



NIEDERSÄCHSISCHE STAATS- UND
UNIVERSITÄTSBIBLIOTHEK GÖTTINGEN

RDF Daten durchsuchen

Alex Jahnke (SUB Göttingen)

Metadaten und Linked Open Data in Bibliotheken und Archiven



GEORG-AUGUST-UNIVERSITÄT
GÖTTINGEN

Triplestore

- eine auf den Umgang mit RDF-Tripeln hin optimierte Datenbanksoftware
- bisweilen spricht man auch von Subjekt-Prädikat-Objekt Datenbanken
- entweder technisch vollständig selbständig oder basierend auf bereits existierenden relationalen oder nicht-relationalen Datenbanken
- es handelt sich dabei nicht um Graph-Datenbanken im strengen Sinn
- eine Übersicht über die wichtigsten Produkte gibt:
https://en.wikipedia.org/wiki/List_of_subject-predicate-object_databases
- ein Quad-Store ist ein Triplestore, der mit benannten Graphen umgehen kann

SPARQL

- **SPARQL Protocol and RDF Query Language**
- aktuelle Version: SPARQL 1.1 (März 2013)
- eher eine "Protokollfamilie"
 - Abfragesprache (Query Language)
 - Datenmanipulationssprache (Update)
 - Beschreibung von SPARQL Endpoints (Service Description)
 - Verteilte Abfragen (Federated Query)
 - Ausgabeformate (Query Results ... (3))
 - Abfragen auf der Basis der Abhängigkeit der Graphen voneinander (Entailment Regime)
 - Verwendung von SPARQL über HTTP (Protocol)
 - Retrieval und Speichern von Graphen über HTTP (Graph Store Protocol)

SPARQL Query Language

- ***SELECT*** Suche nach Tripeln bzw. Teilgraphen
- ***ASK*** Test ob ein best. Triple vorhanden ist
- ***DESCRIBE*** Beschreibung einer Ressource
- ***CONSTRUCT*** RDF Graphen erzeugen

Spezifikation: <https://www.w3.org/TR/sparql11-query/>

SPARQL Syntax

```
1 ▼ PREFIX exp: <http://example.org/property/>
2 PREFIX dbr: <http://dbpedia.org/resource/>
3
4 ▼ SELECT ?person ?hochschule WHERE {
5     ?person exp:isAlumnusOf ?hochschule .
6 }
```

SPARQL-Syntax: Graph Pattern

```
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 .
}
```

Übung

- siehe:

https://www.eromm.org/dcgkb/doku.php?id=schulungen:metadaten_und_linked_data

Verteilte Suche

```
1 SELECT ?person ?geburtsname
2 WHERE {
3     ?person <http://dbpedia.org/ontology/occupation> <http://dbpedia.org/resource/Composer> .
4     SERVICE <http://fr.dbpedia.org/sparql> {
5         ?person <http://fr.dbpedia.org/property/nomDeNaissance> ?geburtsname .
6     }
7 }
```


RDF Daten veröffentlichen

- Dereferenzierung
 - einzelne Beschreibungssätze in einem RDF Serialisierungsformat als Webresource vorhalten: <http://d-nb.info/gnd/117453889/about/lds>
- Data Dump
 - gesamter Datenbestand in einem RDF Serialisierungsformat in einer (oder mehreren) Datei(en): <http://datendienst.dnb.de/cgi-bin/mabit.pl?userID=opendata&pass=opendata&cmd=login>
- SPARQL Endpoint
 - Datenbestand über eine SPARQL-Schnittstelle durchsuchbar machen: <http://bnb.data.bl.uk/sparql>
- RDF in Attributes (RDFa)
 - RDF-Daten in HTML-Seiten einbetten:

```
<div about="http://dbpedia.org/resource/Albert_Einstein">  
  <span property="foaf:name">Albert Einstein</span>  
  <span property="dbp:dateOfBirth" datatype="xsd:date">1879-03-14</span>  
  <div property="dbp:birthPlace" resource="http://dbpedia.org/resource/German_Empire">  
    <span property="dbp:conventionalLongName">the German Empire</span>  
  </div>  
</div>
```

Dereferenzierung

The screenshot shows the Restlet Client interface. The top navigation bar includes 'CLIENT', 'REQUESTS', 'SCENARIOS', 'Settings', 'Pricing', 'Help', and 'Sign in'. The main area is divided into 'REQUEST' and 'RESPONSE' sections.

REQUEST SECTION:

- Method: HTTP
- URL: `:// d-nb.info/gnd/4163853-0`
- Method: GET
- Buttons: Send
- HEADERS: A table with one entry:

Accept	: text/html
--------	-------------
- BODY: XHR does not allow payloads for GET request. or change a method definition in settings.

RESPONSE SECTION:

- Status: 303 See Other
- Cache Detected - Elapsed Time: 77ms
- HEADERS:

Date: Thu, 23 Feb 2017 20:28:29 GMT
Server: Apache
Vary: Accept
Location: <u>http://d-nb.info/gnd/4163853-0/about/html</u>
Content-Length: 52
Keep-Alive: timeout=15, max=100
Connection: Keep-Alive
Content-Type: application/xhtml+xml
- BODY: We are sorry, but Chrome API does not allow to get a response body for redirect.
- length: 80 Bytes

A red arrow points from the 'Accept' header in the request section to the 'Location' header in the response section, illustrating the dereferencing process.

Dereferenzierung

The screenshot shows the Restlet Client interface with the following details:

- REQUEST:** Method: HTTP, URL: `:// d-nb.info/gnd/4163853-0`, Method: GET. A "Send" button is visible.
- HEADERS:** An "Accept" header is set to `application/rdf+xml`. A note indicates "XHR does not allow payloads for GET request. or change a method definition in settings."
- RESPONSE:** Status: **303 See Other**. Headers include: Date: Thu, 23 Feb 2017 20:26:57 GMT; Server: Apache; Vary: Accept; Location: <http://d-nb.info/gnd/4163853-0/about/lds.rdf>; Content-Length: 55; Keep-Alive: timeout=15, max=100; Connection: Keep-Alive; Content-Type: application/rdf+xml.
- BODY:** A red error message states: "Unexpected character (W) at position 0." Below it, a grey message says: "We are sorry, but Chrome API does not allow to get a response body for redirect." The body length is 80 Bytes.

A red arrow points from the "Accept" header in the request section to the "Location" header in the response section.

Dereferenzierung

Restlet Client

chrome-extension://aejoelaoggembaahagimdiliamlcdfm/restlet_client.html

CLIENT REQUESTS SCENARIOS Settings Pricing Help Sign in

HTTP // d-nb.info/gnd/4163853-0/about/lds.rdf GET [0] Send

HEADERS: Accept: application/rdf+xml

BODY: XHR does not allow payloads for GET request. or change a method definition in settings.

RESPONSE: 200 OK

Cache Detected - Elapsed Time: 272ms

HEADERS: Date: Thu, 23 Feb 2017 20:34:24 GMT; Server: Apache; Link: <http://creativecommons.org/publicdomain/zero/1.0/>; rel="license"; Content-Disposition: inline; filename="4163853-0_lds.rdf"; Keep-Alive: timeout=15, max=99; Connection: Keep-Alive, Keep-Alive; Content-Length: 2185; Content-Type: application/rdf+xml;charset=UTF-8

BODY: <?xml version="1.0" encoding="UTF-8" ?><rdf:RDF xmlns:schema="http://schema.org/" xmlns:gndo="http://d-nb.info/standards/elementset" http://purl.org/dc/elements/1.1/"><rdf:Description rdf:about="http://d-nb.info/gnd/4163853-0"><rdf:type rdf:resource="http://d-nb.info/standards/elementset/gnd#SubjectHeadingSensoSt"><gndo:gndIdentifier rdf:datatype="http://www.w3.org/2001/XMLSchema#string"> 4163853-0 </><gndo:oldAuthorityNumber rdf:datatype="http://www.w3.org/2001/XMLSchema#string"> (DE-58<gndo:relatedDdcWithDegreeOfDeterminacy2 rdf:resource="http://dewey.info/class/398.8/"><gndo:variantNameForTheSubjectHeading rdf:datatype="http://www.w3.org/2001/XMLSchema#st"><gndo:preferredNameForTheSubjectHeading rdf:datatype="http://www.w3.org/2001/XMLSchema#><gndo:broaderTermGeneral rdf:resource="http://d-nb.info/gnd/4163822-0"/><gndo:gndSubjectCategory rdf:resource="http://d-nb.info/standards/vocab/gnd/gnd-sc#6.2a"><gndo:gndSubjectCategory r</rdf:Description>

Bibliotheksdaten als Linked Data (eine kleine Auswahl)

		Dump	SPARQL
DDB	http://www.dnb.de/DE/Service/DigitaleDienste/LinkedData/linkedata_node.html	✓	
British Library	http://bnb.data.bl.uk/	✓	✓
BnF	http://data.bnf.fr/	✓	✓
BVB/KoBV	http://lod.b3kat.de/doc/de/startseite/	✓	✓
OCLC	https://www.oclc.org/developer/develop/linked-data.en.html	(✓)	(✓)
STCN (KB Den Haag)	https://www.kb.nl/organisatie/onderzoek-expertise/informatie-infrastructuur-diensten-voor-bibliotheken/short-title-catalogue-netherlands-stcn/zoeken-in-de-stcn-met-sparql		✓
KB Sweden	http://libris.kb.se/sparql		✓
BNE	http://www.bne.es/en/Inicio/Perfiles/Bibliotecarios/DatosEnlazados/	✓	✓

Beispiel: DBpedia

- Gemeinsames Projekt von:
 - Univ. Mannheim
 - Univ. Leipzig
 - Hasso-Plattner-Institut für Softwaresystemtechnik Potsdam
 - OpenLink Software, Burlington Mass.
- Veröffentlicht seit 2007 die in Wikipedia entsprechend ausgezeichneten Informationen (Infoboxen, Kategorien etc.) als RDF Triples (Linked Data), verbunden mit weiteren offenen Datensammlungen
- Gegenwärtig (Sept. 2014) ca. 3 Mrd. RDF-Triples
- SPARQL Endpoints:
 - <https://dbpedia.org/sparql>
 - <http://dbpedia.org/snorql/>
- Sprachspezifische Unterprojekte
 - z.B. <http://fr.dbpedia.org/sparql>

Gemeinsame Normdatei (GND)

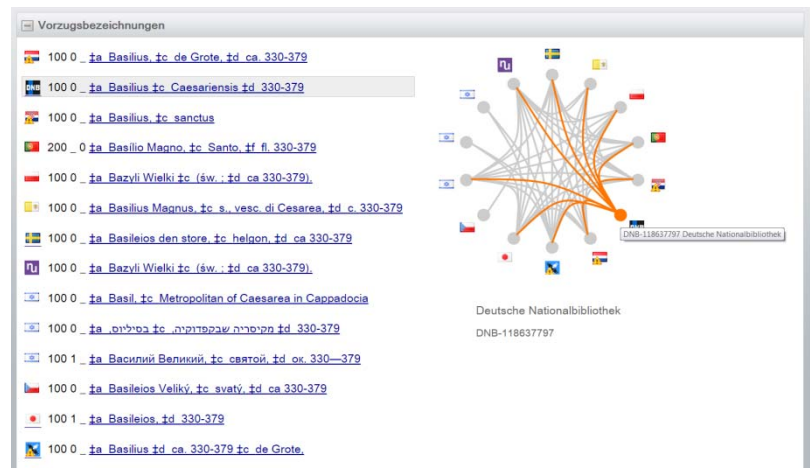
- führt seit April 2012 die
 - Personennamendatei (PND)
 - Gemeinsame Körperschaftsdatei (GKD)
 - Schlagwortnormdatei (SWD)
 - und einige lokale Normdateien

zusammen

- Koordiniert von der Deutschen Nationalbibliothek (DNB)
- Redaktionell betreut von den größeren wissenschaftlichen Bibliotheken in Deutschland und Österreich
- Verwendet von Bibliotheken, Kulturerbeeinrichtungen, verschiedenen DH Projekten und anderen (z. B. Wikipedia)
- Umfang > 10 Mio Datensätze

Virtual International Authority File (VIAF)

- Zusammenführung von derzeit 52 Normdateien aus dem Bibliotheks- und Kulturerbebereich
- Entstanden aus einem gemeinsamen Projekt der Dt. Nationalbibliothek und der Library of Congress
- Betreut und gehostet von OCLC (Dublin/Ohio)
- Enthält:
 - Personen
 - Körperschaften
 - Geographika
 - Werke
- Clustering statt Merging



International Standard Name Identifier (ISNI)

basilius von caesarea [more](#)

Examples: Prince; Princeton University; Ken Follett; Follett, Ken; Follet,

 **isni**
User id: ISNI | [log in](#) | [logout](#)


shortlist	title data	search history	1
results search [or] (Any Phrase (APH:)) basilius von caesarea 1 hits			

labels	sources data	marc21
---------------	--------------	--------

Please help us improve this record

If you have any supplemental information about the identity listed here, please click in this box to go to the contribution form.

Thank you in advance!


-  **ISNI:** 0000 0001 2102 9433
- Name:**
- Barseg (Kesarac'i)
 - Barseg Kesarac'i
 - Barseg (Sowrb)
 - Barsegh (Kesarats'i, Saint)
 - Barsegh, Kesariinsky (Santo)
 - Barsel (Kesaraçi)
 - Basil
 - Basil dari Kaisarea
 - Basil (le minime)
 - Basil (Metropolitan of Caesarea in Cappadocia)
 - Basil (of Caesarea)
 - Basil of Caesarea (Christian Saint)
 - Basil (of Caesarea, saint)
 - Basil of Caesarea (saint; 0329?-0379)
 - Basil (of Caesarea, St)
 - Basil of Caesarea (ок.329-379)
 - Basil (Saint)
 - Basil (saint; 0329?-0379)
 - Basil (Saint, Bishop of Caesarea)

Open Researcher and Contributor ID (ORCID)

- Identifier für wissenschaftliche Autoren
- Weiterentwicklung von Thompson Reuters' ResearcherID
- Registry service seit 2012
- Derzeit (Feb. 2017) > 3 Mio. Identifier vergeben
- Kompatibel mit ISNI
- ORCID ID ist gleichzeitig ein persistenter URI
- Dereferenzierungsservice zu RDF im Moment noch vorläufig

Alexander Jahnke

ORCID ID

 orcid.org/0000-0003-3572-6170



NIEDERSÄCHSISCHE STAATS- UND
UNIVERSITÄTSBIBLIOTHEK GÖTTINGEN

Vielen Dank für Ihre Aufmerksamkeit



GEORG-AUGUST-UNIVERSITÄT
GÖTTINGEN