

Metadata and Content Management Bridging Technical Documentation and Automation Technology

Dana Baro^{1 2}

¹Karlsruhe University of Applied Sciences, Faculty of Information Management and Media, 76133 Karlsruhe, Germany

²SKAN AG, Technical Editorial Department, 4123 Allschwil, Switzerland

Abstract. The use of taxonomic metadata in technical communication is not only transforming more and more to multi-hierarchical ontologies but also gaining attraction and new use cases beyond technical authoring. In the field of technical documentation, by classifying text modules product documentations can be generated automatically. This approach can also be applied to other documents containing classified objects, which should be merged according to certain specifications and requirements. Based on the way of working in content management systems (CMS), in this paper, a use case from pharmaceutical machinery industry (provided by SKAN AG) is shown in which the creation of a plant-specific alarm matrix should be implemented in a partially automated way.

Depending on the complexity of the plant, usually, the process to build an alarm matrix for a project takes an automation engineer approximately two weeks and shows non-standardized manual steps and elements, that make the process complex. By developing comprehensive standards of alarm matrices, the inheritance of properties to newly created matrices should be possible with the help of metadata in order to reduce the initial creation effort. For this purpose, individual components of the alarm matrix are to be managed as objects for the variant management and component identifications are to be automated through the machine processing of the planning documents. By working out the relations between individual documents, elements, and components, the necessary basis was to be created for making the underlying logic machine-interpretable.

1 Introduction

The company specializes in the development and design of process isolators for aseptic filling of biopharmaceutical products.[1]

At the Technical Editorial Department (TED) a CMS with its classification model has long been in use. Metadata stores information about each object that indicates its state, history, validity, and usage. Thus, metadata contain the administrative information about the content that is necessary for search operations and reuse. [2]

Against this background and the fact that documents like instruction handbooks have been created semi-automatically based on their metadata for several years, this approach has not yet been implemented for similar applications beyond technical writing.

2 Technological Background

2.1 CMS

CMS are used to create and manage modular and submodular text modules, which can be merged and published depending on their validity.

With the help of a classification model validities can be assigned to the content on a product- and information-related basis.

The goal of content management is the simplification, systematization, and automated aggregation of content objects. It supports the consistency, the reuse, and publishing process of content objects.

The TED uses the methods of content management in the CMS by dividing the modules according to their content into several content types.

These types are “standard”, “template” and “project-specific”. Modules with the content type “standard” apply without restriction and can be reused without adapting the content. Modules with the content type “template” must be copied and be transferred to a module with content type “project-specific”. The content must be then be adapted to individual project-specific features.

Those modules are collected in a master document to create the publication. In order to create a new project, applicable standard modules are taken over according to the product configuration to be described and template modules are adapted project-specifically.

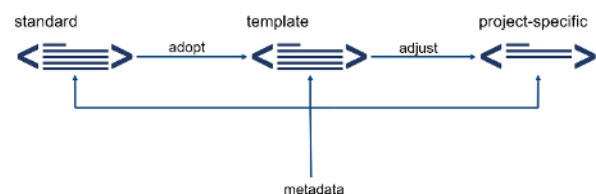


Figure 1 Content management concept TED

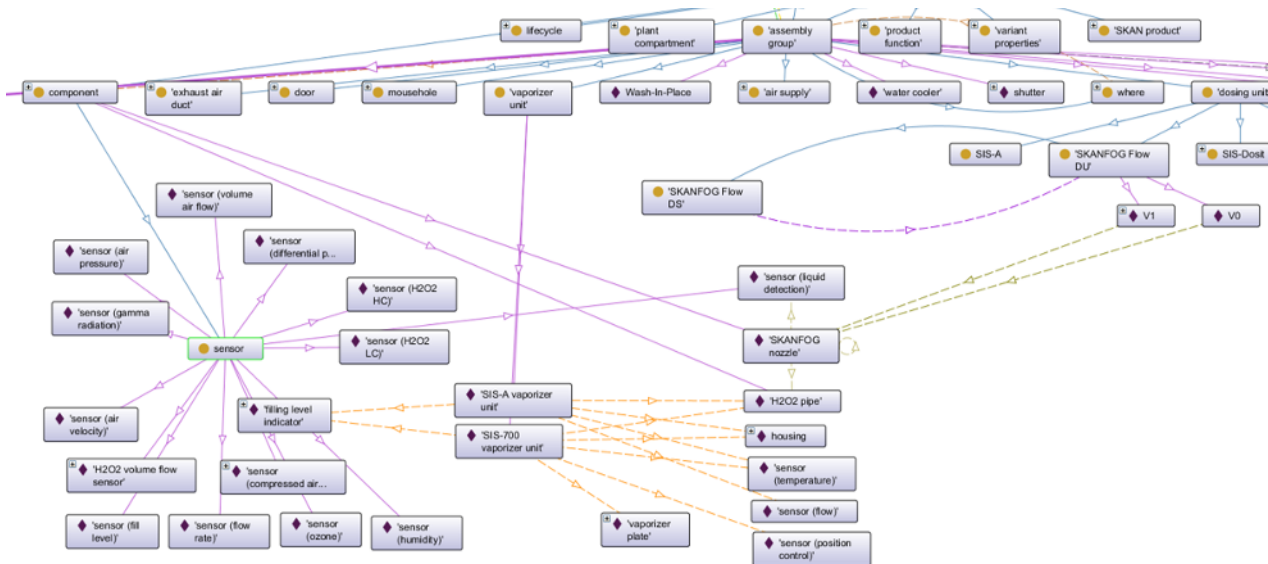


Figure 2 Extract from the ontology

2.2 Semantic Technologies

Semantic designates the study of the meanings of words and phrases in one language. Semantic technology uses formal semantics to help systems understand language and process information the way humans do. Thus, they can store, manage and retrieve information based on meaning and logical relationships. [3]

Semantic metadata provide information about the context of the content and its validity. In the field of technical documentation, we speak about native, augmented and artificial intelligence related to metadata. Native intelligence pursues the goal of process automation and builds on semantic information modeling and semantic metadata. Augmented intelligence aims to model the complexity of real-world information and products and to overcome the typical shortcomings of taxonomic modeling of metadata. The model includes objects with their properties and their relationships to each other. Artificial intelligence is used for automated knowledge and metadata extraction from data and content sources, and thus for system-side assignment of content to a specific ontology or taxonomy. [4]

An ontology represents knowledge by the usage of classes, subclasses, individuals, and defined properties to describe and relate them. Thus, it is possible to create models which contain machine-interpretable statements about things. The open-source software Protégé is used to model the ontology in the semantic web language OWL (Web Ontology Language). OWL is designed to represent comprehensive and complex knowledge about things, groups of things, and relations between them [5].

As shown in **Figure 2** within an ontology many aspects can be considered and modeled. For example, different sensor types as well as different product variants, its components but also different lifecycle phases and a lot more. This is what levels up the intelligence respectively the semantic richness of metadata.

3 Use Case “Automation”

3.1 Department Automation

The Automation Department is responsible for the management of the automation part of the projects from contract review to handover to the customer, for the coordination and cooperation during commissioning in house and at the customer's premises. Additionally, the automation engineers support the customer service with their knowledge and assist the technical editors by providing information for the creation of the user manuals

Furthermore, automation engineers are responsible for the planning and development of software for SPS controls, user interfaces and their documentation. In addition, the department identifies, develops, and maintains programming-supporting applications. The execution and documentation of program tests within the scope of a standard and GMP (Good Manufacturing Practice) -compliant controller qualification is part of the automation department as well.

3.2 Alarm Matrix

An alarm matrix is used to plan and control error messages for individual components within a project by defining at what time and under what conditions which actions are triggered by the system in case an error occurs. It serves as a basis for implementation on the programming side.

The alarm messages, shown on the human machine interface (HMI), are part of the alarm matrix. They are the communication component that is presented to the operator in case an error occurs.

Behind the alarms there is a logic after which they are created (see **Figure 3**). This structure defines three main parts of an alarm:

- tag
- object classes
- fault

The tag, which identifies the component, has its own logical structure which is mostly applied as standard. The first number represents the filling line, the component is part of. The following numbers allow conclusions to be drawn about specific components, assembly groups and on the number of times they are installed inside the plant. Using this tag, it is possible to define reactions for each fault possible for specific components or assembly groups across multiple filling lines and multiple installed components.

The object classes represent answers to the questions what exactly has or is causing a fault in the equipment, where the error originates locally and when, meaning during which plant mode or decontamination phase, it occurred. The fault defines the type of the error. The alarm in **Figure 3** represents a fault in which the air velocity at the outfeed section during the plant mode “high speed” is too high.

Additionally, with specifications regarding the alarm reaction, alarm set points and alarm delays it can be determined what should happen in case of an error, what specific values characterize a critical point during specific plant modes and when an error reaction starts running. Reaction patterns are representative numbers for different pattern sets in which an automation engineer defines which measured values are critical and should cause which reactions of the plant including showing alarms on the HMI and under which conditions they are active or passive.

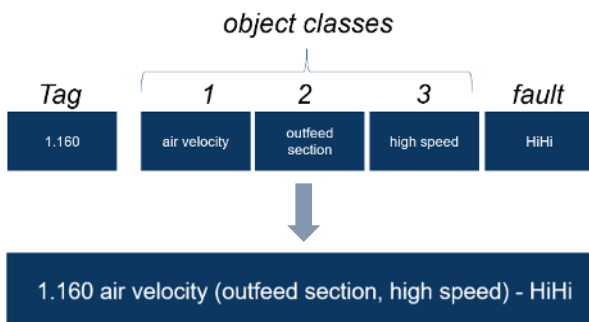


Figure 3 logic and components of an alarm message

Regarding the alarms the automation department and the TED communicate closely, because the technical writers are responsible for the troubleshooting within the instruction handbook and the operating manual. They receive the corresponding information how to proceed in case an alarm occurs from the automation engineers.

3.3 Actual Process

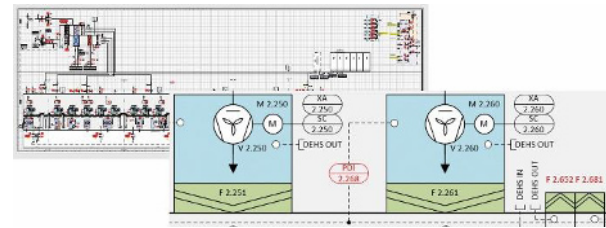


Figure 4 P&I Diagram

Previously, the creation of an alarm matrix starts with the P&I Diagram (Pipe and Instrument Diagram, see **Figure 4**). This diagram is a technical drawing of the plant including the tags an automation engineer uses for the matrix to implement corresponding reactions for the components in case of an error. Using this diagram, the automation engineer creates an element list by reading out the tags.

The element list is used to define the elements and for the assignment to the standardized software structure. After defining the reaction patterns, the alarm matrix is ready to be used for the plant and further programming tasks.

Of course, there are already existing alarm matrices which are used as a template for a new alarm matrix. However, this type of reuse is not the intended one.

3.4 Target-Process

With the help of a software solution the process of creating an alarm matrix should be supported: On the one hand by automizing manual steps and on the other hand by supporting the creation of standards.

Developing a standard can take place by using a top down or bottom up approach. Bottom up means taking existing matrices to compare them and to define the greatest possible common quantity to use as standard. Since previous attempts to specify a standard top down failed due to the large product variance, the bottom up approach was chosen. During this upcoming process, data from existing project documents will be fed into the system and used to try to identify the highest common concordance within the different product groups. The analysis of data to create a standard can be supported by the capabilities of the semantic network. Also, functions have been implemented within the software that allow the comparison of different alarm matrices, making it easier for the user to determine matches.

Thus, it is possible to already use the tool while the standard is further optimized in parallel.

By developing comprehensive standards of alarm matrices for the different product groups, the inheritance of properties and the standard data set to newly created matrices should be possible in order to reduce the initial creation effort. For this purpose, individual components of the alarm matrix are to be managed as objects for the variant management and component identifications are to be automated through the machine processing of the planning documents.

By working out the relations between individual documents, elements, and components, the necessary

basis for making the underlying logic machine-interpretable can be created.

Due to the need to manage specific parts of an alarm matrix specifically, further methods of the TED could probably get applied to the creation of an alarm matrix. When it comes to definable variant combinations the implementation of submodular content variant management can come to play. In technical documentation this method is used to reuse smaller content units (fragments) as modules such as warnings for example. The prerequisite for this is that consequently also the classification takes place on a submodular level in order to be able to specify the corresponding validities. If pattern sets are classified, they can be maintained and reused as submodular entities. This promotes the use case specificity when generating an alarm matrix. Beforehand, however, it must be verified whether the submodular classification is feasible during within the working processes of the engineers and on the one hand whether this very use case specificity cannot already be implemented by the intelligence of the system on the other.

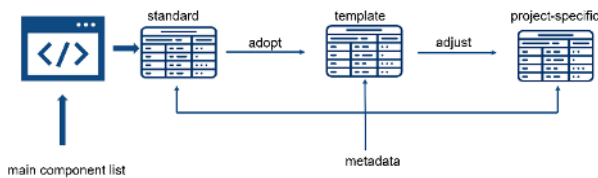


Figure 5 Target process creation alarm matrix

The bridge between technical documentation and automation technology becomes strongly visible when considering the new planned process in **Figure 5**: The automation engineer uploads the main component list or the P&I Diagram from the project to the system and chooses a standard matrix which should be applied to the new data. The assignment can be made automatically by the usage of the component tags. This means that only patterns and alarms get assigned which are present in the uploaded data of the new project. The reuse creates a new matrix template which can be adjusted for the customer specific requirements. Metadata are applied as well so new alarm matrices can inherit metadata from the standard and new validities can be implemented.

Additionally, subsequent changes can be implemented by creating a new version.

4 Project Knowledge Warehouse

The project Knowledge Warehouse focusses on the provision of service-relevant information including knowledge of service employees, existing documents and navigated diagnosis. The Knowledge Warehouse should serve as a central database for information to support troubleshooting processes. In addition to various systems, the main data sources are also and especially knowledge from the service employees and experts. The tool provides templates which can be filled with experiential knowledge as well as guided troubleshooting similar to decision trees to help identifying the problem's cause.

To manage and retrieve the information the ontology is used to classify the content.

Merging the semantic information from the automation use case and the ontology creates an even larger knowledge model that further extends the intelligence of the metadata. This influences the content retrieval positively and ensures the delivery of use-case-specific information.

5 Conclusion and Outlook

By analysing the alarm notifications not only the automation processes was considered, but also a direct connection between the automation processes and the technical documentation was established: the components (object classes and faults) of the standardized alarm notifications (Figure 3) are already integrated or can be represented in the classification concept. The object classes by implementing product metadata, plant modes and decontamination phases, and the faults within the integration of functional metadata including specific events under which the errors were located.

Against this background, it will be possible in the future to avoid deviations due to inconsistent designations or deviating designations or spellings. Because with the stored logic of the alarms and the ontology, the creation of new alarm messages can be guided and supported by the system.

In addition, with regard to the creation of alarm matrices and the technological approaches, it is possible, as the number of alarm matrices increases, to use the data from the semantic network and existing projects to infer which values are the most likely in which contexts, so that the degree of automation can be further increased. Using semantic technologies to map relations and logics, the deposited information can be used to support the search of service-relevant information in connected content delivery applications. Additionally, the tool can interpret the search queries more accurately and present potentially unknown relationships to the user, which supports the information retrieval and reduces the effort of the information retrieval.

Accessing information in real-time will be largely improved with the optimal use of multi-hierarchical ontologies and the development of a well-fitting system is an important application area in technical communication, explaining how a complex information system could be set up in the age of the Internet of Things in the industry.

In the sense of standardization across departments, the development of a classification concept that can be used uniformly to replace existing concepts differing from each other, represents a desirable goal.

The understanding and representation of technical and product-specific backgrounds to link all company-specific knowledge within the ontology can drive many different applications such as process automation, knowledge and terminology management as well as use-case-specific provision of information for employees and customers in general.

The described use case shows how metadata and content management methods can be the bridge between technical documentation and automation technologies. This bridge is desirable in numerous other applications in TED in order to create added value for the company and ultimately also for the customer.

The competencies and qualifications, initially from the Technical Editorial Department, have great cross-departmental potential. In addition, they close the gap between content creation and application engineering by architectural, planning and modelling capabilities and analytical knowledge for process improvement.[6]

Special thanks to the Faculty of Information Management and Media for the opportunity to be part of ETLTC conference and Ulrich Barth and Benjamin Gallmann from SKAN for their support.

References

1. SKAN AG, <https://skan.com/unternehmen/uber-skan/> [last call 11/11/2021]
2. P. Drewer; W. Ziegler (2009): *Technische Dokumentation*, Würzburg: Vogel Buchverlag (2009)
3. Ontotext (n.s.): *What is Semantic Technology?*, <https://www.ontotext.com/knowledgehub/fundamentals/semantic-web-technology/> [last call 14/01/2022]
4. W. Ziegler (2017): *Wie intelligent können unsere Daten denn noch werden?*, https://www.i4icm.de/fileadmin/content/HSKA/03_Vortraege/1981_Wie_intelligent_k_nnen_Informationen_denn_noch_werden_.pdf. [last call 01/27/2022]
5. OWL Working Group *OWL*, <https://www.w3.org/OWL/> (2012) [last call 2021/01/25]
6. W. Ziegler (2022): *New Roles and Competencies in Technical Communication Induced by Semantics and Analytics*, The ACM Chapter International Conference on Educational Technology, Language and Technical Communication ETLTC