



DBpedia

A crystallization point for the Web of Data

Christian Bizer, Jens Lehmann, Georgi Kobilarov, Soren Auer, Christian Becker,
Richard Cyganiak, Sebastian Hellmann

Presented by Jeremy Chen

Motivation



- Most existing knowledge bases cover only specific domains
- Costly to keep existing knowledge bases up-to-date as domains change

At the same time.....

- Wikipedia is a big central knowledge source
- **DBpedia leverages this gigantic source of knowledge by extracting structured information**

Advantages



- Covers many domains
- Evolves automatically as Wikipedia changes
- Truly multilingual
- Accessible on the Web

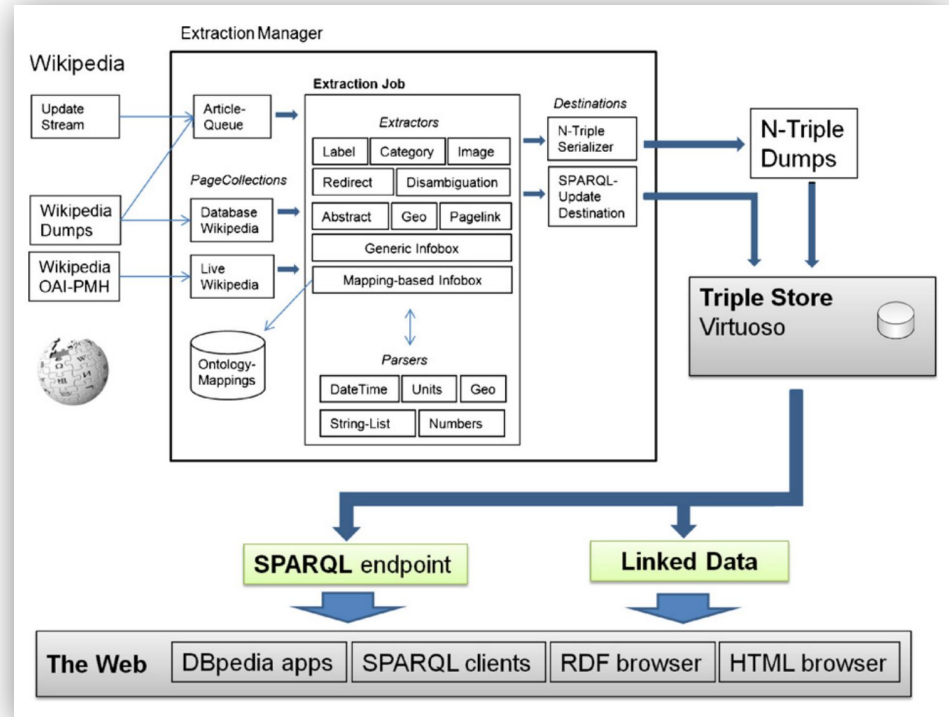
Outline



- DBpedia knowledge extraction framework
- DBpedia knowledge base
- Different access mechanisms
- Interlinked Web content with DBpedia
- Applications

DBpedia Knowledge Extraction Framework

- Workflows
 - Dump-based extraction
 - Monthly SQL dumps
 - Live extraction



DBpedia Knowledge Extraction Framework

- Infobox Extraction
 - Generic infobox extraction
 - *Subject*: DBpedia URI
 - *Predicate*:
<http://dbpedia.org/property>
/ + attribute name
 - *Object*: attribute value
 - Mapping-based infobox extraction



```
{{ Infobox Actor
| birthname = Thomas Jeffrey Hanks
| birthdate = {{birth date and age|1956|7|9}}
| birthplace = [[Concord, California|Concord]],
               [[California]]
| yearsactive = 1979 - present
| occupation = Actor, producer, director,
               [[voice over artist]],
               writer, speaker
```

Fig. 2. Infobox Tom Hanks.

DBpedia Knowledge Base



- Entity Identifier: English article name
- Entity Properties
 - General Properties
 - Label
 - English abstract
 - Links to external Web pages
 - etc.
 - Infobox-specific properties

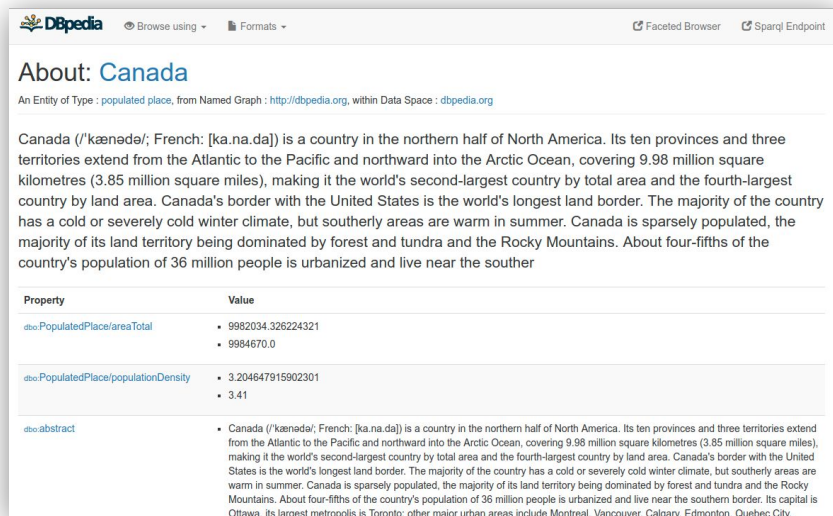
Accessing DBpedia



- 4 access mechanisms
 - Linked Data
 - SPARQL endpoint
 - RDF dumps
 - Lookup index

Accessing DBpedia - Linked Data

- *Semantic Web agents*: RDF descriptions
- *Traditional Web browsers*: simple HTML view



The screenshot shows the DBpedia web interface for the entity 'Canada'. The page title is 'About: Canada'. Below the title, there is a description of Canada in English. A table lists properties and their values, including 'dbpedia:PopulatedPlace/areaTotal' and 'dbpedia:PopulatedPlace/populationDensity'. The table has two columns: 'Property' and 'Value'. The 'Value' column contains lists of values for each property.

Property	Value
dbpedia:PopulatedPlace/areaTotal	<ul style="list-style-type: none">9962034.3262243219984670.0
dbpedia:PopulatedPlace/populationDensity	<ul style="list-style-type: none">3.2046479159023013.41

[dbpedia:abstract](#)

- Canada (/ˈkænədə/; French: [ka.na.da]) is a country in the northern half of North America. Its ten provinces and three territories extend from the Atlantic to the Pacific and northward into the Arctic Ocean, covering 9.98 million square kilometres (3.85 million square miles), making it the world's second-largest country by total area and the fourth-largest country by land area. Canada's border with the United States is the world's longest land border. The majority of the country has a cold or severely cold winter climate, but southerly areas are warm in summer. Canada is sparsely populated, the majority of its land territory being dominated by forest and tundra and the Rocky Mountains. About four-fifths of the country's population of 36 million people is urbanized and live near the southern border. Its capital is Ottawa, its largest metropolis is Toronto; other major urban areas include Montreal, Vancouver, Calgary, Edmonton, Quebec City.

Accessing DBpedia - SPARQL Endpoint

- <http://dbpedia.org/sparql>

Virtuoso SPARQL Query Editor

Default Data Set Name (Graph IRI)

Query Text
select distinct ?Concept where {[] a ?Concept} LIMIT 100

(Security restrictions of this server do not allow you to retrieve remote RDF data, see [details](#).)

Results Format:

Execution timeout: milliseconds (values less than 1000 are ignored)

Options:

- Strict checking of void variables
- Log debug info at the end of output (has no effect on some queries and output formats)
- Generate SPARQL compilation report (instead of executing the query)

(The result can only be sent back to browser, not saved on the server, see [details](#))



Concept
http://www.w3.org/2002/07/owl#FunctionalProperty
http://www.w3.org/1999/02/22-rdf-syntax-ns#Property
http://www.w3.org/2002/07/owl#Thing
http://www.w3.org/2002/07/owl#Class
http://www.w3.org/2002/07/owl#Ontology
http://www.w3.org/2002/07/owl#ObjectProperty
http://www.w3.org/2002/07/owl#DatatypeProperty
http://xmlns.com/foaf/0.1/Organization
http://dbpedia.org/ontology/Company
http://xmlns.com/foaf/0.1/Person
http://dbpedia.org/ontology/Activity
http://dbpedia.org/ontology/Name
http://dbpedia.org/ontology/Person
http://dbpedia.org/ontology/Actor
http://dbpedia.org/ontology/Place
http://dbpedia.org/ontology/Publisher
http://dbpedia.org/ontology/Genre
http://dbpedia.org/ontology/Language
http://dbpedia.org/ontology/Location
http://dbpedia.org/ontology/Software
http://dbpedia.org/ontology/School
http://xmlns.com/foaf/0.1/Document
http://purl.org/ontology/bibo/Book
http://purl.org/vocommons/voaf#Vocabulary

Accessing DBpedia - RDF Dumps



- Sliced DBpedia knowledge base by triple predicate
 - e.g. <https://wiki.dbpedia.org/dbpedia-version-2016-04>

DBpedia version 2016-04

Dataset category: [DBpedia release](#)

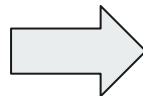
Publication Year: 2016

This release is based on updated Wikipedia dumps dating from March/April 2016 featuring a significantly expanded base of information as well as richer and (hopefully) cleaner data based on the DBpedia ontology.

Accessing DBpedia - Lookup Index

- Label -> DBpedia URI

<http://lookup.dbpedia.org/api/search/KeywordSearch?QueryClass=place&QueryString=berlin>



```
<ArrayOfResult xmlns="http://lookup.dbpedia.org/" xmlns:xsd="http://www.w3.org/2002/07/owl#Thing">
  <Result>
    <Label>Berlin</Label>
    <URI>http://dbpedia.org/resource/Berlin</URI>
    <Description>
      Berlin is the capital city of Germany and one of the 16 states of Germany. It is located in the northeastern part of the country, in the area in the European Union. Located in northeastern Germany, it is influenced by a temperate seasonal climate.
    </Description>
    <Classes>
      <Class>
        <Label>city</Label>
        <URI>http://dbpedia.org/ontology/City</URI>
      </Class>
      <Class>
        <Label>populated place</Label>
        <URI>http://dbpedia.org/ontology/PopulatedPlace</URI>
      </Class>
      <Class>
        <Label>settlement</Label>
        <URI>http://dbpedia.org/ontology/Settlement</URI>
      </Class>
      <Class>
        <Label>place</Label>
        <URI>http://dbpedia.org/ontology/Place</URI>
      </Class>
      <Class>
        <Label>owl#Thing</Label>
        <URI>http://www.w3.org/2002/07/owl#Thing</URI>
      </Class>
      <Class>
        <Label>place</Label>
        <URI>http://schema.org/Place</URI>
      </Class>
      <Class>
        <Label>city</Label>
        <URI>http://schema.org/City</URI>
      </Class>
    </Classes>
  </Result>
</ArrayOfResult>
```

Interlinked Web Content



- More data publishers set RDF links to DBpedia entities
- Together with outgoing edges published by DBpedia, DBpedia becomes a central interlinking hub of Web of Data
 - Web of Data Browsing and Crawling
 - Linked Data browser, Semantic Web search engines
 - Web Data Fusion and Mashups
 - Fusing data from different sources to generate integrated views
 - Web Content Annotation

Interlinked Web Content - Example

- Complementary data about Spain
 - Agents can follow these links to retrieve additional information

```
<http://dbpedia.org/resource/Spain> owl:sameAs  
http://www4.wiwiss.fu-berlin.de/eurostat/resource/countries/Espa%C3%B1a ;  
http://rdf.freebase.com/ns/guid.9202a8c04000641f8000000000034e30 ;  
http://www4.wiwiss.fu-berlin.de/factbook/resource/Spain ;  
http://sw.opencyc.org/2008/06/10/concept/Mx4rvVjowpwpEbGdrcN5Y29ycA .
```

- Annotate a research paper with “Data Integration”

```
<http://data.semanticweb.org/conference/eswc/2008/paper/356>  
swc:hasTopic <http://dbpedia.org/resource/Data_integration> .
```

Applications



- Browsing and exploration
 - DBpedia Mobile
- Querying and search
 - DBpedia Query Builder
 - Relationship Finder
- Content annotation

Applications - DBpedia Mobile

- Location-aware client for the Semantic Web



Fig. 7. DBpedia Mobile running on an iPhone 3G and showing a map view of resources in the user's proximity.

Applications - DBpedia Query Builder

- Users can hardly know the properties/identifiers in the knowledge base

Soccer player with tricot nr. 11, playing for a club having a stadium with >40.000 seats, born in a country with >10M inhabitants

Subject	Predicate	Object
<input type="text" value="?player"/>	<input type="text" value="currentclub"/>	<input type="text" value="?club"/>
<input type="text" value="?player"/>	<input type="text" value="clubnumber"/>	<input type="text" value="11"/>
<input type="text" value="?player"/>	<input type="text" value="countryofbirth"/>	<input type="text" value="?country"/>
<input type="text" value="?club"/>	<input type="text" value="capacity"/>	<input ">40000"="" type="text" value=""/>
<input type="text" value="?country"/>	<input type="text" value=""/>	<input ">10000000"="" type="text" value=""/>
<input type="button" value="+"/>	<input type="text" value="GDP_PPP (11)"/>	
	<input type="text" value="population_estimate (10)"/>	
	<input type="text" value="population_census (9)"/>	
	<input type="text" value="established_date2 (8)"/>	
	<input type="text" value="established_date1 (6)"/>	
	<input type="text" value="established_date3 (5)"/>	
	<input type="text" value="GDP_nominal (3)"/>	
	<input type="text" value="accessionEUdate (2)"/>	

Click on a column head this page Results: 10

10 results found in 0.00

Nr.	?player	?country	>40000	>10000000
1	Cicinho	Brazil	80354	187560000
2	Gonzalo Fierro	Chile	62000	16432674
3	Lukas Podolski	Poland	69901	38536869
4	Mark González	South Africa	45362	47432000
5	Michael Thurk	Germany	52000	82438000
6	Ramón Morales	Mexico	72480	107784179
7	Robin van Persie	Netherlands	60432	16336346
8	Stefano Mauri	Italy	82656	58751711

Fig. 8. Form-based DBpedia query builder.

Applications - Relationship Finder

- Find connections between two different entities in DBpedia

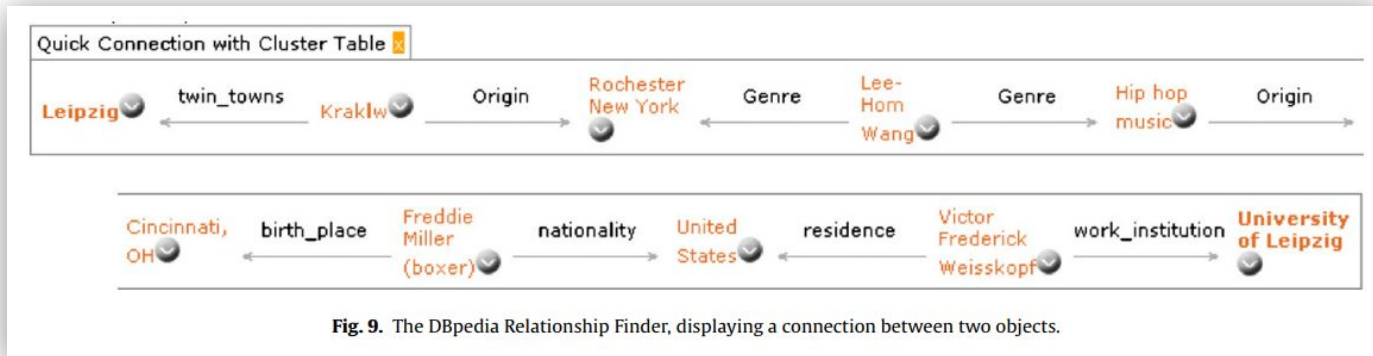


Fig. 9. The DBpedia Relationship Finder, displaying a connection between two objects.



Thank you!