

Semantic Web / web 2.0 Trends, Hypes und Realität

● Dr. Harald Sack
Institut für Informatik, FSU Jena

● Web Montag
[[Jena]]
18. September 2006

● TowerByte eG / c/o Interchen Tower

Semantic Web / web 2.0

Trends, Hypes und Realität

● Das World Wide Web – 4 Thesen

- „Das Web ist riesig“
 - > 55 Milliarden Web Seiten (25 Milliarden im Google Index - 07/2006)
- „ohne Hilfe findet man sich im Web nicht zurecht“
 - Suchmaschinen erschließen den Zugang ins WWW
- „Web-Inhalte sind für den Menschen bestimmt“
 - maschinelle Verarbeitung von Webinhalten birgt viele Probleme
- „Das Web hat sich verändert und wird sich weiter ändern“
 - (1) web 2.0
 - (2) Semantic Web



Dr. Harald Sack, Friedrich-Schiller-Universität Jena, Institut für Informatik

2

Semantic Web / web 2.0

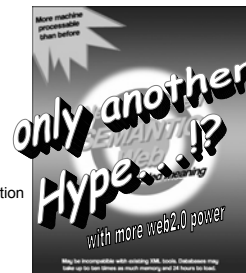
Trends, Hypes und Realität

Teil 1: web 2.0

- Wie sich das WWW verändert hat
- Partizipation
- Syndikation und Aggregation
- kollaborativ erstellte Metadaten als Bindeglied zum

Teil 2: Semantic Web

- Metadaten und Wissensrepräsentation
- WWW für Mensch und Maschine
- Vision des Semantic Web



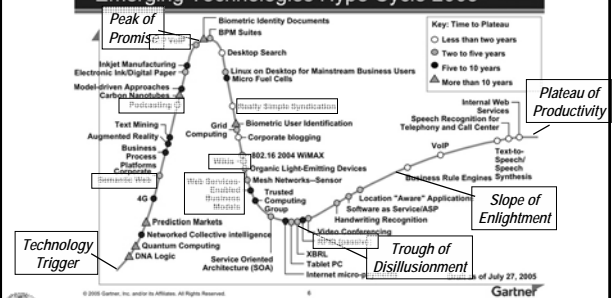
Dr. Harald Sack, Friedrich-Schiller-Universität Jena, Institut für Informatik

3

Semantic Web / web 2.0

Trends, Hypes und Realität

Emerging Technologies Hype Cycle 2005



Dr. Harald Sack, Friedrich-Schiller-Universität Jena, Institut für Informatik

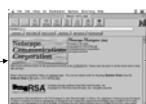
4

Wie sich das WWW verändert hat

1990 – Tim Berners Lee
Geburtsstunde des WWW



1992 – Lynx Browser



1993 – NCSA Mosaic
erster graphischer
Web-Browser



1995 – amazon / ebay
E-Commerce erobert das WWW

Dr. Harald Sack, Friedrich-Schiller-Universität Jena, Institut für Informatik

5

Web 1.0 - Kommerzialisierung des WWW

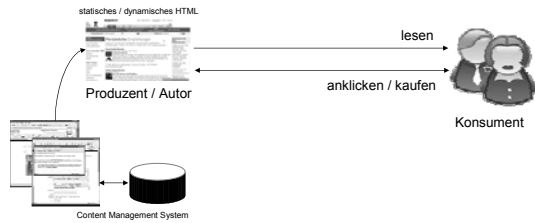
- **Merkmale:**
 - **Privatkommunikation**
 - **Broadcast**
 - **eingeschränkte Interaktivität**
- **Web 1.0** orientierte sich an der Funktionsweise der klassischen Massenmedien → **Broadcast, Einweg-Kommunikation**
- Nutzer ist von den übrigen Teilnehmern des WWW abgekapselt
- Nutzer als anonyme Masse, die lediglich
 - **lesen,**
 - **kaufen und**
 - **Werbekbanner anklicken sollen**

Dr. Harald Sack, Friedrich-Schiller-Universität Jena, Institut für Informatik

6

Web 1.0 - Kommerzialisierung des WWW

- Strikte Trennung zwischen **Produzent** und **Konsument** der angebotenen Information



Dr. Harald Sack, Friedrich-Schiller-Universität Jena, Institut für Informatik

7

Auf dem Weg zum Web 2.0

Neue Anwendungen verändern das Gesicht des WWW

„Reconquista“ oder „Renaissance“

Netzkultur vs. Massenkultur



Dr. Harald Sack, Friedrich-Schiller-Universität Jena, Institut für Informatik

8

Web 2.0 – Partizipation, Syndikation und Aggregation

- **Öffentliche Kommunikation, Soziale Netze, echte Interaktivität**
 - **Web 2.0** bietet dem Nutzer die Freiheit, Inhalte **aktiv mitzugestalten**
 - Nutzer ist **zugleich Konsument und Produzent** von Inhalten
 - Weblogs / Foren / Wikis / P2P-Filesharing...
 - Es bilden sich **soziale Netzwerke** (Communities)
 - Social Tagging Systeme, ...
 - Das **Web ersetzt den Desktop**
 - Web-basierte Mailbrowser / Kalender / Office-Systeme / ...

Dr. Harald Sack, Friedrich-Schiller-Universität Jena, Institut für Informatik

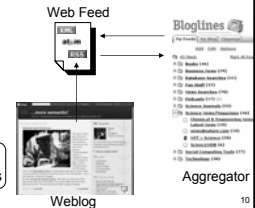
9

Web 2.0 – Partizipation, Syndikation und Aggregation

Weblogs ...

- sind Websites,
- basieren auf einfachen Content-Management-Systemen
- nutzen Basistechnologien (HTTP, URLs, RSS, XML, CSS)
- präsentieren Inhalte a-chronologisch geordnet
- publizieren Beiträge mit genauer Zeitangabe
- werden regelmäßig aktualisiert
- sind (stark) verlinkt mit anderen Weblogs
- sind thematisch universell
- bieten eine Kommentarfunktion
- sind eine Publikations-, Informations- und Kommunikationsform

Nachrichtensyndikation und Aggregation via RSS/Atom Web Feeds



Dr. Harald Sack, Friedrich-Schiller-Universität Jena, Institut für Informatik

10

Web 2.0 – Partizipation, Syndikation und Aggregation

Weblogs und Podcasts



Bertolt Brecht (Radiotheorie: 1927-1932)

„Der Rundfunk ist aus einem Distributionsapparat in einen **Kommunikationsapparat** zu verwandeln. Der Rundfunk wäre der denkbar großartigste Kommunikationsapparat des öffentlichen Lebens, ein ungeheures Kanalsystem, das heißt, er wäre es, **wenn er es verstünde, nicht nur auszusenden, sondern auch zu empfangen**, also den Zuhörer nicht nur zu hören, sondern auch sprechen zu machen und ihn nicht zu isolieren, sondern ihn auch in Beziehung zu setzen.“

Dr. Harald Sack, Friedrich-Schiller-Universität Jena, Institut für Informatik

11

Web 2.0 – Partizipation, Syndikation und Aggregation

Wikis

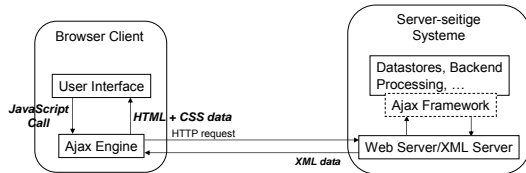
- Sammlungen von Webseiten (Hypermedia-Dokumenten), die von vielen Autoren (oft von jedermann) editiert werden können
- Ward Cunningham (1996) - wikiwiki = [hawaiianisch] schnell
- Webseiten lassen sich direkt beim Lesen über eine spezielle Bearbeitungsfunktion (mit spezieller, einfacher Syntax – Wiki-Syntax) im Browser verändern
- auch unerfahrene Nutzer können in Wiki einfach neue Seiten erstellen und mit bestehenden verlinken
- Seiten sind **nicht** an einem vorgegebene Struktur gebunden
 - **keine Threads** und **keine chronologische Sortierung** der Einträge

Dr. Harald Sack, Friedrich-Schiller-Universität Jena, Institut für Informatik

12

Web 2.0 – Partizipation, Syndikation und Aggregation

AJAX Web-Applikation



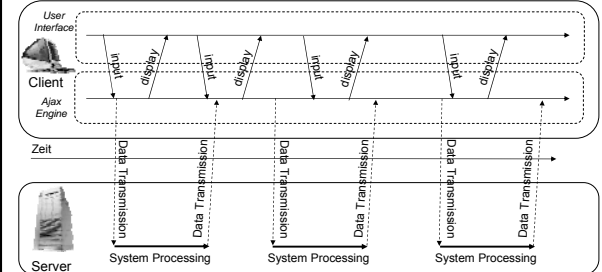
- **Ajax Engine** (in JavaScript) wird beim ersten Aufruf der Webseite geladen
 - zuständig für User Interface Design und Server Kommunikation
 - erlaubt asynchrone Kommunikation unabhängig vom Server

Dr. Harald Sack, Friedrich-Schiller-Universität Jena, Institut für Informatik

19

Web 2.0 – Partizipation, Syndikation und Aggregation

AJAX Web-Applikation



Dr. Harald Sack, Friedrich-Schiller-Universität Jena, Institut für Informatik

20

Die Grenzen des WWW



- was ist wichtig, was nicht?
- was ist Information, was ist Werbung?
- was bedeutet die Information?
- wie glaubhaft ist die Information?
- was gehört zusammen?
- was ist redundant?

Mensch verfügt über **Kontextwissen / Weltwissen** und kann daher Information (meist) richtig deuten...

Dr. Harald Sack, Friedrich-Schiller-Universität Jena, Institut für Informatik

21

Metadaten als Bindeglied zum Semantic Web

Metadaten

- „Metadaten sind Daten, die Informationen über andere Daten enthalten“ und „Metadaten werden benötigt, um in der Lage zu sein, einen bestimmten Zweck zu erfüllen (oder) ein bestimmtes Ergebnis zu erreichen“ (informelle Definition, Wikipedia)
- „Metadaten sind strukturierte, kodierte Daten, die Charakteristika informationstragender Entitäten beschreiben, zum Zweck der Identifikation, Recherche, Beurteilung und der Verwaltung der damit beschriebenen Entitäten.“ (Durell, 1985)
- „Metadaten sind maschinenlesbare Informationen über elektronische Ressourcen oder andere Dinge“ (W3C)



Metadaten werden herkömmlich vom Autor (einer Ressource) selbst oder von einem ausgewiesenen Experten **zentral** erstellt

Dr. Harald Sack, Friedrich-Schiller-Universität Jena, Institut für Informatik

22

Metadaten als Bindeglied zum Semantic Web

Metadaten

- Einfaches Beispiel: Bücher und bibliografische Metadaten



Identifikation über:

- Autor
- Titel
- ISBN / ISSN
- ...

Klassifikation über:

- Kategorien
- Schlüsselworte
- Abstract / Zusammenfassung
- ...

Bibliografie
Recherche

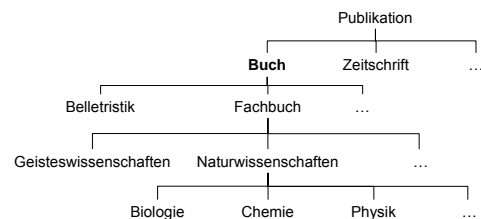
Dr. Harald Sack, Friedrich-Schiller-Universität Jena, Institut für Informatik

23

Metadaten als Bindeglied zum Semantic Web

strukturierte vs. unstrukturierte Metadaten

- z.B. **Taxonomien**



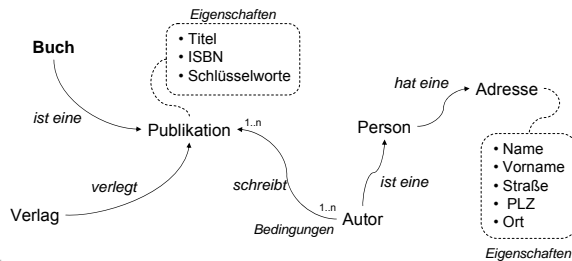
Dr. Harald Sack, Friedrich-Schiller-Universität Jena, Institut für Informatik

24

Metadaten als Bindeglied zum Semantic Web

strukturierte vs. unstrukturierte Metadaten

- z.B. **Ontologien**



Metadaten als Bindeglied zum Semantic Web

strukturierte vs. unstrukturierte Metadaten

- z.B. **Schlagworte (Keywords)**



XSLT RSA HTTP CSS
 Internet Huffman Code WLAN
 MPEG Informatik CGI SSL
 Java HTML Ethernet Kryptografie
 WWW XML MP3
 HDTV XML Netzwerke PHP
 TCP/IP

WWW für Mensch (und Maschine ?)

Metadaten im WWW

- Das WWW ist ausgelegt für die Nutzung durch den Menschen
- **HTML** ist nur geeignet zur Festlegung von Struktur / Formatierung
- **<meta>**-Tag im HTML Header zur Angabe von Autor, Description, Keywords (manipulierbar, kein Standard)

```

<meta name="description"
  content="Homepage of Dr. Harald Sack, ..." >
<meta name="keywords" content="homepage teaching ..." >
<meta name="Autor" content="Harald Sack">
    
```

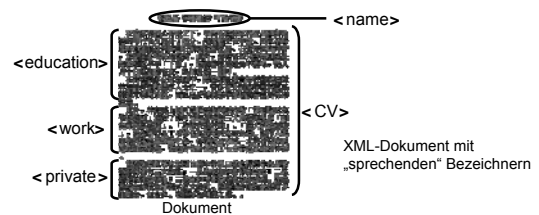
- **XML** erlaubt die Angabe „sprechender“ Bezeichner für semi-strukturierte Daten

```

<adresse>
  <name>
    <vorname>Harald</vorname>
    <name>Sack</nachname>
  </name>
</adresse>
    
```

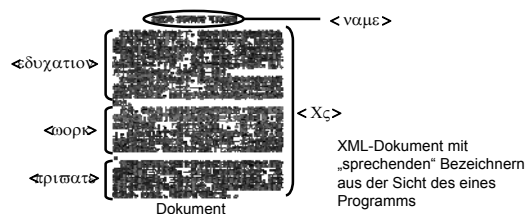
WWW für Mensch (und Maschine ?)

Interoperabilität von Metadaten im WWW



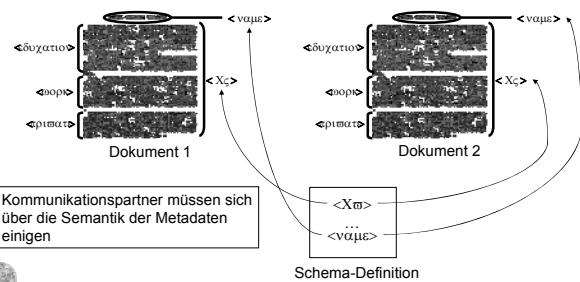
WWW für Mensch (und Maschine ?)

Interoperabilität von Metadaten im WWW



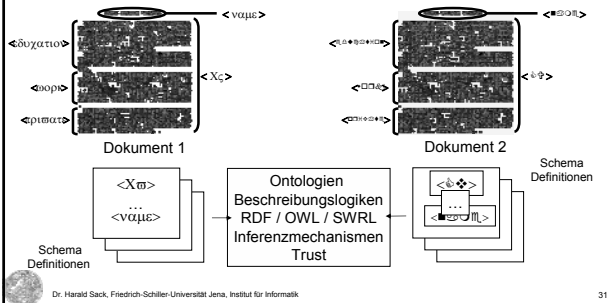
WWW für Mensch (und Maschine ?)

Interoperabilität von Metadaten im WWW



WWW für Mensch (und Maschine ?)

Interoperabilität von Metadaten im Semantic Web



WWW für Mensch (und Maschine ?)

Das Ende des WWW?

- → Das WWW ist zur **Nutzung durch den Menschen** bestimmt
- **Problemfelder:**
 - [1] Informationssuche
→ „wo finde ich...?“
 - [2] Informationsextraktion
→ „was ist wichtig?“
 - [3] Wartung
→ „sind meine Daten (noch) konsistent?“
 - [4] Personalisierung
→ „was mich interessiert...“

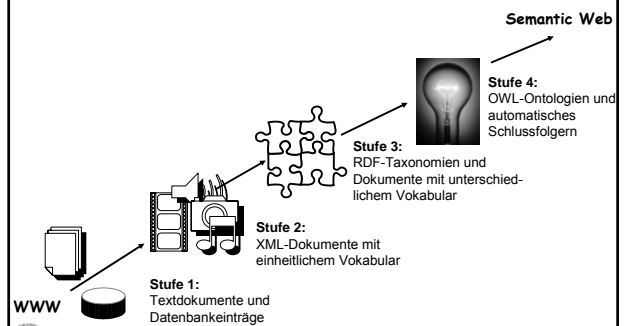
Semantic Web

Was ist das Semantic Web?

- „The Semantic Web is a web of data, in some ways like a global database.“
- „...a universal Web of semantic assertions“
- „...emphasises the intelligent manipulation of the knowledge that is made available via the web.“
- **„The Semantic Web is an extension of the current web in which information is given well-defined meaning, better enabling computers and people to work in cooperation“**

TBL, Semantic Web Roadmap, Sept 1998

Roadmap to the Semantic Web



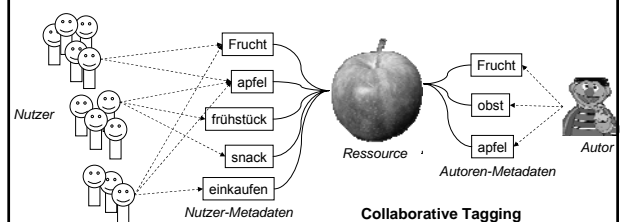
Die Vision des Semantic Web

Ziele des Semantic Web

- **Organisation des Wissens** in einzelne Domänen entsprechend seiner Bedeutung
- Automatisierte Werkzeuge zur **Wartung**, zur **Beseitigung von Inkonsistenzen** und zur **Extraktion neuen Wissens**
- Ersetzen der Schlüsselwortbasierten Suche durch
 - **Inhaltsbasierte Suche** über gezielte Frage-und-Antwort / Visualisierung und
 - benutzerfreundliche, **personalisierte Suche**
- **Extraktion und Präsentation von Wissen**
- Beantwortung von Fragen über verschiedene Dokumente hinweg
- Exakte Festlegung, wer genau welche Teile bestimmter Informationen sehen darf und wie vertrauenswürdig diese sind (**Trust**)

Web 2.0 vs. Semantic Web ?

Kollaborative erstellte Metadaten



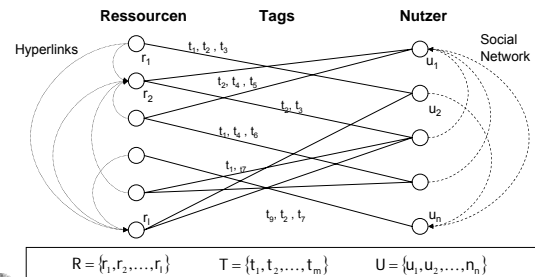
Web 2.0 vs. Semantic Web ?

Collaborative Tagging

- **Tags**
 - Nutzerdefinierte, deskriptive Metadaten (meist) in Form eines einzelnen Wortes (String)
- **Tagging**
 - Vorgang, in dem Nutzer deskriptive Metadaten zu eigenen und fremden Ressourcen zuordnen
- **Social (Collaborative) Tagging System**
 - erlaubt die gemeinsame Nutzung der verwendeten Tags
 - dabei dient jedes Tag als Link zu Ressourcen, die ebenfalls mit diesem Tag versehen sind

Web 2.0 vs. Semantic Web ?

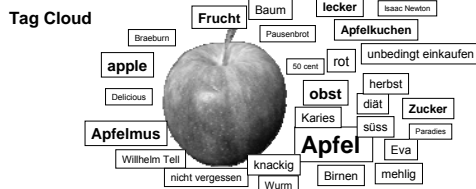
Collaborative Tagging Systems



Web 2.0 vs. Semantic Web ?

Folksonomies

- Nutzer-generierte Klassifikation (Konzept), die aus einer Übereinkunft innerhalb einer Nutzergruppe emergiert
- Sammlung von populären Tags für eine Ressource



Web 2.0 vs. Semantic Web ?

Folksonomies und Tag Clouds

- Tags können **gewichtet** werden
 - Anzahl
 - soziales Netzwerk
 - Reputation / Trust
 - Vorkommen in der damit ausgezeichneten Ressource

Tag Cloud einer Ressource r_k

$$T_k = \{t_i | t_i \in T \wedge \exists (u_i, r_k, t_i)\}$$

$u_i \in U, r_k \in R$

- Tag Cloud besitzt **keine Struktur** = flach
- Struktur kann über **Data Mining Techniken** gewonnen werden
 - z.B. co-occurrence

⇒ **Ontology Learning**

Fazit

- **Web 2.0** bindet den Nutzer partizipativ und interaktiv in die Kommunikationsstruktur des WWW mit ein
- Das **Semantic Web** soll es (autonom agierenden) Anwendungsprogrammen ermöglichen
 - **heterogene, über das Web zugreifbare Daten (Web Data + Web Services) miteinander in Beziehung zu setzen,**
 - diese Daten **gemeinsam zu nutzen (Synergieeffekte),**
 - und aus diesen **neue Informationen zu gewinnen (Inferenzmechanismen)**

⇒ Semantic Web und Web 2.0 können (und müssen) sich gegenseitig ergänzen