



# DH@Madrid Summer School 2016

## EJERCICIOS

Aprovechamiento de las tecnologías semánticas y los recursos enlazados en el análisis de textos multilingües

**Tecnologías digitales aplicadas al estudio de la poesía**

Elena Montiel Ponsoda, Víctor Rodríguez Doncel



European  
Research  
Council

Madrid, 29 junio de 2016



# Ejercicios

- (1) Encuentre en Cervantes Virtual los metadatos en RDF de la obra “Proemio y Carta”. Descargue el archivo y ábralo con un visor de archivos de texto (Notepad++ o equivalente).
- (2) ¿Qué URI identifica al recurso descrito?
- (3) ¿Cuál es el nombre (rdfs:label) del recurso descrito?



# Ejercicios

(4) (Opcional/Después de la sesión) RDF se puede escribir de varias maneras. Transformar el recurso del formato en que se encuentra (XML/RDF) al formato que hemos visto en la sesión (NTRIPLES).

Se recomienda utilizar el conversor que se puede descargar en esta página <http://www.l3s.de/~minack/rdf2rdf/>

(5) Identificar intuitivamente cómo viene representado el idioma de la obra. ¿Mediante un recurso, mediante una constante o mediante ambas?

(6) Si el idioma viene identificado como un recurso, identificar con qué otros datasets está enlazado.



# Ejercicios

(8) Desambiguar con Babelfy la frase “Santillana, jugador del Real Madrid, no fue Marqués de Santillana”. (Babelfy es una herramienta online).

(9) Identificar las entidades nombradas y determinar si se ha desambiguado correctamente o no



# Ejercicios

(10 Avanzado) Desambiguar la frase anterior utilizando la API de babelfy. Tendrá que registrarse por adelantado para obtener una API key

(11 Avanzado) Con los resultados del ejercicio anterior, encontrar la URI que identifica al Marqués de Santillana en Babelnet. Encontrar también la URI que identifica al synset de “jugador”

(12 Avanzado) Descubra la herramienta para enriquecer ontologías con info lingüística en lemon-ontolex  
<http://lemonadetools.linkeddata.es/>





# DH@Madrid Summer School 2016

## EJERCICIOS

Aprovechamiento de las tecnologías semánticas y los recursos enlazados en el análisis de textos multilingües

**Tecnologías digitales aplicadas al estudio de la poesía**

Elena Montiel Ponsoda, Víctor Rodríguez Doncel



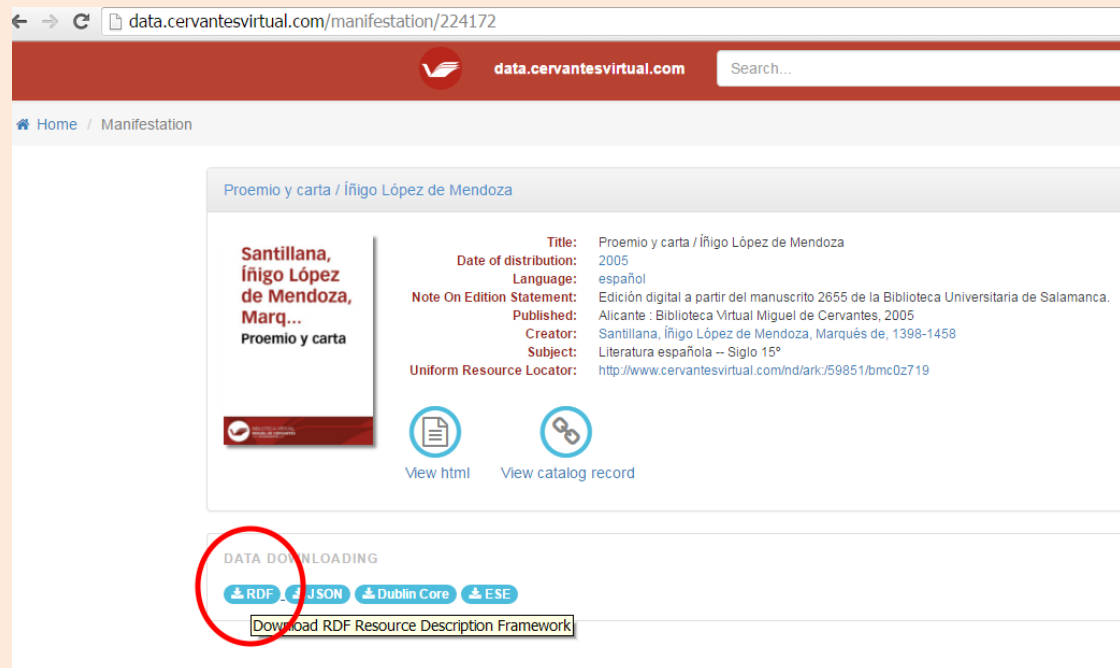
European  
Research  
Council

Madrid, 29 junio de 2016



# Soluciones (algunas)

(1) Encuentre en Cervantes Virtual los metadatos en RDF de la obra “Proemio y Carta”. Descargue el archivo y ábralo con un visor de archivos de texto (Notepad++ o equivalente).



data.cervantesvirtual.com/manifestation/224172

data.cervantesvirtual.com Search...

Home / Manifestation

Proemio y carta / Íñigo López de Mendoza

**Santillana, Íñigo López de Mendoza, Marq...**  
Proemio y carta

Title: Proemio y carta / Íñigo López de Mendoza  
Date of distribution: 2005  
Language: español  
Note On Edition Statement: Edición digital a partir del manuscrito 2655 de la Biblioteca Universitaria de Salamanca.  
Published: Alicante : Biblioteca Virtual Miguel de Cervantes, 2005  
Creator: Santillana, Íñigo López de Mendoza, Marqués de, 1398-1458  
Subject: Literatura española -- Siglo 15?  
Uniform Resource Locator: http://www.cervantesvirtual.com/nd/ark:/59851/bmc0z719

View html View catalog record

DATA DOWNLOADING

RDF JSON Dublin Core ESE

Download RDF Resource Description Framework



# Soluciones (algunas)



(2) ¿Qué URI identifica al recurso descrito?

<http://data.cervantesvirtual.com/manifestation/224172>

(3) ¿Cuál es el nombre (rdfs:label) del recurso descrito?

**Proemio y carta / Íñigo López de  
Mendoza**





# Soluciones (algunas)

(5) Identificar intuitivamente cómo viene representado el idioma de la obra. ¿Mediante un recurso, mediante una constante o mediante ambas?

## COMO CONSTANTE:

```
<inLanguage xmlns="http://schema.org/"  
rdf:datatype="http://www.w3.org/2001/XMLSchema#string">  
es</inLanguage>
```

## COMO RECURSO

(<http://data.cervantesvirtual.com/language/es> ):

```
<language xmlns="http://purl.org/dc/elements/1.1/"  
rdf:resource="http://data.cervantesvirtual.com/language/es"/  
>
```



# Soluciones (algunas)

(6) Si el idioma viene identificado como un recurso, identificar con qué otros datasets está enlazado.

**Podemos navegar por la nube de datos enlazados. Si navegamos a: <http://data.cervantesvirtual.com/language/es>**

**Nos encontramos con enlaces a tres datasets:**

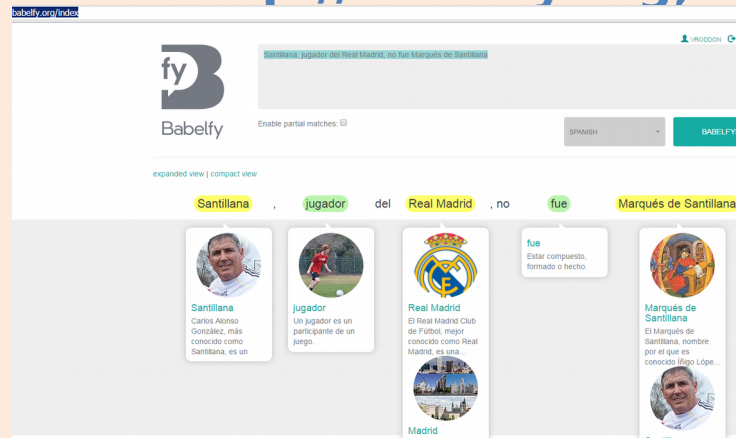
- **Wikipedia:**  
[http://en.wikipedia.org/wiki/Spanish\\_language](http://en.wikipedia.org/wiki/Spanish_language)
- **Library of Congress:**  
<http://id.loc.gov/vocabulary/iso639-1/es>
- **Dbpedia:**  
[http://dbpedia.org/resource/Spanish\\_language](http://dbpedia.org/resource/Spanish_language)



# Soluciones (algunas)

(8) Desambiguar con Babelify la frase “Santillana, jugador del Real Madrid, no fue Marqués de Santillana”. (Babelify es una herramienta online).

**Basta con acudir a la web <http://babelify.org>, seleccionar español e introducir la frase.**



(9) Identificar las entidades nombradas y determinar si se ha desambiguado correctamente o no

**Carlos Alonso (Santillana)**

**Marqués de Santillana**

**Real Madrid**



# Soluciones (algunas)

(10 Avanzado) Desambiguar la frase anterior utilizando la API de babelfy. Tendrá que registrarse por adelantado para obtener una API key

1) Descubra la forma de invocar a la API. Documentación en: <http://babelfy.org/guide>

La API call tiene esta forma:

<https://babelfy.io/v1/disambiguate?text={text}&lang={lang}&key={key}>

2) Utilice algún servicio web para transformar la frase a una codificación válida para ser insertada en la URI. Por ejemplo:

<http://meyerweb.com/eric/tools/dencoder/>

[Santillana%2C%20jugador%20del%20Real%20Madrid%2C%20no%20fue%20Marqu%C3%A9s%20de%20Santillana](http://meyerweb.com/eric/tools/dencoder/)

3) Construya la llamada incluyendo la API key

<https://babelfy.io/v1/disambiguate?text=Santillana%2C%20jugador%20del%20Real%20Madrid%2C%20no%20fue%20Marqu%C3%A9s%20de%20Santillana&lang=es&key=06e8317d-4d2d-49ec-b2ef-13f7889c78cV>

4) Invóquela utilizando un navegador cualquiera



# Soluciones (algunas)

(10 Avanzado) Desambiguar la frase anterior utilizando la API de babelfy. Tendrá que registrarse por adelantado para obtener una API key

```
[{"tokenFragment":{"start":2,"end":2},"charFragment":
{"start":12,"end":18},"babelSynsetID":"bn:00060799n","DBpediaURL":"http://dbpedia.org/resource/Player_(game
)","BabelNetURL":"http://babelnet.org/rdf/s00060799n","score":0.0,"coherenceScore":0.0,"globalScore":0.0,"sourc
e":"MCS"}, {"tokenFragment":{"start":4,"end":4},"charFragment":
{"start":24,"end":27},"babelSynsetID":"bn:01997706n","DBpediaURL":"http://dbpedia.org/resource/Real_Madrid_
C.F.","BabelNetURL":"http://babelnet.org/rdf/s01997706n","score":0.8064516129032258,"coherenceScore":0.555
555555555556,"globalScore":0.23006134969325154,"source":"BABELFY"}, {"tokenFragment":
{"start":5,"end":5},"charFragment":
{"start":29,"end":34},"babelSynsetID":"bn:00015620n","DBpediaURL":"http://dbpedia.org/resource/Madrid","Bab
elNetURL":"http://babelnet.org/rdf/s00015620n","score":0.7446808510638298,"coherenceScore":0.555555555555
5556,"globalScore":0.2147239263803681,"source":"BABELFY"}, {"tokenFragment":
{"start":7,"end":7},"charFragment":
{"start":37,"end":38},"babelSynsetID":"bn:00116360r","DBpediaURL":"","BabelNetURL":"http://babelnet.org/rdf/s
00116360r","score":1.0,"coherenceScore":0.1111111111111111,"globalScore":0.009202453987730062,"source":
"BABELFY"}, {"tokenFragment":{"start":8,"end":8},"charFragment":
{"start":40,"end":42},"babelSynsetID":"bn:00083187v","DBpediaURL":"","BabelNetURL":"http://babelnet.org/rdf/s
00083187v","score":0.0,"coherenceScore":0.0,"globalScore":0.0,"source":"MCS"}, {"tokenFragment":
{"start":9,"end":9},"charFragment":
{"start":44,"end":50},"babelSynsetID":"bn:00053513n","DBpediaURL":"http://dbpedia.org/resource/Marquess","B
abelNetURL":"http://babelnet.org/rdf/s00053513n","score":1.0,"coherenceScore":0.3333333333333333,"globalSc
ore":0.027607361963190184,"source":"BABELFY"}, {"tokenFragment":{"start":9,"end":11},"charFragment":
{"start":44,"end":64},"babelSynsetID":"bn:02538511n","DBpediaURL":"http://dbpedia.org/resource/Íñigo_López_d
e_Mendoza,_1st_Marquis_of_Santillana","BabelNetURL":"http://babelnet.org/rdf/s02538511n","score":1.0,"coheren
ceScore":0.2222222222222222,"globalScore":0.012269938650306749,"source":"BABELFY"}]
```



# Soluciones (algunas)

(11 Avanzado) Con los resultados del ejercicio anterior, encontrar la URI que identifica al Marqués de Santillana en Babelnet.

Encontrar también la URI que identifica al synset de “jugador”

Uri de jugador: <http://babelnet.org/rdf/page/s00060799n>

← → ↻ babelnet.org/rdf/page/s00060799n

| Home | | SPARQL endpoint |

**S00060799n**  
<http://babelnet.org/rdf/s00060799n>

skos: Concept

Property	Value
skos:broader	<ul style="list-style-type: none"><li>bn: s000006759n</li><li>bn: s00018350n</li><li>bn: s00022157n</li><li>bn: s00022164n</li><li>bn: s00037180n</li><li>bn: s00074905n</li></ul>
Is skos:broader of	18667
bn-lemon:dbpediaCategory	dbpedia: Category:Game_terminology
bn-lemon:definition	77
lexinfo:derivedForm	bn: s00091708v
Is lexinfo:derivedForm of	bn: s00091708v
skos:exactMatch	<ul style="list-style-type: none"><li>dbpedia: Player_(game)</li><li>lemon-Omega: OW_eng_Synset_12994</li><li>freebase: m.02vz9</li><li>lemon-WordNet31: 110459618-n</li><li>yago: Player_(game)</li></ul>
lemon:isReferenceOf	137
dcterms:license	< <a href="http://creativecommons.org/licenses/by-nc-sa/3.0/">http://creativecommons.org/licenses/by-nc-sa/3.0/</a> >
skos:narrower	22
Is skos:narrower of	<ul style="list-style-type: none"><li>bn: s000006759n</li><li>bn: s00022164n</li><li>bn: s00037180n</li></ul>
Is lemon:reference of	137
skos:related	61
Is skos:related of	135
bn-lemon:synsetID	bn:00060799n
bn-lemon:synsetType	concept
bn-lemon:wikipediaCategory	13
bn-lemon:wiktionaryPageLink	wiktionary: participant





# DH@Madrid Summer School 2016

## EJERCICIOS

Aprovechamiento de las tecnologías semánticas y los recursos enlazados en el análisis de textos multilingües

**Tecnologías digitales aplicadas al estudio de la poesía**

Elena Montiel Ponsoda, Víctor Rodríguez Doncel



European  
Research  
Council

Madrid, 29 junio de 2016

