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.

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

ISSN: 2434-9526

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 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.

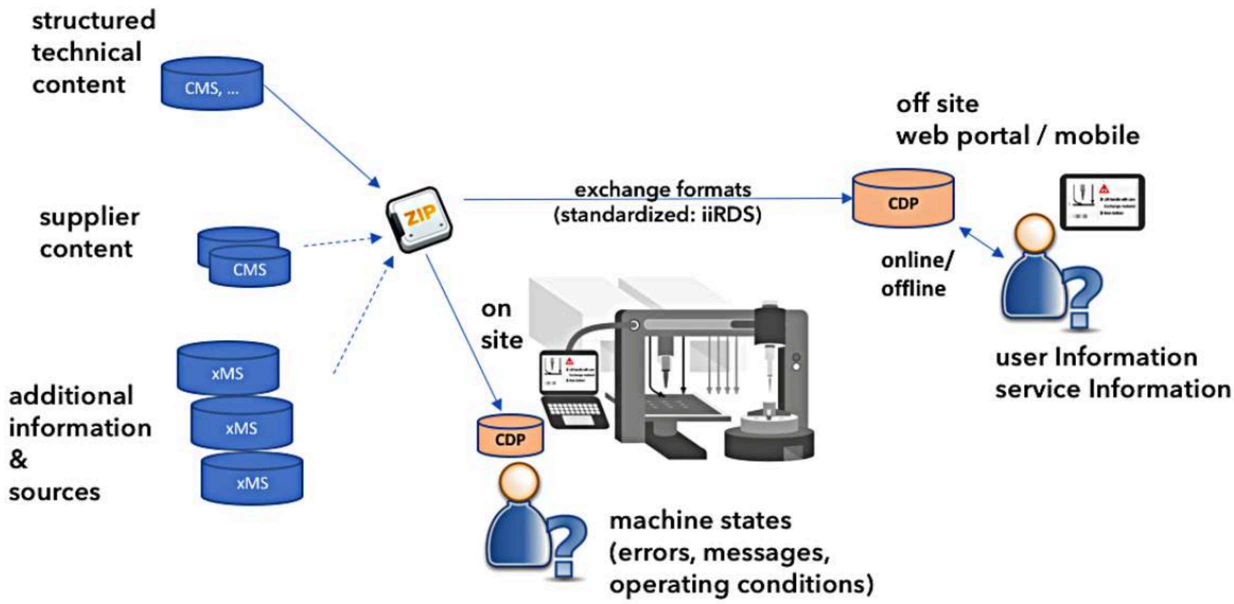

Figure 1. Typical CDP setting including structured and unstructured content sources as well as on-site/off-site access

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

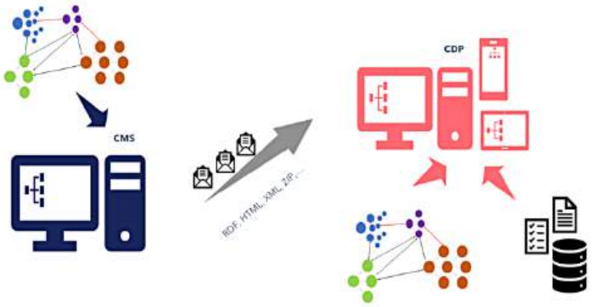

Figure 2. Application areas of ontologies supporting content creation in CMS or content delivery in CDP

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.