

Mejorando la visibilidad de sitios Web usando tecnología semántica

Lenin Montenegro¹, Verónica Ochoa¹, Mauricio Espinoza Mejía²

¹Maestría en Gestión Estratégica de TI, Facultad de Ingeniería, Universidad de Cuenca Av. 12 de abril s/n, Ciudadela Universitaria, Cuenca, Ecuador.

² Departamento de Ciencias de la Computación, Universidad de Cuenca, Av. 12 de abril s/n, Edificio Tecnológico de la Facultad de Ingeniería, Cuenca, Ecuador.

Autores para correspondencia: lenin.montenegro@ucuenca.ec, veronica.ochoa@ucuenca.edu.ec, mauricio.espinoza@ucuenca.edu.ec

Fecha de recepción: 21 de Septiembre de 2014 - Fecha de aceptación: 17 de Octubre de 2014

RESUMEN

La visibilidad de una página Web involucra el proceso de mejora de la posición del sitio en los resultados devueltos por motores de búsqueda como Google. Hay muchas empresas que compiten agresivamente para conseguir la primera posición en los motores de búsqueda más populares. Como regla general, los sitios que aparecen más arriba en los resultados suelen obtener más tráfico a sus páginas, y de esta forma, potencialmente más negocios. En este artículo se describe los principales modelos para enriquecer los resultados de las búsquedas con información tales como fechas o localidades; información de tipo clave-valor que permite al usuario interactuar con el contenido de una página Web directamente desde el sitio de resultados de la búsqueda. El aporte fundamental del artículo es mostrar la utilidad de diferentes formatos de marcado para enriquecer fragmentos de una página Web con el fin de ayudar a las empresas que están planeando implementar métodos de enriquecimiento semánticos en la estructuración de sus sitios Web.

Palabras clave: Web semántica, microformatos, visibilidad sitios Web.

ABSTRACT

The visibility of a website involves the process of improving the position of the site in the results returned by search engines such as Google. There are many companies competing aggressively for the top position in popular search engines. As a general rule, the sites that appear higher on the search results tend to get more traffic in Web pages, and thus potentially more business. This article describes the main models to enrich the search results with information such as dates or locations; key-value type information that allows the user to interact with the contents of a Web page directly from the search results. The main contribution of this paper is to show the usefulness of different markup formats to enrich fragments of a Web page in order to help companies that are planning to deploy semantic enrichment methods on the structuring of their websites.

Keywords: Semantic Web, micro-formats, website visibility.

1. INTRODUCCIÓN

Uno de los principales objetivos de cualquier propietario de un sitio Web es aumentar la visibilidad de su sitio en los motores de búsqueda. Algunos trabajos han demostrado que únicamente los primeros resultados obtenidos por los motores de búsqueda son usados por la mayoría de usuarios (King, 2008; Yu, 2011), de allí la importancia de mejorar la visibilidad de un sitio Web. La visibilidad Web puede ser definida como el grado de facilidad en que un usuario puede alcanzar a través de una referencia en línea el sitio Web de una empresa u organización (Smithson *et al.*, 2010).

Existen diferentes métodos para incrementar la visibilidad de los sitios Web en los motores de búsqueda, y la estructura del sitio es un componente fundamental para hacer que el sitio Web sea legible por una máquina (Graf *et al.*, 2007; Yu, 2011). Una forma de mejorar la estructura de un sitio Web y ofrecer datos más relevantes y convincentes en los resultados de búsqueda, es marcar semánticamente la información contenida en los documentos Web (Yu, 2011). En el proceso de marcado se utilizan etiquetas especiales para designar las secciones y el contenido de una página Web; *fragmentos semánticamente enriquecidos* son el resultado de este proceso. Esto permite que un *crawler* o araña¹ entienda el contenido de la página Web durante el proceso ejecutado por el motor de búsqueda. Con la ayuda del marcado semántico, los motores de búsqueda pueden determinar el tema y relevancia de las diferentes secciones de una página Web (Yu, 2011).

Actualmente, la Web Semántica (Berners-Lee *et al.*, 2001) ha alcanzado un alto grado de madurez en términos de definir estándares tales como Resource Description Framework (RDF)², Resource Description Framework in Attributes (RDFa)³ y microformatos⁴ que permiten a los desarrolladores incluir metadatos en las páginas Web. Otra de las opciones para etiquetar contenido y dar significado a los datos que conforman una página Web son los microdatos⁵.

En este artículo, se describen diferentes formatos de marcado que permiten generar fragmentos enriquecidos dentro de una página Web. Estos fragmentos permiten mejorar los resultados de búsqueda en la Web, generando resúmenes de búsquedas basados en datos estructurados o metadatos asociados a los documentos Web. Los metadatos incluyen información de tipo clave-valor e incluso objetos multimedia como imágenes y video.

El resto de este artículo tiene la siguiente estructura. La Sección 2, ofrece una definición de fragmentos enriquecidos y describe algunas motivaciones del uso de estas piezas de texto semánticamente enriquecidas para mejorar la visibilidad de una página Web. En la Sección 3, se describe brevemente algunos de los principales formatos de marcado usados para enriquecer fragmentos de un sitio Web. En la Sección 4 se exponen algunos criterios que permiten caracterizar los lenguajes de marcado semántico. Entonces se muestra la aplicación de uno de estos formatos de marcado en un caso de uso (Sección 5) antes de concluir en la Sección 6.

2. FRAGMENTOS ENRIQUECIDOS

Los fragmentos enriquecidos son entidades usadas en las páginas Web para describir un tipo concreto de información tal como: una persona, un producto, un evento o una empresa. Cada entidad define sus propias propiedades. Por ejemplo, un individuo puede definir las siguientes propiedades para describir la entidad persona: nombre, dirección, lugar de trabajo, cargo o dirección de correo electrónico. Los formatos enriquecidos representan una forma simple de agregar significado semántico a un contenido legible por un humano y entendible por una máquina. Básicamente, el proceso de enriquecimiento se ejecuta sobre ciertas porciones de código Hypertext Markup Language (HTML)⁶ o Extensible HyperText Markup Language (XHTML)⁷, en donde se inserta contenido semántico aprovechando las características de los atributos *id* o *class* usado por algunas etiquetas HTML.

En la literatura es posible categorizar diferentes formatos de marcado para la creación de fragmentos semánticamente enriquecidos, dependiendo del lenguaje de presentación de la información. Por ejemplo, los microdatos se implementan mediante HTML5⁸ y los microformatos mediante XHTML o HTML. RDFa, utiliza además un mecanismo de aplicación de metadatos similar

¹ Un crawler o araña web (o araña de la web) es un programa que inspecciona las páginas del World Wide Web

² <http://www.w3.org/RDF/>

³ <http://www.w3.org/TR/rdfa-syntax/>

⁴ <http://www.microformats.org>

⁵ <http://www.w3.org/TR/microdata/>

⁶ <http://www.w3.org/html/wg/>

⁷ <http://www.w3.org/TR/xhtml/>

⁸ <http://www.w3.org/TR/html5/>

a los microformatos es decir el proceso de enriquecimiento aprovecha los atributos de algunas etiquetas XHTML. En la Sección 3 se describirá con mayor detalle los formatos mencionados.

El objetivo principal de todos estos formatos de marcado es estandarizar la forma en la que se describen tanto propiedades como la propia sintaxis, todo este código no es visible para las personas sino que está hecho para los distintos robots de los motores de búsqueda como puede ser Google⁹. De hecho, existe mucha evidencia que el marcado semántico puede desempeñar un papel muy importante en el aumento del ranking de un sitio Web y mejorar su visibilidad en los motores de búsqueda (Bröcker & van Ahee, 2008).

Los autores de este artículo creen que mostrando los beneficios y barreras del marcado semántico, y mediante la presentación de ejemplos sobre cómo se puede implementar los fragmentos enriquecidos, más organizaciones pueden fomentar la implementación de esta tecnología en sus sitios Web.

3. FORMATOS DE MARCADO PARA LA CREACIÓN DE FRAGMENTOS ENRIQUECIDOS

Para aplicar fragmentos enriquecidos en un sitio Web, además de definir la temática del mismo y la información que se desea destacar, es necesario elegir un tipo de formato de marcado que permita interpretar adecuadamente los datos indexados por los motores de búsqueda. Google establece algunas directrices de calidad para el correcto uso del marcado, de forma que no sea contraproducente su uso (Google Inc., 2014a; Google Inc., 2014b). Algunas de las recomendaciones principalmente enfocadas al marcado, establecen evitar:

- Marcar contenido que no sea visible para los usuarios.
- Marcar contenido irrelevante o confuso, como comentarios falsos o contenido que no esté relacionado con el tema central de una página.

En las siguientes secciones se describe algunos de los formatos de marcado más relevantes en la actualidad. El método de selección de los formatos descritos en este trabajo no está basado en ningún proceso formal, únicamente se ha tomado en consideración los formatos más buscados en los motores de búsqueda como Google. Para cada uno de los formatos seleccionados se describe los principales atributos usados en el proceso de anotación y se presentan ejemplos con la idea de ilustrar el proceso de creación de fragmentos enriquecidos.

3.1. *Microformatos*

Los microformatos se definen como “una manera de añadir simples marcadores (en las páginas Web) a elementos de datos legibles por una persona tales como: eventos, información de contacto o localidades, tal que la información contenida en estos sitios se puede extraer, indexar, buscar o combinar vía software” (Microformats Wiki, 2013). Este formato de marcado ofrece una forma sencilla de aplicar metadatos, haciendo uso de algunos atributos del lenguaje (X)HTML (Meñdez *et al.*, 2007).

Las principales ventajas de los microformatos son la sencillez de su aplicación, la coherencia con las normas vigentes y el amplio soporte de los desarrolladores. El mayor inconveniente de los microformatos es que no tienen ontologías, modelos formales o esquemas. El vocabulario se puede comprobar de forma manual con las especificaciones, pero no hay manera de ejecutar una comprobación automática.

Modelo

El modelo de etiquetado reutiliza los siguientes atributos (X)HTML:

⁹ <https://www.google.com>

- *class*: indica el tipo de objeto sobre el que se realiza la descripción.
- *rel*: se utiliza en los hiperenlaces para expresar el tipo de relación que se establece entre el objeto y la URL de destino.
- *rev*: el tipo de relación en los hiperenlaces se define en sentido contrario a “*rel*”, desde la URL destino hacia el objeto.
- *title*: ofrece una alternativa entendible por una máquina a un texto etiquetado entendible por una persona (Pastor Sanchez, 2013).

Etiquetado

Los microformatos se clasifican en dos grupos: básicos y compuestos. Los primeros como: *geo* o *relicense*, están destinados a la descripción de cosas pequeñas y conceptos individuales. También se pueden utilizar para la creación de microformatos compuestos. Los microformatos compuestos tienen estructuras anidadas complejas. Ejemplos de microformatos compuestos son *hCalendar* y *hCard* que reflejan los datos de los estándares para el intercambio de información de calendarios, *iCalendar* y el estándar para el intercambio de información personal, *vCard*, respectivamente.

El vocabulario de microformatos es constantemente actualizado y en junio de 2014 se actualizó la definición de este lenguaje de etiquetado a la versión *microformats2*¹⁰. Un ejemplo usando este formato de etiquetado para anotar un evento es mostrado a continuación:

```
<div class="h-event">
  <h1 class="p-name">Reunión de Consejo</h1>
  <p>
    Desde:<time class="dt-start" datetime="2014-12-02 12:00">2 de Diciembre de 2014,
      12:00</time>
    Hasta :<time class="dt-end" datetime="2014-12-02 18:00">18:00</time>
    En:<span class="p-location">Salón de la Ciudad</span>
  </p>
  <p class="p-summary">Convocados para la aprobación de la ordenanza municipal
    emergente de tránsito.</p>
</div>
```

En el ejemplo la clase *h-event* representa el nombre de la clase raíz que indica la presencia de un evento definido mediante el estándar *hCalendar*. Las clases *p-name*, *dt-start*, *dt-end*, *p-location* y *p-summary* definen algunas de las propiedades del evento a etiquetar.

3.2. RDFa

Es el lenguaje de etiquetado estandarizado por el World Wide Web Consortium (W3C) que permite introducir semántica en las páginas Web. Las descripciones de recursos se realizan mediante tripletas RDF, formadas por: sujeto, predicado y objeto. La entidad que se describe es el sujeto y puede clasificarse con taxonomías de clases y subclases. El predicado representa las propiedades del sujeto o relaciones con otros recursos. El objeto es el valor de la propiedad o recurso con el que se relaciona al sujeto.

La ventaja fundamental de RDFa es que permite combinar, mezclar, y extender vocabularios existentes con facilidad. Además, es posible utilizar diferentes espacios de nombres en un documento anotado con este formato, por ejemplo “*dc:*” para el vocabulario “Dublin Core”¹¹ y “*cc:*” para identificar el vocabulario “Creative Commons”¹².

¹⁰ <http://microformats.org/wiki/microformats-2>

¹¹ <http://dublincore.org/>

¹² <https://wiki.creativecommons.org/>

Modelo

Los atributos que se utilizan son:

- *about*: identifica el recurso al que se refiere la descripción del marcado semántico, “sujeto”.
- *datatype*: define el tipo de datos utilizado por un literal.
- *inlist*: permite agrupar varios “objetos” a un mismo “sujeto” a través de un único atributo *rel* o *property*.
- *prefix*: define equivalencias entre prefijos y espacios de nombres.
- *property*: expresa relaciones entre el recurso “sujeto” y cualquiera de los recursos “objeto”.
- *resource*: permite expresar el recurso “objeto” de una relación pero sin utilizar enlaces navegables.
- *typeof*: indica el tipo de clase del recurso “sujeto”.
- *vocab*: permite hacer referencia a vocabularios para utilizar los elementos descriptivos de un modo sencillo para el marcado semántico (Pastor Sanchez, 2013).

Etiquetado

Es posible aplicar RDFa sólo para XHTML versión 2, no para HTML. La versión XHTML 1.1 también puede ser utilizado en la práctica con RDFa, pero con limitaciones considerables (Birbeck, 2009).

```

<div vocab="http://purl.org/dc/terms/">
  <div resource="/alice/posts/trouble_with_bob">
    <h2 property="title">The trouble with Bob</h2>
    <h3 property="creator" resource="#me">Alice</h3>
    ...
  </div>
</div>
...
<div class="sidebar" vocab="http://xmlns.com/foaf/0.1/" resource="#me" typeof="
  Person">
  <p>
    <span property="name">Alice Birpemswick</span>,
    Email: <a property="mbox" href="mailto:alice@example.com">alice@example.com</a
    >,
    Phone: <a property="phone" href="tel:+1-617-555-7332">+1 617.555.7332</a>
  </p>
  ...
</div>

```

En el ejemplo mostrado previamente, la clase *vocab* dentro del primer elemento *<div>* permite identificar el vocabulario usado para anotar el título (*title*) y creador (*creator*) de un documento. Las clases *resource* y *typeof* descritos como atributos dentro del segundo elemento *<div>*, identifican un recurso de tipo persona (*person*) que tiene como propiedades un correo electrónico (*mbox*) y un teléfono (*phone*). Todas estas últimas anotaciones están basadas en el vocabulario FOAF¹³, el cual es usado para describir información sobre personas, los vínculos entre éstas y las cosas que crean y hacen.

¹³ <http://xmlns.com/foaf/spec/>

3.3. Microdatos

Por su naturaleza, los microdatos pueden considerarse como un formato de marcado híbrido a partir de los microformatos, RDFa y HTML5. La especificación HTML5 sobre microdatos es una forma de marcar semánticamente contenido legible tanto para humanos como para máquinas a partir de elementos HTML utilizados en una página Web.

Modelo

De acuerdo a la especificación de microdatos (W3C, 2013), este formato de marcado consiste en grupos de pares nombre-valor. Estos grupos se denominan ítems y cada par nombre-valor es una propiedad. Los microdatos definen cinco atributos en HTML5:

- *itemscope*: delimita un ítem de información del contenido web.
- *itemtype*: indica el tipo de ítem sobre el que se realizará el marcado semántico.
- *itemid*: asocia un identificador al ítem descrito.
- *itemprop*: especifica el elemento de un vocabulario que se refiere a una propiedad o atributo del contenido web que se está marcando.
- *itemref*: permite referirse a ítems definidos en cualquier lugar del documento. Esto permite realizar asignaciones de propiedades a dichos ítems desde otros diferentes (Pastor Sanchez, 2013).

Etiquetado

Los microdatos usan simples atributos en etiquetas HTML, generalmente `` o `<div>` para describir a las entidades y sus propiedades. El ejemplo que se ilustra a continuación tiene como intención anotar los datos de una película.

Título: Europa Report
 Dirección: Sebastián Cordero
 Reparto: Christian Camargo, Anamaria Marinca, Charlto Copley, Daniel Wu, Karolina Wydra
 Año: 2013

Para anotar todo un bloque de información (como el mostrado en el HTML previo) el atributo *itemscope* es utilizado. La información sobre la película identifica cuatro propiedades: *título*, *director*, *reparto* y *año*; para las cuales se utiliza el atributo *itemprop* como mecanismo de etiquetado:

```
<div itemscope>
Título: <span itemprop="titulo">Europa Report</span>
Dirección: <span itemprop="director">Sebastián Cordero</span>
Reparto: <span itemprop="reparto">Christian Camargo, Anamaria Marinca, Charlto
    Copley, Daniel Wu, Karolina Wydra</span>
Año: <span itemprop="año">2013</span>
</div>
```

Para que las etiquetas anotadas tengan sentido, éstas deben estar basadas en un vocabulario estandarizado en el cual se explique y defina que la propiedad “título” por ejemplo representa el título de una película. Varios esfuerzos han sido ejecutados por la comunidad a efectos de poder incluir información sobre los metadatos. Una vez que se comprobó que no era posible incorporar RDFa en HTML5 (W3C, 2009), el editor de la especificación HTML. Hickson (2009) presentó una primera propuesta para microdatos en mayo 2009. Google, Yahoo!, Bing y Yandex presentaron como otra opción el proyecto Shema.org a mediados del 2011 (Ronallo, 2012).

El código mostrado a continuación muestra el proceso de etiquetado del director de la película usada como ejemplo, definiendo que *Sebastián Cordero* es un objeto de tipo persona usando el vocabulario *schema.org*:

```
<div itemscope itemtype="http://schema.org/Movie">
Título: <span itemprop="name">Europa Report</span>
Dirección: <span itemprop="director" itemscope itemtype="http://schema.org/Person"><span itemprop="name">Sebastián Cordero</span></span>
Reparto: <span itemprop="actor"> Christian Camargo , AnamariaMarinca , Charlto Copley , Daniel Wu, Karolina Wydra</span>
Año: <span itemprop="años">2013</span>
</div>
```

Existen situaciones en las que es necesario proporcionar información semántica a los motores de búsqueda pero no necesariamente a los usuarios que acceden a la Web, para este caso, este formato de etiquetado recomienda el uso de la etiqueta *meta*. En el ejemplo que sigue, la etiqueta *meta* es utilizada para ofrecer información sobre la calificación de la película en cuestión:

```
<div itemprop="rating" itemscope itemtype="http://schema.org/Rating">
Puntuación: <span itemprop="ratingValue">8,5</span>
<meta itemprop="bestRating" content="10" />
</div>
```

4. ANALISIS COMPARATIVO DE LOS FORMATOS DE MARCADO

Los formatos de marcado descritos en la sección previa ofrecen un mecanismo para estructurar e introducir semántica en los documentos publicados en la Web. Efectuando un análisis muy general, la selección del lenguaje de marcado puede depender de los tipos de datos.

Tabla 1. Comparación entre los lenguajes de marcado: Microdatos, Microformatos y RDFa.

	Microdatos	Microformatos	RDFa
Modelo de Datos	Árbol	Árbol	Grafo
Complejidad del dominio de la sintaxis	Medio	Bajo	Alto
Transformación a JSON	Si	Si	Si
Interpretación de la página de resultados del buscador	Si	Si	Si
Encadenamiento de ítems	Si (básico)	No	Si
Transclusión	Si	Si	Si
	(copia y referencia)	(copia y referencia)	(copia)
URIs compactas	No	No	Si
Revinculación con prefijos	No	No	Si
Mezcla de vocabularios	No	No	Si
Amplitud del vocabulario	Amplia	Reducida	Sin restricción

Los microformatos, por su baja *complejidad de implementación y dominio de la sintaxis* no permiten la flexibilidad y amplitud requerida por la mayoría de sitios Webs, limitándose a sitios con páginas únicas, poca información por etiquetar o con contenido de poca relevancia. Sin embargo, vale

la pena destacar que algunas de las etiquetas usadas en este formato de anotación son aceptadas por Google¹⁴, Yahoo¹⁵, Yandex¹⁶ y Microsoft¹⁷ como soporte para visualización en las páginas de resultados de búsqueda.

Respecto al *posicionamiento en las páginas de resultados del buscador*, los microdatos tienen cierta ventaja sobre los otros formatos debido al uso de un esquema de marcado de datos estructurado basado en el vocabulario *schema.org*¹⁸ que es compatible con los principales motores de búsqueda. Este vocabulario también soporta el formato de marcado con RDFa. Los microdatos con el uso de *schema.org* proporcionan un vocabulario mucho más extenso y detallado que los microformatos.

Todos los formatos de marcado soportan la *transformación a JSON*, lo cual representa una ventaja al momento de implementar o modificar el código fuente de las páginas Web a ser anotadas, ya que JSON permite agregar semántica de un modo no invasivo al código HTML. En el sitio de JSON-LD¹⁹ es posible encontrar ejemplos de implementación con RDFa, microformatos y microdatos. El proceso de anotación con JSON hace uso de la etiqueta *script*, la cual se puede incluir al inicio de la página Web a ser anotada. Aunque JSON dista de ser un formato para generar fragmentos enriquecidos, sin embargo también es aceptado e interpretado por los motores de búsqueda, de esta manera se independiza la estructura de etiquetado y se mantiene la jerarquía de anidación entre ítems.

Para agregar contenido semántico en un sitio Web sin tener que preocuparse de la *extensión y mezcla de vocabularios* la opción recomendada es RDFa. Con este formato es posible crear un vocabulario propio o hacer uso de vocabularios existentes en la Web tales como DB-pedia²⁰ o Freebase²¹ para ejecutar el proceso de anotación. Con RDFa también es posible mezclar diferentes vocabularios, lo que permite declarar distintos tipos de contenido en una misma página Web, además de que se puede *encadenar ítems para declaraciones en distintas secciones del sitio* y así no afectar o depender del diseño de la página. En la declaración de vocabularios RDFa soporta la *utilización de prefijos* para reemplazar dominios o URI demasiado largas y trabajar con *URIS compactas*. Hay que tener en cuenta que RDFa se utiliza fundamentalmente para agregar semántica a los sitios antes que enriquecer fragmentos de contenido, por lo que no se recomienda para el uso de etiquetado del código HTML para posicionamiento en las páginas de resultados de búsqueda, sino más bien para el entendimiento y comprensión máquina a máquina (M2M).

Todos los formatos analizados permiten ejecutar *transclusión*, esto significa que es posible especificar un conjunto de propiedades una vez en una página, por ejemplo, una dirección postal, y copiar esas propiedades a múltiples elementos de la página. RDFa permite hacer esto por referencia, no haciendo una copia. Ambos Microdatos y Microformatos permiten transclusión tanto por referencia y por copia.

Finalmente, es necesario destacar que las comprobaciones efectuadas con la herramienta de pruebas de datos estructurados de Google²² muestra que no existe variación alguna en los resultados de búsqueda obtenidos con los distintos formatos de marcado: microformatos, microdatos, RDFa o microdatos con JSON-LD. Si el contenido enriquecido es el mismo, la visualización en la página de resultados del buscador se presenta de igual manera.

¹⁴ <https://support.google.com/webmasters/answer/146897>

¹⁵ <https://developer.yahoo.com/searchmonkey/siteowner.html>

¹⁶ <http://help.yandex.ru/webmaster/microformats/what-is-microformat.xml>

¹⁷ <http://www.bing.com/webmaster/help/markin-up-your-site-with-structured-data-3a93e731>

¹⁸ <http://schema.org/>

¹⁹ JSON para Linked-Data - <http://www.w3.org/TR/json-ld/>

²⁰ <http://dbpedia.org/About>

²¹ <http://www.freebase.com/>

²² <http://www.google.com/webmasters/tools/richsnippets>

5. CASO DE ESTUDIO

En esta sección se describe las acciones ejecutadas para anotar una página Web usando los formatos de etiquetado descritos en la sección previa. El objetivo de este ejemplo es mostrar paso a paso las actividades que deben ser ejecutadas por un editor de páginas Web para enriquecer fragmentos de código HTML con información semántica.

5.1. Creación del sitio web

Se creó una página Web acerca de un evento anual local llamado “Tripy Tripy” a realizarse a principios de noviembre con motivo de las fiestas de independencia de Cuenca. En un principio, sin ningún tipo de marcado, se envió el sitio al buscador Google²³ para que lo indexe. El resultado del proceso de indexación fue comprobado ejecutando una búsqueda sobre el tema relacionado a la página (ver Fig. 1).



Figura 1. Página indexada sin microformatos.

Una parte del código HTML usado para describir la página Web es mostrado a continuación. Como se puede observar no existe aún ninguna etiqueta ajena al lenguaje HTML que describa un proceso de etiquetado semántico de los datos.

```
<a href="http://cloud.dixbit.com/tripytripy/halloween">

</a>
<header><h2>TRIPY TRIPY Halloween</h2></header>
<p>Fiesta de Disfraces. Para aquellos diferentes, de espíritu rebelde. Invitados especiales: Maayan Nidam, Tone of Arc(live), Pancho y José, Pedro Aguirre.</p>
<p>Cuenca, viernes 31 de octubre en Turi, frente a la iglesia. $20-$25</p>
```

5.2. Marcado de datos estructurados

Lo primero que se pudo comprobar una vez ejecutado el proceso de etiquetado, es que de cara al usuario final la apariencia de la página no se ve modificada en absoluto. Para definir información sobre un evento dentro de la página, se utilizó la propiedad *itemprop* con el valor *url* y como contenido el valor de la etiqueta *<a href>*. Con el fin de enlazar la propiedad *url* con otro evento declarado en otra sección de la página Web, se utilizó un identificador (*id*). Para asignar una imagen al evento se utilizó nuevamente la propiedad *itemprop* dentro de la etiqueta *img*.

```
<a id="url-hall" itemprop="url" href="http://cloud.dixbit.com/tripytripy/halloween">
</a>
```

Para definir el evento y sus propiedades se crearon secciones anidadas con las etiquetas **, usando la propiedad *itemtype* y asignado como valor el urú del concepto evento dentro del vocabulario

²³ <https://www.google.com/webmasters/tools/submit-url>

(<http://schema.org/Event>). Para hacer referencia a otros contenidos que pertenecen al mismo evento se utilizaron los alias creados en otras secciones de la página Web. Para este caso se usó la propiedad *itemref*. Finalmente, se agregaron las propiedades del evento: nombre (*name*), descripción (*description*) y ejecutante (*performer*). Para la propiedad *performer* fue necesario además anidar el vocabulario *Person* para indicar el nombre (*name*) del ejecutante del evento.

```
<span itemscope itemtype="http://schema.org/Event" itemref="url-hall _fecha-hall _lugar-hall _precios-hall">
<header<h2 itemprop="name">TRIPY TRIPY Halloween</h2</header>
<p<span itemprop="description">Fiesta de Disfraces. Para aquellos diferentes, de
espíritu rebelde</span>.
Invitados especiales:
<span itemprop="performer" itemscope itemtype="http://schema.org/Person"> <span
itemprop="name">Maayan Nidam</span></span>,
<span itemprop="performer" itemscope itemtype="http://schema.org/Person"> <span
itemprop="name">Tone of Arc</span> </span>(live),
<span itemprop="performer" itemscope itemtype="http://schema.org/Person"> <span
itemprop="name">Pancho y José</span></span>,
<span itemprop="performer" itemscope itemtype="http://schema.org/Person"> <span
itemprop="name">Pedro Aguirre</span></span>.
</p></span>
```

Para ofrecer una información más completa y exacta sobre la dirección del evento, fue necesario utilizar el concepto *PostalAddress* con las propiedades localidad (*addressLocality*), calle (*streetAddress*), región (*addressRegion*) y país (*addressCountry*). En este caso, el concepto y sus propiedades son parte del vocabulario *schema.org*.

```
<span id="direccion-hall" itemprop="address" itemscope itemtype="http://schema.org
/PostalAddress">
<span itemprop="addressLocality">Cuenca</span>
<meta itemprop="streetAddress" content="Frente a la Iglesia de Turi">
<meta itemprop="addressRegion" content="Azuay">
<meta itemprop="addressCountry" content="Ecuador"></span>
```

Para definir la fecha del evento se utilizó el formato ISO 8601 (año, mes, día) independiente de la información almacenada en el contenido de la página Web (en el caso del ejemplo, 31 de octubre).

```
<span id="fecha-hall" itemprop="startDate" content="2014-10-31">31 de octubre</
span>
```

Para definir el lugar del evento se utilizó el concepto *Place* perteneciente al vocabulario *schema.org*. La propiedad *name* fue usada para identificar el sitio en donde se desarrollará el evento.

```
<span id="lugar-hall" itemprop="location" itemscope itemtype="http://schema.org/
Place" itemref="direccion-hall"><span itemprop="name">Turi</span></span>
```

Finalmente, para anotar el precio del evento utilizó nuevamente el vocabulario *schema.org*. El concepto *Offer* con la propiedad *price* fueron utilizados para describir el importe a cubrir por los asistentes al evento.

```
<span id=" precios-hall" itemprop=" offers" itemscope itemtype=" http://schema.org /
Offer">
<span itemprop=" price">$20-$25</span></span>
```

5.3. Comprobación de la visibilidad de la página

Una vez finalizado el proceso de enriquecimiento semántico, se volvió a ejecutar el proceso de búsqueda. Se comprobó que esta vez el motor de búsqueda indexa y muestra de mejor manera la información de los tres eventos del caso de uso, incluso la información sobre la fecha del evento es localizado²⁴ usando el formato o idioma de entrada. La Fig. 2 muestra el resultado del proceso de indexación de la página enriquecida semánticamente.

6. CONCLUSIÓN Y TRABAJO FUTURO

En la sociedad en donde la información digital está en crecimiento constante, la innovación en el contenido de páginas Web debe ser prioritario en las entidades o empresas que requieran posicionarse en los motores de búsqueda más utilizados como Google para su comercialización en distintos países.



Figura 2. Página indexada con microformatos.

En este artículo se describieron los principales formatos de etiquetado para enriquecer los resultados de las búsquedas. Uno de los resultados que se pudo observar con la implementación del caso de uso fue que la aplicación de fragmentos enriquecidos mejora el posicionamiento en los resultados del motor de búsqueda. En general, la inclusión de fragmentos enriquecidos tiene impacto sobre los resultados, ubicando a la página Web entre los primeros lugares de la lista dependiendo de la consulta realizada.

Es necesario hacer notar que la implementación de fragmentos enriquecidos en una página Web requiere i) gran capacidad de análisis de información, ii) una definición adecuada del contenido a ser anotado, iii) la selección apropiada del vocabulario que defina los conceptos requeridos, y iv) el uso correcto del esquema o vocabulario seleccionado. Es opinión de los autores que la inclusión de estos fragmentos enriquecidos, permite optimizar el tiempo de navegación de un usuario, pues la información de mayor relevancia puede ser obtenida en una sola consulta.

Como trabajo futuro se tiene planificado valorar la efectividad del enriquecimiento semántico en el contexto de las preferencias de búsqueda de un usuario. La idea es comparar el grado de preferencia de los resultados obtenidos con una búsqueda tradicional frente a una búsqueda que ofrece resúmenes enriquecidos con información semántica.

²⁴ La localización de productos o adaptación de productos a las necesidades de un mercado objetivo contempla generalmente la traducción de software.

REFERENCIAS

- Berners-Lee, T., J., Hendler, O. Lassila, 2001. The semantic web. *Scientific American*, 284(5), 34-43.
- Birbeck, M., 2009. Xforms and internet applications: Using rdfa in xhtml 1. Descargado de <http://internet-apps.blogspot.com/2007/02/using-rdfa-in-xhtml-1.html> el 26 de octubre 2014.
- Bröcker, J., G. van Ahee, 2008. Semantics & search engine optimisation. *Search*, 15 pp.
- Google Inc., 2014a. Directrices de los fragmentos enriquecidos - ayuda de herramientas para webmasters de google. Descargado de <https://support.google.com/webmasters/answer/2722261> el 26 de octubre 2014.
- Google Inc., 2014b. Directrices para webmasters - ayuda de herramientas para webmasters de google. Descargado de <https://support.google.com/webmasters/answer/35769> el 26 de octubre 2014.
- Graf, A., D. Innsbruck, A. Graf, 2007. Rdfa vs. microformats a comparison of inline metadata formats in (x)html. Disponible en <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.86.8800>.
- Hickson, I., 2009. Annotating structured data that html has no semantics for. Descargado de <http://lists.whatwg.org/htdig.cgi/whatwg-whatwg.org/2009-May/019681.html> el 26 de octubre 2014.
- King, A.B., 2008. *Website optimization: speed, search engine & conversion rate secrets*. Disponible en <http://www.ebooksdownloadfree.com/SEO/Website-Optimization-Speed-Search-Engine--Conversion-Rate-Secrets-B11717.html>.
- Méndez, E., A. Bravo, L.M. López, 2007. Microformatos: web 2.0 para el Dublin Core. *El profesional de la información*, 16(2), 107-113.
- Microformats Wiki, 2013. Introduction to microformats. Descargado de <http://microformats.org/wiki/introduction> el 26 de octubre 2014.
- Pastor Sanchez, J.A., 2013. Mercado semántico: tecnologías y aplicación para la representación de sistemas de organización del conocimiento en el contexto Linked Open Data. *Scire: Representación y Organización del Conocimiento*, 19(2), 55-68.
- Ronallo, J., 2012. The code4lib journal - html5 microdata and schema.org. Descargado de <http://journal.code4lib.org/articles/6400> el 26 de octubre 2014.
- Smithson, S., C.A. Devece, R. Lapidra, 2010. Online visibility as a source of competitive advantage for small-and medium-sized tourism accommodation enterprises. *The Service Industries Journal*, 31(10), 1573-1587.
- W3C, 2009. public-rdf-in-xhtml-tf. Descargado de <http://lists.w3.org/Archives/Public/public-rdf-in-xhtml-tf/2009Aug/thread.html> el 26 de octubre 2014.
- W3C, 2013. Html microdata. Descargado de <http://www.w3.org/TR/microdata> el 26 de octubre 2014.
- Yu, L., 2011. *A developer's guide to the Semantic Web*. Springer.