Querying the Wikidata Knowledge Graph

Peter Haase
LDBC Meeting, Barcelona
19.3.2015

metaphacts.
metaphacts GmbH

- Founded 10/2014
- Currently team of four
- Based in Walldorf
- Partnership with Systap
Portfolio: Software, solutions, services for knowledge graphs

- Storing and querying of knowledge graphs
  - Scalable databases for big graphs, building on bigdata/Blazegraph
  - High-performance graph analytics, based on MapGraph / GAS
  - Light-weight reasoning with large-scale knowledge graphs

- Creation and curation of knowledge graphs
  - Semi-automatic creation of knowledge graphs from existing sources
  - Data integration and ontology-based data access
  - Collaborative management of knowledge graphs

- Application development utilizing knowledge graphs
  - Rapid development of end-user oriented applications
  - Visualization of knowledge graphs, semantic search
  - Mobile applications & augmented and virtual reality
Wikidata

- **Collecting structured data.** Unlike the Wikipedias, which produce encyclopedic articles, Wikidata collects data, in a structured form.
- **Collaborative.** The data in Wikidata is entered and maintained by Wikidata editors, who decide on the rules of content creation and management in Wikidata.
- **Integration with Wikipedia:** serving language labels and data.
- **Multilingual.** Editing, consuming, browsing, and reusing the data is fully multilingual. Data entered in any language is immediately available in all other languages.
- **A secondary database.** Wikidata can record not just statements, but also their sources, thus reflecting the diversity of knowledge available and supporting the notion of verifiability.
- **Free.** The data in Wikidata is published under the [Creative Commons](https://creativecommons.org).
Wikidata in RDF

- “Native” Wikidata model data model  (not RDF)
  - 16 million entities
  - 34 million statements
  - 80 million labels
  - 350 languages

- RDF exports available
  - Non-trivial mapping to RDF
  - >400 million triples

- Examples:
  - [http://wikidata.metaphacts.com/sparql](http://wikidata.metaphacts.com/sparql)
Entities in Barcelona
= Entities in Barcelona =

{{#widget: Map | markers = $}}

SELECT ?lat ?lng ?link ?description
WHERE {
  ?entity wd:P131c wd:Q1492;
  wd:P625c ?coordinates;
  wd:P373c ?description .
  BIND(?entity as ?link) .
}

$}}
People born in Barcelona on a Timeline
= People born in Barcelona on a Timeline =

```sparql
{{#widget:Timeline
query ='

FILTER(lang(?label)='en')
}

start='birth'
end='death'
label='label'
link='uri'
interval='DECADE'
}}
```
Semantic Search

Semantic Search

A Barcelona [poem]
Barcelona [municipality of Spain]
Barcelona [film]
Barcelona [city]
Barcelona [village]
Barcelona [song]
Barcelona [administrative territorial entity of Ukraine]
Barcelona [band]
Barcelona [city with millions of inhabitants]
Barcelona [single]
Graph algorithms

Taxonomic Ranks (Biological classification)
Graph algorithms

Taxonomic Ranks (Biological classification)

```xml
<row>
  <col width="12" height="1000px">
    {{#widget: com.metaphacts.widgets.GraphWidget
       query = 'PREFIX gas: <http://www.bigdata.com/rdf/gas#>
       CONSTRUCT { ?pred wd:P361c ?out } WHERE {
       SERVICE gas:service {
             gas:program gas:gasClass "com.bigdata.rdf.graph.analitics.BFS" .
             gas:program gas:in wd:Q146481 . # one or more times, specifies the initial frontier.
             gas:program gas:out ?out . # will be bound to the visited vertices.
             gas:program gas:out1 ?depth . # will be bound to the depth of the visited vertices.
             gas:program gas:out2 ?pred . # will be bound to the predecessor.
             gas:program gas:traversalDirection 'Reverse' .
             gas:program gas:linkType wd:P361c
       }
       | layout = {{
             name = 'breadthfirst' | roots = {{ 'http://www.wikidata.org/entity/Q146481' }}
       }}
     }
   </col>
</row>
```
## Statistics

### Most popular classes

<table>
<thead>
<tr>
<th>class</th>
<th>count</th>
</tr>
</thead>
<tbody>
<tr>
<td>Item</td>
<td>16842535</td>
</tr>
<tr>
<td>human</td>
<td>2693145</td>
</tr>
<tr>
<td>Wikimedia category page</td>
<td>2422875</td>
</tr>
<tr>
<td>taxon</td>
<td>1896799</td>
</tr>
<tr>
<td>GlobeCoordinatesValue</td>
<td>1847718</td>
</tr>
<tr>
<td>Wikimedia disambiguation page</td>
<td>805146</td>
</tr>
<tr>
<td>village-level division in China</td>
<td>588520</td>
</tr>
<tr>
<td>owl:Class</td>
<td>191385</td>
</tr>
<tr>
<td>Wikimedia list article</td>
<td>186733</td>
</tr>
<tr>
<td>Wikimedia template</td>
<td>185499</td>
</tr>
</tbody>
</table>

### Most popular properties

<table>
<thead>
<tr>
<th>property</th>
<th>count</th>
</tr>
</thead>
<tbody>
<tr>
<td>rdf:type</td>
<td>31348748</td>
</tr>
<tr>
<td>instance of</td>
<td>12446770</td>
</tr>
<tr>
<td>label</td>
<td>8534489</td>
</tr>
<tr>
<td>country</td>
<td>2658435</td>
</tr>
<tr>
<td>sex or gender</td>
<td>2492158</td>
</tr>
<tr>
<td>description</td>
<td>2305311</td>
</tr>
<tr>
<td>located in the administrative territorial entity</td>
<td>2036112</td>
</tr>
<tr>
<td>taxon rank</td>
<td>1904288</td>
</tr>
<tr>
<td>coordinate location</td>
<td>1891469</td>
</tr>
<tr>
<td>taxon name</td>
<td>1881588</td>
</tr>
</tbody>
</table>
Qualified Statements and References

Douglas Adams (Q42)

English writer and humorist

Also known as: Douglas Noël Adams, Douglas Noel Adams, DNA, Bop Ad

Date of birth
11 March 1952

1 reference

Wikipedia pages linked to this item (64 entries)

<table>
<thead>
<tr>
<th>Language</th>
<th>Code</th>
<th>Linked page</th>
</tr>
</thead>
<tbody>
<tr>
<td>العربية</td>
<td>arwiki</td>
<td>دوغلاس آدمز</td>
</tr>
<tr>
<td>مصرى</td>
<td>arzwiki</td>
<td>دوجلاس ادامز</td>
</tr>
<tr>
<td>Boarisch</td>
<td>banwiki</td>
<td>Douglas Adams</td>
</tr>
<tr>
<td>беларуская</td>
<td>be x oldwiki</td>
<td>Дуглас Адамз</td>
</tr>
</tbody>
</table>
# Qualified Statements and References

<table>
<thead>
<tr>
<th>spouse</th>
<th>Jane Belson</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>start date</td>
</tr>
<tr>
<td></td>
<td>end date</td>
</tr>
</tbody>
</table>

- **1 reference**

<table>
<thead>
<tr>
<th>reference URL</th>
<th><a href="http://www.nndb.com/people/731/000023662/">http://www.nndb.com/people/731/000023662/</a></th>
</tr>
</thead>
<tbody>
<tr>
<td>original language</td>
<td>English</td>
</tr>
<tr>
<td>title</td>
<td>Douglas Adams</td>
</tr>
</tbody>
</table>
Representing and querying qualified statements and references

```
SELECT ?property ?object ?reference WHERE {
  entity:Q42 ?s_property ?sid .
  <http://www.w3.org/ns/prov#wasDerivedFrom> ?reference .
  FILTER ( REPLACE(str(?s_property), "\.$", "") =
    REPLACE(str(?v_property), "\.$", ""))
}
```

RDR – Reification Done Right

- Simple extension to RDF and SPARQL
- Allows to use a statement as the subject of another statement
- Formalized by Hartig and Thompson in *Foundations of an Alternative Approach to Reification in RDF*

```sparql
@ prefix : <http://www.wikidata.org/entity/> .
@ prefix prov: <http://www.w3.org/ns/prov#wasDerivedFrom> .
<<:Q42 :p26 :Q14623681>> prov:wasDerivedFrom :R801b4ec5 .

SELECT ?property ?object ?reference WHERE {
}
```
Conclusions

- Wikidata as useful knowledge graph with real life use cases
- Wikidata query services currently being developed based on RDF / SPARQL
- Expressive Wikidata data model that poses challenges for representation in RDF
- Reification Done Right as possible approach for dealing with qualified statements and references
- Standardization desired
- Interesting from benchmarking perspective
Contact us!

metaphacts GmbH
Kautzelweg 13
69190 Walldorf
Germany
p +49 6227 8308660
m +49 157 50152441
e info@metaphacts.com
@metaphacts