

XML für Nachrichten

Anwendungsfall RSS- Newsfeed

Florian Wilmshöver

Proseminar Auszeichnungssprachen

27.06.2006



Gliederung

- Begriffe - Was ist ein Newsfeed?
- Geschichte
- Struktur eines RSS Dokuments
- Proseminar RSS Beispiel
- Funktionen innerhalb von RSS
- Fazit



Begriffe

- Was versteht man unter einem Newsfeed?
,feed‘ zu deutsch etwa ,Einspeisung‘ oder
,Zuführung‘ also Nachrichtenzufuhr
- Erste Gedanken
 - RSS
 - Atom
- Beide werden meistens synonym für
Newsfeed verwendet



Richtig Schnelle Schlagzeilen ?

- Große Webseiten mit häufigen Updates als Anbieter
- Newsfeed beinhaltet die neuesten Informationen in einem XML-Exportformat
- Anwender kann selbst entscheiden wie und mit welcher Software er die Informationen verwenden möchte
(klassischer Newsreader, Onlinedienst, integriert in Browser oder E-Mail-Client)
- ‚schneller als selber surfen‘



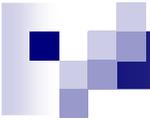
Was ist RSS?

- Definition Syndikation: “Regelmäßiger Austausch aktueller Informationen/Inhalte zwischen Anbieter und Anwender“
- RSS ist ein Syndikationsformat
- Wofür steht RSS:
 - Rich Site Summary (0.9x)
 - RDF Site Summary (1.0, 1.1)
 - Really Simple Syndication (2.0)



Geschichte von RSS(1)

- Erste Version wurde von Netscape für folgende Zwecke eingeführt
 - Beschreibung von Informationskanälen
 - Aggregation(Zusammentragen) von Inhalten
- Version 1.0 erhielt neue, komplexere und RDF-basierte Spezifikation zur Verknüpfung von RSS-Daten mit RDF-Meta-Daten anderer Dokumente (Semantic Web)



Geschichte von RSS(2)

- Version 1.0 (Nachfolger von 0.9) und 2.0 (Nachfolger von 0.91-0.93) sind unterschiedliche Abspaltungen der Ursprungsversion 0.9 und werden heute getrennt voneinander entwickelt
- 1.0 enthält RDF, 2.0 besteht aus reinem XML
- Verschiedene teilweise inoffizielle Weiterentwicklungen (Versionen 1.1 und 3.0)
- Das Weitverbreitetste Format ist heute RSS 2.0 ursprünglich von Dave Winer entwickelt



Struktur eines RSS 2.0 Dokuments(1)

- Klare Baumstruktur, da reines XML Dokument
- RSS 2.0 verlangt XML 1.0
- Je nach Verwendung des Feeds sind eventuell zusätzliche Namespacedeklarationen(Module) notwendig
- RSS Elemente selbst sind jedoch kein Teil eines Namespace (Abwärtskompatibilität)



Struktur eines RSS 2.0 Dokuments(2)

■ Strukturierende Elemente:

- <rss>: root-Element des Dokuments, enthält ein Versions-Attribut
- <channel>:
 - Container für Ressourcen des Dokuments (<item>s)
 - Enthält ausserdem Metadatenelemente die den gesamten Feed beschreiben
- <item>: beschreibt ein Element des Newsfeeds und kann zusätzliche Metadaten zu dem Element enthalten



Notwendige Elemente

- beschreibende Elemente (für den ganzen Channel oder für einzelne Items)
 - <title>: Titel des Dokuments
 - <link>: URL des Webseitenartikels
 - <description>: Beschreibung des Channels oder Nachrichtentext des <item> (z.B.)



Optionale Elemente (1)

■ Elemente für <channel>:

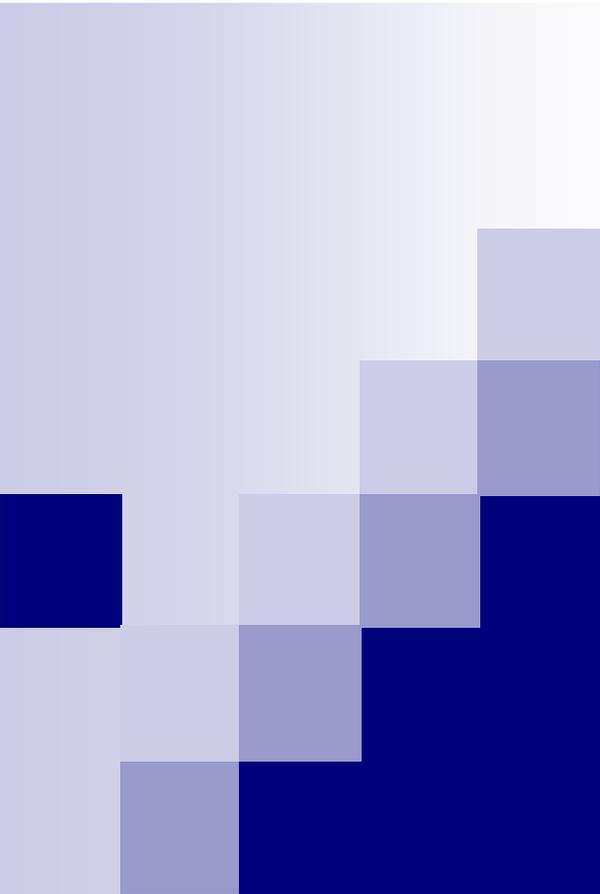
- <category>: Informatik
- <pubDate>: Tue, 27 Jun 2006 14:15:00 CET (RFC 822) - Veröffentlichungen des Inhalts
- <buildDate>: wie pubDate – beschreibt den Zeitpunkt der Generierung der RSS-Datei
- <language>: de (ISO-Kürzel)
- <ttd>: Time-To-Live in Minuten – Zeit bevor der RSS-Feed aktualisiert werden muss (nur für spezielle Anwendungen)



Optionale Elemente (2)

■ Elemente für <item>:

- <category> und <pubDate> wie bei <channel>
- <author>: E-Mail Adresse des Verfassers
- <source>: gibt einen Quell-RSS-Feed an
 - <source url=„http://url.zum.fe.ed“>Quell Feed</source>
- <guid>: eindeutiger String zur späteren Identifikation des Beitrags
 - isPermaLink=„True“
- <comments>: URL zu einer Diskussionseite für den Beitrag



Beispiel



Einbinden von Webservices in RSS

- Definition: Eine rssCloud-Schnittstelle implementiert ein rudimentäres ‚publish/notify‘-Protokoll mittels XML-RPC oder SOAP
- <cloud>: nur für <channel>-Element
Tag realisiert die rssCloud Schnittstelle



Benötigte Parameter von <cloud>

- Domain=„rss.nachricht.en“: gibt den Server an auf dem der entsprechende Webdienst läuft
- Port=„80“: entsprechender TCP-Port
- Path=„/pfadzumrpc“: Pfad zum XML-RPC-Dienst auf dem Server
- registerProcedure=“xmlNachrichten.rssBitteBenachrichtigen“
- Protocol=„xml-rpc“: Alternative hier wäre SOAP



Funktionsweise von <cloud>

- rssCloud muss beim Anbieter realisiert sein
- Newsreader registriert sich beim Anbieter
- Anbieter benachrichtigt nach Aktualisierung seiner Inhalte alle registrierten Newsreader, dass neue Inhalte zur Verfügung stehen
- Newsreader aktualisiert daraufhin den entsprechenden RSS-Feed
- Funktionsweise hat grobe Ähnlichkeit mit MVC-Architektur



Nutzen von Webservices für RSS

- Vermeidet nutzlose Aktualisierungen des selben Feeds
- ‚verteilte Anwendung‘ reduziert Traffic
- Hauptsächlich für Online-Newsreader geeignet, da der Newsreader auf geöffnetem TCP-Port auf Benachrichtigungen warten muss



Binärdaten innerhalb eines Feeds

- `<enclosure>`-Element ermöglicht anhängen von beliebigen Binärdateien
- `<enclosure>` darf nur einmal innerhalb eines `<item>`-Elements verwendet werden

Attribute:

- `url=„http://rss.dat.ei“` - ausschließlich http
- `length=„8192“` – Größe in Bytes
- `type=„audio/mpeg“` – MIME-Typ der eingebetteten Datei



Anwendungen von <enclosure>

- Eine der Bekanntesten Anwendung sind die sogenannten ‚Podcasts‘
- ‚Podcasts‘ sind lediglich RSS 2.0 Feeds mit eingebetteten mp3-Dateien
- Da beliebige Binärdaten eingebettet werden können sind auch ‚Video-Podcasts‘ oder ähnliches denkbar



Erweiterungen durch Module (1)

- RSS 2.0 bietet die Möglichkeit zusätzliche Elemente zu verwenden
- Nicht in der Spezifikation enthaltene Tags müssen in einem W3C-konformen XML-Namespace enthalten sein
- RSS 2.0-eigene Tags sind in keinem Namespace enthalten um Kompatibilität mit den Vorgängerversionen zu wahren



Modul Bittorrent

- Falls nur .torrent verfügbar:

```
<enclosure url=„link.zur.datei.torrent size=„1234“  
type=„x-application-bittorrent“>
```

„size“ ist in diesem Fall gröÙe der .torrent Datei

- Torrent-Datei als alternative Quelle:

```
<bitTorrent:torrent  
bitTorrent:url=„link.zur.datei.torrent“/>
```

zusätzlich zu normalem <enclosure>-Element



Erweiterungen durch Module(2)

- Vielzahl an verschiedenen Modulen
- Jeder Anbieter definiert eigenen, teils sehr umfangreichen Namespace
- Bei intensiverem Gebrauch von Modulen entsteht schnell Unübersichtlichkeit



Fazit

- Mit RSS 2.0 lassen sich schnell und ohne viel Aufwand Newsfeeds erstellen
- RSS lässt sich einfach anpassen
- RSS 2.0 Spezifikation ist in gewissen Details nicht eindeutig
 - generelle Freiheiten beim Verfassen von Newsfeeds
 - Module
 - Namespacedefinitionen
 - Zugehörigkeit der Elemente
- Atom soll hier in Zukunft Abhilfe schaffen und die Vagheiten von RSS beseitigen



Literatur

- Heinz Wittenbrink: *Newsfeeds mit RSS und Atom*, Galileo Computing 2005
- Dr. Cai Ziegler: *Semantic Web: RDF Anwendungen: FOAF, RSS, CC/PP und XUL*, iX 03/2006
- Moritz Sauer: *RSS: Nachrichten Appetizer für jedermann*, ct 12/03
- Jo Bager: *Podcasts: Audio-Inhalte online bereitstellen und empfangen* ct 14/05
- <http://blogs.law.harvard.edu/tech/rss>
- <http://de.wikipedia.org/wiki/RSS>
- <http://www.tagesschau.de>