

Die Einbindung juristischer Informationen in Volltext-Retrievalsysteme

Ulrich Müller

Bei juristischen Volltexten hierarchisch gegliederte Aufbauprinzipien auf dem Vormarsch

Juristische Informationen werden seit geraumer Zeit im vollen Textumfang für eine rechnergestützte Auswertung angeboten. Nachdem bisher vorwiegend datenbankähnliche Aufbereitungen dominierten, beginnen sich zunehmend auch hierarchisch gegliederte Aufbauprinzipien durchzusetzen. Dabei übernehmen CD-ROM-Angebote mit einer zunehmend breiter werdenden Palette von Retrievalsystemen eine gewisse Vorreiterrolle. In diesem Beitrag soll es um praktische Probleme der Vorbereitung und der Handhabung derartiger juristischer Texte gehen, um sie in hierarchisch gegliederte Volltext-Retrievalsysteme einzubinden.

Auffallende Ähnlichkeit bei der Lösung der Hypertextfunktionen

1. Entwicklungstendenzen der Volltext-Retrieval

Zunächst fällt vor allem auf, daß diese Marktlücke "Volltext-Retrieval" offenbar von einer größeren Anzahl renommierter Softwarehäuser entdeckt wurde. Ohne auf Produktnamen eingehen zu wollen, ist doch unverkennbar, daß die typischen Hypertextfunktionen zunehmend gleichartig oder zumindest sehr ähnlich gelöst werden. Trotzdem bleiben noch genügend produkttypische Eigenheiten erhalten, und die Tendenz läßt erkennen, daß die Vielfalt der Darstellungsvarianten unter Beibehaltung bestimmter Grundprinzipien eher zunehmen wird.

Trendbestimmend: buchähnliche Arbeitsweise

Hierarchische Gliederung

Trendbestimmend sind Inhaltsverzeichnisse mit wahlweisen Öffnungs- und Schließmöglichkeiten für die Ebenen dieser Hierarchiestruktur, so daß eine quasi buchähnliche Arbeitsweise angeboten wird. Eher als Ausnahme gibt es noch pull-down orientierte Vorgehensweisen, bei denen zum Tableau der nächst niedrigeren Ebene gesprungen wird. Soweit es sich dann wieder um eine Übersicht mit einem Angebot zum Öffnen der folgenden Ebene handelt, kommt natürlich der gleiche Effekt zustande, aber die Übersicht zum Gesamtbestand – zu den verbleibenden Eintragungen der Vorebene – geht weitgehend verloren.

unbegrenzte Tiefenstaffelung über das Inhaltsverzeichnis

Tiefenstaffelung

Über das Inhaltsverzeichnis ist in der Regel eine unbegrenzte Tiefenstaffelung möglich. Das gilt auch dann, wenn die hierarchische Gliederung über Verweistableaus organisiert wird. Problematisch wird es erst, wenn als letzter Zielverweis ein Textblock angeboten wird, der nur in diesen Blockgrenzen über den Bildschirm scrollen kann. Für die Darstellung eines Gesetzestextes muß man sich dann nämlich entscheiden, ob man das Gesamtdokument anbietet und auf die Gliederung in der §-Ebene verzichtet oder eben die einzelnen "§§" blockweise als Text aufbereitet.

variantenreichere Suchanfragen

Suchanfrage

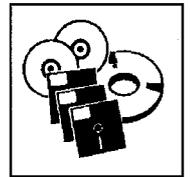
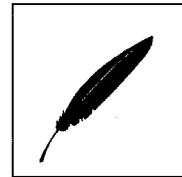
Die meisten Systeme gehen davon aus, daß der Volltext im gesamten Textstrang auf Suchanfragen zu analysieren ist, die in einer einzigen Suchzeile eingegeben werden. Auf Wunsch kann der Textumfang auch eingegrenzt werden. Die Anfrage richtet sich dann auf eine Hierarchieebene, einen ausgewählten Verzweigungsblock usw. Über diese Selektion anhand der Baumstruktur des Textes hinaus können bei einigen Systemen auch vorab markierte Textbereiche separat durchsucht werden, so daß sich die Suchanfrage wesentlich variantenreicher auf mögliche interessierende Textblöcke orientieren läßt.

Die Verknüpfung dieser Varianten in einer Suchmaske – so wie der Nutzer Datenbankarbeit gewohnt ist – ist bisher eher noch die Ausnahme. Die Forderung in diese Richtung wird aber zwingend, wenn mit Textblöcken gearbeitet wird, die allein für sich ein juristisches Dokument darstellen und eine Vielzahl bibliografischer Zusatzinformationen enthalten.

Markierungssprache

Jede Funktionalität bedarf einer entsprechenden Kennzeichnung innerhalb des Textes. Die dafür verwendete Markierungssprache ist derzeit noch überraschend individuell. Eine durchgängige Orientierung auf die Verwendung von SGML-Auszeichnungen ist noch nicht zu erkennen, dürfte aber trotzdem nicht lange auf sich warten lassen. Allein ihr zunehmender Einsatz in Redaktionssystemen und als layoutneutrale Textmarkierung im DTP-Bereich zwingen letztlich zu Entscheidungen in diese Richtung.

Ulrich Müller ist Geschäftsführer der Firma IlmDat GmbH in Ilmenau/Thüringen. Er beschäftigt sich seit ca. drei Jahren mit der Problematik der Volltextaufbereitung für elektronisches Publizieren.



Anfügen

Darunter soll der Service verstanden werden, nachträglich aufbereiteten Text anzufügen oder zusätzliche Textblöcke des Nutzers einzubinden und sie in den wichtigsten Funktionen gemeinsam zu behandeln. Auch dafür gibt es nur erste Ansätze.

Diese hier nur angedeutete, bereits vorhandene und zu erwartende Funktionalität der Retrievalsysteme muß man also im Auge haben, wenn man juristische Texte für elektronisches Publizieren vorbereiten will.

2. Zur Beachtung für den Einstieg ins elektronische Publizieren

Im Endausbau wird ein Verlag selbstverständlich von einem in sich geschlossenen Redaktionssystem ausgehen. D.h. der Text wird, beginnend vom Autor über die Redaktion bis zur Layout-Gestaltung, durchgängig im SGML-Format verwaltet. Dafür stehen seit langem die notwendigen Werkzeuge, wie Editoren, Datenbanken, Parser, Konvertierer und Aufbauhilfen für die Dokumenttypdefinition (DTD) zur Verfügung. Die Ausgabe der Informationen erfolgt dann tatsächlich wahlweise als Druckexemplar oder auf ein elektronisches Medium. Trotzdem gibt es Besonderheiten, die diese Produktkategorie vom Printprodukt unterscheiden:

- Der Text muß gesonderte Markierungen aufweisen, wenn im elektronischen Produkt Suchfunktionen angeboten werden, die sich speziell auf diese Textstellen richten.
- Die vorgesehene Indexierung des Textes muß ebenfalls durch entsprechende Markierungen intelligent unterstützt werden.
- Querverweise sind ohnehin noch zusätzlich anzubringen.

Es ist also Nacharbeit für die Auszeichnung des Textes notwendig, auch wenn gemeinhin von einer "Knopfdruck"-Option gesprochen wird. Der Einstieg ins elektronische Publizieren vollzieht sich aber ohnehin anders als über diese Komplettlösung. Um ältere Bestände aufzubereiten, die Technologie und natürlich auch den Markt zu erproben, wird meist der unabhängige Seitenweg benutzt, und man sucht nach einfacheren Werkzeugen für die Textvorbereitung zum EP: Textmakro-Programme z. B. um SGML-Grundstrukturen rechnergestützt zu erstellen (auch wenn dabei noch sehr viel manuell eingegriffen werden muß).

Ziel: geschlossenes Redaktionssystem, wahlweise Ausgabe im Druck oder elektronisch

Es bleiben aber Besonderheiten für das elektronische Medium ...

... die umfangreiche Nacharbeit verlangen.

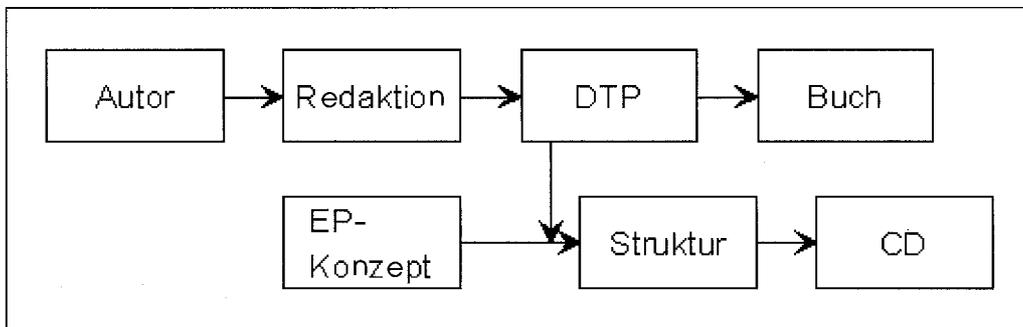


Abb. 1: Übersichtsdarstellung zur Veranschaulichung des EP-Einstiegs

Ein derartiger "Seiteneinstieg" ins EP hat eigentlich nur 2 Besonderheiten, für die entsprechende Voraussetzungen geschaffen werden müssen.

a) Es sind automatische Markierungshilfen notwendig, die zumindest die hierarchische Struktur mit SGML-Tags belegen, möglichst aber auch weitere herausgehobene Textbereiche kennzeichnen.

Als automatische Tagginghilfen bieten sich an:

Konvertierungsprogramme ———> SGML

- aus einer Textverarbeitung durch Umsetzung
 - der Kennzeichnung der hierarchischen Gliederungsebenen
 - der Layout-Kennzeichnung der Titelzeilen

Taggingprogramme ———> SGML

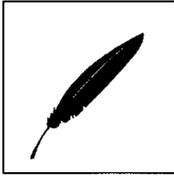
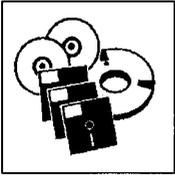
- aus ASCII
 - durch Identifikation typischer Zeichenketten und Umgebungserkennung

b) Das gesamte Strukturpaket ist automatisch zu kontrollieren und zu analysieren:

- Kontrollprogramme dienen der kompletten Überprüfung
 - auf die Einhaltung der Hierarchielogik
 - auf richtige Schreibweise der tag's,
 - hinsichtlich der Querverweisuordnung
 - auf Vorhandensein der Grafikinformatoren

Kontrollprogramme

Darüber hinaus sind weitere Überlegungen im logistischen Bereich anzustellen, um nach der prinzipiellen Einstiegsentscheidung tatsächlich auch ein gutes Produkt zustande zu bringen.



3. Die Modifizierung der SGML-Struktur

Jede Funktionalität im Ergebnis wird durch einen entsprechenden Markierungsaufbau hervorgebracht und muß vorausschauend durch eine Markierungslogistik vorgedacht sein. Allein für den Hierarchieaufbau juristischer Informationen sind folgende funktionale Textbausteine zu beachten:

Gestaltung und Darbietung des Informationskomplexes

		Beispiel
Sachkomplex	Zusammengehörige Gruppe von Dokumenten	Verwaltungsrecht
Dokument	In sich geschlossenes Regelwerk mit eigenen bibliografischen Informationen	Gesetz
Indexblock	Baustein mit dem kleinsten geschlossenen Sinnzusammenhang	Paragraph
unterstes Zielelement	kleinster Zitierbaustein	Absatz

Auf dieser Grundüberlegungen aufbauend ist der Informationskomplex zu gestalten und darzubieten. D.h. das Dokument als geschlossenes Regelwerk sollte als fortlaufender Textstrang über den Schirm scrollen, der "§" muß als Textblock indexiert werden, Zielverweise sind mindestens noch bis zum Absatz anzubringen und inhaltlich zusammengehörige Dokumente sind über den Sachkomplex zu klammern. Der "Knopfdruck"-Version des EP muß spätestens hier große Skepsis entgegengebracht werden.

Weitere Überlegungen zeigen ...

Darüber hinaus zwingen weitere Überlegungen, über einen modifizierten Hierarchieaufbau gegenüber dem Printprodukt nachzudenken. Das sind vor allem kritische Bedenken hinsichtlich

- der Leistungsfähigkeit der Retrievalsoftware, inwieweit sie diesen Anforderungen Rechnung tragen kann.

- der tatsächlichen Belegung der Gliederungsebenen. Für das Printprodukt ist es uninteressant, ob in einigen Textbereichen bis zur letzten Strukturebene tiefengestaffelt ist oder ob mehrere Zwischenebenen im Strukturaufbau fehlen. Am Bildschirm dagegen ruft ein mehrfaches leeres Aufblättern des Inhaltsverzeichnisses schon einige Nervosität hervor..
- der Nutzung der spezifischen Suchmöglichkeiten der Retrievalsoftware. In Rechtssprechungsdatenbanken z. B. ergibt der Textblock "Schlagworte" eben nur dann einen Sinn, wenn dafür eine eigenständige Suche angeboten wird.

... nur ausnahmsweise 1:1 vom Printprodukt zur elektronischen Publikation

Diese Überlegungen bestätigt nochmals, daß die Standard-Markierungen für das EP-Produkt ganz speziell zu überarbeiten und anzupassen sind. Eine 1:1-Übergabe von Print-Produkten zur EP-Applikation ist nur in Ausnahmefällen vorstellbar.

Die notwendigen Eingriffe in den Hierarchieaufbau und die Tag-Markierung sind natürlich wieder Fehlerquellen und bedürfen einer ausgefeilten Kontroll-Analytik oder eines speziellen SGML-Editors.

4. Dokumentenmanagement

Weitere Besonderheiten des EP

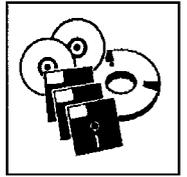
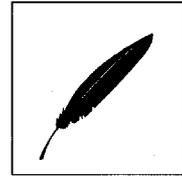
Nach diesem Ausflug in die Problematik der Nachstrukturierung soll der Ersteinstieg in das EP an einem besonderen Detail näher beleuchtet werden: Alle bisherigen Betrachtungen laufen darauf hinaus, daß strukturierte Einzeldokumente vorhanden sind und eine Retrievalsoftware benötigt wird, die

- SGML-Markierungen interpretiert,
- ein Inhaltsverzeichnis mit Öffnungs- und Schließoptionen besitzt,
- eine Zielverweiszuoordnung bis zum kleinsten Zitierbaustein ermöglicht und
- den Aufbau von Suchmasken unterstützt, deren Suchzeile auf jeweils unterschiedliche Textpassagen gerichtet sind und die dann gegenseitige Verknüpfungen auf der Hierarchieebene des Dokuments zulassen, so daß vorwiegend das Dokument an sich aufgrund der bibliografischen Angaben und evtl. weiterer Selektionsvorgaben im Volltext gefiltert werden kann.

redaktionelle Vorüberlegungen

Derart aufgebaute Volltext-Informationssysteme zwingen zu gründlichen redaktionellen Vorüberlegungen:

- zur Reihenfolge des physischen Aufbaus der Dokumente,
- der Zuordnung jedes Einzeldokuments zu Sachkomplexen mit einem eigenständigen Hierarchiebaum und
- der Formulierung dieser Zwischenüberschriften.



Oftmals sind darüber hinaus auch Kurzformen des Dokumententitels zu formulieren, wenn das für den Aufbau des Inhaltsverzeichnisses erforderlich sein sollte. Der Arbeitsgang in seiner ausführlichen Form sieht dann etwa so aus:

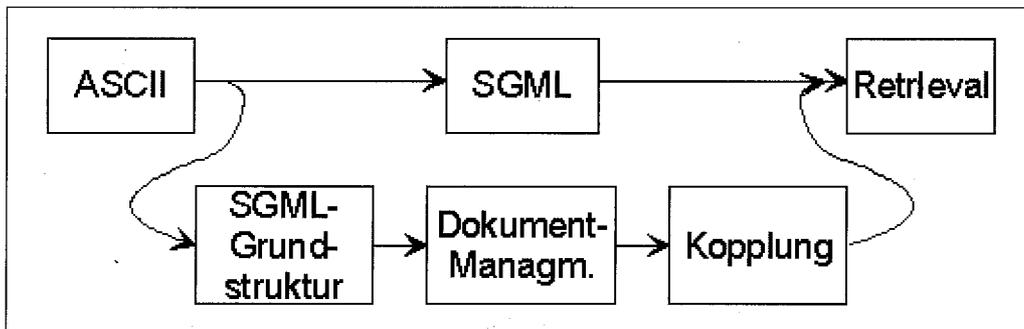


Abb. 2:
Bestandteile des
SGML-Tagging

Dieses Dokumentenmanagement ist das Ergebnis der eigentlichen redaktionellen Arbeit, ist eigenständig und hat mit dem eigentlichen Tagging wenig zu tun. Tatsächlich läßt sich dieser Arbeitsgang auch völlig separat bewältigen und wird letztlich nur über eine programmtechnische Verknüpfung in den Arbeitsablauf der SGML-Markierung eingefügt. Das Dokumentenmanagement ist die Grundlage für den Aufbau des Textstranges, die gegenseitige Anordnung und die hierarchische Verknüpfung der Dokumente vor dem Einlesen in die Retrievalsoftware.

Der Redakteur legt also vorab in einer Zuordnungsliste den Aufbau, die Gliederungsstruktur mit allen Zwischenüberschriften, die Reihenfolge der Einzelbeiträge und die Kurztitel dieser Dokumente für das Elektronikprodukt fest. Die Koppelliste wird damit zum eigentlichen Produktmanager. Das eröffnet wieder neuen Denkspielraum für die Variantenbildung derartiger zusammengestellter EP-Produkte.

Zumindest wird erneut die Erkenntnis bestätigt, daß bei einem Volltext-Retrievalsystem die eigentliche logistische und redaktionelle Arbeit vor dem Einlesen in die Retrievalsoftware zu bewältigen ist.

5. Zusammenfassung

Die elektronische Publikationsvariante eines Volltextsystems ist unter Berücksichtigung der vorgesehenen Retrievalsoftware und auf der Grundlage eines speziell angepassten Publikationskonzeptes gesondert vorzubereiten. Zu diesen Vorbereitungsarbeiten zählt in erster Linie der informationslogistische Teil:

- das Herausarbeiten der funktionalen Textbausteine- und davon abgeleitet,
- welche besondere Beachtung oder Behandlung erfahren diese Textblöcke, und schließlich die Konsequenz für
- die Modifizierung der vorhandenen Tag-Struktur oder das Tagging selbst.

Im Falle der Neuverkettung juristischer Dokumente ist das Management ihres Hierarchieaufbaus, das den eigentlichen redaktionellen Teil ausmacht und dem Aufbau des Textstranges dient, von besonderer Bedeutung.