



Informations- und Wissensmanagement durch strukturierte Dokumente

Jiri Panyr, Werner Zucker

Siemens AG, Abt. ZT IK 1
Otto-Hahn-Ring 6, 81739 München
e-mail: jiri.panyr@mchp.siemens.de / werner.zucker@mchp.siemens.de

Inhalt

1. Einleitung
2. Probleme und Ursachen bei der bisherigen Vorgehensweise
 - 2.1 Seiten- und Dokument-Metapher
 - 2.2 Dokument-Management-Systeme
 - 2.3 Information Retrieval / Informationsfilterung
 - 2.4 Executive Information System
 - 2.5 World Wide Web / Elektronisches Publizieren
 - 2.6 Konsistenzproblematik / Dokumentenaustausch (Protokolle)
3. Randbedingungen
 - 3.1 Kundenforderungen
 - 3.2 Qualitätsmanagement durch ISO 9000
4. Optimierungsbedarf bei der Dokumentation
 - 4.1 Inhaltsorientiertes Strukturieren von Dokumenten und Information
 - 4.2 Konsistenzproblematik
 - 4.3 Auffinden von Information
 - 4.4 Durchgängigkeit im Prozeß durch Dokumentenaustausch
 - 4.5 Automatisiertes Erzeugung von Dokumentation
 - 4.6 Single Source Publishing /Single Source Information
5. Weltweite Aktivitäten in Auswahl
6. Resume

Zusammenfassung

Technische Dokumentation ist ein wichtiger Bestandteil komplexer technischer Systeme. Sie ist als Informationsressource sowohl für den Kunden als auch für den Hersteller von entscheidender Bedeutung. Wesentlich ist dabei ihre inhaltliche Konsistenz, Austauschbarkeit und Wiederverwendung. Je umfangreicher und komplexer die technische Dokumentation ist, um so wichtiger ist ihre inhaltliche Strukturierung. Während sich die Strukturierungstechniken (und die Trennung von



Dieses Dokument wird unter folgender [creative commons](http://creativecommons.org/licenses/by-nc-nd/2.0/de/) Lizenz veröffentlicht:
<http://creativecommons.org/licenses/by-nc-nd/2.0/de/>

Inhalt und Layout) im Dokumentationsbereich zunehmend durchsetzen, fehlen bei der Anwendung strukturierter Dokumente meist Werkzeuge und Verfahren, die diese Strukturen auch nutzen. In der Arbeit wird auf diese Problematik hingewiesen. Zunächst werden die Probleme mit dem bisherigen präsentations- und seitenorientierten Paradigma aufgelistet, dann auf die Verbesserungsmöglichkeit durch die Nutzung der Strukturierung hingewiesen. Schließlich werden einige weltweite Aktivitäten im diesem Bereich genannt.

Abstract:

Technical documentation is an important part of complex technical systems. As an information resource, it is of critical importance for the customer as well as for the producer. Consistency in content, interchangeability and reuse is essential. The more extensive and complex the technical documentation the more important is the content structuring. While techniques of structuring (and the separation of contents and layout) in the documentation area are more and more accepted, tools and methods in application of structured documents are missing that utilize these structures. This paper focuses on these problems. First the problems with the current presentation and page oriented paradigm are listed, then the possibility of improvement through using the structuring is explained. Finally a survey of world wide activities in this area is presented.

1. Einleitung

In der Informationsgesellschaft von heute kommt der Bereitstellung, Verarbeitung und dem Austausch von Informationen eine überragende Bedeutung zu. Dennoch wird der Effizienz und der Effektivität beim Umgang mit ihr nur geringe Bedeutung beigemessen, besonders dann, wenn Information in Dokumentform vorliegt. Dann wird zum Teil erheblicher Aufwand für die Darstellung von Dokumenten (d.h. für Layout und Typographie), nicht aber für ihre inhaltlich logische Strukturierung geleistet. Das schlägt sich nieder in den zum Einsatz kommenden DTP-Werkzeugen und -Formaten und findet im Boom der Web-basierten Informationsdistribution seine erneute Bestätigung.

Auch dem Aspekt der wiederverwendbaren und weiterverarbeitbaren Information wird, von Ausnahmen abgesehen, nach wie vor nur eine geringe Aufmerksamkeit gewidmet. Besonders die Dokumentation von technischen Systemen wird im Wertschöpfungsprozeß eher als notwendiges Übel betrachtet. Anstelle von informationsorientierter Verarbeitung tritt daher die präsentationsorientierte Verarbeitung. Anstelle semantischer Informationseinheiten treten visuelle Einheiten. Die Strukturierung erfolgt mittels Layout auf der grafischen Ebene, nicht aber auf der logisch semantischen Ebene.

Technische Dokumentation ist ein wichtiger Bestandteil komplexer technischer Systeme. Sie ist als Informationsressource sowohl für den Kunden als auch für den Hersteller von entscheidender Bedeutung. Wesentlich ist dabei ihre inhaltliche Konsistenz, Austauschbarkeit und Wiederverwendung. Je umfangreicher und

komplexer die technische Dokumentation ist, um so wichtiger ist ihre inhaltliche Strukturierung. Während sich die Erkenntnis der Strukturierungsnotwendigkeit und die Trennung von Inhalt und Layout zumindest im Dokumentationsbereich zunehmend durchsetzt, fehlen bei der Handhabung strukturierter Dokumente meist Werkzeuge und Verfahren, die diese Strukturen auch nutzen. Mit anderen Worten: Die inhaltlichen Strukturen von Dokumenten sind zwar vorhanden, viele Anwendungssysteme (z.B. Information Retrieval, Suchverfahren in WWW, Desktop Publishing Systeme, Executive Information Systeme) setzen in der Regel weiter auf das alte gesamtdokumentenausgerichtete und präsentations- bzw. seitenorientierte Paradigma auf.

2. Probleme und ihre Ursachen bei der bisherigen Vorgehensweise

Zum o.g. älteren Paradigma werden im folgenden einige ausgewählte Beispiele aufgelistet.

2.1 Seiten- und Dokument-Metapher

Die bisherigen Textverarbeitungssysteme (wie z.B. Word etc.) und auch Desktop Publishing Systeme (DTP-Systeme, wie z.B. Interleaf) arbeiten primär präsentationsorientiert, d.h. die dokumentarische Bezugseinheiten sind ganze Dokumente, während die Präsentationselemente die Dokumentseiten sind. Es wird in diesem Kontext auch über Dokumenten- bzw. Seiten-Metapher gesprochen. Im Vordergrund steht bei allen Ansätzen nicht der Inhalt, sondern primär das Layout. Dies führt u.a. zu folgenden Konsequenzen:

- **Seitenorientierte Navigation / Logische Einheiten nicht als Sprungziel**
Bei der Navigation in Dokumenten wird unabhängig davon, ob Verweise präzise auf eine bestimmte Textstelle zeigen, von einer Reihe von Viewern (wie die z.B. von Adobe, Frame, Interleaf) nach dem "Anklicken" lediglich der Anfang der Seite, die das Querverweisziel enthält, angezeigt. Eine präzisere Positionierung innerhalb einer Seite ist in einem der o.g. Viewer nicht möglich. Heutige Beschreibungsfomalismen für Dokumente erlauben es nicht, logische Einheiten als Sprungziel zu verwenden. Dies führt einerseits zu den o.g. Problemen bei der Navigation, andererseits führt dies zu semantischen Ambiguitäten, wenn beispielsweise die unmittelbare Umgebung eines Sprungzieles bei der Bearbeitung gelöscht wird. Es ist dann nicht klar, ob der Link noch gültig ist oder nicht.
- **Unzureichende Methodik und Systematik**
Obwohl auch DTP-Systeme gewisse Strukturierungshilfe (wie z.B. Komponenten) bieten, sind diese so gut wie nicht genutzt, da eine logische Systematik (Vorschrift) und Methodik bzgl. ihrer Nutzung fehlt. Komponentennamen werden willkürlich vergeben. Komponenten haben auch in gleich gearteten Dokumenten oftmals eine unterschiedliche Benennung und Gestaltung (Eigenschaften), obwohl sie die gleiche Position und Funktion im Text innehaben.

- Informationsverlust bei Konvertierung
Die Konvertierung der von DTP-Systemen verwendeten Formate (wie ps, pdf, ...) führt zu einem erheblichen Informations- und Layoutverlust. Die Erhaltungsfunktionalität basiert in der Regel noch immer auf Formaten (wie z.B. DIF, DCA, RTF, MIF), die inhaltliche Erweiterungen bei der Konvertierung nicht berücksichtigen.

Die o.g. Darstellungsorientierung unter gleichzeitiger Vernachlässigung der inhaltlichen Aspekte beim Textaufbau führt zu Reihe von Problemen bei der inhaltlichen Erschließung von Dokumenten und Wiedergewinnung von Informationen, sowie auch bei ihrer Präsentation in einer Hypertext-Umgebung. Die bis vor kurzem in diesem Bereich dominierende textorientierte Künstliche Intelligenz (KI) hat diese konservative Metapher mehr oder minder unterstützt und keine neue Paradigma gefördert.

2.2 Dokument-Management-Systeme

Konventionelle Dokument-Management-Systeme verwalten Versionen und Varianten, kontrollieren die Zugangsberechtigung und die Freigabe, verfügen über Änderungs- und Revisionsmanagement und meist geringfügige Workflow-Funktionalität. Kennzeichnend für alle Systeme ist das speicher-orientierte Management von Dokumenten. Es werden Files verwaltet jedoch nicht Information. Die kleinste managebare Einheit ist das File und nicht die Information. Der Zugriff auf Information in Dokumenten erfolgt nicht über die logischen Struktur der Dokumente sondern über die Attribute von Files. Information Retrieval findet dokumentenbezogen statt und kann nicht auf bestimmte Teile fokussiert werden. Jegliche Funktionalität von Dokument-Management-Systemen ist File-orientiert. Dokument-Management-Systemen sind somit eher komfortable File-Management Systeme und weniger Informations-Datenbasen.

Desweiteren nutzen konventionelle Dokument-Management-Systeme die Erkenntnisse aus der DB-Technologie so gut wie nicht (Beispiel: Datenstrukturierung, Redundanzfreiheit, zentraler Zugriff, benutzerspezifische Views, Normalisierung), d.h. sie sind lediglich unvollkommene Datenhaltungsinstrumente. Es bestehen Schwierigkeiten z.B. bei Wiederverwendung und Ähnlichkeitsnutzung dieser Dokumente. So wird z.B. auch Ähnlichkeitsplanung bei CAD/CAP-Dokumenten nicht unterstützt, da die im DMS abgelegte Dokumente de facto nicht präzise recherchierbar sind. Die im DMS abgelegten Dokumenten können inhaltlich nur rudimentär beschreiben, d.h. erschlossen. Inhaltliche Extraktion ist bisher nur durch unvollkommene syntaktisch-semantische Werkzeuge möglich. Obwohl immer eine Aufgabe der text-orientierten KI-Forschung, sind die Lösungen bisher unbefriedigend.

Bei allen Anwendungen (z.B. Entwicklungsdokumentation) ist eine elektronische Übernahme von Dokumententeilen wünschenswert, findet jedoch hauptsächlich wegen technischer und organisatorischer Inkompatibilität nicht statt.

2.3 Information Retrieval / Informationsfilterung

Die präsentationsorientierte Dokumentgestaltung verursacht bei der inhaltlichen Erschließung und bei der gezielten Wiedergewinnung (Retrieval) von Informationen eine Reihe von Problemen, insbesondere bei längeren, inhaltlich nicht weiter strukturierten Dokumenten. Nicht ohne Grund sind bis jetzt die besten Retrievalergebnisse mit Dokumenten in Form eines Abstracts erzielt. Eine automatische oder zumindest semi-automatische Erstellung eines signifikanten Textextrakts oder eines redundanzfreien Textabstrakts wurde bisher vor allem durch ungenaue auf den ganzen Text anwendbare syntaktisch-semantische Werkzeuge versucht. Im weiteren werden einige Probleme kurz skizziert:

- Große Probleme gibt es mit der automatischen Indexierung bei Dokumenten, die als nicht weiter differenzierbare Textmasse präsentiert sind. Falls Strukturierung im Dokument vorhanden ist, dann lediglich in Form von Layout. Bei größeren Dokumenten ist daher eine sinnvolle Selektion der bedeutungstragenden Wörter kaum möglich. Semantische Textanalysen allein sind nicht ausreichend. Daher ist vonnöten, die zu erschließende Texte schon bei ihrer Erstellung einer inhaltlichen Gliederung zu unterziehen, mit konsequenter Trennung von Inhalt und Layout.
- Die unvollkommene inhaltliche Erschließung von Dokumenten verursacht auch große Probleme bei ihrer Wiedergewinnung (Retrieval, Recherche), wie z.B.:
 - Suche ist nur im ganzen Dokumenten möglich bzw. in ihren strukturierten Teilen (Titel, Autor, Verlag etc.).
 - Bei größeren Texten entstehen viele überflüssige Treffer, weil keine weitere inhaltliche Textdifferenzierung (s.o. Indexierung) möglich ist.
 - Relevanzfeedback-Verfahren (interaktive Retrievalstrategien) operieren lediglich auf den Beschreibungen von ganzen Dokumenten (bzw. ihrer Abstracts). Sie ermöglichen zwar eine Recall-Erhöhung, liefern jedoch gleichzeitig auch viele Ballastdokumente.
- Ähnliche Probleme entstehen auch bei der Informationsfilterung, die ebenfalls nicht komponenten-orientiert ist, d.h. es werden zu viele Ballastdokumente ausgewählt. Die nutzergerechte Verteilung der gefilterten Informationseinheiten (SDI) wird dadurch erschwert.

Die Entwicklung von neuen Methoden im Bereich der Information-Retrieval-Systeme hat infolge der nichtrealistischen Versprechungen der KI über Jahre stagniert. Die Konzepte derzeit eingesetzter Systeme (auch der neueren) basieren in der Regel immer noch auf den Ansätzen vom Ende der 60iger bis Anfang der 70iger Jahre. Hier helfen nicht mehr einige Änderungen im Detail, sondern eine Neukonzeption der Dokumenterstellung und -erschließung. Projekte, die das mit einem realistischen Zeitrahmen und vernünftigen Ressourcen ermöglichen, sind bisher nicht in Sicht.

2.4 Executive Information System

Seit Ende der 80er Jahre setzt sich zunehmend eine neue Variante der Management-Unterstützungssysteme durch, die sich unter dem Begriff *Executive Information Systems* (EIS) zusammenfassen läßt. Diese DV-Systeme - in BR Deutschland auch als *Führungs-Informationssysteme* (FIS) oder *Chef-Informationssysteme* (CIS) bezeichnet - sind in der Nähe von Berichtssystemen anzusiedeln, wobei sie jedoch aufgrund ihrer großen Flexibilität und Funktionalität (individuelle Drill-Down-Technik und Ausnahme-Berichtsgenerierung, Trendanalysen, Einbindung von e-mail etc.) über die Grenzen traditioneller (DV-gestützten) Berichtssysteme hinausgehen und vom Top-Management selbst benutzt werden sollen. In Abgrenzung zu älteren Entscheidungsunterstützenden Systemen werden EIS nicht für die Lösung schlecht strukturierter Entscheidungsproblematiken herangezogen, sondern sind vielmehr personenbezogen auf bestimmte Entscheidungsträger und deren Informationsbedarf ausgerichtet. Sie umfassen jene aufbau- und ablaufsorganisatorischen Regelungen, Führungsgrößen und -methoden, die zur Initiierung, Planung, Analyse und Kontrolle von Geschäftsprozessen im Unternehmen benötigt werden. Damit EIS bedarfsgerechte Unterstützung (des Top-Management) leisten können, muß ein hoher Integrationsgrad mit den datenliefernden operativen Informations- und Kommunikationssystemen erreicht werden. Für die Realisierung effektiver EIS muß daher, neben der betriebswirtschaftlichen Integration in das Führungssystem, die technische Integration in die informations- und kommunikationstechnische Infrastruktur geleistet werden. Diese Integration muß das firmeneigene Informationsmanagement gewährleisten.

Bei der Realisierung von EIS in verschiedenen Unternehmen (in der Regel gab es meist mehrere Anläufe) haben sich die fehlende konzeptionelle Daten- und Dokumentmodellierung und die daraus folgende Probleme mit der qualitativen und physikalischen Datenbeschaffung (d.h. Schwierigkeiten bei der Identifikation originärer Informationsquellen) als Haupthindernis gezeigt. So sind in einigen Unternehmen die gewünschten Daten aufgrund nichtoptimaler Datenstrukturen (wirtschaftlich) nicht ökonomisch extrahierbar, bzw. es existiert kein konsistenter Datenpool für diese Systeme, so daß EIS auf (teilweise willkürlich gewählten) anwendungsbezogenen Extrakten der individuellen Datenverarbeitung basieren muß. Dadurch sind auch einige wichtige Funktionen nicht oder nur schwer realisierbar, z.B.:

- (Exception-)Reporting, da Dokumente aus verschiedenen Quellen unterschiedlich strukturiert sind;
- Individuelle Drill-Down-Techniken, d.h. zielorientiertes Abrufen von Detailinformationen nach der vorher definierten logischen Abhängigkeit.

Die geschilderten Probleme hängen mit der häufig nicht vorhandenen Einbettung von EIS in übergreifende Konzepte der Bürokommunikation und mit dem Fehlen eines gebietsüberschreitenden Informationsmanagement zusammen.

2.5 World Wide Web / Elektronisches Publizieren

Das World-Wide Web (WWW) besteht aus vernetzten Seiten ohne erkennbare Struktur. Bei der Gliederung der Seiten ist jeder Autor völlig frei, weil sich mit der im WWW verwendeten Sprache HTML (bisher in der Regel HTML 3.0 als Standard) nur die Gestaltung des Inhalts steuern läßt, und nicht die inhaltliche Struktur eines Dokuments. Dies hat sich zwar inzwischen bei HTML 4.0 als eine Reaktion auf das Konzept von XML etwas gebessert, es ist dennoch z.B. nicht möglich, im Rahmen einer Qualitätssicherung zu überprüfen, ob eine Seite ganz bestimmte, dort notwendige Angaben enthält. Werden sich z. B. im ZT-Intranet alle laufenden Projekte auf einer HTML-Seite aufgelistet, kann nicht überprüft werden, ob dort jeweils ein Ansprechpartner mitsamt Telefonnummer und e-mail-Adresse genannt wird. Die Dublin-Core-Initiative versucht zwar dies in Griff zu bekommen, sie scheint jedoch bisher lediglich bei den Bibliothekaren auf gewisse Resonanz zu stoßen (vgl. *Dublin Core Metadata 1997*).

Wegen fehlender inhaltlicher Strukturen in den WWW-Dokumenten gibt es nur zwei Möglichkeiten, auf Informationen zugreifen zu können: durch Navigation (*Browsen*) und durch den Einsatz von Suchmaschinen. Bei der Navigation ist der Leser auf die Qualität der vom Autor vorgegebenen Hypertext-Struktur angewiesen, die nicht immer mit der gewünschten Zugriffsstruktur übereinstimmt. Es gibt daher Suchmaschinen im WWW, mit denen über Stichwörter oder über Kombinationen von Stichwörtern nach Dokumenten mit der gesuchten Information recherchiert werden kann. Suchmaschinen basieren auf dem Prinzip der gewichteten Volltextsuche in möglichst vielen Seiten im WWW. Es ist nicht möglich, die Suche auf bestimmte inhaltliche Strukturen einzuschränken.

Ein großer Vorteil des WWW liegt im elektronischen Publizieren von Dokumenten. Durch Plug-Ins können moderne WWW-Browser inzwischen auch Dokumente mit proprietären Formaten (PDF, WinWord, usw.) darstellen. Damit lassen sich diese zwar anzeigen, lesen und drucken; eine Weiterverarbeitung, wie etwa bei einer automatischen Indexierung durch Suchmaschinen, setzt proprietäre Software zur Konversion des verwendeten Formats voraus. Zusätzlich wird durch den Einsatz von proprietären Browsern, d.h. solche, die i.a. nur bestimmte HTML-Sprachvarianten unterstützen, der weltweite Zugriff auf die gesuchte Information zunehmend eingeschränkt.

Zusammengefaßt ist das WWW ein hervorragendes Medium, Informationen zu publizieren; zur maschinellen Weiterverarbeitung oder gar Verwaltung der publizierten *Informationen* dagegen ist das WWW ungeeignet, weil es keine inhaltlichen Strukturen kennt.

2.6 Konsistenzproblematik / Dokumentenaustausch (Protokolle)

Die inhaltliche Konsistenz von Dokumentation mit inhaltlichen Bezügen kann verletzt werden, wenn Änderungen nicht adäquat propagiert werden. Es gibt hier drei Klassen von inhaltlichen Bezügen in elektronischer Dokumentation:

Wiederverwendung, Querverweise und sonstige Bezüge zwischen *Informationseinheiten*.

Bei der Wiederverwendung existieren die gleichen Informationseinheiten an verschiedenen Stellen im Dokument, evtl. angepaßt an den jeweiligen Kontext. Wird nun eine dieser Stellen verändert, müssen die Änderungen an allen anderen nachgezogen werden, um die inhaltliche Konsistenz beizubehalten.

Bei Änderungen am Absprungspunkt oder am Einsprungsziel eines Querverweises sind verschiedene Folgeaktivitäten denkbar, um wieder eine inhaltlich konsistente Dokumentation zu erhalten. Entweder der Verweis ist inhaltlich obsolet oder Dokument strukturell unvollständig, dann muß er gelöscht werden. Oder es sind keine Folgeänderungen nötig und der Verweis bleibt unverändert. Zuletzt kann es sein, daß durch die anfängliche Änderungen ein umfangreiches Umstrukturieren oder Überarbeiten von Teilen der Dokumentation notwendig wird.

Weitere inhaltliche Bezüge liegen schließlich z.B. zwischen Übersetzungen vor: Nach Änderungen am Original müssen alle bereits existierenden Übersetzungen nachgezogen werden.

Mit der Konsistenzproblematik hängt auch die Frage des Dokumentenaustausches und der Protokolle zusammen. Unter einem Protokoll wird im weiteren die Gesamtheit aller für den Daten- bzw. Dokumentenaustausch über eine Schnittstelle vereinbarten Regeln verstanden. Darunter fällt auch die in der Datenübertragung als Protokoll bezeichnete Verfahrensvorschrift. Mit dieser Vorschrift werden dann sowohl die Formate der "Nachrichten" als auch der Datenfluß bei der Datenübertragung festgelegt (vgl. auch ISO-Schichtenmodell).

Der gängige, auf dem sog. Document Interchange Format (DIF; 7-bit-Codierung) oder auf der Document Content Architecture (DCA; 8-bit-Codierung) bzw. auf den Formaten RTF und MIF basierende Dokumentenaustausch führt durch die Vermischung von logischer und physischer (Layout) Struktur zu einer erheblichen Layoutveränderung bzw. zu einem Layoutverlust und somit auch zu einer Minderung des Informationswertes bei ausgetauschten Informationseinheiten.

3. Randbedingungen

3.1 Kundenforderungen

Technische Dokumentation dient zur Kommunikation des technischen Wissens zwischen Hersteller und Kunden. Gleichzeitig sind die in den Dokumenten enthaltenen Informationen für den Betreiber von langlebigen technischen Systemen wertvoll und bedeutsam, weil zur Wartung und Reparatur auf die benötigten Unterlagen auch noch nach Jahren oder Jahrzehnten zugegriffen werden muß. Angesichts zunehmend kürzerer Innovationszyklen in der Softwareentwicklung müssen daher bei der Archivierung von elektronischer Dokumentation auch die zur Ansicht und Weiterverarbeitung benötigten Applikationen aufgehoben werden.

Dies wurde bereits in den 80er Jahren vom amerikanischen Verteidigungsministerium (DoD) erkannt. Um die allein durch den Besitz eines komplexen technischen Systems anfallenden Kosten ('cost of ownership') zu senken, forderte das DoD im Rahmen der CALS-Initiative u.a. von seinen Zulieferern, die gesamte technische Dokumentation in einem herstellerunabhängigen Format ('CALSDTD') auszuliefern, mit dem auch gleichzeitig die Inhalte der Dokumentation explizit gekennzeichnet wurden. Grundlage dieses Formats ist der ISO-Standard SGML, der seitdem immer häufiger zur Definition von branchenspezifischen Austauschformaten herangezogen wird.

Im folgenden werden vier wichtige Initiativen aus der unmittelbaren Vergangenheit kurz beschrieben:

- **TBL-P Projekt**
Die Deutsche Telekom fordert im Rahmen des TBL-P Projekts, an dem u.a. Alcatel SEL und Siemens ÖN als Zulieferer beteiligt sind, die technische Dokumentation in SGML bezüglich einer vorgegebenen, inhaltlichen Struktur (*Telekom DTD*), in der Informationseinheiten wie z.B. *Kommando*, *Warnungshinweis* usw. gekennzeichnet sind.
- **European Forum for Telecom Industry Information Interchange (EFTI3)**
Aus dem ESPRIT-Projekt *TecPad* (EP 8793) hervorgegangen ist die Initiative *EFTI3*, an der neben den Herstellern Siemens ÖN, Italtel, Ericsson, Alcatel und Nokia auch die Betreibergesellschaften Teledanmark und British Telecom beteiligt sind. Dieses Forum hat inzwischen eine Austauschstruktur verabschiedet, die es einem Kunden ermöglicht, die Dokumentation von verschiedenen Herstellern mit einer stets gleichen Struktur zu beziehen um sie danach intern weiterzuverarbeiten.
- **Telecommunications Industry Forum (TCIF)**
In Amerika gibt es mit TCIF eine ähnliche Initiative wie *EFTI3*, an der u.a. Bellcore, CISCO Systems, Lucent Techn., Newbridge, Northern Telecom und Siemens Stromberg-Carlson auf der Herstellerseite beteiligt sind sowie u.a. AT&T, Bell Atlantic, MCI Communications, Pacific Bell und Sprint Corp. auf der Betreiberseite.

Auch in der Automobilbranche wird zunehmend SGML zum Austausch von inhaltlich strukturierter technischer Dokumentation eingesetzt. In dem sehr ambitionierten Projekt *MSR* (Messen, Steuern, Regeln) z.B. sind Mercedes Benz und Bosch federführend beteiligt, zudem Volkswagen, BMW und Porsche.

- **SAEJ2008**
In Amerika gibt es den Industriestandard SAE J2008 zum Austausch von strukturierter Servicedokumentation für alle umweltrelevanten Bestandteile von Automobilen. Vom California Air Regulations Board (CARB) wird inzwischen gefordert, daß der Austausch von Servicedokumentation konform zur SAE J2008 stattfinden muß, um in Kalifornien ab dem Jahr 2002 Fahrzeuge verkaufen zu können.

3.2 Qualitätsmanagement durch ISO 9000

Die Dokumentation von Verfahrensanweisungen, Ergebnissen und Produkten bilden einen zentralen Aspekt beim Qualitätsmanagement. Darüber hinaus spielen Dokumente eine tragende Rolle im Rahmen der Produkthaftung.

Dies alles findet in den europäischen Normen EN ISO-900x seinen Niederschlag, schwerpunktmäßig im Kapitel 4 *Forderung an die Qualitätssicherung / Qualitätsdarlegung* und speziell in den Unterkapiteln 4.4 *Designlenkung*), 4.5 (*Lenkung der Dokumente*) und 4.16 (*Lenkung von Qualitätsaufzeichnungen*).

Auch wenn die EN ISO 900x keine direkten Vorgaben für Art, Ausführung und Umfang von Dokumenten gibt, so stellt sie aus qualitativer Sicht hohe Ansprüche an Dokumentation als solche, vor allem an das Dokumentenmanagement. Es wird gefordert, daß Dokumente mit korrektem Ausgabestand und richtigen Inhalt zum richtigen Zeitpunkt am richtigen Ort sind. Ungültige und/oder überholte Dokumente sind sofort an allen Stellen zu entfernen. Dokumente, die aus rechtlicher Sicht oder zur Erhaltung des Wissensstandes aufbewahrt werden, sind entsprechend zu kennzeichnen. Änderungen von Dokumenten und Daten sind, wo durchführbar, erkennbar in der Dokumentation selbst oder in geeigneten Anlagen ausgewiesen werden.

Für Designunterlagen gilt insbesondere, daß das Designergebnis in einer Form dokumentiert werden muß, das in Bezug auf die Designvorgaben verifiziert und validiert werden kann. Hiermit wird grundsätzlich eine Beziehungstransparenz gefordert, welche das State-of-the-art Dokumentenmanagement a priori nicht erfüllen kann.

4. Optimierungsbedarf bei der Dokumentation

4.1 Inhaltsorientiertes Strukturieren von Dokumenten und Information

Dokumente, präziser dokumentgebundene Informationen, spielen in allen Wertschöpfungsketten des Hauses eine zentrale Rolle. Sie dienen der Sicherung und Weitergabe von Ergebnissen und Know-how für Produkte und Prozesse. Sie sind Grundlage für Vertrieb und Marketing, Entwicklung und Betrieb, Service und Wartung.

Auf Grund dieser enormen Bedeutung von Dokumentation im Wertschöpfungsprozeß, den stetig wachsenden Kundenanforderungen, den Erfordernissen in Bezug auf die Informationsdistribution und das Qualitätsmanagement ist es notwendig, auch bei Dokumenten die Informationssicht in den Vordergrund zu stellen. Basis hierfür ist eine logisch inhaltliche Strukturierung von Information und Dokumentation, die erforderliche Transparenz und Präzision in allen Phasen beim Umgang mit Information und Dokumentation sicherstellt. Bei der dazu notwendigen technologischen Grundlagen für modulares Design und Reuse von Dokumentation und Information stehen

- die Trennung von logischem Inhalt und grafischer Darstellung
- die Modellierung komplexer Dokument- und Informationsstrukturen
- und Konzepte und Methoden der Wiederverwendung im Mittelpunkt.

Eine Generalisierung über jegliche Art von Dokumentation und Information kann aus dem bisher gesagten nicht abgeleitet werden. Die folgende Auflistung zeigt daher exemplarisch, in welchen Bereichen potentiell Handlungsbedarf besteht.

Technische Unterlagen

- Produktunterlagen
- Angebotsdokumentation
- Lasten-, Pflichtenhefte (bzw. A10, A20, A30)
- Entwicklungsdokumentation, Software- und Hardwareunterlagen
- Bedienhandbücher, Anleitungen
- Service- und Wartungsunterlagen

Prozeß- und projektrelevante Unterlagen

- Projektsteuerung
- Projektunterlagen
- QM-Unterlagen
- E-mail
- Standardisierter elektronischer Briefverkehr

Unternehmensspezifische Informationen

- Informationsbasen zur Know-how-Sicherung
- Betriebswirtschaftliche Informationen und Unterlagen
- Betriebsführungsrelevante Informationen und Unterlagen

4.2 Konsistenzproblematik

Inkonsistenzen entstehen, wenn nach bestimmten Veränderungen am Dokumenteninhalt die dadurch notwendig gewordenen Folgeänderungen unterbleiben (s.o. Abschnitt 2.6). Es gibt bislang kein Werkzeug, um die Autoren beim Erstellen einer inhaltlich konsistenten Dokumentation zu unterstützen, weil mit heutigen Textverarbeitungssystemen i.d.R. nicht der Inhalt von Dokumenten manipuliert werden kann, sondern nur die Darstellung dieser Inhalte. Lassen sich die Inhalte von Dokumenten maschinell verarbeiten, wurden also die *Information* in den Dokumenten kodiert, können die Autoren von Dokumentation jedoch hinsichtlich der oben (im Abschnitt 2.6) beschriebenen Konsistenzproblematik durch semi-automatische Verfahren unterstützt werden.

4.3 Auffinden von Information

Dokumentenarchitekturen, welche die logische und physische (Layout) Struktur von Dokumenten unabhängig von spezieller Hard- und Software (meist in hierarchischer Form) beschreiben, ermöglichen eine effektive und effiziente inhaltliche

Erschließung von Dokumenten und somit auch ein verlust- und ballastarmes Retrieval von Information. Ein Vertreter solcher Architekturen ist ODA. Mit SGML wird mit einem völlig anderen Ansatz das gleiche Ziel, die Trennung von Logik und Layout, verfolgt.

In der Vergangenheit wurde dies schon bei einigen Ansätzen probiert, hierzu einige Beispiele::

- Strukturierung der Informationseinheit (Dokument) durch Aspektierung bei Siemens-Projekt CONDOR (es wurde modifiziert und wesentlich erweitert von GOLEM-System übernommen).
- Textfacetten, d.h. ein Objekt ist nicht als ganzes beschrieben, sondern nach seinen Eigenschaften (z.B. Gift: Chemische Substanz, sinnliche Eigenschaften - wie z.B. Farbe, Geruch, Zustand etc. - Vergiftungssymptome, Gegenmaßnahmen etc.); leider existieren bisher keine Werkzeuge, die diese Beschreibung zumindest semi-automatisch ermöglichen.

Der Benutzer konnte dann durch Angabe von Aspekten und/oder Teilschemata/Textfacetten festlegen, welche Textteile für die folgende Recherche berücksichtigt werden sollen. Ähnlich kann man dann auch bei den Dokumentensammlungen vorgehen, die schon mit Werkzeugen erstellt wurden, die eine Trennung von Inhalt und Layout ab Erstellung ermöglichen. Einerseits werden die Probleme mit der Indexierung von großen Dokumenten in die inhaltliche Beschreibung ihrer thematisch zusammenhängenden Komponenten verlagert, andererseits werden relevante Informationen in einer ausreichenden Qualität und Quantität (ohne den Ballast des sonstigen informationslosen Dokumententextes) schnelles und präziser ermittelt sowie an den richtigen Ort verteilt (Prämissen der Informationslogistik).

Die verbindliche Verwendung von geeigneten, einheitlichen Werkzeugen zur inhaltlichen (logischen) und physischen (Layout) Strukturierung von Dokumenten kann auch den o.g. Engpaß (d.h. die fehlende konzeptionelle Datenmodellierung und daraus folgende Probleme mit der Identifikation originärer Informationsquellen) bei dem Einsatz von Führungsinformationssystemen (EIS) beseitigen helfen.

4.4 Durchgängigkeit im Prozeß durch Dokumentenaustausch

Dokumente werden oft benötigt, um die Informationen zwischen den Beteiligten eines Arbeitsprozesses auszutauschen. Dies trifft auf Verwaltungsabläufe ebenso zu, wie auf die Weitergabe von Entwicklungsunterlagen an die Autoren von Servicedokumentation oder die Weitergabe der Daten in einem CIM-Prozeß (CAD/CAP/CAM - Kette).

Das Konzept elektronischer Verfahrensketten, egal ob beim Publizieren oder in anderen Anwendungsgebieten, wie z.B. bei der o.g. computerintegrierten Fertigung CIM (CAD/CAP/ CAM - Kette) oder Entwicklungsdokumentation, lebt davon,

daß ein Kernbestand von Basisdaten (oder Dokumente bzw. ihrer Teile) weitergereicht und in den nächsten Stufen des Prozesses angereichert wird (Wiederverwendung von Dokumentation oder ihrer Teile und nutzungsgerechte Informationsverteilung). Dokumente als Container für Informationen ermöglichen jedoch keinen durchgängigen Informationsfluß zwischen den beteiligten Funktionen, wenn mit dem Austausch auch ein Konvertierungsaufwand verbunden ist oder sogar ein erneutes Eingeben.

Bei Dokumentformaten, mit denen nicht der Inhalt, sondern die Darstellungsanweisungen kodiert werden, kann kein Informationsaustausch erfolgen, lediglich ein Layout-Austausch. Damit ist es jedoch nicht möglich, die in einem Dokument enthaltenen Informationen wiederzuverwenden, automatisch weiterzuverarbeiten oder für ein nutzergerechtes Verteilen aufzubereiten.

Um beim Dokumentenaustausch einen durchgängigen Informationsfluß zu ermöglichen, bedarf es einerseits eines gemeinsamen Verständnisses der auszutauschenden Information sowie eines Austauschformats mit dem sich die Informationseinheiten im Dokument beschreiben lassen.

4.5 Automatisiertes Erzeugung von Dokumentation

Dokumentation wird bislang fast ausschließlich von Redakteuren geschrieben. Wenn sich Dokumentation auch nicht vollständig automatisiert erzeugen läßt, kann der Autor jedoch weit mehr Unterstützung durch das Textverarbeitungssystem erwarten, als z.B. das Erstellen von Verzeichnissen oder das automatische Numerieren von Überschriften. Den Inhalt der Dokumentation bereitzustellen, bleibt Aufgabe des Redakteurs; er kann jedoch sowohl beim Erzeugen der erforderlichen Dokumentstruktur und beim Auffüllen dieser Struktur mit bereits existierenden Informationen unterstützt werden, als auch bei einem späteren Veredeln der Dokumentation (z.B. mit Verweisen auf andere Quellen, Indizes und Verzeichnissen).

Steht fest, welche Struktur das zu erzeugende Dokument haben muß, welche Informationen möglich und welche notwendig sind, kann eine leere Struktur des Dokuments automatisch erzeugt werden, die der Autor oder ein Generator nur noch mit den jeweiligen Inhalten füllen muß. Die leere Struktur selbst kann auch aus den Informationen in anderen Dokumenten gewonnen werden. Beispielsweise kann die Struktur einer Programmbeschreibung aus den Strukturen des Programms erzeugt werden. Auch lassen sich kundenspezifische Varianten aus einem umfassenden Dokumentationsbestand generieren oder vorliegende Bestandteile automatisch in einem Dokument integrieren. Grundvoraussetzung ist, daß die Struktur des zu erzeugenden Dokuments explizit ist. Gibt es bereits Ausgangsmaterial für das zu erzeugende Dokument, muß spezifiziert sein, welche Informationen an welcher Stelle und in welcher Form im Zieldokument übernommen werden sollen.

Automatisiertes Erzeugen von Dokumentation ist vor allem aus Kosten- und aus Qualitätssicht bedeutsam. Aus Kostensicht, weil Dokumentation gezwungener

Maßen meist als "Abfallprodukt" gesehen wird. Aus Qualitätssicht, weil nur eine weitgehend automatisierte Verarbeitung die erforderliche Qualität sichert. Es kann davon ausgegangen werden, daß die hier betrachteten Information und Dokumentation i.a. logisch inhaltlich gut bis sehr gut strukturierbar ist. Es liegt also die Forderung nach einer automatisierten Erzeugung nahe.

Ein zentraler Schwerpunkt dabei ist die strukturelle Transformation von Dokumenten (Werkzeuge Metamorphosis, Balise, etc.). Damit ist es möglich, Dokumentation und Information für die jeweiligen Bedürfnisse maßzuschneidern. Aus der Distributionssicht führt dies zu *Single-Source-Publishing*, aus konzeptueller Sicht zu *Single-Source-Information*.

4.6 Single Source Publishing /Single Source Information

Technische Dokumentation wird i.d.R. in mehreren Varianten publiziert, entweder um bestimmte Zielgruppen (Sprachvarianten, nutzungsspezifische Varianten, ...) oder um verschiedene Publikationsmedien (HTML, PDF, Papier, ...) zu bedienen. Bislang wird zunächst ein Originaldokument erstellt, aus dem anschließend alle benötigten Varianten erzeugt werden. Dabei besteht zum einen die Gefahr, daß es bei Änderungen an redundant dokumentierten Informationen zu inhaltlichen Inkonsistenzen kommen kann, wie bereits im Abschnitt 'Konsistenzproblematik' besprochen wurde. Zum anderen ist beim Erzeugen gerade von verschiedenen Publikationsvarianten oft ein manueller Konvertierungsaufwand erforderlich, der zudem bei jeder nachfolgenden Veränderung erneut anfällt.

Hier sollte die Information daher redundanzfrei in einem einzigen Dokument verwaltet werden, aus dem - nach Bedarf - das jeweils gewünschte Publikationsformat generiert wird (*Single Source Publishing*). Die Voraussetzung dafür ist erneut, daß die *Informationen* im Dokument zugreifbar sind und daß spezifiziert ist, wie diese Informationen in den verschiedenen Formaten dargestellt werden soll und welche Sichten auf das Dokument jeweils gewünscht werden..

Die Vorteile eines solchen Ansatzes sind:

- Ändern sich die Anforderungen an das Layout, muß das Ausgangsdokument nicht verändert werden, sondern lediglich die Anweisung, wie die Informationseinheiten im jeweiligen Format dargestellt werden sollen. Beispielsweise wird die im WWW verwendete Sprache HTML zur Darstellung von Dokumenten oft für neue Versionen des Browsers erweitert. Damit ältere Versionen nicht antiquiert wirken, muß lediglich der Einsatz der neuen Sprachmittel neu spezifiziert werden; das Generieren der Publikation erfolgt automatisch.
- Aus einer Quelle können über formatspezifische Filter die Publikationen für *verschiedene* Medien bedient werden, ohne manuellen Zusatzaufwand.

Die gewünschten Varianten können als unterschiedliche Sichten aus einem redundanzfreien Dokumentationsbestand generiert werden; die Pflege der Informationen erfolgt zentral.

Qualitativ hochwertige Dokumentation und Information setzt auch korrekte, widerspruchsfreie Primärinformation voraus. Unter Abwägung von Aufwand und Nutzen ist von einem Idealszenario auszugehen, bei dem die Quelle für jedes Dokumentationsfragment und jede Informationseinheit eindeutig verifizierbar ist. Quelle bedeutet in diesem Zusammenhang nicht ausschließlich ein technisches System, wie z.B. eine Datenbank. Quelle meint ebenso den menschlichen Autor. Dies gestattet, sowohl einzelne Dokumente als auch ganze Dokumentverbände als Informationsressource, ähnlich einer Datenbank zu nutzen. Voraussetzung hierfür ist

- die Einführung konzeptuellen Dokument- und Informationsschemata (Document Type Definition - DTD);
- die formale inhaltlich, logische Strukturierung (Logical Markup).

Die Datenintegrität von Datenbank-Management-Systemen kann verständlicherweise nicht übertragen werden.

5. Weltweite Aktivitäten in Auswahl

- Bei der Firma Ericsson läuft seit 1991 ein strategisches Programm zur Geschäftsprozeß-Optimierung mittels durchgängiger, SGML-basierter Dokumentation, mit einem Aufwand von ca. 140 Mio. DM (bis Ende 1995). Dazu folgendes Zitat aus: *A Guide to SGML at Ericsson*, Spring 1995: "By being the first to deliver documentation using electronic media that adheres to the SGML standard, Ericsson can influence or even define the branch standard for this application, and side-step the costly and time-consuming process of converting our documentation".
- Der Deutsche Telekom fordert im Rahmen des TBL-P Projekts, an dem u.a. Alcatel SEL und Siemens ÖN als Zulieferer beteiligt sind, die technische Dokumentation in SGML bezüglich einer vorgegebenen, inhaltlichen Struktur (Telekom DTD), in der Informationseinheiten, wie z.Z. Kommando, Warnungshinweis u.s.w., gekennzeichnet sind.
- Bei der Firma Newbridge wird die gesamte technische Dokumentation auf der Basis von SGML erstellt, verwaltet, gepflegt und publiziert. Zitat (<http://www.newbridge.com/Information/CustDoc/index.html>): "Newbridge has adopted SGML, an ISO standard for the markup of text and definition of document hierarchy, as the basis for its development of online documentation. This standard allows Newbridge not only to create a better, more sophisticated product, but also to deliver it in ways that make sense for customers' specific needs. In the future, Newbridge customers can choose to receive their documentation on CD-ROM, paper or across the Internet."
- Siemens ÖN Breitbandnetze hat eine Allianz mit Newbridge (Kanada) gebildet. Zu diesem Zweck wird die Kundendokumentation bestimmter Produkte in SGML ausgetauscht. Damit Siemens und Newbridge (NB) diese Dokumentation jeweils bei sich in einen Produktionsprozeß einfließen können, wird Siemens-

Dokumentation auf NB-DTD transformiert und umgekehrt (man spricht dabei auch über Allianz-Dokumentation).

- British Telecom betreibt eine zentrale Anlieferungsstelle, bei der die technische Dokumentation in SGML vom Hersteller empfangen, kontrolliert, freigegeben und als SGML selber oder mittels einer Echtzeit-Konvertierung nach HTML im BT-Intranet verteilt wird.
- Verschiedene Fachverlage setzen auf SGML als internes Verwaltungsformat für seine Publikationen auf. So z.B. nutzen die Verlage Dr. Otto Schmidt und Carl Heymanns SGML zur Qualitätssicherung ihrer Erzeugnisse.
- Beim Deutschen Institut für Normung (DIN) gibt es ein Projekt, die Normen mit SGML inhaltlich zu strukturieren. Mit der vom Beuth Verlag im April 1997 publizierten CD-ROM mit SGML-Daten können Firmen die für sie relevanten Normen durch ein automatisches Konvertieren nach HTML in ihrem Intranet publizieren.
- In der zivilen und militärischen Flugzeugindustrie (Airbus, Lufthansa, British Aerospace, Royal Air Force, ...) wird die technische Dokumentation mit SGML ausgetauscht und von den Fluglinien automatisch weiterverarbeitet und publiziert.
- Die Firma Novell erzeugt, verwaltet und publiziert ihre Online-Dokumentation auf der Basis von SGML.
- Seit zwei Jahren gibt es die - inzwischen unter der Schirmherrschaft des *World-Wide Web Consortiums* (W3C) stehende - Initiative *Extensible Markup Language* (XML), mit der sich im Internet zukünftig Informationselemente publizieren lassen, und nicht mehr nur darstellungsorientierte HTML-Seiten. Inzwischen sind Microsoft und Netscape an der Spezifikation von XML beteiligt. Microsoft hat zudem angekündigt, daß XML "im Internet Explorer 4.0 verwendet werden wird".

6. Resume und zugleich Zusammenfassung

Durch die Trennung der inhaltliche und logischen Struktur wird die in den Dokumenten enthaltene Information explizit und maschinell weiterverarbeitbar gemacht. Durch diese Eigenschaften können dann die Grundprämissen der Informationslogistik erfüllt werden. Dabei werden sowohl strukturierte Dokumente als auch Datenbanken als Container für Informationen bzw. Informationseinheiten (Wissenseinheiten) angesehen, die editiert und gepflegt werden sowie recherchierbar und adäquat publizierbar gemacht werden müssen.

Literatur:

Hinweise zu Aktivitäten im Abschnitt 5 und zu Initiativen im Abschnitt 3.1 können alle im Internet gefunden werden. Die im Text erwähnten ISO-Normen sind allgemein bekannt. Zur effektiven Nutzung von strukturierter Dokumentation in verschiedenen Anwendungen (wie z.B. in Information-Retrieval-Systemen oder in Führungsinformationssystemen) findet man in der Literatur bisher so gut wie keine Hinweise (vgl. die im Literaturverzeichnis angegebene Conference Proceedings).

Dublin Core Metadata: <URL:http://purl.org/metadata/dublin_core> 1997.

Heeter, G.: Technical Information - the North American Customer. In: Proceedings der 7. Fachtagung Technische Dokumentation, Siemens AG, München 1998.

Möller, H.: Konsistentes Dokumentenengineering mit protokollierten Dokumenten. Diss. Berlin: FU Berlin, FB Mathematik und Informatik 1997.

SGML Europe '96: Conference Proceedings (Conf. Chair: Pamela Gennusa), München 1996. Alexandria (USA): Graphic Communication Ass. 1996.

SGML / XML '97: Conference Proceedings, Washington D.C. 1997. Alexandria (USA): GCA Graphic Communication Ass. 1997.

SGML / XML Europe '98: Conference Proceedings, Paris 1998. Alexandria (USA): GCA Graphic Communication Ass. 1998.