

Verstehen verstehen

**Eine technikwissenschaftliche
Perspektive**

Vortrag am 18.01.2019 am IWM Leipzig

Prof. Dr. Hans-Gert Gräbe

<http://bis.informatik.uni-leipzig.de/HansGertGraebe>

Eine triviale Vorbemerkung

Verstehen setzt Verstehen in wenigstens folgenden Dimensionen voraus:

- Wer möchte verstehen? – Subjekt
- Was soll verstanden werden? – Objekt

Neben diesem philosophisch sehr komplexen Verhältnis sind weitere Dimensionen *optional* zu berücksichtigen:

- Unter welchen Gegebenheiten geschieht das Verstehen?
– Bedingtheiten
- Warum verstehen? – Interessen

Sprache ist voraussetzungsreich

Ein Beispiel:

- Am 8.11. wurde an der Station Leipzig-Flughafen um 17 Uhr eine Temperatur von 16°C gemessen.
- Am `8.11.` wurde an der *Station* `Leipzig-Flughafen` um `17 Uhr` eine `Temperatur` von `16°C` gemessen.
- Dinge und deren Bezeichnungen.

RDF Basics (1)

Beispiel: Beschreibung einer Lehrveranstaltung

```
@prefix od: <http://od.fmi.uni-leipzig.de/model/> .  
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .  
@prefix odr: <http://od.fmi.uni-leipzig.de/rooms/> .  
@prefix odp: <http://od.fmi.uni-leipzig.de/personal/> .
```

```
<http://od.fmi.uni-leipzig.de/w18/SWS.IADW.1>  
  a od:Vorlesung ;  
  od:beginsAt "11:15" ;  
  od:dayOfWeek "donnerstags" ;  
  od:endsAt "12:45" ;  
  od:locatedAt odr:Hs_19 ;  
  od:servedBy odp:Graebe_HansGert ;  
  rdfs:label "Vorlesung ..." .
```

Identifizier und Literale. Namensräume.

RDF Basics (2)

Auflösung in Drei-Wort-Sätze

Subjekt Prädikat Objekt .

W18:SWS.IADW.1	a	od:Vorlesung .
W18:SWS.IADW.1	od:beginsAt	"11:15" .
W18:SWS.IADW.1	od:dayOfWeek	"donnerstags" .
W18:SWS.IADW.1	od:endsAt	"12:45" .
W18:SWS.IADW.1	od:locatedAt	odr:Hs_19 .
W18:SWS.IADW.1	od:servedBy	odp:Graebe_HansGert .
W18:SWS.IADW.1	rdfs:label	"Vorlesung ..." .

Mehr dazu siehe <http://od.fmi.uni-leipzig.de/>

RDF Basics (3)

Konzeptionelle „Zutaten“:

- UTF-8 als **einheitliche Zeichenbasis** für URIs und Literale.
 - Best Practise: URIs nur aus ASCII-Zeichen, keine Umlaute oder Ähnliches
- URI als „digitale Identitäten“ von Ressourcen, *zeigen* auf Ressourcen
 - Wie digitale Identitäten von Personen sind dies **textuelle Repräsentationen der „Dinge“** in den im Internet kursierenden Textfragmenten.
- Für Computer sind URIs einfach Zeichenketten, für Menschen ist es hilfreich, wenn die URI bereits eine Ahnung von der Semantik des Gezeigten vermittelt.
 - Best Practise: „sprechende Namen“ als URIs

RDF Basics (4)

- **RDF - Resource Description Framework**
 - Konzept zum Aufschreiben von Geschichten über „die Welt“ als *Mengen* von Drei-Wort-Sätzen
 <Subjekt> <Prädikat> <Objekt> .
 - Subjekt und Prädikat müssen URIs sein, als Objekt kann eine URI oder ein Literal (Typ `rdfs:Literal`) stehen. Literale können Typ- und Sprachmarkierungen tragen.
 - Es gibt verschiedene Notationen für dieselbe Menge von RDF-Sätzen (Turtle, `rdf/xml`, `json`, `ntriples`) und Werkzeuge, diese Notationen ineinander umzuwandeln.
 - Redland RDF libraries <http://librdf.org/>
 - Mustersuche als mächtiges Konzept der Analyse derartiger Mengen. SPARQL als Anfragesprache.

RDF - Sätze und Muster

Sätze folgen dabei Mustern:

1) Turtle: Sammle alle Sätze mit demselben Subjekt → Eigenschaften eines Individualsubjekts als Schlüssel-Wert-Paare.

- Schlüssel-Wert = Eigenschaft und Ausprägung

2) Sammle alle Sätze mit demselben Prädikat

A od:beginsAt B

- od:beginsAt nicht nur eine URI (*Syntax*), sondern auch ein zweistelliges Prädikat mit einer gewissen *Semantik*, die in allen Sätzen mit diesem Prädikat als *Instanziierungen des Prädikats* transportiert wird.

3) Andere Muster sind möglich, SPARQL als genereller Standard einer Anfragesprache für Mustersuche.

RDF - Grundlegende Beschränkungen

- Mengensemantik, Reihenfolge der Sätze spielt keine Rolle
- Problem der Kontextualisierung. In welchem raumzeitlichen Kontext ist der Satz gültig? Hier gibt es mehrere Ansätze:
 - Aufbohren zu Quadrupeln mit vierter Komponente als URI auf die Provenienz.
 - Wird der Satz als Instanziierung eines zweistelligen Prädikats interpretiert, so ergibt sich der Kontext oft implizit aus dem Prädikat.
 - Oft ergibt sich der Kontext allgemeiner aus dem Namensraum des Prädikats und steht damit als (explizites oder implizites) Modell für eine ganze Klasse von Begrifflichkeiten. Damit verlagert sich die Frage aber nur in die Beschreibung des Modells und damit eine Abstraktionsebene nach oben.

RDF - Beschreibungen von Beschreibungen

- **Selbstähnlichkeit des Konzepts:** Auch Beschreibungen von Beschreibungen können als RDF-Sätze formuliert werden. Insbesondere kann man RDF verwenden, um RDF zu beschreiben.
 - Eine URI, die in einem Satz als Prädikat auftritt, kann in einem anderen Satz als Subjekt oder Objekt auftreten.
Beispiel:
 - od:beginsAt rdfs:domain od:LV .
 - od:beginsAt rdfs:range rdfs:Literal .
- Damit können auch *Begriffe und Konzepte* RDF-basiert beschrieben werden. → **Universalien**
 - Was sind Universalien? Ideen aus Platons Ideenhimmel oder institutionalisierte Konventionen, „state of the art“?

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

Namensräume und Begriffswelten

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

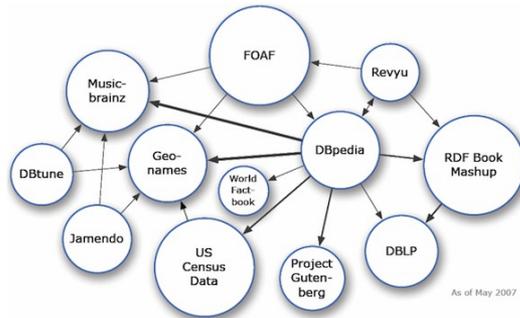
Namensräume erlauben es, überlappungsfrei URIs zu generieren.

- 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.

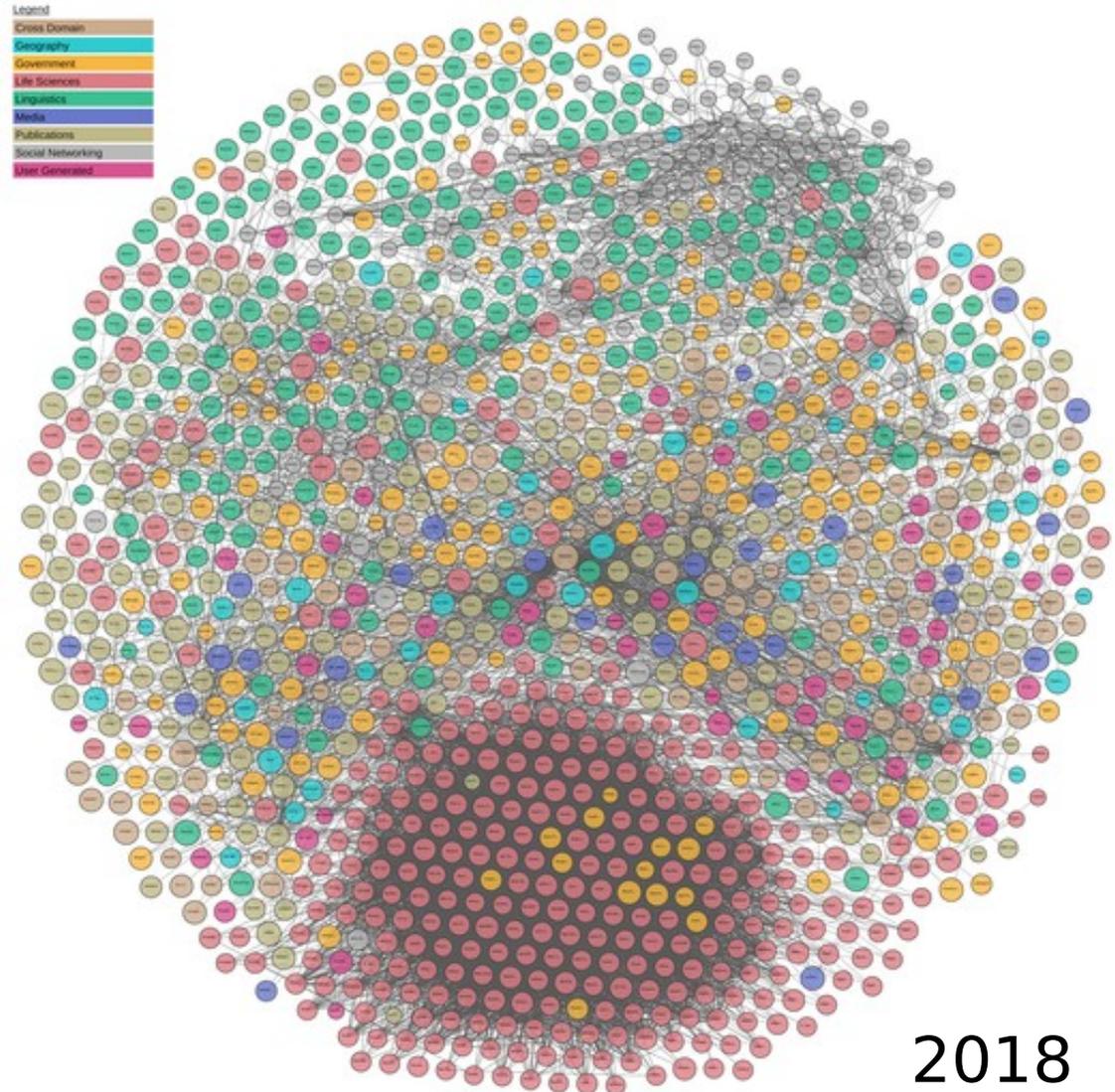
Die Linked Open Data Cloud

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.



2007



2018

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>

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/>
 - Studiere 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.
 - Kann man 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.