of future instances are prejudicial to reuse. Methodologies like DITA benefit here from their standardised constraints. The development of Content Management Systems is reflected in the lately developed methodologies DITA and structured text control. They integrate considerations regar-ding versioning and validities of worked out modules. In the overall-evaluation, DITA performs best.

Evaluation Methodologies for Ontology-Design

Concerning the personalisation criteria, all methodologies reached only 12 of 40 reachable points. As regards the reuse criteria the results differ marginally, the Ontology Development Guide 101 performs hereby best with 20 out of 70 reachable points. In the overall-evaluation, METHONTOLOGY and Ontology Development Guide 101 perform similar. A combination of these two methodologies should cover concise basis for the development of concepts and the considered ontology lifecycle. All methodologies offer very limited possibilities regarding the reuse and personalisation. A further-development, which considers these aspects, is desirable.

9.3 Conclusion

Capture of Requirements for Modeling Methodologies

From the developed semantic communication model (Sieber/Lautenbacher) the criteria for reuse and personalisation of technical documentation have been evolved. This model reflects the communicative frame in which each kind of technical documentation is embedded (cp. „Requirements for Modeling Methodologies“ on page 62). It is therefore appropriate to be considered as a reference model for evaluation purposes in the field of technical documentation. For the creation of the criteria the Quality Function Deployment (QFD) has been adopted for evaluation purposes. Therewith a supporting methodic approach for the definition of the requirement criteria was gained. The obtained matrix has been extended in form of a guideline implementing textual meanings for the points to be given. This helped a lot during the evaluation as well as later for discussing the results.

Results of Evaluation

The evaluation shows that none of the evaluated methodologies has the ability to fulfil all desired criteria that aim in covering and describing in a consistent and formal way the communicative frame of the semantic communication model. This is not astonishing as none of the methodologies was originally developed in alignment to any kind of reference model for technical documentation.

The lack of such a reference model is also noticeable in the field of those methodologies that have been developed for the design of ontologies. Even if their focus should be on guiding the design of reasonable concepts, the methodologies lack in giving concrete guidelines, how to do this.

It is further remarkable, that ontologies are able to describe the desired sender-recipient knowledge lattices, but none of the methodologies instructs to do so. With an integration of this ability of ontologies to capture the formal description of knowledge lattices and role concepts into DITA, the communicative frame would be completely formal covered.

Baseline for the Development of an Reuse Optimised Methodology for Technical Documentation

The developed Reuse Function Deployment constitutes the basic for a further development, which aims in utilising the Quality Function Deployment for anchoring the requirements for reuse optimisations within technical documentation processes. This can be done by the deployment of measures and metrics, which are meeting the deployed requirements.

Part IV Intelligent Content Syndication

> An investment in knowledge pays the best interest.
>
> *Benjamin Franklin, 1706 - 1790*

10 Developing a Lean Documentation Model

The focus of the thesis in Part 2 and 3 has been the development of a reference model for technical documentation and the development of an evaluation method for the evaluation of methodologies in the context of technical documentation with special regard to *reuse* optimisations.

In the following chapters, the focus is on the *cost-efficient processing* of technical documentation along a product development cycle. The author provides in this chapter therefore the mapping of concepts from the Lean Production on technical documentation. The Lean Production consists of several methods for production processes. These are briefly introduced in Section 10.1. In Section 10.2 is shown that technical documentation can also be regarded as a production process, before depicting the measures to be taken within a Lean Production in Section 10.3 and mapping them on measure for Lean Documentation.

10.1 Lean Production

The concept of 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 continuous 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.

10.2 Lean Documentation

Analogue to Lean Production "lean" in the context of documentation can be understood in the first instance as the reduction on the essential by avoidance of all redundant actions. This means not compulsory the reduction of the amount of documentation. The Lean Documentation has to be regarded as a holistic approach for the optimisation of processes, tools and products of documentation and consists - as the Lean Production - of several concepts, which are contributing to the increase of efficiency. The objective of Lean Documentation can be substantiated in gaining a higher output of high qualitative content in a shorter time (cf. Baker, 2005).

In order to define more precise the possible measures to establish Lean Documentation as follows documentation is mapped to a production process.

Documentation as Production Process

The objective of production is the delivery of a defined product to a customer. Therefore, to some raw materials new, desired attributes are added during several assembly steps. These several steps are strung together according to the production specifications. After quality ensurance tests the product is delivered to the customers.

Whereas the production mainly deals with concrete goods, the documentation is attented to the abstract good of "information" (cf. Lemma on page 38). What the raw materials for the production process are the data instances in form of specifications, design papers etc. for the documentation process. For concretising the analogies between the two processes the author opposes in the following table the process relevant terms of the production to those of the documentation (cf. (Schott, 2006)).

Production	Documentation
Raw materials	Specifications, design papers, etc.
Production	Writing/construction of documentation
Production flow	Data/Publishing flow
Supplier	Data instances-suppliers
Production specifications	Editorial guidelines, documentation standards
Tools (mostly mechanical)	Tools (mostly computer-based)
Assembly steps	Editorial actions (research, writing, editing, generating)
Semi finished parts	Raw texts, re-usable text elements
Product	Documentation incl. medium
Quality inspection	Quality inspection
Customer	Intentional group
Stocking	Archiving

Table 3 Production vs. Documentation

10.3 Mapping of Lean Production Measures

Kaizen

> *„Das Unternehmen unterliegt nicht mehr der Illusion, alles perfekt machen zu können, sondern es weiß, dass alles verbesserungsfähig ist.“*
>
> (Biehal, 1993)

[The enterprise cherishes no longer the illusion to be able to make everything perfect, but it knows that everything is improvable.]

The employees are asked to create this process actively by handing in improvement proposals. The realisation of numerous small improvements demands high discipline from the employees to avoid falling back in 'old' patterns. The motivation of the employees is therefore the main key for success. Transferred to the documentation the possibility to hand-in and realise improvement proposals concerning processes, tools or products should be established and assisted.

According to (Ohno, 1993) standardisation is the basic for flexibility. A new standard should be created by improving an established one. This means that standards have to be developed parallel to processes, tools and products and that they are not unalterable. This assures that standards follow the actual requirements.

The Total Quality Management (TQM) applies the Kaizen principle to quality standards. The approach of TQM defines quality as correctness and optimisation of customer conveniences. Whereas the correctness describes the objective quality with regard to standards and specifications, the customer convenience represents the subjective quality sensation. Only with this dual definition of quality, improvements become feasible as quality is no longer interpreted as a long list of defined characteristics, but involves the real needs of the customers (cf. (Bach, 1993)).

As follows, the author summarises measures for supporting the Kaizen principle in production and documentation (in Table 4).

Principle	Explanation	Lean Production	Lean Documentation
Kaizen	Continuous improvement of processes, tools and products	Support and motivation for employees for handing-in of improvement proposals	Support and motivation for employees for handing-in of improvement proposals
		Standardisation by utilisation of normed parts and tools	Standardisation by editorial guidelines and documentation standards
		Total Quality Management	Quality testing

Table 4 Support of Kaizen

Jidoka

Jidoka stands for autonomous automation viz. the production process is stopped via sensors, controllers or operating personal by occurrences of irregularities in order to avoid deliverance of defective parts for following processes. The fault can be localised easily and thereby prospective occurring problems can be solved long-term and in advance by quality ensurance measures. The basic approach of this principle is to optimise the processes in such a way that they become more and more robust to faults. In analogy to measures that serve and support Jidoka in production processes measures are listed which support Jidoka in documentation in Table 5.

Principle	Explanation	Lean Production	Lean Documentation
Jidoka	Intelligent Automation	Automation of product flow	Automation of data/publishing flow
		Development of fault-robust processes	Development of consistent templates and processes

Table 5 Support of Jidoka

Muda

„Hauptziel und Kern der Lean Production ist es, Verschwendung jeglicher Art erfolgreich zu beseitigen. Unter Verschwendung ist hierbei all das zu verstehen, das die Kosten erhöht ohne einen sinnvollen Beitrag zur Produktion oder Dienstleistung zu leisten, also ohne eine sinnvolle Wertschöpfung zu erbringen.“

(Traeger, 1994)

[Main objective and kernel of Lean Production is the elimination of each kind of waste. Waste means thereby everything that increases the costs without serving reasonable contribution to the production or service viz. without performing reasonable value adding.]

Principle	Explanation	Lean Production	Lean Documentation
Muda	Avoidance of waste	Avoidance of over-production	Avoidance of needless, redundant and too long texts
		Avoidance of transport times	Central content/document repository management
		Avoidance of waiting times	Intelligent role and rights management within editorial systems; intelligent data/publishing flow
		Avoidance of needless processing	Avoidance of needless editorial work by usage of templates and streamlining of standards
		Avoidance of needless stock of inventory	Minimisation of time between development and documentation
		Avoidance of needless motion sequences	Automation by usage of appropriate editorial systems
		Avoidance of reworking	Avoidance of reworking

Table 6 Support of Muda

This basic principle can be met within the Lean Documentation by above listed measures (see Table 6).

Just-in-Time

The Just-in-Time principle can be divided into Kanban (intelligent suppliance system) and Heijunka (constant production levelling).

> *„Kanban ist eine Methode zur Umsetzung des Just-in-Time Systems [...] Auf dieser Grundlage beginnen die Arbeiter eigenständig zu arbeiten und entscheiden selbst über ihren Arbeitsrhythmus [...]"*
>
> (Ohno&Taiichi, 1993)

[Kanban is a method for the implementation of the Just-in-Time system [...] On this basic the employees start working independent and decide on their own their working rhythm.]

The basic idea of the Kanban principle is the consideration that the whole production process should be controllable in the reverse direction: the end-assembly is regarded as starting point of the whole production process and controls all upstream processes. These upstream processes are instructed by their downstream processes to produce exactly as much parts as the downstream processes need. The whole production process conforms to the actual need instead to estimated values.

Heijunka describes the sequential work planning and the thereout resulting production levelling.

Principle	Explanation	Lean Production	Lean Documentation
Just-in-Time	Demand orientated process flow	Production of parts on demand	Deliverance of texts and text elements on demand
		Sequential work planning	Reasonable project and process planning: definition of the several documentation-relevant activities (e.g. text writing, editing, reviewing, adapting ...) as processes

Table 7 Support of Just-in-Time

The introduction of Lean Documentation has to move the documentation from the passive into the active role. Data has not to be simply processed but has to be actively asked for by demand. The transfer of this trivial sounding basic principle requires fundamental, internal changings.

> *„Lean-Grundsätze ermöglichen nicht nur Kostensenkungen und Qualitätssteigerungen, sondern stabilisieren auch die Arbeitsabläufe eines Unternehmens und bringen Angebot und Nachfrage in Einklang. Dadurch wiederum wird es möglich, das ständige Krisenmanagement zu beenden und stattdessen die erforderlichen Voraussetzungen für kontinuierliche Verbesserungen zu schaffen."*
>
> (Drew, McCallum&Roggenhofer, 2005)

[Lean-principles enable not only cost reduction and quality increases but stabilise the workflows in the enterprise and align supply and demand. Therewith it becomes feasible to end the permanent crisis management and rather concentrate on managing the necessary pre-requirements for continuous improvements.]

Supplier Relationship and Customer Orientation

This is also reflected in the above listed measures meeting the principles of supplier relationship and customer orientation:

Principle	Explanation	Lean Production	Lean Documentation
Supplier Relations		Supplier delivers only those parts that are needed	Authoring and delivering of raw texts only on demand
		Fostering own solutions by contract placings in form of boundary conditions and not exact construction details	Fostering self-reliant document compilation beyond the intrinsic documentation departments by supporting document guidelines and templates
Customer Orientation		Production process follows customer needs	Documentation process follows the needs of the intentional groups

Table 8 Support of Supplier Relation and Customer Orientation

Andon

Andon means the visual control based on display panels. The background is that each employee can see the actual target-performance comparison. Mapped to the approach of Lean Documentation, editorial systems could demand - according to the captured text status - additional data from other departments. A positive side effect of such a visual control is the visible workflow and the possible exertion of influence on the document compilation process. This increases the awareness for documentation and engages the motivation and the cooperation (Baker, 2005).

Principle	Explanation	Lean Production	Lean Documentation
Andon	Visual control	Panel usage with target-performance comparisons	Usage of text/documentation status for visualising the documentation process

Table 9 Support of Andon

10.4 Operational Aspects

Ohno describes the implementation of the Toyota Production System (TPS) as follows:

> *„Das Topmanagement muss seine Denkart ändern und sich dafür einsetzen, den konventionellen Fluß [sic] der Produktion, des Transports und der Belieferung mit Teilen umzukehren. Diese Forderung stößt meist auf starken Widerstand, und die Umsetzung erfordert Mut. Je engagierter sich das Management jedoch dafür einsetzt, desto erfolgreicher wird die Einführung des Toyota-Produktionssystems sein.“*
>
> (Ohno, 1993)

[The top management needs to change their way of thinking and to campaign for a reversal of the conventional flow of production, transport and parts' supply. This claim evokes great resistance and the transformation requires courage. The more dedicated the management acts the more successful the implementation of the TPS will be.]

This statement underlines the importance of the holistic approach of the Lean Production resp. every other lean application area. As the concept of Lean Production relies on the multidisciplinary cooperation, the usual 'department thinking' needs to be overcome:

> *„Abteilungsdenken hat in der Lean Production nichts zu suchen, ebensowenig [sic] das bewußte [sic] (oder unbewußte [sic]) Zurückhalten von Informationen, um die eigene Position zu stärken.“*
>
> (Traeger, 1994)

[Department thinking will lead to nothing in the Lean Production nor the conscious (or unconscious) retention of information in order to strengthen the own position.]

10.5 Conclusion

Technical Documentation: Ready for Lean Concept

With regard to technical documentation the basic element for such an unhindered data/information flow is the elimination of or intelligent bridging between stand-alone solutions concerning data formats and already or to be applied modeling methodologies.

As analyzed and shown in the Chapters 2, 3, 7 and 9 the actual implementations of processing data is far away from such a desired Lean Documentation optimised workflow: Implemented Content Management Systems or other editorial systems are used stand-alone in the area of external technical documentation with only some touchpoints of integration of product data. The consciousness of preparing and value-adding data instances is mainly missing and not at all supported in the internal technical documentation processing.

Cost-Efficiency and Reuse of Technical Documentation

The mapped principles and to be taken measures for Lean Documentation demand

- the analysis and emphasis of the process characteristics of documentation and
- the consistent avoidance of waste along this process, i.e. the consistent implementation of intelligent reuse strategies.

For the last point, the developed Reuse Function Deployment can serve for analysing suitable modeling methodologies for reuse optimisations. Furthermore, the developed Lean Documentation model constitutes a basic for improving continuously the technical documentation along a product lifecycle by enabling a cost-efficient Lean Production of documentation.

In the next chapter, the background of the Transaction Cost Theory is presented with special regard to possible cost minimisations of technical documentation processing.

11 Transaction Cost Theory

This chapter introduces the *Transaction Cost Theory* and explains how it can be transferred to technical documentation. In Section 11.1, the theory is therefore introduced and the relevant concepts briefly explained. Section 11.2 maps the main principles of the theory to technical documentation. In conclusion, it is argued for the necessity to develop a framework including the approaches from this theory and the Lean Production established on the developed semantic communication model in order to gain the desired optimisation of reuse along a product life cycle. The conceptual framework is presented in the next chapter.

11.1 Transaction Cost Theory

The *Transaction Cost Theory* analyses, what kind of transactions can be most efficiently operated in which institutional arrangements. Transaction cost theorists assert that the total cost incurred by a firm can be grouped largely into two components — transaction costs and production costs. Transaction costs, often known as coordination costs, are well defined as the costs of "all the information processing necessary to coordinate the work of people and machines that perform the primary processes," whereas production costs include the costs incurred from "the physical or other primary processes necessary to create and distribute the goods or services being produced". The focus of the Transaction Cost Theory is to find out those institutional arrangements, in which the cumulated transaction- and production costs are the lowest.

Transaction Costs

Transactions are the core elements of the Transaction Cost Theory (cf. (Williamson, 1985)). A transaction means hereby the transfer of so-called rights of disposal of services. This transfer preceeds the material exchange of goods (cp. (Picot&Dietl, 1990) and (Picot, 1982)): In economics and related disciplines, a *transaction cost* is a cost incurred in making an economic exchange. A number of different kinds of transaction costs exist:

- Ex-ante transaction costs:
  - *Search and information costs* are costs such as those incurred in determining that the required good is available on the market, which has the lowest price, etc.
  - *Bargaining costs* are the costs required to come to an acceptable agreement with the other party to the transaction, drawing up an appropriate contract, etc.

- Ex-post transaction costs:
  - *Policing and enforcement costs* are the costs of making sure the other party sticks to the terms of the contract, and taking appropriate action (often through the legal system) if this turns out not to be the case.

Transaction costs consist of costs incurred in searching for the best supplier/partner/customer, the cost of establishing a supposedly "tamper-proof" contract, and the costs of monitoring and enforcing the implementation of the contract.

Transactions are efficient, when actors organise them in such a way that they have –in comparison to other contract resp. organisational forms – the lowest transaction and production costs (cp. (Williamson, 1991) and (Ebers&Gotsch, 1995). It is assumed that actors select that form of organisation, which is correlated to the lowest overall costs.

Attitude Assumptions

The Transaction Cost Theory acts on the specific idea of man that is expressed in the following assumptions of attitudes (cf. (Williamson, 1985) and (Ebers&Gotsch, 1995)):

- *Limited rationality* - means that actors behave only limited rational due to cognitive restrictions in assimilation and processing.
- *Opportunism* - actors behave strategic against their contract partners, i.e. they try to achieve their own interests. Williamson describes this behaviour as „selfinterest seeking with guile“ (Williamson, 1981)
- *Risk neutrality* - the involved actors stand in a neutral way vis-à-vis and do not differ in taking risks.

Based on this attitude assumptions Williamson advises:

> *„Organize transactions so as to economize on bounded rationality while simultaneously safeguarding them against the hazards of opportunism“*
>
> (Williamson, 1985)

Characteristics of Transactions

> *„The principal dimensions with respect to which transactions differ are asset specificity, uncertainty, and frequency. The first is the most important and most distinguishes transaction cost economics from other treatments of economic organizations, but the other two play significant roles“*
>
> (Williamson, 1985)

With background of the presented attitude assumptions, the costs of a transaction are mainly influenced by three factors:

- *Specificity,*

- *uncertainty,* and
- *frequency.*

Institutional Arrangements

Transaction cost economics suggest that the costs and difficulties associated with market transactions sometimes favour *hierarchies* (or in-house production) and sometimes *markets* as an economic governance structure. An intermediate mechanism, called *hybrid* or relational, between these two extremes has recently emerged as a new governance structure.

- The market is the most efficient institutional arrangement, if transactions are not correlated with high uncertainty or transaction-specific costs.
- In-house solutions lend theirselves, if high transaction-uncertainties exist, particularly, if high gainings are possible out of opportunistic behaviour.
- In-between these boundary solutions so-called hybrid solutions as long-term contracted cooperations come into consideration.

11.2 Conclusion

Requirements for Technical Documentation

Transferred to technical documentation the consideration of the presented Transaction Cost Theory demands precise descriptions of the desired process chain of documentation. For such processes, the transaction cost relevant attributes specificity, uncertainty and frequency can be analysed. Based on such analysed transaction costs and under consideration of the production costs it can be traceably decided, how the value-adding processing of data instances could be executed the most cost-efficient way.

Together with the developed Lean Documentation model, the Transaction Cost Theory presents the approach to a paradigm of technical documentation that is introduced in the next chapter: measures for the *Intelligent Content Syndication* at SAP AG have been implemented as framework for technical documentation, which supports a cost-efficient value-adding chain of technical documentation along a product development cycle.

12 Application for SAP AG

In this chapter, it will be argued for two kinds of paradigm shifts that needs to be integrated and implemented in order to achieve the so-called *Intelligent Content Syndication*. In Section 12.1, the therefore required shift in *editorial reorganisation* is illustrated and in Section 12.2, the necessary shift as regards *operational aspects*.

According to the credo of Akao („copy the spirit and not the form“) the Intelligent Content Syndication constitutes a conceptual framework that combines the cognitions gained in this research work and is strongly influenced by the new developed Lean Documentation model.

12.1 Paradigm Shift: Editorial Reorganisation

Tradeoff between Re-Usable and Customised Communication

As depicted in Chapter 2 today's challenges of technical communication departments are:

- Producing a continuously increasing volume of technical documentation,
- producing more, and faster, the product accompanying documentation,
- end users are not a uniform group:
  - different product knowledge,
  - different backgrounds,
  - different reasons for using a product.

Summarised there is the need of specific, personalised documentation rather than a standard one-size-fits-all document.

There are solutions like Single Sourcing (cf. Section 7.1) embedded in Content Management Systems facing this challenge by offering the reuse of existing content. Nevertheless quite often, there is a remarkable trade-off between re-usable and customised communication.

.

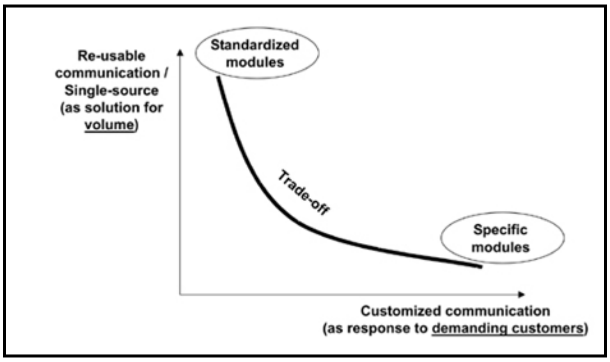


Figure 30 Tradeoff between Re-usable and Customised Communication

Adding Intelligence to Modules

In order to break the trade-off between reuse and personalisation, appropriate methodologies have to be applied that enable

- the production of “universal” modules that can be reused easily
- while maintaining the ability to personalise the modules.

The evaluation of existing methodologies (cf. Chapter 9) has shown that none of the evaluated methodologies fulfils the therefore desired requirements. A combination of DITA as the fundamental basic for modularising technical documentation and the application of Ontology Development Guide 101 promises the desired fulfilment of disposed requirements.

Content Syndication Architecture

The paradigm shift to an Intelligent Content Syndication means therewith

- ontology-based modeling of metadata,
- linguistic-based modeling of data,

and as a consequence of that

- the systematically separation of:
  - standardised acquisition of data instances in modules,
  - (semi-) automised metadata creation and annotation in standardised way (using metadata models and guidelines),
  - reuse supported by (semi-) automised content syndication processes (metadata/model-driven),
  - customised construction of documents (supported by automised content syndication processes), and
  - publication of documents in target format and media (see Figure 31).

.

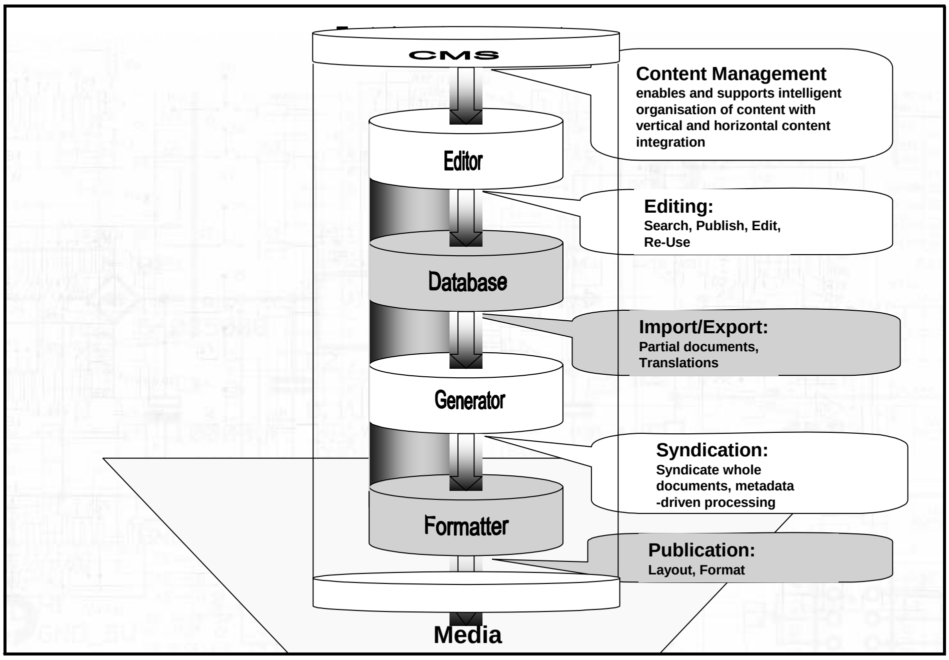


Figure 31 Content Syndication Architecture - CMS

.

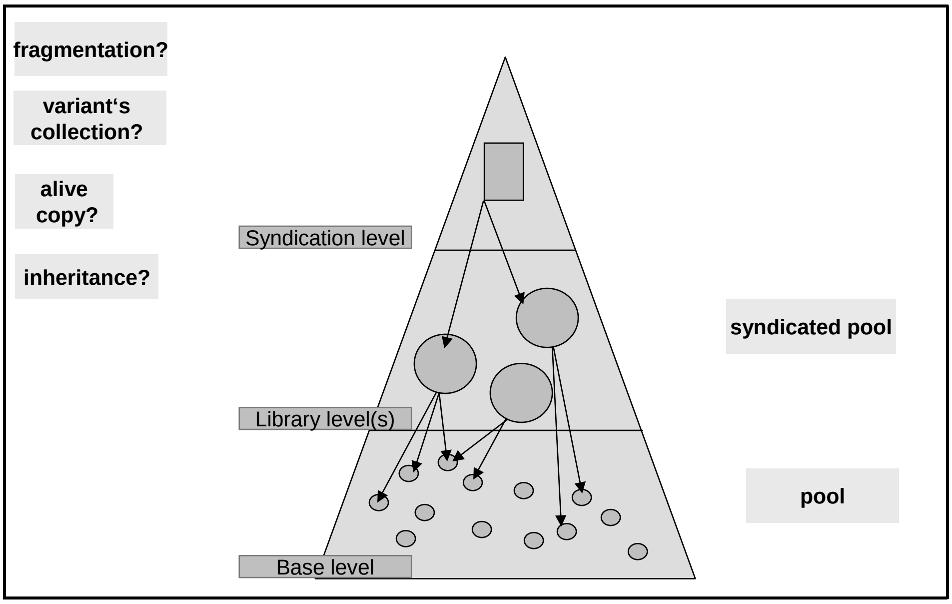


Figure 32 Content Syndication Architecture - Overview

This requires in the first instance the editorial reorganisation from document-writer to knowledge-collector and distributor (cf. Figure 32).

Besides, the supporting system needs to be set up in accordance with the desired content syndication process. To give an example, according to the content syndication architecture in Figure 32 the system has to enable and support the desired processing of pool elements on base and library level according to their assigned reuse characteristics. As shown in Figure 33 the reusable elements have to be combined on the content syndication level and the data instance ’Wählen Sie §path_expression.’ needs to be completed with therefore valid expressions.

.

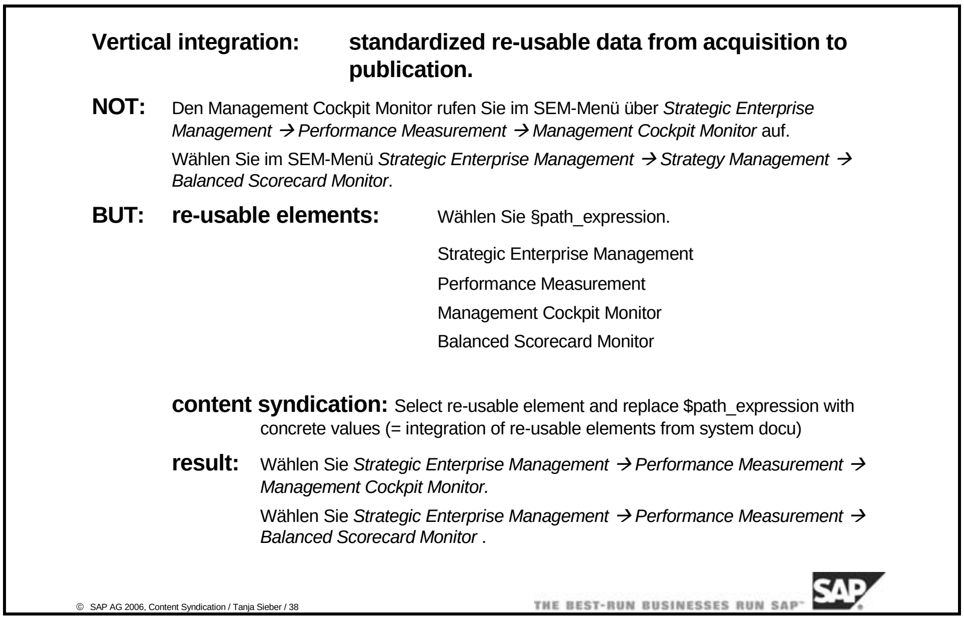


Figure 33 Content Syndication Application Potential@SAP - Example 1

This approach of vertical integration can easily be extended to intelligent horizontal integration by enabling and automising the reuse of already implemented data models and their according data instances (see Figure 34 and Figure 35).

.

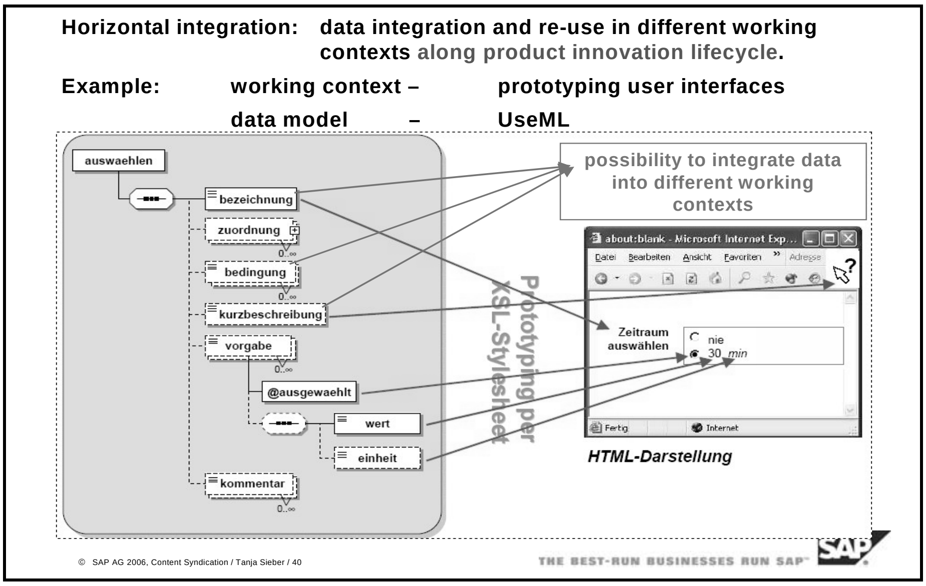


Figure 34 Content Syndication Application Potential@SAP - Example 2

That kind of desired vertical integration taps the potential of using ontologies (see Figure 35) that represent the knowledge and map concepts on the data model level. The author created therefore a taxonomy of relevant terms of the desired Product Innovation Lifecycle and could show the potential for reuse, if this taxonomy is combined with the several existing data models on the different stations along the Product Innovation Lifecycle.

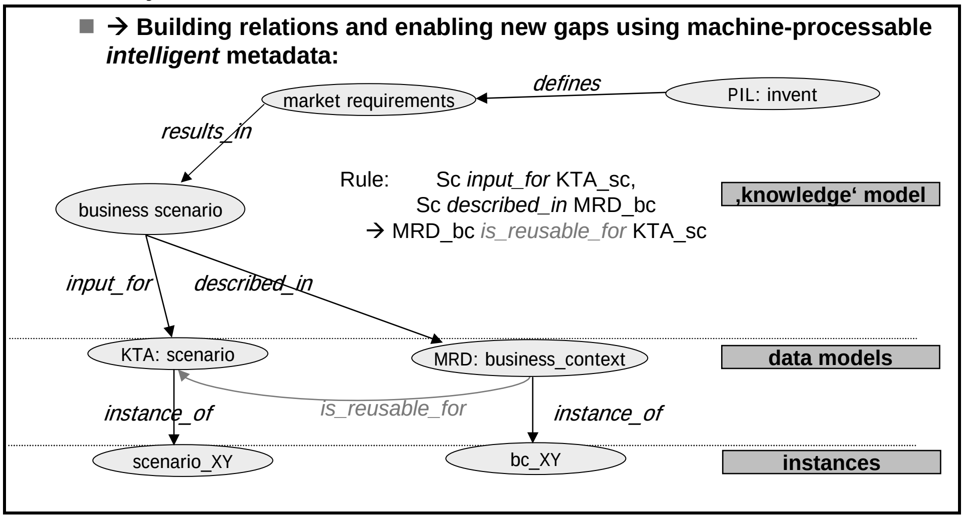


Figure 35 Content Syndication Application Potential@SAP - Example 3

12.2 Paradigm Shift: Operational Aspects

Process Modelling

According to the Transaction Cost Theory it is necessary to describe the desired processing of data instances more detailed. A desired result can be achieved more efficient if tasks and appropriate resources can be guided and administrated as process (Fischer, 2003). According to (Hansen&Neumann, 2005) a process can be defined as follows:

Definition *Process*

A process consists of an amount of correlated tasks that have to been executed in a specific sequence in order to achieve a defined result.

A formal description of a process requires the description of the tasks to be completed, their sequence, pre- and post-conditions and the appropriate role specifications. Transferred to the production of technical documentation the desired processing of data instances can be broken down and for each value-adding process/task it can be defined what kind of input this process needs, what skills are needed for executing this process and what the desirable output is. At SAP AG, the author worked out the according descriptions for processes and desired future reuse processing at the KPS department (cf. Figure 36).

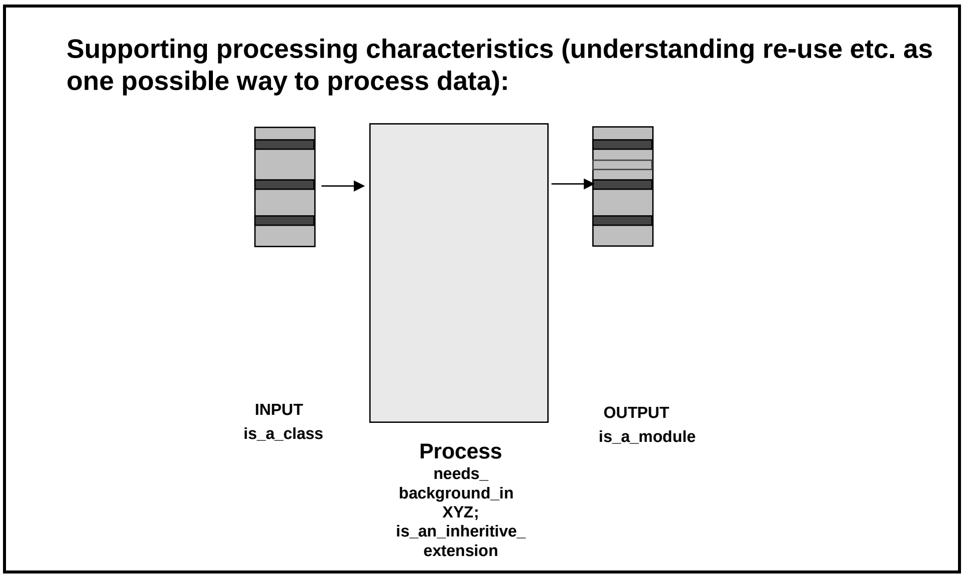


Figure 36 Definition of Processing Characteristics@SAP

Documentation Market

This differentiating process-oriented consideration of technical documentation allows new possibilities of content-automatisation. The formal possibility of addressing a pull-mechanism comes into existence. Implemented rule-based process- and skill definitions enable transaction cost analysis and the cost-efficient arrangement can be found out. If this analysis results in a market situation, the future production of documentation can proceed like exemplary shown in Figure 37.

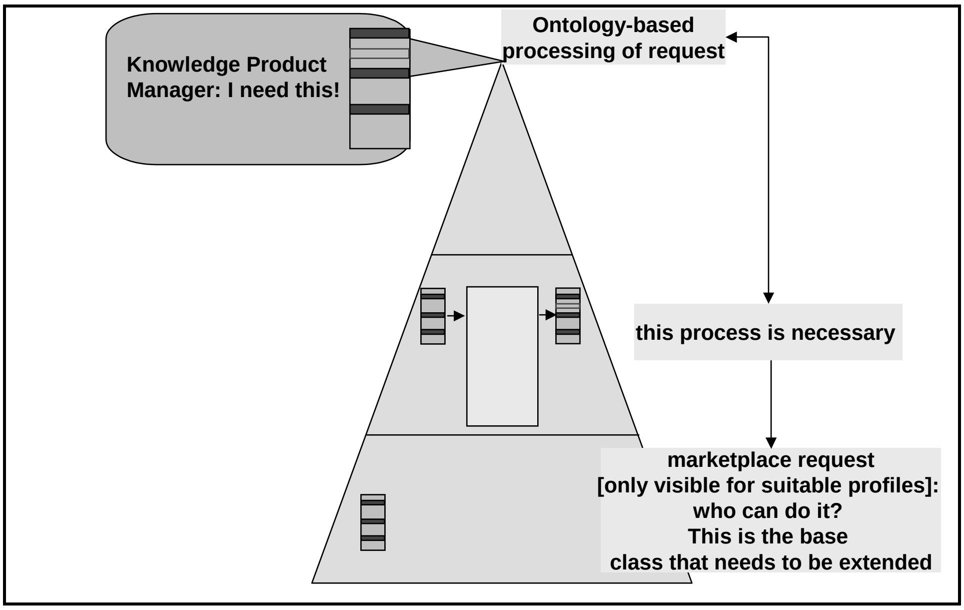


Figure 37 Content Syndication and Pull Principle@SAP - Example

12.3 Conclusion

Summarised the presented conceptual framework *Intelligent Content Syndication* enables and supports

- intelligent organisation of content,
- vertical integration: flow of data from acquisition to publication,
- horizontal integration: flow of data according to product development cycle,
- cost-efficient processing of data

by

- ‚intelligent' ontology- and linguistic-based modularisation through all areas of data capturing from product idea and definition to feedback from customers or support, which allows metadata-based processing of data instances and
- ontology-based descriptions of processing data instances that allow the cost-calculation and transaction cost analysis and results in decisive optimisations as regards cost-efficiency and value-added chain of technical documentation.

This conceptual framework of Intelligent Content Syndication underlied with the idea of Lean Documentation can bring together systems and people who have knowledge resp. data instances with people who need knowledge resp. data instances.

Part V Related Work and Conclusion

There is nothing either good or bad but thinking makes it so.

William Shakespeare, 1564 - 1616

13 Related Work

This chapter provides related work on reuse of technical documentation in Section13.1, on methodologies in Section 13.2 and on paradigms in Section 13.3.

13.1 Related Work on Reuse of Technical Documentation

As depicted in Section 7.2 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. Explanations how to achieve this desired modularisation is missing within the work of Rockley and Ziegler.

13.2 Related Work on Technical Documentation Methodologies

In Section 2.5, related work on methodologies for structuring technical documentation have been presented. The recent amongst them is the structured text control (Ley, 2006): Ley proposed a structured text control as 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. There exist no experience values in transferring this methodology on other application fields.

Ley's methodology is based on speech-act based theories and linguistic understanding of documentation as communicative act. The motivation for Ley for the development of this methodology was the existence of the area of conflict between customer usability on the one side and economic effectiveness of the other side (Ley, 2006). According to Ley, an optimised compliance of both can be gained only by a reasonable structuring of information. The therefore developed text control raises the claim on enabling complete and consistent modeling of a clearly outlined quantity of information. There exists no evaluation of the benefit of this methodology in comparison to others.

13.3 Related Work on Technical Documentation Paradigms

Stock focuses in his work on the development of a new document model and its embedding in the process of documentation (Stock, 2007). 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 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 of typical graph-based causal relations and thereout resulting possible actions. His approach is in creating an all-vehicle-context-containing documentation that can generate at runtime in dependency 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. As shown in Appendix D.1, this kind of product splitting is already implemented in the field of checking procedures since several years.

Although the area of checking procedures would be a perfect candidate for analysing the reuse of knowledge and documentation along a product development lifecycle Stocks research work lacks this.

14 New Scientific Results

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 methodology serves reuse or cost-efficiency requirements.

The driving force for the thesis was therefore the 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, the author provided in this thesis the following new scientific results.

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

One basic approach of the 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.

The author 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.

Terminological statements/definitions and models for the concepts sign, data, information and knowledge have been compared. Based on the gained cognitions, the novel semantic data model has been developed for closing the gap between data and semiotics. Based on the underlying understanding of signs in terms of Peirce the author gave definitions for data and the appropriate terms within this novel model.

This model has been extended 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 - a semantic network based on a lattice of concepts was introduced. This was addressing the integration of conceptual dated knowledge, i.e. ontologies.

- In the second extension – the semantic communication model the integration of basic communication theories has been addressed. The author transferred therefore Austin's speech-act theory into a write-act, that takes place in technical documentation. Therefore, the author analysed several communication theories, and transferred the concepts of message and noise into the model.

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

Thesis 1: *The author has 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. The author has created a multi-layer extension of that and proved a more flexible modelling of concepts. The author has worked out the extension of this multi-layer model in integrating the model into a communicative frame, which represents technical documentation as a communicative act.*

(2) Development of a new evaluation method for methodologies

The method QFD (Quality Function Deployment) developed by Akao has been transferred and implemented for the creation of an evaluation matrix for methodologies for ontology-design and technical documentation with respect to the optimisation of reuse. Therewith a supporting methodic approach for the definition of the requirement criteria has been gained.

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

The obtained evaluation matrix has been extended in form of a guideline implementing textual explanations for the points to be given. For the application of the methodologies, a special use case has been assumed and applied in the evaluation of methodologies for structuring technical documentation and for ontology-design. It could be shown this way, that none of the analysed and evaluated methodologies fulfils 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: *The author transferred the method QFD (Quality Function Deployment) and implemented it for the creation of the novel evaluation method for methodologies for ontology-design and technical documentation. The author 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 the approach to transfer the concept Lean Production to the domain of technical documentation.

The author worked out the basic concepts of Lean Production and transferred the measures supporting these concepts on technical documentation. The to be taken measures have been integrated into the novel model for technical documentation and its engineering – the Lean Documentation.

The author 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.

The author 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: *The author developed a new Lean Documentation model based on the concepts and measures of Lean Production. The author analysed the applicability of Transaction Cost Theory for technical documentation.*

(4) Application and integration of the results into the Intelligent Content Syndication paradigm

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

The lessons learned are:

- 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 pragmatically, otherwise they are unlikely to be understood and to be used at all.
- The existing boundaries between departments should not 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. The author has 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 the author worked out several new measures and splitted them up into novel paradigm shifts in editorial and operational reorganisations.*

Part VI Appendix

A Bibliography

A.1 Own 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 Grundlagenanalye 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/.

A.2 References

Ackoff, R. L. (1989). Zukunftssicherung durch Controlling. Stuttgart: Poeschel.

Abele, T. (August 1993). Lean Documentation. Retrieved March 2008, from http://www.tanner.de/de/abz/1993/abz08_93.pdf.

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

Alexiev, V., Breu, M., Bruijn, J. de, Fensel, D., Lara, R., Lausen, H., & [Ed.]. (2005). Information Integration with Ontologies: Experiences from an Industrial Showcase: Wiley, J.

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

Andaloussi, J., & Maler, E. (1996): Developing SGML DTDs from text to model to markup. New Jersej: Prentice Hall PTR.

Aristotle. (2007). On interpretation, from http://www.netLibrary.com/urlapi.asp?action=summary&v=1&bookid=1085844.

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.

Auer, S., & Lehmann, J. (2007). What have Innsbruck and Leipzig in Common? Extracting Semantics from Wiki Content. Proceedings of European Semantic Web Conference. Springer.

Aussenac-Gilles, N., Biébow, B., & Szulmann, S. (2000). Revisiting Ontology Design: A Method Based on Corpus Analysis. Knowledge Engineering and Knowledge Management. EKAW 2000 (pp. 172–188).

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.

Bach, A. (1993). Qualitätsmanagement - Managementqualität. In F. Biehal (Ed.), Lean Service. Dienstleistungsmanagement der Zukunft für Unternehmen und Non-Profit-Organisationen (pp. 76–96). Bern, Stuttgart Wien: Haupt; Manzsche Verlags- und Universitätsbuchhandlung.

Back, A., Bendel, O., & Stoller-Schai, D. (2001). E-Learning im Unternehmen - Grundlagen - Strategien - Methoden - Technologien. Zürich: Orell Füssli Verlag.

Backhaus, K. (1994). Multivariate Analysemethoden: Eine anwendungsorientierte Einführung ; mit 191 Tabellen (7., vollst. überarb. und erw. Aufl.). Springer-Lehrbuch. Berlin: Springer.

Baker, M. More Content in Less Time - Applying the secrets of Lean Thinking and Agile Programming to the creation of content, from http://www.analecta.com/morecontent.pdf.

Becker, S. (1968). The prospect of rhetoric, from http://zimmer.csufresno.edu/~johnca/spch100/notes.htm.

Berners-Lee, T. (1993). Naming and addressing: URIs, URLs, … W3C Overview, from http://www.w3.org/Addressing/.

Berners-Lee, T. (1998). Cool URIs don't change. W3C Style, from http://www.w3.org/Provider/Style/URI.html.

Bevier, F. F. (2002). Physik der Information: Die Realität als Massstab: Messen, Vergleichen, Speichern. - Formalismus der Information und der Informationsverarbeitung: bussole.

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

Biehal, F. (Ed.). (1993). Lean Service: Dienstleistungsmanagement der Zukunft für Unternehmen und Non-Profit-Organisationen. Bern, Stuttgart Wien: Haupt; Manzsche Verlags- und Universitätsbuchhandlung.

Biehal, F. (1993). Dienstelsitungsmanagement und die schlanke Organisation. In F. Biehal (Ed.), Lean Service. Dienstleistungsmanagement der Zukunft für Unternehmen und Non-Profit-Organisationen (pp. 9–66). Bern, Stuttgart Wien: Haupt; Manzsche Verlags- und Universitätsbuchhandlung.

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

Buder, M. (Ed.). (1997). Grundlagen der praktischen Information und Dokumentation: Ein Handbuch zur Einführung in die fachliche Informationsarbeit (3. völlig neu gefaßte Ausgabe). DGD-Schriftenreihe. München et al.: Saur.

Burkhard, R., & Eppler, M. (2005). Knowledge Visualization. In D. G. Schwartz (Ed.), Encyclopedia of Knowledge Management. B&t Database Management/Cip.

Bühler, K. (1965, 1992). Sprachtheorie: Die Darstellungsform der Sprache. 2. Aufl. Jena: G. Fischer. Neudruck: Stuttgart: Fischer.

Bühler, K. (1969). Die Axiomatik der Sprachwissenschaften. Frankfurt/Main: Klostermann.

Capurro, R. (2000). Einführung in den Informationsbegriff. Retrieved 15. Januar 2007, from http://www.capurro.de/infovorl-index.htm.

Corcho, O., Fernández-López, M., Gómez-Pérez, A., & Lopez-Cima, A. (2005). Building legal ontologies with METHONTOLOGY and WebODE. In B. Selic, R. Benjamins, P. Casanovas, J. Breuker, & A. Gangemi (Eds.), Springer-11645 / Dig. Serial]: Vol. 3369. Law and the Semantic Web. Legal Ontologies, Methodologies, Legal Information Retrieval, and Applications (pp. 142–157). Berlin Heidelberg: Springer-Verlag GmbH.

Dance, Frank E. X., & Larson, Carl. (1976). The Functions of Human Communication: A Theoretical Approach. New York: Holt, Rinehart & Winston.

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.

Drew, J., McCallum, B., & Roggenhofer, S. (2005). Unternehmen Lean: Schritte zu einer neuen Organisation. Frankfurt, New York: Campus Verlag.

Ebers, M., & Gotsch, W. (1995). Institutenökonomische Theorien der Organisation. In A. Kieser (Ed.), Organisationstheorien. 2., überarb. Aufl. (pp. 185–235). Stuttgart: Kohlhammer.

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

Fischer, F., & Scheibeler, A. A. W. (2003). Handbuch Prozessmanagement: Effizienzsteigerung mit ISO 9001+9004 ; 1000 Beispiele aus der Praxis. München: Hanser.

Flensburg, P. (2007). An enhanced communication model. In: Proceedings of the 30th Information Systems Research Seminar in Scandinavia (IRIS) 2007.

Frank, U. (2000): Evaluation von Artefakten in der Wirtschaftsinformatik. In: Häntschel, I; Heinrich, L.J. (Ed.): Evaluation und Evaluationsforschung in der Wirtschaftsinformatik. München, Wien: Oldenburg, pp. 35 - 48.

Fuchs-Kittowski, K. (1976). Informatik und Automatisierung: Theorie u. Praxis d. Struktur u. Organisation d. Informationsverarbeitung ; Band 1. Berlin: Akademie-Verl.

Gómez-Pérez, A., Fernández-López, M., & Corcho, O. (2004). Ontological engineering: With examples from the areas of knowledge managment, e-commerce and the semantic web. Advanced information and knowledge processing. London: Springer.

Göpferich, S. (1998). Interkulturelles Technical Writing: Fachliches adressengerecht vermitteln. Ein Lehr- und Arbeitsbuch; 40: Narr, G.

Gruber, T. (1993). What is an ontology? Retrieved Mar 13, 2006, from www.ksl.stanford.edu/kst/what-is-an-ontology.html.

Gruber, T. (1995). Towards principles for the design of ontologies used for knowledge sharing. International Journal of Human-Computer Studies (pp. 907–928).

Grünninger, M., & Fox, M. (1995). Methodology for the Design and Evaluation of Ontologies.

Hallowell, D. L. (2006). QFD: When and How Does It Fit in Software Development, from software.isixsigma.com.

Hansen, H. R., & Neumann, G. (1987). Wirtschaftsinformatik Bd.1: Grundlagen und Anwendungen: UTB.

Hartley, R. Transmission of Information. Bell System Technical Journal 7 (pp. 535–563).

Herzwurm, G. (1997). Kundenorientierte Evaluierung von Software-Tools zur Unterstützung von Quality Function Deployment. Studien zur Systementwicklung. Köln: Univ. Lehrstuhl für Wirtschaftsinformatik Systementwicklung.

Hevner, A., March, S., Park, J.& Ram, S. (2004): Design Science in Information Systems Research. In: MIS Quarterly, Vol. 28, No. 1, S. 75-105.

Honeycut, L. (2007. Aristotle's Rhetoric, from http://www.public.iastate.edu/~honeyl/Rhetoric/rhet3-1.html#1404a

Hong, S., Goor, G. & Brinkkemper, S. (1993): A Formal Approach to the Comparison of Object-Oriented Analysis and Design Methodologies. In: Nunamaker, J.F.; Sprague, R.H. (Hrsg.): Information Systems of HICCS, Los Alamitos, pp. 689-698.

House, E. R. (1993). Professional evaluation: Social impact and political consequences. Newbury Park: Sage.

IEEE (1990). IEEE Standard Boards: IEEE St. 610.12-1990, IEEE Standard Glossary of Software Engineering Terminology, IEEE Press 1990.

Ingwersen, P. (1986). Information technology and information use: Towards a unified view of information and information technology. London: Graham.

Ingwersen, P. (1992). Information retrieval interaction. London: Taylor Graham.

Ingwersen, P. (1996). Information science: Integration in perspective ; [proceedings CoLIS 2 ; Second International Conference on Conceptions of Library and Information Science - Integration in Perspective ; in Copenhagen, October 13 - 16, 1996]. (CoLIS, Ed.). Copenhagen: Royal School of Librarianship.

Kalmring, D. (Ed.). (2004). Performance Measurement von wissensintensiven Geschäftsprozessen: Ein Beitrag zum Controlling im Wissensmanagement: Deutscher Universitäts-Verlag.

Keller, T., & Tergan, S.-O. (2005). Visualizing Knowledge and Information: An Introduction. In S.-O. Tergan (Ed.), State-of-the-Art-Survey: . Knowledge and information visualization. searching for synergies. Berlin;Heidelberg: Springer.

Kieser, A. (Ed.). (1995). Organisationstheorien (2., überarb. Aufl.). Stuttgart: Kohlhammer.

Kietz, J., Maedche, A.,& Volz, R. (2000): A Method for Semi-Automatic Ontology Acquisition form a Corporate Intranet, ECAW 2000.

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.

Kovács, L. (2004). Adatbázisok tervezésének és kezelésének módszertana. Budapest: ComputerBooks Kiadó Kft.

Kuhlen, R., Seeger, T., & Strauch, D. (Eds.). (2004). Grundlagen der praktischen Information und Dokumentation: Band 1: Handbuch zur Einführung in die Informationswissenschaft- und praxis. Band 2: Glossar: Saur, K G.

Kuhlen, R. (2004). Information. In R. Kuhlen, T. Seeger, & D. Strauch (Eds.), Grundlagen der praktischen Information und Dokumentation. Band 1: Handbuch zur Einführung in die Informationswissenschaft- und praxis. Band 2: Glossar (pp. 3–20). Saur, K G.

Kuhlen, R. (1999). "Knowledge is becoming hypertextified": zur zweiten ACM-Hypertext Konferenz November 1989 Pittsburgh. Konstanz.

Kuntz, A. (2005). Daten, Wissen, Information. Retrieved 19. April 2006, from http://server02.is.uni-sb.de/courses/ident/themen/dat_wiss_info/information.php.

Kunzmann, P., Burkard, F. P., & Wiedmann, F. (1991). dtv-Atlas Philosophie. München: dtv.

Lackes, R., Brandl, W., & Siepermann, M. (1998). Datensicht von Informationssystemen: Datenmodellierung und Datenbanken: Springer Bln.

Lautenbacher, F., Sieber, T., Bauer, B., & Cabral, A. (2007). Linguistic Modeling Methods and Ontologies in Requirements Engineering. Proceedings of the International Workshop on Semantic-Based Software Development. Montreal.

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.

Markus, M. Lynne. (2001). Toward a Theory of Knowledge Reuse: Types of Knowledge Reuse Situations and Factors in Reuse Success. Journal of Management Information Systems (pp. 57–94).

McCarthy, J. (1989). Artificial intelligence, logic and formalizing common sense. In R. H. Thomason (Ed.), Philosophical logic and artificial intelligence. Dordrecht: Kluwer.

Moodey, D.&Shanks, S. (1994): What makes a good data model? Evaluating the Quality of Entity Relationship Models. In: Loucopoulos, P. (Hrsg.): Entity-Relationship Approach - ER'94. Business Modelling and Re-Engineering. Berlin, Heidelberg, pp. 94-111.

Müller, I. (SAP AG). (2006). Quality Gates and Decision Points, from SAP Corporate Portal --> PIL --> Documentation.

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.

Namahn (2007). Information Mapping : a research note, from http://www.namahn.com/resources/documents/note-IM.pdf.

Neumann, G., Strembeck, M., & Zdun, U. (2002) Using Runtime Introspecitble Metadata to Integrate Requirement Traces and Design Traces in Software Components. In: Proc. of ECCOP USE, Malaga, Spain. from: http://wi.wu-wien.ac.at/home/mark/publications/use02.pdf

North. (2004). Knowledge Stair. In D. Kalmring (Ed.), Performance Measurement von wissensintensiven Geschäftsprozessen. Ein Beitrag zum Controlling im Wissensmanagement (p. 13). Deutscher Universitäts-Verlag.

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

Noy, N. F., & Musen, M. A. (2000). PROMPT: Algorithm and Tool for Automated Ontolgy Merging and Mapping.

Noy, N. F., & Musen, M. A. (2003). The PROMPT Suite: Interactive Tools for Ontology Merging and Mapping.

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.

Polanyi, M. (1998). Personal Knowledge: Towards a Post-critical Philosophy: Taylor & Francis Books Ltd.

Probst, G. (2000). Manager Bilanz. (10).

Quality Function Deployment: Kundenorientiert Produkte entwickeln, from www.4managers.de/themen/quality-function-deployment.

Rector, A., Drummond, N., Horrdige, M., Rogers, J., Knublauch, H., Stevens, R., Wang, H., & Wroe, C. (2004). OWL Pizzas: Practical Experience of Teaching OWL-DL: Common Errors & Common Patterns. Engineering Knowledge in the Age of the Semantic Web.

Ried, T. (2006). Strukturierungsmethoden für TD. Wiesbaden.

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

Ruesch, J., & Bateson, G.. (1994). Communication-A systems theoretical point of view (1994). In: Journal Systemic Practice and Action Research, Vol 7 (1).

Rupp, C., & die Sophisten (2007). Requirements-Engineering und Management - Professionelle, iterative Anforderungsanalyse für die Praxis, 4. Auflage, Hanser Verlag, München.

Saatweber, J. Absolute Kundenorientierung durch QFD - Quality Function Deployment. Retrieved 21. November 2006, from www.qm-infocenter.de.

Rekowski, A. von (SAP AG) (2006). Knowledge Re-Use within Knowlege Productization Services, from SAP Corporate Portal --> KPS --> 010 Cross Projects --> IndustrializationConcept --> Content ReUse Workstream --> 10_Projects --> 01_Reuse_Design --> 04_Workshops.

SAP AG. (2002). Allgemeine Standards und Richtlininen für das Schreiben bei SAP. Retrieved November 2006, from SAP Corporate Portal --> Global Service&Support --> Organisation & Departments --> Knowledge Productization Services --> Programs --> Documentation --> Standards & Guidelines.

SAP AG. (2003). 2x10 goldene Regeln für Informationsentwickler. Retrieved November 2006, from SAP Corporate Portal --> Global Service&Support --> Organisation & Departments --> Knowledge Productization Services --> Programs --> Documentation --> Standards & Guidelines.

SAP AG. (2005a). S&G for Generic Information Classes (including quick guide). Retrieved November 2006, from SAP Corporate Portal --> Global Service&Support --> Organisation & Departments --> Knowledge Productization Services --> Programs --> Documentation --> Standards & Guidelines.

SAP AG. (2005b). S&G for 'New' BAdIs. Retrieved November 2006, from SAP Corporate Portal --> Global Service&Support --> Organisation & Departments --> Knowledge Productization Services --> Programs --> Documentation --> Standards & Guidelines.

SAP AG. (2005c). S&G for Release Notes. Retrieved November 2006, from SAP Corporate Portal --> Global Service&Support --> Organisation & Departments --> Knowledge Productization Services --> Programs --> Documentation --> Standards & Guidelines.

SAP AG. (2005d). PIL - Use Cases. Retrieved November 2006, from SAP Corporate Portal --> PIL --> Use Cases.

SAP AG. (2006a). Do's and Dont's for Developers. Retrieved November 2006, from SAP Corporate Portal --> WS Service&Support --> KPS --> Workspace Public --> Service Portfolio --> KDE --> Documentation Standards --> Product Standard --> Overview of Product Standard for Product Documentation --> Documentation Checklist for Components (T203).

SAP AG. (2006b). Matrix of Documentation Types and Target Groups. Retrieved November 2006, from SAP Corporate Portal --> WS Service&Support --> KPS --> Wokspace Public --> Service Portfolio --> KDE --> Documentation Standards --> Product Standard.

SAP AG. (2006c). PIL - Templates. Retrieved November 2006, from SAP Corporate Portal --> PIL --> Templates.

SAP AG. (2006d). PIL -- Use Cases. Retrieved November 2006, from SAP Corporate Portal --> PIL --> Process Overviews.

SAP AG. (2006e). PIL-Documentation. Retrieved November 2006, from SAP Corporate Portal --> PIL --> Documentation.

SAP AG. (2006f). Documentation Checklist for Scenarios/Processes (T293). Retrieved November 2006, from SAP Corporate Portal --> WS Service&Support --> KPS --> Workspace Public --> Service Portfolio --> KDE --> Documentation Standards --> Product Standard --> Overview of Product Standard for Product Documentation --> Documentation Checklist for Components (T203).

SAP AG. (2006g). Documentation Checklist for Solutions/Applications (T202). Retrieved November 2006, from SAP Corporate Portal --> WS Service&Support --> KPS --> Workspace Public --> Service Portfolio --> KDE --> Documentation Standards --> Product Standard --> Overview of Product Standard for Product Documentation --> Documentation Checklist for Components (T203).

SAP AG. (2006h). PIL-Process Overviews. Retrieved November 2006, from SAP Corporate Portal --> WS Service&Support --> KPS --> Workspace Public --> Service Portfolio --> KDE --> Documentation Standards --> Product Standard.

SAP AG. (2006i). Standards and Guidelines for Links Between Information Classes. Retrieved November 2006, from SAP Corporate Portal --> Global Service&Support --> Organisation & Departments --> Knowledge Productization Services --> Programs --> Documentation --> Standards & Guidelines.

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

Searle, John (1969) Speech Acts: An Essay in the Philosophy of language. Cambridge, Mass.: Cambridge University Press

Seeger, T. (1997). Grundbegriffe der Information und Dokumentation. In M. Buder (Ed.), DGD-Schriftenreihe: . Grundlagen der praktischen Information und Dokumentation. Ein Handbuch zur Einführung in die fachliche Informationsarbeit. 3. völlig neu gefaßte Ausgabe. München et al.: Saur.

Shannon, C. E., & Weaver, W. (1949). The mathematical theory of communication. Urbana: Univ. of Illinois Press.

Shannon, C. E., & Weaver, W. (1972). The mathematical theory of communication. Urbana: Univ. of Illinois Press.

Sleeman, D. HONOURS PROJECTS 2006/7. Retrieved October 2006, from http://www.csd.abdn.ac.uk/~sleeman/teaching/projects/ugrad/.

Staab, S., Studer, R., Schnurr, H.-P., & Sure, Y. (2001). Knowledge Processes and Ontologies. IEEE Intelligent Systems.

Stock, I. (2007). Erstellungsprozess neuartiger technischer Dokumentationen.

Stowasser, J. Maria, Petschenig, M., Skutsch, F., & Pichl, R. (1979). Der kleine Stowasser: Lateinisch-deutsches Schulwörterbuch. München: Freytag.

Stuckenschmidt, H., & Harmelen, F. (2005). Information Sharing on the Semantic Web. Springer-11645 /Dig. Serial]. Berlin, Heidelberg: Springer-Verlag Berlin Heidelberg.

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.

Traeger, D. H. (1994). Grundgedanken der Lean Production. Teubner-Studienskripten. Stuttgart: Teubner.

Underwood, M. (2003). CCMS -Communicatioin, Cultural and Media Studies, online verfügbar unter http://www.cultsock.ndirect.co.uk/MUHome/cshtml/index.html.

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

Uschold, M., & King, M. (1995). Toward a Methodology for Building Ontologies. IJCAI.

van Ven, A. H. de (Ed.). (1981). Perspectives on organization design and behavior. Wiley series on organizational assessment and change. New York: Wiley.

Vrandecic, D., Pinto, H., Sure, Y., & Tempich, C. (2005). The DILIGENT Knowledge Processes. Journal of Knowledge Management 9 (5) (pp. 85–96).

Wahren, K. H. (1987). Zwischenmenschliche Kommunikation und Interaktion in Unternehmen: Grundlagen, Probleme und Ansätze zur Lösung: de Gruyter, Walter, GmbH & Co. KG.

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.

Wiegand, H. Ernst. (1989). Der Begriff der Mikrostruktur: Geschichte, Probleme, Perspektiven. 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. 409–462). Berlin: de Gruyter.

Williamson, O. E. (1975). Markets and hierarchies analysis and antitrust implications: Astudy in the economics of internal organization. New York, London: Free Press; Collier Macmillan.

Williamson, O. Eaton, & Ouchi, W. G. (1981). The Markets and Hierarchies Program of Research: Origins, Implications, Prospects. In A. H. de van Ven (Ed.), Wiley series on organizational assessment and change: . Perspectives on organization design and behavior (pp. 347–370). New York: Wiley.

Williamson, O. E. (1985). The economic institutions of capitalism: Firms, markets, relational contracting. New York NY: Free Press.

Williamson, O. E. (1991). Comparative economic organization: The analysis of discrete structural alternatives. Vorträge, Reden und Berichte / Europa-Institut, Sektion Wirtschaftswissenschaft. Saarbrücken.

Willke, H. (2001). Systemisches Wissensmanagement: ; 2047: Lucius & Lucius.

Willke, H. (2004). Einführung in das systemische Wissensmanagement: Carl Auer Verlag.

Ziegler, W. (2005). Fünf Methoden zur Feinarbeit: Variantenverwaltung in CMS. technische Kommunikation, 27(3/2005), 40–44.

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

Please note: The cited URLs were provided by the author in all conscience. They were last cecked in February 2008. However, even though „cool URIs don't change" (cf. (Berners-Lee, 1998)), URIs in the dynamic surrounding of the WWW (typically referred to as URLs (Berners-Lee, 1993)) and the content they represent are subject to change. In future they might differ from the cited sources in this work.

B List of Figures

C List of Tables

D Examples: Technical Documentation

To give a better insight into the existing diversity of technical documentation, in this appendix two examples are given:

Section D.1 shows how external technical documentation for vehicle diagnosis, i.e. so-called fault isolation procedures (FIPs), is developed within service diagnostic engineering at GM Europe GmbH. In Section D.2, an example out of the area of internal technical documentation (i.e. the development and importance of use cases within the requirements engineering process) is illustrated and represented by an example from SAP AG.

References: The following examples have been produced and analysed at General Motors Europe GmbH and SAP AG in the years 2005 and 2006. Section D.1 is mainly based on and extracted from (Sieber&Kovács, 2007), Section D.2 from (Sieber&Lautenbacher, 2007).

D.1 External TD (GM Europe GmbH)

Automotive Engineering - Vehicle Diagnosis

With the rapid development in the electronics sector, and its results spilling over into the automotive sector, vehicle diagnosis presents new challenges to today's technicians. Not only are they faced with a constantly increasing number of electronic systems and sophisticated technology, they also need to fulfil the conflicting tasks of correctly diagnosing those advanced technology vehicles and at the same time spending a minimum amount of time on trouble shooting and repair.

Primary Goal: Fault Localisation

Therefore, the prime goal of a future oriented diagnosis strategy must be an optimised concept, which helps the technician to detect and repair all existing faults in the shortest time possible. The strategy must guide the technician straight to the point of interest, the fault localisation. In order to make a quick fault localisation possible, it is necessary to break down a vehicle into its constituent Electronic Systems. Each of these Electronic Systems can be diagnosed with the help of Checking Procedures and Checking Equipment in order to find the fault.

Since the above-mentioned Checking Procedures are developed to diagnose one Electronic System, the technician evaluates the customer complaint with his experience and knowledge in order to detect that Electronic System, which has to be diagnosed exactly.

All diagnosis on a vehicle follows a logical process: beginning with the evaluation and the attempt to recreate the customer complaint, followed up from preliminary checks as visual inspection and the performance of diagnostic system checks:

Diagnosable vs. Non-Diagnosable

A Checking Procedure (CPR) helps to carry out on a vehicle a guided diagnosis for a certain Electronic System. This diagnosis is based on logically structured measurements, which aim at testing the proper functionality of single components: wiring harness, sensors, actuators, etc. There are two sorts of Electronic Systems with different methods of fault finding:

- *Diagnosable systems*: The Electronic System has an Electronic Control Unit (ECU). Vehicle diagnosis is strongly based on the functionality of the Diagnostic Tester, that communicates with the concerning Electronic Control Unit (ECU). In addition further methods are used (Non ECU supported tests) in order to detect and locate all possible faults.
- *Non-diagnosable systems*: As there is no ECU existing in these Electronic Systems, fault finding with the help of a Diagnostic Tester is not possible, therefore other methods have to be used, which are not supported by an electronic tester.

Requirements of a CPR

The developed CPR has to meet the following requirements: Detection of a single fault and presentation of a logical procedure, which needs little time and guides the technician straight to the fault without many unnecessary works. According to this, it can be realised that it makes sense to break down the Electronic System into smaller parts - so-called Functional Groups.

.

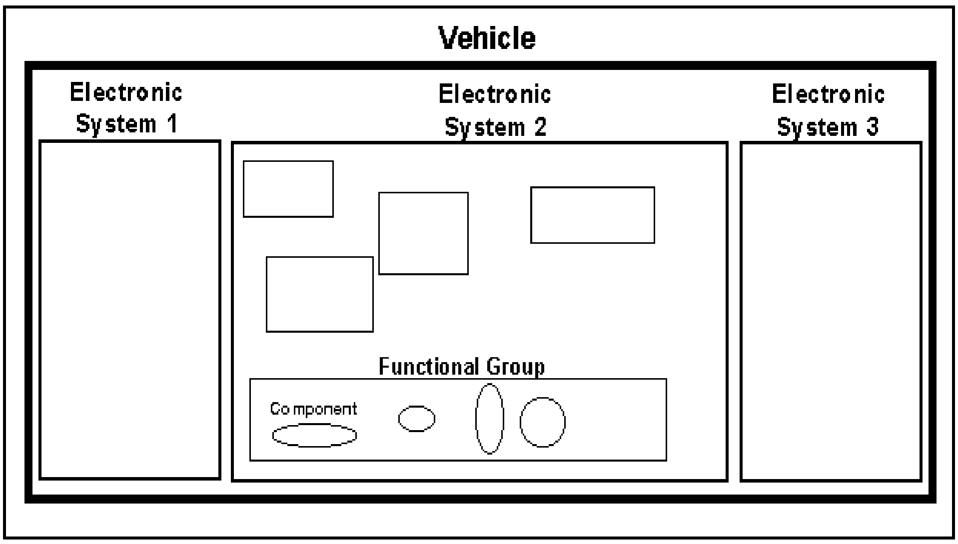


Figure 38 Functional Group Concept@GM

It is faster to surround the fault localisation first by detecting the non-operational Functional Group and searching the fault within this Functional Group than to search within the whole Electronic System.

Tasks of a CPR

Regarding the development of a CPR, two tasks have to be fulfilled connected to this Functional Group concept:

1. All effects of all possible faults have to be recognized AND
2. All faults have to be localised.

A checking procedure at GM is structured as shown in the following picture. It contains different so-called diagnostic levels (A, B and C):

.

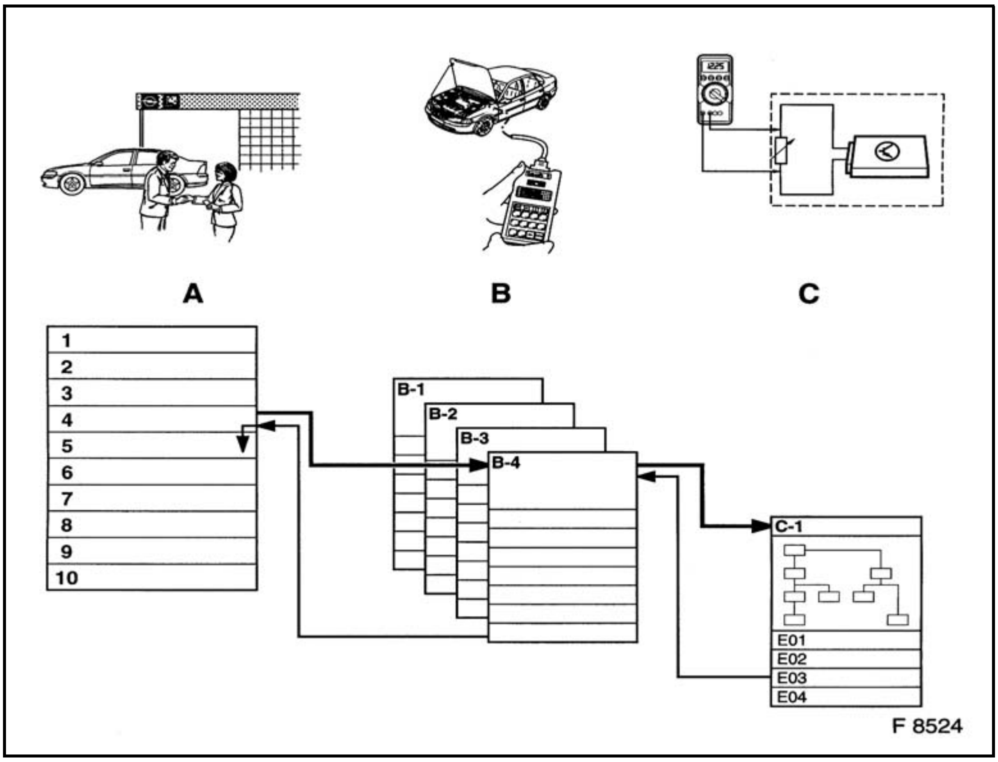


Figure 39 CPR reflects structure of vehicle diagnosis@GM

Different types of data form the content on these levels:

- Diagnostic level A:
  - main diagnostic system check,
  - guides the user through the Checking Procedure,
  - contains several test steps, most of them jumping to according B-Levels or instructing customer complaint record and vehicle end test.
- Diagnostic level B:

  - finds the non-operational Functional Group within an Electronic System,
  - recognizes the effects of faults and jumps in according C-table in order to localise the fault there.
- Diagnostic level C:
  - finds the defective component within a Functional Group,
  - contains Fault Isolation Procedures.

Extraction of Functional Groups

A Functional Group is the smallest part of an Electronic System that can be pinpointed as faulty by available non-intrusive means (i.e DTCs, DLPs, Es Tests, etc.) that can be diagnosed. To subdivide an Electronic System in several Functional Groups first the following information has to be evaluated:

- Wiring Schematic Diagram, Block Diagram or Flow Charts.
- Terminal assignment of the ECU wiring harness connectors.
- Functional description of the concerned Electronic System.

The task of the diagnostic developer is to create conclusive Functional Groups out of this information. This means that the Electronic System has to be subdivided in proper units being aware of the function of the several Functional Groups and their single components. At this point, the diagnostic developer has to pay attention to a special order of priority within the Functional Groups. This order of priority is made up of the knowledge about how these single sections work together in main. This means that the correct function of one single section could be a condition for the performance of another section. E.g., the Functional Group 'Engine Coolant Temperature Sensor Circuit' will not perform correctly if there is a problem within the Functional Group 'System Voltage Circuit'. Next to this, other information like wiring harness has to be taken into consideration as well. This means that information like the connection of sensors, actuators, switches, etc. within the wiring harness have to be considered as well.

After the extraction of Functional Groups, it has to be examined for each Functional Group:

- What kind of faults may occur?
- What kind of fault markers is available?
- Are there enough fault markers, which detect all effects of possibly occurring faults?
- Are there tests, which recognize an effect of a fault, which is already recognized by another test?

Input for Development Process

To answer the important questions during the process of developing a FIP the following input is required:

- What kind of faults is possible?
  - Wiring schematic diagram (internal TD)
  - Functional descriptions (internal TD)
  - Electronic System knowledge
  - Simulator
  - Vehicle
- What kind of fault markers exists already?
  - Tech2 specifications (internal TD)
  - Manufacturer's specifications (internal TD)
- What kind of fault markers needs to be implemented?
  - Comparison: are all effects of possible faults covered?

In order to discover the exact faults within a Functional Group, for each Functional Group one FIP has to be developed. When developing a Fault Isolation Procedure, the elimination logic has to be followed and attention has to be paid to the following priorities:

1. Short to voltage
2. Short to ground
3. Circuit interruption

Besides, the accessibility of the single components has to be regarded and the measuring equipment has to be taken into consideration. Whenever possible, the Diagnostic Tester Functions should be used. The structure of a FIP should also avoid test steps requiring disconnecting, connecting, disconnecting of the same component.

Example: FIP Engine Coolant Temperature Sensor

.

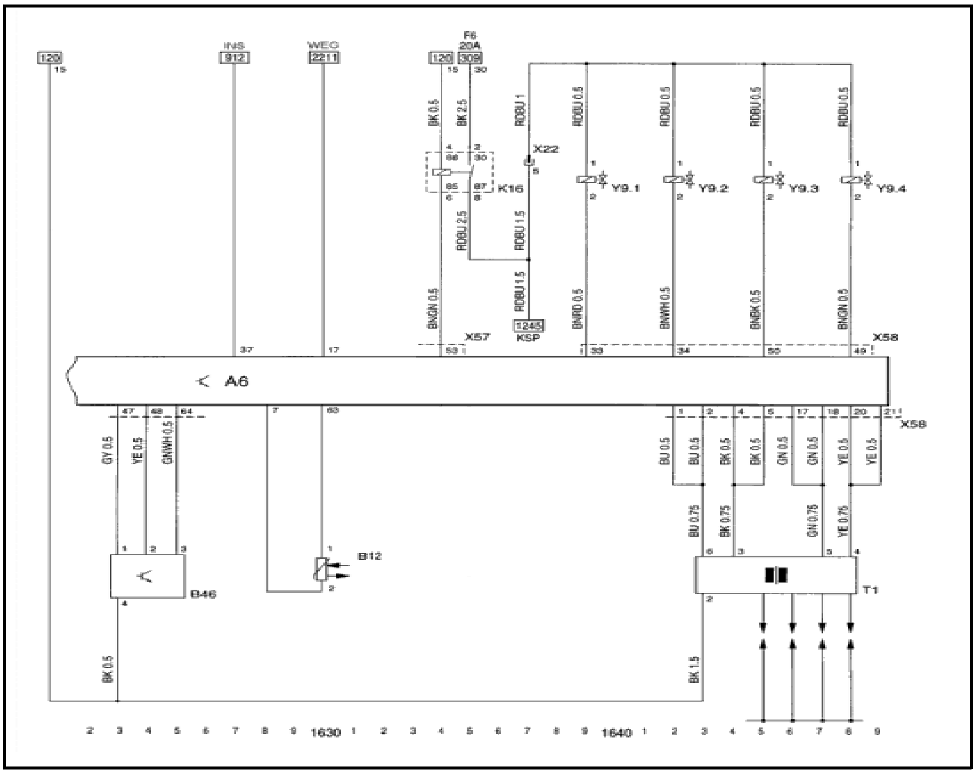


Figure 40 Wiring schematic diagram example@GM

Functionality (Input: wiring schematic diagram and functional description):

- Electrical Connections and Signals:
  - A6, Terminal 7 --> B12, terminal 2: Ground Supply;
  - B12, Terminal 1 --> A6, Terminal 63: Signal (analogue 0,5...4,5V);
  - A6, Terminal 63 --> B12, Terminal 1: Pull Up 5V
- Possible faults:
  - B12 defective,
  - A6 defective,
  - Short to ground/ short to voltage or circuit interruption between A6, terminal 63 and B12, terminal 1,
  - Circuit interruption or short to voltage of ground circuit.

.

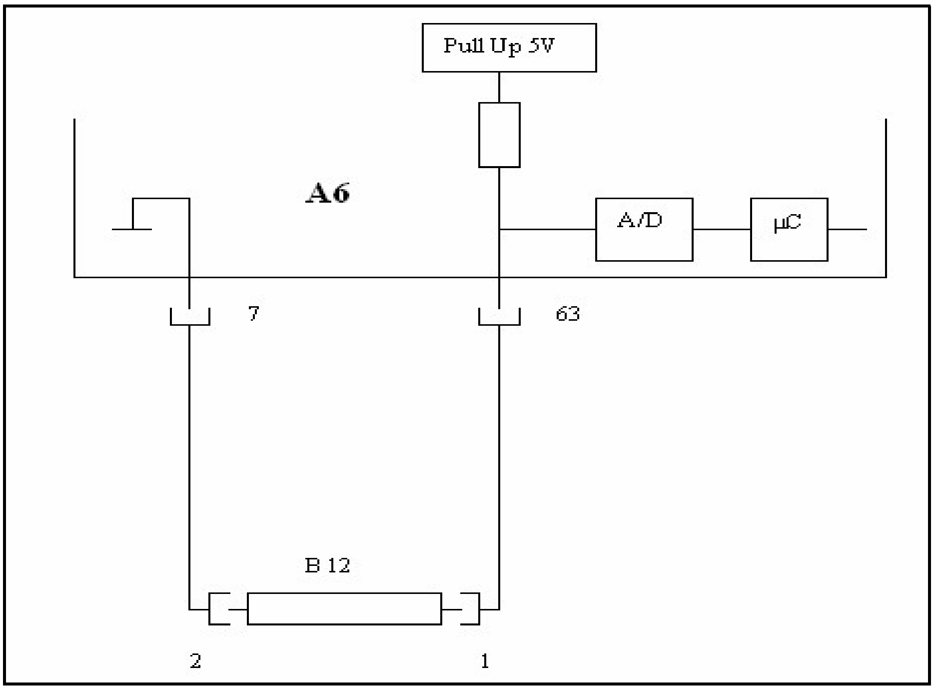


Figure 41 Functional View on Wiring Schematic@GM

- Existing fault markers (Input: Tech2 specification):
  - DTCs: Coolant Temperature Voltage Low, Voltage High
  - DLPs: Coolant Temperature (in V and °C),
- Do the existing fault markers recognize all effects of possible faults?
  - Coolant Temperature Voltage Low --> Short to ground of signal circuit, B12, A6 defective
  - Coolant Temperature Voltage High --> Short to voltage and circuit interruption of signal circuit, B12, A6 defective
  - DLP Coolant Temperature --> Circuit interruption or short to voltage of ground circuit, B12, A6 defective

Out of the given specifications, the following fault isolation procedure can be developed (see Figure 42).

.

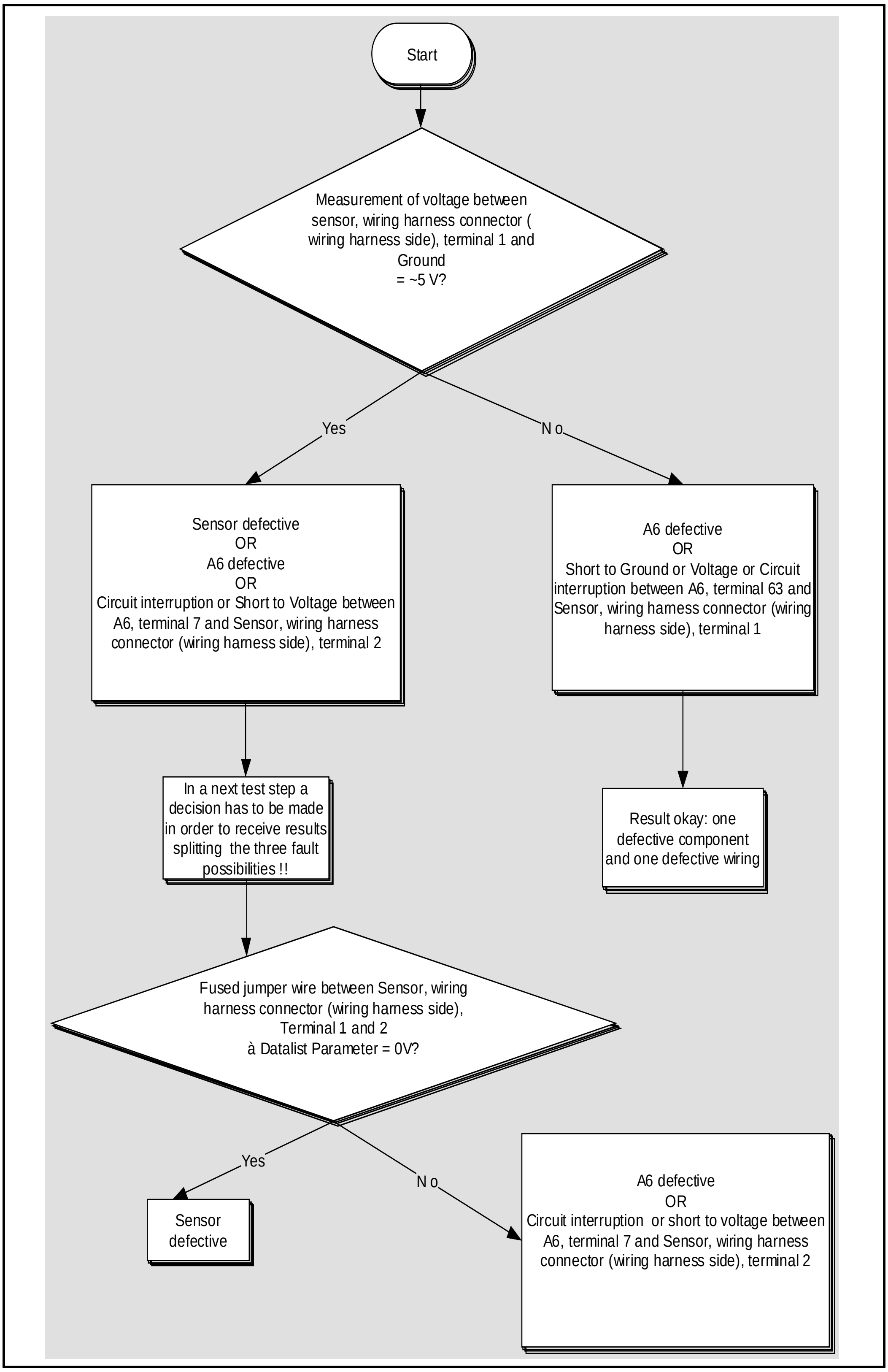


Figure 42 Engine Coolant Temperature FIP@GM

D.2 Internal TD (SAP AG)

Requirements Engineering

Anforderungsanalyse (engl.: Requirements Engineering) beschäftigt sich mit der Analyse des Ist-Zustandes bei einem existierenden System bzw. der Analyse der Kundenwünsche als Soll-Zustand für ein neu zu erstellendes System.

Gemäß IEEE ist eine Anforderung wie folgt definiert:

> *„Eine dokumentierte Darstellung einer Bedingung oder Fähigkeit, die entweder von einem Benutzer zur Lösung eines Problems oder zur Erreichung eines Ziels benötigt wird oder die ein System oder System-Teile erfüllen oder besitzen muss, um einen Vertrag, eine Norm, eine Spezifikation oder andere, formell vorgegebene Dokumente zu erfüllen."*
>
> (IEEE610)

Oftmals wird eine Anforderung auch folgendermaßen definiert:

> *„Eine Anforderung ist eine Aussage über eine Eigenschaft oder Leistung eines Produktes, eines Prozesses oder der am Prozess beteiligten Personen."*
>
> (Rupp, 2007)

Anforderungsarten

Es existieren dabei verschiedene Anforderungsarten:

- funktionale Anforderungen,
- technische Anforderungen,
- Anforderungen an die Benutzerschnittstelle sowie
- Qualitätsanforderungen.

In (Neumann et Al., 2002) werden drei verschiedene Arten von Anforderungen aufgelistet:

- *Kundenanforderungen* beschreiben was der Kunde wünscht und welche Kundenprobleme gelöst werden sollen.
- *Produktanforderungen* fokussieren eher auf das Produkt (bzw. die Software Komponente), welches erstellt werden soll.
- *Projektanforderungen* bestimmen die technische Lösung für eine bestimmte Anforderung und beschreiben wie die Anforderungen des Kunden in eine zufriedenstellende technische Lösung umgewandelt werden kann.

Produktanforderungen sind dabei die Brücke zwischen Kunden- und Projektanforderungen und hinterfragen, wann der Kunde ein bestimmtes Feature im Produkt erhalten soll.

Probleme

Typische Probleme in der Anforderungsanalyse sind unklare Zielvorstellungen, hohe Komplexität, Sprachbarrieren, veränderliche Anforderungen, schlechte Qualität, unnötige Merkmale sowie ungenaue Planung (Rupp, 2007). Außerdem gelten folgende Herausforderungen als besonders kritisch:

- Es gibt keine eindeutige Definition, was eine "Best Practice" ist und es fehlen akzeptierte Referenzmodelle.
- Anforderungen sind nur mangelhaft dokumentiert, zu lösungsorientiert, unvollständig, inkonsistent, nicht implementierbar oder nicht skalierbar strukturiert.
- Verfügbare Werkzeuge sind ineffektiv, bieten nur sehr generische Konzepte, fordern hohen Administrationsaufwand und bieten schlechte Visualisierungskonzepte an.
- Zwischen Produktmanagement, Forschung und Entwicklung, Marketing und Vertrieb gibt es oft keine klar abgestimmten Vorgehensweisen und kein einheitliches Kommunikationsmedium.
- Häufige und späte Änderungen der Anforderungen sind meist nicht vermeidbar und werden insbesondere bei Prozessen mit sequentiellen Vorgehensweisen nicht ausreichend beherrscht.

Dokumente

Derzeit werden häufig folgende wesentlichen Dokumente im Rahmen der Anforderungsanalyse erstellt:

- eine Liste von wesentlichen Funktionalitäten des zu entwickelnden Produkts oder Systems,
- eine Liste der wesentlichen Produktattribute,
- Anwendungsfälle,
- ein konzeptuelles Modell sowie
- ein Glossar (Data Dictionary), welches die verwendeten wesentlichen Begriffe beschreibt.

Die Produkt- oder Systemfunktionalität beschreibt dabei eine Liste von Eigenschaften, welche ein Produkt oder System leisten soll. Die Elemente dieser Liste werden dabei kategorisiert, priorisiert und nach sinnvollen Kriterien gruppiert. Kategorien können entweder evident (notwendig und sichtbar), versteckt oder optional sein. Attribute bezeichnen Charakteristiken des Produkts oder Systems, die keine Funktionalität darstellen, also beispielsweise Größe, Gewicht, Effizienz, Antwortzeiten, Fehlertoleranz, Bedienungskomfort, etc. und werden in gewünschte und verlangte Attribute unterschieden.

Die Anwendungsfälle werden oft auch *Use Cases* genannt.

Use Cases im Software Engineering

Die dokumentierten Anforderungen werden im Rahmen des Software Engineering verwendet, um daraus Analyse- und Designdokumente zu erstellen, Code zu implementieren, diesen zu testen und nach Fertigstellung das fertige Produkt an den Kunden auszuliefern. Dabei gibt es verschiedene Software-Entwicklungsprozesse und Vorgehensmodelle, die das Vorgehen genauer beschreiben (siehe hierzu Sieber&Lautenbacher, 2007).

Einen sehr wichtigen Stellenwert nehmen dabei vor allem die Use Cases ein. Hierin werden die Akteure und Szenarien detailliert in natürlichsprachlicher Form beschrieben. Diese Akteure sind Personen oder Systeme, die mit dem zu erstellenden Produkt oder System später interagieren. Die Szenarien stellen eine Abstraktion über die möglichen Interaktionen dar. Es wird also durch die Anforderungsanalyse nur ein Verständnis des Problembereichs bereitgestellt, die exakte Ausführung in allen Details wird hingegen nicht beschrieben. Allerdings werden zusätzlich bestimmte Dokumente erstellt, die für die Entwicklung von essentieller Notwendigkeit sind. Hier wären z.B. das Vokabular (Glossar) des Produkts oder Systems zu nennen, welches Konzepte der realen Welt, die für die Entwicklung des Produkts wichtig sind, sowie deren Attribute beinhaltet und natürlichsprachlich beschreibt.

Use Cases entstehen primär in der ersten Phase des jeweiligen Software-Entwicklungsprozesses und werde dann in nachfolgenden Phasen erweitert. Ursprünglich textuell beschriebene Anforderungen werden so im Laufe eines Projekts graphisch modelliert und durch verschiedene Diagrammtypen verfeinert (begonnen mit einem Use Case-Diagramm über Aktivitätsdiagramme, Sequenzdiagramme oder Zustandsautomaten bis hin zu Klassendiagrammen). Ein Reuse von Anforderungen, Analyse- oder Designdokumenten über die Grenzen eines einzelnen Projekts hinweg findet allerdings derzeit meist nicht statt.

Hingegen werden auf Implementierungsebene existierende Komponenten, Middleware-Ansätze, Architekturen und Frameworks sehr wohl in unterschiedlichen Projekten immer wieder verwendet. Hier ist allerdings bisher das Fachwissen des Software-Entwicklers gefragt, der genau wissen muss, welche Frameworks denn existieren und welche er für welche Problemstellungen einsetzen kann. Auch auf Testebene werden existierende Frameworks (wie bspw. jUnit) immer wieder verwendet, wobei hier allerdings die Testfälle selbst jeweils auf das Projekt angepasst werden müssen.

Beispiel: SAP AG Use Cases Template

Im folgenden ist eine Use-Case Beschreibung als Beispiel aufgeführt, wie sie von der SAP AG durch ein Template vorgegeben entsteht.

.

Goal	**User/ Actor**	**Scenarios**	**Data**
Manage an open credit order (initiated by superior's request or own initiative)	Credit manager	*Main Case* Credit manager reviews open credit order details, customer credit history, and external customer resources to make a credit order decision. Credit manager approves outright, rejects, increases credit limit and approves, or leaves unchanged, the credit order. *Extensions* 1.Add credit order and account comments *Open Issues* 1.Can the credit manager delete credit orders?	·*Credit order data:* company name, order number, order date, order status, order value, customer credit limit, credit remaining, order comments ·*Internal account data:* company name, account number, credit score, credit risk class, payment history, disputed items, account comments ·*External account data:* Hoovers report, Dun and Bradstreet report, customer financial statement

Table 10 Example: Use Case Description@SAP

E Analysis: Data, Knowledge, Information

This appendix begins by giving and reflecting existing definitions of the concept data.

In Section E.2 some definitions for knowledge are given and analysed in relation to the developped semantic data model (Sieber/Kammerer). The appendix ends with presenting and reflecting definitions for the concept information in Section E.3.

References: All three sections are extracted from (Sieber&Kammerer, 2007a).

E.1 Kritische Reflexion ausgewählter Definitionen des Terminus Daten

Im Folgenden werden einige ausgewählte Definitionen des Terminus Daten (engl.: data items) vorgestellt und kritisch beleuchtet. Aus forschungsstrategischen Gründen wurden solche Definitionen und Aussagen ausgewählt, die besonders leicht zu kritisieren sind und die die Vorteile des hier zu entwickelnden Konzepts von Datum unterstützen.

Daten als beobachtete Unterschiede

> *„Daten sind beobachtete Unterschiede.“ (Z1)*
>
> (Willke, 2004)

Wenn man unter Daten "beobachtete Unterschiede" versteht, impliziert dies zuerst einmal, dass Daten Unterschiede sind. Unterschiede können aber nur durch Vergleiche erkannt werden. Dies impliziert wiederum zweierlei: (1) dass mindestens zwei Objekte der Anschauung miteinander verglichen werden und dass (2) ein wahrnehmendes Subjekt den Akt des Vergleichens ausführt. Dies bedeutet, dass ein jedes Datum ein wahrnehmendes Subjekt voraussetzt. Damit kann es kein Datum ohne wahrnehmendes Subjekt geben. Die Frage, die sich daran anschließt, ist:

Existiert das Datum dann auch nur so lange, wie das wahrnehmende Subjekt den Unterschied konstatiert?

Dies kann - ohne das gemeine Verständnis von dem, was Daten sind, zu missachten - sicherlich nicht bejaht werden. Auch ist es fraglich, ob etwas nur im Unterschied zu etwas anderem als Datum begriffen werden kann. Sicherlich ist die Farbe indigo ein Datum, das als Datum (nämlich: indigo zu sein) bestehen bleibt, auch wenn kein wahrnehmendes und unterscheidendes Subjekt vorhanden ist und wenn kein weiteres Objekt vorhanden ist, zu dem es in Opposition steht (beispielsweise: kobalt). Davon bleibt jedoch unberührt, dass jedes Datum zu allen anderen Daten in vielfältigen Relationen steht (Relation hier in einem durchaus mathematischen Sinne verstanden), dass also m.a.W. Unterschiede zwischen einzelnen Daten durchaus bestehen - es ist aber keinesfalls so, dass ausschließlich die Unterschiede (also: die Relationsterme, wie z.B. x ist größer als y) Daten sind.

Interessant an Z 1 ist außerdem, dass Willke von beobachteten und nicht - verallgemeinernd - von beobachtbaren Unterschieden spricht. Damit reduziert Willke Potenzialität auf Faktizität und schränkt damit die Menge der Daten erheblich ein.

Betrachten wir abschließend zu diesem Zitat ein kleines Beispiel.

Bsp. 1 Zwei Zahlenobjekte - Widersprüchlichkeit von Z1

Gegeben seien die folgenden beiden Objekte (Zahlen) elf und zwölf im nachfolgenden Absatz:

11 12

Ein Subjekt, das die beiden Objekte wahrnimmt und miteinander vergleicht, kann u.a. die folgenden Unterschiede und damit nach Z1 Daten beobachten:

- Das Objekt 11 ist kleiner als das Objekt 12.
- Das Objekt 11 steht links von Objekt 12.
- Das Objekt 12 steht rechts von Objekt 11.
- etc.

Das Subjekt kann aber auch die folgenden Gemeinsamkeiten konstatieren, die jedoch nach Z 1 keine Daten sind:

- Beide Objekte sind arabische Zahlen.
- Beide Objekte stehen auf demselben Blatt Papier.
- Beide Objekte sind in der Schriftart Arial gedruckt.
- etc.

Diese Aussagen, die man nach einem allgemeinen Verständnis ebenfalls als Daten ansehen würde (sofern man die oben konstatierten Unterschiede als Daten ansieht), sind - wie gesagt - nach Willke keine Daten.

Bemerkenswert ist darüber hinaus, dass weder das Objekt "11" oder das Objekt "12" selbst nach Z 1 ein Datum darstellen noch eine der folgenden Aussagen:

- "11" ist eine Zahl.
- "12" ist eine Zahl.

da nach Z1 kein beobachteter Unterschied zugrunde liegt.

Spätestens an dieser Stelle wird deutlich, dass die Konzeption von Datum, wie sie in Z 1 expliziert wurde, weder intuitiv noch in ihrem Ansatz sowie in ihren Konsequenzen haltbar ist, sich vielmehr sogar in Widersprüche verwickelt.

Data as Outputs from Devices

> *„Raw, unrelated numbers or entries, explicated on a medium (e.g., in a database). Data are numbers, characters, images, or other outputs from devices.“ (Z2)*
>
> (Burkhard&Eppler, 2005)

Der Sinn des englischen Wortes "explicated", erschließt sich im Zusammenhang mit "on a medium" nicht ohne weiteres. Hier ist ein Widerspruch zum ersten Satzteil, in dem aufgeführt wird, dass "numbers" oder "entries" "unrelated" sind. Wenn etwas erläutert wird, impliziert das jedoch, dass die Eigenschaft "unrelated" nicht mehr zutrifft. Versteht man "explicated on a medium" als auf einem Medium gespeichert folgt daraus nach Z2, dass (3) Daten nur existieren, wenn sie auf einem Medium abgelegt sind. Daran schließen sich einige Fragen an:

Existiert ein Datum nur so lange, wie es auf einem Medium abgelegt ist?

Wenn man Daten so versteht, wie sie in Z2 beschrieben werden, muss man dies bejahen. Dies wird sicherlich auch von Personen, die im Bereich der EDV Erfahrungen sammeln durften, mit Rechnerabstürzen und dadurch verloren gegangenen "Daten" bestätigt werden.

Darf man sich unter Medium ein Blatt Papier, eine alte Druckerpresse, einen Menschen, einen Computer oder eine CD vorstellen?

"On a medium" kann im Deutschen auf einem Medium/Hilfsmittel/Träger bedeuten. Durch das in Klammern aufgeführte Beispiel "in a database" liegt die Vermutung nahe, dass ein (Speicher-) Medium gemeint ist, es wird jedoch nicht eindeutig behauptet. Dies führt dazu, dass je nach dem, wie eng die Abgrenzung des Terminus Medium gezogen wird, etwas als Datum bezeichnet werden kann, was bei einem anderen Verständnis von Medium nicht darunter fallen würde. Da es mehr als fraglich ist, ob man von einem gemeinen Verständnis für den Terminus Medium ausgehen darf und sollte, ist die Aussage an dieser Stelle nicht eindeutig und lässt nur Spekulationen zu.

Weiterhin interessant ist bei Z2 die weiterführende Definition für Daten im zweiten Satz, wo Daten als "output of devices" bestimmt werden. Wenn Daten als das verstanden werden, was von Geräten (prinzipiell) ausgegeben werden kann, impliziert dies, dass (4) Daten ohne vorherige Ausgabe irgendwelcher Geräte nicht existieren. Dies lässt zum einen vermissen, dass Daten auch ohne das Wirken irgendwelcher Geräte existent sind und kann zum anderen - je nach Abgrenzung des Terminus Medium - zu einem Widerspruch zu (3) führen, wonach Daten nur dann als existent angenommen werden, wenn sie auf einem Medium abgelegt sind. Demnach wäre die logische Folge der Kombination der beiden Aussagen, dass nur dann von einem Datum gesprochen werden kann, wenn die Nummer, das Zeichen, das Bild oder andere semiotische Einheiten von einem Gerät auf ein Medium ausgegeben werden.

Die Widersprüchlichkeit von Z2 kann abschließend an einem Beispiel verdeutlicht werden.

Bsp. 2 Zwei Zahlenobjekte - Widersprüchlichkeit von Z2

Gegeben seien erneut die folgenden beiden Objekte (Zahlen) elf und zwölf im nachfolgenden Absatz:

11 12

Das Objekt "11" werde nun am Rechner erfasst und in einer Datenbank gespeichert; danach ist es nach (3) ein Datum. Dasselbe Objekt werde nun aber von einer Person auf ein Papier geschrieben. Nach (3) und unter der Annahme, dass Papier ebenfalls ein Medium i.S.v. Z2 ist, handelt es sich um ein Datum; nach (4) allerdings trifft die Voraussetzung, dass ein Datum von einem Gerät ausgegeben sein muss, um ein Datum zu sein, offensichtlich nicht zu, weshalb es sich nach (4) nicht um ein Datum handelt.

An dieser Stelle wird wieder deutlich, dass die Aussagen zum Terminus Daten, wie sie in Z2 ausgeführt wurden, in ihren Konsequenzen nicht haltbar sind, da die Widersprüchlichkeit der Aussagen dazu führen, dass sich potentielle Dateneigenschaften eines Objekts gegenseitig ausschließen.

Daten sind in einem Medium gespeichert

> *„Daten sind in einem Medium gespeicherte Zeichen.“ (Z3)*
>
> (Back, Bendel&Stoller, 2001)

Wenn man Daten als Zeichen versteht, die in einem Medium gespeichert sind, folgt daraus (vgl. Z2), dass (3) Daten nur existieren, wenn sie auf einem Medium gespeichert sind und - nach Z3 - (5) Daten medienabhängig sind. Auf die Problematik der auch hier fehlenden Festlegung, was genau unter einem Medium zu verstehen ist, wurde bereits bei der kritischen Analyse von Z2 eingegangen, so dass dies an dieser Stelle nicht wiederholt werden muss. Eine Frage, die sich jedoch stellt, ist folgende:

Sind dieselben Zeichen, wenn sie in unterschiedlichen Medien gespeichert werden, unterschiedliche Daten?

Bsp. 3 Objekt in unterschiedlichen Medien

Zur Veranschaulichung sei folgendes Objekt gegeben:

611497

Dieses Objekt kann auf verschiedene Art und Weise gespeichert sein - hier einige Beispiele:

(a) 611497 wird in eine Datenbank eingetragen.

(b) 611497 wird von einem Subjekt handschriftlich auf ein Stück Papier geschrieben.

(c) Das Papierstück mit handgeschriebener 611497 wird eingescannt und als Bilddatei abgespeichert.

(d) 611497 ist in einem Handy als Telefonnummer gespeichert.

Wenn man den Terminus Medium so weit fasst, dass auch ein Blatt Papier als ein Medium bezeichnet werden kann, haben wir in allen vier Fällen ein Datum vorliegen:

- In (a) ist 611497 in einer Datenbank gespeichert und dadurch nach dem Verständnis von Z3 ein Datum.

- In (b) ist 611497 auf dem Medium Papier geschrieben und damit nach Z3 ein Datum.
- In (c) ist 611497 in der Bilddatei abgespeichert und somit nach Z3 ein Datum.
- In (d) ist 611497 in einem Handy gespeichert und nach dem Verständnis von Z3 ebenfalls ein Datum.

Auffallend ist, dass 611497 an und für sich nach Z3 kein Datum ist; erst nachdem es auf vier verschiedenen Medien fixiert wurde, wurden daraus vier Daten

Des weiteren stellt sich die Frage:

Was ist in Z3 genau unter Zeichen zu verstehen?

Folgen die Autoren beispielsweise Peirce, wonach Zeichen entweder Ikonen, Indizes oder Symbole sind, oder verstehen sie unter Zeichen ausschließlich Schriftzeichen? Eine Erörterung der unterschiedlichen Definitionen dieses Terminus würde hier zu weit führen, es sei jedoch darauf hingewiesen, dass in Z3 aufgrund der fehlenden terminologischen Festlegung des Terminus Zeichen als Konsequenz der Terminus Daten nicht ausreichend definiert ist. Auch dies soll anhand eines Beispiels genauer gezeigt werden.

Welche der folgenden Objekte sind nach Z1 Zeichen und dadurch bei Speicherung auf einem Medium Daten?

(a)

(b) schwarz

(c) ________

(d) 11

(e) elf

(f) XI

Wenn man annimmt, dass ausschließlich Schriftzeichen Zeichen sind, können nur (b), (d), (e) und (f) als potentielle Kandidaten für Daten in Betracht kommen. Nimmt man aber an, dass Zeichen i.S.v. Peirce Ikonen, Indizes und Symbole sind, können nach Z3 all[redacted]geführten Objekte als potentielle Daten angesehen werden. Abschließend kann festgestellt werden, dass aus Z3 nicht hervorgeht, welche Objekte potentiell als Daten zu verstehen sind, da der Terminus Zeichen im Unklaren bleibt.

Data as Raw Symbols

> *„Data are raw. They are symbols or isolated and non-interpreted facts. Data represent a fact or statement of event without any relation to other data. Data simply exists and has no significance beyond its existence (in and of itself). It can exist in any form, usable or not. It does not have meaning of itself.“ (Z4)*
>
> (Keller&Tergan, 2005)

Wenn man "raw" als eine konstitutive Eigenschaft von Daten ansieht, impliziert das, dass (6) Daten etwas sind, das sie nach einer Verarbeitung nicht mehr sind. Dies entspricht sicherlich nicht dem allgemeinen Verständnis von Daten und wirft folgende Frage auf:

Wann fängt eine Verarbeitung eines Datums an und ab wann kann man nicht mehr von einem Datum sprechen?

Wenn man eine Zahl auf ein Papier schreibt und dieses Papier dann einscannt, verarbeitet man die Zahl, und nach Z4 könnte man deshalb nicht mehr von einem Datum sprechen. Bis zu welchem Verarbeitungsschritt könnte man die Zahl i.S.v. Z4 noch ein Datum nennen? Bis zum Einscannen? Oder ist bereits das Aufschreiben eine Verarbeitung? Unserem Verständnis nach bleibt das Datum ein Datum, und zwar sowohl vor, während als auch nach der Verarbeitung.

Daten sind - so weiter Z4 - Symbole oder isolierte und nicht-interpretierte Fakten. Der Terminus Symbol wird in Z4 nicht näher erläutert. Es bleibt unklar, ob dieser i.S.d. Semiotik zu verstehen ist oder auf andere Art und Weise. Ohne eine entsprechende Klärung, in welchem Sinne der Terminus hier verwendet wird, kann nur spekuliert werden, was darunter verstanden werden soll.

Wenn man Daten so versteht, dass sie eine Tatsache oder Aussage über ein Ereignis darstellen ("represent a fact or statement of event"), folgt daraus, dass (7) alle Daten einen überprüfbaren Wahrheitsgehalt haben und (8) ihnen ein Ereignis vorausgegangen sein muss (das sie dann beschreiben). Darunter fallen dann zweifellos Ergebnisse einer Messreihe.

Wie verhält es sich mit Angaben, deren Wahrheitsgehalt man nicht ohne weiteres überprüfen kann?

Nach Z4 ist beispielsweise der Mittelpunkt der Erde kein Datum. Dies widerspricht jedoch dem gemeinen Verständnis von Datum. In der Fortsetzung heißt es nun jedoch, dass Daten einfach existieren, was nun wiederum den Mittelpunkt der Erde als ein mögliches Datum nicht ausschließen würde. In Z4 wird jedoch durch das Verknüpfen des Datenbegriffs mit der Forderung nach Fakten oder vorausgegangenen Ereignissen die Anzahl tatsächlich vorhandener Daten erheblich eingeschränkt.

Eine weitere Einschränkung erfolgt in Z4 durch den Zusatz, dass Daten in keiner Beziehung zu anderen Daten stehen. Hier stellt sich die Frage, was hier mit "relation" gemeint ist.

Die Problematik soll wieder durch ein Beispiel verdeutlicht werden.

Bsp. 4 Objekte und ihre Relation

Gegeben seien die folgenden Objekte:

11

Euro

11 Euro

Nach Z4 können folgende Daten identifiziert werden:

- 11 (repräsentiert eine Tatsache, dass es sich um eine Anzahl von elf handelt)

- Euro (repräsentiert eine Tatsache, dass es sich um die Währung Euro handelt)

In 11 Euro gibt es jedoch eine syntaktische Relation zwischen zwei Daten, und nach Z4 handelt es sich somit um kein Datum. Ob eine solche Relation in Form einer syntaktischen Konstruktion ein Ausschlusskriterium für ein potentielles Datum ist, ist äußerst fragwürdig. Was in Z4 fehlt, ist eine Abgrenzung des Terminus Relation, die eindeutig definiert, welche Art von Relation hier verstanden wird.

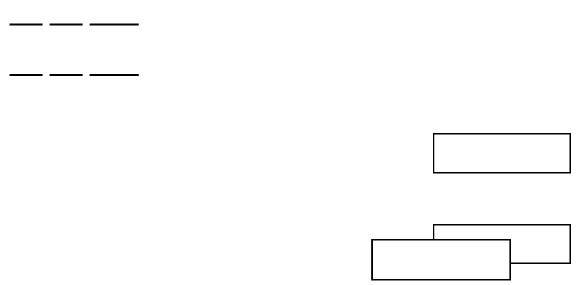
Daten als Menge zusammengehöriger Zeichen

> *„Eine Menge von zusammengehörigen Zeichen, die in einem gegebenen Kontext Sinn machen.“ (Z5)*
>
> (Seeger, 1997)

Wenn man Daten als Zeichenmenge versteht, deren Elemente in einem gegebenen Kontext Sinn machen, impliziert das, dass (9) einzelne Zeichen keine Daten sind und (10) Objekte, die in gegebenen Kontexten sinnlos sind, keine Daten sind. Dies schränkt die Menge der Daten erheblich ein und wirft die folgenden Fragen auf:

Wann kann man von einer Menge zusammengehöriger Zeichen sprechen?

Wie bereits in Z3 fällt auf, dass von Zeichen die Rede ist und nicht genauer definiert wird, was darunter zu verstehen ist. Unter der Annahme, dass Zeichen i.S.v. Peirce verwendet wird, seien die folgenden Objekte gegeben:



Ein wahrnehmendes Subjekt kann folgende Mengen zusammengehöriger Zeichen bilden:

- Drei Rechtecke.
- Fünf Formen.
- Zwei Formen mit jeweils drei waagrechten Strichen.
- Zwei übereinander plazierte Rechtecke.
- etc.

Ausgehend von diesen Zeichenmengen, die je nach analysierendem Subjekt und dessen Erkenntnisinteresse bereits sehr unterschiedlich ausfallen können, stellt sich die nächste Frage:

Wodurch ist ein Kontext gegeben?

Wenn man die Zeichenmenge "Drei Rechtecke" betrachtet, kommt als Kontext in Frage, dass sich im Umfeld andere Objekte befinden, die eine andere Form haben. Die Zeichenmenge ist also insofern sinnvoll, als dass sie sich von den anderen Formen unterscheidet. Diese Zeichenmenge ist damit nach Z5 ein Datum. Interessanterweise kann aber weder "drei" noch "Rechteck" als Datum nach Z5 bezeichnet werden.

Durch die Forderung nach einem Kontext, einem Sinn und einer Menge von zusammengehörigen Zeichen schränkt Seeger die Menge tatsächlich vorhandener Daten in nicht nachvollziehbarem Maße ein.

Daten als materialisiertes dargestelltes Wissen

> *„Bedeutung von Daten: Damit von einem Sender nachgefragtes Wissen als Information über einen Kanal zu einem Empfänger gelangen kann, muss dieses Wissen über den Kanal vermittelbar gemacht werden. Materialisiertes, nach syntaktischen Regeln dargestelltes Wissen wird als "Daten" bezeichnet. Damit diese Daten (oder Teile von ihnen) nach dem Vermittlungsprozess als Informationen handlungsrelevant werden können, müssen ihre semantischen Eigenschaften reinterpretierbar sein." (Z6)*
>
> (Kuhlen, 2004)

Kuhlen definiert an dieser Stelle den Terminus Daten als materialisiertes, nach syntaktischen Regeln dargestelltes Wissen. Wenn man Daten wie in Z6 versteht, besteht (11) im Zusammenhang mit Daten immer ein Sender- und Empfängerverhältnis. Daran schließt sich direkt die folgende Frage an:

Was passiert mit ein- und demselben Objekt, wenn es auf unterschiedliche Empfänger trifft?

Angenommen, Empfänger A hat eine Messreihe gestartet, weil er etwas über den Verlauf einer zu messenden Eigenschaft erfahren will. Die Messreihe sieht wie folgt aus:

1 12225,2

2 12226,8

3 12256,2

4 12222,6

Diese Messreihe repräsentiert das nachgefragte Wissen, es handelt sich also um Daten, wobei hier nach Z6 nicht klar ist, ob nun die gesamte Messreihe als ein Datum anzusehen ist oder die darin enthaltenen Werte.

Empfänger A druckt sich die Messreihe aus und lässt das Papier auf seinem Schreibtisch liegen. Nun kommt zufällig eine andere Person in den Raum (Empfänger B) und blickt auf das Blatt Papier. Empfänger B hat nun aber gar nicht von einem Sender entsprechendes Wissen nachgefragt. Wenn man Daten i.S.v. Z6 versteht, liegen nun keine Daten vor. Es sei denn, es genügt, davon auszugehen, dass beim Entstehungsprozess der Daten auf der Empfängerseite Wissen nachgefragt wurde, um die Zeichen auf dem Papier als Daten bezeichnen zu können. Dies geht aber aus Z6 so nicht hervor.

Kuhlen impliziert mit Z6 eine relativistische Subjektivität von Daten, die nur schwer nachvollziehbar ist. Die Eigenschaft, ein Datum zu sein, hängt sicherlich nicht davon ab, welches Subjekt das Datum wahrnimmt.

E.2 Diskussion ausgewählter Definitionen des Terminus Wissen

Wissen als kognitive Struktur

„Wissen ist der gesicherte Bestand an Modellen über Objekte bzw. Objektbereiche und Sachverhalte, die partiell in einem Individuum (in Form seines Gedächtnisses), in einer gesellschaftlichen Gruppe, aber auch in einer Organisation, einem ganzen Kulturkreis oder in der Menschheit insgesamt als kognitive Struktur vorhanden sind. [...] Wissen ist sozusagen der statische Bestand, die Summe der bisherigen individuellen und kollektiven Erfahrungen oder Erkenntnisse, die mit guten Gründen in gewissem Ausmaß verallgemeinerbar sind, aber nicht nur auf Meinungen beruhen. [...] Im Zusammenhang von Information und Wissen ist dieses stets in einem Zusammenhang der Lösung von Problemen aufzufassen. Der Empfänger einer Information wird durch diese in die Lage versetzt, problemlösungsorientiert zu handeln. Dieses Problemlösungswissen steht in Zusammenhang mit solchem Wissen, das seinerseits im Prozess des Handelns eine Rolle spielt. [...] Angesichts des Aufgabenfeldes der Wissensvermittlung stellt sich selbstverständlich die Frage, wo sich zu vermittelndes Wissen überhaupt befindet. Wissen findet sich explizit in Dokumenten, implizit in den Köpfen der Menschen (nach Guntry bei Capurro). Dokumente können beispielsweise Bücher, Online-Archive oder Zugfahrpläne sein. Somit kann Wissen entweder direkt (explizit à implizit) oder durch menschliches Eingreifen (implizit à implizit bzw. explizit à implizit à implizit (Informationsarbeit)) vermittelt werden. Wissen muss hierzu über Medien (den Kanal) codierbar und transportierbar sein. Es bedarf also der Medialität, ohne die keine für den Wissensvermittlungsprozess notwendige Kommunikation möglich ist. Um Wissen medial zu vermitteln, muss es zunächst an das Medium angepasst werden ("verdatet"), ehe es kommuniziert werden kann. Verdatung bedeutet beispielsweise, wenn Wissen in einer Radiosendung verbreitet werden soll, dass es in akustische Sprachsignale übersetzt wird, da Sender und Empfänger über ein weitgehend gleiches Codier- und Decodiersystem verfügen müssen (z.B. natürliche Sprache)." (Z7)

(Kuntz, 2005)

Wenn man nach Z7 Wissen als kognitive Struktur versteht, bezeichnet (12) Wissen etwas, das intrapersonal vorliegt. Einer Organisation kann man sicherlich keine kognitive Struktur zusprechen, weshalb eine Organisation als Organisation (und nicht z.B. als das Gesamt ihrer Mitarbeiter und Mitarbeiterinnen) über kein Wissen verfügt. Kuntz ist jedoch dieser Meinung, weshalb bei ihm (13) Wissen zu etwas wird, das auch extrapersonal vorliegt. Dieser Widerspruch findet sich dann nochmals, wenn es um die Frage geht, wo sich zu vermittelndes Wissen befindet. Versteht man Wissen nach (12) als etwas, was nur intrapersonal vorliegt, kann sich zu vermittelndes Wissen nur und ausschließlich in den Köpfen befinden. Nach Z7 gibt es aber die Möglichkeit (14), dass Wissen explizit in Dokumenten vorkommt und (15) implizit in den Köpfen der Menschen. Wird Wissen in diesem Sinne verstanden, stellt sich die Frage:

Wenn Wissen nur in den Köpfen der Menschen vorkommt, wie kann es dann überhaupt zu explizitem Wissen, z.B. in den Dokumenten, kommen?

Sobald Wissen extrapersonalisiert wird, also z.B. versprachlicht und damit verdatet wird, liegt kein Wissen mehr vor, sondern - entsprechend der hier entwickelten Terminologie - ausschließlich Dateninstanzen. Dies entspricht dem, was Kuntz unter "Verdatung" versteht.

Wissen als Handlungsbasis

> *„Wissen bezeichnet Kenntnisse und Fähigkeiten, die der Mensch zur Lösung von Problemen einsetzt. Dazu gehören Erwartungen über Ursachen-Wirkungs-Zusammenhänge, praktische Alltags- und Handlungsregeln.“ (Z8)*
>
> (Probst, 2000)

In Z7 und Z8 findet sich ein wichtiger Ansatzpunkt, der bei der Annäherung an den Terminus Wissen weiterhilft, nämlich dass (16) Wissen die Basis für Handlungen und Problemlösungen stellt.

Wissen kann im Rahmen einer Verdatung auf Dateninstanzen abgebildet werden. Dies kann beispielsweise über formale Beschreibungsverfahren erfolgen, die auf diese Eigenschaft von Wissen (nämlich als Handlungsbasis zu dienen) zielen: Auf der Grundlage entsprechend formal beschriebener Fakten und Regeln können Folgerungen logisch abgeleitet werden. Die Regeln decken dabei typische "Wenn-Dann"-Konstrukte (materiale Implikationen) ab und haben dadurch einen stark handlungsorientierten Charakter. Entscheidend ist allerdings, dass dies nur für dasjenige Wissen möglich ist, das als Fakten, Regeln und Folgerungen expliziert werden kann.

Die Art von Wissen, für die man eine derartige Explizierung vornehmen kann, entspricht dem, was hier als explizit bezeichnet werden soll. Es sei aber nochmals expressis verbis darauf hingewiesen, dass hier nicht die Meinung aus Z7 geteilt wird, nach der Wissen in Dokumenten irgendwie stecken würde. Das explizite Wissen befindet sich nach wie vor ausschließlich in den Köpfen der Menschen, es kann jedoch formalisiert und in Dateninstanzen codiert werden.

Wissen als Prozessbeschreibung

> *„If you can't describe what you are doing as a process you don't know what you are doing.“ (W. Edwards Deming, Zitat nach Sure03, S.29) (Z9)*
>
> (Sure, 2003)

Folgt man Z9, müsste man für sehr viele Handlungen, die wir im Alltag ausführen, davon ausgehen, dass wir als Handlungsbasis dafür kein Wissen besitzen. Wenn man Wissen wie in Z9 als (17) die Fähigkeit versteht, das, was man tut, als Prozess beschreiben zu können, stellt sich die Frage: Was ist die Grundlage für Handlungen, für die man nicht in der Lage ist, einen genauen Prozess zu beschreiben?

Deming vernachlässigt hier das implizite Wissen, das ebenso wie das explizite Wissen in den Köpfen der Menschen vorhanden ist, und schränkt damit die Menge dessen, was Wissen darstellt, erheblich ein. Implizites Wissen kann nicht unmittelbar "verdatet" werden: dieses Wissen ist für den Menschen über Reflexion nicht unmittelbar zugänglich. Als Konsequenz kann nicht immer genau angegeben werden, anhand welcher Regeln ein Problem gelöst wurde oder in der Zukunft zu lösen ist. Die Prozessbeschreibung über etwaige darauf basierende Handlungen kann man in solchen Fällen nicht liefern, obwohl sie auf der Grundlage des vorhandenen Wissens erfolgen.

Als Paradebeispiel wird in diesem Zusammenhang oft die Intuition genannt, die hier nicht als Wissen verstanden werden soll. Trotzdem gibt es Wissen, das ad hoc reflexiv nicht für uns zugänglich ist, wie z.B. grammatikalisches Wissen über die eigene Muttersprache. Dies ist etwas, das wir aus der Praxis kennen und wissen, wofür wir aber nicht unmittelbar eine Prozessbeschreibung geben können, wie wir beispielsweise beim Bilden eines Satzes vorgegangen sind. Derartiges Wissen nennen wir implizit vorhandenes Wissen.

E.3 Reflexion ausgewählter Definitionen des Terminus Information

Wie Wissen ist auch Information ein weitläufig verwendeter und schwer abzugrenzender Terminus. Eine einheitliche, allgemein akzeptierte Definition des Terminus Information ist nicht vorhanden, obwohl verschiedene Wissenschaften Informationen und deren Verarbeitung zum Gegenstand haben. Im Folgenden werden ausgewählte Zitate betrachtet, die der Terminusentwicklung auf der Grundlage des entwickelten semantischen Datenmodells und des vorgestellten Sender-Empfänger-Grundmodells entgegenkommen.

Akt des Unterweisens

> *„Men so wise should go and inform their kings.“ (1330) (Z10)*
>
> (Oxford, 2007)

Dem Oxford English Dictionary nach lehnt die historische Bedeutung des Wortes inform im Englischen an (18) den Akt des Unterweisens an. Es bleibt dem Oxford English Dictionary folgend jedoch unklar, was man genau unter dem Begriff der Information versteht - den Prozess, das dabei Übermittelte oder das, was beim Empfänger ankommt.

Information als Vorstellung

> *„Atque ego in summo oratore fingendo talem informabo, qualis fortasse nemo fuit.“ (Z11)*
>
> (Cicero)

[Und wenn ich mir den besten Redner vorstelle, dann werde ich ihn so erschaffen (definieren), wie es vielleicht noch keinen gegeben hat.- Wörtlich: und bei dem sich vorzustellenden (Gerundiv) höchsten Redner werde ich einen solchen formen, wie vielleicht niemand war. Übers. d. Verf.]

> *„informatio, onis, f. (informo) 1. Vorstellung, Begriff: antecepta animo ursprünglich. 2. Erläuterung, Deutung.*
>
> *in-formo l. 1. formen, gestalten, bilden: clipeum V; bildl . animum. met. 2.bilden, unterrichten: ad humanitatem. 3. darstellen, schildern: oratorem, proscriptionem. 4. sich denken, sich vorstellen: deorum notiones, sapientem T.“ (Z12)*
>
> (Stowasser et al., 1979)

Nach Z12 kann Information zwei Bedeutungen zugesprochen werden: zum einen (19) eine Vorstellung (wörtlich: eine vom Geist vorweggenommene Vorstellung (von etwas)), zum anderen (20) eine Erläuterung. Weiter fällt auf, dass das Verb informare vier Bedeutungen haben kann, die sich auf den ersten Blick erheblich voneinander unterscheiden. Die Gemeinsamkeiten zwischen formen, unterrichten, darstellen und sich denken werden erst durch die angegebenen Zusätze verständlich. Es kann sich beispielsweise bei formen nicht nur um das (21) "Formen eines Gegenstandes" (clipeum V. - einen Schild, bei Vergil), sondern auch um das Formen im metaphorischen Sinne von (22) "den Geist bzw. Charakter formen" handeln. Bilden im Sinne von "Unterrichten" ist ebenso eher als (23) "Werte vermitteln" gemeint. In der letzten angegebenen Bedeutung wird das Geistige und Subjektive noch mal ganz deutlich durch die reflexive Verwendung (24) sich (etwas) vorstellen, z.B. im Zusammenhang mit "sich eine Vorstellung von den Göttern machen". Unter diesen Aspekten kann man den Terminus (sich) informieren als (25)ein Aufnehmen der Form eines Erkenntnisgegenstandes ohne dessen materiellen Stoff verstehen.

Es bleibt jedoch unklar, wie der Terminus Information genau verwendet werden soll: Information kann (nach 19) sowohl die "ideelle" Form, die dabei aufgenommen wird, darstellen als auch (nach 20) die "materielle" Form, die einem Erkenntnisgegenstand beigegeben wird, um die Aufnahme der ideellen Form zu erzeugen. Ebenso könnte man (20) so auslegen, dass mit Information der Akt des Informierens gemeint ist. Es wird hier nicht klar, ob Information je nach Kontext unterschiedliche Bedeutungen hat oder ob Information synchron mehrere Bedeutungsnuancen annehmen kann.

Entscheidungsgehalt

> *„It is desirable therefore to eliminate the psychological factors involved and to establish a measure of information in terms of purely physical quantities.“ (Z13)*
>
> (Hartley, 1928)

Nach Hartley wird (26) Information ausschließlich im Sinne eines physikalischen Prozesses betrachtet. Dabei wird ein rein technisches Übertragungssystem zugrunde gelegt, bei dem es einen Sender, einen Zeichenvorrat, einen Auswahlprozess und einen Empfänger gibt. Auswahlprozesse und damit einhergehende Entscheidungen sind dementsprechend notwendig, sobald es mehr als zwei Elemente in einem gegebenen Zeichenvorrat gibt (N = 2). Alle Elemente eines Zeichenvorrats besitzen nach Hartley den gleichen Informationsgehalt. Dies kann man sich an einem Beispiel veranschaulichen.

Bsp.6 Informationsgehalt nach Hartley

Gegeben sei ein Würfel mit den Zahlen 1 bis 6. Für einen Würfelwurf gibt es dementsprechend 6 mögliche Resultate (N = 6): 1, 2, 3, 4, 5, 6.

Der Informationsgehalt der einzelnen Resultate nach Hartley ist dabei immer gleich groß und lässt sich mit H = logN berechnen.

Die Einheit Hartley bezeichnet dabei den Informationsgehalt bei Verwenden der Basis 10. Die Hartley-Formel unter Verwendung der logarithmischen Basis 2 lässt Aussagen der Art zu, wie viele Bits resp. Entscheidungen minimal notwendig sind, um einen bestimmten Zeichenvorrat übertragen zu können. In Bsp. 15 könnte beispielsweise die erste Entscheidung das Aufteilen in zwei Gruppen (1,2,3) und (4,5,6) anhand des Kriteriums "<=4" erfolgen. Eine zweite Entscheidung könnte dann das Splitten in weitere Untergruppen nach dem Kriterium "gerade/ungerade" in (1,3), (2), (4,6) und (5) sein. Dann wäre noch mal eine Entscheidung notwendig, die das letzte Maß an Unsicherheit reduziert. Um die oben genannten Zeichen binär zu codieren, bräuchten wir dementsprechend 3 Entscheidungen bzw. 3 Bits. Somit kann man sich unter dem Hartley-Maß für den Informationsgehalt die grammatische Komplexität einer Zeichensequenz vorstellen, die durch die Länge des kürzesten Algorithmus, der die Sequenz generiert, dargestellt werden kann. Eine Sequenz vorgegebener Länge mit festem Alphabet hat nach Hartley immer eine maximal mögliche Komplexität, die ihrer Informationskapazität entspricht.

Das semantische Datenmodell geht davon aus, dass Daten eine Bedeutung zugeordnet werden kann. Informationsgehalt, wie er bei Hartley verstanden wird, lässt sich mit dem semantischen Datenmodell nicht vereinbaren, da Hartleys Ansatz nur syntaktische und keinerlei semantischen Komponenten berücksichtigt.

Nachrichtentechnischer Informationsbegriff

> *„In particular, information must not be confused with meaning. In fact, two messages, one of which is heavily loaded with meaning and the other of which is pure nonsense, can be exactly equivalent, from the present viewpoint, as regards information. It is this, undoubtedly, that Shannon means when he says, that "the semantic aspects of communication are irrelevant to the engineering aspects." But this does not mean that the engineering aspects are necessarily irrelevant to the semantic aspects.“ (Z14)*
>
> (Shannon&Weaver, 1972)

Während Hartley von einem Zeichenvorrat ausgeht, der sowohl Sender als auch Empfänger bekannt ist, entwickeln Shannon und Weaver das klassische nachrichtentechnische Sender-Kanal-Empfänger-Schema . Der Austausch von Information im technischen Sinn wird in diesem Kommunikationsmodell beschrieben. Dabei kodiert ein Sender Information, um sie über einen Kanal an den Empfänger zu senden, der die Information dekodiert (versteht). Im Zentrum des Interesses stehen bei Shannon und Weaver neben Codierungs- und Decodierungsprozessen auch die Kanaleigenschaften incl. auftretender Störungen und die Wahrscheinlichkeitsverteilung der Zeichen. Shannon verallgemeinerte Hartleys Formel und ging dabei davon aus, dass die Wahrscheinlichkeiten, dass als nächstes ein bestimmtes Zeichen aus einem gegebenen Zeichenvorrat gewählt werden kann, für jedes Zeichen anders sein kann.

Shannon und Weaver unterscheiden explizit drei Ebenen des Kommunikationsproblems - eine technische, eine semantische und eine pragmatische Ebene -, wobei sie sich ausschließlich dem technischen Problem, dessen Herausforderungen und Lösungen widmen und Information auch in diesem Kontext verstanden sehen wollen. Nach Shannon und Weaver kann man Information als etwas verstehen, was sich in der Außenwelt abspielt, unabhängig von einem erkennenden oder beobachtenden menschlichen Subjekt (27). Dies impliziert, dass Information ausgetauscht werden kann und etwas ist, was man von A nach B transportieren kann. Demgegenüber steht das semantische Datenmodell mit dem visualisierten Sender-Empfänger Grundmodell, bei dem davon ausgegangen wird, dass das einzige, was in der Außenwelt transportiert werden kann, Dateninstanzen sind. Nach Z14 ist diese Ebene der Dateninstanzen aber genau das, was als messages bezeichnet wird. Die Nachrichten (i.S.v. Shannon/Weaver) bzw. Dateninstanzen (i.S.v. Sieber/Kammerer) sind aber die Dinge, die in der Außenwelt übertragen werden - wo findet sich nach Shannon/Weaver also die Information? Losgelöst von semantischen und pragmatischen Komponenten, betrachten Shannon und Weaver lediglich die syntaktischen Eigenschaften der Nachrichten. Information kann in dem Kontext als (28) als eine Art Prädikat der übertragenen Nachrichten aufgefasst werden, das eine Aussage darüber trifft, wie überraschend (im Sinne von wie selten hinsichtlich stochastischen Verteilungen) eine Nachricht rein syntaktisch betrachtet ist.

Der nachrichtentechnische Informationsbegriff muss im Kontext unseres entwickelten Datenmodells jedoch revidiert bzw. erweitert werden, da hierbei die semantische Ebene betrachtet wird. Information ist nach diesem Verständnis kein Ding in der Außenwelt, das unabhängig von einem betrachtenden Subjekt existiert oder von einem System aufgenommen wird, ohne sich dabei zu verändern.

Information ist immer subjektiv

> *„Az információ tehát mindig szubjektív fogalom, függ a feldolgozótól. Az informázión igen sok oldalról lehet vizsgálni és elemezni. Az információnak vannak*
>
> *-statisztikai,*
>
> *-szintaktikai,*
>
> *-szemantikai,*
>
> *-pragmatikai,*
>
> *-apobetikai oldalai.“*
>
> (Kovács, 2004)

[Information ist immer subjektiv und hängt von der Interpretation ab. Information kann man von mehreren Aspekten betrachten: statistisch, syntaktisch, semantisch, pragmatisch und intentional . Übers. d. Verf.]

Der Unterschied zwischen pragmatischer und intentionaler Komponente wird an einem Beispiel veranschaulicht.

Bsp.7 Regen

Wenn der Sender weiß, dass der Empfänger einen neuen Regenschirm hat und er diesen sehen möchte, in der Kommunikation aber mitteilt, dass es regnen wird, mit der Absicht, dass der Empfänger sich dadurch veranlasst fühlt, seinen neuen Regenschirm mitzunehmen, dann haben wir den Fall, dass die pragmatische Komponente (Empfänger soll den Schirm mitnehmen) eine andere ist als die intentionale (Sender will den Schirm sehen).

Es stellt sich die Frage, ob und welche der folgenden Aussagen nach Z15 als Information zu bezeichnen ist: "es wird regnen", "ich muss meinen Schirm mitnehmen", "ich will den neuen Schirm sehen" oder "ich will, dass der Empfänger den Schirm mitnimmt"?

Nach Z15 ist Information etwas Subjektives (29): nehmen wir an, dass "es wird regnen" etwas ist, was außen ist, dann kann in einem ersten Abstraktionsschritt eine semantische Komponente abgeleitet werden. Unter der Voraussetzung dessen, dass Sender und Empfänger die gleiche Sprache verstehen, kann ein wahrnehmendes Subjekt die Bedeutung des Satzes "es wird regnen" durch die Bedeutung seiner einzelnen Bestandteile und seiner grammatikalischen Zusammensetzung ableiten. Weitere zusätzliche potentielle Interpretationen - aus Sicht des Empfängers: "ich muss meinen Schirm mitnehmen" und aus Sicht des Senders intentional: "ich will den neuen Schirm sehen" und perlokutional: "ich will, dass der Empfänger den Schirm mitnimmt" -können dann ebenfalls als die beim Subjekt gebildeten Abstraktionen betrachtet werden. Nach Z15 ist Information etwas Subjektives. Wenn man Information so auffasst, muss man davon ausgehen, dass sich Information verändert, je nach dem, welche Subjekte an der Kommunikation beteiligt sind. Es wird nicht klar, was nun genau als Information zu bezeichnen ist. Ist die Bedeutung des Satzes "es wird regnen" eine subjektiv vorhandene Information oder beispielsweise der im Empfängersubjekt ausgelöste Vorsatz, den Schirm mitzunehmen. Oder ist gar der Satz "es wird regnen" nur dadurch eine Information, weil ein wahrnehmendes Subjekt rein theoretisch unter gewissen Voraussetzungen in der Lage ist, daraus eine darauf basierende Handlung abzuleiten? In Z15 finden wir keine eindeutigen Antworten auf diese Fragen, was jedoch auf der anderen Seite sicherlich die erste Aussage (29) bestätigt: Information ist etwas Subjektives und dementsprechend in der Reflexion darüber auch nur dem wahrnehmenden Subjekt zugänglich.

Bedeutung für Empfänger

> *„Wird eine Zeichenfolge übertragen, so spricht man von einer Nachricht. Die Nachricht wird zu einer Information, wenn sie für einen Empfänger eine Bedeutung hat.“ (Z16)*
>
> (Hansen&Neumann, 1978)
>
> *„Information is data that has been given meaning through interpretation by way of relational connection and pragmatic context. This "meaning" can be useful, but does not have to be. Information is the same only for those people who attribute to it the same meaning. Information provides answers to "who", "what", "where", "why", or "when" questions. From there, data that has been given meaning by somebody and, hence, has become information, may still be data for others who do not comprehend its meaning.“ (Z17)*
>
> (Keller&Tergan, 2005)

In Z16 und Z17 ist ein Widerspruch zum semantischen Datenmodell, da im semantischen Datenmodell bereits Daten eine Bedeutung zugeordnet ist und dies nicht das Alleinstellungsmerkmal von Information ist. Auf der anderen Seite kann man den Fokus auf Teilstellen der Zitate legen und durch deren Betonung durchaus auf Grundlage des semantischen Datenmodells Unterscheidungen zwischen Daten und Information finden. Wenn in Z16 beispielsweise die Rede davon ist, dass die Nachricht für einen Empfänger eine Bedeutung haben muss, schließt es nicht aus, dass prinzipiell der Nachricht bzw. den Daten eine Bedeutung zugeordnet wird und nun, wenn es für einen Empfänger eine spezielle Bedeutung hat, eine Information wird. In Z17 kann man ähnlich ansetzen, indem man auf die Art und Weise fokussiert, auf die Daten interpretiert werden sollen: durch relationale Verknüpfung und pragmatische Kontextualisierung. Hier finden sich durchaus Erweiterungen zum Abstraktionsprozess, der für die Datenfindung notwendig ist, und der pragmatische Charakter ist sicherlich etwas, worin sich Informationen von Daten unterscheiden. Unglücklich ist jedoch in beiden Zitaten, dass die Begrifflichkeiten wie Bedeutung nicht erläutert bzw. nicht ausreichend eingeführt werden.

Information als Metakategorie

> *„Der Ausdruck "Information" (singulare tantum) meint, als Metakategorie, dass eine Aussage den Charakter einer geltenden Antwort auf einer "mitgeteilten" Frage hat. Nicht alle Formen des Redens sind also Informationen - eine Bitte oder ein Befehl sind keine Informationen - und auch nicht alle Formen von Aussagen, sondern nur die, die sich als Antwort auf eine Frage verstehen. Schließlich sind auch nicht alle Antworten Informationen, sondern dieses Prädikat kommt ihnen zu, wenn sie innerhalb eines ausdrücklichen sozialen Rahmens - zum Beispiel einer wissenschaftlichen Fachgemeinschaft - sowie eines ausdrücklichen Vorverständnisses und wenn ein Kommunikationsprozess in Form von Sprechen-Hören-Verstehen stattfinden.“ (Z18)*
>
> (Capurro, 2000)

Capurro führt hier zunächst in Z18 den Terminus Information als singulare tantum ein, spricht dann aber von Informationen und lässt den Leser im Ungewissen, was nun unter Information zu verstehen ist, wenn es auch im Plural verwendet wird. Nach Z18 kann man Information als (30) eine Metakategorie einer Aussage verstehen, die die Werte "wahr" oder "falsch" annehmen kann. Zusätzlich gilt die Einschränkung, dass der Wert wahr nur dann angenommen wird, (31) wenn die Aussage den Charakter einer geltenden Antwort auf eine mitgeteilte Frage hat, (32) wenn die Aussage innerhalb eines sozialen Rahmens stattfinden, (33) ein ausdrückliches Vorverständnis vorhanden ist und (34) ein Kommunikationsprozess in Form von Sprechen-Hören-Verstehen stattfindet. Dabei stellt sich aus dem Blickwinkel der Technischen Dokumentation der Schluss, dass im Prozess Schreiben-Lesen-Verstehen nach Capurro der Begriff Information nicht anzuwenden ist, was so nicht nachzuvollziehen ist. Des weiteren wirft Z18 folgende Fragen auf:

Wann können wir von einer geltenden Antwort auf eine Frage sprechen?

Auf welche Art und Weise muss die Frage mitgeteilt sein?

Wann liegt ein ausdrücklicher sozialer Rahmen vor?

Wann kann von einem ausdrücklichen Vorverständnis ausgegangen werden?

Wir veranschaulichen uns das an einem Beispiel.

Bsp. 8 Zugfahrplan

Die konkret gestellte Frage "Wann fährt der nächste Zug nach Berlin?" wird auf verschiedene Arten beantwortet. Das fragende Subjekt befindet sich beispielsweise am Bahnhof in Hamburg und erkundigt sich dort selbst anhand eines dort hängenden Zugabfahrplans nach der genauen Uhrzeit und findet eine Zeile mit der Angabe, die ihm die Antwort auf seine Frage gibt.

Betrachten wir diese Zeile als Aussage, ist diese also nach (31) als Information zu betrachten, bzw. hat die Metakategorie Information=TRUE. Da nach (34) aber kein Kommunikationsprozess stattgefunden hat, müssen wir durch die geforderte UND-Verknüpfung der Einschränkungen die Metakategorie Information=FALSE annehmen.

Alle anderen Angaben auf dem Zugfahrplan haben nach Z18 bereits per se keinerlei Berechtigung, die Metakategorie Information=TRUE zu führen, da sie auf die mitgeteilte Frage nicht die geltende Antwort geben und nach (34) kein entsprechend geforderter Kommunikationsprozess stattfinden wird.

Das fragende Subjekt erkundigt sich nun bei einem Bediensteten der Deutschen Bahn AG und bekommt die Antwort "in zehn Minuten". Diese Aussage hat nun nach (31) ebenfalls die Metakategorie Information=TRUE. Nehmen wir sämtliche einschränkenden Bedingungen ebenfalls als erfüllt an, können wir hier nun also tatsächlich von der Metakategorie Information ausgehen, die für die Aussage "in zehn Minuten" wahr ist.

Es wird im obigen Beispiel nicht klar, warum sich eine gesprochene Aussage den informativen Charakter betreffend von der Zeile auf dem Zugfahrplan unterscheidet. Durch die einschränkende Wirkung der Beschränkung des stattzufindenden Prozesses auf Sprechen-Hören-Verstehen schränkt Capurro die Potenzialität tatsächlich wahrgenommener Informationen erheblich ein.

An folgendem Beispiel wird auf eine andere Problematik in Z18 hingewiesen.

Bsp. 9 Herdplatte

Die Situation ist alltäglich: eine Mutter kocht am Herd, während ihr gerade zweijähriges Kind parallel in der Küche herumtobt. Als das Kind sich mit seinen Händen am Herd hochziehen will, ruft die Mutter "Finger weg vom Herd!"

Nach Z18 sind Befehle nicht befähigt, die Metakategorie Information=TRUE innezuhaben. Im Beispiel wird deutlich, dass hier der Aspekt vernachlässigt wird, dass der informative Charakter einer "Nachricht" von einem wahrnehmenden Subjekt durch Relation mit vorhandenem Vorwissen gebildet und in handlungsfähige Entscheidungen umgesetzt wird. Die Befehlsform ist sicherlich in einigen Situationen angebracht, die zu übermittelnde Nachricht so anzureichern, dass das wahrnehmende Subjekt schnellstmöglich in der Lage ist, den für ihn relevanten pragmatischen Charakter zu erkennen und dementsprechend zu agieren.

Information als Veränderung

> *„Die Teilmenge von Wissen, die aktuell in Handlungssituationen benötigt wird und vor der Informationsverarbeitung nicht vorhanden ist. Information ist demnach entscheidend von den Kontextfaktoren der Benutzer und der Nutzungssituation abhängig." (Z19)*
>
> (Kuhlen, 1999)

> *„Wenn Information wahrgenommen wird, wirkt sie auf die Wissensstruktur des Empfängers, indem sie sie verändert.“ (Z20)*
>
> (Ingwersen, 1992)

Kuhlen betont in Z19 den (35) pragmatischen Charakter von Information, allerdings ist nicht ganz klar, wie man von einer Informationsverarbeitung sprechen kann, wenn Information gleichzeitig die Teilmenge von Wissen ist, die vor der Verarbeitung noch nicht vorhanden ist. Hier wäre evtl. sinnvoller von einer Informationsgenese zu sprechen.

Nach Z20 kommt ein sehr interessanter Aspekt zum Zuge: (36) wahrgenommene Information verändert die Wissensstruktur des wahrnehmenden Subjekts. Da auf kognitionspsychologische Studien an dieser Stelle verzichtet wird, kann aus diesem Blickwinkel keinerlei Expertise beigetragen werden. Dennoch erscheint uns diese Aussage von Ingwersen zwei Kernpunkte zu treffen: zum einen, dass (37) im Zusammenhang mit Information immer ein wahrnehmendes Subjekt vorhanden sein muss und zum anderen, dass (38) im Wahrnehmen von etwas als Information ein Wirken auf die vorhandene Wissensstruktur ausgelöst wird/wurde. Hieran schließt sich auch Kuhlens zweiter Absatz an, indem er in Z19 betont, dass natürlich unter diesem Aspekt der sich verändernden Wissensstruktur eines Subjekts die Informationswahrnehmung entscheidend von den Kontextfaktoren des Subjekts abhängig ist.

Information als geistiges Abbild

> *„Während physikalische Objekte der direkten und unmittelbaren Wirkung ausgesetzt sind, die die Umgebung auf sie ausübt, verfügt ein organisiertes System über die spezifische Fähigkeit, diese Wirkungen zunächst abzubilden und auf diese Weise in systeminterne, in Form von Funktionen auftretende Wirkungen umzusetzen. Die abgebildete Wirkung tritt als Information in Erscheinung. Sie ist 'die untrennbare Einheit von Bedeutung (Repräsentation der Wirkung) und Funktion in der Hierarchie der Lenkung und Regulierung des Systemverhaltens'.“ (Z21)*
>
> (Fuchs-Kitkowski, 1976)

Wird der Informationsprozess als zeichenvermitteltes Widerspiegelungsverhältnis aufgefasst, ist Information (39) eine durch aktive Tätigkeit des Empfängers zustande gekommene Konstruktion. Auf der Ebene menschlicher Informationsverarbeitung betont dies die subjektive Seite von Information. Information ist ein Abbild des Objekts, das durch das Subjekt entworfen wird. "(Objektive) Information" wäre demzufolge ein unzutreffender Begriff, da Information immer auch an einen Empfänger gebunden ist.