

# Semantic Web

Prof. Dr. Klaus Meyer-Wegener

Vortrag in IZED am 22. Oktober 2018



# Semantic Web – was ist das?

- **Ausgangspunkt:**

- Das ubiquitäre **World-wide Web** (WWW).

## 1. Suche im WWW ist **Volltextsuche**.

- Findet Texte, wenn ein Wort in ihnen vorkommt.
- Aber nicht, wenn das Wort nicht vorkommt – obwohl der Text inhaltlich mit dem Begriff zu tun hat.
  - "Negerkuss" – "Schaumkuss"
- Kann man auch nach "Bedeutung" suchen?
- Ja, wenn sie hinzugefügt wird.



# Semantic Web – was ist das?

## 2. Das WWW enthält **Texte**.

- Aber keine strukturierten Daten
  - Wie etwa Tabellen – Statistiken
- Idee:
  - Die ubiquitäre Infrastruktur des WWW auch für die Verbreitung von **Daten** nutzen.
- Abfragen und auswerten
  - Wie in Datenbanken



# Semantic Web

- Die Daten kommen dann von überall her.
- Ihre Bedeutung muss erklärt werden!
- **Ontologien**
  - Das sind Begriffssystematiken oder -systeme.
  - Beziehungen zwischen Begriffen
    - Oberbegriff – Unterbegriff
    - Synonyme, Homonyme
    - Gegensätze
    - Typ – Instanz
- **Zunächst aber: Struktur der Daten beschreiben**



# Resource Description Framework (RDF)

- **Tripel:**
  - **Subjekt – Prädikat – Objekt**
  - Alle sind jeweils **URIs**,
    - also weltweit eindeutige Bezeichner.
  - Subjekt des einen Tripels kann Objekt eines anderen sein und umgekehrt
    - Dadurch Zusammenhänge ausdrücken
  
- **Darstellung als Graph:**
  - Subjekte/Objekte sind Knoten, Prädikate Kanten
  - "Giant Global Graph" (GGG)
    - Zur Abgrenzung vom WWW
    - Hat sich nicht durchgesetzt



## RDF (2)

- **Annahme:**
  - Wissen besteht aus vielen miteinander verknüpften **Elementaraussagen** der Form: Subjekt – Prädikat – Objekt.
- **Beispiel:**
  - *Professor – hält – Vortrag*
- **Ressource:**
  - Oberbegriff für Subjekt, Prädikat und Objekt
  - Also etwas, über das man etwas aussagen will oder das man verknüpfen will
- **Muss eindeutig benannt werden!**
  - Sonst weiß man nicht, worüber eine Aussage gemacht wird.
- **Uniform Resource Identifier (URI)**
  - Weltweit eindeutig, hierarchisch strukturiert
  - Bekannteste Form: URL ("Uniform Resource Locator") im WWW
    - Identifiziert Web-Seiten, auch weltweit eindeutig
  - URI kann auch auf Web-Seite führen, muss es aber nicht



## RDF (3)

- **Tripel von URIs**

- Eine spezielle Form von Tupel:

[Subjekt: "URI1", Prädikat: "URI2", Objekt: "URI3"]

- **Beispiel:**

- [Subjekt: "http://example.org/SemanticWeb",  
Prädikat: "http://example.org/VerlegtBei",  
Objekt: "http://www.springer.com/Verlag"]

- **Subjekt und Prädikat müssen URIs sein, das Objekt darf auch ein **Literal** (ein Wert) sein:**

- [Subjekt "http://www.springer.com/Verlag",  
Prädikat: "http://example.org/Name",  
Objekt: "**Springer-Verlag**"]

- **Subjekt des einen Tripels kann Objekt des anderen sein!**



## RDF (4) – Turtle-Syntax

- **Vereinbarte Schreibweise:**
  - URIs in spitzen Klammern, Literale in Anführungszeichen:

```
<http://www.springer.com/Verlag>
```

```
<http://example.org/Name> "Springer-Verlag" 
```

- **Abkürzung wiederholt vorkommender Teile von URIs:**

```
@prefix ex: <http://example.org/>
```

```
@prefix springer: <http://springer.com/>
```

```
ex:SemanticWeb ex:VerlegtBei springer:Verlag .
```

```
ex:SemanticWeb ex:Titel "SemanticWeb - Grundlagen" .
```

```
springer:Verlag ex:Name "Springer-Verlag" .
```

- **Von Menschen leicht zu interpretieren –  
und von Computern auch**





## RDF (5) – Triple Stores

- Für die Speicherung bieten sich **Tabellen** an:

Subjekt	Prädikat	Objekt
ex:SemanticWeb	ex:VerlegtBei	springer:Verlag
ex:SemanticWeb	ex:Titel	"SemanticWeb - Grundlagen"
springer:Verlag	ex:Name	"Springer-Verlag"

Dann in **Datenbanken** abspeichern – und auswerten!

- SPARQL** ist eine Sprache, in der man die Auswertungen formulieren kann.



# RDF (6) - SPARQL

- **Anfragesprache für RDF:**
  - SPARQL (SPARQL Protocol And RDF Query Language)
- **Beispiel:**

```

PREFIX abc: <http://example.com/exampleOntology#>
SELECT ?capital ?country
WHERE {
  ?x abc:cityname ?capital ;
      abc:isCapitalOf ?y .
  ?y abc:countryname ?country ;
      abc:isInContinent abc:Africa .
}
  
```



# RDFS

- **RDF Schema**
- **Im Prinzip wie in Datenbanken:**
  - Gemeinsame Struktur für eine größere Menge von Ressourcen
  - "Formular":
    - Für Angestellte brauchen wir den Namen, das Geburtsdatum, ...
  - Nicht bei jeder einzelnen Ressource wieder angeben!
  - Und: Nichts vergessen!

- **Zunächst einmal Prädikate mit besonderer Bedeutung:**

```
<ex:SemanticWeb> <rdf:type> <ex:Lehrbuch> .
```

```
<ex:Lehrbuch> <rdf:type> <rdfs:Class> .
```

```
<ex:Buch> <rdf:type> <rdfs:Class> .
```

```
<ex:Lehrbuch> <rdfs:subClassOf> <ex:Buch> .
```



## RDFS (2)

- **Weitere Elemente:**
  - Klassen:  
Class, Resource, Property, Literal
  - Eigenschaften:  
subClassOf, subPropertyOf, domain, range
- **Genügt zur Definition einfacher Ontologien**
  - Mögliche richtige Aussagen  
von ganz sicher falschen trennen:  
`<ex:HansMeier> <ex:hatAlter> "-127" .`
- **Erste Schlussfolgerungen möglich**
- **Aber noch nicht umfassend genug**



# OWL

- **Taxonomien**
- **Formale Semantik**
  - Logische Schlussfolgerungen möglich
- **Klassen und Instanzen**
  - *Julius "ist ein" Mensch.*
- **Subklassen**
  - *Alle Menschen "sind auch" Lebewesen.*
- **Eigenschaften (Properties)**
  - *Ein Mensch "hat" einen Namen.*
- **Beispiel: CIDOC CRM**
  - (Museums-) Objekte
  - `<E12 Production> <P108 has-produced>`  
`<E24 Physical Man-Made Thing> .`



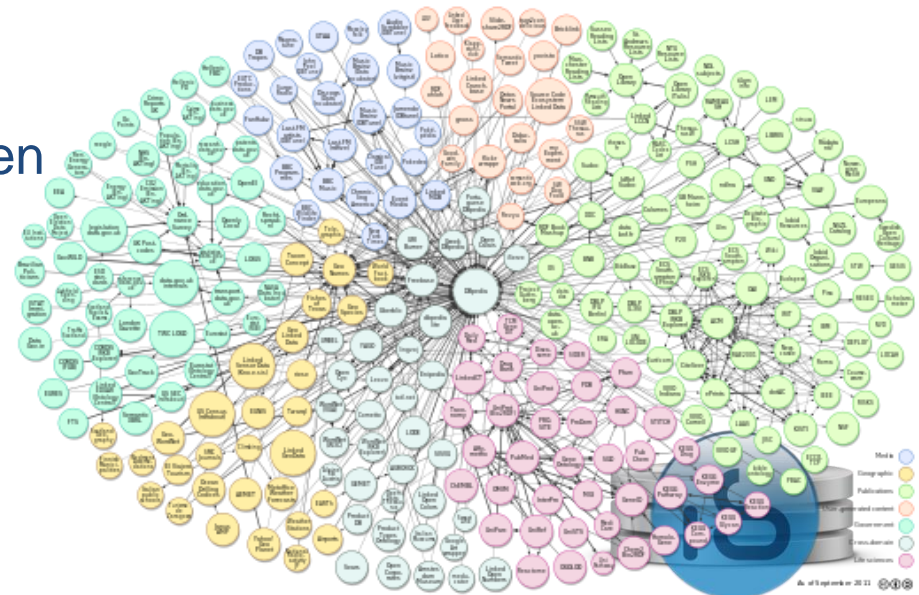
## OWL (2)

- Zuordnung *eigener* Datenobjekte zu den Bestandteilen einer Ontologie
  - "Mona Lisa" <rdf:type> <E24 Physical Man-Made Thing> .
  - "Mittelalter" <rdf:type> <E2 Temporal Entity> .



# Linked (Open) Data

- **Nutzung der Infrastruktur des World-wide Web (WWW) zur Bereitstellung von (freien) Daten im RDF**
  - Teil des Semantic Web
- **Anbieter von Daten:**
  - Regierungen, Zeitungen, Rundfunk- und Fernsehanstalten, Museen, ...
- **Über SPARQL abfragbar**
  - Oder insgesamt herunterladen
- **Einstieg:**
  - [linkeddata.org](http://linkeddata.org)
  - <https://lod-cloud.net/>



# Literatur

- Tim Berners-Lee, James Hendler, Ora Lassila: The Semantic Web: a new form of Web content that is meaningful to computers will unleash a revolution of new possibilities. In: *Scientific American*, 284 (5), May 2001, S. 34-43 (dt.: Mein Computer versteht mich. In: *Spektrum der Wissenschaft*, August 2001, S. 42–49).
- Pascal Hitzler, Markus Krötzsch, Sebastian Rudolph, York Sure: *Semantic Web – Grundlagen*. Springer : Berlin, Heidelberg, 2008. – ISBN 978-3-540-33993-9.
- Dean Allemang, Jim Hendler: *Semantic Web for the Working Ontologist: Effective Modeling in RDFS and OWL*. 2<sup>nd</sup> ed. Morgan Kaufmann : Amsterdam a.o., 2011. – ISBN 978-0-12-385965-5.
- Christian Bizer, Tom Heath, Tim Berners-Lee: Linked Data - The Story So Far. *Int. J. Semantic Web Inf. Syst.* 5(3), 2009: 1-22.

