

Das Resource Description Framework (RDF) - ein neuer Weg zur Verwaltung von Metadaten im Netz

Eckhardt Schön
Technische Universität Ilmenau
Fakultät für Informatik und Automatisierung
Institut für Praktische Informatik und Medieninformatik
Postfach 0565, D-98684 Ilmenau
schoen@prakinf.tu-ilmenau.de

Zusammenfassung

Es wird erläutert, warum die Definition eines Standards für Metadaten in Computernetzen, besonders im Internet, dringend notwendig ist. Als aktueller Vorschlag des W3-Consortiums wird das Resource Description Framework (RDF) mit seinem Datenmodell und seiner auf XML beruhenden Syntax vorgestellt. Ein ausführliches Beispiel demonstriert die Möglichkeiten des RDF.

1 Notwendigkeit

Das Wissen der Menschheit wächst nicht nur ständig, sondern liegt auch in einer zunehmenden Anzahl von Ressourcentypen vor. Waren es früher vor allem Bücher, in denen Informationen gespeichert waren, so sind es heute auch verschiedenartige Textdateien und Multimediaobjekte. Dabei geht die Tendenz dahin, daß das World Wide Web zum weltweiten Wissensspeicher mit einer rasant anwachsenden Zahl von Dokumenten wird. Die Informationsanbieter, wie Bibliotheken, Verlage, Firmen oder Institutionen sind zunehmend über das Internet miteinander verbunden. Die Menge des potentiell verfügbaren Wissens wächst also enorm, die Suche nach einer bestimmten Information wird aber immer schwieriger und aufwendiger.

Abhilfe können Metadaten, d.h. Informationen über Informationen, schaffen, wie sie in Form von Katalogen in Büchereien oder als Indexdateien schon lange bekannt sind. Will man jedoch Metadaten weltweit vernetzter Informationen bereitstellen, treten eine ganze Reihe von Schwierigkeiten auf. So weisen diese Daten keine einheitliche Struktur auf, was mit unterschiedlichen Sprachen, Ressourcentypen und auch Traditionen zu tun hat. Früher hatte jedes Archiv sein eigenes Klassifikations- und Beschreibungsschema. Derartige Katalogdaten sind damit kaum international austauschbar und nicht maschinenverständlich, selbst wenn sie in digitalisierter Form vorliegen.

In einem Zeitalter, in dem die schnelle und zielgerichtete Beschaffung von Informationen immer wichtiger wird, besteht dringender Bedarf nach einem einheitlichen und effizienten Management von Metadaten. Ein Standard sollte folgende Anforderungen erfüllen:

- Die Metadaten müssen (maschinell) durch Applikationen verarbeitet werden können.
- Sie sollten trotzdem auch von einem Menschen gelesen werden können, also in Textform vorliegen.
- Die Metadaten müssen für den Austausch über Computernetze, insbesondere das Internet, geeignet sein.

- Sie sollen durch verschiedenartige Applikation nutzbar sein, wie z.B. Suchmaschinen, digitale Bibliotheken, Softwareagenten oder Datenbanken.
- Die Metadaten-Verwaltung sollte nicht auf eine Anwendungsdomäne (Bibliothek, WWW, Musik-Datenbank, Videoarchiv, ...) zugeschnitten, sondern domänenneutral sein.
- Der Austausch von Klassifizierungsschemata soll unterstützt und eine Erweiterung problemlos möglich sein.

2 Das RDF-Datenmodell

Das World Wide Web Consortium (W3C) hatte vor ca. zwei Jahren eine Arbeitsgruppe eingesetzt, die eine Lösung für das Problem der Metadaten liefern sollte. Diese hat im vorigen Jahr einen ersten Entwurf für das Resource Description Framework (RDF) vorgelegt. Dieser wurde inzwischen diskutiert und mehrfach überarbeitet. Das aktuelle Working Draft vom 19.08.98 soll der zu erwartenden Spezifikation schon sehr nahe kommen und dient deshalb als Grundlage für diese Darstellung.

Das Resource Description Framework stellt eine Infrastruktur zur Codierung, zum Austausch und zur Wiederverwendung von Metadaten zur Verfügung. Dabei wurden verschiedene Entwicklungen in Hinblick auf Metadaten aufgegriffen, so u.a.

- Dublin Core Initiative (CD),
- Platform for Internet Content Selection (PICS) (W3C seit 1995),
- Meta Content Framework Using XML (vorgelegt von Netscape),
- Web Collections using XML (vorgelegt von Microsoft).

Wie die Bezeichnung Framework schon andeutet, ist das RDF nichts Starres, sondern bietet einen erweiterbaren Rahmen für die Behandlung von Metadaten. Die Erweiterbarkeit wird u.a. dadurch erreicht, daß die vorgeschlagene Syntax auf der eXtensible Markup Language (XML) beruht (siehe [1] und [2]), die eine eigene Definition von Tags erlaubt. Das Datenmodell, das hier zunächst beschrieben werden soll, ist jedoch syntaxunabhängig.

Die Aufgabe des RDF ist die Beschreibung von Ressourcen, worunter Objekte verstanden werden, die Informationen enthalten (Texte, Webseiten, Videos, ...). Die Ressourcen werden durch ihre URI (Uniform Resource Identifier) identifiziert. Sie haben benannte Eigenschaften mit Werten.

Ein einfaches Beispiel demonstriert das. Die beiden Sätze „*Der Autor des Dokumentes <http://www.x.net/doc1.html> ist I. R. Gendwer.*“ bzw. „*I. R. Gendwer ist der Autor des Dokumentes <http://www.x.net/doc1.html>.*“ sind für einen Menschen gleichwertig, eine Maschine würde jedoch jeweils ein anderes Objekt als Ressource festlegen. Im RDF-Datenmodell weist deshalb ein Pfeil von der Ressource zum Wert der Eigenschaft. Dieser Pfeil wird durch den Eigenschaftstyp benannt.

[Resource] — PropertyType → Value

Für das Beispiel bedeutet das:

[<http://www.x.net/doc1.html>] — Autor → „I. R. Gendwer“

Auf diese Weise kann die Eigenschaft einer Ressource eindeutig beschrieben werden. Es wird, wie bei Metadaten üblich, zwischen dem Typ einer Eigenschaft (*PropertyType* oder *Attribut* genannt) und dem Wert unterschieden. Der Typ ist bei gleichartigen Ressourcen derselbe,

während sich die Werte i.a. unterscheiden. Mehrere Eigenschaften (Properties) einer Ressource werden zu einer Beschreibung (Description) zusammengefaßt.

Leider sind Verhältnisse oft komplizierter. So ist der Wert des Eigenschaftstyps *Autor* i.a. nicht nur ein String, sondern ein Objekt (Knoten, Node), das selbst wieder Eigenschaften hat und damit bezüglich dieser Eigenschaften die Rolle einer Ressource übernimmt.

[<http://www.x.net/doc1.html>] — Autor → [Autor_001]

[Autor_001] — Name → „I. R. Gendwer“

[Autor_001] — Mail → irgendwer@x.net

[Autor_001] — Phone → 12 34 56 78

Autor_001 ist eine ID, die die Person des Autors eindeutig beschreibt. Auf diese beziehen sich dann die Eigenschaften.

Das RDF definiert auch Konstrukte für erweiterte Relationen, bei denen z.B. einer Eigenschaft mehrere Werte (geordnet, ungeordnet oder alternativ) zugewiesen werden oder bei denen der Wert selbst wieder eine Eigenschaft ist. Es kann damit ein recht kompliziertes Geflecht aus Ressourcen und deren Eigenschaften entstehen. Das Beispiel in Abschnitt 4 wird das etwas andeuten.

Das Resource Description Framework verfolgt hinsichtlich der Eigenschaftstypen eine Art Klassenkonzept, wie es aus objektorientierten Programmiersprachen bekannt ist. Die zu einer Domäne gehörenden Eigenschaftstypen werden in einem Schema (siehe [4]) zusammengefaßt und über das Internet zugänglich gemacht. Sie können damit für die Beschreibung vieler Ressourcen genutzt, aber auch erweitert und angepaßt werden.

Die Spezifikation des RDF sieht verschiedene Möglichkeiten vor, wie die Metadaten gespeichert werden können, sowohl innerhalb der Ressourcen (z.B. bei XML-Dateien) oder auch extern (z.B. bei Bildern).

3 Die Syntax des RDF

Obwohl das Datenmodell der Kern des Resource Description Framework ist, benötigt man für die maschinelle Verarbeitung natürlich eine geeignete Syntax. Entsprechend der oben genannten Ziele soll die Syntax nicht nur von Computern sondern auch von Menschen gelesen werden können. Darüber hinaus müssen die Metadaten leicht über das Netz ausgetauscht werden können.

Das komplexe, mehrdimensionale Datenmodell muß serialisiert werden, um den obigen Forderungen gerecht zu werden. Das World Wide Web Consortium schlägt als Sprache XML vor. Das hat verschiedene Gründe, wozu u.a. die erwartete hohe Akzeptanz und die Erweiterbarkeit zählen. In XML lassen sich bei Bedarf jederzeit neue Eigenschaften definieren, ohne daß dafür die Spezifikation geändert werden müßte.

Für obiges Beispiel lautet die Beschreibung in der Serialisationsyntax:

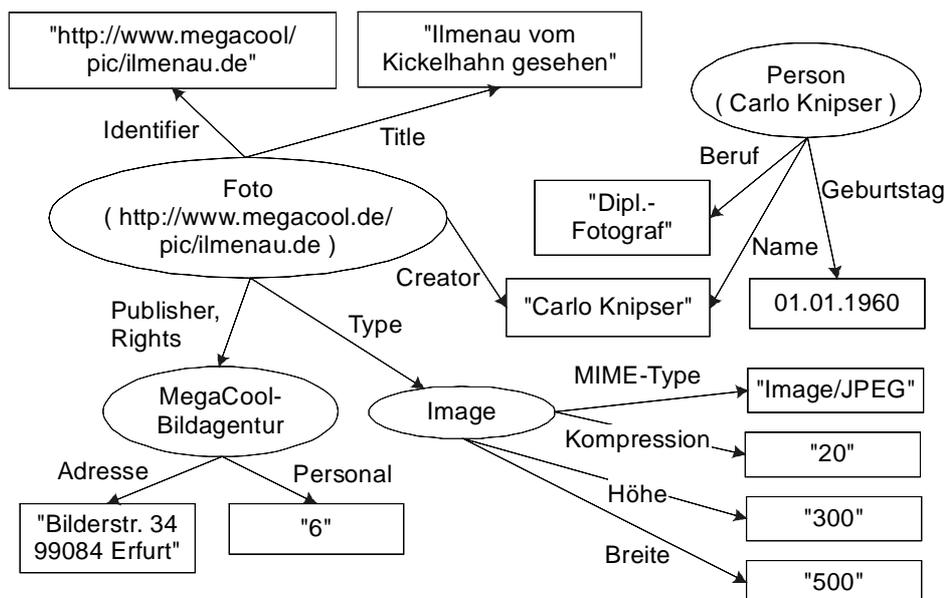
```
<?xml version="1.0"?>
<RDF:RDF
  xmlns:RDF="http://www.w3.org/RDF"
  xmlns:DC="http://metadata.net/rdf/DC">
  <RDF:Description about="http://www.x.net/doc1.html">
    <DC:Author>I. R. Gendwer</DC:Author>
  </RDF:Description>
</RDF:RDF>
```

Der Code beginnt mit einer *Processing Instruction*, die das Dokument als XML-Dokument kennzeichnet. Im darauffolgenden einleitenden RDF-Tag werden die Namensräume in Form von Attributen festlegen. Durch die Nutzung von Namensräumen (Namespace in XML, [5]) können Schemata, in denen Eigenschaftstypen definiert sind, weltweit benutzt werden. Der qualifizierter Name eines RDF-Tags besteht aus dem selbstgewählten Bezeichner für den externen Namensraum, einem Doppelpunkt und dem Elementnamen. Zum Beispiel bedeutet `<DC:Author>`, daß der PropertyType *Author* im Dublin Core Schema unter der angegebenen Adresse festgelegt ist.

Der obige Code kann bequem in eine XML-Datei eingebettet werden oder auch extern abgelegt werden. Probleme gibt es jedoch, wenn man versucht, Metadaten auf diese Weise in HTML einzubetten, der wohl noch auf längere Zeit verbreitetsten Websprache. Ein HTML-Browser kann mit den neuen Tags nicht umgehen. Standardmäßig werden unbekannte Tags nicht dargestellt, was in diesem Fall durchaus gewünscht ist. Die Metadaten sollen schließlich nicht auf dem Bildschirm dargestellt werden. Was dargestellt wird, ist in diesem Fall jedoch der Autorenname „I. R. Gendwer“. Dieser Name und gegebenenfalls weitere Werte von Eigenschaften würden von einem Browser angezeigt. Um das zu vermeiden, schlägt das W3C parallel zu dieser Syntax eine sogenannte *Abbreviated Syntax* vor, bei der alle Metadaten in Form von Attributen innerhalb der Tags formuliert werden. Damit werden sie von einem HTML-Browser bei der Darstellung komplett ignoriert. Die Metadaten sind jedoch in das HTML-Dokument eingebettet und können von geeigneten Applikationen (Suchmaschinen, Softwareagenten usw.) ausgelesen und verarbeitet werden. Obiges Beispiel erhält dann die folgende Form:

```
<RDF:RDF
  xmlns:RDF="http://www.w3.org/RDF"
  xmlns:DC="http://metadata.net/rdf/DC">
  <RDF:Description about="http://www.x.net/doc1.html"
    DC:Author="I. R. Gendwer" />
</RDF:RDF>
```

4 Beispiel



Um die Möglichkeiten des RDF wenigstens andeutungsweise zu demonstrieren, werden in diesem Beispiel die Metadaten eines Medienobjektes, genauer eines digitalisierten Fotos, betrachtet. Die Abbildung 4.1 (vorherige Seite) zeigt das Datenmodell. Hier sind nur einfache Eigenschaften von Ressourcen enthalten, wobei nicht nur das Foto [<http://www.megacool.de/pic/ilmenau.de>] sondern auch [MegaCool-Bildagentur], [Image] und [Person (Carlo Knipser)] als Ressourcen auftreten. Der Code zu diesem Datenmodell lautet:

```
<RDF:RDF
  xmlns:RDF="http://www.w3.org/RDF"
  xmlns:DC="http://metadata.net/rdf/DC"
  xmlns:MeC="http://megacool.de/rdf/MeC">
  <RDF:Description
    about="http://www.metacool.de/pic/ilmenau.jpg">
    <DC:Title>Ilmenau vom Kickelhahn gesehen</DC:Title>
    <DC:Identifizier DC:Scheme="URI">
      http://www.megacool.de/pic/ilmenau.de
    </DC:Identifizier>
    <DC:Creator>Carlo Knipser</DC:Creator>
    <DC:Type>Image</DC:Type>
    <DC:Publisher>MetaCool-Bildagentur</DC:Publisher>
    <DC:Rights>MetaCool-Bildagentur</DC:Rights>
  </RDF:Description>
  <RDF:Description ID="MegaCool-Bildagentur">
    <MeC:Adresse>Bilderstr. 34, 99084 Erfurt</MeC:Adresse>
    <MeC:Personal>6</MeC:Personal>
  </RDF:Description>
  <RDF:Description ID="Image">
    <MeC:MIME-Type>Image/JPEG</MeC:MIME-Type>
    <MeC:Kompression>20</MeC:Kompression>
    <MeC:Höhe>300</MeC:Höhe>
    <MeC:Breite>500</MeC:Breite>
  </RDF:Description>
  <RDF:Description ID="Person (Carlo Knipser)">
    <MeC:Name>Carlo Knipser</MeC:Name>
    <MeC:Beruf>Dipl.-Fotograf</MeC:Beruf>
    <MeC:Geburtstag>01.01.1960</MeC:Geburtstag>
  </RDF:Description>
</RDF:RDF>
```

Die Abbildung 4.2 zeigt erweiterte Relationen, die sich schon etwas komplizierter beschreiben lassen.

Bei der „Stichwort-Sammlung“ handelt es sich um einen Container, der vom Typ *RDF:Bag* ist und damit eine Eigenschaft mit mehreren Werten beschreibt. Diese sind in diesem Fall ungeordnet und werden lediglich durch eine ID (RDF:_1 usw.) beschrieben.

Anders liegt der Fall bei der Eigenschaft vom Typ *Motiv*. Sie hat als Wert keinen Knoten (Ressource oder String) sondern wiederum eine Eigenschaft. Diese Beziehung muß konkretisiert (reify) werden, um sie innerhalb des Datenmodells vernünftig beschreiben und die Syntax formulieren zu können. Die Konkretisierung ist in Abbildung 4.3 dargestellt.

Eine genauere Beschreibung erfolgt im Vortrag.

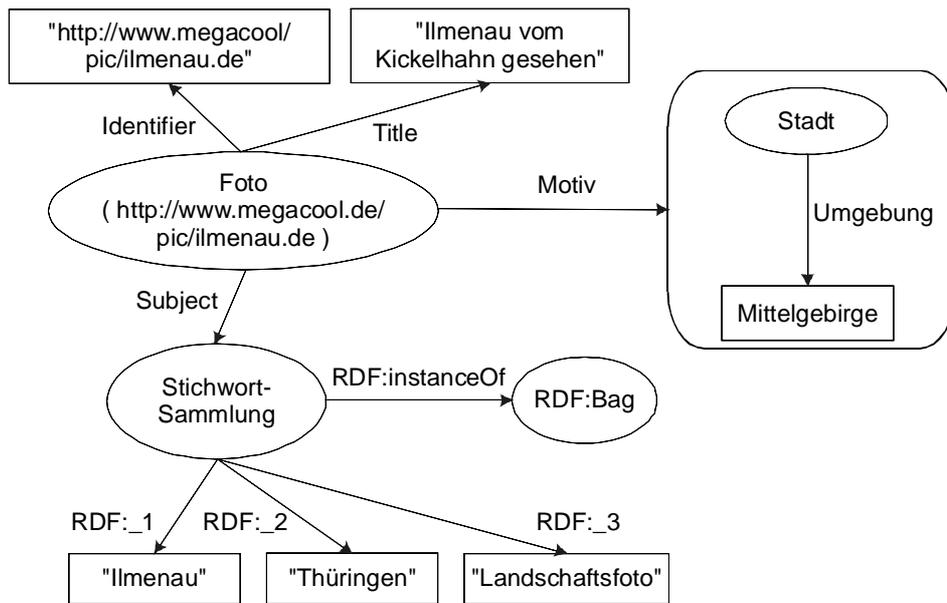


Abbildung 4.2

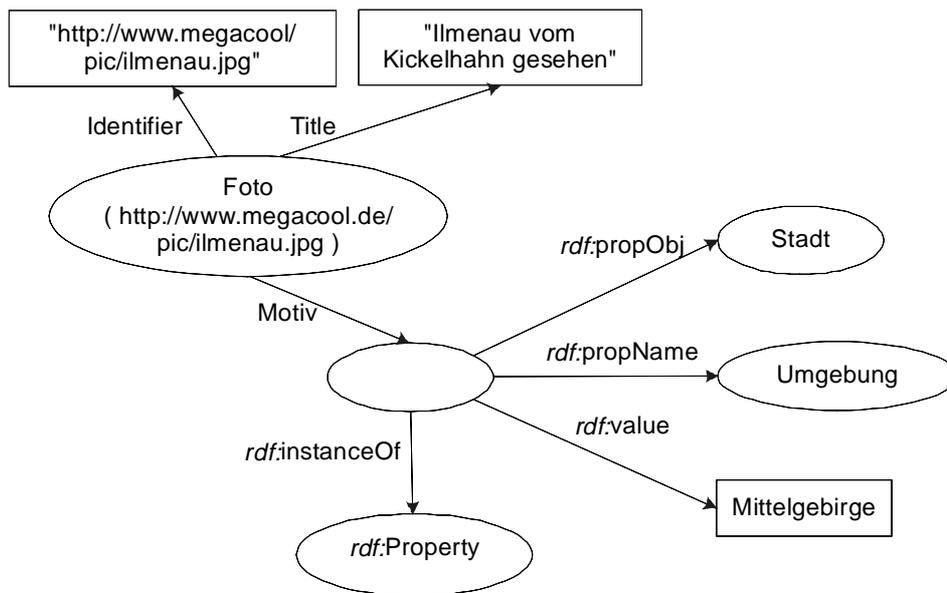


Abbildung 4.3

Die Syntax für die erweiterten Relationen dieses Beispiels lautet:

```

<RDF:RDF
  xmlns:RDF="http://www.w3.org/RDF"
  xmlns:DC="http://metadata.net/rdf/DC"
  xmlns:MeC="http://megacool.de/rdf/MeC">
  <RDF:Description
    about="http://www.metacool.de/pic/ilmenau.jpg">
    <DC:Title>Ilmenau vom Kichelhahn gesehen</DC:Title>
    ...
    <DC:Subject>

```

```

    <RDF:Bag ID="Stichwort-Sammlung">
      <RDF:LI>Ilmenau</RDF:LI>
      <RDF:LI>Th&uuml;ringen</RDF:LI>
      <RDF:LI>Landschaftsfoto</RDF:LI>
    </RDF:Bag>
  </DC:Subject>
  <MeC:Motiv ID="001"/>
</RDF:Description>
<RDF:Description ID="001">
  <RDF:propObj resource="Stadt"/>
  <RDF:propName resource="Umgebung"/>
  <RDF:value>Mittelgebirge</RDF:value>
  <RDF:instanceOf resource="RDF:Property"/>
</RDF:Description></RDF:RDF>

```

Dieses Beispiel sollte kurz andeuten, welche Potenzen im Resource Description Framework enthalten sind. Es zeigt aber auch, daß die Beschreibung komplexer Zusammenhänge zu einem komplexen Code führt, wie das bei formalisierten Sprachen die Regel ist.

5 Ausblick

"The Resource Description Framework (RDF) is used in Navigator 5 for many purposes. Its a swiss army knife and we will use it wherever it makes sense to use the rdf data model as a representation language." [9]

Diese Aussage der Entwickler der nächsten Navigator-Version, macht deutlich, daß das RDF nicht nur akademisch interessant ist, sondern sofort in der Praxis benutzt wird. Bedarf ist vorhanden, da die Dokumentenverwaltung bei der bisherigen Entwicklung des WWW insbesondere der Sprache HTML praktisch keine Rolle gespielt hat.

Jetzt wird unter Führung des World Wide Web Consortium die Arbeit mit Dokumenten im Web auf eine neue zukunftssträchtige Grundlage gestellt. Im Mittelpunkt stehen dabei die eXtensible Markup Language (XML) und das Document Object Model (DOM) [7], eine plattform- und sprachneutrale API zur Manipulation von Dokumenten. In diesen Kontext gehören neben RDF [3] und RDF-Schemata [4] auch die eXtensible Style Language (XSL) [6], Namespaces in XML [5], XLink und XML-Data.

Damit soll die Basis geschaffen werden, um das WWW zukünftig als zentrales Publikationsmedium und effizienten Informationspool nutzen zu können. Die neuen Standards sollen Einzellösungen von Firmen zuvorkommen. Sie bieten mit ihrer Flexibilität und Erweiterbarkeit hoffentlich genügend Potenzen, um auf eine breite Akzeptanz bei den Informationsanbietern zu stoßen.

Literatur

- [1] H. Behme, S. Minert *XML in der Praxis*. Addison-Wesley, Bonn, 1998
- [2] XML 1.0 Recommendation (W3C), <http://www.w3c.org/TR/1998/REC-xml-19980210>
- [3] Resource Description Framework (RDF) Model and Syntax Specification (W3C), <http://www.w3.org/TR/1998/WD-rdf-syntax-19980819>

- [4] Resource Description Framework (RDF) Schema Specification (W3C),
<http://www.w3.org/TR/1998/WD-rdf-schema-19980814>
- [5] Namespaces in XML (Working Draft des W3C),
<http://www.w3.org/TR/1998/WD-xml-names-19980802>
- [6] Extensible Stylesheet Language (XSL), Version 1.0,
<http://www.w3.org/TR/1998/WD-xsl-19980818>
- [7] Document Object Model (DOM) Level 1 Specification, Version 1.0,
<http://www.w3.org/TR/1998/PR-DOM-Level-1-19980818>
- [8] E. Miller. An Introduction to the Resource Description Framework. *D-Lib Magazine*,
Mai 1998
- [9] Guha, R. Churchill, J. Giannandrea Netscape Communications Corporation, Mozilla.Org,
<http://www.mozilla.org/rdf/doc/index.html>