Conceiving the Web of Linked Data as a Database

Olaf Hartig

Database and Information Systems Research Group
Humboldt-Universität zu Berlin
An Analogy ...
Traditional Database Setting
Distributed Database

Olaf Hartig - Conceiving the Web of Linked Data as a Database
SPARQL Endpoint Federation
What's different for Linked Data?
What's different for Linked Data?
What's different for Linked Data?

• No (a priori) information
What's different for Linked Data?

- No (a priori) information
- Infinite number of locations
What's different for Linked Data?

- No (a priori) information
- Infinite number of locations
What's different for Linked Data?

- No (a priori) information
- Infinite number of locations
- Infinite number of “cards”
Research Question

How to query this kind of “database”? 
A Novel Query Execution Paradigm

Link Traversal Based Query Execution
Main Idea

- **Intertwine query evaluation with traversal of data links**

- **We alternate between:**
  - **Evaluate parts** of the query (triple patterns) on a continuously augmented set of data
  - **Look up URIs** in intermediate solutions and add retrieved data to the query-local dataset
Main Idea

- Intertwine query evaluation with traversal of data links

- We alternate between:
  - Evaluate parts of the query (triple patterns) on a continuously augmented set of data
  - Look up URIs in intermediate solutions and add retrieved data to the query-local dataset
Main Idea

- Intertwine query evaluation with traversal of data links

- We alternate between:
  - Evaluate parts of the query (triple patterns) on a continuously augmented set of data
  - Look up URIs in intermediate solutions and add retrieved data to the query-local dataset
Main Idea

- Intertwine query evaluation with traversal of data links

- We alternate between:
  - Evaluate parts of the query (triple patterns) on a continuously augmented set of data
  - Look up URIs in intermediate solutions and add retrieved data to the query-local dataset
Main Idea

- Intertwine query evaluation with traversal of data links

- We alternate between:
  - Evaluate parts of the query (triple patterns) on a continuously augmented set of data
  - Look up URIs in intermediate solutions and add retrieved data to the query-local dataset
Main Idea

- Intertwine query evaluation with traversal of data links

- We alternate between:
  - Evaluate parts of the query (triple patterns) on a continuously augmented set of data
  - Look up URIs in intermediate solutions and add retrieved data to the query-local dataset
Main Idea

- Intertwine query evaluation with traversal of data links
- We alternate between:
  - Evaluate parts of the query (triple patterns) on a continuously augmented set of data
  - Look up URIs in intermediate solutions and add retrieved data to the query-local dataset
Main Idea

- Intertwine query evaluation with traversal of data links

- We alternate between:
  - Evaluate parts of the query (triple patterns) on a continuously augmented set of data
  - Look up URIs in intermediate solutions and add retrieved data to the query-local dataset
Main Idea

- Intertwine query evaluation with traversal of data links
- We alternate between:
  - Evaluate parts of the query (triple patterns) on a continuously augmented set of data
  - Look up URIs in intermediate solutions and add retrieved data to the query-local dataset
Main Idea

- Intertwine query evaluation with traversal of data links
- We alternate between:
  - Evaluate parts of the query (triple patterns) on a continuously augmented set of data
  - Look up URIs in intermediate solutions and add retrieved data to the query-local dataset
Main Idea

- Intertwine query evaluation with traversal of data links
- We alternate between:
  - Evaluate parts of the query (triple patterns) on a continuously augmented set of data
  - Look up URIs in intermediate solutions and add retrieved data to the query-local dataset
Main Idea

- Intertwine query evaluation with traversal of data links

- We alternate between:
  - Evaluate parts of the query (triple patterns) on a continuously augmented set of data
  - Look up URIs in intermediate solutions and add retrieved data to the query-local dataset
Main Idea

- Intertwine query evaluation with traversal of data links

- We alternate between:
  - Evaluate parts of the query (triple patterns) on a continuously augmented set of data
  - Look up URIs in intermediate solutions and add retrieved data to the query-local dataset
Main Idea

- Intertwine query evaluation with traversal of data links
- We alternate between:
  - Evaluate parts of the query (triple patterns) on a continuously augmented set of data
  - Look up URIs in intermediate solutions and add retrieved data to the query-local dataset
Main Idea

- Intertwine query evaluation with traversal of data links

- We alternate between:
  - Evaluate parts of the query (triple patterns) on a continuously augmented set of data
  - Look up URIs in intermediate solutions and add retrieved data to the query-local dataset
Main Idea

- Intertwine query evaluation with traversal of data links
- We alternate between:
  - Evaluate parts of the query (triple patterns) on a continuously augmented set of data
  - Look up URIs in intermediate solutions and add retrieved data to the query-local dataset

```
<table>
<thead>
<tr>
<th>?actor</th>
<th>?loc</th>
</tr>
</thead>
</table>
```

```
http://.../movie2449

?actor -> actor_in -> filmingLocation

{lives_in} ?actor -> ?loc
```

"Discovered data"
“Real World” Example

SELECT DISTINCT ?author ?phone WHERE {
  FILTER regex( str(?topicLabel), "ontology engineering", "i" ) .

  { ?author owl:sameAs ?authorAlt }
  UNION
  { ?authorAlt owl:sameAs ?author }

  ?authorAlt foaf:phone ?phone }

Return phone numbers of authors of ontology engineering papers at ESWC'09.

<table>
<thead>
<tr>
<th>Result size</th>
<th>2</th>
</tr>
</thead>
<tbody>
<tr>
<td># of retrieved graphs</td>
<td>297</td>
</tr>
<tr>
<td># of accessed servers</td>
<td>16</td>
</tr>
<tr>
<td>avg. execution time</td>
<td>1min 30sec</td>
</tr>
</tbody>
</table>
What's different for Linked Data?

- No (a priori) information
- Infinite number of locations
- Infinite number of "cards"