

Semantic Web – Motivation und Entstehungsgeschichte

Tobias Beidermühle

Seminar „*Service-orientierte Architekturen*“

6. Dezember 2006

Begriffseinordnung „Semantic Web“

- *Semantic Web...*
 - ein Konzept, eine Vision, wie das vorhandene *World Wide Web* erweitert und in Zukunft genutzt werden soll.
 - ist eine Initiative, diese Vision in die Realität umzusetzen.
 - ist nicht gleichbedeutend mit „semantisches Netz“.
 - zielt auf die Erschaffung eines konkreten semantischen Netzes.

Inhaltsverzeichnis

1. Begriffseinordnung *Semantic Web*
2. Motivation
 - Probleme bei Nutzung des *World Wide Web*
 - Beispiele für Nutzung des *Semantic Web*
3. Hintergründe
 - Semantik
 - Zielsetzung für *Semantic Web*
4. Entstehungsgeschichte
 - WWW & W3C
 - META & PICS
 - XML
 - CDF & MCF
 - RDF
 - *Semantic Web Roadmap*
 - *Semantic Web Start*
 - DAML, OIL & DAML+OIL
 - OWL
 - RIF
5. Fazit
 - „*The Semantic Web Revisited*“
 - Kritik
 - Beispiele

Motivation – Probleme des World Wide Web

„Todays Web content is suitable for *human* consumption.“

– „A Semantic Web Primer“, Grigoris Antoniou, Frank van Harmelen, April 2004

- Probleme bei Suchmaschinen
 - zu viele Ergebnisse
 - zu wenig Ergebnisse
 - Informationen über verschiedene Seiten verteilt
 - sensibel auf Vokabular: Wortwahl beeinflusst Ergebnis
 - Beispiel: Suche nach „Haus“ + „Bau“, „Hausbau“ oder „Häuserbau“
 - Google: 13 Millionen Treffer für „Hausbau“
 - Wikipedia: kein Artikel „Hausbau“, dafür aber „Wohnungsbau“
 - sprachen-abhängig

Motivation – Beispiele für Semantic Web

„[At the doctor's office, Lucy called her brother Pete:] 'Mom needs to see a specialist and then has to have a series of physical therapy sessions. Biweekly or something. I'm going to have my agent set up the appointments.' [...]

Lucy instructed her Semantic Web agent through her handheld Web browser. The agent promptly retrieved information about Mom's *prescribed treatment* from the doctor's agent, looked up several lists of *providers*, and checked for the ones *in-plan* for Mom's insurance within a *20-mile radius* of her *home* and with a *rating* of *excellent* or *very good* on trusted rating services.

It then began trying to find a match between available *appointment times* (supplied by the agents of individual providers through their Web sites) and Pete's and Lucy's busy schedules. In a few minutes the agent presented them with a plan.“

– „*The Semantic Web*“, Tim Berners-Lee, James Hendler, Ora Lassila
erschienen in „*Scientific American*“, Mai 2001

Motivation – Beispiele für Semantic Web

- *Personal Agents*
- *Knowledge Management* im Unternehmen
 - Wissen ist ein *Asset* für ein Unternehmen.
 - „internes *Semantic Web*“ zum Speichern, Verwalten und vor allem Wiederverwenden von Wissen („*Knowledge Discovery*“, „*Knowledge Retrieval*“)
- *B2C e-commerce*
 - automatisiertes Vergleichen und Filtern von
 - Preisen, Versandbedingungen, AGB
 - „*reputation*“ über „*independent rating agencies*“
 - Ziel: „*automated negotiations*“

Hintergründe – Semantik

- **Kommunikation**: Übertragung von Symbolen
- **Information** ergibt sich, wenn beide Seiten einer Kommunikation die Bedeutung der Symbole kennen.
- **Semantik**
 - ergibt sich beim Menschen aus der Verknüpfung eines Symbols mit einem Konzept, das etwas beschreibt.
 - Beispiel: Menschen verknüpfen das Symbols „rote Ampel“ mit dem Konzept „Stehen-Bleiben“.
 - wird „übertragen“, wenn Symbole auf gleiche Weise interpretiert werden.
- **World Wide Web**
 - für Menschen: Kommunikations- und Informationsmedium
 - für Maschinen: **nur** Kommunikationsmedium
 - Maschinen können miteinander kommunizieren, aber sich nicht informieren, da sie Daten nicht interpretieren können.

Hintergründe – Semantik

- **unbekannte** oder **mehrdeutige** Symbole
 - können nicht interpretiert werden (weder von Mensch noch Maschine)
 - aber: Menschen können meist eine Verbindung zu einem bekannten Symbol herstellen oder nachfragen, wie es zu interpretieren ist.
- Fragestellung: Wie können Maschinen herausfinden, wie sie Daten zu interpretieren haben?
 - erster Ansatz: *text-processing of Web content*
 - basiert auf künstlicher Intelligenz und Computerlinguistik
 - bisher wenig erfolgreich bzw. zu aufwändig
 - „I am a professor of computer science.“
 - „I am a professor of computer science, you may think. Well ...“
 - zweiter Ansatz: *easily machine-processable Web content*
 - Web wird durch maschinen-verarbeitbare und -verständliche Daten ergänzt.

Hintergründe – Zielsetzung für Semantic Web

- *Semantic Web*
 - Versuch, Maschinen zum „Verstehen“ von Sachverhalten zu bringen
 - Ziel: Maschinen sollen selbstständiger handeln, Menschen weniger ausprogrammieren müssen.
 - Herausforderung: Semantik so eindeutig formulieren, dass menschliche Interpretation verzichtbar wird.

Inhaltsverzeichnis

1. Begriffseinordnung *Semantic Web*
2. Motivation
 - Probleme bei Nutzung des *World Wide Web*
 - Beispiele für Nutzung des *Semantic Web*
3. Hintergründe
 - Semantik
 - Zielsetzung für *Semantic Web*
4. Entstehungsgeschichte
 - WWW & W3C
 - META & PICS
 - XML
 - CDF & MCF
 - RDF
 - *Semantic Web Roadmap*
 - *Semantic Web Start*
 - DAML, OIL & DAML+OIL
 - OWL
 - RIF
5. Fazit
 - „*The Semantic Web Revisited*“
 - Kritik
 - Beispiele

Entstehungsgeschichte – WWW & W3C

- März 1989: *"Information Management: A Proposal"*, Tim Berners-Lee, CERN
- 1989-1992: Entwicklung von HTML und Web-Browsers am CERN
- 1993: CERN erlaubt freie Verwendung von Web-Protokollen
- 1993/1994: Tim Berners-Lee und andere Wissenschaftler erwägen Gründung eines Gremiums für WWW
- Sept. 1994: First „*International World Wide Web Conference*“, CERN
 - Vortrag von Tim Berners-Lee, darin u.a. *„The Need of Semantics in the Web“*
 - Gründung des W3C wird angekündigt
- Okt. 1994: Gründung des W3C

Entstehungsgeschichte – META & PICS

- **Juni 1993:** „HTML“ (Internet Draft), Tim Berners-Lee, Daniel Connolly
 - keine Meta-Daten vorgesehen
- **Aug. 1995:** „Platform for Internet Content Selection (PICS)“
 - Motivation: „Labeln“ von Web Content, um Selbstkontrolle zu ermöglichen und „Zensur“ durch US-Gesetzgebung zu verhindern (speziell: Filtern von jugendgefährdenden Inhalten).
- **Nov. 1995:** „HTML 2.0“ (Standard), Tim Berners-Lee, Daniel Connolly
 - **META**-Element definiert
 - „to provide a means to discover that the data set exists and how it might be obtained or accessed; and
 - to document the content, quality, and features of a data set, indicating its fitness for use.“
 - <http://www.ietf.org/rfc/rfc1866.txt>
- **Okt. 1996:** „PICS 1.1“ wird *W3C Recommendation*
 - „self-rating, third-party rating & easy-to-use“
 - <http://www.w3.org/PICS/principles.html>
 - integriert über HTML-**META**-Elemente

Entstehungsgeschichte – XML

„XML is the ASCII of the future.“

– „XML Files“, Tim Bray, Oktober 1998,
<http://www.w3.org/2000/01/ala2349/all>

- **Nov. 1996:** „XML“, First W3C Working Draft

Anmerkungen:

- **Dez. 1997:** „XML“, W3C Proposed Recommendation
- **Jan. 1998:** „Namespaces in XML“, Note
- **Feb. 1998:** „XML“, W3C Recommendation (First Edition)
- **März 1998:** „Namespaces in XML“, First W3C Working Draft
- **Jan. 1999:** „Namespaces in XML“, W3C Recommendation

Entstehungsgeschichte – CDF & MCF

- **März 1997:** „*Channel Definition Format (CDF)*“, Castedo Ellerman
 - „CDF is an open specification that permits a web publisher to offer frequently updated collections of information, or channels, from any web server for automatic delivery to compatible receiver programs on PCs or other information appliances.“
 - <http://www.w3.org/TR/NOTE-CDFsubmit.html>
 - XML-basiert
 - *Submission request to W3C* von Microsoft
- **Juni 1997:** „*Meta Content Framework (MCF)*“, Ramanathan V. Guha
 - „specification for a data model for describing information organization structures (metadata) for collections of networked information.“
 - <http://www.w3.org/TR/NOTE-MCF-XML/>
 - XML-basiert
 - *Submission request to W3C* von Netscape

Entstehungsgeschichte – RDF

- **Okt. 1997:** „*Resource Description Framework (RDF)*“, *Working Draft*
 - W3C-eigenes Projekt
 - Nachfolger von PICS
 - XML-basiert

Anmerkungen:

- **Feb. 1999:** „*RDF*“, *W3C Recommendation*, Ora Lassila, Ralph R. Swick
- **Feb. 2004:** „*RDF Primer*“, *W3C Recommendation*
 - besteht mittlerweile aus mehreren Spezifikationen:
 - „*RDF Concepts and Abstract Syntax*“
 - „*RDF/XML Syntax, Specification*“
 - „*RDF Vocabulary Description Language 1.0: RDF Schema*“
 - „*RDF Semantics*“
 - Arbeit der *RDF Core Working Group* damit abgeschlossen
 - *RDF Interest Group* wird eingestellt.

– <http://www.w3.org/RDF/>

Entstehungsgeschichte – Semantic Web Roadmap

- Sept. 1998: „Semantic Web roadmap“, Tim Berners-Lee
 - erste Publikation mit Begriff „Semantic Web“
 - erste Vorstellung einer Architektur („layer-cake“):
 1. „basic assertion model“
 2. „The Schema Layer“
 3. „Conversion language“
 4. „The logical layer“
 5. „Digital Signature“
- <http://www.w3.org/DesignIssues/Semantic.html>
– <http://www.w3.org/DesignIssues/Architecture.html#SemanticWeb>

Entstehungsgeschichte – Semantic Web Start

- **Jan. 2000:** „*Seeds of the Semantic Web*“, Dan Connolly,
Ala Midwinter Meeting, San Antonio
 - “Where we are: on the verge of another revolution [...]”
 - The RDF Revolution: rise of the *Semantic Web*“

– <http://www.w3.org/2000/01/ala2349/all>
- **Dez. 2000:** „*Semantic Web on XML*“, Tim Berners-Lee,
XML 2000 Conference, Washington DC
 - erste Präsentation zu *Semantic Web*
 - Konkretisierung der Techniken, Layers
- **2001:** „Initiative ‚*Semantic Web*‘ startet“
 - W3C „*Semantic Web Activity*“ (ersetzt „*Metadata Activity*“)
 - „*Semantic Web Interest Group*“ (ersetzt „*RDF Interest Group*“)
 - Artikel „*The Semantic Web*“ im „*Scientific American*“
 - diverse weitere Vorträge & Präsentation

– <http://www.w3.org/2001/sw/EO/talks>

Entstehungsgeschichte – DAML, OIL & DAML+OIL

- **Okt. 2000:** „DAML-ONT“, *Initial Release*
 - „effort to help bring the ‘semantic web‘ into being, focusing on the eventual creation of a web logic language.“
– <http://www.daml.org/2000/10/announcement.txt>
 - „*DARPA Agent Markup Language Program*“
 - 22 US-Unternehmen, -Universitäten (auch MIT/W3C)
- **Nov. 2000:** „*The Ontology Inference Layer (OIL)*“, *Informal Description*
 - „Ontologies provide a shared and common understanding of a domain that can be communicated between people and across application systems“
– <http://www.ontoknowledge.org/oil/TR/oil.long.html>
 - fünf europäische Universitäten und Institute
- **Jan. 2001:** „*DAML+OIL*“
 - „joint EU/US committee [...] to examine the DAML-ONT language release and produce a more stable version, as well as making it more consistent with the European OIL project.“
– <http://lists.w3.org/Archives/Public/www-rdf-logic/2001Jan/0041.html>

Entstehungsgeschichte – OWL

- **Nov. 2001:** W3C startet „*Web Ontology Working Group*“
 - Teil der *W3C Semantic Web Activity*
 - Nachfolger von DAML+OIL
- **Juli 2002:** „*Web Ontology Language (OWL)*“, *W3C Working Draft*
 - „OWL facilitates greater machine readability of web content than XML, RDF, and RDF-S support by providing a additional vocabulary for term descriptions.“
 - <http://www.w3.org/TR/2002/WD-owl-features-20020729/>
- **Feb. 2004:** „*OWL*“, *W3C Recommendation*
- **Mai 2004:** W3C schließt „*Web Ontology Working Group*“
 - „We have achieved all that we were chartered to do, and I believe our work is being quite well appreciated.“, Jim Hendler
 - <http://www.w3.org/2001/sw/WebOnt/>

Entstehungsgeschichte – RIF

- **April 2005:** W3C organisiert Workshop
 - „bring together rule system vendors, rule users [...] to work toward developing a standard rule language for the Web.“
- **Nov. 2005:** W3C startet „*Rule Interchange Format (RIF) Working Group*“
 - Teil der *Semantic Web Activity*
- **März 2006:** „*RIF Use Cases and Requirements*“, *First Working Draft*
 - „Synthesized from nearly fifty use cases, the document specifies use cases and requirements for a format that allows rules to be translated between rule languages and thus transferred between rule systems.“ – <http://www.w3.org/2005/rules/>
- **Mai 2007:** „RIF Phase 1“ Recommendation
 - Core für RIF soll fertiggestellt werden

Inhaltsverzeichnis

1. Begriffseinordnung *Semantic Web*
2. Motivation
 - Probleme bei Nutzung des *World Wide Web*
 - Beispiele für Nutzung des *Semantic Web*
3. Hintergründe
 - Semantik
 - Zielsetzung für *Semantic Web*
4. Entstehungsgeschichte
 - WWW & W3C
 - META & PICS
 - XML
 - CDF & MCF
 - RDF
 - *Semantic Web Roadmap*
 - *Semantic Web Start*
 - DAML, OIL & DAML+OIL
 - OWL
 - RIF
5. Fazit
 - „*The Semantic Web Revisited*“
 - Kritik
 - Beispiele

Fazit – *The Semantic Web Revisited*

Aussagen aus *“The Semantic Web Revisited”*:

- Voller Nutzen erst wenn alle Standards “well established” sind
- Standards werden kontinuierlich entwickelt und werden schon genutzt.
- Ontologien in bestimmten Anwendungsgebieten werden entwickelt, z.B. *e-science*.
- Kosten für Ontologien-Entwicklung sinken mit der Zeit.

Problematiken:

- URI-Konzept häufig noch nicht umgesetzt.
- Ontologien ändern sich über die Zeit („*Ontology management*“).

In Zukunft:

- Immer mehr Daten werden veröffentlicht: „collaborative science, commercial incentives, regulatory requirements“
- Fragestellungen: effektive Verwaltung von dezentralisierten Informationen; Verbinden von Ontologien; effektives automatisches Navigieren auf Metadaten-Graph; Trust

– „*The Semantic Web Revisited*“, Tim Berners-Lee, Nigel Shadbolt, Wendy Hall, „*IEEE Intelligent Systems*“, Juni 2006

Fazit – Kritik

- „What Do Skeptics Say about the Semantic Web?“
 - Bad precedent: failure of early artificial intelligence research
 - Fear, uncertainty, and doubt (FUD): difficulty of implementation details; too costly
 - Status quo: we don't need a Semantic Web; distraction from linear progress

aus: „*The Semantic Web*“, Michael C. Daconta, Leo J. Orbst, Kevin T. Smith, Wiley Publishing, 2003
- andere Kritikpunkte
 - „Metacrap“: menschliches Verhalten verhindere vollständige Umsetzung
 - „Censorship“: automatisierte Zensur werde erleichtert
 - „Privacy“: persönliche Meta-Daten erlaubten wenig Anonymität in z.B. Blogs
 - „Doubling output formats“
 - ein Format für Mensch, eins für Maschine; Publishing-Prozess wird verzögert
 - Ansätze: automatisierte Erzeugung der Meta-Daten on request; *Microformats*

aus: „*Semantic Web*“, *Wikipedia*, http://en.wikipedia.org/wiki/Semantic_Web#Criticism

Fazit – Beispiel „e-science“

- e-science
 - Einsatz moderner Informationstechnologie im wissenschaftlichen Umfeld, speziell durch verteilte Systeme – <http://de.wikipedia.org/wiki/E-Science>
- Beispiel aus Biowissenschaften
 - *“Open Biomedical Ontologies (OBO)”* – <http://obo.sourceforge.net>
 - Entwicklung von freien Ontologien
 - menschenlesbare Beschreibungen von Begriffen, eingeordnet in Ontologie, gespeichert in maschinenlesbaren Dateien (OWL)
 - Beziehungen zwischen Ontologien, globaler eindeutiger Namespace
 - seit März 2003 bei Sourceforge registriert
 - *“Ontology Lookup Service (OLS)”* – <http://www.ebi.ac.uk/ontology-lookup/>
 - Suche nach biomedizinischen Begriffe in Ontologien
 - Suche: “von Hand” im Browser oder per Webservice (WSDL verfügbar)
 - Ergebnis: Einordnung des Begriffs in eine Ontologie, untergeordnete Begriffe
 - benutzt Ontologien von der OBO-Cummunity

Fazit – Beispiele

- Großbritannien: „*Integrated Public Sector Vocabulary*“ – www.esd.org.uk/standards/ipsv
 - Taxonomie mit Vokabeln aus dem öffentlichen Sektor, inkl. internen Bezeichnungen und Abkürzungen
 - diverse Formate, auch XML und RDFS
- „*Friend of a Friend (FOAF)*“ – www.foaf-project.org
 - „creating a Web of machine-readable pages describing people, the links between them and the things they create and do.“
 - „*FOAF Vocabulary Specification*“, basiert auf XML & RDF – xmlns.com/foaf/0.1/
 - Tools zum Erstellen einer Personen-Beschreibung in FOAF-Spezifikation
 - FOAF-Datei wird unter URI im Web verfügbar gemacht: „Diese Web-Ressource beschreibt mich.“
- Creative Commons – www.creative-commons.org
 - Lizenz für Fotos und andere Werke im Web wird mit RDF beschrieben

Vielen Dank für Ihre Aufmerksamkeit