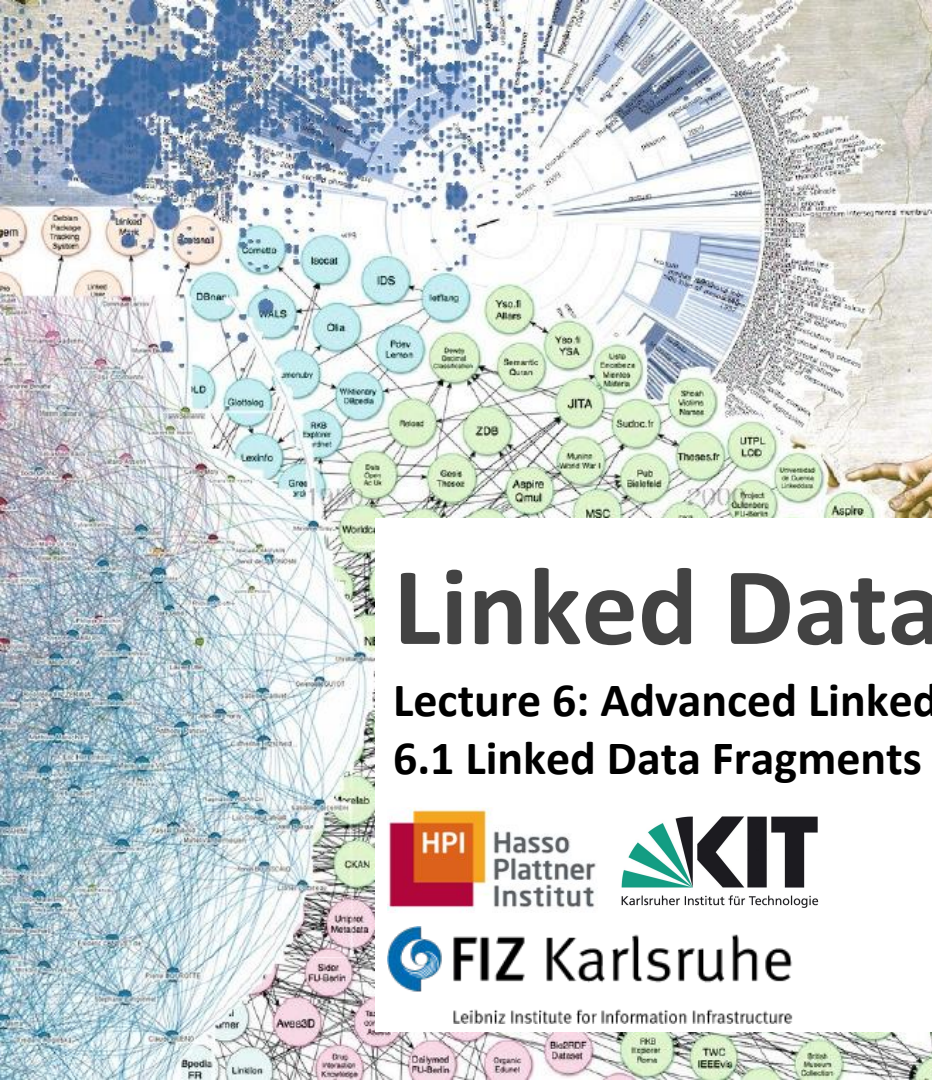




Linked Data Engineering

Lecture 6: Advanced Linked Data Based Applications

6.1 Linked Data Fragments



Leibniz Institute for Information Infrastructure

Prof. Dr. Harald Sack
FIZ Karlsruhe

Leibniz Institute for Information Infrastructure
Karlsruhe Institute of Technology

Autumn 2016



SPARQL revisited....

- A **SPARQL endpoint** allows clients to send any SPARQL query for evaluation.
[/sparql?query={query}](#)
 - What if clients send rather **complex queries**?
 - What if **many clients** send medium complex queries?
 - How to **mirror the server's data**?

- Most SPARQL endpoints have less than **95% of availability**
 - Average **downtime** ca. 1.5 days per month
 - It does not work for **(reliable) productive applications** to query publicly available Linked Data resources

Linked Data Availability on the Web

- Available **SPARQL Endpoints lack the necessary uptime**
 - High uptime possible, but only at high costs
- For most datasets no public SPARQL Endpoints exist
 - Data publishers provide **data dumps** instead, which **cannot be queried live**
- Users implement **local mirrors** of publicly available Linked Data resources to maintain availability
 - Synchronization problems
 - Is this still a “web of data”?

- A **Linked Data Fragment** of a Linked Data dataset is a set of RDF triples that consists of three parts:
 - **Data**
all triples of this dataset that match a specific selector
 - **Metadata**
triples that describe the dataset and/or the Linked Data Fragment
 - **Controls**
hypermedia links and/or forms that lead to other Linked Data Fragments

Linked Data Fragments

- **Data Dumps**

- **Data** all dataset triples
- **Metadata** dataset size, number of triples
- **Controls** (none)

Linked Data Fragments

- **SPARQL Endpoint**
 - **Data** result sets matching the query
 - **Metadata** (none)
 - **Controls** (none)

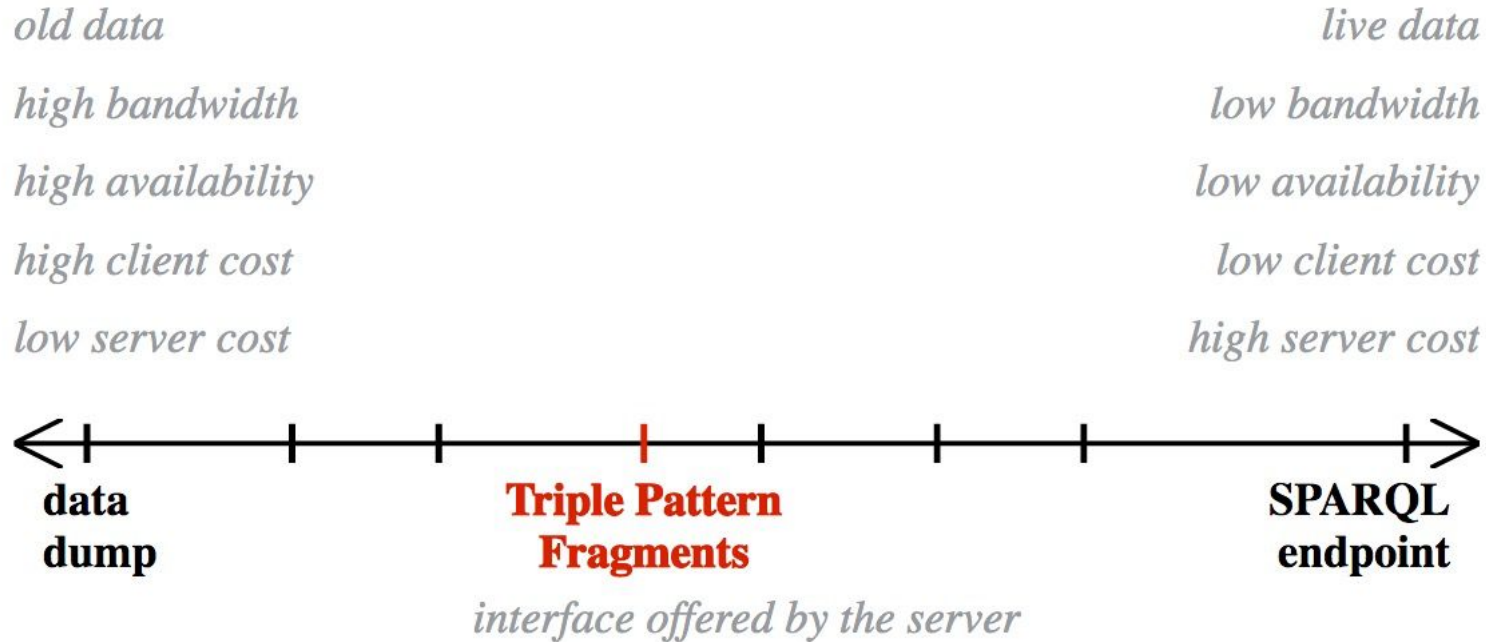
Linked Data Fragments

- **Triple Pattern Fragments**

{ ?subject ?predicate ?object. } - each component can be a variable or a constant

- **Data** matches of a triple pattern (paginated)
- **Metadata** total number of matches
- **Controls** access to other triple pattern fragments of the same dataset

Linked Data Interfaces and Triple Pattern Fragments



Triple Pattern Fragments

- **Triple Pattern Fragments** showing things in France

DBpedia – Linked Data Fragments

DBpedia 2016-04

Query DBpedia 2016-04 by triple pattern



subject:

predicate: <http://dbpedia.org/ontology/country>

object: <http://dbpedia.org/resource/France>

Find matching triples

Matches in DBpedia 2016-04 for { ?s <<http://dbpedia.org/ontology/country>> <<http://dbpedia.org/resource/France>> }

Showing triples 1 to 100 of ±43,343 with 100 triples per page. [next](#)

```
&ME country France.
01.002_Fighter_Squadron_%22Storks%22 country France.
01.003_Fighter_Squadron_%22Navarre%22 country France.
01.004_Fighter_Squadron_%22Dauphiné%22 country France.
01.007_Fighter_Squadron_%22Provence%22 country France.
01.091_Fighter_Squadron_%22Gascogne%22 country France.
02.003_Fighter_Squadron_%22Champagne%22 country France.
02.004_Fighter_Squadron_%22La_Fayette%22 country France.
02.005_Fighter_Squadron_%22Île-de-France%22 country France.
03.003_Fighter_Squadron_%22Ardennes%22 country France.
03.011_Fighter_Squadron_%22Corsica%22 country France.
1_rue_Sésame country France.
101st_Infantry_Division_(France) country France.
102nd_Infantry_Regiment_(France) country France.
10th_District_Court country France.
10th_Parachute_Division_(France) country France.
10th_arrondissement_of_Marseille country France.
10th_arrondissement_of_Paris country France.
11Eleven_Project country France.
11th_Army_Corps_(France) country France.
11th_Foreign_Infantry_Regiment_(France) country France.
```

[Link to this Triple Pattern Query](#)

- **Triple Pattern Fragments** allow for a less expensive data infrastructure
 - A SPARQL server can serve as a possible back-end
 - Queries are relatively simple
 - The Header-Dictionary-Triple format (HDT) can be used for compressed storage of RDF triples (<http://www.rdfhdt.org/>)
 - Complex queries can be managed in a more intelligent client
- **Triple Pattern Fragments** allow for
 - **Simple servers** and **smart clients**

Triple Pattern Fragments - Query Execution

- **Smart Client**
 - receives a Query and the URI of a Triple Pattern Fragment (TPF) of a dataset
 - uses TPF controls to determine access to other TPFs
 - uses TPF metadata to determine query execution

Triple Pattern Fragments - Query Execution

- **Example:**

Search for artists who have something in common with Neil Armstrong

```
PREFIX dbr: <http://dbpedia.org/resource/>
PREFIX dbo: <http://dbpedia.org/ontology/>
PREFIX dct: <http://purl.org/dc/terms/>

SELECT DISTINCT ?artist ?common
WHERE {
    dbr:Neil_Armstrong dct:subject ?common.
    ?artist a dbo:Artist;
            dct:subject ?common.
}
```

[Link to SPARQL query](#)

Triple Pattern Fragments - Query Execution

- GET first page of all Triple Pattern Fragments and inspect their metadata

```
dbr:Neil_Armstrong dct:subject ?common
```

[39 triples](#)

```
?artist a dbo:Artist
```

[145,878 triples](#)

```
?artist dct:subject ?common
```

[43,723,493 triples](#)

Triple Pattern Fragments - Query Execution

- Execution starts with smallest Fragment because it is most selective

dbr:Neil_Armstrong dct:subject dbr:Category:1930_births .

[39 triples](#)

```
:Neil_Armstrong dct:subject dbr:Category:1930_births.  
:Neil_Armstrong dct:subject dbr:Category:1966_in_spaceflight.  
:Neil_Armstrong dct:subject dbr:Category:1969_in_spaceflight.
```

?artist a dbo:Artist

[145,878 triples](#)

?artist dct:subject ?common

[43,723,493 triples](#)

Triple Pattern Fragments - Query Execution

- Execution continues recursively until all options have been tested

dbr:Neil_Armstrong dct:subject dbr:Category:1930_births .

[39 triples](#)

~~:Neil_Armstrong dct:subject dbr:Category:1930_births.~~
:Neil_Armstrong dct:subject dbr:Category:1966_in_spaceflight.
:Neil_Armstrong dct:subject dbr:Category:1969_in_spaceflight.

dbr:10,000_Maniacs a dbo:Artist .

[145,878 triples](#)

dbr:10,000_Maniacs rdf:type dbo:Artist.
dbr:100_Demons rdf:type dbo:Artist.
dbr:100_Kila rdf:type dbo:Artist.

?artist dct:subject ?common

[43,723,493 triples](#)

Triple Pattern Fragments - Query Execution

- Execution continues recursively until all options have been tested

dbr:Neil_Armstrong dct:subject dbr:Category:1930_births .

[39 triples](#)

~~:Neil_Armstrong dct:subject dbr:Category:1930_births.~~
:Neil_Armstrong dct:subject dbr:Category:1966_in_spaceflight.
:Neil_Armstrong dct:subject dbr:Category:1969_in_spaceflight.

dbr:10,000_Maniacs a dbo:Artist .

[145,878 triples](#)

~~dbr:10,000_Maniacs rdf:type dbo:Artist.~~
dbr:100_Demons rdf:type dbo:Artist.
dbr:100_Kila rdf:type dbo:Artist.

dbr:10,000_Maniacs dct:subject dbr:Category:1930_births .

0 triples

Triple Pattern Fragments - Query Execution

- It gets the first page of the fragment, which provides mappings for a solution.

```
dbr:A._T._Mahmud dct:subject dbr:Category:1930_births .      1 triples
```

- Found a solution:
 - **?artist** dbr:A._T._Mahmud
 - **?common** dbr:Category:1930_births

[Link to LDF query](#)

Query Linked Data on the Web

Live in your browser, powered by Triple Pattern Fragments.



Choose datasources:

DBpedia 2016-04 ✕



Type or pick a query:

Award ceremonies in Dutch speaking countries ▼

```
SELECT DISTINCT ?artist ?common
WHERE {
  dbr:Neil_Armstrong dct:subject ?common.
  ?artist a dbo:Artist;
    dct:subject ?common.
}
```

Execute query

14 results in 1.9s

Query results:

?artist	http://dbpedia.org/resource/A._T._Mahmud
?common	http://dbpedia.org/resource/Category:1930_births

?artist	http://dbpedia.org/resource/Abbey_Lincoln
?common	http://dbpedia.org/resource/Category:1930_births

?artist	http://dbpedia.org/resource/Abdul_Rahim_Sarban
?common	http://dbpedia.org/resource/Category:1930_births

?artist	http://dbpedia.org/resource/Ada_Madssen
?common	http://dbpedia.org/resource/Category:1930_births

?artist	http://dbpedia.org/resource/Adel_Rootstein
?common	http://dbpedia.org/resource/Category:1930_births

?artist	http://dbpedia.org/resource/Adge_Cutler
----------------	---

[Link to LDF query](#)

<http://linkeddatafragments.org/>



Linked Data Fragments

Query the Web of data on *Web-scale* by moving intelligence from servers to clients.

Watch this talk to learn the foundations of Linked Data Fragments.

A huge amount of Linked Data is available on the Web. But can live applications use it?
SPARQL endpoints are expensive for the server, and not always available for all datasets.
Downloadable dumps are expensive for clients, and do not allow live querying on the Web.

Thanks a lot to
Ruben Verborgh
from Ghent University/imec
<https://ruben.verborgh.org/>



Next: 02 Metadata and Semantic Annotation

OpenHPI - Course Linked Data Engineering - Lecture 6: Advanced Linked Data Based Applications