

技術情報のためのコンテンツマネジメント技術の深化

The Evolution of Content Management towards Intelligent Delivery Systems for Technical Communication

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

コンテンツデリバリーポータル(Content Delivery Portal - CDP)は、コンテンツマネジメントの概念を動的かつ状況に応じた情報アクセスに拡張する。CDPは、ウェブ上の技術情報を粒度の細かいレベルで検索するために、さまざまなタイプの検索戦略をサポートしている。システムのインテリジェンスは、本来、拡張または人工知能の根本的な方法論とは異なる。

Content delivery portals (CDP) extend the concepts of content management towards a dynamical and situational information access. CDP support different types of search strategies in order to retrieve technical information on the web on a deep granular level. The intelligence of systems differs in the underlying methodologies of native, augmented or artificial intelligence.

過去数十年間に、コンテンツマネジメントシステム (Content Management Systems - CMS) と、より一般的な構造化オーサリング技術が、コンテンツ作成プロセスを最適化するために世界中の多くの地域で導入された。これと対応するコンテンツ再利用シナリオを介して、文書管理者はいくつかの目標を目指した:

- コンテンツ作成コストの削減
- 翻訳コストの削減
- コンテンツと文書の品質改善と品質保証
- 顧客が必要とする情報、または企業間のコミュニケーションに必要な情報の適時の作成および納品

対応するシステムに関わる技術には、ほとんどの場合 XML ベースのデータ形式が含まれる。より正確には、意味構造の程度が異なる XML 情報アーキテクチャーによって与えられた規則に従ってコンテンツが作成される。このコンテキストでは、テクニカルコミュニケーションの分野は、複数のレベルでコンテンツを標準化するための先進的な領域となった。これは、用語や文体のための言語規則、標準化された意味論的に構造化されたデータ形式、または企業の出版のためのレイアウト規則を対象にする。

しかし、上記の CMS のアプローチと原動力のほとんどは、基本的に情報の内部および製品中心の視点に焦点が当てられている。これは、作成された文書のコスト効率と法令遵守が、システム導入の成功の定義にとって最も重要であることを意味する。

コンテンツに関する新たな視点

最近の技術開発と、ユーザーの実際のニーズに適したより正確なコンテンツへの期待の高まりは、コンテンツ作成、コンテンツ処理、および情報提供に関する視点を変えている。さらに、情報の取り扱い、人間、機械、ソフトウェアのインタラクションや環境を考慮する新しい技術によってますます影響を受けている。さらに具体的には、情報に関する新しい視点のための重要な原動力は、その結果としてのコンテンツマネジメントに関するお互いに相互依存しているさまざまな技術とアプリケーションに見出すことができる。

- 民間の分野で使用されるモバイルデバイスや対応するアプリであり、結果として企業でも使用が求められている
- コンシューマー製品のためのファセットとフィルタリングされた検索を含む Google および他の検索アプリケーションからの検索駆動型情報アクセス
- 追加情報によって視覚環境を豊かにする仮想現実 (Virtual Reality) と拡張現実 (Augmented Reality)
- 民間および企業の状況における機械翻訳とその受け入れの拡大
- さまざまなデータソースからの意思決定に関連する情報と知識の自動抽出を可能にする人工知能技術
- 販売プロセスや機械の状態の追跡など、顧客の行動から生成された大量データを分析するための大規模なデータ技術
- 世界の IoT (Internet of Things) とインダストリー 4.0 のイニシアチブが製品のリアルタイム動作を制御して対応することを目指している。これらは、製品のライフサイクル内のさまざまな段

階で適用し、たとえばプロセスの生産、使用、維持において適用可能である。

テクニカルコミュニケーション分野における最近の認識は、多くの企業が依然としてかなり保守的で、ともかく静的である。つまり、文書中心であり、紙ベースであるということである。それにもかかわらず、さまざまな分野において、上記の新しい技術の動向に沿って製品情報が変化してきた。

公開の場でよく議論されているアプローチの1つは、「インテリジェントコンテンツ（知的コンテンツ）」という概念である。コンテンツやその他の種類の情報の処理には、ITアプリケーションとそれに対応するインターフェイスと基礎となる特殊なプロセスが必要なため、一般的な用語「インテリジェントシステム」も使われる。ここで取り組んでいるシステムタイプは、「コンテンツデリバリーポータル」(Content Delivery Portal - CDP)として知られている。これらは、使用者にコンテンツを動的に配信するために使用されているし、情報の量とメディアは実際の要件に適合している。時には、これは使用者自身が情報の今後の必要性を知る前でさえも情報提供される。

コンテンツデリバリーポータルの定義

CDPは、技術的には、検索機能を介して、さまざまなユーザーグループのために、モジュールや集約されたコンテンツまたは他の情報へのウェブをベースとしたアクセスを提供するシステムとして定義される。

これには、他のシステムからコンテンツにアクセスしたり、コンテンツをインポートしたり、コンテンツのライフサイクル内で管理および更新するための基本的な機能が含まれる。第1の機能は、ポータル内の関連するデータソースおよび対応するシステムに由来する技術的な配信および統合プロセスであると理解することができる。第2の機能は、使用者によるコンテンツ検索のためのユーザーインターフェイスを含む検索機能を提供する。さらに、CDPは通常、コンテンツ要求やその他のイベントを配信するウェブサービスを提供する。これらの機能は、通常、コンテンツ配信の文字どおりのコアとして理解されている。それにもかかわらず、実際のアプリケーションでは、上述のコンテンツ統合プロセスは、コンテンツを使用者に純粋に配信することよりも、技術的にはるかに洗練されている。

さまざまな種類の検索

コンテンツ配信のさまざまなアプローチを理解するには、まずCDPで使用される3つの基本的な

検索方法を特定すると役に立つ。

● 構造によるナビゲーション:

マニュアルベースの情報の伝統的な電子表現は、PDFまたはHTML文書である。ブックマークや文書ロジックを反映するツリー構造などのナビゲーション構造が含まれている。すべてのオンラインヘルプメディア、電子書籍、および最近のウェブアプリケーションは、伝統的な章やネストされたトピック理念をサポートするナビゲーション機能を提供する。

● メタデータによるファセット検索:

単一のトピックまたは情報断片に直接アクセスするには、メタデータを利用する。それらは、CDP内のすべての種類の情報オブジェクトを充実させ、必要なサブセットにフィルタリングすることを許す。それらは、コンシューマー製品や製品構成タスクのための製品ポータルから知られているように、しばしばファセットとして実装される。

● 直接検索:

最初の2つの戦略は構造化された検索と見なすことができるが、直接検索では通常、製品の構造やメタデータに関する知識はほとんど必要ない。直接検索は、通常、検索された情報の関連性を表示するため、ランキングアルゴリズムにより最大量の検索結果をもたらす。

このような独特の検索戦略のセットとは別に、使用者が一連の連続した方法を必要とする状況がある。たとえば、ファセットは直接検索の結果の数を絞り込むために使用されるか、関連のないセクションやトピックを非表示にすることにより、Webドキュメントの量を減らす。

異なる種類のインテリジェンス

前述の検索戦略は、異なるシステムタイプによって特定のレベルでサポートされている。これらのレベルと結果を理解するために、ここではインテリジェントコンテンツとインテリジェントシステムの前述の概念を参照する。より詳細には、以下のタイプのインテリジェンスを区別する。

● 本来的なインテリジェンス:

この概念は、ライターによる周知の意味論的構造化を対象にする。コンテンツ構造およびコンテンツは、アクセス可能であり、主にXML構造化および対応する情報モデルによってシステムの理解が可能となる。

さらに、すべてのメタデータは、インテリジェンスへの本来の貢献として理解できる。文書構造でさえ、低レベルのメタデータ記述、入れ子構造、モジュールトピックのつながりなど、これらの意味に反映している。高レベルの本来のインテリジ

エンスの例として、PI 分類方法のような方法を使用して、製品部品および情報分類に関するトピックの有効性を体系的に表現することができる。この意味で、製品および対応する情報オブジェクトを分類するために使用されるあらゆるタイプのインダストリー分類は、固有のインテリジェンスのアプローチに属する。

● 拡張インテリジェンス:

多くの場合、メタデータの実質的な集合およびコンテンツオブジェクト間の関係は、データソースによって提供される文書またはトピックに含まれない。したがって、特別なシステムは、コンテンツを「外部」メタデータスキームと、そのデータがオーサリングプロセスと独立にリンクされるように充実させることができる。これは、コンテンツ配信アプリケーションの内部または外部で実行される。後者の場合、特殊なメタデータプラットフォームが関与する。上位レベルのアプローチでは、コンテンツオブジェクト間の関係のモデリングを重要視する。

● 人工知能 (Artificial Intelligence)

テクニカルコミュニケーションに使用される人工知能の技術は、統計的または言語的分析を利用して、コンテンツから知識およびメタデータを抽出しようと試みる。つまり、自動的にコンテンツオブジェクトにメタデータを割り当てようとする。割り当てられたメタデータスキームは、拡張の性質を持つことができる。

データとシステムの生来のインテリジェンスがナビゲーションとファセット検索を直接サポートすることは明らかである。拡張されたインテリジェントアプローチもまたこれらの検索方法に貢献するが、メタデータの豊富化のため、人手による労力を必要とする。メタデータの深さとロジックが使用者の期待と知識に対応していない場合、生来の分類学的メタデータで問題が発生することがある。これは特に、企業とその製品の使用者との間の用語の不統一の場合に当てはまる。このような問題は、使用者が使用するさまざまな用語を企業が利用する内部用語にマッピングできる CDP の用語拡張によって対処できる。これはすでに拡張インテリジェンスアプローチの一部である。配信された情報の異なるターゲットグループ間の分類法または存在論的マッピングについても同じことが言える。最近の CMS の開発では、存在論ベースのトピック関係のモデリングを通じてメタデータ機能を拡張しようとしている。しかし、この存在論的開発は製品のモデリングの洞察に関して要求が非常に高いので、近い将来、存在論的モデリングを CMS の標準機能として使用することはありそうにない。

人工知能は、たとえば、テキスト分類を自動化することによって、拡張インテリジェンス概念に基づく人手を減らすために使用されることになるかもしれない。モデル化されていない隠れた関係も明らかにすることができる。これは、非 CMS ソースからの非構造化コンテンツの表示およびそのアクセスに最も関連するものとなる。もちろん、これは、インテリジェントデリバリーシステムにおける直接検索機構の能力を改善し、コンテンツオブジェクト間の自動リンクを導き出すことを可能にする。

コンテンツ関連性分析とコンテンツインテリジェンス

Amazon の販売ポータルのような、コンシューマー製品業界の製品間の自動リンクが知られている。彼らは分析に際し、典型的な消費者行動とリンク情報を活用する。テクニカルコミュニケーションでは、少なくとも 2 つの分析手法が「コンテンツインテリジェンス」という用語の中に含まれることがある。

第 1 の方法は、CMS 内のコンテンツ再利用の効率を分析し、コンテンツ作成プロセスを最適化するために使用される。そこでの分析は、モジュラーコンテンツの再利用率を明らかにし、たとえば、ドキュメントのコストモデルからの情報を引き出す。第 2 の方法はデリバリーシステムに焦点を当て、システムへの使用者アクセスを追跡することによってコンテンツの関連性を測定しようとする試みである。技術的には、通常のウェブ解析は CDP に適用できるが、測定対象のプロパティに適合させる必要がある。典型的な分析プロパティは、たとえば、最も関連性の高い(または無関係な)トピックのアクセス頻度、または使用者の閲覧時間およびナビゲーション行動を測定する。検索用語を分析することで、使用者と情報ニーズのより現実的なイメージを得ることができる。特に、顧客または他の特別なターゲットグループによって使用される上記用語の用語ベースを分析することができ、それによって検索プロセスをさらに改善するのに役立つ。

さらに、結果をもたらさない検索プロセスは、必要とされる情報の潜在的な欠如を明らかにする。したがって、コンテンツ関連性分析は、今後、技術情報の全体的な品質とアクセシビリティを向上させるのに役立つ。CDP 使用者統計から得られたフィードバックを使用することにより、コンテンツ管理の実装の成功に関する事前の疑問に対しても答えられる。

結論

この記事では、動的な方法で使用者にコンテンツを提示するインテリジェントデリバリーシステムの概念を紹介した。使用者は、ナビゲーション、ファセッティング、直接検索、またはそれらの組み合わせの基本的な検索戦略によってコンテンツにアクセスすることができる。たとえば、サービス技術者、エンドユーザーまたは販売プロセスのためのポータルのような実用的なアプリケーションに応じて、前述の検索戦略はさまざまな重要性を有する。

CDPは、オンライン情報の複雑な構造とパッケージングを考慮した交換および更新メカニズムを介して、コンテンツ管理システムと従来のポータルアプリケーションとの間の溝を埋める。この目的のために、コンテンツとメタデータのための標準的な交換フォーマットを開発し実装することが必要である。さらに、CDPには、人工知能からさまざまなソースの非構造化データにアクセスする方法がいくつか含まれている。

CDPはIoT（インターネットオブシングス）やIndustry 4.0シナリオで重要な役割を果たすかもしれない。そこで、情報は機械とコンポーネントに接続され、機械の状態や特定のイベントに反応して表示され、交換される。このようなシナリオでは、CDPはウェブサービスを通じて状況に応じてコンテンツを配信し、直接的な使用者のやり取りではなく、機械やソフトウェアのやり取りによって情報を交換する。

訳者：(株)情報システムエンジニアリング (ISE)

In the past decades, content management systems and, even more general, structured authoring technologies have been introduced in many regions in the world to optimize content creation processes. Via this and corresponding content reuse scenarios, documentation managers aimed at several goals: the reduction of content creation costs, reduction of translation costs, quality improvements and quality assurance of content and documentation, in-time creation and in-time delivery of information required by customers or for business to business communication.

Technologies involved in the corresponding systems most often include XML-based data formats. More precisely, content is authored according to rules given by XML information architectures with different degrees of semantic structuring. In this context, the field of Technical Communication has been a progressive area for standardizing content at multiple levels. This covers linguistic rules for the terminology and writing styles, standardized and semantically structured data formats or layout rules for corporate publishing.

However, most of the approaches and drivers of CCMS mentioned above have been focusing basically on internal and product-centric views of information. This means that, cost efficiency and legal compliance of the created documents have been most crucial to the definition of successful system implementation.

A new perspective on content

Recent technological developments and the increased expectations of users towards more precise content suitable to their actual needs are changing perspectives on content creation, content processing and information provisioning. In addition, information handling is influenced increasingly by emerging technologies which consider the interactions and environments of humans, machines, and software.

More specific, important drivers for a new view on information and consequently on content management can be found in a variety of technologies and applications being also partially interdependent of each other:

- mobile devices and corresponding apps used in private spheres, and that are consequently being demanded also in business use
- search driven information access from Google and other retrieval applications which includes faceted and filtered searching for consumer goods
- virtual reality and augmented reality enriching visual environments by additional information
- machine translation and its growing acceptance in private and business contexts
- artificial intelligence technologies which allow the automated extraction of decision-relevant information and knowledge from various data sources
- big data technologies for analyzing mass data generated for example from customer behavior in sales processes or from the tracking of machine states
- global IoT (Internet of Things) and Industry 4.0 initiatives which aim to control and react to the real-time behavior of products. They are applicable at different stages within the lifecycle of products, e.g. in the production, use and maintaining processes.

Recent real-life perceptions in the field of technical communication are that many companies are still quite traditional and somehow static, which means that they are document-centered and paper-based. Nevertheless, in various areas, product information has changed along the new technological lines drawn above. One approach often discussed in the public is the concept of "intelligent content". Because the processing of content and any other kind of

information requires IT applications and corresponding interfaces and underlying specialized processes, we will use also the general term "intelligent systems". The system types we are focusing on here are known as "content delivery portals" (CDP). They are used to dynamically deliver content to users, while the amount of information and media are adapted to their actual requirements. Sometimes this occurs even before they themselves know about their upcoming need for information.

A definition of Content Delivery Portals

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. This includes basic mechanisms to access or import content from other systems and to manage and update within the content lifecycle. The first mechanism can be understood as a technical delivery and integration process from relevant data sources and corresponding systems into the portal. The latter mechanism then offers retrieval functionalities including user interfaces for content searching by users. In addition, CDP usually offers web services which deliver content requests and other events. These functionalities are usually understood as being the literal core of content delivery. Nevertheless, in practical applications, the mentioned content integration processes are technically much more sophisticated than the pure delivery of content to the user.

Different types of searches

To understand the different approaches to content delivery, it is also helpful to first identify the three basic search strategies used in CDP.

- **Navigation by structure:**

The traditional electronic representations of

manual-based information are PDF or HTML documents. They include navigation structures as bookmarks or other tree-like structures which reflect the document logics. All of the online help media, E-books, and recent web-apps offer these types of navigation functionalities which support the traditional chapter or nested topic philosophy.

- **Facetted searching by metadata:**

To access directly single topics or information fragments, one takes advantage of metadata. They enrich all kinds of information objects in the CDP and allow the total set of information to be filtered into the required subset. They are often implemented as facets as they are known from product portals for consumer goods or for product configuration tasks.

- **Direct searching:**

While the first two strategies can be seen as structured searches, direct searching usually requires only little knowledge of the product structure or metadata logics. Direct searching usually yields the largest amount of results which thereby require a ranking algorithm in order to display the relevance of the retrieved information.

Apart from this distinctive set of search strategies, there are many situations where users need a sequence of consecutive methods. For example, the facets are used to narrow down the number of results from direct searching, or they reduce the volume of web-documents by hiding irrelevant sections and topics.

Different types of intelligence

The mentioned search strategies are supported at certain levels by different system types. To understand these levels and consequences, we will refer to the mentioned concepts of intelligent content and intelligent systems. In more detail, we will

differentiate between the following types of intelligence.

- Native Intelligence:

This concept covers the well-known semantic structuring of content by authors. The content structure and the content are accessible and machine readable mostly by XML-structuring and corresponding information models. In addition, all metadata, enriching and identifying modular content topics can be understood as being native contributions to intelligence. Even the document structures reflect in these senses the low-level metadata stating, nesting, and sequencing of modular topics. As an example of high-level native intelligence, one can use methods like the PI-classification scheme to systematically express the validity of topics with regard to product components and information classes. In this sense, all type of industry taxonomies used to classify products and corresponding information objects belong to approaches of native intelligence.

- Augmented Intelligence:

In many cases, a substantial set of the metadata and the relations between content objects are not included in the documents or the topics provided by the data sources. Special systems therefore allow the content to be enriched by "external" metadata schemes and for that data to be independently linked of the authoring process. This is done within or outside the content delivery applications. In the latter case, specialized metadata platforms are involved. A high-level approach uses ontologies for the modeling of the relations between the content objects.

- Artificial Intelligence:

The technologies of artificial intelligence used in technical communication attempt to extract knowledge and metadata from content by using

statistical or linguistic analysis. This means, they try to assign metadata to content objects in an automated way. The assigned metadata scheme can be of native or augmented nature.

It is obvious, that the native intelligence of data and systems directly support navigation and faceted searching. Augmented intelligence approaches also contribute to these search methods but require a more manual effort for the enrichment of metadata. Problems may generally arise with the native taxonomic metadata if the depth and logic of the metadata does not correspond to the user's expectations and knowledge. This is especially the case for a terminology mismatch between companies and the users of their products. Such problems can be addressed via terminology extensions in CDP that will allow the different terms used by the user to be mapped to the internal terms used by companies. This is then already part of an augmented intelligence approach. Similar statements hold generally for taxonomy or ontology mapping between different target groups of delivered information. Most recent CMS developments attempt to extend their metadata functionalities via the ontology-based modelling of topic relations. However, as this ontological development is highly demanding in regard to the modeling insight of products, it is unlikely that ontology modeling will be used as a standard functionality for CMS in the near future.

Artificial intelligence can be used to reduce the manual effort of augmented intelligence concepts by, for example, automating text classification. It can also reveal hidden relations which have not been modeled before. This becomes most relevant for the presenting of and for the accessing of unstructured content from non-CMS sources. Of course, this generally improves the power of the direct searching mechanisms in the intelligent delivery systems, and allows automated

linking to be derived between content objects.

Content Relevance Analytics and Content Intelligence

Automated linking between products in the consumer goods industry is known because of sales portals like Amazon. They use the analytics of consumer behavior for recommendations and linking. In technical communication, there are at least two analytics approaches which are sometimes subsumed under the term "content intelligence".

The first method analyzes the efficiency of content reuse within CMS and is used to optimize content creation processes. The analytics there reveal the reuse rates of modular content and allow example cost models to be derived for documentation. The second method focusses on delivery systems and attempts to measure the relevance of content by tracking user access to the system. Technically, normal web analytics can be applied to CDP but they need to be adapted to the properties being measured. Typical analytic properties measure for example the access frequency of the most relevant (or irrelevant) topics or the reading times and navigation behaviors of users. By analyzing search terms, one can get a more realistic picture of users and their information needs. Especially the above-mentioned terminology base of terms used by customers or other special target groups can be analyzed which thereby helps to further improve searching processes.

In addition, searching processes which lead to no results reveal a potential lack of required information. Therefore, content relevance analytics might help in the future to improve the overall quality and accessibility of technical information. The question raised earlier about the success of implementing content management is also answered via the feedback obtained from CDP user statistics.

Summary

In this article, we presented some concepts of intelligent delivery systems that present content to users in dynamic ways. Users can access content via the basic searching strategies of navigation, faceting, direct searching or combinations thereof. Depending upon the practical applications like for example portals for service technicians, end-users or for sales processes, the mentioned searching strategies are of varying importance.

CDP bridges the gap between content management systems and traditional portal applications via exchanging and updating mechanisms which take into account the complex structures and packaging of online information. For this purpose, there are initiatives to develop and implement standard exchange formats for content and metadata. In addition, CDP partially include methods from artificial intelligence to access unstructured data from various sources.

CDP might play a crucial role in IoT (Internet of things) or Industry 4.0 scenarios. There, information is connected to machines and components and is displayed and exchanged as it reacts to machine states or certain events. In such scenarios, CDP will deliver content situationally through web services and it will be triggered by machines and software instead of by direct user interactions.