

Metadaten für intelligenten Content

1 Einleitung

Seit einigen Jahren ist es in Mode, Content unter gewissen Voraussetzungen als intelligent zu bezeichnen. Dies geschieht weniger aus technischen Gründen als, um – ggf. auch aus marketingtechnischen Gründen – besondere und noch zu beschreibende Konzepte der Technischen Kommunikation zu betonen. Denn die zugrundeliegenden Technologien existieren bereits sehr lange, sollen und können nun aber in einem gewissen neuen Licht erscheinen. In diesem Beitrag sollen daher die Voraussetzungen betrachtet werden, unter denen die Bezeichnung intelligent gerechtfertigt sein kann. Dabei ist klar, dass es nur zum Teil um die Inhalte selbst geht. Zu einem großen Teil sind die Prozesse der Verarbeitung, der Präsentation, des Zugriffs und der Nutzung derjenigen Inhalte gemeint, die letztlich in der Technischen Kommunikation eingesetzt werden. Ein – wenn nicht sogar der – Schlüssel zur Intelligenz von Inhalten sind dabei Metadaten. Bei Metadaten handelt es sich abstrakt um prozessrelevante Zusatz- bzw. Verwaltungsinformationen zu den Objekten, die in dem jeweiligen IT-System abgelegt und mithilfe der Metadaten verarbeitet werden können.

Bei einer differenzierten Betrachtung der angesprochenen Voraussetzungen für intelligenten Content werden verschiedene Arten bzw. Stufen der Intelligenz betrachtet. In den entsprechenden Kapiteln finden sich dazu wichtige Aussagen jeweils grau hinterlegt zusammengefasst.

Die unterschiedlichen Stufen lassen sich als natürliche, erweiterte und künstliche Intelligenz der Inhalte sowie der damit verbundenen Prozesse verstehen. Entsprechend wären dann die englischen Bezeichnungen als Native, Augmented und Artificial Intelligence zu sehen. Nicht zufällig finden sich in diesen Bezeichnungen auch Bezüge zu aktuell diskutierten Themen wie Augmented bzw. Virtual Reality und eben auch die sich rasant in viele Bereiche des Lebens ausbreitende künstliche Intelligenz.

Intelligenter Content stammt zu großen Teilen aus Content-Management-Systemen (CMS) oder zumindest aus strukturierten Anwendungen. Daher ist es notwendig, die in den CMS angewandten

Methoden zu betrachten, um Content intelligent zu erzeugen. Ein Kern der nachfolgenden Betrachtungen wird dazu die mittlerweile häufiger verwendete Methode der PI-Klassifikation sein.

Darüber hinaus geht es aber immer mehr um die intelligenterere Verteilung von Informationen. Die entsprechenden Content-Delivery-Portale müssen neben den strukturierten ggf. auch unstrukturierte Inhalte auffindbar machen. Für diese wird man erweiterte Konzepte oder sogar Methoden der künstlichen Intelligenz nutzen müssen, um sie für die Suche zu erschließen. Eins der Konzepte findet sich im neuen iRDS-Format der tekcom-AG Information 4.0. Auch dort war die PI-Klassifikation einer der Ausgangspunkte für die ontologische Modellierung der Metadaten.

2 Der Ursprung der Intelligenz

Mit der Einführung von unternehmensspezifischen oder standardisierten Informationsmodellen in der Technischen Kommunikation, insbesondere aber seit der Nutzung von Content-Management-Systemen (CMS) auf XML-Basis, lassen sich technische Inhalte nach vordefinierten Regeln strukturieren und kontrollieren (Krüger/Ziegler 2008). Diese Regeln können je nach Informationsmodell und angewandter Technologie unterschiedlich tief strukturiert sein bzw. kontrolliert werden. So können auch DTP-Werkzeuge wie Word, FrameMaker oder InDesign innerlich strukturierte, aber bereits gelayoutete Informationen erzeugen. Mithilfe von XML-Werkzeugen werden die Inhalte dagegen medienneutral strukturiert. Das Layout wird dann erst im Nachgang an die Strukturen gebunden. In diesem Zusammenhang ist auch die sog. semantische Modellierung wichtig: Wenn die Strukturelemente sprechend benannt sind (z.B. Voraussetzungen, Handlungsschritt, Funktion, Wartungsinformation), lässt sich der Inhalt letztlich bei Bedarf sehr differenziert im Layout gestalten.

Dies ist natürlich nur ein Aspekt der feingliedrigen Strukturierung durch semantische Informationsmodelle, er ist aber in diesem Beitrag der wichtigste. Auf die Unterstützung der Standardisierung von Inhalten durch Strukturvorgaben sei daher nur am Rande verwiesen (Krüger/Ziegler 2008). Zudem erleichtern sprechend benannte Strukturen auch eine automatisierbare Qualitätskontrolle durch z.B. Controlled Language Checker (CLC), indem linguistische Regeln klarer an semantische Elemente gebunden werden können.

Wesentlich für die vorliegende Betrachtung ist also das Erkennen der Strukturen und die automatische Verarbeitung, z.B. für das jeweilige Publikationsmedium. Semantische Strukturelemente lassen eine spezielle Bedeutung der Inhalte durch Menschen oder Program-

me erkennen, wodurch sie sich spezifisch verarbeiten lassen. In diesem Sinne sind semantische Strukturelemente ebenfalls als implizite Metadaten für Inhalte auf sehr feiner Ebene zu interpretieren.

In Content-Management-Applikationen werden die Inhalte üblicherweise in modulare Einheiten aufgeteilt. Hintergrund ist die erhoffte Wiederverwendung der modularen Inhalte und die spezifische Aggregation der Module zu Dokumenten entsprechend den zu dokumentierenden Produkten bzw. den Produktvarianten. Um diese modularen Einheiten wiederzufinden bzw. identifizieren zu können, nutzt man in den CMS Metadaten. Im einfachsten Fall helfen sie, die Module manuell, aber systematisch zu suchen und korrekt zu verwenden¹. Es können aber auch hochkomplexe, automatisierte, vernetzte und damit „intelligente“ Prozesse der Aggregation und Publikation in den Systemen ablaufen. In der Vergangenheit war somit die hocheffiziente Erstellung von Dokumentation ein Treiber für die Definition und Nutzung von Metadaten.

Eine ähnliche Motivation brachte auch das komplexe Konzept der modularen Dokumentation im europäischen militärischen und Luftfahrt-Umfeld hervor (S1000D). Dort ist die Organisation der umfangreichen Zulieferinformationen und deren Integration in ein Gesamtprojekt eine der wichtigen Anforderungen. Entsprechend fein gestalten sich daher die Metadaten, die weitgehend die potenziellen Komponenten des militärischen Geräts umfassen. Das entsprechende Informationsmodell ist dabei auch auf inhaltlicher Ebene tief strukturiert.

Neben der dargestellten Content-Management-Systematik hat sich seit relativ kurzer Zeit ein weiteres Feld entwickelt, das auf der Intelligenz der Informationen aufbaut (Ziegler 2017). Content-Delivery-Portale sollen die Inhalte endlich nicht mehr nur statisch zur Verfügung stellen, sondern eine neue Sichtweise auf die Informationen ermöglichen. Zum Einen sollen unterschiedliche Inhalte dem Nutzer in Abhängigkeit vom jeweiligen Kontext angezeigt werden können. Der Kontext kann dabei über vielfältigste Anwendungsszenarien definiert sein: in einer Lernsituation, bei der Simulation von Betriebszuständen, z.B. bei der Einstellung, Justage oder Bedienung von Produkten, als Reaktion auf Maschinen- oder Software-Zustände, für Fehlerfälle, in Service-Situationen etc. Zum Anderen sollten die Inhalte über verschiedene technische Methoden gut durchsuchbar sein und wirklich passende Ergebnisse liefern. Natürlich beruht das Delivery massiv auf der Qualität der Daten, die in ein Delivery-

¹ Da semantische XML-Elemente auf allen Ebenen verwendet werden können, gibt es auch einen fließenden Übergang in der Modellierung zwischen semantischen Wurzelementen von Modulen (z.B. Task, Maintenance-Plan) und der expliziten Vergabe von Metadaten an neutralen Modulen. Das Zweite ist zu bevorzugen, insbesondere wenn sich die Informationsmodelle der Module nicht unterscheiden.

System eingespeist werden und damit stark auf den Metadaten, die direkt oder indirekt am Content hängen.

In den nachfolgenden Teilen des Beitrags wird immer wieder der Bezug zu CMS und CDP hergestellt, um derzeit relevante Anwendungsfälle zu diskutieren. Selbstverständlich lassen sich viele weitere zukünftige oder schon existierende Anwendungsfälle auch außerhalb dieser Systeme realisieren.

Intelligenter Content umfasst Inhalte, deren Verarbeitung automatisierbar oder zumindest stark systematisierbar sein soll.

Die intelligente Automatisierung in CMS verlangt die Vergabe und Auswertung von Metadaten. Ziel der Automatisierung ist die Erzeugung von Dokumenten passend zu Produktvarianten. Die semantische Strukturierung von CMS-Inhalten ist der Spezialfall einer impliziten Vergabe von Metadaten. Sie kann genutzt werden, um z.B. Publikationsprozesse zu automatisieren. In CDP werden unterschiedlichste Inhalte dynamisch mithilfe von Metadaten zur Verfügung gestellt. Als Vision der Industrie 4.0 soll der Produkt- und Nutzungskontext die Inhalte intelligent, d.h. automatisiert selektieren und dem Nutzer anzeigen. Der Ursprung der Intelligenz ist also generell der Wunsch nach Automatisierbarkeit der mit dem Content verbundenen Prozesse.

3 Die natürliche Intelligenz durch Klassifikationen und andere Metadaten

CMS wurden für die Erstellung und Verwaltung von modularen Informationsobjekten – d.h. im einfachsten Fall granulare Texte und darin referenzierte Medien – eingeführt. Die Informationsobjekte bilden den Content, der in den zu erstellenden Dokumenten durch Referenzierung der modularen Objekte aggregiert wird. Die Systeme bieten Funktionalitäten für die Basisprozesse des Versions-, Varianten- und Übersetzungsmanagements sowie die crossmediale Publikation der erstellten Dokumente. Als Datenformat wird üblicherweise XML eingesetzt, auch wenn andere Datenformate prinzipiell möglich und im Einsatz sind. Die modularen Informationen werden, wie oben beschrieben, gemäß den in den Systemen implementierten Informationsmodellen strukturiert, wobei das jeweils verwendete Modell und damit Semantik und Tiefe der Strukturierung gerade im mitteleuropäischen Raum sehr system- bis kundenspezifisch ist. Im englischsprachigen Raum findet eine in Teilen stärkere Standardisierung auf dieser inhaltlich-strukturellen Ebene durch das DITA-Informationsmodell statt.

CMS bieten den Nutzern in der Regel eine bereits vordefinierte Menge an Metadaten, um die Informationsobjekte zu verwalten. Sie

sind üblicherweise aus zahlreichen Kundenprojekten entstanden und bilden eine Sammlung typischer bzw. meist genutzter Metadaten. Sie lassen sich in den meisten Systemen kundenspezifisch anpassen und erweitern. Die Metadatenkategorien und deren Nutzung waren in der Vergangenheit somit nicht von einem Standard vorgegeben.

Eine Ausnahme stellt die stark reglementierte Branche militärischer und Luftfahrt-Produkte dar. In dieser existiert für die Dokumentationserstellung ein enges Konzept für die modulare Erstellung nach einem vorgegebenen Informationsmodell einschließlich der Vorgabe von Metadatenkategorien bis hin zu den vordefinierten Werten. Der Hintergrund ist die notwendige Organisation von Entwicklungsprojekten mit zahlreichen Zulieferern, die dann auch die modularen Informationen zu ihren gelieferten Komponenten bereitstellen müssen. Auch wenn es immer wieder Versuche gab, diese Systematik für andere Branchen zu nutzen, war diese doch in der Regel zu komplex, um sie vollständig zu übertragen. Eine weitere Ausnahme bildet die Dokumentation von Kraftwerkssystemen mit ihren Metadaten systemen KKS bzw. RDS-PP, die ebenfalls ein Schlüssel system gemäß der Anlagenstruktur zur Verfügung stellt (Ruchnow 2014).

Weitere eher allgemeine Ansätze finden sich z.B. bei der Metadata-Initiative des Dublin Core (Dublincore 2012). Dessen in anderen Domänen vielfach verwendete Definitionen von Metadaten sind allerdings eher dokumentorientiert angelegt und werden derzeit weniger für die bisherigen modularen und variantenspezifischen Dokumentationsprozesse und deren feingranulare Objektverwaltung genutzt.

Eine derzeit in den weiter verbreiteten CMS häufig verwendete Methode ist hingegen die PI-Klassifikation. Sie ist im Kern unabhängig von Informationsmodellen, Systemen und Produkten. Historisch wurde sie mit dem Informationsmodell PI-Mod zuerst verwendet (Steurer/Ziegler 2010). Durch die Namensähnlichkeit kommt es dadurch auch zu gelegentlichen Verwechslungen. Die PI-Klassifikation teilt die für die modularen Informationsobjekte verwendeten Metadaten zuerst in die vier bekannten Klassifikationsbereiche ein (Drewer/Ziegler 2013), danach folgen aber noch weitere Bereiche²:

- Intrinsic informationsbezogen
beschreibt die Informationsklassen, und damit die konkrete Art der Informationsarten, die in den jeweiligen Modulen enthalten sein dürfen.
- Intrinsic produktbezogen
beschreibt die Klassen der Produktkomponenten, aus denen die Gesamtprodukte bestehen.

² Die Metadaten, die wie die Versionierung den Content Lifecycle beschreiben, werden in der vorliegenden Betrachtung außer Acht gelassen. Sie sind aber ebenfalls für die Prozesssteuerung von Relevanz.

- **Extrinsisch informationsbezogen**
beschreibt die Dokumentarten, Zielgruppen, Zielmärkte etc., für die die Module verwendet werden können.
- **Extrinsisch produktbezogen**
beschreibt die Endprodukte, die dokumentiert werden und für die die modularen Objekte genutzt werden können.

Die intrinsischen Klassifikationen haben eine besonders wichtige Bedeutung für die Trennschärfe der Informationen. Sie definieren gleichzeitig das primäre Modulkonzept und die entsprechenden Metadaten müssen für Redakteure eindeutig zu jedem Modul zuzuordnen sein. Sie machen also die Module inhaltlich trennscharf und grenzen diese voneinander ab. Die extrinsischen Klassifikationen hingegen beschreiben die potenzielle Verwendbarkeit für Produkte oder Dokumentarten und sind daher eher für den nachgelagerten Prozess der Dokumenterstellung sowie für Rechercheprozesse relevant.

Häufig sind die Klassifikationen hierarchisch aufgebaut und sind dann in den Systemen in Form einer auswählbaren Baumstruktur zur Vergabe oder zur Suche nach Metadaten visualisiert. Im Sinne der PI-Klassifikation dürfen bei der Vergabe die intrinsischen Baumstrukturen auch nur einen Wert erhalten. Die extrinsischen Metadaten sind dagegen mehrwertig und beschreiben die mehrfache Verwendung, die ja auch das Ziel einer CMS-Anwendung ist.

Bei den intrinsisch-informationsbezogenen Klassifikationen findet sich die Einteilung in grundlegende Basisklassen. Meist sind darunter prozedurale und deskriptive Informationsarten sowie weitere nach Bedarf. Auch bestehende Informationsmodelle wie S1000D oder DITA können hier ihre Modulbasistypen einordnen.

4 Der PI-Fan als Referenzmodell

Um die Methode an einem anschaulichen Beispiel nachvollziehen zu können, wurde der PI-Fan geschaffen. Es handelt sich um ein freizunutzbares Referenzmodell, mit dem Systemhersteller ihre Implementierung eines Produktes demonstrieren können, dessen Dokumentation mithilfe der PI-Klassifikation erstellt wird³ (Ziegler 2016).

Generell ist die PI-Klassifikation eine Methode, die den anwendenden Unternehmen die Freiheit gibt – es aber auch notwendig macht, ihre eigenen spezifischen Produkt- und Informationsstrukturen abzubilden. Sie bleibt damit bewusst eine Abstraktionsebene über den konkreteren Vorgaben der S1000D- bzw. der KKS-/RDS-PP-Systematiken. Sie hat aber mit diesen natürlich auch viele Parallelen

³ Soll eine reale Implementierung nachweislich mit der Methode arbeiten, kann sie als zertifizierte PI-Class[®]-Anwendung überprüft und freigegeben werden.

in der methodischen Definition der Metadaten. So gibt es auch in vielen weiteren Implementierungen und Projekten der Vergangenheit verwandte Ansätze. Die PI-Klassifikation hat derartige Ansätze mit einem methodischen Rahmen versehen und breit in den vorhandenen CMS anwendbar gemacht. Sie ist dabei auch nicht auf bestimmte Branchen festgelegt, wenn man den Produkt- bzw. Komponentengedanken weiter fasst und auf immaterielle Gebiete wie z.B. Dienstleistungen, funktionale Baugruppen oder Software überträgt.

Produktklasse 1 (Baugruppe/ Funktionsgruppe)	Produktklasse 2 (Komponente/ Funktionseinheit)	Produktklasse 3 Bauteil
Gesamtgerät		
Antrieb	Antrieb Anschluss Elektromotor Getriebe	
Beleuchtung	Abdeckung Leuchtmittel	
Heizung	Heizelement	
Rotor	Flügel Flügelhalter Lauftrad	
Schutz	Schutzgitter	
Anzeige-_Bedienelemente	Geschwindigkeitsregler Schwenksteuerung Temperaturfühler Temperatursteuerung Display	
Halterung	Bodenständer	

Abb. 1: Intrinsische und mehrstufig-hierarchische P- und I-Klassifikationen des PI-Fans (Version 1.2). Zu dunkel unterlegten Einträgen gibt es im Referenzmodell noch keine Module. Module müssen dann aus jedem Taxonomiebereich einen Wert tragen.

Informationsklasse 1 (Modultyp)	Informationsklasse 2 (Produktlebenszyklus)	Informationsklasse 3 (Detaillierter Produktlebenszyklus)
Anleitung	Bedienung	Gebälseeinstellung Höheneinstellung Neigungseinstellung Schwenkaktivierung Temperatureinstellung
	Inbetriebnahme	
	Lagerung	
	Montage	
	Wartung	Prüfung Reparatur
Beschreibung	Aufbau Entsorgung Funktion Tech.Daten	
Plan	Diagnose	Fehlercode Manuell
Sicherheit	Allgemeine Sicherheit Bestimm. Verwendung Spezifische Sicherheit Vorhers. Fehlgebrauch	

Neben den anfangs genannten vier Klassifikationsbereichen enthält eine vollständige PI-Klassifikation potenziell aber auch noch die sog. Variantenmerkmale. Diese werden teilweise gar nicht separat behandelt und hängen mit der Art des Variantenmanagements zusammen, das im jeweiligen CMS genutzt wird. Die Variantenmerkmale waren in der Vergangenheit in den meisten Fällen identisch mit extrinsischen Klassifikationen. D.h. es genügte häufig, ganze Module oder Teile der Inhalte in diesen auszuzeichnen und damit eine Zuordnung zu den Produktvarianten vorzunehmen. Damit kann bei Bedarf dann automatisiert eine spezifische Dokumentationsvariante erstellt werden. Die Technologien dahinter arbeiten dann z.B. mit der Filterung von Masterdokumenten (modulare und submodulare Variantensammlungen) oder mit der Generierung nach intrinsisch klassifizierten Strukturvorgaben. Im letzteren Fall geben also intrinsische Klassifikationen die Sequenzierung im Dokument vor, die extrinsischen Klassifikationen lassen die zu der Produktvariante passenden Module selektieren. In beiden Fällen werden intelligente Verfahren mithilfe von passenden Metadaten eingesetzt, um das Variantenmanagement abzubilden. Je mehr und je feiner die Klassifikationen modelliert werden, desto hochintelligenter und feiner können die Automatisierungsprozesse ablaufen. Das Problem ist dann häufig eher die Übersichtlichkeit bei der Erfassung, die Usability der Systeme und die Abstraktionsmöglichkeit der Auswirkungen von Metadaten bei der Erfassung.

Durch die zunehmende Konfigurationsabhängigkeit vieler Produkte sind aber in Theorie und Praxis anstelle von oder auch ergänzend zu den extrinsischen Klassifikationen weitere Variantenmerkmale von wachsender Bedeutung. In den Modulen können spezielle Merkmale (Anschlussspannung, Abmessungen, Steuerungsmerkmale, Materialien und Überzüge, Nutzungssituationen, Bussysteme etc.) für Varianten in den Modulen verantwortlich sein. D.h. es können Merkmale für das Variantenmanagement erforderlich werden, die z.B. nur an einzelnen Komponenten hängen oder auch übergreifend für alle Komponenten relevant sein können. Das Endprodukt wird dann häufig nur in Zusammenhang mit der Merkmalskombination eindeutig definiert. Hier kommen in der Realität der Produktentwicklung vielfältigste konzeptionelle Ausprägungen der Variantenmerkmale vor. Ein spezieller Anwendungsfall sind stücklistenbasierte Artikelnummern, die aus den ERP-Systemen stammen. Sie sind nicht intrinsisch, da ein Modul für mehrere Nummern gelten kann; sie sind aber – obwohl sie also mehrwertig sind – auch nicht extrinsisch, da sie ja keine Endprodukte beschreiben. Unternehmen, die sich mit einer voll stücklistengenerierten Dokumentation auseinandersetzen, müssen derartige Konzepte einbeziehen.

Im PI-Fan sind der Übersichtlichkeit halber die Variantenmerkmale direkt auch im extrinsischen Produktnamen enthalten. So ist ein

„T5-DH2“-Ventilator eigentlich eine Abkürzung für einen 5-stufig einstellbaren Tischventilator mit Display und zweistufiger Heizung. In den Demonstratoren zum PI-Fan der Systemanbieter kann man dann die Schwerpunkte des Variantenmanagements (extrinsisch und/oder merkmalsbezogen) erkennen, sodass der PI-Fan als Referenzmodell für derartige intelligente Auswertungsprozesse dienen kann.

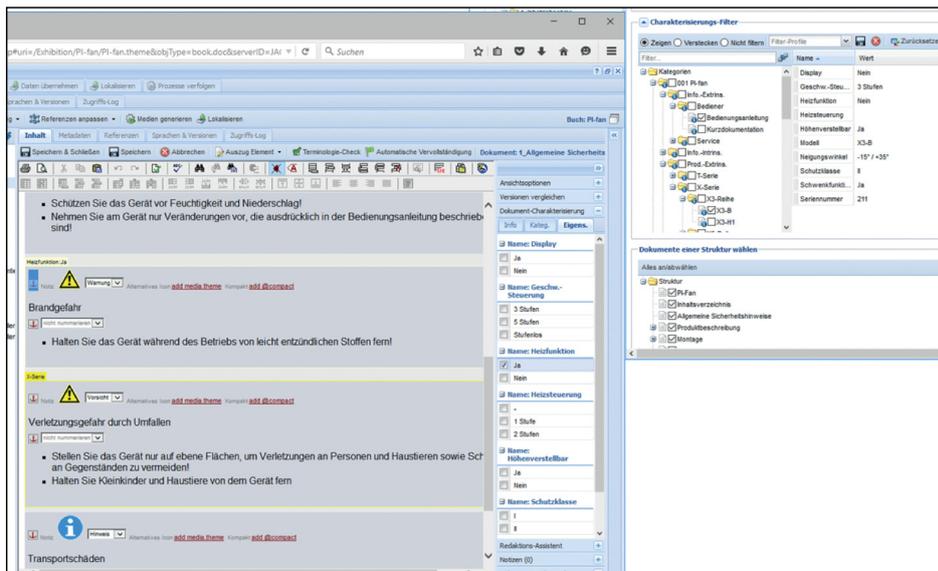


Abb. 2: Variantenmerkmale – hier die Heizfunktion, im System als Eigenschaften bzw. Charakterisierung benannt – lassen sich an einzelne submodulare Teile des Contents hängen. Bei der Dokumenterstellung lassen sich diese Merkmale global auswählen und sie filtern das Dokument als Ganzes ggf. modular und submodular. Gleichzeitig kann auch extrinsisch – hier die X-Serie – ausgezeichnet und gefiltert werden. (Quelle: Expert Communication Systems)

Als weitere – bei Bedarf in PI-Class[®] enthaltene – Metadaten lassen sich noch sog. funktionale Metadaten definieren. Diese sollen wiederum spezielle Funktionen oder Automatisierungsschritte bei der Publikation und zwischen Modulen steuern. Als Beispiel können die mit dem PI-Mod-Informationsmodell vorgestellten Wartungsintervalldaten, Werkzeuglisten oder Fehlermeldungs-codes dienen. Diese hängen zusätzlich explizit an Modulen und können im Publikations- oder Delivery-Prozess ausgewertet werden. Prominentes Beispiel ist die automatisierte Erzeugung von Wartungsplänen aus den in einer Dokumentation verwendeten Modulen. Hier sind vielfältigste Mechanismen denkbar und im Einsatz. Gerade im CDP-Bereich lassen sich viele logische und physische Verknüpfungen von Content auf der Basis von Klassifikationen automatisch erstellen. Die scheinbare Intelligenz des Contents beruht also, wie dargestellt, auf der Automatisierung und Nutzung von vordefinierten Metadaten.

Betrachtet man nun die aktuellen Content-Delivery-Portale, so erscheint es natürlich, dass die grundlegenden Suchkonzepte aus den intelligenten CMS-Mechanismen übertragen werden. Die Suchprozesse können über die Navigation (Dokumentstrukturen), über Facettierung (Filterung und Selektion nach Klassifikationen) oder über die direkte Suche (Volltext oder erweiterte Verfahren) ausgeführt werden. Insbesondere Navigation und Facettierung sind also direkte

Nutznießer einer für den Anwendungsfall passenden Metadatenmodellierung. Berücksichtigt man, dass CDP auch als IT-Service gedacht sind, die z.B. auf Anfrage eines Ereignisses ohne menschliches Zutun die benötigte Informationseinheit zur Verfügung stellen sollen, dann ist klar, dass eine notwendige Präzision von Suchergebnissen einer feingranularen Klassifikation und einem methodisch sauberen Konzept folgen muss.

Klassifikationssysteme wie die PI-Klassifikation erlauben es im Sinne einer natürlichen Intelligenz, modulare Informationen mit relevanten Metadaten auszustatten. Diese können ihre eindeutige inhaltliche Bedeutung (intrinsisch) und ihre mehrfache Verwendbarkeit (extrinsisch) beschreiben. Zusätzlich erlauben weitere Variantenmerkmale, komplexere Steuerungen von Produktvarianten abzubilden und damit konfigurationsspezifische Ausgaben zu erzeugen. In Content-Delivery-Applikationen unterstützen die vergebenen Metadaten die grundlegenden Suchprozesse der Navigation und der facettierten Suche. Die Suche umfasst Module, daraus aggregierte Dokumente und ggf. weitere Inhalte, die nicht aus CMS stammen. Darüber hinaus können Inhalte durch Klassifikationen und funktionale Metadaten automatisiert vernetzt und ausgewertet werden. Zur Demonstration der Funktionalitäten eines CMS oder CDP kann der PI-Fan, der alle relevanten Klassifikations- und Variantenmechanismen enthält, als Referenzmodell genutzt werden.

5 Erweiterte Intelligenz durch Beziehungen

In der Praxis der PI-Klassifikation gibt es typische Modellierungssituationen, die z.B. zu Problemen mit der taxonomischen Eindeutigkeit der intrinsischen Klassifikationen führen. Ursache ist in nicht wenigen Fällen, dass Variantenmerkmale mit den intrinsischen Klassifikationen vermischt oder als Kreuzprodukt ausmultipliziert werden. Hierdurch sollen auch die Komponentenvarianten für Redakteure direkt im Baum der intrinsischen Komponentenklassen explizit auswählbar sein. Was in einfachen Fällen funktionieren kann, wird aber leicht bei mehreren einzubeziehenden Merkmalen unübersichtlich bis nicht umsetzbar. Dies ließe sich technisch lösen, wenn, wie bereits angesprochen, Variantenmerkmale als weitere Dimension an z.B. einzelne Komponentenklassen gebunden werden könnten. Dies ist aber so in wenigen CMS der Fall.

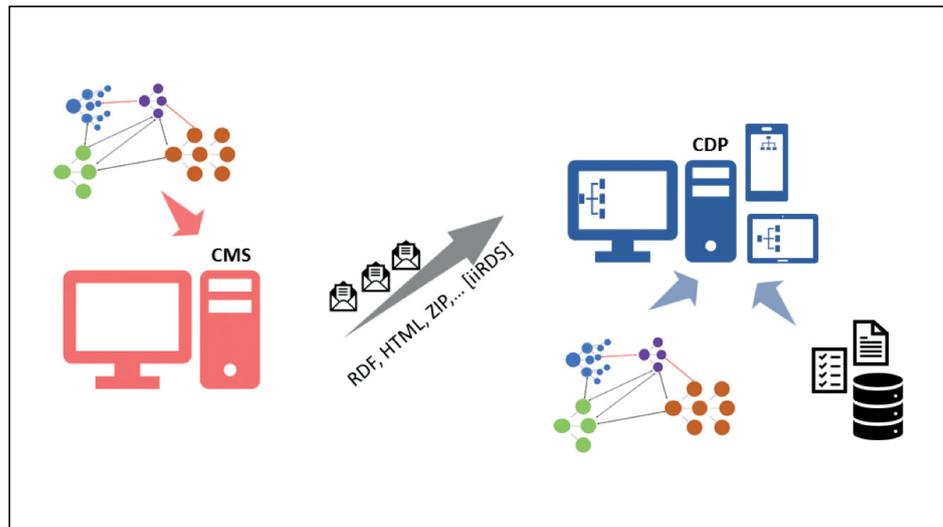
Eine weitere Besonderheit ist darin zu sehen, dass die intrinsische Produktklassifikation zwar tendenziell wie ein partitives (Be-

stands-)System wirkt, es aber nicht vollständig ist. Teile davon haben zwar Bestandscharakter, aber man darf nicht vergessen, dass es sich um eigenständige Komponentenklassen handelt, die nur aus Gründen der Übersichtlichkeit und logischen, den Redakteuren bekannten Teile-Gliederungen als Baum dargestellt werden. Denn wenn Komponenten an mehreren Stellen physisch oder, wie in der Software, funktional verbaut sind, heißt das nicht, dass sie auch in der intrinsischen P-Klassifikation an mehreren Stellen vorkommen dürfen. Dieses widerspricht dem Verständnis der intrinsischen Eindeutigkeit. Es gibt hierfür verschiedene Lösungsansätze bei der Definition der Taxonomien, um (wieder) zu einer Eindeutigkeit zu gelangen, u.a. das Verschieben und Zusammenlegen von Klassen, um die rein partitive Sichtweise aufzulösen. Hier soll jedoch gezeigt werden, dass zukünftig auch andere Wege beschritten werden können.

Hintergrund dieser und anderer Situationen ist die Tatsache, dass sich die Metadaten bei einer ganz grundlegenden Betrachtung zu modularen Content-Objekten eigentlich aus einem Netz von Beziehungen ergeben oder ergeben sollten. Derartige Beziehungs- oder Wissensnetze sind seit langem in verschiedenen Disziplinen wie in der Linguistik oder Informatik als Ontologien in Benutzung. Ontologische Beziehungsnetze waren in der Vergangenheit immer auch im Bereich des Semantic Web in der Diskussion. Es gab in den letzten Dekaden auch vielfältige technische Umsetzungen durch XML Topic Maps (XTM) und andere formale Beschreibungen wie das Resource Description Framework (RDF) oder Web Ontology Language (OWL). Gerade RDF hat sich aber über die Zeit als Beschreibungssprache für Beziehungen zwischen allgemein betrachteten Ressourcen, d.h. im Speziellen für Informationsobjekte etabliert.

Wie können nun diese Wissensnetze für die Technische Kommunikation eingesetzt werden? Im Sinne der zu Beginn eingeführten erweiterten Intelligenz gibt es dafür zwei Bereiche: zum Einen durch die Erweiterung der Modellierung und der Metadatenlogik bei der Erfassung von Informationen im CMS, zum Anderen durch die Verknüpfung von Informationen bei der Suche und bei der Darstellung von Suchergebnissen in CDP. Obwohl es sinnvoll erscheint, beides gleichzeitig zu nutzen, wird dies derzeit in vielen realen Fällen noch nicht möglich sein. Dies liegt z.B. an der derzeit noch geringen Unterstützung von Ontologien in CMS, aber auch daran, dass CDP auch Daten zugreifbar machen sollen, die nicht aus den CMS stammen und somit ggf. auch unabhängig vom CMS verknüpft werden müssen.

Abb. 3: Grundprinzipien von Augmented Intelligence. Links: Nutzung einer Ontologie als Beziehungsnetz im CMS. Mitte: Übergabe der mit Standardtechnologien paketierte CMS-Inhalte und der Metadaten an CDP. Rechts: Nutzung von Ontologien im CDP mit den verschiedenen Ausgabemedien. Es werden damit Informationen aus CMS und ggf. aus weiteren Datenquellen und ggf. Begriffssystemen für die Suche vernetzt.



Bei der Modellierung der Beziehungen als Ontologie oder als sonstiges Netz können nun auch die in der PI-Klassifikation definierten Beziehungen ein vollständigeres Gefüge der vorhandenen Informationen ergeben. Die zweidimensionalen Taxonomien werden so zu einem räumlichen Geflecht von Beziehungen. Jede Informationseinheit kann damit eine Menge von unterschiedlichen Beziehungen zu anderen Objekten haben. Die Arten der Beziehungen sind ebenfalls modellierbar, es finden sich aber sicher die bekannten Typen („besteht aus“, „ist ein“) sowie produkt-, informations- oder branchen- bzw. firmenspezifische Beziehungen und Klassen. Bei der Übertragung der Logik der PI-Klassifikation auf den entstehenden iiRDS-Standard zur Übergabe von paketierte Inhalte finden sich daher sowohl die Beziehungen zu intrinsischen („hat Topictyp“) und extrinsischen Objekten („hat Dokumentart“) als auch z.B. die funktionale Beziehung („erfordert Hilfsmittel“) und Variantenmerkmale („bezieht sich auf Produktmetadaten/Produktmerkmale“).

Es existieren am Markt bereits erste Lösungen für die Modellierung von Beziehungsnetzen, die sich in CMS integrieren und/oder als Zusatzkomponente von CMS nutzen lassen. Treiber sind hier z.B. auch die Wünsche, den Produktentwicklungsprozess und die Produktkomplexität enger mit der Dokumentation zu verzahnen und Variantenabhängigkeiten auch für Redakteure systematisch erschließbar zu machen. Parallel dazu existieren Lösungen, die Ontologien als Teil von CDP nutzen oder diese ergänzend für die Informationsvernetzung in CDP nutzbar machen. Zusätzlich gibt es Konzepte und Systeme, die aus dem Terminologiemanagement heraus entwickelt werden. Die dabei definierten Begriffssysteme sind die Basis für ontologische Modelle der jeweiligen Unternehmen und können damit auch dem internen Wissensmanagement dienen (Drewer u.a. 2017). Dadurch können auch die Mechanismen der Mehrsprachigkeit und der Synonymie für z.B. CDP erschlossen werden (Wetzel 2014).

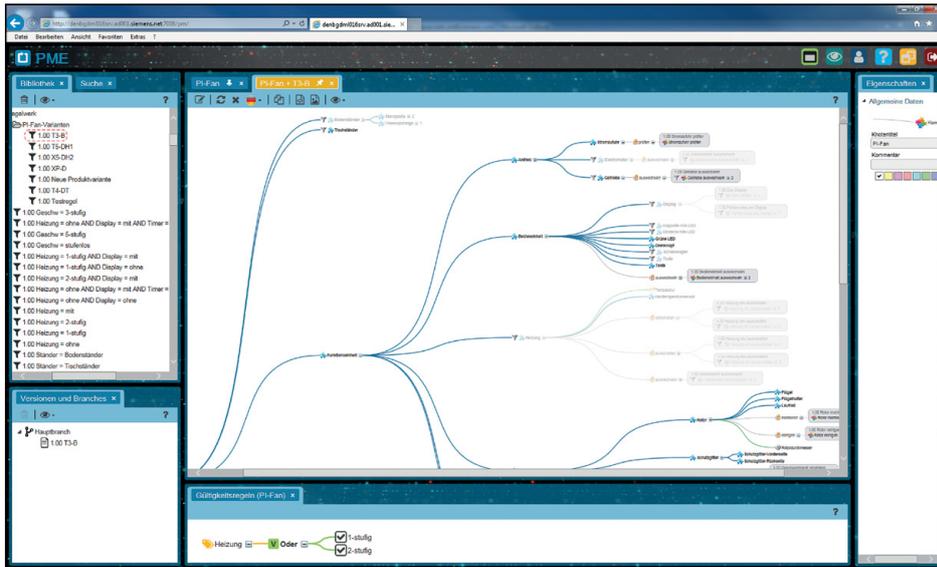


Abb. 4: Ontologische Modellierung von Produktstrukturen zur Steuerung von CMS-Prozessen. Im Bild wurde der PI-Fan prototypisch funktional sowie partitiv modelliert und für eine spezielle Ausprägung eines Produkts T3-B gefiltert (ausgegraut). (Quelle: Ontolis)

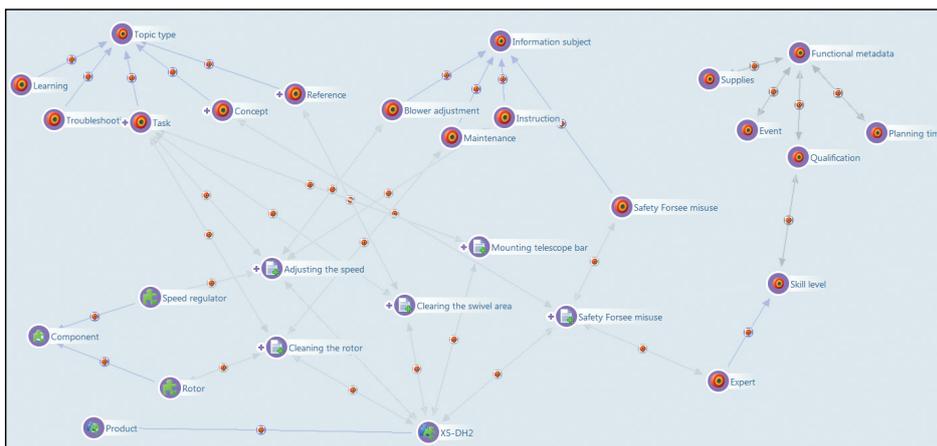


Abb. 5: Ausschnitt aus der Entwicklung einer Ontologie für den PI-Fan zur Vernetzung von Modulen/Topics (Mitte und rechts unten mit Textsymbol). Auf Basis der ontologisch genutzten PI-Klassifikation können CDP mehrdimensionale Verknüpfungen für die Suche nutzen. (Quelle: Intelligent Views)

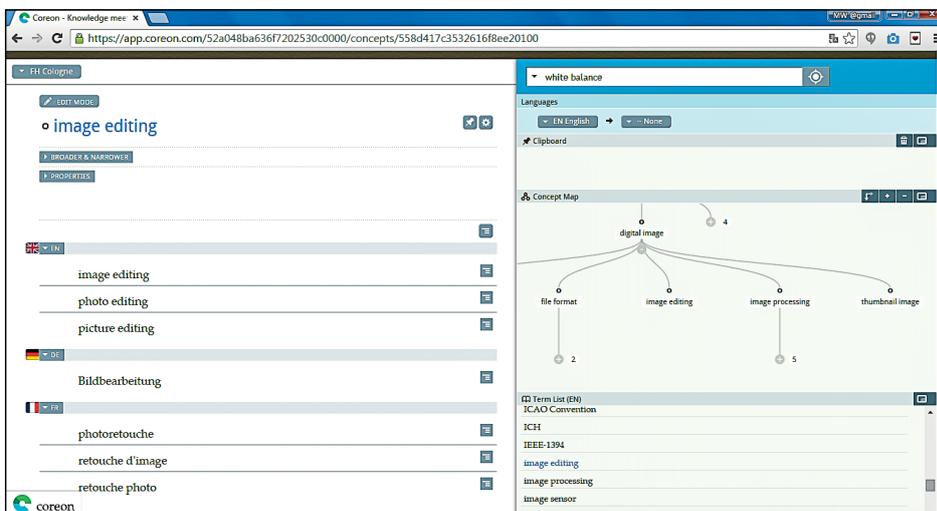


Abb. 6: Mehrsprachig aufgebautes Begriffssystem, das als Ontologie, bzw. Taxonomie visualisiert wird. Die Ontologie wird im Bereich der Terminologie als Concept Map bezeichnet. (Quelle: Coreon)

Die Erweiterung der taxonomischen und hierarchischen Klassifikationen führt zu den schon länger bekannten ontologischen Modellen. Mit derartigen Netzen von Beziehungen können – neben dem Aufbau des Terminologie- und Wissensmanagements – im Speziellen die Prozesse der Dokumenterstellung systematisiert werden. Die Informationslogik soll damit möglichst weitgehend kongruent mit den Funktions- und Produktmodellen von Software- und Hardwareentwicklung sein. Im Content Delivery sollen dann Inhalte über alle relevanten Verbindungen recherchierbar gemacht werden und intelligent – da automatisierbar – verknüpft sein.

6 Künstliche Intelligenz oder: Wie intelligent soll der Content denn noch werden?

Es muss gefragt werden, welche Rollen mit den Metadaten und intelligentem Content verbunden sind. Die Modellierung und die Befüllung von Beziehungsnetzen, wie auch die PI-Klassifikation, ist eine sehr spezielle administrative Tätigkeit und verlangt Produktwissen, Kommunikation mit der Produktentwicklung und ein tieferes Modellierungsverständnis. Letztlich sollen diese Modelle und Technologien die Dokumentationsentwicklung noch weiter systematisieren, prozesssicherer sowie konfigurations- und variantenspezifischer gestalten lassen. Eine noch weitere Automatisierung der Erstellungsprozesse selbst ist mit diesen Technologien wahrscheinlich nicht verbunden. Ontologien sollen z.B. bei der Abschätzung und Planung der Content-Varianten unterstützen, die Anwender bzw. Redakteure müssen aber ggf. nicht direkt mit Ontologien konfrontiert werden. Als Visualisierung und Informationsmedium sollten sie aber hilfreich sein. Im CDP-Bereich ist die intelligente Vernetzung von Content, d.h. das Auffinden von fachlichen und thematischen Bezügen an der Suchoberfläche das Entscheidende. Auch hier ist es allerdings fraglich, ob, außer für Experten, eine Ontologiedarstellung zielführend ist. Vielmehr geht es um die Unterstützung und Verbesserung bzw. Systematisierung der Suchprozesse.

Allerdings besteht auch bei den bisher dargestellten Konzepten die Notwendigkeit, die zu erstellenden bzw. die im CDP zu präsentierenden Informationsobjekte in das erweiterte Beziehungsnetz einzuordnen. Dem entspricht bisher die analoge Verortung von Modulen in die Bäume der (PI-)Klassifikationen. Der nächste Schritt der Automatisierung ist die Einordnung von Content in die Beziehungsnetze mithilfe von analytischen Verfahren der Künstlichen Intelligenz (KI). Diese beruhen auf unterschiedlichen Technologien der Textanalyse, nutzen aber im Grundsatz statistische und teilweise linguistische Ver-

fahren zur Explizierung der im Modell implizit zugeordneten Klassen und Beziehungen. Diese aus anderen Domänen bekannten Verfahren lassen sich auf die Technische Kommunikation und die CMS/CDP-Anwendungen übertragen (Oevermann/Ziegler 2016). Derartige Technologien könnten somit für die Bestandsdatenmigration von unklassifizierten CMS-Inhalten in neu eingeführte CMS genutzt werden. Alternativ könnte damit auch die Qualitätssicherung unterstützt werden, um zu prüfen, ob die vergebenen Klassen mit den Inhalten übereinstimmen.

Auf CDP-Seite können mit KI auch unstrukturierte Informationen und monolithische Dokumente aus Drittsystemen ohne bestehende Klassifikationssystematik verarbeitet werden. Dies kann z.B. genutzt werden, um in den Dokumenten Quasi-Module zu identifizieren, feingranulare Metadaten (Klassenzuordnungen in Ontologien oder im Taxonomibaum) zu vergeben und dann über die Facetensuche zugreifbar zu machen. Natürlich kann die KI aber darüber hinausgehen und zur Analyse und Interpretation der Suchanfragen, zur Sprachanalyse bei natürlichsprachigen Eingaben oder zur Muster- und Objekterkennung bei der visuellen Suche genutzt werden. Auch können Beziehungen und damit Modellerweiterungen automatisiert gefunden werden, die gar nicht durch die manuelle Modellierung vorgenommen wurden. Die Grenzen der Methoden verschieben sich hier ständig; bis hin zur Neurorobotik (Gehirn-Computer-Vernetzung) mit der Verknüpfung von menschlicher und künstlicher Intelligenz, z.B. bei einer zukünftigen gedanklichen Suche (Hanel 2016).

Ontologische Ansätze für die Klassifikation von Informationsobjekten und für die Beschreibung der Beziehungen zwischen diesen sollen helfen, die Informationsobjekte prozesssicher zu erstellen und präzise finden zu können. Die Methoden der Künstlichen Intelligenz können helfen, das Netz der Beziehungen automatisch zu befüllen, d.h. Informationen automatisiert mit Metadaten auszustatten. Sie kann auch dazu genutzt werden, jetzige und zukünftige Arten der Suchanfragen zu interpretieren und die korrekten Informationen automatisiert zu selektieren.

7 Literatur

- Drewer, Petra/Massion, François/Pulitano, Donatella (2017): Was haben Wissensmodellierung, Wissensstruktur, künstliche Intelligenz und Terminologie miteinander zu tun? 2017 DIT (Deutsches Institut für Terminologie e.V. http://dttev.org/images/img/abbildungen/DITeV_org_Terminologie_und_KI_2017_03_22_v2.pdf).
- Drewer, Petra/Ziegler, Wolfgang (2013): Technische Dokumentation: Übersetzungsgerechte Texterstellung und Content-Management. 2. Aufl. Würzburg: Vogel.
- Dublincore (2012): Dublin Core Metadata Element Set, Version 1.1. <http://dublincore.org/documents/2012/06/14/dces/>.
- Hanel, Stephanie (2016): Künstliche Intelligenz und Neurorobotik. Lindau Nobel Laureate Meetings. <http://www.lindau-nobel.org/de/kuenstliche-intelligenz-und-neurorobotik/>.
- Krüger, Manfred/Ziegler, Wolfgang (2008): Standards für strukturierte technische Informationen – Ein Überblick. In: Muthig, Jürgen (Hrsg.): Standardisierungsmethoden für die Technische Dokumentation. (=tekom Hochschulschriften 16), 11-40.
- Oevermann, Jan/Ziegler, Wolfgang (2016): Automated Intrinsic Text Classification for Component Content Management Applications in Technical Communication. In: Proceedings of the 2016 ACM Symposium on Document Engineering. DocEng '16, Vienna, New York City: ACM Press, 95-98.
- Richnow, Jörg/Rossi, Clemens/Wank, Helmut (2014): Kennzeichnung von Windkraftwerken mit dem Reference Designation System for Power Plants – RDS-PP®. VGB Powertech 7, 1-9. <https://www.vgb.org/vgbmultimedia/RICHNOWPT714DEU-p-8846.pdf>.
- S1000D (o.J.): International specification for technical publications using a common source database. <http://www.s1000d.de/>.
- Wetzel, Michael (2014): Multilinguale Taxonomien mit Coreon – Wissens- und Sprachmanagement in einer Lösung. In: Drewer, Petra/Mayer, Felix/Schmitz, Klaus-Dirk (Hrsg.): Rechte, Rendite, Ressourcen. Wirtschaftliche Aspekte des Terminologiemanagements.
- Ziegler, Wolfgang (2013): Content Management in der Technischen Kommunikation. Ein Überblick. In: Hennig, Jörg/Tjarks-Sobhani, Marita (Hrsg.): Content Management und Technische Kommunikation. (=tekom Schriften zur Technischen Kommunikation 18), 11-25.
- Ziegler, Wolfgang (2016): Ein Fan von Klasse. In: technische kommunikation 4, 38-45.
- Ziegler, Wolfgang (2017): Verteilen leicht gemacht. In: technische kommunikation 3, 31-34.
- Ziegler, Wolfgang/Steurer, Stephan (2010): Mit PI-Mod dokumentieren. Standardisiertes Informationsmodell für den Maschinen- und Anlagenbau. In: technische kommunikation 6, 51-55.