

# **Modellierung nachhaltiger Systeme und Semantic Web**

## **Modellierung von Begriffswelten**

**Vorlesung im Modul 10-202-2330 im Master  
und Lehramt Informatik sowie  
im Modul 10-202-2309 im Master Informatik**

Prof. Dr. Hans-Gert Gräbe

<http://informatik.uni-leipzig.de/~graebe>

## RDF – Zentrale Konzepte

### Zusammenfassung der zentralen Konzepte

- Zentrale Idee: Speichere textuelle Beschreibungen auf uniforme Weise als Tripel und verwende Standardkonzepte und -werkzeuge zur Verwaltung dieser Daten
- *Ressourcen*: URI, HTTP access
  - URI = Unique Resource Identifier
  - Damit kann auf einen weltweit verteilten Datenbestand auf uniforme Weise über ein gängiges Protokoll zugegriffen werden.
- *Resource Descriptions*: Gib auf Anfrage an die HTTP-Adresse ein nützliches Stück Information im RDF-Format zurück, das mit anderen solchen Informationseinheiten zu neuen RDF-Sätzen kombiniert werden kann.
- Betreibe *RDF Triple Stores* als Teil einer weltweiten verteilten Datenspeicher-infrastruktur, etwa <http://od.fmi.uni-leipzig.de/>
- (Verteilte) Anfragesprache SPARQL
  - Stelle *SPARQL Endpunkte* auf RDF Triple Stores bereit

## RDF – Sprachformen und Praxen

### Verfahrenswissen → Verfahrensweisen

- Korrespondenz zwischen der Kohärenz der Sprachform und der Kohärenz der Praxen
- Die Etablierung kohärenter Praxen als Verfahrensweisen erlaubt es, Prädikate zu substantivieren. „Zum Beginn der Schulstunde werden die Pausenbrote weggepackt“. Erst nach einer solchen Transposition des Prädikats an die Subjekt-Position werden normative Sätze möglich.
  - Parallelen zum Konzertbeispiel in der ersten Vorlesung.
- Grundweisheiten der Informatik:
  1. Eine Funktion kann erst aufgerufen werden, nachdem sie definiert wurde.
  2. Eine Funktion, die definiert wird, aber nicht aufgerufen, deutet auf einen Designfehler hin.
- *Verfahrenswissen* ist die Beschreibungsform, *Verfahrensweise* die Ausführungsform.

## Die Linked Open Data Cloud

### Noch einmal zur Grundidee

- Auf diese Weise entsteht eine weltweit vernetzte dezentrale *offene* Datenbank, die Linked Open Data Cloud, in der alle *öffentlichen* Informationen freizügig und maschinenlesbar verfügbar sind.
  - Der Spieß des Datensammelns wird umgekehrt und an einer Welt gebaut, in der ein *Kernbestand öffentlich verfügbarer Informationen* als Public Domain eine wesentliche kulturelle Konstituante ist.
  - Siehe <http://lod-cloud.net/>
  - Wachsen der LOD Cloud: <http://lod-cloud.net/versions/>
- Kontext von Industrie 4.0 und allen großen Datenprojekten, auch Google ist ohne diese Bemühungen nicht denkbar.
- Basis für innerbetriebliche und zwischenbetriebliche Informationssysteme wie ERP und CRM.

## Namensräume und Begriffswelten

Kommunikation wird durch die Einführung von **Namensräumen** als URI-Präfix unterstützt.

Namensräume erlauben es, überlappungsfrei URIs zu generieren.

- Damit können Beschreibungen erzeugt werden, welche die *Fiktionen* MEINE Welt, MEINE Begriffe, ICH-Kern, Welt und Wirklichkeiten, Wirklichkeitskonstruktion bedienen, ohne diese gedanklich transzendieren zu müssen.
- Aber wir wollen ja mehr: Kooperation mit konkreten Anderen.
- Semantik = pragmatisch kontextualisierte Bildung von *Modellen* als Basis für gemeinsame *Verfahrensweisen*.
- Wir brauchen Sprache, um über die Modelle selbst sprechen zu können, und damit Wege der Formalisierung von Semantik.

## Konzeptualisierungen und Begriffswelten

### Aber: **Das Turmbau von Babel Phänomen**

- Was bedeutet es, dass jeder kommunikative Zusammenhang hierbei zusammen mit den eigenen Verfahrensweisen auch seine eigenen Modelle und damit seine eigene Sprache entwickelt und spricht?
- Welche Konzepte können Übersetzungsleistungen unterstützen?
- **Ontologien** (oder Vokabulare): WIR einigen uns auf die Verwendung gemeinsamer Namensräume (foaf, skos, org, sioc usw.) für *spezielle* Zwecke und damit auf gemeinsame *Teilmodelle* von Welt.
  - Phänomen der Kohärenz privater und kooperativer Sprachpraxen.
- Wie geht das genau?
  - **Inhalt:** Pragmatisch kontextualisierte Bildung von *Modellen* als Basis für gemeinsame *Verfahrensweisen*.
  - **Form:** *Ontologisierung* als pragmatisch kontextualisierte Semantifizierung von Syntax.
  - Beispiele auf der nächsten Folie

## Konzeptualisierungen und Begriffswelten

- **Beispiel foaf: Friend of a Friend**
  - foaf: <<http://xmlns.com/foaf/0.1/>>
  - Weiterleitung auf <http://xmlns.com/foaf/spec/>
  - Wir studieren das dort entwickelte Modell und die Beschreibungsformen von Semantik und Syntax
- **Beispiel skos: Simple Knowledge Organization System**
  - skos: <<http://www.w3.org/2004/02/skos/core#>>
  - Weiterleitung auf tabellarische Übersicht  
<https://www.w3.org/2009/08/skos-reference/skos.html>
  - Ganz am Ende der Seite drei Referenzen
- **Beispiel org: The Organization Ontology**
  - org: <<http://www.w3.org/ns/org#>>
  - Weiterleitung auf eine Turtle-Datei. Herunterladen und anschauen
  - rdfs:seeAlso <<https://www.w3.org/TR/vocab-org/>>

## Konzeptualisierungen und Begriffswelten

- Sozial ein extrem schwieriger Prozess, aber das ist **der Kern semantischer Technologien**: Die *Institutionalisierung* maschinenlesbarer gemeinsamer Begriffswelten *als sozialer Prozess*.
- Damit verbunden sind *Modellbildungen*, Bedingtheiten (Kontextualität verschiedener Wirklichkeiten) und der Prozess des Transzendierens von Kontexten, wenn Ontologien nicht wie ursprünglich vorgesehen angewendet werden.
  - Miteinander sprechen – Vereinbarung von Ontologien
  - Weiterentwicklung von Ontologien
  - Große Datenbanken von Ontologien: <http://prefix.cc> oder <http://lov.okfn.org> (Linked Open Vocabularies)
  - Kreativität im kooperativen Kontext. Formalisierungserfordernis, um Informationen als Daten auszutauschen. Noch einmal das Konzertbeispiel.



## Zwei Beispiele

- DBPedia – Extrahiert strukturierte Information aus Wikipedia
  - DBpedia is a crowd-sourced community effort to extract structured information from Wikipedia and make this information available on the Web. ... We hope that this work will make it easier for the huge amount of information in Wikipedia to be used in some new interesting ways. ...
  - Beispiel: <http://dbpedia.org/page/Leipzig>
- Linked-Data-Service der Deutschen Nationalbibliothek
  - Die Deutsche Nationalbibliothek erstellt einen Linked-Data-Service, der langfristig eine Nutzung der gesamten nationalbibliografischen Daten inklusive aller Normdaten durch die Semantic-Web-Community erlaubt. Sie ist bestrebt, durch diesen Datenservice einen Beitrag zur Informationsinfrastruktur weltweit zu leisten und damit eine Voraussetzung für moderne kommerzielle und nicht-kommerzielle Webdienstleistungen anzubieten.
  - <http://www.dnb.de/lds>

## Schema.org

- Anderer Zugang: <http://schema.org> - Googles Ontologisierung der Welt und Einbau in Webseiten statt Aufbau einer verteilten Datenbank wie in der Linked Open Data Cloud.
- Schema.org und Microdata: <https://schema.org/docs/gs.html>
  - itemscope, itemtype und itemprop und die Verbindung zu RDF.
- Auszeichnung von Webseiten mit diesem Markup erhöht deren Sichtbarkeit bei Google.

## Googles Knowledge Graph

- **Googles Knowledge Vault:** Extrahiert durch supervised learning aus den untersuchten Webseiten entsprechende Fakten als Googles Wissensbasis.
  - Enthielt 2014 über 1.6 Milliarden Fakten, die mit einem probabilistischen Konfidenzwert bewertet sind.
- **Google Knowledge Graph:** Konsolidierung und Anreicherung mit strukturierten Fakten aus Freebase (2007 gegründet, 2010 von Google aufgekauft), Wikipedia und Wikidata.
  - Enthielt 2016 über 70 Mrd. Fakten.
  - Ende 2015 wurde die Google Knowledge Graph API veröffentlicht, über die Webentwickler auf den Bestand zugreifen können.

Das ist aber nur ein Teil des **Giant Global Graph** (Tim Berners-Lee, 2007).

## Wolfram Alpha

- Ebenfalls Suchmaschine, die auf Fakten aufbaut, die aus eigener Recherche gewonnen wurden. Zusammen mit *Mathematica* als Compute Engine lassen sich komplexere Präsentationen und Visualisierungen erstellen. Ziel ist die Vernetzung von mathematischem Wissen und Allgemeinwissen.
- <https://www.wolframalpha.com>
  - Beispiel „Leipzig“.