

RDF und XML - Möglichkeiten für digitale Publikation und Archivierung

Alexander Schreiber
TU Chemnitz
Fakultät für Informatik
als@informatik.tu-chemnitz.de

18. April 2000

Zusammenfassung

Welche Möglichkeiten bieten XML und RDF für das digitale
Publizieren und Archivieren ?

Was ist XML ?

- XML - Sprache zur Definition von Sprachen
- Mittel zur Strukturierung von Informationen
- abgeleitet von SGML (da SGML sehr komplex)
- *wohlgeformte* XML-Dokumente sind auch gültige SGML-Dokumente
→ mit SGML-Tools bearbeitbar
- hochflexibel → sehr viele Anwendungen
- reines Textformat → *plattformunabhängig*
- mittlerweile in sehr vielen Projekten verwendet

XML-Beispiel

```
<?xml version="1.0" standalone="yes"?>
  <document>
    <author>Alexander Schreiber</author>
    <title>
      RDF und XML - Moeglichkeiten fuer digitale Publikation
      und Archivierung
    </title>
    <date>2000-04-18</date>
    <language>GER</language>
    <publisher>
      Fakultaet fuer Informatik
    </publisher>
  </document>
```

Warum mit XML publizieren ?

- plattformunabhängige, offene Formate
- layoutneutrale Erstellung der Publikation
- einheitliche Datenbasis für verschiedene Publikationsformen
- maschinenlesbare Strukturierung möglich
- vereinfacht automatische Bearbeitung von Publikationen
- vereinfacht Realisierung besonderer Darstellungsformen (z.B. für Blinde)

XML und digitale Publikationen

- DocBook XML = XML Version von DocBook (SGML)
<http://nwalsh.com/docbook/>
- TEI Text Encoding Initiative <http://www.hti.umich.edu/docs/TEI/>
- Quark eStage: Publishing-System auf XML-Basis
- XHTML - Redefinition von HTML 4.0 auf XML-Basis
<http://www.w3.org/TR/xhtml1/>
- digitale Buchformate

XSL

- Umformungsmittel von Struktur auf Layout notwendig
- XSL = eXtended Stylesheet Language
- auf CSS und DSSSL basierend
- ordnet XML-Elementen Layout zu
- 2-stufiger Prozess : XML \rightarrow XSL fo \rightarrow Ausgabeformat
- generiert aus XML beliebiges Ausgabeformat (WML, Webseiten, Papier, . . .)

MONARCH und XML

- derzeit in MONARCH akzeptiert: Text, HTML, PS, DVI, PDF
- Probleme:
 - aus Archivsicht: u.a. Indizierbarkeit bei PDF und PS
 - aus Nutzersicht: Anzeigesoftware
 - bisher kein XML möglich
- Diplomarbeit zu Integration von XML (Bert Auerbach), abgeschlossen
- Archivierung von XML und daraus erzeugtem HTML

RDF

- Metadaten
- Beschreibung von Ressourcen (URI)
- maschinen-verständlich
- Idee: „Indexkarten“ für Web-Ressourcen
- maschinelles Suchen und Einordnen von Ressourcen
- Ziel: einfaches Erzeugen, Austauschen und Auswerten von Metadaten

RDF-Beispiel

```
<? xml version="1.0" ?>
<RDF xmlns = "http://w3.org/TR/1999/PR-rdf-syntax-19990105#"
      xmlns:DC = "http://purl.org/DC#" >

  <Description about = "http://dstc.com.au/report.html" >
    <DC:Title> The Future of Metadata </DC:Title>
    <DC:Creator> Jacky Crystal </DC:Creator>
    <DC>Date> 1998-01-01 </DC>Date>
    <DC:Subject> Metadata, RDF, Dublin Core </DC:Subject>
  </Description>
</RDF>
```

RDF und DublinCore

- erweitert XML um DublinCore Namespace
- Möglichkeit der Einbettung der RDF-Daten in HTML-Seiten (HEAD)
- Features von RDF/XML ermöglichen Verbesserungen gegenüber reinem DublinCore Modell (xml:lang)
- Erweiterungen:
 - DCQ: Dublin Core Qualification Namespace
 - DCT: Dublin Core Terms Namespace
 - ermöglichen detailliertere Beschreibung der DC-Bezeichner

Einsatz von XML/RDF in MONARCH

- RDF in automatisch generierte `index.html` einbetten
- Archivierung von Publikationen als XML
- jedoch nicht für internes Metadaten-Format - SOIF
- RDF als Basis für Technologie „aggregierte Dokumente“ in MONARCH
- Indexseite von Publikationen in XML für Suchmaschinen ?

Aggregierte Dokumente in MONARCH

- Aggregat: (thematische) Gruppierung von Einzelpublikationen
- Gruppierung der Publikationen auf Metadaten-Ebene
- alle Komponenten bereits archiviert
- Zusammenstellung des Aggregats
- Archivierung des Aggregats (Metadaten-Satz) als eigene Publikation
- Generierung von SOIF aus Metadaten-Satz, eigener Publikationstyp
- Integration in Recherchesystem

Abschliessendes

Anmerkungen ?

Fragen ?