

Práctica 9 - Obtención de Información Estructuradas y uso de Ontologías.

Hugo Pérez Fernández. U0250708@uniovi.es

Sistemas de Información para la Web. Grado de Ingeniería Informática. EII.
universidad de Oviedo. Campus de los Catalanes. Oviedo.

1 Introducción

En esta práctica se verán las diferencias entre la obtención manual y automática de la información estructurada para tres textos dados, además se tratarán temas como por qué hay tantas ontologías y como tratar el problema de alinearlas a la hora de su uso.

2 Obtención manual de información estructurada

Primero obtendremos toda la información estructurada de los textos indicados en el Anexo.4.1, usando para ello la web Schema.org:

Texto 1: En el caso de este texto se está mencionando a una persona (*schema:Person*), cuyo nombre es Miles Davis, además esta persona tiene como nacionalidad *schema:Nacionalidad* Americano (Estadounidense), y su profesión (*schema:Occupation*) es músico concretamente de Jazz.

Texto 2: En este caso la información se puede estructurar de la siguiente forma:

- Hay una persona (*schema:Person*) que se llama Barack Obama cuya ocupación (*schema:Occupation*) es presidente.
- Hay una organización (*schema:Organization*) que se llama Unión Europea.
- Hay una ciudad (*schema:City*) que se llama Washington.
- Hay un tipo de cambio de moneda (*schema:ExchangeRateSpecification*) que se llama Euro y tiene un valor de 1.3.
- Hay una reunión (*schema:BusinessEvent*) entre Barack Obama y la Unión Europea sobre el valor del tipo de cambio del Euro.

Texto 3: En este último se puede estructurar la información del siguiente modo:

- Hay una organización (*schema:Organization*) que se llama The New York Times.
- Hay una persona (*schema:Person*) que se llama Jhon McCarthy.
- Jhon McCarthy está muerto.

- El New York Times informa (*schema:Report*) que Jhon McCarthy esta muerto.
- LISP es un lenguaje de programación.
- John McCarthy inventó LISP.

Con los datos obtenidos anteriormente se, usando la notacion Turtle, la información estructurada queda de la siguiente manera:

Listing 1.1. Archivo Turtle con la información estructurada de los textos.

```

1 @prefix wikidata: <http://www.wikidata.org/> .
2 @prefix foaf: <http://xmlns.com/foaf/0.1/> .
3
4 # Texto 1:
5
6 # Persona de nombre Miles Davis y ocupación músico de
  Jazz
7 <#Miles_Davis> wikidata:Q21503252 foaf:Person ;
8                foaf:name "Miles Davis" ;
9                wikidata:Property:P27 wikidata:Q30 ;
10               wikidata:Property:P106 wikidata:
11               Q15981151 .
12
13 # Texto 2:
14
15 # Persona Barack Obama Presidente
16 <#Barack_Obama> wikidata:Q21503252 foaf:Person ;
17                foaf:name "Barack Obama" ;
18                wikidata:Property:P39 wikidata:Q11696
19                .
20
21 # Organización llamada Unión Europea
22 <#UE> wikidata:Q21503252 wikidata:Q43229 ;
23      wikidata:Q82799 "Unión Europea" .
24
25 # Ciudad llamada Washington
26 <#AVDC> wikidata:Q21503252 wikidata:Q515 ;
27        wikidata:Q82799 "Washington" .
28
29 # Tipo de cambio del euro a 1.3
30 <#Tipo_euro> wikidata:Q21503252 wikidata:Q4916 ;
31              wikidata:Q21503252 wikidata:Q4916 ;
32              wikidata:P1181 "1.3" .
33
34 # Encuentro entre Unión Europea y Barak Obama
35 <#Meeting> wikidata:Q21503252 wikidata:Q2761147 ;
36           wikidata:Property:P710 wikidata:Q458 ;
37           wikidata:Property:P710 wikidata:Q76 ;

```

```

36          wikidata:Property:P921 <#tipo_euro> .
37
38 # Texto 3:
39
40 # Organizacion New York Times
41 <#New_York_Times>    wikidata:Q21503252    wikidata:Q43229 ;
42                      wikidata:Q82799 "New York Times" .
43
44 # Persona John McCarthy muerta
45 <#John_McCarthy>    wikidata:Q21503252    foaf:Person ;
46                    foaf:name "John McCarthy" ;
47                      wikidata:Q21503252    wikidata:Q48422
48
49 # Lenguaje de programación desarrollado por John McCarthy
50 <#LISP> wikidata:Q21503252    wikidata:Q9143 ;
51        wikidata:Q82799 "LISP" ;
52        wikidata:P178 <#John_McCarthy> .
53
54 # The New York Times informa sobre la muerte de John
    McCarthy
55 <#News> wikidata:Q21503252    wikidata:Q732577 ;
56        wikidata:Property:P921 <#John_McCarthy> ;
57        wikidata:Property:P123 <#New_York_Times> .

```

Ahora verificaremos la estructura de la sistaxis RDF de del archivo anterior mediante la herramienta RDFSShape:

RDF Data info

The screenshot shows the RDFSShape web application interface. On the left, the 'RDF data' input pane shows the original Turtle file content. On the right, the 'RDF data' output pane shows the same content after validation. The bottom status bar indicates 'Well formed RDF' and 'Data format: turtle'.

Input RDF data (Left Pane):

```

1 @prefix wikidata: <http://www.wikidata.org/> .
2 @prefix foaf: <http://xmlns.com/foaf/0.1/> .
3
4 <#UE>    wikidata:Q21503252    wikidata:Q43229 ;
5         wikidata:Q82799    "Unión Europea" .
6
7 <#New_York_Times>    wikidata:Q21503252    wikidata:Q43229 ;
8         wikidata:Q82799    "New York Times" .
9
10 <#Tipo_euro>    wikidata:P1181    "1.3" ;
11              wikidata:Q21503252    wikidata:Q4916 .
12
13 <#Miles_Davis>    wikidata:Property:P106    wikidata:Q15981151 ;
14              wikidata:Property:P27    wikidata:Q30 ;
15              wikidata:Q21503252    foaf:Person ;
16              foaf:name    "Miles Davis" .

```

Output RDF data (Right Pane):

```

43
44 # Persona John McCarthy muerta
45 <#John_McCarthy>    wikidata:Q21503252    foaf:Person ;
46                    foaf:name    "John McCarthy" ;
47                    wikidata:Q21503252    wikidata:Q48422 .
48
49 # Lenguaje de programación desarrollado por John McCarthy
50 <#LISP>    wikidata:Q21503252    wikidata:Q9143 ;
51          wikidata:Q82799    "LISP" ;
52          wikidata:P178    <#John_McCarthy> .
53
54 # The New York Times informa sobre la muerte de John McCarthy
55 <#News>    wikidata:Q21503252    wikidata:Q732577 ;
56          wikidata:Property:P921    <#John_McCarthy> ;
57          wikidata:Property:P123    <#New_York_Times> .

```

Status Bar:

- Well formed RDF
- Data format: turtle

Fig. 1. Resultado de la validación del ttl con RDFSShape.

Una vez verificado podemos traducir el formato turtle a JSONLD, usando para ello la herramienta RDF-Translator, de forma que quedaria de la siguiente manera:

```
1 {
2   "@context": {
3     "foaf": "http://xmlns.com/foaf/0.1/" ,
4     "rdf": "http://www.w3.org/1999/02/22-rdf-syntax-ns#"
5     "rdfs": "http://www.w3.org/2000/01/rdf-schema#" ,
6     "wikidata": "http://www.wikidata.org/" ,
7     "xsd": "http://www.w3.org/2001/XMLSchema#"
8   },
9   "@graph": [
10    {
11      "@id": "file:///base/data/home/apps/s%7Erdft-
12        translator/2.408516547054015808/#Tipo-euro" ,
13      "wikidata:P1181": "1.3" ,
14      "wikidata:Q21503252": {
15        "@id": "wikidata:Q4916"
16      }
17    },
18    {
19      "@id": "file:///base/data/home/apps/s%7Erdft-
20        translator/2.408516547054015808/#Meeting" ,
21      "wikidata:Property:P710": [
22        {
23          "@id": "wikidata:Q458"
24        },
25        {
26          "@id": "wikidata:Q76"
27        }
28      ],
29      "wikidata:Property:P921": {
30        "@id": "file:///base/data/home/apps/s%7Erdft-
31          translator/2.408516547054015808/#tipo-euro"
32      },
33      "wikidata:Q21503252": {
34        "@id": "wikidata:Q2761147"
35      }
36    },
37    {
38      "@id": "file:///base/data/home/apps/s%7Erdft-
39        translator/2.408516547054015808/#Miles_Davis" ,
40      "foaf:name": "Miles Davis" ,
41      "wikidata:Property:P106": {
```

```

38         "@id": "wikidata:Q15981151"
39     },
40     "wikidata:Property:P27": {
41         "@id": "wikidata:Q30"
42     },
43     "wikidata:Q21503252": {
44         "@id": "foaf:Person"
45     }
46 },
47 {
48     "@id": "file:///base/data/home/apps/s%7Erdf-
        translator/2.408516547054015808/#
        New_York_Times" ,
49     "wikidata:Q21503252": {
50         "@id": "wikidata:Q43229"
51     },
52     "wikidata:Q82799": "New York Times"
53 },
54 {
55     "@id": "file:///base/data/home/apps/s%7Erdf-
        translator/2.408516547054015808/#John_McCarthy"
        ,
56     "foaf:name": "John McCarthy" ,
57     "wikidata:Q21503252": [
58         {
59             "@id": "wikidata:Q48422"
60         },
61         {
62             "@id": "foaf:Person"
63         }
64     ]
65 },
66 {
67     "@id": "file:///base/data/home/apps/s%7Erdf-
        translator/2.408516547054015808/#WDC" ,
68     "wikidata:Q21503252": {
69         "@id": "wikidata:Q515"
70     },
71     "wikidata:Q82799": "Washington"
72 },
73 {
74     "@id": "file:///base/data/home/apps/s%7Erdf-
        translator/2.408516547054015808/#UE" ,
75     "wikidata:Q21503252": {
76         "@id": "wikidata:Q43229"

```

```

77     },
78     "wikidata:Q82799": "Unión Europea"
79 },
80 {
81     "@id": "file:///base/data/home/apps/s%7Erdf-
        translator/2.408516547054015808/#Barack_Obama"

82     "foaf:name": "Barack Obama" ,
83     "wikidata:Property:P39": {
84         "@id": "wikidata:Q11696"
85     },
86     "wikidata:Q21503252": {
87         "@id": "foaf:Person"
88     }
89 },
90 {
91     "@id": "file:///base/data/home/apps/s%7Erdf-
        translator/2.408516547054015808/#LISP" ,
92     "wikidata:P178": {
93         "@id": "file:///base/data/home/apps/s%7Erdf-
            translator/2.408516547054015808/#
            John_McCarthy"
94     },
95     "wikidata:Q21503252": {
96         "@id": "wikidata:Q9143"
97     },
98     "wikidata:Q82799": "LISP"
99 },
100 {
101     "@id": "file:///base/data/home/apps/s%7Erdf-
        translator/2.408516547054015808/#News" ,
102     "wikidata:Property:P123": {
103         "@id": "file:///base/data/home/apps/s%7Erdf-
            translator/2.408516547054015808/#
            New_York_Times"
104     },
105     "wikidata:Property:P921": {
106         "@id": "file:///base/data/home/apps/s%7Erdf-
            translator/2.408516547054015808/#
            John_McCarthy"
107     },
108     "wikidata:Q21503252": {
109         "@id": "wikidata:Q732577"
110     }
111 }

```

```
112     ]
113   }
```

Listing 1.2. Archivo Resultado de la traduccion de turtle a JSONLD.

3 Obtención automática de información estructurada

3.1 Obtención de RDF's.

En esta sección obtendremos los documentos RDF para cada texto, par ello usaremos las siguientes herramientas: Open Calais y FRED. En el caso de la herramienta recomendada en el guión de prácticas DBPedia-spotlight no se hará uso de ella debido a que no exporta un archivo rdf de modo que la generación de este no sería automático como se indica en el titulo de la sección.

Los archivos resultantes se encuentran en el fichero comprimido adjunto en la tarea llamado *informacionEstructuradaAutomatica.zip*, este fichero contendrá tanto los archivos RDF de cada texto como los que posteriormente se habrán traducido a formato turtle y a JSONLD.

Dada esta información ahora podremos dar respuesta a las siguientes cuestiones:

¿Qué ontologías usa cada servicio para "tipar" las instancias detectadas en el texto?

Las ontologías que usa cada servicio para "tipar" las instancias son las siguientes:

Open Calais: En el caso de esta herramienta para todos los textos usa las siguientes ontologías, estas son fáciles de obtener de los archivos en formato turtle obtenidos de la traducción de los rdf.

- <http://s.opencalais.com/1/pred/>
- <http://www.w3.org/1999/02/22-rdf-syntax-ns>
- <http://www.w3.org/2000/01/rdf-schema>
- <http://www.w3.org/XML/1998/namespace>
- <http://www.w3.org/2001/XMLSchema>

FRED: En el caso de esta herramienta presenta diferencias en las ontologías usadas en funcion del texto puesto que en unos necesita mas que en otros. El conjunto de todas las ontologías usadas es el siguiente:

- <http://www.ontologydesignpatterns.org/ont/dul/DUL.owl>
- <http://www.essepuntato.it/2008/12/earmark>
- <http://www.ontologydesignpatterns.org/ont/fred/domain.owl>
- <http://www.ontologydesignpatterns.org/ont/vn/abox/role/>
- <http://www.ontologydesignpatterns.org/ont/cnlp/dependencies.owl>
- <http://ontologydesignpatterns.org/cp/owl/semiotics.owl>
- <http://www.ontologydesignpatterns.org/ont/fred/pos.owl>

- <http://www.ontologydesignpatterns.org/ont/boxer/boxer.owl>
- <http://www.ontologydesignpatterns.org/ont/boxer/boxer.owl>
- <http://www.ontologydesignpatterns.org/ont/fred/quantifiers.owl>
- <http://www.ontologydesignpatterns.org/ont/fred/quantifiers.owl>
- <http://www.w3.org/2002/07/owl>
- <http://www.w3.org/1999/02/22-rdf-syntax-ns>
- <http://www.w3.org/2000/01/rdf-schema>
- <http://schema.org/>
- <http://www.w3.org/XML/1998/namespace>
- <http://www.w3.org/2001/XMLSchema>

¿Existe algún tipo de dichas ontologías que pudiera considerarse equivalente a otro tipo en schema.org? Señala todos los casos que te hayas encontrado con cada servicio, indicando también (si fuera posible) qué propiedades de un tipo mapearían sobre propiedades del otro.

Existen muchos casos en los que un tipo de entidad de una de las ontologías usadas mapea con otro de Schema.org, algunos ejemplos son los siguientes:

En el caso de Open Calais nos encontramos con:

- La entidad "<http://s.opencalais.com/1/type/em/e/Person>" que en Schema.org sería "<https://schema.org/Person>", además en este caso mapearían directamente los siguientes atributos: name.
- La entidad "<http://s.opencalais.com/1/type/em/e/City>" que en Schema.org sería "<https://schema.org/City>", además en este caso mapearían directamente los siguientes atributos: name.
- La entidad "<http://s.opencalais.com/1/type/em/e/Organization>" que en Schema.org sería "<https://schema.org/Organization>", además en este caso mapearían directamente los siguientes atributos: name y additionalType.
- La entidad "<http://s.opencalais.com/1/type/tag/Industry>" que en Schema.org sería "<https://schema.org/industry>".
- La entidad "<http://s.opencalais.com/1/type/er/Company>" que en Schema.org sería "<https://schema.org/Corporation>", además en este caso mapearían directamente los siguientes atributos: name.
- La entidad "<http://s.opencalais.com/1/type/em/r/Quotation>" que en Schema.org sería "<https://schema.org/Quotation>", además en este caso mapearían directamente los siguientes atributos: about, spokenByCharacter.
- La entidad "<http://s.opencalais.com/1/type/em/e/PublishedMedium>" que en Schema.org sería "<https://schema.org/Periodical>", además en este caso mapearían directamente los siguientes atributos: name.
- La entidad "<http://s.opencalais.com/1/type/em/e/ProgrammingLanguage>" que en Schema.org sería "<https://schema.org/ComputerLanguage>", además en este caso mapearían directamente los siguientes atributos: name.

Por otro lado en el caso de FRED nos encontramos con los siguientes casos:

- La entidad <http://www.w3.org/2002/07/owl/Class> que en Schema.org sería <https://schema.org/Class>, además en este caso mapearían directamente los siguientes atributos: subObject.

- La entidad "http://www.ontologydesignpatterns.org/ont/fred/domain.owl/ProgrammingLanguage" que en Schema.org sería https://schema.org/ComputerLanguage, además en este caso mapearían directamente los siguientes atributos: name.
- La entidad "http://www.ontologydesignpatterns.org/ont/fred/domain.owl/Report" que en Schema.org sería https://schema.org/Report.

Cabe destacar que en el caso de la herramienta FRED ya se usa Schema.org como ontología para la generación de los archivos.

Reflexiona acerca de los motivos que pueden llevar a cada equipo de desarrolladores a producir una ontología propia. Investiga (somergidamente) sobre el problema de alineación de ontologías.

Puede haber diversas razones por las que un equipo de desarrolladores produzcan una ontología propia; la primera es que esta se cree por mero desconocimiento de la existencia previa que satisfaga sus necesidades, otra razón podría ser que quisieran que la ontología fuera hecha a medida de sus necesidades, de modo que añadirían a cada entidad los atributos que desearan para su propósito, además otra razón sería que si la creasen ellos podrían asegurar el mantenimiento de esta asegurando así que no se discontinuase, como podría ocurrir con otras como wikiData o Schema.org, por último cabría destacar la posibilidad de que sea un requerimiento del proyecto la creación de dicha ontología porque esta no exista o se necesita que sea cerrada.

En cuanto al problema de alineación entre ontologías esto se produce en el momento que la misma entidad está definida en más de una ontología y:

- La nomenclatura de los atributos es distinta.
- En unas ontologías aparecen unos atributos y en otras no.
- La definición es ligeramente distinta.

Esto produce que trabajar con varias de estas ontologías al mismo tiempo sea complejo y tedioso, por ejemplo en el caso de realizar consultas SPARQL, que mediante el uso de consultas federadas podríamos transformar las entidades de una ontología en las mismas de otra esto no se puede hacer directamente por lo anteriormente mencionado. Por ello en muchas ontologías se incluye en todas las entidades el atributo sameAs en el que se recoge una lista de equivalencias con otras ontologías, hay que mencionar que en el caso de wikiData esto se realiza de diferente manera de modo que las entidades en vez de tener un atributo sameAs con una lista de equivalencias poseen un atributo de equivalencia por cada ontología que tenga registrada para ello.

Utiliza el servicio sameAs.org para localizar equivalencias para todos los tipos fundamentales de Open Calais, DBPedia Spotlight y FRED, así como los homólogos de Schema.org que detectaste. ¿Existe algún servicio para el que no se disponga de información en sameAs.org?

El único caso en el que he conseguido alguna equivalencia ha sido en el caso de la ontología de Schema.org, algunas de estas son las siguientes:

1 [

2 {

```

3      "uri": "http://schema.org/Thing" ,
4      "numDuplicates": "2" ,
5      "duplicates": [
6          "http://schema.org/Thing" ,
7          "http://www.w3.org/2002/07/owl#Thing"
8      ]
9  },
10 {
11     "uri": "http://purl.org/vocabularies/princeton/wn30/
        synset-person-noun-1" ,
12     "numDuplicates": "12" ,
13     "duplicates": [
14         "http://dbpedia.org/ontology/Person" ,
15         "http://dbpedia.org/ontology/PokerPlayer" ,
16         "http://proton.semanticweb.org/protonu#Profession" ,
17         "http://purl.org/vocabularies/princeton/wn30/synset
        -person-noun-1" ,
18         "http://schema.org/Person" ,
19         "http://sw.opencyc.org/2009/04/07/concept/en/
        _http---dbpedia_org_ontology_PokerPlayer_" ,
20         "http://umbel.org/umbel#PersonTypes" ,
21         "http://umbel.org/umbel/rc/PersonWithOccupation" ,
22         "http://www.w3.org/2000/10/swap/pim/contact#Person"
23     ,
24     "http://xmlns.com/foaf/0.1/Person"
25 ]
26 },
27 {
28     "uri": "http://schema.org/Organization" ,
29     "numDuplicates": "8" ,
30     "duplicates": [
31         "http://dbpedia.org/ontology/Organisation" ,
32         "http://proton.semanticweb.org/protonu#Organization
        " ,
33         "http://proton.semanticweb.org/protonu#
        SportOrganization" ,
34         "http://purl.org/vocabularies/princeton/wn30/synset
        -organization-noun-1" ,
35         "http://schema.org/Organization" ,
36         "http://umbel.org/umbel/rc/Organization" ,
37         "http://umbel.org/umbel/rc/Organizations" ,
38         "http://xmlns.com/foaf/0.1/Organization"
39     ]
40 },
41 {

```

```

41     "uri": "http://schema.org/City" ,
42     "numDuplicates": "507" ,
43     "duplicates": [
44         "http://dbpedia.org/ontology/City" ,
45         "http://dbpedia.org/resource/City" ,
46         "http://dbpedia.org/resource/Town" ,
47         "http://dbpedia.org/resource/Citie" ,
48         "http://dbpedia.org/resource/Shahr" ,
49         "http://dbpedia.org/resource/Cities" ,
50         "http://dbpedia.org/resource/Cittie" ,
51         "http://dbpedia.org/resource/CittA " ,
52         "http://dbpedia.org/resource/Ciudad" ,
53         "http://dbpedia.org/resource/Citties" ,
54         "http://dbpedia.org/resource/Pilseta" ,
55         "http://dbpedia.org/resource/City_work" ,
56         "http://dbpedia.org/resource/Porkberry"
57     ]
58 }
59 ]

```

Listing 1.3. Equivalencias de sameAs.org para Schema.org.

Al realizar este apartado no he conseguido para el resto de ontologías que la web sameAs.org funcione, he probado con los tipos básicos de cada ontología asegurandome que tuviese un atributo sameAs pero aun asi no funciona como se observa en la siguientes imagenes.



Fig. 2. Resultado la búsqueda de <http://dbpedia.org/page/Thing> en SameAs.org.



Fig. 3. Resultado acceder de <http://dbpedia.org/page/Thing> y buscar el atributo sameAs.

4 Anexos

4.1 Textos de Prueba

Texto 1: “Miles Davis was an american jazz musician.”

Texto 2: “President Barack Obama and European Union leaders huddled in Washington amid growing fears over the future of the euro, which closed greater than 1.3 dollars.”

Texto 3: “The New York Times reported that John McCarthy died. He invented the programming language LISP.”