


Ressourcen-Beschreibung im Semantic Web

Cristina Vertan

Inhaltsübersicht

- Wie sollen die Ressourcen für Semantic Web annotiert werden ? 
- Was ist und wie funktioniert RDF ?
- Wie kodiert man RDF-Statements in XML-Syntax ?
- Wie beschreibt man komplexere Ausdrücke in RDF ?
- Wie wenden wir RDF im Praktikum an?

Wie werden die Ressourcen zur Zeit im Web beschrieben?

- Das Web wurde zuerst für menschliche Benutzung entworfen, deswegen sind alle Ressourcen :
 - zwar maschinell lesbar (durch HTML, DHTML, XHTML)
 - Aber **NICHT maschinell verstehbar**.
- Man muss ausreichend Information geben:
 - nicht nur darüber, wie z.B. die Maschine den Inhalt einer Webseite darstellen soll (Fonts, Farben, Hintergrund, links, etc.)
 - sondern auch worüber die Webseite handelt.
- Die vorgeschlagene Lösung ist: Benutzung von Metadaten

14.04.2004

SoSe'04

3

Was sind Metadaten ?

- Metadaten sind „data about data“
- Z.B. besteht ein Bibliothekskatalog aus Metadaten, weil er die Publikationen beschreibt (besonders deutlich bei Stichwörtern).
- Es gibt keinen allgemeinen und klaren Unterschied zwischen Daten und Metadaten. Der Unterschied ist meistens anwendungsspezifisch (d.h. was in eine Anwendung Metadaten sind, sind in einen anderen nur Daten)
- Um Metadaten zu beschreiben braucht man ein Modell.
- Erste Versuch wurde mit XML gemacht.

14.04.2004

SoSe'04

4

Was ist XML ?

- XML (eXtensible Markup Language) ist ein Framework für die Definition von Markupsprachen
- Es gibt keine feste Sammlung von Markup-tags (für jede Anwendung kann man andere Tags definieren)
- Es gibt einen Pool von generischen Werkzeugen, die XML-Dokumente bearbeiten können (z.B. XML-Parser)
- XML wurde entwickelt als Informationstrukturierungs - Standard.

14.04.2004

SoSe'04

5

XML Beispiel

`<description sprache=„deutsch“>`

Das historische Zentrum von Sighisoara (Schaessburg) wurde von der UNESCO auf die Liste des Weltkulturerbes gesetzt. Und das zurecht. In der Mitte steht wie einst die mächtige Burg ,...

`</description>`

Tags

Attribut

Attributwert

Die Syntax-Korrektheit der tags wird in einer DTD bzw. in einer XML-Schema Datei beschrieben

14.04.2004

SoSe'04

6

Warum kann man mit XML keine Metadaten beschreiben ?


- Mit XML kann man den Inhalt einer Ressource beschreiben aber
- Man kann **keine Beziehungen** zwischen verschiedenen Ressourcen beschreiben.

14.04.2004

SoSe'04

7

Inhaltsübersicht

- Wie sollen die Ressourcen für Semantic Web annotiert werden ?
- Was ist und wie funktioniert RDF ? 
- Wie kodiert man RDF-Statements in XML-Syntax ?
- Wie beschreibt man komplexere Ausdrücke in RDF ?
- Wie wenden wir RDF im Praktikum an?

14.04.2004

SoSe'04

8

Was ist RDF ?

- RDF = **R**esource **D**escription **F**ramework
- RDF basiert auf der Idee, dass Sachen, die beschrieben werden sollen, Merkmale haben und diese Merkmale Werte. D.h. man kann Ressourcen beschreiben, indem man ihre Merkmale und deren Werte beschreibt.



RDF ist ein Modell mit dessen Hilfe man Merkmale und Werte von Merkmalen darstellen kann

Terminologie

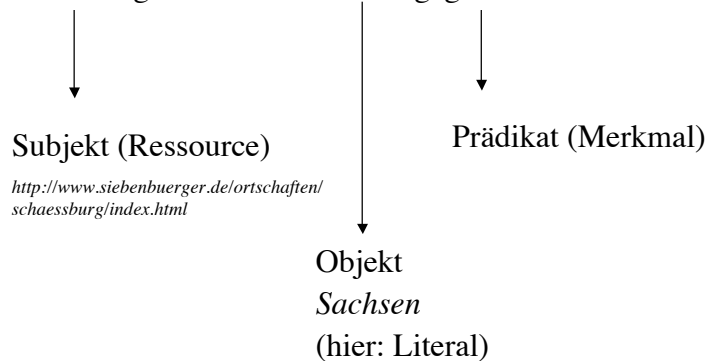
- **Ressourcen** = alles was man mit RDF -Ausdrücken beschreiben kann (z.B. eine Webseite, Teile einer Webseite, ein HTML-Element, ein Objekt, das man nicht via Web erreichen kann)
- **Merkmal** = eine spezifische Charakteristik, ein Attribut oder eine Beziehung, die für die Beschreibung einer Ressource benutzt wird. Jedes Merkmal hat eine Semantik, hat erlaubte Werte /-Bereich, und kann eine Beziehung zu anderen Merkmalen haben)
- **RDF-Ausdruck** = Eine Ressource + Ein Merkmal + der Merkmalswert

Wie identifiziert man Ressourcen ?

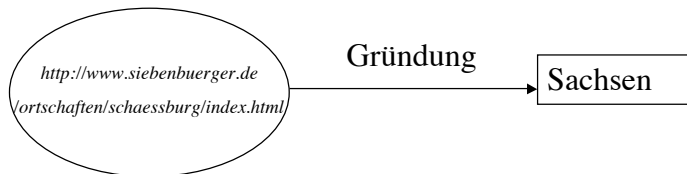
- Jede Ressource wird durch ein URI (Uniforme Resource Identifier) eindeutig identifiziert.
- Web-Ressourcen haben als URI eine URL Adresse

Wie beschreibt man mit RDF einfache Ausdrücke ?

Schässburg wurde von Sachsen gegründet.



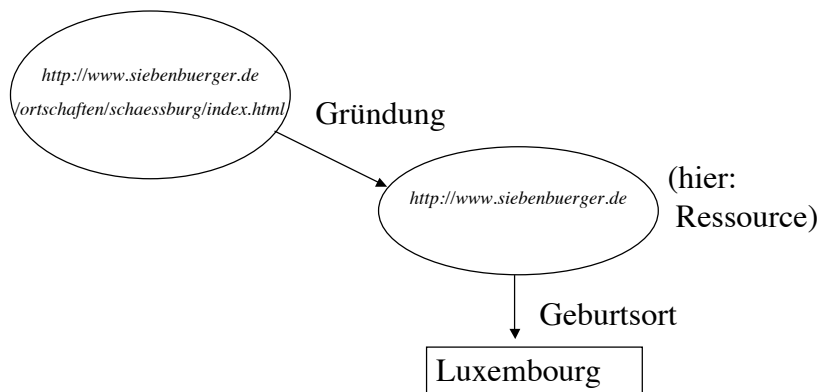
RDF-Ausdrücke als Graph



Die Pfeilrichtung ist **wichtig** : sie geht immer vom Subjekt zum Objekt

Spezifischer RDF-Ausdruck


Schässburg wurde von Sachsen aus Luxemburg gegründet.



Bei der Repräsentation mit RDF zu beachten

- Korrekte Identifizierung von Subjekt und Objekt. Man wählt als Subjekt normalerweise das logische Subjekt einer Proposition (worüber es geht) und nicht das syntaktische Subjekt.
- Z.B. : ein Satz wie :
Die Sachsen haben Schässburg gegründet,
könnte als RDF-Subjekt auch "Schässburg" haben.
- Sorgfältige Überlegung darüber, was man als Literal und was als Ressource beschreibt. Eine Ressource ist immer ein Verbindungspunkt zu anderen Ressourcen!

Inhaltsübersicht

- Wie sollen die Ressourcen für Semantic Web annotiert werden ?
- Was ist und wie funktioniert RDF ?
- Wie kodiert man RDF-Statements in XML-Syntax ? 
- Wie beschreibt man komplexere Ausdrücke in RDF ?
- Wie wenden wir RDF im Praktikum an?

Warum XML-Kodierung für RDF ?

- XML ist nur **EINE** (von mehreren möglichen) Kodierungsweisen für RDF
- XML wird benutzt weil:
 - es das Standard-Austausch-Format für Web-Ressourcen ist
 - es Möglichkeiten bietet für die eindeutige Identifizierung von Merkmalen, auch wenn von zwei Benutzern derselbe Bezeichner benutzt wird (d.h. Benutzer gruppieren ihre Merkmale in "Namespaces")

XML Syntax für RDF -1-

* mehrere
? kann fehlen
| oder

- Jeder RDF-Ausdruck hat die Form:
- `<rdf:RDF> description* </rdf:RDF>`

↓
Zeigt den Namespace,
in dem das tag "rdf"
definiert wurde

↘
Alles über die Ressource.
Normalerweise mehrere
Ausdrücke

XML Syntax für RDF -2-

* mehrere
? kann fehlen
| oder

Description:

- `<rdf:Description idAboutAttr? > propertyElt*
</rdf:Description>`

Subjekt

idAboutAttr : ID=„ein XML-erlaubtes Symbol“ (neue Ressource)

oder

idAboutAttr : ,about=„URI -reference“ (schon vorhandene Ressource)

XML Syntax für RDF -3-

* mehrere
? kann fehlen
| oder

- propertyElt
 - `<propName> description | string </ propName >`
- oder

Objekt

- `<propName resourceAttr >`

Prädikat

resource = „URI-reference“

ein XML Namespace-Präfix:Name

XML Syntax für RDF - Beispiel 1 -

- *Schässburg wurde von Sachsen gegründet.*

```
<rdf:RDF>
  <rdf: Description
    about=„http://www.siebenbuergen.de/ortschaften/schaessb
    urg/index.html“>
    <praktikum:Gruendung> Sachsen
  </praktikum:Gruendung>
</rdf:Description>
</rdf:RDF>
```

- `praktikum` ist der Benutzer-Namespace. Man definiert ein Namespace mit:
- `xmlns:praktikum=URL`

14.04.2004

SoSe'04

21

XML Syntax für RDF - Beispiel 1 Fortsetzung-

Vollständiges XML-Dokument:

```
<?xml version="1.0"?>
<rdf:RDF
  xmlns:rdf=„http://www.w3.org/1999/02/22-rdf-syntax-ns#“
  xmlns:praktikum=„http://nats-www.informatik.....“>
  <rdf: Description
    about=„http://www.siebenbuergen.de/ortschaften/schaessburg/index.html“>
    <praktikum:Gruendung> Sachsen
  </praktikum:Gruendung>
  </rdf:Description>
</rdf:RDF>
```

14.04.2004

SoSe'04

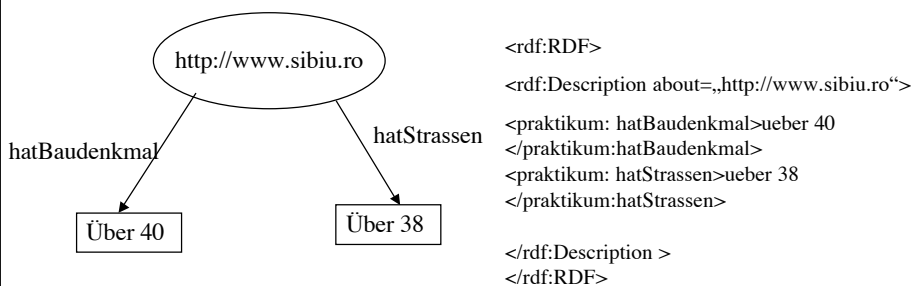
22

Inhaltsübersicht

- Wie sollen die Ressourcen für Semantic Web annotiert werden ?
- Was ist und wie funktioniert RDF ?
- Wie kodiert man RDF-Statements in XML-Syntax ?
- **Wie beschreibt man komplexere Ausdrücke in RDF ?** ←
- Wie wenden wir RDF im Praktikum an?

Darstellung multipler Prädikate (für dasselbe Subjekt)

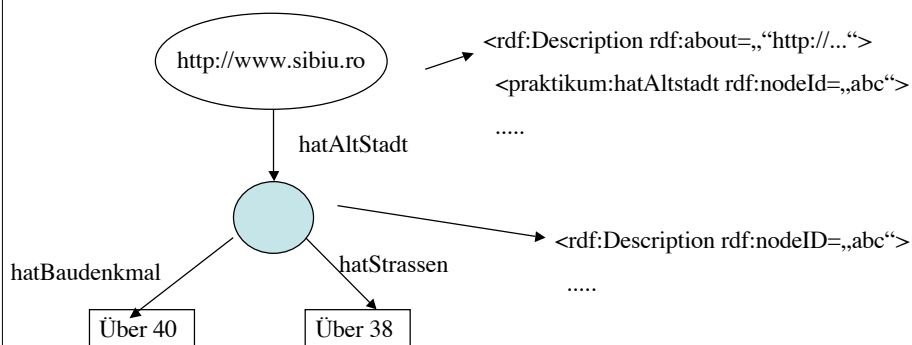
*In Sibiu gibt es ueber 40 Bauten, die zum **Baudenkmal** erklärt worden sind und ueber 38 **Strassen und Plaetze** von geschichtlichem Interesse.*



```
<rdf:RDF>
<rdf:Description about="http://www.sibiu.ro">
<praktikum: hatBaudenkmal>ueber 40
</praktikum: hatBaudenkmal>
<praktikum: hatStrassen>ueber 38
</praktikum: hatStrassen>
</rdf:Description >
</rdf:RDF>
```

Künstliche Ressourcen (blank nodes)

- Man erzeugt „blank nodes“ um Ressourcen darzustellen, die kein URI haben



14.04.2004

SoSe'04

25

Fortgeschrittene Elemente

- Man kann Ressourcen oder Literale in **Container** gruppieren:
 - **rdf:Bag**: Reihenfolge der Ressourcen ist nicht wichtig
 - **rdf:Seq**: die Reihenfolge ist festgelegt (z.B. Beschreibung von Objekten, die alphabetisch sortiert sein müssen)
 - **rdf:Alt**: Alternative für dasselbe Wert (z.B. alternative Übersetzungen für denselben Ort)
- Noch kompliziertere Strukturen kann man mit **rdf:List** aufbauen (damit kann man festlegen, dass nur diese Ressourcen zu der Gruppe gehören)

14.04.2004

SoSe'04

26

Inhaltsübersicht

- Wie sollen die Ressourcen für Semantic Web annotiert werden ?
- Was ist und wie funktioniert RDF ?
- Wie kodiert man RDF-Statements in XML-Syntax ?
- Wie beschreibt man komplexere Ausdrücke in RDF ?
- Wie wenden wir RDF im Praktikum an?



Aufgaben während des Praktikums

- Die Texte in beiden Sprachen (Deutsch und Englisch) müssen analysiert und wichtige Merkmale extrahiert werden.
- Die Analyse sollte sinnvoll auf Paragraphen/Sätzen/Wörtern-Ebene vorgenommen werden.
- Die Merkmale sollen in einem eigenen Namespace gespeichert werden.
- Man muss auch sich überlegen ob allgemeine Merkmale wie „description“, „creator“ usw. nicht direct vom "Dublin Core Metadata" übernommen werden können
- Die Merkmale auf englisch sollten nicht die Übersetzung von deutschen sein. Es wird die Aufgabe der Ontologie Gruppe sein, eine Hierarchie unter den Merkmalen zu erzeugen.
- Alle Texte müssen mit RDF-Ausdrücken im XML-Syntax annotiert werden.

Aufgaben für heute

- Je eine Gruppe von 2 Studenten sollte für einen Text (deutsch oder englisch) eine erste RDF-Annotierung erstellen.
- Die annotierten Texte sollen zunächst in Netz, als Ausgangsmaterial für die Annotierungs-Gruppe verfügbar gemacht werden.

Web-Adressen

- RDF -Syntax
<http://www.w3.org/TR/REC-rdf-syntax>
- Dublin Core Metadata
<http://dublincore.org/documents/dcmes-xml/>