

Erfahrungen beim Aufbau der ersten landesrechtlichen CD-ROM der Neuen Bundesländer

Die Landesgesetzgebung von Sachsen und Thüringen

Ulrich Müller

Das Projekt

Für die vorliegende CD-ROM-Applikation bestand von vornherein die Zielstellung, mit einem Leistungsangebot aufzuwarten, das über das Standardniveau hinausgeht und auch anspruchsvollere Nutzererwartungen bedient. Das vorliegende Gesetzespaket sollte so dargeboten werden, daß weitestgehend die Vorzüge von strukturierten Datenbanken und hierarchisch gegliederten Volltext-Aufbereitungen miteinander verbunden werden. Was lag also näher, als sich schon frühzeitig mit der HyperText-Materie auseinanderzusetzen. Die geänderten Wirtschaftsstrukturen hier im Osten taten ein Übriges, das Nachdenken über neue Leistungsprofile zu beschleunigen und sich modernsten Informationstechniken zuzuwenden.

Es hat sich aber gezeigt, daß es nicht genügt, sich als spezialisierter Dienstleister in Vorbereitung solcher Applikationen umfassend theoretisch mit der Materie HyperText an sich auseinanderzusetzen. Die Zeitschrift jur-pc hat solche vorbereitenden Gedankenkonstruktionen schon im vergangenen Jahr durch die Bereitstellung ihres Podiums hervorragend unterstützt.

Zusammenarbeit mit den Verlagen

Bei der praktischen Umsetzung kamen zusätzlich noch Probleme der wirtschaftlichen Gestaltung des Arbeitsprozesses, der redaktionellen Vorbereitung und vor allem der effektiven Zusammenarbeit mit den Verlagen hinzu, die in ihrer Tragweite für das Projekt nicht zu unterschätzen waren.

I. HyperText vom Fließband

Einstieg der Verlage in das elektronische Publizieren

Die erste landesrechtliche CD-ROM der neuen Bundesländer ist weitgehend ein Eigenprodukt von Firmen dieser Region.

Für die Verlage in Dresden Sächsisches Druck- und Verlagshaus GmbH und in Erfurt Verlag und Druckerei Fortschritt Erfurt GmbH stellt es den Einstieg in dieses Gebiet des elektronischen Publizierens dar. Für die Firma in Ilmenau waren es die ersten praktischen Erfahrungen bei der Aufbereitung eines Datenkomplexes für ein HyperText-Produkt. Nun sind HyperText-Anwendungen zwar vielfach im Gespräch, aber im gesamtdeutschen Rahmen gibt es noch relativ wenige spezialisierte Unternehmen, die derartige Volltextaufbereitungen für eine CD-ROM-Applikation anbieten.

Statistik der Datenaufbereitung

Um Charakter und Umfang der Aufgabe zu verdeutlichen, sollten einige anschauliche, statistische Angaben vorangestellt werden:

So wurde in nur 10 Wochen eine Textmenge von über 10.000 Seiten und fast 1000 Grafikinformatoren für die CD-ROM-Applikation gestaltet, strukturiert und an die Retrieval-Software angepaßt. Der gesamte Text mußte mit ca. 85.000 Zusatzmarkierungen versehen werden, um die HyperText-Arbeitsweise zu ermöglichen.

Als Retrievalsoftware wurde CD-Answer HyperText der Firma Dataware Technologies verwendet. Das Strukturierungskonzept und die Informationsaufbereitung wurden von einem Systemhaus in Ilmenau auf der Basis eigener logistischer Studien erarbeitet und umgesetzt.

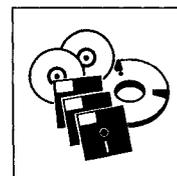
Dieses unabhängige Leistungssegment der Vorbereitung von CD-ROM-Applikationen wurde damit erstmals kommerziell angeboten und findet offensichtlich auch bei anderen Informationsanbietern und Verlagen zunehmend Interesse.

Erfahrungen

Ulrich Müller hat als Themenleiter in einem Systemhaus in Ilmenau die erste CD-ROM der neuen Bundesländer entwickelt und im Juli 1993 die Firma ImDat mitgegründet, die vorwiegend im Vorfeld von CD-ROM-Applikationen wirksam sein wird.

1. Dem Informationspaket ist grundsätzlich ein Navigationskonzept beizufügen.

Die Navigationsangebote im HyperText-Bereich ergeben sich nicht automatisch aus der Leistungsfähigkeit der Software oder dem redaktionellen Aufbereitungsstand, beides muß aneinander angepaßt, vorbereitet und abgestimmt werden. Das "Geheimnis" besteht in der strukturellen und logistischen Aufarbeitung des Textstranges. Anders als in einer festgefü-



ten Datenbankstruktur gibt es im fortlaufenden Textstrang keine Einschränkungen der Zuordnung; es steht vielmehr die volle Variationsbreite der Textgliederung, des hierarchischen Aufbaus und der gegenseitigen Verkopplung zur Verfügung.

Aus diesen Möglichkeiten läßt sich noch wesentlich mehr machen als nur die Navigation über ein Inhaltsverzeichnis und die Suchanfrage mit Unterstützung von Indexdateien. Eine quasi mundgerechte Aufarbeitung über spezielle Themen oder einen Handlungsablauf unter Verwendung der angebotenen Informationen sollten künftig schon zum Service eines komfortablen Komplett-Paketes gehören. Dieser Denkansatz geht davon aus, daß viele Nutzer die Information nicht zum Selbstzweck benötigen, sondern nur für eine ganz spezifische Problemlösung (Vorschriften zur Abfallbeseitigung, Bauanträge usw.). – Folglich kann eine Vielzahl typischer Anfragen bereits als Koppelpaket aufbereitet sein mit

- Ausgangsinformationen,
- Behördenwegen,
- Gebühren und Terminen,
- Beachtung weiterer Vorschriften,
- Formularen usw.

Der eigentliche Text steht dann unmittelbar für den direkten Zugriff mit weiteren zusätzlichen Navigationsangeboten zur Verfügung und ist andererseits nur Hintergrundinformation für Arbeitsalgorithmen. Die Verknüpfung erfolgt über Querverweis-Links.

Querverweise sind dann einerseits Bestandteil des Textes, um schnell zur Verweisposition zu gelangen, andererseits aber auch Hilfsmittel, um die relevante Hintergrundinformation anzufordern.

Die Erfahrung zeigt aber vor allem, daß es nicht genügt, solche Arbeitsweisen nur aufzuzeigen; das Angebot muß von den Redaktionen angenommen und umgesetzt werden. In der Regel werden im Ergebnis dieser Überlegungen dem ursprünglichen Informationspaket zusätzliche Navigationshilfen hinzugefügt.

2. Die Belegung des Textes mit Strukturmarkierungen setzt eine fundierte konzeptionelle Vorarbeit voraus.

Die Strukturierung des Textes ist – auch wenn es gelegentlich so scheint – keinesfalls nur die Abarbeitung rein formaler Algorithmen. Tatsächlich wird dadurch ein ganz konkretes Logistikkonzept praktisch umgesetzt. Der Gesamttext wird durch folgende Markierungen eingeteilt:

- Kennungs-Tags zum Hervorheben von Textbereichen
- Hierarchie-Tags zur hierarchischen Gliederung des Textes
- Klammer-Tags für eine Zusammenfassung innerhalb der gleichen Ebenen oder auch ebenübergreifend

Dieser Einteilung liegen entsprechende Fragestellungen zugrunde:

- Welche Textbereiche sind besonders zu kennzeichnen (für ganz unterschiedliche spätere Funktionen)?
- Welche Informationshierarchie wird für das Gesamtpaket aufgebaut?
- Welche Abfragemöglichkeiten sollen angeboten werden?
- Über welche Teilgebiete des Gesamttextes ist unabhängig von seiner Hierarchie eine Verknüpfungsklammer zu schaffen?

Übliche Strukturierungskennzeichnungen für das Druckformat oder für Layout-Informationen können dabei selbstverständlich als Markierungshilfe mit genutzt werden. Eine identische Übernahme ist aber nicht zu erwarten. Das ergibt sich allein schon aus der Logik, daß eine Druckaufbereitung eben anderen Gestaltungsgesetzen folgen muß als die HyperText-Präsentation.

3. Das Strukturierungskonzept des Textes ist weitgehend unabhängig von der verwendeten Retrievalsoftware, muß aber sowohl bei der Syntax der Markierungen als auch der Positionsvorschriften jeweils angepaßt werden.

Selbstverständlich sind auch funktionale Besonderheiten zu berücksichtigen, so daß sich eine Wechselbeziehung zwischen der Leistungsfähigkeit der Retrievalsoftware, dem Aufbau des Informationspaketes und der konkreten Ausführung des Strukturierungsplanes ergibt.

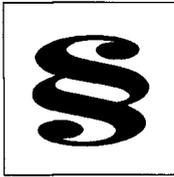
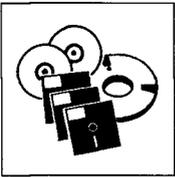
Typische Benutzeranfragen vorbereiten

Querverweise

Einteilungskriterien

Layout-Information als Markierungshilfe

Retrievalsoftware-unabhängiges Datenmodell



"Potentielle" Hypertextziele

4. Für die Grundstrukturierung des Textes, d. h. für die Markierung der Hierarchieebenen und für die Kennzeichnung exponierter Textbereiche (Titel, Datum, usw.) lassen sich Werkzeuge für eine automatisierte Arbeitsweise schaffen, Voraussetzung ist nur die gute und eindeutige Erkennbarkeit der zu markierenden Textstellen.

5. Querverweis-Links lassen sich grundsätzlich nur manuell einfügen

Eine Vereinfachung wird nur dann möglich, wenn die Retrievalsoftware eine kombinierte Bezeichnung der Strukturebenen und der Zielmarkierungen des Querverweises zuläßt. Alle Strukturmarkierungen, auf die ein Verweis möglich ist, erhalten dann bereits im automatisierten Programmdurchlauf eine zusätzliche Zielkennzeichnung, die natürlich eindeutig sein muß. Die individuellen Merkmale sind z.B. die Kurzbezeichnung des Gesetzes, die Nr. des Abschnitts, der Paragraph usw. Für jeden dieser Strukturbereiche wird dann grundsätzlich ein Zielpunkt definiert, unabhängig davon, ob ein Verweis auf diese Position existiert. Die vollständige Zielpunktdefinition setzt sich dann aus der Kurzbezeichnung des Dokuments und den hierarchischen Gliederungsmerkmalen zusammen.

Beispiel einer solchen Markierung:

<GREFID=SaVerf-0.16.2>

als Zielmarke für: Sächs. Verfassung Artikel 16, Absatz 2.

Der zusätzliche Aufwand lohnt sich aus zwei Gründen:

- Bei der verbalen Beschreibung der Querverweiszielposition im Text sind i. d. R. eben diese individuellen Gliederungs-Merkmale benannt, so daß sich die Formulierung der Zielmarke ohne weitere Kontrolle logisch daraus ergibt.
- Bei periodischen Updates müssen die vorangegangenen älteren Texte nicht erneut bearbeitet werden.

6. Die Strukturierung großer Textmengen erfordert eine Ablauforganisation, wie sie für Produktionsprozesse typisch ist.

Ablauforganisation

Nach diesem Prinzip sind

- Arbeitsschritte zu definieren,
- Verantwortung und Übergabeformate festzulegen,
- Produktionsfortschrittskontrolle und Qualitätskontrolle zu organisieren,
- und parallele Arbeitsprozesse zu koordinieren.

Qualitätssicherung

Die wichtigsten Hilfsmittel dieses Produktionsprozesses sind gute Prüfprogramme in den verschiedensten Bearbeitungsstufen, die das Einlesen der Daten in die Retrievalsoftware ohne Fehler ermöglichen und vorab bereits die wichtigsten funktionalen Arbeitsschritte kontrollieren oder bereits schon simulieren.

Die Wirkung der rein formal erscheinenden Textmarkierungen sollten dem Operator für diesen Arbeitsgang plastisch vorgeführt werden können. Das ist notwendig für Hintergrundinformationen und schafft zudem einen notwendigen Spielraum eigener Entscheidungsfindungen in Zweifelsfällen.

7. Alle geplanten funktionalen Merkmale oder besonderen Effekte der Arbeitsweise bedürfen unbedingt einer Bestätigung durch eine Probeinstallation in einem Demonstrationsdatenbestand.

Diese Überlegungen sollen letztlich dazu führen, den Vorbereitungsaufwand für HyperText-CD-ROM-Applikationen auf ein Minimum zu reduzieren. Das ist allein im Dienstleistungsbereich natürlich nicht zu schaffen. Es ist völlig klar, daß diese Komponente noch kein vollständiges Bild ergeben kann.

II. HyperText als geschlossenes Verlagskonzept

Die Verlegerperspektive

Um die Problematik abzurunden und alle Aspekte einer effektiven Arbeit auszuloten, ist es notwendig, gerade auch die verlegerische Seite mit zu betrachten.

Bisherige Überlegungen gehen oft noch davon aus, daß die CD-ROM ein Abfallprodukt der Verlagsarbeit darstellt und folglich alle weitergehenden Aktivitäten erst nach dem Druck einsetzen (Abb. 1, auf der Folgeseite).

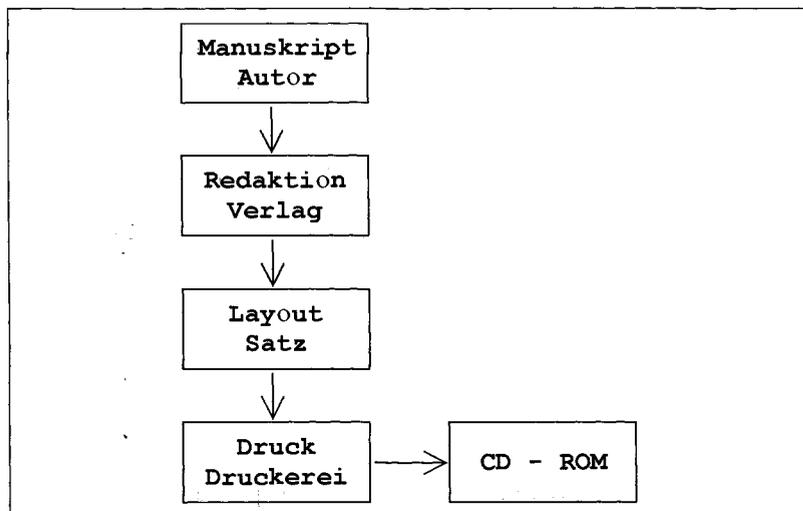
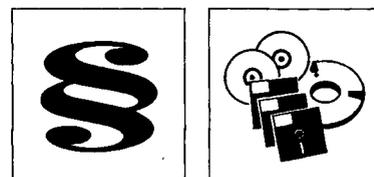


Abb. 1:
Verbreitete (traditionelle) Ablauforganisation

Derart drastische Fehlentscheidungen sind heute kaum noch anzutreffen. Allein die Aufwandskosten zwingen zu Überlegungen, zumindest die nochmalige Übernahme auf Datenträger einzusparen. Der Rohtext läßt sich mit vertretbarem Aufwand unmittelbar den Satzbändern entnehmen.

Trotzdem liegt dann als Ergebnis nur die identische Nachbildung der Druckversion vor. Das kann für ein elektronisches Informationsangebot nur die Ausgangsform sein. Für HyperText-Applikationen gilt das um so mehr.

Elektronische Produkte bedürfen einer datentechnischen und redaktionellen Vorarbeit, die die gleichen Etappen wie das Printprodukt durchläuft, aber speziell auf diese Medien zugeschnitten ist.

Die dargestellten Schritte bleiben schon etwa gleich, nur ihr Inhalt ändert sich wesentlich (Abb. 2).

“Text-Recycling”

*Medienspezifische
Materialaufbereitung*

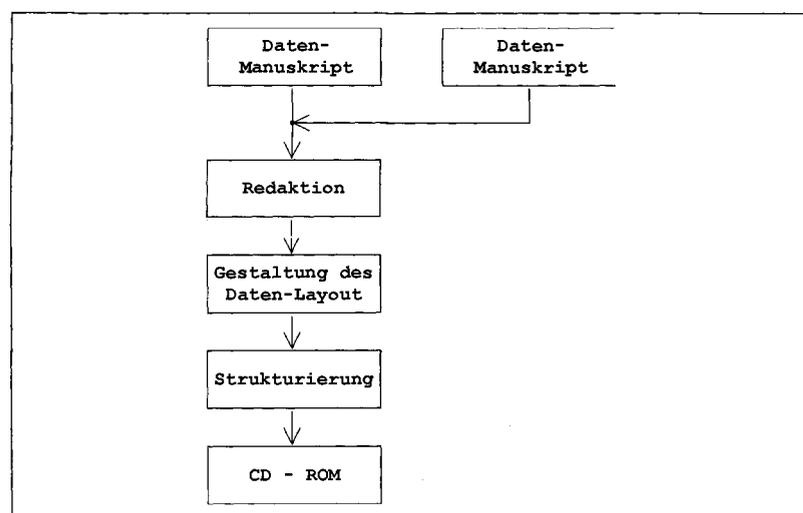


Abb. 2:
EP orientierte
Ablauforganisation mit einer
Erweiterung des
Angebots

Vom Autor wird nicht nur die Bereitstellung des Textes auf Datenträger, sondern selbstverständlich auch die Einarbeitung der Verweise, der Grafikinformatoren und die notwendigen Verknüpfungen erwartet. Alle Zusatzinformationen, wie sie für HyperText-Angebote typisch sind, sind natürlich von Anfang an mit aufzubereiten.

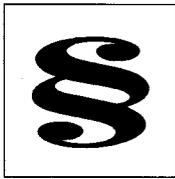
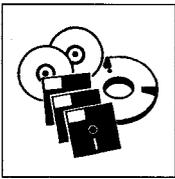
Von den Redaktionen müssen die notwendigen Überlegungen ausgehen, um das Produkt mit einem Leistungszuwachs zu versehen, der sich aus der Informationsverarbeitung im Hintergrund ergibt. Daß deren Verarbeitungsbesonderheiten zu berücksichtigen sind, ist selbstverständlich.

Viel stärker trägt die Redaktion die Verantwortung für die Zusammenstellung des Informationspakets. Alle zusätzlichen Navigationshilfen müssen bereits von hier aus entschieden und vorbereitet werden.

Der Text selbst muß evtl. sogar völlig andersartig aufgebaut werden. Ein Lexikon z. B. kann abweichend zur alphabetischen Begriffsanordnung durchaus auch eine hierarchisch gegliederte Sachgebietszuordnung erhalten. Selbstverständlich ist dann auch eine Mehr-

Die Arbeit der Autoren

Die Aufgabe der Redaktionen



Datenlayout

fachzuordnung möglich. Der Vorteil besteht darin, daß der Nutzer auch zu Eintragungen geführt wird, die mit dem eigentlichen Suchbegriff gar nicht belegt sind.

Das Datenlayout orientiert sich gegenüber dem gedruckten Werk an anderen Grundsätzen und richtet sich vorwiegend nach dem Leistungsprofil der Software (graphische oder zeichenorientierte Oberfläche) und den Darstellungsmöglichkeiten des Rechners.

Dazu gehören Besonderheiten wie

- keine Wechsel von Schriftart und Schriftgröße,
- sparsamer Umgang mit Leerzeilen,
- kein spaltenweiser Satz,
- Einschränkungen beim Einrücken des Textes oder
- Besonderheiten bei der Darstellung des Zwischentitels.

Die Strukturierung des Textes ist bei dieser Rangfolge tatsächlich nur ein letzter formaler Akt mit einem hohen Automatisierungsanteil. Das betrifft dann natürlich auch den Querverweis, wenn der Autor als der eigentliche Urheber des Verweises bereits entsprechende Markierungen gesetzt hat.

Die anfangs viel beschworene Informationslogistik ist also ebenfalls Teil der redaktionellen Vorarbeit und damit eigentlich Verlagsarbeit.

"Electronic publishing"-Produkte sind andersartig und von ihrem Angebotsprofil gründlicher vorzubereiten als Printproduktionen.

Der Kostenaufwand läßt sich ohne Qualitätseinbuße nur dadurch minimieren, daß der Ablaufprozeß innerhalb des Verlages auf dieses Produkt zugeschnitten wird.

Andererseits sind zwei unabhängige Ablaufstränge von vornherein nicht vertretbar, so daß letztlich eine möglichst weitgehende Zusammenführung beider Komplexe das Ziel der Überlegungen sein muß.

Die Lösung liegt in einer SGML-Strukturierung des Datenpaketes (SGML = Standard Generalized Markup Language), d. h. die Textverarbeitung erfolgt durchgängig vom Autor bis zur Druckerei rechnergestützt. Alle Texte enthalten nur Strukturmarkierungen. Die Layout-Entscheidungen werden erst mit der Festlegung über die Ausgabe ganz individuell getroffen – je nachdem, ob ein Printprodukt oder eine CD-ROM entstehen soll.

Darüber ist bereits umfassend berichtet worden (vgl. Müller, Neue Wege für Hypertext-Anwendungen auf CD-ROM, jur-pc 9/92, 1758 ff.). Hier sollte lediglich die Einbettung in das Verlagskonzept herausgearbeitet werden. Natürlich wandelt sich dann auch der Charakter eines Dienstleistungsunternehmens von der Auftragsrealisierung hin zu beratenden Komponenten.

Die Retrievalsoftware setzt bereits heute ganz bewußt auf einer SGML-Strukturierung auf, so daß von dort her alle Voraussetzungen für die beschriebene rationelle Arbeitsweise gegeben sind.

*SGML-Strukturierung und
medienspezifische Exportfilter*