

RDF basierte semantische Annotationsansätze

Art der Arbeit

Betreuer: ...

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

Matrikelnummer: ...

Bearbeitungszeitraum: ... -

Abgabetermin: ...

Inhaltsverzeichnis

Formelverzeichnis	V
Abkürzungsverzeichnis	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
2.2.1 Serialisierung	5
3 RDF und Linked Data	7
Literaturverzeichnis	IX
Eidesstattliche Erklärung	XI

Abbildungsverzeichnis

Abb. 1 Kernkonzepte des RDFs	4
---	---

Tabellenverzeichnis

Tab. 1 Die Entwicklung von RDF	3
---	---

Formelverzeichnis

Abkürzungsverzeichnis

1. Einleitung

1.1. Motivation

Im Zuge der Globalisierung sind viele Unternehmen vom hohen Wettbewerbsdruck betroffen (cmt: Handbuch Interorganisationssysteme). Um ihren Ertrag langfristig zu sichern und innovativ gegenüber der Konkurrenz zu bleiben, bieten viele Industrieunternehmen, zum Beispiel „Value Added Services“, um die steigende Dienstleistungsnachfrage ihrer Kunden zu befriedigen (vgl. Meffert u. a. 2015, S. 4). Das Erbringen dieser Dienstleistungen heutzutage erfordert Geschäftsprozesse die zwischenbetrieblich ausgerichtet sind (vgl. Fleisch 2001, S.19). Im Kontext einer vernetzten Welt (2013, vgl.) ist es von besonderer Bedeutung, dass Datenintegration während der Unterstützung zwischenbetrieblicher Tätigkeiten auch erfolgen kann, wenn diese Daten über das World Wide Web ausgetauscht werden. Dies kann insbesondere Hilfreich sein um Daten- und Funktionsintegration zwischen kleinen und mittelständischen Unternehmen zu unterstützen, im Fall das die Implementierung von Electronic Data Interchange (EDI) Alternativen zu teuer und unrentabel geschätzt würden. (cmt: empty cite)

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 kleine und mittelständische Unternehmen. Das Review wird anhand der Methodik von Fettke durchgeführt. Eine strukturierte Literatursuche wurde mithilfe der Leitlinien 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 Suchschlüssel wurden während der Literatursuche benutzt:

- Primärschlüssel: `Resource Description Framework`, `RDF`, `JSON-LD`, `RDFa`
- Sekundärschlüssel: `semantic annotation`, `semantic web`, `linked data`

Außerdem wurden die „Key Figures“ in der RDF und Semantic Web Szene, die auf Social Media wie Twitter (z. B. @lanthaler) aktiv sind, verfolgt um auf neue Entwicklungen und Literatur zu stoßen. 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. Der dritte Abschnitt erläutert die Rolle von RDF in Linked-Data-Anwendungen. Das Konzept des Semantic-Webs wird vorgestellt und das Resource Description Framework in Attributes (RDFa) wird als exemplarischer Ansatz beschrieben. Ferner wird die Anwendung von Linked-Data erweitert, indem neue Ansätze gezeigt werden. Die Rolle von JSON-Linked-Data (JSON-LD) als alternative RDF-Serialisierungssyntax, um REST APIs erweiterbar zu gestalten, wird erkundet. 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 domänen-spezifische 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 Verbinden 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.

¹vgl. 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 1 veranschaulicht diesen Aspekt. Das Beispiel³ drückt die Beziehung zwischen einem Kartograph und seinen produzierten Werken aus.

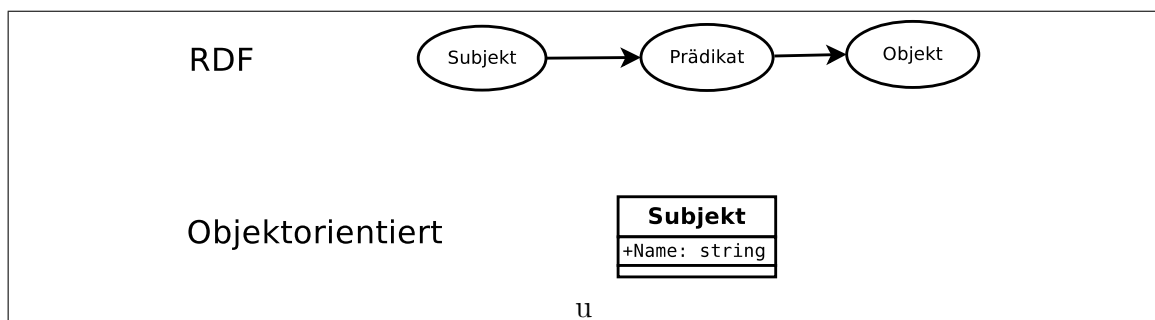


Abb. 1 Kernkonzepte des RDFs anhand eines Beispielgraphs

Die konkrete RDF-Serialisierungssyntaxen werden im Unterunterabschnitt 2.2.1 erörtert. 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 Raymond, 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).

Ein *Literale* ist eine konkrete Ausprägung eines Datentyp (wie z. B. eine Zeichenket-

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.

te, 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.

2.2.1. 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. Abschnitt 3 für Definition) in Anwendungsfelder wo es vorher mit klassischen 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 Web-Seiten 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 (vgl. Abschnitt 3). Eine RDFa Serialisierungsmöglichkeit für das Datenmodell in Abbildung 1 wurde in Listing 1 veranschaulicht.

```
1 <html>
2 <body>
3 <p>test</p>
4 </body>
5 </html>
```

Listing 1 Datenmodell in RDFa

⁵<http://www.w3.org/TR/xmlschema11-2/>

⁵wie in Norm ISO/IEC 10646 und in „[The Unicode Standard](#)“ vorgegeben ist

JSON-LD wurde ursprünglich entwickelt mit der Absicht die Interoperabilität von Webservices für Linked Data (vgl. Definition in Abschnitt 3) auszubauen und Linked Data in JSON-basierte Datenbanksysteme abzuspeichern (vgl. Lanthaler u. a. 2014, Abs. 1).⁶ Die Möglichkeit Linked Data (genauer gesagt RDF-Graphen) in der JavaScript Object Notation⁷ (JSON) zu serialisieren lässt eine reibungslose Upgrade-Pfad für eingesetzte Systeme zu. Beispiel X veranschaulicht wie die existierende JSON-basierte REST-APIs des Globmaplabs mit JSON-LD als RDF-Graphen ausgedruckt werden können.

⁶Vincent zeigt auch wie JSON-LD benutzt werden kann um Webseiteninhalt semantisch zu annotieren.

⁷vgl.

3. RDF und Linked Data

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 und Vokabularen zum Einsatz. Nach May, S. 488 kann man Ontologie im Kontext der 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.

Obwohl ist nach S.17 kein einheitliche Verwendung des Begriffes Ontologie im Kontext des Semantic Webs gibt, wird der gängige Unterschied (laut W3C Vocabularies) in dieser Arbeit (klassifizierung und seine Beziehungen) benutzt. Dieser hsd Vision hat sich nicht durchgesetzt ist der Fokus zurückgegangen um 2006 hat Tim Berners-Lee das Konzept von Linked Data introduced. Linked Data beschreibt eine nach De Wilde, S. 61 bescheidene Version des Semantic Webs (auch manchmal „Web of Data“ genannt) indem Web Konzepte und RDF Standards benutzt werden um disparate Datasets im Netz mit einander dezentralisiert verknüpft werden können. der sagt eine Maschine kann das Semantic Web (auch Web of Data im Beitrag bezeichnet) und vier Regeln (auch als Erwartungen) gesetzt:

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.

Um sicher zu stellen das der Begriff einer URI von einem unfassener "Community" von Anwender gleich ist (Linked Data „Regel“ #1), ist der Ansatz von Vokabulare (oder Ontologien) essenziell. Es gibt eine umfassende Zahl von Vokabulare die im Web benutzt werden können, die am häufigsten benutzten laut LOD <footnote> sind FOAF, SKOS, Dublincore. Da es eine vielfältige Menge an Vokabulare gibt, es ist wichtig das sie miteinander Älignedßind (soweit es geht) um die Interopabilität zu gewähren<cmt: Semiotik Discussion> <cite>. Schema.org ist auch ein Vokabulare die von den große Suchmaschineanbieter zusammenentwickelt wurde und stellt eine vereinfachte Vokabulare dar. <cite: nächste Abbildung> Als Beispiel dazu, hat das Globmaplab⁸ seine Datenbestände weitgehend an das GND-Ontologie älnedünd mithilfe von RDFa und JSON-LD seine HTML und JSON APIs mit Meta-Daten versehen.se Möglichkeit se Möglichkeit

⁸<http://kartenlabor.uni-erfurt.de>

Literaturverzeichnis

- Beckett D (2004), *RDF/XML Syntax Specification (Revised)*, W3C Recommendation, <http://www.w3.org/TR/2004/REC-rdf-syntax-grammar-20040210/>, W3C.
- 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 Keyser 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 Bruxelles.
- 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: Strategiein 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.
- Lanthaler M, Sporny M, Kellogg G (2014), *JSON-LD 1.0*, W3C Recommendation, <http://www.w3.org/TR/2014/REC-json-ld-20140116/>, 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).

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/NOTE-rdf11-primer-20140624/>, W3C.

The JSON Data Interchange Format (2013), Standard ECMA-404.

Vincent J (2015), *Video SEO Just Got Better: Meet Wistia's New Standard Embed*.

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