

SIW PL8. Información semántica.

Hugo Fonseca Díaz
uo258318@uniovi.es

Escuela de Ingeniería Informática. Universidad de Oviedo.

13 de noviembre de 2020

1. Ejercicio 1

Para este ejercicio se ha utilizado la noticia de la *BBC* sobre el retorno del expresidente Evo Morales a Bolivia [1]. Los párrafos utilizados se encuentran en texto plano en el archivo *noticia.txt*. Al pasar dicho texto por la herramienta de *PermID* disponible en [2], se obtiene una serie de entidades. Una vez obtenidas, procedemos a complementarlas con nuestro propio conocimiento, por lo que nos quedan las siguientes entidades:

- **Evo Morales:** Person(familyName: Morales, givenName: Evo, hasOccupation: Role(hasOccupation: Occupation(name: President), startDate: 2006, endDate: 2019), nationality: Country(Bolivia), knows: Person(Luis Arce), member: Organization(Movimiento al Socialismo), url: <https://www.wikidata.org/wiki/Q4241>)
- **Luis Arce:** Person(familyName: Arce, givenName: Luis, hasOccupation: Occupation(name: President), nationality: Country(Bolivia), knows: Person(Evo Morales), member: Organization(Movimiento al Socialismo), url: <https://www.wikidata.org/wiki/Q16149892>)
- **Bolivia:** Country(name: Bolivia, url: <https://www.wikidata.org/wiki/Q750>)
- **Argentina:** Country(name: Argentina, url: <https://www.wikidata.org/wiki/Q414>)
- **Movimiento al Socialismo:** Organization(name: Movimiento al Socialismo, legalName: Movimiento al Socialismo, url: <https://www.wikidata.org/wiki/Q1142239>)
- **Elección presidencial:** VoteAction(name: Presidential election, participant: Organization(Movimiento al Socialismo), candidate: Person(Luis Arce), url: <https://www.wikidata.org/wiki/Q74596528>)

¿Deberían tener valor todos los *itemid*? Ya que una de las grandes ventajas de etiquetar los documentos HTML con microdatos es el interlazado de los mismos, todos las entidades deberían tener un *id* único que les identifique.

¿Qué valor debería asignarse? Como la forma en la que nos referimos a una entidad varia según diversos factores, entre los que se encuentran cosas como el idioma, es mejor escoger un identificador

que sea independiente a estas cuestiones. En nuestro caso se ha escogido la entrada de dichas entidades en la *Wikidata* [3], un grafo de conocimiento de datos abiertos y de ámbito colaborativo. Wikidata es ampliamente utilizada y por tanto la información se revisa a menudo, por lo que incluso si hubiese falta de información o información incorrecta las ventajas a largo plazo son mayores. Además, los identificadores de Wikipedia están específicamente diseñados para nuestro problema, por lo que es una opción recomendable.

Inconveniencias a la hora de incrustar los metadatos El mayor problema al definir los metadatos para este ejercicio ha sido el tener que respetar la estructura original de la noticia. Esto ha causado que ciertas propiedades y relaciones descritas anteriormente en el documento (cuando hablábamos de las entidades definidas) no hayan podido etiquetarse correctamente en el documento HTML. Es por esto también por lo que ciertas propiedades, como la ocupación de las personas, den lugar a errores al validar el documento con la herramienta de *Google Structured Data Testing Tool* [4]. Esto se debe al hecho de que no disponemos de los datos para poder etiquetar ciertos metadatos considerados necesarios por la herramienta.

Dicha validación puede observarse en la **Figura 1**.

Figura 1: Validación del documento HTML del ejercicio 1.

The screenshot shows the Google Structured Data Testing Tool interface. On the left, there is a text area containing HTML code with JSON-LD snippets. On the right, a table lists validation errors and warnings. The errors include missing values for 'estimatedSalary', 'occupationLocation', and 'description'. Warnings indicate that 'description' and 'mainEntityOfPage' are recommended but not provided. The table also shows detected entities like 'nationality' and 'hasOccupation' with their respective types and names.

Icon	Property	Message
✖	estimatedSalary	Se requiere un valor para el campo estimatedSalary.
✖	occupationLocation	Se requiere un valor para el campo occupationLocation.
⚠	description	El campo description está recomendado. Proporciona un valor si está disponible.
⚠	mainEntityOfPage	El campo mainEntityOfPage está recomendado. Proporciona un valor si está disponible.
	nationality	
	@type	Country
	name	Bolivia
	memberOf	
	@type	Thing
	name	Mas
	hasOccupation	
	@type	Occupation
	name	first indigenous president
✖	estimatedSalary	Se requiere un valor para el campo estimatedSalary.
✖	occupationLocation	Se requiere un valor para el campo occupationLocation.
⚠	description	El campo description está recomendado.

2. Ejercicio 2

En el ejercicio 2 hacemos una serie de mejoras en el fichero *json-ld* disponible en [5] y junto al contenido de [6] despojado de microdatos creamos un documento HTML llamado *exercise2.html*.

Referencias

- [1] Noticia de la BBC sobre el retorno del expresidente Evo Morales a Bolivia, <https://www.bbc.com/news/world-latin-america-54872827>. Última vez accedido 12 de noviembre de 2020.
- [2] *Intelligent Tagging Demo*, herramienta de *PermID*, <https://permid.org/onecalaisViewer>. Última vez accedido 12 de noviembre de 2020.
- [3] Página de la *Wikidata*, https://www.wikidata.org/wiki/Wikidata:Main_Page. Última vez accedido 12 de noviembre de 2020.
- [4] Página de la *Google Structured Data Testing Tool*, <https://search.google.com/structured-data/testing-tool/>. Última vez accedido 12 de noviembre de 2020.
- [5] Página de Daniel Gayo-Avello donde encontramos el fichero *json-ld*, <http://danigayo.info/ejemplos-SIW/03-json-ld-v3.json>. Última vez accedido 12 de noviembre de 2020.
- [6] Página de Daniel Gayo-Avello donde encontramos el documento HTML con microdatos incrustados, <http://danigayo.info/ejemplos-SIW/microdata.html>. Última vez accedido 12 de noviembre de 2020.