Web Developer Gera #3

Donnerstag, 7. November 2019

Über mich

- Johannes
- 34
- aus Gera
- Fachinformatiker für Systemintegration
- Bachelor Betriebliche IT-Systeme
- Master Informatik
- Arbeite seit mehreren Jahren als Webentwickler

Elsterkind GmbH

- Digitalagentur
- Online Marketing, Ads, Affiliate, Analytics, Webentwicklung…
- 2018 in Leipzig gegründet
- 10 Mitarbeiter + 6 Freelancer

Triple Stores, DBpedia & SPARQL

Thema

Disclaimer

Dieser Vortrag soll **informativ** sein und gegebenenfalls **Interesse für das Thema wecken**.

Ich bin darin **kein Profi**. Ich benutze das aktuell **nicht produktiv**. Ich verfolge nicht aktiv die Entwicklung der vorgestellten Techniken. ABER ich finde das Thema spannend.

Fragen werden nach **bestem Wissen und Gewissen** beantwortet.

Agenda

- Triple Store Datenbanken
- Projekt DBpedia
- Intro SPARQL mit Anwendungsbeispiel

Triple Stores

Triple Store

- Wikipedia:
  - Ein Tripelspeicher oder RDF-Speicher (RDF = Resource Description Framework) ist eine speziell entwickelte Datenbank für das Speichern und Abrufen von Tripeln durch semantische Abfragen.

- Triple:
  - Datenentität, die aus Subjekt, Prädikat und Objekt besteht

Wer → Hat → Was

Aufbau eines Triples 1/2

- Subjekt:
  - Das Element, über das eine Aussage getroffen wird
- Prädikat:
  - Die Eigenschaft, über die eine Aussage getroffen wird
- Objekt:
  - Der Wert, den die Eigenschaft bekommt

Aufbau eines Triples

- Subjekt:
  - Beispiel: Das Kutscherhaus Gera
- Prädikat:
  - Beispiel: liegt am
- Objekt:
  - Beispiel: Mühlgraben

Aufbau eines Triples 2/2

- Die einzelnen Attribute werden mit URIs gekennzeichnet
  - URI = Uniform Resource Identifier
- Ähnlich zu URLs, aber muss nicht auf ein Webdokument zeigen
- Dienen zur weltweit eindeutigen Bezeichnung von Ressourcen
- Beispiel: http://gera.de/orte/Kutscherhaus

Aufbau eines Triples

- Subjekt:
  - Beispiel: Das Kutscherhaus Gera
  - URI: <http://gera.de/orte/Kutscherhaus>
- Prädikat:
  - Beispiel: liegt am
  - URI: <http://gera.de/geo/liegt_am>
- Objekt:
  - Beispiel: Mühlgraben
  - URI: <http://gera.de/gewässer/Mühlgraben>

RDF Graph

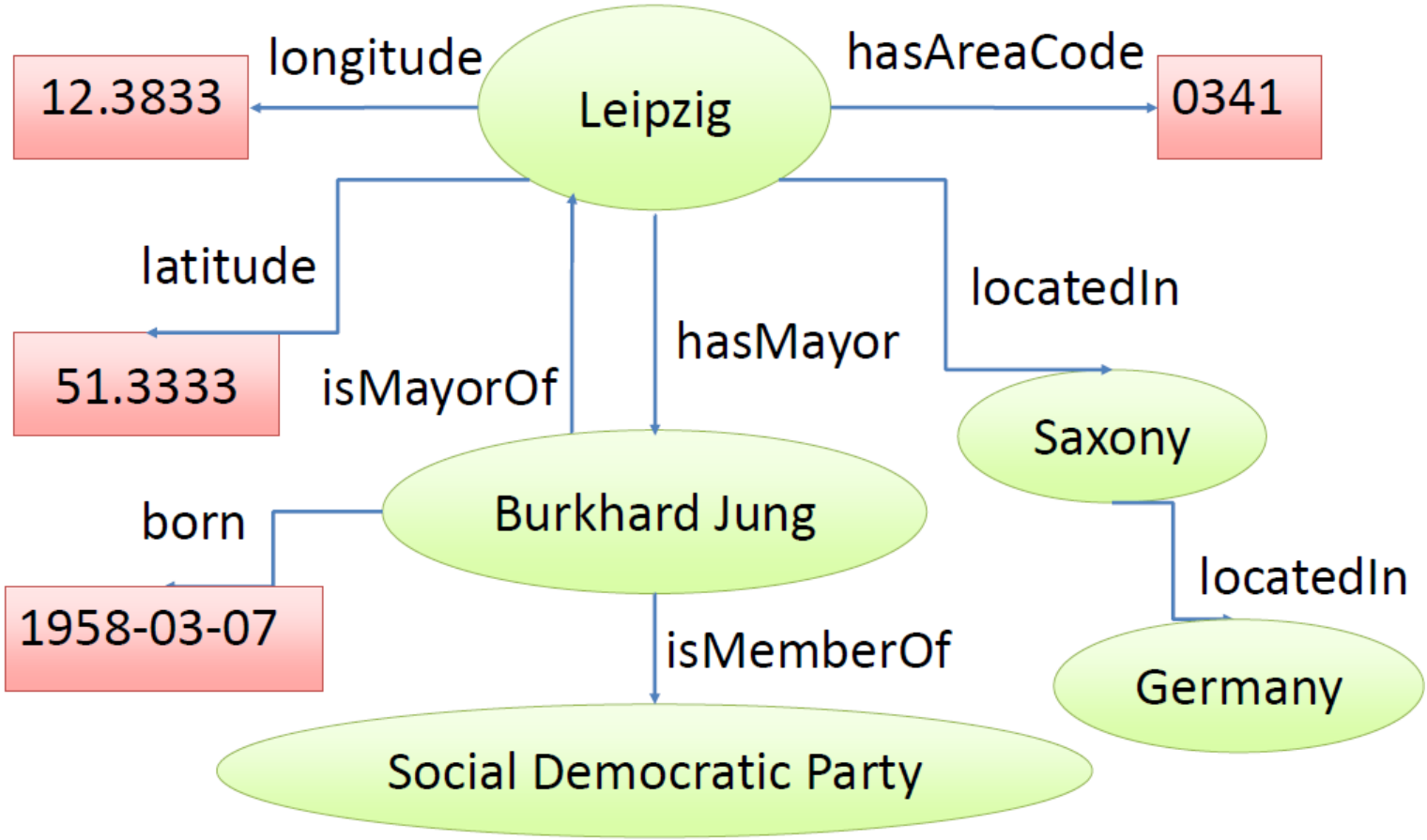


Bild: Uni Leipzig – Vorlesung Semantic Web 2012

Ontologie

- Die Definition eines solchen Graphen, wird Ontologie bezeichnet
- Dies ähnelt einem Datenbankschema
  - Ort hat Longitude und Latitude als Literal
  - Ort hat Bürgermeister als Objekt
  - Bürgermeister hat Name als String
  - usw. usw.

- Auch können Regeln für diese Schemata festgelegt werden
  - Ort hat nur 1 Bürgermeister
  - Latitude und Longitude müssen Float sein
  - Es kann mehrere Orte geben

Notation

- Es gibt verschiedene Möglichkeiten Triple-Daten zu definieren
- Eine einfache Notation ist die Turtle Syntax

```
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
@prefix xsd: <http://www.w3.org/2001/XMLSchema#> .
@prefix dbr: <http://dbpedia.org/resource/> .
@prefix dbp: <http://dbpedia.org/property/> .
@prefix geo: <http://www.w3.org/2003/01/geo/wgs84_pos#> .

dbr:Leipzig     dbp:hasMayor    dbr:Burkhard_Jung .
dbr:Leipzig     rdfs:label      "Leipzig"@de .
dbr:Leipzig     geo:lat         "51.333332"^^xsd:float .
dbr:Leipzig     geo:long        "12.383333"^^xsd:float .
```

Zusammenfassung

- Speicherung semantischer Daten
- Subjekt → Prädikat → Objekt
- Schema definiert Struktur
- Adressierung mit Hilfe von URIs
- Objekte können URIs oder Literale sein

DBpedia

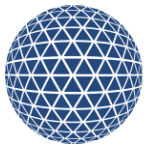
DBpedia

- 2007 ins Leben gerufen
- Unter anderem von der Uni Leipzig
- Gemeinschaftsprojekt verschiedener Einrichtungen

- Ziel:
  - Wikipedia-Artikel nach den Konzepten von **Linked Open Data** für das **Semantic Web** transformieren

DBpedia

- Einfach gesagt:
  - Man nehme Wikipedia-Artikel und forme die Informationen so um, dass die Daten strukturiert und maschinenlesbar abgerufen werden können.



InfAI®
Institut für Angewandte Informatik

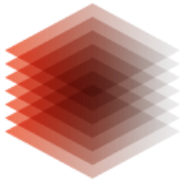
FIZ Karlsruhe
Leibniz-Institut für Informationsinfrastruktur

Fraunhofer
IAIS

OPEN KNOWLEDGE
FOUNDATION

SEMANTIC WEB COMPANY

Linked Open Data Initiative
LinkedOpenData.jp ★★★★★



TIB
LEIBNIZ-INFORMATIONSZENTRUM
TECHNIK UND NATURWISSENSCHAFTEN
UNIVERSITÄTSBIBLIOTHEK

Knowledge Graph™

Datenbasis

- Als Datenbasis dienen nicht die Fließtexte
- Ausgewertet werden:
  - Infoboxen, Tabellen, geographische Daten, Weblinks, etc.
- Die extrahierten Daten werden als RDF-Daten gespeichert
- Die RDF-Datenbank wird in längeren Abständen aus Wikipedia Daten neu erstellt

Wappen	Deutschlandkarte
	

Basisdaten	
Bundesland:	Thüringen
Höhe:	194 m ü. NHN
Fläche:	152,19 km^2
Einwohner:	94.152 *(31. Dez. 2018)*[1]
Bevölkerungsdichte:	619 Einwohner je km^2
Postleitzahlen:	07545–07557
Vorwahlen:	0365, 036695
Kfz-Kennzeichen:	G
Gemeindeschlüssel:	16 0 52 000
Stadtgliederung:	40 Stadtteile
Adresse der Stadtverwaltung:	Kornmarkt 12 07545 Gera

Deutsches DBpedia

- Projektseite: http://de.dbpedia.org/
- Deutsche Live-Datenbasis leider aktuell nicht verfügbar ☹

- Vortrag vom 12.09.2019 – 14th DBpedia Community Meeting
  - „The Return of German DBpedia“
  - Man will wieder aktiv(er) an dem Projekt arbeiten
  - Verbesserung der deutschen Extraktionsdaten
  - Suche nach aktiven Mitgliedern/Unterstützung
  - Deutsches Wikipedia ist das 4. größte weltweit

DBpedia Datenbasis (2016-10 de)

Anzahl von Triplen	104.989.619
Anzahl von Entitäten (owl:Thing)	1.374.894
Anzahl von Personen (dbo:person)	627.264
Anzahl von Organisationen (dbo:organizations)	62.054
Anzahl von Orten (dbo:locations)	406.943
Mapped Templates	17,48 %
Mapped Properties	7,21 %

SPARQL

SPARQL

- Wikipedia:
  - SPARQL ist eine graphenbasierte Abfragesprache für RDF. Der Name ist ein rekursives Akronym für SPARQL Protocol and RDF Query Language.
- 2008 W3C-Standard
- Grobe Struktur ähnlich SQL

Arten von SPARQL Anfragen

- SELECT
  - Gibt eine Tabelle mit Daten aus, die die Bedingungen erfüllen
- CONSTRUCT
  - Wie SELECT, nur setzt Daten in neuer Form zusammen
- DESCRIBE
  - Gibt alle Anweisungen im Datensatz aus
- ASK
  - Prüft, ob es Daten gibt, die einem SELECT entsprechen

Struktur

- PREFIX
  - Definitionen von Präfixen, die in der Anfrage benutzt werden (ähnlich Namespaces)
- SELECT/…
  - Anfragetyp, was gesucht wird
- WHERE
  - Filter der Anfrage
- Dazu noch:
  - FILTER, ORDER, LIMIT, OFFSET, …

SPARQL Beispiel

- Einfaches Anwendungsbeispiel
- Nachvollziehbar auf http://dbpedia.org/sparql/
  - Einfach Query Text einfügen und „Run Query“ ausführen
- Hinweis:
  - Die genutzten Prefixe sind vordefinierte Prefixe für die genutzte Ontologie. Daher müssen die hier nicht explizit definiert werden

SPARQL Beispiel

```
SELECT ?person
WHERE {
    ?person  dbo:birthPlace  dbr:Gera .









}
LIMIT 10
```

SPARQL Beispiel

person
http://dbpedia.org/resource/Max_Frankel
http://dbpedia.org/resource/Konrad_Weise
http://dbpedia.org/resource/Tina_Liebig
http://dbpedia.org/resource/Thomas_Blaschek
http://dbpedia.org/resource/Jürgen_Simon
http://dbpedia.org/resource/Erik_Balnuweit
http://dbpedia.org/resource/Prince_Heinrich_VIII_Reuss_of_Köstritz
http://dbpedia.org/resource/Helga_Königsdorf
http://dbpedia.org/resource/Dietrich_Peltz
http://dbpedia.org/resource/Heike_Drechsler

SPARQL Beispiel

```
SELECT ?person
WHERE {
    ?person  dbo:birthPlace  dbr:Gera .
    ?person  foaf:gender  "female"@en .




}
LIMIT 10
```

SPARQL Beispiel

person
http://dbpedia.org/resource/Tina_Liebig
http://dbpedia.org/resource/Helga_Königsdorf
http://dbpedia.org/resource/Heike_Drechsler
http://dbpedia.org/resource/Kathrin_Zimmermann
http://dbpedia.org/resource/Heidi_Eisenschmidt
http://dbpedia.org/resource/Marlies_Göhr
http://dbpedia.org/resource/Margitta_Pufe
http://dbpedia.org/resource/Patricia_Polifka
http://dbpedia.org/resource/Karin_Hübner
http://dbpedia.org/resource/Bianca_Schmidt

SPARQL Beispiel

```
SELECT ?person
WHERE {
    ?person  dbo:birthPlace  dbr:Gera .
    ?person  foaf:gender  "female"@en .



    MINUS {
          ?person  dbo:deathDate  ?deathDate .
    }
}
LIMIT 10
```

SPARQL Beispiel

person
http://dbpedia.org/resource/Tina_Liebig
http://dbpedia.org/resource/Heike_Drechsler
http://dbpedia.org/resource/Kathrin_Zimmermann
http://dbpedia.org/resource/Heidi_Eisenschmidt
http://dbpedia.org/resource/Marlies_Göhr
http://dbpedia.org/resource/Margitta_Pufe
http://dbpedia.org/resource/Patricia_Polifka
http://dbpedia.org/resource/Bianca_Schmidt
http://dbpedia.org/resource/Sabrina_Schmutzler
http://dbpedia.org/resource/Hanka_Kupfernagel

SPARQL Beispiel

```
SELECT ?person
WHERE {
    ?person  dbo:birthPlace  dbr:Gera .
    ?person  foaf:gender  "female"@en .
    ?person  rdf:type  dbo:Athlete .
    ?person  dct:subject  dbc:Olympic_athletes_of_East_Germany .


    MINUS {
          ?person  dbo:deathDate  ?deathDate .
    }
}
LIMIT 10
```

SPARQL Beispiel

person
http://dbpedia.org/resource/Heike_Drechsler
http://dbpedia.org/resource/Margitta_Pufe

SPARQL Beispiel

```
SELECT ?person, ?weight, ?height
WHERE {
    ?person  dbo:birthPlace  dbr:Gera .
    ?person  foaf:gender  "female"@en .
    ?person  rdf:type  dbo:Athlete .
    ?person  dct:subject  dbc:Olympic_athletes_of_East_Germany .
    ?person  dbo:Person\/weight  ?weight .
    ?person  dbo:Person\/height  ?height .
    MINUS {
          ?person  dbo:deathDate  ?deathDate .
    }
}
LIMIT 10
```

SPARQL Beispiel

person	weight	height
http://dbpedia.org/resource/Heike_Drechsler	"68.0"^^<http://dbpedia.org/datatype/kilogram>	"181.0"^^<http://dbpedia.org/datatype/centimetre>
http://dbpedia.org/resource/Margitta_Pufe	"83.0"^^<http://dbpedia.org/datatype/kilogram>	"180.0"^^<http://dbpedia.org/datatype/centimetre>

Beispiel für semantische Suchmaschine

- Fragestellung:
  - „Wie groß ist Heike Drechsler?“

```
SELECT ?height
WHERE {
    ?person  rdfs:label  "Heike Drechsler"@de .
    ?person  dbo:Person\/height  ?height .
} LIMIT 1
```

height
"181.0"^^<http://dbpedia.org/datatype/centimetre>

Beispiel für semantische Suchmaschine

- Fragestellung:
  - „Welche ist die größte Stadt Thüringens?“

```
SELECT ?city, ?population
WHERE {
    ?state rdfs:label "Thüringen"@de .
    ?city dbo:federalState ?state .
    ?city dbo:populationTotal ?population .
} ORDER BY DESC(?population)
 LIMIT 1
```

city	population
http://dbpedia.org/resource/Erfurt	"204880"^^<http://www.w3.org/2001/XMLSchema#nonNegativeInteger>


ENDE
imgflip.com