

Informationen werden intelligent – ein Überblick

Dr. Jan Oevermann, plusmeta GmbH, Karlsruhe

Konzepte wie Intelligente Informationen, Content Delivery und iIRDS verbreiten sich rasant im Umfeld der Technischen Dokumentation – doch wie optimiert man den zielgerichteten und semantischen Zugriff auf die eigenen Inhalte? Der Beitrag soll einen Überblick geben, um die eigenen Ansätze, den semantischen Informationszugriff zu optimieren, sowohl einordnen als auch bewerten zu können.

Bedarf an intelligenten Informationen

Die digitale Transformation in der Industrie, neue Zielmedien und Endgeräte sowie immer größer werdende Datenmengen machen die gezielte Bereitstellung von Informationen zu einer Herausforderung. Technische Dokumentation wird oft in Formaten bereitgestellt, die keinen granularen und semantischen Zugriff auf die benötigten Informationen zulassen. In den meisten Fällen sind das PDF-Dokumente oder statische Online-Hilfen. Dadurch können viele Inhalte nicht in moderne Digitalisierungsszenarien integriert werden und sind aus Sicht einer effizienten Informationsrecherche schwer zugänglich.

Etablierte Methoden zur Erfassung, Strukturierung und Annotation von Technischer Dokumentation können bereits heute einen granularen semantischen Zugriff auf Inhalte über mehrere Quellen hinweg ermöglichen (vgl. Content Delivery) – ihre Umsetzung ist in der Regel jedoch mit einem erheblichen Arbeitsaufwand verbunden. Die nachträgliche „Semantifizierung“ von Bestandsdaten ist deshalb für die Industrie von großem Interesse, bringt aber auch viele Herausforderungen mit sich. Durch Innovationen im Bereich der Künstlichen Intelligenz und durch die Methoden des semantischen Webs ist es möglich, Teile dieser Aufbereitung zu automatisieren und so personellen Arbeitsaufwand zu verringern.

Optimierungen des Informationszugriffs

Auf Basis von etablierten Konzepten innerhalb der Technischen Kommunikation lassen sich vier Ebenen definieren, auf denen die Eigenschaften von Inhalten auf die Bedürfnisse des Information Retrieval optimiert werden können und damit einen gezielteren Zugriff auf Informationen zulassen. Daraus ergeben sich vier grundlegende Ebenen für die Optimierung: Content, Struktur, Metadaten, Ontologien & Semantische Netze oder kurz: **CoSMOS** (vgl. Oevermann, 2019).

Content

Eine sprachliche Inhaltsebene befasst sich mit dem Text – unabhängig von seiner Makrostruktur: dem Content. Neben der Standardisierung einzelner Informationseinheiten ist auch die Konsistenz innerhalb des Inhaltsbestands wichtig, wozu z. B. die Datenqualität zählt.

Struktur

Die Strukturierung Technischer Dokumentation ist speziell in Verbindung mit Content-Delivery-Konzepten von enormer Bedeutung für den granularen Zugriff auf relevante Inhalte. Neben der strukturierten Erstellung in Informationsmodellen ist hier besonders die nachträgliche Aufbereitung monolithischer Dokumente von Interesse.

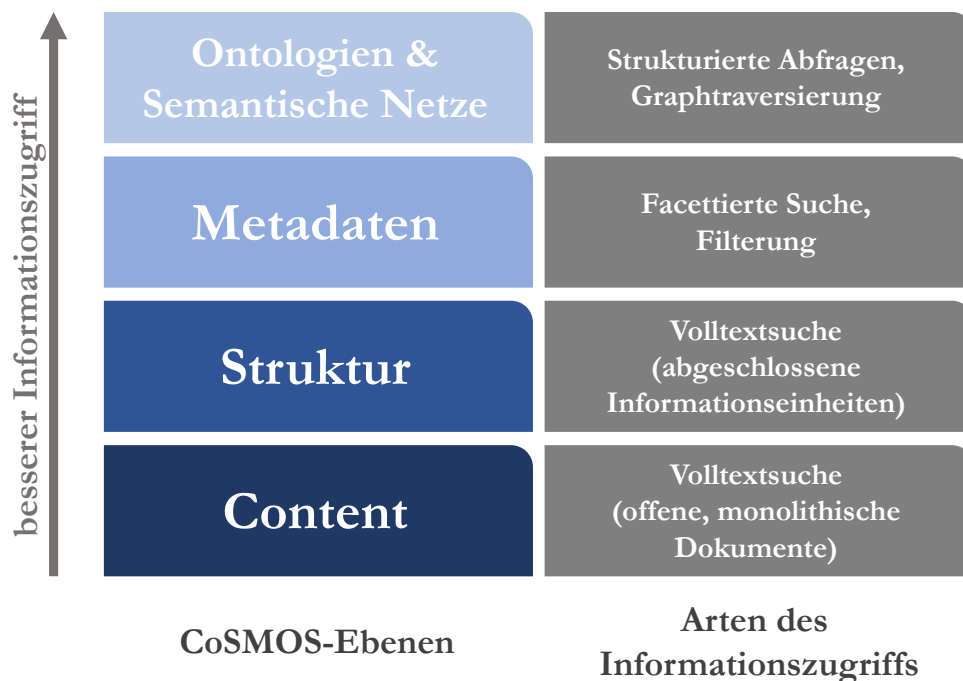
Metadaten

Metadaten für Content sind in der Technischen Kommunikation immer schon dazu verwendet worden, Content besser zu organisieren und über eine gezieltes Information Retrieval (z. B. in einem CMS) zugänglich zu machen. Eine systematisierte Entwicklung dieser Klassifikationen oder Metadatenmodelle sowie die korrekte und effiziente Zuweisung sind die größten Herausforderungen in diesem Bereich.

Ontologien und semantischen Netze

Die logische Fortführung dieser zusätzlichen Annotation von Inhalten sind die Prinzipien von Ontologien und semantischen Netzen: Metadaten können zueinander und mit Inhalten in verschiedenen Beziehungen stehen und sind universal definiert, um einen einfachen Austausch zu ermöglichen. Durch die netzartigen Relationen kann z. B. neues Wissen gewonnen werden und alternative Zugriffsmöglichkeiten auf die Informationen sind durch strukturierte Abfragen möglich.

Übersicht (CoSMOS)



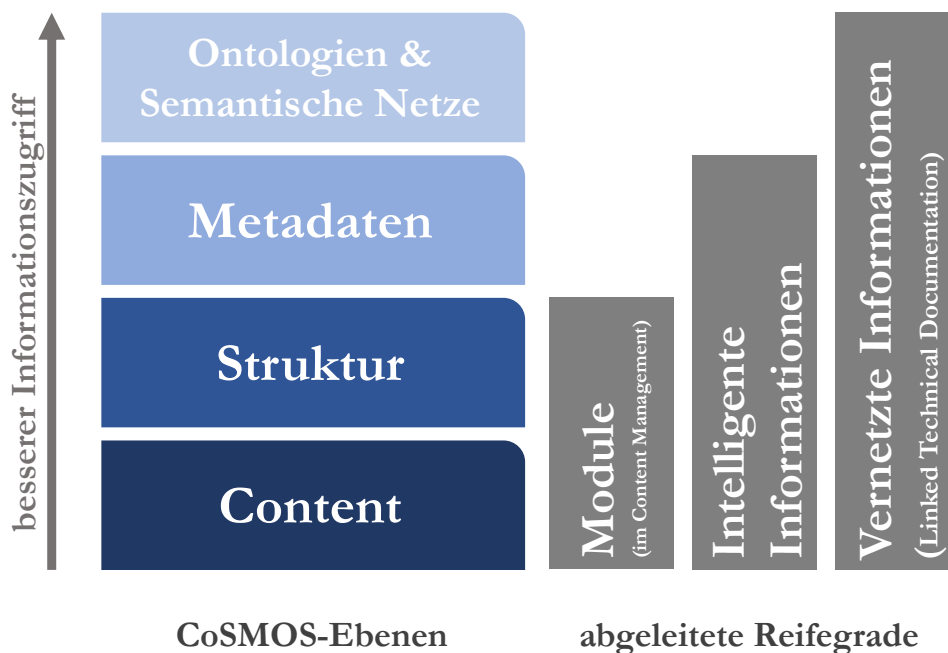
CoSMOS-Übersicht mit zugeordnetem Informationszugriff (übernommen aus Oevermann, 2019)

Beispielhafte Optimierungen

In der Technischen Dokumentation sollten zunächst empfohlene und einfache Optimierungen (sog. „low hanging fruits“) über alle der oben genannten Ebenen hinweg umgesetzt werden, um schnell zu einem optimalen semantischen Informationszugriff zu gelangen. Das sind beim Content eine benutzerorientierte Terminologie und standardisierte Formulierungsmuster mit Hilfe von TD-Standardisierungsmethoden sowie SEO (Suchmaschinenoptimierung), bei der Struktur eine topic-basierte Modularisierung (PI-Mod, DITA etc.), bei den Metadaten eine systematische Klassifizierung (z. B. basierend auf der PI-Klassifikation) und auf der Ontologie-Ebene die Verwendung einer Domänenontologie zur Verknüpfung und Annotation der Inhalte (z. B. iIRDS) und entsprechender Systeme.

Reifegrade intelligenter Informationen

Das CoSMOS-Modell kann auch zur qualitativen Einordnung von Technischer Dokumentation verwendet werden, um so Unternehmen eine Möglichkeit zu geben, die eigenen Informationsprozesse zu beurteilen.



CoSMOS-Ebenen mit abgeleiteten Reifegraden für TD (übernommen aus Oevermann, 2019)

Literatur

- Oevermann, Jan (2019): Optimierung des semantischen Informationszugriffs auf Technische Dokumentation. tekomp Hochschulschriften (Band 25). Stuttgart : tcworld.

Kontakt:
jan@plusmeta.de