



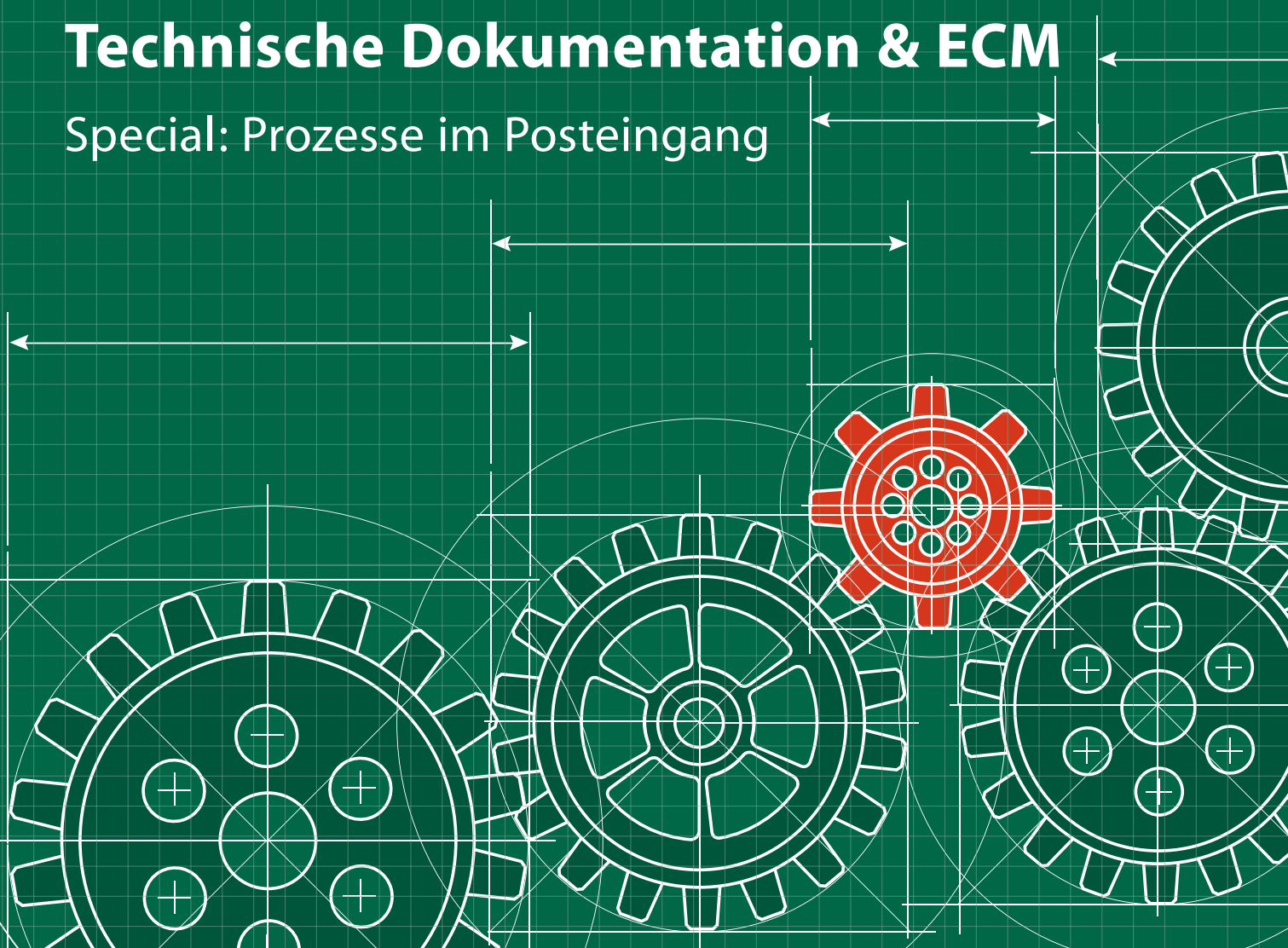
DOK

Technologien, Strategien & Services für das digitale Dokument

Wikis als Intranet 2.0 - ein Erfahrungsbericht
E-Mail-Richtlinien in der betrieblichen Praxis
Softwaregestützte E-Discovery-Prozesse

Das Beste aus zwei Welten: Technische Dokumentation & ECM

Special: Prozesse im Posteingang



„Pimp my ECM“ - thematische Suche für intelligentes Enterprise Content Management

Enterprise Content Management (ECM), Enterprise Search, Social Tagging, Tag Clouds, „facettierte Suche“, Ontologien

Enterprise Content Management (ECM) soll den effizienten Umgang mit den Informationsbeständen des Unternehmens sichern oder, wie der frühere Hewlett-Packard CEO Lew Platt es ausdrückte: „If HP knew what HP knows, we would be three times as profitable“. Auf Deutsch: Wenn Siemens wüsste, was Siemens weiß. ECM Werkzeuge und Strategien sollen gemäß der Definition durch die Association for Information and Image Management (AIIM) das Management der unstrukturierten Information eines Unternehmens ermöglichen. Eine vollständige ECM-Architektur muss Erfassung, Management, Speicherung, Archivierung und Auslieferung von Inhalten und Dokumenten leisten, die sich auf die Prozesse der Organisation beziehen. Obwohl in jedem Unternehmen ein ECM im weiteren Sinne existiert, werden meist nur Insellösungen für einzelne Teilbereiche statt einer umfassenden ECM-Architektur realisiert.

Geradezu stiefmütterlich wird das Problem des Auffindens und Auslieferens von Dokumenten behandelt, die bei der Informationsrecherche gesucht werden. Zwar wird der intelligente Zugriff auf Dokumente als eine der Hauptfunktionen in ECM-Lösungen angesehen, etwa wenn im Unterschied zu einem Filesystem von einer Klassifizierung und Kategorisierung gesprochen wird, die das Finden der Dokumente erleichtert. Jedoch ist substantiell nur wenig Innovatives zu finden. Das Gros der angebotenen Systeme verlässt sich auf die etablierte Volltextsuche, wo für jedes Dokument eine Menge von unverbundenen Indextermen abgelegt wird. Das Problem an dieser Form des Zugangs ist: Sie vermittelt keinerlei Bild des Dokumentenbestandes. Der Anwender wird bei der Auswahl der Suchbegriffe mit dem weißen leeren Suchfenster allein gelassen.

Suchfenster

Eine weitere Unzulänglichkeit der Standardsuche betrifft das Ranking der Dokumente, das in vielen internen Informationssys-

www.topiczoom.de

Christoph Ringstetter ist wissenschaftlicher Mitarbeiter am Center for Language and Information Processing der Ludwig-Maximilians-Universität München und freier Mitarbeiter von **TopicZoom**. Das 2008 gegründete Spin-off der Universität München entwickelt Lösungen, in denen semantische Sprachtechnologien für die Wissenspräsentation und Recherche genutzt werden. TopicZoom beschäftigt 5 Mitarbeiter.

temen eher zufällig erscheint und nur wenig hilfreich ist. Die organisationsinterne Suche kann nicht wie eine Internetsuchmaschine auf intelligente Algorithmen zurückgreifen, welche die zu bevorzugende Dokumente anhand der Zahl ein- bzw. ausgehender Links sowie der gesammelten Nutzerdaten erkennt. Auch wenn von den großen Suchmaschinenanbietern Fortschritte bei der Enterprise Search durch Filter und Möglichkeiten des Customizing zu erwarten sind, werden solche Verbesserungen nicht ausreichen, den effektiven Informationszugang innerhalb einer Organisation sicher zu stellen.

Tagging hilft nur begrenzt weiter

Die dargestellten Schwächen der Volltextsuche können auch nicht durch das viel diskutierte Social Tagging beseitigt werden, bei dem die Anwender ein Dokument mit einem oder mehreren Schlagwörtern versehen, eine Funktionalität, die viele populäre Webseiten nutzen. Der Software-Anbieter Jumper Networks strickt sogar seine ECM-Lösung um dieses Prinzip herum. Liefert kollaboratives Tagging immerhin die Verbindung eines Dokuments zu einem oder mehreren Zentralbegriffen, so ist neben den Ungenauigkeiten, die aus Schreibfehlern und Benennungsvarianten resultieren, das eigentliche Hauptproblem, dass dieses Tagging „flach“ ist. Es werden damit eben keine tiefgehenden thematischen und hierarchischen Verbindungen zwischen den Begriffen hergestellt, eine unterstützende Navigation innerhalb des Dokumentenraums wird ebenfalls nicht erreicht. Immerhin kann durch Visualisierungstechniken wie Tag Clouds, einer Darstellung von Tags verbunden mit einer grafischen Hervorhebung gemäß ihrer Häufigkeiten, eine grobe Idee zu den in der Dokumentensammlung vorhandenen Inhalten vermittelt werden.

Zielführend adressiert wird das Informationsproblem, das sich bei einem fortgeschrittenen Enterprise Content Management

stellt, nur durch eine inhaltlich strukturierte Suche und Navigation. Der Dokumentenraum wird durch eine geeignete Metaresource, in diesem Fall eine Themenhierarchie erschlossen, indem die wichtigen Indexterme der Dokumente an thematische Knoten einer Wissensstruktur gekoppelt werden. Ein Knoten kann, insbesondere wenn er höher in der Hierarchie steht, auf sehr viele Indexterme verweisen. Die Knoten selbst sind durch Relationen, im gezeigten Fall durch eine hierarchische Relation verbunden. Der Benutzer kann nun entweder durch eine graphische Navigation oder durch eine mit Vorschlägen unterstützte Suche einen der Knoten auswählen und erhält als Ergebnis alle Dokumente, deren Indexterme an den ausgewählten Knoten oder einen seiner Nachfolger gekoppelt sind. Das Ranking der Dokumente erfolgt nach der Zahl der relevanten Indexterme und deren semantischen Abstand zum gewählten Knoten der Wissensstruktur. Als äußerst hilfreich hat sich die Anreicherung der Ergebnisdokumente durch thematische Profile und der relevanten Knoten mit Trefferzahlen erwiesen, als Alternative zu den aus den Suchmaschinen bekannten Vorschau-Fenstern, den Snippets. Der mit dem thematischen Vokabular vertraute Anwender kann mithilfe solcher Profile sofort die Relevanz der Dokumente beurteilen.

Neben dieser themenzentrierten Suche stellt die sogenannte „facetiierte Suche“ eine interessante Erweiterung dar. Hier wird der Dokumentenraum zunächst durch die Auswahl eines thematischen Knoten eingeschränkt, etwa das Thema „ECM“. Anschließend werden alle mit diesem Thema verbundenen Themen und deren Trefferdokumente angegeben. So kann schnell ein Überblick zu verbundenen Themen erreicht werden.

Bei allen Unterschieden im Detail zwischen den Lösungen der verschiedenen Anbieter im Bereich wissensbasierter ECM-Systeme: Sie greifen immer auf eine meist mit dem Begriff der Ontologie bezeichnete Wissensstruktur zurück, die Relationen zwi-

schen den relevanten Informationseinheiten einer Organisation herstellt. Da bislang automatisierte Verfahren zur Bildung des Kerns einer solchen Wissensstruktur nicht zur Verfügung stehen, müssen Zusammenhänge zwischen den Schlüsselkonzepten des Informationsraumes eines Kunden entweder manuell neu erstellt oder, falls sie schon einmal modelliert wurden, aus einer vorhandenen Struktur übertragen werden. Verschiedentlich wurde etwa die so genannte „Upper Ontology“, die eher allgemeine Zusammenhänge repräsentiert, als Ausgangspunkt für eine fallbezogene Spezialontologie verwendet. Eine andere Vorgehensweise ist die, eine sehr große enzyklopädische Ressource zu benutzen, um einen Dokumentenraum vorzustrukturieren und gemeinsam mit den späteren Anwendern die Binnensemantik aufgetretener Konzepte anzupassen und wo notwendig, die entstandene Ressource weiter zu spezialisieren, sie etwa mit Zusatzlayern und tiefergehenden Konzepten auszustatten.

Die Investition in Technologie wird eine Investition in Wissen

In jedem Fall werden in einem wissensbasierten ECM-Projekt umfangreiche organisationsspezifische Anpassungsarbeiten erforderlich. Schnell können hunderte von Arbeitsstunden anfallen, um eine erste Version der Wissensstruktur aufzubauen. Auch der spätere Wartungsaufwand sollte großzügig budgetiert werden: Sobald dafür nicht ausreichende Mittel vorgesehen sind kann der Nutzen der Maßnahme je nach informationeller Dynamik der Geschäftsprozesse des Kunden schnell abnehmen. Die Einführung einer solchen Technologie in der Organisation muss in jedem Fall als strategische Entscheidung betrachtet werden. Kosten und Risiken eines intelligenten Enterprise Content Managements wie auch dessen Erträge sollten in Bezug zu den klassischen Reorganisationsmaßnahmen gesetzt werden.

Auf der Seite der Aktiva ist nicht zu unterschätzen, dass die Entscheidung für eine explizite Modellierung der Informationsbedürfnisse, einschließlich einer zugrundeliegenden Wissensstruktur, vom ersten Projekttag an zu einer Vertiefung des Wissens der Organisation über sich selbst führen kann. In wissensintensiven Unternehmen wie in der Beratungsbranche oder in solchen mit einem überproportionalen Forschungs- und Entwicklungsanteil kann ein Projekt zur Integration eines wissensbasierten ECM von Anfang an erkennbaren Mehrwert erbringen. Steht erst einmal die Struktur zur Verfügung, ergeben sich durch Verzahnung mit einem ebenfalls wissensbasierten System zum Personalmanagement und einer gezielten Presse- und Internetbeobachtung vollkommen neue Möglichkeiten des strategischen Managements. Es bleibt zu hoffen, dass die Möglichkeiten wissensbasierter ECM-Systeme erkannt und genutzt werden, damit die Unternehmen nicht einen Quantensprung innerhalb der Informationsökonomie des 21. Jahrhunderts verpassen. ■