



Escuela de
Ingeniería
Informática
Universidad de Oviedo

Sistemas de Información para la Web

Práctica 8 - Información semántica : Microdatos

Esteban Montes Morales - UO246305

Ejercicio 1: Cuestiones

¿Deberían tener valor para itemid todos los items?

No tiene por qué, sintácticamente hablando, pero es altamente recomendable, ya que normalmente sirven para identificar de forma única las entidades, y de esta forma podrá obtenerse la información necesaria de ella.

¿Qué valor debería asignarse?

Una URI/URL/URN que identifique de forma concisa dicha entidad y que sirva de fuente de información para la misma. Dependiendo del contexto, es muy posible que la decisión del itemid sea crucial.

¿Si hubiera varios candidatos cuál escogerías? ¿Por qué?

Tal y como comentaba antes, depende de la situación. Tal vez si el texto que estoy “tagueando” se sitúa en un contexto bibliográfico, prefiera utilizar identificadores que tengan por ejemplo como fuente la wikipedia, mientras que en caso de estar “tagueando” un texto meramente informativo, como por ejemplo, una guía turística, prefiera utilizar los enlaces a los sitios oficiales, o a páginas relevantes dedicadas al turismo.

¿Cuál crees que es la solución de compromiso que podría resultar menos polémica en la mayoría de los casos?

Creo que fuentes como wikipedia serían posiblemente las más correctas genéricamente, debido al contenido variopinto que hay para cada entrada.

¿Qué inconvenientes concretos te ha supuesto la obligación de incrustar los metadatos allí donde te “forzaba” la estructura del texto?

- Falta de expresividad
 - En mi opinión se hace muy verboso y es muy complejo anidar y relacionar información y entidades.
- Ambigüedad a la hora de identificar el tipo de la entidad
 - Aunque esto también es un problema común a RDF o a JSON-LD, y es que es muy importante elegir correctamente el tipo, ya que vas a definir información del texto en base a las propiedades disponibles para ese tipo concreto.
- Duplicado de entidades
 - Una misma entidad puede aparecer repetida en múltiples ocasiones, y tener que definir una entidad con el mismo tipo y el mismo identificador, en lugar de poder reutilizarlo, me parece un punto desfavorable.

- No se puede añadir información adicional no visible (metainformación)
 - Debes ceñirte a la estructura del texto, y no puedes añadir información que no esté presente como tal en el mismo. Esto es una desventaja clara, ya que tal vez al lector humano, no le interese toda esa información adicional no visible, pero una máquina puede utilizarlo para relacionar información y generar nueva.

Ejercicio 1: Herramienta de pruebas de Google

Google Herramienta de pruebas de datos estructurados

doi

NUEVA PRUEBA

```

1 <div itemscope itemtype="http://schema.org/Country"
  itemid="http://www.spain.info/">
2   <p>Durante su visita a <span itemprop="name">España</span> el 6 de
  Noviembre,</p>
3 </div>
4 <div itemscope itemtype="http://schema.org/Person"
  itemid="http://government.ru/en/gov/persons/15/events/">
5   <p>el <span itemprop="jobTitle">ministro </span><span
  itemprop="nationality">ruso </span><span itemprop="name">Lavrov</span>
  logró que su homólogo </p>
6 </div>
7 <div itemscope itemtype="http://schema.org/Person"
  itemid="http://www.exteriores.gob.es/Portal/en/Ministerio/Ministro/Pag
  inas/Biografia.aspx">
8   <p><span itemprop="nationality">español</span>, <span
  itemprop="name">Josep Borrell</span>, se comprometiera a crear un
  grupo de trabajo conjunto
9   para luchar contra la desinformación, algo que fue recibido con
  escepticismo.</p>
10 </div>
11 <div itemscope itemtype="http://schema.org/Organization"
  itemid="http://www.nato.int/">
12   <p>&quot;De momento no se ha firmado nada, pero desde luego la
  imagen de España
13   entre los socios de la <span itemprop="name">OTAN</span> ha
  quedado por los suelos&quot;; asegura</p>
14   <div itemscope itemtype="http://schema.org/Person"
  itemid="http://www.realinstitutoelcano.org/wps/portal/rielcano_es/biog
  rafia/?WCM_GLOBAL_CONTEXT=/elcanov2_es/elcanov2/investigadores/mira-
  milosevich-juaristi">
15     <p>
16       <span itemprop="name">Mira Milosevich</span>,
17       <span itemprop="jobTitle">analista principal</span> del
18     </p>
19     <div itemscope itemtype="http://schema.org/Organization"
  itemid="http://www.realinstitutoelcano.org/">
20       <p>
21         <span itemprop="legalName">Real Instituto Elcano</span>.
22       </p>
23     </div>
24   </div>
25 </div>
26 <div itemscope itemtype="http://schema.org/City"

```

Detectado 0 ERRORES 0 ADVERTENCIAS 19 ELEMENTOS

Organization	0 ERRORES	0 ADVERTENCIAS	2 ELEMENTOS
Country	0 ERRORES	0 ADVERTENCIAS	6 ELEMENTOS
City	0 ERRORES	0 ADVERTENCIAS	2 ELEMENTOS
LandmarksOrHistoricalBuildings	0 ERRORES	0 ADVERTENCIAS	1 ELEMENTO
Person	0 ERRORES	0 ADVERTENCIAS	7 ELEMENTOS
Place	0 ERRORES	0 ADVERTENCIAS	1 ELEMENTO

Ejercicio 2: Herramienta de pruebas de Google

Google Herramienta de pruebas de datos estructurados

NUEVA PRUEBA

```
1 <!DOCTYPE HTML>
2 <html>
3 <head>
4   <meta http-equiv="Content-Type" content="text/html; charset=UTF-8"
5 />
6   <title>Ejercicio 2</title>
7   <script type="application/ld+json">
8     {
9       "@context": {
10         "cc": "http://creativecommons.org/ns#",
11         "ctag": "http://comontag.org/ns#",
12         "dc": "http://purl.org/dc/elements/1.1/",
13         "dcl": "http://purl.org/dc/terms/",
14         "dcl1": "http://purl.org/dc/elements/1.1/",
15         "dcat": "http://www.w3.org/ns/dcat#",
16         "dcterms": "http://purl.org/dc/terms/",
17         "foaf": "http://xmlns.com/foaf/0.1/",
18         "gr": "http://purl.org/goodrelations/v1#",
19         "grddl": "http://www.w3.org/2003/g/data-view#",
20         "hcalendar": "http://microformats.org/profile/hcalendar#",
21         "hcard": "http://microformats.org/profile/hcard#",
22         "ical": "http://www.w3.org/2002/12/cal/icaltzd#",
23         "ma": "http://www.w3.org/ns/ma-ont#",
24         "md": "http://www.w3.org/ns/md#",
25         "og": "http://ogp.me/ns#",
26         "org": "http://www.w3.org/ns/org#",
27         "owl": "http://www.w3.org/2002/07/owl#",
28         "prov": "http://www.w3.org/ns/prov#",
29         "qb": "http://purl.org/linked-data/cube#",
30         "rdf": "http://www.w3.org/1999/02/22-rdf-syntax-ns#",
31         "rdfa": "http://www.w3.org/ns/rdfa#",
32         "rdfs": "http://www.w3.org/2000/01/rdf-schema#",
33         "rev": "http://purl.org/stuff/rev#",
34         "rif": "http://www.w3.org/2007/rif#",
35         "rml": "http://www.w3.org/ns/r2ml#",
36         "schema": "http://schema.org/",
37         "sd": "http://www.w3.org/ns/sparql-service-description#",
38         "sioc": "http://rdfs.org/sioc/ns#",
39         "skos": "http://www.w3.org/2004/02/skos/core#",
40         "skosxl": "http://www.w3.org/2008/05/skos-xl#",
41         "v": "http://rdf.data-vocabulary.org/#",
42         "vcard": "http://www.w3.org/2006/vcard/ns#",
43         "void": "http://rdfs.org/ns/void#",
44         "wdr": "http://www.w3.org/2007/05/powder#",
45         "wdrs": "http://www.w3.org/2007/05/powder-s#",
```

Detectado 0 ERRORES 3 ADVERTENCIAS 11 ELEMENTOS

Periodical	0 ERRORES	0 ADVERTENCIAS	4 ELEMENTOS
Chapter	0 ERRORES	0 ADVERTENCIAS	1 ELEMENTO
ScholarlyArticle	0 ERRORES	3 ADVERTENCIAS	5 ELEMENTOS
Person	0 ERRORES	0 ADVERTENCIAS	1 ELEMENTO