

RDF and Linked Data

Olaf Hartig

Linköping University

TDDD43
Advanced Data Models and Databases

Sep. 25, 2017

Outline

(1) RDF

(2) SPARQL

(3) Linked Data

Sep. 25, 2017

Olaf Hartig - RDF and Linked Data - TDDD43

2/41

RDF in General

- **Resource Description Framework**
- A **resource** may basically be everything
 - e.g. persons, places, Web documents, abstract concepts
- **Descriptions** of resources
 - Attributes
 - Relationships
- The **framework** contains:
 - A data model, and
 - Languages and syntaxes



Sep. 25, 2017

Olaf Hartig - RDF and Linked Data - TDDD43

3/41

RDF Data Model

- Data comes as a set of **triples** (subject, predicate, object)
- **Subject**: resources
- **Predicate**: properties
- **Object**: literals or resources
- **Examples**:
 - (Mount Baker , last eruption , 1880)
 - (Mount Baker , location , Washington)

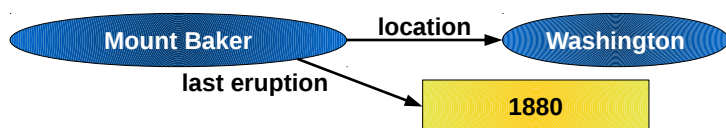
Sep. 25, 2017

Olaf Hartig - RDF and Linked Data - TDDD43

4/41

RDF Data Model (cont'd)

- **RDF based data may be understood as a graph**:
 - Triples as directed edges
 - Subjects and objects as vertices
 - Edges labeled by predicate
- **Example**:
 - (Mount Baker , last eruption , 1880)
 - (Mount Baker , location , Washington)



Sep. 25, 2017

Olaf Hartig - RDF and Linked Data - TDDD43

5/41

Uniform Resource Identifier (URI)

- **URIs extend the concept of URLs**
 - Globally **unique identifier** for resources
 - URL of a Web document usually used as its URI
 - Attention: URIs identify not only Web documents
- **Example**:
 - Me:
<http://olafhartig.de/foaf.rdf#olaf>
 - RDF document about me:
<http://olafhartig.de/foaf.rdf>
 - HTML document about me:
<http://olafhartig.de/index.html>

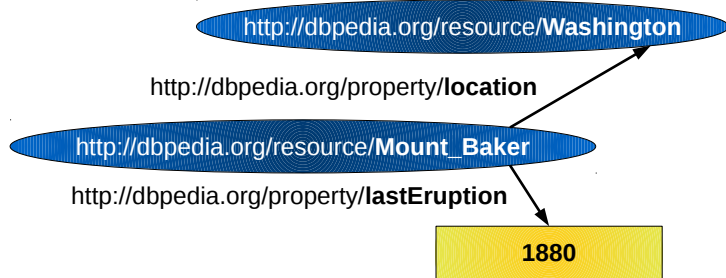
Sep. 25, 2017

Olaf Hartig - RDF and Linked Data - TDDD43

6/41

Example (revisited)

- (`http://dbpedia.org/resource/Mount_Baker`,
`http://dbpedia.org/property/lastEruption`, 1880)
- (`http://dbpedia.org/resource/Mount_Baker`,
`http://dbpedia.org/property/location`,
`http://dbpedia.org/resource/Washington`)



Sep. 25, 2017

Olaf Hartig - RDF and Linked Data - TDDD43

7/41

Compact URI (CURIE)

- **Abbreviated Notation for URIs**
- **Syntax:**
 - Prefix name (references a prefix of the URI)
 - Colon character (":")
 - Reference part
- **URI by concatenating the prefix and the reference part**
- **Examples:**
 - `dbpedia:Mount_Baker` for
`http://dbpedia.org/resource/Mount_Baker`
 - `myfoaf:olaf` for
`http://olafhartig.de/foaf.rdf#olaf`

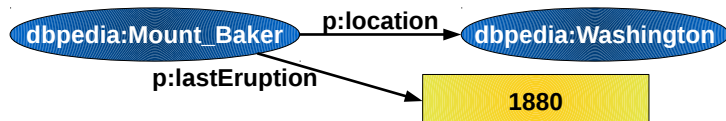
Sep. 25, 2017

Olaf Hartig - RDF and Linked Data - TDDD43

8/41

Example with CURIES

- **Using**
 - `dbpedia` for prefix `http://dbpedia.org/resource/`
 - `p` for prefix `http://dbpedia.org/property/`
- **we have**
 - (`dbpedia:Mount_Baker`, `p:lastEruption`, 1880)
 - (`dbpedia:Mount_Baker`, `p:location`, `dbpedia:Washington`)



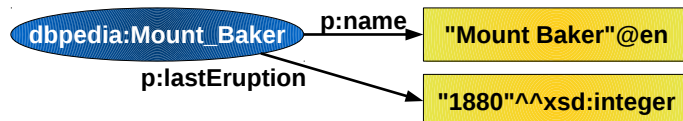
Sep. 25, 2017

Olaf Hartig - RDF and Linked Data - TDDD43

9/41

Literals

- **Literals may occur in the object position of RDF triples**
- **Represented by strings**
- **Literal strings interpreted by datatypes**
 - Datatype identified by a URI
 - Common to use the XML Schema datatypes
 - If no datatype, then interpreted as `xsd:string`
- **Untyped literals may have language tags (e.g. @de)**



Sep. 25, 2017

Olaf Hartig - RDF and Linked Data - TDDD43

10/41

Turtle: A Readable Syntax for RDF

- **Simple, human-readable notation to list RDF triples:**
 - Triples separated by a period (".") character
 - Example:

```
<http://dbpedia.org/resource/Mount_Baker>
  <http://dbpedia.org/property/lastEruption>
    "1880"^^xsd:integer .
<http://dbpedia.org/resource/Mount_Baker>
  <http://dbpedia.org/property/location>
    <http://dbpedia.org/resource/Washington> .
```

Sep. 25, 2017

Olaf Hartig - RDF and Linked Data - TDDD43

11/41

Turtle supports CURIES

- **@prefix** directive binds a prefix to a namespace URI

```
@prefix dbpedia : <http://dbpedia.org/resource/> .
@prefix p : <http://dbpedia.org/property/> .
@prefix xsd: <http://www.w3.org/2001/XMLSchema#> .

dbpedia:Mount_Baker p:lastEruption "1880"^^xsd:integer .
dbpedia:Mount_Baker p:location dbpedia:Washington .

dbpedia:Washington p:boundingstates dbpedia:Oregon .
dbpedia:Washington p:boundingstates dbpedia:Idaho .
```

Sep. 25, 2017

Olaf Hartig - RDF and Linked Data - TDDD43

12/41

Turtle provides some syntactic sugar

- Property lists separated by a semicolon (“;”) character
- Object lists separated by a comma (“,”) character

```
@prefix dbpedia : <http://dbpedia.org/resource/> .
@prefix p : <http://dbpedia.org/property/> .
@prefix xsd: <http://www.w3.org/2001/XMLSchema#> .

dbpedia:Mount_Baker p:lastEruption "1880"^^xsd:integer ;
                    p:location      dbpedia:Washington .

dbpedia:Washington p:borderingstates dbpedia:Oregon ,
                    dbpedia:Idaho .
```

Sep. 25, 2017

Olaf Hartig - RDF and Linked Data - TDDD43

13/41

Turtle, some more syntactic sugar

- Shortcuts for number literals

```
dbpedia:Mount_Baker p:lastEruption "1880"^^xsd:integer ;
                    geo:lat  "48.777222"^^xsd:float ;
                    geo:long "-121.813332"^^xsd:float .
```

Equivalent:

```
dbpedia:Mount_Baker p:lastEruption 1880 ;
                    geo:lat  48.777222 ;
                    geo:long -121.813332 .
```

Sep. 25, 2017

Olaf Hartig - RDF and Linked Data - TDDD43

14/41

RDF/XML – An XML Syntax for RDF

```
@prefix dbpedia : <http://dbpedia.org/resource/> .
@prefix p : <http://dbpedia.org/property/> .
@prefix xsd: <http://www.w3.org/2001/XMLSchema#> .

dbpedia:Mount_Baker p:lastEruption "1880"^^xsd:integer .
dbpedia:Mount_Baker p:location      dbpedia:Washington .
```

Turtle

```
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:p="http://dbpedia.org/property/">
  <rdf:Description rdf:about="http://dbpedia.org/resource/Mount_Baker">
    <p:lastEruption >
      rdf:datatype="http://www.w3.org/2001/XMLSchema#integer" >
        >1880</p:lastEruption>
    <p:location rdf:resource="http://dbpedia.org/resource/Washington"/>
  </rdf:Description>
</rdf:RDF>
```

RDF/XML

Sep. 25, 2017

Olaf Hartig - RDF and Linked Data - TDDD43

15/41

Outline

(1) RDF ✓

(2) SPARQL

(3) Linked Data

Sep. 25, 2017

Olaf Hartig - RDF and Linked Data - TDDD43

16/41

SPARQL in General

- A family of W3C recommendations
- **SPARQL Query**
 - Declarative query language for RDF data
 - Our focus today
- **SPARQL Update**
 - Declarative update language for RDF data
- **SPARQL Protocol**
 - Communication between SPARQL processing services (a.k.a. SPARQL endpoints) and clients
- **SPARQL Query Results XML Format**
 - XML format for serializing query results

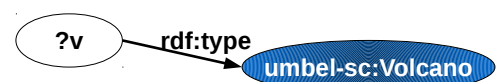
Sep. 25, 2017

Olaf Hartig - RDF and Linked Data - TDDD43

17/41

Main Idea of SPARQL Queries

- **Pattern matching:**
 - Describe subgraphs of the queried RDF graph
 - Subgraphs that match the description yield a result
 - Mean: **graph patterns** (essentially, RDF graphs with variables)



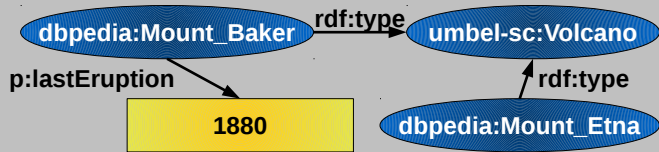
Sep. 25, 2017

Olaf Hartig - RDF and Linked Data - TDDD43

18/41

Main Idea of SPARQL Queries

Queried RDF graph:



Result:

?v
dbpedia:Mount_Baker
dbpedia:Mount_Etna

Sep. 25, 2017

Olaf Hartig - RDF and Linked Data - TDDD43

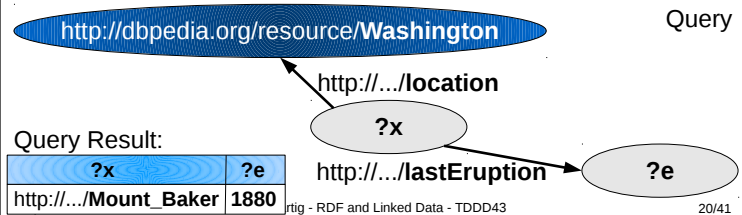
19/41

Another Example



Data

Query



Query Result:

?x	?e
http://.../Mount_Baker	1880

Olaf Hartig - RDF and Linked Data - TDDD43

20/41

Components of a SPARQL Query

```

PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX umbel-sc: <http://umbel.org/umbel/sc/>
SELECT ?v
FROM <http://example.org/myGeoData>
WHERE {
  ?v rdf:type umbel-sc:Volcano .
}
ORDER BY ?name
  
```

Sep. 25, 2017

Olaf Hartig - RDF and Linked Data - TDDD43

21/41

Components of a SPARQL Query

```

PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX umbel-sc: <http://umbel.org/umbel/sc/>
SELECT ?v
FROM <http://example.org/myGeoData>
WHERE {
  ?v rdf:type umbel-sc:Volcano .
}
ORDER BY ?name
  
```

• Prologue:

- Prefix definitions for using compact URIs (CURIEs)

Sep. 25, 2017

Olaf Hartig - RDF and Linked Data - TDDD43

22/41

Components of a SPARQL Query

```

PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX umbel-sc: <http://umbel.org/umbel/sc/>
SELECT ?v
FROM <http://example.org/myGeoData>
WHERE {
  ?v rdf:type umbel-sc:Volcano .
}
ORDER BY ?name
  
```

• Result form specification:

- SELECT for projection (similar to projection in relational algebra)
- Other forms: DESCRIBE, CONSTRUCT, and ASK

Sep. 25, 2017

Olaf Hartig - RDF and Linked Data - TDDD43

23/41

Components of a SPARQL Query

```

PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX umbel-sc: <http://umbel.org/umbel/sc/>
SELECT ?v
FROM <http://example.org/myGeoData>
WHERE {
  ?v rdf:type umbel-sc:Volcano .
}
ORDER BY ?name
  
```

• Dataset specification:

- Specify the RDF dataset to be queried (use URIs that identify particular RDF graphs in your RDF database)

Sep. 25, 2017

Olaf Hartig - RDF and Linked Data - TDDD43

24/41

Components of a SPARQL Query

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX umbel-sc: <http://umbel.org/umbel/sc/>
SELECT ?v
FROM <http://example.org/myGeoData>
WHERE {
  ?v rdf:type umbel-sc:Volcano .
}
ORDER BY ?name
```

• Query Pattern:

- WHERE clause specifies the graph pattern to be matched

Sep. 25, 2017

Olaf Hartig - RDF and Linked Data - TDDD43

25/41

Basic Graph Pattern

- Set of triple patterns (i.e., RDF triples with variables)
- Variable names prefixed with “?” (or “\$”)
- Turtle syntax
 - Including syntactic sugar (e.g., property and object lists)

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX umbel-sc: <http://umbel.org/umbel/sc/>
SELECT ?name
WHERE {
  ?v rdf:type umbel-sc:Volcano ;
    rdfs:label ?name .
}
```

Sep. 25, 2017

Olaf Hartig - RDF and Linked Data - TDDD43

26/41

Basic Graph Pattern (Example)

```
dbpedia:Mount_Etna rdf:type umbel-sc:Volcano ;      Data*
                    rdfs:label "Etna" .
dbpedia:Mount_Baker rdf:type umbel-sc:Volcano .
dbpedia:Beerenberg rdf:type umbel-sc:Volcano,
                    umbel-sc:NaturalElevation ;
                    rdfs:label "Beerenberg"@en ;
                    rdfs:label "Бееренберг"@ru .
```

• Question: What are the names of all (known) volcanos?

```
SELECT ?name WHERE {      Query*
  ?v rdf:type umbel-sc:Volcano ;
    rdfs:label ?name . }
```

Result:

?name
"Etna"
"Бееренберг"@ru
"Beerenberg"@en

*Prefix declarations omitted

Sep. 25, 2017

Olaf Hartig - RDF and Linked Data - TDDD43

27/41

Optional Graph Pattern

```
dbpedia:Mount_Etna rdf:type umbel-sc:Volcano ;      Data
                    rdfs:label "Etna" .
dbpedia:Mount_Baker rdf:type umbel-sc:Volcano .
dbpedia:Beerenberg rdf:type umbel-sc:Volcano ;
                    rdfs:label "Beerenberg"@en .
```

• Question: What are *all* (known) volcanos and their names?

```
SELECT ?v ?name WHERE {      Query
  ?v rdf:type umbel-sc:Volcano ;
    rdfs:label ?name . }
```

- Problem: Mount Baker *missing* (b/c no name in the data)

?v	?name
dbpedia:Mount_Etna	"Etna"
dbpedia:Beerenberg	"Beerenberg"@en

Sep. 25, 2017

Olaf Hartig - RDF and Linked Data - TDDD43

28/41

Optional Graph Pattern

• Keyword OPTIONAL indicates optional patterns

```
SELECT ?v ?name WHERE {      Query
  ?v rdf:type umbel-sc:Volcano .
  OPTIONAL { ?v rdfs:label ?name }
}
```

?v	?name
dbpedia:Mount_Etna	"Etna"
dbpedia:Mount_Baker	
dbpedia:Beerenberg	"Beerenberg"@en

• Optional patterns may result in unbound variables

Sep. 25, 2017

Olaf Hartig - RDF and Linked Data - TDDD43

29/41

Constraints on Solutions

• Syntax: Keyword FILTER followed by filter expression

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX umbel-sc: <http://umbel.org/umbel/sc/>
PREFIX p: <http://dbpedia.org/property/>
SELECT ?v
WHERE {
  ?v rdf:type umbel-sc:Volcano ;
    p:lastEruption ?le .
  FILTER ( ?le > 1900 )
}
```

- Filter expressions contain operators and functions
- Operators and functions operate on RDF terms

Sep. 25, 2017

Olaf Hartig - RDF and Linked Data - TDDD43

30/41

Unary Operators in Constraints

Operator	Type(A)	Result type
! A	xsd:boolean	xsd:boolean
+ A	numeric	numeric
- A	numeric	numeric
BOUND(A)	variable	xsd:boolean
isURI(A)	RDF term	xsd:boolean
isBLANK(A)	RDF term	xsd:boolean
isLITERAL(A)	RDF term	xsd:boolean
STR(A)	literal / URI	simple literal
LANG(A)	literal	simple literal
DATATYPE(A)	literal	simple literal

Sep. 25, 2017

Olaf Hartig - RDF and Linked Data - TDDD43

31/41

Binary and other Operators

- **Logical connectives && and ||**
 - for xsd:boolean
- **Comparison operators =, !=, <, >, <=, and >=**
 - for numeric datatypes, xsd:dateTime, xsd:string, xsd:boolean
- **Comparison operators = and !=**
 - for other datatypes
- **Arithmetic operators +, -, *, and /**
 - for numeric datatypes
- **Furthermore:**
 - REGEX(*String*,*Pattern*) or REGEX(*String*,*Pattern*,*Flags*)
 - langMATCHES(*A*,*B*)
 - etc.

Sep. 25, 2017

Olaf Hartig - RDF and Linked Data - TDDD43

32/41

Constraints (Example)

```
dbpedia:Mount_Etna rdf:type umbel-sc:Volcano ;
                    rdfs:label "Etna" .
dbpedia:Beerenberg rdf:type umbel-sc:Volcano,
                           umbel-sc:NaturalElevation ;
                    rdfs:label "Beerenberg"@en ;
                    rdfs:label "Бееренберг"@ru .
```

- **Question:** What volcanos have an "e" in their name?

```
SELECT ?v WHERE {
  ?v rdf:type umbel-sc:Volcano ;
     rdfs:label ?name .
  FILTER( REGEX(STR(?name),"e") )
}
```

```
?v
dbpedia:Beerenberg
dbpedia:Beerenberg
```

Sep. 25, 2017

Olaf Hartig - RDF and Linked Data - TDDD43

33/41

Constraints (Example, cont'd)

```
dbpedia:Mount_Etna rdf:type umbel-sc:Volcano ;
                    rdfs:label "Etna" .
dbpedia:Beerenberg rdf:type umbel-sc:Volcano,
                           umbel-sc:NaturalElevation ;
                    rdfs:label "Beerenberg"@en ;
                    rdfs:label "Бееренберг"@ru .
```

- **Question:** What volcanos have an "e" in their name?

```
SELECT ?v WHERE {
  ?v rdf:type umbel-sc:Volcano ;
     rdfs:label ?name .
  FILTER( REGEX(STR(?name),"e","i") )
}
```

```
?v
dbpedia:Mount_Etna
dbpedia:Beerenberg
dbpedia:Beerenberg
```

Sep. 25, 2017

Olaf Hartig - RDF and Linked Data - TDDD43

34/41

Components of a SPARQL Query

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX umbel-sc: <http://umbel.org/umbel/sc/>
SELECT ?v
FROM <http://example.org/myGeoData>
WHERE {
  ?v rdf:type umbel-sc:Volcano .
}
ORDER BY ?name
```

- **Solution modifiers:**

- Only for SELECT queries
- Modify the [result set](#) as a whole (not single solutions)
- Keywords: DISTINCT, ORDER BY, LIMIT, and OFFSET

Sep. 25, 2017

Olaf Hartig - RDF and Linked Data - TDDD43

35/41

SPARQL 1.1

- **New features of SPARQL 1.1 Query:**

- Aggregate functions (e.g., COUNT, SUM, AVG)
- Sub-queries
- Negation (EXISTS, NOT EXISTS, MINUS)
- Assignments (e.g., BIND, SELECT expressions)
- Property paths (navigation à la XPath)
- Basic query federation (SERVICE, BINDINGS)

- **SPARQL 1.1 Update:**

- Graph update (INSERT DATA, DELETE DATA, INSERT, DELETE, DELETE WHERE, LOAD, CLEAR)
- Graph management (CREATE, DROP, COPY, MOVE, ADD)

Sep. 25, 2017

Olaf Hartig - RDF and Linked Data - TDDD43

36/41

Outline

- (1) RDF ✓
- (2) SPARQL ✓
- (3) Linked Data

Sep. 25, 2017

Olaf Hartig - RDF and Linked Data - TDDD43

37/41

Data on the Web

The image shows two screenshots side-by-side. On the left is the IMDb page for the movie 'War Child' (1999), with a yellow arrow pointing from the IMDb database icon below to a central yellow box. On the right is the CIA World Factbook page for Albania, with a yellow arrow pointing from the CIA World Factbook database icon below to the same central yellow box. The central yellow box contains the text: 'Traditionally, published in HTML documents that are designed for human consumption'.

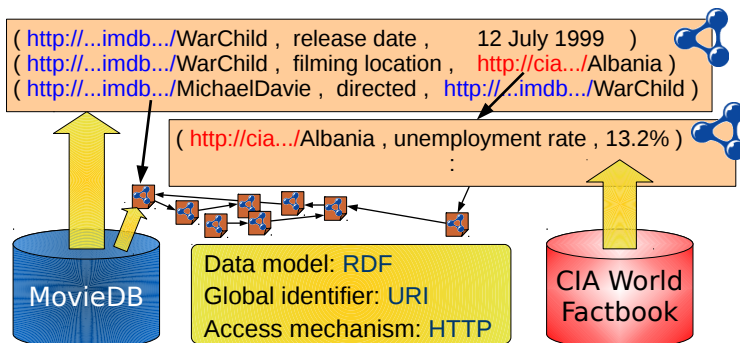
Sep. 25, 2017

Olaf Hartig - RDF and Linked Data - TDDD43

38/41

Linked Data Publishing Principles

- Goal: publishing structured data on the WWW in a standardized, machine-readable manner



Sep. 25, 2017

Olaf Hartig - RDF and Linked Data - TDDD43

39/41

Adoption

- Started as a grassroots community effort in 2007
 - Publish existing, open license datasets as Linked Data
 - Interlink things between different data sources
- Prominent publishers joined the effort
 - e.g., BBC, NY Times, Library of Congress, Thomson Reuters, Springer, Nature, Best Buy, Sears, Renault, UK Government
- Numbers
 - BTC 2014 Crawl (February – June 2014): ca. 4.1B triples in ca. 44M docs from ca. 48K sites
 - Iodstats (as of September 24, 2017): ca. 149B triples in ca. 3.0K datasets



Sep. 25, 2017

Olaf Hartig - RDF and Linked Data - TDDD43

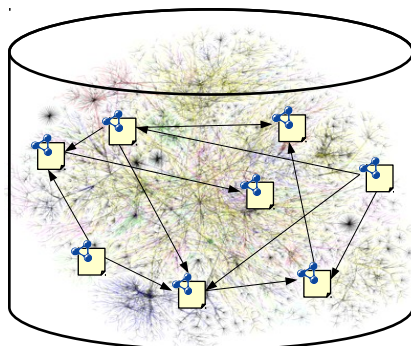
40/41

The Web of Linked Data

...a globally distributed network of data
...which we may understand as a huge distributed database

How do we enable applications to query this data?

- Active area of research!
- Looking for a thesis topic in this area? Contact me!



Sep. 25, 2017

Olaf Hartig - RDF and Linked Data - TDDD43

41/41