Wikidata with SPARQL

David Arroyo Menéndez

October 1, 2019

Wikidata Definition

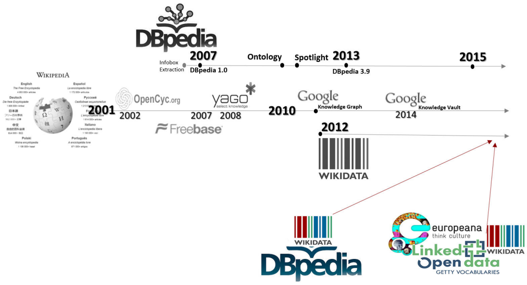
Wikidata is a collaboratively edited knowledge base hosted by the Wikimedia Foundation. It is a common source of open data that Wikimedia projects such as Wikipedia can use, and anyone else, under a public domain license. This is similar to the way Wikimedia Commons provides storage for media files and access to those files for all Wikimedia projects, and which are also freely available for reuse. Wikidata is powered by the software Wikibase.
See: `https://www.wikidata.org`

SPARQL definition

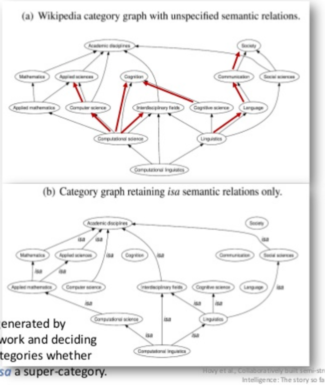
SPARQL is a recursive acronym for SPARQL Protocol and RDF Query Language) is an RDF query language—that is, a semantic query language for databases—able to retrieve and manipulate data stored in Resource Description Framework.

Linked Open Data. History

Knowledge Bases: Timeline



Linked Open Data. History



(a) Wikipedia category graph with unspecified semantic relations.

(b) Category graph retaining *isa* semantic relations only.

WikiTaxonomy is generated by traversing the network and deciding for each pair of categories whether the sub-category *isa* a super-category.

Hovy et al., Collaboratively built semi-structured content and Artificial Intelligence: The story so far, *Artificial Intelligence* (2012)

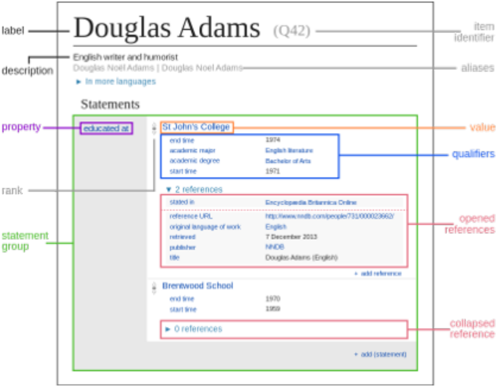
SPARQL in Python, dbpedia as example

```
from SPARQLWrapper import SPARQLWrapper, JSON

sparql = SPARQLWrapper("http://dbpedia.org/sparql")
sparql.setQuery("""
    PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
    SELECT ?label
    WHERE { <http://dbpedia.org/resource/Asturias> rdfs:label ?
""")
sparql.setReturnFormat(JSON)
results = sparql.query().convert()

for result in results["results"]["bindings"]:
    print(result["label"]["value"])
```

Wikidata: semantic model in a wikipedia page (I)



Douglas Adams (Q42)

English writer and humorist

Douglas Noel Adams | Douglas Noël Adams

▸ In more languages

Statements

educated at

St John's College

end time	1974
academic major	English literature
academic degree	Bachelor of Arts
start time	1971

▾ 2 references

stated in	Encyclopaedia Britannica Online
reference URL	http://www.nndb.com/people/731/000023662/
original language of work	English
retrieved	7 December 2013
publisher	NNDB
title	Douglas Adams (English)

+ add reference

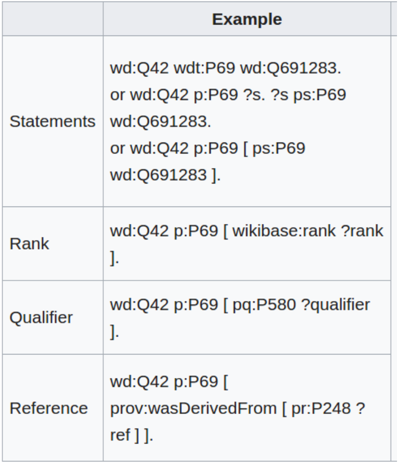
Brentwood School

end time	1970
start time	1959

▸ 0 references

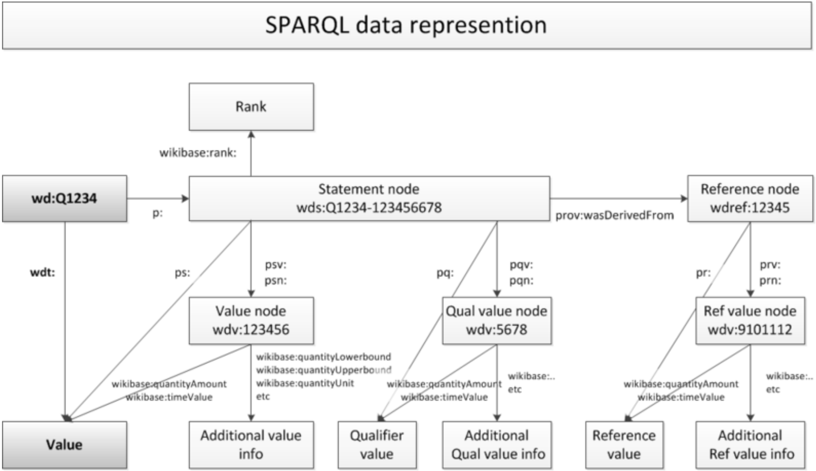
+ add [statement]

Wikidata: semantic model in a wikipedia page (II)



	Example
Statements	wd:Q42 wdt:P69 wd:Q691283. or wd:Q42 p:P69 ?s. ?s ps:P69 wd:Q691283. or wd:Q42 p:P69 [ps:P69 wd:Q691283].
Rank	wd:Q42 p:P69 [wikibase:rank ?rank].
Qualifier	wd:Q42 p:P69 [pq:P580 ?qualifier].
Reference	wd:Q42 p:P69 [prov:wasDerivedFrom [pr:P248 ?ref]].

Wikidata: semantic model in a wikipedia page (III)



Wikidata as Language

If you take a look at Germany (Q183), then you can see a whole host of properties like population (P1082), median income (P3529) or even images with the image (P18) property.

```
SELECT
  ?country ?countryLabel ?population ?area ?medianIncome
WHERE {
  ?country wdt:P463 wd:Q458.
  ?country wdt:P1082 ?population.
  ?country wdt:P2046 ?area.
  ?country wdt:P3529 ?medianIncome.

  SERVICE wikibase:label { bd:serviceParam wikibase:language "
}
```

Wikidata in Python (I)

```
import requests
url = ’https://query.wikidata.org/sparql’
query = """
SELECT
  ?country ?countryLabel ?population ?area ?medianIncome
WHERE {
  ?country wdt:P463 wd:Q458.
  ?country wdt:P1082 ?population.
  ?country wdt:P2046 ?area.
  ?country wdt:P3529 ?medianIncome.
  SERVICE wikibase:label { bd:serviceParam wikibase:language "e
}
"""
r = requests.get(url, params = {’format’: ’json’, ’query’: quer
data = r.json()
print(data)
```

Wikidata in Python (II)

Print ten females in json.

```
import requests

url = "https://query.wikidata.org/sparql"

query = """
SELECT ?name ?nombre ?sexo_o_g_nero ?sexo_o_g_neroLabel WHERE {
  ?human wdt:P31 wd:Q5.
  OPTIONAL { ?human wdt:P21 ?nombre. }
  OPTIONAL { ?human wdt:P21 ?sexo_o_g_nero. }
}
LIMIT 10"""

r = requests.get(url, params = {’format’: ’json’, ’query’: quer
data = r.json()
print(data)
```

Wikidata in Python (III)

Print ten cats in json:

```
import requests
url = "https://query.wikidata.org/sparql"
query = """#added before 2016-10
#Cats
SELECT ?item ?itemLabel
WHERE
{
  ?item wdt:P31 wd:Q146.
  SERVICE wikibase:label { bd:serviceParam wikibase:language "[
}
LIMIT 10
"""
r = requests.get(url, params = {'format': 'json', 'query': quer
data = r.json()
print(data['results']['bindings'])
```