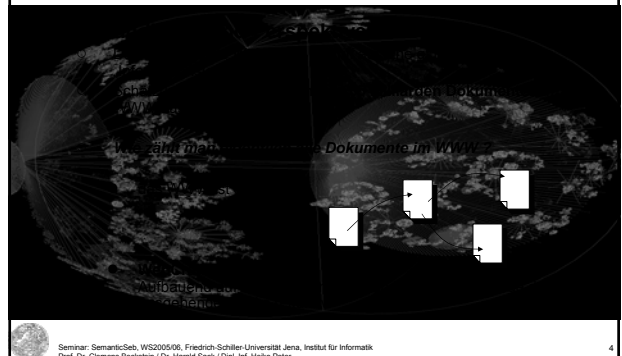


Semantic Web

- Seminar
- Wintersemester 2005/2006
- Prof. Dr. Clemens Beckstein
Dr. Harald Sack
- Dipl. Inf. Heiko Peter
- Friedrich-Schiller-Universität Jena

Die Grenzen des World Wide Web



Semantic Web

Die Grenzen des World Wide Web und Die Vision des Semantic Web

Dr. rer. nat. Harald Sack
Institut für Informatik
FSU Jena

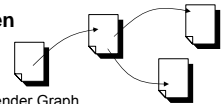
Seminar: SemanticSeb, WS2005/06, Friedrich-Schiller-Universität Jena, Institut für Informatik

2

Die Grenzen des World Wide Web

● Grundlagen und Perspektiven

- das WWW ist ein Graph
- Problem:
das WWW ist KEIN zusammenhängender Graph
- Konnektivität des WWW
 - **Schwach verbundene Komponenten**
 - Hyperlinks werden als **ungerichtete Kanten** interpretiert
 - einzelne Komponente mit ca. **90%** aller untersuchten Dokumente
 - **Stark verbundene Komponenten**
 - Hyperlinks werden als **gerichtete Kanten** interpretiert
 - **28%** der untersuchten Dokumente bilden eine stark verbundene **Kernkomponente (SCC)**



Seminar: SemanticSeb, WS2005/06, Friedrich-Schiller-Universität Jena, Institut für Informatik
Prof. Dr. Clemens Beckstein / Dr. Harald Sack / Dipl. Inf. Heiko Peter

5

Semantic Web

Die Grenzen des World Wide Web und Die Vision des Semantic Web

- Grundlagen und Perspektiven
- Das Ende des WWW?
- Die Vision des Semantic Web
- Wie das Semantic Web die Welt verändert

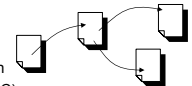
Seminar: SemanticSeb, WS2005/06, Friedrich-Schiller-Universität Jena, Institut für Informatik

3

Die Grenzen des World Wide Web

● Grundlagen und Perspektiven

- Die Struktur des WWW
 - 90 % schwach verbundene Komponenten
 - 28 % stark verbundene Komponente (SCC)
 - Durchmesser SCC: > 27
- Durchmesser des Gesamtgraphen: > 500
- WS, dass zwischen zwei zufällig ausgewählten Dokumenten ein Pfad besteht: 0,24



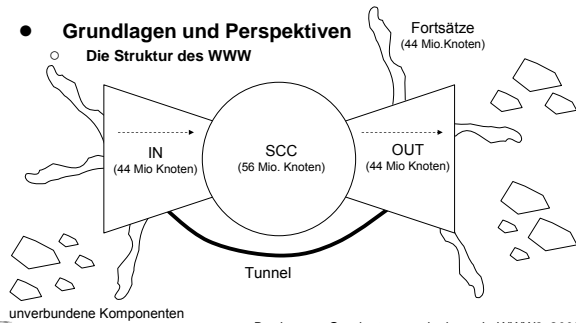
Broder e.a.: Graph structure in the web, WWW9, 2000

Seminar: SemanticSeb, WS2005/06, Friedrich-Schiller-Universität Jena, Institut für Informatik
Prof. Dr. Clemens Beckstein / Dr. Harald Sack / Dipl. Inf. Heiko Peter

6

Die Grenzen des World Wide Web

- **Grundlagen und Perspektiven**
Die Struktur des WWW



Broder e.a.: Graph structure in the web, WWW9, 2000

Seminar: SemanticSeb, WS2005/06, Friedrich-Schiller-Universität Jena, Institut für Informatik
Prof. Dr. Clemens Beckhaus / Dr. Marco Stock / Dipl.-Inf. Heiko Peine

7

Die Grenzen des World Wide Web

- **Das Ende des WWW?**

- Was nutzt uns die ganze Informationsfülle?

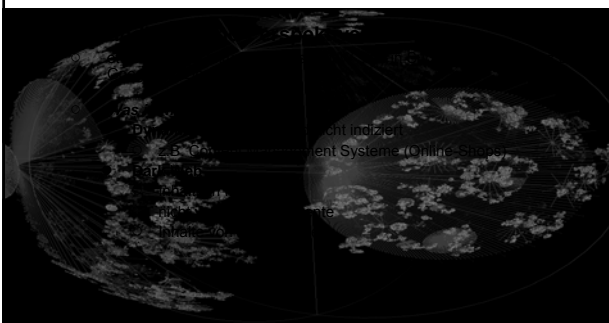


- was ist wichtig, was nicht?
- was ist Information, was ist Werbung?
- was bedeutet die Information?
- wie glaubhaft ist die Information?
- was gehört zusammen?
- was ist redundant?

Seminar: SemanticSeb, WS2005/06, Friedrich-Schiller-Universität Jena, Institut für Informatik
Prof. Dr. Clemens Beckhaus / Dr. Marco Stock / Dipl.-Inf. Heiko Peine

10

Die Grenzen des World Wide Web



Seminar: SemanticSeb, WS2005/06, Friedrich-Schiller-Universität Jena, Institut für Informatik
Prof. Dr. Clemens Beckhaus / Dr. Marco Stock / Dipl.-Inf. Heiko Peine

8

Die Grenzen des World Wide Web

- **Das Ende des WWW?**

- Was nutzt uns die ganze Informationsfülle?



Mensch verfügt über **Kontextwissen / Weltwissen** und kann daher Information (meist) richtig deuten...

Seminar: SemanticSeb, WS2005/06, Friedrich-Schiller-Universität Jena, Institut für Informatik
Prof. Dr. Clemens Beckhaus / Dr. Marco Stock / Dipl.-Inf. Heiko Peine

11

Die Grenzen des World Wide Web

- **Grundlagen und Perspektiven**

- Dokumentenbestand im WWW verdoppelt sich etwa **alle 6 Monate**
- ein Ende dieser Entwicklung ist zunächst nicht absehbar
- **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?*

→ Ist das WWW bereits am Ende?

Seminar: SemanticSeb, WS2005/06, Friedrich-Schiller-Universität Jena, Institut für Informatik
Prof. Dr. Clemens Beckhaus / Dr. Marco Stock / Dipl.-Inf. Heiko Peine

9

Die Grenzen des World Wide Web

- **Das Ende des WWW?**

- Was nutzt uns die ganze Informationsfülle?
- → Das WWW ist zur **Nutzung durch den Menschen** bestimmt
- WWW basiert auf der Markupsprache HTML
- HTML beschreibt
 - **wie** Informationen **dargestellt** werden sollen,
 - **wie** Informationen miteinander **verknüpft** werden können,
 - aber **nicht, was diese Informationen bedeuten...**



Semantik ???

Seminar: SemanticSeb, WS2005/06, Friedrich-Schiller-Universität Jena, Institut für Informatik
Prof. Dr. Clemens Beckhaus / Dr. Marco Stock / Dipl.-Inf. Heiko Peine

12

Die Grenzen des World Wide Web

- **Semantik**

Die **Semantik (Bedeutungslehre)** ist das Teilgebiet der Sprachwissenschaft (*Linguistik*), das sich mit

- Sinn und
- Bedeutung

von **Sprache** beziehungsweise **sprachlichen Zeichen** befasst.

Die Semantik kümmert sich um die Frage, wie Sinn und Bedeutung von komplexen Begriffen aus denen von einfachen Begriffen abgeleitet werden können und stützt sich dabei in der Regel auf die **Syntax**.

Die Grenzen des World Wide Web

- **Das Ende des WWW?**

- → Das WWW ist zur **Nutzung durch den Menschen** bestimmt
- **Problemfelder:**
 - [1] Informationssuche
→ „wo finde ich...?“
 - [2] Informationsextraktion
→ „was ist wichtig?“
 - [3] Wartung
→ „sind meine Daten (noch) konsistent?“
 - [4] Personalisierung
→ „was mich interessiert...“

Die Grenzen des World Wide Web

- **Das Ende des WWW?**

- → Das WWW ist zur **Nutzung durch den Menschen** bestimmt



Programm (z.B. Robot) kann nicht unterscheiden zwischen wichtiger / unwichtiger Information

Die Grenzen des World Wide Web

- **Problemfeld 1: Informationssuche**



- Keyword-Suche führt zu **vielen nicht relevanten Ergebnissen**
 - Bedeutungsunterschiede
 - Homonyme
 - unterschiedlicher Kontext
- Keyword-Suche findet **nicht alle relevanten Ergebnisse**
 - Synonyme
 - fehlende Kontextpräzisierung

⇒ Einbeziehung von formalen Terminologien / **Ontologien**

Die Grenzen des World Wide Web

- **Das Ende des WWW?**

- → Das WWW ist zur **Nutzung durch den Menschen** bestimmt



Wir benötigen Information über die dargebotene Information → **Metadaten (!)**

Werbung ?!

HTML bietet nur sehr limitierte Möglichkeiten zur Auszeichnung von Metadaten ...

Die Grenzen des World Wide Web

- **Problemfeld 2: Informationsextraktion**



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

⇒ **Aggregation und Integration von Informationsquellen aus unterschiedlichen Ressourcen**

Die Grenzen des World Wide Web

Problemfeld 3: Wartung



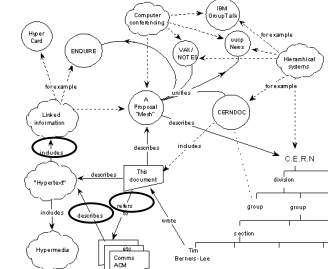
• je komplexer und größer die WebSite, desto aufwändiger die Wartung der oft nur schwach-strukturierten Daten

- **Problem:**
- Konsistenz
 - Korrektheit
 - Aktualität

• maschinelle Wartung benötigt **Information über SEMANTIK**

Die Vision des Semantic Web

Die ursprüngliche Vision des WWW



• ursprüngliche Beschreibung des WWW **beschreibt unterschiedliche Beziehungen zwischen Dokumenten**

- describes
- includes
- refers to
- ...

• diese Beziehungen werden heute im WWW als nicht-attribuiertes Hyperlink ausgedrückt

→ Semantik ging verloren!

Die Grenzen des World Wide Web

Problemfeld 4: Personalisierung



• Anpassung des dargestellten Informationsinhalts an die **persönlichen Bedürfnisse und Anforderungen**

- **Problem:**
- „woher kommen die benötigten Informationen?“

• Informationsanbieter vs. Informationskonsument

• erfordert Möglichkeit der **Repräsentation von SEMANTIK**

Die Vision des Semantic Web

Die ursprüngliche Vision des WWW



• **Grundgedanke:**
WWW als universeller Wissensraum

Alles kann miteinander verlinkt werden

• **Zweck: Zusammenarbeit**

“Communication by shared knowledge”

benötigt intuitiv zu bedienendes Interface!

Die Vision des Semantic Web

Die ursprüngliche Vision des WWW



„The Web was designed as an information space, with the goal that it should be useful not only for human-human communication, but also that machines would be able to participate and help...“

Tim Berners-Lee, Semantic Web Roadmap, Sept. 1998

HTML als Sprache zur Informationsdarstellung im WWW besitzt **keine Möglichkeit, die Bedeutung** der dargestellten Information **auszudrücken**



Das volle Potenzial des WWW wird nicht genutzt...

Tim Berners-Lee

Die Vision des Semantic Web

Die Vision des Semantic Web

- 1997, WWW7 in Brisbane, Australien
 - TBL stellt seine Vision eines „Global Reasoning Web“ vor
- 1998, WWW8 in Toronto, Kanada
 - TBL präsentiert seine „Semantic Web Roadmap“
- Was ist das Semantic Web?
- Was kann das Semantic Web?
- Was ist das Semantic Web nicht?
- Wird das Semantic Web unser Leben verändern?
- Gibt es eine Semantic Web Killer-Applikation?



Tim Berners-Lee

Die Vision des Semantic Web

- Was ist das Semantic Web?
 - „The Semantic Web is a web of data, in some ways like a global database.“
 - „...a universal Web of semantic assertions“
 - „...emphasises the intelligent manipulation of the knowledge that is made available via the web.“
 - „The Semantic Web is an extension of the current web in which information is given well-defined meaning, better enabling computers and people to work in cooperation“

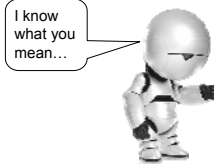
Die Vision des Semantic Web

- Was kann das Semantic Web?

- **Stufe 1: Textdokumente und Datenbankeinträge**
Daten sind Anwendungsbezogen (proprietär), Semantik steckt in den jeweiligen Anwendungen, nicht in den Daten
 - z.B. Warenkatalog bei Amazon, Index von Google, Datenbestand bei Ebay, ...
- **Stufe 2: XML-Dokumente mit einheitlichem Vokabular**
Innerhalb bestimmter vorgegebener Domäne erreichen Daten **Unabhängigkeit von der Anwendung** und können so **zwischen Anwendungen ausgetauscht** werden
 - z.B. Dublin-Core Bibliografien zur unabhängigen Beschreibung von Dokumenten, elektronische Krankenakte, etc...

Die Vision des Semantic Web

- Was ist das Semantic Web?
 - „The Web can reach its full potential only if it becomes a place where data can be shared and processed by automated tools as well as by people. For the Web to scale, tomorrow's programs must be able to share and process data even when these programs have been designed totally independently.“



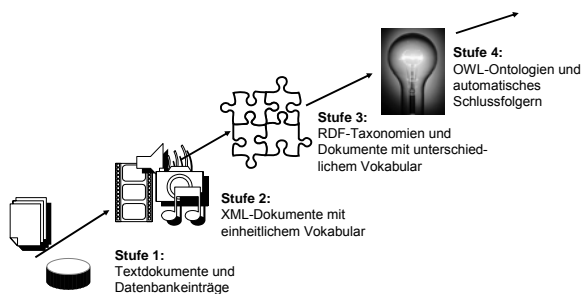
Die Vision des Semantic Web

- Was kann das Semantic Web?

- **Stufe 3: RDF-Taxonomien und Dokumente mit unterschiedlichem Vokabular**
Daten stammen aus **unterschiedlichen Domänen** und werden durch hierarchisch strukturierte **Taxonomien** klassifiziert. Einfache Beziehungen zwischen den einzelnen Kategorien dienen der Kombination von Daten.
 - z.B. TGI ist eine Vorlesung ist eine Lehrveranstaltung
 - z.B. Theoretische Informatik gehört zum Institut für Informatik gehört zur Fakultät Mathematik und Informatik gehört zur Universität Jena

Die Vision des Semantic Web

- Was kann das Semantic Web?



Die Vision des Semantic Web

- Was kann das Semantic Web?

- **Stufe 4: OWL-Ontologien und automatisches Schlussfolgern**
 - **Neue Daten** können aus bestehenden Daten gewonnen werden durch die Anwendung von **Schlussfolgerungsregeln**. **Daten, Eigenschaften von Daten und Beziehungen zwischen Daten** lassen sich semantisch beschreiben.
 - z.B. automatisches Übersetzen eines Dokuments von einer Domäne in eine andere
→ Anwendung „versteht“ Daten

Die Vision des Semantic Web

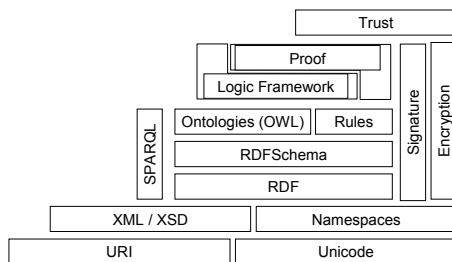
- Was kann das Semantic Web?
 - Kann man den semantischen Daten auch trauen?
 - Authentizität
Ist der mutmaßliche Urheber (Autor) auch der Tatsächliche?
 - Integrität
Wurden die Daten manipuliert?
- ⇒ **Public Key Kryptografie und digitale Signaturen**
- Semantische Daten müssen signiert werden, damit Inferenzsysteme deren Zuverlässigkeit mit in Betracht ziehen können
 - Semantisches Dokument = Sammlung von Aussagen + zugehörige Signaturen

Die Vision des Semantic Web

- Was ist das Semantic Web nicht?
 - Nicht Teil des Semantic Webs sind (jetzt und in naher Zukunft)
 - assoziatives Schlussfolgern
 - räumliches Denken
 - Erkennung von Text, Bildern, Gesten, etc...
 - Komplexe Entscheidungsprozeduren
 - Schlussfolgern aus unvollständigen oder kontextabhängigen Informationen
 - ...
 - „Just as Prolog is not AI but merely a useful tool for it, SW might be just a good tool for AI“

Die Vision des Semantic Web

- Was kann das Semantic Web?



Die Vision des Semantic Web

- Werkzeugkasten für das Semantic Web
 - Sprachstandards, mit denen Semantik der Information im Web ausgedrückt werden kann
 - Werkzeuge, um die Semantikinformation im Web nutzen zu können
 - Beteiligung vieler unterschiedlicher Disziplinen der Informatik:
 - Künstliche Intelligenz
 - Computer Linguistik
 - Kryptografie
 - Datenbanktechnik
 - Theoretische Informatik
 - Rechnerarchitektur
 - Systemtheorie
 - Computernetzwerke



Die Vision des Semantic Web

- Was ist das Semantic Web nicht?
 - „The Semantic Web is not AI“
 - RDF und OWL sind (im Vergleich zur Künstlichen Intelligenz) relativ simple Beschreibungssprachen
 - einfache Art untereinander austauschbare Daten zu formulieren und zu beschreiben
 - eine Möglichkeit, einfache Aussagen zu treffen
 - bieten primitive Möglichkeit der Inferenz innerhalb eines beschränkten Rahmens
 -and that is it!
 - „One goal in SW is to keep things relatively simple and not necessarily seek absolute completeness“

Die Vision des Semantic Web

- Semantic Web Anwendungsbeispiele
 - Suchmaschinen
 - Eingabe in Suchmaschine: **“GOLF von 2000“**
 - Nachfrage zu **“GOLF“**:
PKW oder SPORTART oder SONSTIGES ?
Verfeinerung: **PKW**
 - Nachfrage zu **“2000“**:
Baujahr oder technische Parameter?
Verfeinerung: **Baujahr**
 - Nachfrage zu **“GOLF von 2000“**:
Verkaufsangebote, Dokumentation oder sonstige Artikel ?
Verfeinerung: **Dokumentationen**
 - Nachfrage zu **“Dokumentationen“**: ...



Die Vision des Semantic Web

● Semantic Web Anwendungsbeispiele

- Suchmaschinen (II)
 - Eingabe in Suchmaschine: "GOLF von 2000"
- Sonstige (persönliche) verfügbare Informationen:
 - Benutzer hat in den vergangenen Monaten regelmäßig die Ergebnisse von Profi-Golfturnieren abgerufen
 - Benutzer hat im vergangenen Jahr eine Garnitur Golfschläger (via Online-Shopping) gekauft
 - Benutzer hat die RSS-Feeds zu Golf-Fachpublikationen im Web abonniert
 - Benutzer hat den Bildschirm-Hintergrund „Pebble-Beach“ gewählt
 - Benutzer besitzt die Rechte an der Domain www.golfing-fanatics.tv
 - ...



Die Vision des Semantic Web

● Semantic Web Anwendungsbeispiele

- Stundenplan
 - Student möchte an Seminar „**Semantic Web**“ teilnehmen
 - Semantisch annotierte Web-Seite zum Seminar ermöglicht
 - Eintrag / Abgleich des Termins mit Terminkalender
 - Eintrag der Adresse(n) des/der Seminarleiter(s) in das eigene Adressverzeichnis
 - Wenn der Seminarleiter selbst eine semantisch annotierte Webseite besitzt
 - » Beschaffung relevanter Zusatzinformation
 - Eintrag der Adressen der übrigen Seminarteilnehmer in das eigene Adressverzeichnis
 - Eintrag des Seminarorts (wenn nötig) in den Routenplaner



Die Vision des Semantic Web

● Semantic Web Anwendungsbeispiele

- Urlaubsreisen
 - Eingabe an den persönlichen Agenten:
Urlaub, mind. 10 Tage, mit Familie, am Meer < 2000€
 - Agent überprüft **Terminkalender** von **jedem Familienmitglied** und identifiziert entsprechende Freiräume
 - Agent überprüft **Urlaubsangebote** innerhalb der gefundenen Termine entsprechend den Anforderungen
 - Orte einschränken
 - Familienfreundlichkeit?
 - Gibt es in der Nähe einen Golfplatz?
 - ...
 - Agent erfragt **freie Kapazitäten** und stellt eine Auswahl der ermittelten Angebote entsprechend bekannter Nutzerpräferenzen zusammen



Die Vision des Semantic Web

● Semantic Web Anwendungsbeispiele

- Stundenplan (II)
 - Semantisch annotierte Web-Seite zum Seminar ermöglicht
 - automatische Einschreibung in das Seminar
 - » Prüfung von eventuell zu erfüllenden Vorbedingungen
 - » Prüfung, in wie weit Seminar für Studienplanung (Creditpoints, Studienschwerpunkt, Studien-/Diplomarbeit) sinnvoll
 - Semi-automatische Seminarvorbereitung
 - » Suche nach gleichen/ähnlichen bereits bearbeiteten Seminarthemen im Web (Präsentation+Ausarbeitung)
 - » Überprüfung der Vertrauenswürdigkeit/Qualität der gefundenen Arbeiten...



Die Vision des Semantic Web

● Semantic Web Anwendungsbeispiele

- Urlaubsreisen (II)
 - Eingabe an den Agenten: "Buche Reise xy"
 - Agent führt notwendige **Buchungen** durch
 - Fluggesellschaft / Bahn / ...
 - Hotel
 - Mietwagen (der notwendigen Größe...)
 - Ausflüge vor Ort
 - ...
 - Agent veranlasst **Finanztransaktionen**
 - Agent trägt Termin in die **Kalender** der Familienmitglieder ein
 - Agent erstellt **Routenplanung** für das Navigationsgerät im familieneigenen PKW für die Fahrt zum Flughafen und reserviert einen günstigen Platz im Flughafenparkhaus



Die Vision des Semantic Web

● Das Semantic Web verändert die Welt...

- Ziele des Semantic Web
 - **Organisation des Wissens** in einzelne Domänen entsprechend seiner Bedeutung
 - Automatisierte Werkzeuge zur **Wartung**, zur **Beseitigung von Inkonsistenzen** und zur **Extraktion neuen Wissens**
 - Ersetzen der Schlüsselwortbasierten Suche durch **Inhaltsbasierte Suche** über gezielte Frage-und-Antwort / Visualisierung
 - Nutzerfreundliche und personalisierte Suche
 - **Extraktion und Präsentation von Wissen**
 - Beantwortung von Fragen über verschiedene Dokumente hinweg
 - Exakte Festlegung, wer genau welche Teile bestimmter Informationen sehen darf (**Vertrauen**)

Die Vision des Semantic Web

- Das Semantic Web verändert die Welt...
 - **Praktische Umsetzung ?**
 - Jede Information im Semantic Web muss **eindeutig identifiziert** werden können
→ **URI**
 - Informationen im Semantic Web müssen so formuliert werden, dass Sie von Agenten verstanden werden können. Dies erfordert eine **einheitliche Syntax**
→ **XML und XMLSchema**
 - Einfache **Beziehungen** zwischen Informationen müssen semantisch ausgedrückt werden können
→ **RDF mit RDFSchemata**

Die Vision des Semantic Web

- Das Semantic Web verändert die Welt...
 - **Praktische Umsetzung ?**
 - Unterschiedliche **Semantik muss erklärt werden**, damit scheinbar nicht vereinbare Informationen miteinander vereint werden können → Ontologien/Regeln
→ **OWL / SWRL**
 - Aus vorhandener Information soll **neue Information** geschlussfolgert werden
→ **Inferenzmechanismen**
 - Zugriffsrestriktionen und Vertrauenswürdigkeit müssen gewährleistet sein
→ **XMLEncryption und XMLSignature**

Die Vision des Semantic Web

- **Zusammenfassung**
 - 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 ergänzen sich gegenseitig
 - *Gibt es eine Semantic Web Killer-Applikation?*
