### **Vorlesung Semantic Web**

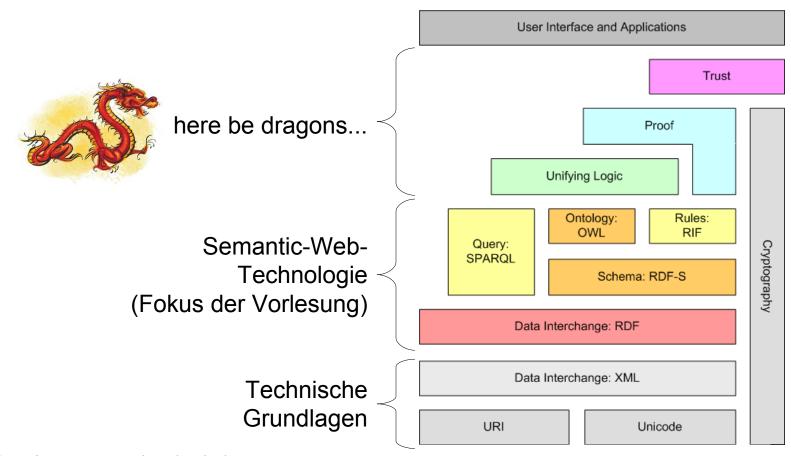


Vorlesung im Wintersemester 2012/2013 Dr. Heiko Paulheim Fachgebiet Knowledge Engineering



#### **Semantic Web - Aufbau**





Berners-Lee (2009): Semantic Web and Linked Data http://www.w3.org/2009/Talks/0120-campus-party-tbl/



#### Was bisher geschah



- RDF: eine Sprache zur Beschreibung von Informationen
  - Graphenbasiert
  - Menge von Aussagen: Subjekt, Prädikat, Objekt
  - Notationen: RDF/XML, N3
- Besondere Sprachmittel
  - Leere Knoten
  - Mengen und Listen
  - Reifikation
- Semantik
  - Non-unique name assumption
  - Open world assumption



#### Was heute passiert



- Verarbeitung von RDF
  - Speicherung
  - Darstellung
- Kombination von RDF und HTML
- Einige Beispiele von RDF in der freien Wildbahn





- RDF ist eine Menge von Tripeln
- RDF-Speicher werden daher auch "Triple Store" genannt
- Unterschiedlichste Systeme
  - native Implementierung
  - Nutzung einer relationalen Datenbank





- naive Implementierung mit einer Tabelle
  - drei String-Spalten

Subjekt	Prädikat	Objekt
<http: foo.bar="" peter=""></http:>	<http: foo.bar="" vatervon=""></http:>	<http: foo.bar="" stefan=""></http:>
<http: foo.bar="" peter=""></http:>	<http: foo.bar="" name=""></http:>	"Peter Müller"
<http: foo.bar="" peter=""></http:>	<rdf:type></rdf:type>	<http: foo.bar="" person=""></http:>
<http: foo.bar="" stefan=""></http:>	<rdf:type></rdf:type>	<http: foo.bar="" person=""></http:>
<http: foo.bar="" peter=""></http:>	<http: foo.bar="" vatervon=""></http:>	<http: foo.bar="" julia=""></http:>
<http: foo.bar="" peter=""></http:>	<http: foo.bar="" kennt=""></http:>	_:genID01
_:genID01	<http: foo.bar="" vatervon=""></http:>	<http: foo.bar="" markus=""></http:>





- Abfragemechanismus
  - "Wer ist der Vater von Julia?"

```
SELECT subjekt FROM triples
WHERE praedikat = '<http://foo.bar/vaterVon>'
AND objekt = '<http://foo.bar/Julia>'
```

Wer kennt den Vater von Julia?

```
SELECT t1.subjekt FROM triples AS t1, triples AS t2
WHERE t1.praedikat = '<http://foo.bar/kennt>'
AND t1.objekt = t2.subjekt
AND t2.praedikat = '<http://foo.bar/vaterVon>'
AND t2.objekt = '<http://foo.bar/Julia>'
```





Jede komplexe Abfrage führt zu einem Self-Join

```
SELECT t1.subjekt FROM triples AS t1, triples AS t2
WHERE t1.praedikat = '<http://foo.bar/kennt>'
AND t1.objekt = t2.subjekt
AND t2.praedikat = '<http://foo.bar/vaterVon>'
AND t2.objekt = '<http://foo.bar/Julia>'
```

t1.Subjekt	t1.Prädikat	t1.Objekt	t2.Subjekt	t2.Prädikat	t2.Objekt





- Abfragemechanismus:
  - "Wer ist Vater, Mutter, Bruder, Schwester von Julia?"

```
SELECT praedikat, objekt FROM triples
WHERE (         praedikat = '<http://foo.bar/vaterVon>'
         OR praedikat = '<http://foo.bar/mutterVon>'
         OR praedikat = '<http://foo.bar/bruderVon>'
         OR praedikat = '<http://foo.bar/schwesterVon>')
AND objekt = '<http://foo.bar/Julia>'
```

Derartige Abfragen führen zu vielen gereihten OR-Statements





- Abfragemechanismus
  - Welche Informationen gibt es über Julia?

```
SELECT subjekt,praedikat,objekt FROM triples
  WHERE subjekt='<http://foo.bar/Julia>'
SELECT subjekt,praedikat,objekt FROM triples
  WHERE objekt='<http://foo.bar/Julia>'
```





- Property Table
  - Ein Eintrag pro Subjekt (statt pro Tripel)
  - eine Spalte pro Prädikat

Subjekt	rdf:type	foo:vaterVon	foo:kennt
foo:Peter	foo:Person	foo:Stefan	NULL
foo:Peter	NULL	foo:Julia	NULL
foo:Stefan	foo:Person	NULL	_:genID01
_:genID01	foo:Person	foo:Markus	NULL





- Abfragemechanismus:
  - "Wer ist der Vater von Julia?"

```
SELECT subjekt FROM properties WHERE
foo_vaterVon = "http://foo.bar/Julia"
```

"Wer kennt den Vater von Julia?"

SELECT pl.subjekt FROM properties AS pl, properties AS p2 WHERE

```
p1.foo_kennt p2.subjekt AND
p2.foo_vaterVon = "http://foo.bar/Julia"
```





- Abfragemechanismus:
  - "Welche Informationen gibt es über Julia?"
  - Mit Julia als Subjekt ist das trivial:

```
SELECT * FROM properties WHERE
subjekt = "http://foo.bar/Julia"
```

Mit Julia als Objekt nicht:

```
SELECT 'vaterVon',foo_vaterVon FROM properties WHERE
  objekt = "http://foo.bar/Julia"
SELECT 'kennt',foo_kennt FROM properties WHERE
  objekt = "http://foo.bar/Julia"
SELECT ...
```





- Geschwindigkeit
  - kürzere Tabelle → z.T. effizientere Abfragen
  - gut für Anfragen von ausgehenden Kanten
- Speicherbedarf
  - sehr kompakte Speicherung für ähnliche Ressourcen
  - schlecht bei unähnlichen Ressourcen
    - viele NULL-Werte
    - Probleme mit Relationen von hoher Kardinalität
      - z.B. "kennt"





- Vertikale Partitionierung
  - Variante von Property-Table
  - eine Tabelle pro Property

Subjekt	rdf:type
foo:Peter	foo:Person
_:genID01	foo:Person

Subjekt	foo:vaterVon
foo:Peter	foo:Stefan
foo:Peter	foo:Julia
_:genID01	foo:Markus

Subjekt	foo:kennt
foo:Stefan	_:genID01





- Abfragemechanismus:
  - "Wer ist der Vater von Julia?"

```
SELECT subjekt FROM foo_vaterVon WHERE
    objekt = '<http://foo.bar/Julia>'
```

"Wer kennt den Vater von Julia?"





- Abfragemechanismus:
  - "Welche Informationen gibt es über Julia?"

```
SELECT subjekt,'foo:vaterVon',objekt FROM foo_vaterVon
WHERE
        subjekt = '<http://foo.bar/Julia>'
SELECT subjekt,'foo:vaterVon',objekt FROM foo_vaterVon
WHERE
        objekt = '<http://foo.bar/Julia>'
SELECT subjekt,'foo:kennt',objekt FROM foo_kennt
WHERE
...
```





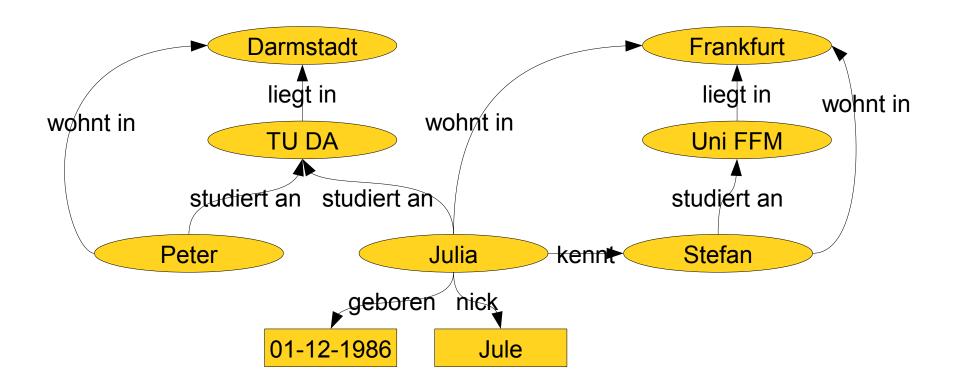
- Geschwindigkeit
  - sehr kleine Tabellen → schnelle Joins
  - dafür oft viele einzelne Abfragen nötig
- Speicherbedarf
  - kompakte Speicherung bei sehr verschiedenen Ressourcen
  - gut bei Relationen hoher Kardinalität
  - etwas schlechter für ähnliche Ressourcen



#### Beispiel



Ein sozialer Graph soll gespeichert werden





#### **Beispiel**



- Ein sozialer Graph soll gespeichert werden
  - 10000 Personen mit Wohnort, Uni, Nickname, Geburtsdatum
  - Jede Person kennt im Mittel 50 andere Personen
  - 100 Universitäten mit Ort
- Wie groß ist der Speicherbedarf?
- Wie aufwändig sind die Abfragen...
  - Finde alles über eine Person heraus
  - Erstelle eine Liste aller Nicknames und Geburtstage zu Freunden einer Person
  - Finde alle Personen, die in der Stadt wohnen, in der sie studieren



#### **Beispiel: Naiver Triple Store**



- Speicherbedarf
  - 10000 Personen mit Wohnort, Uni, Nickname, Geburtsdatum
    - → 10000 \* 4 Tripel
  - Jede Person kennt im Mittel 50 andere Personen
    - → 10000 \* 50 Tripel
  - 100 Universitäten mit Ort
    - → 100 Tripel
- Gesamt: Tripel-Tabelle mit ~540.000 Einträgen
  - 1,6 Mio Strings
  - keine Null-Werte



#### **Beispiel: Naiver Triple Store**



- Finde alles über eine Person heraus SELECT pred obj FROM triples WHERE subj="person" SELECT subj pred FROM triples WHERE obj="person"
- Erstelle eine Liste aller Nicknames und Geburtstage zu Freunden einer Person

```
SELECT tab2.obj AS nick, tab3.obj AS geburtstag
FROM triples AS tab1, triples AS tab2, triples AS tab3
WHERE
tab1.subj = "ich"; tab1.pred = "kennt";,
tab1.obj = tab2.subj, tab2.pred = "nick",
```

tab1.obj = tab3.subj, tab3.pred = "geburstag"



#### **Beispiel: Naiver Triple Store**



Finde alle Personen, die in der Stadt wohnen, in der sie studieren

```
SELECT tab1.subj
FROM tripel AS tab1, tripel AS tab2, tripel AS tab3
WHERE
tab1.pred = "studiertAn",
tab1.obj = tab2.subj, tab2.pred = "liegt_in",
tab2.obj = tab3.obj, tab3.pred = "wohntIn"
tab1.subj = tab3.subj
```



#### **Beispiel: Property Table**



- Sechs verschiedene Prädikate
- 10000 Personen: je 50 Einträge
- 100 Universitäten: je 1 Eintrag
- Tabelle mit ~500000 Einträgen
  - ~3,5 Mio. Strings, davon viele Null-Werte



#### **Beispiel: Property Table**



- Finde alles über eine Person heraus SELECT studiert\_an, geboren, nick, kennt, wohnt\_in FROM person WHERE subj="person" SELECT kennt FROM person WHERE obj="person"
- Erstelle eine Liste aller Nicknames und Geburtstage
   zu Freunden einer Person
   SELECT tab2.geburtstag, tab2.nickname FROM triples AS tab1, triples
   AS tab2 WHERE
   tab1 subi = "ich" tab1 konnt = tab2 subi

tab1.subj = "ich", tab1.kennt = tab2.subj



#### **Beispiel: Property Table**



Finde alle Personen, die in der Stadt wohnen, in der sie studieren

SELECT tab1.subj FROM tripel AS tab1, tripel AS tab2, tripel AS tab3 WHERE

```
tab1.studiertAn = tab2.subj
```

tab2.liegtIn = tab1.wohntIn



#### **Beispiel: vertikale Partitionierung**



- Tabelle wohnt\_in, studiert\_in, geboren, nick: je 10000 Einträge
- Tabelle kennt: 500000 Einträge
- Tabelle liegt\_in: 100 Einträge
- 6 Tabellen à 2 Spalten, ~540000 Einträge
- ~1,1 Mio Strings, keine Null-Werte



#### **Beispiel: vertikale Partitionierung**



Finde alles über eine Person heraus

```
SELECT nick, FROM nick WHERE subj="person"
SELECT geboren FROM geboren WHERE subj="person"
SELECT studiertAn FROM studiertAn WHERE subj="person"
SELECT wohntIn FROM wohntIn WHERE subj="person"
SELECT kennt FROM kennt WHERE subj="person"
SELECT subj FROM kennt WHERE kennt="person"
```

- Erstelle eine Liste aller Nicknames und Geburtstage zu Freunden einer Person
  - SELECT nickname.object, geburtstag.object FROM kennt, nickname, geburtstag
     WHFRF

```
kennt.subj = "ich", kennt.obj = nickname.subj,
kennt.obj = geburtstag.subj
```



#### **Beispiel: vertikale Partitionierung**



Finde alle Personen, die in der Stadt wohnen, in der sie studieren

```
SELECT studiertAn.subj FROM studiertAn, liegtIn, wohntIn WHERE
```

```
studiertAn.obj = liegtIn.subj,
liegtIn.subj = wohnortVon.subj,
wohntInSubj = studiertAn.subj
```



#### Vergleich der Lösungen



- Speicherbedarf:
  - Naiver Triple Store: 1,6 Mio Strings
  - Property Table: 3,5 Mio Strings
  - Vertikale Partitionierung: 1,1 Mio Strings
- Abfragen
  - Naiver Triple Store: große Joins, einfache "offene" Abfrage
  - Property Table: wenige Joins
  - Vertikale Partitionierung: viele Joins, schwierige "offene" Abfrage
- Fazit:
  - Trade-off zwischen Speicherplatz und Geschwindigkeit
  - Kein optimales Modell für alle Abfragen





- Gemischte Lösungen
  - aus allen vorher kennen gelernten
  - z.B. Einzeltabellen für häufig verwendete Prädikate

Subjekt		Prädikat		Objekt	
<http: foo.bar="" peter=""></http:>		<http: foo.bar="" vatervon=""></http:>		<http: foo.bar="" s<="" td=""><td>tefan&gt;</td></http:>	tefan>





- Performance-Verbesserung mit Indexierung
  - Problem: Vergleiche über URIs/Literale aufwändig
  - Lösung z.B.: Nutzen von Hashes
    - String-Vergleiche können durch int-Vergleiche ersetzt werden
  - Beispiel 3Store: URI-/Literal-Objekte werden mit Flag unterschieden

Subjekt (int6	4)	Prädikat (int64)	Objekt (int64)	Objl	₋iteral (bool)
0A7B22F834A	A043BC	423ABEF890432ABC	432FE8A8BE390A32	FAL	SE
432FE8A8BE3	hash (ir	nt64)	URI		βE
B879EA890B	•		<http: foo.bar="" peter=""></http:>		SE
	423ABE	F890432ABC	<rdf:type></rdf:type>		
	432FE8	A8BE390A32	<a href="http://foo.bar/Person:">http://foo.bar/Person:</a>	>	
			•••		





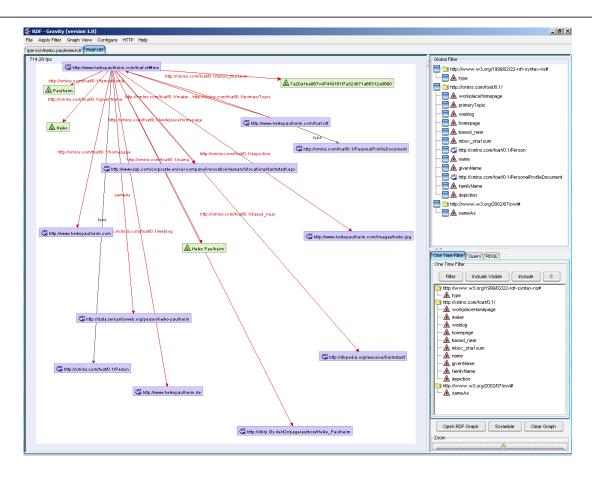
- Hashing verhindert "teure" String-Vergleiche:
  - "...WHERE Subjekt = '<http://foo.bar/Peter>'" wird zu
  - "...WHERE Subjekt = hash('<http://foo.bar/Peter>')"
  - Hash-Funktion effizient zu berechnen

Subjekt (int6	4)	Prädikat (int64)	Objekt (int64)	ObjL	₋iteral (bool)
0A7B22F834A	\043BC	423ABEF890432ABC	432FE8A8BE390A32	FAL	SE
432FE8A8BE3	hash (ir	nt64)	URI		SE
B879EA890B			<a href="http://foo.bar/Peter">http://foo.bar/Peter&gt;</a>		SE
		F890432ABC	<rdf:type></rdf:type>		
	432FE8	A8BE390A32	<http: foo.bar="" person:<="" td=""><td>&gt;</td><td></td></http:>	>	



#### **RDF-Tools: Visualisierung**





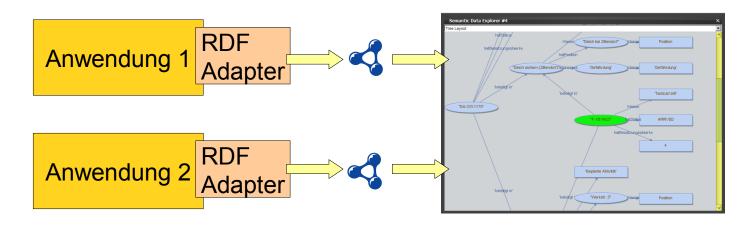
RDF Gravity http://semweb.salzburgresearch.at/apps/rdf-gravity/



#### **RDF-Tools: Visualisierung**



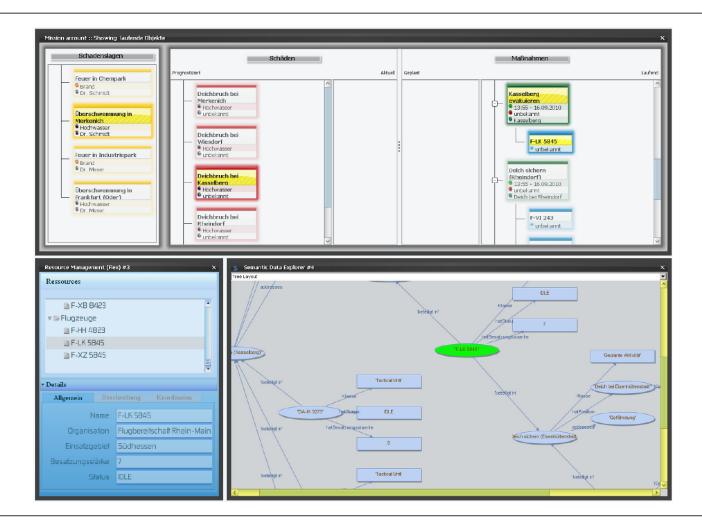
- Experiment bei SAP Research 2010:
  - Daten aus Anwendungen als RDF zur Verfügung stellen
  - Den Graph direkt anzeigen
  - Verschränkte Interaktion:
    - Objekt im Graph wählen → wird in der Anwendung hervorgehoben
    - Objekt in der Anwendung wählen → wird im Graph hervorgehoben





#### **RDF-Tools: Visualisierung**







#### **RDF-Tools: Visualisierung**



- Experiment bei SAP Research 2010:
  - Nutzer müssen verschieden komplexe Aufgaben lösen
  - einmal mit, einmal ohne RDF-Visualisierung
  - Ergebnis:
    - Für komplexe Tasks signifikant geringere Bearbeitungszeit
    - Signifikant höhere Nutzerzufriedenheit



# RDF-Tools: Validierung und Konvertierung



- W3C RDF Validator:
  - W3C RDF Validator: http://www.w3.org/RDF/Validator/
  - Gibt RDF/XML in Tripeln und als Graph aus
- RDF About Validator:
  - http://www.rdfabout.com/demo/validator/
  - kann auch N3



#### **RDF-Tools** (weitere)



- Programmierframeworks
  - zum Entwickeln RDF-basierter Anwendungen
  - z.B. JENA, RDFReactor, ...
- Reasoner
  - können Schlussfolgerungen auf RDF-Graphen ziehen
  - können Anfragen beantworten
- Werden wir noch kennen lernen



#### RDF und HTML



- Das semantische Web verwendet also RDF
- Das "klassische" Web verwendet HTML

• Muss man jetzt jede Information doppelt codieren?

```
<html>
...
    <b>Dr. Mark Smith</b>
    <i>Physician</i>
    Main St. 14
    Smalltown
    Mon-Fri 9-11 am
    Wed 3-6 pm
...
</html>
```

```
:p a :Physician .
:p :hasDegree "Dr." .
:p :hasName "Mark Smith" .
:p :hasAddress :a .
:a :street "Main Street" .
:a :number "14"^^xsd:int .
:a :city "Smalltown" .
:p :hasOpeningHours [
   a rdf:Bag ;
   [ :day :Monday;
        :from "9"^^xsd:int;
        :to "11"^^xsd:int;
]
...
```



# Gemeinsame Verwendung von HTML und RDF – Variante 1



Verweis auf RDF-Beschreibung – haben wir schon mal gesehen



</rdf:Description>

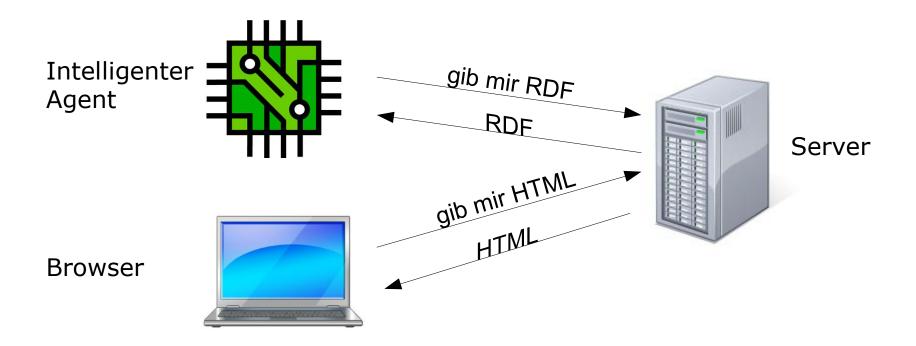
</rdf:RDF>

<dc:relation rdf:resource="http://www.w3.org/2001/sw/" />

## Parallele Verwendung von HTML und RDF – Variante 2



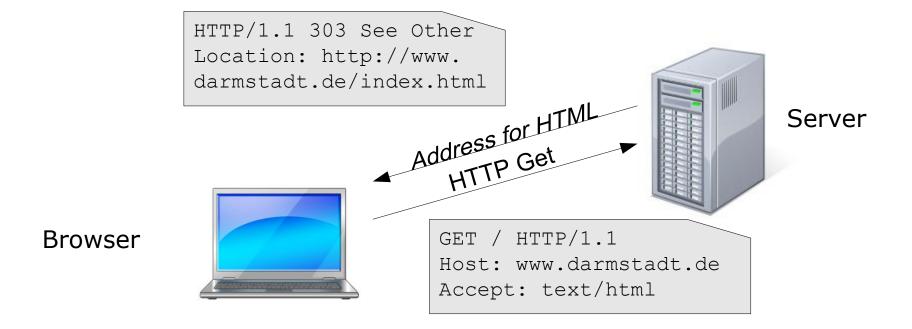
Content Negotiation





### **Content Negotiation im Detail**

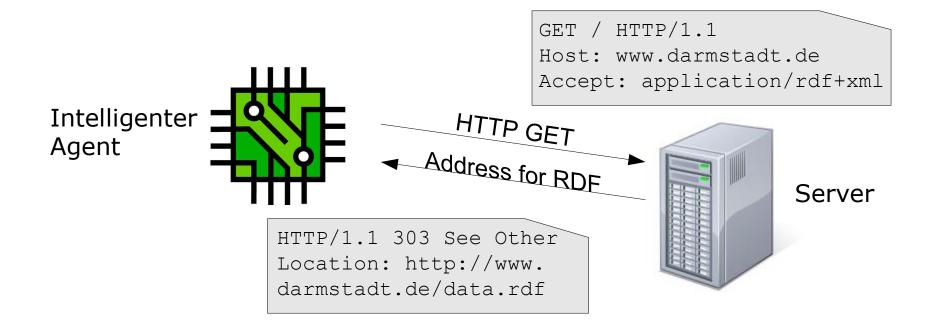






#### **Content Negotiation im Detail**







#### **Content Negotiation: MIME Types**



- MIME: Multipurpose Internet Mail Extensions
- Ursprünglich zur Kennzeichnung von Mail-Anhängen
  - Text, PDF, ...
- Erste Version: 1996
- Werden von der IANA verwaltet
- Für Semantic Web wichtige MIME Types:
  - application/rdf+xml
  - text/turtle
  - text/n3
  - application/sparql-query
  - application/sparql-results+xml



## Gemeinsame Verwendung von HTML und RDF



- Link auf RDF-Dokument
  - Umsetzung mit HTML-Bordmitteln möglich
- Content Negotiation
  - Aufwändigere Maßnahme
  - Ein URI für verschiedene Repräsentationen möglich
- In beiden Fällen
  - Zwei getrennte Repräsentationen nötig
  - "Doppelte Buchführung"
  - Potentielle Inkonsistenz





- Idee hinter RDF in Attributes (RDFa):
  - Warum nicht HTML und RDF in einem Dokument gemeinsam nutzen?
  - Information muss nur einmal codiert werden



RDFa kombiniert XHTML mit RDF



### Sprachmittel von RDFa



- about="http://foo.bar/aSubject"
  - Legt das Subjekt einer Seite oder eines Unterabschnittes fest
- property = "http://foo.bar/aProperty"
  - Definiert eine Datenrelation
  - Inhalte des Tags werden als Literal interpretiert
- rel = "http://foo.bar/aRelation"
  - Definiert eine Relation zu einer anderen Ressource
- href = "http://foo.bar/aResource"
  - Definiert das Objekt einer Relation
  - kann wieder Subjekt werden
- typeof = "http://foo.bar/aType"
  - legt den Typ (rdf:type) einer Ressource fest

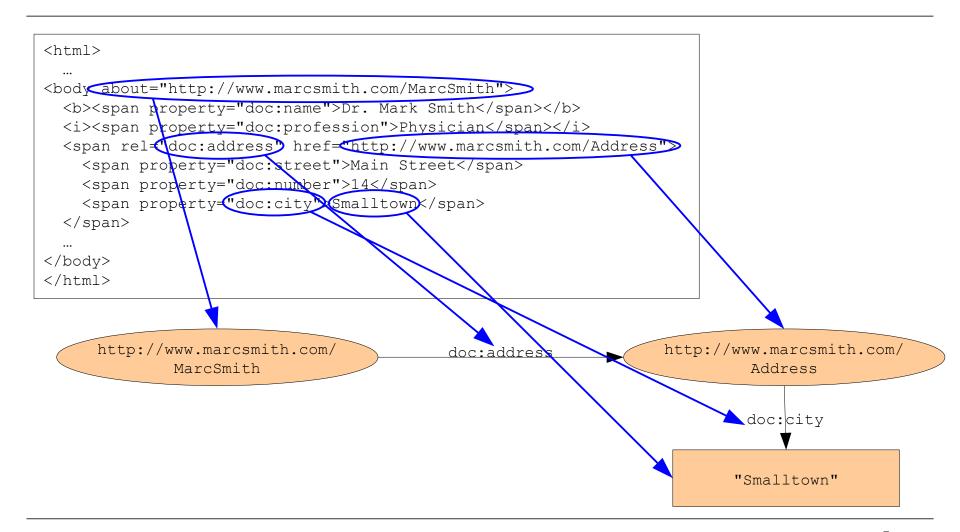




```
< ht.ml>
<body about="http://www.marcsmith.com/MarcSmith">>>
          <b><span property doc:name \not name \not
          <i><span property="doc:profession">Physiclan</span></i>
          <span rel='doc:address' href="http://www.marcsmith.com/Address">
                     <span property="doc:street">Nain Street</span>
                     <span property="doc:number">14</span>
                     <span property="doc:city">Smalltown</span>
          </span>
</body>
</html>
                              http://www.marcsmith.com/
                                                                                                                                                                                                                                                           doc:name
                                                                                                                                                                                                                                                                                                                                                                                       "Dr. Marc Smith"
                                                                          MarcSmith
```

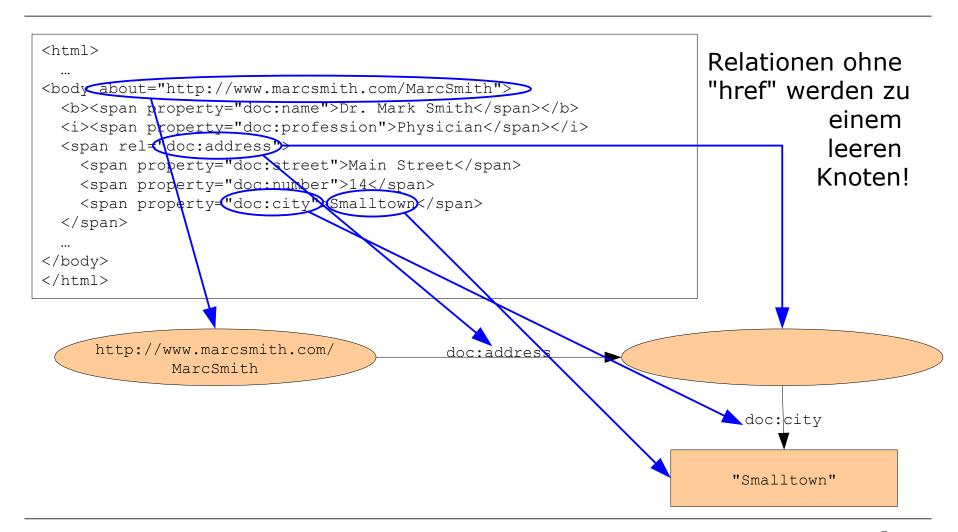








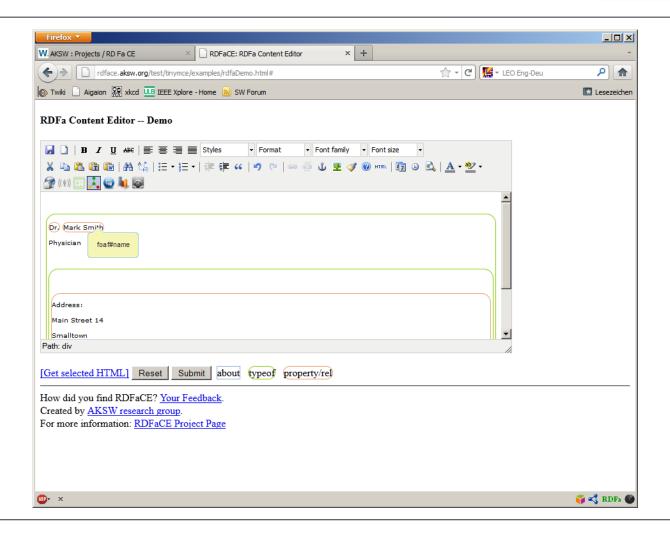






#### **RDFaCE**

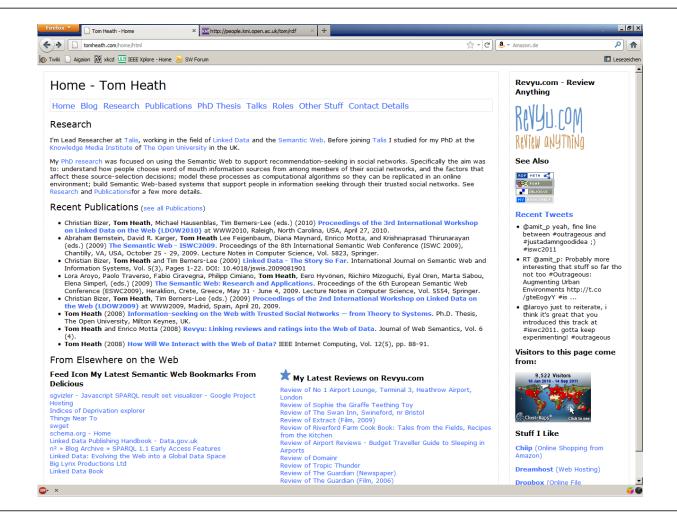






# RDF in der freien Wildbahn: FOAF-Profile

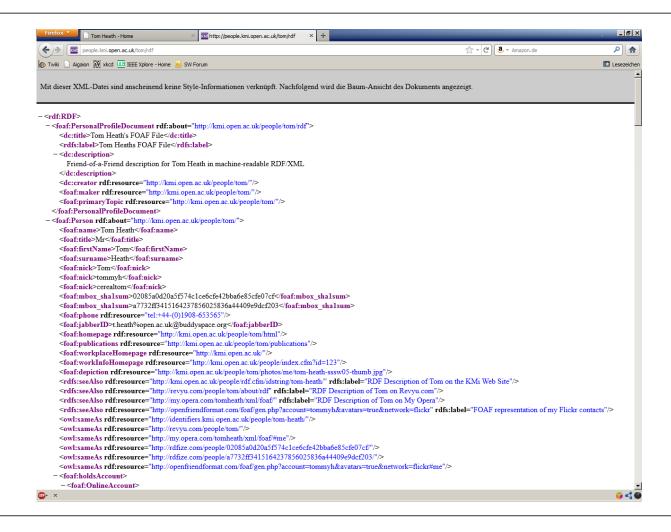






# RDF in der freien Wildbahn: FOAF-Profile

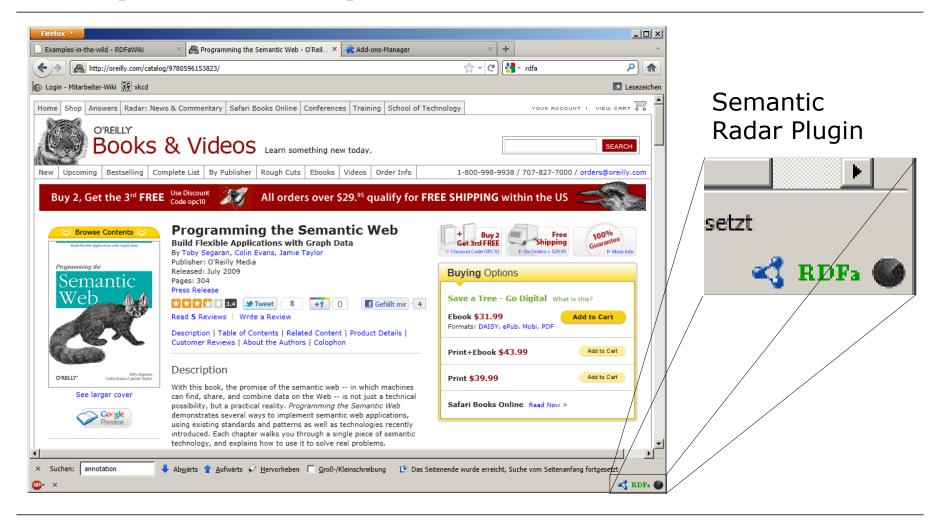






# RDFa in der freien Wildbahn: Beispiel O'Reilly

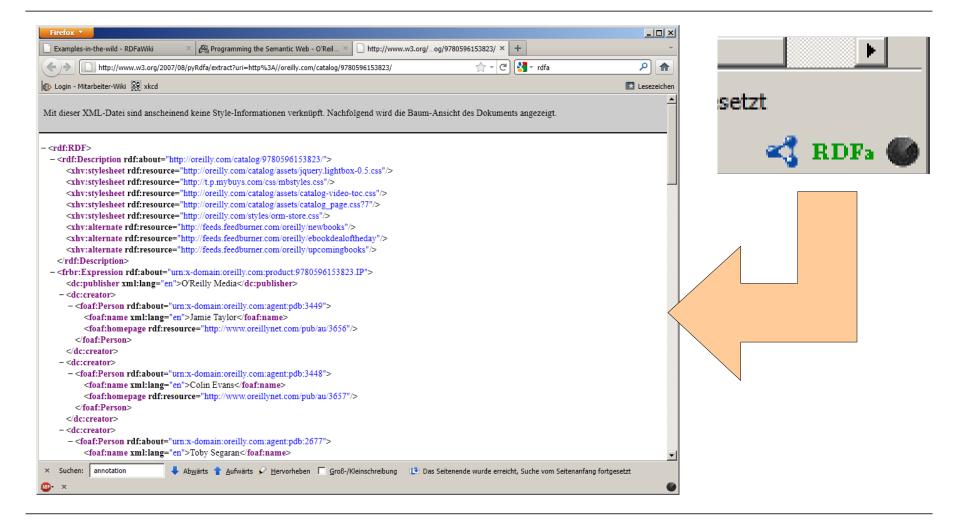






### RDFa in der freien Wildbahn: Beispiel O'Reilly







### RDFa in der freien Wildbahn: Beispiel O'Reilly



Normale Suche ohne RDFa:

#### Programming the Semantic Web - O'Reilly Media

shop.oreilly.com/product/9780596153823.do +1

With this book, the promise of the **semantic web** — in which machines can find, share, and combine data on the **Web** — is not just a technical possibility, but a ...

Verbessertes Ergebnis mit RDFa:



#### Programming the Semantic Web [Book] [10]

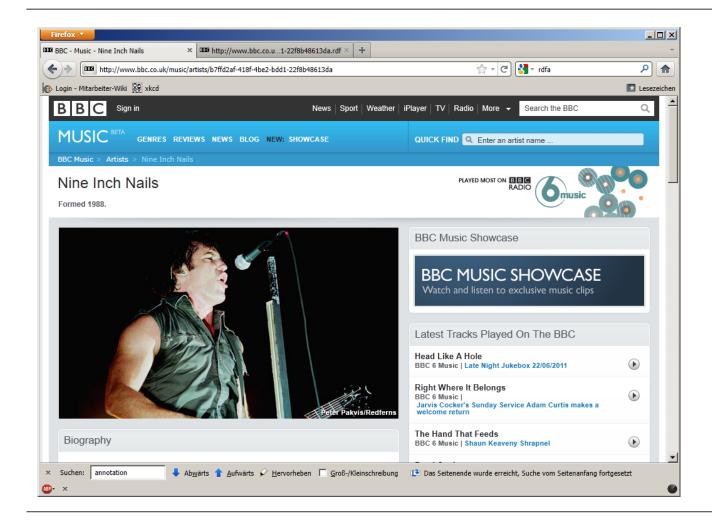
by Toby Segaran, Colin Evans, Jamie Taylor - O'Reilly (2009.07.01) - paperback - 280 pages
With this book, the promise of the **Semantic Web** -- in which machines can find, share, and combine data on the **Web** -- is not just a technical ...

★★★★★ 8 reviews - Add to Shopping List



### RDFa in der freien Wildbahn: Beispiel BBC Music

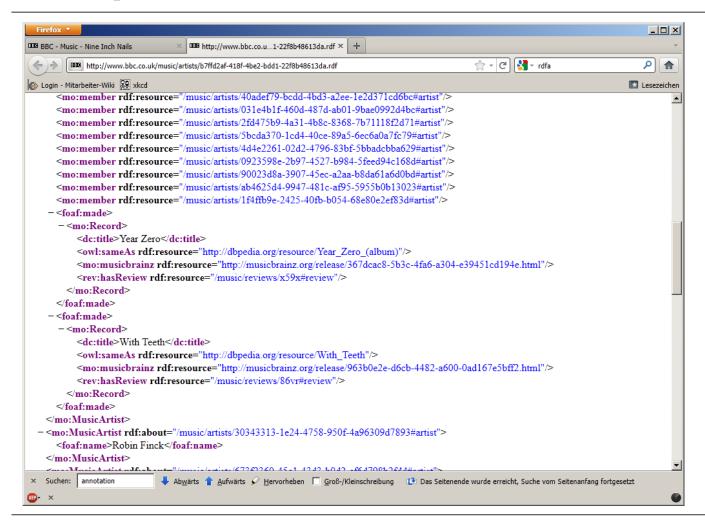






### RDFa in der freien Wildbahn: Beispiel BBC Music

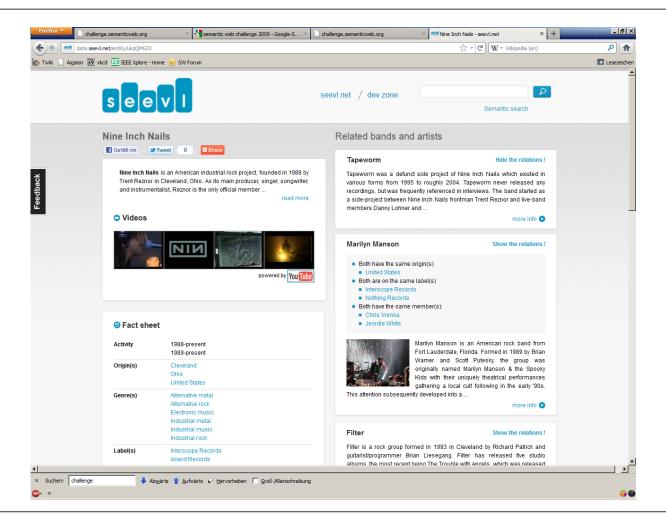






# Musikempfehlungen auf Basis von freilaufendem RDFa

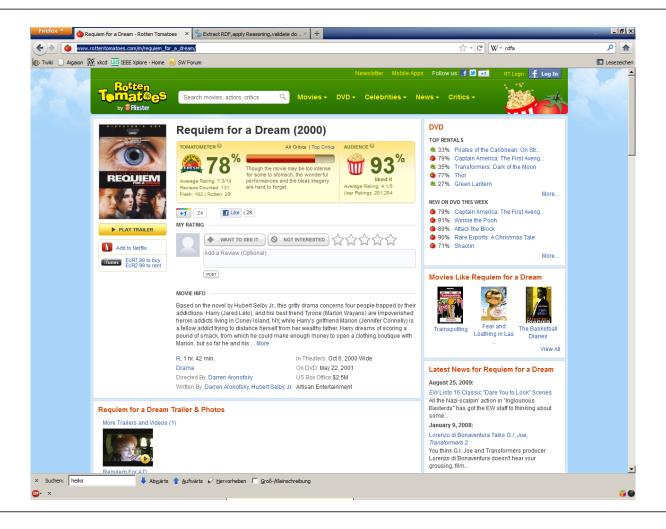






### RDFa in der freien Wildbahn: Beispiel Rotten Tomatoes

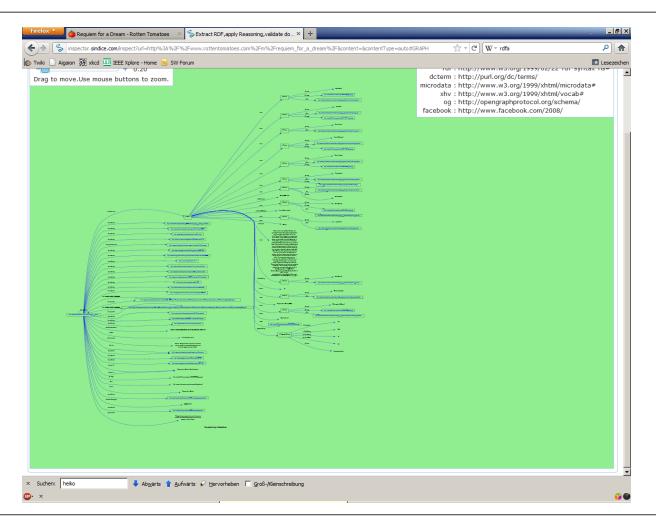






### RDFa in der freien Wildbahn: Beispiel Rotten Tomatoes

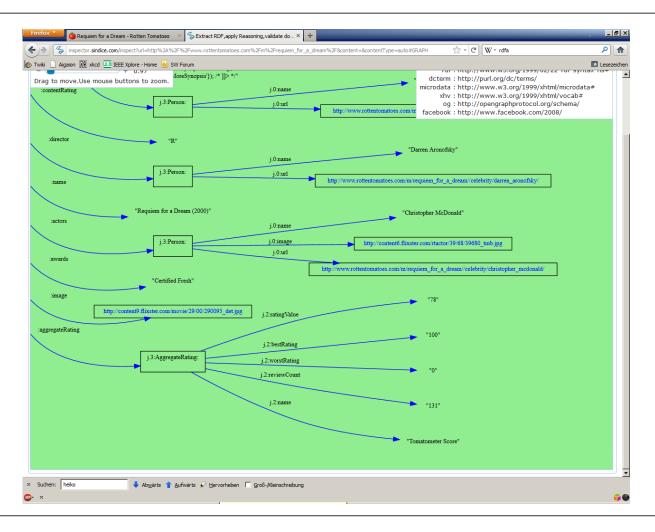






# RDFa in der freien Wildbahn: Beispiel Rotten Tomatoes







#### **Zusammenfassung RDF II**



- Verschiedenste Tools für RDF existieren
  - Speicherung
  - Visualisierung
  - und vieles mehr
- RDFa ermöglicht die Verbindung von HTML und RDF
  - ohne Redundanz
- Einige Beispiele für RDF in der freien Wildbahn



#### Zusammenfassung



- RDF ist ein Standard zur Beschreibung beliebiger Informationen
  - Graphenbasiert
  - Serialisierungen in Triples und XML
- Semantische Prinzipien
  - Non-unique Naming Assumption
  - Open World Assumption
- Tools zur Speicherung



#### Ein kritischer Blick zurück



- Kann RDF mehr als XML?
- XML ist eine Auszeichnungssprache für Informationen
- In XML kann man beliebige Tags und Attribute definieren
- XML-Tagnamen haben für den Computer keine Bedeutung

- RDF ist eine Auszeichnungssprache für Informationen
- In RDF kann man beliebige Klassen / Relationen definieren
- RDF-Bezeichner haben für den Computer keine Bedeutung



#### Ein kritischer Blick zurück



Warum haben wir das dann eigentlich gelernt?



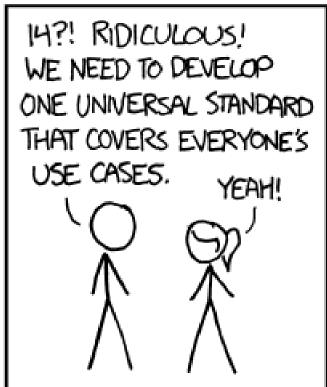


#### Ein kritischer Blick zurück



HOW STANDARDS PROLIFERATE: (SEE: A/C CHARGERS, CHARACTER ENCODINGS, INSTANT MESSAGING, ETC.)

SITUATION: THERE ARE 14 COMPETING STANDARDS.



SOON:

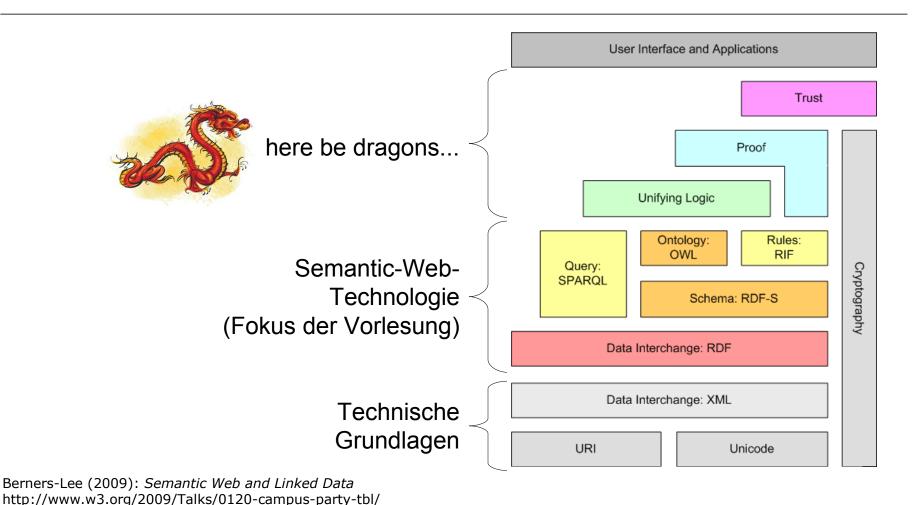
SITUATION: THERE ARE 15 COMPETING STANDARDS.

http://xkcd.com/927/



#### Semantic Web - Aufbau





30.10.12 | Fachbereich 20 | Knowledge Engineering | Heiko Paulheim | 69



### **Vorlesung Semantic Web**



Vorlesung im Wintersemester 2012/2013 Dr. Heiko Paulheim Fachgebiet Knowledge Engineering

