

International Conference on Dublin Core and Metadata Applications

Tutorial II Einführung in Semantic Web



DC-2008
BERLIN



Gliederung des Vortrags

1 Einführung: Die Idee des Semantic Web

- 1.1 Eine kurze Übersicht zum Thema
- 1.2 Ein anschauliches Beispiels

2 Konzepte des Semantic Web

- 2.1 Metadaten
- 2.2 Taxonomien
- 2.3 Ontologien

3 Technologien des Semantic Web

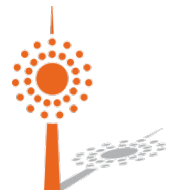
- 3.1 XML
- 3.2 RDF
- 3.3 OWL

4 Anwendungen des Semantic Web

- 4.1 Ein aktuelles Projekt

5 Fazit und Ausblick

6 Tutorial



1 Einführung Die Idee des Semantic Web

- Das Semantische Web
 - Erweiterung des gegenwärtigen Webs
 - Informationen auch in maschinenlesbarer Form
- Worum geht es?
 - durch spezielle Definition und Verknüpfung von Web-Daten
 - effektivere Such-, Automatisierungs- und Integrationsprozesse
 - über mehrere Anwendungen wiederverwendbar



1.1 Eine kurze Übersicht zum Thema

- Entwicklung
 - bisher WWW hauptsächlich Archiv für (vom Menschen lesbare) Dokumente
 - automatisch zu verarbeitende Daten kaum vertreten
 - durch das Hinzufügen computergerechter Informationen zu bestehenden Websites wird Web zu semantischem Netz
 - Computer werden Bedeutung semantischer Daten erkennen
 - Nachschlagen von Definitionen und Schlussregeln über Hyperlinks (sogenannte URIs)



1.1 Eine kurze Übersicht zum Thema

- Ziele
 - Entwicklung von universellen Softwareagenten u.a. neuen Funktionen des Webs
 - jeder Benutzer wird semantische Websites selbst erstellen können
- Wozu kann man das brauchen?
 - explizite Informationen im Web vorhanden, aber nicht in maschinenverständlicher Form
 - bisher immer Mensch nötig, der Relevanz der Informationen prüft



1.1 Eine kurze Übersicht zum Thema

- Lösung
 - Daten so bereitstellen, dass Programme die relevante (Teil-) Information bekommt
 - Dateien generieren, die der Maschine Beziehung zwischen verschiedenen Datenbeständen erklärt
 - z.B. „semantischer Link“ zwischen Datenbank mit Spalte von Postleitzahlen und „PLZ-Feld“ in Online-Formular)
 - Maschinen folgen diesen Links
 - erleichtert Datenintegration aus verschiedenen Quellen



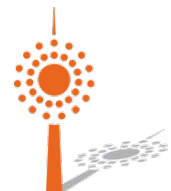
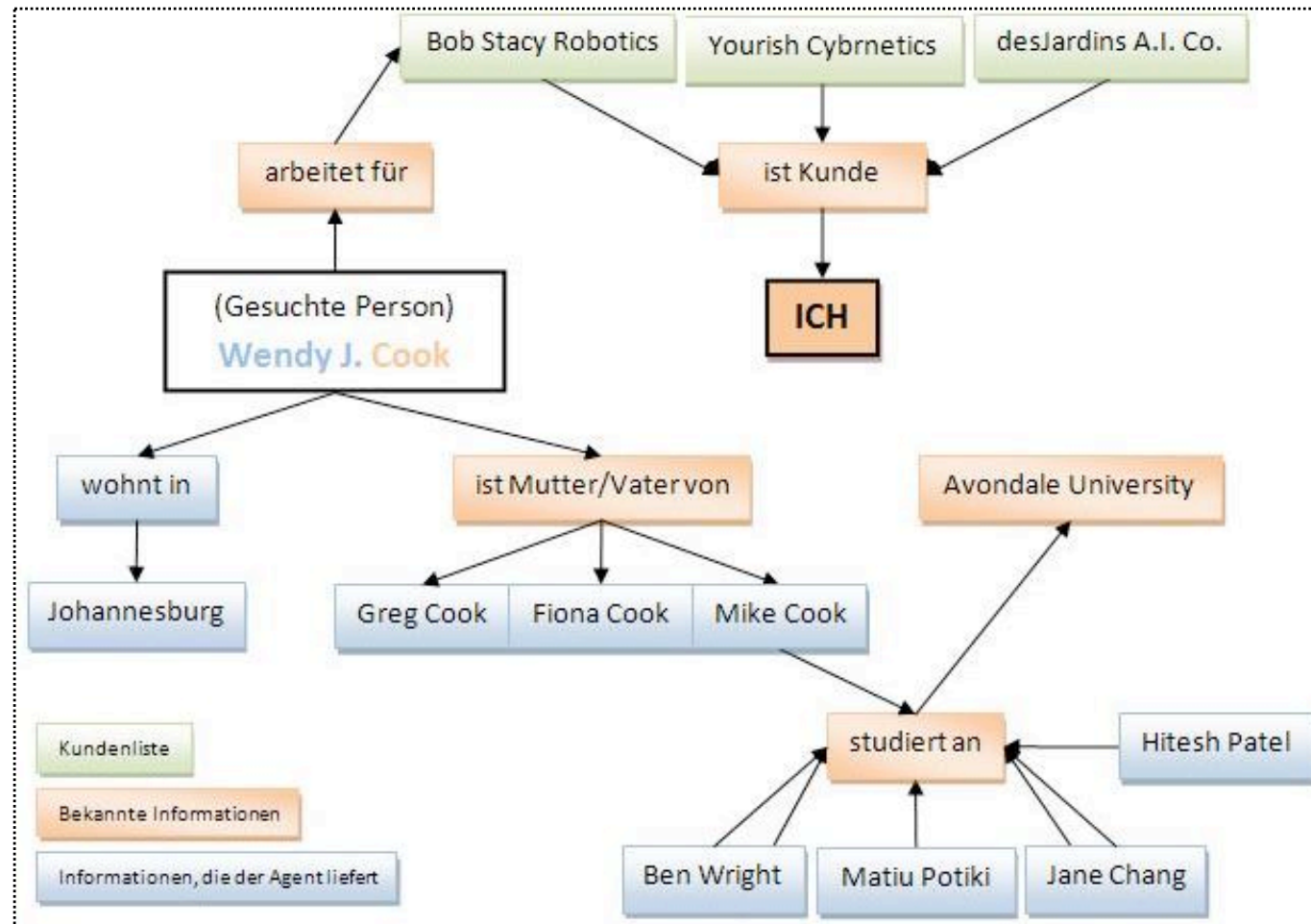
1.2 Ein anschauliches Beispiel

- Beispiel „Wendy Cook“
 - Person mit Nachname "Cook"
 - arbeitet für Firma auf Ihrer Kundenliste
 - hat Sohn (Student der Avondale University)
- Suchprogramm setzt Inhalte (verschiedener) Websites in Beziehung zueinander
- versteht Relationen wie "Mike Cook ist Kind von Wendy Cook"
- zieht aus all dem die richtigen Schlüsse



1.2 Ein anschauliches Beispiel

- Das Beispiel „Wendy Cook“



2 Konzepte des Semantic Web

- semantische Verknüpfung verschiedene Ressourcen (Dokumente, Bilder, Menschen, Konzepte etc.)
 - Web wird schrittweise Bedeutung hinzugefügt
 - vollständig neuen Satz von Beziehungen (istDort, arbeitetFür, istAutorVon, hatZumThema, istAbhängigVon etc.)
- effizientere Informationsintegration bzw. –verwaltung
- neue automatische Dienstleistungen



2.1 Metadaten

- Definition:
 - ("Daten über Daten")
 - strukturierte Daten
 - Beschreibung Informationsressource (besser auffindbar)
 - liefern Grundinformationen über Dokument (z.B. Autor, Titel, Zeitpunkt der Veröffentlichung)
 - Erschließungsarbeit in Bibliotheken



2.1 Metadaten

- Tim Berners-Lee (Erfinder des WWW, Direktor des W3C):

„Metadaten sind maschinenlesbare Informationen über elektronische Ressourcen oder andere Dinge.“

- effektiver Einsatz von Metadaten setzt gewissen Standardisierungsgrad voraus
- Suche nach effizienten und kostengünstigen Ansätzen der Ressourcenbeschreibung und entsprechenden Verfahren



2.2 Taxonomien

- Definition:

Die Taxonomie ist eine Hierarchie von Begriffen.

- Altertum: Versuch Flora und Fauna zu ordnen
- Begriff nicht auf Informationstechnologie beschränkt
- keine Standardisierung der Technologie
- weit verbreitete Anwendung: Windows-Dateisystem mit Ordnerstruktur



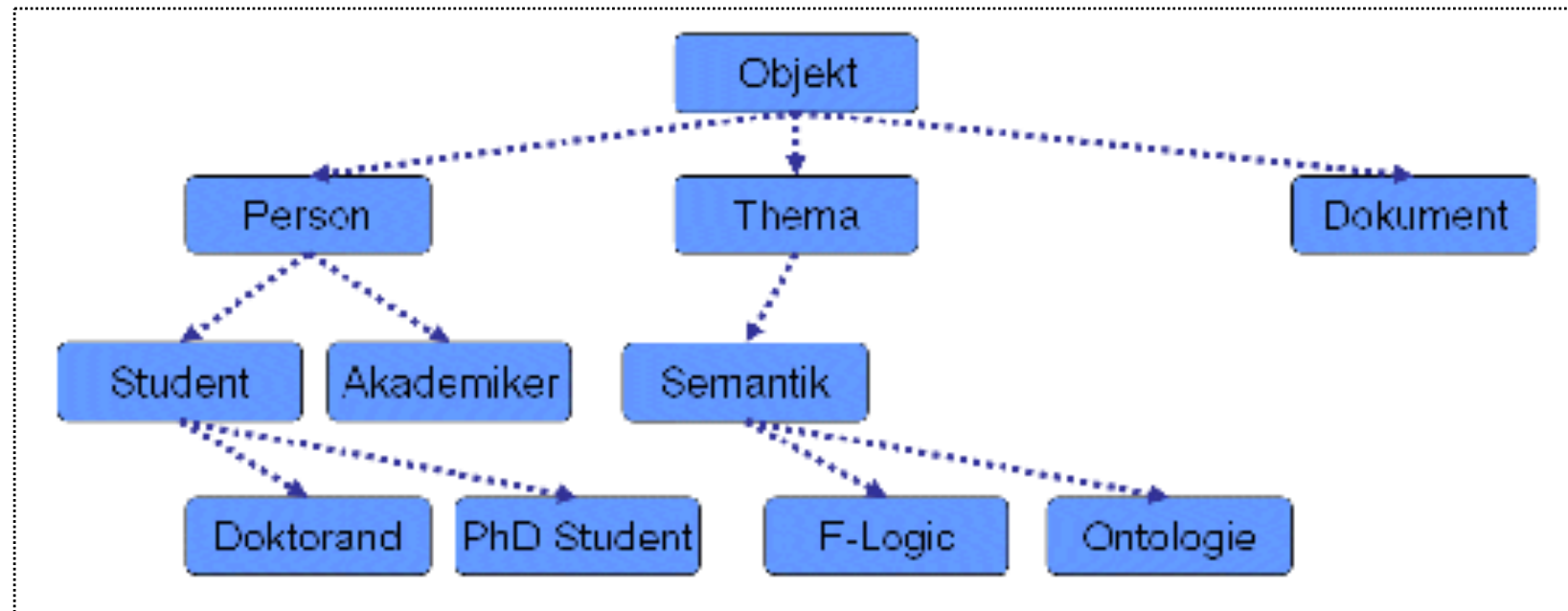
2.2 Taxonomien

- Taxonomie definiert Objektklassen und zwischen ihnen bestehende Relationen
- Adresse = spezielle Art von Ortsangabe
- Formal: Klasse der Adressen = Teilklasse der Klasse Ortsangaben
- "erbt" alle Eigenschaften der übergeordneten Klasse
- Eigenschaft, die für Ortsangaben gilt, gilt auch für Adressen



2.2 Taxonomien

- Beispiel für eine Taxonomie
 - **Objekt** → **Person**, Thema oder Dokument
 - Person** → **Student** oder Akademiker
 - Student** → Doktorand oder PhD Student



2.3 Ontologien

- Definition:
 - formale Beschreibung von Ausdrücken und deren Relationen untereinander
 - auch Wissensbereich genannt
 - Taxonomien = Hilfsmittel, während
- Ontologien für anspruchsvollere Aufgaben geeignet (erweiterungsfähig in Richtung Integration und regelbasierte Systeme)



2.3 Ontologien

- Trennung zwischen Schema (Meta-Modell) und Inhalten
- Konzepte definieren das Schema
- Instanzen stellen die Ausprägungen dar
- Zusammenhänge zwischen Objekten (oder auch anderer Ontologien) werden mittels
 - Wenn-Dann Beziehungen, Zuweisungen,
 - Negationen, logischen Verknüpfungen,
 - mathematischen Operationen usw.
ausgedrückt



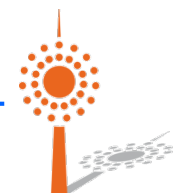
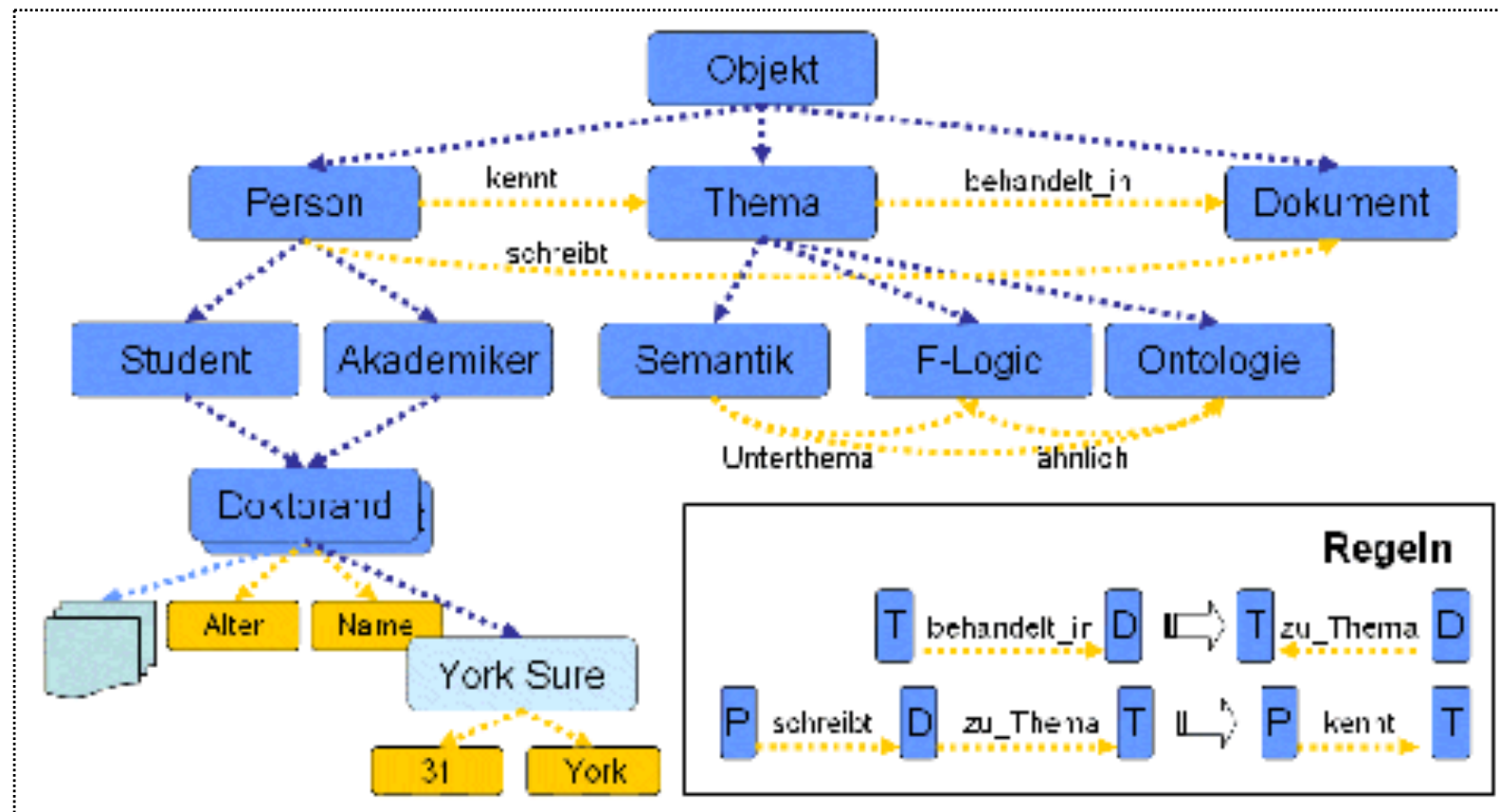
2.3 Ontologien

- mittels Abfragesprache kann Ontologie in beliebige Richtungen angefragt werden
- Verbesserung von Suchsystemen, Navigation und Visualisierung
- Weitere Einsatzgebiete: Frage- und Antwortsysteme, sowie Integration heterogener Datenquellen
- Verwendung für
 - genauere Suchmaschinen,
 - intelligentere Software-Agenten und
 - Wissensmanagement



2.3 Ontologien

- Ein Beispiel für eine Ontologie



3 Technologien des Semantic Web

- Zwei wichtige Technologien für die Entwicklung des Semantic Web sind bereits verfügbar:
 - die eXtensible Markup Language (XML) und
 - die Resource Description Framework (RDF).
- Eine andere, ebenso wichtige Technologie wird gerade entwickelt:
 - die Web Ontology Language (OWL).



3 Technologien des Semantic Web

- XML

- Interoperables, syntaktische Fundament

- XML lässt jeden seine eigenen Tags erstellen

HTML-Tag <body>...</body>

XML-Tag <vorname>...</vorname>

- Scripte oder Programme können Gebrauch von Tags machen

- XML erlaubt es, Dokumenten eigenmächtige Strukturen zu verleihen, sagt nichts über Bedeutung der Struktur aus



3 Technologien des Semantic Web

- RDF
 - Bedeutung (Semantik) wird mittels RDF ausgedrückt
 - Datenintegration, -sammlung und Interoperabilität möglich
- OWL
 - Interoperabilität auf ausdrucksvollerem Niveau nötig
 - Web Ontology Working Group entwickelt (basierend auf RDF)
Sprache für Definition von strukturierten, web-basierten Ontologien



3.1 XML

- Definition
 - XML = Satz an Regeln (Richtlinien oder Konventionen)
 - Erstellung von Textformaten zur Strukturierung von Daten
 - XML ist KEINE Programmiersprache
 - man muss kein Programmierer zu sein, um XML zu benutzen oder zu lernen



3.1 XML

- XML erleichtert es dem Computer, Daten zu generieren oder zu lesen
- sorgt dafür, dass bestimmte Datenstruktur eindeutig bleibt
- XML ist
 - erweiterbar,
 - plattformunabhängig und
 - unterstützt Internationalisierung / Lokalisierung
 - Unicode



3.2 RDF

- Definition
 - W3C's Resource Description Framework
 - XML Textformat
 - beschreibt Ressourcen und unterstützt Metadatenanwendungen
 - Bsp.: Personen in Web-Photoalbum aufgrund von Informationen aus Kontaktliste identifizieren und automatisch Email senden
 - integriert Anwendungen und Agenten in semantisches Web



3.2 RDF

- Beispiel

```
<rdf:Property rdf:about="http://localhost/2008/05/19/izat#PU">
```

```
<rdfs:label xml:lang="de-DE">PU</rdfs:label>
```

```
<rdfs:comment xml:lang="de-DE">Publikationsart</rdfs:comment>
```

```
<rdfs:isDefinedBy rdf:resource="http://localhost/2008/05/19/izat#"/>
```

```
</rdf:Property>
```



3.2 RDF

- Computer benötigen Mechanismen für die Vereinbarung von Bedeutungen, um effektiv zu kommunizieren
- In RDF stimmt Dokument zu, dass
 - bestimmte Dinge (Menschen, Webseiten,...)
 - Eigenschaften haben („ist Schwester von“, „ist Autor von“)
 - mit bestimmten Werten
- natürlicher Weg große Mehrheit der Daten (Verarbeitung durch Maschinen) zu beschreiben



3.2 RDF

- Semantik wird mittels Dreierpärchen ausgedrückt
- Dreierpärchen: wie Subjekt, Verb und Objekt
 - können geschrieben werden, indem XML-Tags verwendet werden
- Erkennung von Subjekt, Verb und Objekt anhand eines Universal Resource Identifier (URI)
 - Befähigt jeden, neues Konzept oder Verb zu definieren, indem URI dafür definiert wird



3.2 RDF

- „Die menschliche Sprache gedeiht, wenn der selbe Ausdruck verwendet wird, um etwas anderes auszudrücken, aber Automatisierung tut das nicht.“
- RDF-Dreierpärchen gestalten Informationsnetze über verwandte Dinge
- URI gewährleistet, dass Konzepte
 - nicht einfach nur Wort in Dokument, sondern
 - an einmalige Definitionen gebunden



3.2 RDF

- Beispiel: Vielzahl von Datenbanken mit Informationen über Menschen, einschließlich Adressen
- Suche nach Menschen in speziellem Postleitzahlengebiet
→ Kenntnis über Felder in jeder Datenbank
- RDF-Lösung:

(Feld 5 in der Datenbank A) (ist ein Feld des Types) (Postleitzahl)



3.3 OWL

- Definition Ontologie
 - bezieht sich auf Wissenschaft, die unterschiedlichste Arten von Entitäten und ihre Beziehungen untereinander beschreibt
- Definition OWL
 - Ontologie Web Language
 - Sprache zur Definition und Realisierung von Web Ontologien



3.3 OWL

- Beschreibungen von Klassen, Eigenschaften und ihren Instanzen
- Fakten, zunächst nicht explizit in Ontologie vorhanden, jedoch aus deren Semantik abgeleitet
- Basis: ein oder mehrere Dokumente
- Ziel: Anwendungen verarbeiten Inhalte von Informationen anstatt Informationen nur zu präsentieren



3.3 OWL

- durch zusätzliches Vokabular stärkere Interpretationsmöglichkeiten von Web-Inhalten möglich
- drei Untersprachen mit steigender Ausdrucksmächtigkeit:
 - OWL Lite,
 - OWL DL und
 - OWL Full
- jede Untersprache = Erweiterung ihres einfacheren Vorgängers



3.3 OWL

- Vorteile gegenüber XML
 - Ontologie = Repräsentation von Wissen, nicht nur Datenformat
 - Formate wurden mit Semantik, wie "Bei Empfang dieser *Bestellung*, transferiere die *Menge* Euro auf *Konto*" versehen
 - keine Schlussfolgerungen ausserhalb dieses Kontexts möglich
 - Tools, die aus Informationen Schlussfolgerungen ziehen, die
 - nicht auf ein spezielles Anwendungsgebiet beschränkt sind



4 Anwendungen des Semantic Web

- Ein aktuelles Projekt

SKOS (Simple Knowledge Organization System)

- W3C Working Draft
- Kodierungs- und Austauschformat für Anwendungen im Semantic Web
- RDF-basiertes Vokabular zur Repräsentation/Kodierung von Thesauri, Klassifikationen, etc.



4 Anwendungen des Semantic Web

- Veröffentlichung und Kombination kontrollierter, strukturierter und maschinenlesbarer Vokabulare für das Semantische Web
- einfache, intuitiv zu bedienende Sprache für Entwicklung und Nutzung neuer Knowledge Organisation Systems
- SKOS-Datenmodel besteht aus OWL-Ontologie
- definiert durch SKOS-RDF-Schema, Elemente: Klasse und Eigenschaft
- SKOS-Daten werden über RDF-Triple ausgedrückt



4 Anwendungen des Semantic Web

- einfache, flexible, maschinenlesbare Repräsentation von
 - Thesauri und Taxonomien,
 - Subject Headings
 - Klassifikations-Schemata
 - u. a. kontrolliertem Vokabular
- Ziel: Anwendung klassischer Information Retrieval-Verfahren



4 Anwendungen des Semantic Web

- Nutzung:
 - mehrsprachige Übersetzungen von Termen und Konzepten,
 - Textsuchmaschinen,
 - Import und Export von Ontologien,
 - Management von Terminologien uvm.
- Transformation von MeSH (Medical Subject Headings)
- wird in finnischem Projekt dazu benutzt, um im Gesundheits-
-Informations-Portal „TerveSuomi“ ontologische Annotationen
zu unterstützen



4 Anwendungen des Semantik Web

- Zusammenfassung
 - SKOS = interessantes Vokabular zur Repräsentation von unterschiedlichen Begriffsschemata, insbesondere von inhaltlichen Metadaten
 - kein Vokabular für die Individuenebene, aber zukunftsfähig und vielseitig
 - permanente Weiterentwicklung und Überarbeitung der SKOS-Spezifikationen



5 Fazit und Ausblick

Die nötigen Konzepte und Technologien können dazu verwendet werden, das Web in ein semantisches Web zu verwandeln.

- Metadaten zur Beschreibung der Ressourcen
- XML zur freien Benennung der Strukturelemente
- Taxonomien zur Klassifikation der Entitäten
- RDF als Datenmodell für maschinenlesbare Metadaten
- Ontologien zur Strukturierung und zum Datenaustausch
- OWL zur formalen Beschreibung von Termen und deren Relationen



6 Tutorial

Tutorial:

<http://fabdax.fh-potsdam.de/dcmoodle>

(<http://ecampus.fh-potsdam.de/moodle/mod/book/view.php?id=7334>)

Quiz:

(<http://ecampus.fh-potsdam.de/moodle/mod/quiz/attempt.php?q=20>)



Weiteres Programm

13.30 – 15.00

Einführung in RDA (Resource Description and Access)

Dierk Eichel

15.30 – 17.00

Einführung in Social Tagging/Computing

Johannes Hercher



1.2 Zwei anschauliche Beispiele

- Das Beispiel „Physiotherapie“
 - Die Stereoanlage spielt: "We can work it out" von den Beatles, als das Telefon klingelt. Sowie Peter den Hörer abnimmt, sendet das Telefon eine Nachricht an alle lokalen Geräte mit Lautstärkeregler, woraufhin die Beatles verstummen.
 - Peters Schwester Lucy ist am Apparat, von der Arztpraxis aus. "Sie muss zu einem Spezialisten und dann alle zwei Wochen in die Physiotherapie. Ich lasse meinen Agenten die Termine organisieren."
 - Peter erklärt sich bereit, einen Teil der Chauffeurdienste zu übernehmen. Noch in der Praxis programmiert Lucy mit einem portablen Webbrowser ihren Webagenten.



1.2 Zwei anschauliche Beispiele

- Das Beispiel „Physiotherapie“
 - Dieser holt sich sogleich vom Agenten des Arztes die Daten des Rezepts, das der Arzt ausgestellt hat, und geht mehrere Listen mit Physiotherapeuten durch. Daraus wählt er diejenigen aus, die für die gewünschte Krankenversicherung eine Zulassung haben, sich weniger als 30 Kilometer entfernt von ihrer Wohnung befinden und von vertrauenswürdigen Verbraucherorganisationen die Einstufung "ausgezeichnet" oder "sehr gut" erhalten haben.
 - Schließlich gleicht er die freien Termine der ausgewählten Anbieter (aufgelistet von deren Agenten auf den deren Webseiten) mit Peters und Lucys dicht gedrängtem Terminplan ab. Ein paar Minuten später präsentiert der Agent einen Plan.



1.2 Zwei anschauliche Beispiele

- Das Beispiel „Physiotherapie“
 - Peter ist nicht begeistert: er müsste im dichtesten Berufsverkehr zur Universitätsklinik am anderen Ende der Stadt fahren.
 - Er weist seinen eigenen Agenten an, die Suche zu wiederholen und dabei die Kriterien Standort und Zeit wichtiger zu nehmen.
 - Lucys Agent, der für Aufgaben dieser Art volles Vertrauen in Peters Agenten hat, unterstützt diesen, indem er Zugangsausweise für gewisse Websites und halb aufbereitete Daten an ihn weiterreicht.
 - Kaum eine Sekunde später wird ein neuer Plan vorgeschlagen, mit einer näher gelegenen Praxis und günstigeren Terminen, allerdings mit zwei Warnhinweisen:



1.2 Zwei anschauliche Beispiele

- Das Beispiel „Physiotherapie“
 - Erstens müsste Peter ein paar weniger wichtige Verabredungen verlegen.
 - Zweitens hat die Praxis zwar eine Zulassung von der Versicherung, aber nicht für Physiotherapie.
 - Aber der Agent weiß Beruhigendes: "Behandlungstyp und Abdeckung durch Versicherung auf anderem Wege zuverlässig sichergestellt. (Einzelheiten?)"
 - Lucy stimmt dem Plan auf der Stelle zu, Peter murmelt nur: "Spar mir die Einzelheiten", und damit ist alles geregelt.
 - Natürlich kann Peter es nicht lassen, später am Abend seinen Agenten zu fragen, wie der diese Praxis gefunden hat, obwohl sie nicht in der richtigen Liste verzeichnet ist.

