

Institut für Informatik der Universität Zürich
Sommersemester 2001

Seminar Datenbanktechnologie fürs Web

Semantic Web

Prof. Dr. K.R. Dittrich

Betreuende Assistenten:
Ruxandra Domenig
Farshad Hakimpour

vorgelegt von:

Urs Odermatt
Schörlistrasse 5, 8050 Zürich
urs.odermatt@access.unizh.ch
Matrikelnummer 98-717-416

Abgabetermin: 2. Juli 2001

Inhaltsverzeichnis

1. EINLEITUNG	2
2. SEMANTIK IM WEB HEUTE	3
2.1. Das Internet heute.....	3
2.2. Suchen und Finden im WWW.....	3
2.2.1. Beispiel: Bildersuche.....	3
2.2.2. Maschinerverständliche Information.....	4
2.3. Metainformation.....	4
2.3.1. Wie man Dinge findet.....	4
2.3.2. Metainformationen in HTML.....	5
3. RESOURCE DESCRIPTION FRAMEWORK (RDF)	7
3.1. Überblick.....	7
3.2. Ziele.....	7
3.3. Beschreibung.....	7
3.3.1. RDF Datenmodell.....	7
3.3.2. RDF Syntax.....	8
3.3.3. RDF Schema (RDFS).....	9
3.4. Verbreitung von RDF Metadaten.....	11
3.5. Abfrage und Interpretation von RDF Metadaten.....	11
3.6. Fazit.....	13
4. ONTOLOGIEN	14
5. AUSBLICK	15
6. LITERATURVERZEICHNIS	16

1. Einleitung

“Semantic Web“ ist neuerdings zum Schlagwort für die nächste Generation des World Wide Webs geworden. Das Semantische Web ist eine Vision von Tim Berners-Lee [1,2], „Erfinder“ des WWW und Präsident des W3-Consortiums (W3C)¹, in dem alle Webressourcen mit semantischen Metadaten versehen sind. Die Semantik ist die Lehre von der Bedeutung sprachlicher Zeichen².

Das Semantic Web ist kein neues, separates Web, sondern eine Erweiterung des derzeit existierenden Webs. Die dahinterliegende Idee ist, einen Mechanismus zur Verfügung zu stellen, so dass Daten im Web durch maschinelle Akteure erarbeitet und interpretiert werden können. Es handelt sich deshalb auch nicht einfach um eine Anwendung, sondern vielmehr um eine generelle Vision und Architektur für die Entwicklung von webbasierten Anwendungen der nächsten Generation³. [5,6]

Angesichts der steigenden Informationsflut im Internet zeigen sich immer deutlicher die Grenzen der heutigen Publikationsmethoden. Da HTML (Hypertext Markup Language) primär als Präsentationsformat dient, bestehen Web-Seiten aus einem Gemenge von Layout-Informationen und den eigentlichen Inhalten. Beschreibungen, worum es in einzelnen Dokumenten überhaupt geht, reduzieren sich bestenfalls auf ein paar "Meta"-Tags⁴. Entsprechend schwierig ist es für Softwareroboter, daraus gezielt Daten zu extrahieren oder die Inhalte zu taxieren. Diese Beschränkungen bekommen Benutzer regelmässig zu spüren, wenn sie von Suchmaschinen mit irrelevanten Suchergebnissen überschüttet werden.

Das W3C will ein "semantisches Web". Die Initiative des W3C für ein solches Web [14] beruht auf hauseigenen und darauf aufbauenden Standards zur Formulierung von Metadaten, die zu einem übergreifenden Framework zusammengefasst werden. Dies mit dem obersten Ziel, Web-Inhalte möglichst gut maschinell verarbeiten zu können und so neuen Content-bezogenen Dienstleistungen Auftrieb zu geben. Dreh- und Angelpunkt für die Darstellung von Metadaten soll das Resource Description Framework (RDF) sein, auf das ich im dritten Kapitel näher eingehen möchte. [10]

¹ <http://www.w3.org>

² DUDEN, Die Deutsche Rechtschreibung

³ <http://www.semanticweb.org/>

⁴ siehe 2.3.2.

2. Semantik im Web heute

2.1. Das Internet heute

Die Welt ist zusammengedrückt. Das globale World Wide Web bietet neue Möglichkeiten und Chancen zum Austausch und Beschaffen von Informationen: Homepages, Newsgroups, Email, Chatrooms, Internet-Telefonie, E-Business...

Die Webgemeinde vergrößert sich von Tag zu Tag. Die Anzahl der Webseiten, die ins Internet gestellt werden, vermehrt sich unaufhaltsam. Es gibt in der Zwischenzeit sicherlich fast nichts mehr, das es nicht im World Wide Web zu finden gäbe, doch die Frage ist nur, wie man es findet.

2.2. Suchen und Finden im WWW

2.2.1. Beispiel: Bildersuche

Die Suchmaschine Altavista⁵ bietet die Möglichkeit, speziell nach Bildern zu suchen. Anhand des folgenden Beispiels soll das semantische Problem des World Wide Webs, wie es heute besteht, betrachtet werden:

Ein Lehrer behandelt in der Schule mit seiner Klasse das Thema Raubtiere und ist unter anderem auf der Suche nach einem Bild eines Jaguars. Er tippt das Wort „Jaguar“ bei Altavista ein und erhält eine Liste gefundener Bilder.



Er klickt auf die erste Datei und erhält eine schwarz weisse Zeichnung des gesuchten Raubtiers. Da er aber nach einem farbigem Foto sucht, klickt er auf die nächste Bilddatei in der Liste.



Das ist das zweite Bild und es hat direkt nichts mit seiner Suche zu tun. Es ist ein Auto der Marke „Jaguar“. Also versucht er es noch mit der dritten Datei.



Auch dieses Bild hat keinen Bezug zu einem Raubtier. Es ist ein Kampfflugzeug vom Typ „Jaguar“.

Wenn man weiss, wie Suchmaschinen arbeiten, erstaunt dieses Ergebnis keinesfalls. Die Suchmaschine suchte einfach nach einer Bilddatei mit dem Namen „Jaguar“. Das heisst, es werden alle Bilder mit der Bezeichnung „jaguar.gif“ oder „jaguar.jpg“ gesucht und falls gefunden

⁵ <http://ch-de.altavista.com/searchimg?stype=simage>

aufgelistet. Wenn das Wort jedoch mehrere Bedeutungen hat, wie in diesem Fall, ist die Suchmaschine nicht in der Lage, die verschiedenen Aspekte zu unterscheiden.

2.2.2. Maschinenverständliche Information

Das Konzept der maschinenverständlichen Dokumente beinhaltet keine künstliche Intelligenz welche die Maschinen befähigen soll, die Menschen zu verstehen. Es bezieht sich lediglich auf die Möglichkeit, dass Maschinen fähig sein sollen, ein definiertes Problem mit definierten Operationen auf existierenden, definierten Daten zu lösen. [2]

Somit muss also der Mensch einen Weg finden, die Daten so im Web bereit zu stellen, dass sogenannte „Agenten“ die gewünschten Informationen finden und auch weiterverarbeiten können.

Die verfügbaren Daten im Internet werden immer weiter zunehmen. Wenn wir nicht im totalen Informationschaos enden wollen, müssen jedoch die Voraussetzungen geschaffen werden, dass Maschinen für uns den Überblick behalten. Das heisst, das Web soll sich vom „maschinenlesbaren“ zum „maschinenverständlichen“ Medium wandeln. Es müssen somit Informationen über Informationen geschaffen werden, sogenannte Metainformationen, wie sie im nächsten Kapitel besprochen werden.

2.3. Metainformation

2.3.1. Wie man Dinge findet

Das Internet, wie es heute besteht, dient hauptsächlich als Forum für menschliche Interaktion. Da die meisten Dokumente für den menschlichen Gebrauch geschrieben sind, ist die einzige Möglichkeit der Internetsuche, nach übereinstimmenden Wörtern oder Sätzen in den Dokumenten zu suchen. Jeder, der schon einen Suchdienst wie Altavista oder Google benutzt hat, weiss, dass das Eintippen von ein paar Schlüsselwörtern und das Erhalten von Hunderten von „Treffern“ nicht sehr nützlich ist. Man muss anschliessend das Gesuchte selbst herausfiltern. Zudem kann es auch sein, dass das gesuchte Schlüsselwort, gar keine bedeutende Stellung im relevanten Dokument einnimmt.

Es muss also eine Möglichkeit bestehen, die Dinge im Web exakt zu beschreiben. Genau das ist die Aufgabe von Metainformationen. Metadaten sind „Daten über Daten“ oder in unserem aktuellen Kontext: „Daten die Webressourcen beschreiben“.

Wir alle haben schon, wenn auch nicht bewusst, Gebrauch von Metainformationen gemacht: [3]

Beispiel 1: Die Bibliothek

Man ist in der Bibliothek und sucht in all diesen zahlreichen Gestellen nach einem bestimmten Buch. Heutzutage haben die meisten Bibliotheken ein Computersystem zur Suche eines Buches, grundlegend einfach eine elektronische Version der alten Karteikarten. Dieses System erlaubt es nun, die Bücher nach Autor, Titel, Begriffen usw. zu suchen. Die Liste enthält Datum, Autor, Titel und viele weitere nützliche Informationen, inklusive der wichtigsten von allen, nämlich wo das Buch zu finden ist.

Beispiel 2: Der Videoladen

Man ist in einem Videoladen und möchte einen Spielfilm von James Cameron. Ein grosser Videoladen besitzt auch ein Computersystem das der Bibliothek ähnlich ist, doch sind die Suchkriterien anders. Aber im Grunde genommen ist es dasselbe.

Beispiel 3: Das Telefonbuch

Man ist spät am Abend noch im Büro an der Arbeit. Um effizient weiter arbeiten zu können, müsste der Hunger nun aber mit einer Pizza gestillt werden. In den „Gelben Seiten“ findet man schnell einen Pizzakurier, um eine Bestellung aufzugeben.

Was haben diese drei Situationen gemeinsam und was unterscheidet sie? Erstens, jedes dieser Systeme basiert auf Metadaten. In jedem Beispiel benötigt man eine kleine Information (der Ort des Buches, der Name des Videos, die Telefonnummer des Pizzakuriers), die man nicht hat. Bei jedem Beispiel benutzen wir Metadaten, um die gewünschten Informationen zu bekommen. Theoretisch sind Metainformationen nicht notwendig. Man könnte einfach durch die Bibliothek oder durch den Videoladen gehen, bis man den gewünschten Titel gefunden hat oder einfach herumtelefonieren, bis man einen Pizzakurier erreicht. Das wäre aber sehr ineffizient und zudem nicht sehr intelligent. Im Internet wäre es sogar ein Ding der Unmöglichkeit alle Webseiten zu durchschauen, bis man das gesuchte Dokument gefunden hat. Metainformationen kommen im World Wide Web also eine grosse Bedeutung zu.

2.3.2. Metainformationen in HTML

Seit der HTML Version 3.2 erfolgt die Ablage von Metadaten durch die im Kopfbereich eines Dokumentes unterzubringenden "Meta"-Tags. Dieser Metatag besitzt zwei Attribute, den `name` und den `content` Operator.

Die möglichen Inhalte für diesen TAG können beispielsweise den Autor des Dokumentes, das Erstellungsdatum, die Sprache verzeichnen. Eine genaue Auflistung der vorgesehenen Inhalte erfolgte im Dublin Core⁶.

```
<html>
  <head>
    <title>Graubuenden Ferien</title>
    <META name="author" content="Graubuenden Ferien">
    <META name="publisher" content="Graubuenden Ferien">
    <META name="keywords" content="Graubuenden, Ferien, Tourismus">
    <META name="description" content="Die Fereinecke der Schweiz">
    <META name="copyright" content="Graubuenden Ferien">
    <META name="robots" content="all">
  </head>
  <body>
    <p>
  </body>
</html>
```

Ein solches mit Metatags beschriebenes Dokument offenbart seinen Inhalt auf eindeutige Weise und gibt damit von sich aus Suchmaschinen, Katalogsystemen und dem Nutzer der Seite die Möglichkeiten, Informationen über das Dokument zu erhalten und so eine Einordnung der Inhalte vorzunehmen.

Leider ist das nur ein eingeschränktes Mittel zur Verwendung von Metadaten. Zudem haben wir im vorangehenden Bildersuchbeispiel bei Altavista gesehen, dass dasselbe Wort verschiedene Bedeutungen haben kann. Für eine Maschine ist es nicht ohne weiteres klar, ob wir unter dem Begriff „Jaguar“ ein Tier, ein Auto oder eben ein Kampfflugzeug suchen. Wenn wir etwas be-

⁶ siehe 3.3.2.

schreiben durch Metainforamtionen, brauchen wir Begriffe, die wieder eine Beschreibung benötigen, das heisst Metainformationen über Metainformationen. Es entsteht ein Netzwerk mit Beziehungen untereinander, das uns als Entity-Relation-Model (ERM) bekannt ist. [2]

ER ist ein generelles Modell, das Metainformationen im World Wide Web gut darstellen kann. Das W3C entwickelte ein spezielles Modell für Web, das Resource Description Framework, kurz RDF, was Gegenstand des nächsten Kapitels ist.

3. Resource Description Framework (RDF)

3.1. Überblick

RDF ist seit Februar 1999 eine W3C-Empfehlung [16] und geht unter anderem auf Eingaben und Vorschläge von Netscape und Microsoft zurück. Allerdings fand die Technik bisher geringe Verbreitung. Auf ihrer Basis wurden kaum Beschreibungssprachen entwickelt.

Das RDF legt ein abstraktes Verfahren fest, um die Zusammenhänge zwischen Entitäten und deren Merkmalen zu beschreiben. Es ist einerseits an keine bestimmte Syntax gebunden, andererseits setzt der Austausch von konkreten Metadaten eine solche aber voraus. Innerhalb des W3C wurde deshalb eine RDF Spezifikation [16] auf Basis der Extensible Markup Language (XML) entwickelt. Ergebnis dieser Standardisierungsbemühungen ist unter anderem "RDF Schema" [15]. Dieses definiert ein System, mit dem sich Grammatiken für bestimmte Arten von Metadaten entwickeln lassen. Dazu müssen Klassen von Ressourcen und Eigenschaften spezifiziert sowie deren Beziehungen untereinander dargelegt werden. Damit kommt RDF Schema bei der Definition von RDF Anwendungen die gleiche Aufgabe zu, die XML Schema für XML-Applikationen übernehmen soll. Um unnötige Inkonsistenzen zwischen den beiden Spezifikationen zu vermeiden, sollen zumindest gemeinsame Aspekte wie die Festlegung von primitiven Datentypen einheitlich gehalten werden. [13]

3.2. Ziele

RDF ist als Metasprache für Metadaten charakterisierbar. Dem Anwender soll dabei die Arbeit mit dem Netz erleichtert werden, wobei zwei Ziele im Vordergrund stehen: [7,8]

- Informationen sollen besser katalogisiert werden, wobei die Abhängigkeiten zwischen Dokumenten transparenter werden sollen.
- Es soll eine Prozedur entwickelt werden, die die Metadaten eines Dokumentes verarbeiten kann, um somit automatische Prozesse ablaufen zu lassen. Zwar sind alle Informationen im WWW maschinenlesbar, sie können aber nicht von Maschinen verstanden, semantisch erfasst werden. Automatisierte Auswertungen können dazu beitragen, dass dem Anwender nur solche Informationen angezeigt werden, die er als interessant klassifiziert hat.

3.3. Beschreibung

3.3.1. RDF Datenmodell

Das grundlegende RDF Datenmodell zur Erreichung der genannten Ziele besteht aus folgenden drei Objekten: [3]

- **Ressourcen** sind alle Objekte, die durch RDF Ausdrücke beschreibbar sind. Dieses könnte zum Beispiel eine HTML-Seite, irgend eine Datei im Web oder eine Email-Adresse sein.
Ressourcen sind immer durch **URIs** (Uniform Resource Identifier) bezeichnet und wie gesagt, alles kann einen URI haben.
- **Eigenschaften** sind im Sinne von RDF spezielle Aspekte, Attribute oder Charakteristika, die ein Objekt näher spezifizieren. Eigenschaften werden ebenfalls durch eine Ressource beschrieben und beinhalten einen Wert, der ein Literal (Text) oder wiederum eine Ressource sein kann.

- **Aussagen** sind der beschreibende Inhalt des durch die beiden vorherigen Aspekte spezifizierten Objektes.

Eine Ressource, die bezeichnende Eigenschaft und der beschreibende Inhalt (Wert) dieser Eigenschaft, bilden im RDF Kontext eine Einheit bestehend aus **Subjekt**, **Prädikat** und **Objekt** - sie bilden eine RDF Aussage.

Ein RDF Aussage kann als gerichteter Graf dargestellt werden: Subjekte und Objekte sind Knoten und jedem Prädikat (Eigenschaft) entspricht eine Kante, die vom Subjekt zum Objekt weist. [7,8]



Der obige Graf kann folgendermassen gelesen werden:
 "Urs Odermatt ist der Gestalter von home.html"

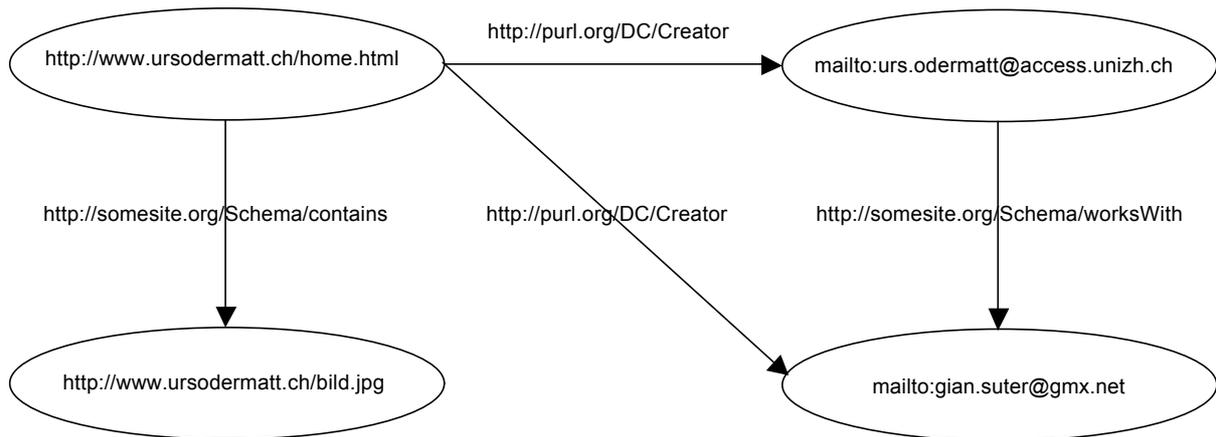
3.3.2. RDF Syntax

RDF beschreibt ein einfaches, aber mächtiges Modell zum Beschreiben von Ressourcen. Eine Syntax zur Darstellung dieses Modells wird benötigt, um maschinenverständliche Informationen übers Web auszutauschen und um die Interoperabilität zwischen Applikationen zu unterstützen. RDF benutzt die formale Struktur von XML zur Unterstützung der konsistenten Repräsentation der Semantik.

RDF bietet die Möglichkeit, die Semantik zu definieren. Nun ist es aber wichtig, diese Semantik zu vereinheitlichen. Das Prädikat „Creator“ zum Beispiel mag verschiedene Bedeutungen haben, abhängig von der Gemeinschaft, die das Wort benutzt. Problematisch wird es, wenn verschiedenen Gemeinschaften das selbe Prädikat benutzen, aber unterschiedliche Dinge darunter verstehen. Um das zu verhindern, benutzt RDF die XML Namespace-Methode zur einheitlichen Identifizierung von Prädikaten. Mittels Namespaces in XML lässt sich das Vokabular eines Prädikates eindeutig zuweisen.

Als eines der wenigen bekannten Beispiele für eine RDF-Anwendung dient "Dublin Core" (DC)⁷. Sie entspringt dem Bibliotheksumfeld und legt ein Vokabular fest, mit dem sich der Autor, das Thema oder die Urheberrechte eines Dokuments beschreiben lassen. Um somit auf eine entsprechende Ressource zu verweisen, die die semantische Bedeutung (Schema) des Dublin Core-Vokabulars festlegt, wird ein sogenannter **PURL** (Persistent Uniform Resource Locator) verwendet. [8]

⁷ <http://dublincore.org/>



Die entsprechende syntaktische Schreibweise des obigen RDF Statements [4] in XML würde folgendermassen lauten:

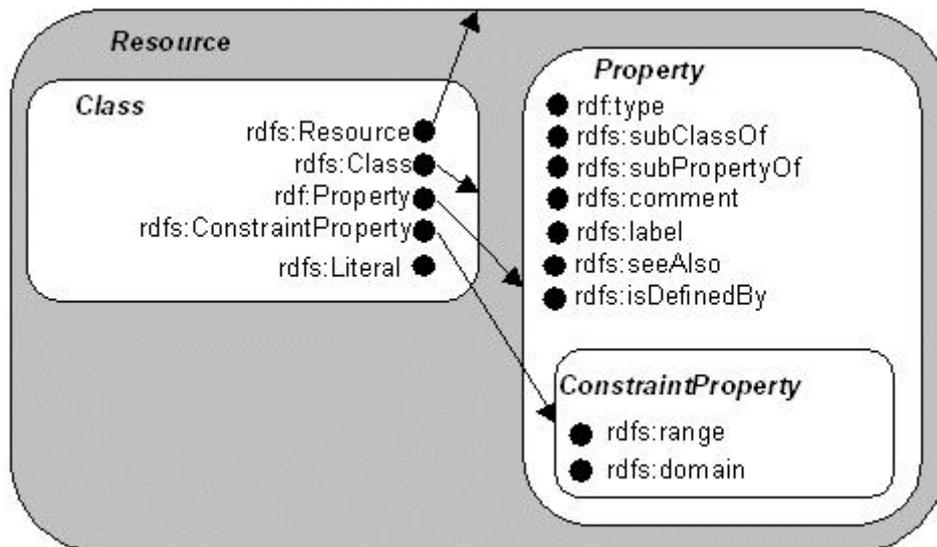
```
<?xml version="1.0" encoding="UTF-8"?>
<rdf:RDF xmlns:rdf="http://www.w3.org/2001/05/01-rdf-syntax-ns#"
  xmlns:dc="http://purl.org/metadata/dublin_core#"
  xmlns:ss="http://somesite.org/Schema">
  <rdf:Description about="http://www.ursodermatt.ch/home.html">
    <dc:Creator rdf:resource="mailto:urs.odermatt@access.unizh.ch">
      <dc:contains>
        rdf:resource="http://www.ursodermatt.ch/bild.jpg"
      </dc:contains>
    </dc:Creator>
    <rdf:Description about="mailto:urs.odermatt@access.unizh.ch">
      <ss:worksWith rdf:resource="mailto:gian.suter@gmx.net">
      </rdf:Description>
    </dc:Creator>
  </rdf:Description>
</rdf:RDF>
```

3.3.3. RDF Schema (RDFS)

RDF selbst bietet keine Möglichkeit, die Beziehungen zwischen Eigenschaften und Ressourcen zu definieren und gegebenenfalls Restriktionen aufzuerlegen. Um diese Beziehungen und Restriktionen zu deklarieren, werden RDF Schemas benutzt. Das heisst, es wird zum Beispiel definiert, welche Bedeutung „Creator“ hat und dass der Wert des Prädikates vom Typ „Person“ sein muss. Diese Schema-Vokabulare (einfache Ontologien⁸) definieren das gültige Prädikat in einer gegebenen RDF Beschreibung, sowie die Charakteristika des Prädikatwertes selbst.

⁸ siehe 4.

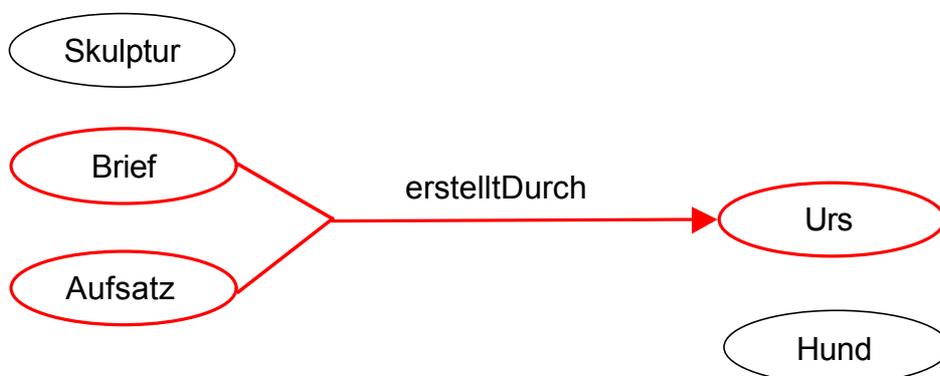
Das RDF Schema wird ausgedrückt durch das RDF Datenmodell. Der Aufbau des Schemas basiert auf einem Klassensystem und ist vergleichbar mit demjenigen einer objekt-orientierten Sprache wie zum Beispiel Java. RDF definiert aber, im Gegensatz zu den objekt-orientierten Programmiersprachen, nicht die Eigenschaften der Instanzen einer Klasse. Ein RDF Schema definiert Eigenschaften von Klassen einer Ressource zu welcher sie gehören.



Quelle: W3C [15]

Klassen werden als abgerundete Rechtecke dargestellt. Eine Ressource entspricht einem schwarzen Punkt. Die Pfeile definieren, welche Ressource zu welcher Klasse gehört, resp. welche Ressource von welcher Klasse definiert wird. Subklassen werden innerhalb von Klassen dargestellt. In unserem Beispiel wäre „ConstraintProperty“ eine Subklasse von „Property“.

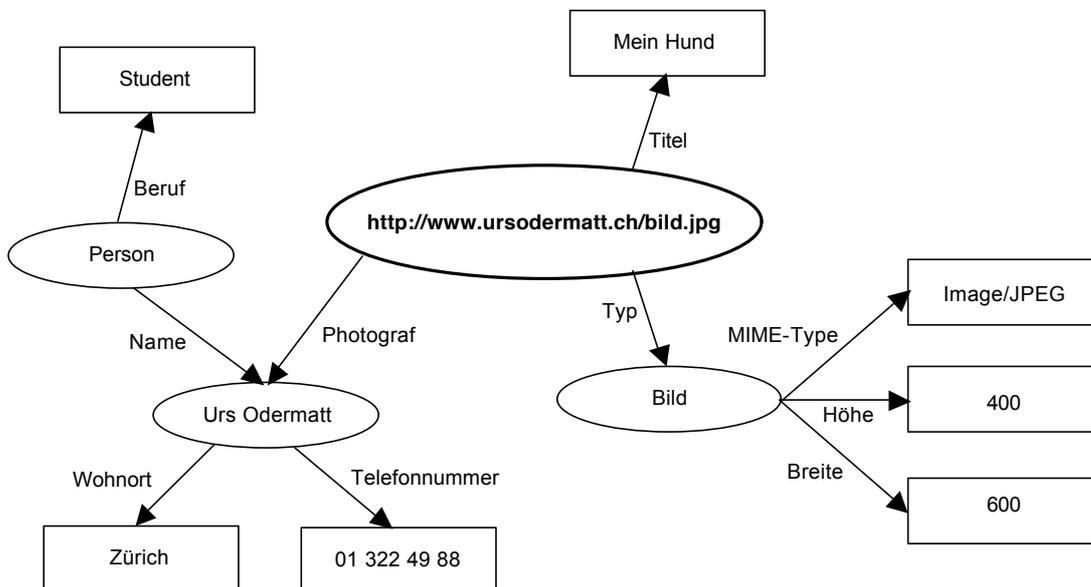
In diesem objektorientierten Klassensystem des RDF Schemas können unter anderem auch Einschränkungen definiert werden: Die Instanz `rdfs:range` der Klasse „ConstraintProperty“ wird eingesetzt, um die gültigen Werte der Eigenschaft einzugrenzen. Die Instanz `rdfs:domain` andererseits, schränkt die Menge der Ressourcen ein, die eine Eigenschaft haben kann.



Die entsprechende Syntax des vorhergehenden Beispiels würde folgendermassen lauten:

```
<rdf:Description ID="erstelltDurch">
  <rdf:type resource="http://www.w3.org/TR/WD-rdf-syntax#Property" />
  <rdfs:domain rdf:resource="#Document" />
  <rdfs:range rdf:resource="#Person" />
</rdf:Description>
```

Als abschliessende Betrachtung möchte ich die Möglichkeiten von RDF durch die Metadaten eines digitalen Bildes zumindest andeutungsweise aufzeigen:



3.4. Verbreitung von RDF Metadaten

Für die Verbreitung von RDF Metadaten bieten sich letztendlich drei Wege an. So können sie beispielsweise konventionell als Overhead in die Dokumentdatei eingeschrieben sein, gemeinsam mit dem Dokument selbst über einen Mechanismus innerhalb des HTTP-Protokolls übertragen werden, und schließlich können sie, ebenfalls über HTTP, von einem Drittanbieter bezogen werden, zum Beispiel einem Katalog oder einem weltweiten Index.

3.5. Abfrage und Interpretation von RDF Metadaten

RDF ermöglicht die standardisierte Darstellung von klar definiertem Wissen im Web. Doch nur die Präsentation von Wissen und Informationen reicht nicht aus: Benutzer sowohl Softwareagenten müssen die Daten abfragen und in verschiedener Weise benutzen können. Eine RDF Abfragesprache und entsprechende Werkzeuge werden benötigt.

Weiter sollte eine RDF Abfragemaschine einfach in die Webumgebung und in andere Softwarekomponenten integriert werden können. Eine RDF Abfrage ist keine Aufgabe für sich, sondern eine Teilaufgabe von weiterführenden Techniken und Methoden im World Wide Web.

Da die RDF Syntax selbst nicht die geeignete Basis für eine konkrete Repräsentation der Abfragesprache bildet, wäre es nützlich, wenn diese Anforderungen einer Abfragesprache in einer bereits existierenden, deklarativen Sprache (zuminderst teilweise) eingebettet werden könnten. Dies würde es uns erlauben, bereits existierende Maschinen weiter zu benutzen. Ein Beispiel einer solchen Sprache ist Frame-Logic (F-logic). [12]

Nachfolgend zeige ich Gemeinsamkeiten und Unterschiede zwischen RDF/RDFS und F-logic in zwei einfachen Beispielen auf:

Der Ersteller der Ressource

`http://www.w3.org/Home/Lassila" ist Ora Lassila.`

```
<rdf:RDF>
  <rdf:Description about="http://www.w3.org/Home/Lassila">
    <s:Creator>Ora Lassila</s:Creator>
  </rdf:Description>
</rdf:RDF>
```

Repräsentation in F-logic

`"http://www.w3.org/Home/Lassila"[Creator->>"Ora Lassila"]`

Das obige Beispiel zeigt eine der einfachsten RDF und F-logic-Äusserungen; diese sind gleichbedeutend: Eine beschriebene Ressource (Objekt) ist definiert durch ein Prädikat (Attribut) mit einem Wert („Ora Lassila“).

RDF Schema-Aussagen können auch in F-logic ausgedrückt werden. Das folgende Beispiel definiert zwei Klassen (Employee als ein Subklasse von Person und Researcher als Subklasse von Employee) und eine Eigenschaft cooperatesWith. Die Einschränkung Range und Domain dieser Eigenschaft ist bei beiden die Klasse Researcher.

```
<rdf:RDF
  <rdfs:Class rdfs:ID="Employee">
    <rdfs:subClassOf rdf:resource="http://ontology.org/human-ontology#Person"/>
  </rdfs:Class>
  <rdfs:Class rdfs:ID="Researcher">
    <rdfs:subClassOf rdf:resource="#Employee"/>
  </rdfs:Class>
  <rdf:Property ID="cooperatesWith">
    <rdfs:domain rdf:resource="#Researcher"/>
    <rdfs:range rdf:resource="#Researcher"/>
  </rdf:Property>
</rdf:RDF>
```

Repräsentation in F-logic:

```
Employee :: Person.
Researcher :: Employee[cooperatesWith=>>Researcher].
```

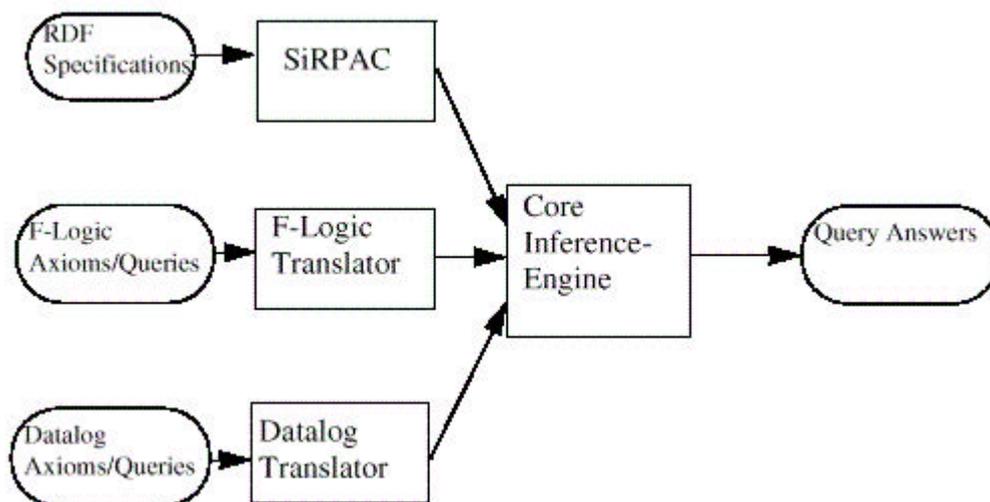
Dieses Beispiel zeigt auch einen der Unterschiede zwischen RDFS und F-logic: Wie in den meisten objektorientierten Sprachen definiert F-logic eine Klasse mit einem Attribut, welches

einen bestimmten Wert besitzt. Im Gegensatz dazu definiert RDFS eine Eigenschaft mit einer bestimmten Domain und Range, was aber kein Problem bei der Repräsentation in F-logic darstellt.

Ein Beispiel eines solchen RDF Interpreters in F-logic ist SiLRI (Simple Logic-based RDF Interpreter) [9]. Die Applikation ist in Java implementiert, bietet eine angemessene Performance und ist relativ klein in der Grösse. Die Applikation ist für fast alle Plattformen erhältlich und, noch viel wichtiger, einfach in andere Javaprogramme zu integrieren⁹.

Die Abfragemaschine benutzt SiRPAC¹⁰ um RDF Spezifikationen in Tripel zu übersetzen, welche mittels F-logic-Ausdrücke abgefragt werden können.

Die Architektur dieses Systems sieht folgendermassen aus:



Quelle: W3C [12]

3.6. Fazit

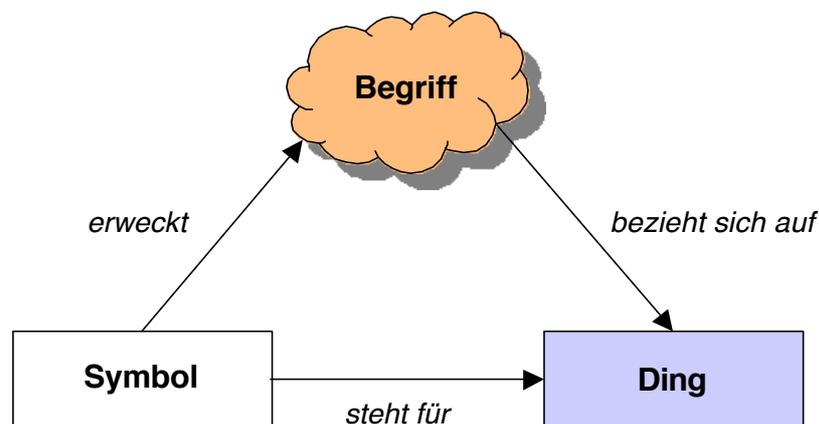
Als vorteilhaft an RDF mag eindeutig die Unabhängigkeit der Metainformation von den Inhalten eines Dokumentes gelten. Das Gegenteil von dem, was XML zum Vorteil gereicht hat und der Metasprache ihre Mächtigkeit einbringt, ist der grosse Pluspunkt einer reinen Metasektion, die bei, auf oder im Dokument allokiert ist. Die Metainformation ist strikt von den Inhalten des Dokumentes getrennt, was dazu führt, dass das Format keine Rolle mehr spielt. Hierdurch wird es möglich, auch veraltete oder brandneue Formatstandards zu verwenden und trotzdem die Dokumente für den Kreislauf der Dokumentenrecherche und Verarbeitung bereitzustellen.

⁹ Die Software kann gratis heruntergeladen werden: <http://ontobroker.semanticweb.org/silri/>

¹⁰ Simple RDF Parser & Compiler, <http://www.w3.org/RDF/Implementations/SiRPAC/>

4. Ontologien

Ontologien sind formale, semantische Modelle, die ebenfalls dazu dienen, den Austausch und das Teilen von Wissen zu erleichtern, insbesondere zwischen menschlichen und maschinellen Akteuren. Die Kommunikationssituation, in der sich die Akteure befinden, wird durch das semiotische Dreieck beschrieben. [11]



Das semiotische Dreieck illustriert die Interaktion zwischen Worten (oder allgemeiner: Symbolen), Begriffen und realen Dingen in der Welt. Worte, die benutzt werden, um Informationen zu übertragen, können die Essenz einer Referenz, das ist der Begriff oder das referenzierte Ding in der Welt, nicht vollständig erfassen. Dennoch gibt es eine Korrespondenz zwischen Wort, Begriff und Ding. Die Auswahl einer bestimmten Korrespondenz aus der Vielzahl a priori möglicher Korrespondenzen geschieht durch den Empfänger einer Nachricht. Hierbei benutzen verschiedene Empfänger unter Umständen verschiedene Begriffsbildungen und haben einen variierenden Erfahrungshintergrund, was wiederum zu verschiedenen Resultaten bezüglich der Korrespondenz zwischen einem Wort und den möglichen Begriffen und Dingen in der Welt führen kann. Ontologien erfordern die Einigung einer Gruppe von Anwendern auf die jeweiligen Begriffe und deren Zusammenhänge.

Eine Ontologie wird ausgedrückt durch eine logische Theorie, die sich zusammensetzt aus einem Vokabular und einer Menge von logischen Aussagen zu der jeweils interessierenden Anwendungsdomäne. Die logische Theorie spezifiziert Beziehungen zwischen Worten (Symbolen) und schränkt dabei die Menge der möglichen Interpretationen für Worte und ihren zugehörigen Beziehungen ein. Auf diese Weise reduziert eine Ontologie die Anzahl möglicher Korrespondenzen zwischen Worten und Dingen, die der Empfänger einer Nachricht, der sich auf eine Ontologie festgelegt hat, als gültig interpretieren kann. Idealerweise bleibt im Kontext von Kommunikationssituation und Ontologie für jedes Wort aus dem Vokabular genau eine Korrespondenz mit Begriffen und Dingen in der Welt übrig.

Ontologien unterstützen demnach auch die Vision eines semantischen Webs, in dem sie maschinellen Akteuren durch ihr formales Modell die Verarbeitung und Interpretation von Daten im Web ermöglichen.

5. Ausblick

Der Erfolg der Initiative für ein semantisches Web hängt bei allem Know-how, das in die Formulierung von Empfehlungen gesteckt wird, davon ab, dass Hersteller diese Vorgaben in konkrete Produkte umsetzen. Um der Implementierung der Standards auf die Sprünge zu helfen und ihre Praxistauglichkeit zu beweisen, will das W3C im Rahmen der Semantic Web-Aktivitäten entsprechende Open-Source-Projekte ins Leben rufen. [10]

Was können wir erwarten, wenn das Web der nächsten Generation reichlich mit Metadaten versehen sein wird? Es wird auf jeden Fall viel einfacher sein, etwas im Internet zu finden, da den Suchmaschinen viel mehr Informationen zur Verfügung stehen werden. Eine Suchmaschine wird in der Lage sein, relevante Dokumente von unwichtigen zu unterscheiden. Dies wird automatischen Softwareagenten neue Wege öffnen, das Web zu durchsuchen und uns die gewünschten Informationen zu präsentieren, sei das nun für E-Business-Anwendungen oder für unseren eigenen Bedarf. [1,7]

Das Web von heute, eine unstrukturierte Menge von Informationen, wird durch die Umwandlung in ein Semantic Web überschaubar werden. Kurz, das Web von morgen wird viel nützlicher sein.

Die Zukunft wird es zeigen...

6. Literaturverzeichnis

- [1] Berners-Lee, Tim (Sept. 1998), *Semantic Web Road map*:
<http://www.w3.org/DesignIssues/Semantic.html>
- [2] Berners-Lee, Tim (Sept. 1998), *What the Semantic Web can represent*:
<http://www.w3.org/DesignIssues/RDFnot.html>
- [3] Bray, Tim (24. Jan. 2001), *What is RDF?*:
<http://www.xml.com/pub/a/2001/01/24/rdf.html>
- [4] Champin, Pierre-Antoine (5. April 2001), *RDF Tutorial*:
<http://www710.univ-lyon1.fr/~champin/rdf-tutorial/>
- [5] Dumbill, Edd (1. Nov. 2000), *The Semantic Web: A Primer*.
<http://www.xml.com/pub/a/2000/11/01/semanticweb/index.html>
- [6] Dumbill, Edd (7. März 2001), *Building the Semantic Web*:
<http://www.xml.com/pub/a/2001/03/07/buildingsw.html>
- [7] Lassila, Ora (13. Nov. 1997), *Introduction to RDF Metadata*:
<http://www.w3.org/TR/NOTE-rdf-simple-intro>
- [8] Miller, Eric (Mai 1998), *An Introduction to the Resource Description Framework*:
<http://www.dlib.org/dlib/may98/miller/05miller.html>
- [9] Ontoprise GmbH, *SiLRI (Simple Logic-based RDF Interpreter)*:
http://www.ontoprise.de/co_silri.htm
- [10] Sommergut, Wolfgang (28. Feb. 2001), *Das WWW soll intelligenter werden*:
http://www.e-businessworld.com/english/crd_die_457452.html
- [11] Universität Karlsruhe: Institut AIFB:
<http://www.aifb.uni-karlsruhe.de/WBS/sst/Research/Publications/stichwort.pdf>
- [12] W3-Consortium, *A Query and Inference Service for RDF*:
<http://www.w3.org/TandS/QL/QL98/pp/queryservice.html>
- [13] W3-Consortium, *Semantic Web Activity: Resource Description Framework (RDF)*:
<http://www.w3.org/RDF/>
- [14] W3-Consortium, *Semantic Web Activity*:
<http://www.w3.org/2001/sw/>
- [15] W3C Candidate Recommendation (27. März 2000), *Resource Description Framework (RDF) Schema Specification 1.0*:
<http://www.w3.org/TR/2000/CR-rdf-schema-20000327/>
- [16] W3C Recommendation (22. Feb. 1999), *Resource Description Framework (RDF) Model and Syntax Specification*:
<http://www.w3.org/TR/1999/REC-rdf-syntax-19990222/>