

# Thema: „State of the Art“ des Semantic Web

im Fach Hauptseminar Informationsmanagement, Wintersemester 2004/05

von Mathias Paul Eckert aus Bischofswerda

Matrikelnummer: 2767731

Fach: Wirtschaftsinformatik

## Zusammenfassung:

Die vorliegende Arbeit widmet sich dem aktuellen Entwicklungsstand des Semantic Web. Ausgehend von einer Charakterisierung der Schwächen des traditionellen WWW wird dabei zunächst die Vision des Semantic Web dargestellt. Anschließend werden die im Schichtenmodell des Semantic Web enthaltenen Sprachen und Techniken kurz erläutert, sowie auf ihren Reifegrad hin überprüft. Nach einer kurzen Vorstellung von aktuellen Anwendungen des Semantic Web erfolgt abschließend eine Zusammenfassung der gewonnenen Erkenntnisse zum Entwicklungsstand des Semantic Web und dessen zukünftiger Entwicklung.

## Schlagnworte:

Semantic Web, XML, RDF, Ontologien, Web of Trust, Web Services, Unicode, URI, Namespaces, RDF Schema, RDF Site Summary, Cyc, FOAF.

## Inhalt

1	Einleitung .....	2
2	World Wide Web vs. Semantic Web .....	2
2.1	Das traditionelle World Wide Web .....	2
2.2	Vision des Semantic Web .....	3
3	Techniken des Semantic Web .....	4
3.1	Unicode und Uniform Resource Identifier .....	5
3.2	XML-Familie und Namespaces .....	6
3.3	RDF und RDF Schema .....	7
3.4	Ontologien .....	8
3.5	Vertrauens- und Sicherheitsinfrastruktur .....	11
3.6	Web Services .....	12
4	Anwendungen des Semantic Web .....	14
4.1	RDF Site Summary .....	14
4.2	Die Ontologie Cyc .....	15
4.3	Friend of a Friend .....	16
5	Gegenwart und Zukunft des Semantic Web .....	17
5.1	Aktueller Entwicklungsstand .....	17
5.2	Hemmnisse .....	17
5.3	Zukünftige Entwicklung .....	18
	Literaturverzeichnis .....	19

# **1 Einleitung**

Die vorliegende Arbeit beschäftigt sich mit dem aktuellen Entwicklungsstand des Semantic Web. Ausgangspunkt ist dabei Kapitel 2, welches das traditionelle World Wide Web (WWW) in seiner heutigen Form, samt seinen bekannten Schwachstellen, erläutert, bevor auf die grundsätzliche Philosophie des Semantic Web eingegangen wird. Anschließend wird in Kapitel 3 das Schichtenmodell des Semantic Web vorgestellt. Dabei werden die verschiedenen Techniken der einzelnen Schichten kurz beschrieben. Hauptaugenmerk liegt auf einer Bewertung des aktuellen Reifegrades der jeweiligen Technologie. Kapitel 4 stellt exemplarisch aktuelle Anwendungen des Semantic Web vor, die bereits den Status von wissenschaftlichen Prototypen verlassen haben und der breiten Öffentlichkeit zur Verfügung stehen. Abschließend wird in Kapitel 5 der aktuelle Stand des Semantic Web zusammengefasst. Dabei wird ebenfalls auf Hemmnisse für eine erfolgreiche Ausbreitung eingegangen sowie ein Ausblick auf die kurz- und mittelfristig zu erwartenden Weiterentwicklungen gewagt.

## **2 World Wide Web vs. Semantic Web**

Im folgenden Kapitel soll zunächst das WWW in seiner heutigen Form kurz charakterisiert werden. Besonders Wert wird dabei auf das Aufzeigen von Schwachstellen gelegt. Anschließend wird die Philosophie des Semantic Web als Weiterentwicklung des WWW beschrieben. Insbesondere werden dabei Vorteile gegenüber dem traditionellen WWW, sowie Schritte zur dessen Realisierung erläutert.

### **2.1 Das traditionelle World Wide Web**

Das WWW in seiner gegenwärtigen Form ist ein multimediales hypertextbasierendes Informationssystem im Internet. Es entstand 1989 am Europäischen Kernforschungszentrum in Genf (CERN) ([Schn+00], S. 696). Obwohl viele der von seinen Entwicklern vorgeschlagenen Techniken und Konzepten nicht oder nur teilweise verwirklicht wurden, hat es in kurzer Zeit eine ungeahnte Popularität erlangt und sich rasch zu einer alltäglichen Anwendung für breite Bevölkerungsschichten entwickelt ([Fens+03], S. XIVf.). Aufgrund seiner Fähigkeiten auch andere Anwendungen des Internets abzubilden, erfolgt heutzutage oft sogar eine Gleichsetzung von Internet und WWW. Das Web basiert auf dem Hypertext-System. Dokumente enthalten dabei Querverweise zu beliebigen anderen Dokumenten. So entsteht eine unkomplizierte, weltweite Vernetzung von Informationen. Diese Hypertext-Dokumente können mit Bildern, Sounds, Animationen und sogar Programmen angereichert werden und bieten somit ein multimediales Informationsangebot. Durch seine heterogene und dezentrale Struktur bietet das WWW eine große Vielfalt, beinhaltet jedoch auch gravierende Schwachstellen.

Die Fähigkeiten des traditionellen WWW beschränken sich auf die eines Transport- und Präsentationsmediums für Dokumente, da es Rechnern hier lediglich möglich ist die Syntax von Dokumenten zu lesen und zu verarbeiten. Die Semantik der Dokumente, also das Verstehen oder Interpretieren der Dokumentinhalte, ist durch Rechner nicht interpretierbar ([Hölß03], S.1). Diese Aufgabe bleibt bisher dem Mensch vorbehalten. Die Informationsflut unstrukturierter HTML-Dokumente ist jedoch nur schwer zu bewältigen. Dies führt u.a. zu erheblichen Problemen sowie beachtlichen Zeitaufwänden bei der Suche nach Dokumenten (vgl. [Daco+03]. S. 4ff.).

Aufgrund dieser fehlenden maschinellen Verarbeitbarkeit können Informationen aus verschiedenen Quellen nicht automatisch integriert werden. Aus Quellen wie etwa Datenbanken, Programmen oder Sensoren ist es nicht möglich Informationen zu generieren ([Köni02], S. 481). So ist in der WWW-Gemeinde Ernüchterung eingetreten, denn das WWW ist, in seiner heutigen Form, bereits jetzt an seine Grenzen gestoßen. Es bestehen kaum noch Möglichkeiten es essentiell zu verbessern. Im Gegenteil, angesichts der heterogenen und dezentralen Organisation, sowie eines exponentiellen Wachstums, wird es in Zukunft schwierig sein, das gegenwärtige Niveau überhaupt zu halten (vgl. [Fens+03], S.1ff.).

## 2.2 Vision des Semantic Web

Das Semantic Web ist Teil der großen Vision Tim Berners-Lee's von einem „einzigem Bedeutungsnetz von allem und jedem“ ([Frau04], S. 52). Die Philosophie des Semantic Web basiert auf der Idee, Daten so zu definieren und zu verbinden, dass sie für eine effektivere Informationssuche verwendet, automatisch integriert, sowie in verschiedenen Applikationen wiederverwendet werden können. Das Web stellt dann eine universell zugängliche Plattform zur Verfügung, auf der Daten bereitgestellt und vor allem verarbeitet werden können, sowohl von automatisierten Werkzeugen, als auch von Menschen (vgl. [Hend+02], S 676; [W3C04a]).

Um das Semantic Web zu realisieren muss das traditionelle Web erweitert werden. Der wichtigste Aspekt ist dabei die semantische Strukturierung. Die über das WWW übertragenen Daten müssen dafür mit wohldefinierten Bedeutungsinformationen (Semantik) angereichert werden. Solche Daten können dann einfacher maschinell gesucht, bzw. neu kombiniert werden (vgl. [Daco+03], S. 1ff.). Somit kann das Semantic Web, neben der Automatisierung von bisher menschlichen Aufgabenträger vorbehaltenen Tätigkeiten, ein weiteres Hauptproblem des traditionellen Web, die mangelhafte Computer-Mensch-Interaktion, beheben.

Für die Entwicklung des Semantic Web sind grundsätzlich die folgenden Aufgaben zu bewältigen (vgl. [Fens+03], S. 2f.):

- Entwicklung von Sprachen für maschinell lesbare Metadaten (Daten über Daten) über Dokumente, sowie die Entwicklung von Terminologien, die diese Sprachen nutzen und sie im Web verfügbar machen,
- Entwicklung von Werkzeugen und neuen Architekturen die solche Sprachen und Terminologien nutzen um Informationsressourcen zu finden, zugänglich zu machen, zu präsentieren und aktuell zu halten,
- Anreicherung der vorhandenen Daten und Dokumente im WWW mit Metadaten, sowie
- Realisierung von Applikationen, die den menschlichen Nutzern des Semantic Web ein neues Level von Nutzen zur Verfügung stellt.

Triebkraft für die Weiterentwicklung des Semantic Web ist, wie bereits beim traditionellen WWW, das World Wide Web Consortium (W3C). Das W3C ist ein internationales Industriekonsortium bzw. eine Interessenvertretung für Webtechnologien ([Böhm+03], S. 73). Hauptziel dieser Organisation ist die Schaffung von ausgereiften internationalen Standards die als Grundlage für die Entwicklung des Semantic Web zwingend nötig sind ([Fens+03], S. XVf.).

### 3 Techniken des Semantic Web

Das Semantic Web basiert auf einer hierarchischen Struktur von Sprachen, Mechanismen und Technologien. Diese werden im Schichtenmodell des Semantic Web dargestellt (vgl. Abb. 3.1). In der Literatur existieren zahlreiche Versionen dieses Modells. In der vorliegenden Arbeit wird zur Komplexitätsreduzierung ein, im Vergleich zu Tim Berners-Lee's Originalversion, vereinfachtes Modell (vgl. Abb. 3.1) verwendet. Die ursprünglichen Schichten *Ontology Vocabulary* und *Logic* wurden dabei genauso zu einer Schicht (die Schicht Ontologien) zusammengefasst, wie die Schichten *Proof*, *Trust* und *Digital Signature* zur Schicht Vertrauens- und Sicherheitsinfrastruktur. Zusätzlich wurden Web Services als flankierendes Element in das Modell aufgenommen.

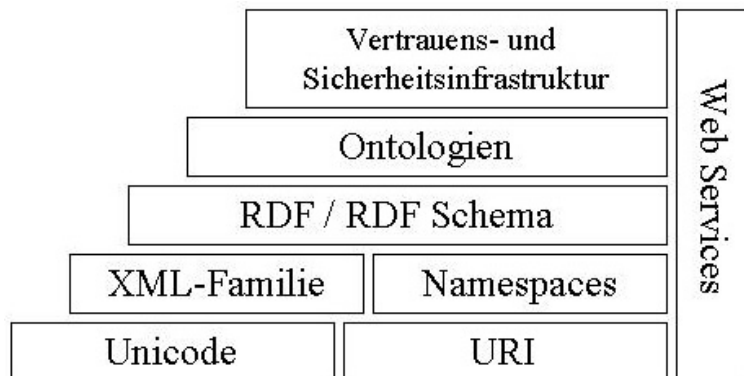


Abb. 3.1: Schichtenmodell des Semantic Web (in Anlehnung an: [Hölß03], S. 2)

Tabelle 3.1 gibt einen Überblick über die grundlegenden Funktionen der Schichten des in der vorliegenden Arbeit verwendeten Modells des Semantic Web.

Tab. 3.1: Schichten des Semantic Web und ihre Funktion

Schicht	Funktion
Unicode / URI	Standards für einheitliche Bezeichnung
XML-Familie / Namespaces	Schaffung von Datenstrukturen
RDF / RDF Schema	Klassifikation sowie semantische Anreicherung von Dokumenten
Ontologien	Daten interpretieren und zueinander in Beziehung setzen
Vertrauens- und Sicherheitsinfrastruktur	Garantie der Wahrheit von Daten und Informationen
Web Services	Kommunikationsmechanismus zwischen Anwendungen

Im folgenden Kapitel werden die einzelnen Techniken im Schichtenmodell des Semantic Web in ihrer hierarchischen Reihenfolge vorgestellt. Neben einer bewusst kurz gehaltenen grundsätzlichen Erläuterung der einzelnen Aspekte wird besonders deren aktueller Stand bzw. Reifegrad dargestellt.

### 3.1 Unicode und Uniform Resource Identifier

Unicode ist ein internationaler Standard, der jedem Zeichen eine einheitliche Nummer zuordnet, und zwar plattform-, programm- und systemabhängig. Vor dessen Einführung existierte eine Vielzahl solcher Kodierungssysteme die untereinander nicht verträglich waren. Dies führte zu Problemen im weltweiten Datenaustausch. Mit Unicode dagegen, kann Text weltweit ohne Informationsverlust ausgetauscht werden. Der Unicode-Standard ist eine Implementierung der internationalen Norm ISO/IEC 10646. Verantwortlich für den Standard ist das Unicode-Konsortium, eine gemeinnützige Organisation mit dem Ziel den Unicode-Standard weiterzuentwickeln und zu verbreiten. Die letzte Hauptversion, Unicode 4.0, stammt aus dem Jahr 2003. Der Standard ist weltweit anerkannt und etabliert. Er ist stabil und ausgereift und unterliegt somit nur noch punktuellen Veränderungen bzw. Verbesserungen (vgl. [Unic04]).

Uniform Resource Identifier (URI) bezeichnet einen Gesamtbegriff für eine Menge von Einzelkonzepten. Diese haben das Ziel sämtliche Ressourcen innerhalb eines unsortierten und dezentral organisierten Informationsraumes, wie etwa dem WWW, eindeutig und unverwechselbar zu identifizieren. Dies geschieht mit Hilfe einer kompakten Zeichenkette mit der sowohl abstrakte also auch physische Ressourcen adressiert werden können. Die Ressourcen selbst werden durch URIs nicht beeinflusst. Im Jahr 1998 hat das World Wide Web Consortium (W3C) URIs durch den Request for Comments (RFC) 296 standardisiert (vgl. [Daco+03], S. 44; [W3C04b]).

Das wichtigste Einzelkonzept im Rahmen der URI ist das der Uniform Resource Locator (URL). Diese wurden bereits 1993 im RFC 1738 standardisiert und sind somit von Beginn an eine der tragenden Säulen des traditionellen WWW gewesen. Sie stellen Syntax und Semantik zur formalisierten Darstellung von Ablageorten und zum Zugriff auf Ressourcen über das Internet zur Verfügung. Alle über das Internet auf Ressourcen zugreifende Programme nutzen das Konzept der URL (vgl. [Kara04]).

Ein weiteres Einzelkonzept der URI ist das Konzept der Uniform Resource Names (URN). URN beschränken sich auf die eindeutige und unverwechselbare Bezeichnung von Ressourcen. Informationen zum Standort werden dagegen nicht geliefert. Eine verbreitete Anwendung von URN ist der ISBN-Code (International Standard Book Number) für Bücher. Die Uniform Resource Characteristics (URC) gehören ebenfalls zur Gruppe der URI und haben für das Konzept des Semantic Web eine essentielle Bedeutung. Sie bieten die Möglichkeiten beliebige Informationen bzw. Eigenschaften von Ressourcen zu beschreiben. Es können so Metadaten einheitlich dargestellt werden. Konkrete Anwendungen auf diesem Gebiet beschränken sich derzeit auf Entwicklungen und Tests, da sich die Umsetzung für einen dynamischen und komplexen Informationsraum wie das WWW schwierig gestaltet ([Kara04]). Der bekannteste Ansatz stammt von der Dublin Core Metadata Initiative (DCMI), die eine einheitliche Systematik zur Beschreibung von Internet-Ressourcen entwickelt hat (vgl. [Dubl04]).

Probleme im Zusammenhang mit URIs entstehen durch die dezentrale Organisation des Internets. Jedem ist es möglich beliebige Ressourcen mit beliebigen URIs zu adressieren. Es ist also durchaus möglich, dass eine Ressource durch mehrere URIs gleichzeitig gekennzeichnet ist.

Unicode und URI bilden zusammen die unterste Schicht im Modell des Semantic Web und fungieren somit als Fundament für die darüber liegenden Schichten. Beide Standards sind zur Vereinheitlichung eines globalen Informationsraumes wie dem Internet unerlässlich. Die genannten Standards sind bereits tragenden Säulen des traditionellen WWW und haben durch ihre jahrelange Nutzung und Verbesserung inzwischen einen stabilen und ausgereiften Stand erreicht. Sie können somit ohne Abstriche als Basis des Semantic Web genutzt werden.

### 3.2 XML-Familie und Namespaces

Die Extensible Markup Language (XML), sowie weitere ergänzende Standards bilden zusammen die XML-Technologiefamilie. Diese stellt die zweite Schicht im Modell des Semantic Web dar und ist zugleich dessen elementarste Komponente. Die Wichtigkeit von XML wird auch durch die Bezeichnung des Semantic Web als XML-based Web verdeutlicht. Grundsätzlich nimmt XML im Semantic Web die Aufgaben einer maschinenlesbaren Transportsprache wahr.

XML strukturiert Daten und Dokumente in einer Hierarchie- oder Baumstruktur. Es bietet die Syntax um semantische Markupsprachen, sogenannte XML-Applikationen, zu definieren. Eine solche Sprache ist bspw. die Hypertext Markup Language (HTML). Grundsätzlich werden in der XML die Inhalte von den Strukturierungsinformationen (Markup) getrennt. Das Markup wird dabei durch sogenannte Tags gekennzeichnet, die jeweils konkrete Inhalte umschließen. So wird eine einfache und standardisierte Syntax zur Verfügung gestellt, um Daten mit Metadaten zu versehen.

Das XML-Konzept hat sich mittlerweile zu einem allgemein angewendeten Standard entwickelt (vgl. [Daco+03], S. 27f.). Dieser Erfolg hat mehrere Gründe. Einerseits verwendet XML eine vom Menschen lesbare Syntax in Klartextform und ist somit plattform- und programmunabhängig. Die in XML gespeicherten Daten sind somit genauso unkompliziert zu durchsuchen, wie etwa HTML-Dateien. In Hinblick auf die Vielfalt von Software-Agenten und Systemen im Semantic Web ist dies eine unabdingbare Notwendigkeit. Andererseits ist die XML-Technologie bereits ausgereift. Die theoretischen Wurzeln liegen in der bereits 1969 geschaffenen Standardized Generalized Markup Language (SGML). Diese wurde im Laufe der Jahre kontinuierlich perfektioniert und getestet ([Daco+03], S. 31). XML selbst wurde 1998 von der W3C standardisiert.

Es existieren eine Vielzahl von Standards die XML ergänzen und für die Realisierung des Semantic Web ebenfalls von Bedeutung sind. Die Wichtigsten werden im Folgenden kurz erläutert.

XML Schema definiert Syntax und Semantik zur Formulierung von Grammatiken beliebiger XML-Vokabulare und ist dabei selbst ein XML-Vokabular. Es können so XML-Dokumente erzeugt werden, die einem definierten Vokabular sowie einer definierten hierarchischen Struktur entsprechen. Dabei können auch verschiedene Datentypen innerhalb der XML-Dokumente festgelegt werden (vgl. [Fens+03], S. 103ff.). XML Schema wurde im Jahr 2002 vom W3C standardisiert. Die Validierung von XML-Dokumenten, also die Überprüfung ob ein Dokument einem Schema entspricht, ist angesichts der Komplexität von XML-Schema immer noch mit Problemen behaftet. Mit zukünftigen Versionen sollen diese jedoch behoben werden. Neben XML Schema existieren weitere weniger verbreitete Schema-Sprachen. Zu diesen gehören die Document Type Definition (DTD), RELAX NG sowie Schematron ([Daco+03], S. 37ff.).

Das Konzept der Namespaces liefert einen einfachen Mechanismus um einen global einmaligen Namen für die Elemente und Attribute einer Markupsprache zu schaffen. Dies ist aus zweierlei Gründen wichtig. Einerseits können so Konflikte durch gleiche Elementnamen in verschiedenen Sprachen gelöst werden und andererseits mehrere Markupsprachen kombiniert werden. Namespaces verwenden URIs um Markupsprachen eindeutig zu kennzeichnen. Seit 1999 sind Namespaces vom W3C standardisiert (vgl. [Daco+03], S. 42ff.). Weitere ebenfalls relevante Standards der XML-Familie sind (vgl. [Daco+03], S. 119ff.):

- Document Object Model (DOM) – erlaubt den programmtechnischen Zugriff, sowie die Manipulation von von XML-Dokumenten,
- XPath – Sprache zur Adressierung von bestimmten Teilen von XML-Dokumenten,

- Style-Sheet-Familie (XSL, XSLT, XSLFO) – Transformation von XML Dokumenten und Darstellung in verschiedenen Medienformaten,
- XQuery – Abfrage von Daten in XML-Format,
- XLink – komplexe Sprache für Links zwischen Dokumenten, der Zuordnung von Metadaten sowie der Verbindung zu externen Ressourcen,
- XPointer – Pointer-Sprache für XML-Ressourcen,
- XML Base – Spezifizierung einer Basis-URI zur Verwendung von relativen URIs,
- Extensible Hypertext Markup Language (XHTML) – strukturiertere Form von HTML,
- XForms – neue Funktionalitäten, sowie mehr Flexibilität für Formulare, sowie
- Scalable Vector Graphics (SVG) – Sprache zur Beschreibung zweidimensionaler Grafiken in XML.

Der XML-Standard an sich, ist durch seine lange Entwicklungsgeschichte (beginnend mit der SGML 1969) als weltweit anerkannt und ausgereift zu bezeichnen. Er bietet so ein stabile und mächtige Basis für die weiteren Schichten des Semantic Web. Andere Teile der XML-Familie, wie etwa das Konzept der Namensräume oder XQuery sind dagegen recht jung und befinden sich noch in der Entwicklung. Von ausgereiften Versionen kann somit noch nicht gesprochen werden. Die Verbreitung dieser Konzepte ist deshalb noch beschränkt. Deutliche Fortschritte sind hier jedoch in den nächsten Jahren zu erwarten.

### 3.3 RDF und RDF Schema

Die dritte Schicht im Modell des Semantic Web bildet das Resource Description Framework (RDF) sowie die zugehörige Schema-Beschreibungssprache RDF Schema. Grundsätzlich ist RDF eine Sprache zur Präsentation von Informationen über Ressourcen im WWW. Es ist somit Grundlage für die Beschreibung und den Austausch von Metadaten. Verschiedene Anwendungen können diese maschinenlesbaren Daten nutzen um zusammenzuarbeiten ([Daco+03], S. 85).

Das syntaxunabhängige, grundlegende RDF-Datenmodell besteht aus drei Teilen und wird deshalb als Tripel bezeichnet (vgl. Abb. 3.2). Ressourcen kennzeichnen die Objekte die durch RDF-Aussagen beschrieben werden sollen. Dabei kann es sich sowohl um eine Web-Ressource, als auch um eine Ressource aus der realen Welt handeln. Eine Ressource wird grundsätzlich durch eine URI identifiziert. Eigenschaften dagegen sind ein spezieller Aspekt, eine Charakteristik oder eine Beziehung einer Ressource. Der Wert dieser Eigenschaft kann aus einem Literal oder einer anderen Ressource bestehen. Die Werte, welche eine Eigenschaft annehmen kann sind dabei genauso definiert wie die Ressourcen die sie beschreibt und die Verbindungen zu anderen Eigenschaften die sie eingeht. Eine Ressource und deren Eigenschaft, sowie einem bestimmten Wert für diese Eigenschaft bilden zusammen eine Aussage (vgl. [Mano+04]). Des Weiteren existieren in RDF verschiedene Konstrukte zur Gruppierung von Ressourcen, sowie Möglichkeiten zur Schaffung von Aussagen höherer Ordnung, also von Aussagen über Aussagen (Reification) ([Daco+03], S. 92ff.). Für die Kodierung von RDF hat das W3C XML vorgeschlagen und dafür 1999 eine entsprechende Syntax standardisiert (RDF/XML) ([Hölß03], S. 3).

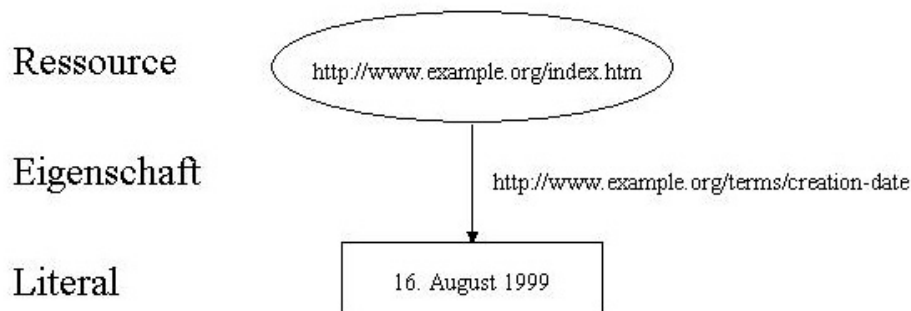


Abb. 3.2: Beispiel eines RDF-Tripels (in Anlehnung an [Mano<sup>+</sup>04])

Mit RDF selbst ist es nicht möglich Beziehungen zwischen Ressourcen und Eigenschaften festzulegen. Zu diesem Zweck wurde vom W3C im Jahre 2000 RDF Schema standardisiert. RDF Schema definiert einen einfachen Satz von RDF-Ressourcen und Eigenschaften. Mit diesem ist es möglich RDF-Vokabulare zu kreieren. Ein solches Vokabular beschreibt die Bedeutung von Eigenschaften, die Art von Ressourcen welche eben diese Eigenschaft besitzen und Beziehungen zwischen Eigenschaften. Das Datenmodell von RDF Schema orientiert sich am objektorientierten Paradigma. Daher ist es möglich Gruppen von Ressourcen mit ähnlichen Eigenschaften zu Klassen zusammenzufassen. Diese Klassen können in hierarchische Beziehungen zueinander gesetzt werden (vgl. [Daco+03], S. 104ff.; [Fens+03], S. 200).

Das grundsätzliche RDF-Konzept ist als ausgereift zu bezeichnen. Deutlich wird dies vor allem durch die nur leichten Nachbesserungen an dem Standard seit seiner Erstveröffentlichung im Jahre 1999. Trotzdem kann man von einer umfangreichen Verbreitung und Anwendung nicht sprechen, im Gegenteil, RDF wird von der breiten Öffentlichkeit nicht wahrgenommen. Die Gründe hierfür sind vielschichtig. Einerseits ist das mangelhafte Zusammenspiel von RDF mit XML-Dokumenten zu kritisieren. Bis in das Jahr 2002 war es wegen der offenen Grammatik von RDF nicht möglich, in XML- oder XHTML-Dokumente eingebettetes RDF zu validieren. Dieses Problem wurde vom W3C inzwischen jedoch behoben. Die im Vergleich zu XML höhere Komplexität des RDF-Konzepts behindert ebenfalls eine erfolgreiche Ausbreitung. In diesem Zusammenhang sind vor allem die Verwendung von Metaphern aus den unterschiedlichsten Bereichen (objektorientiertes Paradigma, Linguistik, relationale Daten), die Wahlmöglichkeiten zwischen Darstellungen durch Attribute oder Eigenschaften, die Schwierigkeiten bei der händischen Erstellung von RDF/XML-Syntax und die Idee der Reification zu nennen. Letztlich konnten auch die ersten verfügbaren Anwendungen, wie etwa Dublin Core (DC) oder RDF Site Summary (RSS) (vgl. 4.1), die Vorteile von RDF nicht deutlich genug hervorheben, und so die Anwender nicht vom Nutzen des Konzepts überzeugen ([Daco+03], S. 96ff.).

Von einer zukünftigen Zunahme der Verbreitung und Bedeutung von RDF ist jedoch auszugehen. Die steigenden Anzahl von Publikationen wird dabei genauso zur Diffusion in breitere Anwenderschichten beitragen wie die verbesserte Unterstützung durch Software-Werkzeuge, die verbesserte XML-Schema-Integration oder die Fortschritte im Bereich von Ontologien ([Daco+03], S. 102f.).

### 3.4 Ontologien

Ontologien bilden im Schichtenmodell des Semantic Web die Schicht oberhalb von RDF und RDF Schema. Sie geben in abstrakter Weise das akzeptierte Wissen unserer Welt wieder. Genauer sind Ontologien Konstrukte um Wissen über bestimmte Gebiete zu repräsentieren,

auszutauschen und wiederzuverwenden (vgl. [Hölß03], S. 12). Dies soll dabei sowohl durch Menschen, also auch durch Maschinen möglich sein. Ausgedrückt wird eine Ontologie durch eine logische Theorie, die aus einem definierten Vokabular sowie einer Menge von logischen Aussagen besteht. Methodisch werden dabei verschiedene Techniken der objektorientierten Modellierung verwendet. Außerdem ist es nötig, dass sich eine Gruppe von Anwendern auf die jeweiligen Begriffe sowie deren Zusammenhänge innerhalb des relevanten Gebietes einigt ([Mädc+01], S. 393). Für die Anwendung im Semantic Web stellen Ontologien semantische Modelle dar, die Daten interpretieren und zueinander in Beziehung setzen (vgl. [Fens+03], S. 173ff.).

Es existieren verschiedene Modelle von Ontologien. Diese reichen von Taxonomien, welche nur schwache semantische Zusammenhänge darstellen, bis hin zur Darstellung von starken semantischen Zusammenhängen durch logische Theorien. Abbildung 3.3 zeigt die einzelnen Bestandteile dieses sogenannten ontologischen Spektrums.

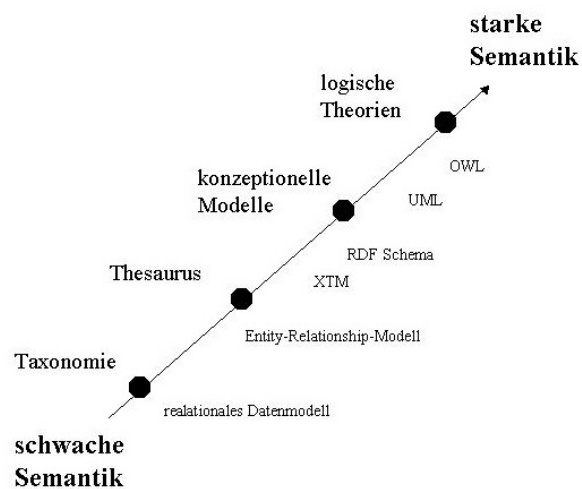


Abb. 3.3: ontologisches Spektrum (in Anlehnung an [Daco<sup>+</sup>03], S. 157)

Taxonomien bilden den unteren Endpunkt des ontologischen Spektrums. In Taxonomien werden bestimmte Objekte in einer Baumstruktur klassifiziert. Die Semantik von Beziehungen zwischen Vater- und Kindelementen sind hier relativ unspezifiziert. Beispiele für Taxonomien sind das relationale Datenmodell oder XML ([Daco+03], S. 145ff.).

Mehr semantische Zusammenhänge können durch Thesauri dargestellt werden. Dabei werden Homonyme, Synonyme, sowie hierarchische Beziehungen von Begriffen abgebildet. Beispiele für Thesauri sind XML Schema oder das Entity-Relationship-Modell (ERM) ([Daco+03], S. 159ff.).

Am oberen Ende des ontologischen Spektrums liegen konzeptionelle Modelle und logische Theorien. Diese können starke semantische Zusammenhänge darstellen. In einem konzeptionellen Modell werden die Inhalte sowie Beziehungen innerhalb eines bestimmten Wissensgebietes dargestellt, ebenso wie deren Attribute samt zugehörigen Werten. Außerdem existieren sogenannte Regeln zwischen den genannten Einzelteilen eines konzeptuellen Modells.

Logische Theorien sind durch Software direkt semantisch interpretierbar. Sie bauen auf Axiomen und Inferenzregeln auf, die zusammen verwendet werden um bestimmte Theoreme zu bestätigen. In die Gruppe der konzeptionellen Modelle und der logischen Theorien sind bspw. die Sprachen Unified Modelling Language (UML), Web Ontology Language (OWL) oder RDF Schema einzuordnen ([Daco+03], S. 166).

Ebenfalls in diese Gruppe gehören Topic Maps. Aufgrund ihrer Bedeutung für die Philosophie des Semantic Web werden diese im Folgenden näher beschrieben. Mit Hilfe von Topic Maps können Informations-Indizes erstellt werden. Es wird somit ein Metaschicht gebildet und diese mit Hilfe von URIs mit den eigentlichen Informationsressourcen verbunden. Dies erleichtert das Finden von relevanten Informationen, da primär in der Metaschicht, welche die Bedeutungen der Informationen enthält, navigiert wird, anstatt in den eigentlichen Informationsressourcen ([Widh+02], S. 6).

Topic Maps haben ihre Ursprünge im Bereich der SGML und wurden im Jahr 2000 von der International Organization for Standardization (ISO) standardisiert (ISO/IEC 13250). Mit XML Topic Maps (XTM) existiert seit 2001 eine Syntax für Topic Maps in XML, die sich seitdem als die am weitesten verbreitete Syntax etabliert hat (vgl. [Hölß03], S. 11). Die folgenden Ausführungen beziehen sich auf diese XTM-Spezifikation.

Herzstück des Konzeptes sind sogenannte Topics. Topics sind Repräsentationen von Subjekten die von Interesse sind. Beziehungen zwischen Topics werden durch Associations modelliert, wobei jeder dieser Associations ein bestimmter Typ zugeordnet ist. In diesem Zusammenhang ist außerdem zu erwähnen, dass den Topics die durch Associations verbunden sind, jeweils eine bestimmte Rolle innerhalb der Association zugeordnet wird. Außerdem existieren Occurrences, die Informationen darstellen, welche für eine bestimmte Topic relevant sind. Occurrences sind ebenfalls typisiert und stellen dem Nutzer somit schon beim navigieren innerhalb der Metaebene Informationen über die eigentlichen Informationsressourcen zur Verfügung (vgl. [Böhm+03], S. 88ff.; [Mück+01], S. 298).

Das Konzept der Topic Maps scheint dem des RDF zunächst sehr ähnlich. Grundsätzlich verfolgen beide Ansätze jedoch verschiedene Ziele. Topic Maps stellen vorrangig inhaltsbasierte Indizes über Dokumente dar und ermöglichen so eine verbesserte Mensch-Maschine-Kommunikation. Während Topic Maps eher unter dem Stichwort Personalisierung zu kategorisieren sind, wird RDF im Bereich der Automatisierung eingeordnet, da hier Metadaten standardisiert, beschrieben und verbunden werden um eine bessere maschinelle Auswertung dieser Daten zu gewährleisten. Beide Paradigmen stehen also nicht in direkter Konkurrenz, sondern ergänzen sich ([Daco+03], S. 178f.).

Ontologien gelten als die „Killerapplikationen“ des Semantic Web. Umfangreiche und schlüssige Ontologien sowie anwendungsfreundliche Werkzeuge zu deren Nutzung werden zum Durchbruch des Semantic Web führen. Trotz erheblicher Fortschritte in den letzten Jahren kann davon heute jedoch noch nicht gesprochen werden. Derzeit existiert eine Vielzahl von Ontologien zu spezifischen Wissensgebieten. Es gibt ebenfalls Versuche allgemeine Ontologien zu erstellen. Die wichtigsten Vertreter auf diesem Gebiet sind WordNet und CYC (vgl. 4.2). Trotz dieser Versuche wird es, aufgrund von Komplexität und Dynamik, jedoch nie möglich sein das gesamte Wissen der Menschheit in einer Ontologie darzustellen. Für die effektive Nutzung von Ontologien bezüglich der Idee des Semantic Web wäre es wünschenswert, dass sich einzelne Ontologien zu bestimmten Wissensgebieten durchsetzen und somit als Quasi-Standards fungieren. Seit seiner Standardisierung durch die W3C hat sich die OWL als Ontologiesprache für das Semantic Web etabliert ([Haus+04], S. 223ff.). Andere verbreitete Sprachen werden schrittweise zurückgedrängt. Für die weitere Entwicklung und Verbreitung von Ontologien ist dies als positiv zu bewerten. Es existieren ebenfalls erste ausgereifte Anwendungen zur Erstellung und Verwaltung von Ontologien. Zu nennen sind in diesem Zusammenhang Projekte wie Protege oder OntoEdit. Werkzeuge zur effektiven Nutzung von Ontologien im Web befinden sich dagegen noch im Stadium der wissenschaftlichen Erprobung. Ein Beispiel hierfür ist der semantische Webbrowser Magpie (vgl. [Dzbo+03], S. 690ff.).

Konnten die bisher beschriebenen Schichten des Semantic Web als ausgereift und nahezu uneingeschränkt nutzbar bezeichnet werden, so trifft dies auf die Ontologie-Schicht im

Modell des Semantic Web nicht zu. Es existieren im Bereich der Ontologien immer noch grundlegende konzeptionelle Unzulänglichkeiten wie etwa das Problem der Vereinigung von Ontologien oder die eingearbeitete Subjektivität der jeweiligen Ersteller. Weitere Hemmnisse stellen Komplexität der Konzeptes sowie die noch mangelhafte Unterstützung durch Werkzeuge bzw. Anwendungsprogramme dar.

### 3.5 Vertrauens- und Sicherheitsinfrastruktur

Die oberste Schicht im Modell des Semantic Web bildet die Vertrauens- und Sicherheitsinfrastruktur. Durch diese wird das Semantic Web gleichzeitig zu einem „Web of Trust“.

Schon im traditionellen WWW besteht keine Garantie für die Wahrheit von Informations-Ressourcen. Das Semantic Web an sich ändert dies nicht. Im Gegenteil, das Problem wird weiter verschärft, da die Wahrheit von Informationen hier auch von maschinellen Aufgabenträgern bewertet werden muss. Maschinen entscheiden dabei über die Glaubwürdigkeit von Informationen indem der Kontext dieser Ressource analysiert wird. Zu diesen Kontextdaten gehören bspw. Autor, Zeitpunkt oder Beglaubigungen.

Die Vertrauensentscheidung im Semantic Web läuft im Modell konzeptionell auf zwei Ebenen ab (vgl. Abb. 3.4). Zuerst wird versucht zu klären, ob die Informationen wirklich vom angegebenen Autor stammen, bevor anschließend die Vertrauenswürdigkeit des Autors überprüft wird. Für beide Ebenen existieren verschiedene Ansätze die im Folgenden kurz skizziert und bewertet werden.

Der erste Ansatz auf der Ebene der Identitätsprüfung des Autors bezieht sich auf die Integration von XML-Signaturen (vgl. [Daco+03], S. 79; [Haus+04], S. 137ff), unter Verwendung bereits existierender Public-Key-Infrastrukturen. Dieser Ansatz wurde von der W3C im Jahr 2002 standardisiert. Problematisch ist dabei die noch geringe Verbreitung von Schlüsseln und Zertifikaten.

Ein alternativer Ansatz verwendet die Fundorte von RDF-Modellen im Web um die Identität des Autors zu überprüfen. Dieser Ansatz ist zwar aufwandsarm zu realisieren, bietet jedoch nur einen geringen Grad an Sicherheit. Außerdem ist die genaue Zuordnung der RDF-Modelle nicht immer möglich ([Bize03], S. 11).

Ebene	Ansatz
Stammen die Informationen wirklich vom Autor ?	XML-DSIG Signaturen + vorhandene Public Key Infrastrukturen verwenden
	Fundorte der RDF-Modelle im Web
Ist der Autor vertrauenswürdig ?	Web of Trust mit expliziten Vertrauensaussagen
	Netzwerkanalyse mit expliziten Vertrauensaussagen
	Netzwerkanalyse mit impliziten Vertrauensaussagen
	Pragmatik

Abb. 3.4: Ebenen und Ansätze der Vertrauensentscheidung

Auf der Ebene der Prüfung der Vertrauenswürdigkeit des Autors werden insgesamt vier Ansätze unterschieden (vgl. Abb. 3.4). Ansatz 1 wird als „Web of Trust mit expliziten Vertrauensaussagen“ bezeichnet und von der W3C gestützt. Dabei bestimmt der Nutzer des Semantic Web selbst die Anbieter von Informationen, denen er vertraut. Nur Informationen dieser Anbieter, bzw. von Anbietern denen diese Anbieter vertrauen, werden berücksichtigt. Zur Umsetzung wurde das Friend-of-a-Friend-Konzept (FOAF-Konzept) (vgl. 4.3) vorgeschlagen, welches RDF Schema um Vertrauensaussagen erweitert. Der umfassenden Realisierung dieses Konzeptes stehen jedoch einige Hindernisse im Wege. Einerseits ist eine enge Community von Nutzern erforderlich, mit Mitgliedern welche sich gegenseitig kennen. Andererseits muss jeder Nutzer für jede Anwendungsdomäne selber Vertrauensaussagen treffen. Der Ansatz ist somit mit einem hohen Aufwand verbunden und lässt sich in einem offenen System, wie dem Semantic Web, nur schwer realisieren ([Bize03], S. 12).

Der zweite Ansatz auf dieser Ebene wird als „Netzwerkanalyse mit expliziten Vertrauensaussagen“ bezeichnet. Dabei vertraut ein neuer Nutzer einem bestehenden Web of Trust, anstatt selbst Vertrauensaussagen zu treffen ([Ratn+02], S 1ff.). Als mögliche Webs of Trust werden z.B. Reputationsdienste auf elektronischen Marktplätzen genannt. Vorteile entstehen bei diesem Ansatz durch die Nichterforderlichkeit einer engen Community. Außerdem muss nicht jeder Nutzer Vertrauensaussagen abgeben. Erforderlich ist aber ein Vertrauen in die Aussagen anderer Nutzer. Nachteilig wirkt sich die Anfälligkeit gegenüber Täuschungsversuchen, sowie die mangelnde Motivation der Nutzer, eigene Vertrauensaussagen zu treffen, aus ([Bize03], S. 13).

Die „Netzwerkanalyse mit impliziten Vertrauensaussagen“ bildet den dritten relevanten Ansatz. Dabei wird grundsätzlich den Informationen vertraut, welche von vertrauenswürdigen Autoren stammen. Grundlage hierfür ist eine umfangreiche Datenbasis mit Informationen zu sämtlichen Personen, die als Autoren im Semantic Web auftreten. Die sich aus der Erstellung einer solchen Datenbasis ergebenden Schwierigkeiten sind gleichzeitig der große Nachteil dieses Ansatzes. Positiv ist dagegen der geringe Aufwand für den Nutzer, da keine expliziten Vertrauensaussagen zu treffen sind ([Bize03], S. 14).

Der letzte Ansatz zur Prüfung der Vertrauenswürdigkeit von Autoren, ist der Ansatz der „Pragmatik“. Dabei werden Informationen auf Widersprüche im Vergleich zu einer bestimmten Ontologie geprüft. Im Detail bezieht sich dieser Vergleich bspw. auf Widersprüche zwischen Eigenschaften von Objekten oder unterschiedliche URIs von Objekten bei gleichen Eigenschaften ([Bize03], S. 15). Dieser hohe Aufwand kann jedoch nur eine sehr grobe inhaltliche Prüfung gewährleisten. Eine Anwendung in einem offenen Netz wie dem Semantic Web ist nach Meinung des Autors nicht möglich.

Die Vertrauens- und Sicherheitsinfrastruktur bildet die höchste Schicht im Modell des Semantic Web und ist dabei gleichzeitig die am wenigsten entwickelte ([Daco+03], S. 233; [Hölß03], S. 16). Derzeit existieren kaum umfassenden Konzepte und Standards. Die Existenz geeigneter Applikationen beschränkt sich auf wissenschaftliche Prototypen. Bedeutende Verbesserungen in diesem Gebiet sind für die nähere Zukunft nicht zu erwarten.

### **3.6 Web Services**

Web Services nehmen im Schichtenmodell des Semantic Web eine Sonderstellung ein. Sie bilden grundsätzlich ein eigenständiges, hoch dynamisches Umfeld, mit eigener Vision, Standardisierung und Anwendung. Für die Verwirklichung der Ideen des Semantic Web sind sie jedoch ein wichtiges flankierendes Element. Sie stellen, auf Grundlage von XML, einen definierten Kommunikationsmechanismus zwischen verteilten, lose gekoppelten Anwendungen zur Verfügung. Ein Web Service kann dabei vorhandene Geschäfts- und Systemfunktionalitäten oder neu entwickelte, auf Komponenten basierende, Anwendungen,

beschreiben und lässt sich im Internet, Extranet oder Intranet einsetzen und registrieren ([Capg02], S. 1). Web Services werden als Transportmedium für reine Maschine-zu-Maschine-Kommunikation verwendet und fungieren somit als Basis für die maschinelle Kommunikation über Applikations- und Organisationsgrenzen hinweg (vgl. [Haus+04], S. 12ff.). Die in Web Services verwendeten Technologien werden ebenfalls durch ein Schichtenmodell (vgl. Abb. 3.3) dargestellt, welches im Folgenden kurz vorgestellt wird.

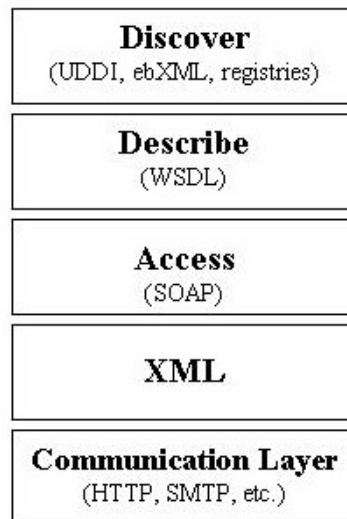


Abb. 3.5: Schichtenmodell von Web Services ([Daco<sup>+</sup>03], S. 59)

Das Hypertext Transport Protocol (HTTP) sowie andere Internetprotokolle werden von Web Services als Transportprotokoll (Communication Layer) verwendet. Die Inhalte des Web Service werden durch XML dargestellt. Dieser Nutzung von Standardprotokollen, sowie die Übertragung in XML garantieren dabei die geforderte Interoperationalität. Der SOAP-Standard stellt die einheitliche Syntax für die Versendung und den Empfang von Web Services zu Verfügung (Access) und basiert ebenfalls auf XML. Die so versendeten Daten werden durch die Web Service Definition Language (WSDL) beschrieben (Describe). WSDL definiert dabei wo der Web Service zu finden ist, welcher Service zur Verfügung gestellt wird, sowie in welcher Form der Service angesprochen werden kann. Die oberste Schicht im Modell der Web Services bildet Universal Description, Discovery and Integration (UDDI). UDDI ist ein öffentliches Register in dem Unternehmen die von ihnen zur Verfügung gestellten Web Services beschreiben (Discover) (vgl. [Daco+03], S. 58).

Der Zusammenhang zwischen dem Semantic Web und Web Services besteht in der Erweiterung von traditionellen Web Services hin zu Semantic Web Services. Werden bisher durch Web Services rein syntaktische Inhalte übertragen, kommt in Semantic Web Services die Ebene der Bedeutung (Semantik) hinzu ([Haus+04], S. 217). RDF und Ontologien spielen dabei eine gewichtige Rolle. Außerdem soll die Suche nach Web Services auf Techniken des Semantic Web basieren und somit erleichtert werden (vgl. [Agar+03], S. 211ff.; [Daco+03], S. 83f.).

Web Services haben in jüngster Vergangenheit eine rasante Entwicklung genommen die sich in einem regelrechten „Hype“ niedergeschlagen hat. Die Technologie für Web Services kann, bis auf die UDDI-Schicht, als standardisiert und ausgereift angesehen werden. Die Verbreitung und kommerzielle Nutzung spiegelt dies jedoch nicht wieder. Laut Studien nutzt nur ein geringer Teil der Unternehmen, trotz vorhandener Infrastruktur, Web Services ([Capg02], S. 8). Die Überlegungen hinsichtlich Semantic Web Services sind dagegen bisher

nur theoretischer Natur, da Teile der dafür notwendigen Technologien des Semantic Web (Ontologien, Vertrauens- und Sicherheitsinfrastruktur) noch nicht ausgereift sind.

## **4 Anwendungen des Semantic Web**

Im Folgenden werden aktuelle Anwendungen des Semantic Web vorgestellt. Dabei wurden bewusst Anwendungen ausgewählt, die den Status eines wissenschaftlichen Prototyps bereits verlassen haben und der breiten Öffentlichkeit zur Verfügung stehen. Die Auswahl erfolgte anhand des Schichtenmodells des Semantic Web. Ziel war es, für jede der höher gelegenen Schichten eine repräsentative Beispielanwendung auszuwählen. Die unteren Schichten des Semantic Web können aufgrund ihres anwendungsfernen Grundlagencharakters vernachlässigt werden.

### **4.1 RDF Site Summary**

RSS (RDF Site Summary) soll im folgenden Abschnitt als Beispielanwendung für die RDF/RDF-Schema-Schicht des Semantic Web beschrieben werden. Grundsätzlich ist RSS ein Dateiformat für den XML-basierten Austausch von Nachrichten aller Art. Daten werden dabei strukturiert abgelegt und für die maschinelle Verarbeitung durch einen RSS-Reader bereitgestellt. Inzwischen hat sich RSS zu einem verbreiteten Standard für den Austausch von Nachrichten und menschlicher Kommunikation, vor allem in Diskussionsforen und Weblogs im WWW entwickelt (vgl. [Charl04]; [Eisen04]). Der Nutzen für den Anwender liegt dabei in der deutlichen Zeitersparnis. Anstatt mehrere Internetseiten täglich persönlich auf aktuelle Nachrichten zu überprüfen, wird der Nutzer nach automatischer Abfrage und Aufbereitung durch sogenannte RSS-Ströme auf neue Inhalte hingewiesen ([Volz03], S. 1). Der Nutzer erhält somit bspw. ein kommentiertes Inhaltsverzeichnis der neuesten Beiträge eines Diskussionsforums. RSS kann auch als Weiterentwicklung des Newsletter-Konzepts interpretiert werden.

Die grundlegende Funktionsweise des RSS-Konzepts gestaltet sich recht einfach. Betreiber einer Webseite erstellen eine RSS-Datei mit den entsprechenden Daten. Die Inhalte sind dabei an die jeweiligen Bedürfnisse anpassbar. Die ursprüngliche Idee besteht darin die wichtigsten Metadaten über neue Informationen bereitzustellen, die Einbindung kompletter neuer Beiträge ist jedoch ebenfalls möglich. Für Diskussionsforen bzw. Weblogs ist die Erstellung solcher RSS-Dateien automatisierbar, für herkömmliche Webseiten jedoch derzeit nur manuell möglich ([Charl04]). Zur Nutzung der daraus erzeugten RSS-Ströme benötigt der Anwender einen RSS-Reader. Beim Besuch einer relevanten Webseite kann der Anwender den RSS-Strom der Seite durch einfügen in seinen RSS-Reader abonnieren. Ändern sich die Inhalte der Webseite wird automatisch ein RSS-Strom erzeugt, an die Abonnenten verschickt und im RSS-Reader dargestellt.

Die Syntax einer RSS-Datei basiert auf einem RDF-Schema und wird in XML abgebildet. Zunächst wird dabei ein sogenannter Channel als Header der Datei definiert, der allgemeine Informationen über den RSS-Strom enthält. Dazu gehören der Titel und die URI der zugrundeliegenden Webseite oder in die Darstellung im RSS-Reader einzubindende Grafiken. Einzelnen Datenblöcke, mit Informationen über neue Inhalte der Webseite, werden anschließend als sogenannte Items aufgezählt. Diese Items können bspw. den Titel, Erstellungszeit, den Autor und eine kurze Zusammenfassung enthalten. Zur Standardisierung dieser übermittelten Metadaten können verschiedene Module in RSS eingebunden werden. Ein weit verbreitetes Modul ist der Dublin Core Metadata Set, welcher eine standardisierte Gruppe von gängigen Metadaten-Elementen zur Verfügung stellt ([Volz03], S. 8ff.).

RSS stellt eine einfache auf XML und RDF basierende Anwendung für das Semantic Web dar. Das Konzept wurde bereits Ende der neunziger Jahre entwickelt und hat sich mittlerweile zu einer Anwendung für breite Nutzerschichten entwickelt. Es ist eine Realisierung der Grundidee des Semantic Web, dem Anwender die Informationssuche durch Automatisierung zu erleichtern (vgl. [Eisen04]). Es existiert eine Vielzahl von ausgereiften RSS-Readern sowie Tools zur automatischen Generierung von RSS-Dateien. Die Nützlichkeit, bei gleichzeitiger Einfachheit, der Technologie hat auch zu ersten Verwendungen in eigentlich fremden Anwendungsgebieten geführt. So existieren Implementierungen von RSS zur Publikation von Änderungsmeldungen in Unternehmenssystemen wie bspw. ERP- und CRM-Systemen ([Volz03], S. 4). Hauptschwachpunkt im aktuellen RSS-Konzept, neben der derzeitigen Versionenvielfalt, ist nach Meinung des Autors die Beschränkung der automatischen RSS-Erzeugung auf Diskussionsforen und Weblogs. Tools zur automatischen Erzeugung von RSS-Dateien für herkömmliche Webseiten in der nahen Zukunft könnten den endgültigen Durchbruch von RSS hin zu einer Standardanwendung für das WWW bedeuten.

## 4.2 Die Ontologie Cyc

Als Beispielanwendung für die Ontologie-Schicht, im Modell des Semantic Web, soll im Folgenden die Ontologie Cyc vorgestellt werden. Das Projekt Cyc wurde bereits im Jahre 1984 gestartet um im großen Maßstab Alltags- und enzyklopädisches Wissen in einem einheitlichen System zu formalisieren. Formalisiert heißt, dass die Ontologie speziell auf die Anwendung durch Computer ausgerichtet ist. Im Gegensatz zu Expertensystemen, die sich auf eine enge Wissensdomäne beschränken, wird mit Cyc versucht eine möglichst große Spanne des allgemeinen menschlichen Wissens abzudecken (vgl. [Fisc04], S. 2). Zum heutigen Zeitpunkt umfasst die Cyc-Ontologie circa drei Millionen Faustregeln, sowie 300 000 Terme oder Konzepte aus dem täglichen Leben. Abbildung 4.1 zeigt einen Ausschnitt aus dem obersten Teil der Ontologie.

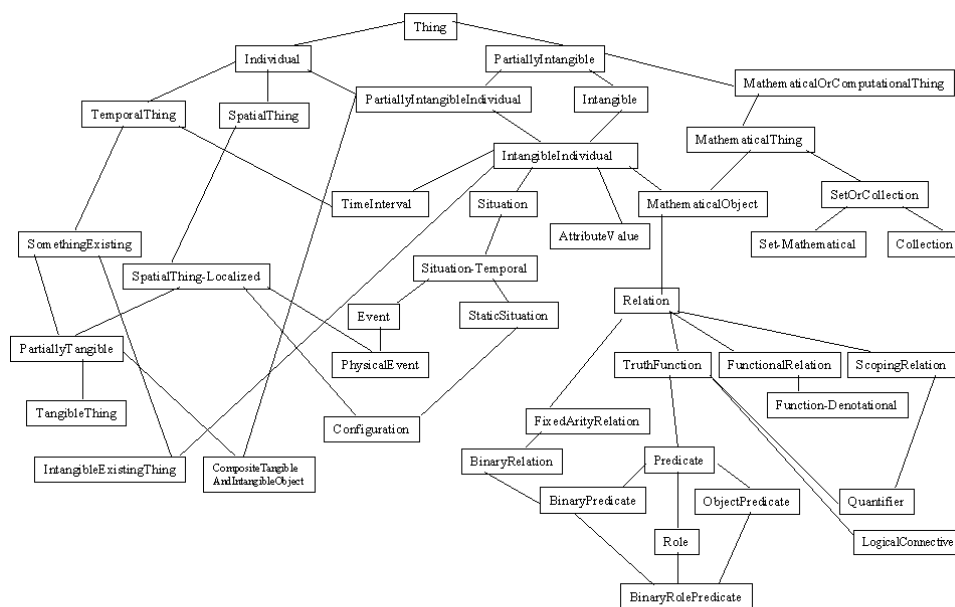


Abb. 4.1: Cyc-Ausschnitt ([Cyc04])

Das ursprüngliche Ziel von Cyc, die Fähigkeit des Systems selbständig zu lernen oder dem Nutzer mitzuteilen wenn nicht genügend Informationen zur Verfügung stehen, konnte bisher nicht erreicht werden. Cyc wird heute u. a. angewendet um große Bildbestände semantisch zu verwalten oder um Datenbanken abzugleichen bzw. zu bereinigen. Das System ist dabei fähig automatisch semantische Fehler aufzuspüren oder passende Bilder zu bestimmten Aussagen zu finden. Speziell im WWW wird Cyc durch die Suchmaschine Lycos genutzt, welche die Ontologie zur Verbesserung ihrer Suchergebnisse benutzt. Seit 1994 wird Cyc kommerziell vermarktet. Ein Teil des Systems steht den Nutzern jedoch als OpenSource zur Verfügung. Cyc nutzt keine der Standard-Ontologiesprachen wie etwa OWL, sondern eine eigene Entwicklung, CyL.

Trotz erster sinnvoller Anwendung von Cyc, sind deren Schöpfer von ihren selbst gesetzten Zielen weit entfernt (vgl. [Egge02], S. 17). Die bereits beschriebenen Probleme bei der Entwicklung von Ontologien werden bei eine umfassenden Variante wie Cyc besonders deutlich. Fragen nach der Explikation von impliziten menschlichen Wissen oder dem Einfluss durch Subjektivität der Ersteller sind bis heute nicht endgültig beantwortet. So spiegelt sich in Cyc, einer us-amerikanischen Entwicklung, deutlich der amerikanische Kulturraum wieder. Ein weiteres Hindernis auf dem Weg zur Vollendung von Cyc ist die komplizierte Schnittstelle zur Dateneingabe. Diese ist nur durch Experten auf dem Gebiet der Logik oder Informatik möglich. Menschen die nicht zu dieser Gruppe gehören, jedoch relevantes und wertvolles Wissen aus anderen Gebieten zusteuern könnten, wird der Zugang zu Cyc so quasi verweigert. Zum Durchbruch könnte Cyc eine natürlichsprachliche Zugangsschnittstelle verhelfen, deren Verwirklichung zur Zeit jedoch noch nicht möglich ist.

Nach anfänglicher Euphorie ist mittlerweile Ernüchterung im Bezug auf Cyc eingezogen. Wissenschaftler zweifeln inzwischen an der grundsätzlichen Machbarkeit einer allumfassenden Ontologie und schlagen stattdessen die Zusammenfassung von Ontologien, welche spezielle Wissensdomänen abbilden, vor (vgl. [Egge02], S. 22).

### **4.3 Friend of a Friend**

Für den hierarchisch am höchsten gelegenen Teil des Modells des Semantic Web, die Schicht der Vertrauens- und Sicherheitsinfrastruktur, existieren bisher keine ausgereiften Anwendungen für die Öffentlichkeit (vgl. 3.5). Im Folgenden wird deshalb, mit „Friend of a Friend“ (FOAF), eine RDF-Anwendung beschrieben, die grundsätzlich auch Erweiterungsmöglichkeiten im Hinblick auf die Vertrauens- und Sicherheitsinfrastruktur bietet.

FOAF ist ein einfaches RDF-Vokabular mit dem es möglich ist persönliche Informationen und Beziehungen auszudrücken. Es werden Metadaten über Personen erstellt. Die FOAF-Dateien werden auf Webservern dezentral abgelegt und können maschinell ausgewertet werden. Durch die nichtzentrale Verwaltung ergeben sich für den Nutzer Vorteile im Hinblick auf die Kontrolle über seine veröffentlichten Daten, sowie die jederzeit mögliche Aktualisierung. Derzeitige Anwendungen beziehen sich vor allem auf den Aufbau und die Betreuung von Online-Communities (vgl. [Dumb04]). Interessante Perspektiven ergeben sich durch die Aggregation von FOAF-Dateien und der Extraktion von bspw. Querverbindungen. Durch Erweiterungen des herkömmlichen FOAF-Vokabulars um bestimmte Vertrauensaussagen kann FOAF ebenfalls als eine der ersten Anwendungen für ein zukünftiges Web of Trust interpretiert werden (vgl. [Reag04]). Es wird dabei der bereits diskutierte Ansatz „Web of Trust mit expliziten Vertrauensaussagen“ verfolgt (vgl. 3.5). Teile der Daten können dabei öffentlich zugänglich gemacht werden, während andere Teile verschlüsselt und somit nur für bestimmte Personen lesbar abgelegt werden. Die W3C-Standards XML Signature und XML Encryption werden dabei angewendet. Auf dem Markt

existieren bereits zahlreiche Software-Applikationen die es bspw. erlauben FOAF-Dateien einfach und ohne Kenntnis der RDF/XML-Syntax zu erstellen.

FOAF hat sich als interessante Anwendung im Bereich der sozialen Netzwerke etabliert. Erste Versuche mit FOAF auch die Ideen eines Web of Trust zu verwirklichen, sind dagegen bisher wenig aufgereift und verbreitet. FOAF bietet in diesem Bereich jedoch grundsätzlich diverse Ansatzpunkte. Das starke Interesse an sozialen Netzwerken im Allgemeinen könnte somit auch zu Weiterentwicklungen der Sicherheits- und Vertrauensinfrastruktur im Semantic Web führen.

## **5 Gegenwart und Zukunft des Semantic Web**

In dem folgenden, diese Arbeit abschließenden, Kapitel wird zunächst der aktuelle Entwicklungsstand des Semantic Web noch einmal zusammenfassend beschrieben. Anschließend werden die Faktoren aufgelistet, die eine schnellere Weiterentwicklung des Semantic Web derzeit behindern. Abschließend wird versucht die kurz- und mittelfristigen Entwicklung vorherzusagen, indem die Meinungen verschiedener Experten zusammengefasst werden.

### **5.1 Aktueller Entwicklungsstand**

Das Semantic Web ist heute keine Fiktion mehr. Es existiert zumindest teilweise und wird auch genutzt. Die Situation ist dabei durchaus vergleichbar mit der in den frühen Tagen des traditionellen WWW ([Köni02], S. 483). Die Tätigkeiten beschränken sich nicht mehr nur auf das akademische Umfeld, den auch Unternehmen nutzen bereits limitierte Anwendungen des Semantic Web (vgl. [Viss03a], S. 39). Derzeit verfügbare Applikationen fungieren dabei jedoch nur als Insellösungen.

Von einem ausgereiften Semantic Web kann somit nicht gesprochen werden. Die Entwicklung befindet sich noch immer im Stadium des Aufbaus der Infrastruktur. In den letzten Jahren hat das W3C, mit der Standardisierung der grundlegenden Schichten (bspw. XML, RDF) des Semantic Web, ein stabiles Fundament geschaffen. Die darauf aufsetzenden Techniken befinden sich jedoch noch im frühen Entwicklungsstadium oder haben trotz eines ansprechenden Reifegrades mit einem mangelndem Bekanntheitsgrad und geringen Anwenderzahlen zu kämpfen.

Aktuelle, dem Semantic Web eigentlich nicht verwandte, Trends wie Weblogging oder Social Networks, führen derzeit indirekt zu einer Verbreitung von Technologien des Semantic Web in breitere Nutzerschichten (vgl. 4.1; 4.3).

### **5.2 Hemmnisse**

Bei dem Semantic Web handelt es sich zwar um eine Weiterentwicklung des traditionellen WWW, jedoch mit revolutionären Veränderungen. Tim Berners-Lee spricht in diesem Zusammenhang sogar von einem Paradigmenwechsel ([Frau04], S. 52). Ein solcher Wandel ist grundsätzlich schwierig. Für die langsame Verbreitung des Semantic Web existieren zahlreiche Gründe. Die wichtigsten Ursachen werden im Folgenden aufgezählt (vgl. [Schn03], S. 62):

- frühes Entwicklungsstadium der Technologien der oberen Schichten im Modell des Semantic Web (Ontologien, Sicherheits- und Vertrauensinfrastruktur) (vgl. 3.4; 3.5),
- höhere Komplexität der Techniken des Semantic Web im Vergleich zum traditionellen WWW,

- mangelnde Nützlichkeit der ersten Anwendungen des Semantic Web für breite Nutzerschichten,
- Fehlen einer „Killer-Applikation“ für das Semantic Web,
- Mangel an nichtwissenschaftlichen Publikationen und Anleitungen zu Techniken des Semantic Web, sowie
- geringe Netzwerkeffekte auf Grund niedriger Anwenderzahlen.

### **5.3 Zukünftige Entwicklung**

Führende Experten sehen insgesamt eine positive Entwicklung des Semantic Web in den kommenden Jahren voraus. Für die nähere Zukunft wird vor allem eine große Anzahl an Prototypen erwartet, welche die Fähigkeiten des Semantic Web deutlicher zu Geltung bringen werden als bisherige Anwendungen. Die Nutzung von Ideen des Semantic Web in Standardsuchmaschinen des WWW ist in den kommenden Jahren ebenfalls zu erwarten ([Schn03], S. 62). Deutliche Fortschritte in den bisher noch im Entwicklungsstadium befindlichen Schichten des Semantic Web werden auch diese Teilbereiche zu einem Reifegrad führen, welcher Standardisierung und effektive Nutzung ermöglicht. Zu diesen Gebieten zählen vor allem Ontologien und die Sicherheits- und Vertrauensinfrastruktur.

In der folgenden Phase wird es zu einer zunehmenden Integration der bisherigen Insel-Applikationen kommen sowie zur Verbreitung von Werkzeugen und Applikationen, welche auch von Nicht-Experten genutzt werden können. Fortschritte sind dann ebenfalls im Bereich der Semantic Web Services zu erwarten. Im Allgemeinen wird es zu einer Kommerzialisierung des Semantic Web, einhergehend mit einer zunehmenden Durchdringung von breiten Nutzerschichten, kommen (vgl. [Viss03a], S. 39).

Mittelfristig wird das Semantic Web als allgegenwärtiger Teil des Webs wahrgenommen werden. Ein Unterschied zwischen WWW und Semantic Web ist dann nicht mehr existent (vgl. [Schn03], S. 62; [Viss03b], S. 41).

## Literaturverzeichnis

- [Agar<sup>+</sup>03] Agarwal, S., Handschuh, S., Staab, S. Surfing the Service Web. In: Fensel, D., Sycara, K., Mylopoulos, J. (Hrsg.), The Semantic Web – ISWC 2003. Springer Verlag, Berlin, Heidelberg, New York, 2003, S. 211-226.
- [Bize03] Bizer, C. Bausteine einer Vertrauens- und Sicherheitsinfrastruktur für das Semantic Web. [http://www.wiwiss.fu-berlin.de/suhl/bizer/pub/XMDX03\\_SemTrust.pdf](http://www.wiwiss.fu-berlin.de/suhl/bizer/pub/XMDX03_SemTrust.pdf), Abruf am 2004-12-12.
- [Böhm<sup>+</sup>03] Böhme, R., Michel, K.-U. Einsatzpotentiale von Topic Maps: Konzeption und prototypische Umsetzung eines verteilten Redaktionssystems. Diplomarbeit, Lehrstuhl für Wirtschaftsinformatik, insbesondere Informationsmanagement, Prof. Dr. E. Schoop, TU Dresden, 2003.
- [Capg02] Capgemini: Der Markt für Web Services: Erwartungen, Treiber, Investitionsabsichten. <http://www.de.capgemini.com/servlet/PB/show/1004620/Web-Services.pdf>, Abruf am 2004-12-13.
- [Charl04] Charlier, M. Surfen ist out: RSS knüpfen Netze im Netz – und das auch noch vollautomatisch. <http://www.webwriting-magazin.de/inter/rss.php>. Abruf am 2004-12-13.
- [Cyc04] Cycorp. OpenCyc Selected Vocabulary an Upper Ontology. <http://www.cyc.com/cycdoc/vocab/upperont-diagram.html>, Abruf am 2004-12-16.
- [Daco<sup>+</sup>03] Daconta, M. C., Obrst, L. J., Smith, K. T. The Semantic Web: A Guide to the Future of XML, Web Services and Knowledge Management. Wiley Publishing, Indianapolis, 2003.
- [Dubl04] Dublin Core Metadata Initiative. About the Initiative. <http://www.dublincore.org/about/>, Abruf am 2004-12-09.
- [Dumb04] Dumbill, E. XML Watch: Finding Friends with XML and RDF. <http://www-106.ibm.com/developerworks/xml/library/x-foaf.html>, Abruf am 2004-12-13.
- [Dzbo<sup>+</sup>03] Dzor, M., Domingue, J., Motta, E. Magpie – Towards an Semantic Web Browser. In: Fensel, D., Sycara, K., Mylopoulos, J. (Hrsg.), The Semantic Web – ISWC 2003. Springer Verlag, Berlin, Heidelberg, New York, 2003, S. 690 – 705.
- [Egge02] Egger, C. Cyc – Repräsentation von Commonsense Wissen. Seminar Dr. Kai-Uwe Carstensen, Aspekte der Wissensrepräsentation in der Computerlinguistik, Institut für Computerlinguistik, Universität Zürich, <http://www.ifi.unizh.ch/cl/carstens/seminare/wisrep2002/EggerCycArbeit.pdf>, Abruf am 2004-12-13.
- [Eisen04] Eisenzopf, J. Making Headlines with RSS – Using Rich Site Summaries To Draw New Visitors. <http://www.newarchitectmag.com/archives/2000/02/eisenzopf/>, Abruf am 2004-12-13.

- [Fens<sup>+</sup>03] Fensel, D., Hendler, J., Liebermann, H., Wahlster, W. Spinning the Semantic Web: Bringing the World Wide Web to Its Full Potential. Massachusetts Institute of Technology, Cambridge, London, 2003.
- [Fisc04] Fischer, D. Ein Lehrbeispiel für eine Ontologie: OpenCyc. <http://www.ipsi.fraunhofer.de/orion/pubFulltexts/MetaOpenCyc20Feb04.pdf>, Abruf am 2004-12-13.
- [Frau04] Frauenfelder, M. Semantic Web. In: Technology Review – Das M. I. T. – Magazin für Innovation, 11 (2004), S. 50-56.
- [Haus<sup>+</sup>04] Hauser, T., Löwer, U. M. Web Services: Die Standards. Galileo Press GmbH, Bonn, 2004.
- [Hend<sup>+</sup>02] Hendler, J., Berners-Lee, T. Integrating Applications on the Semantic Web. In: Journal of the Institute of Electrical Engineers of Japan, 122 (2002) 10, S. 676-680.
- [HölB03] Hölbling, A. Technologien des Semantic Web. In: Tolksdorf, R., Eckstein, R. (Hrsg.), XML Clearinghouse Report. <http://141.20.27.87/webportal/reports/Technologien%20des%20Semantic%20Web.pdf>, Abruf am 2004-12-09.
- [Kara04] Karadeniz, B. Uniform Resource Identifiers. <http://www.netplanet.org/adressierung/uri.shtml>, Abruf am 2004-12-09.
- [Köni02] König, W. Interview with James Hendler on the Semantic Web. In: Wirtschaftsinformatik, 44 (2002) 5, S. 481-483.
- [Mädc<sup>+</sup>01] Mädche, A., Staab, S., Studer, R. Ontologien. In: Wirtschaftsinformatik, 43 (2001) 4, S. 3939-395.
- [Mano<sup>+</sup>04] Manola, F., Miller, E. RDF Primer. <http://www.w3.org/TR/rdf-primer/>, Abruf am 2004-12-12.
- [Mück<sup>+</sup>01] Mück, T. A., Widhalm, R. Topic Maps. In: Wirtschaftsinformatik, 43 (2001) 3, S. 297-300.
- [Ratn<sup>+</sup>02] Ratnakar, V, Gil, Y. Trusting Information Sources One Citizen at a Time. <http://www.isi.edu/%7Egil/papers/GilRatnakarISWC2002.pdf>, Abruf am 2004-12-12.
- [Reag04] Reagle, J. FOAF Spheres of Privacy. <http://reagle.org/joseph/2003/09/foaf-spheres.html>, Abruf am 2004-12-13.
- [Schn<sup>+</sup>00] Schneider, U., Werner, D. Taschenbuch der Informatik. 3. Auflage, Fachbuchverlag Leipzig im Carl Hansen Verlag, München, Wien, 2000.
- [Schn03] Schneeberger, J. Interview mit Mario Jeckle. In: Künstliche Intelligenz, 03/03, S. 62.
- [Unic04] Unicode Consortium: Was ist Unicode ?. <http://www.unicode.org/standard/translations/german.html>, Abruf am 2004-12-09.
- [Viss03a] Visser, U. Interview with James Hendler. In: Künstliche Intelligenz, 03/03, S. 39-40.

- [Viss03b] Visser, U. Interview with Patrick Hayes. In: Künstliche Intelligenz, 03/03, S. 41-42.
- [Volz03] Volz, R. RSS kurzgefasst. [http://www.vrtprj.de/content/istandards/rssguide\\_de.pdf](http://www.vrtprj.de/content/istandards/rssguide_de.pdf), Abruf am 2004-12-13.
- [W3C04a] World Wide Web Consortium. Semantic Web – Activity Statement. <http://www.w3c.org/2001/sw/Activity>, Abruf am 2004-12-09.
- [W3C04b] World Wide Web Consortium. Uniform Resource Identifier (URI) – Activity Statement. <http://www.w3.org/Addressing/Activity>, Abruf am 2004-12-09.
- [Widh+02] Widhalm, R., Mück, T. Topic Maps. Springer-Verlag, Berlin, Heidelberg, New York, 2000.