

Semantic Web

Sebastian Iwanowski
FH Wedel

Kap. 7:

Strategien zur Verbindung des Semantic Web mit dem World Wide Web

Quelle: Liyang Yu: A Developer's Guide to the Semantic Web (div. Kap.)

Ziel:

Maschinenlesbare semantische Infos auf normalen Webseiten

2 gebräuchliche Standards:

1) Microformats

- benutzt vCard-Ontologie für Personen. im Unterschied zu RDF eher objektorientiert
- Über das class-Attribut von HTML können eigene Unterklassen gebildet werden.

2) RDFa (**RDF** in HTML **a**tttributes)

- Beliebige RDF-Ontologie wird mit Hilfe von HTML-Attributen definiert.

W3C

- Die Prädikate müssen zu den HTML-Attributen rel, rev, property gehören.
- Das zum Prädikat gewählte HTML-Attribut legt fest, zu welchem Attribut das Subjekt und Objekt gehören müssen.

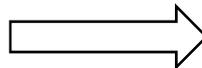
Ziel:

Maschinenlesbare semantische Infos auf normalen Webseiten

Standard 1: Microformat

- vCard ist nicht nur Ontologie, sondern auch **Syntax-Standard** 1996: älter als XML
- vordefinierte properties für personelle, adressale, geographische, organisatorische und kommentierende Infos
- **Integration in HTML über class-Attribut: Jede ontologische Entität ist ein eigenes HTML-Element**

Liyang Yu
Delta Air Lines
work 404.773.8994
1030 Delta Blvd.
Atlanta, GA 30354
USA
liyang.yu@delta.com



```
<div id="hcard-liyang-yu" class="vcard">
  <a class="n fn url" href="http://www.liyangyu.com">
    <span class="given-name">Liyang</span>
    <span class="family-name">Yu</span>
  </a>
  <div class="org">Delta Air Lines</div>


---


  <div class="tel">
    <span class="type">work</span>
    <span class="value">404.773.8994</span>
  </div>
  <div class="adr">
    <div class="street-address">1030 Delta Blvd.</div>
    <span class="locality">Atlanta</span>,
    <span class="region">GA</span>
    <span class="postal-code">30354</span>
    <div class="country-name">USA</div>
  </div>
  <a class="email" href="mailto:liyang.yu@delta.com">
    liyang.yu@delta.com
  </a>
</div>
```

Ziel:

Maschinenlesbare semantische Infos auf normalen Webseiten

Standard 2: RDFa

- Ontologien in RDF, Syntax in HTML
- Die Prädikate müssen Werte der HTML-Attribute rel, rev, property sein.
- rel-Prädikat: Subjekt zu about, Objekt zu href
- rev-Prädikat: Objekt zu about, Subjekt zu href
- property-Prädikat: Subjekt zu about, Objekt zu content

Bsp.:

```
<div xmlns:foaf="http://xmlns.com/foaf/0.1/">
  <a about="http://www.liyangyu.com#liyang"
    rel="foaf:homepage"
    href="http://www.liyangyu.com/">Liyang Yu</a>
  <span property="foaf:firstName" content="Liyang"/>
  <span property="foaf:lastName" content="Yu"/>
</div>
... <!-- other HTML code -->
```

RDFa-Anwendung soll die passenden Einträge „in der Nähe“ suchen.

Standard Ontologies

Dublin Core

`http://www.purl.org/metadata/dublin-core#`

(Präfix dc:)

- für Beschreibung von Dokumenten
- reiner RDF-Standard

Kap. 2.6

FOAF

`http://xmlns.com/foaf/spec/index.rdf`

(Standard)

`http://xmlns.com/foaf/0.1/`

(Präfix foaf:)

- für Beschreibung von sozialen Netzwerken
- OWL in RDF-Notation

Kap. 7

Focusing Search Engines with Semantic Markup

Kap. 8

Google: Rich Snippets

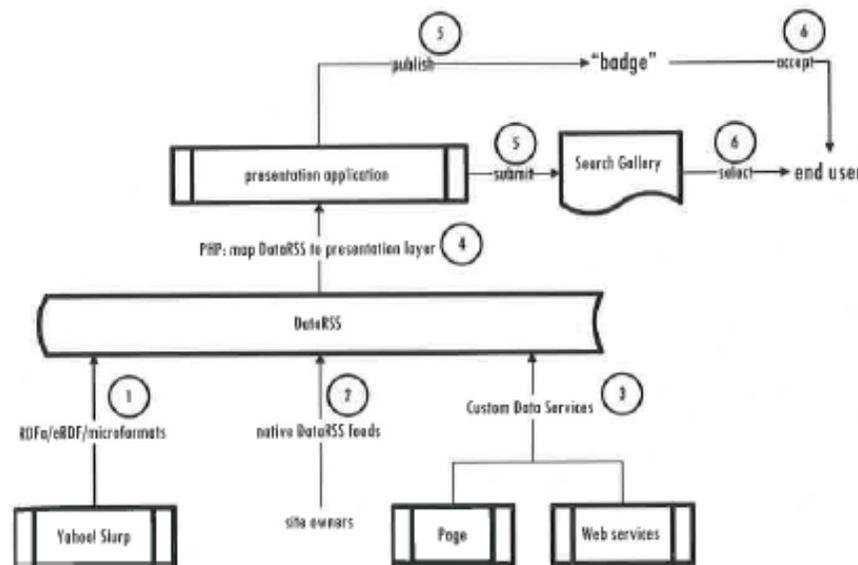
- **Markup über Microformat oder RDFa möglich**
- **Als Ontologien sind nur FOAF oder vCard zugelassen.**
- **Entsprechend angereicherte Webseiten werden automatisch in der Suche berücksichtigt. Die Darstellung der Ergebnisse kann aber nicht vom Webseitenanbieter beeinflusst werden.**

Focusing Search Engines with Semantic Markup

Kap. 8

Yahoo: Search Monkey

- unterstützt beliebige Ontologien
- Für eine Berücksichtigung in der Suchmaschine muss der Webseitenanbieter einige Einstellungen vornehmen. Dafür kann er die Darstellung der Ergebnisse auch beeinflussen.



Beeinflussung der Darstellung

Cache für Yahoo Search Engine

Alternative Zugangswege

3 grundsätzliche Strategien:

1) Manuelles Markup

- Eine ausgewählte Domäne von Webseiten wird systematisch mit Markup-Infos versehen.
- Für die Webseiten werden per Hand Ontologien erstellt.

Bsp.: Semantic Wiki

2) Automatisches Markup

- Webseiten mit einer vordefinierten Struktur werden automatisch mit semantischen Informationen in einer vordefinierten Ontologie versehen.

Bsp.: DBpedia

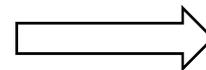
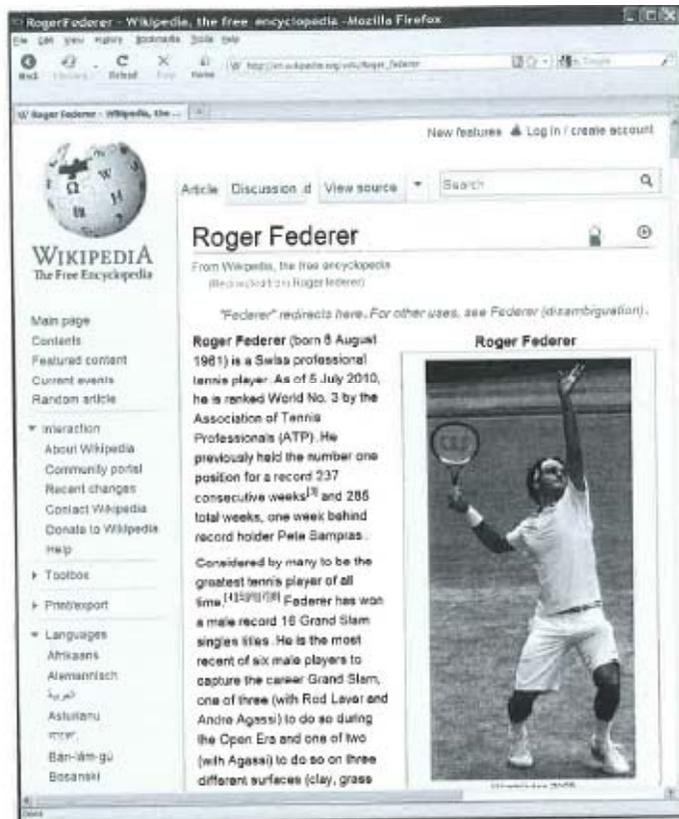
3) Neues maschinenlesbares Web

- „Projekt“ Open Linked Data

2) Automatisches Markup: Projekt DBPedia

<http://wiki.dbpedia.org/Ontology>

- automatisches Markup für Wikipedia
- nutzt festgelegte Annotationen von Wikipedia-Seiten aus:



```
{{Infobox Tennis player
|playername = Roger Federer
|image = [[File:Roger Federer (26 June 2009, Wimbledon) 2
new.jpg|200px]]
|caption = Wimbledon 2009
|country = [[Switzerland]]
|nickname= 'Swiss Maestro' '<ref>...</ref><br />
'Federer Express' '</br>
'Fed Express' '</br>
'FedEx' '<ref>...</ref><br/>
|residence = [[Wollerau]], [[Switzerland]]
|datebirth = {{birth date and age|df=yes|1981|08|08}}
|placebirth = [[Basel]], [[Switzerland]]
|height = {{height|m=1.86}}
|weight = {{convert|85.0|kg|lb st|abbr=on}}<ref>...</ref>
|turnedpro = 1998<ref>...</ref>
|plays = Right-handed; one-handed backhand
|careerprizemoney = [[US$]] 53,362,068<br />+ [[ATP Tour
records#Earnings|All-time leader in earnings]]
|singlesrecord = 678-161 (80.8%)<ref>...</ref>
|singlestitles = 61
|highestsinglesranking = No. ' ' '1' ' ' (February 2, 2004)
|currentsinglesranking = No. ' ' '1' ' ' (July 6, 2009)
|AustralianOpenresult = ' ' 'W' ' ' {
[[2004 Australian Open - Men's Singles|2004]],
[[2006 Australian Open - Men's Singles|2006]],
[[2007 Australian Open - Men's Singles|2007]]}
|FrenchOpenresult = ' ' 'W' ' ' {
[[2009 French Open - Men's Singles|2009]]}
|Wimbledonresult = ' ' 'W' ' ' {
[[2003 Wimbledon Championships - Men's Singles|2003]],
[[2004 Wimbledon Championships - Men's Singles|2004]],
[[2005 Wimbledon Championships - Men's Singles|2005]],
[[2006 Wimbledon Championships - Men's Singles|2006]],
[[2007 Wimbledon Championships - Men's Singles|2007]],
[[2009 Wimbledon Championships - Men's Singles|2009]]}
|USOpenresult = ' ' 'W' ' ' {
```

10.1 Roger Federer's wiki page in Wikipedia

2) Automatisches Markup: Projekt DBPedia

- automatisches Markup für Wikipedia <http://wiki.dbpedia.org/Ontology>
- Jede Wikipediaseite bekommt URI: <http://dbpedia.org/resource/pageName>
- DBPedia stellt für Elemente der Infobox von Wikipedia Entitäten bereit und transformiert die Inhalte automatisch in Individuals von vordefinierten Klassen

Table 10.1 Mapping Wikipedia's infobox templates to classes defined in DBpedia ontology

Wikipedia infobox template	DBpedia class mapping	Example page
Tennis player	<code>dbclass:TennisPlayer</code>	Roger Federer
Officeholder	<code>dbclass:officeHolder</code>	Bill Clinton
Person	<code>dbclass:Person</code>	Tim Berners-Lee
German Bundesland	<code>dbclass:City</code>	Berlin
Film	<code>dbclass:Film</code>	Forrest Gump
Company	<code>dbclass:Company</code>	Ford Motor Company
University	<code>dbclass:University</code>	Tsinghua University

prefix dbclass: <<http://dbpedia.org/ontology/>>

- DBPedia stellt für Elemente der Infobox von Wikipedia Entitäten bereit und transformiert die Inhalte automatisch in Individuals von vordefinierten Klassen
- Abfragen werden durch Anbindung an SPARQL unterstützt.

3) Neues maschinenlesbares Web: Linked Open Data

- definiert durch Minimal-Standards, die erfüllt werden müssen (wohl nicht genormt)
- unterscheidet zwischen Informationsquellen und Nichtinformationsquellen
Webseiten: `application` Konzepte: `rdf+xml`
- Jede Entität hat eine Webseitendarstellung und eine RDF-Darstellung (mit verschiedenen URIs)
- Anfragen müssen automatisch an die passende URI weitergeleitet werden

2 alternative Techniken: 303 URIs oder Hash URIs

Offene Frage: Wie verbreitet ist dieser Standard ?