



# XML als Archivierungsformat: Tools und News

Susanne Dobratz & Matthias Schulz  
Humboldt-Universität zu Berlin  
Arbeitsgruppe elektronisches Publizieren  
Rechenzentrum/Universitätsbibliothek



# Inhalt

- Historie
- XML als Archivierungsformat
- Die DIML DTD
- Schreiben in XML?
- XML Schreibwerkzeuge
- XML Konvertierung



# Historie 1

## ■ 1987

- erste Dissertations DTD (ETD) von Juri Rubinsky

## ■ 1997

- Entscheidung HU: SGML als Archivierungsformat
- Überarbeitung der DTD zur Version 1.1 von Neil Kipp
- Anpassung an deutsche Verhältnisse durch Matthias Schulz Benennung der DTD "DIML" (DissertationMarkupLanguage)
- Promovendenbefragung zur Akzeptanz Digitale Dissertation und Nutzung Schreibwerkzeuge
- Konvertierung mit SGML Author for Word
- Entwicklung einer Word-Dokumentvorlage



# Historie 2

## ■ 1998

- Beitritt NDLTD (erste dt. Universität)
- Etablierung der Dokumentvorlage
- Autorenschulungen an der HU Berlin

## ■ 1999

- Einbringen der HU- Erfahrungen in das Projekt Dissertation Online

## ■ Seit 2000

- Unterstützung weiterer Publikationsvorhaben in der Universität auf Basis von SGML/XML (Öffentliche Vorlesung, Zeitschriften)



# SGML/XML als Archivformat

- Langzeitarchivierung, d.h. 10 Jahre und länger
- Verfügbarkeit des Datenformates auf verschiedenen Hardwareplattformen und Betriebssystemen
- Internationale Normierung durch ISO8879
- Einfache Transformation in
  - neue Präsentationsformen oder
  - Druckformate (ohne Datenverluste frei wählbar)
- Integration von Multimediaobjekten
- Migration kann kostengünstiger realisiert werden



# SGML/XML als Rechercheformat

- Vereinheitlichung von Wissensstrukturen (DTDs)
- Semantische Kodierung
- Zielgenauere Recherchen im Vergleich zu
  - Katalogsuche
  - Volltextrecherche
  - Recherche in Metadaten (z.B. Dublin Core)
- Veredelung der Information durch Informationsextraktion
- Automatisierte Katalogisierung möglich



# Was ist ein Dokument?

• **Struktur**    **DTD: DIML, MATHML, CML, CALS**  
– Aufteilung und Abfolge der Information

• **Inhalt**    **Volltext - XML Instanz**  
– zu vermittelnde Information in Schrift, Bild, Ton

• **Layout**    **Stylesheets: XSL:FO, XSLT**  
– Visualisierung des Inhaltes und der Struktur des Dokumentes

• **Metadaten**    **Dublin Core, RDF**  
– Beschreibung des Dokuments



# XML als Archivformat für Dissertationen

- ... ABER darin schreibt kein Autor

MS Word	75%
LaTeX	22%
WordPerfect	2,6%
FrameMaker und Andere	1,4%
SGML/XML	keine

von 300 eingereichten Arbeiten an der Humboldt-Univ. im Zeitraum von Januar 1998 - Juni 2001



# XML als Archivformat für Dissertationen

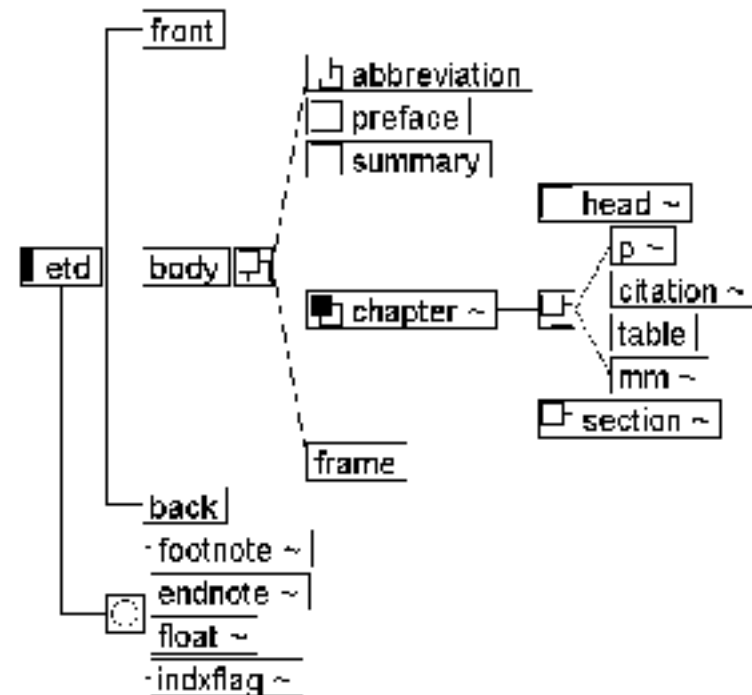
- ... ABER darin schreibt kein Autor, weil
  - Tools entsprechen nicht den Schreibgewohnheiten
  - Benutzung erfordert sehr viel Vorwissen
  - daher: „Abholen der Autoren dort, wo sie stehen“ (Prof. Diepold)
- FAZIT:
  - Definition eines Abgabeformates: Ziel: Basisstrukturierung des Textes durch Autoren
  - Aufbau einer Konvertierung
  - Aktive Autorenunterstützung
    - Dokumentvorlage
    - Online Hilfe zur Dokumentvorlage und FAQ
    - Kurse zur Dokumentvorlage



# Die Diml-DTD

- eine SGML-DTD
- DiML-Instanzen sind mit der XML-Spezifikation kompatibel, da nur diejenigen Eigenschaften von SGML verwendet wurden, die auch in XML enthalten sind

## DiML-Aufbau





# Umstieg von SGML nach XML

- Was musste getan werden?
  1. DTD nach XML-Spezifikation formulieren
  2. DTD weiter modularisieren, um neue Standard-DTD's zu integrieren
  3. Neue Konvertierungswerkzeuge  
Word??->XML
- Problem Qualität der Umsetzung
- Ausgabe in XML muss nach xDiML und nicht nach einer Anbieter-DTD



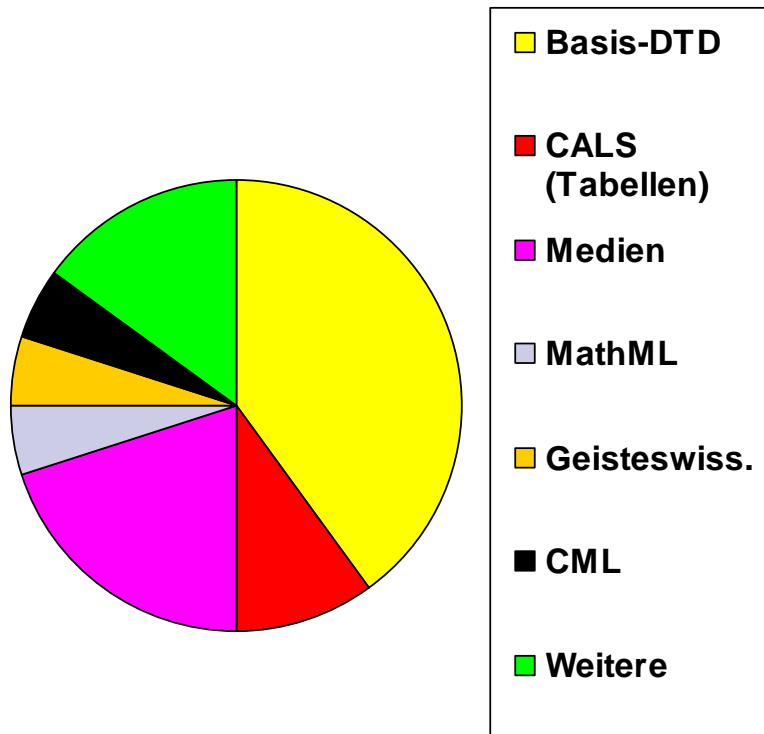
# Umstieg von SGML nach XML

## ■ Was ist **stabil geblieben**?

- Anforderungen an die Autoren in der Textverarbeitung Word, LaTeX
  - Dokumentvorlage
  - Formeln mit Formeleditor
  - Literatur mit Reference Manager o.ä.
  - Behandlung von Abbildungen, etc. als externe Dateien
- Schulungsinhalte



# xDiML3.0 "Pizzamodell"



- Basis-DTD
  - Front, Body, Back
  - Titelblatt
  - Bibliographie
- Medien
  - SVG, SMIL, weitere
  - Bitmapformate
- Geisteswiss.
  - Gedichte, Rezension
- Weitere
  - Offene definierte Schnittstelle z.B. TEI-Module, Dublin Core



# Warum Modularisierung

- Anpassung an internationale Entwicklungen
  - Mai 1999 DTD-Workshop HU Berlin
  - Vereinheitlichung von Standards, um Strategien, Konvertierungen, Tools, Schulungsunterlagen mehrfach zu nutzen (Migration nach XML)
  - Integration einer Recherche



# Warum Modularisierung? 2

Partner	HU Berlin	Virginia Tech	Univ. of Iowa	Univ. of Montreal/Lyon	Tech. Univ. Helsinki	Univ. of Michigan	Univ. Oslo
<b>DTD</b>	DiML	ETD	TDM	TEI-Light	HutPubl	TEI-Light	ISO-Book
<b>Konv. nach SGML/XML</b>	SGML-Author for Word	SGML-Author	Majix	Omnimark rtf2sgml	Frame-Maker+ SGML	Omni-mark rtf2sgml	Balise
<b>Von SGML nach HTML, PDF</b>	Perl-script, DSSSL	Perl-script	CSS	XSL	DSSSL, Frame-Maker	-	-

siehe <http://dochost.rz.hu-berlin.de/epdiss/dtd-workshop>



# Tools für XML

## ■ Editoren

- XMLNotepad, XMLSpy, XMetal, TurboXML, XMLCooktop, EPCEdit, ...

## ■ Textverarbeitungssysteme, DTP-Systeme

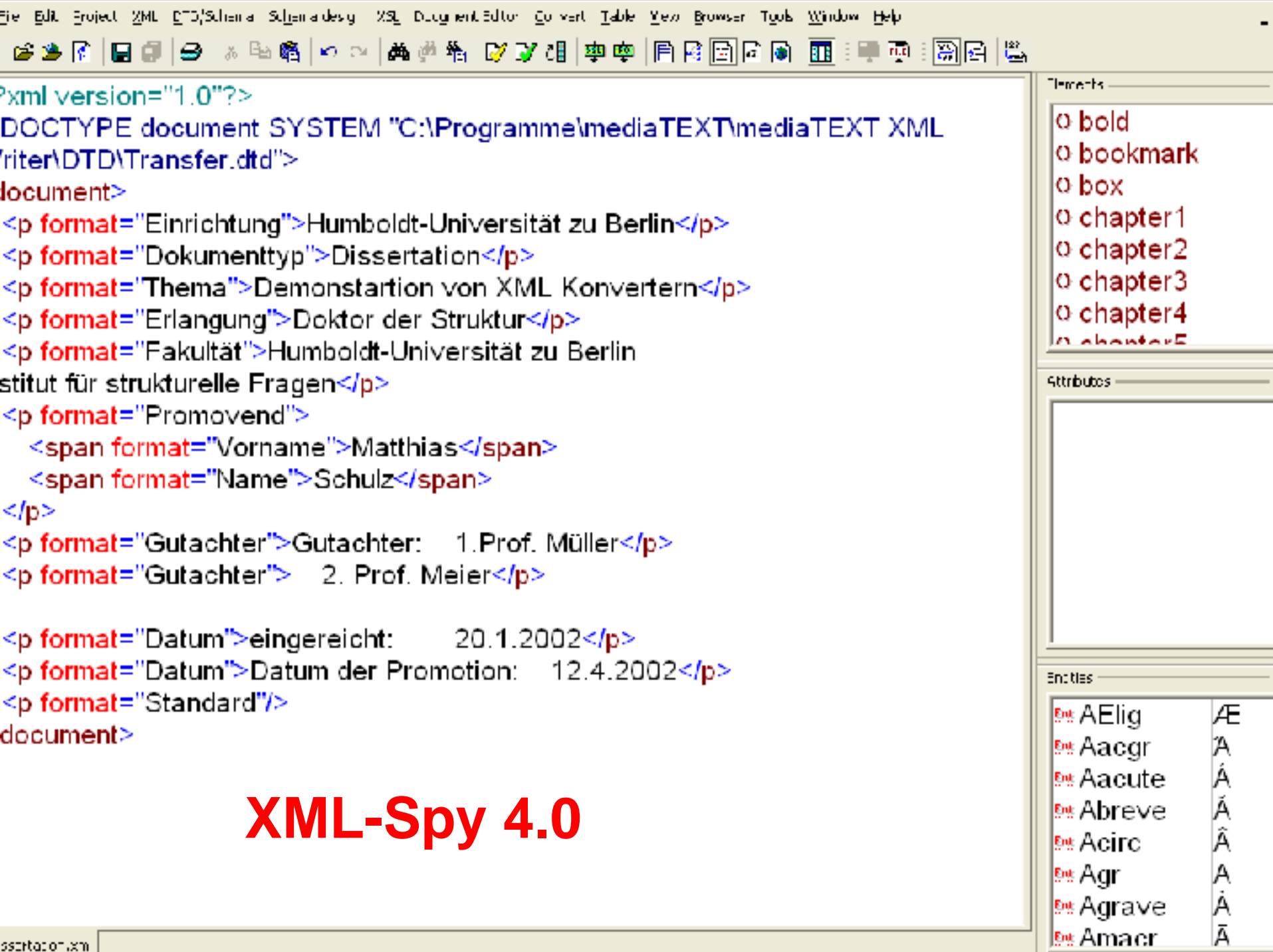
- Abiword, Staroffice, FrameMaker, QuarkXpress + Avenue, InDesign

## ■ Konverter

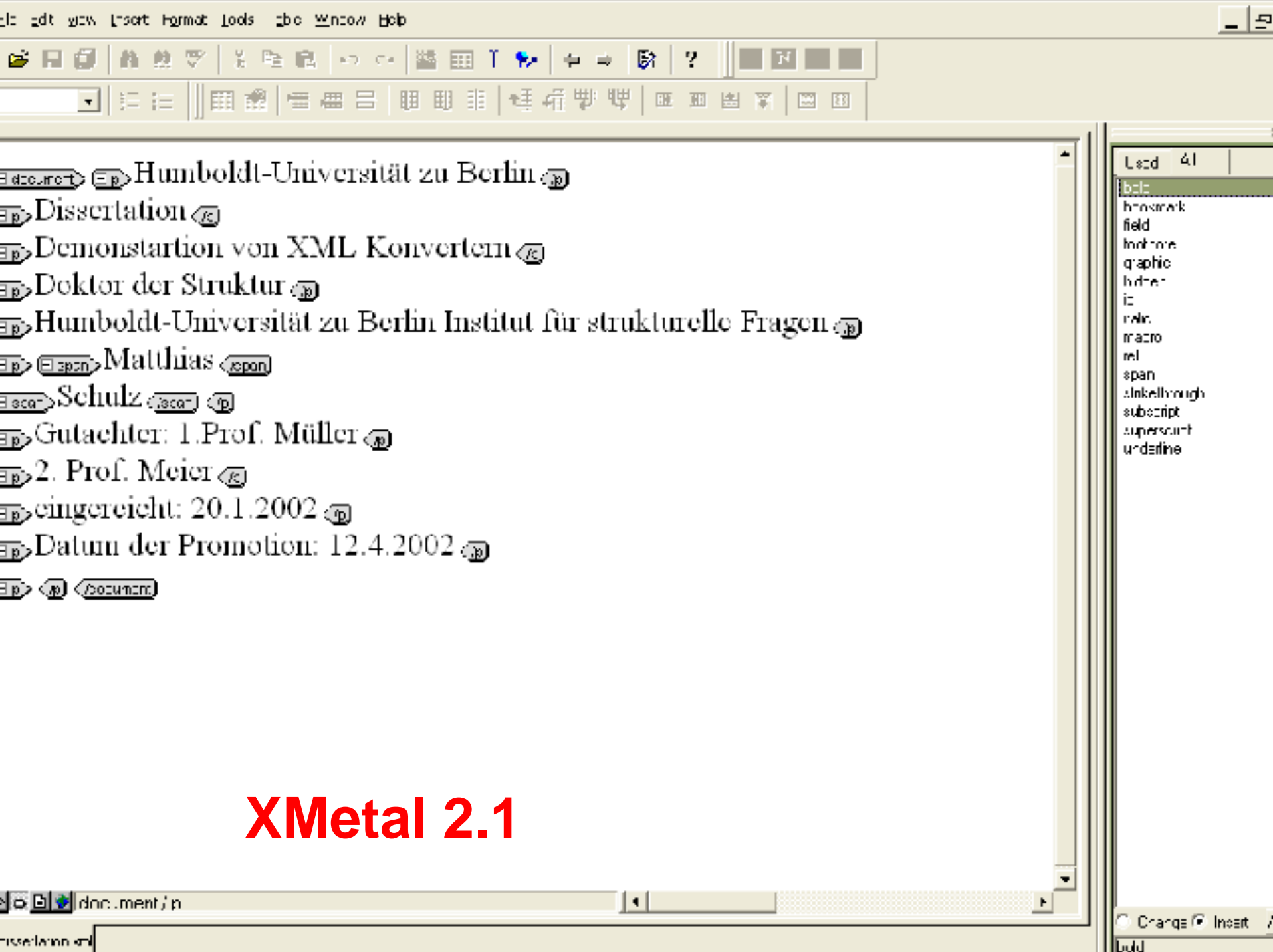
- Majix, SCHEMAMarkupKit, XMLWriter, ...

## ■ Layouter / Styleengines Satzsysteme

- Powerpublisher, 3B2



# XML-Spy 4.0



# XMetaL 2.1



Humboldt-Universität zu Berlin

# Staroffice 6.0

DISSERTATION

## Demonstration von XML Konvertern

Doktor der Struktur

Humboldt-Universität zu Berlin  
Institut für strukturelle Fragen

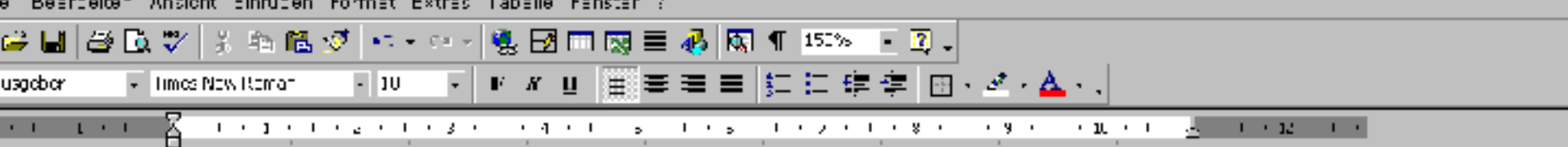
Matthias Schulz

```

?xml version="1.0" encoding="UTF-8"?>
!DOCTYPE office:document-content PUBLIC "-//OpenOffice.org//DTD OfficeDoc
office:document-content xmlns:office="http://openoffice.org/2000/office" :
  <office:script/>
  <office:font-decls>
    <style:font-decl style:name="Symbol" fo:font-family="Symbo
    <style:font-decl style:name="Courier New" fo:font-family="
    <style:font-decl style:name="Arial Unicode MS" fo:font-fam
    <style:font-decl style:name="HG Mincho Light J" fo:font-fa
    <style:font-decl style:name="Thorndale" fo:font-family="Th
    <style:font-decl style:name="Times" fo:font-family="Times"
    <style:font-decl style:name="Times New Roman" fo:font-fami
    <style:font-decl style:name="Arial" fo:font-family="Arial"
  </office:font-decls>
  <office:automatic-styles>
    <style:style style:name="P1" style:family="paragraph" styl
      <style:properties fo:margin-top="7.001cm" fo:margi
    </style:style>
    <style:style style:name="P2" style:family="paragraph" styl
      <style:properties fo:margin-top="6.001cm" fo:margi
    </style:style>
  </office:automatic-styles>
  <office:body>
    <text:sequence-decls>
      <text:sequence-decl text:display-outline-level="0"
      <text:sequence-decl text:display-outline-level="0"
      <text:sequence-decl text:display-outline-level="0"
      <text:sequence-decl text:display-outline-level="0"
    </text:sequence-decls>
    <text:p text:style-name="Einrichtung">Humboldt-Universität
    <text:p text:style-name="P1">Dissertation</text:p>

```

**XML-Output aus Staroffice 6.0**



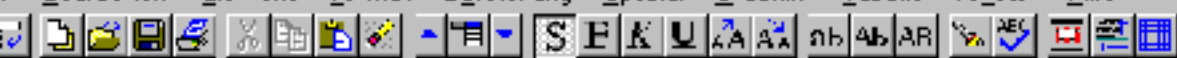
Kurt-Victor Selge  
Reimer Hansen  
Christof Gerstrich

**FrameMaker + SGML 6.0**

**Phillip Melanchthon 1497 - 1997**

Drei Reden, vorgetragen  
am Melanchthon-Dies der Theologischen Fakultät  
in der Humboldt-Universität zu Berlin

23. April 1997



- Auto: Titelseite
- Titelseite
  - Blocktitel
  - Blocktitel eingerückt
  - Body
  - Copyright
  - Datum
  - Datum Titelseite
  - Fakultät
  - Header
  - F. Br. ue
  - Graphikseite
  - Gross\_Link
  - Haup.titel
  - Head Anmerkung
  - Head Bibliographie
  - Head Impressum
  - Head Lebenskur
  - Head Vorw.
  - Header

X:\DVL\kml-5\test.fm



Kurt-Victor Selge¶  
 Reimer Hansen¶  
 Christof Gerstrich¶

**Phillip Melanchthon 1497 - 1997¶**

Drei Reden, vorgelesen  
 am Melanchthon-Denk der Theologischen Fakultät  
 in der Humboldt Universität zu Berlin¶

23. April 1997¶

**FrameMaker + SGML 6.0**

# FrameMaker + SGML 6.0

Wrap this object or objects	In this element	With this qualifier
P Autor_Titelseite	Autor	Autor_Titelseite
7 Haupttitel	Haupttitel	Haupttitel
P Untertitel	Untertitel	Untertitel
P Universität	Universität	Universität
7 Fakultät	Fakultät	Fakultät
P Head_Vorwort	Übersicht	Head_Vorwort
P Standard_Zentriert	Standard_Zentriert	
7 Datum_Titelseite	Datum	Datum_Titelseite
P Herausgeber	Herausgeber	Herausgeber
P Copyright	Copyright	
P Left-Nr	Left	Left-Nr
P Datum	Datum	Datum
P Graphikseite	Graphikseite	
P hmsg_text	hmsg_text	
P hmsg_Präsident	hmsg_Präsident	
7 hmsg_Name	hmsg_Name	
P re_Name	re_Name	
P re_Abteilung	re_Abteilung	
7 re_Address	re_Address	

Kurt-Victor Selge  
Reiner Hansen  
Christof Gerstrich

**Phillip Melanchthon 1497**

Die Reden, vorgelesen  
am Melanchthon-Dies der Theologischen Fakultät  
an der Humboldt Universität zu Berlin

23. April 1997

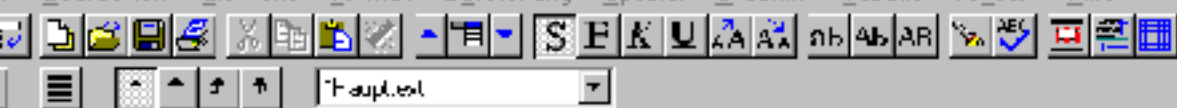
Aktuelles Dokument > Strukturieren

Konvertierungsabellen: Textwert wordEvol III

Struktur hinzufügen

Abrechen

**FrameMaker + SGML 6.0**

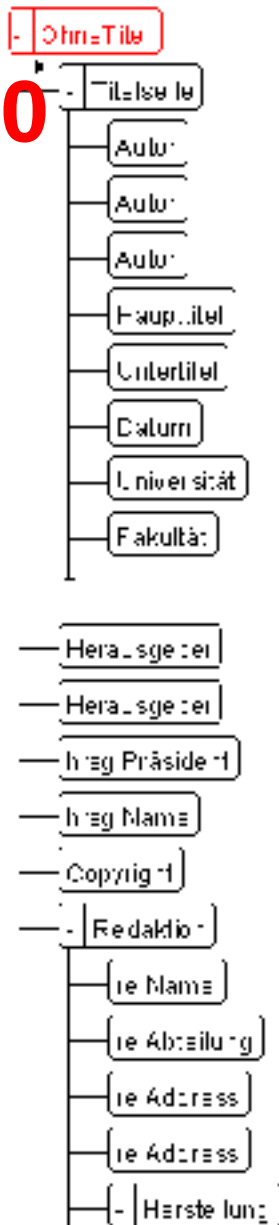


und titell.fm

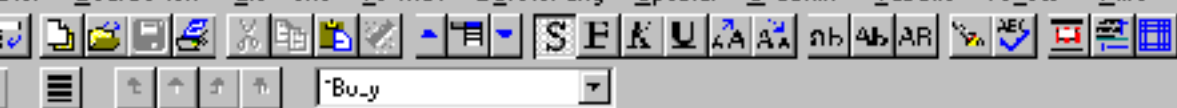
# FrameMaker + SGML 6.0

Kurt-Victor Selge  
 Reinier Hansen  
 Christof Gerssich  
 Philip Melancthon 1497 - 1597  
 Drei Reden, vorgetragen  
 am Melancthon-Dies der Theologischen Fakultät  
 an der Humboldt-Universität zu Berlin  
 23. April 1997  
 Humboldt-Universität zu Berlin  
 Theologische Fakultät  
 Herausgeber: Herausgeber  
 Der Präsident der Humboldt-Universität zu Berlin  
 Prof. Dr. Dr. h.c. Hans Meyer  
 Copyright: Alle Rechte liegen beim Verfasser  
 Gudrun Kramer  
 Forschungsabteilung der Humboldt-Universität  
 Unter den Linden 6  
 10099 Berlin  
 Lince DREI, Agentur für Satz und Graphik  
 Wühlschstr. 33  
 10245 Berlin  
 Fax 874  
 Redaktionsschluss 18.07.1997  
 Verantwortlich  
 Die Theologische Fakultät der Humboldt-Universität hat am  
 April 1997 zu einem Dies academiens in ihrem großen Hörsaal  
 Berlin-Damm eingeladen. Anlat war die 500. Wiederkehr  
 Geburtstag des Philologen Melancthon am 14. Februar 1597

## Strukturansicht



Kurt-Victor Selge  
 Reinier Hansen  
 Christof Gerssich  
 Philip Melancthon 1497 - 1927  
 Drei Reden, vorgetragen am Melanctho...  
 23. April 1997  
 Humboldt-Universität zu Berlin  
 Theologische Fakultät  
 Herausgeber: Herausgeber  
 Der Präsident der Humboldt-Universität...  
 Prof. Dr. Dr. h.c. Hans Meyer  
 Copyright: Alle Rechte liegen bei...  
 Gudrun Kramer  
 Forschungsabteilung der Humboldt-...  
 Unter den Linden 6  
 10099 Berlin



# FrameMaker + SGML 6.0

```

[EDD-Version 6.0]
-----
[SGML-Anwendung: TV1]

[Element (Container): CV]
  [[[Gültig als Element auf höchster Ebene.]]
  [Allgemeine Regel:Titelblatt, Impressum, (Toc? Floating|Vorwort?)*, Vorlesung*]]

[Element (Container): Floating]
  [Allgemeine Regel:(Paragraf|img)*]

  [Attributliste
    1. [[Name: ID ]]           Einzigartige ID [[ Optional ]]
    2. [[Name: Farbe ]]       Zeichenfolge [[ Optional ]]]]

[Element (Container): Cit]
  [Allgemeine Regel:<TEXT>]

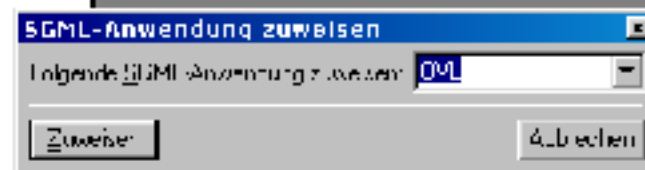
  [Attributliste
    1. [[Name: ID ]]           Einzigartige ID [[ Optional ]]
    2. [[Name: Farbe ]]       Zeichenfolge [[ Optional ]]]]

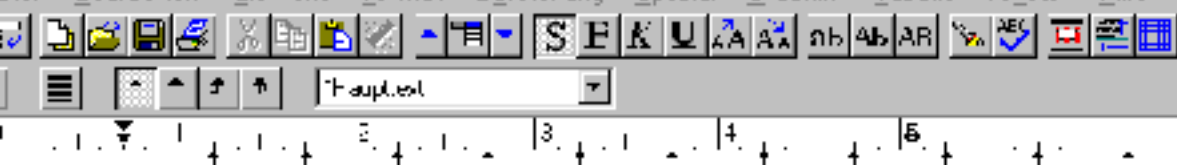
[Element (Container): Citycode]
  [Allgemeine Regel:<TEXT>]

  [Attributliste
    1. [[Name: ID ]]           Einzigartige ID [[ Optional ]]
    2. [[Name: Farbe ]]       Zeichenfolge [[ Optional ]]]]
  
```

# FrameMaker + SGML 6.0

{Kurt Victor Selge¶  
{Lerner Hansch¶  
{Christof Gersticht¶  
{Philipp Melanchthon (1523-1591)¶  
{Drei Reden, vorgelesen¶  
{am Melanchthon-Dies der Theologischen Fakultät¶  
{in der Humboldt-Universität zu Berlin¶  
{23. April 1957¶  
{Humboldt-Universität zu Berlin¶  
{Theologische Fakultät¶  
{¶  
{Herausgeber, Herausgeber,¶  
{Der Präsident der Humboldt-Universität zu Berlin¶  
{Prof. Dr. Dr. h. c. h. Hans Meyer¶  
{Copyright: Alle Rechte liegen beim Verfasser¶  
{Gunter Kramer¶  
{Forschungsabteilung der Humboldt-Universität¶  
{Unter den Linden 6¶  
{10059 Berlin¶  
{Lilie D. 211, Agentur für Satz und Graphic¶  
{Wüchelsch 39¶  
{10245 Berlin¶  
{Hell 87¶  
{Postfachschrift 15 07 1097¶  
{¶  
{Vorwort¶  
{Die Theologische Fakultät der Humboldt-Universität hat am 23.  
{April 1957 zu einem Dies academicus in ihren großen Hörsaal im  
{Berliner Dom eingeladen. Anlaß war die 100. Wiederkehr des  
{Geburtstages von Philipp Melanchthon am 16. Februar Melan-  
{chthons in seiner Zeit und in der wechselnden Bedeutung des





# FrameMaker + SGML 6.0


*Karl Victor Selge*  
*Ribi von Hansen*  
*Ulrich von Gersdorff*

## Phillip Melanchthon

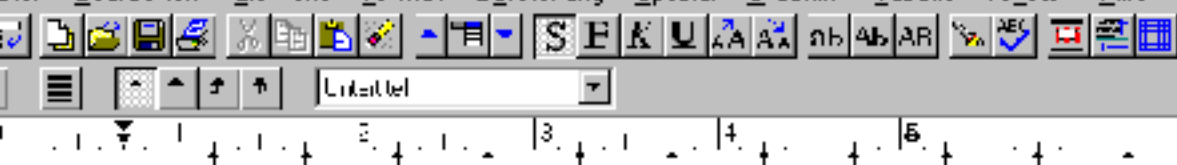
Die Rede, vorgelesen  
am Melanchthon-Dienstag der Theologie  
an der Humboldt Universität

23. April 1997

FrameMaker+SGML

 Elemente müssen werden aus der EDD importiert...

OK



# FrameMaker + SGML 6.0

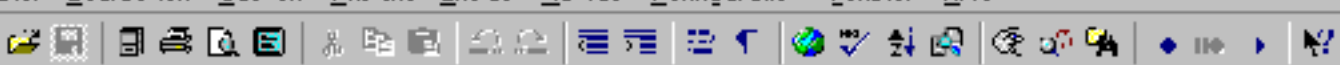
Umschlag > Titelkarte > Autor > *Kurt-Viktor Selig* < Autor ¶  
Autor > *Reiner Hansen* < Autor ¶  
Autor > *Christof Gerstlich* < Autor ¶

Haupttitel > **Phillip Melanchthon 1497 -  
1997** < Haupttitel ¶

Untertitel > Dies Buch, vorgelesen  
am Melanchthon-Dies der theologischen Fakultät  
in der Humboldt-Universität zu Berlin < Untertitel ¶

Datum > 23. April 1997 < Datum ¶





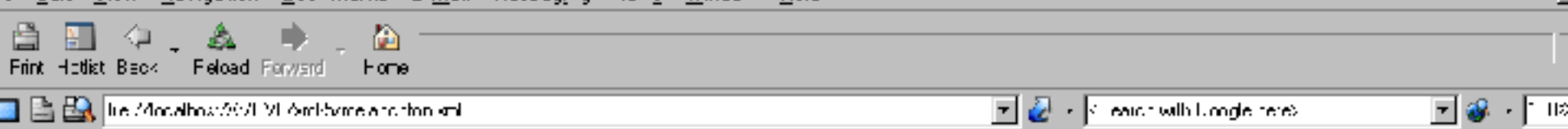
```

<?xml version='1.0' encoding='UTF-8'?>
<?xml-stylesheet href='Ovl.css' type='text/css'?>
<!DOCTYPE owl SYSTEM 'Ovl.dtd' [

<owl:
  <titlepage>
    <author>Kurt Victor Selge</author>
    <author>Hanser Hansens</author>
    <author>Christof Gestrich</author>
    <title>Philipp Melancthon 1497 - 1997</title>
    <subtitle>Drei Reden, vorgetragen am Melancthon Dies der Theologischen
      Fakultät an der Humboldt-Universität zu Berlin</subtitle>
    <date>23. April 1997</date>
    <university>Humboldt-Universität zu Berlin</university>
    <faculty>Theologische Fakultät</faculty>
  </titlepage>
  <impressum id='impressum'>
    <heading>Impressum</heading>
    <hrsg>
      <text>Herausgeber:</text>
      <affiliation>Der Präsident der Humboldt-Universität zu
        Berlin</affiliation>
      <name><addition>Prof. Dr. Dr.
        h. c. </addition><firstname>Hans</firstname> <lastname>Neger</lastname></name>
    </hrsg>
    <copyright>Copyright: Alle Rechte liegen beim Verfasser</copyright>
    <redaktion>
      <name><firstname>Andrun
        </firstname> <lastname>Erkner</lastname></name>
      <abteilung>Forschungsabteilung der Humboldt-Universität</abteilung>
      <address> <street>Unter den Linden</street>
        <city/> <country>10099</country> <tel>Hilf mir helfen
        </tel> </address> <date>Redaktionschluß: 18.04.1997</date>
    </redaktion>
    <no>Heft 87</no>
  </impressum>
  <foreword>
    <heading id='vorwort'>Vorwort</heading>
    <p>Die Theologische Fakultät der Humboldt Universität hat am 23.
      April 1997 zu einem Dies academicus in ihrem großen Mensaal in
      Berliner Dom eingeladen. Anlaß war die 500. Wiederkehr des Geburtstages
      von Philipp Melancthon am 16. Februar. Melancthons in seiner Zeit und in der
      wechselnden Beurteilung der Nachwelt vornehmlich des 19. und 20. Jahrhunderts
      sollte vor einer breiteren Publikum gedacht werden. Dieses Ziel dient auch die
      Veröffentlichung der Reden in der Schriftenreihe 'Öffentliche
      Vorlesungen' der Humboldt-Universität.</p>
    <p>Diese Universität hat Melancthons vor 100 Jahren, am 16. Februar
      1997, schon gedacht, und der Abstand zu ihm war schon damals kaum geringer als
      heute. Andererseits gibt es eine Linie der Kontinuität. Sie hängt mit
      dem Wesen der gelehrten Bildung und der Schulbildung und der Frage der
      Bedeutung der Religion für beide wie der Bedeutung der Bildung für

```

# XML-Output aus FrameMaker + SGML 6.0



**Kurt Victor Selge**

**Reiner Hansen**

**Christof Gestrich**

## **Philipp Melanchthon 1497 - 1997**

Drei Reden, vorgelesen am Melanchthon-Dies der Theologischen Fakultät an der Humboldt-Universität zu Berlin

23. April 1997

Humboldt-Universität zu Berlin

Theologische Fakultät

### **Impressum**

Herausgeber:

Der Präsident der Humboldt-Universität zu Berlin

Prof. Dr. Dr. h.c. Hans Meyer

Copyright: Alle Rechte liegen beim Verfasser

Gudrun Krämer

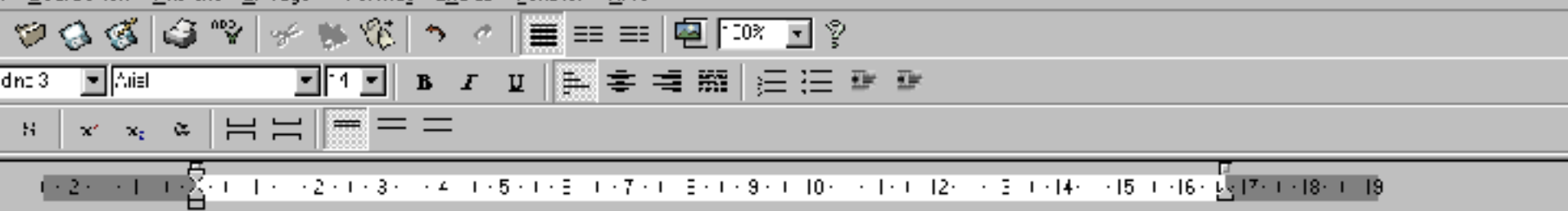
Forschungsabteilung der Humboldt-Universität

Unter den Linden 6 10099 Berlin

Redaktionschluss: 18. 07. 1997

Heft 87

**XML-Output  
mit XSLT-Stylesheet  
aus FrameMaker + SGML 6.0  
Im Opera-Browser**



# AbiWord

## Die Sprache der Vögel

### **Der Engpaß**

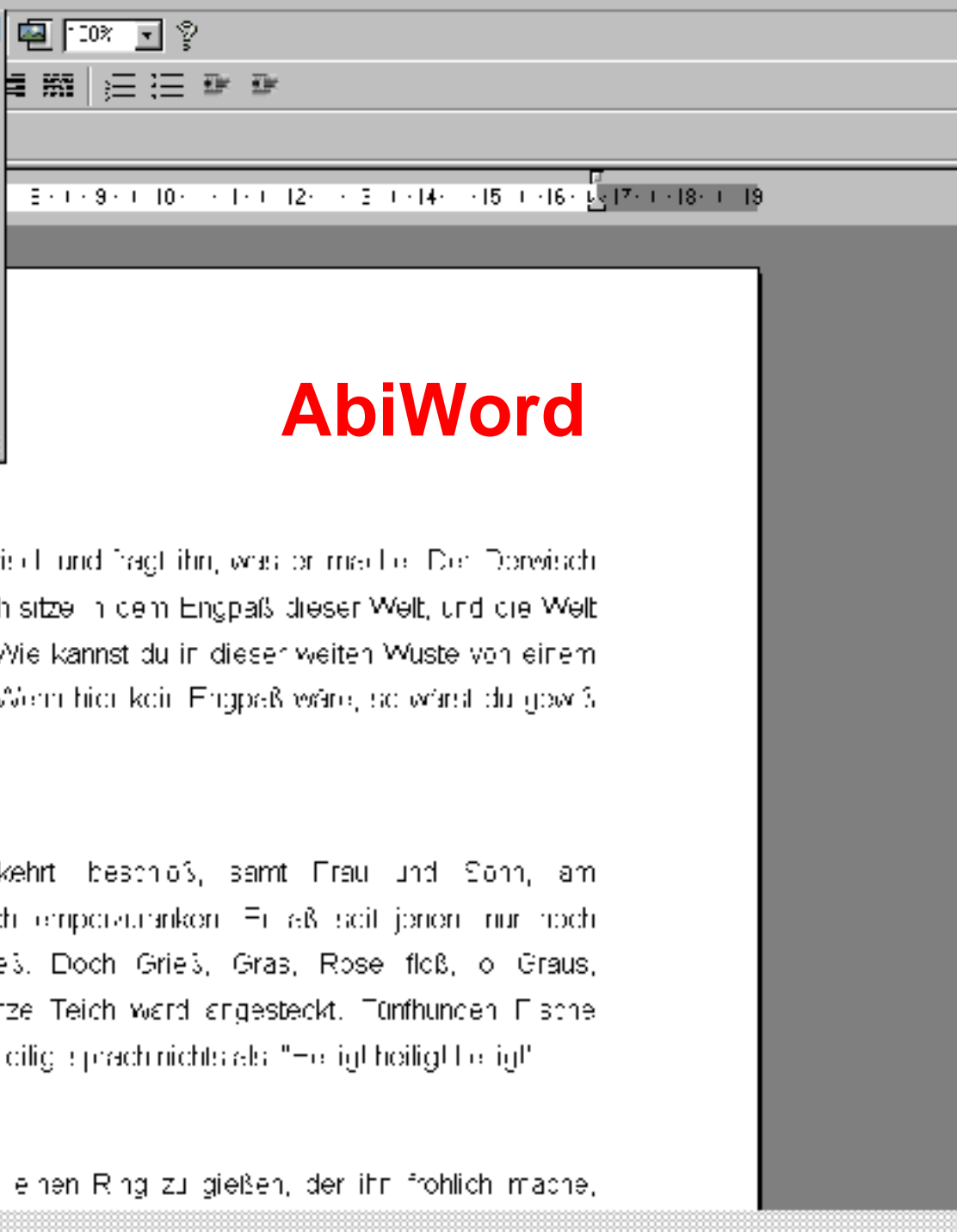
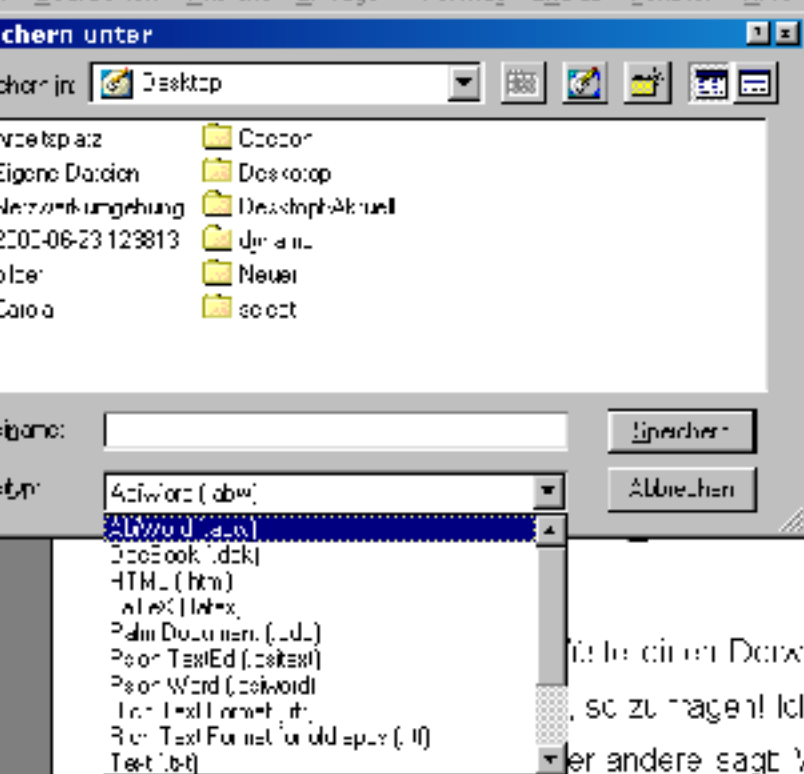
Ein Mann trifft in der Wüste einen Dornich und fragt ihn, was er mache. Der Dornich antwortet: Schäme dich, so zu fragen! Ich sitze in dem Engpaß dieser Welt, und die Welt ist mir jetzt gar enge. Der andere sagt: Wie kannst du in dieser weiten Wüste von einem Engpaß sprechen? Der Dornich sagt: Wenn hier kein Engpaß wäre, so wärest du gewiß nicht auf mich gestoßen.

### **Der Hecht**

Ein Hecht, vom heiligen Anton bekehrt beschoß, samt Frau und Sohn, am vogelreichen Gedankenmonat sich emporkranken. Er aß seit jenem nur noch dies: Seegras, Seerose und Seegrieß. Doch Grieß, Gras, Rose floß, o Graus, entsetzlich wieder hinter aus. Der ganze Teich ward ergesteckt. Tausendfische sind vondeckt. Doch sank Anton, gerufen eilig sprach nichts als: "Heilig heilig heilig!"

### **Der Ring des Königs**

Ein König befiehlt seinen Weisen, ihm einen Ring zu gießen, der ihn frohlich mache,



# AbiWord

Erste einen Dorfwirt und fragt ihn, was er mache. Der Dorfwirt antwortet: „Ich sitze in dem Engpaß dieser Welt, und die Welt stößt hier an.“  
Der andere sagt: Wie kannst du in dieser weiten Wüste von einem Engpaß sprechen? Der Dorfwirt sagt: Wenn hier kein Engpaß wäre, so wärest du gewiß nicht auf mich gestoßen.

## Der Hecht

Ein Hecht, vom heiligen Anton bekehrt beschoß, samt Frau und Sohn, am vogelreichen Gedanken moralisch sich emporkranken. Er aß seit jenen nur noch dies: Seegras, Seerose und Seegrieß. Doch Grieß, Gras, Rose floß, o Graus, entsetzlich wieder hinter aus. Der ganze Teich ward ergesteckt. Tausendfische sind verreckt. Doch sank Anton, gerufen eilig sprach nichts als: „Heilig heilig heilig!“

## Der Ring des Königs

Ein König befiehlt seinen Weisen, ihm einen Ring zu gießen, der ihn frohlich mache,

```
<?xml version='1.0' encoding='ISO-8859-1'?>
<abiword version='0.7.13' fileformat='1.0'>
<!--
----->
<!-- This file is an AbiWord document. -->
<!-- AbiWord is a free, Open Source word processor. -->
<!-- You may obtain more information about AbiWord at www.abi-source.com -->
<!-- You should not edit this file by hand. -->
<!--
----->

<!--      Build ID          = (none) -->
<!--      Build_Version     - 0.7.13 -->
<!--      Build_Options     = Licensed:Off Debug:Off Icons:Off Spell:Off -->
<!--      Build_Target      = /cygdrive/c/abiword/abi/src/WIN32_1.1.8_i386_OB/obj -->
<!--      Build_CompTime    = 10:22:02 -->
<!--      Build_CompDate    = Feb 22 2001 -->

</!-->

<lists>
<list id='15467' parentid='0' type='0' start-value='1' list-delim='&#x2013;' list-decimal='.'>
</lists>
<pagesize pagetype='Letter' orientation='portrait' width='8.500000in' height='11.000000in' units='inch' page-scales='1.000000in'>
<section props='page-margin-bottom:0.7875in, page-margin-left:0.9840in, page-margin-right:0.9840in, page-margin-top:0.9840in'>
<p style='Heading 1'>Die Sprache der Vögel</p>
<p style='Heading 2'>Der Engpaß</p>
<p props='line-height:1.5; margin-bottom:6pt; margin-top:6pt; orphans:0; text-align:justify; widows:0'><!-->
<p style='Heading 3'>Der Hecht</p>
<p props='line-height:1.5; margin-bottom:6pt; margin-top:6pt; orphans:0; text-align:justify; widows:0'><!-->
<p style='Heading 3'>Der Ring des Königs</p>
<p props='line-height:1.5; margin-bottom:6pt; margin-top:6pt; orphans:0; text-align:justify; widows:0'><!-->
<p level='1' listid='15467' parentid='0' style='Numbered List' props='bgcolor:ffffff; color:000000; field-font:NULL; font-family:Arial; font-size:10pt; list-style-type:disc; margin-bottom:6pt; margin-top:6pt; orphans:0; text-align:justify; widows:0'>1. Die Sprache der Vögel</p>
<p level='1' listid='15467' style='Numbered List' props='bgcolor:ffffff; color:000000; field-font:NULL; font-family:Arial; font-size:10pt; list-style-type:disc; margin-bottom:6pt; margin-top:6pt; orphans:0; text-align:justify; widows:0'>2. Der Engpaß</p>
<p level='1' listid='15467' style='Numbered List' props='bgcolor:ffffff; color:000000; field-font:NULL; font-family:Arial; font-size:10pt; list-style-type:disc; margin-bottom:6pt; margin-top:6pt; orphans:0; text-align:justify; widows:0'>3. Der Hecht</p>
<p level='1' listid='15467' style='Numbered List' props='bgcolor:ffffff; color:000000; field-font:NULL; font-family:Arial; font-size:10pt; list-style-type:disc; margin-bottom:6pt; margin-top:6pt; orphans:0; text-align:justify; widows:0'>4. Der Ring des Königs</p>
</section>
</abiword>
```

## XML-Output aus AbiWord



```
AbiWord Markup Language DTD
Current as of 2-07-00
```

```
Currently, this document is intend for reference purposes only,
and should not be interpreted as an absolute guide to the
AbiWord file format. A more comprehensive view can be gained
from the code in files ie_exp_AbiWord_1.cpp and
is in ap_AbiWord_1.cpp. These are currently definitive. Other
documentation is available in docs/AbiWord_DocumentFormat.abw
```

```
Sam Tobin-Hochstadt
```

# AbiWord-DTD

```
<!ELEMENT abiword (styles?, lists?, section+, data?)*>
<!ATTLIST abiword
  version CDATA 'unnumbered'>
<!-- This is the root element, with an attribute
representing the version number. -->
```

```
<!ELEMENT styles (s)*>
<!ELEMENT s EMPTY>
<!ATTLIST s
  basison CDATA #IMPLIED
  name CDATA #REQUIRED
  type CDATA #IMPLIED
  props CDATA #IMPLIED>
<!-- Styles allow for paragraph level formatting.
The basison attribute specifies inheritance
and the props attribute specifies features
of the style. -->
```

```
<!ELEMENT lists (l)*>
<!ELEMENT l EMPTY>
<!ATTLIST l
  id ID #REQUIRED
  parentid CDATA #REQUIRED
  type CDATA #REQUIRED
  startvalue CDATA #REQUIRED
  listidlin CDATA #REQUIRED>
<!-- Styles allow for paragraph level formatting
The basison attribute specifies inheritance
and the props attribute specifies features
of the style. -->
```

```
<!ELEMENT section (p+)>
<!ATTLIST section
  props CDATA #IMPLIED
```



# XML unterstützende Schreibwerkzeuge

- Nicht linearer Schreibvorgang beim Autor
  - ständige Überprüfung kann hinderlich sein
  - verändertes Verfahren beim Schreiben : hohe Kenntnis der Struktur ist vorauszusetzen
- anderes Schulungskonzept ist erforderlich
- entsprechende Software ist in den einzelnen Fächern anzuwenden



# Demonstration von XML Konvertern

Doktor der Struktur

Humboldt-Universität zu Berlin  
Institut für strukturelle Fragen

Matthias Schulz



Startet die Konvertierung

**Word + XMLWriter Add-IN Mediatext Jena**

```
TextPad - [C:\Dokumente und Einstellungen\Matthias Schulz\Desktop\Word-Beispiel\dissertation.xml]
Datei Bearbeiten Suchen Ansicht Extras Makros Konfigurator Fenster Hilfe
[Icons]
<?xml version="1.0"?>
<!DOCTYPE document SYSTEM "C:\Programme\mediatEXT\mediatEXT XML Writer\DTD\
<document>
  <p format="Einrichtung">Humboldt Universität zu Berlin</p>
  <p format="Dokumenttyp">Dissertation</p>
  <p format="Thema">Demonstration von XML Konvertern</p>
  <p format="Erlangung">Doktor der Struktur</p>
  <p format="Fakultät">Humboldt-Universität zu Berlin
Institut für strukturelle Fragen</p>
  <p format="Promovend">
    <span format="Vorname">Matthias</span>
    <span format="Name">Schulz</span>
  </p>
  <p format="Gutachter">Gutachter: 1. Prof. Müller</p>
  <p format="Gutachter">2. Prof. Meier</p>
  <p format="Datum">eingereicht: 20.1.2002</p>
  <p format="Datum">Datum der Promotion: 12.4.2002</p>
  <p format="Standard"/>
</document>
```

# Output aus Word + XMLWriter Add-IN Mediatext Jena



# Konvertierung LaTeX

- Latex-Dokumente relativ gut strukturierbar
  - Autorenrichtlinien, STY-Files
  - Problem: Autoren erstellen math. Formeln nicht semantisch bezogen (Bsp.: Müh)
- Projekt: Univ. BW München
  - Herr Piruschka
- Projekt in Planung: CalTech
  - mit Mitgliedern der Tech User Group (Sebastian Rahtz u.a.)



# Fragen?

Herzlichen Dank für Ihre Aufmerksamkeit!

- Matthias Schulz

- [matthias.schulz.1@rz.hu-berlin.de](mailto:matthias.schulz.1@rz.hu-berlin.de)

- Susanne Dobratz

- [dobratz@rz.hu-berlin.de](mailto:dobratz@rz.hu-berlin.de)