

Projekt Semantic Web (SoSe 2004) 19565 PJ

Prof. Dr.-Ing. Robert Tolksdorf
Freie Universität Berlin
Institut für Informatik
Netzbasierende Informationssysteme
mailto: tolk@inf.fu-berlin.de
<http://www.robert-tolksdorf.de>



[1] © Robert Tolksdorf, Berlin

Zielsetzung

- Das Semantic Web ist eine Initiative zur Entwicklung von Standards mit denen sich semantische Zusammenhänge im Web repräsentieren lassen und maschinell verarbeitet werden können. In dem Projekt erproben wir diese Technologien in einer Anwendung.
- Dabei soll die Berliner Hochschullandschaft - mit besonderer Berücksichtigung der Informatik - in einem Semantic Web modelliert und erfasst werden.

[2] © Robert Tolksdorf, Berlin

Grundlagen

- Das Grundlagenwissen zu Semantic Web soll selber erarbeitet werden.
- Dafür stehen umfangreiche Quellen zur Verfügung:
 - Material des Infotags Semantic Web des XML Clearinghouse (nur vom FU-Netz aus zugänglich)
 - Die XML Clearinghouse Reports
 - Technologien des Semantic Web
 - Software zum Ontologiemanagement mit OWL
 - Die Ergebnisse des Projekts Semantic Web vom SoSe 2003 (nur vom FU-Netz aus zugänglich)

[3] © Robert Tolksdorf, Berlin



Infotag Semantic Web

Robert Tolksdorf, FU Berlin
Rainer Eckstein, HU zu Berlin
XML Clearinghouse für Berlin und Brandenburg
www.xml-clearinghouse.de

Semantic Web

- Wieso Semantic Web?
- Überblick
- Wesentliche Technologien

Ausgangslage: Wissen schwierig im Web zu ermitteln

- Aktuelles World Wide Web
- Auffinden von Wissen schwierig
- Vorhandene Technologien zur Suche: Suchmaschinen, im Kern: Information Retrieval
- Dienste: Google, Yahoo, AltaVista, MetaCrawler etc.
- Probleme:
 - Recall limitiert: Alles relevante gefunden?
 - Precision limitiert: Wieviele gefundene Dokumente sind relevant?
 - Relevanz fragwürdig: Nur auf Wort-Vektoren basierend

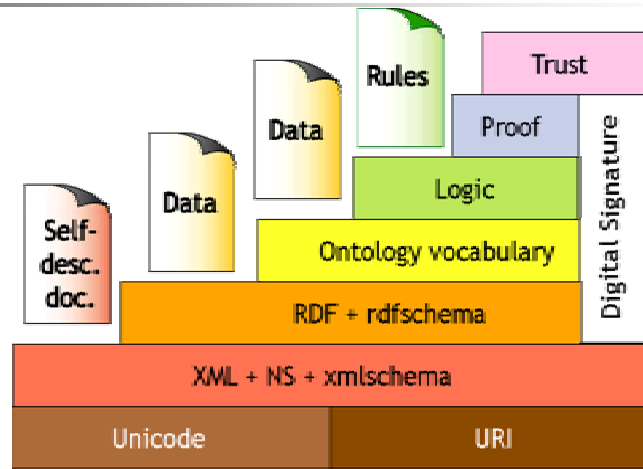
Zielsetzung

- Bedeutung von Informationen maschinenverarbeitbar markieren
- Damit Unterstützung von
 - Entdeckung gehaltvollerer Informationen
 - Datenintegration
 - Bessere Navigation
 - Automatisierung von Aufgaben

Aktueller Ansatz: Semantic Web

- Idee: Definition von standardisierten Technologien zur maschinenverarbeitbaren Beschreibung von Wissen in einem großen Netz
- Prinzipien
 - Jede Information kann mit URIs identifiziert werden
 - Ressourcen und Beziehungen sind semantisch typisiert
 - Partielle Informationen sind akzeptabel
 - Absolute Wahrheit wird nicht benötigt
 - Evolution als Entwicklungsprinzip
 - Minimalistisches Design der Technologien

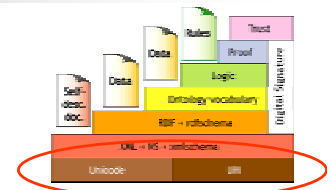
Schichten des Semantic Web



Aus: Tim Berners-Lee: <http://www.w3.org/2000/Talks/1206-xml2k-tbl/slide10-0.html>

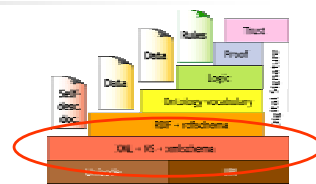
Bestandteile des Semantic Web

- Unicode
 - Universelles Zeichenrepertoire
 - Umfasst alle Zeichen der Welt
 - Ist fortlaufend in Erweiterung
 - Als ISO 10646 normiert
- URI
 - Universeller Bezeichner
 - Klare Syntax
 - Bei Interpretation im Netz: URL



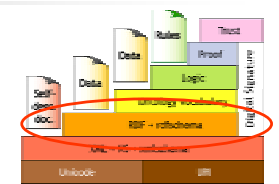
Bestandteile des Semantic Web

- XML
 - Sprache zur Definition von Auszeichnungssprachen
- Namespaces
 - Konzept zur Verwendung mehrerer XML Sprachen in einem Dokument ohne Namenskonflikte
- XML Schema
 - Modellierungssprache, die XML um Datentypen, Strukturierungs- und Wiederverwendungsmöglichkeiten erweitert



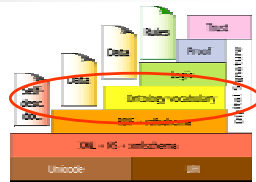
Bestandteile des Semantic Web

- RDF
 - Modell und Syntax zur Annotation von Metadaten
- RDF Schema
 - Sprache zur Definition von Metadatenformaten und einfachen Ontologien



Bestandteile des Semantic Web

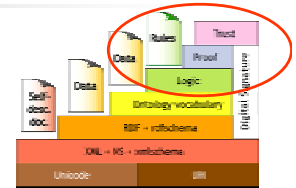
- Ontologien
 - Vokabulare
 - Begriffsbeziehungen (Unterkategorie, Untereigenschaft, Wertebereiche, ..., selbstdefinierte)
 - Sprachen für Web-Ontologien: DAML+OIL, OWL



Bestandteile des Semantic Web

- Logik
 - Semantik auf logischer Basis
 - Ableitungsregeln
- Proof
 - Konsistenz
 - Ableitung (Inferenz)
- Trust

- Am Beginn der Entwicklung



Maschinelle Auswertung

- Metadaten: menschliche Nutzer und intelligente Anwendungen
- Ziele
 - Sammlung von dezentralen Webinhalten
 - Informationsverarbeitung
 - Ergebnisaustausch mit anderen Programmen
=> gemeinsame Ontologien für Semantikbeschreibungen
- Framework
 - Konsistenzüberprüfung (Seiteninhalt, Metadaten, Ontologie(n))
 - Klassifikation (Informationsreduktion)
 - Anfragen (Inferenz-Service: logisches Schließen)

Mögliche Anwendungen

- Semantische Suche:
 - Anhand von Ontologien mittels Inferenz sinnvolle(re) Dokumente und Informationen finden
- B2C-Commerce: (auch B2B)
 - Kunde: Produkt mit bestimmten Anforderungen (Gestaltung, Fähigkeiten, Leistung, Preis, ...) gesucht
Was passt am besten?
=> U.U. interaktive Anpassung der Spezifikation
- ...

Zielvision

Tim Berners-Lee:

- Automatische Koordination von Terminen mit Suche von geeigneten Dienstleistern:
 - Geschwisterpaar will Vater zum Arzt und zur Behandlung begleiten
 - Wo sind geeignete Ärzte, KG-Praxen? (in der Nähe)
 - Wann sind Termine passend frei? (Präferenzen, Wegezeit)
 - Müssen andere Termine dafür umgelegt werden? (Priorität)

Zustand heute

- Das Semantic Web existiert nicht (Aussage vom 18.2.2003)
- Es ist allgemein akzeptiert, dass es existieren sollte
- Wege zu einem Semantic Web
 - Entdeckung der "Killer Applikation"
 - Bottom-Up Entwicklung getrieben von Anwendern
 - Von Firmennetzwerken bzw. speziellen Anwendungsdomänen zu allgemeiner Verfügbarkeit
- Viele Methoden / Technologien / Werkzeuge existieren schon (KI)
=> für das Semantic Web verfügbar machen / integrieren

Herausforderungen

- Wissen aufbereiten
 - Entwicklungsumgebung
 - Wissensmanagement
 - Organisatorische Einbindung
 - Retrospektive Erschließung von Wissen
- Wissen nutzen
 - Inferenzen
 - Skalierbare Tools
 - Endnutzer

Die Hochschullandschaft Berlins als Semantic Web

- Diese Dokument beschreibt die Aufgaben des Projekts, das in der Lehrveranstaltung Projekt Semantic Web bearbeitet wird.
- Die vorliegende Projektvereinbarung beschreibt das Vorhaben des Baus eines Semantic Web, das die Berliner Hochschulen als Informationssystem erfasst. Dabei soll die Informatik mit besonderem Detail erfasst sein. Auftraggeber des Projekts ist die AG Netzbasierte Informationssysteme, vertreten durch Prof. Tolksdorf.

1. Arbeitsziele des Vorhabens

- Das Projekt soll ein Online-Informationssystem auf Basis der Technologien des Semantic Web erstellen.
- Der Inhalt des Informationssystems sollen semantisch erschlossene Angaben über die Hochschullandschaft Berlins unter besonderer Berücksichtigung der Informatik sein.
- Der Hochschulkompass liefert für die 17 Hochschulen Berlins 27 grundständige und 9 weiterführende Studiengänge der Informatik.
- Das Online-Informationssystem soll die Hochschulen und die Informatik-Studiengänge inhaltlich erschließ- und vergleichbar machen.

- Notwendige Komponenten sind dazu
 - Modellierungen zur Beschreibung von Hochschuleinrichtungen, beispielsweise auf Basis der Educational Resource Ontology (EduOnto) , ISO/IEC SC36 und anderen.
 - Fachmodelle der Informatik, beispielsweise der ACM CS Classification
 - Modellierungen der genannten Studiengänge und Anreicherung durch inhaltliche Bezüge und Vergleichsmöglichkeiten
 - Eine unterliegende Infrastruktur für das System aus Basis vorhandener, frei erhältlicher Tools.

[21] © Robert Tolksdorf, Berlin

1. Arbeitsziele des Vorhabens

- Notwendige Arbeiten im Projekt umfassen
 - die inhaltliche Erarbeitung der genannten Modelle und ihre Kodierung nach Semantic Web Standards
 - die Erfassung von Hochschulen und relevanter Studiengänge
 - die Auswahl und Implementierung von Software der Infrastruktur
 - die Realisierung des Semantic Web als Online-Informationssystem
 - das Management des Projekts

- Das Informationssystem soll qualitativ geeignet sein, als ein "Berliner Hochschulkompass Informatik" zu dienen.

[22] © Robert Tolksdorf, Berlin

1. Arbeitsziele des Vorhabens

- Das Projekt wird extern begutachtet; als Gutachter werden die Teilnehmer/innen des an der TU Berlin stattfindenden Seminars "XML und Semantic Web" fungieren.
- Der Arbeitsplan sieht Meilensteinberichte vor, auf deren Grundlage ein Review-Report als Feedback erstellt wird.
- Am Ende des Semesters findet eine Abschlusspräsentation vor Interessierten aus der FU Informatik und anderen Hochschuleinrichtungen und insbesondere vor Teilnehmern des an der TU Berlin stattfindenden XML und Semantic Web statt.

[23] © Robert Tolksdorf, Berlin

2. Beschreibung des Arbeitsplans

Arbeitspaket	A1: Modellbildung	Dauer	W1- W12
Aufgaben	•Erstellung von Modellen für Hochschuleinrichtungen und Informatikthemen auf der Grundlage vorhandener Ansätze •Modellierung von Informatik-Studiengängen damit •Dokumentation der Arbeiten		
Ergebnisse	Dokumentiertes Modell von Berliner Informatik-Studiengängen und dessen Repräsentation mit Semantic Web Technologien		
Meilensteine	1. W7: Modell 2. W13: Überarbeitetes Modell mit Instanzen		

[24] © Robert Tolksdorf, Berlin

2. Beschreibung des Arbeitsplans

Arbeitspaket	A2: Infrastruktur	Dauer	W1- W12
Aufgaben	<ul style="list-style-type: none"> •Auswahl von Komponenten für eine Infrastruktur zum Bau eines Semantic Web •Konzeption und Umsetzung geeigneter Navigationshilfen im Informationssystem •Erstellung der Onlineversion des Informationssystems •Dokumentation der Arbeiten 		
Ergebnisse	Infrastruktur für die Erstellung und Nutzung des Informationssystems und dessen Online-Version		
Meilensteine	<ol style="list-style-type: none"> 1. W5: Komponenten der Infrastruktur 2. W9: Navigationsdienste 3. W13: Online-Version des Informationssystems 		

[25] © Robert Tolksdorf, Berlin

2. Beschreibung des Arbeitsplans

Arbeitspaket	A3: Projektmanagement	Dauer	W0- W13
Aufgaben	<ul style="list-style-type: none"> •Organisation der Projektarbeit •Qualitätssicherung (Sicherstellung der Erfüllung von Arbeiten und der Einhaltung von Deadlines) •Dokumentation des Projekts •Außenvertretung des Projekts 		
Ergebnisse	Erreichen der Projektziele		
Meilensteine	<ol style="list-style-type: none"> 1. W3: Projektorganisation und Arbeitsverteilung 2. W13: Projektabschluss 		

[26] © Robert Tolksdorf, Berlin

2. Beschreibung des Arbeitsplans

Termin	Projekt- woche	A1	A2	A3	Thema	Vortragend e/r
21.4.	W0				Einleitung, Organisation, Arbeitsorganisation und -planung I	RT
28.4.	W1				Arbeitsorganisation und -planung II	RT/NN
5.5.	W2					
12.5.	W3					
19.5.	W4					
26.5.	W5					
2.6.	W6					
9.6.	W7					
16.6.	W8					
23.6.	W9					
30.6.	W10					
7.7.	W11					
14.7.	W12					
21.7.	W13				Abschlusspräsentation	

[27] © Robert Tolksdorf, Berlin

Zu tun

- Weitere Organisatorische Fragen etc.
- Selbstorganisation der Projektteilnehmer
 - Entscheidung über Projektmanagement
 - Zuordnung zu APs
 - Start AP3 durch Projektmanagement
 - Entscheidungen zur Kommunikationsinfrastruktur des Projekts
 - ...

[28] © Robert Tolksdorf, Berlin