ACM CHAPTER PROCEEDINGS ON EDUCATIONAL TECHNOLOGY, LANGUAGE AND TECHNICAL COMMUNICATION

THEME:
INFORMATION DESIGN AND MANAGEMENT

2019
VOLUME 1

EDITED BY:
DEBOPRIYO ROY

THE UNIVERSITY OF AIZU, JAPAN

PUBLISHER:
ACM CHAPTER ON ELEARNING AND TECHNICAL COMMUNICATION

ISSN: 2434 - 9526

MESSAGE FROM THE EDITOR

The 1st Annual ACM Chapter Research Seminar on Technical Communication and Information Design was held at the University of Aizu, Aizuwakamatsu from January 28-February 1st, 2019. It was a highly successful event in the middle of the snowy Aizu winter, attended by over 75 participants from our university, and other universities from Japan and aboard. The theme of the conference focused on information design and management was reflective of our times and the emerging new trends in the discipline of technical communication. Research by content management experts in the software industry suggests that AI, personalization, analytics, open services platforms and vendor wars will be top trends in technical communication research in 2019. The presentations made in this conference by German faculty, and the master's students from Karlsruhe University of Applied Sciences (HSKA) helped us see the changing tides and the emerging importance of content management and delivery portals. The full research paper publications in this proceeding will help readers with an in-depth analysis, while the experience reports will allow a glimpse through how our student participants felt about the initiative. The central theme of content management and delivery was aptly supported by complementary presentations on AI, virtual reality, technical illustrations, and psychology in technical communication, biomedical engineering and smart cities.

The conference enabled participants to attend presentations on a variety of topics from a wide spectrum in the fields in technical communication, educational technology and applications and language teaching and learning. The interesting feature of the conference was it's five-day layout, including a focus on intensive interactions between the different multicultural student groups over a sustained period of time, multiple workshops spread across the week, student presentations, exhibitions and posters, and importantly informal and formal lunch and dinner meetings to promote intercultural communication and activity. Student exhibitions allowed groups to showcase the idea of delivery portals, ontology, XML metadata modeling, virtual reality, 3D Printing, Android-based pronunciation app, IoT-based lighting, and other topics on which the student teams collaborated.

Heartiest thanks to Prof. Wolfgang Ziegler (HSKA), Prof. Miyazaki, Prof. Abderazek, and Prof. Shigaku Tei from UoA; Prof. Kuroda (JTCA & Osaka University) and Prof. Kishi (JTCA & Tokyo Gakugei University); colleagues at the Center for Language Research namely Prof. Wilson, Prof. Blake, Prof. Younghyon, Prof. Yasuta, Prof. Nicholas and Prof. Perkins; the student groups and volunteers at the University of Aizu and HSKA, Germany, and colleagues from the University of Monterrey, Mexico namely Prof. Romero and Prof. Campos. It was an all-round comprehensive effort and we could make it happen through intense collaboration and support.

This edition of the proceeding truly stands out due to the commitment shown by the volunteer reviewers, authors who showed so much interest in reviewing papers from peers in the community, providing scholarly feedback for improvement.

We will look forward to carrying on with this initiative and develop a robust partnership for research and teaching collaboration in years to come. We will continue to work on our ideas and expand on this collaboration in years to come.

May 23rd, 2019
Debopriyo Roy
Editor-in-Chief, ACM Chapter Proceedings on Educational Technology, Language and Technical Communication 2019

Website: https://www.acmchapterseminarjapan.com/

Articles from the 2019 ACM Chapter Research Seminar on Technical Communication and Information Design

The University of Aizu, Aizuwakamatsu, Fukushima, Japan

ACM CHAPTER CONFERENCE
ELearning and Technical Communication

ISSN # 2434-9526

This proceedings published by the ACM Chapter on ELearning and Technical Communication are a scientific journal for primarily technical communication, educational technology, information design and management and CALL-focused genres. However, other related topics and genres are welcome as well (based on editorial decision). This journal will be published annually in June-July as online version and as print on demand.

ACM Chapter on ELearning and Technical Communication
The University of Aizu, Aizuwakamatsu, Fukushima, 965-8580, Japan.
Group ID: 80944
Client Number: 9977421
Chapter Type: Professional
Chapter Status: Active as of June 15, 2010
Charter Date: June 15, 2010

TECHNICAL SPONSORS & PARTNERS

Table of Contents

SECTION # 1: Research Articles

SECTION # 2: Extended Abstracts

Copyright Information

The *Japanese National Centre for ISSN* at the National Diet Library, Japan will maintain the catalog record for this journal.
https://www.ndl.go.jp/en/data/issn/index.html

SECTION 1

RESEARCH PAPERS

ACM Chapter Proceedings on Educational Technology, Language & Technical Communication - 2019

FROM INFORMATION MODELING OF TECHNICAL CONTENT TO INTELLIGENT DELIVERY

2019 Volume 1

Anita Ahmadpour
Faculty of Information Management and Media, Karlsruhe University of Applied Sciences
Amalienstr. 81-87, 76133 Karlsruhe
GERMANY
ahan1011@hs-karlsruhe.de

Anja Wilhelm
Faculty of Information Management and Media, Karlsruhe University of Applied Sciences
Amalienstr. 81-87, 76133 Karlsruhe
GERMANY
wian1031@hs-karlsruhe.de

ISSN: 2434-9526

PAGES: 3-9

ACM Chapter Proceedings on Educational Technology, Language & Technical Communication

From Information Modeling of Technical Content to Intelligent Delivery

Anita Ahmadpour
Faculty of Information Management and Media,
Karlsruhe University of Applied Sciences
Amalienstr. 81-87, 76133 Karlsruhe
GERMANY
ahan1011@hs-karlsruhe.de

Anja Wilhelm
Faculty of Information Management and Media,
Karlsruhe University of Applied Sciences
Amalienstr. 81-87, 76133 Karlsruhe
GERMANY
wian1031@hs-karlsruhe.de

ABSTRACT
This paper allows a basic insight into the current processes of content management in technical communication. The research topic of this paper is the comparison between using standardized information models and working with individually designed information models from a didactic perspective. The goal of this comparison is to examine advantages and disadvantages of each model type, as well as of the use of content management systems (CMS) and content delivery portals (CDP). The approach of students at Karlsruhe University of Applied Sciences was to use both model types during their semester project, which involved creating, managing, and delivering content to a CDP with and without the use of a CMS. Simultaneously, the students underwent various learning processes and explored technical and methodological aspects of information modeling. As a result, they concluded that developing an individually designed information model improves comprehending the functioning of information models and of the delivery procedures. Furthermore, they demonstrated that working with a CMS simplified documentation and delivery processes in comparison to working without a CMS. While the students experienced the state-of-the-art workflow of content creation and delivery, current developments indicate that the introduction of an exchange standard on the market will enhance the process of content delivery.

CCS Concepts
• **Information systems → Hierarchical data models**
• **Information systems → Database utilities and tools**
• **Information systems → Extensible Markup Language (XML)**

Keywords
XML; information model; component content management system; content delivery portal; intelligent information; metadata.

ACM Chapter Research Seminar on Technical Communication & Information Design 2019, January 28-Feb 1, 2019, Aizuwakamatsu, Japan

ISSN: 2434-9526

1. INTRODUCTION
This paper focuses on the comparison between using standardized information models and working with individually designed information models from a didactic perspective. This topic is important for two reasons. First, there is a general need for companies to create, manage, and deliver significant amounts of intelligent information for their customers. Second, there has been a change in user behavior in the digital world nowadays in which users prefer to receive only those pieces of information that will help them solve their individual problems instead of reading through extensive documentation. There are different ways for companies to address and satisfy these needs, such as using suitable information models as well as content management systems (CMS) and content delivery portals (CDP). This paper demonstrates advantages and disadvantages of two model types, as well as of the use of CMS and CDP.

2. METHODS AND APPROACH
2.1 Context of Research
The context of this research study is a collaborative semester project of the students from Karlsruhe University of Applied Sciences, Germany, and students from the University of Aizu, Japan. During this semester, in the course *XML-Based Information and Content Management* by Prof. Dr. Ziegler, the main tasks of the German project team were to create an information and metadata model for content creation and management, and to find solutions for content delivery. For that purpose, they used content, mainly from the field of 3D printing, created by the project teams from the University of Aizu. The research team from Karlsruhe was provided access to the CDP SCHEMA Content Delivery Server and the CMS SCHEMA ST4 by the company SCHEMA Holding GmbH. Thus, the research team refers mainly to the options and possibilities of working with this provider in the examples presented here.

2.2 Working with Content Management Systems
In the world of technical documentation, content creation and management leans heavily on the use of CMS. CMS are used to create XML-based content, often referred to as modules, to

manage and assign metadata to the modules as well as to create document outputs for delivery to the target audience.

Standardized working procedures allow the user of a CMS to create modular content units (modules) by working with an XML-based editor. CMS editors use predefined structures and sequencing matrices, based on a Document Type Definition (DTD) or an XML Schema Definition (XSD), to standardize content and improve the quality of the resulting documentation. Further improvements in the quality of outputs are often achieved by combining the editor with language and terminology checkers. These system extensions verify the correct use of spelling and grammar, and check the produced content for proper application of technical or company terminology. In addition, CMS like SCHEMA ST4 provide different editing views during content creation. The content creator can thus use the editor view best suited for their knowledge level. Users with knowledge of XML might prefer working with an active view of the XML element tags, whereas users without this knowledge might prefer a more familiar editor view, one similar to standard text processing programs like Microsoft Word.

Created content units are managed using metadata to specify the modules. This can be achieved, for example, by adding certain life cycle or product-related data to the modules. Modules can be put together into content packages representing the documents for delivery to the target audience. An important feature of a CMS is that modules can be organized into different projects representing different target documentations. CMS are thus designed to allow modules to be reusable, enabling different target documentations to use the same modules without having to edit or recreate the module or its content. At the same time, necessary changes to a specific module can be performed on a single source module and are automatically updated in every reuse of the module. A CMS also makes it possible to manage different variants of a module. The need for variant management could arise, for example, from the development of a newer version of a product that leads to necessary modifications (such as additional information) within the earlier version of a module. Although variant management is another important feature of CMS in company operations, it will not be further introduced in this paper, as it was not one of the research topics of this student project.

CMS provide opportunities for system-standardized and -automated document creation and delivery that does not need any form of manual transformation. Content packages, consisting of modules managed in the CMS, can be exported from the CMS as different file formats, such as PDF for static digital outputs and print, or HTML for dynamic digital outputs for online or mobile documentation [1].

2.3 Content Delivery Portals: The Extension of Content Management Systems

CDP are the logical extension to the standardized content creation within a CMS. CDP enable content delivery based on intelligent information. The need for CDP resulted from a change in user habits of accessing information. Printed documentation is still a standard in technical communication due to legal requirements. Nonetheless, in the era of digitalization and Industry 4.0, it has become outdated as a standalone medium for content delivery. Customers demand digital and, increasingly, mobile documentation. Furthermore, they expect information tailored to their specific needs in a specific situation. As a result, documentation must evolve into selectively accessible information with various user-specific searching, filtering and navigational possibilities to improve usability and user experience, as well as to increase user productivity [2].

2.3.1 Intelligent Information: The Key to Content Delivery

Intelligent information provides the basis for content delivery. The classical uses of metadata in a CMS are to tag variants of a module, to mark a target audience, and to automatically create documents out of individual modules. Metadata are only used inside of the CMS; they are not included in the publication, but intelligent information delivery requires that metadata be included in the publication of content. Modules and their related metadata must be delivered to the content delivery server to make them accessible for specific user requests via different devices. To retrieve specific modules or specific information from a specific module, the CDP uses metadata or taxonomies for the implementation of different navigational, searching, and filtering options (see Figure 1).

A semantic search limits the resulting information provided by the CDP. Facets, based on the metadata imported into the CDP, are used to filter content and reduce the focus of the search to the information that is required by the user. On SCHEMA Content Delivery Server, facets can be directly imported from the CMS SCHEMA ST4, they can be created directly on the CDP, or they can be imported from other sources that fit the SCHEMA file structure for metadata navigational files [4].

Metadata are the key to intelligent information. The combination of a delivery platform and intelligent information is the basis for meeting new user requirements of Industry 4.0. Metadata describe exactly which content should be provided for which user in which situation. The need of the target audience as a future focal point of content delivery demands intelligent information to be defined as:

- the right information: correct and relevant to context and product,
- for the right person: personalized and tailored to the user’s role,
- in the right situation: appropriate for the user's context and access options,
- on any device: displayed via browser, on mobile devices (online and offline), or via augmented reality devices [5].

A frequently used approach to make information more intelligent is PI-Class ®. PI-Class ®, which was also used for this research, is a method for the classification of modular contents based on metadata that can be sorted via product-specific and information-specific criteria and are further differentiated into intrinsic and extrinsic categories [6].

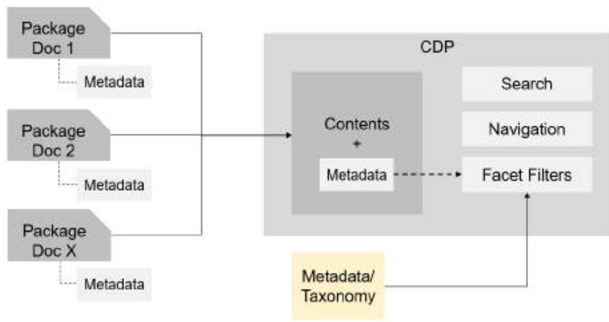


Figure 1: Context and assignment of metadata in content delivery, adapted from Kreutzer and Parson [3, pp. 25–26]

2.3.2 Content Delivery Starts from the User's Perspective

The main question to be considered for intelligent content delivery and modeling of intelligent information is: How do users formulate their information request? The standard information gathering process consists of three main steps: (1) The user sends a request for information to the server of the CDP. (2) The CDP understands the request and (3) provides the user with the needed information, tailored to their use case.

The use cases of content delivery services depend on user search behavior. The first use case concentrates on the point of view of service providers, like technicians. Usually, they focus on one specific component or aspect of a product. They do not need the context of the entire product. Therefore, the document structure of products is unnecessary to them. Consequently, the content needs to contain both the extrinsic and the intrinsic metadata so they can find the relevant information for the task they want to accomplish. The second use case revolves around the way end users search for information. For end users, the product's document structure is important because it is their way to orient themselves in order to find the information they need. Once the end users find the relevant document structure, they proceed to look for the specific information they need. This means that content packages or documents need to receive only extrinsic metadata. However, the individual modules within the content packages contain the intrinsic metadata that enable the filtering functions within the document structures.

Easy navigation, efficient searches, and filtering options are the factors that limit the amount of information for the user on the CDP. A CDP's navigational, searching, and filtering options depend on whether extrinsic, intrinsic, or both data are added to the content for the CDP. The research team adapted two different approaches on how to design the CDP's navigation, search features, and facet filters. The first approach is based on a single question arising out of the product-related context. In this case, from the user's perspective, only *one* question about a specific feature of the product must be answered. The user does not have any need for information about the product's context. This approach was executed by applying both intrinsic and extrinsic metadata to the modules uploaded to the CDP. Therefore, this approach covers the first use case explained above. The second approach focuses on a question regarding the product as a whole, that is, a user request based on the product documentation as a whole. Here the user does need the product context. As a result, only the extrinsic information for content packages were included in the delivery and were usable for navigational and searching features. Therefore, this approach comprises the second use case explained above. The choice of the respective approach to be taken for content delivery depends on user behavior.

2.4 Comparative Approach

In order to reflect and examine the processes of content creation, management, and delivery, the research team at Karlsruhe University of Applied Sciences decided to create, manage, and deliver content to a CDP in two different ways. One way was to use a CMS, namely SCHEMA ST4, to create content by means of the standardized CMS-specific information model, and to deliver it to a compatible CDP. The other way was to design their own information model from square one, to create content with it, to manage it manually, and to deliver it to a CDP by using manually programmed transformation processes.

2.5 Creation of the Individually Designed Information Model

The research team adopted a top-down approach to the individually designed information model. This is not the standard approach to such models. Usually, companies take the bottom-up approach, which is to examine, analyze, and categorize their information first and then create or decide on an information model that fulfils their needs. However, at the beginning of the semester, the research team did not know whether their research partners from the University of Aizu would send them descriptive or instructional content in text, video, 3D, or image format. As the research team did not know what kind of content they would receive, they had to design an information model that was able to cover all of those types of information and provide the basic structure to create the content in a modular fashion. The basis for the research team's modeling was XML Schema Definition (XSD). Using XSD, they defined a logical structure for their information model. All XML modules with which they subsequently created content were based on this information model. They also designed the information model for a reference structure that was the basis for creating a module that structured other modules into content packages. The research team designed both information models following the steps described in the next sections.

2.5.1 Identify Information Units

The research team identified various possible information units for the target content. For instance, they contemplated whether they would need to display content in lists or in tables and whether content required headings and instructional steps. After thorough consideration and discussions, they determined the necessary information units.

2.5.2 Create Elements

After determining the necessary information units, they created elements within the XSD based on the previously defined information units. For the elements, they used both layout-oriented and semantic names. Examples of layout-oriented names are *table*, *list*, and *title*. Examples of semantic ones are *action* and *preview object*. They also defined the rate of occurrence, that is, how frequently an element could be used in a module later on.

2.5.3 Create Element for Metadata Assignment

The research team created an element that enabled them to make subsequent content more intelligent: They created an element named *meta* with which they could determine the metadata for the subsequent content.

2.5.4 Assign Attributes to Elements

After having created all the necessary elements within the XSD, the research team assigned attributes to certain elements and decided whether those attributes were required or optional. For instance, they set the attributes *type* and *source* of the element *object* as required, whereas they set the attribute *scale* to be optional.

Whether attributes should be added to an element or not depended on the respective element and its possible use in future modules. For example, in the case of the element *object*, the research team decided to use it in order to reference media objects, such as videos, 3D models, text files, and images. Therefore, they added the attribute *type*, which can adopt predefined values, such as *image*, *video*, *cad*, *pdf*, and *thumbnail*. In the case of the element *object*, the research team decided to create only one element that could then be used to reference different types of media objects. They made this decision based on the fact that the overall structure and the principle of including media objects were identical. Therefore, attributes were necessary to differentiate between the various types of media objects referenced in the element *object*, as can be seen in Figure 2. Another approach could have been to create different elements for each media object; that is, to create individual elements for an image, video, 3D model, or text file. In this case, the assignment of attributes would not have been necessary in order to distinguish between the different types of media objects.

Defining some attributes as required and others as optional was contingent on two factors. On the one hand, it was influenced by the stylesheets of the automated transformation procedure. On the other hand, it depended on whether extra information in form of attributes was needed for the element to function. For instance, in the case of the element *object*, the attribute *source* was always necessary, because it contains the path to the respective media object. Without this path, referencing a media object would have been impossible. In contrast to that, there was a default size of all media objects defined in the stylesheets; therefore, assigning the attribute *scale* to the element was only required if a different size was necessary. For this reason, they set the attribute *scale* as optional.

```
<media>
    <object
        type="image"
        source="media/T_E_1_4_1.JPG"
        scale="small">
    </object>
    <caption>Illustration of powder
    bed fusion</caption>
</media>
<media>
    <preview_object
        type="thumbnail"
        source="media/T_E_1_4_2.jpg"
        scale="medium">
    </preview_object>
    <object
        type="video"
        source="media/T_E_1_4_3.MP4"
        scale="medium">
    </object>
    <caption>Selective Laser Sintering
    (SLS) Technology</caption>
</media>
```

Figure 2: A part of a module in which the element *object* was used with different values for the attribute *type*.

2.5.5 Assign Datatypes to Elements and Attributes

The research team assigned suitable datatypes to all elements and attributes within the XSD, such as the datatype *string* and *idref*.

2.5.6 Reuse Elements and Attributes

If the research team wanted to reuse certain elements or attributes within the XSD, they set them to global. Setting the elements or attributes to global declared the element or attribute for the entire XSD, which meant the research team could reference them at other places within the XSD, making it unnecessary to recreate the same element. Elements and attributes that were reused were, for instance, *meta*, *title*, and *p*, as well as *source*, *scale*, and *id* (see Figure 3).

2.6 Working with Individually Designed Information Models

After taking the steps in the sections 2.5.1 –2.5.6, the research team had designed three information models. In the file *main.xsd*, they defined the hierarchical structure and possibilities to embed different modules into one another. In the file *sub.xsd*, they defined the internal structure of every module type, such as the structure of the descriptive *concept* modules and of the instructional *task* modules. They included the file *sub.xsd* into the file *main.xsd* in order to bring the overall hierarchical structure together with the internal module structure. These were the two main files for content creation. In the file *ref.xsd*, the research team arranged the basis for the reference structure that was needed to assemble the modules into content packages and to organize them for transformational purposes. In the end, the research team designed three information models: one for an overall hierarchical document structure (*main.xsd*), one for the internal module structure (*sub.xsd*), and one for the reference structure (*ref.xsd*).

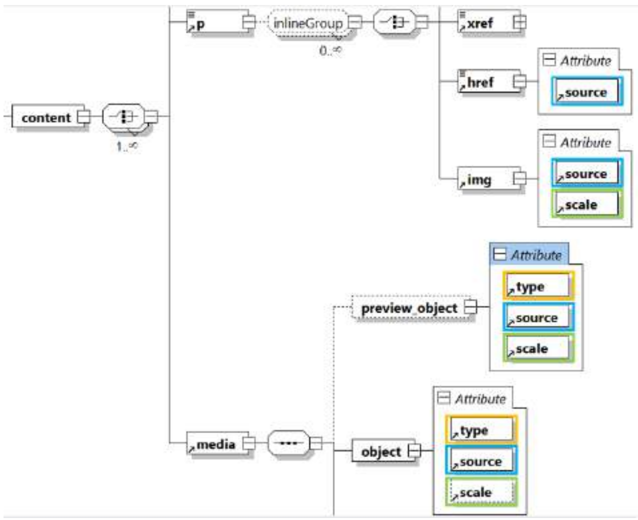


Figure 3: A part of the tree structure of the *sub.xsd* with reused attributes highlighted in different colors.

This tripartite structure of information models reflects the typical universal compilation of structures employed by systems that are oriented on book-format delivery. Depending on the tool used, however, there are certain differences in the use of information models or structures. For instance, the CDP TopicPilot by DOCUFY GmbH, comprises the overall hierarchical document structure and the internal module structure. It does not need any reference structure. Unlike TopicPilot, the CDP used in the context of this research project, SCHEMA Content Delivery Server, needs the reference and internal module structure to import content. Since the research team set out to understand the interdependency of the three elements of this tripartite structure as well as the overall structure's purpose in the currently used delivery processes, they applied the above-mentioned information models to create and deliver content.

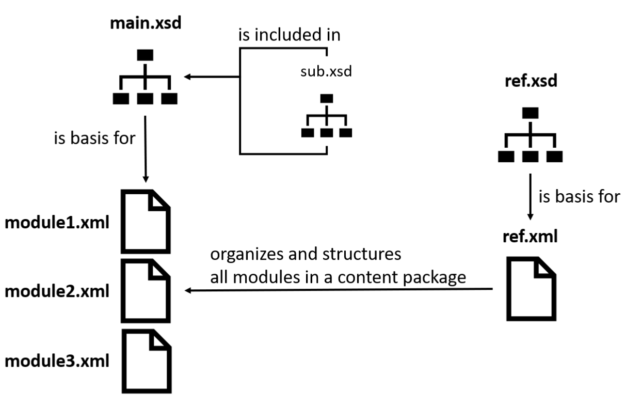


Figure 4: A simplified overview of the interdependencies of the modules and information models.

Figure 4 explains in more detail how the research team worked with the three information models. They created modules in XML format based on the information model *main.xsd* (with the file *sub.xsd* included). After transferring the content from their research partners into modules, they created an XML file (*ref.xml*) based on the information model *ref.xsd*. In the file *ref.xml*, they defined the structure of the respective content package and organized the order and hierarchy of every module. Afterwards, the modules and the reference structure *ref.xml* had to be transformed so that the content package could be uploaded on the CDP.

2.7 Working with Standardized CMS-Specific Information Models

Apart from working with their our own individually designed information model, the research team also used the CMS SCHEMA ST4 to create, edit, manage, and deliver content. They created modules in the CMS and edited them with the integrated editors, such as XMetal XML Editor. They uploaded media onto the CMS and referenced these media in different modules. They created taxonomies and assigned the metadata to the respective modules to make them more intelligent and thus enable future filtering on the CDP. They also assigned modules to projects to directly deliver them as content packages to the CDP SCHEMA Content Delivery Server.

3. RESULTS AND DISCUSSION

3.1 Comparison

While creating, managing, and delivering intelligent content to a CDP, the research team was able to experience both the advantages and disadvantages of using a standardized CMS-specific information model and of working with an individually designed information model.

3.1.1 Individually Designed Information Model

The research team discerned several notable advantages inherent in designing and using their own information model to create, edit, manage, and deliver content.

One such advantage is that they were free to design their own information model. That is, they could tailor the information model to their individual needs and wishes. They could thus specify or generalize certain information units as they pleased while omitting unnecessary units. During the design phase, the research team tried out different ways of determining different information units, for instance, by using one element with various attributes or creating several elements without any attributes (as explained in greater detail in Section 2.5.4). This experimentation taught them a great deal about the technical and methodological aspects of information modeling.

From a didactic point of view, moreover, designing their own information model helped them to understand what standardized information models are based on. They obtained the relevant background information and achieved deep insights into the basic pillars of information modeling.

Furthermore, the research team found that an additional advantage of modeling with XSD was the visualization of the information model as a tree structure. This graphic representation provided them with an overview of the sequencing of elements and attribute assignments, thus simplifying the development of the information model.

There were also several disadvantages to working with their own individually designed information model.

One of them was that because they intended to deliver the content on a CDP, they needed to develop an automated transformation procedure to transform the XML-based modules into a format compatible with the CDP, such as HTML or PDF. Programming such a transformation procedure turned out to be a complex task and needed a considerable amount of time and work.

Furthermore, in the case of changes within the information model, modules had to be updated and modified manually. Depending on the modifications, this could be time-intensive work. Over the course of the semester, several changes in the metadata concept caused modifications within the information model, which resulted in all modules having to be updated again.

As the research team did not work with a CMS, they needed to manage and store their content and resources, such as media files, in folders. That means that they had to invent their own logical systematic folder structure to keep the growing number of files in order. This manual management of content organization and storage also required considerable time investment. A related final disadvantage that stemmed from them not using a CMS was the lack of automated variant management, which in turn was another task that they had to perform manually.

3.1.2 Standardized Information Model with CMS

There were also several notable advantages to using a CMS and creating and editing content with the standardized CMS-specific information model.

Use of a CMS enabled exchanging, reusing, and delivering content into systems that are also based on or that support the same information model. Therefore, the research team was able to deliver content and metadata out of the CMS SCHEMA ST4 to the CDP SCHEMA Content Delivery Server simply by adjusting some settings in the CMS.

The CMS's integrated editor also assisted the research team in creating and editing content. For instance, it always showed what other possible elements could be selected at a certain position in the XML code of the modules. Thus, the research team did not need to know the information model by heart or look it up somewhere else, because the editor provided them with the relevant information at the right time.

The research team was able to make content intelligent by creating taxonomies in the CMS and adding the metadata to the respective modules by drag and drop. The research team considered this method as more time-efficient as compared to the manual metadata assignment in the case of the individually designed information model.

Another advantage of using a CMS was that the modules and media resources were stored in an automated fashion. Unlike working with the individually designed information model, while working with the CMS, the research team did not need to manage the content manually, resulting in time savings that they could use for further productive content creation.

There were, of course, also disadvantages to using a CMS for content creation, editing, managing, and delivery.

First, they felt constrained by the CMS's standardized information model. There were few possibilities to adapt the standard information model to their needs and wishes. As a result, in some cases, the information model was not general enough for their needs, while in other cases, it was not specific enough. For instance, the research team could not mark product names with an individual element because, in SCHEMA ST4, there was no such element. They thus had to invest time into finding a workaround. Consequently, they had to use another element, namely *italic*, to mark product names. The research team would have preferred a simple way of adding new elements to the standardized information model of the CMS to fulfil their needs.

While working with a CMS, the research team depended on certain systems and providers that were compatible with the CMS and the standardized information model. As mentioned above, the research team originally considered this to be an advantage because it enabled them to deliver content to a CDP with minimal effort. However, they also realized the potential difficulty that this dependency could cause: It would have been problematic and time-intensive work to deliver the content of SCHEMA ST4 to a CDP that does *not* support the same information model.

Another source for possible complications is the probability that necessary system updates might entail unexpected changes in the procedures. It would take time to get used to those changes.

3.2 Conclusion

All in all, the research team concluded that developing and using an individually designed information model improves comprehension of the functioning of information models and delivery procedures. In contrast to that, working with a CMS and a standardized information model simplifies documentation and delivery procedures, because the automated processes are time-efficient. Therefore, the research team suggests that learners of information modeling design their own information model, because this work returns deep insights into the field of information modeling. For companies that need to produce and deliver extensive documentation and intelligent information, however, the research team recommends the use of a CMS. The research team believes that the ideal situation would be for companies to invest in designing their own individual information model and integrating it within a CMS. This would enable companies to exploit the advantages of each format while using the advantages inherent in each approach to offset the corresponding disadvantage in the other. In brief, companies would benefit from both the advantages of automated processes and having an information model tailored to their exact needs and wishes.

3.3 Content Delivery: Current Challenges and Future Prospects

With the aid of a CDP, content providers can make optimum use of their modelled, intelligent information. Depending on the implemented CMS used during the documentation process, a restructuring of contents might be needed to make the content packages compatible for an export to a CDP. If the CMS provider has no connected CDP, the transformation of the content structures can become time- and labor-intensive, depending on the amount of changes to the existing content structure. At the present time, there is no common standard for content and metadata exchange established on the market. However, the need for such a standard has been recognized and is currently in development by tekom, the Association for Technical Communication, to take the next step in the field of content delivery in technical communication [3, pp. 22–24].

The future of content delivery relies on intelligent information modeling and metadata concepts. However, not every content package format and not every metadata model is supported by different CDP. On the one hand, considerable changes must often be applied to content packages and metadata created within a CMS to enable uploading them to a CDP from a different vendor. On the other hand, content and metadata creators working without a CMS must follow a CDP provider's often strict structuring rules during content creation and transformation. As a result, the next step in making intelligent information more easily accessible is the development of a standard to support a vendor-independent exchange and delivery of content. In April 2018, tekom published the first version of the iiRDS, the intelligent information Request and Delivery Standard. It aims to create a cross-vendor standard for delivering and exchanging intelligent information by defining a standardized vocabulary for metadata in technical communication and a common package format for content delivery widely supported by many CMS. The iiRDS consists of three main objectives [3, p. 24]:

- A common vocabulary for metadata in the form of a defined ontology, which is to be delivered with the content. This ontology is limited to the field of technical communication.
- An exchange package format to ease the exchange of intelligent information across CDP, web portals, and CMS.
- Detailed specifications describing the structure of the metadata model and the package format.

The iiRDS is designed for the delivery and exchange of intelligent information, not as a standard for the creation of intelligent information. Therefore, the ontology for technical communication cannot include company-specific vocabulary. However, the metadata model of the iiRDS can provide companies with a basis to develop their own company-specific metadata concept to complement the iiRDS metadata with the metadata of the company's field of expertise [7].

4. ACKNOWLEDGMENTS

The authors would like to express their gratitude to the International Office at Karlsruhe University of Applied Sciences, the University of Aizu, the Faculty of Information Management and Media, and the Master's program of Communication and Media Management at Karlsruhe University of Applied Sciences for the generous financial support that made this project and exchange possible. Furthermore, they would like to thank Prof. Dr. Debopriyo Roy for supervising the project groups, for supporting them throughout the semester, and for organizing and coordinating the ACM Chapter Conference. The authors would also like to thank SCHEMA Holding GmbH for providing their software for them to work with, and especially Tim Rausch for helping them to tackle technical questions about the CMS and CDP. Last but not least, they would like to thank Dr. Anthony Santoro for his constructive criticism and for proofreading this paper.

5. REFERENCES

[1] Drewer, P. and Ziegler, W. (2014). *Technische Dokumentation. Übersetzungsgerechte Texterstellung und Content-Management.* 2nd ed. Würzburg: Vogel Communications Group GmbH & Co. KG.

[2] Ziegler, W. (2018). Man muss auch austeilen können, *Technische Kommunikation*, *04*, 15-21.

[3] Kreutzer, M. and Parson, U. (2018). Intelligente Lieferung, *Technische Kommunikation*, *04*, 22-29.

[4] SCHEMA Holding GmbH. (2017). *SCHEMA Content Delivery Server: Product guide*. Nuremberg, Germany: SCHEMA Holding GmbH.

[5] Fritz, M. (2017). Was ist Intelligente Information?, *Intelligente Informationen. Schriften zur Technischen Kommunikation*, *22*, 11-25.

[6] Ziegler, W. (n.d.). PI Classification. *I4ICM – Institute for Information and Content Management*. [Online]. Retrieved from: https://www.i4icm.de/en/research-transfer/pi-classification/ (Accessed: 21 February 2019).

[7] Parson, U. (2017). Das Datenmodell der Technischen Dokumentation in iiRDS, *Intelligente Informationen. Schriften zur Technischen Kommunikation*, *22*, 26-39.

MODELING OF COMPLEX METADATA IN TECHNICAL COMMUNICATION BY ONTOLOGIES

2019 Volume 1

Heidrun Fischer
Karlsruhe University of Applied Sciences
Moltkestraße 30, 76133 Karlsruhe
Germany
fihe1014@hs-karlsruhe.de

Lara Krägel
Karlsruhe University of Applied Sciences
Moltkestraße 30, 76133 Karlsruhe
Germany
krla1016@hs-karlsruhe.de

ISSN: 2434-9526

PAGES: 11-16

ACM Chapter Proceedings on Educational Technology, Language & Technical Communication

Modeling of Complex Metadata in Technical Communication by Ontologies

Heidrun Fischer
Karlsruhe University of Applied Sciences
Moltkestraße 30, 76133 Karlsruhe
Germany
fihe1014@hs-karlsruhe.de

Lara Krägel
Karlsruhe University of Applied Sciences
Moltkestraße 30, 76133 Karlsruhe
Germany
krla1016@hs-karlsruhe.de

ABSTRACT
Managing content efficiently is one of the challenges companies face regarding information overload. Reasons for this overload are many different products, variants of these products and the corresponding documentation. There is a great demand for methods, which allow for an easy retrieval of information. Therefore, information is managed modular and metadata is used to classify these modules. Concepts behind metadata are becoming more and more complex due to the fact that there are many different relationships and dependencies between metadata. As a result, hierarchical models are insufficient in many cases. To solve this problem, some companies have started to use ontologies, because this opens up the possibility to create even non-hierarchical models. This paper's objective is to explain aspects and principles of ontology modeling and its importance for Content Management. There is various software that can help with the modeling of ontologies. ONTOLIS, a software platform for semantic networks, will be used as a software example for ontology modeling. Furthermore, other possible applications of ONTOLIS will be presented. This includes e.g. product, content or terminology modeling. The paper provides an insight into different ways of combining these models with each other. Advantages and disadvantages of ontology modeling using ONTOLIS as well as limits and potential of the software will be discussed. As a conclusion the relevance of ontology modeling and its implementation in ONTOLIS will be presented.

CCS Concepts
Information systems → Information systems → Information retrieval → Document representation → Ontologies

Keywords
Metadata; Technical Communication; Ontology; ONTOLIS; Data Modeling; Content Management; Content Delivery

1. INTRODUCTION
In recent years, there is a rise in the complexity of products due to high customers' expectations and the resulting customization of products. In addition to that, laws and guidelines regarding product liability make product documentation in many different languages necessary, which leads to an extensive amount of documentation. [3, p. 17]

Companies face the challenge to manage this large amount of documentation corresponding to their different products and the product variants in today's multilingual market. Consequently, one of the most pressing questions is: How can information be efficiently managed and easily retrieved?

A fundamental approach in Content Management is to divide the content into modules. Advantages of this method are the reusability of content sections (modules) in different contexts and the automated aggregation of modules into the required documents. In order to do that, metadata is used to classify the content modules for a professional Content Management. [13, p. 53]

The use of metadata (=data about data) is key, because it describes attributes or properties of the information object, which is to be managed and organized in a standardized way [5, p. 11-12].

Content Management Systems (CMS) as well as Content Delivery Portals (CDP) make use of metadata. CMS for an intelligent automation of product documentation and CDP one for basic search processes and search with facets to deliver information dynamically to the user [13, p. 54, 60]. The metadata in the CDP is either imported from the CMS or can be imported manually for content that does not originate from a CMS. This topic is explained in more detail in other contributions of the ACM chapter conference.

The metadata, which is used to classify the content, is based on classification systems like, e.g., the PI classification. These systems help to structure the metadata in a hierarchical way and are a useful tool for companies to classify their product information. Depending on the complexity of the products, the metadata concept tends to get very complex as well. Taxonomies and hierarchies can only offer a two-dimensional presentation of the product, whereas in reality metadata derives or should be derived from the underlying network of relationships between content objects. A new development in Content Management is, to make use of these knowledge networks in the form of ontologies. By using ontologies, metadata concepts can be extended and provide a more complete presentation of the information object and its various relations. [13, p. 60-62]

The relevance of ontologies in Technical Communication is one of the topics, which were explored in the cooperation project

ACM Chapter Research Seminar on Technical Communication & Information Design 2019, January 28-Feb 1, 2019, Aizuwakamatsu, Japan

ISSN: 2434-9526

between the University of Aizu (UoA) and the Karlsruhe University of Applied Science (HsKA) as part of the ACM chapter conference on Technical Documentation and Information Design. As one of the results of the cooperation project, this paper gives an overview of different aspects of ontologies and their use in Technical Communication.

2. ASPECTS AND PRINCIPLES OF ONTOLOGY MODELING

2.1 Ontology Definition

In the field of computer science, an often-quoted definition of ontology is the one from T. Gruber, saying that is as an "explicit specification of a conceptualization" [4, p. 1]. What does this mean exactly? Studer et al. offer an extended definition based on the first one, adding the attributes "formal" and "shared", making it a "formal, explicit specification of a shared conceptualisation" [11, p. 184]. They explain the different aspects of the definition as follows. The **conceptualization** is an abstract model of a part of the world, which is created by identifying the relevant concepts of that part. It is **explicit**, because the concepts, their use and their constraints are explicitly defined. Due to the fact, that the ontology is **formal**, a machine can also read it. Lastly the ontology consists of consensual knowledge, which makes it a **shared** conceptualization [11, p. 184]. The following sections give an overview over some of the most relevant aspects of ontology modeling in regard to the application use case, which is presented in this paper.

2.2 Elements of an Ontology

Ontology is set in a **domain**. The domain is the part of the world, which the model refers to. By using **classes**, the concepts of this domain can be described. The classes themselves are determined by **properties** (also roles or slots), which describe their attributes and features. The individual concepts of a class of concepts are called instances. There can be different **relationships** between the classes and instances, which can also be seen in Figure 1. [6]

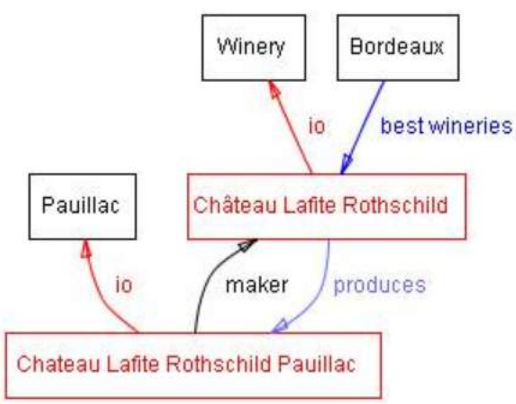


Figure 1. Example for classes, instances, and relations in the domain "wine industry" (black=class, red=instance, direct links=properties/instance-of/subclass-of) [6]

In this example (see Figure 1) "Château Lafite Rothschild" as a specific winery is an instance of the class "Winery". It has a relationship to "Chateau Lafite Rothschild Pauillac", because it is a manufacturer of this specific wine. The wine itself is an instance of the class "Pauillac".

Even though the example only shows a small extract of an entire knowledge network, it gives an idea on how ontology can represent a certain domain of the real world and how complex such knowledge networks can become eventually.

2.3 Ontology modeling

In order to develop ontology, it is necessary to define the classes of the domain in question. The classes should be arranged in a hierarchy, creating subclasses. Properties and allowed values need to be defined as well. [6] The modeling of various classes and relations develops into a network that sometimes can become very complex and confusing. For this reason, there is a great demand for software, which assists with the modeling and visualizing of ontologies. Mind mapping software packages, which are good for visualization, however, are not sufficient enough to model a real ontology that is machine-readable and can therefore be used for automated processes. Hence, it is necessary to describe the ontology in a certain description language such as RDF (Resource Description Framework) or OWL (Web Ontology Language) [1, p. 149-150]. There are software solutions, which specialize in ontologies and offer visual ontology editors that simplify their creation. Examples for software providers are protégé, iViews or ONTOLIS, which will be described in detail later on.

2.4 Ontologies in CMS and CDP

Ontologies have been widely used in different applications related to knowledge management or in the Semantic Web for many years [2, p. 415]. It is only a logical conclusion that the benefits of ontologies can be used in Technical Communication as well. As a result, there are efforts to use ontologies in Content Management and Delivery.

The ontological modeling of metadata leads to an augmented intelligence of the content, which in general helps automating processes related to Content Management and Content Delivery. [13, p. 51, 54] In CMS, ontologies can help to extend the metadata model and create more intelligent content. Unfortunately, many CMS do not fully support ontologies yet, but there are first software solutions and components that make it possible to use ontologies in CMS. [13, p. 61-62]

Figure 2. Using ontologies in CMS and CDP for augmented intelligence of information [Ziegler, 62]

There are also software solutions for using ontologies in CDP. The objective is to distribute information with the CDP intelligently in different contexts and use cases. This means, that the right information is available to the user at the right time in the right medium. Irrelevant information will not be shown. A manual search request is not always needed, since the request of an event could also lead to the delivery of the content. [13, p. 53-54, 60] Ontologies help to establish relevant relations between content objects, which the CDP can use to find the objects through these (multidimensional) relations. This improves the manual and automated search processes and allows a new and dynamic view on the information. [13, p. 53, 63-64]

In addition to the previously explained use of ontologies in Content Management, there are also efforts to combine ontologies with artificial intelligence (AI). One of the goals is to achieve further automation in Content Management related processes like the automated classification of content. By using analytical approaches of AI, namely text analysis, linguistic procedures, etc., the information can be classified according to the ontology. This helps with the import of originally unclassified content into a CMS or quality management, because it assures that classes are compliant with the given content. Moreover, it is helpful for CDPs, since it helps with the processing and classification of unstructured information with the classes of a certain ontology. Due to the fact, that the network of the ontology is filled automatically, the assigned metadata can be used in the facetted search. Another benefit is the improved interpretation of current or future search requests. [13, p. 64-65]

3. ONTOLOGY MODELING USING SOFTWARE

3.1 Approach of ONTOLIS

One software solution for modeling and visualizing ontologies is ONTOLIS, a web-based platform for semantic networks which is mainly used for product modeling, but can also be used for content, metadata or terminology models. The main software is called **ontolis base**, which is used to create ontologies. There are further extensions available for special requirements, e.g. the **Content Management Extensions** or the **Terminology and Localization Manager**. Figure 3 gives an overview over the existing and future software from ONTOLIS. [9, p. 12]

Figure 3. ONTOLIS's software structure [9, p. 12]

ONTOLIS defines ontology as an "abstract model of a part of the real world (domain)" that consists of **nodes** (instances), which are connected to each other by **relations** [7, p. 4]. Every node in ONTOLIS has certain **properties**, which describe it and make it easy to find the node in the entire network. ONTOLIS's approach is to put nodes in **node classes**, which means that all nodes in the same node class have the same set of properties (not the same values for these properties). [7, p. 4-6]

Furthermore, every model in ONTOLIS is based on a schema that defines what the models look like and specifies the rules for the creation of these models: Which node classes, relation classes and properties are available and how can they be combined? ONTOLIS views the schema like a "grammar for the modeling language" [8, p. 4]. There can be many models (or only one model) based on one schema, but it is also possible to have various schemas for different purposes.

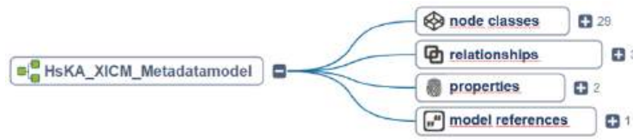


Figure 4. Schema definition in ONTOLIS

To combine node classes, relations and properties in the schema, it is necessary to define general node classes, relations and properties at first and to assign the defined relations and properties to the node classes afterwards. In addition, one can define what kind of model references there are. [8, p. 5]

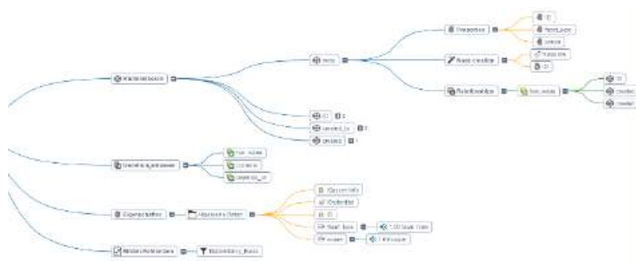


Figure 5. Combination of defined schema elements

As previously indicated in Figure 4 and Figure 5, ONTOLIS emphasizes the hierarchical structure visualized by tree structures, in order to make the semantic networks easily readable for human beings. There are different views available, which allow to see more features like non-hierarchical relations (see Figure 8). Nevertheless, every view is based on a hierarchical structure. Figure 6 shows the different views.

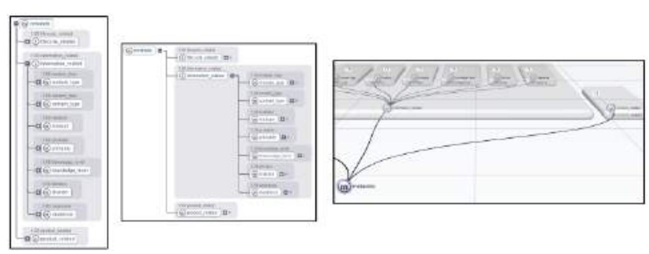


Figure 6. Different views in ONTOLIS

3.2 Modeling Metadata

Even though the software focuses on modeling products, there are many functions in ONTOLIS, which are useful for modeling metadata concepts. There are different approaches to model metadata concepts in an ontology. This paper describes an example of a possible approach using ONTOLIS on the basis of an existing metadata concept.

The metadata used in this example was developed as a concept for 3D printing during the cooperation project between UoA and HsKA. The metadata concept is described in more detail in another paper of this conference.

In consideration of single sourcing and reusability ONTOLIS provides two different possibilities to reference models in other models. One way is the **simple reuse** of models in other models; the other way is referencing models by **proxies**. The difference between these two options is, that a proxy reference only picks the root node of the referenced model; hence substructures are not included, while the reuse selects the root node and its whole substructure. The proxy reference allows changes in properties and the substructure, whereas the reuse does not allow any changes. These possibilities help to guarantee compliance with principles like single sourcing and reusability of elements for a high granularity level in the network.

The given metadata concept was implemented in ONTOLIS as follows: There is a single model for every metadata and its possible values. These single models are reused in larger-sized models, which are based on the PI classification, namely lifecycle metadata, information related metadata and product related metadata. Finally, these three models are also reused in a larger-sized model, which represents the entire metadata concept. The result is a big metadata network, which consists of three submodels classified by PI and which themselves are structured into submodels as well, the smallest unit being the single metadata with its values. Since the substructures and properties of all models should be fixed in this case, the embedding can be solved with the simple reuse technique explained above. The embedding of the models is shown in Figure 7.

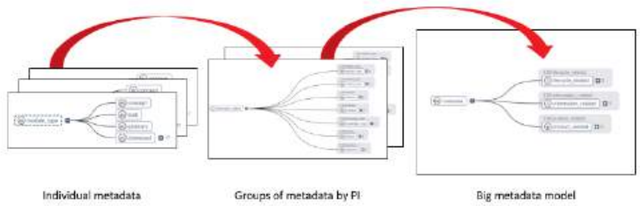


Figure 7. Embedding of models by simple reuse

This embedding gives a good overview and creates a technical connection of hierarchical relations in the metadata network. However, as already mentioned, ontologies are complex: Not every relation is hierarchical, this also applies to metadata ontologies. The selection of a specific metadata value can have the effect, that values of other metadata are no longer available – even if that data is not hierarchically related. This shows, that there is also a need to connect nodes in a non-hierarchical way. To solve this problem, ONTOLIS offers the possibility to create **cross connections** by automated relations between data. The visualization of these automated relations is not fully sophisticated yet, but Figure 8 shows an example of a cross connection and how it how it could be visualized. It is shown that the possible values for the metadata "software" change depending on which value is selected for the metadata "hardware".

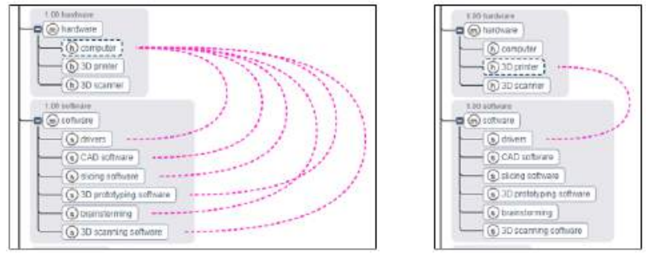


Figure 8: Cross connections in ONTOLIS

3.3 Other Applications of Ontologies in Technical Documentation

As mentioned before, ONTOLIS offers other applications in addition to modeling metadata. It is also a helpful tool to create ontologies in other contexts like product, content or terminology modeling. With all these applications, ONTOLIS tries to create a platform that combines all ontologies, which are necessary and useful for a company. ONTOLIS's goal is to provide a software that exceeds the usual CMS. That is why it combines product, content, terminology, metadata management and Content Delivery in one system.

3.3.1 Product modeling

ONTOLIS's main purpose is product modeling. Ontologies, in context of product modeling, can be used to show what versions and variants of the product exist. In addition, it helps to see how the product is built and how different parts of the product are connected and related to each other. Therefore, these product models help to share knowledge, so that everyone who works on the product, its development and the corresponding documentation has the same knowledge level. Another advantage of the ontologies is, that they are useful for automated processes, so that machines can select information from it. In case of product modeling, ONTOLIS provides possibilities to show almost every piece of information about the product like components, functions, properties or technologies due to the given flexibility when creating the schema of a model.

Figure 9. Product model in ONTOLIS [10, p. 26]

3.3.2 Content Modeling

ONTOLIS can also be used for content modeling. For this use case, the content is divided in topics (or modules) by principles of Content Management. Once the content and the product are modeled in ONTOLIS, it is possible to combine these models, e.g., by connecting a specific content object to the corresponding product component. If the metadata is also modeled like explained in section 3.2, there is another advantage: When the metadata model is applied to the content as well as to the product model, there is automatically a connection between product and content by metadata. With the metadata model it is possible to classify the product and content and also filter models by the metadata. Figure

10 shows how the modeled metadata (on the left side) can be used to classify the content (on the right side).

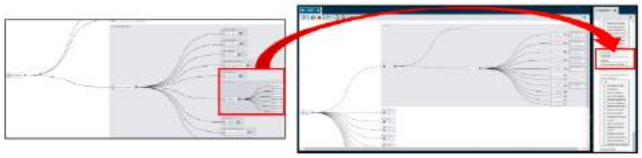


Figure 10. Application of metadata model to content model

When applying a metadata model to other models, ONTOLIS allows creating **dependency rules** for the process of assigning metadata to nodes. Dependency rules create the possibility to support users who assign metadata to nodes by hiding values, which are not valid, when certain values of other metadata are chosen. In the case of the example in Figure 8, the dependency rule would hide the values "CAD software", "slicing software", "3D prototyping software", "brainstorming" and "3D scanning software", if the user selected "3D printer" as the hardware, since these types of software cannot be used on a 3D printer.

The classification of content (or products) by metadata helps to filter the models, e.g., to see variants. ONTOLIS has a feature to create **validity rules**, which allow filtering the model by these validity rules. These rules specify which property values have to be chosen. If a model is checked by a rule, ONTOLIS hides the nodes, which are not compliant with the rule.

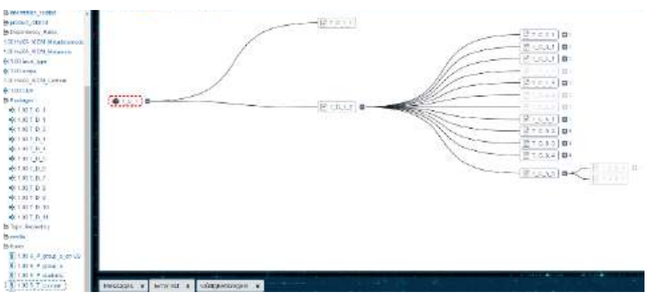


Figure 11. Filtering of models by validity rules

One validity rule in ONTOLIS can combine as many values as needed by operators like AND or OR and there are many other functions for specifying validity rules.

3.3.3 Terminology modeling

Another application of ONTOLIS is terminology management, which gives a knowledge overview by conceptual systems in consideration of (multilingual) terminological principles. The conceptual systems are ontologies built from proxies (one per term). Each proxy is an individual model for each term to which the user can navigate to from the entire conceptual system. When navigating to such a proxy, there are many functions and properties that allow for a professional terminology management.

Figure 12. Conceptual system in ONTOLIS [12]

Figure 13. Terminology management in ONTOLIS [12]

3.4 Advantages and Disadvantages of ONTOLIS

ONTOLIS offers a many more useful functions that a usual CMS cannot offer. Due to user-defined schemas, there is a high flexibility and every user can create models tailored to his own needs and for different purposes. For companies with an advanced Technical Documentation it is a good opportunity to combine ontologies for every information model inside the company. This allows creating an intelligent information network. Additionally, the tree-structure makes the entire network easy to read for human beings, which can be motivating for those working with it.

What could be improved is the delivery aspect of ONTOLIS. At the moment, the Content Delivery in ONTOLIS is not sufficient, especially because it's impossible to connect ONTOLIS to another Content Delivery Portal. A visualized preview of xml or html files, which can be uploaded to the node, that represents a content module, would be helpful. In this way, there would be a node that not only represents the content but also includes the content. An even better option would be a feature that allows creating and editing the xml content directly in ONTOLIS and getting the desired preview as well. ONTOLIS already offers a context box for creating and editing xml content, but this is not a sufficient editor and additionally, the delivery part is not implemented yet. At this point, there are helpful models and ontologies, which could be used for proper Content Delivery, but the interface for the Content Delivery is not satisfactory.

Moreover, complex relations and dependencies between nodes, which are not hierarchical, are very complex to create. When we think of the reason why ontologies are necessary for intelligent information, namely that hierarchical taxonomies are not sufficient to create whole data models, it is a great disadvantage that these complex dependencies are difficult to create and especially to display, although the visualization aspect is strongly emphasized by ONTOLIS.

ONTOLIS is still working on these aspects, especially to improve the Content Delivery aspect as well as the inclusion of xml data and is eager to help its customers to find a solution for every need. This includes user specific customizations, e.g., for the visualization of complex dependencies and relations or a preview of the content. At the same time, this means that the user cannot create these things on his own and has to rely on the help of the manufacturer.

4. DISCUSSION AND CONCLUSION

This paper gave an overview of basic aspects and principles of ontology modeling and its applications in Technical Communication. Using metadata and classifying content accordingly has become inevitable in professional Content Management. Making use of the advantages, which ontologies offer, is a newer development and not fully matured yet. Although there are different software solutions, that support the implementation of ontologies in CMS and CDP, they still have not reached their full potential yet. One of the existing software solutions to create ontologies in the field of Technical Communication is ONTOLIS, which was presented in detail in section 3. With the combination of tree-like structures, referencing possibilities, cross connections and dependencies, it is possible to create a complex ontology that is still comprehensible for human beings. The metadata concept, which was developed during the UoA and HsKA cooperation project, could be modeled as ontology inside the ONTOLIS software. In addition to that, the content itself could be modeled in ONTOLIS and classified by combining the content and metadata model. This helps users to manage and filter their content in a way that could be compared to the functionality of a CMS. Content preview, Content Delivery and some functions for precise ontology modeling are aspects, which need to be improved in the future and could not be used sufficiently for this project. That shows that the software is still relatively new and that there is more untapped potential. Keeping this in mind, companies should evaluate whether they need an ontology and if the additional expense is within reasonable limits. For larger companies, which manage a considerable amount of information and aim for an automated Content Management and Delivery process, the advantages will most likely outweigh the effort of modeling and maintaining an ontology. Another interesting aspect are AI mechanisms to classify content automatically or to make search results more significant and dynamic. If these efforts to combine ontologies and AI are successfully developed, it can change the way users access Technical Documentation in the future.

5. ACKNOWLEDGEMENTS

We would like to thank the International Office, AAA of HsKA, University of Aizu (UoA), Faculty IMM of HsKA and the Master Program KMM of HsKA, which funded the cooperation project between the University of Aizu and the Karlsruhe University of Applied Science. Thanks to this funding, we were able to participate in the project and gain important experiences.

6. REFERENCES

[1] Arroyo, Sinuhé and Siorpaes, Katharina. 2014. Ontologies and Ontology Languages. In Sicilia, Miguel-Angel (Ed.). Handbook of Metadata, Semantics and Ontologies. 2014. World Scientific, Singapore.

[2] Bürger, Tobias and Hausenblas, Michael. 2014. Metadata Standards and Ontologies for Multimedia Content. In Sicilia, Miguel-Angel (Ed.). Handbook of Metadata, Semantics and Ontologies. 2014. World Scientific, Singapore.

[3] Drewer, Petra and Ziegler, Wolfgang. 2014. Technische Dokumentation: Übersetzungsgerechte Texterstellung und Content-Management. 2nd Edition. Vogel, Würzburg, Germany.

[4] Gruber, Thomas R. 1993. Toward Principles for the Design of Ontologies Used for Knowledge Sharing. In N. Guarino, N. and Poli, R. (Ed.), Formal Ontology in Conceptual Analysis and Knowledge Representation. Kluwer Academic Publishers, Deventer, The Netherlands, 1993. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.89.5775&rep=rep1&type=pdf

[5] Mendez, Eva and van Hooland, Seth. 2014. Metadata Typology and Metadata Uses. In Sicilia, Miguel-Angel (Ed.). Handbook of Metadata, Semantics and Ontologies. 2014. World Scientific, Singapore.

[6] Noy, Natalya F. and McGuinness, Deborah L. (n. d). Ontology Development 101: A Guide to Creating Your First Ontology. Stanford University, Stanford, CA, 94305. https://protege.stanford.edu/publications/ontology_development/ontology101-noy-mcguinness.html

[7] ONTOLIS GmbH. 2018. ONTOLIS-Schulung. Einführung in die Produktmodellierung. Training Document.

[8] ONTOLIS GmbH. 2018. ONTOLIS-Schulung. Schema-Funktionen. Training Document.

[9] ONTOLIS GmbH. 2018. ONTOLIS-Schulung. Vorstellung ONTOLIS. Training Document.

[10] ONTOLIS GmbH. 2018. Ontologie-Modellierung in ONTOLIS. Training Document.

[11] Studer, Rudim, Benjamins, V. Richard, and Fensel, Dieter. 1997. Knowledge Engineering: Principles and methods. Data Knowl. Eng. 25 (1-2), 161–197 (1998). http://dx.doi.org/10.1016/S0169-023X(97)00056-6

[12] Weinmann, Thomas. ONTOLIS GmbH. 2018. AW: AW: Zitieren der ONTOLIS Unterlagen. E-Mail to Lara Krägel.

[13] Ziegler, Wolfgang. 2017. Intelligente Informationen. Schriften zur Technischen Kommunikation,Volume 22, p. 51-66, Hennig J. and Tjarks-Sobhani M. (Ed.) (2017), tcworld GmbH. https://www.i4icm.de/fileadmin/user_upload/TK22_Metadaten_fuer_intelligenten_content_Ziegler.pdf

THE EVOLUTION OF METADATA CONCEPTS FROM METADATA HARVESTING TO ONTOLOGY MODELING

2019 Volume 1

Alessa Fleischer
Karlsruhe University of Applied Sciences
Moltkestraße 30
76133 Karlsruhe
Germany
flal1015@hs-karlsruhe.de
alessafleischer@web.de

Julien Combelles
Karlsruhe University of Applied Sciences
Moltkestraße 30
76133 Karlsruhe
Germany
coju1013@hs-karlsruhe.de
julien.combelles@web.de

ISSN: 2434-9526

PAGES: 18-23

ACM Chapter Proceedings on Educational Technology, Language & Technical Communication

The Evolution of Metadata Concepts From Metadata Harvesting to Ontology Modeling

Alessa Fleischer
Karlsruhe University of Applied Sciences
Moltkestraße 30
76133 Karlsruhe
Germany
flal1015@hs-karlsruhe.de
alessafleischer@web.de

Julien Combelles
Karlsruhe University of Applied Sciences
Moltkestraße 30
76133 Karlsruhe
Germany
coju1013@hs-karlsruhe.de
julien.combelles@web.de

ABSTRACT

Content Management and Content Delivery, like many other areas in today's life, are heavily reliant on metadata. In order to ensure the long-term usefulness of metadata concepts and to further develop them into ontologies, it is essential to define the metadata clearly in the first place. This requires careful research and planning. This paper will focus on different methods of collecting and visualizing metadata in order to develop a stable and reliable metadata concept which can be used in working with ontologies. The paper shows the entire process of the development at the Karlsruhe University of Applied Sciences in cooperation with the University of Aizu. In the project many different assignments from the students of Aizu were transformed into XML-topics and enriched them with metadata. Afterwards they were uploaded into a Content Delivery Portal where they were accessible through facets and filters.

The first part of the paper covers a theoretical overview about the importance of metadata as well as the concept of ontologies. Afterwards, the focus lies on the process recommended by us to create a metadata concept including critical evaluations of the steps. The final step is the visualization of metadata through an ontology and the possibility to publish ontologies by means of Content Delivery Portals. A final conclusion is given to summarize the findings.

Figure 1. Process of the project

CCS Concepts

- **Information Systems → Document metadata**
- **Information Systems → Novelty in Information Retrieval**
- **Information Systems → Ontologies**

ACM Chapter Research Seminar on Technical Communication & Information Design 2019, January 28-Feb 1, 2019, Aizuwakamatsu, Japan

ISSN: 2434-9526

Keywords
XML; content management; metadata; visualization; content delivery; ontologies; intelligent information.

1. THEORETICAL OVERVIEW

Although widely known within technical communication, this paper will give a short introduction on the usage of metadata, its relevance in general and especially for the academic project of HSKA and UoA in context of ontology usage and content delivery portals. Also, the approaches to gain and collect metadata in the process will be briefly compared and a short overview about ontologies and their meaning in technical communication will be given. Modern products are becoming more and more complex and customers are demanding individually customized products. When buying a car for example, people can choose different options from the rims to the steering wheel. In the sector of machinery and plant engineering, in many cases, every single product is different. Therefore, all documentation considering this product has to be different as well, resulting in a vast number of topics and information parts. To manage this information metadata, data about this data, is needed.

1.1 Importance of metadata

Metadata is essential for using and providing information in an efficient and reliable way e.g. in Content Management Systems or Content Delivery Portals. "Metadata can enrich modular or document-based content via additional information. This is used for the retrieval and reuse processes within Content Management Systems or subsequently outside of Content Management Systems within the delivery processes." [1]. Usually metadata can be divided into product-related and information-related metadata. The product-related metadata gives information about the relation of the topic to the product, e.g. for which product the topic is needed and in which part of the product. The information-related metadata gives information about the topic itself e.g. whom it was created by and what sort of information (task, explanation, safety-advice etc.) it is dealing with. Ziegler developed the PI-Class-Method and states that in "more detail, the PI classification method defines semantic metadata according to four basic classes building up the dimensions of an information space. These classes are organized with respect to products (P) and to information (I):" The four classes are intrinsic and extrinsic product classes and intrinsic and extrinsic information classes. [2]

1.2 Generating metadata

To generate metadata there are two different approaches: inductive (Top-Down) or deductive (Bottom-up). The Bottom-up-approach is used, when there is already content and information which can be used to derive metadata from. This means modeling the metadata according to the existing content. The opposite is the Top-Down-approach. Using the Top-Down approach, the metadata concept is already predefined e.g. by an industrial standard such as “S1000D for military and aviation products”. A Top-Down-approach is also necessary when there is no existing content to derive metadata from [3].

In practice both approaches are used together most times, as there is often some form of content but not enough to model a reliable metadata concept. Sometimes there are also changes and new requirements that make an optimization of the concept necessary.

1.3 Ontologies

The concept of Ontologies might not be well known even though its importance has been acknowledged around the world for a long time, starting with its important role in Tim Berners Lee’s concept of the Semantic Web in 2001 [4].

In the field of computing and knowledge ontologies can be described as “an explicit specification of a conceptualization” [5]. Specification means “a formal and declarative representation” already implying that “an ontology should be machine readable” [6]. The term conceptualization describes an “abstract, simplified view of the world” which “is based on concepts, objects, and other entities […] and the relationships that exist among them” [6]. These definitions underline the most important criteria for ontologies with focus on “Clarity”, “Coherence” and “Extendibility” [5] as well as the fact that ontologies capture “consensual knowledge” not the “subjective knowledge” of individuals [6].

It is vital to understand that an ontology might be “used to structure datasets and metadata, but […] is […] not a piece of metadata itself.” [8] There is also no single consensual ontology engineering methodology [7] as well as no standard language to document it – OWL and RDF are two of the most prominent modeling languages [6]. Compared to other methods of capturing information, an ontology is defined by its formal structure as well as its semantical understandable structure, as it is shown in Figure 1.

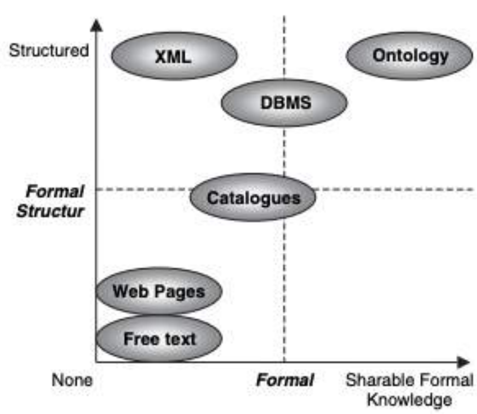


Figure 2. Structure of different methods to manage information in comparison. [8]

Enabling machines to understand the semantics of data underlines the great advantage of ontologies. Even though humans might understand the correlation between sets of data “the […] facts would remain unknown to a typical DBMS and indeed to any other information system, for which the model of the world is limited to data structures […] with no automatically interpretable semantics.” [8] The visual presentation of ontologies, of course, is helpful to human beings too; knowledge networks, for example, show the connections in a net and are explained in more detail in chapter 3.2.

In the field of Technical Communication ontologies are a relatively new and their usage is at a very early stage. But as the demands for easy-to-use and intuitive CDPs thrive, the need for a way to document and display metadata and the content in more manageable ways while still keeping track of enormous amounts of data network becomes more and more visible. More detailed documentation and information on the subject of Ontologies can be found for example in Gaševic, Djuric and Devedžic [6] or Davies [8].

2. DEVELOPMENT OF A METADATA CONCEPT ALONGSIDE A REAL PROJECT

This chapter explains ways and possibilities to create a metadata concept in reference to a project, by using an academic project as an example. The project was realized in 2018 and 2019 in cooperation with the University of Aizu, Japan and the Karlsruhe University of Applied Sciences, Germany with the goals of Content Management and Content Delivery on the subject of 3D printing.

During the project we received several assignments on 3D printing from students of the University of Aizu, which we gathered, split and transformed into XML-tagged topics for the usage in Content Management Systems or Content Delivery Portals. To use these topics in a Content Delivery Portal, we had to enrich them with the relevant metadata first.

As the project was started ahead of the Aizu students’ lectures, we had to use the Top-Down approach to generate metadata because there was no content to work with from the beginning. Later on, Bottom-Up methods were integrated as well to optimize the Top-Down concept. The project can be split into three different parts, which can be seen as a useful model for developing a metadata concept with the goal of manual Content Delivery.

Table 1. Relevant steps for creating a metadata concept

<table>
<tr><td>Step 1</td><td>research and documentation of possible relevant metadata</td><td rowspan="3">Continuous refinement process: make relevant changes</td></tr>
<tr><td>Step 2</td><td>narrow down to relevant data and group it (lifecycle, product, information)</td></tr>
<tr><td>Step 3</td><td>develop ID-concept and document metadata in XML</td></tr>
</table>

The following subchapters will give a detailed overview about each individual step.

2.1 Research and documentation

In order to get a first impression of the possible metadata, we researched the subject ‘3D Printing’ and made up a broad concept of various metadata, because there was no clear main focus of the content at this point.

A good way to obtain an overview about the different types of metadata is to create a mind-map which can help to document and structure all possible metadata after the first research and brainstorming phase. There are different tools on the market for displaying mind-maps like wiseMapping, lucid chart and Mindjet Mindmanager. In our project the tool Mindmanager was used which can be seen in Figure 3.

In the mind-map the metadata was already distinguished by product-related and information-related to gain a better overview about the subject and to find out were certain values might be missing. We used the first hierarchy level to collect all relevant keywords and and second and third level to write down potential values. In Mindmanager connections showing dependencies of hierarchical values can be made. In our case for example certain printing processes were connected to special types of printing material (because a printer printing with plastic based process will only be able to use plastic as a material and will not be able to print with metal). When used in Content Management it is very useful if the Management System knows about these connections an only suggests the correct metadata to the user. In order to make this step easier later on, we connected the values with orange arrows, which is visualized in Figure 3.

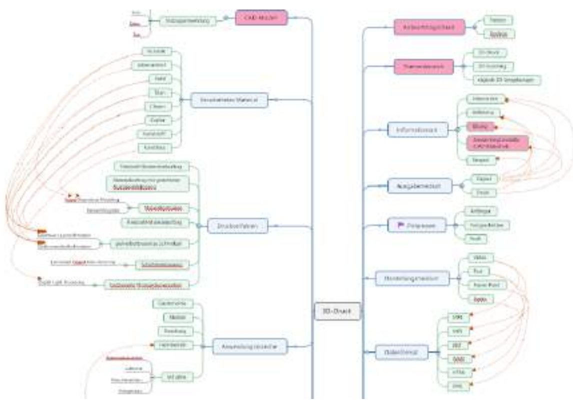


Figure 3. Dependent metadata values visualized in a mind-map.

After the first round of documentation, the final result was a broad mind-map containing a lot of possible metadata and potential values.

2.2 Narrowing down

Once there is a satisfying overview about the subject and all possible related metadata, the next step is to narrow down the metadata to the relevant values. This step included more research about the data to find potential duplicates and irrelevant information and it also included Bottom-up work as described in chapter 2.4

Another important fact is that the purpose of the mind-map was to collect various sorts of metadata and to show dependencies but there was not much information about the individual metadata itself, e.g. whether it is single- or multivalued or whether its usage in the topic is required. The graphical approach to document this 'meta-metadata' was not sufficient, which is why a tabular method was used. One well known and comprehensive tool to create tables is Microsoft Excel.

Three different sheets were used to document ID and live-cycle, information-related and product-related metadata. The sheets contained columns with the exact term of the metadata, the data-type such as "integer", "decimal" or "string", all possible values, whether the value will be created automatic e.g. in the usage of a Content Management Systems, if the value is required, generic and if it's single- or multivalued.

The table was very helpful during the creation of the XML structure and the development of the ID concept as described in the following chapter.

2.3 Developing and structuring

Developing ID's for each metadata value and documenting the entire metadata concept in a structured XML-format was a crucial step in integrating the metadata into our workflows and making it readable for the Content Management and Delivery systems.

In order to be addressed by the systems, each metadata value needed its own unique identifier. To make the ID's more understandable for people working with the metadata we focused on semantic ID's showing the hierarchical position of the metadata value. The value for the language 'English', for example, had the ID **m_lan01** while the value for the locale 'American English' (which was subordinated to the language) had the ID **m_lan01_loc01**. While this process was very helpful in our case, we also had to recognize its downsides since changes in the hierarchy of the metadata values resulted in a completely new ID-structure which had to be re-implemented in a very time-consuming process. The entire ID structure and its semantical background was documented in a text editor in order to pass on the knowledge to others and to have a strong theoretical basis before continuing on to the next steps.

After the IDs were set, the entire metadata concept was transferred into an XML document. This allowed us to combine all the information collected in the last step – the values, their IDs and the relevant information about metadata ('meta-metadata') mentioned in the previous subchapter. The structured form was machine-readable and could therefore be uploaded into the Content Delivery portal to be used as filters and facets.

```
<!--Language and locale-->
<meta class="language" value="language" id="m_lan" scope="all" facet_type="extrinsic">
    <meta class="language" value="en" id="m_lan01" >

    <!--subclass of language="en" is locale-->
        <meta class="language" value="en-AU" id="m_lan01_loc01"></meta>
        <meta class="language" value="en-CA" id="m_lan01_loc02"></meta>
        <meta class="language" value="en-GB" id="m_lan01_loc03"></meta>
        <meta class="language" value="en-IN" id="m_lan01_loc04"></meta>
        <meta class="language" value="en-US" id="m_lan01_loc05"></meta>
    </meta>

    <!--subclass of language="de" is locale-->
    <meta class="language" value="de" id="m_lan02">
            <meta class="language" value="de-DE" id="m_lan02_loc06"></meta>
            <meta class="language" value="de-CH" id="m_lan02_loc07"></meta>
            <meta class="language" value="de-AT" id="m_lan02_loc08"></meta>
    </meta>

    <meta class="language" value="ja" id="m_lan03"></meta>
</meta>
```

Figure 4. Metadata and its additional information documented in XML

2.4 Continuous refinement process

Throughout the entire development described in this chapter a continuous refinement process was important to keep track of relevant changes and to make the resulting adjustments. With every new assignment that we received from the students in Aizu we found missing as well as irrelevant parts in our metadata and made the corresponding changes. This is also likely to happen in a commercial environment, as product ranges de- or increase and new requirements for products or the documentation itself can occur.

This is where the Bottom-up approach mentioned earlier becomes relevant. It underlines the importance of never choosing one option alone but combing both to get a full overview on the topic and shows why a reasonable amount of time should be spent in on the development of the metadata.

3. REALIZATION OF METADATA CONCEPTS THROUGH ONTOLOGIES IN A BUSINESS APPLICATION

Besides the already mentioned ways and tools to develop, document and implement a metadata concept, there is an additional form of working with metadata, which is gaining more and more attention: Ontologies. As chapter 1.3 already gave a theoretical overview, this chapter will focus on the practical implementation of a metadata concept through ontologies and the associated benefits and limitations.

There are several providers with different approaches, which are specifically focused on creating, managing and using ontologies in the context of Technical Communication. This chapter focusses on the company i-views and their tool "Smart Data Engine" with the Knowledge-Builder and the related Content Delivery Portal "i-viewscontent".

3.1 Introduction in i-views

The company intelligent views gmbh (i-views) was founded in 1997 as a spin-off of the Fraunhofer Society and is located in Darmstadt, Germany [9]. The i-views tool Knowledge-Builder was used to create a knowledge network to display our metadata concept. With the help of i-views we also managed to transform our content, which was tagged with our metadata, into the Content Delivery Portal "i-viewscontent".

The "Smart Data Engine" is, as of today, the only ontology-modeling tool that supports their own Content Delivery Portal (further information on this topic can be found in chapter 3.3) and allows a direct connection between the ontology and the content displayed in the Content Delivery Portal. This makes i-views interesting for the challenges of Technical Communication.

3.2 Creation of a knowledge network

Before further parts of i-views are shown, the internal terminology of i-views has to be explained, as the general terminology concerning ontologies varies widely. The different entities in the Knowledge-Builder are called types and objects. The types classify the superior categories and the objects are parts of these categories. The type "city" e.g. can have the objects "Frankfurt", "New York" and "Paris". Types and objects can be connected with defined relationships. The objects can also have different attributes which can include e.g. passwords, files or colors and can be defined by the user. Figure 5 visualizes the general structure in the Knowledge-Builder.

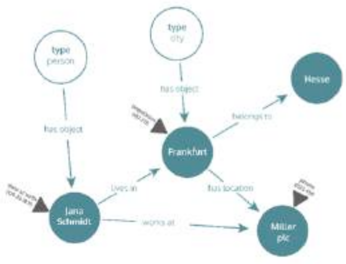
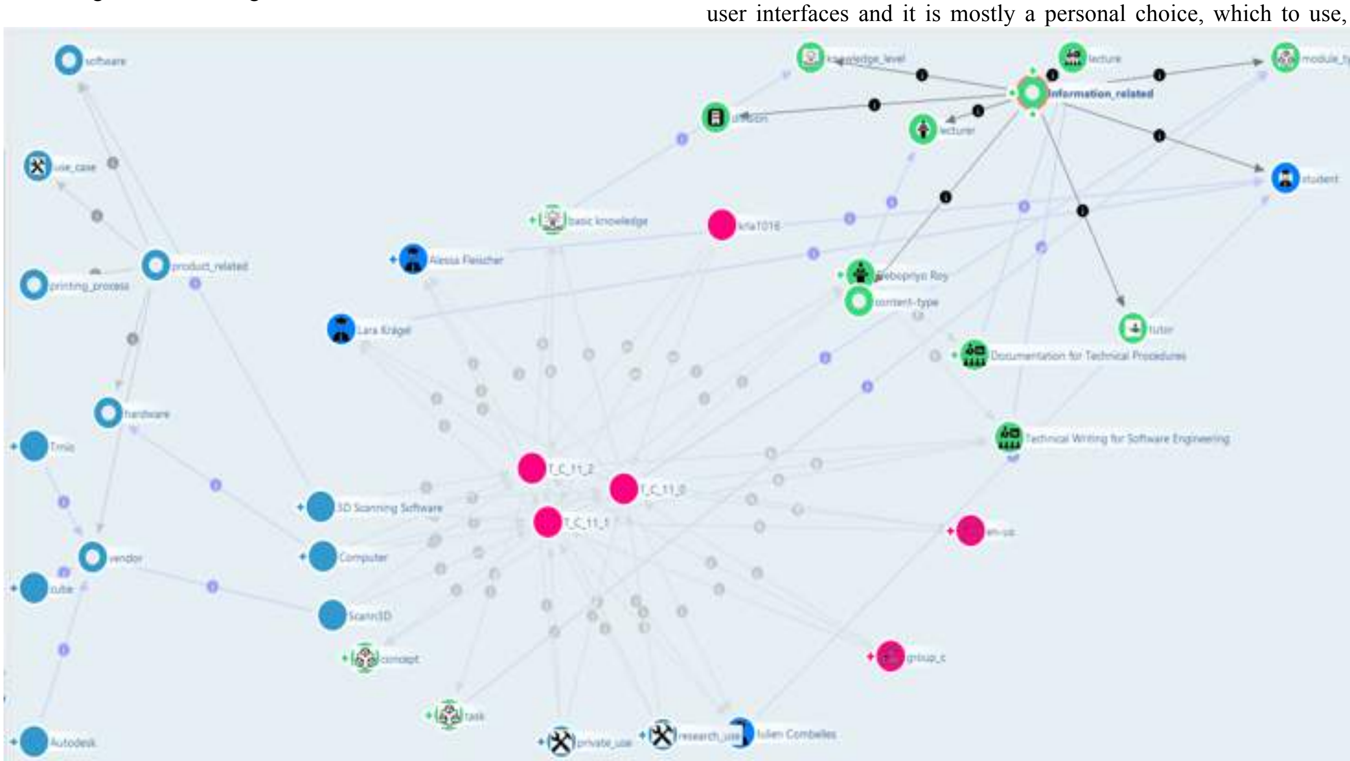


Figure 5. Entities and their relations in the Knowledge-Builder [10]

The objects and types can be created within the tool or can be imported from a csv-file or a Microsoft Excel table. The hierarchy plays an important role: subtypes of types can be created so that a multilevel metadata concept can be realized. The lowest hierarchy level is always an object.

The relationships are defined centrally in the tool for all objects. This can be confusing at first, depending on how many relationships have been created, but this ensures that no relationship is created twice. In the relationship it can be defined which objects can be connected through a relation as well as the name of the relationship and whether it is a reversed or a symmetrical relationship.

With the Knowledge-Builder it is possible to work in two different user interfaces: the textual user interface and the graphical user interface. Most tasks can be accomplished in both user interfaces and it is mostly a personal choice, which to use,



although one user interface supports certain aspects better than the other.

By creating all necessary types and objects and connecting them to each other, it is possible to build up the metadata concept as an ontology step by step. So far, the user is able to use the ontology as a knowledge network to find information about the subject. But there is a final step missing to use the network as basis for a Content Delivery Portal: adding the content in form of topics into the network.

We created a special type named “Topic-ID” with objects named after the different topic-IDs, which had been determined in the ID concept (as explained in chapter 2.3). Afterwards the HTML-files containing the topics were added as attributes and the resulting connections to the objects in the Knowledge-Builder according to the metadata structure of the topic were made. With this done, the knowledge network was ready to be uploaded in the Content Delivery Portal. The Content Delivery Portal will be the subject of the next chapter.

It can be summarized, that after a short training by i-views the Knowledge-Builder is an easy-to-use tool. The visualization of the metadata concept as an ontology makes it less complex and more lucid than in text form and the direct link to the Content Delivery Portal is very useful. The graphical visualization can become complex when a lot of connections are shown but the tool offers the possibility to enable and disable certain types and objects to make desired relations clearer.

On the downside there are several limitations concerning the work with the Knowledge-Builder. The processing of big amounts of topics and metadata is time-consuming and inefficient. There is no native batch-processing integrated, so if needed, it has to be implemented additionally through Java Script. If a new metadata occurs e.g. and all topics need to receive this data, the metadata needs to be drawn to every single topic separately if no additional script is implemented which often has to be done by a professional.

Furthermore, we perceived the strong distinction between types and objects as obstructive, especially with deeply structured hierarchies and the advantages were not conclusive. Sometimes we had the problem that an entity that was created and designed as an object X of the type Y, had subobjects, which made it necessary to transform it into a type. But now the new type was on the wrong hierarchical level as it should still be a subtype of Y and not on the same level as Y.

Another difficulty that resulted from the strong distinction of types and objects was the different handling of relations between types and objects as the Knowledge-Builder is mainly focused on relations between objects.

3.3 Content Delivery with i-views

One of the major advantages of the Knowledge-Builder from i-views is the direct link to the Content Delivery Portal i-viewscontent. As we worked with a tryout-version of the Knowledge-Builder, help was needed from i-views to transform our knowledge network into an active version of the Knowledge-Builder. Once this was done and we had access to the active version, we could make changes to adapt the network to the requirements of the Content Delivery Portal itself.

The portal supports the differentiation of the product-related and the information-related content. The retrieval of the topics can be done through the search engine or via filters and facets efficiently.

As the CDP called Content Delivery Suite by the company SCHEMA was mainly used in the project, we had a small issue

Figure 7. Information- and product-related filters on “i-viewscontent” start page

with the HTML-files that needed to be imported into i-viewscontent. i-viewscontent supports generic HTML, but since our proprietary processing-script transformed the XML topics into CDP-specific HTML, modifications had to be made to our HTML to display it in the i-viewscontent.

After making small modifications, the Delivery Portal supported the content to a satisfying extent and the topics could be found through the information- and product-related filters as well as the search function. Thus we were able to give the project partners in Japan a good impression of how their researches were realized in an easy-to-access medium like a Content Delivery Portal.

4. CONCLUSION

The process of creating a suitable metadata concept from gathering all information in a mind-map to displaying it in a sophisticated way like an ontology is long and enduring. But working with large amounts of content without using proper metadata concepts is even more time consuming and troubling.

When creating a metadata concept, it is important to take time and work accurate. As described in this paper small steps should be taken rather than trying to fit the whole process into one day of work with no possibility to do any refinements.

The metadata concept has to be planned carefully and all perspectives must be taken into account, as bigger changes in the end will most likely be very time consuming. However, this does not mean that the metadata is written in stone and smaller changes and optimizations should and must be done continuously.

Currently, there is no tool to realize the whole process from harvesting metadata to displaying the concept efficiently, although tools like the Knowledge-Builder by i-views are a good approach for the final visualization. Until then, we would recommend using the individual steps with different tools to develop a successful metadata concept.

After developing the metadata concept, ontologies can be the final step of visualization. Although they have a high potential, it should always be considered whether an ontology is the right form for the certain use case. For a classical user documentation, the existing Content Management Systems and publication channels are more suitable than ontology-based systems. Nevertheless, being able to connect the ontology to a Content Delivery Portal to omit the classical Content Management System and publish the information directly from the ontology can be a chance in the future though so far ontology-based system are not optimized to create the content in XML or HTML format themselves.

What makes an ontology-based system stand out is the possibility to react to the context the user is in and to present him the correct information as well as related information. One example is the i-views and BMW cooperation, which aims to develop a web portal in which the costumer can enter personalized information about himself and get matched to the most suitable car through the ontology-based system. Every car is connected to certain entities like “outdoor fan” or “family person” and the car with the most matches is proposed.

Another use case for ontology-based system is the maintenance and assembly of highly complex and unique plants and machinery, where regular documentation is too extensive for being managed by a technician. Ontologies offer a possibility to present the right information at the right time, which is important as periods of inactivity can be extremely costly for these machines.

Ontologies can also be used in context of automated tagging through artificial intelligence. The idea is that the knowledge of the ontology can be used to identify special types of content and their subjects and the tagging software can determine all classifying metadata and add them automatically. It is important to highlight the fact that the ontology can’t be created by the AI itself, it can only be used to tag new content. The ontology itself has to be created manually because no artificial intelligence can see the relations between parts of the content so far – though no one knows what future innovations will be made in the field of Content Management and Content Delivery.

5. ACKNOWLEDGMENTS

At first, we would like to thank all the institutions which enabled us through their generous financial support to make the journey to Japan possible and realize this project: we thank the International Office, AAA of HSKA, the University of Aizu (UoA), the faculty IMM of HSKA and the masters-program KMM of HSKA.

We would especially like to thank Prof. Dr. Ziegler of the Karlsruhe University of Applied Sciences, Germany and Prof. Dr. Roy of the University of Aizu, Japan for guiding us through this whole project and for all the work that went into the organization of this exchange and the conference.

We would also like to thank Dr. Steinacker from i-views who supported us with all our questions concerning the i-views tools and who helped us to transform our content into the Content Delivery Portal.

We also thank Mrs. Berumen-Roden of HSKA for proofreading this paper.

And finally, a special thanks to the students of the University of Aizu for providing us with all the content about 3D printing and supporting us throughout the seminars and workshops.

6. REFERENCES

[1] Ziegler, W. 2017. Basic Concepts Which Support the Management and Delivery of Intelligent Content. In *Frontier, Official Journal of Japan Technical Communicators Association JTCA*, 89.

[2] Parson, U., Sapara, J. and Ziegler, W. 2018. iiRDS for Technical Writers - Introduction to the Metadata. In *Tagungsband zur tekom Jahrestagung*, 225.

[3] Drewer, P and Ziegler, W. 2010. *Technische Dokumentation*. Vogel Communications Group, Würzburg. 373

[4] Berners-Lee, T., Hendler J. and Lassila, O. 2001. The Semantic Web. In *Scientific American* (NYC, Nature Publishing Group, Volume. 5). DOI= 10.1038/scientificamerican0501-34

[5] Gruber, T. 1995. Toward Principles for the Design of Ontologies Used for Knowledge Sharing. *In International Journal of Human-Computer Studies* (Orlando, Academic Press, Volume 43, Issue 5-6), 907-928.

[6] Gaševic, D., Djuric, D. and Devedžic, V. 2009. *Model Driven Engineering and Ontology Development*. Springer Verlab, Berlin, Heidelberg. 28,46,62,244.

[7] Forbes, D., Wongthongtham, P., Terblanche, C. and Pakdeetrakulwong, U. (2018): Ontology engineering applications in healthcare and workforce management systems. In *Studies in Systems, Decision and Control* (Cham, Springer, Volume 123). 27.

[8] Davies, J., Studer, R. and Warren, P. 2006: *Semantic Web technologies. Trends and research in ontology-based systems*. Wiley, Chichester.117-122.

[9] *About intelligent views gmbh*. 2018. Retrieved February 17, 2019 from https://i-views.com/en/about-us/

[10] *Tryout Tutorial – The knowledge network in the Knowledge-Builder*. 2018. Retrieved February 17, 2019 from http://documentation.i-views.com/manual/misc/dffe3fce-c967-4528-a7d1-62e4d1fc32e4#

XML CONTENT TRANSFORMATION AND IMPORT INTERFACES FOR CONTENT DELIVERY PORTALS

2019 Volume 1

Karla-Brigitte Hanczaryk
Karlsruhe University of Applied Sciences 76133 Karlsruhe
Germany
haka1018@hs-karlsruhe.de

ISSN: 2434-9526

PAGES: 25-34

ACM Chapter Proceedings on Educational Technology, Language & Technical Communication

XML Content Transformation and Import Interfaces for Content Delivery Portals

Karla-Brigitte Hanczaryk
Karlsruhe University of Applied Sciences
76133 Karlsruhe
Germany
haka1018@hs-karlsruhe.de

ABSTRACT
Modern processes and technologies in Technical Communication aim at the provision of information to users of industrial and consumer products, which is adapted to their situational needs and their actual knowledge level [2, 3]. For this purpose, retrieval mechanisms are used by Content Delivery Portals (CDP) to dynamically deliver content to users [3].

The research topic of this paper is to convey possibilities for importing the content into the CDP, depending on whether a homogeneous or a heterogeneous system landscape is given, and provide an in-depth description of a self-developed import process with programmed transformations. In this paper, we describe the main functionalities of a selected CDP and highlight the requirements to integrate input data into the CDP focusing on retrieval mechanisms. In this context, we elaborate on different requirements for input data and functionalities of other CDPs from various vendors. Furthermore, we introduce an option for automating the generation of content packages.

The goal of this paper is therefore to present an alternative process of delivering content to a CDP without the usage of a content management system (CMS)[1]. We discuss advantages and disadvantages of self-developed data import. Additionally, we will refer to the upcoming standard iiRDS, which has the objective to standardize the interchange format for data imported into CDPs.

CCS Concepts
• Information systems→ XML query languages; Information integration; Resource Description Framework (RDF); Extensible Markup Language (XML); Hypertext languages; Document filtering; Ontologies; Document structure; Document topic models; Users and interactive retrieval; Distributed retrieval; Retrieval on mobile devices; Structured text search; Multimedia and multimodal retrieval.

ACM Chapter Research Seminar on Technical Communication & Information Design 2019, January 28-Feb 1, 2019, Aizuwakamatsu, Japan

ISSN: 2434-9526

[1] In other literature and commercial applications content management systems are also called component content management systems (CCMS).

Keywords
Technical Documentation; Content Management; Content Delivery; Content Transformation; Content Retrieval; Extensible Stylesheet Language Transformation (XSLT); Content Classification; Batch Processing; Intelligent Information; Semantic Web; Ontology; Resource Description Framework (RDF); iiRDS

1. INTRODUCTION
Users of industrial and consumer products have a tendency to avoid reading classical and lengthy paper-based documentation [1]. Technical documentation consisting of hundreds of pages are demotivating for users because it requires enormous expenses to find the information, which is the answer to the question of the user in the current situation. To receive the necessary information, many users submit search queries to search engines. Another popular search method based on filtering is widely used in websites, for example of the successful online retailer Amazon. Customers find the information needed from the enormous data inventory of Amazon by selecting appropriate predefined filter criteria. “Companies are increasingly discovering the value of well-structured and addressable content for supporting a variety of content-driven scenarios mainly for after-sales information processes” [1].

“The definition of CDP is given technically as systems, offering web based access to modular, aggregated content or other information for various user groups by related retrieval mechanisms.” [3]. By providing facetted search, navigation by structure and direct search, CDPs offer access to this content. These search strategies in CDPs are described extensively in [3].

The CDP distributes content, which originates from different management systems like document management systems (DMS), product information management (PIM) or mainly from XML-based CMS. The CMS supports the creation, management, modular reuse and publication of content for cross media usage in different media formats [4, p.289].

In practice, many companies of industrial and consumer products are using a CDP of any vendor without having a management system, which has a direct interface for importing their content to the CDP. The configuration of the CDP import interface by the CDP vendor is expensive in this case. The alternative is to create a self-developed import process, which requires programmed transformations.

Within the cooperation project between the Karlsruhe University of Applied Sciences and the University of Aizu, we have simulated the situation and challenges of companies, which have to import their content into their CDP self-dependently. We focused on these alternative possibilities of content import.

We analyzed the CDP called SCHEMA Portal with the backend component SCHEMA Content Delivery Server (SCDS) provided by SCHEMA Holding GmbH and its import interface. SCHEMA Portal is a commonly-used CDP for industrial applications. An implementation alternative for the import interface will be shown with the CDP TopicPilot from DOCUFY GmbH.

2. REQUIREMENTS FOR THE DATA IMPORT INTO THE CDP

Currently, vendors of CDPs have different requirements, how the input data has to be provided so that the respective CDP can interpret, process and display it correctly. The reasons behind it are the historic lack of a standard and maybe the fact that the CDP vendors want to avoid to be interchangeable and loose customers, which decide providing their data by another CDP from another vendor. In principle the different CDPs require similar input data, which differs in detail e.g. in the form how the data has to be provided.

2.1 Requirements of SCDS

The SCHEMA Portal is document-oriented. For this reason, the SCDS requires individual information items e.g. topics with a document structure.

2.1.1 Content Package

For each document a content package has to be created. The composition rules of the content package are based on the packages of Microsoft Help Viewer Format [6]. The content package is a ZIP file containing a mandatory file in Microsoft Help Container (MSHC) format, a manifest file in Microsoft Help Asset (MSHA) format and optionally a thumbnail. The topics and resources are archived in a ZIP file, which uses the file extension .mshc. Currently, the SCHEMA Portal expects the content in the MSHC archive in XHTML or HTML 5, which may contain graphics, videos and audio content. Additionally, the SCHEMA Portal supports the PDF format as well as files in Microsoft Word, Excel and PowerPoint format. These formats have to be packaged with the MSHA file in place of the MSHC file to create the content package.

2.1.2 Manifest File

The package manifest is a XHTML file with the extension .msha, containing all metadata, which will be used to retrieve the individual package by the corresponding facetted search in the SCHEMA Portal.

```
<?xml version="1.0" encoding="utf-8"?>
<html xmlns="http://www.w3.org/1999/xhtml">
    <head>
        <title>Manifest T_E_1</title>
    </head>
    <body class="vendor-book">
        <div class="details">
            <span class="id">T_E_1-en-US</span>
            <span class="vendor">Karlsruhe University of Applied Sciences</span>
            <span class="locale">en-US</span>
            <span class="product">3D printer</span>
            <span class="name">Introduction to 3D Printing</span>
            <div class="metadata-list">
                <span class="meta" title="SCHEMA.CDS.Concept">m_id</span>
                <span class="meta" title="SCHEMA.CDS.Concept">m_crea_by</span>
                <!-- [...] -->
            </div>
            <a class="thumbnail" href="thumb/Icon_3D.png">Icon_3D.png</a>
        </div>
        <div class="package-list">
            <div class="package">
                <span class="name">T_E_1-en-US</span>
                <span class="mime-type">application/xhtml+xml</span>
                <a class="current-link" href="T_E_1-en-US.mshc">T_E_1-en-US.mshc</a>
            </div>
        </div>
    </body>
</html>
```

Figure 1. Example for a manifest file (SCDS)

The package manifest optionally includes metadata about e.g. title, author, language, product, product name and a description for describing the respective content packages. These Microsoft Help Viewer specific metadata are not used in the CDP for filtering. Strictly speaking, these metadata are certainly important metadata to which the user should be able to filter. It is redundant to assign them in the intended element tags as well as filter options in the manifest file. The indented element tags for describing the package could be used at the very most for internal management and for describing the relation of a manifest file to a content package to avoid confusion. In addition, the manifest file references all additional components in the archive [5, p.88-89]. For each content package a thumbnail in PNG or JPG format can be applied, which functions as an icon of the content package and appears in the content collection of the SCHEMA Portal UI [5, p.88].

2.1.3 Facet File

For the usage of the facetted search the SCHEMA Portal requires an XML file with all filter criteria. Because the mechanism of filtering is controlled by metadata, the XML file has to be the collection of all possible metadata, called facets in this context, which can be filtered in the SCHEMA Portal to retrieve the individual topics. The structure and content of the facet file in XML is based on specific rules of an XML SCHEMA Definition (XSD), which is defined by the vendor SCHEMA [8, p.132].

There is only one facet definition file, which applies to all content packages in the Schema Portal. The metadata classes and values in all language variants are maintained in this facet file. The metadata structure, e.g. taxonomy, is mapped by the nested element tags in the facet file. Furthermore, it is possible to specify in the facet file e.g. for which search scopes the facets should be valid [7, p.8].

3. IMPORT OPTIONS OF DATA INTO THE CDP

The options of importing content into the SCDS differ depending on whether a homogeneous or a heterogeneous system landscape is given.

3.1 Homogeneous System Landscape

The vendors supply standard solutions for importing content directly to the CDP when the customers are using their respective content management system. In the example of SCHEMA Holding GmbH, the CMS SCHEMA ST4 gets enabled to directly import content packages to the SCDS via the configured direct import interface [9, p.7, p.17].

3.2 Heterogeneous System Landscape

In a heterogeneous system landscape, where software systems of different manufactures are used, it is necessary to import packages in the CDP with a self-developed import process, which requires programmed transformations.

4. TRANSFORMATION PROCESS FOR THE IMPORT OF CONTENT INTO THE CDP

4.1 Relevance of Content Intelligence

The selective access of content is applicable on condition that the content is intelligent. Currently, there are three levels of content intelligence: Native intelligence, augmented intelligence and

artificial intelligence [3]. For selective access of content in the CDP this content has to be natively intelligent at least. This means the native or natural intelligence, which is inherent in the content. Therefore, a metadata model and a semantic information model is required on which the content is based [10].
The semantic structure explicitly reveals the semantic function of a proposition or term by tagging them in XML element tags like action, result and notice and makes the content machine-readable. The semantic structures are usually given and enforced by XML-based information models, which are implemented as XML Schema Definitions (XSDs) or Document Type Definitions (DTDs) [2]. The semantic tagging of content could realize filtering of content more granular down to the level of text fragments and terms with semantic functions. The information model in form of a DTD or XSD ensures the consistency of XML instances, which is essential for efficient transformation and automation.
For the possibility of content retrieval, it is required that the information objects are enriched by metadata. All metadata contributes to the native intelligence by identifying and classifying the products and the corresponding information objects to control the variance [3, p.74].

4.2 Initial Situation

Our colleagues at the University of Aizu delivered content mainly about technologies like 3-D printing and IoT in various formats like PDF and HTML. The project participants of the Karlsruhe University of Applied Sciences processed the content in XML based on a self-developed semantic information model. The next step was to enrich the individual XML topics with unique IDs and appropriate metadata.
Because the SCHEMA Portal is web-based, the XML files have to be transformed to HTML with a transformation script written in XSLT. The following transformations are based on a solution developed by Tim Rausch in his master thesis [11], which we extended and configured for the use cases in our project.
Two transformation processes have been implemented. The first process generates the necessary files for the content package and the second one generates the facet file.

4.3 Facet Definition

The facet file in XML format is the collection of all possible metadata for filtering in the CDP. An example for a facet definition in a facet file is given in figure 2.

```
<cdsnav:facet use="filter"
              id="d1e112"
              type="taxonomy"
              valid-for-scopes="downloads manuals products topics default">
    <cdsnav:title xml:lang="en">module_type</cdsnav:title>
    <cdsnav:concepts>
        <cdsnav:concept id="m_mo02">
            <cdsnav:title xml:lang="en">task</cdsnav:title>
            <cdsnav:concepts/>
        </cdsnav:concept>
        <!-- [...] -->
    </cdsnav:concepts>
</cdsnav:facet>
```

Figure 2. Example for the definition of a facet in the facet file (SCDS)

The facet file consists of facets, which correspond to metadata classes whose names are given by titles in the respective languages. Each facet contains concepts, which correspond to metadata values. The names of the metadata values are also given by titles. The linear or hierarchical structure of metadata values is mapped by the corresponding arrangement of the concept elements either in nested or linear structure. Each facet and concept must have a unique ID, which will be referenced by the individual HTML topics for classifying [8 p.132; 11 p.42]. To control the valid search scopes of the facets, the respective search scopes have to be listed in the attribute `valid-for-scopes`. If the `valid-for-scopes` attribute contains the value `topics` then the facet will be used to filter concrete topics inside of content packages, otherwise the facet will be used in the CDP for filtering the content packages.
The metadata concept can be created directly as a facet file or a transformation from an external metadata format into the facet format is required. The facet file in XML is easier to generate when there is an appropriate metadata file in an analog structure. In our project the metadata concept is mapped in a separate metadata file in a hierarchical structure already containing metadata IDs, which could have been generated by the transformation script as well.
If a metadata class should be used for filtering either the content packages or topics in the CDP, this has to be specified in the metadata file so it can be transferred to the facet file by the transformation. It is recommended to add the information about the search scope of the metadata class in an attribute of the respective metadata class.
Because of the recursive structure of the facet definition in the XSD of SCHEMA, the transformation script can be programmed in a recursive way.

4.3.1 Generation of Topics and the Manifest File

Finally, the metadata file will be transformed by the XSLT script to generate the facet file as result, which can be uploaded to the SCHEMA Portal.

4.4 Basic Structure of the Transformation Script Generating Files

A particular challenge developing the transformation results from the mass processing of XML files simultaneously where each file requires different content and has to contain unique information, e.g. IDs. The code in figure 3 shows a part of the transformation script for generating the necessary files, which can be considered as the macrostructure of the transformation.

```
<xsl:template match="/">
    <xsl:call-template name="index"/>
    <xsl:for-each select="collection(concat($goToProjectFolder,'/',
                                        '?select=*.xml;recurse=no;on-error=warning'))">
        <xsl:result-document encoding="utf-8" method="xhtml" omit-xml-declaration="no"
                             href="{concat($packageID, '-', $langCode)}/{$langCode}/
                                   {concat(//meta[@class='ID']/@value,'.html')}">
            <html>
                <xsl:call-template name="headContent"/>
                <xsl:call-template name="bodyContent"/>
            </html>
        </xsl:result-document>
    </xsl:for-each>
    <xsl:call-template name="MSHA"/>
</xsl:template>
```

Figure 3. Macrostructure of the transformation script for generating files

This template processes the index file first which functions as the cover page of each content package. In the next step, the `collection()` function requires an URI as an argument to return a collection of nodes. In this case, the argument consists of a composed path to navigate to the package project folder and to access all files with extension .xml to retrieve the collection of the input XML files. `<xsl:result-document>` is used to generate multiple output files. The attribute `href` is used to specify the result documents. A content package folder named by its predefined ID and language is generated. The ID for each XML file is queried, which is used to create the unique HTML file name in specific subfolders. The transformation of content required in the `<head>` and `<body>` area of each HTML file is specified in respective templates, which are subject of discussion in the following chapters. Additionally, the transformation script will also generate the necessary manifest file.

4.5 Topics

The CDP of SCHEMA requires each topic in an XHTML file, which contains valid XHTML markup.

4.5.1 Metadata of Topics

The `<head>` element as the container for metadata must include the metadata required in Microsoft Help Viewer Format [6].

```
<meta name="Microsoft.Help.Id" content="T_E_1-T_E_1_1" />
<meta name="Microsoft.Help.TocParent" content="T_E_1-index" />
<meta name="Microsoft.Help.TocOrder" content="1" />
<meta name="SCHEMA.CDS.Concept" content="m_id" />
<meta name="SCHEMA.CDS.Concept" content="m_crea_by" />
```

Figure 4. Example of required metadata in HTML files (SCDS)

The CDP of SCHEMA needs to arrange the HTML files in a logical hierarchical document structure. To facilitate this, each HTML file includes the ID of the topic (`Microsoft.Help.Id`), the ID of the hierarchically superordinate topic (`Microsoft.Help.TocParent`) as well as the order of the topic (`Microsoft.Help.TocOrder`).

It is recommended to define a reference structure with all this information in a separate XML file, which is used to build up the document structure for each package to support the generation of necessary data in each HTML file. This logic of reference structure has been established in many other online media like Microsoft HTML Help and ePUB for e-books. The reference structure is comparable to the table of contents of a document with the difference that it contains IDs instead of titles. The reference structure is structured in nested chapter elements with attributes containing the ID of the topics to be integrated in the document.

```
<xsl:function name="pixsl:getParent">
    <xsl:param name="value"/>
    <xsl:value-of select
     ="if (document($goToProjectRefFile)/document//chapter
          [child::chapter[@ID=$value]]/@ID !='')
      then (document($goToProjectRefFile)/document//chapter
          [child::chapter[@ID=$value]]/@ID)
      else ('index')"/>
</xsl:function>
<xsl:variable name="topicID" select="//meta[@class='ID']/@value"/>

<meta name="Microsoft.Help.TocParent" content
 ="{concat($packageID, '-', pixsl:getParent($topicID))}"/>
```

Figure 5. Possible solution for querying the ID of superordinate topic

A function is used to add the ID of the correct parent topic in each HTML topic. It is defined with `<xsl:function>` and is used in XPath expressions like variables. Each function needs a namespace and a name, e.g. `pixsl:getParent`. A function can be reused in different places in the code and with different arguments according to the required context. For example, the `pixsl:getParent` is used to generate the information about the parent topic of the currently processed XML and also for generating the breadcrumb tail.

The function queries the ID of the currently processed XML topic using the variable `topicID` and if the ID is not null, it looks up the corresponding ID in the reference structure. At last the function queries the ID of the hierarchically superordinate topic. If there is no hierarchically superordinate topic, the parent topic is the index page, which will display the name of the content package.

The topic ID assigned in the HTML topics is based on the concatenation of the package ID to which the topic is assigned and the ID, which is given in the XML topic. The reason for this approach is the fact that the CDP of SCHEMA is unable to handle different topics with the same ID. To make it technically possible to reuse any XML topic in various content packages, the HTML topics get the ID with information in which package the HTML topic is used and from which XML file it emerges. The concatenated IDs in HTML files are also needed for the index file, which functions as the cover page of each content package. In contrast to the topics, which are based on XML topics, the index files will be generated from scratch. The index IDs is composed by the package ID to which they are assigned and the string "index". In principle, XSLT offers the function `generate-id()`, which facilitates the generation of unique IDs. But the approach of controlled and concatenated IDs presented supports easy traceability and assignment to content packages.

```
<meta name="Microsoft.Help.TocOrder" content
 ="{count(document($goToProjectRefFile)
  //$hierTopicID/preceding-sibling::chapter)+1}"/>
```

Figure 6. Expression to determine the topic order

To indicate the topic order, which must be specified in each HTML topic, the transformation script queries the ID of the currently processed XML topic and looks up the corresponding ID in the reference structure and counts the preceding-siblings and adds 1 for the current topic.

To keep the efforts for classifying the XML files low, it is recommended to copy the required metadata of the metadata file and paste them into the XML topics.

The defined metadata in the XML topics, which are used for classifying are transformed in the corresponding HTML topics within the head container in specific syntax (see `SCHEMA.CDS.Concept` in figure 4). It is notable that only the metadata ID, which is defined in the facet file has to be referenced in the HTML topics. The CDP will match the metadata IDs in the HTML topics by comparing them with the data in the facet file to display the respective metadata names. If taxonomical metadata is used in the XML topics, it is not necessary to add all the ancestor metadata. Instead, it is sufficient to add just the metadata value of the lowest hierarchical level of the taxonomy. If the rate of changes in the metadata concept is very high or the technical writers feel uneasy, it is better to map all hierarchical levels because it reduces the degree of abstraction and ensures better data traceability.

4.5.2 Links between Topics

To insert a link in an XML file to another topic the element `<xref>` has to be used in compliance with the information model. `<xref>` contains an attribute `idref`, which references the ID of the topic as link target.

```
<xsl:function name="pixsl:genMSHelpLink">
    <xsl:param name="value"/>
    <xsl:variable name="MSHelpLink" select="'ms-xhelp:///?Id='"/>
    <xsl:value-of select="concat($MSHelpLink, $packageID, '-', $value)"/>
</xsl:function>

<xsl:template match="xref">
    <a href="{pixsl:genMSHelpLink(@idref)}" rel="help">
        <xsl:apply-templates/>
        <xsl:text> </xsl:text>
        <xsl:value-of select="document(concat(
         $goToProjectFolder, '/', @idref, '.xml'))/*/title"/>
    </a>
</xsl:template>
```

Figure 7. Possible solution for linking topics

In the template matching the `<xref>` element, the `idref` attribute is queried and is used as the argument value of the function `pixsl:genMSHelpLink`. Because of the MS Help Viewer specification, the link has to be introduced with `ms-xhelp:///?Id=’`. To map the HTML topic ID of the target topic, the ID of the currently processed package has to be queried, which will be concatenated with the queried `idref` specified in the XML file. To specify the topic title as link name, the `document()` function is used to access the XML topic file by its topic ID in the project folder. From this XML file the content of the `<title>` element of the topic is selected.

4.5.3 *Interactive 3-D Data*

As we are dealing with 3-D printing, which contains 3-D data, we wanted to show a way to display 3-D data in the CDP of SCHEMA. SCHEMA does not provide a solution for the integration of 3-D data in web pages. Our most important requirement is to find a solution to display 3-D data in web pages on each computer without efforts like changing browser settings or installing plugins.

X3DOM, which consists of X3D (Extensible 3-D) and DOM (Document Object Model), “is an open source JavaScript framework, used to create declarative 3-D scenes in Web pages”. [13] “Node descriptions are placed in the HTML document or included via the inline tag, as regular XML / HTML elements.” [14] Only a minimal amount of code is necessary.

```
<div style="width:100%; height:600px">
    <x3d style="width:800px; height:600px">
        <scene>
            <inline url="media/3Dprinter.x3d" />
        </scene>
    </x3d>
</div>
```

Figure 8. Example for a referenced X3D model in HTML

It is recommended to integrate the X3D data in a distinct `<div>` container in the `<body>` of the HTML topic. In the X3D context the `<scene>` element contains the `<inline>` `url` with the reference to the external X3D file by using relative path specification [15] (see figure 8). In the head of HTML topic, it is required to reference the X3DOM JavaScript and stylesheet files (see figure 9).

```
<link rel="stylesheet" type="text/css" href="../assets/x3dom/x3dom.css" />
<script type="text/javascript" src="../assets/x3dom/x3dom.js"/>
<meta name="viewport" content="width=device-width, minimum-scale=1.0, maximum-scale=1.0" />
```

Figure 9. Required specification in HTML head for displaying X3D model

For an optimal view of the X3D object, the viewport can be adjusted by inserting some specification in `<meta name=”viewport”>` in the `<head>` of the HTML topic or in the X3D file itself. Obviously, the 3-D models in X3D format as well as the X3DOM library has to be included in the content package e.g. in the assets folder. It should be mentioned that the Firefox browser behaves reliable in displaying 3-D data in web.

4.6 Layout

Links to stylesheets and JavaScript files for the layout of the topics have to be added in `<meta>` elements in the HTML topics. It is possible to include self-developed layout stylesheets. Accordingly, a layout stylesheet could be included for the corporate design of a company. Alternatively, the default CSS of SCHEMA can be deployed to reduce the effort. The Bootstrap CSS Framework in version 3.0.2 [12] as well JQuery in version 1.10.2 is provided [11, p.33]. The stylesheet specification of Bootstrap can be extended arbitrarily to customize the layout of the content of the topic.

The transformation script for generating the topics adds the links to the CSS and JavaScript by relative path specification for each topic. For each element type defined in the information model another transformation script specifies the layout and technical functions like links to websites and topics, counting captions as well as preview images for videos. The layout transformation script is included in the main transformation script. The separation of these transformation scripts has the advantage that the files can be changed independently from each other. For example, different stylesheets for various target groups or devices may be created, which can be exchanged by just referencing a different layout transformation stylesheet in the main transformation script.

4.7 Manifest File

The package manifest file has to contain e.g. all valid metadata in the `<div class=“details“>` area, which is used to retrieve the individual package by filtering in the CDP. This type of metadata is known as extrinsic metadata in Technical Communication (see [4]). It is appropriate to collect and define these extrinsic metadata in one file, which is valid for the individual content package. This file is queried by the main transformation script to generate the package manifest. To avoid the complexity of too many files, it is recommended to collect and define these metadata as well as the required package ID and the title of the package in the already existing reference structure. The easiest way for the technical writer is to copy the needed extrinsic metadata for the content package from the metadata file and insert them in the reference structure. Because this principle is the same as in assigning metadata in the XML topics, technical writers do not have to learn a new procedure and syntax and as a further benefit, the programming code in the transformation script can be reused. The transformation script generates the required syntax for metadata in the manifest by assigning the metadata ID from the reference structure (see figure 10).

```
<meta name="SCHEMA.CDS.Concept" content="m_id" />
```

Figure 10. Example for a metadata assignment in manifest file (SCDS)

Because the extrinsic metadata in a content package is valid for all contained topics, the extrinsic metadata could be defined in the reference structure once. The transformation script queries these metadata and adds them to each output topic. Generating the extrinsic metadata in all HTML topics has advantages like reducing manual efforts for technical writers and possible errors from manual work.

Furthermore, a unique package ID and a package name as the title of the document have to be defined in the reference structure, which will also be transferred to the manifest file.

The <div class="package-list"> area in the manifest file references the content of the content package. In this area the thumbnail picture can be referenced by relative path specification as an icon of the content package, which appears in the content collection of the SCHEMA Portal UI. In package-list area the package ID is required and also the format of the content like XML, PDF or DOCX has to be defined. Finally, the content archive with the extension .mshc has to be referenced by relative path specification [5, p.88-92]. All these data in the manifest will be generated by the main transformation script.

4.8 Generation of Topics and the Manifest File

Figure 11 gives an overview of the transformation for processing files.

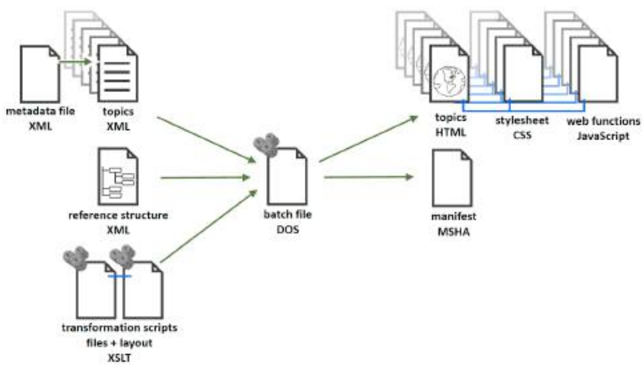


Figure 11. Transformation process for generating topics and the manifest file

The XML topics receive data from the metadata file and the reference structure. Executing the batch file, the transformation scripts will be triggered to generate the HTML topics with links to the CSS stylesheets based on the input XML topics and the manifest file based on the reference structure.

4.9 Composition of Content Packages

Usually the archive in MSHC format includes a directory structure for all types of content, which is organized in specific folders. The structure of the folders defined by SCHEMA is extensible. It is essential that all relative path specification and references are valid within the content package. In our project environment we indented to create a separate folder for each language variant in the MSHC archive containing the index and topics in HTML. Additionally, we created an assets folder in the MSHC archive containing files for the representation of the HTML like layout stylesheets, fonts, JavaScript files and media objects, including the thumbnail. All topics of a content package in one language variant in XHTML format have to be zipped with all the additional data. The generated file with the extension .zip has to be renamed to a file with extension .mshc. In the next step, the MSHC file has to be zipped together with the MSHA file to create the content package to be uploaded to the CDP [11, p.32].

For uploading content in other formats like PDF or a ZIP with 3-D data, it is necessary to package these files instead of the MSHC file with the corresponding MSHA file. It has to be considered that the MSHA file specifies the respective content formats.

The development of the transformations is an iterative process. Combined with the number of documentation packages, the effort of manually composing the content packages for SCHEMA is huge and not justifiable. Therefore, this process has to be automated to be feasible. Below an optional process for automating the composition of all created packages will be presented, which can be applied if no CMS is used for automation. The need to configure the transformation script for each content package should be avoided. A batch file can be implemented, which has access to files and folders and which controls the generation of all content packages.

There are some components like warning symbols for safety advices in form of pictures or layout stylesheets, which are relevant across content packages. These data, which are used across are collected in a folder, which is called “standard” in our case.

Let us assume that several authors are writing XML topics for content packages and that the folders of these package projects do not have unique names and are in different directories. The batch file will access individual package projects by path specification and the projects will be copied to a temporary folder acting as a pool of all projects. The copied projects get their package ID as folder name. By starting the batch file, the transformation script will be triggered for each input package project in the pool. This will iteratively create a new folder with HTML files and the MSHA file for the respective content package. In the next step, the batch file copies the content of the folder called “standard” into each content package folder in the correct place. In the following step the batch script copies the project-specific data like media objects in the assets folder of the respective content package folder. The batch script will zip the HTML topics per language variant together with the additional assets data and renames the archive automatically into a file with extension .mshc, which finally will be zipped together with the generated manifest file.

With just one mouse click to execute the batch file, all content packages can be automatically generated in a form that the packages can be uploaded into the CDP.

4.10 Different Approaches of Importing Data into the CDP of SCHEMA and DOCUFY

The currently available CDPs usually have the same core functionality. Below some comparable import specifications and different technical approaches between the CDPs of SCHEMA and DOCUFY will be highlighted.

As the name of TopicPilot suggests, the CDP is topic-orientated [16, p.6], compared to the CDP of SCHEMA, which takes a document-orientated approach. TopicPilot also requires content packages in a ZIP format and supports not only individual topics in the content package, but also documents consisting of many topics within a document structure. The content package for TopicPilot consists of media objects, individual topics files and a manifest file as for the CDP of SCHEMA. It also requires a reference structure, in contrast to the CDP of SCHEMA. The content packages for TopicPilot do not support assets folders with stylesheets and JavaScript files.

While the CDP of SCHEMA requires topics in HTML, TopicPilot in Version 3.2.1 expects the topics in DYXML2 format [11, p.47], which has to conform to the corresponding DTD [16, p.15]. DYXML is a markup language based on XHTML, which is developed by DOCUFY [16, p.15; 17]. By importing the topics in TopicPilot, the DYXML2 input topics will be automatically transformed to HTML for web display [16, p.15]. With the use of HTML as input format in SCHEMA Portal, it is easier to control the layout and to use advanced features like JavaScript or 3-D data. Transforming XML topics to HTML has the advantage that the markup language is widely used. By comparison, the transformation of XML topics to DYXML2 can get complex. The specification of DYXML2 is about one hundred pages long. Examples of the use of DYXML2 are hard to find.

The required manifest files of both CDPs differ in content. While the manifest of the SCHEMA Portal principally expects the metadata for filtering the content packages, the manifest file for TopicPilot just declares version of the manifest and the version of DYXML2 used for the content package [11, p.47].

A reference structure, which maps the document structure in form of the table of contents is required by both CDPs for assigning the individual topics to the document structure. While the reference structure for the data integration into the CDP of SCHEMA can be developed freely for supporting the transformation script, TopicPilot requires the reference structure to be compliant to its rules in each content package. TopicPilot calls its reference structure “topic tree”, which has to be defined in a separate file in DYXML2 format.

```
<node id="d3e21" name="Safety">
    <section id="d3e22" data-type="chapter">
        <p id="d3e23" data-type="title" data-role="heading">Safety</p>
    </section>
    <node id="d3e34" name="General Safety Instructions">
        <topicref tid="00000000000000000000ba47d9211cac" version="1" xml:lang="en-US"
         href="../topics/00000000000000000000ba47d9211cac_1_en_US.xml"/>
    </node>
</node>
```

Figure 12. Example for the topic tree (TopicPilot)

The reference structures for both CDPs use the topic IDs to link the topics. The “topic tree” in TopicPilot has additional chapter headings, compared to the reference structure for the CDP of SCHEMA, and the hierarchy is mapped by nested elements including `href` attributes referring to the respective topics by relative path specification, the topic IDs, their language and version [16, p.22; 11, p.51].

Advantages of the reference structure for the CDP of SCHEMA are the freely definable structure and content. The possibility of defining chapter headings in the reference structure for TopicPilot, which will be automatically generated in the CDP is practical. In SCHEMA it is necessary to create individual topics with chapter headings, which requires administrative effort.

Both CDPs have in common that they require facet definition in separate files. One difference relies in the fact that the facet definition of SCHEMA is used across all content packages within the CDP and it is not possible to have more than one facet file in the CDP. On the contrary, TopicPilot requires a facet file in each content package, which may differ across the content packages. The definition of the facets is similar for both CDPs. Each facet is defined by a class name and a unique ID and contains facet values for respective languages variants with unique IDs.

Because TopicPilot requires a facet file for each content package, it is necessary to use same IDs for the same facets across the facet files, otherwise there will be redundant facets in TopicPilot. The sum of all facets defined in the respective facet file map the overall set of facets, which will be displayed in TopicPilot. Each facet defined in these files will be added to this overall set of facets [11].

The facet file of TopicPilot is divided in two parts, the definition of the facets and the assignment of facets to topics of the content package [11 p.57]. All necessary metadata for the topics of one content package are declared in the first part of the facet file. Thus, the topics specified for TopicPilot do not include the classifying metadata for respective variant of the topic, in contrast to the CDP of SCHEMA. In the second part of the TopicPilot facet file all topics are listed, which are included in the content package referenced by relative path. One or more facets are assigned to these topics by referencing the respective facet IDs [11, p.57-59;
17 p.10-14] . As in the CDP of SCHEMA, it is sufficient to reference the lowest value of the hierarchical level of the facet [11]. Figure 13 shows a code snippet of the facet file of TopicPilot including a topic with two assigned facets.

```
<topic href="topics/00000000000000000000000ba47d9250_1_en_US.xml">
    <facet-ref ref-id="111111111111111111111111111d1e299"/>
    <facet-ref ref-id="111111111111111111111111111d1e139"/>
</topic>
```

Figure 13. Example for the assignment of metadata to a topic (TopicPilot)

The definition of one facet occurs in multiple facet files in TopicPilot. It is a potential source of errors, defining the same facet with same ID in several places because the contradictions, which can arise, lead to duplicate display of facets in the front end.

The approach of SCHEMA with the idea of maintaining all possible facets in a master document as the single source of facets has the benefit that there is a comprehensible overview and necessary changes to the facet definition have to be done in just one place. In TopicPilot, the overall facet tree is not visible at a glance within the content packages, merely parts of it.

With our experience in the project, it is not recommended to create the content packages directly in the required formats and structures of the import specifications of the respective CDPs. Originally, the concept of importing data of each vendor was optimized for efficient transfer of data stored in their own CMS to the respective CDP. Finally, the import specifications are not designed to be human-readable and comprehensible.

4.11 Advantages and Disadvantages of self-developed Transformations for Data Integration

There are several advantages in regard to self-developed data transformation for integrating the content to CDPs.

In some CMS, e.g. SCHEMA ST4, just one information model is implemented, which has to be used for structuring the XML without any alternatives. In different CMS the implemented information models can differ significantly in the degree of semantic structures [2]. By developing a transformation process one gets the freedom to write XML topics based on any information model. It should not be underestimated that one also gets completely freedom in designing various layouts for specific user groups and devices. Often the CDP vendors offer just a few layout stylesheets without the option of customization. With a self-developed transformation, different web technologies like interactive 3-D display can be integrated in the CDP, which the CDP vendor does not support by default. Requesting such extensions from the CDP vendor might involve large costs for customization.

The large effort of developing the transformation is the main disadvantage. New releases or updates of the CDP could involve a changed import specification. Consequently, the transformation scripts have to be adapted.

5. THE iiRDS STANDARD

In March 2017 the “European Association for Technical Communication” (tekom) published a first working draft of the “intelligent information Request and Delivery Standard” (iiRDS) [18,30]. The purpose of the open source standard iiRDS is to request and deliver intelligent information between individual and independent companies [18]. “iiRDS enables the exchange and assembly of digital user assistance content across systems and manufacturers.” [18]

iiRDS standardizes two aspects of intelligent information. First, the method, how use cases can be transformed to valid requests. Secondly, how valid requests can be answered by delivering appropriate information objects [19, p.18].

The standard specifies a package format for documents similar to those of DOCUFY and SCHEMA: Furthermore, it introduces a data model for the annotation of technical documentation in Resource Description Framework (RDF) [18].

“iiRDS defines a taxonomy of information types and describes relations between information units as a basic ontology.” [18] An ontology is a set of classifications of concepts in a conceptual hierarchy and which are connected to each other by associative relations. Ontologies have the objective to represent the knowledge within a domain [19, p.27]. Developers and technical writers use and profit from ontologies mapping the dependencies of product components and variants to provide an overview of the domain and traceability of the concepts. Thus, iiRDS provides a vendor neutral vocabulary specialized for and restricted to the domain of technical documentation and user assistance. Due to the network structure, which characterizes an ontology, it is not appropriate to map the logic hierarchically like implemented for the metadata concept in a taxonomy structure without considering their logical relations. Taking a look at the specification of the ontology in “iirds-core.rdf”, it is obvious that the structure is not deep; it does not exceed the third hierarchy level. Classes, types, type of relations or references to instances will be expressed by semantic tagging of elements (see figure 14).

The content files have to be compliant to the iiRDS XHTML5 format, which predefine a set of XHTML elements and media formats to reach maximum compatibility of content files for delivery packages between the iiRDS Consumer [18]. iiRDS XHTML5 is a subset of XHTML5 [19, p.32-33]. The elements of XHTML5 are layout-orientated. To express the semantic of elements, it is recommended to add the attribute `data-role` on HTML5 elements. Semantic tagging with `data-role` enables iiRDS Consumers to render information correctly on a display device [18]. Combining machine-readable content by using semantic tagging and machine-readable RDF, the content can be queried and connected with other data [19, p.29].

```
<iirds:Topic rdf:about="http://myCompany.com/io/1312">
    <iirds:title>Mounting the rotor</iirds:title>
    <iirds:language>en</iirds:language>
    <iirds:revision>2</iirds:revision>
    <iirds:has-topic-type rdf:resource="http://iirds.tekom.de/iirds#GenericTask"/>
    <iirds:has-rendition>
        <iirds:Rendition>
            <iirds:format>application/html</iirds:format>
            <iirds:source>rendition/mounting_en_1.html</iirds:source>
        </iirds:Rendition>
    </iirds:has-rendition>
</iirds:Topic>
```

Figure 14. Example for a topic with metadata compliant to iiRDS specification

With iiRDS the metadata is not specified within the content files. The `<iirds:informationUnit>` is the abstract base class for intelligent information in the iiRDS metadata file. The information unit can be understood as a kind container, which represents a specific set of metadata about a piece of information.

Only subclasses of the `<iirds:informationUnit>` can be used, e.g. `<iirds:Topic>` [18]. Therefore, the metadata like ID, format or language are specified in subclasses of the abstract information unit. The example in figure 14 shows the usage of an `<iirds:Topic>` for a topic with the title “Mounting the rotor”. The English variant has two revisions. The type of the topic in the example is a generic task, which is expressed by an URI in `rdf:resource`. The URI at identifies an abstract or physical resource. `rdf:resource` refer to the metadata URI defined in the metadata file “iiRDS.core.rdf”:

```
<iirds:Task rdf:about="http://iirds.tekom.de/iirds#GenericTask">
    <rdfs:label xml:lang="en">Task</rdfs:label>
</iirds:Task>
```

Figure 15. Example for a metadata definition in iiRDS.core.rdf

The topic has additional metadata referring to the content lifecycle status. An information unit refers to a physical instance like a document, topic, fragment within a topic or a iiRDS package [18, 19]. In this example, `<iiRDS:Rendition>` refers to a HTML file in the respective content package, which is stored in the folder rendition.

The content of an instance gets classified by the metadata of the information unit in which it is referenced. This approach should facilitate that several information units can refer to a single physical file as an instance to use the content in various contexts. Thus, the possibility of reusing content is given, which is not feasible currently in the CDP of SCHEMA because the metadata and ID are defined directly in the instance.

Using ontologies enables the automatic connection of content by known relations or use case relevant relations to display corresponding related links to additional information, which supports user guidance. Thus, manual efforts as well efforts to program them manually in transformation scripts, like we did within the project, will be saved.

iiRDS has also a method to assemble a content package in a linear or hierarchical structure using nested semantic tagging and references to the content.

The standard describes possibilities for realizing requests in the context of IoT use cases. Content stored in RDF format within a semantic database can be queried by using SPARQL (SPARQL Protocol and RDF Query Language) to manipulate or retrieve specific content [18].

With the adaption of iiRDS as a standard to import packages the relevance of the individual import specifications of the different vendors of CDPs will diminish.

6. CONCLUSION

The strength of intelligent Content Delivery Portals is based on targeted delivering of tailored information depending on the framework considering e.g. the current use case, situation, language environment, geographic positions, target group or processing system, knowledge level, display device and necessary display format. All these factors are not considered by providing the user a document consisting of a thousand pages in which he has to cumbersomely search the information to close his knowledge gap. Users have the option to access selectively and dynamically the specific content applying the search strategies. Each CDP vendor has its own requirements of how the input data has to be provided so that the respective CDP can interpret, process and display it correctly. The current situation leads to costs switching from one CDP to another and therefore may result in vendor lock-in. If there is no interface, which automatically transforms the content into the format, which is accepted from the CDP, the requirement of a self-developed transformation of content and additional required data arises. Therefore, it is necessary to investigate the specification of the import interface of the respective CDP. Self-developed transformations for importing data to the CDP require some efforts indeed, but e.g. working with the CDP of SCHEMA one gets the freedom to base topics on any information model and to manipulate and extend the HTML files arbitrary. This way, we enable the integration of interactive 3-D data in X3D format into the CDP. Tekom offers a proposal of standardizing the information delivery with the iiRDS standard. iiRDS considers the upcoming requirements and use cases in digital transition and describes e.g. methods of interacting and exchanging relevant and required information among machines, devices and users. The augmented intelligence support in iiRDS by using ontologies opens links and relations of product components and functions to find related topics automatically and to optimize the user guidance. The delivery of adapted content for a specific goal is a justified requirement of customers and simultaneously a challenge for Technical Communication in digital transition, which can be mastered by using CDPs and intelligent delivery concepts.

7. OUTLOOK

The integration of iiRDS in the CDP has strong impacts on self-developed transformations for importing data to the CDPs. For example, because the facet file no longer exists, there is no use of a transformation script for generating facets. Additionally, it is expected that the basic ontology of iiRDS might be extended with a GUI, which visualizes the network. Depending on how the iiRDS standard is implemented by the CDPs and how it will be developed further, self-developed transformation processes loose relevance.

iiRDS is a promising and sophisticated approach for standardizing metadata and their logical relations as well as for requesting intelligent information in context of use cases e.g. in IoT. Augmented intelligence approaches like in iiRDS have the potential of optimizing search and retrieval methods and didactical user guidance by using extensive and logical metadata, which correspond to user expectations and knowledge. On the other hand, more metadata have to be assigned because of the additional relations, which are also a kind of metadata [3]. Furthermore, the complexity increases. “Artificial intelligence (AI) can be used to reduce the manual effort of augmented intelligence concepts by, for example, automating text classification [see 20]. It can also reveal hidden relations, which have not been revealed before.” [3a]. AI also has the potential to automate the classification of content in the ontologies [19].

The kind of intelligence, which exceeds the native intelligence, enables many automation processes to reduce the manual effort related to programming self-developed transformation processes. For example, links between content objects can be generated automatically based on the ontology-based modelling of relations. Therefore, it is no longer necessary to generate related links by hand.

8. ACKNOWLEDGMENTS

Special thanks to Tim Rausch (SCHEMA Holding GmbH) for insightful discussions and support. We would also like to thank our project colleagues at the University of Aizu for the enriching experience and input data about 3-D printing and IoT and Prof. Debopriyo Roy, Ph.D. (University of Aizu) for the organizational support.

9. REFERENCES

[1] Ziegler, W.2018. Content Delivery and Content Analytics for Industrial Applications. The Upcoming Era of Digital Information Services. In Frontier, Official Journal of Japan Technical Communicators Association JTCA.

[2] Ziegler, W. 2017. Basic Concepts Which Support the Management and Delivery of Intelligent Content, p. 84-91.

[3] Ziegler, W. 2017. The Evolution of Content Management towards Intelligent Delivery Systems for Technical Communication. In Frontier, Official Journal of Japan Technical Communicators Association JTCA, p. 68-75.

[4] Drewer, P., and Ziegler, W. 2011. Technische Dokumentation. Übersetzungsgerechte Texterstellung und Content-Management. (2nd ed.). Vogel, Würzburg.

[5] SCHEMA Holding GmbH. 2018. SCHEMA CDS. API Reference. Version 2018.9. Reference documents for SCDS of SCHEMA Holding GmbH from 2018-10-17.

[6] Microsoft. 2016. Microsoft Help Viewer SDK (Apr. 2016). Retrieved 2019-02-08 from https://docs.microsoft.com/en-us/visualstudio/extensibility/internals/microsoft-help-viewer-sdk?view=vs-2017.

[7] SCHEMA Holding GmbH. 2018. SCHEMA Portal Anwendung. SCHEMA CDS. Version 2018.9. Reference documents for SCDS of SCHEMA Holding GmbH from 2018-10-15.

[8] SCHEMA Holding GmbH. 2018. Administration. SCHEMA CDS. Version 2018.9. Reference documents for SCDS of SCHEMA Holding GmbH from 2018-10-15.

[9] SCHEMA Holding GmbH. 2018. SCHEMA Content Connector. SCHEMA CDS. Version 2018.9. Reference documents for SCDS of SCHEMA Holding GmbH from 2018-10-15.

[10] Ziegler, W. 2017. Wie Intelligent können Informationen denn noch werden? Die Zukunft des Content Management und Delivery. Intelligence? What’s Intelligence in TC?. In Conference Proceedings for Annual Conference tekom.

[11] Rausch, T. 2018. Technologische Einführung – xICM. Technologische Grundlagen von Content-Delivery-Portalen. Eine Einführung in die Verarbeitung von Daten- und Austauschformaten. Master Thesis. Karlsruhe University of Applied Sciences.

[12] Bootstrap. 2016. Getting started. Retrieved 2019-02-08 from http://bootstrapdocs.com/v3.0.2/docs/getting-started/.

[13] The Fraunhofer Institute for Computer Graphics Research IGD. (n.d.). official x3dom documentation. Background: What is X3DOM, and what can it do for me? Retrieved 2019-02-08 from

https://doc.x3dom.org/gettingStarted/background/index.html.

[14] The Fraunhofer Institute for Computer Graphics Research IGD. (n.d.). official x3dom documentation. Retrieved 2019-02-08 from https://doc.x3dom.org/index.html.

[15] The Fraunhofer Institute for Computer Graphics Research IGD. (n.d.). official x3dom documentation. Include an external X3D-File. Retrieved 2019-02-08 from https://doc.x3dom.org/tutorials/models/inline/index.html.

[16] DOCUFY GmbH. 2017. TopicPilot 2.5. Importspezifikation. Reference documents for TopicPilot of DOCUFY GmbH.

[17] DOCUFY GmbH. 2018. Spezifikation. DYXML2. Austauschformat. Reference documents for TopicPilot of DOCUFY GmbH from 2018-02-16.

[18] European Association for Technical Communication – tekom e.V. 2017. iiRDS Specification - intelligent information Request and Delivery Standard – First Public Working Draft, 03 April 2017. (2017). Retrieved 2019-02-08 from https://iirds.tekom.de.

[19] Henning, J., and Tjarks-Sobhani, M. 2017. Intelligente Informationen. tekom. Schriften zur Technischen Kommunikation. Proceeding 22. tcworld GmbH, Stuttgart.

[20] Oevermann, J., and Ziegler, W. 2018. Automated Classification of Content Components in Technical Communication. Computational Intelligence. Volume 34, Issue 1. February 2018. Pages 30-48. Wiley, New Jersey, USA

PRACTICAL APPROACHES TOWARDS STEADY INFORMATION WORKFLOWS: FROM MARKETING AUTOMATION TO CONTENT CONFIGURATION AND THE CONNECTED SUB-PROCESSES

2019 Volume 1

Dominik Lächler
Bürkert Fluid Control Systems
Head of Bürkert Information Management (bIM)
74653 Ingelfingen
Germany
dominik.laechler@burkert.com

ISSN: 2434-9526

PAGES: 36-39

ACM Chapter Proceedings on Educational Technology, Language & Technical Communication

Practical Approaches towards Steady Information Workflows: From Marketing Automation to Content Configuration and the Connected Sub-processes

Dominik Lächler
Bürkert Fluid Control Systems
Head of Bürkert Information Management (bIM)
74653 Ingelfingen
Germany
dominik.laechler@burkert.com

ABSTRACT

With the increased demands which affect the development of new products in terms of connectivity and networking with other products towards the "digital factory", not only the products themselves and how they are developed change, but also the information belonging to the product and its application must keep pace in this context to keep up with more and more complex products and devices – in an B2B environment. It is clear that individual departments of a company cannot cope with these challenges on their own and so it's the management of a company that has to define it as a part of its company strategy. This highlights the importance to think out of the box – within the management and the departments. This is what Bürkert has recognized and therefore has begun to set the course for the information of the future, which will meet the increased demands on information, the delivery, and retrieval of it. The boundaries between the individual departments are becoming increasingly blurred because the creation of information has to be seen as a process and not as a departmental task: this affects the technical editing as well as the marketing of a company as the main producers of information today.

Keywords

Technical Communication; Technical Documentation; Information Modeling, Information Architecture; Content Management; Component Content Management; Content Delivery; Content Retrieval; Content Classification; Intelligent Information; Digital Information Service; Semantic Metadata; PI-Classification; Ontology; Augmented Intelligence, iiRDS; Artificial Intelligence

ACM Chapter Research Seminar on Technical Communication & Information Design 2019, January 28-Feb 1, 2019, Aizuwakamatsu, Japan

ISSN: 2434-9526

1. INTRODUCTION

The realization that the future not only belongs to the individual product but also to the specific content and not to the static document itself is only gradually gaining ground in classical mechanical engineering companies: What people actually want today is not "Illustration shows special equipment" but "Illustration shows what you actually ordered". So what companies are struggling with is how to implement it and how to set up the infrastructure to really make a difference in terms of information in the context of digitalization. So they face two major challenges here: What kind of digital content do customers need and how can we generate this kind of content as efficient and automated as possible? Is it only a matter of publishing a PDF as online help and thus only changing the format, but not the content, or is it a matter of fundamentally realigning communication with the customer, which produces completely new ways of thinking, processes and business models and thus breaking with previous publication structures and channels. In the following, we will summarize the recent situation of content management and delivery from a technical and methodological point of view. Additionally, we will generally discuss the use of these systems for creating and implementing digital information services as use cases of digitization and as business models of content delivery technologies. We will also consider more elaborated aspects like augmented and artificial intelligence, which have to be given more consideration in the future of technical communication. Bürkert in particular as a manufacturer of products with a high degree of variance faces particular challenges: On the one hand, the classical approach to the structuring and classification of content in the form of product and information-related metadata is not sufficient if the appropriate product information for the product is to be generated ad-hoc. Therefore, the evaluation of technical features of a product has to be considered what has not been the focus of most editors and goes far beyond the work of the ordinary technical editor. From pure Content Management, the way is being prepared to Content Configuration in Content Management System (CMS) On the other hand; it is an immense effort to deliver sales information that really supports the sales department in their daily work by providing the right information at the right time in a language that the user understands. Building up ontologies to build intelligent searches to deliver precise content through sharp classifications is what Bürkert is striving for at the moment.

Additionally, a consistent language in source and target languages delivers the base for all multi-language content-related activities at Bürkert. A state-of-the-art Authoring Assistant Tool and a Translation Memory System (TMS) form the necessary IT backbone [1].

The respective decision is important and affects tools, content and the necessary classifications. This paper will try to give an insight into the activities in a mid-sized company.

Structure of an information logistics chain

At the beginning it is essential to look at the basic information flows in the company in order to gain an understanding of where and how information is created, which target groups and which systems they use and what is done to ensure that this information is retrievable and is not maintained redundantly. It is to be considered with these considerations of course that not every information in the connection is content in the sense of the communication of an information, but that it very frequently rather concerns essential master data, without which today no enterprise can function no more in any specialist area. However, it can be stated without doubt that in the reverse conclusion content is always an information - this both physically as a data object as well as semantically on the basis of its inner structure, which in principle is again based on e.g. XML schemata, with its structure pursuing a certain purpose. Content as an information without purpose inevitably abolishes itself. No purpose, no content.

If one now follows these thoughts and observes the information flows from a higher level, then the four pillars of information logistics literally impose themselves on one: Content authoring, content delivery, content retrieval and content analytics. The interlocking of the individual columns runs transversely and an exact demarcation is only possible in individual cases, but the areas of delivery and retrieval become blurred in some places to such an extent that an exact demarcation neither makes sense nor is possible. It is important to understand that no pillar can exist on it's own. But be careful: In a document-based world this is even common. A world which, with regard to the individual links in the chain of information logistics in the context of digitization in the field of content, is becoming increasingly unimportant and perhaps at some point will only exist in special, highly regulated areas.

Authoring

For the sake of simplicity, instead of content authoring, I will speak of authoring in the following; the same applies to the other pillars, but is not explicitly mentioned there again. On the production side of information one finds in a company today a whole zoo of different IT systems, production processes and granularity levels of content. This of course has to do with the completely different objectives: Selling and explaining, i.e. in the classical sense, technical documentation and the individual marketing areas play a decisive role in the production of content - whether in text, image or other form. Every employee and every department does what he or she thinks is right, and the coordination of the different areas often falls by the wayside. In addition to the "classic" content creators and author groups, further players will join in the context of Industry 4.0. And that means content creators who are not human, but therefore do not play a lesser role. We are talking about increasingly "intelligent" products whose data logs and analyses enable us to draw conclusions about the condition of the product, the machine or the complete plant from these records. The interpretation and evaluation is then in human hands again. Therefore I plead at this point also impressively for the fact that the term "automation" should be handled very carefully in the context of content creation, because nothing automates itself at all, unless first of all the prerequisites are created manually to automate the content creation. For example, by creating classifications that not only look at the product side, but also at the customer and application side. A product always has certain functions, as in our example the so-called "batch filling", where certain quantities of liquid are filled into certain containers within a certain time. The fact that the product has this function results from the customer requirement, which in turn results from its applications. I.e. only the linkage of internal and external view leads at all to create standardized classification and then to develop the content relevant for this purpose. This content must always be able to answer the following questions: What am I, where do I belong and what purpose do I pursue? In the field of technical documentation, this is referred to as feature-based documentation creation, i.e. the corresponding operating information should match the ordered and configured product 1 to 1. Information models try to help and establish standards [2,3]. This task can be solved today with modern Content Management Systems (CMS), which give the creators the possibility to classify XML-based information modules accordingly or also to "tag" them. The operating information can then be generated via the interface to a configurator, provided that mapping of the data and data structures are synchronized. This may not sound too complicated in the first step - it shouldn't be. But the big challenge here is the topic of "change". With the changed perspective on information and the corresponding IT systems, the occupational profile of the corresponding groups of people is also changing. The actual editor transforms into an information architect, whereby this designation in detail also requires further gradations, because in addition to all conceptual work, content must also be created for the first time in the final consequence. There has to be still someone who ultimately sets the individual stones on top of each other alongside the architect who does all the planning. The only difference is that not everyone does everything, but has a special area of responsibility according to their qualifications. In the next section we will increasingly turn to the area of content in marketing, which in the sense of content management in the area of technical communication has so far been neglected by the industry, both conceptually and in terms of content, because marketing is marketing and ticks differently.

1.1 Delivery & Retrieval

Distributing content and finding it again afterwards - presumably no witchcraft, if one understands it as the provision of print document on the company website and a classic full text search to get suggestions. If that were our understanding, the paper would end at this point. But of course we have here another, a new understanding of delivery: As the supply of information and in granularity appropriate to the purpose. The retrieval of course is for better or worse connected to the delivery side, so it is now considered within a section. This new type of delivery and retrieval can only ever be considered within the scope of the authoring page, because the preliminary work in the field of classification there is indispensable for all subsequent steps. The term "Content Delivery" appears above all frequently in the field of technical communication, namely in the form of Content Delivery Portals (CDP) [4]. That makes use of the classified nodes in the editorial system and filter the information on the basis of the classification. Content delivery is a prerequisite for retrieval. Actually a logical step and technologically only moderately spectacular and well conceivable as a tool for e.g. the service technician who is specifically looking for a remedy for a specific fault for a specific device variant. The inclined reader already finds here under indication of the appropriate search terms in the Internet a wealth of information and conversion examples. Basically, a piece of cake, if everything was done correctly on the authoring side up to the creation of the digital information twin.

The delivery side is methodically more interesting and demanding if it is to be used as a sales enablement platform - a kind of retrieval for sales. There is no doubt that there are countless software providers who offer a module similar in one way or another. As I said: From a purely technological point of view it's just a website that I call up and somehow look at information. And that "somehow" is the crux of the matter. Because the introduction to a product or a solution in the B2B area does not follow the same laws as we know it from the consumer area: Much less emotion, but much more facts and tough competition, price pressure and the complexity of the devices. The wealth of knowledge and information that has to be acquired in order to survive in the tough supplier and component market is enormous and a permanent process in sales that is never complete. Do glossy brochures, flyers or huge catalogues then help at this point? They have their justification, but not when it comes to providing information quickly and purposefully. At Bürkert we have therefore gone one step further and developed a "taxonomy of the market" by using methods of classical classification engineering on the one hand and by defining, on the basis of use cases, which information needs arise in which sales step and how this information has to look like on the other hand.

This results in a link between customer and company that cannot be mapped by simple taxonomies. Ontologies and semantic networks therefore play a central role in this context, because they can be used to model classifications that are linked to each other multidimensionally - the result is a network that allows different entries and knows no dead ends. Content can then be attached to each of these individual objects via a Content Management System, for example, but the search then functions via the network. It is easy to imagine, for example, that a certain process such as gas regulation for laser cutting systems always requires certain media inserts, which in turn force certain materials into the products, and so on and so forth. Now, at the latest, it is no longer about delivery, but about the "how" of retrieval and we see that classification is not everything, but without classifications everything is nothing. Depending on my previous knowledge and my interests, I choose the entry level. But here too, the information found is only as valuable as it would serve the purpose for which it was intended. So if you're thinking about a combined search and content delivery platform, you should put the tool selection at the bottom of the list and think carefully in advance about what you're doing it for and for what purpose. If that's clear, the work can begin. However, it is also valid here that documents are no solution and the creators must think and create accordingly granularly and modularly. This may even be the biggest challenge..

Analytics

At this point, no fundamentally new findings are provided, because the Analytics area is a very broad, very well researched area and the experts for this are usually located in the corresponding departments "Online Marketing" or similar. Nevertheless, it should be said at this point that all activities in the area of authoring, delivery and retrieval should be evaluated objectively and therefore measured as well as possible, time and time again. Which modules are viewed how often in my CDP? Which search terms are used in the Sales Enablement Portal and how often does the search run into emptiness? Which paths in the semantic net are followed and which are not? In addition, targeted customer surveys are a valuable source for continuously improving the quality of content, especially when it comes to operating information.

Basic Requirements

The beautiful new world of information, in which it is available on request and helps the user at the specific moment, will only work if all the prerequisites have been met on the linguistic side and are continuously maintained. We are talking here about a uniform language, a corporate terminology that reflects as many facets of the company, its products and its employees as possible. What does something like this look like? In the first step all relevant terms have to be identified and then all synonyms, i.e. terms that denote the same thing. This can only be achieved by implementing terminology circles consisting of people with the necessary knowledge - classically R&D, product management and sales - the mixture makes the difference and in the end there should only be one name for each thing and these names must be accessible to all employees centrally via a terminology database. Terminology management for control purposes is recommended, so that new terms can be coordinated and approved in these circles. Bürkert has even gone so far in this respect that terminology development is an integral part of the product development process. But why is a uniform language so important? That's because of two things: Without a consistent use of terminology, people quickly talk past each other and perhaps create different understandings of one and the same subject matter, which weakens the quality of the content and thus the external image. Secondly, any inconsistency in the source language inevitably affects the target languages, i.e. Translation Management also benefits from these activities. Speaking of Translation Management: Without central coordination of all translation-relevant activities, you quickly run into a confusion of languages and all the work in the source language was in vain. Translation is not the task of the individual specialist departments, but rather an overarching corporate function and central to every company that does not only act locally.

2. Summary and Outlook

The aim of this article was to summarize the activities that the digital revolution has almost forced on a medium-sized company in connection with content. Each aspect discussed, from the conception of the classifications to the requirements for the corresponding software, can and fills entire volumes. For a deepening of the topics a manual research and the constant further training are highest requirement.

What becomes clear, however, is that the upheavals in the world of work and business not only affect the products, but all associated processes, such as authoring, delivery, retrieval and analytics processes.

So what can we expect in the coming years? Nobody can look into the future, but it can not be overseen that data in the form of content are no longer only documents but valuable ressources. Because through them also new business models open up: Ad-hoc provision of maintenance recommendations on the basis of measurement data - previously unimaginable but today already in implementation and in case of doubt a tangible sales argument. If we manage to provide sales-relevant information in the right granularity in the future and avoid page-long PowerPoint presentations, we have gained a lot on the cost side and on the user side. In the long term, we will also completely get away from the document, which automatically forces us to think about the way we consume information - because transferring the print-based world to the digital world does not bring us any advantage. Sure, not all solutions are set: E.g. how are individual pieces of information archived in a legally compliant manner if there is no document left, but rather virtually thousands of product-specific operating information pieces? Nevertheless, the framework conditions are set, certain methods, procedures and tools will prevail and others will quickly disappear into insignificance. But as long as we keep checking our own activities for meaningfulness, it is no disgrace to do one or two more laps in order to be early adopters or even first movers to implement a global Digital Information Service for all content that matters.

.

References

[1] Drewer, P. and Ziegler, W. (2014). Technische Dokumentation. Übersetzungsgerechte Texterstellung und Content-Management. 2nd ed. Würzburg: Vogel Communications Group GmbH & Co. KG.

[2] Parson U., Sapara J., Ziegler W. 2017. iiRDS for Technical Writers - Introduction to the Metadata. In Conference Proceedings of the Annual tekom Conference.

[3] European Association for Technical Communication – tekom e.V. 2017. iiRDS Specification - intelligent information Request and Delivery Standard – First Public Working Draft, 03 April 2017. (2017).

[4] Ziegler W. 2013. Alles muss raus! Content Delivery für Informationsportale, In Proceedings of the annual tekom conference, p. 47-48.

ACKNOWLEDGEMENTS

Special thanks to Prof. Dr. D. Roy, University of Aizu (UoA), responsible for the conference and project organization at UoA.

.

BASICS OF METADATA & APPLICATION IN TECHNICAL DOCUMENTATION

2019 Volume 1

Marie Christin Poggemann
Karlsruhe – University of Applied Sciences
Moltkestraße 30
76133 Karlsruhe
Germany
poma1023@hs-karlsruhe.de

Evelyn Heller
Karlsruhe – University of Applied Sciences
Moltkestraße 30
76133 Karlsruhe
Germany
heev1012@hs-karlsruhe.de

ISSN: 2434-9526

PAGES: 41-46

ACM Chapter Proceedings on Educational Technology, Language & Technical Communication

Basics of Metadata & Application in Technical Documentation

Marie Christin Poggemann
Karlsruhe – University of Applied Sciences
Moltkestraße 30
76133 Karlsruhe
Germany
poma1023@hs-karlsruhe.de

Evelyn Heller
Karlsruhe – University of Applied Sciences
Moltkestraße 30
76133 Karlsruhe
Germany
heev1012@hs-karlsruhe.de

ABSTRACT
Managing numerous product variants, retrieving information easily and rapidly, and handling intelligent content are major challenges in Technical Documentation. Currently, there is only one possibility to meet these requirements: the effective classification of content with metadata using a metadata concept.

Organizing content in Technical Documentation by applying metadata concepts is essential, but it is also a great effort. To be precise, a lot of time is invested in implementing a metadata concept or transferring an existing content structure into a metadata concept. However, if the wrong strategy is chosen, the potential of the content to quickly assemble individualized output or to offer intelligent search and filter functions is left unused. Low rates of content use demotivate companies to invest more effort into Content Management, thus making their organizational processes prone to error. To ensure that the invested time provides a suitable solution, there are several approaches to meet the implementation-specific requirements.

This paper covers the basics of metadata and metadata concepts in terms of development, deployment and presentation based on the current state of knowledge in Technical Documentation. The goal is to give an overview of the current range of possibilities and approaches on the market and clarify the additional benefits of metadata concepts as well as their cases of application. At this point, both bottom-up and top-down approaches are examined. Reference is made to the, in some areas, commonly used metadata concepts in the Technical Documentation: PI-Classification and iiRDS. In addition, an example of a different approach to develop or refine a metadata concept by the tagging service of IntraFind is provided.

CCS Concepts
- **Information systems – Information retrieval – Specialized information retrieval**
- **Information systems – Information retrieval – Document representation – Ontologies**

ACM Chapter Research Seminar on Technical Communication & Information Design 2019, January 28-Feb 1, 2019, Aizuwakamatsu, Japan

ISSN: 2434-9526

Keywords
Content Management System; Content Delivery Portal; Metadata; Metadata concept; Technical Documentation; Intelligent information; IntraFind; PI-Classification; iiRDS

1. INTRODUCTION
The constantly changing market demands and higher customer expectations interplay with the internal possibilities of a company in product development. For example, products need to be more individual and adapted to the specific needs of the customer. To meet this requirement, to address certain target groups, to provide specific publication formats or document types and to differentiate from market competitors, companies expand their product range to numerous variants. But to produce more individualized products, more complex development processes are required [1].

Additionally, the progress in technology and production leads to more complex products. In contrast to this, the time to introduce a product to the market needs to be as short as possible to keep up with the frequent releases of new product versions on the market. These expectations also have to be considered globally, which means that information products have to be created in multiple languages. [2]

These requirements imply that production and documentation have to work more efficiently. Content Management offers the possibility to create documentations in multiple languages and variants in a shorter time by dividing the content of documentations in several content components called modules and reusing them. To find, manage and reuse these modules and even entire documents they hold metadata. Thus, Content Management leads to a quicker creation of documentations that fit to the specific requirements of the customer. [2]

In the course of digitization and Industry 4.0, it is expected that information is delivered dynamically and intelligent. Information is considered to be intelligent when the required information is provided individualized and context-dependent on the preferred device with a didactically appropriate presentation medium. The creation of intelligent information requires the definition, assignment and administration of metadata. [3]

In terms of the role of metadata for search and reuse, metadata can be considered as indispensable key element of Content Management.

2. DEFINITION

Stable operating Content Management Implementations can consist of thousands, or in rare cases millions, of objects to be managed per language [4]. Therefore, usefulness and reusability of the objects depend on the administrative properties of the objects and systems. The motivation for users to create content with a CMS should be that it's easier to find and reuse an object instead of creating a new one. This gives search and reuse mechanisms a central meaning in CMS, as well as their base: the metadata of the objects. [2]

Metadata hold information in form of structured data about objects which can be text modules, images, videos or whole documents. The data indicates state, history of development, validity and deployment of objects. This is very useful, when the aim is to summarize large amounts of data in a structured way. Hence, metadata enable process orientated working and automatization. The administrative data is necessary for searches and the reuse of the content in the CMS. In order to work systematically with metadata, they need to be categorized according to the content of the objects. [5]

2.1 Properties of metadata

When starting to classify metadata it becomes apparent that metadata can have different properties. Looking at the type and use of the content, it needs to be decided what properties the metadata should have.

First of all, metadata can be monovalent or polyvalent. Polyvalent means that metadata can have several values. Monovalent indicates that it can only have one value. Furthermore, metadata can either depend on each other in form of a hierarchy or be independent of each other. When metadata are monovalent and hierarchical, it is called a taxonomy. [2]

Another differentiation is between typified and untypified metadata. Untypified metadata means that the author can indicate free text. Typified metadata provides a predefined set of values, where the author has to choose the fitting. It is strongly recommended not to use untypified metadata because it takes away the option to control what the authors indicate, and the benefit of standardization is lost. Additionally, the opportunity to use metadata for specific searching disappears. Also, the advantage of typified metadata is its' use for controlled terminology. [2]

2.2 Benefits of metadata

In terms of the benefits of metadata there are two main categories. On the one hand, they provide the opportunity of automation, reuse and variant management, which are the focus of CMS.

Processes can be automatized in a CMS with the evaluation of the metadata. In terms of change management, automation is used in applying changes that are made in a module or document to all previous versions. Furthermore, documents can be generated automatically, when the metadata of the modules indicates to which document they belong. [6]

Variant management is used to transform uncontrollable redundancies in non-modular information products with modular reuse in controllable redundancies. A distinction can be made between redundancies on a document, modular and submodular level. Depending on the content that is used, it has to be decided on which level the metadata is needed. [2]

Redundancies between documents are maintained in the configuration management. The variants of information products contain modules of different compositions. The modules are aggregated in the required configuration without considering the module variants. That's why there are many content redundancies between the documents but not necessarily a structural redundancy. [2]

On a modular level, there are different modules in multiple variants. These variants have metadata that indicate to which product variant the module variants belong. To create a document for a product variant, the modules with the metadata for that product variant are chosen and placed together automatically. Without analyzing the modules in depth, there are still uncontrollable redundancies between the modules. [2]

On the submodular level the variants that differ on a sentence or word level are identified and systematized as well. These small information units are called fragments. The metadata of the fragment variants indicate to which variant they belong. Besides the reuse of fragment variants, the reuse of module variants is organized on this level as well. [2]

However, metadata enable precise retrievals and filtration of information units which is the fundamental of Content Delivery Portals (CDP). Their use becomes more and more popular because single topics or information fragments can be accessed directly instead of searching through entire documents. [7]

2.3 Metadata concept

To describe and predefine the obligatory structure and scope of metadata, metadata concepts are used. They define the properties, values and the structure of metadata as well as the metadata categories that should be used. [8]

In the field of Technical Documentation there have been many years of experience in dealing with metadata and specific metadata concepts have been developed [9]. They have a central meaning for the exchange of metadata between different technical systems. If information from different sources should be set in relation with each other so that they can be analyzed together, the entered values have to be linked and compared on the level of their semantic meaning. This implies, that in addition to their actual value, the attribute which is the property that describes the value, has to be indicated as well. [8]

At the beginning of a project, it should be determined whether relevant metadata concepts already exist or are even predefined and should be taken into account. Later, when creating the metadata concept, it must be considered how fine-granular the classification should be. With the granularity of the classification the content components are described more detailed and can be identified more clearly. However, the complexity of the metadata concept increases, and it becomes more laborious to classify content components. [9] Besides the granularity of the classification, the appropriate data format for the metadata value has to be determined. For example, if "ID" has to be an integer or a string.

In a metadata concept it is also defined, if a value has to be set for a metadata or if the author can choose not to set a value for the metadata.

3. DEPLOYMENT

In Technical Documentation metadata are used in the discipline of Content Management. In CMS or CDP metadata are the key element for reuse, search and intelligent information. [9]

In CMS, content is divided in modules to reuse and aggregate them into documents according to products or product variants. To find and identify the content components in the CMS, metadata is allocated to them. This enables to automatize the process of aggregation and publication of documents in the system. Therefore, in the past the creation of documents was the main focus and use of metadata. CMS usually offer users an already predefined amount of metadata, which has established itself in the course of numerous projects. This metadata can be adapted and extended in most CMS [9].

Besides CMS, a new area that focuses on working with the intelligence of information has developed. CDP try to not show the information statically, but to present content to the user depending on the respective context. Additionally, the information should be easily searchable so that the portal delivers relevant results. Thus, the delivery heavily depends on the data quality that is fed into the system and on the metadata, that is directly or indirectly connected with the content. [9]

In a CDP information can be searched via navigation, faceting and full-text search. The assigned metadata can be used to search and filter content. Often, they are implemented as facets which are known from product portals for consumer goods or for product configuration tasks. This type of search is called facetted searching. [7]

The company Schema offers a CDP that differentiates between outer and inner facets. Figure 1 shows the outer facets that can be used to filter different documents in the CDP. Figure 2 demonstrates the inner facets that are used to filter modules within a document.

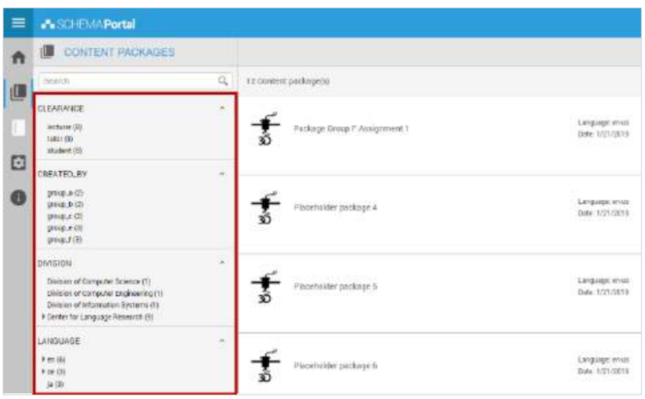


Figure 1. Outer facets that filter documents in the CDP of Schema

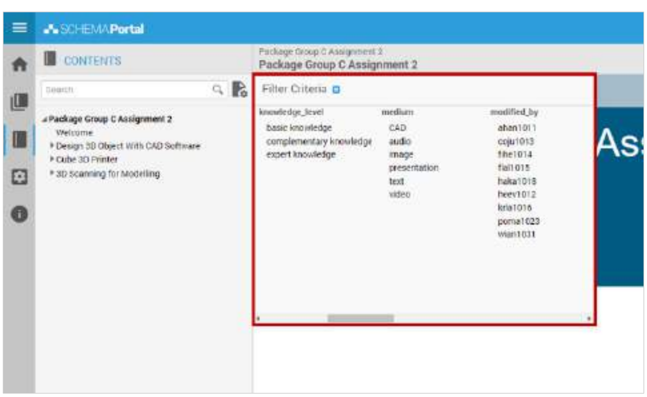


Figure 2. Inner facets that filter modules within a document in the CDP of Schema

There are two different options to use metadata and content in CDP: Metadata and content can be imported from a CMS into a CDP or metadata and content can be manually entered into XML files, transformed and imported into a CDP.

For example, if metadata is manually integrated into the Schema Portal, it can be implemented by using an XML file containing all metadata. Metadata name, value and ID are allocated by attributes. Within the XML file, the tags, containing the metadata information are nested hierarchically, depending on the metadata concept. Tag and attribute names can be defined individually. The prerequisite, however, is that the XML file is well shaped.

4. CREATION

When modelling metadata, there are two possible methods, which can either be conceptualized inductive or deductive. They are called bottom-up-methods (inductive) and top-down-methods (deductive). While there is always one primary method, both, bottom-up and top-down, complete each other and must build a consistent overall concept. [2] In addition, target group, use cases and technical deployment should be considered as well when creating a metadata concept [10].

4.1 Bottom-up-method

Bottom-up-methods are based on the analysis of existing documents and the extraction of metadata [2].

To generate metadata and separate different types of metadata, the classification analysis is used. The metadata found in the existing content must be divided into different classes and subclasses. Afterwards, they are structured according to their relation to each other e. g. hierarchically. To set the criteria for the definition of the different classes, the product and document specific separation of information as well as the normatively predefined structures of documents are considered. [2]

Another approach is the variant analysis to identify properties which indicate differences between content on modular and submodular level. This goes hand in hand with the content standardization and filters the uncontrollable redundancies between content components. It requires an in-depth product knowledge to find technical motivated differences and derivate the corresponding metadata. [2]

4.2 Top-down-method

Top-down-methods use predefined rules for the creation of metadata. These rules can be defined by methods, standards, norms, guidelines or previous bottom-up-processes.

Standards, norms and guidelines are usually valid for specific branches, products or document types. For example, the Machinery Directive defines the structure of an instruction manual. The demand could be metadata, that indicates to which information of the structure a module belongs. Beside official standards of standardization institutes, companies often have own standards for information structuring that can be used for the metadata creation. [2]

Requirements that are based on the company strategies can connect the documentation concept with other information systematics or the product development. An example would be the development-related component-oriented information gathering as part of a product-lifecycle-management-strategy. [2]

In addition, rules for the creation of metadata can be derived using bottom-up methods. The application of these rules equals the top-down procedure.

4.3 Bottom-up with IntraFind

If there is a sufficient number of source texts, potential metadata can be detected bottom-up. The tagging service of the company IntraFind Software AG offers an automatable solution for this.

IntraFind is one of the market leaders for search technologies in Germany, Switzerland and Austria. Since 2000 they have been operating as a provider in the field of enterprise search, information access and text mining. Therefore, they offer a range of software products, solutions and product-independent consultation. [11]

One of their solutions is a tagging service to generate metadata from unstructured text. Information classes, like names of people, locations, companies or products, are the meaningful content building blocks in texts and could be (potential) metadata. [11] Figure 3 illustrates how the key wording of a text works in the tagging service demo of IntraFind.

Figure 3. Example for tagging with the tagging service of IntraFind [12]

Initially, the tagging service recognizes and generates keywords from the text and allocates them to the respective information class. Then, the metadata can be extracted and, for example be reused in CMS or Document Management Systems (DMS). In the following course of normalization, different word forms and spellings of the metadata are identified and harmonized, for example names like “Angela Merkel”, “A. Merkel” and “Frau Merkel”. [13]

By automatically recognizing and categorizing keywords, metadata and metadata classes can be derived. In general, metadata, that is created with the IntraFind tagging service, can be used for the following purposes:

- Enrichment of information
- Improvement of the search
- Classification tasks
- Control of workflows
- Identification compliance-relevant data
- Identification of documents, which are to be transferred to a DMS or archive for audit reasons. [11]

Regarding the metadata concept, the automatic detection of keywords can provide initial impulses in the creation process or help to optimize and expand existing metadata concepts. Due to the elimination of manual key wording, the effort of the employees can be reduced, and valuable time is saved. [13]

Currently, IntraFind works in cooperation with the company intelligent views GmbH (i-views) that is focused on ontologies. They are trying to implement a solution to automatically extend the metadata in the i-views ontology based on the extracted metadata with the tagging service.

4.4 Standards and Methods for Metadata Concepts in Technical Documentation

In the field of Technical Documentation there are several established methods and standards for the classification of information with metadata.

Standards mainly exist in highly regulated industries or product areas. They predefine (nearly) complete metadata models that are used top-down. Certainly, these standards are very tailored to the respective industry. For example, there is a standard (S1000D) for the modular documentation of the operation and service of military aviation products which clearly specifies which metadata and values may be used. With KKS or RDS-PP encryption, there is also a standard for classification in the documentation of power plants. [2]

In addition to the specific standards, there are also several general approaches like the PI-Classification and iiRDS. Both, PI-Classification and iiRDS, are described in the following sections.

4.4.1 PI-Classification

The PI-Classification is a widely used, logical method to classify modular content in order to create, manage and use it in Content Management [14]. The great advantage of PI-Classification is, that it is mostly independent from information models, products and systems. At the same time, the method offers the possibility of individual adaptation. Therefore, it can be used for various products and different industries. [2]

In principle, the method distinguishes between product-related (P) and information-related (I) criteria on one hand, and between intrinsic and extrinsic metadata on the other. [14]

This disposition results in four classes of semantic metadata (see figure 4):

Intrinsic information-related

The intrinsic information class precisely characterizes the information type of an information component. [15]

Intrinsic product-related

The intrinsic product class includes the product components with which the information component is directly related. [15]

Extrinsic information-related

The extrinsic information class describes the assignment of an information component to certain document types and output media. This metadata can also contain an attribution to target groups or markets. [15]

Extrinsic product-related

The extrinsic product class expresses the product groups, products and product configurations of a respective company. This metadata regularly provides a hierarchical representation of all product groups. [15]

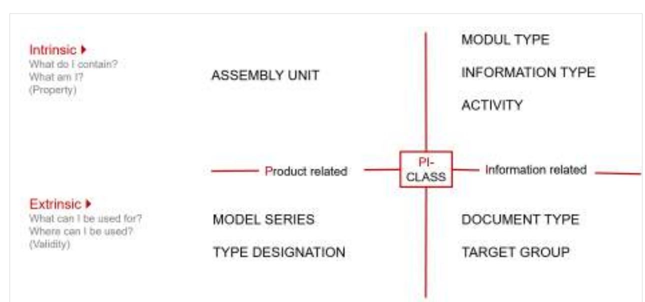


Figure 4. Metadata classes of the PI-Classification

Basically, intrinsic metadata describes the properties and the subject of an information component. The extrinsic metadata, however, contains information about the validity for using the information components. Accordingly, intrinsic metadata form taxonomies that clearly distinguish information units from each other. While extrinsic metadata can be multi-valued in the sense of reuse. [16]

In addition to these four essential classification classes, the PI-Classification can be extended with metadata that describes the life cycle of the information components (e. g. status, version, date of creation) as well as variant properties and functional metadata. [17]

Variant properties are characteristics that are responsible for creating variants and that have not been depicted by extrinsic metadata up to now. Typical examples for variant characteristics are product properties like the material or dimension. It requires in-depth product knowledge and an understanding of modeling to recognize the variant characteristics. Due to the fact, that products are increasingly configured by the customer, variant characteristics are becoming very important. [2]

Functional metadata are very useful for the dynamic aggregation and delivery of information. They offer the possibility to control automation steps between modules and during the publication. Functional metadata can be tool lists, spare parts, maintenance interval and error codes. A typical example for the automation is the generation of maintenance plans based on the modules used in a documentation. With functional metadata logical and physical links can be derived automatically, especially in CDPs. These links seem to be intelligent, but they are generated by the automation and use of predefined metadata. [9]

4.4.2 iiRDS (intelligent information request and delivery standard)

Besides the PI-Classification there is another, new standard called iiRDS which was initiated and developed by the tekom in 2016. [18] The standard is to be used as an exchange and deployment mechanism and consists of a package format for delivery exchange as well as a metadata model that is based on the PI-Classification. The aim of iiRDS is to develop a common standard for the delivery of intelligent information. Therefore, the standard is an important part of the evolution of Technical Documentation towards Industry 4.0. [19]

The iiRDS metadata model is the first standardized vocabulary for Technical Documentation that can be used by several suppliers and devices. To formulate the metadata, RDF, a proven W3C standard is used. iiRDS represents a (first) approach for the metadata standardization in technical documentation and provides the basis for extension of individual metadata concept. [19]

5. PRESENTATION

For creating, understanding and using a metadata concept, it is necessary to organize and present metadata clearly and structured. For this purpose, there are different presentation forms that depend on the application context. Some forms of presentation have the aim to help people, who work with the metadata, to understand the metadata concept. These representations clarify the basic structure, present the different types of metadata with their values and properties (e.g. monovalent/polyvalent) and reveal the relationships and dependencies between the different types of metadata. These presentation forms help to develop and understand structures. For that reason, they are particularly important during the creation of a suitable metadata concept. Popular forms of representation are hierarchical lists, tables or mind maps. Usually, these forms of representation cannot be interpreted by machines.

The choice of the presentation form depends on the application and the aim of the presentation. Hierarchical lists provide an easy to understand and well-structured overview of all metadata and their values. However, they are not suitable for the presentation of properties, detailed information or dependencies. Compared to a list, in a table much more information can be stored. A disadvantage of the tabular presentation is that it can be very confusing and unclear because of numerous values, long columns and rows. Moreover, the (hierarchical) structure is less visible and dependencies can only be visualized laboriously. Accordingly, a table is optimal for managing detailed information but unsuitable for providing a quick overview. A compromise between a list and a table is the mind map. All metadata is displayed hierarchically and arranged clearly and ordered. Furthermore, it is possible to add additional information in form of notes, links, images or icons. Individual nodes or information can be shown or hidden by the viewer as needed.

In addition to the forms of presentation that have been already mentioned, there is the presentation of metadata in the systems that they are used in. The different systems offer a wide range of metadata presentation with various interfaces when creating metadata, assigning metadata (input mask) and using metadata (e. g. search and directories). Typical forms of representation are hierarchical tree structures or taxonomies as shown in figure 5. Depending on the system, these tree structures can be used to assign metadata and/or to search & filter content components. [2]

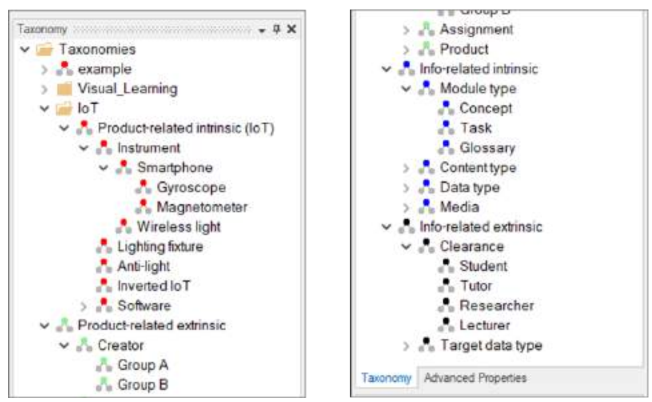


Figure 5. Taxonomies in CMS Schema ST4

The forms of presentation specified above have a fundamental feature in common: They represent the metadata two-dimensionally and hierarchically. Reality, however, is more complex than a simple hierarchical representation. Metadata can not only be hierarchically structured, they can also be in different relationships and dependencies to each other. In order to map these interconnected relationships and dependencies adequately, multidimensional representation forms are required, e. g. network-liked structures. [9]

Most mind map software (e. g. Mindmanager, Free Mind, coggle) offer the possibility to draw connections between the different elements of the mind map. This function is very helpful to make these dependencies clearer, but it also includes the risk of unclear information and complexity. Furthermore, the mind map cannot be interpreted by machines or used for search mechanisms.

When focusing on connections and dependencies, the previously described methods are insufficient. A more convenient form of representation would be a knowledge network in the form of an ontology. Ontology systems such as ontolis or i-views provide the possibility to visualize dependencies and relationships. For Technical Documentation, ontologies offer the additional value to enhance search mechanisms in CDP or to extend metadata modeling in CMS. While ontologies are readable by machines, their complexity makes it difficult for people to understand them. Therefore, ontologies are only partly suitable in helping to understand a metadata concept. [9]

6. CONCLUSION AND PROSPECT

In the past as well as presently, metadata is a determining factor for Technical Documentation. Until now, metadata was mainly used to indicate variants, to manage information components for reuse and to automatically aggregate documents in CMS. With regard to the intelligent delivery of information and CDP, metadata is becoming increasingly important, because they are the key element for the intelligent retrieving and information delivery. In the field of Technical Documentation, there are already approved procedures, methods and standards for the creation of a metadata concepts. Contrary to the adaption to these guidelines, the metadata concepts usually are adapted to the specific needs of a company and their products. This requires detailed knowledge about the company, the products and metadata modeling.

Regarding the future development of metadata, three specific trends are particularly notably: the progression towards intelligent information and iiRDS, the linking of ontologies with search mechanisms and content delivery and the possibility of automated classification of information components via AI.

7. ACKNOWLEDGMENTS

We would like to thank the International Office (AAA) of Karlsruhe – University of Applied Sciences (HSKA), the University of Aizu (UoA), the Faculty Information and Media Management of HSKA and the master program Communication and Media Management of HSKA for enabling us to participate in the ACM Chapter Conference through their financial support. We would like to show our gratitude to Prof. Debopriyo Roy for his efforts and support. We also want to thank Rudger Lörch of IntraFind Software AG for providing material and insights into the functionality and use of the IntraFind tagging service.

8. REFERENCES

[1] Lindemann, U. 2003. *Entwicklung individualisierter Produkte*. In: Zäh, M. F. 2003.

[2] Drewer, P. and Ziegler, W. 2014. *Technische Dokumentation: Übersetzungsgerechte Texterstellung und Content-Management.* Würzburg: Vogel Buchverlag.

[3] tcworld GmbH (ed.). 2019. *Intelligent Information*. [Online]. Available: https://www.technical-communication.org/technical-communication/intelligent-information.html

[4] Straub, D. and Ziegler, W. 2008. *Effizientes Informationsmanagement durch spezielle Content-Management-Systeme.* Stuttgart: tcworld GmbH.

[5] Parson, U. 2018. *Turning technical documentation into intelligent information*. Stuttgart: tcworld GmbH [Online]. Available: http://www.tcworld.info/e-magazine/technical-communication/article/turning-technical-documentation-into-intelligent-information/

[6] Ziegler, W. 2016. *European Perspectives on Technical Communication: Drivers and Concepts of Content Management Systems in the Age of Globalization and Mass Customization*. [Online]. Available: https://www.i4icm.de/fileadmin/content/HSKA/02_Publikationen/ICM/TCSymposium2016English.pdf

[7] Ziegler, W. 2017. *The Evolution of Content Management towards Intelligent Delivery Systems for Technical Communication.* [Online]. Available: https://www.i4icm.de/fileadmin/user_upload/Ziegler_JTCS_2017.pdf

[8] Bibi, D. et al. 2017. *Strukturierung von Metadaten*. Berlin: IANUS - Forschungsdatenzentrum Archäologie & Altertumswissenschaften. [Online]. Available: https://www.ianus-fdz.de/it-empfehlungen/metadaten-struktur

[9] Ziegler, W. 2017. *Metadaten für intelligenten Content*. Stuttgart: tcworld GmbH.

[10] Parson, U. 2017. *Der Kern intelligenter Informationen: die Metadaten.* [Online]. Available: https://intelligent-information.blog/de/blog/2017/07/03/der-kern-intelligenter-informationen-die-metadaten/

[11] IntraFind Software AG (ed.). 2019. *IntraFind Tagging Service for automatic tagging*. München. [Online]. Available: https://www.intrafind.de/products/tagging-service

[12] IntraFind Software AG (ed.). 2019. *IntraFind Tagging Service-Demo*. München. [Online]. Available: http://intrafind.org/demo/taggingservice/.

[13] IntraFind Software AG (ed.). 2013. *Tagging Service*. München. [Online]. Available: https://www.intrafind.de/tl_files/documents/Produkte/INTRAFIND_Datenblatt_Tagging-Service_DE.pdf

[14] Ziegler, W. PI Classification. [Online]. Available: https://www.i4icm.de/en/research-transfer/pi-classification/

[15] Ziegler, W. and Parsen, U. and Sapara, J.2017.*iiRDS for Technical Writers – Introduction to the metadata.* [Online]. Available: https://www.i4icm.de/fileadmin/content/HSKA/03_Vortraege/iiRDS03_iiRDS_for_Technical_Writers_-_Introduction_to_the_Metadata_Ziegler.pdf

[16] Ziegler, W. 2016. *Content management systems tested: Introducing the PI-Fan.* [Online]. Available: http://www.tcworld.info/e-magazine/technical-communication/article/content-management-systems-tested-introducing-the-pi-fan/

[17] Ziegler, W. 2017. *Intelligence? What's Intelligence in TC*?. Stuttgart: tcworld GmbH. [Online]. Available: https://www.i4icm.de/fileadmin/content/HSKA/03_Vortraege/1981_Wie_intelligent_k_nnen_Informationen_denn_noch_werden_.pdf

[18] Hallwachs, J. 2019. *What Is iiRDS, Exactly? A Short Introduction*. [Online]. Available: https://iirds.org/iirds-a-short-introduction/

[19] tekom Working Group Information 4.0. 2017. *iiRDS Specification*. [Online]. Available: https://iirds.tekom.de/fileadmin/iiRDS_specification/20171020-0.9-rfc/index.html

DRIVERS OF DIGITAL INFORMATION SERVICES: INTELLIGENT INFORMATION ARCHITECTURES IN TECHNICAL COMMUNICATION

2019 Volume 1

Prof. Dr. Wolfgang Ziegler
Karlsruhe University
of Applied Sciences
76133 Karlsruhe
GERMANY
wolfgang.ziegler@hs-karlsruhe.de

ISSN: 2434-9526

PAGES: 48-52

ACM Chapter Proceedings on Educational Technology, Language & Technical Communication

Drivers of Digital Information Services: Intelligent Information Architectures in Technical Communication

Wolfgang Ziegler
Karlsruhe University
of Applied Sciences
76133 Karlsruhe, Germany
wolfgang.ziegler@hs-karlsruhe.de

ABSTRACT

Technical communication has recently gained increased attention in the context of industrial digitization. In this article, we provide an introduction into recent concepts of intelligent content with respect to the information systems of content management and content delivery. We describe how different technologies, which can be defined within a content-driven intelligence cascade, support the most relevant use cases for requesting and delivering content in product and information space. These use cases and technologies feature the key concepts of intelligent information architecture and will be – from the viewpoint of technical communication – an important driver for business-relevant digital information services.

CCS Concepts

- **Information systems~Document representation**
- **Information systems~Document topic models**
- **Information systems~Ontologies**
- Computing methodologies~Ontology engineering
- Software and its engineering~Extensible Markup Language (XML)

Keywords

Technical Communication; Technical Documentation; Information Modelling, Information Architecture; Content Management; Component Content Management; Content Delivery; Content Retrieval; Content Classification; Intelligent Information; Digital Information Service; Semantic Metadata; PI-Classification; Ontology; Augmented Intelligence, iiRDS; Artificial Intelligence

1. INTRODUCTION

From the perspective of information management, there is a boost from standardized content creation in technical communication towards more elaborated information modeling approaches and information technologies. Semantic modeling has been used before for content structures, mainly to increase and maintain information quality. Semantic approaches have then been extended towards semantic metadata modeling, leading nowadays to ontology modeling and semantic networks of information [1,2]. These technologies are hereby either used for modeling the complexity of products in combination with content creation processes or for the enhancement of search processes and corresponding increased precision of search results. Finally, applications of artificial intelligence are included in these technologies to enhance search and delivery processes.

In the following, we will summarize the recent situation of content management and delivery from a technical and methodological point of view. Additionally, we will generally discuss the use of these systems for creating and implementing digital information services as use cases of digitization and as business models of content delivery technologies. We will also consider more elaborated aspects like augmented and artificial intelligence that have to be given more consideration in the future of technical communication.

2. CONTENT MANAGEMENT

Content management systems (CMS) build a well-established technological backbone of technical communication. There are almost three decades of history of CMS facing the challenges of product communication in global environments. They have been used to optimize the internal processes of documentation departments with a strong focus on machinery industry and their suppliers as well as related areas like automotive or aviation industries. CMS try to solve and handle the complexity of product documentation arising from product variants, product change and globalization. For this purpose, they rely on methodologies and technologies like topic-based content, content variants, version management and sophisticated functionalities supporting, for example, translation management on granular level and database handling of nested reusable object and their correlated content lifecycle.

2.1 Semantic Content by Information Models

In order to ensure quality and to reduce costs, content is standardized on a linguistic level, e.g. by terminology management, writing guidelines and authoring memory

ACM Chapter Research Seminar on Technical Communication & Information Design 2019, January 28-Feb 1, 2019, Aizuwakamatsu, Japan

ISSN: 2434-9526

functionalities. Moreover, content is structured by using XML-information models and is then published in highly automated processes creating standardized and reliably styled documents. The semantics of XML-based information models and the standardization of these models have been a task of the last decades and led, so far, to different results on a global scale. Especially on the US market, there is a strong tendency towards structuring using the DITA model. Whereas in central Europe, because of the long-term history of CMS, customer tend to use CMS-related XML-structures. Beside semantic content, many companies rely on sophisticated content management processes as variant management and change management depending on product lifecycles or even automated document generation driven by parts lists.

2.2 Semantic Metadata by Classification

Content can become more "intelligent" in order to automatize processes in CMS. For this purpose, technical writers have to enrich topics and other content objects by using semantic metadata. One of the spreading concepts for defining semantic metadata is the method of PI-classification [3,1] and standards derived from this [4,5]. In a basic implementation of this methodology, reusable information units (topics or even content fragments) can be addressed in a virtual information space and can be made up by the dimensions of the so-called intrinsic and extrinsic metadata. They basically describe the uniquely contained (intrinsic) information and the (extrinsic) potential usages of the topics. Both can have P- and I-branches, meaning in the intrinsic dimension, that topics can be assigned uniquely to product components and to information types. In extrinsic dimensions, topics can be assigned to multiple end product types delivered to customers. These products contain the intrinsically defined product components along the physical, functional or software structure. If needed, topics can be assigned extrinsically to the information products (document types) where they should be contained.

Applying this, or comparable classification methods, makes it possible to address topics in a unique way and supports, for example, the mentioned automated document aggregation. This can be achieved by document templates with a placeholder structure of intrinsic classification selectors. Metadata projections in the extrinsic dimension of product space will then populate the template and yield the required documents containing the topics reused from the database. It is important to note that, so far, all classification branches and standard metadata dimensions of the PI-classification are denoted by a company-specific set and usually by hierarchies of product classes, respectively information classes. This is in contrast to more specific industry standards like S1000D or Ispec2200 used in military and aviation industry where complete metadata value sets are predefined.

The PI-classification and correlated approaches intend to provide a concise scheme for writers in relation to the sizes, the contained information and content delimitations of the topics. Using the classification consistently for a manually retrieval of objects within CMS often helps to overcome typical problems of folder-based storing and searching due to sometimes non-deterministic folder structures and the manifold of object deposition. In recent industrial applications there are even more dimensions of an extended PI-classification necessary, describing additional variant properties of products and topics, or, so-called functional metadata. The latter will be relevant to connect topics, for example, to service processes. Considering this, functional metadata can cover concepts like linking content to error codes and system events, to work times and with tools. All these extended metadata might be necessary to describe the handling of complex configurable systems of software and hardware within their life cycle.

3. CONTENT DELIVERY

In most cases, product documentation from CMS has been widely produced in the past – at least for end users – to comply with legal requirements and international standards. Documents have been produced and delivered as classical manuals (information for use) in pdf format or as more or less static electronic documents like online-help formats or linked HTML on storage media like CDs. Even more, the need of technicians for structured, detailed and specific information for installing and maintaining complex products has often been neglected. More generally, the requirements of all kind of target groups and users for an interactive and more situational information delivery have not been met for a long time. This has started to change in recent years. Coming up with modern web technologies and system architectures, content can be delivered increasingly by so-called content delivery portals (CDP) [6]. The basic definition of CDP can be given as "systems, offering web-based access to modular and aggregated content or other information types for various user groups by related retrieval mechanisms".

3.1 Content Access

CDP and users thereof can benefit from the granular topic-based nature of content originating from CMS and also from the assigned metadata. As mentioned, the latter has been used mostly for document creation and retrieval in CMS. Nowadays, three basic access types can be defined:

- structured search by facets and filters corresponding to complex classification taxonomies or more simple metadata sets from CMS
- navigation along classical document structures of nested topics
- direct search based on full-text database indexes by search teams and phrases

The corresponding search functionalities offering, of course, also subsequent mixtures of the basic access types, are often considered as exclusively human driven, i.e. as manually performed access types. But CDP can also be implemented as a middle ware system of web services delivering on request. Using this approach, content can be addressed manually or via web services in many ways and for many use cases. Corresponding applications are even more initiated by recent IoT and Industry 4.0 approaches from engineering, production, product monitoring, and service management. There, content can be delivered to users (technicians or end users) depending on events, actual machine states and operating conditions. It is clear, that in such a case, the identification of the relevant content crucially depends on the concise location of topics in an information space described by standard and extended (PI-)classifications.

Regarding the devices being used for accessing content, all known types of responsive web interfaces or any other mobile apps are available. Limitations apply for offline usage including synchronizing mechanisms due to the limited storage capacity of applications on mobile devices. Therefore, offline capabilities are offered only by certain vendors. Other applications are run due to security reasons only on-site on machines as local services.

3.2 Content Sources and Use Cases

So far, only content originating from CMS has been considered as deliverables from CDP. In this case, systems will have to import topics based on the XML-format and they will be displayed as transformed to the HTML Web format. Web documents made up of nested topics can then be navigated similar to online help information. Intrinsic metadata can then allow for filtering within documents. Other use cases could cover non-granular documents being downloaded as printable deliverables within a low-level delivery platform using mostly extrinsic metadata.

In many industrial use cases, there are other data sources involved containing unstructured documents without explicit metadata. This situation is prevailing, for example, for information from product service departments including field reports and descriptions of error handling by technicians or many other internal bulletin information. The more these types of unstructured documents or records-based sources like technical parts data from engineering are involved, the more direct search capabilities will be needed. On the other hand, technicians can also benefit from the detailed facetted search because of their extensive technical product knowledge. By using the corresponding search facets reflecting the functional, logical or physical structure of products, they can drill down search results and access quickly the required information. End users without deeper product knowledge, on the other side, might have - even in case they will receive structured information from CMS - primarily a tendency towards direct search and also navigational access. Hence, the actually required and future search behavior of CDP users is due to important future research.

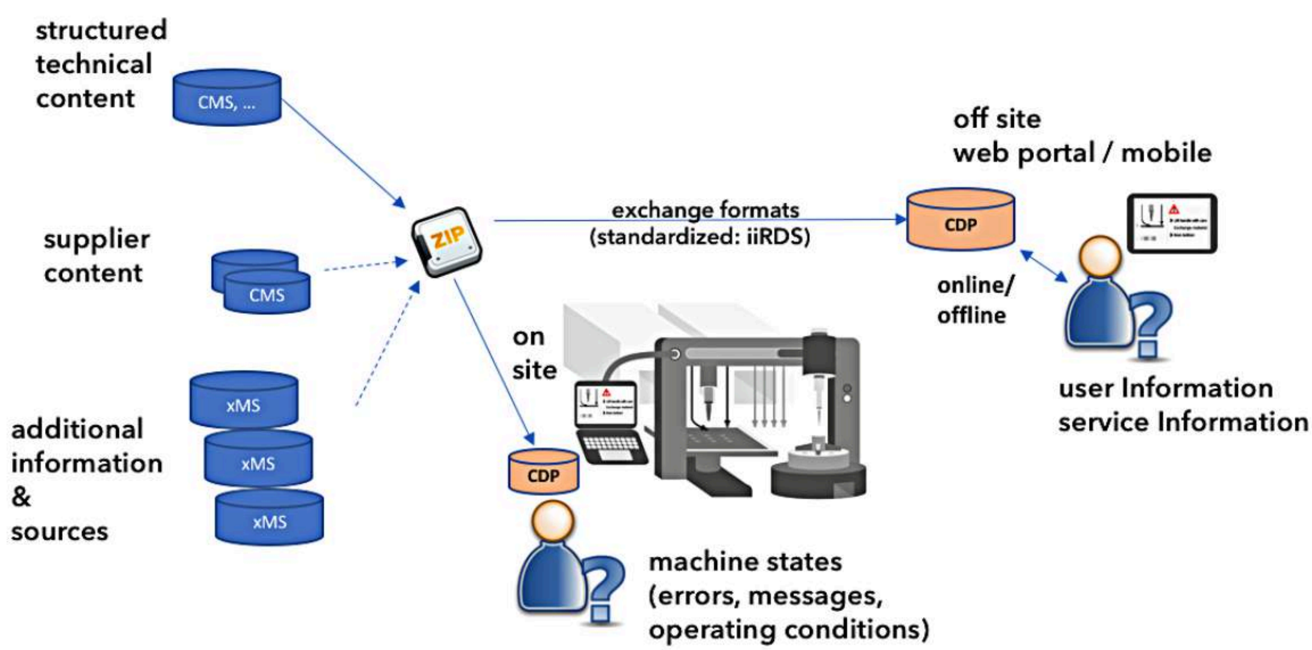


Figure 1. Typical CDP setting including structured and unstructured content sources as well as on-site/off-site access

Another open issue left for the near future is the standardization of the information interchange between source systems and delivery systems. Enterprise search systems usually access databases directly and file repositories and allow users for direct search. In the case of CDP, being based more on predefined structured search facets, the corresponding packaged content and metadata deliverables from CMS have to be transformed and imported according to the receiving CDP specifications. The required transformation processes are costly and sometimes require complex customization efforts. The recently defined iiRDS standard for CDP-ready packaging is addressing this task [5]. It is based on the metadata side as well as on the logics of PI-classification, but predefines a basic set of intrinsic information classes, whereas the product classes can still be defined as product specific taxonomies. Packaging on CMS side, as well as import processes on CDP side, might rely in the future, at least in parts, on this interchange data format.

CDP can belong to different system categories. As explained before, they can arise from CMS and structured content creation domains. On the other side, as long as they deliver content according to the basic definition and corresponding access types, CDP can be seen as special types of, for example, enterprise search systems, document management and enterprise content management systems, web content management systems, service management systems, web shops or any other generalized web portal. These system types vary in the emphasis on search capabilities, display capabilities of content and metadata handling. Within companies, the amount of potentially structured content from technical communication and documentation domains is presumably much smaller than the amount of unstructured content. This means, defining the actual use case for requested, retrieved and delivered content will define, simply put, the weight and importance of direct search processes vs. structured faceted search. Additional decision criteria consist in the mentioned offline capabilities or in the easiness and the standardized data handling for interchange, import or crawling of content objects including various media data. Another specialized use case is driven at the moment by mixed reality applications.

There, object recognition can be performed by a variety of different user interfaces and technologies. In a PI-formulation, the object recognition defines the intrinsically defined components, whereas the specific use case or manual selection by users defines the required information type. This causes a web request to the CDP, which, in turn, will deliver content via web services to the interface. In such cases, the CDP will therefore not be accessed directly, but can be integrated in other environments and services.

4. DIGITAL INFORMATION SERVICES

Content delivery implementations, the types of content access therein and the discussion of content interoperability can be seen as technical aspects of more general business concepts and their recent flow of change. In the era of digitization, business models are shifting from a purely product-centric view to customer- and use-case-centered services connected to products [7]. In manufacturing and heavy machinery industries, the measuring and sensing of machine data become most valuable for machine service models or use-based payment plans. In such an environment, situational and "intelligent" information becomes part of the product value chain and a core task for companies. In this sense, there is an increasing number of companies which connect the information creation to product models and the digital product twin. Hereby, relevant granular information can be pushed in real time to users, operators, analysts or technicians. The triggers for the pushed information delivery are, as already mentioned before technically, special events, recent or changed machine states or operating conditions. Due to sensing and analyzing machine data, preventive maintenance tasks can be announced and coupled to operational or to servicing task content.

In such industrial scenarios, suppliers get increasingly involved in the value chain and will have to deliver components and corresponding granular information topics as an integral product set. Their information has to be integrated from their delivery output into larger delivery environments containing information from multiple source and suppliers. Such a situation is, of course, a well-known situation in the context of military and aviation industries. But for other industrial areas, it is now becoming a growing precondition for suppliers. In addition to this development, the procurement processes will, and already are, changing for component suppliers of soft- and hardware products. These components are increasingly configurable and replaceable by competing components. Therefore, in the future it will be a definite requirement for suppliers to offer information access for selecting the appropriate component from a configurational online platform. Also, to deliver most detailed technical background information to the potential customer, specific for a selected component configuration and relevant for a technical integration. And finally, as already explained, to offer content for integrating into the digital twin or into an integral delivery platform. For this reason, delivery is considered as a digital information service also from a sales point of view. Corresponding delivery projects therefore combine and focus use-case-driven sales information with configurational technical information.

A comparable situation is prevailing for end products in industrial applications or in consumer industry: companies will have to offer additional (digital) services in combination with the product to keep attractivity on the markets, which is one of the basic facts of digitization. Digital information services for product use and servicing will be part of services. A more complex concept of digitization in industry will include the features of measuring, controlling – and – informing. The corresponding digital information services, either as push or as searchable pull implementation, will then also require a highly configuration-specific character. Examples of digitization and digital services can be seen in smart-farming or mobility solutions. In the first case, seeds and pesticides are the physical part of a product, whereas digital services consist of measuring location, condition and the amount of used products for supporting and informing farmers for gaining efficiency. Or, of course, controlling in the future automatizes processes. In the second case, cars and other means of transportation will be the physical product part – more often used and not owned – but the use is being optimized (or controlled) by informing users and by measuring traffic and other conditions. But even if traditional products with lower-level information scenarios will continue to exist for a longer period of time, dynamically improved searchable and situational intelligent content from CDP might already offer more suitable and satisfactory information services than classical documentation.

5. THE INTELLIGENCE CASCADE

The concepts of intelligent content have been heavily discussed in the recent years and they have found their technical use case in content delivery portals and related applications. The business use cases have been focused on in the previous section. It is widely accepted, that "intelligence" in the context of technical communication is mainly based on:

- granular content
- semantically (XML-)structured content
- semantic metadata assigned to content

Of course, all other aspects of technical communication from standardized, comprehensible writing and localization to (web-)design and media integration, are still of basic importance. But from the point of view of content delivery and related information services, the searchability and accessibility in information space as well as the processability by information systems are the most crucial aspects. But recent developments took further steps to guarantee and to improve these key factors. In terms of an intelligence cascade, one can find approaches along the lines of

- native intelligence
- augmented intelligence, and
- artificial intelligence

Native intelligence covers the use and implementation of classified content by metadata in CMS or other authoring systems. The discussed methods, like the PI-classification, lead to a concise schema for developing and using metadata which can be displayed directly as facets or processed in other ways in CDP. As a result, recent CDP can act as digital information services for manual or machine-induced search connected to deep linking or communicating web services.

Augmented intelligence has become of recent interest as ontology modeling of semantic networks. These networks can be used in various ways, for example to:

- derive (CMS-)classification from the logical of physical model of products and support content creation
- extend search and retrieval in CDP using semantically modelled relations between content, i.e. improving quality of search results according to use cases
- implement semantic middleware for managing information from and between various domains or data sources, allowing for mapping mechanisms and search across those domains

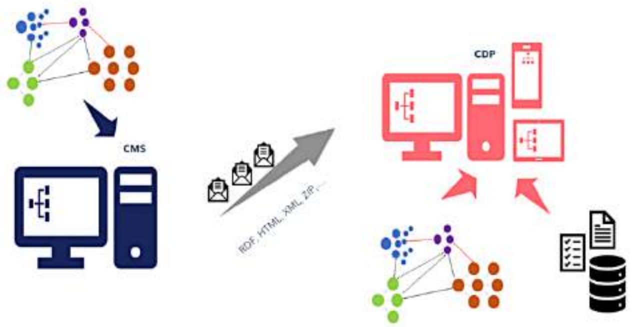


Figure 2. Application areas of ontologies supporting content creation in CMS or content delivery in CDP

The first two cases are sketched in figure 2, implying the support of ontologies for CMS (including hierarchical metadata) or for CDP (including facets). Recent implementations usually focus on one of these sides while there is no general restriction for supporting both sides. A future use cases derived from ontologies will, therefore, be more dynamic or logic driven compositions of semantically linked topics. Information requests can benefit from predefined semantic models of mesoscopic structures which could be seen as micro-documents suitable for specific problems with limited context as connected tasks or workflows or any other product relations. This would also bridge the logical gap between monolithic documents and single topic delivery.

Artificial intelligence (AI) is also diffusing into technical communication and content-related technologies. In the given context, AI can be used for automated content classification of legacy data or of unstructured data [8]. In these cases, one needs a set of training data including high-quality classified and, therefore, native intelligent content. Further applications deal with content generation, i.e. delivering content which is dynamically created from training data corresponding to (CDP-)requests. On the front-end side, one can include AI technologies like object and speech recognition to identify the requested object and information types and to facilitate search processes

The latter is generally true for all levels of the intelligence cascade. Companies and software vendors of information systems will have to rethink their methodologies and architecture towards support of one or more of these levels.

6. SUMMARY AND OUTLOOK

This paper intended to give a foundation for the discussion of content management and content delivery based on recent technologies and information architectures. We have explained the key methods of this content-related information systems and the associated concept of intelligent content including semantic modelling of structured content and classification metadata. On the delivery-side, we were focusing on content access types, content sources in industry and various technical use cases covering standardized data exchange. This led to future business use cases, described as more elaborated digital information services and focusing on the side of the supplier and product manufacturing. But we have shown, that even in lower-level and more recent stages of such information scenarios, intelligent and highly searchable content from CDP can provide improved information services. These services can be offered and used more efficiently when the underlying concepts are addressing the diverse levels of an intelligence cascade. This cascade can be described in terms of native, augmented and artificial intelligence. In actual implementations, all these diverse approaches require, on the conceptional side, a deep knowledge of products, relations, processes and the required volume of information space. This also requires, on the one hand, collaboration and communication between groups in product development and information development. On the other hand, it requires, that academic and professional education of information manager in technical communication will also have to cover these new fields and levels of intelligent information architectures.

7. ACKNOWLEDGEMENTS

Special thanks to Prof. Dr. D. Roy, University of Aizu (UoA), responsible for the conference and project organization at UoA. Participating students of the Karlsruhe University of Applied Sciences, Hochschule Karlsruhe (HSKA) have been financially supported by the international office of HSKA, by the administration of UoA, by the faculty "Information and Media Management" (IMM) at HSKA and by the Master's Program Communication and Media Management (KMM) of IMM. Special thanks also to K. Tsumura and S. Shreshta for additional organizational support.

8. REFERENCES

[1] Ziegler, W. 2017. Metadaten für intelligenten Content. In Schriften zur Technischen Kommunikation, Vol 22, p. 51-66, tcworld

[2] Ziegler, W. 2017. The Evolution of Content Management towards Intelligent Delivery Systems for Technical Communication. In Frontier, Official Journal of Japan Technical Communicators Association JTCA, p. 68-75.

[3] Ziegler W. 2017. Basic Concepts Which Support the Management and Delivery of Intelligent Content. In Frontier, Official Journal of Japan Technical Communicators Association JTCA, p. 84-91

[4] Parson U., Sapara J., Ziegler W. 2017. iiRDS for Technical Writers - Introduction to the Metadata. In Conference Proceedings of the Annual tekom Conference.

[5] European Association for Technical Communication – tekom e.V. 2017. iiRDS Specification - intelligent information Request and Delivery Standard – First Public Working Draft, 03 April 2017. (2017).

[6] Ziegler W. 2013. Alles muss raus! Content Delivery für Informationsportale, In Proceedings of the annual tekom conference, p. 47-48

[7] Ziegler, W. 2018. Content Delivery and Content Analytics for Industrial Applications. The Upcoming Era of Digital Information Services. In Frontier, Official Journal of Japan Technical Communicators Association JTCA.

[8] Oevermann J., Ziegler W. 2016. Automated Intrinsic Text Classification for Component Content Management Applications in Technical Communication. In Proceedings of the ACM Symposium on Document Engineering. p. 95-98.

SECTION 2

EXTENDED ABSTRACTS

ACM Chapter Proceedings on Educational Technology, Language & Technical Communication - 2019

Intelligent Information Architecture of Content Management Systems and Delivery Portals in Technical Communication

Wolfgang Ziegler
Karlsruhe University
of Applied Sciences
76133 Karlsruhe, Germany
wolfgang.ziegler@hs-karlsruhe.de

Content management systems (CMS) build a well-established technological backbone of technical communication. Especially in Central Europe, there is a long history of CMS facing the challenges of product communication in global environments. They have been used to optimize the internal processes of documentation departments, mainly in the machinery industry and suppliers as well as related areas like the automotive or aviation industries. CMS try to solve and handle the complexity of product documentation arising from product variants, product change, and globalization. For this purpose, they rely on methodologies and technologies like topic-based content, content variants, version management and sophisticated functionalities supporting for example translation management. In order to ensure quality and to reduce costs, content is standardized on a linguistic level e.g. by terminology management, writing guidelines and authoring memory functionalities. Moreover, content is structured by using XML-information models and is then published in highly automated processes creating standardized and reliably styled documents. Many companies use even more sophisticated processes to automated document generation driven by parts lists or other product properties from engineering.

In most cases, product documentation from CMS has been produced – at least for end users – to comply with legal requirements and international standards. Documents have been produced and delivered as classical manuals (information for use) in pdf format or as more or less static electronic documents. The needs of users for an interactive and more situational information delivery have been neglected for a long time. This has changed in recent years. Coming up with modern web technologies and system architectures, content can be delivered increasingly by so-called content delivery portals (CDP). They can make use of the granular topic-based nature of content from CMS and also from metadata, which has been used before only for document creation and can now be used for faceted search processes. But this, content can be addressed manually or via web services in many ways and for many use cases. Corresponding applications are even more initiated by recent IoT/Industry 4.0 approaches from engineering, production, product monitoring, and service management. There, content can be delivered to users (technicians or end users) depending on actual machine states and operating conditions. Therefore, content development and provisioning from TC have developed towards a key component of modern product architectures and correlated digital (information) services.

From the perspective of information management, there is subsequently a boost towards more elaborated information modeling approaches and information technologies. Semantic modeling used only for content structures is extended towards semantic metadata modeling, ontology modeling, and semantic networks. These technologies are either used for modeling the complexity of products in combination with content creation or for the enhancement of search processes and corresponding increased precision of search results. Finally, applications of artificial intelligence are included in these technologies. They can be used to support automated content classification or to enhance search and delivery processes. The concept of content delivery will, therefore, include technologies like object and speech recognition to facilitate search, communication with CDP and navigation in product and information space.

3D Printing Architecture: Cities that Empower

Professor Dr. Ana Cristina García-Luna Romero
Department of Architecture
University of Monterrey
Mexico
ana.garcialuna@udem.edu

Keywords: design, architecture, digital technologies, 3D printing, cities.

Digital technologies provide enticing opportunities to increase accessibility, maximize public standing, decrease costs, and preserve heritage in new, more affordable ways.

The field of architectural design, practice, fabrication, and construction is increasingly becoming aided and dependent on digital technology. The proliferation of computers in design education and practice has resulted in a major paradigm shift and a reorientation in theoretical and conceptual assumptions considered to be central to traditional design education and practice.

Information technology has become ever more pervasive in architectural education and practice and has revolutionized the way we design, practice, evaluate, teach and produce architecture. Digital technology has reconditioned the design process and how we operate as architects. The introduction of new computational tools to architectural design pushed the limits of conventional design models and methods. From inception to the representation of designs, and from production to life-cycle management of buildings, architects had to develop exceptional approaches.

The digital age has also radically reconfigured the relationship between design and production, creating a direct digital connection between what can be imagined and designed and what can be built through "file-to factory" processes of computer numerically controlled (CNC) fabrication (Kolarevic 2003). The ability to generate construction information directly from design information is one aspect of digital architecture. Architectural design information is slowly becoming building information particularly with the emergence of Virtual Building solutions, and Building Information Modeling (BIM) in particular. In fact, BIM implementation could streamline the design and construction processes, and eventually may lead to the re-establishment of the architect as the master builder.

Digital technology also radically changed the way we teach and learn architecture (Gross and Do 1999, Al-Qawasmi 2005). New computerized studios such as the paperless studio and the virtual design studio have been introduced in many architectural schools as new ways of practicing and teaching architectural design.

It is evident that digital media has fundamentally changed the way we design, practice, and product architecture. These changes have given rise to a discourse and debate on the relationship between digital technology and architecture. Despite the extensive literature on the subject, the impact of digital technology on how we design, practice, teach, fabricate and produce architecture has not been sufficiently examined.

It is proposed to reflect on the forward-thinking nature of the use of 3D printing in construction from 3 specific processes of application:

1) Robotic arm extrudes,
2) Sand layers linked together,
3) Metal for solid structures.

This debate is argued by recognizing that digital technologies as digital fabrication are ushering in the third era of construction technology. Prior to the industrial revolution, hand-production methods were abundant and craft defined everything. The craftsman had an almost phenomenological knowledge of materials and intuited how to vary their properties according to their structural and environmental characteristics.

The coming of the industrial revolution saw the triumph of the machine over the hand and the machine was used to standardize everything. Now, however, digital technologies such as additive manufacturing allow craft and industry to merge.

The question is, which technologies are best suited to architecture? Further, can 3D printing help solve the housing crisis?

References.

KOLAREVIC, B. 2003, Architecture in the Digital Age: Design and Manufacturing, Taylor and Francis, UK.
GROSS, M. D., & DO, E. Y. 1999, Integrating Digital Media in Design Studio: Six Paradigms, In Conference Proceedings of the Association of Collegiate Schools of Architecture, Minneapolis, pp. 144-148.
AL-QAWASMI, J. 2005, "Digital Media in Architectural design education: reflections on the e-studio pedagogy," Art, Design & Communication in Higher Education, 4(3), pp. 205-222.

Biomedical Engineering & Healthcare

Professor Dr. Hiram Alejandro Cantú Campos
Department of Biomedical Engineering
University of Monterrey
MEXICO
hiram.cantu@udem.edu

Biomedical engineering is a discipline seeking to reduce the gap between Engineering and Medicine in order to provide solutions to healthcare problems. Along the different specialties areas of Biomedical Engineering, there is a field known as Rehabilitation Engineering that involves the design, development, and implementation of solutions and devices for assisting people with disabilities, promoting inclusion and allowing disabled people to have a more independent life.

Emerging technologies have affected the development of the different disciplines, in a positive way in most of the cases, and Rehabilitation Engineering is not the exemption. An example of these emerging technologies is three-dimensional (3D) printing, an additive manufacturing process able to fabricate three dimensional and solid objects of the desired geometry. 3D printing is considered a convenient technique to manufacture devices capable of assist and/or replace body functions.

Additive manufacturing, such as 3D printing, combined with mechanics and/or electronics, core subjects of a Biomedical Engineering program, permits the creation of more robust and sophisticated rehabilitation devices. Such is the case of different innovative projects that have been developed in the undergraduate program by the students of the Biomedical Engineering program at UDEM: a myoelectric-controlled hand prosthesis was fabricated by 3D printing with the ability to react to muscle signals of the amputated arm; a scanning navigation system was designed, 3D printed, and implemented into a wheelchair for children with cerebral palsy in order to give them autonomy while moving; a wristband was developed and 3D printed to recollect frequency ranges of tremor in patients with Parkinson's disease in order to better control the dose of dopaminergic medication.

The development of these types of projects at the undergraduate level has been a unique and enriching experience for the students at UDEM. It allows them to understand the actual needs of the society and, more importantly, it engages them to focus their knowledge and work in the improvement of the quality of life of others.

Rule-Based Writing - Technical Documentation in the English Language for Engineers

Michael Tewes
Applied German Linguistics and Intercultural Communication
Dean, Faculty of Information Management and Media
Karlsruhe University of Applied Sciences
GERMANY
Michael.Tewes@hs-karlsruhe.de

Engineers and authors of technical communication all over the world have to write standardized English texts, although they are not native speakers of English. This trend is constantly increasing due to the progressive globalization of services, trade, and industry. For cost reasons, however, native speakers are rarely available in companies to proofread English texts.

Until a study by tekom, the European Association for Technical Communication, in 2013, little was known about the needs of authors in writing English texts and the tools they use to write them. Available aids such as John Kohl's "Global English Style Guide" are aimed at native speakers of English. They do not take into account the specific problems of non-native speakers.

As a result of the tekom study, a publication was produced which specifically combined the experience of the 190 authors participating in the study with the scientific findings of linguists. In addition, software producers for language technologies were involved. The result of these efforts is the tekom guideline "Ruled-Based Writing – English for Non-Native Writers" (2013).

The presentation aims to show how this guideline provides engineers and technical writers with a set of rules to help them improve the quality of English technical documentation texts. The rules are available on the linguistic levels of the text (32 rules), the syntax (sentence, 57 rules) and the lexicon (word, 30 rules). On the basis of the rules and the corresponding practical examples, it will be demonstrated how these rules can be applied in everyday business life and how they can increase the quality standard of the texts. In addition, the possibilities offered by the guideline for developing one's own writing rules would be considered.

The aim of these efforts is to continuously improve the quality of English texts in a company, to introduce “Ruled-Based Writing” for all authors and to implement a Style Guide for quality assurance.

Multimedia Knowledge Transfer based on 3D Models used in Augmented Reality, Virtual Reality and for the Creation of Instructional Videos and Supplemented with Dynamic, Multilingual, and Auditory Information

Martin Schober
Professor of Internet and Media Technology
Karlsruhe University of Applied Sciences
Germany
Martin.Schober@hs-karlsruhe.de

3D models are ideally suited for creating learning applications, training applications and technical documentation with new trend-setting technologies such as virtual reality, augmented reality and videos. For didactic reasons, these applications should be interactive, multimedia and multilingual.

Augmented Reality and Virtual Reality are technologies to convey knowledge and to support complex actions visually and dynamically. This results in interesting perspectives for technical documentation, e-learning, and service. Which knowledge and tools are needed to create augmented reality and virtual reality knowledge media? Which hardware and which software tools are available?

Augmented reality is the enrichment of the real world with computer-generated additional objects. The existing reality is supplemented by virtual objects. These can be of a visual, auditory and haptic nature. An everyday example: In a football broadcast, offside lines and the goal distance of a free kick are digitally inserted into the broadcast for the television viewer. Today, glasses are used in service and maintenance, which, for example, display information about a fault in a machine via Bluetooth. The spare part to be used can then be organized immediately. The glasses also use GPS to detect when the fitter is standing in front of the machine and display the troubleshooting instructions.

In contrast to augmented reality, virtual reality is the representation and simultaneous perception of artificial reality and its physical properties in a real-time computer-generated, interactive virtual environment. The real environment is excluded. In the technical environment, this technology can be used to walk through large facilities without danger or, for example, to save the construction of prototypes by using virtual 3D models in development. Augmented reality" means that in a device (mobile devices, AR glasses, wearables) the current image of the environment, supplemented by artificial extensions, can be seen. These artificial extensions can be media such as texts, images, films or 3D models. The real image can either be transferred to a screen via a camera and supplemented with media, or the media can be projected onto the real image that is seen through a glass pane, as is already the case in many state-of-the-art motor vehicles.

One characteristic that distinguishes AR from VR and must be taken into account is its portability. With Virtual Reality, the user is bound to a place in the real world; the application only works if the user is at a certain place, e.g. if the sensors for tracking are fixed at a place. With Augmented Reality, it is the user's intention to move around in his environment, to walk around and to have much greater freedom of movement. Nevertheless, the user must be at the point where the AR system's task can be fulfilled.

Dynamic multilingualism and text to speech in HTML5 documents and Adobe Captivate applications provide a cognitive advantage.

Texts can be dynamically exchanged from XML files using AJAX (Asynchronous JavaScript And XML) and JavaScript in HTML5 and Adobe Captivate. In addition, they can be read out directly in multiple languages using appropriate TextToSpeech software. It has been scientifically proven that it is advantageous to present additional information on images and films not as text, but auditory. Since the visual sensory channel is already occupied with images, it is better to record additional information in an audio format so that there is no overload, e.g. by watching a film at the same time and reading explanatory texts that are faded in. Some multimedia learning theories describe that visual and auditory information is processed in two different subsystems of working memory. By using the auditory subsystem, the visual subsystem can be relieved and learning performance improved.

Practical approaches towards steady information workflows: From Marketing Automation to Content Configuration and the connected sub-processes

Mr. Dominik Lächler
Head, Department of Bürkert Information Management (bIM)
Bürkert Fluid Control Systems
GERMANY

Regardless of what the actual terms mean, it is undisputed that change processes are currently running which are changing the work itself, and above all, the way we actually work.
With the increased demands, which affect the development of new products in terms of connectivity and networking with other products towards the "digital factory", not only the products themselves and how they are developed change, but also the information belonging to the product must keep pace in this context to keep up with more and more complex products and devices.
The realization that the future not only belongs to the individual product but also to the specific content and not to the static document itself is only gradually gaining ground in classical mechanical engineering companies: What people actually want today is not "Illustration shows special equipment" but "Illustration shows what you actually ordered". So what companies are struggling with is how to implement it and how to set up the infrastructure to really make a difference in terms of information in the context of digitalization. So they face two major challenges here: What kind of digital content do customers need and how can we generate this kind of content as efficient and automated as possible? Is it only a matter of publishing a PDF as online help and thus only changing the format, but not the content, or is it a matter of fundamentally realigning communication with the customer, which produces completely new ways of thinking, processes and business models and thus breaking with previous publication structures and channels.
The respective decision is important and effects tools, content and the necessary classifications.
It is clear that individual departments of a company cannot cope with these challenges on their own and so it's the management of a company that has to define it as a part of its company strategy. This highlights the importance to think out of the box – within the management and the departments. This is what Bürkert has recognized and therefore has begun to set the course for the information of the future, which will meet the increased demands on information, the delivery, and retrieval of it. The boundaries between the individual departments are becoming increasingly blurred because the creation of information has to be seen as a process and not as a departmental task: this affects the technical editing as well as the marketing of a company as the main producers of information today.
Bürkert in particular as a manufacturer of products with a high degree of variance faces particular challenges: On the one hand, the classical approach to the structuring and classification of content in the form of product and information-related metadata is not sufficient if the appropriate product information for the product is to be generated ad-hoc. Therefore, the evaluation of technical features of a product has to be considered what has not been the focus of most editors and goes far beyond the work of the ordinary technical editor. From pure Content Management, the way is being prepared to Content Configuration in Content Management System (CMS) On the other hand; it is an immense effort to deliver sales information that really supports the sales department in their daily work by providing the right information at the right time in a language that the user understands. Building up ontologies to build intelligent searches to deliver precise content through sharp classifications is what Bürkert is striving for at the moment - it's Content Delivery.
Additionally, a consistent language in source and target languages delivers the base for all multi-language content-related activities at Bürkert. A state-of-the-art Authoring Assistant Tool and a Translation Memory System (TMS) form the necessary IT backbone.

Technical Communication from the Perspective of Cognitive Psychology - Including Technical Communication Comprehensive Program

Manabu Kishi
Director, Japan Technical Communication Association: JTCA
Professor Emeritus, Tokyo Gakugei University
JAPAN
manabu@u-gakugei.ac.jp

Technical Communication from the view of Cognitive Psychology is to explain it from the following four points of view:

The Vertical Axis: it tries to grasp your comprehensive ability, or not to do.

The Horizontal Axis: it tries to inform you of all the knowledge, while carefully selecting.

The points of the above views are as follows:
1. Accuracy and Understandability are different.
2. Changes of receivers' minds are important.

These points can be concretized in the Technical Communication as follows: "Technical communicators convey necessary technical, products' information including services to the receivers (readers, listeners) accurately and understandably in order to raise the receivers' understandability and satisfaction of achievement."

The Technical Communication Comprehensive Program is the one to authorize a person acquired basic knowledge and ability being necessary for Technical Communication as a Completed Graduate.

This program is aiming at making studying contents at universities useful in the business world.

Four Missions about Re-design for Technical Communication

Satoshi Kuroda
Councilor, Chairperson of Planning Committee of Public Activities, JTCA
Guest Associate Professor
Center for the Study of Co* Design
Osaka University
satoshi.kuroda@ise.co.jp

Technical Communication has been continually changing for its 60 years of development while always responding to the social requirements and to the changes in the industries. We can look down over these changes every 10 years. Now, we face the changing epoch where the Technology of Technical Communication has been rapidly changing to cope with the social changes toward the aging society and the technical changes based on the AI (Artificial Intelligence) and the IoT (Internet of Things).

Technical communicators have been trying to enlarge the applicable targets of the Technical Communication and to change the Output Package Form from Document to Information; for this purpose, we Technical Communicators have been actively developing New Elemental Technologies while looking for the optimal combinations between the knowledge and technology, etc.

Such changes prove that the technologies of Technical Communication would continually shoulder the social missions in the future as well and they would continually contribute to the compatibility of heightening users' value of experiences and users' protections; they show the future possibility as ones of practical business affairs.

About the Editor:

Prof. Dr. Debopriyo Roy is a professor of technical Communication and usability, teaching specialized English language courses to computer science majors at the Center for Language Research – University of Aizu (UoA), Japan. He has over 20 years of teaching and research experience in technical documentation, UX Design, usability testing methods, computer-assisted language teaching (CALL), and task-based language teaching (TBLT). Prof. Roy is the director of the CLR Technical Communication Laboratory; supervises the JTCA technical communication certificate program at the UoA; supervises the academic exchange initiative between the UoA and the Karlsruhe University of Applied Sciences (HSKA), Germany; and assists the Silicon Valley Internship Program at the UoA.

He teaches advanced graduate and undergraduate courses in technical writing, interface design, and technical documentation procedures besides other writing courses. His research projects are focused on language studies based on usability testing methods, technical procedures and manuals design, technical presentations, and interface design and business communication strategies. Prof. Roy has published in premier journals and conference proceedings on topics related to technical communication and usability in computer assisted language learning and were actively involved with the IEEE Professional Communication Society and its Japan Chapter as the chair. He is also an active board member of the ACM Professional Chapter on ELearning and Technical Communication and organizes international conferences and partnerships.

About the ACM Chapter on ELearning and Technical Communication:

This chapter is broadly focused on the use of technical documentation and use of communication patterns for e-learning purposes. Much of the research discussions related to this chapter is associated with the use, design and testing of software and other interfaces for e-learning perspectives. Further, research projects in technical documentation emphasize the user-friendly application of text and graphics for commercial products.

A major focus of this chapter is also to organize international conferences in educational technology, information design, language studies and technical communication, thereby building a community of academic and industry professionals who are interested in cross-disciplinary and cross-cultural initiatives related to the above-mentioned fields of study.

ACM CHAPTER PROCEEDINGS ON EDUCATIONAL TECHNOLOGY, LANGUAGE AND TECHNICAL COMMUNICATION

https://www.acmchapterseminarjapan.com/acmchapterproceedings

SUPPORTED BY

ISSN: 2434-9526