Hochschule für Angewandte Wissenschaften Hamburg
Hamburg University of Applied Sciences

AW1 - Ausarbeitung

Julia Pressburger

semantische Informationssysteme

Abgabedatum: 25. Juli 2008
Prüfer: Prof. Dr. Neitzke
Zweitgutachter: Prof. Dr. von Luck
Fakultät Technik und Informatik
Department Informatik

Inhaltsverzeichnis

1 Motivation **4**

2 Semantic Web **5**

3 Ontologien **6**
3.1 Eigenschaften von Ontologien 7
3.2 Ebenen von Ontologien 8
3.3 Darstellungsformen 9

4 Zusammenfassung **11**

5 Ausblick **12**

A Literatur **13**

B Anhang **15**

Abbildungsverzeichnis

1 google - Screenshot 4
2 Freebase - Screenshot 5
3 Layercake des Semantic Web Miller, W3C 7
4 Das semiotische Dreieck, wiki:2008a 8
5 Beispielontology, wiki:2008c (2008) 9
6 semantische Treppe, Pellegrini und Blumauer (2006) 10
7 RDF Beispiel, grafisch, twine:2008 10

1 Motivation

Das Internet ist ein riesiger Pool von Informationen (in Text, Bildern und Film) und Dienstleistungen. Es gibt bestimmt nichts worüber nicht schon jemand geschrieben hat, oder von dem Jemand nicht schon ein Foto gemacht hat. Es gibt Dienste, die schon einzeln dem Einen oder Anderen eine Hilfe im Alltag sind. Meist muss der Nutzer genau wissen was er sucht und wo er es finden kann um die Informationen und Dienst zu nutzen. Er hangelt sich von Informationshäppchen zu Häppchen und kombiniert die Dienste die er nutzen will selbst und besitzt unzählige Accounts bei den Anbietern. Es gibt keine Kontrollinstanz die den Wahrheitsgehalt einer Information bestätigt. Man kann viel Zeit verlieren dabei sich ein umfassendes Bild zu einem Thema zu machen, kann an falsche Informationen geraten. Ist es möglich eine Informationsplattform zu schaffen die die Recherchearbeit abnimmt. Man vertraut ihr das Themengebiet an zu dem man sich informieren will und sie gibt einem einen umfassenden Pool an Informationen zurück. Die Informationen könnten aus unterschiedlichen Medien bestehen, nur Texte, auch subjektive Informationen (Blogeinträge), dazu noch Fotos oder vielleicht nur audiovisuele Medien und die Informationen wären auf ihren Vertrauensgehalt geprüft.

Die perfekte Informationsplattform

Bei der Eingabe von Lima in das Eingabefeld von google, bekommt man eine Ergebnisliste zurück, wie sie in der Abbildung 1 zu sehen ist. Wenn man sich über Lima, die Hauptstadt von Peru in Südamerika informieren wollte, ist man über die ersten beiden Elemente der Liste erfreut. Die restlichen Elemente ignoriert man.

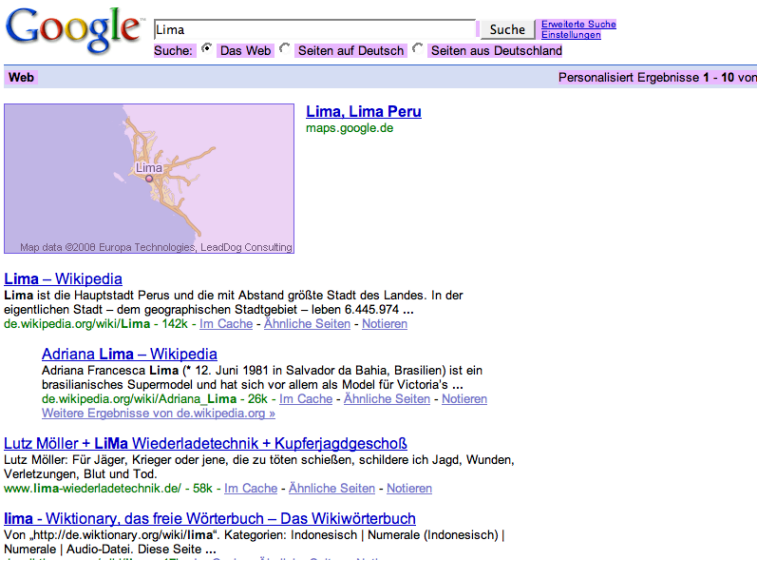

Abbildung 1: google - Screenshot

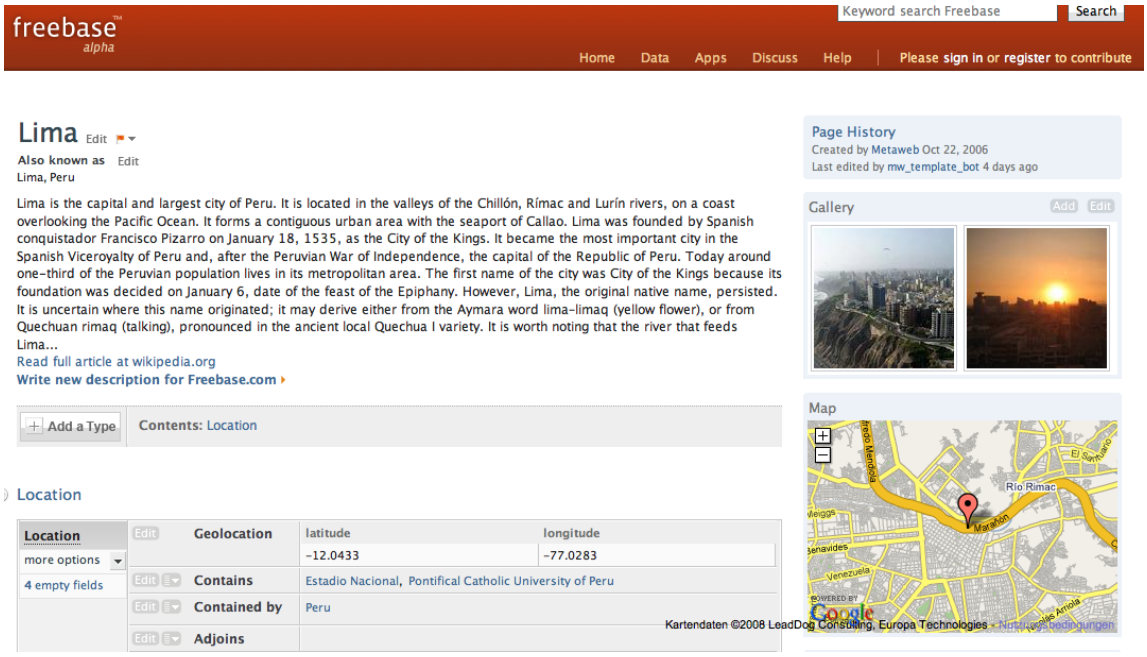

Abbildung 2: Freebase - Screenshot

Wenn man an die Plattform Freebase die Anfrage "Lima" stellt und dann in der erscheinenden Auswahl spezifiziert auf Lima, Stadt in Peru, bekommt man das in der Abbildung 2 zu sehende Ergebniss.

Die Idee

Bei der Vorstellung der AW1 Themen der Professoren gab es auch einen Vortrag über Semantik und Ontologien und Agenten. Als eine semantische Plattform, wurde Freebase vorgestellt. Es entstand die Frage ob es möglich wäre die Ordnung in der Plattform auch automatisch zu generieren oder ob Objekte ßich" automatisch in die schon vorhandene Ordnung einsortieren könnten? Um diese Frage beantworten zu können musste man sich zuerst mit einigen Grundlagen auseinandersetzten wie Semantik und Ontology. In den folgenden Kapitel gibt es eine Einführung.

2 Semantic Web

Das Semantic Web wird schon lange als die nächste Generation des Internets bezeichnet (vgl. Luck u. a. (2005))es hat jedoch immer noch einige Anlaufschwierigkeiten. Nach einem der grössten Beführworter des Semantic Web, Berners-Lee, wäre der Unterschied vom Web

wie wir es kennen, dem Dokumenten Web, zum Semantic Web, dass die Informationen des Internets nutzbar wären wie Objekte in einer Datenbank.

Def.: **Semantic Web *nach: Tim Berners-Lee*** I have a dream for the Web [in which computers] become capable of analyzing all the data on the Web - the content, links, and transactions between people and computers. A " Semantic Web" which should make this possible, has yet to emerge, but when it does, the day-to-day mechanisms of trade, bureaucracy and our daily lives will be handled by machines talking to machines. The " intelligent agents" people have touted for ages will finally materialize. wiki:2008b

Das Dokumenten Web muss maschinenlesbar werden, damit seine Informationen nicht nur für uns Menschen zugänglich sind. Lesbar für Maschinen, Agenten, werden die vorhanden Informationen umfangreicher nutzbar sein.

> "The Semantic Web is not a separate Web but an extension of the current one, in which information is given well-defined meaning, better enabling computers and people to work in cooperation."Tim Berners-Lee (2001)

Video - Vision des Semantic Web
Berners-Lee - Vision of the Semantic Web [1]

Um das Web maschinenlesbar zu machen ist noch viel Anstrengung nötig. Das Internet ist ein Fülle von Informationen und nicht alle sind vertrauensvoll (wahr). Die wenigsten Autoren nutzen die Methoden des Semantic Web. Das W3C übt grosse Anstrengungen, um Werkzeuge und Methoden für die Realisierung des Semantic Web zu standardisieren.

Die Abbildung 3 zeigt den Layer-Cake, veröffentlicht beim W3C. Ziel des W3C ist es, Ebene für Ebene die 7 Schichten des Semantic Web zu standardisieren. Für die Ebenen 1- 4 existieren schon Standards. Die 1. Ebene des Semantic Web bilden URI und Unicode, zu eindeutigen Adressierung und Codierung eines Objekts. Die 2. Ebene ist XML, einfache Strukturierung der Objekte. Als 3. Ebene RDF und RDFS für die formale Spezifikation des Objekts und seiner Eigenschaften. Die 4. Ebene enthält die Ontology als Darstellungsform für ganze Domänen. Die nächsten 3 Ebenen werden die Standards beinhalten die es ermöglichen sollen einer Aussage oder Schlussfolgerung als richtig, vertrauensvoll zu bewerten. Berners-Lee (2006)
Im folgenden Kapitel wird näher auf Ontologien eingegangen.

3 Ontologien

Ontologien sind ein wichtiger Bestandteil im Semantic Web (siehe Abbildung 3) dieses Kapitel gibt eine kurze Einführung in die Eigenschaften von Ontologien, ihre Darstellungsformen und ihre Bedeutung für das Semantic Web.

[1] Ein Interview mit Berners-Lee zur seiner Vision des Semantic Web. Berners-LeeMIT: Technology Review

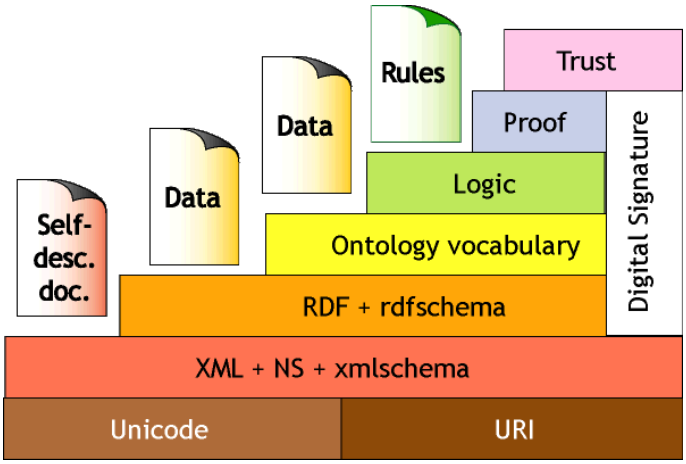

Abbildung 3: Layercake des Semantic Web Miller, W3C

Definitionen

Die Wissenschaft zur Beschreibung von Wissensdomänen ist viel älter als das Internet. Schon Aristoteles beschäftigte sich mit dem einsortieren seiner Umwelt in Kategorien.

Def.: **Ontologie nach Aristoteles:** The question of *what exists?* leads to studying general categories for all things that exist

Übertragen auf das Semantic Web ist eine Definition sein,

Def.: **Ontologie: im Zusammenhang mit Semantic Web** An *ontology* is a formal explicit specification of a shared conceptualisation of a domain of interest.

3.1 Eigenschaften von Ontologien

Grundlegende Karakteristiken von Ontologien sind:

- formal (Formality)
- explizit (Explicitness)
- verteilt (Being shared)
- konzeptioniert (conceptuality)
- Domänen spezifisch (Domain specific)

Studer u. a. (2007)
Ontologien bestehen aus folgenden Elementen:

- Klasse (concept)

- Beziehung (relation)
- Instanz (instance)

Studer u. a. (2007)

Die drei Elemente findet man schon bei Aristoteles, in der Abbildung 4 ist das semiotische Dreieck von Aristoteles dargestellt. Die Klasse entspricht dort dem Wort, die Instanz dem Ding und die Seelenregung der Beziehung, wobei die nicht ganz zu übertragen ist. Aristoteles soll von dem Ding ausgegangen sein, welches eine Seelenregung beim Betrachter oder dem Erzähler ausgelöst hat und das Wort ist das womit man das Ding beschreibt. Man könnte unterschiedliche Dinge mit dem selben Wort beschreiben (Eiche und Buche, sind beides Bäume).

Abbildung 4: Das semiotische Dreieck, wiki:2008a

3.2 Ebenen von Ontologien

Es gibt Ontologien für unterschiedliche Abstraktionsebenen:

Top-level ontologies beschreiben sehr allgemeine Konzepte wie Raum, Zeit, Materie, Ereignis, Aktion

Domain ontologies beschreiben das Vokabular für ein generischen Problembereich (z.B. Medizin), sie spezifizieren Konzepte der *Top-level ontology* weiter .

Task ontologies , beschreiben das Vokabular für eine generische Aufgabe oder Aktivität (z.B. Diagnostizieren)

Application ontologies beinhalten die Definitionen, die zum Modellieren des Wissens für eine bestimmte Anwendung notwendig sind. Sie erweitern und spezialisieren oft das Vokabular einer *Domain* oder *Task ontology*.

Durch die Benennung ist es leicht zu erkennen für welche Abstraktionsebene eine Ontology entworfen wurde. In vielen Wissendomänen gibt es bereits Ontologien oder es wird an ihnen gearbeitet. Ontologien sind komplex auch ihre Darstellungsform ist komplex. Werkzeugen für das Erstellen und Bearbeiten von Ontologien gelingt es nicht die Komplexität heraus zu nehmen. So muss man Spezialist für Ontologien sein um eine zu erstellen, doch die Herausforderung bei einer guten Domänen-Ontology ist, dass sie das Fachwissen einer Domäne repräsentiert. Die Kommunikation und Zusammenarbeit zwischen Ontology-Fachleuten und Domänen-Wissenschaftlern ist wichtig.

3.3 Darstellungsformen

Es gibt verschiedene Darstellungsformen für Ontologien, eine davon ist grafisch. Des weiteren haben sich mehrere Standards für die Beschreibung von Ontologien durchgesetzt. Auf einige wird im folgenden Abschnitt eingegangen, wie RDF und OWL die beide von W3C standardisiert sind. Beispiel für eine Ontologie

grafische Darstellungsformen

Dies (Abbildung 5) ist ein Beispiel für die grafische Darstellung einer Ontology die sich mit der Domäne Kunstwerk auseinander setzt. Die Knoten sind die Klassen und Instanzen, die Kanten sind die Relationen.

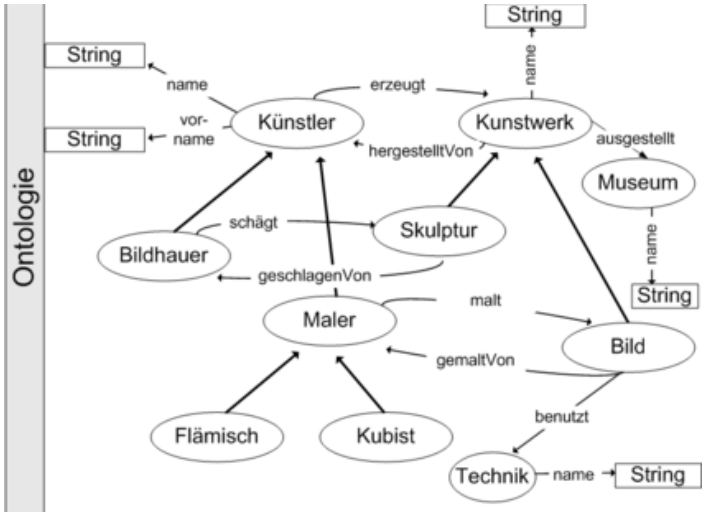

Abbildung 5: Beispielontology, wiki:2008c (2008)

Sprachen

Die Beschreibungssprachen die in diesem Abschnitt beschrieben werden haben unterschiedliche Ordnung. In Abbildung 6 ist die semantische Treppe dargestellt. Am oberen Ende der Treppe findet man Darstellungsformen mit der höchsten semantischen Ordnung, dort befinden sich die Ontologien. Als Sprachen findet wir dort OWL und RDF, wobei OWL eine höhere semantische Reichhaltigkeit aufweist.

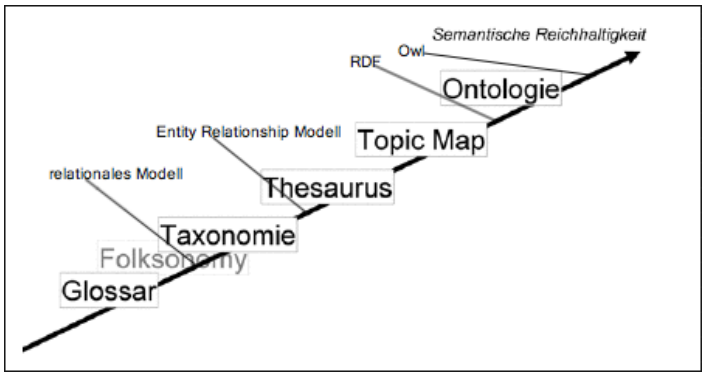

Abbildung 6: semantische Treppe, Pellegrini und Blumauer (2006)

RDF - Resource Description Language Die wichtigsten Fakten zu RDF kurz zusammen gefasst:

- RDF besteht aus so genannten Tupeln
- Subjekt, Prädikat und Objekt
- Erweiterung Schlüssel-Werte-Paar um das Prädikat, die Eigenschaft der Beziehung zwischen Key und Value

ein Beispiel: Buch –> hat einen –> Autor

RDF Modelle können grafisch oder im XML Format dargestellt werden. In Abbildung 7 sehen wir ein grafisch dargestelltes RDF Beispiel von der Seite twine.com (Im Anhang B ist das Beispiel noch als RDF im XML Format).

RDFS -RDF-Schema - RDF Vocabulary Description Language RDF-Schema wurde eingeführt um Gruppen ähnlicher Objekte mit gleichen Eigenschaften auszuzeichnen.

Am Beispiel eines Buches: Bücher haben alle einen Titel und einen Autor, nicht bei jedem Buch muss wieder definiert werden dass es diese Eigenschaften besitzt

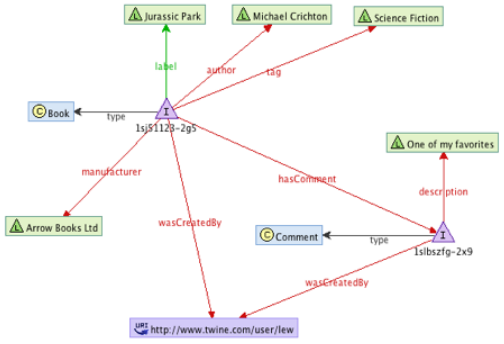

Abbildung 7: RDF Beispiel, grafisch, twine:2008

OWL- Web Ontology Language

OWL ist unterteilt in OWL Lite, OWL DL und OWL Full.

Full bietet einfache Prädikatenlogik inklusive deren Problematik nicht immer voll entscheidbar zu sein.

Lite ist perfekt für einfache Ontologien mit einfachen Taxonomien ohne komplexe Strukturen und Klassen.

DL enthält alle Sprachelemente von Full, bleibt aber voll entscheidbar.

Carsten Schulz (2008)

Warum denkt man über den Einsatz von komplexeren Sprachen als RDF nach, vor allem weil das Problem der Unentscheidbarkeit mit der Nutzung einiger Ausprägung von OWL auftritt?

RDF kann die folgende Zusammenhänge nicht darstellen:

Transitivität *ist größer als*

Eindeutigkeit *ist Mutter von*

Inversität *ist Elternteil von, ist Kind von*

4 Zusammenfassung

Ontologien und Semantik Web sind komplexe und grosse Themen. An der HAW hat es schon Arbeiten gegeben die sich mit Teilaspekten der Fragestellung dieser Ausarbeitung beschäftigt haben. Artem Khvat hat sich in seiner Arbeit "Integration der verschiedenartigen Services"mit den Schwierigkeit des Ontology Matching beschäftigt. Wann sind zwei Ontologien gleich, wie erkenne ich dass sie Aspekte der selben Domäne beschreibenKhvat (2006). Ebenfalls interessant ist die Arbeit von Andreas Christensen, Semantische Anreicherung von Suchanfragen auf Basis von Topic Maps, in der ein System entwickelt wurde mit der Suchanfragen an Hand von Topic Maps semantisch angereichert werdenChristensen (2005). Zur Frage nach der Herkunft der Ontologien die man für eine Plattform nutzen könnte, bei semanticweb.org gibt es eine Menge Ontologien aus vielen Wissensdomänen.

Zum Abschluss des Ausarbeitung ist klar, dass die Idee wie sie in der Motivation beschrieben wird seht komplex und umfassend ist. Sinnvoll wäre es sich in einem engen Rahmen mit klaren Anforderungen eine erste praktische Anwendung um zusetzten. Wie eine solche Anwendung aussehen könnte, darauf wird im Ausblick näher eingegangen.

5 Ausblick

Eine Anwendung im kleinen Rahmen könnte in der iFlat stattfinden. In der iFlat sollen unterschiedliche Geräte selbstständig miteinander kommunizieren um ihre kombinierten Dienste dem Menschen zur Verfügung zu stellen. Wie wird die Kommunikation ablaufen? Wenn zwei Geräte den selben Dienst anbieten (z. B. Anzeigen eine Information), wie wird entschieden welches Gerät die Aufgabe zugewiesen bekommt ? Als Middelware in der iFlat ist Jade (Java Agent Development Framework) vorgesehen. Die Agenten müssen anhand der Ihnen zur Verfügung stehenden Informationen ihr Verhalten aushandeln. Zu diesen Informationen gehören Kontext Informationen, welche in einem einheitlichen Format vorliegen müssen. Eine Standard für Kontext Informationen ist RDF/XMLWeindorf (2008).

Für das kommende Semester wäre es interessant das Thema "Context Awareness" in der iFlat mit Hilfe von Onthologien zu betrachten.

A Literatur

Literatur

[wiki:2008a] : *Grafik: semiotisches Dreieck*. – URL http://de.wikipedia.org/wiki/Bild:DasSemiotischeDreieckBeiAristoteles.png

[wiki:2008b] : *Semantic Web*. – URL http://en.wikipedia.org/wiki/Semantic_Web. – Zugriffsdatum: 24.07.2008

[twine:2008] : *Twine and RDF*. – URL http://www.twine.com/tour/semantic#rdf

[wiki:2008c 2008] : *Ontologie (Informatik)*. 2008. – URL http://de.wikipedia.org/wiki/Ontologie_(Informatik)

[Berners-Lee] BERNERS-LEE, Tim: *Video: Tim Berners-Lee on the Semantic Web*. – URL http://www.technologyreview.com/Infotech/18451/?a=f. – Zugriffsdatum: 27. März 2007

[Berners-Lee 2006] BERNERS-LEE, Tim: *Artificial Intelligence and the Semantic Web*. July 2006. – URL http://www.w3.org/2006/Talks/0718-aaai-tbl/

[Bojā andrs u. a. 2008] BOJĀ ANDRS, Uldis ; BRESLIN, John G. ; PERISTERAS, Vassilios ; TUMMARELLO, Giovanni ; DECKER, Stefan: Interlinking the Social Web with Semantics. In: *Intelligent Systems, IEEE* 23 (2008), May-June, Nr. 3, S. 29–40. – ISSN 1541-1672

[Carsten Schulz 2008] CARSTEN SCHULZ, Alexei Perel Andreas Krohn Axel Barth Jan-Philipp Schucher Jens Ellenberg Roman Ausländer Sören Voskul Thorsten Hillebrand Tobias B. ; NEITZKE, Prof. Dr. M. (Hrsg.): SS 2008 Semantic Wiki / HAW Hamburg. 2008. – Projektbericht

[Chandrasekaran u. a. 1999] CHANDRASEKARAN, B. ; JOSEPHSON, J.R. ; BENJAMINS, V.R.: What are ontologies, and why do we need them? In: *Intelligent Systems and Their Applications, IEEE [see also IEEE Intelligent Systems]* 14 (1999), Jan/Feb, Nr. 1, S. 20–26. – ISSN 1094-7167

[Christensen 2005] CHRISTENSEN, Andreas: *Semantische Anreicherung von Suchanfragen auf Basis von Topic Maps*, HAW Hamburg, Diplomarbeit, Juni 2005

[Khvat 2006] KHVAT, Artem: *Ontologiebasierte Integration der verschiedenartigen Services*, HAW Hamburg, Masterarbeit, 2006

[Luck u. a. 2005] LUCK, M. ; MCBURNEY, P. ; SHEHORY, O. ; WILLMOTT, S.: *Agent Technology: Computing as Interaction (A Roadmap for Agent Based Computing)*. AgentLink, 2005

[Miller] MILLER, Eric: *Digital Libraries and the Semantic Web*. – URL `http://www.w3.org/2001/09/06-ecdl/slide17-0.html`

[Pellegrini und Blumauer 2006] PELLEGRINI, Tassilo ; BLUMAUER, Andreas ; PELLEGRINI, Tassilo (Hrsg.) ; BLUMAUER, Andreas (Hrsg.): *Semantic Web: Wege zur vernetzten Wissensgesellschaft*. Springer, 2006. – ISBN 3540293248

[Studer u. a. 2007] STUDER, Rudi ; GRIMM, Stephan ; ABECKER, Andreas: *Semantic Web Services: Concepts, Technologies, and Applications*. Secaucus, NJ, USA : Springer-Verlag New York, Inc., 2007. – ISBN 3540708936

[Tim Berners-Lee 2001] TIM BERNERS-LEE, Ora L.: *The semantic web*. Mai 2001. – URL `http://www.sciam.com/article.cfm?id=the-semantic-web`

[Weindorf 2008] WEINDORF, Maik: *Ein Framework für Context-Awareness auf mobilen Geräten*, HAW Hamburg, Diplomarbeit, April 2008

B Anhang

RDF - Beispiel, Text

```
<rdf:RDF
    xmlns:basic="http://www.radarnetworks.com/2007/09/12/basic#"
    xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
    xmlns:s="http://www.w3.org/2000/01/rdf-schema#">
    <rdf:Description about="http://www.twine.com/item/1sj51123-2g5">
            <s:label>Jurassic Park</s:label>
            <basic:author>Michael Crichton</basic:author>
            <basic:manufacturer>Arrow Books Ltd</basic:manufacturer>
            <basic:tag>Science Fiction</basic:tag>
            <rdf:type rdf:resource="http://www.radarnetworks.com/2007/09/12/basic#Book"/>
            <radar:wasCreatedBy rdf:resource="http://www.twine.com/user/lew"/>
            <app:hasComment rdf:resource="http://www.twine.com/item/1slbszfg-2x9"/>
            </rdf:Description>
    <rdf:Description about="http://www.twine.com/item/1slbszfg-2x9">
            <rdf:type rdf:resource="http://www.radarnetworks.com/shazam#Comment"/>
            <basic:description>One of my favorites</basic:description>
            <radar:wasCreatedBy rdf:resource="http://www.twine.com/user/lew"/>
     </rdf:Description>
</rdf:RDF>
```

twine:2008