

BACHELORARBEIT

# Untersuchung von eingebettetem RDF in unternehmensbezogenen Webseiten

von  
Heiko Burkhardt

eingereicht am 10.04.2012 beim  
Institut für Angewandte Informatik  
und Formale Beschreibungsverfahren  
des Karlsruher Instituts für Technologie

Referent: Prof. Dr. Rudi Studer  
Betreuer: Dipl.-Inform. Benedikt Kämpgen

Heimatanschrift:  
Hofäcker 17/1  
75210 Keltern

Studienanschrift:  
Gerwigstr. 18  
76131 Karlsruhe

# Inhaltsverzeichnis

<b>1</b>	<b>Motivation und Ziele</b>	<b>1</b>
1.1	Motivation . . . . .	1
1.2	Ziele . . . . .	2
<b>2</b>	<b>Grundlagen</b>	<b>3</b>
2.1	Allgemeine Grundlagen des Semantic Web . . . . .	3
2.1.1	RDF . . . . .	3
2.1.2	RDFS . . . . .	4
2.1.3	OWL . . . . .	5
2.1.4	SPARQL . . . . .	5
2.2	Möglichkeiten der Einbettung von RDF in HTML . . . . .	5
2.2.1	RDFa . . . . .	5
2.2.1.1	Syntax . . . . .	6
2.2.1.2	Beispiele . . . . .	7
2.2.1.3	RDFa 1.1 . . . . .	8
2.2.1.4	RDFa Lite 1.1 . . . . .	9
2.2.2	Microformats . . . . .	9
2.2.3	Microdata . . . . .	10
2.2.4	Schema.org . . . . .	10
<b>3</b>	<b>Vergleich und Anwendungen semantischer Technologien</b>	<b>12</b>
3.1	Vergleich der Technologien . . . . .	12
3.1.1	Anzahl verwendeter Attribute . . . . .	12
3.1.2	URIs . . . . .	13
3.1.3	Namespaces . . . . .	13
3.1.4	Umandlung in RDF . . . . .	13
3.1.5	Standardisiert . . . . .	13
3.1.6	Datentypen für Literale . . . . .	13
3.1.7	Erweiterbarkeit . . . . .	14
3.1.8	DRY . . . . .	14
3.1.9	HTML Versionen . . . . .	14
3.2	Mögliche Anwendungen von semantischen Annotationen . . . . .	14
3.2.1	Suchmaschinen . . . . .	14
3.2.1.1	Google Rich Snippets . . . . .	14
3.2.1.2	Microsoft Bing . . . . .	15
3.2.2	Soziale Netzwerke . . . . .	16
<b>4</b>	<b>Analyse</b>	<b>17</b>
4.1	Allgemeine Analyse . . . . .	17
4.1.1	Crawl . . . . .	18
4.1.2	Verbreitung der einzelnen Technologien . . . . .	19
4.1.3	Prädikate . . . . .	20

---

4.1.4	Namespaces . . . . .	21
4.1.5	Klassenbezeichner . . . . .	22
4.1.6	Subjekte und Objekte . . . . .	24
4.2	Ontologien . . . . .	25
4.2.1	Open Graph Protokoll . . . . .	25
4.2.2	Dublin Core . . . . .	26
4.2.3	hCard . . . . .	26
4.2.4	XHTML Vocabulary . . . . .	27
4.2.5	Schema.org . . . . .	28
4.3	Rich Snippets Annotationen im Datensatz . . . . .	30
<b>5</b>	<b>Diskussion</b>	<b>32</b>
<b>6</b>	<b>Verwandte Arbeiten</b>	<b>34</b>
	<b>Literaturverzeichnis</b>	<b>36</b>



# 1. Motivation und Ziele

## 1.1 Motivation

Die Idee, strukturierte Information in Webseiten einzubetten ist bereits einige Jahre alt. In dieser Zeit wurden verschiedene Technologien entwickelt, welche diese Idee umsetzen. Die Technologie die in der Bachelorarbeit hauptsächlich betrachtet werden soll ist RDFa, da es sich bei RDFa im Gegensatz zu den anderen Technologien um einen etablierten W3C<sup>1</sup> Standard handelt.

Dabei geht es darum die Bedeutung des Inhalts einer Webseite Maschinen verständlich zu machen. Ein wichtiger Vorteil hiervon ist beispielsweise, dass es durch den Einsatz von semantischen Annotationen auf Websites Suchmaschinen möglich ist bessere Ergebnisse zu liefern.

Ein Beispiel für den erfolgreichen Einsatz von eingebetteten Semantic Web Technologien ist das im Verkauf von Unterhaltungselektronik tätige US-amerikanische Unternehmen BestBuy. BestBuy setzt RDFa auf seiner Webseite ein, um beispielsweise Öffnungszeiten und Adressen auszuzeichnen. Außerdem nutzt BestBuy die Good Relations Ontology. Hierbei handelt es sich um ein Vokabular mittels dem es möglich ist Informationen über Produkte semantisch zu annotieren.<sup>2</sup>

Der primäre Grund für den Einsatz von Semantic Web Technologien bei BestBuy ist, die Erhöhung der Sichtbarkeit. Diese erhöhte Sichtbarkeit drückt sich nicht in für den Nutzer sichtbaren Veränderungen der Webseite aus, sondern in den Ergebnissen von Suchmaschinen. So hat seit der Einführung von Semantic Web Technologien der durch Suchmaschinen verursachte Datenverkehr auf BestBuy.com um 30% zugenommen.<sup>3</sup>

Ein weiterer Beleg für die Relevanz von Semantic Web Technologien ist Schema.org. Bei Schema.org handelt es sich um einen Zusammenschluss der größten Suchmaschinenanbieter Google, Microsoft und Yahoo. Ziel dieses Zusammenschlusses ist es, den Entwicklern von Webseiten eine einheitliche Möglichkeit zu bieten, ihre Seite in, einer für die Suchmaschinen verständlichen Weise, semantisch auszeichnen zu können. Hierzu wird auf der Webseite Schema.org eine Sammlung von Schemata angeboten mit denen es möglich ist, die eigene Webseite mit semantischen Annotationen anzureichern.<sup>4</sup> Schema.org setzt hier-

---

<sup>1</sup>World Wide Web Consortium, <http://www.w3.org/>

<sup>2</sup>[http://www.heppnetz.de/projects/goodrelations/primer/#1.2\\_What\\_Does\\_GoodRelations\\_Contribute.3F](http://www.heppnetz.de/projects/goodrelations/primer/#1.2_What_Does_GoodRelations_Contribute.3F)

<sup>3</sup>[http://www.readwriteweb.com/archives/how\\_best\\_buy\\_is\\_using\\_the\\_semantic\\_web.php](http://www.readwriteweb.com/archives/how_best_buy_is_using_the_semantic_web.php)

<sup>4</sup><http://schema.org>

bei jedoch nicht auf RDFa sondern auf Microdata.<sup>5</sup>

Die Frage ist, warum man eigentlich RDF-Tripel in Webseiten einbettet und nicht einfach ein separates RDF/XML-Dokument anbietet. Dies liegt unter anderem daran, dass dieses Vorgehen eine große Menge an redundanten Daten hervorbringen würde. Diese Redundanz würde die Wartbarkeit einschränken.[ABMP08] Außerdem ermöglichen semantische Annotationen in HTML eine direkte Zuordnung zwischen RDF-Tripeln und HTML-Code. Diese Zuordnung wäre bei einer getrennten Veröffentlichung von RDF/XML-Dokumenten und HTML-Dokumenten nicht gegeben. Besonders für Suchmaschinen ist dies hilfreich, da diese nun "verstehen" um was es sich bei dem semantisch annotierten HTML-Code handelt.

Dass Webseiten durch die Nutzung von semantischen Annotationen profitieren können, wird auch in [STH10] deutlich. Hier wird gezeigt wie Suchmaschinen, insbesondere Google, die semantischen Annotationen auf Webseiten nutzen. Diese Annotationen werden nicht nur genutzt um bessere Suchergebnisse zu erzielen, sondern ermöglichen es auch den Suchmaschinen ihren Nutzern direkt bessere Informationen zu den Suchergebnissen zu geben. Sucht ein Nutzer zum Beispiel nach einer Band, so können direkt auf der Seite der Suchmaschine Konzerte dieser Band angezeigt werden.

## 1.2 Ziele

In dieser Bachelorarbeit sollten unter anderem die Technologien zur semantischen Anreicherung von HTML-Seiten untersucht werden. So soll beispielsweise der Frage nachgegangen werden, wie weit verbreitet die einzelnen Technologien sind und ob es Anzeichen dafür gibt, dass sich eine der Technologien durchsetzen kann. Außerdem soll analysiert werden, wie groß der Unterschied zwischen dem in den Annotationen enthaltenem Wissen und dem Wissen, welches ein Mensch durch das Betrachten des Dokuments erlangen kann, ist. Um diesen Unterschied herauszuarbeiten werden die semantischen Annotationen von den Webseiten extrahiert, in einen Triple Store geladen und mittels SPARQL-Abfragen analysiert. Bei dieser Analyse wird darauf geachtet, welche Vokabulare genutzt werden, in welchem Umfang die Vokabulare genutzt werden und wie viele Tripel auf den einzelnen Seiten eingebettet sind.

---

<sup>5</sup>[http://schema.org/docs/gs.html#microdata\\_why](http://schema.org/docs/gs.html#microdata_why)

## 2. Grundlagen

### 2.1 Allgemeine Grundlagen des Semantic Web

Die Idee für ein Semantic Web stammt von Tim Berners-Lee, dem Begründer des World Wide Web. In seinem Artikel [BLHL01] erläutert er diese Idee. Hierbei stellt sich Berners-Lee das Semantic Web als eine Erweiterung des bestehenden Web vor. Bei dieser Erweiterung soll es darum gehen, den bisher nur für den Menschen verständlichen Inhalt eines Dokuments auch für Computer interpretierbar zu machen. Um dies zu erreichen werden die Informationen in einer maschinenlesbaren Form publiziert.

Eine wichtige Rolle spielen hierbei die Uniform Resource Identifier (URI), mit deren Hilfe die einzelnen Ressourcen identifiziert werden können. Bei URIs handelt es sich um eine weltweit eindeutige Zeichenkette.[BLFM05]

Außerdem benötigt man Ontologien um komplexe Wissensbeziehungen modellieren zu können. Laut [HKRS08, S. 12] versteht man im Bereich des Semantic Web unter einer Ontologie ein Dokument in dem Wissen für ein bestimmte Anwendungsdomäne modelliert wird. Zum Beispiel ist es mit der Friend of a Friend (FOAF) Ontologie möglich Personen und Organisationen zu beschreiben.

Im folgenden Abschnitt werden Standards vorgestellt welche im Semantic Web Verwendung finden.

#### 2.1.1 RDF

Bei dem Resource Description Framework (RDF) handelt es sich um eine vom W3C standardisierte Sprache, welche Beziehungen zwischen Ressourcen im World Wide Web modelliert. Diese Tripel bestehen aus einem Subjekt, einem Prädikat und einem Objekt. Eine Menge dieser Tripel bezeichnet man als RDF-Graph.[MM04]

Bei einer grafischen Repräsentation eines RDF-Graphen werden das Subjekt und das Objekt als Knoten und das Prädikat als gerichtete Kante dargestellt.[HKRS08, S. 36]

Bei einem Prädikat handelt es sich immer um eine Ressource, welche mittels einer URI identifiziert wird. Das Subjekt ist entweder eine Ressource, die auch durch eine URI identifiziert wird, oder ein "Blank Node". Das Objekt kann auch eine Ressource sein oder eine Zeichenkette, welche man als Literal bezeichnet.[HKRS08, S. 37-39, S. 57]

Eine einfache Darstellungsform für RDF-Graphen ist die Terse RDF Triple Language (Turtle). Bei Turtle handelt es sich um eine vereinfachte Form der RDF-Serialisierung N3, welche unter anderem von Tim Berners-Lee entwickelt wurde.[BBLP11] Laut W3C

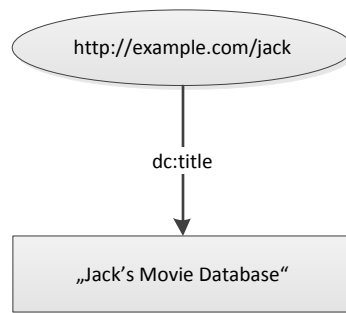


Abbildung 2.1: RDF-Graph Visualisierung

ist jedoch RDF/XML die Standardsyntax für RDF.[MM04] Wie der Name andeutet ist RDF/XML eine RDF-Serialisierung in der Form eines XML-Dokuments.

Folgendes Beispiel zeigt einen Graphen in der Turtle-Syntax:

```

@prefix ex: <http://example.org/movie/> .
@prefix dc: <http://purl.org/dc/elements/1.1/> .

ex:StarWars      dc:title      "Star Wars" .
ex:StarWars      ex:DirectedBy  ex:GeorgeLucas .
ex:Inception     dc:title      "Inception" .
ex:Inception     ex:DirectedBy  ex:ChristopherNolan .
  
```

Das gleiche Beispiel nun in RDF/XML:

```

<?xml version="1.0"?>
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:ex="http://example.org/movie/"
  xmlns:dc="http://purl.org/dc/elements/1.1/">
  <rdf:Description
    rdf:about="http://example.org/movie/StarWars">
    <dc:title>Star Wars</dc:title>
    <ex:DirectedBy
      rdf:resource="http://example.org/movie/GeorgeLucas"/>
    </rdf:Description>
  <rdf:Description
    rdf:about="http://example.org/movie/Inception">
    <dc:title>Inception</dc:title>
    <ex:DirectedBy
      rdf:resource="http://example.org/movie/ChristopherNolan"/>
    </rdf:Description>
</rdf:RDF>
  
```

### 2.1.2 RDFS

RDF Schema (RDFS) ist ein RDF-Vokabular, welches auch Bestandteil der RDF Recommendation des W3C ist. Mit Hilfe von RDFS können einfache Ontologien modelliert werden. RDFS stellt hierbei kein auf eine Anwendungsdomäne spezialisiertes Vokabular dar, sondern ermöglicht es Klassen, Properties und deren Beziehung zueinander zu beschreiben.[GB04]



So ist es mit Hilfe von RDFS zum Beispiel möglich auszudrücken, dass ein Western eine Unterklasse der Klasse Film ist. Hierzu muss zunächst definiert werden, dass sowohl Western als auch Film vom Typ Class sind.

ex:Film	rdf:type	rdfs:Class
ex:Western	rdf:type	rdfs:Class

Nun kann man definieren, dass ein Western eine Unterklasse der Klasse Film ist.

ex:Western	rdfs:subClassOf	ex:Film
------------	-----------------	---------

### 2.1.3 OWL

Auch bei der Web Ontology Language (OWL) handelt es sich um eine W3C Recommendation. Durch OWL können ebenfalls Ontologien beschrieben werden, jedoch ist OWL eine ausdrucksstärkere Sprache als RDFS. Eine OWL-Ontologie kann auch als RDF-Graph dargestellt werden. Daher ist es auch möglich eine OWL-Ontologie in RDF/XML zu formulieren. Laut W3C Recommendation ist RDF/XML sogar die einzige Syntax, die von OWL Tools unterstützt werden muss.[KPSR<sup>+</sup>09]

### 2.1.4 SPARQL

SPARQL ist eine Abfragesprache für RDF. Als Ergebnis einer Abfrage wird ein RDF-Graph oder ein Teil davon zurückgegeben. Eine SPARQL-Abfrage besteht meist aus Tripeln welche denen eines RDF-Graphen ähneln, mit dem Unterschied, dass die SPARQL-Tripel Variablen enthalten können.[PS08]

Folgende Abfrage soll die Namen aller Filme und die dazugehörigen Regisseure, die in einem RDF-Graphen enthalten sind, zurückgeben.

```

PREFIX ex: <http://example.org/movie/>
PREFIX dc: <http://purl.org/dc/elements/1.1/>
SELECT ?name ?director
WHERE {
    ?x      dc:title      ?name .
    ?x      ex:DirectedBy ?director .
}

```

Das Ergebnis einer solchen Abfrage könnte dann folgendermaßen aussehen:

name	director
Star Wars	George Lucas
Inception	Christopher Nolan

## 2.2 Möglichkeiten der Einbettung von RDF in HTML

### 2.2.1 RDFa

Bei RDFa handelt es sich um eine W3C Recommendation, mit der es möglich ist RDF-Tripel in XHTML einzubetten. Ein erster Entwurf für RDFa [Bir04] wurde im Jahr 2004 von Mark Birbeck veröffentlicht. 2008 erlangte RDFa dann den W3C Recommendation Status.[ABMP08]

### 2.2.1.1 Syntax

Die W3C Recommendation geht nur auf die Verwendung von RDFa im Zusammenhang mit XHTML ein, da XHTML laut Spezifikation erweiterbar ist. Dies ist bei HTML4 nicht der Fall. Jedoch werden die meisten RDFa Parser auch RDFa-Annotationen in HTML erkennen und verarbeiten können, womit der Einsatz von RDFa auch in HTML möglich ist.[AB08]

Attribut	Beschreibung	RDF-Entsprechung
<b>rel</b>	wird verwendet um die Beziehung zwischen zwei Ressourcen zu beschreiben	predicates
<b>rev</b>	wird verwendet um eine invertierte Beziehung zwischen zwei Ressourcen zu beschreiben	predicates
<b>content</b>	ein String, welcher maschinenlesbaren Inhalt für ein Literal bietet	plain literal object
<b>href</b>	eine URI, welche die Partnerressource einer Beziehung wiedergibt	resource object
<b>src</b>	eine URI, welche die Partnerressource einer Beziehung wiedergibt, falls die Ressource eingebettet ist	resource object

Tabelle 2.1: Bestehende XHTML-Attribute welche bei RDFa Verwendung finden

Für die Einbettung der Tripel in Webseiten werden bestehende und neue XHTML-Attribute verwendet. Tabelle 2.1 bietet einen Überblick über die bestehenden und Tabelle 2.2 über die neuen XHTML-Attribute, welche bei RDFa Verwendung finden.[ABMP08]

Attribut	Beschreibung	RDF-Entsprechung
<b>about</b>	gibt an um was es sich bei den Daten handelt	subject
<b>property</b>	wird verwendet um eine Beziehung zwischen einem Subjekt und einem Literal zu beschreiben	predicates
<b>resource</b>	eine URI, welche die Partnerressource einer Beziehung wiedergibt, auf die nicht "geklickt" werden kann	object
<b>datatype</b>	stellt ein Datentyp eines Literals dar	
<b>typeof</b>	deklariert ein neues Element von einem bestimmten Typ	

Tabelle 2.2: Neue XHTML-Attribute welche bei RDFa Verwendung finden

### 2.2.1.2 Beispiele

In folgendem Beispiel wird mittels RDFa ausgedrückt, dass die Webseite `http://example.com/jack` den Titel "Jack's Movie Database" hat. Hierzu wird das Attribut `property` verwendet, welches eine Beziehung (hier: `dc:title`) zwischen einem Subjekt (hier: `http://example.com/jack`) und einem Literal (hier: "Jack's Movie Database") beschreibt. Auf die gleiche Weise wird ausgedrückt, dass Jack Miller der Inhaber der Rechte an `http://example.com/jack` ist.

Bei `dc:title` handelt es sich um eine Kurzschreibweise. Diese Kurzschreibweise wird mittels XML-Namensräumen realisiert. Hierfür wird durch das Attribut `xmlns` einem Präfix (hier: `dc`) ein Namensraum (hier: `http://purl.org/dc/elements/1.1/`) zugewiesen.

```
<div xmlns:dc="http://purl.org/dc/elements/1.1/">
  <h1 property="dc:title">Jack's Movie Database</h1>
  <h2 property="dc:rightsHolder">Jack Miller</h1>
  ...
</div>
```

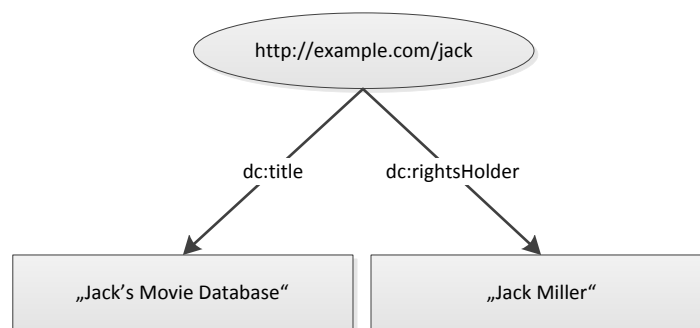


Abbildung 2.2: RDFa Beispiel 1

Hat man nun mehrere Subjekte auf einer Seite, welche beschrieben werden sollen, kann man dies mittels des Attributs `about` erreichen. Dieses Attribut gibt die URI des Subjekts an. In Beispiel 2 wurden somit die Ressourcen mit den URIs `http://example.com/jack/movie/starwars` und `http://example.com/jack/movie/inception` als Subjekt definiert. Hierbei ist zu beachten, dass das innere Subjekt das äußere überschreibt.

```
<div xmlns:dc="http://purl.org/dc/elements/1.1/">
  ...
  <div about="http://example.com/jack/movie/starwars">
    <h2 property="dc:title">Star Wars</h1>
  </div>
  ...
  <div about="http://example.com/jack/movie/inception">
    <h2 property="dc:title">Inception</h1>
  </div>
  ...
</div>
```

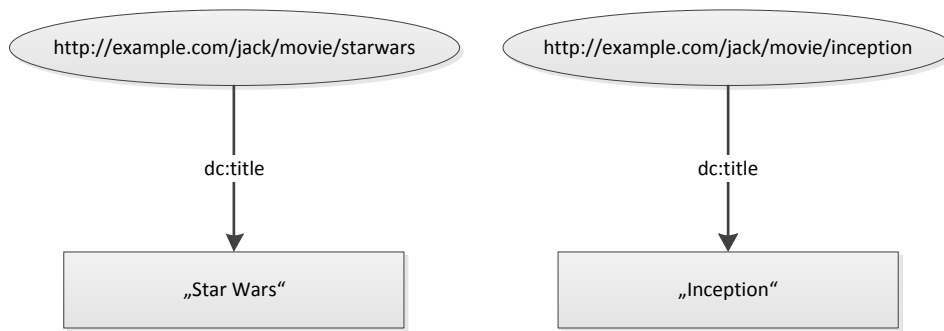


Abbildung 2.3: RDFa Beispiel 2

Durch das Attribut `typeof` ist es möglich den Typ eines neuen Elements zu definieren. In folgendem Beispiel wird definiert, dass es einen neuen Knoten gibt, welcher vom Typ `foaf:person` ist. Jedoch wird für diesen Knoten keine URI spezifiziert, etwa durch `about`. Daher handelt es sich bei dem Knoten um einen "blank node".

```
<div about="http://example.com/jack/movie/starwars/"
  rel="http://example.com/elements/actors/">
  ...
  <div xmlns:foaf="http://xmlns.com/foaf/0.1/"
    typeof="foaf:person" >
    <a property="foaf:name" rel="foaf:mbox"
      href="mailto:info@harrison-ford.com" >
      Harrison Ford
    </a>
  </div>
  ...
</div>
```

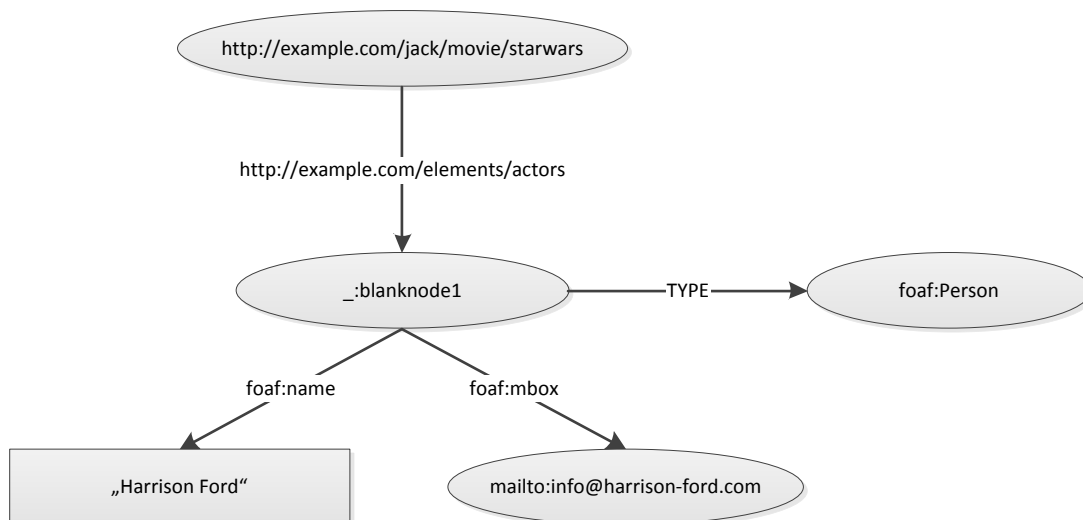


Abbildung 2.4: RDFa Beispiel 3

### 2.2.1.3 RDFa 1.1

Im Dezember 2011 hat die RDF Web Applications Working Group ein Working Draft für RDFa 1.1 veröffentlicht.<sup>1</sup>Den Status "Candidate Recommendation" erlangte RDFa 1.1

<sup>1</sup><http://www.w3.org/News/2011#entry-9288>

dann im März 2012.[ABMH12] Laut Ivan Herman, dem Semantic Web Activity Lead beim W3C, gibt es bei RDFa 1.1 drei erwähnenswerte Änderungen im Vergleich zu RDFa 1.0.<sup>2</sup> So wurde beispielsweise die Profile-Funktion, auf Grund von Problemen bei der Implementierung, aus dem Standard entfernt. Mit der Profile-Funktion war es möglich die Zuordnung zwischen URIs und einem Ausdruck oder einem Präfix in ein separates Profile-Dokument auszulagern. Des Weiteren unterstützt RDFa 1.1 mit dem Attribut `inlist` nun geordnete Listen. Außerdem wurde die Bedeutung des `src` Attributs geändert. Ursprünglich verhielt sich das `src` Attribut wie `about`, nun soll es jedoch wie das `resource` Attribut behandelt werden.

#### 2.2.1.4 RDFa Lite 1.1

RDFa verfügt über einen relativ großen Funktionsumfang. Die aus dem Funktionsumfang resultierende Komplexität stellt jedoch eine Hürde für Anfänger dar. Aus diesem Grund wurde von dem W3C eine Spezifikation [Spo12] für RDFa Lite 1.1 veröffentlicht. RDFa Lite besteht nur aus den Attributen `vocab`, `typeof`, `property`, `about` und `prefix`. Diese abgespeckte RDFa Version soll laut dem W3C für 80% der Webentwickler ausreichende Möglichkeiten zur semantischen Annotation ihrer Webseite bieten.<sup>3</sup>

### 2.2.2 Microformats

Bei Microformats handelt es sich um einen Ansatz mit dem es ebenfalls möglich ist, Webseiten semantisch zu annotieren. Für die Annotation der Webseiten wird hierbei hauptsächlich das Attribut `class` verwendet.[All07, S. 317] Außerdem werden noch die Attribute `title`, `rel` und `href` verwendet. In folgendem Beispiel werden mittels hCard Kontaktdaten semantisch annotiert, was zum Beispiel den automatischen Import der Daten in ein Kontaktverzeichnis ermöglicht.

```
<div class="vcard">
  <div class="fn">John Miller</div>
  <div class="org">Miller Ltd.</div>
  <a class="url" href="http://miller.example.com/">
    http://miller.example.com/
  </a>
</div>
```

In Tabelle 2.3 werden die laut microformats.org stabilen Microformats aufgelistet.<sup>4</sup>

<sup>2</sup><http://www.w3.org/blog/SW/2011/09/19/recent-changes-in-rdfa-1-1/>

<sup>3</sup><http://www.w3.org/News/2011#entry-9288>

<sup>4</sup>[http://microformats.org/wiki/Main\\_Page#Specifications](http://microformats.org/wiki/Main_Page#Specifications)

Bezeichnung	Beschreibung
hCalendar	ist eine Repräsentation des iCalendar Standards und ermöglicht die semantische Auszeichnung von Events
hCard	ist eine Repräsentation des vCard Standards und ermöglicht die semantische Auszeichnung von Kontaktdaten
rel-license	gibt die Lizenz an, unter der, die Webseite steht
rel-nofollow	stellt eine Empfehlung für Suchmaschinencrawler dar diesem Link nicht zu folgen
rel-tag	ordnet dem Zieldokument ein Schlüsselwort zu
VoteLinks	bewertet das Zieldokument positiv oder negativ
XFN	wird verwendet um Beziehungen zwischen Menschen durch Hyperlinks zu modellieren
XMDP	definiert ein Metadata Profile

Tabelle 2.3: Stabile Microformats

### 2.2.3 Microdata

Microdata ist ein Teil des noch in der Entwicklung befindlichem HTML5 Standards. Mittels Microdata ist es möglich in ein HTML-Dokument eine Gruppe von Name/Wert-Paaren einzubetten. Eine Gruppe von Name/Wert-Paaren bezeichnet man als Item. Die einzelnen Name/Wert-Paare stellen eine Property dar.[Hic12]

Tabelle 2.4 gibt eine Übersicht über die Microdata-Attribute.

Attribut	Beschreibung
itemscope	erstellt ein neues Item
itemtype	weist einem Item ein Typ zu
itemid	weist einem Item eine URI zu
itemprop	fügt einem Item ein Property hinzu
itemref	weist ein Property einem Item zu

Tabelle 2.4: Microdata Attribute

Im folgenden Beispiel wird ein Element vom Typ `http://data-vocabulary.org/Person` erzeugt, welches die Eigenschaft "name" mit dem Wert "George Lucas" hat. Außerdem hat das Element noch die Eigenschaft "role" und dem Wert "director".

```
<div itemscope itemtype="http://data-vocabulary.org/Person">
  <h1 itemprop="name">George Lucas</h1>
  Beruf: <span itemprop="role"> director </span>
</div>
```

### 2.2.4 Schema.org

Bei schema.org handelt es sich um einen Zusammenschluss der größten Suchmaschinenanbieter Google, Microsoft und Yahoo. Seit November 2011 wird schema.org auch von dem russischen Suchmaschinenanbieter Yandex unterstützt. Diese Unterstützung drückt sich dadurch aus, dass Yandex Schema.org-Annotationen berücksichtigen wird. Außerdem wird Yandex bei der Weiterentwicklung von schema.org mitwirken.<sup>5</sup>

<sup>5</sup><http://blog.schema.org/2011/11/yandex-now-supports-schemaorg-markup.html>

Ziel dieses Zusammenschlusses ist es, den Entwicklern von Webseiten eine einheitliche Möglichkeit zu bieten, ihre Seite in einer für die Suchmaschinen verständlichen Weise semantisch auszeichnen zu können. Hierzu wird auf der Webseite Schema.org eine Sammlung von Schemata angeboten, mit der es möglich ist die eigene Webseite mit semantischen Annotationen anzureichern.<sup>6</sup>

Als Technologie zum annotieren der Webseiten setzt Schema.org auf Microdata.<sup>7</sup> Microformats wurden abgelehnt, da sie nicht leicht zu erweitern sind und es aufgrund der Verwendung des Class Tags zu Konflikten mit CSS kommen kann. Gegen RDFa sprach nach Ansicht der Suchmaschinenanbieter, dass es sich hierbei um eine zu komplexe Technologie handelt.<sup>8</sup> Jedoch wurde auf dem offiziellen Blog von Schema.org angekündigt, dass neben Microdata auch RDFa Lite als eine mögliche Syntax für die Einbettung von Schema.org-Annotationen unterstützt wird.<sup>9</sup>

Tabelle 2.5 zeigt einen Auszug aus der Typ-Hierarchie von schema.org.<sup>10</sup>

1. Ebene	2. Ebene	Beispiele 3. Ebene
Thing	CreativeWork	Article, Book, Movie, Photograph
	Event	BusinessEvent, Festival, SportsEvent
	Intangible	JobPosting, Offer, Quantity, Rating
	Organization	Corporation, LocalBusiness, NGO
	Person	
	Place	AdministrativeArea, Landform, TouristAttraction
	Product	

Tabelle 2.5: Auszug aus der Typ-Hierarchie

Im nachfolgenden Beispiel wird mittels Microdata und Schema.org ein Film beschrieben.

```
<div itemscope itemtype="http://schema.org/Movie">
  <h1 itemprop="name">Star Wars</h1>
  <div itemprop="director" itemscope
    itemtype="http://schema.org/Person">
    <h2 itemprop="name">George Lucas</h2>
  </div>
</div>
```

<sup>6</sup><http://schema.org>

<sup>7</sup>[http://schema.org/docs/gs.html#microdata\\_why](http://schema.org/docs/gs.html#microdata_why)

<sup>8</sup><http://schema.org/docs/faq.html#14>

<sup>9</sup><http://blog.schema.org/2011/11/using-rdfa-11-lite-with-schemaorg.html>

<sup>10</sup><http://schema.org/docs/full.html>

## 3. Vergleich und Anwendungen semantischer Technologien

In diesem Kapitel findet die theoretische Betrachtung der Technologien zur semantischen Anreicherungen statt. Außerdem wird hier auf mögliche Anwendungen von in HTML-Seiten eingebetteten RDF-Tripeln eingegangen.

### 3.1 Vergleich der Technologien

In diesem Abschnitt werden die Möglichkeiten, zur semantischen Anreicherung von HTML, verglichen. Dies geschieht anhand einiger Kriterien aus [Gra07]. Tabelle 3.6 gibt einen ersten Überblick über die Unterschiede der einzelnen Technologien. Im Folgendem wird dann genauer auf die Unterschiede eingegangen.

	Schema.org	Microdata	RDFa 1.1	Microformats
<b>Anzahl verwendeter Attribute</b>	10	10	13	4
<b>URIs</b>	Ja	Ja	Ja	Nein
<b>Namespaces</b>	Nein	Nein	Ja	Nein
<b>Umwandlung in RDF</b>	Ja	Eingeschränkt möglich	Ja	Eingeschränkt möglich
<b>Standardisiert</b>	Nein	Ja (W3C)	Ja (W3C)	Nein
<b>Datentypen für Literale</b>	Nein	Nein	Ja	Nein
<b>Erweiterbar</b>	Eingeschränkt möglich	Ja	Ja	Nein
<b>DRY</b>	Ja	Ja	Ja	Ja
<b>HTML Version</b>	HTML 5	HTML 5	HTML 4/5, XHTML	HTML 4/5, XHTML

Tabelle 3.1: Vergleich der Technologien

#### 3.1.1 Anzahl verwendeter Attribute

Um einen ersten Eindruck von der Benutzerfreundlichkeit und der Komplexität der Technologien zu bekommen, kann man die Anzahl der verwendeten Attribute heranziehen.



Verwendet eine Technologie viele Attribute, so dürfte es mit dieser Technologie möglich sein komplexere Strukturen zu erzeugen. Diese komplexen Strukturen können dann mit anderen Technologien, welche weniger Attribute verwenden, wahrscheinlich nicht erzeugt werden. Jedoch nimmt auch mit zunehmender Komplexität die Benutzerfreundlichkeit ab. Mit Microdata werden fünf neue Attribute eingeführt und fünf bestehende Attribute verwendet.[Hic12] Da Schema.org auf Microdata basiert, werden hier auch zehn Attribute verwendet. Bei RDFa 1.1 werden acht neue Attribute eingeführt und ebenfalls fünf bestehende Attribute verwendet.[ABMH12] Microformats verzichtet hingegen ganz auf neue Attribute und verwendet ausschließlich vier bestehende Attribute.[Gra07]

### 3.1.2 URIs

Eine weitere Eigenschaft bei Technologien zur semantischen Annotation ist, dass Knoten und Kanten eines Tripels anhand von URIs identifiziert werden können. Dies ist sowohl bei Microdata, Schema.org als auch bei RDFa möglich, nicht jedoch bei Microformats.[ABMH12, Hic12]

### 3.1.3 Namespaces

Um den Code einfacher lesbar zu machen, werden in der Programmierung Namespaces verwendet. Im Bezug auf das Web bedeutet dies meist, dass in einem XML-Dokument eine URI durch ein Präfix abgekürzt wird.[BHL<sup>+</sup>09] Bei den hier betrachteten Technologien wird diese Funktion nur von RDFa unterstützt.[ABMH12]

### 3.1.4 Umandlung in RDF

Um beispielsweise die semantischen Annotationen von einer Webseite in einen Triple Store zu laden ist es notwendig die Annotationen in RDF umzuwandeln. Bei RDFa ist dies ohne Probleme möglich.[ABMH12] Für die Umwandlung von Schema.org Annotationen in RDF stellt das "Schema.RDFS.org"-Projekt Werkzeuge zur Verfügung.<sup>1</sup>Bei Microdata kann diese Umwandlung zu Problemen führen. Das W3C ist jedoch dabei, eine Spezifikation zu erstellen, welche beschreibt wie ein Abbildung von Microdata in RDF zu erstellen ist.[HKTH12] Es gibt keine einheitliche Möglichkeit Microformats in RDF umzuwandeln, daher muss für jedes einzelne Microformat festgelegt werden, wie es umzuwandeln ist.[Gra07]

### 3.1.5 Standardisiert

RDFa hat den Status einer Recommendation beim W3C. Microdata[Hic12] hat den Status "Working Draft" und RDFa 1.1 [ABMH12] den Status "Candidate Recommendation". Schema.org und Microformats sind hingegen nicht von einem offiziellen Gremium standardisiert.

### 3.1.6 Datentypen für Literale

Die Funktion den Datentyp eines Literals angeben zu können ist bei Schema.org, Microdata und Microformats nicht vorhanden, lediglich RDFa bietet diese Möglichkeit.[ABMH12]

---

<sup>1</sup><http://schema.rdfs.org/index.html>

### 3.1.7 Erweiterbarkeit

Ein weiterer Punkt beim Vergleich der Einbettungsmöglichkeiten ist die mögliche Erweiterbarkeit dieser Technologien. Um RDFa zu erweitern muss ein neues RDF-Vokabular angelegt werden. Hierfür muss eine URL ausgewählt werden, mit der ein RDF Dokument identifiziert werden kann, welches das neue Vokabular beschreibt.[ABMH12] Microdata ist ebenfalls erweiterbar, so kann jeder ein eigenes Vokabular definieren und es mittels des `itemtype` Attributes einbinden.[Hic12] Microformats bieten keine direkte Erweiterungsmöglichkeit [Adi08]. Durch den Extension Mechanism<sup>2</sup> ist es möglich die Schemata von Schema.org zu erweitern, jedoch erlaubt es der Extension Mechanism nicht völlig neue Vokabulare zu definieren.

### 3.1.8 DRY

DRY steht für "do not repeat yourself" und ist ein Prinzip, welches darauf abzielt Redundanzen zu vermeiden [ES07, S. 5]. Übertragen auf das Semantic Web bedeutet dies, dass Daten, welche sowohl menschen- als auch maschinenlesbar sind, nur in einer Version existieren sollten [Gra07]. Sowohl RDFa als auch Microformats befolgen das DRY-Prinzip [Adi08]. Microdata und Schema.org erfüllen ebenfalls das DRY-Prinzip, da es auch hier keine getrennten Versionen der Daten gibt.

### 3.1.9 HTML Versionen

Aufgrund dessen, dass Microdata ein Teil des HTML5-Standards ist gibt, es offiziell auch nur eine Unterstützung für HTML5. Folgt man der Empfehlung und nutzt Schema.org Schemata mittels Microdata, so gibt es auch hier nur eine Unterstützung für HTML5. Für RDFa 1.1 gibt es sowohl eine Spezifikation für die Einbettung in HTML4/5[ABP12] als auch für die Einbettung in XHTML[McC12]. Microformats können in HTML4/5 und in XHTML eingebettet werden. [Gra07]

## 3.2 Mögliche Anwendungen von semantischen Annotationen

In diesem Abschnitt wird auf einige mögliche Verwendungen von semantischen Annotationen eingegangen.

### 3.2.1 Suchmaschinen

Wie schon Anfangs beschrieben ist ein wichtiger Grund für die semantische Anreicherung von HTML-Seiten, dass Suchmaschinen ihren Nutzern dann ausführlichere Suchergebnisse präsentieren können. In diesem Abschnitt wird beschrieben, wie Google und Bing diese Annotationen für die Darstellung der Suchergebnisse nutzen.

#### 3.2.1.1 Google Rich Snippets

Google Rich Snippets sind zusätzliche Informationen zu einem Suchergebnis, welche es den Nutzern der Google Suchmaschine ermöglichen soll sich einen besseren Eindruck über den Inhalt einer in dem Suchergebnis enthaltenen Seite zu machen.<sup>3</sup> Google hat die Rich Snippets am 12. Mai 2009 präsentiert.<sup>4</sup> Damit es Google möglich ist Rich Snippets zu erstellen und anzuzeigen, müssen die Webseiten semantisch annotiert sein. Für diese semantische Annotation empfiehlt Google Microdata zu verwenden.

Interessant in diesem Zusammenhang ist auch, dass es bei Google eine Debatte darüber

<sup>2</sup><http://schema.org/docs/extension.html>

<sup>3</sup><http://support.google.com/webmasters/bin/answer.py?hl=de&answer=99170>

<sup>4</sup><http://googlewebmastercentral.blogspot.com/2009/05/introducing-rich-snippets.html>

gab, welche Technologien unterstützt werden sollten. Man gelangte dann jedoch zu der Auffassung, dass es notwendig ist mehrere Technologien zu unterstützen und nicht nur eine.<sup>5</sup> Somit werden auch RDFa und Microformats unterstützt aber nicht empfohlen. Google verpflichtet sich jedoch nicht dazu, die Rich Snippets auch immer anzuzeigen.<sup>6</sup> Mit dem Rich Snippets-Tester<sup>7</sup> bietet Google eine Möglichkeit seine Seite auf die korrekte Verwendung von semantischen Annotationen hin zu untersuchen, hierbei wird dann auch angezeigt wie das Rich Snippet aussehen würde. Momentan ist es unter anderem möglich, Seiten welche Beurteilungen, Rezepte, Personen, Events, Produkte, Musik und Unternehmen behandeln, in Form eines Rich Snippets darzustellen.<sup>8</sup> Folgendes Beispiel zeigt eine mit Microdata eingebettete Beurteilung und das daraus resultierende Rich Snippet.

```
<div class="star-box giga-star" itemprop="aggregateRating"
  itemscope="" itemType="http://schema.org/AggregateRating">
  ...
  <span itemprop="bestRating">10</span>
  ...
  <span itemprop="ratingCount">87,111</span>
  ...
  <span itemprop="reviewCount">156</span>
  ...
</div>
```

### [The Big Bang Theory \(TV Series 2007–\) - IMDb](http://www.imdb.com/title/tt0898266/)

[www.imdb.com/title/tt0898266/](http://www.imdb.com/title/tt0898266/) - Diese Seite übersetzen

★★★★★ Bewertung: 8.6/10 - 87,111 Stimmen

A woman who moves into an apartment next door to two brilliant but socially awkward physicists shows them how little they know about life outside of...

Mit [Johnny Galecki](#), [Jim Parsons](#), [Kaley Cuoco](#).

Abbildung 3.1: Google Rich Snippet

#### 3.2.1.2 Microsoft Bing

Ebenso wie Google nutzt auch die Microsoft Suchmaschine Bing semantische Annotationen um die Suchergebnisse mit mehr Information anzureichern. Im Gegensatz zu Google gibt es von Microsoft keine Empfehlung welche Technologie zur semantischen Anreicherung verwendet werden soll.<sup>9</sup> Unterstützt werden Microdata, Microformats und RDFa, außerdem wird noch Schema.org und das Open Graph Protokoll von dem Crawler berücksichtigt.<sup>10</sup> Vergleicht man in der Dokumentation von Google und Bing, wie Webseiten semantisch anzureichern sind damit sie von den Suchmaschinen berücksichtigt werden, so stellt man fest, dass es hierbei fast keinen Unterschied zwischen den Suchmaschinen gibt. Dies ist natürlich für die Webmaster von Vorteil, da diese ihre Webseite nicht für die verschiedenen Suchmaschinen unterschiedlich annotieren müssen. Jedoch gibt es bei Bing im Gegensatz zu Google momentan noch keine Unterstützung für die GoodRelations Ontology.<sup>11</sup>

<sup>5</sup><http://googlewebmastercentral.blogspot.com/2009/05/introducing-rich-snippets.html>

<sup>6</sup><http://support.google.com/webmasters/bin/answer.py?hl=en&answer=1093493>

<sup>7</sup><http://www.google.com/webmasters/tools/richsnippets>

<sup>8</sup><http://support.google.com/webmasters/bin/answer.py?hl=en&answer=99170>

<sup>9</sup><http://onlinehelp.microsoft.com/de-de/bing/hh207238.aspx>

<sup>10</sup><http://onlinehelp.microsoft.com/de-de/bing/hh207238.aspx>

<sup>11</sup><http://onlinehelp.microsoft.com/de-de/bing/hh207242.aspx>

### 3.2.2 Soziale Netzwerke

Neben Suchmaschinen sind auch soziale Netzwerke wichtige Konsumenten von semantischen Anreicherungen in Webseiten. Wie Suchmaschinen nutzen soziale Netzwerke semantische Annotationen um daraus Snippets zu generieren, welche dann in dem Message Stream der sozialen Netze dargestellt werden. Zur Veranschaulichung wird in Abbildung 4.3. ein Snippet aus Facebook dargestellt, das Informationen über ein Video bietet. Damit aus auf Webseiten enthaltenen Objekten, wie zum Beispiel Artikel oder Videos, Snippets generiert werden können, hat Facebook die Ontologie "Open Graph Protocol" entwickelt, welche in Abschnitt 4.2.1 näher beschrieben wird. Das soziale Netzwerk Google+ empfiehlt hingegen Schema.org Annotationen zu nutzen um daraus Snippets zu generieren. Jedoch kann Google+ auch Open Graph Annotationen nutzen.<sup>12</sup>



Abbildung 3.2: Facebook Snippet

<sup>12</sup><https://developers.google.com/+/plugins/+1button/#plus-snippet>

## 4. Analyse

Um Aussagen über die Verwendung von Semantic Web Technologien in Webseiten mit Unternehmensbezug treffen zu können, ist eine genauere Untersuchung einer Menge von Webseiten, die diesen Bezug aufweisen, von Nöten. In diesem Kapitel werden die Ergebnisse einer solchen Untersuchung präsentiert.

### 4.1 Allgemeine Analyse

Als Ausgangspunkt für diese Analyse diente ein Datensatz mit RDF-Tripel. Dieser Datensatz wurde im Januar 2011 mittels Yahoo Search BOSS<sup>1</sup> von Benedikt Kämpgen und Daniel Herzig erstellt. Hierfür wurde nach Unternehmen und Branchen gesucht. Das Problem dieses Datensatzes ist jedoch, dass alle Tripel nur in RDF/XML vorliegen und es somit nicht mehr möglich ist festzustellen, welche Technologie zur Einbettung dieser Tripel verwendet wurden.

Um dieses Problem zu lösen wurden die URIs aus dem Datensatz extrahiert. Anschließend wurden die Webseiten, auf welche die URIs verweisen nach semantischen Annotationen durchsucht. Mit den gefunden Annotationen wurde dann ein neuer Datensatz angelegt. Im Unterschied zum ursprünglichen Datensatz wurde hier auch festgehalten, welche Technologie zur Einbettung verwendet wurde und wie die URIs der einzelnen Webseite lauten. Dies ermöglicht nun eine genaue Analyse. Folgendes Beispiel zeigt eine Einheit des ursprünglichen Datensatzes.

```
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:h="http://www.w3.org/1999/xhtml">
  ...
  <rdf:Description rdf:nodeID="id5993223">
    <vcard:fn xmlns:vcard="http://www.w3.org/2006/vcard/ns#">
      1pointsize
    </vcard:fn>
  </rdf:Description>
  <rdf:Description rdf:nodeID="id5993223">
    <vcard:org xmlns:vcard="http://www.w3.org/2006/vcard/ns#"
      rdf:nodeID="id1579120790"/>
  </rdf:Description>
</rdf:RDF>
```

<sup>1</sup><http://developer.yahoo.com/search/boss/>

```

</rdf:Description>
<rdf:Description rdf:nodeID="id1579120790">
  <vcard:organization -name
    xmlns:vcard="http://www.w3.org/2006/vcard/ns#">
    1pointsize
  </vcard:organization -name>
</rdf:Description>
...
</rdf:RDF>

```

An diesem Beispiel kann man erkennen, dass in den einzelnen Einheiten des Datensatzes keine Informationen über die Technologie, welche zur Einbettung dieser Tripel verwendet wurde, vorhanden sind.

#### 4.1.1 Crawl

Der Crawl wurde mit Hilfe von ARC2 realisiert. ARC2 ist ein freies Softwareprojekt das PHP-Entwicklern bei dem Umgang mit Semantic Web Technologien unterstützt.<sup>2</sup> Unter anderem ist es mit ARC2 möglich RDFa und Microformats aus Webseiten zu extrahieren. Jedoch fehlt bei ARC2 die Unterstützung für Microdata. Daher wurde für die Extraktion von Microdata die Any23 Bibliothek verwendet.<sup>3</sup>

Zunächst extrahierte ein PHP-Skript die URIs aus dem ursprünglichen Datensatz. Aus den einzelnen Webseiten, auf die die URIs verweisen wurden dann die eingebetteten Tripel extrahiert. Die Extraktion wurde durch ein weiteres PHP-Skript durchgeführt. Anschließend wurden die Tripel in Form der JavaScript Object Notation (JSON) gespeichert. Für jedes Tripel wurde das Subjekt, Prädikat, Object, Subjekt Typ, Objekt Typ, Objekt Datentyp und die Technologie, mit welcher das Tripel eingebettet wurde, festgehalten. Folgendes Listing zeigt einen Ausschnitt aus einer der JSON-Dateien:

```

{"uri": "http://www.nycgo.com/?event=view [...]",
 "triple": [
  {"s": "e-mail address",
   "s_type": "uri",
   "p": "http://www.w3.org/1999/02/22-rdf-syntax-ns#type",
   "o": "http://poshrdf.org/ns/mf#Email",
   "o_type": "uri",
   "o_lang": "",
   "o_datatype": ""},
  {"s": "http://www.nycgo.com/",
   "p": "http://purl.org/dc/elements/1.1/title",
   "o": "The Official New York City Guide to NYC Attractions,
        Dining, Hotels and Things to Do \\/ nycgo.com",
   "s_type": "uri",
   "o_type": "literal",
   "o_datatype": "",
   "o_lang": ""},
  {"extractor": "dc"}
 ]
}

```

<sup>2</sup><https://github.com/semsol/arc2/wiki>

<sup>3</sup><http://incubator.apache.org/any23/>

Um SPARQL-Anfragen an den Datensatz stellen zu können, war es notwendig die Tripel aus den JSON-Dateien in einen Triple Store zu laden. Damit die Information, von welcher Webseite die Tripel stammen, nicht verloren geht, wurden Named Graphs verwendet. Die Graphen wurden dabei nach der URI der Webseite, von der die Tripel stammen, benannt. Um die Named Graphs anlegen zu können wurde die Named Graphs API for Jena (NG4J) verwendet.<sup>4</sup>

Von den ursprünglichen 4839 Einheiten konnten nur 2971 für die Auswertung verwendet werden, dies liegt unter anderem daran, dass bei 958 Einheiten keine verwertbare URI gefunden wurde. Außerdem traten bei 854 Einheiten Extraktionsfehler auf, zum Beispiel konnte bei einigen Einheiten die zu der URI gehörige Ressource nicht gefunden werden.

Einheit	Anzahl	%
Ursprüngliche Einheiten	4839	100
Einheiten ohne verwertbar URI	958	19,80
Doppelte URIs	18	0,37
Extraktionsfehler	854	17,65
Sonstige Fehler	38	0,79
Verwertbare Einheiten	2971	61,40

Tabelle 4.1: Fehler

Abbildung 4.1 gibt einen Überblick darüber wie viele Tripel auf den Webseiten eingebettet waren.

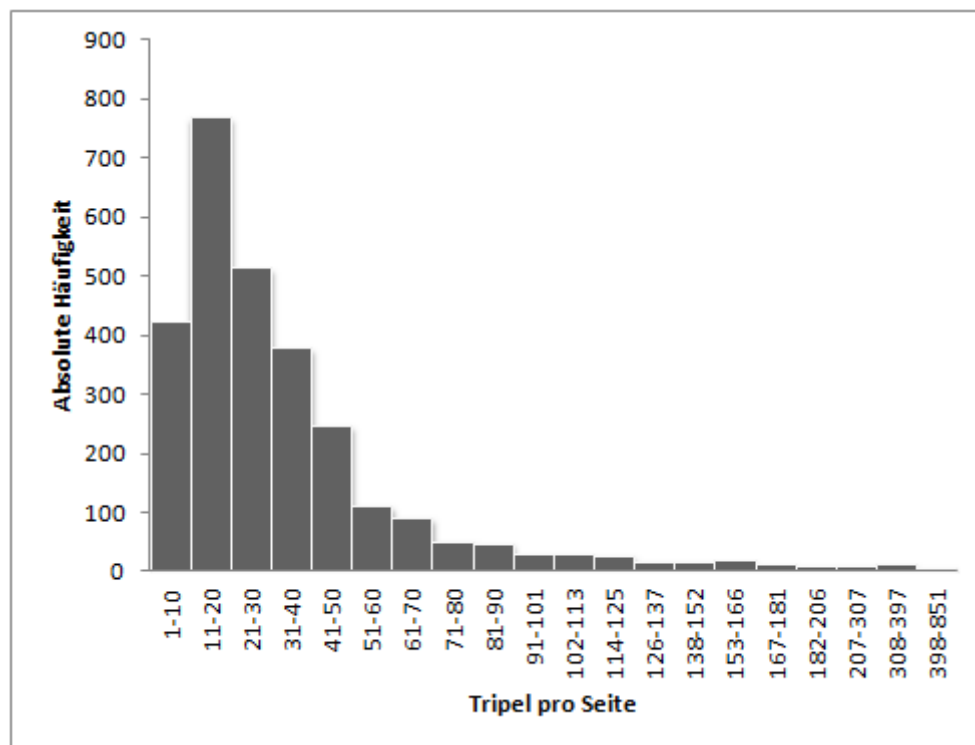


Abbildung 4.1: Anzahl der eingebetteten Tripel pro Webseite

#### 4.1.2 Verbreitung der einzelnen Technologien

In diesem Abschnitt soll der Frage nachgegangen werden, wie weit verbreitet die einzelnen Technologien sind. Um diese Verbreitung zu messen, wurden die einzelnen JSON-Dateien

<sup>4</sup><http://www4.wiwi.fu-berlin.de/bizer/ng4j/>

mittels eines JAVA-Programms daraufhin untersucht, welche Technologien in den einzelnen Einheiten genutzt wurden. Tabelle 4.2 gibt für die einzelnen Technologien an, in wie vielen Einheiten des Datensatzes sie verwendet wurden. Die Erhebung zeigt, dass Microformats mit 71% am häufigsten verbreitet ist. Am zweit häufigsten wurde RDFa verwendet, gefolgt von Microdata. Es dürfte interessant sein vor allem den Anteil von Microdata weiter zu verfolgen, da durch die Schema.org-Empfehlung, Microdata zu verwenden, diese Technologie in nächster Zeit immer populärer werden dürfte.

Technologie	Anzahl	% im Bezug auf alle Einheiten
RDFa	1469	49,44
Microformats	2103	70,78
Microdata	156	5,45

Tabelle 4.2: Semantic Web Technologien

Ein weiteres Maß für die Verbreitung der einzelnen Technologien ist die Anzahl der Tripel, die mit einer Technologie eingebettet wurden. Das Ergebnis dieser Untersuchung, welche ebenfalls mit einem JAVA-Programm durchgeführt wurde, wird in Tabelle 4.3 dargestellt.

Technologie	Anzahl	Anteil an allen Tripel in %
Microformats	72876	68,16
RDFa	28009	26,19
Microdata	6041	5,65

Tabelle 4.3: Tripel

### 4.1.3 Prädikate

Da das Prädikat eines RDF-Tripels die Beziehung zwischen zwei Ressourcen ausdrückt, ist es interessant, welche Prädikate am häufigsten verwendet wurden. Daher findet sich im folgendem Abschnitt eine Aufstellung über die am meisten verwendeten Prädikate, aufgeschlüsselt nach der Technologie, mit der sie eingebettet wurden.

Prädikat	Anzahl	%
<a href="http://www.w3.org/1999/xhtml/vocab#stylesheet">http://www.w3.org/1999/xhtml/vocab#stylesheet</a>	6351	22,67
<a href="http://www.w3.org/1999/xhtml/vocab#alternate">http://www.w3.org/1999/xhtml/vocab#alternate</a>	1722	6,15
<a href="http://www.w3.org/1999/xhtml/vocab#bookmark">http://www.w3.org/1999/xhtml/vocab#bookmark</a>	1330	4,75
<a href="http://www.w3.org/1999/xhtml/vocab#icon">http://www.w3.org/1999/xhtml/vocab#icon</a>	1198	4,28
<a href="http://www.w3.org/1999/02/22-rdf-syntax-ns#type">http://www.w3.org/1999/02/22-rdf-syntax-ns#type</a>	1095	3,91
<a href="http://www.facebook.com/2008/fbmlapp_id">http://www.facebook.com/2008/fbmlapp_id</a>	760	2,71
<a href="http://rdf.data-vocabulary.org/#title">http://rdf.data-vocabulary.org/#title</a>	670	2,39
<a href="http://rdf.data-vocabulary.org/#url">http://rdf.data-vocabulary.org/#url</a>	669	2,39
<a href="http://opengraphprotocol.org/schema/site_name">http://opengraphprotocol.org/schema/site_name</a>	654	2,33
<a href="http://opengraphprotocol.org/schema/title">http://opengraphprotocol.org/schema/title</a>	618	2,21

Tabelle 4.4: Häufige RDFa Prädikate

Betrachtet man Tabelle 4.4, welche die am häufigsten vorkommenden RDFa-Prädikate auflistet, so sieht man, dass es sich bei den vier am häufigsten vorkommenden Prädikaten um XHTML Metainformationen<sup>5</sup> handelt. Außerdem werden Prädikate des Open Graph Protokolls häufig verwendet. Das Open Graph Protokoll ist eine von Facebook entwickelte

<sup>5</sup><http://www.w3.org/TR/2006/WD-xhtml2-20060726/mod-metaAttributes.html>



Ontologie, welche es erlaubt Webseiten in einen sogenannten "Social Graph" zu integrieren.<sup>6</sup>

Prädikat	Anzahl	%
<a href="http://poshrdf.org/ns/mf#nofollow">http://poshrdf.org/ns/mf#nofollow</a>	31834	43,68
<a href="http://www.w3.org/1999/02/22-rdf-syntax-ns#type">http://www.w3.org/1999/02/22-rdf-syntax-ns#type</a>	12662	17,37
<a href="http://poshrdf.org/ns/mf#fn">http://poshrdf.org/ns/mf#fn</a>	3042	4,17
<a href="http://poshrdf.org/ns/mf#bookmark">http://poshrdf.org/ns/mf#bookmark</a>	2739	3,76
<a href="http://poshrdf.org/ns/mf#tag">http://poshrdf.org/ns/mf#tag</a>	2571	3,53
<a href="http://poshrdf.org/ns/mf#org">http://poshrdf.org/ns/mf#org</a>	2048	2,81
<a href="http://poshrdf.org/ns/mf#hentry">http://poshrdf.org/ns/mf#hentry</a>	1064	1,46
<a href="http://poshrdf.org/ns/mf#license">http://poshrdf.org/ns/mf#license</a>	1042	1,43
<a href="http://poshrdf.org/ns/mf#url">http://poshrdf.org/ns/mf#url</a>	977	1,34
<a href="http://poshrdf.org/ns/mf#adr">http://poshrdf.org/ns/mf#adr</a>	951	1,30

Tabelle 4.5: Häufige Microformats Prädikate

Tabelle 4.5 listet die zehn am häufigsten vorkommenden Microformats-Prädikate auf. Mit Abstand am häufigsten wurde das `nofollow`-Prädikat verwendet, es wird verwendet um Suchmaschinen anzuweisen, dass sie einen Hyperlink nicht beachten sollen. Dieses Prädikat wurde notwendig, da durch die vermehrte Platzierung von Links versucht wurde die Linkpopularität von manchen Seiten zu erhöhen.<sup>7</sup>

Prädikat	Anzahl	%
<a href="http://www.w3.org/1999/xhtml/vocab#nofollow">http://www.w3.org/1999/xhtml/vocab#nofollow</a>	1525	25,24
<a href="http://www.w3.org/1999/xhtml/vocab#ALTERNATE-STYLE SHEET">http://www.w3.org/1999/xhtml/vocab#ALTERNATE-STYLE SHEET</a>	617	10,21
<a href="http://www.w3.org/1999/02/22-rdf-syntax-ns#type">http://www.w3.org/1999/02/22-rdf-syntax-ns#type</a>	331	5,48
<a href="http://www.w3.org/1999/xhtml/microdata#item">http://www.w3.org/1999/xhtml/microdata#item</a>	259	4,29
<a href="http://www.w3.org/1999/xhtml/vocab#icon">http://www.w3.org/1999/xhtml/vocab#icon</a>	172	2,85
<a href="http://purl.org/dc/terms/title">http://purl.org/dc/terms/title</a>	161	2,67
<a href="http://data-vocabulary.org/Breadcrumb/title">http://data-vocabulary.org/Breadcrumb/title</a>	159	2,63
<a href="http://www.w3.org/1999/xhtml/vocab#description">http://www.w3.org/1999/xhtml/vocab#description</a>	155	2,57
<a href="http://data-vocabulary.org/Breadcrumb/url">http://data-vocabulary.org/Breadcrumb/url</a>	142	2,35
<a href="http://www.w3.org/1999/xhtml/vocab#keywords">http://www.w3.org/1999/xhtml/vocab#keywords</a>	128	2,12

Tabelle 4.6: Häufige Microdata Prädikate

In Tabelle 4.6 befindet sich eine Übersicht über die am häufigsten verwendeten Microdata Prädikate. Auch hier wird das `nofollow`-Prädikat am häufigsten verwendet.

#### 4.1.4 Namespaces

Um einen Rückschluss auf die verwendeten Ontologien ziehen zu können, kann man die verwendeten Namespaces betrachten. Die Namespace-Definitionen wurden jedoch bei der Extraktion aufgelöst, daher war es notwendig die Namespaces wieder aus den URIs herzuleiten. Für die Herleitung aus den URIs wurden alle bekannten Namespaces von `prefix.cc`<sup>8</sup> in eine Textdatei geladen. Mit Hilfe eines Java-Programms wurde dann untersucht, welche URIs aus dem Datensatz die Namespaces von `prefix.cc` enthalten. Tabelle 4.7 zeigt, welche

<sup>6</sup><http://ogp.me/>

<sup>7</sup><http://microformats.org/wiki/rel-nofollow>

<sup>8</sup><http://prefix.cc/>

Namespaces bei den Tripeln Verwendung fanden, die mit RDFa eingebettet wurden. Insgesamt wurden 34 verschiedene Namespaces gefunden. Am häufigsten wurde das XHTML Vokabular<sup>9</sup> verwendet.

Prefix	Namespace	Anzahl
xhv	<a href="http://www.w3.org/1999/xhtml/vocab#">http://www.w3.org/1999/xhtml/vocab#</a>	11356
og	<a href="http://opengraphprotocol.org/schema/">http://opengraphprotocol.org/schema/</a>	3353
dv	<a href="http://rdf.data-vocabulary.org/#">http://rdf.data-vocabulary.org/#</a>	2982
rdf	<a href="http://www.w3.org/1999/02/22-rdf-syntax-ns#">http://www.w3.org/1999/02/22-rdf-syntax-ns#</a>	1095
vcard	<a href="http://www.w3.org/2006/vcard/ns#">http://www.w3.org/2006/vcard/ns#</a>	1026
ogp	<a href="http://ogp.me/ns#">http://ogp.me/ns#</a>	632
dc	<a href="http://purl.org/dc/elements/1.1/">http://purl.org/dc/elements/1.1/</a>	424
dbtpl	<a href="http://dbpedia.org/resource/Template:">http://dbpedia.org/resource/Template:</a>	400
commerce	<a href="http://search.yahoo.com/searchmonkey/commerce/">http://search.yahoo.com/searchmonkey/commerce/</a>	192
media	<a href="http://search.yahoo.com/searchmonkey/media/">http://search.yahoo.com/searchmonkey/media/</a>	135

Tabelle 4.7: Häufige RDFa Namespaces

Bei Microformats wurden nur zwei Namespaces gefunden. Dies hängt jedoch damit zusammen, dass es mit Microformats eigentlich nicht möglich ist Namespaces zu verwenden, jedoch kann man die verwendeten URIs nachträglich einem Namespace zuordnen. Hierbei können aber keine Rückschlüsse auf die verwendete Ontologien gezogen werden, da fast allen Microformats der selbe Namespace zugeordnet wurde.

Prefix	Namespace	Anzahl
mf	<a href="http://poshrdf.org/ns/mf#">http://poshrdf.org/ns/mf#</a>	73637
rdf	<a href="http://www.w3.org/1999/02/22-rdf-syntax-ns#">http://www.w3.org/1999/02/22-rdf-syntax-ns#</a>	11356

Tabelle 4.8: Microformats Namespaces

Wie bei Microformats gibt es bei Microdata keine Namespaces. Tabelle 4.9 stellt das Ergebnis der nachträglichen Zuordnung von Namespaces zu URIs, dar. Wie bei RDFa wurde auch hier das XHTML Vokabular am häufigsten verwendet.

Prefix	Namespace	Anzahl
xhv	<a href="http://www.w3.org/1999/xhtml/vocab#">http://www.w3.org/1999/xhtml/vocab#</a>	4289
rdf	<a href="http://www.w3.org/1999/02/22-rdf-syntax-ns#">http://www.w3.org/1999/02/22-rdf-syntax-ns#</a>	331
dc	<a href="http://purl.org/dc/terms/">http://purl.org/dc/terms/</a>	161
schema	<a href="http://schema.org/">http://schema.org/</a>	151
rich	<a href="http://rdf.data-vocabulary.org/">http://rdf.data-vocabulary.org/</a>	15

Tabelle 4.9: Microdata Namespaces

#### 4.1.5 Klassenbezeichner

Ein weiterer interessanter Punkt um zu verstehen, was semantisch annotiert wird, ist zu untersuchen, welche Klassenbezeichner verwendet wurden.

Für die Zuweisung von Subjekten zu Klassenbezeichner wird das Prädikat `rdf:type` verwendet. Wird ein Subjekt einem Klassenbezeichner zugewiesen, so bedeutet dies, dass die Ressource, auf welche das Subjekt verweist, eine Instanz der Klasse ist, auf die der Klassenbezeichner verweist. [HKRS08, S.60]

<sup>9</sup><http://www.w3.org/1999/xhtml/vocab/>

Um nun die verwendeten Klassenbezeichner auszugeben, wurde folgende SPARQL-Anfrage verwendet:

```
PREFIX rdf:<http://www.w3.org/1999/02/22-rdf-syntax-ns#>
SELECT ?o WHERE {
    ?s rdf:type ?o
}
```

Typ	Anzahl	%
http://rdf.data-vocabulary.org/#Breadcrumb	677	61,94
http://www.w3.org/2006/vcard/ns#VCard	88	8,05
http://search.yahoo.com/searchmonkey/commerce/Business	82	7,50
http://rdf.data-vocabulary.org/#Review	37	3,39
http://rdf.data-vocabulary.org/#Review-aggregate	25	2,29
http://rdf.data-vocabulary.org/#Organization	21	1,92
http://www.purl.org/stuff/rev#Review	19	1,74
http://rdfs.org/sioc/ns#UserAccount	13	1,19
http://xmlns.com/foaf/0.1/Document	11	1,01
http://xmlns.com/foaf/0.1/Image	10	0,91

Tabelle 4.10: Häufige RDFa Typen

Bei mit RDFa eingebetteten Tripeln wurden 66 verschiedene Klassenbezeichner gefunden. Eine Übersicht der am häufigsten verwendeten Klassenbezeichner bietet Tabelle 4.10. Am häufigsten wurde die Klasse `dv:Breadcrumb` instanziiert. Bei Breadcrumbs handelt es sich um eine Abbildung der Hierarchie einer Webseite die dem Nutzer einen Überblick über die Seite geben soll.<sup>10</sup>

Typ	Anzahl	%
http://posh rdf.org/ns/mf#Vcard	3610	29,43
http://posh rdf.org/ns/mf#Item	1340	10,92
http://posh rdf.org/ns/mf#Category	1318	10,74
http://posh rdf.org/ns/mf#Hentry	1076	8,77
http://posh rdf.org/ns/mf#Adr	945	7,70
http://posh rdf.org/ns/mf#Author	834	6,80
http://posh rdf.org/ns/mf#Vevent	603	4,92
http://posh rdf.org/ns/mf#N	477	3,89
http://posh rdf.org/ns/mf#Experience	445	3,63
http://posh rdf.org/ns/mf#Affiliation	412	3,36

Tabelle 4.11: Häufige Microformats Typen

In allen Microformats Tripeln fanden sich 19 verschiedene Klassenbezeichner. Am meisten wurde der Typ `Vcard` verwendet.

Tabelle 4.12 zeigt die zehn meist verwendeten Microdata-Klassenbezeichner, insgesamt wurden hier 42 unterschiedliche Klassenbezeichner gefunden. Wie bei RDFa wurde auch hier die Klasse `Breadcrumb` am häufigsten instanziiert.

<sup>10</sup><http://support.google.com/webmasters/bin/answer.py?hl=de&answer=185417>

Typ	Anzahl	%
http://data-vocabulary.org/Breadcrumb	157	47,58
http://schema.org/Article	19	5,76
http://schema.org/VideoObject	18	5,45
http://schema.org/Person	13	3,94
http://schema.org/WebPage	11	3,33
http://schema.org/LocalBusiness	8	2,42
http://schema.org/Organization	7	2,12
http://schema.org/Product	7	2,12
http://schema.org/MusicGroup	6	1,82
http://data-vocabulary.org/person	6	1,82

Tabelle 4.12: Häufige Microdata Typen

#### 4.1.6 Subjekte und Objekte

Einen Rückschluss auf die Komplexität der semantischen Struktur in dem Datensatz kann man ziehen, indem man die Art der Subjekte und Objekte analysiert. So sind für manche komplexeren Strukturen Blank Nodes von Nöten, zum Beispiel sind sie unerlässlich, wenn man mehrwertige Beziehungen darstellen will.[HKRS08, S. 56] Tabelle 4.13 zeigt, dass es sich bei den Subjekten nur in 5% um Blank Nodes handelt und bei den Objekten sind nur 1% Blank Nodes.

Ressourcenart	Ressourcentyp	Anzahl	% an der Ressourcenart
Subjekt	URI	102118	95,50
	BNode	5020	4,69
Objekt	URI	75451	70,56
	Literal	30166	28,21
	BNode	1309	1,22

Tabelle 4.13: Ressourcentypen

Bei der Analyse des Datensatzes wurde auch erhoben, welche Datentypen bei den Literalen Verwendung fanden. Wird ein Datentyp angegeben, so erleichtert dies die Verarbeitung durch die Software, welche die semantischen Annotationen extrahiert. Aus Tabelle 4.14 geht jedoch hervor, dass bei 97% der Literale kein Datentyp definiert wurde.

Datentyp	Anzahl	% an allen Literalen
Literale	30166	100,00
Literale ohne Datentyp	29275	97,05
http://www.w3.org/1999/02/22-rdf-syntax-ns#XMLLiteral	863	2,86
http://www.w3.org/2001/XMLSchema#dateTime	13	0,04
http://www.w3.org/2001/XMLSchema#integer	7	0,02
http://www.w3.org/2001/XMLSchema#string	4	0,01
http://www.w3.org/2001/XMLSchema#float	3	0,01
http://www.w3.org/2001/XMLSchema#date	1	0,00

Tabelle 4.14: Literal Datentypen

## 4.2 Ontologien

In diesem Abschnitt werden die häufig verwendeten Ontologien näher beschrieben und auf ihre Verwendung eingegangen.

### 4.2.1 Open Graph Protokoll

Das Open Graph Protokoll bietet die Möglichkeit Webseiten so semantisch anzureichen, dass damit soziale Netzwerke diese Anreicherungen nutzen können. Das Open Graph Protokoll wurde von Facebook für RDFa entwickelt. Man hat kein bestehendes Vokabular genutzt, da es kein einzelnes Vokabular gibt, welches den Bedürfnissen von Facebook genügt.<sup>11</sup>

Mittels folgender Abfrage wurde ermittelt von welchem Typ die Open Graph Annotationen sind. Das Ergebnis der Abfrage findet sich in Tabelle 4.15.

```
PREFIX og:<http://opengraphprotocol.org/schema/>
PREFIX ogp:<http://ogp.me/ns#>
PREFIX rdf:<http://www.w3.org/1999/02/22-rdf-syntax-ns#>
SELECT * WHERE {
  {
    ?s og:type ?o.
  } UNION {
    ?s ogp:type ?o.
  }
}
```

og: bzw. ogp:	Anzahl
article	229
company	148
website	103
game	19
product	16
video.other	14
restaurant	14
album	11
blog	10
actor	8

Tabelle 4.15: Häufige Open Graph Klassenbezeichner

Damit die Open Graph Annotationen genutzt werden können, müssen mindestens die Prädikate `og:title`, `og:type`, `og:image` und `og:url` vorhanden sein.

<sup>11</sup><http://ogp.me/#intro>

og: bzw. ogp:	Anzahl
title	728
site_name	726
url	654
type	647
image	451
description	286
locality	55
longitude	54
latitude	54
postal-code	52

Tabelle 4.16: Häufige Open Graph Prädikate

### 4.2.2 Dublin Core

Dublin Core ist eine Zusammenfassung von Spezifikation für Metadaten, welche eine einfache Beschreibung von Ressourcen ermöglichen. Dublin Core wurde von der "Dublin Core Metadata Initiative" (DCMI) entwickelt.<sup>12</sup>

In dem analysierten Datensatz fanden sich vor allem Prädikate aus dem "Dublin Core Metadata Element Set", diese besteht aus 15 Elementen. Mit diesen Elementen soll es möglich sein eine große Palette an Ressourcen beschreiben zu können.<sup>13</sup> Tabelle 4.17 zeigt die in dem Datensatz verwendeten Prädikate des "Dublin Core Metadata Element Set" und wie häufig diese vorkamen.

Prädikat	Anzahl
<a href="http://purl.org/dc/elements/1.1/title">http://purl.org/dc/elements/1.1/title</a>	5652
<a href="http://purl.org/dc/elements/1.1/format">http://purl.org/dc/elements/1.1/format</a>	4993
<a href="http://purl.org/dc/elements/1.1/relation">http://purl.org/dc/elements/1.1/relation</a>	126
<a href="http://purl.org/dc/elements/1.1/rights">http://purl.org/dc/elements/1.1/rights</a>	41
<a href="http://purl.org/dc/elements/1.1/creator">http://purl.org/dc/elements/1.1/creator</a>	30
<a href="http://purl.org/dc/elements/1.1/publisher">http://purl.org/dc/elements/1.1/publisher</a>	26
<a href="http://purl.org/dc/elements/1.1/editor">http://purl.org/dc/elements/1.1/editor</a>	19
<a href="http://purl.org/dc/elements/1.1/date">http://purl.org/dc/elements/1.1/date</a>	15
<a href="http://purl.org/dc/elements/1.1/created">http://purl.org/dc/elements/1.1/created</a>	10
<a href="http://purl.org/dc/elements/1.1/issued">http://purl.org/dc/elements/1.1/issued</a>	9
<a href="http://purl.org/dc/elements/1.1/description">http://purl.org/dc/elements/1.1/description</a>	9
<a href="http://purl.org/dc/elements/1.1/type">http://purl.org/dc/elements/1.1/type</a>	6
<a href="http://purl.org/dc/elements/1.1/source">http://purl.org/dc/elements/1.1/source</a>	2
<a href="http://purl.org/dc/elements/1.1/identifier">http://purl.org/dc/elements/1.1/identifier</a>	1

Tabelle 4.17: DC Prädikate

### 4.2.3 hCard

Bei hCard handelt es sich um eine Umsetzung des vCard-Standards für das Semantic Web.<sup>14</sup> Der vCard-Standard spezifiziert ein Dateiformat für elektronische Visitenkarten. [DH98] Mit folgender SPARQL-Abfrage soll geklärt werden, welche hCard-Prädikate am häufigsten verwendet wurden. Das Ergebnis findet sich in Tabelle 4.18.

<sup>12</sup><http://dublincore.org/metadata-basics/>

<sup>13</sup><http://dublincore.org/documents/dces/>

<sup>14</sup><http://microformats.org/wiki/hcard>

```

PREFIX vcmf:<http://poshrdf.org/ns/mf#>
PREFIX vcrdfa:<http://www.w3.org/2006/vcard/ns#>
PREFIX rdf:<http://www.w3.org/1999/02/22-rdf-syntax-ns#>
SELECT ?p WHERE {
  {
    ?s rdf:type vcmf:Vcard;
    ?p ?o.
  } UNION { ?a rdf:type vcrdfa:VCard;
    ?p ?o.
  }
}

```

Prädikat	Anzahl
http://poshrdf.org/ns/mf#fn	2816
http://poshrdf.org/ns/mf#org	1983
http://poshrdf.org/ns/mf#url	914
http://poshrdf.org/ns/mf#photo	783
http://poshrdf.org/ns/mf#adr	776
http://poshrdf.org/ns/mf#title	760
http://poshrdf.org/ns/mf#category	621
http://poshrdf.org/ns/mf#summary	559
http://poshrdf.org/ns/mf#dtstart	538
http://poshrdf.org/ns/mf#n	485

Tabelle 4.18: Häufige hCard Prädikate

#### 4.2.4 XHTML Vocabulary

Eine häufig eingesetzte Ontologie ist das XHTML Metainformation Vokabular. Entwickelt wurde diese Ontologie von der W3C XHTML 2 Working Group. In diesem Vokabular werden Elemente definiert, welche zusammen mit den Attributen `rel` und `rev` eingesetzt werden können.<sup>15</sup>

Prädikat	Anzahl
http://www.w3.org/1999/xhtml/vocab#stylesheet	6357
http://www.w3.org/1999/xhtml/vocab#alternate	1722
http://www.w3.org/1999/xhtml/vocab#nofollow	1525
http://www.w3.org/1999/xhtml/vocab#icon	1370
http://www.w3.org/1999/xhtml/vocab#bookmark	1365
http://www.w3.org/1999/xhtml/vocab#ALTERNATE-STYLE SHEET	617
http://www.w3.org/1999/xhtml/vocab#next	208
http://www.w3.org/1999/xhtml/vocab#prev	192
http://www.w3.org/1999/xhtml/vocab#keywords	155
http://www.w3.org/1999/xhtml/vocab#index	128

Tabelle 4.19: Häufige XHTML Vocabulary Prädikate

<sup>15</sup><http://www.w3.org/1999/xhtml/vocab/>

### 4.2.5 Schema.org

In diesem Abschnitt wird darauf eingegangen, inwieweit Schema.org bisher Verwendung findet. Tabelle 4.20 zeigt die zehn am häufigsten verwendeten Schema.org Prädikate, welche mit Microdata eingebettet wurden. Insgesamt wurden 150 verschiedene Schema.org Prädikate verwendet.

Prädikat	Anzahl	%
<a href="http://schema.org/Person/performerIn">http://schema.org/Person/performerIn</a>	24	4,01
<a href="http://schema.org/WebPage/photo">http://schema.org/WebPage/photo</a>	21	3,51
<a href="http://schema.org/Person/jobTitle">http://schema.org/Person/jobTitle</a>	19	3,18
<a href="http://schema.org/Article/name">http://schema.org/Article/name</a>	18	3,01
<a href="http://schema.org/VideoObject/description">http://schema.org/VideoObject/description</a>	18	3,01
<a href="http://schema.org/VideoObject/embedURL">http://schema.org/VideoObject/embedURL</a>	18	3,01
<a href="http://schema.org/VideoObject/height">http://schema.org/VideoObject/height</a>	18	3,01
<a href="http://schema.org/VideoObject/name">http://schema.org/VideoObject/name</a>	18	3,01
<a href="http://schema.org/VideoObject/playerType">http://schema.org/VideoObject/playerType</a>	18	3,01
<a href="http://schema.org/VideoObject/thumbnailUrl">http://schema.org/VideoObject/thumbnailUrl</a>	18	3,01

Tabelle 4.20: Häufige Microdata Schema.org Prädikate

Tabelle 4.21 listet die mit RDFa eingebetteten Schema.org Prädikate auf. An der Tatsache, dass nur 11 Tripel mit RDFa eingebettet wurden kann man sehen, dass sich größtenteils an die Empfehlung, Microdata zu verwenden, gehalten wird.

Prädikat	Anzahl	%
<a href="http://schema.org/name">http://schema.org/name</a>	6	54,55
<a href="http://schema.org/articleBody">http://schema.org/articleBody</a>	5	45,45

Tabelle 4.21: RDFa Schema.org Prädikate

Schema.org bietet momentan 440 verschiedene Klassenbezeichner, in dem analysierten Datensatz fanden sich 27 was bedeutet, dass 6% der Klassenbezeichner verwendet wurden. Um einen Überblick darüber zu bekommen, welche Klassenbezeichner verwendet wurden, werden alle in Tabelle 4.22 aufgelistet. Neben den Klassenbezeichnern ist noch interessant, welche Properties verwendet wurden, da ein Klassenbezeichner nur ausdrückt, von welchem Typ eine Ressource ist. Mit Properties können den Ressourcen noch weitere Eigenschaften zugeordnet werden. So kann beispielsweise einer Person mit Hilfe der Property <http://schema.org/Person/name> ein Name zugeordnet werden. Daher wird im Folgenden analysiert, wie viele und welche Properties die drei am häufigsten vorkommenden Schema.org-Properties haben.



Thing				1		
	CreativeWork			3		
		Article			19	
		Blog			3	
		MediaObject			0	
			VideoObject			18
		Movie			3	
		Review			5	
		TVEpisode			1	
		TVSeries			2	
		WebPage			11	
	Event				2	
	Intangible				0	
		JobPosting			1	
		Offer			5	
		Rating			5	
			AggregateRating			5
		StructuredValue			0	
			ContactPoint			0
	PostalAddress					6
	Organization				7	
		LocalBusiness			8	
			FoodEstablishment			0
				Restaurant		
		LodgingBusiness			0	
			Hotel			1
		PerformingGroup			6	
	MusicGroup				6	
	Person				13	
	Place				2	
		LocalBusiness			8	
	Product				7	

Tabelle 4.22: Verwendete Schema.org Klassenbezeichner

Der Schema.org-Typ `Article` hat insgesamt 44 Properties, 41 davon wurden von in der Hierarchie höher stehenden Typen vererbt. Insgesamt wurden in dem analysierten Datensatz 4 verschiedene Properties verwendet, dabei wurden nur vererbte Properties verwendet. Außerdem fällt beim Betrachten von Tabelle 4.23 auf, dass der eigentliche Inhalt eines Artikels nie annotiert wurde, was mit `articleBody` möglich gewesen wäre.<sup>16</sup>

<sup>16</sup><http://schema.org/Article>

Property	Anzahl
http://schema.org/Article/name	18
http://schema.org/Article/description	16
http://schema.org/Article/image	5
http://schema.org/Article/author	1

Tabelle 4.23: Schema.org Article

Tabelle 4.24 gibt eine Übersicht über die Properties, welche im Zusammenhang mit dem `ViedoObject` Typ in dem analysierten Datensatz verwendet wurden. Auffällig ist hier, dass alle Properties gleich häufig vorkommen, was daran liegt, dass das `ViedoObject` nur auf einer Webseite eingesetzt wurde. Auch bei dem `ViedoObject`-Typ wurden nur vererbte Properties verwendet. Insgesamt gibt es 62 Properties von denen 8 genutzt wurden, dies entspricht 12 %.

Property	Anzahl
http://schema.org/VideoObject/description	18
http://schema.org/VideoObject/embedURL	18
http://schema.org/VideoObject/height	18
http://schema.org/VideoObject/name	18
http://schema.org/VideoObject/playerType	18
http://schema.org/VideoObject/thumbnailUrl	18
http://schema.org/VideoObject/url	18
http://schema.org/VideoObject/width	18

Tabelle 4.24: Schema.org VideoObject

Um Informationen über Personen semantisch anzureichern bietet Schema.org den Typ `Person`. Dieser Typ bietet 100 Properties, von denen 10 in dem Datensatz Verwendung fanden.

Property	Anzahl
http://schema.org/Person/performerIn	24
http://schema.org/Person/jobTitle	19
http://schema.org/Person/name	13
http://schema.org/Person/birthDate	6
http://schema.org/Person/image	4
http://schema.org/Person/deathDate	2
http://schema.org/Person/description	2
http://schema.org/Person/url	2
http://schema.org/Person/address	1
http://schema.org/Person/memberOf	1

Tabelle 4.25: Schema.org Person

### 4.3 Rich Snippets Annotationen im Datensatz

In Abschnitt 3.2.1 wurden die Rich Snippets vorgestellt. In diesem Abschnitt wird nun untersucht, wie viele der untersuchten Webseiten semantische Annotationen enthalten, aus denen Rich Snippets erzeugt werden können. Um dies zu bewerkstelligen, wurde der Datensatz mittels SPARQL-Anfragen analysiert. Beispielsweise wurde mit folgender Anfrage die Anzahl der Annotationen ermittelt, welche den Anforderung von Google und Bing entsprechen, um Rich Snippets vom Typ `Organisation` bilden zu können.

```

PREFIX v:<http://rdf.data-vocabulary.org/#>
PREFIX rdf:<http://www.w3.org/1999/02/22-rdf-syntax-ns#>
SELECT ?s WHERE {
    ?s v:name ?o .
    ?s rdf:type v:Organization .
}

```

Insgesamt wurden auf 328 untersuchten Webseiten Annotationen gefunden, aus denen Rich Snippets gebildet werden können, dies entspricht 11% der untersuchten Webseiten. Tabelle 4.26 listet die verschiedenen Typen und die Anzahl der Annotationen auf, welche den Anforderungen<sup>17</sup> von Google und Bing entsprechen.

	<b>RDFa</b>	<b>Microformats</b>	<b>Microdata</b>
<b>Organizations</b>	19	6	3
<b>People</b>	2	0	1
<b>Reviews (aggregated)</b>	25	0	5
<b>Reviews</b>	37	0	1
<b>Events</b>	0	0	0
<b>Breadcrumbs</b>	677	-	157
<b>Recipes</b>	0	0	0

Tabelle 4.26: Rich Snippets

Die Tatsache, dass keine semantische Annotationen gefunden wurden, aus denen Snippets vom Typ Events und Recipes gebildet werden können, dürfte damit zusammenhängen, dass der Datensatz auf den Unternehmensbereich fokussiert ist.

<sup>17</sup><http://support.google.com/webmasters/bin/answer.py?hl=de&answer=99170&topic=21997&ctx=topic>

## 5. Diskussion

In dieser Arbeit wurden Webseiten mit eingebettetem RDF und die Technologien, welche die Einbettung ermöglichen, untersucht. Als Ausgangspunkt diente ein Datensatz mit unternehmensbezogenen Webseiten. Zuerst wurden die in dem Datensatz eingesetzten Technologien zur semantischen Anreicherung von HTML-Seiten betrachtet. Anschließend wurden dann die einzelnen Tripel unter anderem dahingehend untersucht, welche Vokabulare verwendet wurden, welche Prädikate am häufigsten verwendet wurden und von welchem Typ die Subjekte bzw. Objekte waren.

Bei der Betrachtung der Technologien, mit denen eine semantischen Anreicherung von Webseiten möglich ist, hat sich gezeigt, dass RDFa die größte Kompatibilität zu RDF aufweist. Der Aufwand bei der Nutzung von RDFa ist jedoch größer als bei den anderen Technologien. Microformats lassen sich mit dem geringsten Aufwand einbetten, allerdings eignen sich Microformats nicht dazu den gesamten Inhalt von Webseiten semantisch zu annotieren, da sie nicht erweiterbar sind und die momentan vorhandenen Microformats nur eine überschaubare Menge an Inhaltstypen abdecken. Microdata ist im Gegensatz zu den Microformats erweiterbar und der Aufwand Microdata zu nutzen ist geringer als bei RDFa, jedoch ist die Umwandlung in RDF hier nicht ganz unproblematisch.

Die Analyse des Datensatzes hat gezeigt, dass momentan Microformats mit Abstand am weitesten verbreitet sind. Auf 71% der untersuchten Webseiten wurde mindestens ein Microformat verwendet. Die weite Verbreitung von Microformats dürfte damit zusammenhängen, dass Microformats für die Webseitenbetreiber am einfachsten einzusetzen sind. RDFa, die ausdrucksstärkste Technologie, wurde auf der Hälfte der analysierten Webseiten gefunden. Microdata wurde nur auf 5% der Webseiten eingesetzt. Für die Zukunft ist jedoch zu erwarten, dass sich der Anteil von Microdata, im Gegensatz zu den anderen Technologien, weiter erhöhen wird. Für die steigende Popularität von Microdata dürfte unter anderem Schema.org sorgen, da Schema.org den Einsatz von Microdata empfiehlt.<sup>1</sup> Bei der Untersuchung der einzelnen Tripel wurde deutlich, dass viele Betreiber, von unternehmensbezogenen Webseiten, die semantische Anreicherung von HTML nutzen um ihre Webseite für Suchmaschinen zu optimieren. Deutlich wird dies, wenn man die am häufigsten verwendeten Klassenbezeichner betrachtet. Bei dieser Betrachtung zeigt sich, dass Klassenbezeichner, die für die Erstellung von Rich Snippets Verwendung finden, mit am häufigsten vorkamen, zum Beispiel wurden die Klassen `dv:Breadcrumb` und `dv:Review` besonders häufig instanziiert.

Die Analyse des Datensatzes hat ergeben, dass Schema.org im Bereich der Unternehmens-

---

<sup>1</sup><http://www.schema.org/docs/gs.html>

---

webseiten noch nicht sonderlich weit verbreitet ist. Auch wird der gebotene Umfang an Typen nur zum Teil genutzt. In Zukunft wird sich dies wahrscheinlich ändern, da viele Webseitenbetreiber semantische Annotationen nutzen um von Suchmaschinen besser "verstanden" zu werden und Schema.org ist ein Ansatz der von den wichtigsten Suchmaschinenanbietern unterstützt wird.

Ein ebenfalls wichtiger Grund seine Webseite semantisch anzureichern ist, diese in sozialen Netzwerken sichtbar zu machen, was durch die Popularität des Open Graph Protokolls deutlich wird.

Die Untersuchung hat gezeigt, dass auf vielen Webseiten mit Unternehmensbezug nur allgemeine Informationen, wie zum Beispiel der Titel der Seite oder eine kurze Beschreibung, semantisch annotiert sind. Unter anderem ist die geringe Zahl der Tripel, die durchschnittlich auf einer Webseite eingebettet wurden, ein Indiz dafür, dass meist nur allgemeine Informationen semantisch annotiert sind und nicht der gesamte Inhalt der Webseiten.

Bei den Gründen für den Einsatz von semantischen Annotationen zeigt sich, dass für die meisten Betreiber von unternehmensbezogenen Webseiten wahrscheinlich die erhöhte Sichtbarkeit in sozialen Netzwerken und Suchmaschinen ausschlaggebend war und weniger der Versuch, den Inhalt ihrer Webseiten für Maschinen verständlich zu machen. Zwar schließt das eine das andere nicht aus, aber um den gesamten Inhalt einer Webseite für Maschinen verständlich zu machen müssten die Möglichkeiten zur semantischen Annotation ausgiebiger genutzt werden.

Zu den Technologien kann man sagen, dass für eine maschinenlesbare Darstellung eines möglichst großen Teils von HTML-Seiten nur RDFa und Microdata in Frage kommen. Legt man viel Wert auf eine einfache Einbettung und die Unterstützung durch Suchmaschinen aber weniger Wert auf eine unproblematische Umwandlung der Annotationen in RDF, sollte Microdata zum Einsatz kommen. Hat jedoch die Kompatibilität zu RDF die höchste Priorität, sollte RDFa eingesetzt werden.

## 6. Verwandte Arbeiten

Diese Arbeit gibt einen allgemeinen Überblick über die in dem analysierten Datensatz vorkommenden Ontologien. Es gibt Arbeiten die sich eine nähere Betrachtung der einzelnen Ontologien zum Ziel gesetzt haben, so beschäftigt sich beispielsweise [ACOH11] eingehend mit der GoodRelations Ontologie und [DZFJ05] mit der FOAF-Ontologie. Für die Untersuchungen in [ACOH11] wurde aus 105 verschiedenen Quellen die GoodRelations Tripel extrahiert und in ein Triple Store geladen. Anschließend wurde mittels SPARQL-Abfragen die eigentlichen Untersuchungen durchgeführt, so wurde beispielsweise mit einer Abfrage ermittelt, wie oft die einzelnen Konzepte, welche von der GoodRelations Ontologie zur Verfügung gestellt werden, verwendet wurden.

Die Arbeit [WPH06] von Wang, Patsia und Hendler beschäftigt sich hingegen nicht mit einer Ontologie sondern analysiert die formale Beschreibung von etwa 1300 Ontologien. Hierbei wurde unter anderem erhoben, welche Sprachebenen von OWL (Lite, DL, Full) Verwendung fanden.

Des Weiteren stellt diese Arbeit auch einen Vergleich der verschiedenen Möglichkeiten zur Einbettung von RDF-Tripeln in HTML an. Auch in [Gra07] werden die Technologien verglichen, jedoch ist diese Arbeit älteren Datums und daher wurde Microdata nicht betrachtet. Graf vergleicht die Technologien in seiner Arbeit anhand von Kriterien wie beispielsweise der Namespace Unterstützung oder der Erweiterbarkeit.

In [IUBH10] wird der Crawler LDSpider vorgestellt, welcher speziell dafür entwickelt wurde das "Web of Linked Data" zu durchsuchen und die RDF-Tripel in ein Triple Store zu laden. Für den Crawl folgt LDSpider den RDF Links. Neben der nativen Unterstützung von RDF/XML-Dokumenten ist mit LDSpider auch die Extraktion von in HTML eingebetteten Tripeln möglich, dies wird durch die Verwendung eines Any23<sup>1</sup>-Servers ermöglicht. Bei der Extraktion der Tripel mittels LDSpider geht jedoch die Information verloren mit welcher Technologie die Tripel eingebettet wurden, daher konnte LDSpider für diese Arbeit nicht genutzt werden.

Außerdem beschäftigt sich diese Arbeit damit, wie Suchmaschinen die semantischen Annotationen momentan nutzen. In [STH10] wird darüber hinaus noch ein Ausblick darauf gegeben, was auf diesem Gebiet in Zukunft noch möglich ist.

In dieser Arbeit wird unter anderem der Unterschied zwischen dem Wissen, welches ein Mensch durch das Betrachten einer Webseite erlangen kann, und dem Wissen, welches eine Maschine aus den semantischen Anreicherungen ableiten kann, untersucht. In [MMZ09] hingegen wird untersucht, welches Wissen die Nutzer nachfragen und welches Wissen das

---

<sup>1</sup><http://any23.org/>

Semantische Web bietet. Mika et al. nutzten Suchmaschinenabfragen als Repräsentation für das nachgefragte Wissen. Als Grundlage für diese Analyse fanden Abfragen Verwendung, welche an die Suchmaschine Yahoo gerichtet waren. Eine Stichprobe aus diesen Abfragen wurde mittels der Yahoo Search BOSS API<sup>2</sup> erneut ausgeführt. Im Anschluss konnten dann die Suchergebnisse auf ihre semantische Annotationen hin untersucht werden.

In [AHH<sup>+</sup>10] wird der Service "Epiphany" vorgestellt, mit dem es möglich ist Webseiten automatisch mit semantischen Annotationen zu versehen. Als Grundlage für die semantische Anreicherung kann jeder beliebige RDF-Datensatz verwendet werden. Für die automatische Anreicherung wird zunächst ein RDF-Graph erstellt, welcher aus dem Inhalt der Webseite und dem RDF-Datensatz generiert wird. Anschließend kann dann eine mit RDFa annotierte Version der Webseite ausgegeben werden. In dieser Arbeit wird die Lücke zwischen dem menschenlesbaren und dem maschinenlesbaren Inhalt einer Webseite analysiert. Adrian et al. hingegen stellen eine Möglichkeit vor diese Lücke zu verkleinern.

---

<sup>2</sup><http://developer.yahoo.com/search/boss/>

# Literaturverzeichnis

- [AB08] ADIDA, B. ; BIRBECK, M.: *RDFa Primer. Bridging the Human and Data Webs.* <http://www.w3.org/TR/2008/NOTE-xhtml-rdfa-primer-20081014/>. 2008, Abruf am 07.04.2012
- [ABMH12] ADIDA, B. ; BIRBECK, M. ; MCCARRON, S. ; HERMAN, I.: *RDFa Core 1.1 - Syntax and processing rules for embedding RDF through attributes.* <http://www.w3.org/TR/2012/CR-rdfa-core-20120313/>. 2012, Abruf am 07.04.2012
- [ABMP08] ADIDA, B. ; BIRBECK, M. ; MCCARRON, S. ; PEMBERTON, S.: *RDFa in XHTML: Syntax and Processing – A Collection of Attributes and Processing Rules for Extending XHTML to Support RDF.* <http://www.w3.org/TR/2008/REC-rdfa-syntax-20081014/>. 2008, Abruf am 07.04.2012
- [ABP12] ADIDA, B. ; BIRBECK, M. ; PEMBERTON, S.: *HTML+RDFa 1.1 - Support for RDFa in HTML4 and HTML5.* <http://www.w3.org/TR/2012/WD-rdfa-in-html-20120329/>. 2012, Abruf am 07.04.2012
- [ACOH11] ASHRAF, J. ; CYGANIAK, R. ; ORIAIN, S. ; HADZIC, M.: *Open eBusiness Ontology Usage: Investigating Community Implementation of GoodRelations.* <http://events.linkedata.org/ldow2011/papers/ldow2011-paper12-ashraf.pdf>. 2011, Abruf am 07.04.2012
- [Adi08] ADIDA, B.: *hGRDDL: Bridging microformats and RDFa.* <http://imap.websemanticsjournal.org/index.php/ps/article/viewFile/134/132>. 2008, Abruf am 07.04.2012
- [AHH<sup>+</sup>10] ADRIAN, B. ; HEES, J. ; HERMAN, I. ; SINTEK, M. ; DENGEL, A.: *Epiphany: Adaptable RDFa Generation Linking the Web of Documents to the Web of Data.* [http://www.dfki.de/web/forschung/publikationen/renameFileForDownload?filename=paper.pdf&file\\_id=uploads\\_914](http://www.dfki.de/web/forschung/publikationen/renameFileForDownload?filename=paper.pdf&file_id=uploads_914). 2010, Abruf am 07.04.2012
- [All07] ALLSOPP, J.: *Microformats: Empowering Your Markup for Web 2.0.* New York, 2007
- [BBLP11] BECKETT, D. ; BERNERS-LEE, T. ; PRUDHOMMEAU, E.: *Turtle - Terse RDF Triple Language.* <http://www.w3.org/TR/2011/WD-turtle-20110809/>. 2011, Abruf am 07.04.2012
- [BHL<sup>+</sup>09] BRAY, T. ; HOLLANDER, D. ; LAYMAN, A. ; TOBIN, R. ; THOMPSON, H. S.: *Namespaces in XML 1.0.* <http://www.w3.org/TR/2009/REC-xml-names-20091208/>. 2009, Abruf am 07.04.2012
- [Bir04] BIRBECK, M.: *XHTML and RDF.* <http://www.w3.org/MarkUp/2004/02/xhtml-rdf.html>. 2004, Abruf am 07.04.2012



- [BLFM05] BERNERS-LEE, T. ; FIELDING, R. ; MASINTER, L.: *Uniform Resource Identifier (URI): Generic Syntax*. <http://www.ietf.org/rfc/rfc3986.txt>. 2005, Abruf am 09.04.2012
- [BLHL01] BERNERS-LEE, T. ; HENDLER, J. ; LASSILA, O.: *The Semantic Web*. <http://www.jeckle.de/files/tblSW.pdf>. 2001, Abruf am 07.04.2012
- [DH98] DAWSON, F. ; HOWES, T.: *vCard MIME Directory Profile*. <http://www.ietf.org/rfc/rfc2426.txt>. 1998, Abruf am 09.04.2012
- [DZFJ05] DING, L. ; ZHOU, L. ; FININ, T. ; JOSHI, A.: *How the Semantic Web is Being Used: An Analysis of FOAF Documents*. [http://ebiquity.umbc.edu/\\_file\\_directory\\_/papers/120.pdf](http://ebiquity.umbc.edu/_file_directory_/papers/120.pdf). 2005, Abruf am 07.04.2012
- [ES07] EILEBRECHT, K. ; STARKE, G.: *Patterns kompakt - Entwurfsmuster für effektive Software-Entwicklung*. 2. Aufl. Heidelberg, 2007
- [GB04] GUHA, R. V. ; BRICKLEY, D.: *RDF Vocabulary Description Language 1.0: RDF Schema*. <http://www.w3.org/TR/2004/REC-rdf-schema-20040210/>. 2004, Abruf am 07.04.2012
- [Gra07] GRAF, A.: *RDFA VS. MICROFORMATS A COMPARISON OF INLINE METADATA FORMATS IN (X)HTML*. <http://citeseerx.ist.psu.edu/viewdoc/download;jsessionid=6DC0C70F2C23495C4B8FBF8CF17E5F97?doi=10.1.1.86.8800&rep=rep1&type=pdf>. 2007, Abruf am 07.04.2012
- [Hic12] HICKSON, I.: *HTML Microdata*. <http://www.w3.org/TR/2012/WD-microdata-20120329/>. 2012, Abruf am 09.04.2012
- [HKRS08] HITZLER, P. ; KRÖTZSCH, M. ; RUDOLPH, S. ; SURE, Y.: *Semantic Web : Grundlagen*. 1. Aufl. Berlin, 2008
- [HKTH12] HICKSON, I. ; KELLOGG, G. ; TENNISON, J. ; HERMAN, I.: *Microdata to RDF - Transformation from HTML+Microdata to RDF*. <http://www.w3.org/TR/2012/NOTE-microdata-rdf-20120308/>. 2012, Abruf am 07.04.2012
- [IUBH10] ISELE, R. ; UMBRICH, J. ; BIZER, C. ; HARTH, A.: *LDspider: An Open-source Crawling Framework for the Web of Linked Data*. <http://ceur-ws.org/Vol-658/paper495.pdf>. 2010, Abruf am 07.04.2012
- [KPSR<sup>+</sup>09] KRÖTZSCH, M. ; PATEL-SCHNEIDER, P. F. ; RUDOLPH, S. ; HITZLER, P. ; PARSIA, B.: *OWL 2 Web Ontology Language Primer*. <http://www.w3.org/TR/2009/REC-owl2-primer-20091027/>. 2009, Abruf am 07.04.2012
- [McC12] MCCARRON, S.: *XHTML+RDFa 1.1 - Support for RDFa via XHTML Modularization*. <http://www.w3.org/TR/2012/CR-xhtml-rdfa-20120313/>. 2012, Abruf am 07.04.2012
- [MM04] MANOLA, F. ; MILLER, E.: *RDF Primer*. <http://www.w3.org/TR/2004/REC-rdf-primer-20040210/>. 2004, Abruf am 07.04.2012
- [MMZ09] MIKA, P. ; MEIJ, E. ; ZARAGOZA, H.: *Investigating the Semantic Gap through Query Log Analysis*. <http://www.hugo-zaragoza.net/academic/pdf/iswc09-querylog.pdf>. 2009, Abruf am 07.04.2012
- [PS08] PRUD'HOMMEAUX, E. ; SEABORNE, A.: *SPARQL Query Language for RDF*. <http://www.w3.org/TR/2008/REC-rdf-sparql-query-20080115/>. 2008, Abruf am 07.04.2012
- [Spo12] SPORNY, M.: *RDFa Lite 1.1*. <http://www.w3.org/TR/2012/CR-rdfa-lite-20120313/>. 2012, Abruf am 07.04.2012

- [STH10] STEINER, T. ; TRONCY, R. ; HAUSENBLAS, M.: *How Google is using Linked Data Today and Vision For Tomorrow*. <http://CEUR-WS.org/Vol-700/Paper5.pdf>. 2010, Abruf am 07.04.2012
- [WPH06] WANG, T. D. ; PARSIA, B. ; HENDLER, J.: *A Survey of the Web Ontology Landscape*. <http://www.mindswap.org/papers/2006/survey.pdf>. 2006, Abruf am 07.04.2012