Universidad de Oviedo

Prácticas de laboratorio

Sesión 7

Sistemas de Información para la Web

Información semántica



Omar Teixeira González, UO281847

23/11/2023

Tabla de contenidos

Tabla de contenidos	1
Documentación sobre Información semántica	2
Tipo de contenido a etiquetar.	2
Ejemplos de ilustración.	2
Etiquetado en JSON-LD.	2
Texto planoFichero JSON-LD	2
Fichero JSON-LD	3
Validación del etiquetado	4
Demo de Dandelion.	5
¿Qué diferencias hay entre los datos estructurados que se ofrecen con JSON-LD en realidad a Google y sus algoritmos—y el tipo de información obtenida media	•
¿Es viable etiquetar en todo o en parte el contenido de tu caso usando el	vocabulario de
Schema.org para anotar la información detectada mediante entity linking	g?5

SIW 1 | Página

Documentación sobre Información semántica

Tipo de contenido a etiquetar.

 $3892095 \ mod \ 5 = 0 \rightarrow Libro$

Ejemplos de ilustración.

- El ladrón del Rayo (Percy Jackson y los dioses del Olimpo I), (Rick Riordan).
- El universo, (Isaac Asimov).
- Los juegos del Hambre, (Suzanne Collins).

Etiquetado en JSON-LD.

Texto plano.

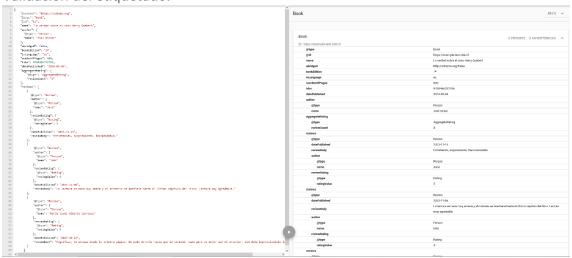
- > Título: La verdad sobre el caso Harry Quebert.
- > Autor: Joel Dickler.
- Editorial: DEBOLSILLO.
- > Idioma: Castellano.
- № Nº de páginas: 800.
- > ISBN: 9788466332286.
- ➤ Edición: 1ª.
- Edición limitada: No.
- Fecha de lanzamiento: 04/02/2016.
- Reviews:
 - o Review1:
 - Autor: Jone.
 - Puntuación: 5.
 - Fecha de publicación: 14/11/2023.
 - Contenido: Entretenido, sorprendente. Recomendable.
 - o Review2:
 - Autor: Inés.
 - Puntuación: 4.
 - Fecha de publicación: 06/11/2023.
 - Contenido: La lectura se hace muy amena y el misterio se mantiene hasta el último capítulo del libro. Lectura muy agradable.
 - o Review3:
 - Autor: María Jesús Albarés Carrasco.
 - Puntuación: 5.
 - Fecha de publicación: 22/10/2023.
 - Contenido: Magnífico, te atrapa desde la primera página. No pude dejarlo hasta que lo terminé. Cada giro es mejor que el anterior. Sin duda imprescindible leerlo.

SIW 2 | Página

```
"@context": "https://schema.org",
  "@type": "Book",
  "@id": "id",
  "name": "La verdad sobre el caso Harry Quebert", "author": {
   "@type": "Person",
   "name": "Joel Dicker"
 },
"abridged": false,
  "bookEdition": "1ª",
  "inLanguage": "es",
  "numberOfPages": 800,
  "isbn": 9788466332286,
  "datePublished": "2016-02-04",
  "aggregateRating": {
    "@type": "AggregateRating",
    "reviewCount": "3"
  "reviews": [
    {
     "@type": "Review",
     "author": {
      "@type": "Person",
"name": "Jone"
     "reviewRating": {
      "@type": "Rating",
      "ratingValue": 5
     },
"datePublished": "2023-11-14",
     "reviewBody": "Entretenido, sorprendente. Recomendable."
      "@type": "Review",
      "author": {
       "@type": "Person",
       "name": "Inés"
      "reviewRating": {
       "@type": "Rating",
       "ratingValue": 4
      "datePublished": "2023-11-06",
      "reviewBody": "La lectura se hace muy amena y el misterio se mantiene hasta el último capítulo del libro. Lectura
muy agradable."
    },
      "@type": "Review",
      "author": {
       "@type": "Person",
       "name": "María Jesús Albarés Carrasco"
      "reviewRating": {
       "@type": "Rating",
"ratingValue": 5
      "datePublished": "2023-10-22",
      "reviewBody": "Magnífico, te atrapa desde la primera página. No pude dejarlo hasta que lo terminé. Cada giro es
mejor que el anterior. Sin duda imprescindible leerlo."
   ]
```

SIW 3 | Página

Validación del etiquetado.



SIW 4 | Página

Demo de Dandelion.



¿Qué diferencias hay entre los datos estructurados que se ofrecen con JSON-LD a las máquinas—en realidad a Google y sus algoritmos—y el tipo de información obtenida mediante *entity linking*?

Principalmente, se podría decir que será más preciso para la clasificación el uso de datos estructurados, como es el caso de JSON-LD, ya que se está haciendo en base a una idea, en el caso del *entity linking*, busca a lo largo de un texto posibles entidades que tengan relación para tratar de categorizarlo, pudiendo dar caso a errores.

Mismamente, en el ejemplo de la izquierda, se puede ver como identifica el nombre del libro, algo que puede ser útil en un texto donde simplemente se haga mención, pero en este caso, que no deja de ser un texto enfocado por entero a dicho libro, es un tanto innecesario, por lo que realmente no es tan preciso como lo que sería si se hubiera realizado un buen etiquetado.

Por este motivo considero que el uso de datos estructurados como *JSON-LD* o RDF es preferible en general, mientras que herramientas como el *entity linking*,

puedan tener preferencia para comprobaciones en ciertas ocasiones.

¿Es viable etiquetar en todo o en parte el contenido de tu caso usando el vocabulario de Schema.org para anotar la información detectada mediante entity linking?

Considero que no, en este caso todo el contenido del texto plano se encuentra ya etiquetado en el fichero JSON-LD, y realmente, mediante *entity linking* solamente se ha detectado el propio libro a raíz del nombre (y bueno, el idioma de este, pero lo obvio para esto), por lo tanto creo que carecería de sentido etiquetar todo o parte del contenido para anotar dicha información, en base a que no deja de ser información ya contenida en el propio texto.

SIW 5 | Página