

企業からの配信情報の管理と配信を支援する基本コンセプト

Basic Concepts Which Support the Management and Delivery of Intelligent Content

By Prof. Dr. Wolfgang Ziegler
Karlsruhe University of Applied Sciences, Germany

コンテンツマネージメントシステム（Content Management Systems - CMS）を首尾よく実装して使用するためには、テクニカルコミュニケーションの分野で知られている基本的な概念とプロセスを理解しなければならない。最も重要な概念の1つは、バリエーション管理である。モジュラーコンテンツの分類、再利用、組み立てに基づき、プロセスの自動化と動的配信に本来のインテリジェンスを提供する。

In order to successfully implement and use content management systems, one has to understand the underlying concepts and processes that prevail in technical communication. One of the most important concepts is variant management. It relies on the classification, reuse and assembly of modular content and provides native intelligence for automation processes and dynamic delivery.

構造化オーサリングと IT システムベースのコンテンツマネージメントは、テクニカルコミュニケーションの標準ツールセットとなっている。もちろん、グローバルな視点から見ると、個々の国や地域で使用の程度は非常に異なる。さらに、これらの技術のユーザーグループは、外部委託するサービスの程度や、企業内のテクニカルコミュニケーションの組織的責任の範囲がどのようになっているかによって異なる。

しかし、そこには学術教育の基礎とビジネス専門家の訓練のための基礎を構築するための共通の理論的側面がある。

基本的な方法

技術製品の情報開発は、製品開発および生産と同様の規則に従う。具体的には、コンテンツの標準化とモジュラー化、およびドキュメントの体系的な構造化のためのルールとプロセスが存在する。テクニカルコミュニケーションとそれに対応するコンテンツ作成プロセスに関して、これはライターが、たとえば、言語理論や認知理論から導き出されたいくつかのオーサリング規則に従わなければならないことと同様のことを意味する。これらのルールは、使用者の理解を容易にし、明確かつ概略的に導かれたコンテンツの構造化に基づき、ライターが標準化文書を作成するうえでのサポートをすることを目的にしている。これは、多くの場合、すべてのタイプの製品情報の一貫性を保証する企業のターミノロジーをベースにした場合に継続される。機械産業の場合、ターミノロジーの一貫性は、販売前資料からユーザーマニュアル、サービス情報、および予備部品カタログに至る広い範囲で実行される。

ほとんどの場合、言語の標準化は、XML データ形式とコンテンツ構造に対応するルール定義を行う際に同時に行われる。これらの構造は、通常、拡張可能なスキーマ定義（Extensible Schema Definition - XSD）またはドキュメント型定義（Document Type Definition - DTD）を使用する XML ベースの情報モデルによって与えられ、実施される。意味のモデル化は、構造要素の意味を著者に明示し、構造化されたコンテンツを系統的に理解できるようにする。

前提条件情報、行動、結果、技術データ、必要なツールと部品、保守間隔情報、エラーメッセージ、ソフトウェアメニューなどを含む意味構造と（XML）要素が頻繁に使用される。これらの意味構造は、段落、リスト、表などの要素からなる汎用モデル化と対比させることができる。意味のモデル化は、より一般的な要素を指定する XML 属性を使用して行うこともできる。意味のモデル化では、意味要素が文脈上の重要性に応じて具体的にフォーマット化されるクロスメディアハフティングプロセスなどのより特定のデータ処理が可能になる。出力メディアをフォーマット化の際の重要な役割のほかに、関連コンテンツの自動リンクや情報収集の生成などのコンテンツ駆動型プロセスを始動することができる。たとえば、機械製品のためのエラーメッセージ、ツールリストおよび部品リストまたは自動的に生成されるサービス計画情報の収集などが含まれる。最後に、意味のモデル化はコンテンツの品質を促進することができる。言語的記述ガイドラインは、別個の意味モデル化の場合に対応する構要素に適用することができる。これは、古典的な再検討プロセスによって、または特殊な制言

確認ソフトウェアによる自動確認によって行うことができる。したがって、意味情報モデルは、より多くのコンテンツ品質と一貫性を高めることができる。

要約すると、意味要素と属性によって強化された情報モデルは、洗練されたコンテンツ処理に使用することができるインテリジェントコンテンツへの重要な第一歩を創出する。CMS アプリケーションでは、一般に3つの基本的なタイプの情報モデルを選択できる：公開（DITA、S1000D、iSpec2200、PI-Mod など）、カスタム、ベンダーベースおよびシステムベース。それらは、意味の程度と、モジュラーベースの記述をサポートする能力が異なる。

CMS は、これらのタイプのサポートに関し大きな違いを示している。これは、多くのシステムが標準モデルを排他的に実装しているのに対し、他のシステムは独自の自己開発モデルしかサポートしていないことを意味している。しかし、カスタム情報モデルに必要なモデルを自由に統合できるシステムは少ない。最近においては、後者は、必然的な統合と保守コストの関係で CMS 開発の初期段階にあったほど一般的ではなくなっている。これらの考察の結果、どんな特別なタイプの情報モデルの事前選択でも、システム実装を可能にする CMS に影響を及ぼす。

メタデータと PI 分類

生来のインテリジェンスを CMS データに追加する重要な方法は、メタデータにある。メタデータは、追加情報を使用してモジュラーまたはドキュメントベースのコンテンツを充実することができる。これは、CMS 内、デリバリープロセス内の CMS 外でのプロセスの取得および再利用に使用される。理論的なメタデータのアプローチとその適用性を説明するためには、ほとんどの CMS 実装に異なる範囲で適用される PI 分類スキーマを使用する。

このスキーマ内では、基本的に4つの基本クラスのメタデータにモジュラーコンテンツを割り当てることができる。これらのクラスは、製品 (P) および情報 (I) に関して編成されている：

- 内的な製品クラス
情報が直接接続されている物理的または仮想的な製品コンポーネントを記述する。複雑な製品の場合、彼らは大きな分類法を構築することができる。
- 内的な情報クラス
これらは、モジュラートピックに含まれる情報タイプを正確に定義する。通常、手続き型、記述型、概念的、安全性の情報、およびすべての

対応するサブクラスのようなクラスの分類法を構築する。

- 外的な製品クラス
それらは、CMS を使用することを検討している会社の個々の製品によって与えられる。それらは通常、すべての対象製品グループの階層を構築する。
- 外的な情報クラス
関連する文書タイプと出力メディアを記述するが、ターゲットグループや市場などのプロパティも対象にすることができる。

内的なクラスと外的なクラスの違いを理解することが重要である。内的なクラスは、固有のモジュールを定義するための概念を反映する。したがって、内的な値は、分類法から単一の値として各コンテンツモジュールに対して選択されなければならない。定義上、同じ内的クラスを持つモジュールは、バリエーションでなければならない。ここで留意すべきは、モジュラーコンテンツのために頻繁に使用される用語「トピック」は、情報クラスの内的固有の側面を反映するだけである。

外的なクラスに使用されるメタデータリストおよび階層は、複数のコンテンツの複数の再利用を反映する複数值になる。実際のアプリケーションでは、これらのクラスのすべてが必要なわけではない。しかし、一般的に、それらは基本的なメタデータディメンションの一般的なセットを形成する。コンテンツオブジェクトの言語は、内的情報プロパティであると理解することができる。しかし、組織上の理由から、言語は通常、翻訳プロセスにリンクされた別のメタデータ次元として保持される。

PI 分類の4つのよく知られた基本的な次元のほかに、この方法は追加のメタデータを考慮する。このようなメタデータのセットは、コンテンツおよび製品のバリエーションのさらに多くのパラメーターを特徴付けるためにより重要になってきている。バリエーション管理についてさらに詳しく説明する。

製品と情報の管理

テクニカルコミュニケーションの管理において今日企業が直面している課題には、複数の原因と側面がある。それらを簡単に説明する：

- 変更管理
製品開発は、オーサリングプロセス中に多くの変化が生じる高いダイナミクスを示す。製品の配送後も、製品の変更や更新がある。最近のアジャイルソフトウェア開発は、たとえ製品機能の時間差開発を可能にする。
- メディア管理

ウェブやアプリベースの情報など、複数のメディアやフォーマットには成果物が必要である。

- バリエーション管理
製品の複雑さと非常に精巧な構成の管理は、使用者の個別のニーズに基づいている。
- 翻訳管理
グローバル化された市場と製品には、費用対効果の高い翻訳、複数言語の提供と市場の固有の特殊性に適したコンテンツが必要である。
- 工程管理と品質保証
すべての成果物は、内部の工程プロセス毎の品質管理システムに基づき、内容および構造の正確さを検証しなければならない。責任とタイムラインは明確で拘束力のあるものでなければならない。プロセスの有効性は全体的な目標である。

これらの要件のために、CMS は対応する概念と機能を提供する。

- バージョン管理により、単一のモジュールコンテンツまたは文書のすべての改訂を追跡することができる。
- クロスメディアパブリッシング、マルチメディア統合、および配信プラットフォームへのインターフェイス。この中にはさまざまな出力形式のメディアオブジェクトの管理が含まれる。必要に応じて、書式設定を自動化することができる。手動で修正する必要はない。
- 特定の製品構成に応じたフィルタリングや自動集計に基づく複雑なコンテンツアセンブリ機構によるドキュメントの作成。
- 翻訳管理機能は、モジュール単位で翻訳量を分析および削減する。CMS の参照資料は、翻訳者を支援し、容易な作業ができるようにする。
- 使用者の権利と役割の管理は、ワークフローおよび状態管理と組み合わせることができる。ターミノロジー管理、検討、制限言語チェックツール、およびオーサリングメモリー機能を介して、コンテンツの言語品質管理を行うことができる。

バリエーション管理の詳細

インテリジェント情報およびインテリジェントシステムの関係では、上記のバリエーション管理は重要な役割を果たす。機械産業における製品バリエーション管理は、定義された構造に従って最終製品に構成要素を構成することからなる。これらの構造は、エンタープライズリソースプランニング (Enterprise Resource Planning - ERP) または製品ライフサイクル管理 (Product Lifecycle Management - PLM) システムにおいて、組み込ま

れた状態で、または、動的に維持された構造で管理される。ソフトウェア業界では、機能的なソフトウェアモデルまたはソフトウェア仕様を利用することができる。

CMS は、上記で定義したメタデータスキーマを利用することができるようになった。ソフトウェアおよびハードウェアコンポーネントは、内的製品クラスにマッピングされる。通常、ドキュメントに関連するコンポーネントは、製品全体のサブセットのみで構成されている。各コンポーネントには、個別の情報クラスに起因するモジュラーコンテンツの情報セットがある。バリエーション管理は、2 つの主要な機能によって実現することができる。フィルタリングあるいはドキュメントの生成によってである。

最初のケースでは、ドキュメントのツリー構造はモジュールの参照とネストによってマスタードキュメントとして手動で構築される。モジュールまたはそのサブモジュール部分が外的製品クラスを有する場合、特定の製品に応じてマスタードキュメントを容易にフィルタリングすることができる。したがって、構成はモジュラーコンテンツの有効性 (外的クラス) を使用して間接的に取得される。たとえば、DITA マップの概念は、このタイプのバリエーション機構に属する。もちろん、このジェネリックフィルタリングには他のすべての外的クラスを使用することができる。

第2のケースでは、特定の内的クラスを運ぶネストされ、シーケンスされたオブジェクトのプレースホルダー構造として参照構造を事前定義することができる。システムは、特定の製品および関連するモジュラートピックのすべて (外的および内的に) を選択し、それらを自動的に組み合わせる。

どちらの場合も、コンポーネントのフィルタリングが可能である。これは通常、構成管理が複雑であるため、実装されていない。ほとんどの場合、モジュールを明示的に分類すると必要な結果が得られる。副作用として、これは通常、個別の製品グループのために、個別のマスタードキュメントまたは参照構造を明示する。

上述したように、追加されたバリエーション特性によってさらに複雑さが生じる。それらは、追加パラメーター (たとえば、制御タイプ、材料、コーディング、電圧およびプラグ) または部品番号によって与えられる製品のバリエーションを記述することができる。この意味で、追加のフィルターとして機能することができる。どちらの場合も、バリエーション特性だけが異なるモジュラートピックのセットがある。

リイ リジェンスと配信

CMSは、インテリジェントで自動化された方法でドキュメントを作成するための精巧な機構を持つことができる。本来のインテリジェンスは、簡潔なメタデータスキーマに由来し、この機構は、たとえこれが明示的に命名されていなくても、基礎となる (PI-) 分類機能の観点から理解することができる。

インテリジェントデリバリーでは、CMSの機構に基づいてドキュメントの作成とプロビジョニングを行うことができる。動的コンテンツデリバリーの最新の議論では、さらに進歩を進めたいと考えている。使用事例に応じてモジュラートピックを要求または取得し、状況に応じて表示することを目的としている。関連メディアは、あらゆる種類のウェブベースのオンラインポータルまたはモバイルアプリを対象としている。そのようなアプリケーションでは、特定の情報の検索にはコンテンツの分類が不可欠であることは明らかである。

しかし、使用者は依然として従来のドキュメント構造を使用したいと考えている。したがって、コンテンツデリバリーポータル (Content Delivery Portal - CDP) は、CMSの本来のインテリジェンスのすべての要素を必要とする。それは、メタデータの分類と分類されたモジュールのセットを含めたドキュメントのロジックを表示するための参照構造を表示するための参照構造である。インテリジェントシステムは、少なくとも CMS および関連システムに由来する構造化コンテンツに対して、これらの基本概念を将来使用することになる。

訳者: (株) 情報システムエンジニアリング (ISE)

Structured authoring and IT system based content management have become a standard toolset in technical communication. Of course, from a global perspective, they are used to very different extents in the individual countries or regions. In addition, the user groups of these technologies vary depending on the degree of services being outsourced or on the organizational responsibilities for technical communication within the companies.

However, there are some common theoretical aspects that build the basis for academic education and for the training of business professionals.

Basic methodologies

Information development for technical products follows similar rules as product development and production. More specifically, there are rules and processes for the standardization and modularization of content and for the systematic assembly of documents.

In regard to technical communication and the corresponding content creation processes, this means, that writers must follow several authoring rules derived for example from linguistic or cognitive theories. These rules are intended for the creating of standardized documents that facilitate user understanding and support writers via a clear and schematically guided structure of content. This is continued in many cases by a corporate terminology base which ensures consistency in all types of product information. In the case of the machinery industry, term consistency is carried out ideally ranging from the pre-sales materials to the user manuals, service information, and spare part catalogues.

Linguistic standardization is in most cases accompanied technically by using the XML data format and corresponding rule definitions for content structures. These structures are usually given and

enforced by XML-based information models that use extensible schema definitions (XSD) or document type definitions (DTD). The information models implemented in CMS can differ significantly in the degree of semantic structures. Semantic modeling reveals the meaning of structure elements explicitly to authors and makes the structured content machine-readable.

There are frequently used semantic structures and (XML-)elements containing for example prerequisite information, actions, results, technical data, required tools and parts, maintenance-interval information, error messages, software menus and furthermore. These semantic structures can be contrasted to a generic modeling consisting of elements like paragraphs, lists, tables. Semantic modeling can also be done by using XML-attributes that specify more generic elements. It is obvious, that semantic modeling allows for more specific data processing like for example cross media publishing processes where semantic elements can be specifically formatted depending on their contextual importance. Beside their important role in formatting output media they can trigger content driven processes like automated linking to related content or the generation of information collections. Examples thereof include collections of error messages, of tools lists and parts lists or automatically generated service plans for heavy machinery products.

Finally, semantic modeling can facilitate the quality assurance of content. Linguistic writing guidelines can be more easily applied to the corresponding structure elements in the case of distinct semantic modelling. This can be done by classical review processes or by automated checking through specialized controlled language checking software. Therefore, semantic information models can enhance even more content quality and consistency.

In summary, information models, enriched by semantic elements and attributes are an important means for creating a first step towards intelligent content which can be used for sophisticated content processing. In CMS applications, there are in general three basic types of information models which can be selected: public (like DITA, S1000D, iSpec2200, PI-Mod), custom, or vendor- and system-based. They differ in the degree of semantics and their capability to support module based writing.

CMS shows large differences in the support of one of these types. This means that a number of systems implement exclusively standard models, while others support only their own self-developed models. But only a few systems are free to integrate any model, which would be necessary for custom information models. The latter is recently not so common as it used to be in the early days of CMS development due to inevitable integration and maintenance costs. As a result of these considerations, a pre-selection of any special type of information model will influence the CMS set being relevant for possible system implementation.

Metadata and PI-classification

An important way to add more native intelligence to CMS data stems from metadata. Metadata can enrich modular or document-based content via additional information. This is used for the retrieval and reuse processes within CMS or subsequently outside CMS within the delivery processes. To gain a theoretical approach to metadata and to explain their applicability, we will utilize the PI-classification schema which can be applied to different extents to most CMS implementations.

Within this schema, modular content can be assigned in principle to four basic classes of metadata. These classes are organized with respect to products (P) and

to information (I):

- **intrinsic product classes**
They state the physical or virtual product components the information is directly connected to. For complex products, they can build a large taxonomy.
- **intrinsic information classes**
They define precisely the information type contained in the modular topic. Usually they build a taxonomy of classes like procedural, descriptive, conceptual, safety information and all corresponding sub classes.
- **extrinsic product classes**
They are given by the individual products of the company which is considering using CMS. They usually build a hierarchy of all the considered product groups.
- **extrinsic information classes**
They describe the relevant document types and output media but can also cover properties like target groups or markets.

It is important to understand the differences between the intrinsic and extrinsic classes. Intrinsic classes reflect the core concept for defining unique modules. Therefore, intrinsic values have to be selected for each content module as a single value from a taxonomy. Per definition, modules which have the same intrinsic classes, have to be variants. Note that the heavily used term “topic” for modular content just reflects the intrinsic and therefore unique aspects of the information classes.

The metadata lists and hierarchies used for extrinsic classes, will be multi-valued which reflects the multiple reuse of content. In practical applications, sometimes not all of these classes are needed. But in general, they form a general set of basic metadata dimensions. The language of content objects can be

understood as being an intrinsic information property. But for organizational reasons, the language is usually kept as being a separate metadata dimension that is linked to the translation processes.

Beside the four well known basic dimensions of the PI-classification, this method considers additional metadata. Such sets of metadata are becoming increasingly important for characterizing even more parameters for content and product variants. We will come back to this when we describe variant management in some more detail.

Managing products and information

The challenges companies are facing today in managing technical communication have multiple causes and aspects. They can be briefly described in terms of

● **Change management**

Product development shows high dynamics which many changes during the authoring processes. Even after product delivery, there are product changes and updates. Recent agile software developments can generate for example time shifted developments in product functions.

● **Media management**

Deliverables are required for multiple media and formats, including web- and app-based information.

● **Variant management**

Product complexity and highly elaborated configuration management accompanies the expectations of users for most specific information.

● **Translation management**

Globalized markets and products require cost-efficient translations, and the delivery of multiple languages, and market specific content adaptations.

● **Process management and quality assurance**

All deliverables have to undergo internal release

cycles and quality control for content and structural correctness. Responsibilities and timelines have to be clear and binding. Process effectivity is an overall goal.

For these requirements, CMS offer corresponding concepts and features:

- Version management to keep track of all revisions of every single modular content or of documents.
- Cross media publishing, multimedia integration and interfaces to delivery platforms. This includes management of media objects for different output formats. If needed, formatting can be automated and requires no manual corrections.
- Document creation via complex content assembly mechanisms like filtering and automated aggregation in accordance with specific product configurations.
- Translation management functionalities analyze and reduce the translation volume on a modular basis. Reference material from CMS supports translators to facilitate their work.
- Users' rights and roles management can be combined with workflow and status control. Linguistic quality control of content can take place via terminology controls, reviews, controlled language checking tools and authoring memory functionalities. Reuse efficiency of modular content can be measured and tracked.

Variant management in more detail

In the context of intelligent information and intelligent systems, the above mentioned variant management plays a crucial role. Product variant management in the machinery industry consists of configuring components into end-products according to defined structures. These structures are managed as-built or even dynamically as as-maintained structures in

enterprise resource planning (ERP) or product lifecycle management (PLM) systems. In the software industry, one can make use of functional software models or software specifications.

CMS can now make use of the above defined metadata schema. Software and hardware components are mapped to intrinsic product classes. Usually, the components relevant for documentation consist of only a subset of the entire product. For each component, there is an information set of modular content, attributed by distinct information classes. Variant management can now be achieved by two major mechanisms: filtering or the generating of documents.

In the first case, document trees are built manually as master documents by the referencing and nesting of modules. If modules or submodular parts thereof carry extrinsic product classes one can easily filter master documents according to specific products. The configuration is obtained therefore indirectly using the validities (extrinsic classes) of modular content. For example, the concept of DITA maps belongs to this type of variant mechanism. Of course, all other extrinsic classes can be used for this generic filtering.

In the second case, one can pre-define a reference structure as a placeholder structure for nested and sequenced objects which carry specific intrinsic classes. The system selects them for a certain product and all (extrinsically and intrinsically) of the relevant modular topics, and assembles them automatically.

In both cases, a filtering of components would be possible. This is usually not implemented because of the complexity of configuration management. In most cases, the explicit classification of modules yields the required results. As a side effect, this usually defines the separate master documents or reference structures for the distinct product groups.

As stated above, an even higher complexity arises by

additional variant characteristics. They can describe product variants given by additional parameters (e.g. control types, materials, coatings, voltages and plugs) or even parts numbers. In this sense, they can act as additional filters. In both cases, there can be a set of modular topics which differ only in variant characteristics.

More intelligence and delivery

CMS can have elaborated mechanisms for creating documents in an intelligent and automated way. The native intelligence stems from a concise metadata schema, and the mechanisms can be understood in terms of an underlying (PI-) classification scheme even if this is not explicitly named.

Intelligent delivery can cover the creation and provisioning of documents based on CMS mechanisms. In the most recent discussion of dynamic content delivery one wants to proceed one step further. One aims to request or retrieve modular topics and display them situationally, depending on the use cases. Relevant media cover all types of web-based online portals or mobile apps. It is clear, that for such applications, the classification of content is indispensable for the retrieval of specific information. But users also still want to use traditional document structures. Content delivery portals (CDP) therefore need all of the ingredients of native intelligence from CMS: reference structures for displaying document logics, the set of classified modules together with the taxonomies of metadata. Intelligent systems will make future use of these basic concepts at least for structured content that originates from CMS and related systems.