

Datenbank, Hypertext oder linearer Text? – Softwaretypen¹

Rainer Dechsling

*Tagung der Arbeitsgruppe
"CD-Rom/informationsorientiert
" des Arbeitskreises
"Elektronisches Publizieren"*

*Die Leipziger Empfehlungen
(vgl. in diesem Heft S. 2696–2699)*

Strukturformen der Information

Am 18. März 1994 tagte in Leipzig die Arbeitsgruppe CD-Rom/informationsorientiert des Arbeitskreises Elektronisches Publizieren. Neben rechtlichen Fragen, die gemeinsam mit der Arbeitsgruppe Recht desselben Arbeitskreises diskutiert wurden, ging es auch um die "Vorstellung und Diskussion der Empfehlungen der Arbeitsgruppe für die Anforderungen an EP-Produkte". Diese von einem Unterausschuß erarbeiteten Empfehlungen wurden am selben Tag vorgelegt, diskutiert und dann – angesichts der Tagesordnung durchaus überraschend – auch sogleich verabschiedet. Der Beschlußfassung ging eine engagiert geführte Debatte voraus, deren wesentliches Ergebnis eine genauere Bereichsdefinition der Empfehlungen war. Anstelle von "Leipziger Empfehlungen zum Elektronischen Publizieren" wurde am Ende sinngemäß von "Leipziger Empfehlungen zur Umsetzung von Büchern auf elektronische Medien" gesprochen.

Völlig zutreffend gehen die "Leipziger Empfehlungen" davon aus, daß die Verlage die mit der Kulturtechnik Lesen entwickelten Methoden der Informationsaufbereitung in die neuen Informationstechnologien bestmöglich zu übertragen und jedenfalls nicht ohne Not über Bord zu werfen haben. Elektronische Publikationen seien in die Kontinuität der gedruckten Werke zu stellen. Sie hätten – und da beginnt auch schon die Diskussion – an die Funktionalität des Buches und damit auch an die Lese- und Arbeitsgewohnheiten der Buchnutzer anzuknüpfen. Deshalb müßten buchnahe funktionelle Anforderungen wie der Zugang über Inhaltsverzeichnisse und Gliederungen, eine Notizfunktion oder die Einblendung lebender Kolumnentitel am Bildschirm zur Verfügung stehen.

In der ursprünglich vorgelegten Fassung der Empfehlungen wurde übersehen, daß es neben der buchkompatiblen Darstellung sequentieller Information zwei weitere Strukturformen der Information zu beachten gilt:

- Die "Kulturtechnik Lesen" umfaßt nicht nur das Lesen von Büchern, sondern ebenso das Studium von Katalogen und Verzeichnissen ("VLB", "Wer liefert was?") und von Fachzeitschriften. Gedruckte Nachweisbände und jahrgangsweise gebundene Fachzeitschriften sind zur sequentiellen Lektüre regelmäßig schlecht geeignet. Es handelt sich um nicht-linearen Text, der über ausgefeilte Register erschlossen werden muß. Derartige Informationen sind in Datenbanksystemen gut aufgehoben. Insofern besteht ein deutlicher Unterschied zum klassisch-sequentiellem Buch, das beim Abfassen der "Leipziger Empfehlungen" Pate gestanden haben dürfte.
- Viele Verlage wollen nicht lediglich gedruckte Information im Verhältnis eins zu eins auf das elektronische Medium übertragen, sondern eigenständige elektronische Informationsprodukte entwickeln. Dabei werden Texte ganz bewußt als kurze und in sich verständliche Einheiten konzipiert und sinnvoll miteinander verknüpft. Die Erstellung und bildschirmgerechte Darbietung der Texte ist nicht-linear. Es handelt sich um klassischen Hypertext, der heute meist mit Multimedia-Elementen (Bild, Ton, Video) angereichert wird.

*Auf der Suche nach der
allumfassenden
Retrieval-Software*

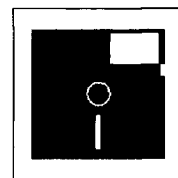
Somit zeigt sich, daß textuelle Information grundsätzlich in zwei vorwiegend nicht-linearen Formen (Datenbank, Hypertext) sowie in linearer Gestalt (Buch) auftreten kann. Jede dieser drei Textgestalten findet ihre Entsprechung in spezifischer Abfragesoftware. Die allumfassende marktgängige Software, die alle drei Ausprägungen der Information vollständig abdeckt, gibt es derzeit nicht. Die nachfolgenden Ausführungen sollen Informationsanbietern und Fachverlagen helfen, den nach dem heutigen Stand der Technik richtigen Softwaretyp zu finden.

Klassische Datenbanken

Der Kategorie der Nachweisdatenbanken dürfte derzeit im Bereich der elektronischen Fachinformation die größte wirtschaftliche Bedeutung zukommen. Kataloge und Verzeichnisse wie das "VLB", "Wer liefert was", "Perinorm" (mehrsprachiger Nachweis technischer Regeln) oder die Leitsatzkartei des deutschen Rechts sind CD-Rom-Projekte, die

Dr. Rainer Dechsling leitet das Lektorat für elektronisches Publizieren des Verlags C. H. Beck. Er ist u. a. für das Projekt "Schönfelder plus" auf CD-ROM verantwortlich.

¹ Nachdruck mit freundlicher Genehmigung des Börsenblatts aus: Börsenblatt Nr. 50 vom 24. Juni 1994, S. 19–22.



Märkte geöffnet haben. Der durchschlagende Erfolg dieser Nachweisdatenbanken beruht unter anderem auf der weitverbreiteten und zutreffenden Erwartungshaltung der Käufer, daß Kataloge und Verzeichnisse zur Umsetzung auf den PC bestens geeignet seien. Aus der Sicht der Verlage lassen sich derartige Projekte relativ unproblematisch als Database Publishing organisieren. Die Substitution der zugrundeliegenden Druckwerke ist erheblich.

Meist erst im zweiten Schritt werden textorientierte Fachzeitschriften auf CD-Rom zur Verfügung gestellt. Es handelt sich um Volltextdatenbanken, die den Abonnenten des gedruckten Werks als raum- und zeitsparendes elektronisches Archiv dienen. Eine nennenswerte Substitution der gedruckten Abonnements durch das elektronische Angebot findet regelmäßig nicht statt, weil es der CD-Rom an Aktualität und Umfeld (zum Beispiel Anzeigen und Begleitinformationen im Umschlagteil) fehlt. Die verlegerischen Kosten für die datenbankgerechte Digitalisierung der Altjahrgänge sind erheblich. (Für die aktuelle elektronische Zeitschrift, auf die hier nicht eingegangen werden kann, gelten ganz andere Mechanismen. Dafür bieten sich Tools wie Adobe Acrobat an.)

Die Abfragesoftware bei Nachweis- und Volltextdatenbanken sollte folgende Charakteristika aufweisen:

- Abbildung einer differenzierten *Feldstruktur* in Suchmaske, Trefferliste und Vollanzeige.
- Möglichkeit des Verlags, den Anwendern mehrere Suchmasken, Trefferlisten und Vollanzeigen zur Auswahl anzubieten, um trotz komplizierter Feldstruktur eine einfache und transparente Recherche zu ermöglichen.
- Felder vom Anwender einzeln ein- und auszublenden.
- Indizes jeweils mit einem Tastatur- oder Mausklick aufzuschlagen.
- Einfache Navigation im Index, zum Beispiel Positionierung des Cursors aufgrund Benutzereingabe.
- Markierung im Index. Direkte Übernahme von Suchbegriffen aus dem Index für die Suche.
- Synonymfunktion und Thesaurus. Suchoperatoren: UND, ODER, OHNE, NICHT, >, >=, <, <=, =, Bereich, Wort-, Satz- oder Absatzabstand.
- Flexible Datumssuche (Mai 1985, 5.85, 1.5.85, 1985-05-01 etc.).
- Schnelle Links-, Mitte- und Rechtstrunkierung.
- Benutzerkontrollierte Trunkierungsauflösung.
- Grundformenerkennung.
- Geschachtelte Klammersetzung, feldübergreifende Verknüpfung.
- Wahlweise unscharfe Suche (fuzzy logic).
- Automatische Rechtstrunkierung. Numerisch auch in Kombination mit Wort, Zeile, Feld.
- Zugriff auf frühere Suchen.
- Indizierung nach Wort, Zeile oder Feld.
- Sprache (Deutsch, Englisch, Amerikanisch, Französisch etc.).
- Stopwortliste.
- Differenzierte Behandlung von Interpunktion und Akronymen.
- Bindestrichbegriffe ein-, zwei- oder dreifach indizierbar.
- Optimierung der Suchgeschwindigkeit nach dem Zahlenverhältnis der unterschiedlichen Indexeinträge und der nachgewiesenen Datensätze.
- Gemeinsames Indizieren mehrerer Felder in beliebiger Kombination.
- Mehrere Indizes pro Feld.
- Paralleles Suchen in mehreren Fenstern.
- Querverweissuche aus dem gefundenen Datensatz heraus (soft wired hypertext).
- Verfolgung fester Verweise (hard wired hypertext) mit einem Klick.
- Jederzeitiges Umschalten zwischen Suchmaske, Trefferliste und Vollanzeige sowie zwischen verschiedenen Suchen unter vollständiger Erhaltung des jeweiligen Status.
- Sortierung oder Ranking der gefundenen Datensätze nach verschiedenen Kriterien.
- Markierung einzelner Datensätze in positiver oder negativer Selektion für die Weiterverarbeitung (Anzeige, Ausdruck, Speicherung, Zwischenablage).
- Kopf- und Fußzeile in der Vollanzeige.
- Präzises Highlighting entsprechend der eingegebenen Suche.
- Präziser Direktsprung zu den gesuchten Passagen des Datensatzes.
- Expliziter Sprung zum nächsten oder vorigen gefundenen oder benachbart gespeicherten Datensatz (logisches und lineares beziehungsweise physisches Blättern).
- Volle Zitierfähigkeit nach Seiten- oder Blattzahl.
- Zügiges Blättern auch in sehr langen Datensätzen (Volltext größer 64 KB).

Volltextdatenbanken für textorientierte Fachzeitschriften

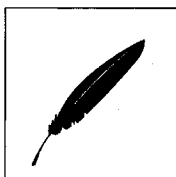
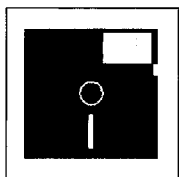
Software-Eigenschaften

Feldbezogene Suchoptionen

Weitere Indexoptionen des Verlags

Navigation

Vollanzeige



Datenbank, Hypertext oder linearer Text?

Suchgeschwindigkeit

Produktionsprozeß

Empfehlenswerte Software

Weniger empfehlenswerte Software

Primär für das elektronische Medium produzierte Texte

Der klassische Hypertext-Ansatz

Typische Hypertext-Systeme

- Unterstützung von Fließtext und geschützten Bereichen (Tabellen).
- Typografische Hervorhebungen auf Feldebene und innerhalb von Feldern (bei zeichenorientierten Dos-Anwendungen wenigstens Highlighting und Invertierung).
- Darstellung von Formeln und Grafiken. Zoomfunktion für Grafiken.
- Für den Verlag: Ausgefeilte Anzeigelogik auf Feldebene.
- Im Sekundenbereich auch in CD-Rom-Indizes mit mehreren Millionen Einzelvorkommnissen.
- Dies auch bei Trunkierung und Abstandssuche.
- Meldung der Trefferzahl vor Aufbau der Trefferliste.
- Zügiger Aufbau der Trefferliste.
- Möglichkeit zur Vollanzeige vor Aufbau der Trefferliste.
- Unterstützung diverser Text-, Grafik- und Datenbankformate beim Einlesen der Daten (SGML, RTF, BMP, TIF, dBase, Oracle, ASCII oder ANSI mit frei definierbaren Delimitern etc.).
- Umsetzung von Zeichen und Zeichenketten.
- Schnelle Indizierung auch bei sehr großen Datenbeständen.
- Menügeführte Anwendungsentwicklung ohne Programmierung.
- Umfangreiche Statistiken und Testmöglichkeiten.
- Vollautomatische Erzeugung des Installationsprogramms und des Premasters für die CD-Rom.

Ohne Anspruch auf Vollständigkeit seien die in Deutschland marktgängigen Softwareprodukte genannt, welche die genannten Anforderungen weitgehend erfüllen: CD Answer / CD Author (Dataware Technologies), Cobra (EPS), OptiSearch (MPW Lasec), CDIS (Makrolog).

Grundsätzlich abzuraten ist von relationalen Datenbanken ohne Volltextkomponente (Clipper und andere dBase-Compiler, MS Access Distribution Kit etc.), denn die meisten Kataloge und Verzeichnisse enthalten bei genauerem Hinsehen Felder, in denen eine Volltextsuche bessere Ergebnisse liefert als die dBase-übliche indizierte Feldbeziehungsweise nicht-indizierte sequentielle Suche. Auch die nachfolgend aufgeführten Hypertextsysteme und buchkompatible Software sind für Nachweis- und Volltextdatenbanken regelmäßig ungeeignet, da ihnen wichtige Funktionen wie differenzierte Suchoptionen oder Trefferlisten fehlen.

Hypertextsysteme für originär elektronische Produkte

Informationsorientierte elektronische Medien müssen nicht eins zu eins aus einem gedruckten Werk gewonnen werden. Es gibt beispielsweise im juristischen Bereich Verlage, die primär für das elektronische Medium nicht-linearen Text produzieren. Anhand einer detailliert vorgegebenen Systematik werden zahlreiche Einzelthemen von externen, per Verlagsvertrag gebundenen Autoren bearbeitet. Jedes Thema muß aus sich heraus verständlich sein und darf nicht länger als zwei bis drei Bildschirmseiten (mit jeweils zirka 20 Zeilen) sein. Die elektronisch dezentral erfaßten Texte werden von einem Herausgeber oder Wissensingenieur überprüft und miteinander verknüpft. Multimediale Erweiterungen mit Bild, Ton und Video sind möglich.

Damit ist auch schon der klassische, auf nicht-linearem Text basierende Hypertext-Ansatz im Grundsatz beschrieben. Die Navigationsprobleme in einem derart dezentral vernetzten Textraum sind beträchtlich. Dennoch kann es sinnvoll sein, dem Ausruf "Lost in hyperspace!" zum Trotz nicht-lineare Informationssysteme zu schaffen. Besonders im institutionellen Bereich (Großunternehmen, Museen etc.) spielen Hypertextsysteme schon heute eine bedeutende Rolle. Auch im Verlagsbereich gibt es Projekte, bei denen ein originäres Hypertextsystem der schlichten Umsetzung vorhandener Druckwerke vorzuziehen ist. Dabei kann durchaus auf den bestehenden Autorenstamm zurückgegriffen werden, soweit er dem PC aufgeschlossen gegenübersteht.

Als typische Hypertextsysteme sind zu nennen: Hypercard (Apple), Toolbook, Knowledge Pro, Multimedia Viewer (Microsoft). Der Übergang zum Multimedia-Bereich ist fließend. Dort gibt es weitere Tools.

Buchkompatible Software für linearen Text

Bislang wurden die Nachweis- und Volltextdatenbanken sowie der klassische Hypertext als überwiegend nicht-lineare Textgattungen behandelt. Von der ersten bis zur letzten Seite linearer Text findet sich herkömmlich vor allem in Büchern, sei es als Shakespeares "Ham-

let" oder als technisches Handbuch. Wenn nun von buchkompatibler Software gesprochen wird, ist damit also Software gemeint, die zur Darstellung von linearem Text besonders geeignet ist. Die eingangs erwähnten "Leipziger Empfehlungen" zielen genau darauf ab. Wichtige Anforderungen an buchkompatible Software sind:

- Buchtypografie am Bildschirm und im Ausdruck
- Zugang auch über Gliederung (Inhaltsverzeichnis)
- sequentielle Textdarstellung
- komfortable Fußnotenverwaltung
- Notizfunktion, Lesezeichen, exzerpieren
- lebender Kolumnentitel beziehungsweise mitlaufende Gliederung
- dreidimensionales Blättern (von Abschnitt zu Abschnitt außerhalb der Treffermenge – zum nächsten oder vorigen gefundenen Datensatz – Direktsprung von Treffer zu Treffer).

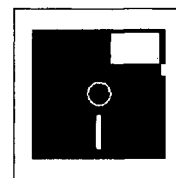
Für die oben beschriebenen nicht-linearen Informationssammlungen sind diese Anforderungen teilweise wenig bedeutsam oder sogar unpassend. So ist etwa der Zugang über eine Gliederung bei Nachweis- und Volltextdatenbanken häufig nur um den Preis intensiver redaktioneller Nachbearbeitung realisierbar, weil die zugrundeliegenden Kataloge, Verzeichnisse und Fachzeitschriften lediglich alphabetisch oder chronologisch geordnet und bestenfalls mit wenig interessanten, ständig wiederkehrenden Rubriken versehen sind. Auch die sequentielle beziehungsweise lineare Textdarstellung ist in nicht-linearen Datenbanken regelmäßig belanglos. Sinn macht darin allenfalls das lineare (beziehungsweise sequentielle beziehungsweise physische) Blättern außerhalb der Treffermenge, um die ursprüngliche alphabetische, chronologische oder sonstige Ordnung am Bildschirm nachvollziehen zu können. Auch das kann aber durch eine parallele Suche zum Beispiel nach den Seiten 800 bis 850 eines bestimmten Zeitschriftenjahrgangs ersetzt werden, wenn nur die Datensätze in der Folge des gedruckten Vorbilds auf dem elektronischen Medium gespeichert sind. Wenig sinnvoll wäre es, sämtliche Datensätze einer nicht-linearen Datenbank als Endlostext darzustellen, da das kaum merkliche Überfliegen von Datensatzgrenzen die konzentrierte Arbeit in heterogenen Datenbeständen erschwert.

Folgende in Deutschland gängige Softwareprodukte verdienen das Etikett buchkompatibel: CD Answer Hypertext (Dataware Technologies), Folio Views, Dynatext. Mit der Einschränkung, daß der lineare Text am Bildschirm nicht sequentiell, sondern datensatzbezogen mit der Möglichkeit jederzeitigen linearen Blätterns dargestellt wird, sind auch CDIS (Makrolog) und Multimedia Viewer (Microsoft) zu nennen. Auch diese Nennungen können nicht den Anspruch auf Vollständigkeit erheben.

Konvergenz und Konnektivität

Eine für alle linearen und nicht-linearen Textgattungen gleichermaßen perfekt geeignete Abfragesoftware gibt es derzeit nicht. Der in verschiedenen Gremien seit Jahren immer wieder unternommene Versuch, das textorientierte elektronische Publizieren auf ein einziges Softwareprodukt zu vereinheitlichen, ist beim jetzigen Stand der Technik zum Scheitern verurteilt, weil kein Verlag seinen Kunden den Verzicht auf werkbezogen sinnvolle und längst selbstverständliche Funktionalität zumuten kann. Das ist indes kein Grund zur Resignation:

- Die Anwender haben unter den immer stärker dominierenden grafischen Benutzeroberflächen (Windows, Apple etc.) viel weniger Aufwand, sich in neue Anwendungen hineinzufinden, als dies unter Dos der Fall ist (war). Die Bedienerfreundlichkeit ist deshalb kein besonders starkes Argument für eine Festlegung der Verlage auf eine bestimmte Abfragesoftware.
- Die Nachfragemacht der Verlegerseite gegenüber den Herstellern und Distributoren von Abfragesoftware ist gewiß dann am größten, wenn sie gebündelt auftritt. Andererseits brächte die konsequente Einigung auf ein bestimmtes Abfragetool auch einen Verlust an Konkurrenz, der die technische Entwicklung zu behindern geeignet wäre. Es fragt sich deshalb, ob es eine gute Strategie wäre, durch die gezielte Förderung eines – zumindest heute noch – suboptimalen Softwareprodukts auf der Anbieterseite oligopolistische oder sogar monopolistische Strukturen zu fördern.
- Unterhalb der Programmoberflächen tut sich so einiges. Produkte wie CD Answer, Folio Views und MS Multimedia Viewer erlauben schon heute, dateiübergreifend zwischen CD-Rom-Anwendungen verschiedener Verlage zu springen, sofern diese unter



Anforderungen an buchkompatible Software

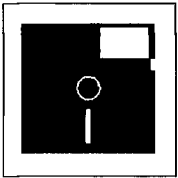
Unterschiedliche Software-Anforderungen für lineare/nicht-lineare Informationssammlungen

Beispiele für buchkompatible Software

Vereinheitlichungsbestrebungen z. Z. zum Scheitern verurteilt

Die Rolle graphischer Benutzeroberflächen

Connectivity



derselben Software laufen und der Anwender über die nötige Laufwerkskapazität (CD-Rom oder Festplatte) verfügt. Für CD Answer (Dos, Windows, Unix, Apple), CD Answer Hypertext (Dos, Windows) und den Microsoft Viewer (Windows) existieren Schnittstellen, die es erlauben, mit Hilfe von Programmiersprachen oder Autoren-systemen auf die Datenbestände zuzugreifen. (So weit ist Folio Views noch nicht.) Praktisch alle Softwareanbieter geloben, ihre Produkte konsequent auf eine plattform-, das heißt betriebssystemübergreifende Client/Server-Architektur umzustellen, die eine umfassende verlagsübergreifende Connectivity erlauben wird.

Angesichts der faktischen Aussichtslosigkeit einer Einigung auf eine bestimmte Abfragesoftware, jedenfalls zum jetzigen Zeitpunkt, eröffnen sich für die Verlage zwei Szenarien, die sich gegenseitig durchaus nicht ausschließen:

Die umfassende Software mit konfigurierbarer Funktionalität

Szenario 1

Es gibt CD-Rom-Softwarehäuser, die daran arbeiten, die Dualität zwischen nicht-linearem und linearem Datenbankkern mit einem umfassenden Produkt zu überwinden. Da die komplette Funktionalität einer entsprechenden Abfragesoftware verwirrend wäre, geht der Trend zur konfigurierbaren Funktionalität. Damit ist gemeint, daß der Verlag im Produktionsprozeß entscheidet, wieviel und vor allem welche Funktionalität er für das konkret geplante elektronische Produkt benötigt. Der Anwender wird dann nur die Funktionen, die für den betreffenden Datenbestand Sinn machen, in der Abfragesoftware vorfinden.

Offenlegung der Schnittstellen

Szenario 2

Die Verlage gehen daran, die Schnittstellen ihrer elektronischen Produkte offenzulegen. Softwareanbieter, die hier technische oder lizenzrechtliche Barrieren aufbauen, werden von den jeweils interessierten Verlagen dazu bewegt, diese Barrieren abzubauen. Nach und nach werden alle elektronischen Produkte für gegenseitige Zugriffe geöffnet. Nachweis- und Volltextdatenbanken, Hypertext- und Multimediasysteme sowie buchkompatible Produkte bis hin zu den hier noch gar nicht erwähnten Lexika, Lernprogrammen, Berechnungsprogrammen und Expertensystemen werden in einigen Jahren so durchlässig sein, wie es unter Windows heute schon die gängigen Büroprogramme (Textverarbeitung, Tabellenkalkulation etc.) sind.

... auf die richtige Software kommt es an!

Fazit

Verlage, die im Begriff sind, in das textorientierte elektronische Publizieren einzusteigen oder eine bereits existierende Produktpalette zu erweitern, sollten genau prüfen, welcher Textgattung ihre Daten zuzurechnen sind. Handelt es sich um Kataloge, Verzeichnisse oder Fachzeitschriften? Dann empfiehlt sich eine marktgängige Volltextdatenbank. Soll ein originär elektronisches Informationsprodukt erstellt werden? In diesem Fall ist wahrscheinlich ein klassisches Hypertextsystem das richtige Tool. Geht es um die Umsetzung sequentiell strukturierter Bücher? Dann sollten die "Leipziger Empfehlungen zur Umsetzung von Büchern auf elektronische Medien" zu Rate gezogen werden. Nur bei Lexika, Nachschlagewerken und Anthologien dürfte es regelmäßig unerheblich sein, ob Datenbanksoftware oder buchkompatible Software (jeweils mit Volltextsuche und Hypertextfunktion) zum Einsatz kommt. Wer schon heute den Gedanken der Konnektivität zur Geltung bringen will, muß sich beim Softwarehaus nach den beiden oben genannten Szenarien erkundigen. Darüber hinaus – was allerdings nicht Thema dieses Beitrags war – sollten die Kriterien prompter Service, Verlagsunabhängigkeit, kontinuierliche und nachhaltige Weiterentwicklung der Software sowie eine darauf abgestimmte, akzeptable Preispolitik erfüllt sein.