

RDF basierte semantische Annotationsansätze

Art der Arbeit

Betreuer:	
vorgelegt von:	
	Straße
	PLZ Ort
	Telefonnummer
	E-Mail
Matrikelnummer:	
Bearbeitungszeitraum:	
Abgabetermin:	•••

Inhaltsverzeichnis

F	rmelverzeichnis	\mathbf{V}
A	kürzungsverzeichnis	VI
1	Einleitung	1
	1.1 Motivation	1
	1.2 Methodik	1
	1.3 Aufbau	2
2	RDF Grundlagen	3
	2.1 Entwicklung	3
	2.2 Beschreibung	3
3	Linked Data und Serialisierung	6
	3.1 Semantic Web	6
	3.2 Linked Data	7
	3.2.1 Serialisierung	8
4	Schlussbemerkungen	10
	4.1 Zusammenfassung	10
	4.2 Kritische Würdigung	10
	4.3 Ausblick	10
Li	eraturverzeichnis	IX
\mathbf{E}	lesstattliche Erklärung	ΚII

Abbildungsverzeichnis

$\mathbf{Abb.}\ 1$ Kernkonzepte des RDFs $\ \ldots\ \ldots\ \ldots$.						4
$\mathbf{Abb.}\ 2$ Linked Data mit RDF $\ \ldots\ \ldots\ \ldots$.						7
Abb. 3 Karte und Zugehörige Autoren						8

1	1	1					•		1		•	
ar	el	ш	en	\mathbf{v}	$\mathbf{e}_{\mathbf{l}}$	r7	e_1	C	h	n	1	S

Tab. 1 Die Entwicklung von RDF	Tab.	1 Di∈	e Entwicklung	von RDF																						٠
--------------------------------	------	--------------	---------------	---------	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	---

Formelverzeichnis

Abkürzungsverzeichnis

1. Einleitung

1.1. Motivation

Im Zuge der Globalisierung sind viele Unternehmen vom hohen Wettbewerbsdruck betroffen <todo: "Handbuch Interorganisationssysteme noch zitieren». Um ihren Ertrag langfristig zu sichern und innovativ gegenüber der Konkurrenz zu bleiben, bieten viele Industrieunternehmen "Value Added Services", um die steigende Dienstleistungsnachfrage ihrer Kunden zu befriedigen (vgl. Meffert u. a. 2015, S. 4). Das Erbringen dieser Dienstleistungen erfordert Geschäftsprozesse die zwischenbetrieblich ausgerichtet sind (vgl. Fleisch 2001, S.19). Im Kontext einer vernetzten Welt (2013, vgl.) ist es von besonderer Bedeutung, dass zwischenbetriebliche Daten- und Funktionsintegration auch erfolgen kann, wenn Daten über das World Wide Web ausgetauscht werden.

1.2. Methodik

Das Ziel der vorliegenden Arbeit ist ein Literatur-Review durchzuführen, um die relevante Literatur über das Resource Description Framework auszuwählen, zu analysieren und zusammenzufassen im Hinblick auf den möglichen Einsatz von RDF für semantische Annotation von Daten innerhalb kleiner und mittelständischer Unternehmen. Das Review wird anhand der Methodik von Fettke durchgeführt. Eine strukturierte Literatursuche wurde mithilfe der Richtlinien von Webster und Watson, S. XVI umgesetzt, um eine Liste von relevanten Literatur zu erstellen.

- 1. Im ersten Schritt wurden EBSCOHost, Google Scholar und W3C Recommendations verwendet, um eine Literaturliste zu erstellen.
- 2. Danach wurden alle Quellen der Literaturliste untersucht, um weitere relevante Literatur zu finden (sogenannte "Go Backward" Ansatz).
- 3. Schließlich wurden Google-Scholar und Semantic-Scholar benutzt, um weitere Publikationen zu identifizieren, die auf der Literaturliste von Schritt eins und zwei verweisen (sogenannte "Go Forward" Ansatz).

Die folgenden Suchschlüsseln wurden während der Literatursuche benutzt:

• Primärschlüssel: Resource Description Framework, RDF, JSON-LD, RDFa

• Sekundärschlüssel: semantic annotation, semantic web, linked data

Zum Schluss wurde die Liste der relevanten Literatur mit Blick auf der Zielstellung verarbeitet und nach wichtige Konzepte systematisiert (vgl. Webster und Watson 2002, S. XVI).

1.3. Aufbau

Nach der Einleitung wird die Entwicklung von RDF im zweiten Abschnitt kurz beschrieben, und die benötigte RDF-Grundlagen und zusammenhängende Begrifflichkeiten werden herausgearbeitet. Die Syntax von RDF wird auch anhand einiger Beispiele demonstriert. Das Konzept des Semantic Webs und der Linked Data wird im dritten Abschnitt vorgestellt. Das Resource Description Framework in Attributes (RDFa) wird als Serialisierungssyntax ausgewählt um die Konzepte der semantischen Annotation im Web of Data exemplarisch zu zeigen. Im letzten Abschnitt werden die Kernkonzepte der Arbeit zusammengefasst und es wird beschrieben, welche Weiterentwicklungen künftig von Interesse sein könnten.

2. RDF Grundlagen

2.1. Entwicklung

Das Resource Description Framework (RDF) wurde ursprünglich 1999 vom World Wide Web Consortium (W3C) als Empfehlung verabschiedet (Lassila und Swick 1999). 2004 wurde diese Version aktualisiert und in mehrere Dokumente aufgeteilt (Beckett 2004). Die aktuelle, erweiterte Version (RDF 1.1) wurde 2014 veröffentlicht (Schreiber und Raimond 2014). RDF 1.1, im Weiteren nur als RDF bezeichnet sofern nicht anders festgelegt, hat das Ziel die Unterstützung neuer Anwendungsfelder für RDF zu stärken. Tabelle 1 zeigt die Erweiterung von RDF 1.0 auf RDF 1.1 (vgl. Klyne und Carroll, Abs. 2; Schreiber und Raimond, Abs. 2; Wood). Davon kann man ableiten, dass RDF sich in die Richtung entwickelt, immer mehr domänenspezifische Anwendung zu unterstützen. Das bedeutet, dass kleinere Unternehmen mit vereinfachte JSON basierte REST-APIs Datenintegration¹ auf die syntaktischen und semantischen Ebene anhand RDF wirksam einsetzen können.

Anwendungsfall	RDF 1.0	RDF 1.1
Web-Metadaten	RDF/XML	HTML5+RDFa 1.1, JSON-LD
Datenaustausch zw. Datenbanken ²	RDF/XML	JSON-LD, TriG, N-Quads
API-Feeds Verbin- den ²	RDF/XML	JSON-LD, RDF/XML

Tab. 1 Die Entwicklung von RDF

2.2. Beschreibung

Laut Schreiber und Raimond in RDF 1.1 Primer:

RDF is intended for situations in which information on the Web needs to be processed by applications, rather than being only displayed to people. RDF provides a common framework for expressing this information so it can be exchanged between applications without loss of meaning.

 $^{^{1}\}mathrm{vgl.}$ die semiotische Ebenen der Integration in Schissler u. a.

²vgl. Klyne und Carroll, Abs. 2; Schreiber und Raimond, Abs. 2; Wood

Wie der Name vermuten lässt, bietet das Resource Description Framework ein Gerüst (Modell, Sprachen und Syntaxen) für die Beschreibung von Attributen, Funktionen und Beziehungen der Ressourcen. Ressourcen können alles sein, was einen einzigartigen Identifier (URI oder IRI) hat (vgl. De Keyzer u. a. 2013, Folie 6). RDF legt eine abstrakte Syntax fest um Zusammenhänge zwischen Ressourcen als eine Menge gerichteter Graphen darzustellen. Ein gerichteter Graph (genannt "Triple" in RDF) hat eine Knote (Subjekt), die über eine gerichtete Kante (Prädikat) mit einer anderen Knote (Objekt) in Verbindung steht. Abbildung 2 veranschaulicht die abstrakte Syntax von RDF. Das Beispiel³ drückt die Beziehung zwischen einem Kartograph und eine gezeichnete Karte aus.

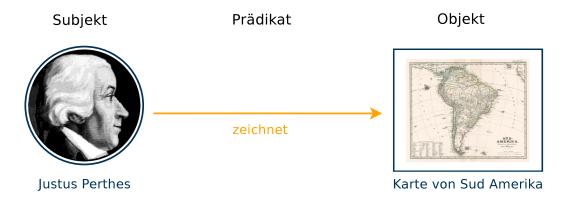


Abb. 1 Kernkonzepte des RDFs

Jeder Komponent eines RDF-Tripels kann ein von drei Ausprägungen haben (vgl. Wood u. a. 2014, Abs. 3.1).

- Ein Subjekt ist ein IRI oder ein Blank Node.
- Ein Prädikat ist ein IRI.
- Ein Objekt ist entweder ein IRI, ein Literale oder ein Blank Node.

Ein IRI (International Resource Identifier, festgelegt in RFC 3987) ist eine Generalisierung eines URIs (Uniform Resource Indicator, RFC 3986), die Zeichen aus der nicht-ASCII Bereich der Universal Character Set⁴ erlauben (vgl. Schreiber und Raimond, Abs 3.2; Duerst und Suignard). IRIs können die Eigenschaften haben, durch die Internet und World Wide Web (TCP/IP + DNS + HTTP) global eindeutig und "dereferenzierbar" zu sein (vgl. Jacobs und Walsh 2004, Abs. 2).

³Das Datenmodell basiert auf einer Web-Applikation (Globmaplab), die für die Arbeit mit den historischen Beständen der Sammlung Perthes konzipiert wurde. Offenlegung: der Autor dieser Arbeit war Lead-Entwickler dieses Projekts.

⁴wie in Norm ISO/IEC 10646 und in "The Unicode Standard" vorgegeben ist

Ein *Literal* ist eine konkrete Ausprägung eines Datentyp (wie z. B. eine Zeichenkette, Zahl oder Datum) und kein IRI. Ein String-Literal kann wahlweise mit einem "language tag" assoziiert sein um die beinhaltete Sprache zu kennzeichnen. RDF Literals können alle Datentypen, die in der XML Schema Definition Language⁵ definiert sind, verwenden (vgl. Wood u. a. 2014, Abs. 5).

Blank Nodes sind von IRIs und Literals disjunkt. Diese Abgrenzung macht es möglich, Ressourcen die keine URI haben in RDF abzubilden.

⁵http://www.w3.org/TR/xmlschema11-2/

3. Linked Data und Serialisierung

3.1. Semantic Web

Einer der Vorreiter des World Wide Webs Tim Berners-Lee hat zusammen mit anderen Autoren 2001 ein Artikel im Scientific American Journal mit dem Titel, "The Semantic Web. A new form of Web content that is meaningful to computers will unleash a revolution of new possibilities" publiziert. In diesem Artikel haben die Autoren der Konzept des Semantic Webs eingeführt und eine Vision der Zukunft formuliert: "The Semantic Web will bring structure to the meaningful content of Web pages, creating an environment where software agents roaming from page to page can readily carry out sophisticated tasks for users."

Um diese Vision gerecht zu werden ist nach Blumauer und Pellegrini der Begriff "Semantic Web" genauer als "Semiotic Web" zu verstehen. Aus Sicht der Semiotik setzt dieser Art von Interoperabilität zwischen Akteure im Semantic Web voraus, dass sie sich auf die syntaktischen, semantischen und pragmatischen Ebene verständigen können. Das heißt, wenn ein Sender eine Nachricht zum Empfänger Schickt, ist der Empfänger in der Lage die Nachricht richtig zu lesen (Syntax), zu interpretieren (Semantik), und schließlich richtig darauf zu reagieren (Pragmatik) (Voigtmann und Zeller 2002, vgl.).

RDF unterstützt die Kommunikation auf die syntaktischen Ebene und auf die semantischen Ebene kommen Ontologien zum Einsatz. Laut May, S. 488 kann man Ontologie im Kontext des Semantic Webs wie folgt charakterisieren:

Eine Ontologie beschreibt Wissen über Konzepte und ihre Zusammenhänge so, dass z. B. einerseits eine Klassifizierung eines Objektes anhand dessen Eigenschaften möglich ist, und andererseits aus dem Wissen über die Konzeptzugehörigkeit eines Objektes weitere Schlüsse über das Objekt und Beziehungen zu seiner Umwelt möglich sind.

Erst wenn geeignete Ontologien existieren und die beteiligte Akteure sich an festgelegte Standards halten ist es möglich auf einer Nachricht richtig zu reagieren (Pragmatik). Dennoch ist Integration auf diesen pragmatischen Ebene aufgrund der stetigen, wachsenden Anzahl an Standards (z. B. ebXML, RosettaNet, Biztalk, etc.) und häufig ändernde Geschäftsprozesse in der globalisierten Wirtschaft für viele KMUs mit zu hohen Kosten verbunden (vgl. Rebstock u. a. 2008, S. 4ff).

3.2. Linked Data

In sein 2006 erschienen Artikel hat Tim Berners-Lee das Semantic Web, nach De Wilde, etwas beschiedener formuliert als Linked Data. Linked Data zielt darauf hin, ein "Web Of Data" zu schaffen indem es vier Prinzipien festlegt um Inhaltsinhaber zu ermuntern, ihre Datensätze untereinander im World Wide Web zu verlinken (Berners-Lee 2006):

- 1. Use URIs as names for things
- 2. Use HTTP URIs so that people can look up those names.
- 3. When someone looks up a URI, provide useful information, using the standards (RDF*, SPARQL)
- 4. Include links to other URIs. so that they can discover more things.

Abbildung 2 zeigt eine erweiterte Darstellung des Datenmodells der ersten Abbildung. Der lila RDF Graph links, der Justus Perthes abbildet, und der grün RDF Graph rechts, der die gezeichnete Karte abbildet, nutzen URIs von bekannten Ontologien, die die vier Prinzipien der Linked Data implementieren.

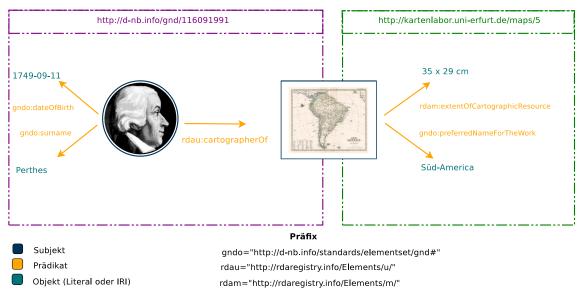


Abb. 2 Linked Data mit RDF

3.2.1. Serialisierung

Mit der Freigabe von RDF 1.1 wurden vier nicht XML-basierten Serialisierungssyntaxen eingearbeitet (vgl. Wood 2014, Abs. 3). In dieser Arbeit werden Resource Description Framework in Attributes (RDFa) Syntax und JSON Linked Data (JSON-LD) behandelt. Die beiden Serialisierungssyntaxen ermöglichen semantische Annotation (vgl. Unterabschnitt 3.2 für Definition) in Anwendungsfelder wo es vorher mit klassichen RDF/XML nicht ideal war. Semantische Annotation ist nach Reif, S. 405f, "den Prozess des Hinzufügens von semantischen Meta-Daten zu Dokumenten, die den Inhalt eines Dokuments in maschinen-verarbeitbarer Form beschreiben".

RDFa macht es möglich maschinenlesbare Metadaten in Webseiten einzubinden indem es neue HTML-Attributen für diesen Zweck festlegt. Diese "Anreicherung" der Metadaten einer Website führt dazu, dass Fremdsoftware in der Lage sind die Metadaten der Webseite automatisch verarbeiten zu können, und dass Suchmaschinen eine gezielter Darstellung des Websiteinhalts für Suchergebnisse anbieten können (vgl. Schreiber und Raimond 2014, Abs. 2), wobei der letztere Punkt insbesondere dann der Fall ist, wenn die Metadaten an weit verbreitete Ontologien und Vokabulare angeglichen werden. Eine RDFa Serialisierungsmöglichkeit für das Datenmodell in Abbildung 3 wurde in Listing 1 veranschaulicht.

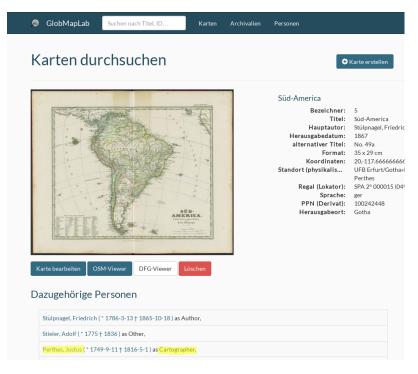


Abb. 3 Karte und Zugehörige Autoren (Screenshot vom Globmaplab Webseite)

```
<body prefix="gndo: http://d-nb.info/standards/elementset/gnd#</pre>
        rdau: http://rdaregistry.info/Elements/u/"
2
  >
3
4
  <div resource="http://kartenlabor.uni-erfurt.de/maps/5"</pre>
       typeof="http://schema.org/Map">
6
    <h4 property="gndo:preferredNameForTheWork">Süd-America</h4>
    <dt>Bezeichner:</dt>
8
    <dd>5</dd>
9
10
  </div>
11
12
  <h3>Dazugehörige Personen</h3>
13
    14
      15
16
        <div resource="http://d-nb.info/gnd/116091991"</pre>
17
             typeof="gndo:cartographer">
18
          19
              resource="http://kartenlabor.uni-erfurt.de/maps/5">
20
            >
21
              Perthes, Justus ( * 1749-9-11 † 1816-5-1 )
22
              as Cartographer
23
            24
          25
        </div>
26
27
```

Listing 1 Datenmodell in RDFa

Auf die erste Zeile im Listing 1 sieht man das prefix Attribut, das zwei Abkürzungen zu bekannten Ontologien festlegt. Das Attribut gndo verweist auf die Gemeinsame Normdatei (GND) Ontologie der Deutsche National Bibliothek und rdau auf das Resource Description & Access (RDA) Vokabular. Zeile 5 bis 10 bildet das HTML Definition List in der rechten Hälfte der Abbildung 3 ab. Mittels RDFa kann man zum Ausdruck bringen, dass die Definitionsliste eine Karte (typeof="http://schema.org/Map") mit einem IRI Bezeichner innerhalb die Uni-Erfurt Domain und dem Namen "Süd-America" beschreibt. Zeile 13 bis 29 stellt die Tabelle unten in Abbildung 3 dar. Hier wird die Ressource Justus Perthes (http://d-nb.info/gnd/116091991) mit der Süd-America Karte verlinkt, indem er die Rolle des Kartographs übernimmt.

4. Schlussbemerkungen

4.1. Zusammenfassung

Die weiterentwicklung von RDF in den letzten zehn Jahren hat den Ansatz von RDF in einer Vielfalt von Anwendungsszenario ermöglicht. Ein möglichen Anwendungsszenario, nämlich das Hinzufügen von Metadaten im Webseiten mittels RDFa, wurde durch faktische Datensätze von das Globmaplab Web-Applikation erörtert. Dabei wurden die vier Prinzipien der Linked Data demonstriert. Schließen sollte diese Arbeit den Leser die grundlegendene Konzepte der RDF basierten semantischen Annotationssprachen vermittelt haben.

4.2. Kritische Würdigung

Aufgrund Zeitmangel war es nicht möglich Vergleiche zwischen den neuen Serialisierungssyntaxen von RDF 1.1, wie zum Beispiel zwischen JSON-LD und RD-Fa, herzustellen oder Anwendungsansätze mit JSON-LD zu zeigen. Darüber hinaus wurde das dritte Prinzip der Linked Data "When someone looks up a IRI, provide useful information, using the standards (RDF*, SPARQL)" strenggenommen im Abbildung 2 verletzt. Die Prädikat zwischen Justus Perthes und der Karte von Sud-Amerika drückt eine "ist Kartograph von" Beziehung aus, aber der IRI http://rdaregistry.info/Elements/u/cartographerOf ist tatsächlich nicht dereferenzierbar da es nur einen lexikalischen Alias des kanonischen IRIs⁶ ist. Aus illustrative Grunden wurde aber den lexikalischen IRI benutzt.

4.3. Ausblick

Wie im Abschnitt "Kritische Würdigung" erwähnt wurde, war es nicht möglich JSON-LD basierte semantische Annotationsansätze zu zeigen. Nach bestem Wissen wurde es bisher noch nicht erforscht ob den Ansatz von JSON-LD die überbetriebliche Datenintegration zwischen KMUs, die JSON basierte REST-APIs über XML-basierte Webservices bevorzugen, unterstützen könnten⁷. Solcher Ansatz von JSON-LD könnte im Zukunft eine vielversprechende Entwicklung sein.

⁶http://rdaregistry.info/Elements/u/P60654

⁷Vgl. Benslimane u. a. weshalb solcher Ansatz für KMUs von Interesse sein könnten.

Literaturverzeichnis

- Beckett D (2004), RDF/XML Syntax Specification (Revised), W3C Recommendation, http://www.w3.org/TR/2004/REC-rdf-syntax-grammar-20040210/, W3C.
- Benslimane D, Dustdar S, Sheth A (2008) Services Mashups: The New Generation of Web Applications. IEEE Internet Comput. 12(5): 13–15.
- Berners-Lee T (2006) Linked Data Design Issues. ().
- Berners-Lee T, Hendler J, Lassila O (2001) The Semantic Web. A new form of Web content that is meaningful to computers will unleash a revolution of new possibilities. Scientific American 284(5): 1–5.
- Blumauer A, Pellegrini T (2006), "Semantic Web und semantische Technologien: Zentrale Begriffe und Unterscheidungen", German, In: *Semantic Web* Pellegrini T, Blumauer A (Hrsg), X.media.press, Springer Berlin Heidelberg, S. 9–25.
- De Keyzer M, Loutas N, Goedertier S (2013), Einführung in RDF & SPARQL.
- De Wilde M (2015), "From Information Extraction to Knowledge Discovery: Semantic Enrichment of Multilingual Content with Linked Open Data", (unveröffentlichte Dissertation), Diss., Université libre de Bruxelle.
- Duerst M, Suignard M (2005), Internationalized Resource Identifiers (IRIs), RFC 3987, http://www.rfc-editor.org/rfc/rfc3987.txt, RFC Editor.
- Fettke DWIP (2006) State-of-the-Art des State-of-the-Art. Wirtschaftsinformatik 48(4): 257–266.
- Fleisch E (2001) Das Netzwerkunternehmen: Strategein und Prozesse zur Steigerung der Wettbewerbsfähigkeit in der "Networked economy" Springer-Verlag.
- Jacobs I, Walsh N (2004), Architecture of the World Wide Web, Volume One, W3C Recommendation, http://www.w3.org/TR/2004/REC-webarch-20041215/, W3C.
- Klyne G, Carroll J (2004), Resource Description Framework (RDF): Concepts and Abstract Syntax, W3C Recommendation, http://www.w3.org/TR/2004/REC-rdf-concepts-20040210/, W3C.

- Lassila O, Swick RR (1999), Resource Description Framework (RDF) Model and Syntax Specification, W3C Recommendation, http://www.w3.org/TR/1999/REC-rdf-syntax-19990222/, W3C.
- May W (2006), "Reasoning im und für das Semantic Web", German, In: Semantic Web Pellegrini T, Blumauer A (Hrsg), X.media.press, Springer Berlin Heidelberg, S. 485–503.
- Meffert H, Bruhn M, Hadwich K (2015) Dienstleistungsmarketing Springer Science + Business Media.
- Monitoring-Report Digitale Wirtschaft 2013: Digitalisierung und neue Arbeitswelten (2013).
- Rebstock M, Janina F, Paulheim H (2008) Ontologies-based business integration Springer Science & Business Media.
- Reif G (2006), "Semantische Annotation", German, In: Semantic Web Pellegrini T, Blumauer A (Hrsg), X.media.press, Springer Berlin Heidelberg, S. 405–418.
- Schissler M, Zeller T, Mantel S (2004), "Überbetriebliche Integration von Anwendungssystemen: Klassifikation von Integrationsproblemen und -lösungen", In: Überbetriebliche Integration von Anwendungssystemen Bartmann D, Mertens P, Elmar J. S (Hrsg), FORWIN-Tagung, Aachen, S. 1–20.
- Schreiber G, Raimond Y (2014), *RDF 1.1 Primer*, W3C Note, http://www.w3.org/TR/2014/NOT rdf11-primer-20140624/, W3C.
- Voigtmann P, Zeller T (2002), "Enterprise Application Integration und B2B Integration im Kontext von Electronic Business und Elektronischen Marktplätze Teil I: Grundlagen und Anforderungen", In: Überbetriebliche Integration von Anwendungssystemen, FORWIN-Bericht-Nr.: FWN-2002-013, Nürnberg, S. 16.
- Webster J, Watson RT (2002) Analyzing the past to prepare for the future: Writing a literature review. Management Information Systems Quarterly 26(2): 3.
- Wood D (2014), What's New in RDF 1.1, W3C Note, http://www.w3.org/TR/2014/NOTE-rdf11-new-20140225/, W3C.

Wood D, Lanthaler M, Cyganiak R (2014), RDF 1.1 Concepts and Abstract Syntax, W3C Recommendation, http://www.w3.org/TR/2014/REC-rdf11-concepts-20140225/, W3C.

Eidesstattliche Erklärung

Ich versichere an Eides statt durch meine Unterschrift, dass ich die vorstehende Arbeit selbstständig und ohne fremde Hilfe angefertigt und alle Stellen, die ich wörtlich oder annähernd wörtlich aus Veröffentlichungen entnommen habe, als solche kenntlich gemacht habe, mich auch keiner anderen als der angegebenen Literatur oder sonstiger Hilfsmittel bedient habe. Die Arbeit hat in dieser oder ähnlicher Form noch keiner anderen Prüfungsbehörde vorgelegen.

Ilmenau,	Tag. Monat Jahr	
		Unterschrift