

Evolutionäres Vorgehen erforderlich

Erschließung, Nutzung und Bewahrung von Information durch DMS

Die effektive Erstellung, Auswertung, Verteilung und die effiziente Archivierung von Information wird zunehmend Voraussetzung für Unternehmenserfolge. Document-Management-Systeme (DMS) können diese Voraussetzungen schaffen. Erforderlich ist eine bedarfsgerechte, den Erfordernissen des gesamten Unternehmens exakt angepaßte Lösung unter Einsatz zeitgemäßer DV-Technologie.

Vor wenigen Jahren wurde die Vision des papierlosen Büros in der DV-Branche propagiert. Sie war in vielen Fällen der Anlaß, Arbeitsplätze mit PC's auszustatten. Der PC als Arbeitsmittel sollte einen wichtigen Schritt zum papierlosen Büro darstellen. Textverarbeitungssysteme sollten die Digitalisierung bei der Schriftguterstellung ermöglichen. Die Vision des papierlosen Büros ist in vielen Fällen eine Vision geblieben und wird es wohl auch auf absehbare Zeit bleiben. Die Probleme, die das papierlose Büro lösen sollte, sind ebenfalls geblieben und wurden unter anderem durch die Modernisierung der Arbeitsmittel für die Schriftguterstellung noch verschärft. Document-Management bedeutet nicht das Wiederbeleben dieser Vision in neuer Form, sondern verfolgt den Ansatz, die Probleme papiergebundener Information zu lösen.

Der PC am Arbeitsplatz, leistungsfähige Datenbanken und moderne Kommunikationsmittel führen dazu, daß immer mehr Information erstellt werden kann. Die moderne Informationsgesellschaft erstickt in ihrer eigenen Informationsflut. Papier ist trotz des technischen Fortschritts weiterhin der verbreitetste Datenträger am Arbeitsplatz.

Leere Kassen der öffentlichen Hand, Kostendruck in Industrie und Wirtschaft, Verdrängungswettbewerb und die Globalisierung der Märkte führen dazu, daß der Umgang mit Information einer kritischen Betrachtung unterzogen werden muß. Information ist Kapital und dieses Kapital muß möglichst gewinnbringend eingesetzt werden können. Die Hauptaufgabe von Document-Management ist also, dieses Kapital sinnvoll zu erschließen, zu nutzen, zu bewahren und zu vermehren.

Das Szenario einer papiergebundenen Informationsverarbeitung ist im Wesentlichen unverändert. Die Führung von Papierakten mit ihrem großen Raumbedarf bei der Lagerung, der unmögliche gleichzeitige Zugriff mehrerer Nutzer auf die gleiche Information, redundante Informationen durch die unkontrollierbare Anzahl möglicher Kopien und damit die doppelte Ablage von Information, die hohen Durchlaufzeiten von Informationen im Unternehmen bei Rundschreiben und die Unterbrechung von Informationswegen durch Abwesenheitszeiten der Mitarbeiter bringen die Unternehmensorganisation an den Rand des Kollaps. Das Ablegen von Information auf Einzelplatz-PCs verhindert sogar einen arbeitsplatzübergreifender Zugriff auf digital vorliegende Informationen.

Evolutionäres Vorgehen erforderlich

Erschließung, Nutzung und Bewahrung von Information durch DMS

Der Einsatz von DV-Technik bei der Schriftguterstellung führt dazu, daß die Schriftgutverwaltung immer aufwendiger wird, da die Papierberge in kürzerer Zeit stärker anwachsen können, als dies in der Vergangenheit möglich war.

Die DV-Branche bietet seit geraumer Zeit technische Lösungen an, die helfen sollen, die verschiedenen Probleme der Informationsgewinnung, -verteilung, -auswertung und -archivierung beherrschbar zu machen. Im Wesentlichen handelt es sich hierbei um Archivierungs-, Scan-, Text-Retrieval- und SGML-Systeme (Standard Generalized Markup Language). Alle diese Systeme werden unter dem Begriff DMS vermarktet. Für den DV-Entscheider, der Gesamtlösungen sucht, ist diese Begriffsunschärfe nicht hilfreich. Es fehlt eine allgemeingültige Begriffsdefinition von DMS, die es dem DV-Entscheider ermöglicht, das für seine Bedürfnisse passende DMS zusammenzustellen. Ein geeigneter Denkansatz ist es, die Problematik des Document-Managements in die Problemfelder Erschließung, Nutzung und Bewahrung von Information zu zerlegen. Darüber wird es möglich, die technischen Systeme Problemfeldern zuzuordnen und funktional von einander abzugrenzen. Nur über eine solche Zuordnung und Abgrenzung kann verhindert werden, daß z.B. ein Archivierungssystem mit den Anforderungen an ein Text-Retrieval-System bemessen und bewertet wird.

Was kann ein DMS bei gesamtheitlicher Betrachtung leisten? Die Leistung eines DMS liegt in der Aufbereitung, Bereitstellung und Auswertung sowie Aufbewahrung von Information. Die klassischen Methoden und Techniken der Schriftgutverwaltung bieten diese Leistungen an, so daß es dem Grunde nach keine Organisation ohne DMS gibt. Allerdings reicht die Leistungsfähigkeit dieser DMS nicht aus, die heutigen Anforderungen an die Informationsverarbeitung zu erfüllen.

Die angebotenen DV-Systeme im Umfeld von DMS können den Bereichen Informationsgewinnung, Informationsverteilung, Informationsretrieval und Informationsarchivierung zugeordnet werden. Bei dieser Klassifizierung ergeben sich zum Teil funktionale Überschneidungen, die durch die Leistungen der marktverfügbaren DV-Systeme bedingt sind. So beinhaltet beispielsweise eine Datenbank neben der Retrieval-Komponente eine Archivierungskomponente. Diese Überschneidungen sollten bei der Gestaltung eines DMS immer berücksichtigt werden.

Informationsgewinnung

Damit Information DV-technisch verarbeitet werden kann, muß eine Digitalisierung erfolgen, soweit sie nicht schon digitalisiert vorliegt. Eine durchgängige DV-technische Lösung muß Medienbrüche im Arbeitsablauf vermeiden und auf das unumgängliche Minimum reduzieren. Die Digitalisierung von papiergebundener Information kann durch Scan-Systeme geleistet werden.

Die Einführung eines Scan-Systems hat weitreichende Auswirkungen auf die Unternehmensorganisation. Es sind Festlegungen zu treffen, welche

Evolutionäres Vorgehen erforderlich

Erschließung, Nutzung und Bewahrung von Information durch DMS

Arbeitsabläufe von Papier als Informationsmedium „befreit“ werden. Die Einführung eines Scan-Systems bedingt neben der Investition in das Scan-System unter Umständen Folgeinvestitionen in die betroffenen Arbeitsplätze, da dann alle Arbeitsplätze innerhalb des festgelegten Arbeitsablaufs mit entsprechenden Arbeitsmitteln ausgestattet sein müssen. Es sind Lösungen erforderlich, die den Umgang mit den bereits bestehenden Papierarchiven regeln (Digitalisierung des gesamten Archivs oder von Teilen des Archivs oder aller Dokumente ab dem Tag der Einführung des Scan-Systems). Bei der Entscheidung für ein Scan-System sind neben dem Umfang der einzulesenden Informationen und der Vorlagenqualität, die direkten Einfluß auf die zu beschaffenden Hardware haben, Entscheidungen über die weitere Verwendung der Information zu treffen. Es ist nicht immer ausreichend, die anfallende Bildinformation weiterzuverarbeiten. Durch Einsatz von OCR-Software kann aus der von der Scan-Software gelieferten Bildinformation eine Textinformation gewonnen werden. Der OCR-Prozeß ist relativ zeitaufwendig, da jede Scan-Vorlage erneut gelesen und auf das Vorhandensein von Text geprüft werden muß. Die Texterkennung selbst ist nicht fehler- oder zweifelsfrei. Selbst wenn das OCR-System eine 98%ige Texterkennung leistet, verbleiben 2% Textinformationen, die nicht, zweifelhaft oder fehlerhaft erkannt werden. Besondere Problemfelder stellen hier Umlaute und Sonderzeichen dar. Eine Handschriftenerkennung ist derzeit nahezu unmöglich. Das bedeutet, daß die Texterkennung ebenfalls in ihrer Einsatzmöglichkeit stark von der Qualität der Vorlage abhängig ist.

Die Aufbereitung von Information bedeutet aber nicht nur, papiergebundene Information zu digitalisieren. Vielmehr muß eine konzeptionelle Unterstützung zur Bereitstellung von Information erfolgen, damit entsprechend einfache und wirksame Suchmechanismen implementiert werden können.

Informationsverteilung

Eine optimale Bereitstellung von Information liegt vor, wenn die richtige Information zum richtigen Zeitpunkt dem richtigen Mitarbeiter zur Verfügung steht. Dazu werden leistungsfähige DV-Systeme benötigt. Es besteht die Notwendigkeit, die richtige Information wiederfinden und heraussuchen zu können. Diese Leistung kann durch konventionelle Datenbank- und/oder Text-Retrieval-Systeme erbracht werden. Gleichzeitig soll die gefundene Information ohne Zeitverlust den richtigen im Sinne von zuständigen Mitarbeitern zugänglich gemacht werden. Das DMS muß also Kenntnis über instanzuelle Zuständigkeiten und Arbeitsabläufe haben, um diese Leistung zu erbringen. Die DV-technische Umsetzung kann durch Einsatz eines Workflow-Management-Systems erfolgen.

Die Aufgabe eines Workflow-Management-Systems sollte sich dabei auf den Informationstransport innerhalb der bestehenden Arbeitsabläufe beschränken. Die Vermischung mit Retrieval-Funktionen bedeutet immer einen technischen

Evolutionäres Vorgehen erforderlich

Erschließung, Nutzung und Bewahrung von Information durch DMS

Kompromiß, da das Workflow-Management-System nicht die Leistungsfähigkeit einer Datenbank oder eines Text-Retrieval-Systems bieten kann.

Informationsretrieval

Die Suche nach Informationen kann von Datenbanken, Text-Retrieval- und SGML-Systemen DV-technisch unterstützt werden. Das DMS muß dabei eine möglichst umfangreiche Suche in den einzelnen Informationsquellen ermöglichen. Während Datenbanken in der Regel die Suche innerhalb von Texten nicht unterstützen, eignen sich reine Text-Retrieval-Systeme in der Regel nicht dazu, Massendaten, wie z.B. Adreßdateien sinnvoll auszuwerten. Die Suchunterstützung von SGML-Systemen beschränkt sich hauptsächlich auf den logischen Aufbau eines Textes. Der textuelle Inhalt bleibt dem SGML-System verborgen.

Am Markt sind DV-Systeme erhältlich, die Text-Retrieval und Datenbankfunktionen miteinander verbinden. Die Entscheidung für ein solches System ist davon abhängig, welche Schwerpunkte für das Informationsretrieval bestehen. Falls textuelle Informationen kurzlebig sind und nach der Auswertung keine Notwendigkeit zur Aufbewahrung mehr besteht, ist der Einsatz einer Text-Retrieval-Datenbank verzichtbar. Falls allerdings Informationen langlebig sind und deshalb redundanzfrei aufbewahrt werden sollen, ist der Einsatz einer Text-Retrieval-Datenbank unverzichtbar.

Der Einsatz von Datenbanken, die keine textuelle Information verarbeiten können, sollte auf die klassischen Bereiche von Massendaten wie Adreßdateien und Stücklisten beschränkt werden oder bleiben.

Beim Einsatz von Text-Retrieval-Systemen entstehen Effizienzprobleme, die konzeptionell gelöst werden müssen. Die Volltext-Retrieval-Möglichkeit führt in der Regel dazu, daß eher zu viel als zu wenig Information gefunden wird. Eine Volltext-Suche, die im Ergebnis alle Dokumente findet, die auch nur am Rande mit der gewünschten Information zu tun haben, ist für den Nutzer nicht zumutbar. Er müßte zu viele Dokumente sichten, bis die gewünschte Information ausgewählt werden kann. Es muß von daher nach Möglichkeiten gesucht werden, eine Verbesserung der Suchgenauigkeit zu erreichen. Eine zu genau formulierte Suchabfrage kann dagegen bewirken, daß in der Treffermenge Dokumente, die für den Nutzer wesentlich sind, nicht mehr enthalten sind.

Dieses Problem kann durch Einsatz von Concept-based Retrieval oder Thesaurusfunktionalitäten gemildert werden. Concept-based Retrieval verwendet Wortbäume, in denen eine Gewichtung für jedes Wort innerhalb des Wortbaums festgelegt ist. Ein Thesaurus stellt ein sprachsemantisches Gerüst von einzelnen Worten untereinander dar. Dieses Wortgerüst entsteht aus der Definition von Wortrelationen, wie Obergriff-Unterbegriff, Langform-Abkürzung, usw. für jedes Wort innerhalb des Thesaurus. Sowohl Concept-based Retrieval als auch Thesaurusunterstützung reduzieren das dargestellte Problem. Allerdings müssen

Evolutionäres Vorgehen erforderlich

Erschließung, Nutzung und Bewahrung von Information durch DMS

die benötigten Wortbäume bzw. Thesauri definiert und implementiert werden. Hier sind fachliche und sprachwissenschaftliche Kenntnisse von Nöten, damit ein sinnvolles Sprachgerüst geschaffen werden kann. Das Angebot an branchenorientierten Thesauri oder gewichteten Wortbäumen ist dürftig, so daß nicht davon ausgegangen werden darf, für jeden Anwendungsbereich ein passendes Wortgerüst zu finden. Thesaurus und Wortbäume setzen voraus, daß das Vokabular innerhalb des Unternehmens vereinheitlicht wird. Die bei der Festlegung eines unternehmenseinheitlichen Sprachgebrauchs bestehenden Schwierigkeiten und entstehenden Aufwände werden durch die Vorteile eines Thesaurus immer aufgewogen, weil dadurch eine intuitive Suche im Informationsbestand gefördert wird und die Suchergebnisse inhaltlich qualitativ verbessert werden.

Zusätzlich sollte das einzusetzende Text-Retrieval-System die Möglichkeit bieten, zu jeder Information bibliographische Daten, wie Verfasser, Erstellungsdatum, Zusammenfassung und dergleichen, abzulegen. Diese bibliographischen Informationen dienen dann als weiterer Filter und vereinfachen die freie Suche in Texten. Der Umgang mit einem Text-Retrieval-System wird durch diese Art Filter wesentlich vereinfacht.

SGML-Systeme dienen hauptsächlich zur Problemlösung der editorbezogenen Formatgebundenheit von Informationen. SGML ist eine Sprache zur editorunabhängigen Beschreibung von Dokumenten. Neben Formatierungsinformationen, wie Zeichensatz und -größe usw., wird der logische Aufbau eines Dokuments beschrieben. Die Beschreibung des Dokuments wird in der Document-Type-Definition (DTD) abgelegt. Jeder Editor, welcher SGML interpretieren kann, ist in der Lage das Dokument ohne Inhalts- und Formatierungsverlust anzuzeigen, zu überarbeiten und zu drucken. Diese Eigenschaft ist wesentlich bei der Langzeitarchivierung von Dokumenten. Bereits heute zeigt sich, daß -durch die Vielfalt und Weiterentwicklung von Editoren und aufgrund deren proprietärer Formatierung der Dokumente- eine Übernahme von Dokumenten in andere Systeme kaum möglich ist. Dieses Problem kann durch Einfrieren des Hard- und Softwarebestands auf den Zeitpunkt der Einführung des DMS gelöst werden. Die Zukunftssicherheit der Investition in ein DMS ist dann aber nicht gegeben.

Die Verwendung von SGML und einem Text-Retrieval-System, welches in der Lage ist, die Document-Type-Definition zu interpretieren, verbessert die Suchmöglichkeiten nach Informationen, weil dann zusätzlich zu den bibliographischen Daten und dem Text die Struktur des jeweiligen Textes zur Informationsgewinnung verwendet werden kann. SGML bietet ferner Unterstützung bei der Texterstellung. Für jeden Text ist die inhaltliche Struktur abschließend festgelegt. Der Autor ist dann von seiner Funktion als Layouter befreit. Seine Verantwortung erstreckt sich auf die Inhalte und nicht auf die Form

Evolutionäres Vorgehen erforderlich

Erschließung, Nutzung und Bewahrung von Information durch DMS

des Dokuments. Als Nebenprodukt wird dadurch Corporate Design bei der Schriftguterstellung gewährleistet.

Die Festlegung der DTD bedingt hohe Aufwände für die Analyse und Klassifikation des Schriftguts. Für einige Branchen werden bereits heute unternehmensübergreifend Document-Type-Definitions entwickelt, damit ein digitaler Dokumentenaustausch allgemein möglich ist.

Informationsarchivierung

Die Archivierungskomponente ist Kern eines DMS. Entsprechend sorgfältig müssen hier die angebotenen Technologien betrachtet werden.

Die Archivierungskomponente muß kostengünstig sein, eine hohe Ausfallsicherheit bieten, überall verfügbar sein, bedarfsgerechte Aufbewahrungsfristen ermöglichen und vor ungewolltem Informationsverlust schützen.

Die derzeit gängige und kostengünstigste Archivierungstechnologie ist das Ablegen von Information auf optischen Speichermedien (CD-ROM, WORM-Platten). Sowohl die Kosten für die Speichermedien, als auch die Aufwände für die physikalische Pflege des Archivs sind von anderen Systemen nicht zu unterbieten. Für die Lesbarkeit optischer Platten ohne Datenverlust geben die Hersteller Garantien von mehr als 20 Jahren. Das bedeutet, daß die Pflege der archivierten Daten in größeren Zeitabständen erfolgen kann, als dies z.B. für Magnetbänder als Sicherungsmedium gilt. Noch heute sollte jedes Magnetband im 3-Monats-Zyklus erneut kopiert werden, damit ein Datenverlust wegen der physikalischen Eigenschaften des Datenträgers ausgeschlossen werden kann.

Der Zugriff auf die optische Platte ist aber in der Regel nur über proprietäre Software möglich. Dies führt zu weitreichenden Einschränkungen, wenn die Archivierung durch eine Text-Retrieval-Datenbank unterstützt werden soll.

Bei der Verwendung einer Text-Retrieval-Datenbank hat die Datenbank die komplette Kontrolle über die abgelegte Information. Die Vorteile einer datenbankbasierten Archivierungskomponente liegen klar auf der Hand: Integrität der Information, Zugriff auf die Information, Verlustfreiheit der Information und Transaktionskontrolle bei der Änderung von Information sind durch die Eigenschaften des Datenbanksystems gegeben. Die Integration mit Text-Retrieval-Funktionen erfolgt über die Datenbanksprache. Die Pflege von Indizes erfolgt für jede Neuaufnahme, Änderung oder Löschung durch die Datenbank. Die Backup- und Recovery-Mechanismen der Datenbank schließen einen Informationsverlust weitgehend aus. Der Zugriff auf Information erfolgt immer unter der Kontrolle des Datenbank-Kernels.

Der Zugriff auf Information durch Verwendung eines Datenbank-Kernels bedingt, daß der Kernel alle Systemressourcen kennen und nutzen können muß. Da die optische Platte als Systemressource dem Kernel in der Regel unbekannt bleibt,

Evolutionäres Vorgehen erforderlich

Erschließung, Nutzung und Bewahrung von Information durch DMS

kann dieses kostengünstige Medium für eine datenbankorientierte Lösung als Archivierungskomponente nicht verwendet werden.

Bei der Verwendung von Adreßreferenzen zwischen Datenbank und optischem Archiv geht im schlimmsten Falle die Integrität der Information verloren, weil Änderungen an der Information am Datenbanksystem vorbei möglich sind. In diesem Falle ist das gesamte DMS nicht mehr tauglich, verlässliche, aktuelle und vollständige Information bereitzustellen.

Fazit

Ein DMS stellt immer eine unternehmensspezifische DV-Lösung dar. Das DMS deckt die Anforderungen ab, die sich aus einer sorgfältigen Bedarfsanalyse ergeben. Nur wenn ein effektives und effizientes DMS eingesetzt wird, können die positiven Effekte aus einem Business Process Re-engineering vollständig erreicht werden, da Business Process Re-engineering davon ausgeht, daß Information als Ressource uneingeschränkt vorliegt.

Ein Problemfeld, welches konzeptionell gelöst werden kann, sind rechtliche Randbedingungen für den Umgang mit digitalisierter Information. Dies betrifft neben gesetzlichen Aufbewahrungsfristen von Dokumenten in der Hauptsache die Themenfelder von Beweiskraft digitalisierter Information im Rechtsverkehr und die Wirksamkeit von Willenserklärungen bei Electronic Data Interchange (EDI). Der Gesetzgeber hat bisher die Papierform als Informationsspeicher fest verankert. Im Gerichtsverfahren kann eine digitalisierte Information nicht als Urkunde Beweiskraft erlangen. Die Zulassung digitalisierter Information als Beweis des Augenscheins ist vom Ermessen des jeweiligen Richters abhängig. Dieser Aspekt ist bei der Festlegung des Umfangs der Digitalisierung von Information nicht unwesentlich. Für EDI-Anwendungen stellt die Wirksamkeit von Willenserklärungen einen eigenen Problembereich dar, der durch individuelle Absprachen mittels eines EDI-Vertrages zwischen den Austauschpartnern gelöst werden muß. Da bei digitalem Datenaustausch der Zugang einer Willenserklärung quasi sofort nach Datenfreigabe gegeben ist, kann diese wegen eines Irrtums des Erklärenden nur bedingt angefochten werden, da Voraussetzung ist, daß die angefochtene Willenserklärung den Adressaten noch nicht erreicht hat. Die Frage ist also, wer verantwortet den Fehler, wenn der Zulieferer seinem Auftraggeber 50.000 Schrauben liefert, obwohl der Auftraggeber aus seiner Sicht nur 5.000 Stück angefordert hat? Ist die Abweichung der Bestellmenge auf eine fehlerhafte Datenübertragung zurückzuführen und in wessen Verantwortung fällt die Wahrung der inhaltlichen Konsistenz beim Datenaustausch? Kann eine Manipulation der Bestelldaten beim Empfänger, sprich Zulieferer, ausgeschlossen werden?

Wegen der Komplexität der dargestellten Themenkreise Informationsgewinnung, -verteilung, -auswertung und -archivierung kann die Einführung eines DMS nur in einer evolutionären Vorgehensweise erfolgen. Zweck dabei muß sein, ein

Evolutionäres Vorgehen erforderlich

Erschließung, Nutzung und Bewahrung von Information durch DMS

Gesamtkonzept zu entwickeln, welches mehrere definierte Entwicklungsschritte erlaubt. Andernfalls kann nicht ausgeschlossen werden, daß im Rahmen der Einführung eines digitalen DMS Teillösungen entstehen, die einer effektiven und effizienten Gesamtlösung entgegenstehen. Nur wenn Document-Management als unternehmenskritische Aufgabe verstanden wird und das Fachwissen über Rechtsangelegenheiten, Organisation, Fachaufgaben und Datenverarbeitung in einem gemischten Projektteam zusammengezogen wird, besteht die Aussicht, die Herausforderungen, die mit Document-Management verbunden sind, vollständig zu begreifen und zukunftsicher zu lösen.