

Web Semântica

Dados Semânticos:
Fontes de Dados

Wikidata



- A Wikidata é um projeto da Fundação Wikimedia, que tem por objetivo reunir dados para suporte da Wikipedia, Wikimedia e outros.
 - <https://www.wikidata.org>
 - Livre – os dados são publicados sob a licença *Creative Commons* (CCO 1.0) e podem ser usados livremente, mesmo para efeitos comerciais.
 - Colaborativo – os dados são inseridos e mantidos por editores da wikidata.

Wikidata (ii)



- Multilingue – a navegação, edição e consumo dos dados é automaticamente multilingue.
- Segundada por Bases de Dados – para além dos dados, encontra-se também informação sobre as suas fontes o que aumenta a possibilidade da sua verificação.
- Dados Estruturados – ao contrário da Wikimedia e da Wikipedia, a Wikidata reúne dados estruturados de uma certa forma.

Wikidata (iii)



• Estrutura dos Dados

- https://www.mediawiki.org/wiki/Wikidata_Query_Service/User_Manual
- O conjunto de dados consiste num grafo de triplos RDF que é pesquisável através de SPARQL
- Utiliza um conjunto de códigos para se referir diretamente a entidades, propriedades e objetos
- Exemplo de um triplo:

```
<http://www.wikidata.org/entity/Q30>  
<http://www.wikidata.org/prop/direct/P36>  
<http://www.wikidata.org/entity/Q61> .
```

Wikidata (iv)



- Este triplo é ainda comum ser representado com os seguintes prefixos:

```
PREFIX wd: <http://www.wikidata.org/entity/>
```

```
PREFIX wdt: <http://www.wikidata.org/prop/direct/>
```

```
wd:Q30 wdt:P36 wd:Q61 .
```

- O seu significado:

- Subject Predicate Object
- (United States) (capital is) (Washington DC)

Wikidata – Testar Queries



- Para testar *queries* é possível aceder a um editor online:

- <https://query.wikidata.org/>

- Exemplo:

```
#capital
SELECT ?item ?itemLabel
WHERE
{
    wd:Q30    wdt:P36    ?item .
    SERVICE wikibase:label { bd:serviceParam
wikibase:language "en". }
}
```

- Resultado

```
wd:Q61      Washington, D.C
```

Wikidata – Testar Queries



- Teste de *queries* :
 - Descrição total do resultado obtido anteriormente:

```
#capital  
PREFIX wd: <http://www.wikidata.org/entity/>  
DESCRIBE wd:Q61
```

- Os resultados obtidos são muitos. Será preferível pesquisar por predicados específicos como: o tipo da entidade (rdf:type) e a sua label (rdfs:label)

Wikidata – Acesso



- Para aceder à Wikidata, de forma programática pode-se utilizar a biblioteca SPARQLwrapper.
 - <https://github.com/RDFLib/sparqlwrapper>
- Esta, não é mais que um simples Wrapper em torno de um serviço SPARQL que permite a execução de *queries*, sem a preocupação com os detalhes técnicos de acesso.
- Instalar o package SPARQLwrapper
 - `pip install sparqlwrapper`

Wikidata – Acesso (ii)



- Uso do SPARQLwrapper:

```
sparql = SPARQLWrapper2("https://query.wikidata.org/sparql")
sparql.setQuery("""
PREFIX wd: <http://www.wikidata.org/entity/>
PREFIX wdt: <http://www.wikidata.org/prop/direct/>
SELECT ?item ?itemLabel
WHERE
{
    wd:Q30    wdt:P36    ?item .
    SERVICE wikibase:label { bd:serviceParam
wikibase:language "en". }
}
""")
for result in sparql.query().bindings:
    print('%s: %s' % (result["item"].value,
result["itemLabel"].value))
```

Wikidata – Acesso (iii)



- Resulto obtido do SPARQLwrapper:

`http://www.wikidata.org/entity/Q61: Washington, D.C.`

DBpedia



- A DBpedia é um esforço comunitário baseado em *crowdsourcing* para a extração de informação criada em projetos da Wikimedia.
- A informação é estruturada na forma de um grafo aberto de conhecimento (OKG – *Open Knowledge Graph*).
- O que permite o uso da elevada quantidade de informação da Wikimedia de modos mais estruturados e interessantes.
- Link: <https://wiki.dbpedia.org/>

A Ontologia da DBpedia



- O grafo de conhecimento da DBpedia é classificado na forma de uma ontologia.
- Esta ontologia atravessa múltiplos domínios de conhecimento e foi criada manualmente a partir da informação usada nas *infoboxes* da Wikipedia.
- Atualmente, esta ontologia possui centenas de classes e milhares de propriedades.
- Link:
 - https://mappings.dbpedia.org/index.php/Main_Page

Acesso à DBpedia



- O acesso aos dados contidos na DBpedia pode ser feito através da sua página web de pesquisa como através do seu *endpoint*, usando SPARQL.
- A página web pode ser usada tanto para extração de informação como para a validação de *queries*.
- Enquanto o *endpoint* é mais adequado para a consulta programática da informação.
- Em qualquer dos casos, e numa fase inicial, é necessária a utilização dos termos da ontologia para descobrir os diversos URIs associados aos recursos.
- Link: <https://dbpedia.org/sparql>

Acesso à DBpedia



- Exemplo de *query* na página web:

```
SELECT ?s ?fn
WHERE {
    ?s a dbo:Film .
    ?s rdfs:label ?fn .
    FILTER (lang(?fn) = 'en')
}
LIMIT 100
```

Acesso à DBpedia



- Exemplo de *query* no *endpoint* SPARQL:

```
from SPARQLWrapper import SPARQLWrapper2

sparql = SPARQLWrapper2("https://dbpedia.org/sparql")
query = """
    select ?furi ?fname
    where {
        ?furi a dbo:Film .
        ?furi rdfs:label ?fname .
        FILTER (lang(?fname) = 'en')
    }
    limit 1000
    """

sparql.setQuery(query)
results = sparql.query()
for r in results.bindings:
    print('%s: %s' % (r["furi"].value, r["fname"].value))
```

Acesso à DBpedia



- Resultado obtido do *endpoint*:

http://dbpedia.org/resource/4000_Miles: 4000 Miles
http://dbpedia.org/resource/The_Man_Who_Came_Back: The Man Who Came Back
http://dbpedia.org/resource/12_Monkeys: 12 Monkeys
<http://dbpedia.org/resource/Airplane!>: Airplane!
[http://dbpedia.org/resource/American_Beauty_\(1999_film\)](http://dbpedia.org/resource/American_Beauty_(1999_film)): American Beauty (1999 film)
<http://dbpedia.org/resource/Amélie>: Amélie
http://dbpedia.org/resource/Andrzej_Wajda: Andrzej Wajda
[http://dbpedia.org/resource/Animal_Crackers_\(1930_film\)](http://dbpedia.org/resource/Animal_Crackers_(1930_film)): Animal Crackers (1930 film)
http://dbpedia.org/resource/Animated_cartoon: Animated cartoon
http://dbpedia.org/resource/Anne_of_the_Thousand_Days: Anne of the Thousand Days
[http://dbpedia.org/resource/Anywhere_but_Here_\(film\)](http://dbpedia.org/resource/Anywhere_but_Here_(film)): Anywhere but Here (film)
[http://dbpedia.org/resource/Armageddon_\(1998_film\)](http://dbpedia.org/resource/Armageddon_(1998_film)): Armageddon (1998 film)
http://dbpedia.org/resource/Arms_and_the_Man: Arms and the Man
http://dbpedia.org/resource/Balducci_levitation: Balducci levitation
[http://dbpedia.org/resource/Beautiful_Girls_\(film\)](http://dbpedia.org/resource/Beautiful_Girls_(film)): Beautiful Girls (film)
<http://dbpedia.org/resource/Berry>: Berry
http://dbpedia.org/resource/Black_Narcissus: Black Narcissus
http://dbpedia.org/resource/Black_tie: Black tie
[http://dbpedia.org/resource/Breakfast_at_Tiffany's_\(film\)](http://dbpedia.org/resource/Breakfast_at_Tiffany's_(film)): Breakfast at Tiffany's (film)
[http://dbpedia.org/resource/Carry_On_\(franchise\)](http://dbpedia.org/resource/Carry_On_(franchise)): Carry On (franchise)
[http://dbpedia.org/resource/Casablanca_\(film\)](http://dbpedia.org/resource/Casablanca_(film)): Casablanca (film)