

XDIML – Dokumentstandard für Dissertationen in XML

Matthias Schulz

Computer- und Medienservice
der Humboldt-Universität zu Berlin

Agenda

- Was ist XDIML?
- Was bietet XDIML?
- Wie funktioniert XDIML?

Was ist XDIML?

- XDIML ist eine DTD (DokumentTypDefinition)
- XDIML ist in einzelnen Modulen aufgebaut, dadurch können:
 - zusätzliche Module von Autoren verwendet werden
 - nicht benötigte Module können gelöscht werden
- XDIML beinhaltet auch Standard DTDs wie MathML oder für Tabellen die CALS-Table.DTD

Warum ein XDIML Format

- Langzeitverfügbarkeit von Dissertationen und wissenschaftlichen Schriften
- qualifizierte Recherche von Volltexten und Extraktion von Teilen des Dokumentes
- Medienunabhängigkeit, sowohl Bearbeitung als auch die Verarbeitung mit standardisierten Werkzeugen
- standardisiertes Format, sowohl in der Definition des Formates (XML-Standard) als auch in der Definition in Form einer DTD/Schema

Vorteile

- Archivierung: durch die standardkonforme Definition Grundvoraussetzung für eine effektive Archivierung
- Recherche: zielgenaue Suche und Entnahme von Informationsbestandteilen (automatisierte Katalogisierung, OAI)
- Medienunabhängige Herstellung mit jedem XML fähigen Werkzeug

Wie funktioniert XDIML

- XDIML DTD ist eine Verbindung von Autoren- und Archiv-DTD
 - z.B. **pagenumber** (ist für das Schreiben nicht notwendig)
- Verfügbarkeit:
 - DTD kann von edoc.hu-berlin.de/diml heruntergeladen werden
 - Dokumentation zur DTD in Englisch verfügbar



XDIML für Autoren

- ein direktes Schreiben tritt sehr selten auf
 - die meisten Autoren schreiben mit einer Textverarbeitung
- „alle“ heute am Markt befindlichen Programme können in einem eigenen Format XML abspeichern (StarOfficeML, WordML)

Textverarbeitung StarOffice7

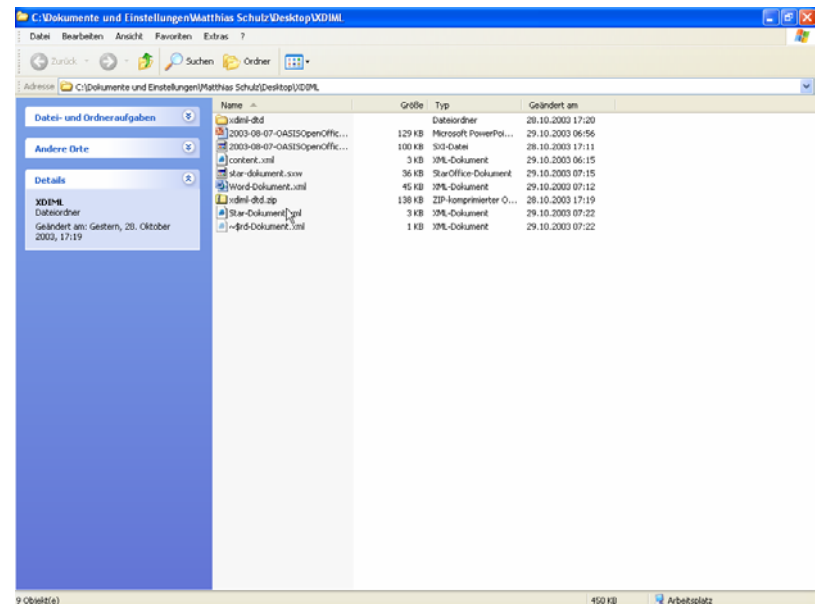
- Dateiformat ist vollständig in XML umgesetzt
- als Standard bei der OASIS eingereicht und Ende 2003 als Standard verabschiedet
- freie Textverarbeitungsprogramme wie KOffice, OpenOffice unterstützen dieses Format
- mit der Integration selbstdefinierter Filter können Erweiterungen vorgenommen werden (Speichern unter XDIML)
- DTD basiert

OASIS

- OASIS = Organization for the Advancement of Structured Information Standards
- Non-Profit, global consortium for e-business standards
 - Web services, business transactions, electronic publishing, interoperability within and between marketplaces, etc.
- Lightweight and open processes
 - expressly designed to promote industry consensus and unite disparate efforts
- More than 600 corporate and individual members

Textverarbeitung Word2003

- Dateiformat nicht vollständig in XML
 - Zeichnungsobjekte und Bilder als binärer Blob gespeichert
- WordML-Schema nicht einfach zugänglich
- Transformation und direktes Schreiben in anwenderspezifisches Format möglich
- Schema basiert



Strukturierung des Textes

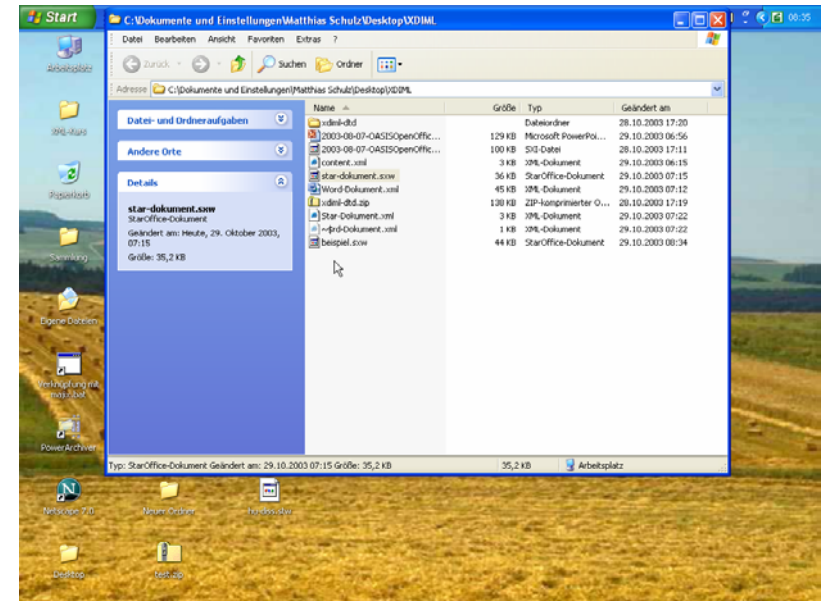
- um die Vorteile des XDIML-Formates zu nutzen, muß der Autor strukturiert schreiben
- das heißt:
 - Verwendung von Dokumentvorlagen
(Überschrift 1 = `<chapter> <head>`)
 - Transformationen sind nur so möglich
- Kontrolle und Hilfe für die Formatierung des Textes

StarOffice7 - Word2003

- vollständige XML-Unterstützung (alle Applikationen)
- OASIS genormtes Format
- Unterstützung von MathML, SVG, DublinCore, XLink
- Transformation von anwenderspezifischen Formaten möglich
- OpenOffices freiverfügbare Textverarbeitung
- nicht vollständige XML-Unterstützung
- kein genormtes Format
- keine Unterstützung von MathML, SVG...
- Transformationen und Erstellung von anwenderspezifischen Formaten möglich
- vollständig kommerzielles Produkt

StarOffice7

- Dokumentvorlage
dissertation-hu.stw
 - Menü Dissertation
- Beschreibung für Autoren
- Filter zur Kontrolle der
Formatvorlagen
- Filter zur Transformation in
das XDIML-Format
- Unterstützung der
Betriebssysteme Windows,
Macintosh, Linux



Aussichten

- Dokumentvorlage und Beschreibung für Autoren für StarOffice wird November 2003
- Filter und Konvertierung Mitte November 2003
- Ende 2003 Unterstützung für Word2003
- 2004 Umsetzung der DTD in ein Schema
 - Autorenhilfe zur DTD/Schema
 - Überarbeitung der Word-Dokumentvorlage

**Vielen Dank
für Ihre Aufmerksamkeit**

matthias.schulz.1@cms.hu-berlin.de

<http://edoc.hu-berlin.de>