

Vom AWW Arbeitskreis 6.1, Projektgruppe 5, Redaktion: Dr. Hubert Salm

Informationsmanagement in der Praxis

Vom Dokumentenmanagement zum ganzheitlichen Informations- und Wissensmanagement

Die digitale Revolution und das „papierlose Büro“

Die erwartete „digitale Revolution“ hat stattgefunden. Nach dem Ende des Wirtschaftsbooms, der nicht zuletzt auch durch die Erwartungen auf neue Märkte durch neue Technologien angetrieben wurde, führen dennoch die Innovationskräfte der Informations- und Kommunikationstechnologien zu tiefgreifenden Veränderungen, deren Auswirkungen wir oft erst deshalb spät zur Kenntnis nehmen, weil die Geschwindigkeit der Innovationszyklen unsere Anpassungsfähigkeit oftmals überfordert.

So ist die vielbeklagte „Informationsflut“ nicht zuletzt darauf zurückzuführen, dass die digitalen Produktions- und Kommunikationsmechanismen eine unerhörte Produktivität freigesetzt haben, denen aber kein Äquivalent bei der Verwaltung der Informationen entgegengesetzt wurde.

Vielfach haben sich auch Prognosen und Erwartungen nicht bestätigt. Und das nicht etwa wegen Mängeln in der Realisierung, sondern weil die Erwartungen eben schlicht falsch waren. So gehört das Schlagwort vom „papierlosen“ Büro ebenso zu den überholten Relikten einer zurückliegenden Epoche der Technikentwicklung wie die 5 1/4“ „Floppy“ oder das 8“ Magnetplattenlaufwerk.

Während sich die Produktion und Verbreitung von „Schriftsätzen“ schon längst auf digitale Mechanismen verlagert hat, sind wir für die kurzfristige Visualisierung und

Speicherung und oft auch für die Verfolgung der Bearbeitung von Schriftsätzen auf das Papier angewiesen. Das Nebeneinander der digitalen und der analogen Verarbeitungsformen rührt nicht etwa daher, dass die technischen Plattformen nicht über ausreichende Kapazitäten verfügen. Es fällt uns einfach schwer, die über Jahrhunderte erprobten Bahnen zu verlassen und neue Konzepte zu entwickeln, die den analogen Verfahren und deren intuitiver Zugänglichkeit zumindest ebenbürtig sind.

Denn erstens erschienen die analogen Mechanismen ausreichend, solange sie funktionierten, und zweitens waren und sind die angebotenen digitalen Verfahren bezüglich ihrer Funktionalität und Wirtschaftlichkeit nicht immer überzeugend.

Für die Realisierung wirklich überzeugender digitaler Verfahren ist ein Geniestreich nicht in Sicht. Die Anforderungen an Software, die im Umfeld des Informationsmanagement gestellt werden müssen, sind anscheinend widersprüchlich: Eine mächtige Funktionalität soll sich mit einfachster Handhabung verbinden und sich dabei ständig neuen Organisationsformen anpassen, ohne veraltete Organisationsstrukturen zu zementieren. Sie soll den Anwender anleiten und steuern, jedoch nicht bevormunden. Sie soll sich an das persönliche Wissens- und Funktionsprofil anpassen und dabei gleichzeitig den Belangen des Daten- und Persönlichkeitsschutzes entsprechen. Dabei sollte Software trotz der kurzen Innovationszyklen aus

Investitionsschutzgründen langlebiger als andere Investitionsgüter sein.

Die Verfahren zur Ordnung und Strukturierung von Informationen sind immer noch sehr nah an den damit klassischerweise verbundenen Medien wie Papier oder Mikrofilm ausgerichtet. Doch heutzutage wird der weitaus überwiegende Teil der Informationen bereits elektronisch erzeugt. Dementsprechend sind nunmehr Verfahren gefordert, die die Verwaltung von digitalen Objekten zulassen und flexibel auf unterschiedliche Anforderungen eingehen, ohne dabei aufwendig in der Pflege oder komplex in der Bedienung zu sein.

IT-Hardware, Netzwerke, digitale und analoge Speichermedien wie Festplatten und Bänder, optische und fotografische Medien, spielen für die Organisation von Daten nur eine (bedeutende) Nebenrolle. Für die Erschließung der Goldadern und Abraumhalden unserer Datenüberflusgesellschaft sind wir auf die kontinuierliche Weiterentwicklung der Softwaresysteme angewiesen. Die Entwicklung des geeigneten Instrumentariums kann nur in den möglichst kurzen Zyklen zwischen Innovation und Markterprobung erfolgen. In dieser Prozesskette wird nicht nur von den Herstellern von Softwaresystemen erwartet, dass sie innovative Produktentwicklungen mit hohen Qualitätsstandards kostengünstig vorantreiben. Innovationsbereitschaft, Flexibilität der Organisation und Bereitschaft zu ständigem Lernen muss Teil der Unternehmenskulturen sein, die

sich in diesem innovativen Feld behaupten wollen.

Dokumentenerfassung

Bei elektronisch erzeugten, ggf. sogar selbst erstellten digitalen Objekten kann man mit einfachen Methoden für eine entsprechende neue und flexible Organisationsform sorgen. Bei der Übernahme extern erzeugter digitaler Information und spätestens bei der Betrachtung von eingehenden Papier- und Fax-Dokumenten wird deutlich, dass der Medienbruch zwischen analoger Welt (Papier, Fax, Mikrofilm) und der digitalen Welt der erste Hinderungsgrund ist, eine gemeinsame Basis zur Verwaltung aller Informationen zu schaffen.

Durch die Einführung automatisierter Verfahren sinken die Erfassungskosten signifikant. Dabei muss nach wie vor, je nach Art der zu erfassenden Dokumente und Qualität der automatisierten Erfassungsverfahren, die manuelle Nachbearbeitung in eine Kostenbetrachtung mit einbezogen werden.

Volltext-Systeme werten den Inhalt von textlichen digitalen Objekten in der Form aus, dass der beinhaltete Text gegen Stop-Wortlisten abgeglichen wird und die danach als sinntragend erkannten Worte in einer Volltext-Datenbank abgelegt werden. Der Vorteil von Volltextsystemen liegt darin, dass sie grundsätzlich ohne eine vorgegebene Metadatenstruktur auskommen und daher sehr flexibel sind. Über die Definition von Stopwortlisten kann verhindert werden, dass die Volltextdatenbank mit Worten aufgefüllt wird, auf die keine sinnvolle Suche möglich erscheint (z.B. Worte mit weniger als vier Buchstaben, Präpositionen, etc.)

Die Schwächen von Volltextsystemen werden dann deutlich, wenn Worte sprachlich mehrfach belegt sind (z.B. die Eingabe des Wortes

„Bank“ führt zu Treffern aus den Bereichen Finanzinstitute, Sitzgelegenheiten und Sandanschwemmungen) und damit eine Eingrenzung schwierig wird. Weiterhin scheitern die Systeme daran, dass das eigentliche Suchwort im Text nicht enthalten ist, obwohl der Text sich mit dem gesuchten Thema beschäftigt (z.B. findet das Wort „Straßenbahn“ nicht die Dokumente, die das Wort „Trambahn oder Tram“ benutzen).

Thesaurussysteme sind dazu in der Lage, sprachliche Gleichheiten auszuwerten und damit zu einer höheren Treffergenauigkeit zu kommen. Somit werden alle Anfragen zum Suchwort „Wendel“ automatisch in das Suchwort „Schnecke“ transformiert und auch hiermit der Bestand durchsucht. Die Pflege von Thesaurussystemen ist allerdings aufwendig und kann nur in einem bestimmten Umfeld eingerichtet werden, in denen die sprachlichen Entsprechungen klar definiert werden können.

Verfahren, die aus nicht textbasierten digitalen Objekten Metadaten extrahieren, sind in einigen, wenigen Spezialanwendungen zu finden. Auch hier werden Verfahren der Mustererkennung eingesetzt, um z.B. aus Audiodaten bestimmte Worte oder Geräusche zu identifizieren, oder aus einer Sammlung von Bilddaten diejenigen heraus zu suchen, auf denen ein Baum abgebildet ist. Der Rechenaufwand für die Extrahierung der Metadaten ist in diesen Verfahren immens, die Treffergenauigkeit abhängig vom Muster, das gesucht werden kann.

Die oben kurz beschriebenen Verfahren können sich, je nach Anwendungsgebiet, gegenseitig sinnvoll ergänzen. Im Informationsmanagement sollten daher alle beschriebenen Verfahren verfügbar sein, um sie im jeweiligen Kontext der Organisation zielgerichtet einsetzen zu können.

Die Strukturierung von Informationen und Dokumenten

Während die vorgenannten Verfahren Metadaten direkt aus dem digitalen Objekt beziehen, werden Metadaten, die das digitale Objekt beschreiben, auf andere Arten gewonnen.

Metadaten wie Verfasser, Erstellungszeitpunkt, Dateiformat, Dateiname, Speicherort, Zugriffsbeschränkungen können über das Dateisystem zur Verwaltung der digitalen Objekte bezogen werden. Inhaltsbeschreibende Daten sind oftmals auch Bestandteil des Dateinamens, der somit auch einen Rückschluss auf den Inhalt zulässt.

Weitere Informationen sind aus dem Aufbau von Dokumenten zu beziehen, so zum Beispiel die verwendete Softwareversion oder die erzeugende Quellapplikation. Mit Hilfe dieser Metadaten können digitale Objekte zusätzlich beschrieben und in Relation gesetzt werden.

Ein weiteres, auf Grund der heutigen Rechnerperformance mögliches Verfahren, ist die Bestimmung von Ähnlichkeiten zwischen digitalen Objekten. Hiermit kann bestimmt werden, in wie fern digitale Objekte ähnlich im Sinne von inhaltlich deckungsgleich sind. Diese Verfahren eignen sich zum Beispiel zur Klassifikation von unbekannten digitalen Objekten in Klassen. Die Klassifikationsverfahren ermitteln die Ähnlichkeiten mit bereits bekannten digitalen Objekten und klassifizieren die unbekannteren digitalen Objekte bei signifikanter Übereinstimmung als zu dieser Klasse zugehörig.

Moderne Wissens-Management-Systeme nutzen diese Technologien, um dem Benutzer Informationen zur Verfügung zu stellen, die in einem großen Informationsspeicher vorgehalten werden. In der Regel sind diese Informations-

speicher die Fileserver der Unternehmen oder elektronische Archiv- und Dokumenten-Management-Systeme.

Eine besondere Herausforderung bei der Konstruktion des „richtigen“ Ordnungsschemas stellt die Dynamik der Informationsbestände und ihres jeweiligen Anwendungskontextes dar:

Ordnungsschemata verändern sich über die Zeit entsprechend der Interessen der Informationssuchenden und -verarbeitenden Personen.

Das Retrieval, also das Wiederfinden der „richtigen“ Informationen, gestaltet sich trotz der vorgegebenen Ordnungsschemata oft schwierig. Das inhärente Problem kennt jeder, der beispielsweise über einen längeren Zeitraum ein Dateisystem als „Austauschplattform“ für die Dokumente einer Arbeitsgruppe organisiert oder in einem Archivbestand, der nach einer überholten Aktensystematik abgelegt ist, Informationen gesucht hat: Ordnungsschemata „veralten“ mit zunehmender Lebensdauer bzw. verändern sich. Da große Informationsbestände auch nicht „mal eben schnell“ manuell reorganisiert werden können, nahm man in der Vergangenheit von einer Anpassung, z. B. einer Reorganisation des Dateisystems, meist Abstand. An die Anpassung von Schemata an die Interessen einzelner Informationssuchender war dabei schon gar nicht zu denken.

Gerade der letzte Punkt, also die Ordnung von Informationen nach individuellen Bedürfnissen („persönliche Wissensportale“), stellt eine interessante Perspektive in unserer informationsüberfluteten Zeit dar. Hier ermöglichen es automatische Verfahren, sich von vorgegebenen starren Ordnungsschemata zu lösen und statt dessen problembezogene Informati-

onsfilter aufzubauen, die dem Informationssuchenden schnell eine individuelle Sicht auf Informationen bieten. Durch die Automatisierung der Klassifizierung wird es zudem möglich, dass Informationen aus dem täglichen Strom automatisch gefiltert oder, wenn man einem solchen System doch nicht

formationen grundsätzlich bereits strukturiert sein werden.

Proprietäres „Dokumentenmanagement“ wird infolge der WEB- und XML-Mechanismen durch das Management von Informationsstrukturen abgelöst, die als Teilmenge auch das klassische Doku-

#	Format	Layout	Medium
1	PDF, TIFF etc.	Fest	Papier
2	PDF, TIFF etc.	fest	Mikrofilm
3	PDF, TIFF etc.	fest	Opt. Speicher, Magnetspeicher
4	ASCII, TIFF etc.	variabel	Opt. Speicher, Magnetspeicher
5	XML, TIFF etc.	variabel	Opt. Speicher, Magnetspeicher

Sofort lesbar

Weiterverarbeitung

ganz „trauen“ will, zumindest nach Interessenlage gewichtet werden. Dadurch erreichen uns zuerst die Informationen (Push-Technologie), die auch tatsächlich unseren Interessen entsprechen.

Technische Strukturierung von Informationen

Das Ziel der nachträglichen Strukturierung besteht darin, entweder Nutz-Daten für die Weiterverarbeitung (Buchungssatz einer Rechnung erfassen) oder Meta-Daten für die Indexierung (Retrieval in einem DMS) aus den digitalen Objekten zu extrahieren und diese Daten soweit möglich und sinnvoll um zusätzliche bereits vorhandene Informationen (aus ERP, CRM,) zu ergänzen.

Die Aufgaben einer nachträglichen Strukturierung verlieren an Bedeutung, da durch den Einsatz von Standards wie EDI, Edifact und insbesondere XML die digitalen In-

formationen einschließlich der Handhabung von Geschäftsprozessen beinhalten. Für die Speicherung von Dokumenten stehen Formate und Speichermedien zur Verfügung, die, abhängig von den jeweiligen Anforderungen, sowohl die Veröffentlichung als auch die Archivierung in idealer Weise unterstützen.

Dieser Artikel bietet einen Einstieg in grundlegende Fragestellungen aktueller und zukünftiger Systeme des Informationsmanagement. Eine ausführlichere Darstellung finden Sie in der AWW-Schrift „Informationsmanagement in der Praxis – Vom Dokumentenmanagement zum ganzheitlichen Informations- und Wissensmanagement“, die im Internet unter der Adresse http://www.awv-net.de/schriften/online/inf_man.html abgerufen werden kann.

Dr. Hubert Salm, MACH I.T. Unternehmensberatung GmbH, Düsseldorf ist Vorsitzender des AWW-Fachausschusses 6 „Informationswirtschaft“.