Anwendungen für Nachrichten (RSS)

Proseminar Auszeichnungssprachen

26.06.2007

Olivier Dounla – Jan Knipp

Gliederung

- Was ist RSS?
- Wofür steht RSS ?
- Beispiele f
 ür die Nutzung von RSS
- Woran erkenne ich RSS im Internet ?
- Wie kann man sich RSS Inhalte anschauen ?
- Die Versionen von RSS im Überblick
 - RSS 0.90, RSS 0.91
 - RSS 1.0, RSS 0.92
 - RSS 2.0
- Elemente eines RSS 2.0 Dokuments
- Ein minimaler RSS 2.0 Feed
- RSS Feed zum Proseminar
- Transformation per XSLT
- Erweiterungen durch Namespaces
- Namespaces selber erstellen
- Wie bietet man RSS an ?
- Vorteile, Probleme & Nachteile von RSS
- Literatur

Was ist RSS?

- RSS ist ein plattformunabhängiges, auf XML basierendes Format, welches entwickelt wurde, um Nachrichten und andere Webinhalte auszutauschen.
- RSS erlaubt es Benutzern Online Content zu abonnieren und somit kommen die abonnierten Inhalte zum Benutzer, ohne das dieser dutzende von Webseiten besuchen muss.
- RSS Dateien sind logisch strukturiert, auf den Inhalt reduziert und zunächst frei von Layout und Design. Somit sind sie von Maschinen interpretierbar

Wofür steht RSS?

- Je nach Version hat RSS eine unterschiedliche Bedeutung:
 - Rich Site Summary
 - RDF Site Summary
 - Really Simple Syndication
- Wichtige Stichworte bei RSS sind Syndikation und Aggregation
 - Syndikation bezeichnet den regelmäßigen Austausch aktueller Informationen zwischen Websites (Content Syndication) und somit das Zugänglichmachen von Daten für weitere Übertragung, Aggregierung und Online-Publikation.
 - Aggregation bezeichnet die Vereinigung von mehreren Teilinformationen bzw. Newsfeeds.

Beispiele für die Nutzung von RSS

- Newsseiten (<u>www.cnn.com</u>, <u>www.tagesschau.de</u>, <u>www.reuters.com</u>)
 verwenden RSS, um tagesaktuelle Nachrichten, sortiert nach
 Themenbereichen zu veröffentlichen.
- Blogs und Podcasts verwenden RSS, um die Leser über neue Einträge zu informieren.
- Im Onlinemarketing wird RSS verwendet, um Kunden über neue Produkte zu informieren.
- In Bibliotheken können Benutzer Titeldaten über Schlagwörter abonnieren.

• ...

Woran erkenne ich RSS im Internet?



Standardisiertes Icon, welches von der Mozilla Foundation in Firefox verwendet wird und von Microsoft für den Internet Explorer 7 übernommen wurde.



Diverse Icons auf Webseiten um RSS Inhalte anzuzeigen

Wie kann man sich RSS Inhalte anschauen ? (1)

Newsreader

- Programme (FeedDemon, Feedreader, BottomFeeder, RSS Bandit, ...)
- Browser (Firefox, IE7, Opera, ...)
- Email-Programme (Thunderbird, Outlook, ...)
- Internetfähige Geräte (Mobiltelefone, Spielekonsolen, ...)

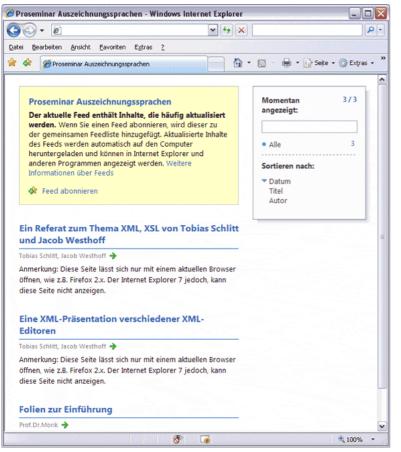
Feedverzeichnisse

z.B. http://www.newsisfree.com,
 ...

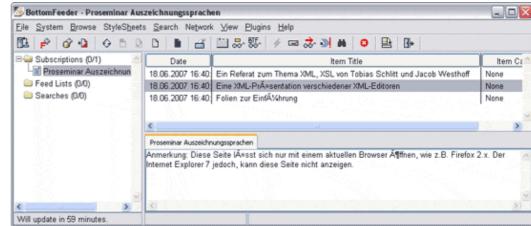
Suchmaschinen

http://www.feedster.com, http://www.daypop.com, ...

Wie kann man sich RSS Inhalte anschauen ? (2)







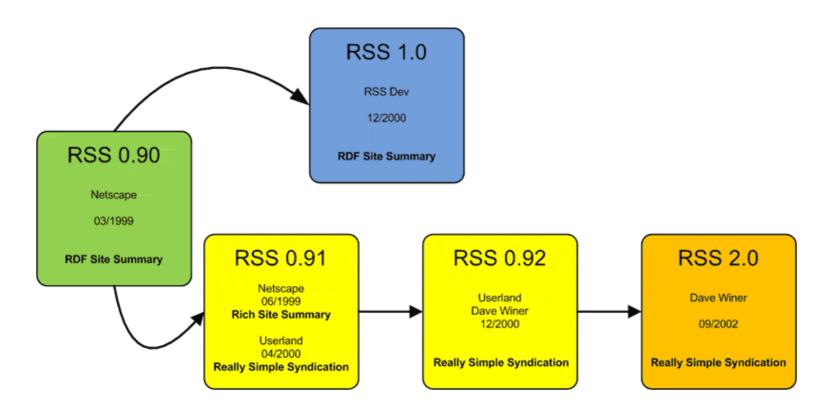
Wie kann man sich RSS Inhalte anschauen ? (3)







Und viele weitere ...



Nach der Entwicklung von RSS 0.90 erfolgte eine Aufspaltung in zwei Lager, in denen unterschiedliche Ziele verfolgt wurden:

- Dan Libby (RSS Dev): bessere Benutzung von RDF in RSS
- Dave Winer (Userland): Vereinfachung von RSS

Spezifikationen existieren parallel!

RSS 0.90

- März 1999 RDF Site Summary
- Von Netscape (Dan Libby) für MyNetscape entwickelt um Informationskanäle zu beschreiben und Inhalte zu aggregieren ("zusammenführen")
- Resource Description Framework (RDF) basiert, allerdings auf einem early working draft, welches nicht voll kompatibel zur finalen RDF Spezifikation ist

RSS 0.91

- Juli 1999 (Netscape) Rich Site Summary
- April 2000 (Userland) Really Simple Syndication
- RDF frei
- Einfache Dokumenttypdefinition auf XML Basis, somit erstmals Validierung gegen DTD möglich (Netscape)
- Nicht abwärtskompatibel zu RSS 0.90

RSS 1.0

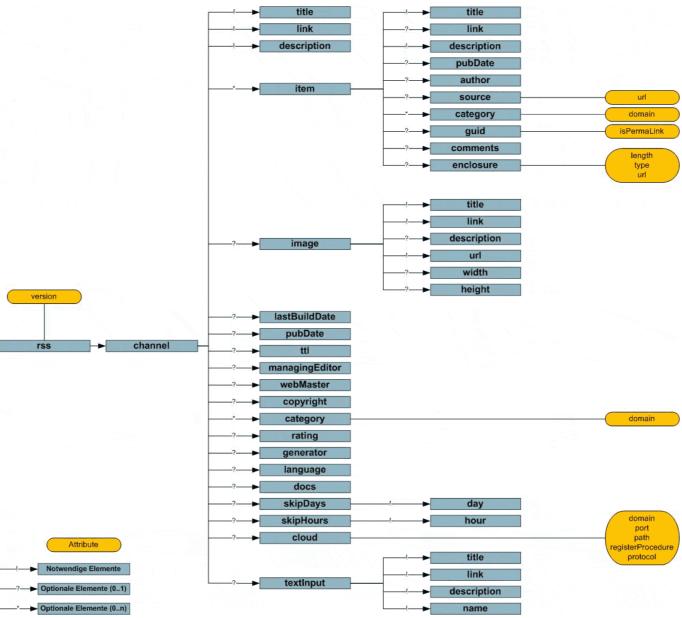
- Dezember 2000 RDF Site Summary
- Von einer nicht kommerziellen, unabhängigen Entwicklergruppe entwickelt (Rael Dornfest, O'Reilly u.a.)
- Beruht auf der finalen RDF Spezifikation und unterstützt XML Namespaces / Module
- Formal genauer spezifiziert als 0.91
- Nicht voll kompatibel zu RSS 0.90
- Inkompatibel zu RSS 0.91

RSS 0.92

- Dezember 2000
- Dave Winer / Userland
- Erweiterung von RSS 0.91

• RSS 2.0 [0.94]

- September 2002 Really Simple Syndication
- Veröffentlicht von Dave Winer
- Rechte wurden übertragen an die Harvard University
- Weiterentwicklung von RSS 0.91, ...
- Kleine Gruppe speziell definierter XML Elemente & Attribute
- Kein Schema verfügbar
- Muß XML 1.0 Deklaration des W3C entsprechen
- RSS 2.0 frozen (aktuell RSS 2.0.9)
- Elemente außerhalb der Spezifikation dürfen nur vorkommen, wenn sie in einem Namespace definiert wurden
- Elemente von RSS 2.0 sind aus kompatiblitätsgründen nicht in Namespaces definiert -> 0.91 oder 0.92 sind valide 2.0 Dateien



Strukturiende Elemente

Element	Funktion	Beispiele
<rss> @version</rss>	Rootelement des RSS-Dokuments Angabe der RSS-Versionsnummer	<rss version="2.0"> <channel> <item></item></channel></rss>
<channel></channel>	Sammlung von Informationsobjekten. Beschreibt und repräsentiert den ganzen Informationskanal. Beinhaltet beliebig viele <item> Elemente</item>	<item> </item>
<item></item>	Neues Informationsobjekt, repräsentiert die Teile des Informationskanals. Kann auf ein vorhandenes Informationsobjekt verweisen oder ein neues erschaffen.	

Element	Funktion	Beispiele
<title></td><td>Der Name des Channels
(Falls zu der Seite eine HTML Seite existiert mit
entsprechendem Inhalt, dann sollte <title> dem Titel der
Webseite entsprechen, HTML sollte möglichst nicht
verwendet werden)</td><td rowspan=4><rss version="2.0"> <channel> <title>Proseminar Auszeichnungssprachen </title> link>http://www <description>Mit dem Proseminar </description> <lastbuilddate> Tu, 26 Jun 2007 14:15:00 GMT </lastbuilddate> <item></item> 		
	Adresse des (i.d.R. HTML) Dokuments, dem der Channel entspricht	
<description></description>	Kurzbeschreibung des Feeds (HTML Markup muss geschützt werden z.B. CDATA Blocks)	
<lastbuilddate></lastbuilddate>	Zeitpunkt der letzten Änderung am Feed RFC 822 (http://asg.web.cmu.edu/rfc/rfc822.html)	

Element	Funktion	Beispiele
<pubdate></pubdate>	Datums- und Zeitangabe der Publikation RFC 822 (http://asg.web.cmu.edu/rfc/rfc822.html)	<rss version="2.0"> <channel></channel></rss>
<ttl></ttl>	"time to live" in Minuten Gibt an, wie lange sich der Inhalt des Feeds nicht verändern kann, so dass die Daten in dieser Zeit gecacht werden können	<pre>cyubDate>Tu, 26 Jun 2007 14:15:00 GMT <ttl>180</ttl> <managingeditor> xyz@def.com (John Doe) </managingeditor> <webmaster> abc@def.com (Jim Knopf) </webmaster> </pre>
<managingeditor></managingeditor>	E-Mail Adresse der für den Inhalt verantwortlichen Person, z.B. verantwortlicher Redakteur Bsp: Name@Example.com (Vorname Nachname)	
<webmaster></webmaster>	E-Mail Adresse der für technische Fragen zuständigen Person Bsp: Name@Example.com (Vorname Nachname)	

Element	Funktion	Beispiele
<copyright></copyright>	Copyrightangabe für den Inhalt des Feeds	<rss version="2.0"> <channel></channel></rss>
<category> @domain</category>	Zuordnung zu einer oder mehreren Kategorien zur Einordnung des Feeds Angabe eines Strings zur Identifizierung	<copyright> Uni Dortmund 2007</copyright> <category>UNI/INFORMATI K</category> <rating>adult</rating> <generator>My Hands 2.0</generator>
<rating></rating>	Einstufung des Feeds gemäß der Jugendfreiheit entsprechend der Platform for Internet Selection (PICS) (http://www.w3.org/PICS/)	
<generator></generator>	Angabe des Programms, welches zur Erzeugung des Feeds verwendet wurde	

Element	Funktion	Beispiele
<language></language>	Sprache des Feeds RFC 1766 (http://www.ietf.org/rfc/rfc1766.txt) http://cyber.law.harvard.edu/rss/languages.html	<rss version="2.0"> <channel></channel></rss>
<docs></docs>	Verweis auf die Dokumentation des Formats RSS 2.0 (<u>http://cyber.law.harvard.edu/rss</u>)	<pre> <language>de</language> <docs>http://cyber.law.harv ard.edu/rss</docs> <skipdays> <day>Saturday</day></skipdays></pre>
<skipdays> <day></day></skipdays>	Gibt an, an welchen Tagen der Feed nicht auf Aktualisierungen geprüft werden soll	<skiphours> <hour>4</hour> <hour>6</hour> </skiphours>
<skiphours> <hour></hour></skiphours>	Gibt an, in welchen Stunden der Feed nicht auf Aktualisierungen geprüft werden soll	

Element	Funktion	Beispiele
<pre><cloud> @domain @port @path @registerProcedure @protocol</cloud></pre>	Möglichkeit, den Feed entsprechend dem vom Userland eigenführten Protokoll zu subskribieren. Der angemeldete Newsreader wird bei Aktualisierung des Feeds sofort benachrichtigt. [publish/notify-Protokoll, wird implementiert über XML-RPC oder SOAP]	<rss version="2.0"> <channel> <cloud domain="www.example.co m" path="/SOAP" port="80" protocol="soap" registerprocedure="Notify Me"></cloud> </channel> </rss>
<image/>	Angabe eines Bildes, welches den Feed repräsentieren soll	
<textinput></textinput>	gibt dem Benutzer die Möglichkeit ein HTML Formular Eingabefeld Textobjekt an eine URL zu senden [Wird von den meisten Aggregatoren ignoriert!]	

Subelemente von <item>

Element	Funktion	Beispiele
<title></td><td>Der Titel des Informationsobjekts
(Auch hier sollte möglichst kein HTML verwendet
werden)</td><td rowspan=4><rss version="2.0"> <item> <title>Ein Referat</title> <link/> http://www <description> Ein Referat von XY über das Thema ABC </description> <pubdate> Tu, 26 Jun 2007 14:15:00 </pubdate> /item>		
	URL des Informationsobjekts	
<description></description>	(Kurz)beschreibung des Informationsobjekts (HTML Markup muss geschützt werden z.B. CDATA Blocks)	
<pubdate></pubdate>	Datums- und Zeitangabe der Publikation (Kann in der Zukunft liegen) RFC 822 (http://asg.web.cmu.edu/rfc/rfc822.html) [beim Fehlen gilt <pubdate> des Kanals]</pubdate>	

Subelemente von <item>

Element	Funktion	Beispiele
<author></author>	Autor des Dokuments, nur nötig bei unterschiedlichen Autoren, sonst <managingeditor> (i.d.R. Email Adresse, evtl. mit Namen des Autors in Klammern)</managingeditor>	<rss version="2.0"> <author>Jim@Knopf.de (Jim Kopf) </author> <source url="http://www.abc.de"/>F eeds <category>UNI/INFORMATI K</category> <guid ispermalink="true">http:// www.example.com/p534 </guid> </rss>
<source/> @url	Gibt den Namen des RSS Channels an, aus dem eine Nachricht entnommen wurde Die URI des Channels	
<category> @domain</category>	Zuordnung des Informationsobjekts zu einer oder mehreren Kategorien Angabe eines Strings zur Identifizierung	
<guid> @isPermalink</guid>	Eindeutiger Identifier eines Eintrags (Globally Unique Identifier) true(URI)/false	

Subelemente von <item>

Element	Funktion	Beispiele
<comments></comments>	URI, unter dem man Kommentare zu einem Eintrag findet	<rss version="2.0"> <channel> <item> <comments>http://example .com/p2423</comments> <enclosure length="1024" type="audio/mpeg" url="http://www.musik.com</td></tr><tr><td><enclosure></td><td>Verweis auf binäre Dateien
Basis für Podcasting</td><td>/musik.mp3"></enclosure> </item></channel></rss>
@length @type @url	Umfang der eingebetteten Dateien in Byte MIME-Typ der eingebetteten Datei URL der eingebetteten Datei	

Subelemente von <image>

Element	Funktion	Beispiele
<title></td><td>Bildbeschreibung ALT Attribut im HTML Tag </td><td><rss version="2.0"></td></tr><tr><td>k></td><td>URL die beim anklicken des Bildes geöffnet werden soll</td><td rowspan=5><irange> <title>Bild</title> <link/> http://www.bild.com/ bild.jpg <description>abc <width>123</width> <height>123</height> </description>		
<description></description>	TITLE Attribut im HTML Tag 	
<url></url>	URL einer GIF, JPG oder PNG Datei, die den Channel repräsentiert	
<width></width>	Breite des Bildes in Pixeln maximum = 144 default = 88	
<height></height>	Höhe des Bildes in Pixeln maximum = 400 default = 31	

Subelemente von <textInput>

Element	Funktion	Beispiele
<title></td><td>Das Label des Submit Buttons</td><td></td></tr><tr><td></td><td>URI eines CGI Skripts</td><td rowspan=3><rss version="2.0"> <textInput> <title></title> <link/> <description></description> <name></name>		
<description></description>	Erklärung der Funktion des <textinput> Feldes</textinput>	
<name></name>	Der Name des Textobjekts, welches an das CGI Skript weitergeleitet wird	

Ein minimaler RSS 2.0 Feed

```
<?xml version="1.0" encoding="utf-8"?>
<rss version="2.0">
 <channel>
  <title>Ein einfacher Feed</title>
  <link>http://example.org/index.html</link>
  <description>Der simpelste RSS 2.0 Feed</description>
  <item>
   <description>Sehr kurz</description>
  </item>
 </channel>
</rss>
```

RSS 2.0 Feed zum Proseminar

```
<?xml version="1.0" encoding="ISO-8859-1"?>
<?xml-stylesheet type="text/xsl" href="proseminar-rss.xsl" ?>
<rss version="2.0">
     <channel>
              <title>Proseminar Auszeichnungssprachen</title>
              <link>http://www-ai.cs.uni-dortmund.de/LEHRE/SEMINARE/PROSEMINAR/2007/index.html
              <description>Mit diesem Proseminar werden zwei Ziele verfolgt. Zum einen sollen Auszeichnungssprachen wie SGML und
              XML mit ihren Anwendungen für eLearning, Autorenumgebungen von im Internet publizierten Büchern und
              Wissensmanagement vermittelt werden. Zum anderen soll aber auch gezeigt werden, wie man sich in kurzer Zeit einen
              Überblick über ein neues Gebiet verschafft. Angesichts der Flut von Veröffentlichungen sehr unterschiedlicher Qualität fällt es
              oft schwer, sich selbstständig in ein Thema einzuarbeiten. Es gibt aber Techniken, die relevanten Zeitschriften und
              Konferenzen zu finden und die Struktur des Gebiets zu erfassen. Diese Techniken werden im Seminar eingeführt und
              geübt.</description>
              <language>de</language>
              <copyright>Uni Dortmund</copyright>
              <lastBuildDate>Tue, 26 Jun 2007 09:07:43 GMT</lastBuildDate>
              <generator>manual
              <docs>http://cyber.law.harvard.edu/rss</docs>
               <image>
                             <url>http://www-ai.cs.uni-dortmund.de/GRAFIKEN/ls8logo.gif</url>
                             <title>Lehrstuhl 8 - Proseminar Auszeichnungssprachen</title>
                             /www-ai.cs.uni-dortmund.de/LEHRE/SEMINARE/PROSEMINAR/2007/index.html
                             <description>LS 8 Logo</description>
                             <width>84</width>
                             <height>69</height>
              </image>
```

RSS 2.0 Feed zum Proseminar

```
<item>
                              <title>Business Process Execution Language</title>
                              <link>http://www-ai.cs.uni-dortmund.de/LEHRE/SEMINARE/PROSEMINAR/2007/bpel.pdf</link>
                              <description>Ein Referat über die BPEL</description>
                              <publi><pubDate>Tue, 19 Jun 2007 09:07:43 GMT</pubDate>
                              <author>Christian Vollmer, Oliver Garbe</author>
              </item>
               <item>
                              <title>Service Oriented Architecture</title>
                              <link>http://www-ai.cs.uni-dortmund.de/LEHRE/SEMINARE/PROSEMINAR/2007/soa.pdf</link>
                              <description>Ein Referat über die SOA</description>
                              <pubDate>Wed, 20 Jun 2007 09:07:43 GMT</pubDate>
                              <author>Torben Merhofe</author>
              </item>
     </channel>
</rss>
```

Transformation per XSLT

in HTML

- i.d.R. nur sinnvoll für den Einsatz im Browser
- Integration in Design der Webseite möglich
- Eigenbewertung der Elemente von RSS
- Verwenden von Namespace-Erweiterungen im eigenen Layout

Beispiel Proseminar

Erweiterung durch Namespaces

Itunes

- Ermöglicht das Einbinden von Podcasts in ITunes
- xmlns:itunes="http://www.itunes.com/dtds/podcast-1.0.dtd"
- Elemente : z.B. <itunes:owner/>, <itunes:category/>, <itunes:duration/>, <itunes:summary/>,<itunes:keywords/>, ...

Bittorrent

- Ermöglicht den automatischen Download von Bit-torrent Dateien, bzw. Informiert über neue Dateien
- xmlns:bt="http://lolifox.com/-/btrssx/"
- Elemente: z.B. <bt:infohash/>, <bt:descriptionuri/>, <bt:torrenturi/>, <bt:size/>, <bt:file/>, ...

Media RSS

- Erweitert die Einbindung von Medienobjekten in RSS und definiert neue Möglichkeiten Feeds mit Metadaten auszustatten
- xmlns:media="http://search.yahoo.com/mrss"
- Elemente : z.B. <media:content url="" filesize="" type="" medium="" ... framerate="" bitrate="" ...>,
 <media:adult/>, <media:rating/>, <media:keywords/>, <media:credits/>, ...

RSSTv

- Ermöglicht zeitgesteuerte Videoaufnahmen (Tivo, DIY PVRs)
- xmlns:tv="http://www.grumet.net/rsstv"
- Elemente : z.B. <tv:channel/>, <tv:start/>, <tv:stop/>, <tv:sub-title/> Oliver Dounla - Jan Knipp

Namespaces selber erstellen

- Alle nicht RSS Elemente gehören in einen Namespace!
- Nur sinnvoll, wenn Software vorhanden ist, die die neuen Sprachelemente verarbeiten kann
- Erweiterungselemente sollten nur Nachkommen von <channel> oder
 <item> sein, da Dokumente sonst fehlerhaft verarbeitet werden können
- Vorhandene Elemente werden nicht verändert, aber es kann ein "gleiches" Neues kreiert werden und dieses verändert werden
- Erweiterungen sollten nur da verwendet werden, wo es wirklich notwendig ist: less is more!
- Namespace URI sollte für die "Ewigkeit" sein
- Vernünftige Dokumentation und Beispiele sollten bei der Namespace URI hinterlegt sein
- Element- und Namespaceprefixe sollten verständlich sein, <qywc1:abcde>
 für Menschen unverständlicher als <auto:cabrio>

Wie bietet man RSS an?

- RSS Datei manuell schreiben
- RSS Datei mit einem RSS Editor (z.B. RSS Creator, Feedspring, ...) erzeugen
- RSS Feed durch Analyse der HTML Tags einer Webseite generieren (z.B. <u>RSS Wizard</u>, <u>FeedFire</u>, ...)
- RSS aus vorhandener XML Datei/Datenbank automatisch generieren (z.B. durch serverseitige XSLT Transformation, PHP, ...)
- Content Managment Syteme (CMS) verwenden, die automatisch RSS Feeds aus dem erstellten Inhalt erzeugen
- Evtl. Feed-Validatoren verwenden, um die Gültigkeit des Feeds zu prüfen (z.B. W3C Feed Validation Service, Feed Validator, ...)
- Auf den Webserver laden und auf der Seite verlinken
- RSS Feed in Suchmaschinen und Feedverzeichnisse eintragen

Vorteile von RSS

- Konzentration auf Inhalte
- Geringere Bandbreite
- Eindeutige Semantik
- Zeitersparnis
- Selbständige Nutzer
- Vergrößerte Informationsreichweite
- Aktualität
- Erleichterung
- Austausch und Weiterpublikation
- Auswertbar durch Maschinen
- Universales Publikationsformat
- "Spam & Phishing" nicht möglich -> hoher Glaubwürdigkeitsfaktor
- Benutzer kann für sich interessante Feed abonnieren und bei nichtgefallen einfach wieder kündigen
- Kann im Browser verwendet werden
- RSS Feeds können automatisch generiert werden

Probleme & Nachteile von RSS

 "RSS" is one of the coolest things you've never heard of when it comes to the Internet.

[MacMillan, Robert: Untangle the World Wide Web with RSS. 01/2007, Reuters, http://www.reuters.com/article/technologyNews/idUSN2919201420070102?sp=true]

- Spezifikation teilweise nicht eindeutig
- Bereitschaft für neue Technologien bei vielen Benutzern gering, bei Verwendung im Browser Optik evtl. zu "schlicht"
- (Noch) keine Personalisierung möglich
- Zahl der Abonnenten und tatsächlich Lesenden kaum feststellbar
- HTML in RSS wird nicht immer richtig dargestellt
- Viele Newsreader zeigen nur die Überschrift und eine kurze Beschreibung
- Evtl. erhöherter Server Load
- Evtl. Informationsüberladung

Literatur



Heinz Wittenbrink: Newsfeeds mit RSS und Atom.
 Galileo Computing. 2005

- http://www.rssboard.org/rss-specification
- •http://de.wikipedia.org/wiki/RSS
- http://www.w3schools.com/rss/
- •http://www.feedicons.com/
- http://www.vrtprj.de/
- •http://www.apple.com/de/itunes/store/podcaststechspecs.html
- http://rss-extensions.org

