

Vorlesung Informatik
Dr. rer. nat. Harald Sack
Institut für Informatik
Friedrich Schiller Universität Jena

Wintersemester 2006/2007

Semantic Web - Vorlesungsinhalt

- Einführung
- 2. Die Sprachen des Semantic Web
- 3. Wissensrepräsentation
- 4. Ontology Engineering
- 5. Web of Trust
- 6. Semantic Web Anwendungen

Semantic Web

1 2 3 4 5 6 7 8 9 10 11 12

29.01.2007 – Vorlesung Nr. 13

6. Semantic Web Anwendungen

- 6.1 Semantic Web und das WWW?
- 6.2 Semantic Wiki
- 6.3 Semantische Suche
- 6.4 War's das...?

6.1 Semantic Web und das WWW

- Das WWW heute (Stand: 09/2006)
 - O Das World Wide Web bietet Zugriff auf ein gigantisches "Informationsuniversum"
 - O Schätzungen gehen von mehr als 55 Milliarden Dokumenten im WWW aus (Surface Web)
 - etwa 25 Mrd. Dokumente des WWW sind in Suchmaschinen wie z.B. Google indiziert (Stand 06/2006)
 - DeepWeb (Darkweb) schätzungsweise bis zu 550 mal größer als Surface Web (Bergman, 2001)

O Probleme:

- Wie finde ich die richtigen (wichtigen) Informationen (vollständig)?
- Wie sorge ich dafür, dass jemand die von mir angebotenen Informationen tatsächlich findet?
- Kann die Informationssuche (zuverlässig) automatisiert werden?
- Kann ich der angebotenen Information auch tatsächlich trauen?

6.1 Semantic Web und das WWW

- Das Ende des WWW?
 - → Das WWW ist zur Nutzung durch den Menschen bestimmt
 - O Problemfelder:
 - [1] Informations suche
 - → "wo finde ich…?"
 - [2] Informationsextraktion
 - → "was ist wichtig?"
 - [3] Wartung
 - → "sind meine Daten (noch) konsistent?"
 - [4] Personalisierung
 - → "was mich interessiert…"

6.1 Semantic Web und das WWW

- Wie kann das Semantic Web helfen?
 - O das Semantic Web soll dem Menschen helfen,
 - genau die Informationen zu bekommen, die er tatsächlich haben möchte,
 - viele bislang manuell auszuführende Einzelaufgaben, die im Rahmen der Erledigung komplexer Aufgaben anfallen, zu integrieren
 - und diese, soweit möglich, zu automatisieren.
 - O dazu bietet das Semantic Web
 - ein einheitliches Adressierungs- und Kodierungsschema,
 - eine standardisierte Formalisierung und Kodierung von Semantik auf unterschiedlichen Ebenen,
 - eine Möglichkeit der Formalisierung und Kodierung der Zuverlässigkeit der dargebotenen Informationen.

Anzeigen

Neuheiten & Angebote fürs Golfspie

Schläger und mehr, günstig online!

(W)hole in one im Golfland Kärnten

Golfbälle sammeln

golfballsammler.de

www.sporthouse24.de

Gratiskatalog Golfen 2005

www.golfland.kaernten.at

Internationale Golf Card

www.golfcards.com

Golfferien in Thailand

www.golfthailand.ch

9 Championship Golfplätze, unschlagbar günstiger Preis

Sie kennen alle im Club?

www.Golfkontakte.de

Discount Golf Shop

Jetzt schon online bestellen!

Super-günstig! bis 40% billiger

Golf Ausrüstung zu niedrig Preisen

reduz, bis zu 50%Greenfee bei 5.940

Golfclubs, Golfresorts, Hotelkosten

Treffen Sie Golfer aus Ihrem Club &

Ihrer Region. Anmelden & Verabreden

www.onlineaolf.co.uk/willkommen

stilvolle Präsentation mit den

Golfballsetzkästen von mdesign+

6.1 Semantic Web und das WWW

[1] Informations suche



 traditionelle Keyword-basierte Suche führt zu vielen nicht relevanten Ergebnissen = mangeInde ,precision'

- Homonyme
- Kontext
- unpräzise Anfrage
- traditionelle Keyword-basierte Suche findet nicht alle relevanten Ergebnisse = mangeInder ,recall'
 - Synonyme
 - Metaphern und Umschreibungen
 - unvollständige Anfrage

Semantic Web

6.1 Semantic Web und das WWW

[1] Informations suche

- O Semantische Suche = kontext-basierte Suche
- O Dazu muss es möglich sein, diesen Kontext zu bestimmen und/oder diesen zu präzisieren
- O Was ist notwendig zur semantischen Suche?
 - Semantisch annotierte Ressourcen
 - Semantisch annotierten Such-Index
 - ,intelligente Suchstring-Auswertung (Query Evaluation)
 - ,intelligente' Suchstring-Erweiterung und -Präzisierung (Query Expansion)
 - Vorschlag von alternativen, verwandten Suchergebnissen (Tag-Browsing)
 - Visualisierung und Navigation im Suchraum / Suchergebnisraum

6.1 Semantic Web und das WWW

[2] Informationsextraktion



- kann nur von menschlichem
 "Agenten" korrekt durchgeführt und bewertet werden
- Software-Agent verfügt nicht über
 - (kulturelles) Kontextwissen
 - Weltwissen
 - Wissen über Nutzerpräferenzen
- um Informationen aus der Text- / Bilddarstellung zu lösen

6.1 Semantic Web und das WWW

[2] Informationsextraktion

- O Semantisch annotierte Ressourcen ermöglichen
 - Eindeutige Bestimmung der beschriebenen Konzepte
 - Bestimmung von Beziehungen und Abhängigkeiten der beschriebenen Konzepte
 - (eingeschränkte) Bestimmung des Kontextes
- O Die semantische Annotation ist dabei unabhängig von der medialen Ausprägung der Ressource (Text / Bild / Grafik)
- O Die Qualität der extrahierten Information hängt von der Qualität der semantischen Annotation ab

6.1 Semantic Web und das WWW

[3] Wartung



- je komplexer und größer die WebSite, desto aufwändiger die Wartung der oft nur schwach-strukturierten Daten
- Problem:
 - Konsistenz
 - Korrektheit
 - Aktualität
- insbesondere in Bezug auf deren Semantik

6.1 Semantic Web und das WWW

[3] Wartung

- O Semantisch annotierte Ressourcen ermöglichen
 - Automatische Überprüfung der Konsistenz Ändert sich der Inhalt einer vernetzten Ressource kann überprüft werden, ob dieser Inhalt (sofern er semantisch annotiert wurde) zu dessen (semantischen) Beschreibung der auf diesen verweisenden Ressource passt

Problem:

Inkonsistenzen können zwar (automatisch) erkannt, aber nur in beschränktem Umfang (automatisch) korrigiert werden.

6.1 Semantic Web und das WWW

[4] Personalisierung



- Anpassung des dargestellten Informationsinhalts an die persönlichen Bedürfnisse und Anforderungen des Nutzers
- **Problem:**"woher kommen die benötigten
 Informationen?"

Informationsanbieter vs. Informationskonsument

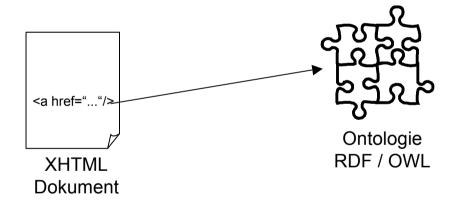
6.1 Semantic Web und das WWW

[4] Personalisierung

- O Persönliche Daten des Nutzers können
 - semantisch annotiert werden,
 - differenziert (je nach Wunsch oder Berechtigung) an einen entsprechend autorisierten Kommunikationspartner weitergegeben werden,
 - zuverlässig authentifiziert werden
- O Semantisch annotierte persönliche Daten können
 - zur Kontextpräzisierung dienen,
 - helfen, bedarfsgerecht zwischen den Bedürfnisse des Informationskonsumenten mit den Angeboten des Informationsproduzenten zu vermitteln

6.1 Semantic Web und das WWW

- Bindeglieder zwischen WWW und Semantic Web
 - O Semantische Annotationen können mit WWW-Dokumenten verbunden werden über
 - Hyperlinks



- RDFa
- GRDDL

15

6.1 Semantic Web und das WWW

RDFa

- O Erweiterung von XHTML
- O standardisiert durch W3C (Working Draft, 10/2006)
- O ermöglich einfache semantische Annotation von XHTML Dokumenten
- O generalisiert Attribute von XHTML <meta> und <link> Elementen, Attribute können so in jedem XHTML-Element genutzt werden
- O einfaches Mapping zur automatischen Extraktion von RDF-Tripeln aus RDFa-annotierten XHTML-Dokumenten

6.1 Semantic Web und das WWW

RDFa

O RDFa Attribute:

about
 URI, die Angaben zur vorliegenden Ressource

enthält

• rel, rev, href spezifiziert Relation zu einer anderen Ressource

property
 spezifiziert property für den Inhalt eines Elements

content optional

überschreibt den Inhalt eines Elements (falls property-Attribut verwendet wird)

datatype optional

spezifiziert Datentyp des Inhalts eines Elements

17

6.1 Semantic Web und das WWW

RDFa - Beispiele

```
<div xmlns:dc="http://purl.org/dc/elements/1.1/"
    about="http://www.minet.uni-jena.de/~sack/WWWBuch">

    <span property="dc:title">
        WWW - Kommunikation, Internetworking, Webtechnologien
    </span>
    <span property="dc:author">Christoph Meinel</span>
        <span property="dc:author">Harald Sack</span>
        <span property="dc:author">Harald Sack</span>
        <span property="dc:date">2004-01-01</span>
</div>
```

6.1 Semantic Web und das WWW

RDFa - Beispiele

```
RDF-Tripel:

("http://www.minet.uni-jena.de/~sack/WWWBuch", dc:title,
    "WWW - Kommunikation, Internetworking, Webtechnologien")

("http://www.minet.uni-jena.de/~sack/WWWBuch", dc:author,
    "Christoph Meinel")

("http://www.minet.uni-jena.de/~sack/WWWBuch", dc:author,
    "Harald Sack")

("http://www.minet.uni-jena.de/~sack/WWWBuch", dc:date,
    "2004-01-01")
```

6.1 Semantic Web und das WWW

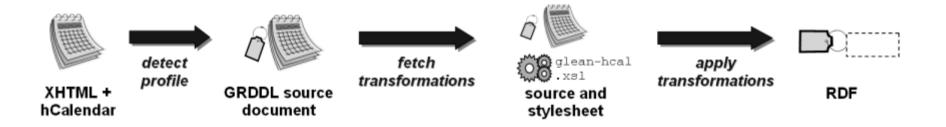
RDFa - Beispiele

```
In ihrem Buch <em property="dc:title"> WWW - Kommunikation,
Internetworking, Webtechnologien </em>, das am
<span property="dc:date">2004-01-01</span> erschienen ist,
unternehmen die beiden Autoren
<span property="dc:author">Christoph Meinel</span> und
<span property="dc:author">Harald Sack</span> den Versuch,
das gesamte WWW zu beschreiben.
```

6.1 Semantic Web und das WWW

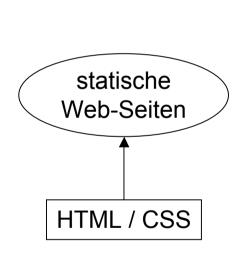
GRDDL

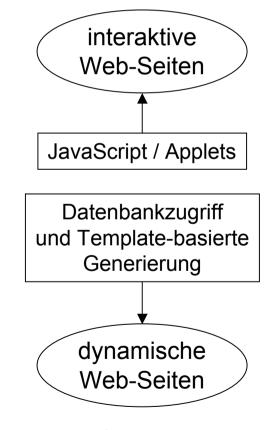
- O Gleaning Resource Descriptions from Dialects of Languages
- O Extraktion von RDF Daten aus XML- und XHTML-Dokumenten (z.B. microformats) mit Hilfe von explizit definierten Transformationsalgorithmen (XSLT)
- O im W3C Standardisierungsprozess (Working Draft, 10/2006)

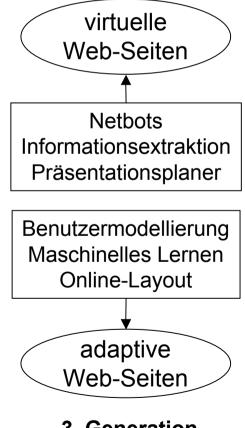


6.1 Semantic Web und das WWW

Die drei Generationen von Web-Dokumenten





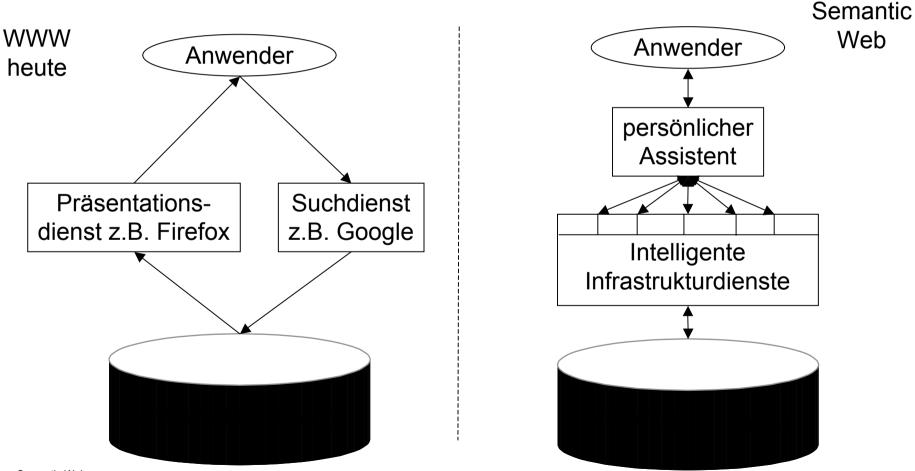


1. Generation

2. Generation

6.1 Semantic Web und das WWW

Intelligente Agenten im Semantic Web



Semantic Web

Semantic Web

1 2 3 4 5 6 7 8 9 10 11 12

29.01.2007 – Vorlesung Nr. 13

6. Semantic Web Anwendungen

- 6.1 Semantic Web und das WWW?
- 6.2 Semantic Wiki
- 6.3 Semantische Suche
- 6.4 War's das...?

24

6.2 Semantic Wiki

- Wo kommen die notwendigen semantischen Daten her?
 - O ,klassischer' Weg: manuelle Annotation durch ausgewiesene Spezialisten

(Indexierer)

O ,praktikabler' Weg: manuelle Annotation durch die Ressourcenautoren

(oft nicht Spezialist)

O ,billiger' Weg: Nutzer annotieren Ressourcen gemeinsam

(Spezialisten und Laien)

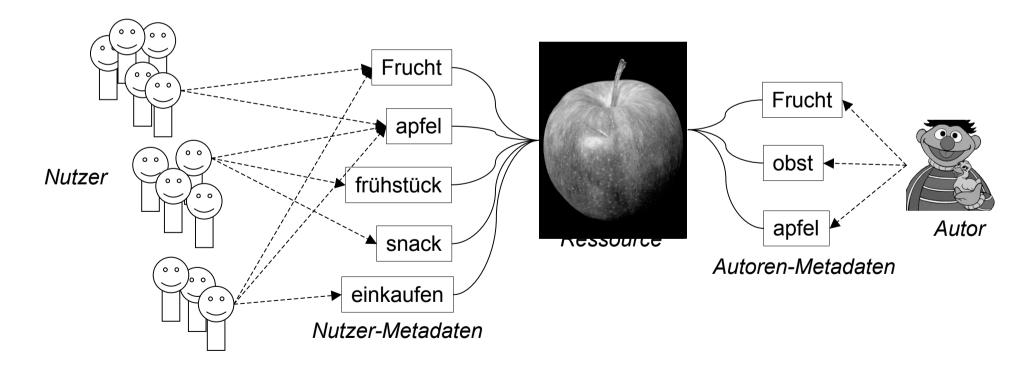
- O Probleme:
 - Aufwand (Kosten) und Skalierbarkeit
 - Qualität der Annotation
 - Objektivität
 - Personalisierung

6.2 Semantic Wiki

- Gemeinsame Erstellung von Metadaten
 - O ,billige' Variante: Social Tagging-Systeme
 - Beliebige Ressourcen können mit Tags (Schlüsselworten) annotiert werden
 - Problem: Tags meist beliebig aus freiem Vokabular wählbar (Homonyme, Synonyme, unterschiedliche Schreibweisen)
 - Komplexere Zusammenhänge können nicht modelliert werden
 - Möglichkeit der Personalisierung
 - O ,premium' Variante: Semantic Wikis
 - Gemeinschaftliche Erstellung von komplexen semantischen Annotationen (Ontologien)

6.2 Semantic Wiki

Social Tagging Systeme



6.2 Semantic Wiki

Social Tagging Systeme

○ Tags

 Nutzerdefinierte, deskriptive Metadaten (meist) in Form eines einzelnen Wortes (String)

○ Tagging

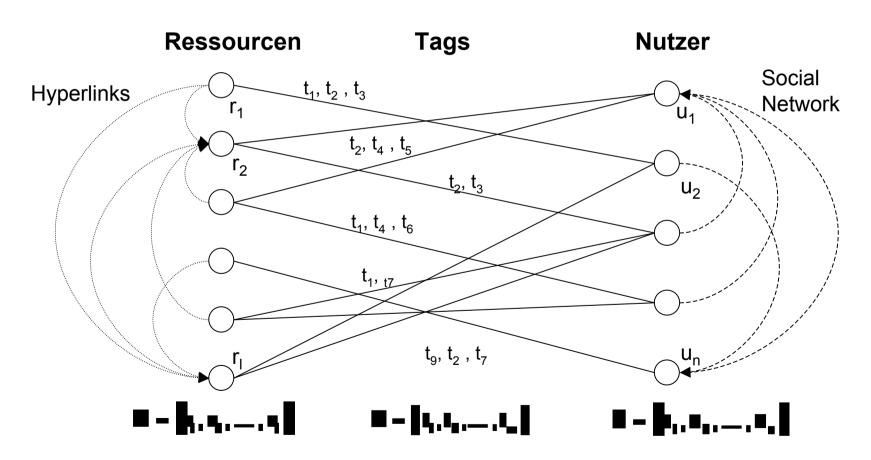
 Vorgang, in dem Nutzer deskriptive Metadaten zu eigenen und fremden Ressourcen zuordnen

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

6.2 Semantic Wiki

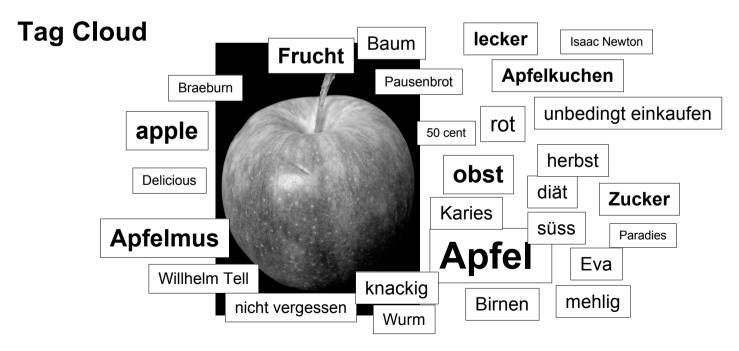
Social Tagging Systeme



6.2 Semantic Wiki

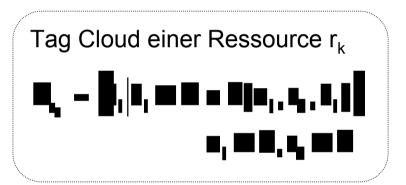
Folksonomies

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



6.2 Semantic Wiki

- Folksonomies und Tag Clouds
 - O Tags können **gewichtet** werden
 - Anzahl
 - soziales Netzwerk
 - Reputation / Trust
 - Vorkommen in der damit ausgezeichneten Ressource



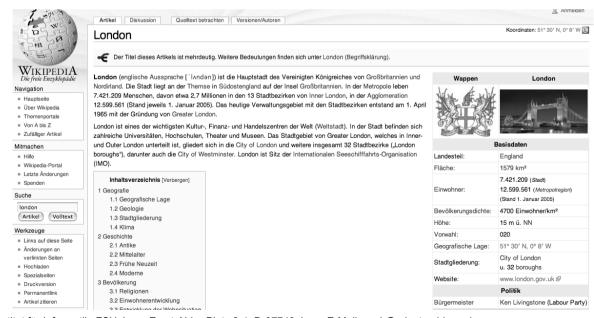
- O Tag Cloud besitzt **keine Struktur** = flach
- O Struktur kann über **Data Mining Techniken** gewonnen werden
 - z.B. co-occurrence



31

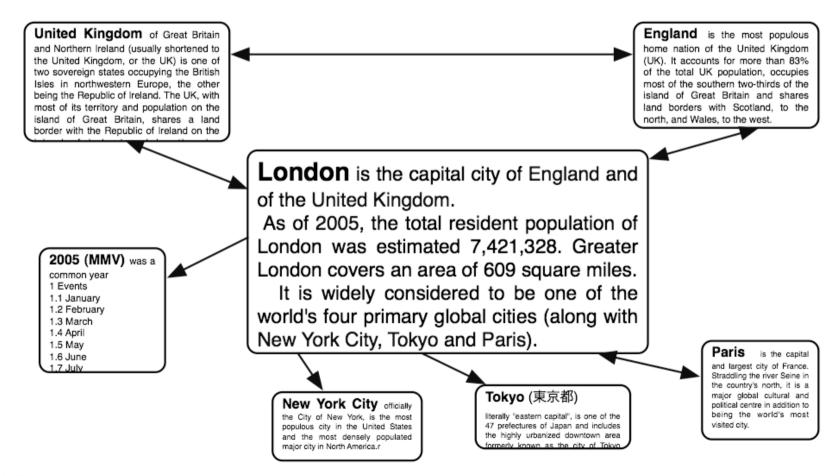
6.2 Semantic Wiki

- Semantic Wikipedia
 - O Dokumente der Wikipedia
 - enthalten unstrukturierten freien Text und Multimedia-Objekte
 - sind über Hyperlinks miteinander verbunden
 - können selbstdefinierten Kategorien angehören
 - können strukturierte Daten enthalten



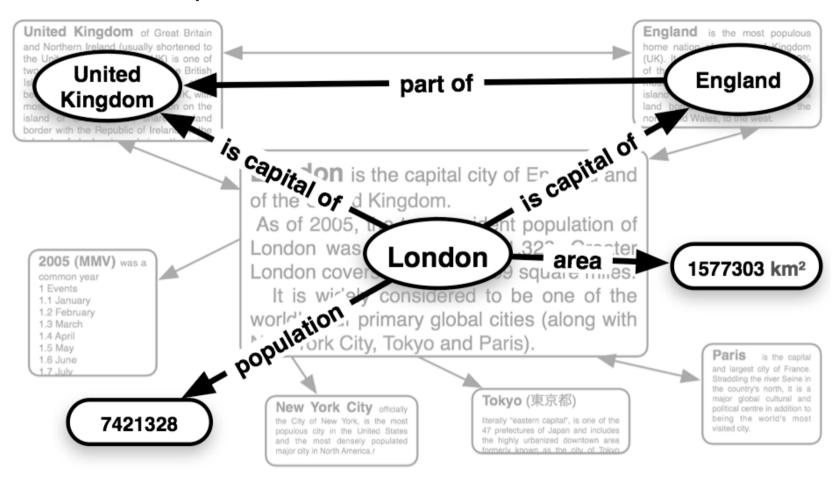
6.2 Semantic Wiki

Semantic Wikipedia



6.2 Semantic Wiki

Semantic Wikipedia



6.2 Semantic Wiki

- Semantic Wikipedia
 - O Einfache Erweiterung der WikiSyntax für
 - Typisierte Links zur Klassifikation von Links zwischen Artikeln entsprechend ihrer Bedeutung
 - Kategorien klassifizieren Artikel entsprechend ihrem Inhalt (z.T. bereits in Wikipedia integriert)
 - Attribute zur Angabe einfacher Eigenschaften (Properties), die sich auf den Inhalt der Artikel beziehen

6.2 Semantic Wiki

Semantic Wikipedia



http://ontoworld.org/wiki/London

'''London''' is the capital city of [[England]] and of the [[United Kingdom]]. As of [[2005]], the total resident population of London was estimated 7,421,328. Greater London covers an area of 609 square miles. [[Category:City]]

Mediawiki Syntax

```
'''London''' is the capital city of [[capital of::England]] and of the [[is capital of::United Kingdom]]. As of [[2005]], the total resident population of London was estimated [[population:=7,421,328]]. Greater London covers an area of [[area:=609 square miles]]. [[Category:City]]
```

Semantic Mediawiki Syntax

6.2 Semantic Wiki

Semantic Wiki Projekte

COW

<u>IkeWiki</u> <u>Rise Semantic Wikipedia</u>

Rhizome

KaukoluWiki SeMediaWiki

<u>KawaWiki</u> <u>SemPerWiki</u> KendraBase SemWiki

<u>KendraBase</u> <u>SemWiki</u> Makna SweetWiki

<u>OntoWiki</u> <u>Wekiwi</u>

<u>OpenRecord</u> <u>WikiOnt</u>

<u>PlatypusWiki</u> <u>WikSAR</u>

<u>POWL</u>

Semantic Web

1 2 3 4 5 6 7 8 9 10 11 12

29.01.2007 – Vorlesung Nr. 13

6. Semantic Web Anwendungen

- 6.1 Semantic Web und das WWW?
- 6.2 Semantic Wiki
- 6.3 Semantische Suche
- 6.4 War's das...?

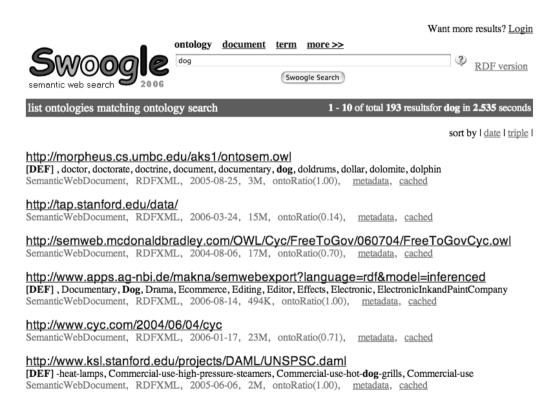
38

- Was ist eine semantische Suchmaschine?
 - O Unterscheide zwei Varianten:
 - Suche nach semantischen Daten
 - z.B. SWOOGLE
 Suche nach RDF-Daten, OWL-Ontologien, etc.
 - 2. Inhaltsbasierte, durch Semantic Web Technologien unterstützte Suche
 - z.B. SWSE, NPBibSearch, ...
 Suche nach Web-Dokumenten, wobei das Auffinden / Filtern der Ergebnisse mit Hilfe von Semantic Web Technologie erfolgt

6.3 Semantische Suche

Suche nach semantischen Daten

Suchmaschine SWOOGLE



- Suche dan RDF / OWL Daten
- Crawler durchsucht HTML-Dokumente nach eingebettetem RDF / OWI
- RDF Export und Schnittstelle für SW-Agenten
- eigenes Ranking-Verfahren (ähnlich PageRank)
- an der University of Maryland entwickelt

- Wie funktioniert eine semantische Suchmaschine?
 - O Inhaltsbasierte, durch Semantic Web Technologien unterstützte Suche
 - Semantic Web Unterstützung in allen einzelnen Phasen des Suchmaschinen-Prozesses
 - Crawling, Indexing, Nutzerinterface, Filterung, Ranking, ...
 - ermöglicht das Auffinden von neuen Querverbindungen zwischen Ergebnisdokumenten
 - automatische Disambiguierung und Kategorisierung von Suchergebnissen
 - über geeignetes semantisches Nutzerprofil können Filterung und Ranking der Suchergebnisse personalisiert werden
 - ermöglicht erweiterte Navigation innerhalb des Suchraumes

- Wie funktioniert eine semantische Suchmaschine?
 - Crawling
 - gezieltes Auffinden von Webdokumenten nach
 - Themengebieten
 - Vorgegebenen Kategorien, z.B.
 - Personendaten
 - Geografische Daten
 - Produktdaten
 - Finanzdaten
 - verteilte Crawler können nicht nur nach geografischer Lage, sondern auch nach vorgegebenen Themengebieten aufgeteilt werden

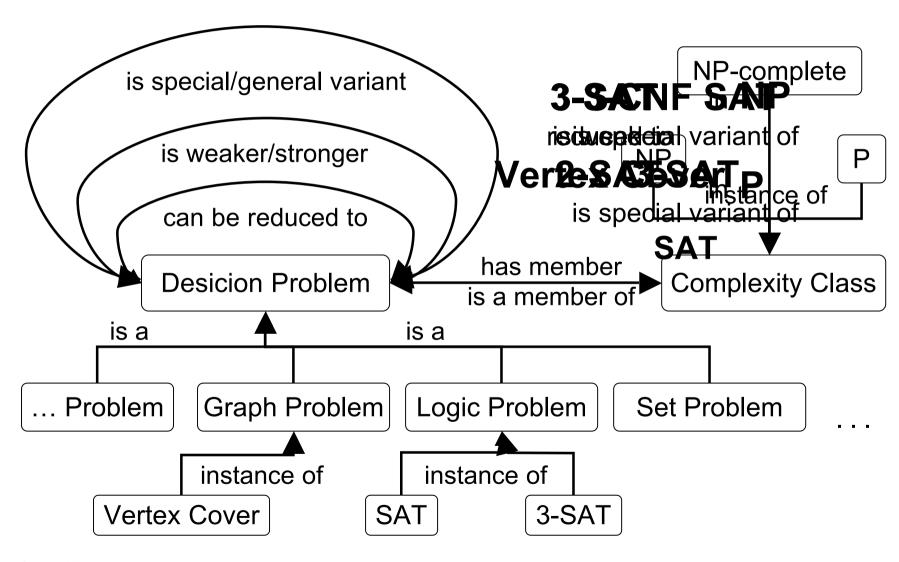
- Wie funktioniert eine semantische Suchmaschine?
 - Indexing
 - Aufbau eines semantischen Indexes
 - Webdokumente werden im Index Ontologien zugeordnet
 - Webdokumente mit eingebetteten semantischen Daten müssen einander zugeordnet werden
 - Ontology Mapping / Ontology Merging
 - Suchmaschine kann eigene Upper-Ontology für den Index verwalten
 - ermöglicht das Auffinden von neuen Querverweisen
 - Authentifikation der im Index verwalteten Webdokumente / Ontologien
 - Ermittelung der Reputation von Webdokumenten

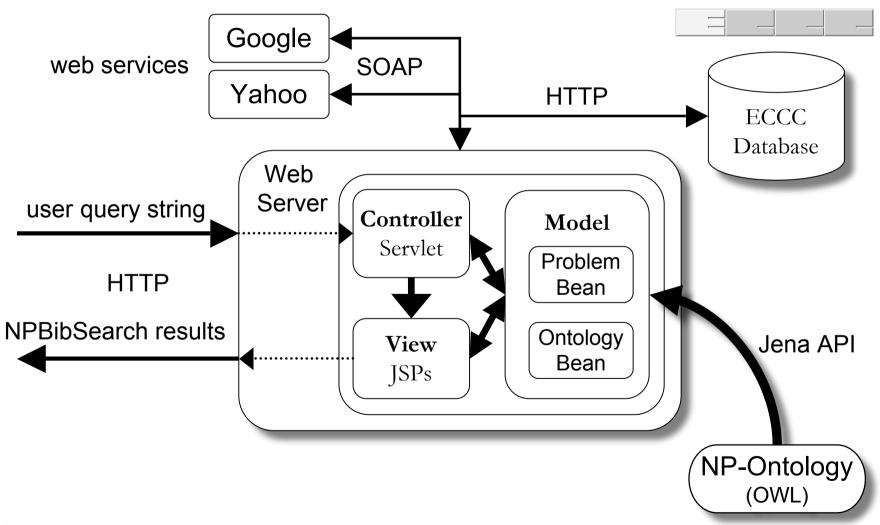
- Wie funktioniert eine semantische Suchmaschine?
 - Nutzerinterface
 - Query String Evaluation
 - Ermittelung der Absicht / Suchanfrage des Nutzers
 - Welche Konzepte sollen ermittelt werden?
 - Kann ein Kontext bestimmt werden?
 - Natural Language Processing (NLP) bei natürlichsprachlichem Nutzerinterface
 - Disambiguierung
 - Identifikation von Synonymen, Metaphern, Umschreibungen, Akronymen, alternativen Schreibweisen
 - Fehlertoleranz
 - Ähnlichkeitssuche

- Wie funktioniert eine semantische Suchmaschine?
 - Nutzerinterface
 - Query String Refinement
 - Soll die Suchabfrage präzisiert (eingeengt oder erweitert) werden?
 - Ermittlung des Suchkontextes
 - Vorschlag von Suchalternativen
 - Navigation im Suchraum
 - Ähnlichkeitssuche

- Wie funktioniert eine semantische Suchmaschine?
 - Filterung und Ranking
 - Personalisierung
 - Nutzerpräferenzen müssen als semantische Annotation vorliegen (z.B. werden in der Suchabfrage automatisch an die Suchmaschine übermittelt)
 - Entsprechend des übermittelten Nutzerprofils lassen sich Suchergebnisse filtern und deren Reihenfolge (Ranking) bestimmen
 - Browsing
 - Der Fortgang der Suche wird auf Basis der ermittelten (und vom Nutzer ausgewählten) Suchergebnisse bestimmt
 - Recommendation Systeme

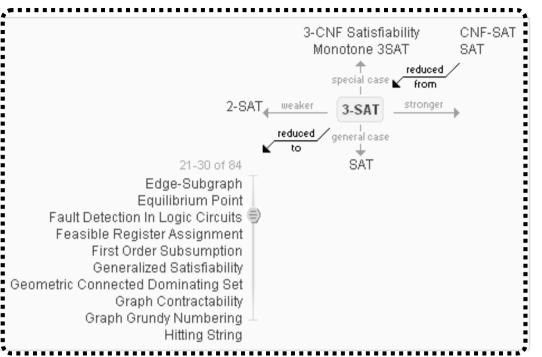
- Semantische Suchmaschinen
 - NPBibSearch
 - Semantische Suchmaschine für den Bereich NP-vollständiger Probleme (Komplexitätstheorie, theoretische Informatik)
 - Suche in eingeschränkter Domain (bibliografische Daten)
 - Kombination aus
 - regulärer Suchmaschine (Google + Yahoo Web Service Interace),
 - Semantic Web Technologien (NP-Ontologie/T-Box, >350 Instanzen/A-Box aus Garey-Johnson, JENA API)
 - und spezialisierten Datenbanken (ECCC, CiteSeer, dblp)

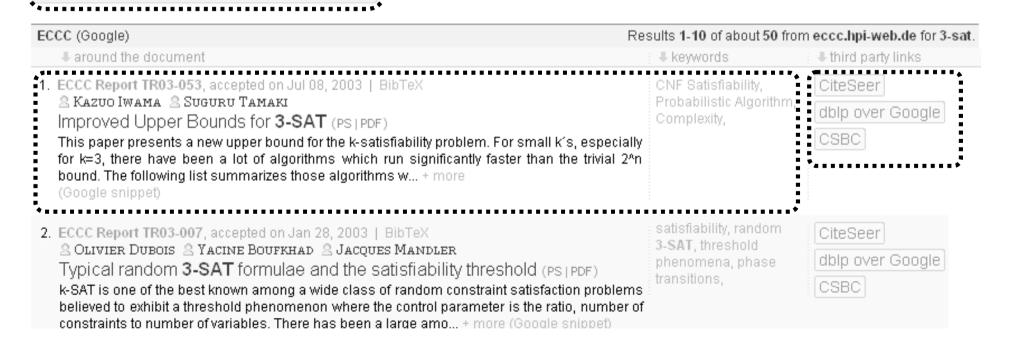






New York, 151-158, 1971.





Semantic Web

1 2 3 4 5 6 7 8 9 10 11 12

29.01.2007 – Vorlesung Nr. 13

6. Semantic Web Anwendungen

- 6.1 Semantic Web und das WWW?
- 6.2 Semantic Wiki
- 6.3 Semantische Suche
- 6.4 War's das...?

6.4 War's das...?

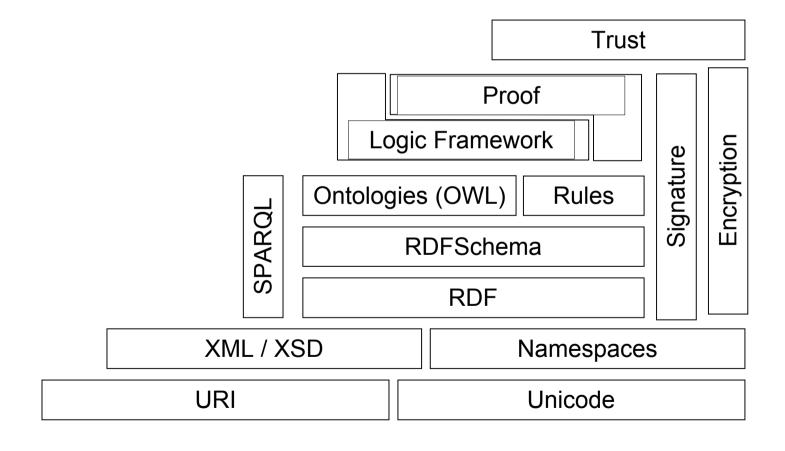
6.4 War's das...?

- O 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 WWW werden sich gegenseitig ergänzen



"Now! ... That should clear up a few things around here!"

Semantic Web Architecture



Semantic Web - Vorlesungsinhalt

- 1. Einführung
- 2. Die Sprachen des Semantic Web
- 3. Wissensrepräsentation
- 4. Ontology Engineering
- 5. Web of Trust
- 6. Semantic Web Anwendungen

Literatur



 Materialien-Webseite http://www.informatik.uni-jena.de/~sack/WS0607/semanticwebmaterialien.htm



bibsonomy - Bookmarks
 http://www.bibsonomy.org/user/lysander07