

SIW PL9. Extracción de información semántica.

Hugo Fonseca Díaz
uo258318@uniovi.es

Escuela de Ingeniería Informática. Universidad de Oviedo.

19 de noviembre de 2020

1. Información estructurada

En esta primera sección se obtendrá información estructurada de tres textos mediante el uso de tipos definidos en `schema.org` [1]. Dicha información se define a continuación:

- **Miles Davis:** es una entidad de tipo **Person**. Sus propiedades son:
 - `familyName`: Davis
 - `givenName`: Miles
 - `hasOccupation`: `Occupation(name: jazz musician)`
 - `nationality`: `Country(name: United States of America)`
 - `url`: `https://www.wikidata.org/wiki/Q93341`
- **Barack Obama:** es una entidad de tipo **Person**. Sus propiedades son:
 - `familyName`: Obama
 - `givenName`: Barack
 - `hasOccupation`: `Occupation(name: President)`
 - `url`: `https://www.wikidata.org/wiki/Q76`
- **European Union:** es una entidad de tipo **Organization**. Sus propiedades son:
 - `name`: European Union
 - `legalName`: European Union
 - `url`: `https://www.wikidata.org/wiki/Q458`
- **Washington:** es una entidad de tipo **City**. Sus propiedades son:
 - `name`: Washington DC

- url: <https://www.wikidata.org/wiki/Q61>
- **Euro a Dólar:** es una entidad de tipo `ExchangeRateSpecification`. Sus propiedades son:
 - currency: EUR
 - currentExchangeRate: `UnitPriceSpecification(priceCurrency: USD, price: 1.3)`
 - url: https://www.ecb.europa.eu/stats/policy_and_exchange_rates/euro_reference_exchange_rates/html/index.en.html
- **The New York Times:** es una entidad de tipo `Newspaper`. Sus propiedades son:
 - name: The New York Times
 - url: <https://www.wikidata.org/wiki/Q9684>
- **John McCarthy:** es una entidad de tipo `Person`. Sus propiedades son:
 - familyName: McCarthy
 - givenName: John
 - hasOccupation: `Occupation(name: computer scientist)`
 - url: <https://www.wikidata.org/wiki/Q92739>
- **LISP:** es una entidad de tipo `ComputerLanguage`. Sus propiedades son:
 - name: LISP
 - url: <https://www.wikidata.org/wiki/Q132874>

2. Modelado RDF

En esta sección se muestra el modelado *RDF* en formato *Turtle* con el que se ha representado la información listada previamente. Dicho modelado es el siguiente:

```
@prefix schema: <https://schema.org/> .
@prefix wikidata: <https://wikidata.org/> .
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .

# Primer texto
wikidata:Q93341 rdf:type schema:Person ;
  schema:familyName "Davis" ;
  schema:givenName "Miles" ;
  schema:hasOccupation [
    rdf:type schema:Occupation ;
    schema:name "jazz musician"
  ] ;
  schema:nationality wikidata:Q30 .
```

```

# Segundo texto
wikidata:Q76 rdf:type schema:Person ;
  schema:familyName "Obama" ;
  schema:givenName "Barack" ;
  schema:hasOccupation [
    rdf:type schema:Occupation ;
    schema:name "president"
  ] .

wikidata:Q458 rdf:type schema:Organization ;
  schema:name "European Union" ;
  schema:legalName "European Union" .

wikidata:Q61 rdf:type schema:City ;
  schema:name "Washington DC" .

<https://www.ecb.europa.eu/stats/policy_and_exchange_rates/euro_reference_exchange_rates/html/index.en.htm
  rdf:type schema:ExchangeRateSpecification ;
  schema:currency "EUR" ;
  schema:currentExchangeRate [
    rdf:type schema:UnitPriceSpecification ;
    schema:priceCurrency "USD" ;
    schema:price 1.3
  ] .

# Tercer texto
wikidata:Q9684 rdf:type schema:Newspaper ;
  schema:name "The New York Times" .

wikidata:Q92739 rdf:type schema:Person ;
  schema:familyName "McCarthy" ;
  schema:givenName "John" ;
  schema:hasOccupation [
    rdf:type schema:Occupation ;
    schema:name "computer scientist"
  ] .

wikidata:Q132874 rdf:type schema:ComputerLanguage ;
  schema:name "LISP" .

```

En la **Figura 1** se observa el output de la herramienta *RDFShape* al introducir el fichero en formato *Turtle*. Se observa que el fichero está bien formado.

Referencias

- [1] Página de *Schema.org*, <https://schema.org/>. Última vez accedido 19 de noviembre de 2020.

Figura 1: Validación del fichero *Turtle* mediante la herramienta *RDFShape*.

The screenshot displays the RDFShape web application interface. The browser address bar shows the URL: `https://rdfshape.weso.es/dataInfo?activeTab=%23dataTextArea&data=%40prefix%20schema%3A%3Chttps%3A%3Awww.w3.org/1999/02/22-rdf-syntax-ns%3E%3E`. The application has a blue navigation bar with the following menu items: RDFShape, Data, Endpoint, ShEx, SHACL, ShapeMap, Wikidata, Examples, and Help.

RDF Data info

RDF input

by Input | **By URL** | By File

```
39 # Tercer texto
40 wikidata:Q9684 rdf:type schema:Newspaper ;
41   schema:name "The New York Times" .
42
43 wikidata:Q92739 rdf:type schema:Person ;
44   schema:familyName "McCarthy" ;
45   schema:givenName "John" ;
46   schema:hasOccupation [
47     rdf:type schema:Occupation ;
48     schema:name "computer scientist"
49   ] .
50
51 wikidata:Q132874 rdf:type schema:ComputerLanguage ;
52   schema:name "LISP"
```

Data format

turtle

Inference

Well formed RDF

```
3 @prefix rdf:<http://www.w3.org/1999/02/22-rdf-syntax-ns#>
4
5 # Primer texto
6 wikidata:Q93341 rdf:type schema:Person ;
7   schema:familyName "Davis" ;
8   schema:givenName "Miles" ;
9   schema:hasOccupation [
10     rdf:type schema:Occupation ;
11     schema:name "jazz musician"
12   ] ;
13   schema:nationality wikidata:Q30 .
14
15 # Segundo texto
16 wikidata:Q76 rdf:type schema:Person ;
```

- Number of statements: 34
- DataFormat: turtle

Detalles

```
{
  "msg": "Well formed RDF",
}
```

Status Bar: 1 2 3 | cpu: 05% | 422,8 GiB | no lan | 100% | 2,3 GiB | 5,0 GiB | 19.11. 20:08