

# RDF basierte semantische Annotationsansätze

Art der Arbeit

Betreuer: ...

vorgelegt von: ...  
Straße  
PLZ Ort  
Telefonnummer  
E-Mail

Matrikelnummer: ...

Bearbeitungszeitraum: ... - ....

Abgabetermin: ...

# Inhaltsverzeichnis

<b>Formelverzeichnis</b>	<b>V</b>
<b>Abkürzungsverzeichnis</b>	<b>VI</b>
<b>1 Einleitung</b>	<b>1</b>
1.1 Motivation . . . . .	1
1.2 Methodik . . . . .	1
1.3 Aufbau . . . . .	2
<b>2 RDF Grundlagen</b>	<b>3</b>
2.1 Entwicklung . . . . .	3
2.2 Beschreibung . . . . .	3
<b>3 Linked Data und Serialisierung</b>	<b>5</b>
3.1 Semantic Web . . . . .	5
3.2 Linked Data . . . . .	6
3.2.1 Serialisierung . . . . .	6
<b>4 Schlussbemerkungen</b>	<b>8</b>
4.1 Zusammenfassung . . . . .	8
4.2 Kritische Würdigung . . . . .	8
4.3 Ausblick . . . . .	8
<b>Eidesstattliche Erklärung</b>	<b>IX</b>
<b>Literaturverzeichnis</b>	<b>IX</b>

# Abbildungsverzeichnis

<b>Abb. 1</b> Kernkonzepte des RDFs . . . . .	4
<b>Abb. 2</b> Linked Data mit RDF . . . . .	6
<b>Abb. 3</b> Karte und Zugehörige Autoren . . . . .	7

# Tabellenverzeichnis

<b>Tab. 1</b> Die Entwicklung von RDF . . . . .	3
---	---

## **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.][S. 4]Meffert2015. Das Erbringen dieser Dienstleistungen erfordert Geschäftsprozesse die zwischenbetrieblich ausgerichtet sind [vgl.][S.19]fleisch2001netzwerkunternehmen. Im Kontext einer vernetzten Welt [vgl.]bmwi2013 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 fettke2006state durchgeführt. Eine strukturierte Literatursuche wurde mithilfe der Richtlinien von [S. XVI]webster2002analyzing 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 Suchschlüssel 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.][S. XVI]webster2002analyzing.

### **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:99:RMS. 2004 wurde diese Version aktualisiert und in mehrere Dokumente aufgeteilt Beckett:04:RSS. Die aktuelle, erweiterte Version (RDF 1.1) wurde 2014 veröffentlicht Schreiber:14:RP. 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. [Abs. 2]Klyne:04:RDF; [Abs. 2]Schreiber:14:RP; Wood:14:WNR). 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 Datenintegration<sup>1</sup> 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 <sup>2</sup>	RDF/XML	JSON-LD, TriG, N-Quads
API-Feeds Verbinden <sup>2</sup>	RDF/XML	JSON-LD, RDF/XML

**Tab. 1** Die Entwicklung von RDF

### 2.2. Beschreibung

Laut **Schreiber:14:RP** in **Schreiber:14:RP**

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.

---

<sup>1</sup>vgl. die semiotische Ebenen der Integration in **Schissler2004**

<sup>2</sup>vgl. [Abs. 2]Klyne:04:RDF; [Abs. 2]Schreiber:14:RP; Wood:14:WNR

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.][Folie 6]Dekeyzer2013. RDF legt eine abstrakte Syntax fest um Zusammenhänge zwischen Ressourcen als eine Menge gerichteter Graphen darzustellen. Ein gerichteter Graph (genannt „Triple“ in RDF) hat einen Knoten (Subjekt), der über eine gerichtete Kante (Prädikat) mit einem anderen Knoten (Objekt) in Verbindung steht. Abbildung 2 veranschaulicht die abstrakte Syntax von RDF. Das Beispiel<sup>3</sup> drückt die Beziehung zwischen einem Kartograph und eine gezeichnete Karte aus.

**Abb. 1** Kernkonzepte des RDFs

Jede Komponente eines RDF-Tripels kann eine von drei Ausprägungen haben [vgl.][Abs. 3.1]Wood:

- 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 Set<sup>4</sup> erlauben (vgl. [[Abs 3.2]Schreiber:14:RP; [rfc3987]). IRIs können die Eigenschaften haben, durch die Internet und World Wide Web (TCP/IP + DNS + HTTP) global eindeutig und „dereferenzierbar“ zu sein [vgl.][Abs. 2]Jacobs:04:AWW.

Ein *Literal* ist eine konkrete Ausprägung eines Datentypen (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 Language<sup>5</sup> definiert sind, verwenden [vgl.][Abs. 5]Wood:14:RCA.

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

---

<sup>3</sup>Das 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.

<sup>4</sup>wie in Norm ISO/IEC 10646 und in „[The Unicode Standard](#)“ vorgegeben ist

<sup>5</sup><http://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 **berners2001semantic** ein Artikel im Scientific American Journal mit dem Titel, **berners2001semantic** 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 **blumauer2006semantic** 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)[vgl.]voightmann2002enterprise.

RDF unterstützt die Kommunikation auf die syntaktischen Ebene und auf die semantischen Ebene kommen Ontologien zum Einsatz. Laut **may2006semantic** 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.][S. 4ff]rebstock2008ontologies.

### 3.2. Linked Data

In sein 2006 erschienen Artikel hat Tim Berners-Lee das Semantic Web, nach **dewilde2015inform** 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 **berners2006linked**:

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.][Abs. 3]Wood:14:WNR. 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 klassischen RDF/XML nicht ideal war. Semantische Annotation ist nach **reif2006semantic** „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.][Abs. 2]Schreiber:14:RP, 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)

**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 Kartographen 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 RDFa, 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 referenzierbar da es nur einen lexikalischen Alias des kanonischen IRIs<sup>6</sup> ist. Aus illustrative Gründen 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 könnten<sup>7</sup>. Solcher Ansatz von JSON-LD könnte im Zukunft eine vielversprechende Entwicklung sein.

---

<sup>6</sup><http://rdaregistry.info/Elements/u/P60654>

<sup>7</sup>Vgl. Benslimane2008 weshalb solcher Ansatz für KMUs von Interesse sein könnten.

## **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