

Kosten-Nutzen Überlegungen bei XML-Projekten (BGBL2000)

Franz Reinisch

*LIS Reinisch OEG
A-7441, Steinbach 49
office@lis-oeg.com
www.lis-oeg.com*

Schlagworte: Bundesgesetzblatt, BGBl, XML, Archiv, CD-ROM, Index, MS-Word, HTML, Konvertierung, Druckvorlagen, Info-Provider, Datenbank

Abstract: Kann ein modernes CD-ROM- oder Online-Archiv mit XML-Technologie unter dem Gesichtspunkt vorhandener Methoden und Softwareprodukte erstellt werden? Am Beispiel der BGBl-CD (ab 2000) wird beispielhaft eine derartige Umsetzung gezeigt. Die detaillierte Aufbereitung der Dokumente für den Druck konnten zur „Zerteilung“ des Dokumentes in XML-Elemente genutzt werden. Die Vorlagen-Elemente: Titel, Kurztitel, Verweise, Unterzeichner und Ausgabedatum konnten so für eine exakte Suche zur Verfügung gestellt werden. Die eingesetzten Produkte sind beispielhaft: MS-WORD als Autorensystem, MS-Internet-Explorer als Anzeigesystem, und der Info-Provider als Datenbank und Indexierungsmaschine.



1. Komplexität des Projektes

Wie komplex zeigt sich das Projekt oder die Fragestellung? Ist ausreichend Budget vorhanden, um überhaupt an eine XML-Lösung denken zu können? In der Vergangenheit wurden viele SGML-Projekte gar nicht erst gestartet, weil die Kosten nicht abschätzbar oder tragbar waren. Durch den „Siegesszug“ von XML ist der Markt mit Shareware- und Freewarekomponenten geradezu überschüttet worden. Wer sucht, der findet. Ein alter Bibliothekars-Grundsatz bewahrheitet sich auch bei der Suche nach kostengünstigen und flexiblen Softwarekomponenten. Ein möglicher Lösungsweg wurde mit der BGBL-CD 2000¹ (bzw 2001) aufgezeigt.

2. Lebensdauer des Projektergebnisses

Wie lange muss das Projektergebnis „Dienst tun“? Ein Langzeitarchiv hat andere Anforderungen als eine Online-Tagezeitung. Ein System zur automatischen Einbringung von Änderungen, ein Kunsttext-Erzeugungssystem², ist komplexer als ein Archiv eines statischen „Papierabbildes“.

3. Warum XML?

Entscheidend für alle Organisationsprojekte ist die Frage: wozu? Kann diese Frage nicht befriedigend beantwortet werden, erübrigt sich eine weitere Beschäftigung mit dem Thema.

Im Falle der BGBL-CD sollte effektiver als bloß im Volltext recherchiert werden können. So wurde gefordert, dass in den Textbereichen:

Ausgegeben am: *bzw* Teil I, II, III
 Kurztitel
 Langtitel
 Verweise (zB: CELEX)
 Unterzeichner

gesucht werden kann. Wobei die Suche in Titeln *bzw* Verweisen (zB CELEX-Nummern) durchaus Sinn macht, die Suche nach dem Unter-

¹ BGBL2000/2001 CDs, erhältlich im Verlag Österreich.

² *Arnold-Moore, T.*, Automatic Generation of Amendment Legislation, in: ICAIL'97, 56-62.

zeichner wohl eher die Leistungsfähigkeit der XML-Technologie aufzeigen sollte. Als kostengünstige Anzeigesoftware sollte ein Internet-Browser verwendet werden.

4. Datenformat der Dokumente

Das Layout des BGBL (in der veröffentlichten Papierform) ist allgemein bekannt. Weniger bekannt ist, dass eine Unmenge von Formatvorlagen für die Drucklegung verwendet wird, die dem BGBL ein homogenes Aussehen gibt.

BUNDESGESETZ	
FÜR DIE REPUBLIK ÖSTERREICH	
Jahrgang 2000	Ausgegeben am 29. Dezember 2000
142. Bundesgesetz: Budgetbegleitgesetz 2001 (NR: GP XXI RV 311 AB 369 S.45.BR: 6250)	
Standard 000_Bg 000_Bg_1	
142. Bundesgesetz, mit dem das Rundfunkgesetz, das He buchgesetz, das Gesetz über Gesellschaften mit Gerichtsgebührengesetz, das Wohnungsgewer Einkommensteuergesetz 1988, das Körperschaft Umsatzsteuergesetz, das Umsatzsteuergesetz 1994, d Grundsteuergesetz 1955, das Bodenwertabgabengesetz, d	

BUNDESGESETZBLATT	
FÜR DIE REPUBLIK ÖSTERREICH	
Jahrgang 2000	Ausgegeben am 29. Dezember 2000
Abschnittswechsel (Fu 142. Bundesgesetz: Budgetbegleitgesetz 2001 (NR: GP XXI RV 311 AB 369 S. 45. BR: 6: 6240 S. 670.)	
Abschnittswechsel (Fu	
142. Bundesgesetz, mit dem das Rundfunkgesetz, das das Firmenbuchgesetz, das Gesetz über Gesellschaften Heftung, das Gerichtsgebührengesetz, das Wohnungsgew das Einkommensteuergesetz 1988, das Körperschaftsteu	

Die mehr als 100 Format-Vorlagen wurden analysiert und in Indexgruppen zusammengefasst. Mehr als sechs bis acht Suchfelder sind erfahrungsgemäß die Obergrenze beim Einsatz von Suchmasken.

5. Konvertierung und „enrichment“

Die vorhandenen Originaldokumente wurden mittels WORD2000 in XML/HTML umgewandelt. Die nicht notwendigen zusätzlichen Informationen (Styles, Rück-Konvertierinformationen usw) sollten entfernt werden (zB mit dem Microsoft Office HTML-Filter 2.0).

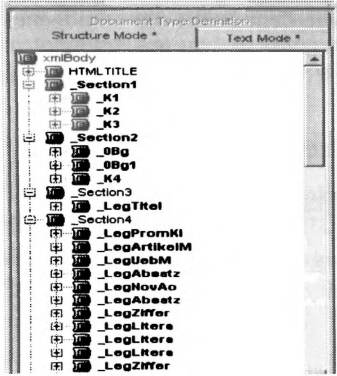
5.1. Vorlagen-Normierung

Leider konnten hier einige Standard-Funktionen nicht genutzt werden, weil die vorhandene Dokumentvorlage eben nicht auf XML-Funktionalität ausgerichtet war. Als Beispiel sei hier nur die Verwendung von Umlauten in Druckvorlagenbezeichnungen genannt. Auch <spitze

Klammern> und vorlaufende Ziffern (00Bg) machen den diversen Filtern Probleme.

5.2. Linkgenerierung

Anschließend wurde der HTML-Text „enriched“, dh um einige einfache Links erweitert. Hyperlinks zum Original-PDF-Dokument sowie zu den Ausschussberichten (AB) und Regierungsvorlagen (RV) wurden automatisch erzeugt. Damit kann der Benutzer das Original-Layout des Dokumentes einsehen. Im Original-PDF Dokument ist auch der gesamte Umfang des Dokumentes sichergestellt.



Das XML-Dokument bedarf zur Einbringung in die Datenbank keiner DTD³, es reicht dem Parser, dass das Dokument „valid“ ist. Die vom WORD-to-HTML Konverter zusätzlich erzeugten Sections stören bei der Indizierung nicht weiter. Dass keine DTD notwendig ist, erweist sich hier zwar als

Vorteil, ist aber für den Prozess der internen Normung äußerst schmerzlich. Analoges gilt bei der Festlegung von Dokumentenvorlagen.

6. Anzeige

Die Anzeige erfolgt mittels MS-Internet-Explorer (>5.5). Die Produktfestlegung erfolgte wegen der notwendigen XML-Komponenten zur schnellen Umwandlung von XML (aus der Datenbank) in HTML (am Bildschirm). Zur „best-try“ Darstellung der Dokumente ist ein ausgeprägtes XSL (zur Darstellung der XML-Elemente an der richtigen Stelle) und ein CSS (zur Darstellung der Schrifttypen und Fontgrößen) notwendig. Die Treffer-Markierung erfolgt als Script im XSL-File.

³ DTD. Document Type Definition, Beschreibung der Struktur eines Dokumentes.

BUNDESGESETZ

FÜR DIE REPUBLIK ÖSTER

Jahrgang 2001

Ausgegeben am 9. Jänner 2001

Teil I

**3. Bundesgesetz:Finanzausgleichsgesetz 2001 – FAG 2001 und Änderung des Finanzau
Zweckzuschussgesetzes 1989
(NR: GP XXI RV 379 und Zu 379 AB 405 S. 52. BR: 6289 AB 6282 S. 671.)**

**3. Bundesgesetz, mit dem der Finanzausgleich für die Jahre 2001 bis 2004 geregt
Bestimmungen getroffen werden (Finanzausgleichsgesetz 2001 – FAG 2001) und
Wohnbauförderungs-Zweckzuschussgesetz 1989 geändert werden**

Der Nationalrat hat beschlossen:

7. Nutzen

7.1. Word als XML-Editor

Bei der Erstellung der BGBl2000-XML-Serie konnten kostengünstige bzw vorhandene Softwarekomponenten genutzt werden. Der Einsatz von MS-WORD als XML-Editor ist hier wohl provokativ genannt worden. Nichts desto trotz kann mit gut strukturierten WORD Dokumenten eine Konvertierung in XML zu guten, brauchbaren Ergebnissen führen.

7.2. Info-Provider als Indexmaschine

Als Datenbank wurde das deutsche Produkt Info-Provider, ein OLE-DB-Provider der Iuk-Rieth GmbH Freiburg⁴ verwendet. Der technische Vorteil: extrem schnelle Suche mit Links- und Rechtstrunkierung, Index, Word-Wheel usw Kaufmännischer Vorteil: Vernünftige Lizenzpolitik.

7.3. Internet-Explorer 5.5 als Anzeigeprogramm

Die Anzeige erfolgt mittels integriertem xml-fähigen HTML-Browser. Bei der BGBl2000-Serie wurde MS-IE5.5 verwendet. Höhere Versionen sind unter Win 95 nicht mehr lauffähig.

⁴ www.ever-iuk.de.

7.4. Vorlagenoptimierung

Die Erfahrungen bei der Überführung eines Druckproduktes in ein XML-(quasi)Online-Produkt sind für die Erstellung zukünftiger Dokumentenvorlagen von großem Wert. Durch die automatische Granularisierung ist letztendlich auch an eine automatische Einbringung der Texte in weitere Zielsysteme denkbar (RIS).

7.5. Entmythologisierung von XML

Schließlich und endlich hat dieses Projekt auch zur internen Beschäftigung mit der Technologie XML geführt. Leider hat der Vorteil der „XML-Produkt-Schwemme“ auch zu einer enormen Beunruhigung im „Normierungsprozess“ geführt. Wer auf eine Vereinfachung des komplexen SGML-Standards gehofft hat, wird durch XML alleine noch nicht das goldene Zeitalter errichtet finden. Systeme, die, wie im Fall der BGBL2000 XML-Anwendung relativ große Spielräume bieten, ermöglichen diese auch in komplexen Systemen. Wer je eine WORD-Dokumentenvorlage für mehr als 2 Abteilungen erstellen musste, weiß wovon die Rede ist. *<Tag>*