
Estudio del impacto del SEO semántico en los motores de búsqueda: aplicando microdatos y RDFa Lite 1.1 en el ámbito de *Schema.org*

Analysis of the impact of semantic S.E.O on search engines: applying microdata and RDFa lite 1.1 in the domain of Schema.org

Samantha Bernardis

Universidad de Murcia, Campus Espinardo, Murcia, samantha.bernardis@um.es

Resumen

Este trabajo se centra en el estudio de *Schema.org*, uno de los ejemplos más llamativos de SEO Semántico. Se trata de un vocabulario común creado por Google, Yahoo y Bing, que usa los microdatos como única sintaxis de marcado semántico recomendada y que favorece la visibilidad de un sitio web y la encontrabilidad de su contenido. Nuestro interés no es solo evaluar la eficacia de *Schema.org*, sino también averiguar si funcionaría también con otra sintaxis de marcado semántico, RDFa Lite 1,1. Además, queremos analizar si la actualización de los contenidos puede favorecer el posicionamiento de sitios web enriquecidos con datos estructurados. Tras evaluar si RDFa Lite 1,1 posee las características para ser aplicada a *Schema.org*, se han instalado tres sistemas web: uno optimizado solo a nivel SEO, otro que incluye *Schema.org* junto con los microdatos y un tercero que usa RDFa Lite 1.1 en lugar de los microdatos, con el fin de averiguar cual de las tres se posiciona mejor. Tras lanzar tres tomas de datos en tres periodos distintos se observa que la web optimizada sólo SEO y la con los microdatos resultan bien posicionadas, y que el mayor beneficio que aporta el uso de *Schema.org* es que, al actualizar los contenidos, favorece el rápido posicionamiento de sitios web enriquecidos con los microdatos, produciendo buenos resultados ya al cabo de una semana.

Palabras clave:

Schema.org, Microdatos, RDFa Lite 1,1, SEO, Posicionamiento Web, Google

Abstract

This paper focuses on *Schema.org*, one of the most remarkable examples of Semantic SEO. That is a share vocabulary created by Google, Yahoo and Bing, which use Microdata as the only recommended **semantic markup format** and enable the visibility of a website and the searchability of its content. Our purpose is not just to evaluate the efficiency of *Schema.org*, but also to check if *Schema.org* can work with RDFa Lite 1.1. Furthermore, we want to analyse if the updating of the contents enables the positioning of aa website enriched with structured data. After evaluating if RDFa Lite 1,1 has got the characteristics for being applied to *Schema.org*, we implemented three websites: one optimized just with traditional SEO, another one enriched with *Schema.org* and Microdata, and the third one with RDFa Lite 1.1 instead of Microdata. After launching three searchings in three different periods of time, we observe that the links from the website optimized just with traditional SEO and that one enriched with Microdata appear on the first page of results. Furthermore, the best benefit of *Schema.org* is that enable the fast positioning of the websites enriched with Microdata, producing good results after one week.

Keywords

Schema.org, Microdata, RDFa Lite 1,1, Search Engine Optimization, Google.

1. Introducción.

En el momento actual, la mejora de las herramientas de búsqueda en la web ocupa un papel fundamental para facilitar la recuperación de los contenidos digitales y aumentar la visibilidad de los sitios web. Sin embargo, la mejora de los resultados de búsqueda resulta una tarea compleja a desarrollar.

En efecto, la presencia de contenido duplicado en diferentes páginas y polisémico, puede generar ambigüedades durante la recuperación, y además dicho contenido no es entendido automáticamente por los motores de búsqueda. Para solucionar estos problemas, en los últimos años, se han desarrollado nuevas tecnologías y proyectos para facilitar la recuperación de contenido "semánticamente" relevante, fiable y de calidad.

Una prueba de ello es el *Schema.org*, desarrollado conjuntamente por los principales motores de búsqueda de la Web, *GOOGLE*, *YAHOO*, *BING* y el buscador ruso *Yandex* (García-Marco, 2012). El uso de vocabularios de descripción y la introducción del marcado semántico en el código HTML ha modificado la manera de entender el posicionamiento en los motores de búsqueda y ha aportado un nuevo concepto: el SEO semántico.

La tarea principal de un motor de búsqueda es básicamente la realización de dos funciones: la indexación y la recuperación de información (Gil Leiva, 2008).

La indexación permite construir las estructuras que facilitaran luego la recuperación de información, es decir, identificar los términos o *las palabras clave* que mejor representan el contenido indexado. Los resultados son ordenados por relevancia basándose en el grado de semejanza de cada uno de los documentos disponibles con la consulta hecha por el usuario.

Los motores de búsqueda web utilizan programas, denominados *crawlers*, que identifican y recopilan documentos para el motor de búsqueda, rastreando toda la web en busca de nueva información.

Los motores de búsqueda aplican un algoritmo para calcular la relevancia de los documentos recuperados. En el caso de Google (el buscador utilizado en este trabajo) se utiliza el algoritmo denominado *Pagerank*, (Page, Brin, Motwani, y Winograd, 1999) basado principalmente en la topología de los hiperenlaces de los documen-

tos web recuperados, así como otros factores que permiten optimizar los resultados de una búsqueda.

Sin embargo, los resultados obtenidos durante la búsqueda se calculan a partir de un modelo de recuperación basado en el principio de composicionalidad del contexto enunciado por Götlob Frege (1) por el que podemos entender tanto un documento como una consulta con un conjunto de términos. Es decir, en los modelos basados en este enfoque, prima la sintaxis y las combinaciones específicas de términos ("playa San Juan").

Una de las limitaciones de este modelo es que no considera las relaciones semánticas entre los términos. Por lo tanto, es posible que durante una búsqueda no se recuperen documentos relevantes cuyo contenido es semánticamente cercano al de la consulta expresada por el usuario. En el caso de los motores de búsqueda, esto significa que pueden quedar fuera de los resultados páginas que traten el tema de los términos de la consulta y que no tengan una combinación de palabras específica.

Para solucionar este problema se diseñan ontologías para que las aplicaciones informáticas puedan entender tanto el contenido semántico de los términos como las relaciones semánticas entre ellos. La definición más conocida de ontología es la de Gruber que fue extendida por Studer (1998) según la cual se entiende con ontología como "una especificación explícita y formal sobre una conceptualización compartida". Las ontologías ayudan a definir los términos que se usan para describir formalmente y representar un cierto dominio y que se relacionan entre ellos.

El buscador entenderá las relaciones entre las palabras, devolviendo un resultado relevante, como se puede ver en la Figura 1. Un resultado que en este caso está conectado con el relacionar semánticamente la película *Avatar*, que es un producto artístico y específicamente una película, con su director, que es una persona, y con un recurso online donde encontrar informaciones sobre ella.

Este resultado se realiza incluyendo en el código HTML metadatos, legibles por máquinas, que describen y organizan la información contenida en el documento a través de una serie de etiquetas. El problema es que hacen falta muchísimas ontologías para representar todos los dominios existen o una "mega-ontología" que abarca el mayor número de dominios posibles.

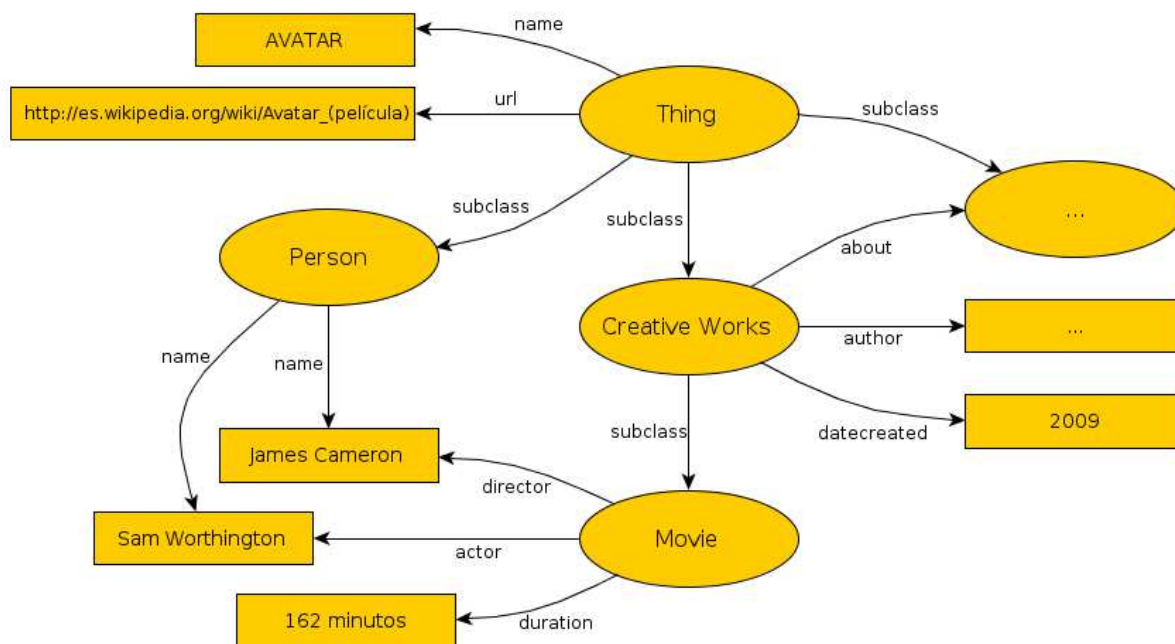


Figura 1: Ejemplo de Ontología elaborada a partir de Schema.org y relación semántica entre términos.
Fuente: elaboración propia

Existen algunas ontologías que son bastante difusas, como FOAF para describir personas y las relaciones entre ellas, o el vocabulario *Goodrelations* que se está difundiendo como un estándar para el comercio electrónico.

Por cuanto concierne el aplicación de técnicas SEO (2), esas pueden marcar las partes más relevantes de un documento HTML usando al-

gunas etiquetas, como <title>, <h1> y muchas otras que ayudan los crawlers a entender cuál es el contenido de un sitio y cuáles son las palabras clave más relevantes que según el creador del sitio web ayudarían a recuperar esta página web cuando se efectúa una petición. Aplicando adecuadamente estas técnicas se pueden conseguir que una pagina aparezca en las primeras posiciones en los listados de los resultados de los motores de búsqueda al ejecutar una determinada búsqueda (Vasta, 2010).

El límite del SEO tradicional es que permite recuperar resultados basados en una búsqueda sintáctica (recuperación de una combinación de palabras). El SEO semántico permite crear asociaciones entre palabras que se pueden conectar semánticamente entre sí, a través de sistemas de búsqueda semántica.

La búsqueda semántica se puede realizar mediante el uso de metadatos (a través de una representación), los cuales permiten especificar elementos personalizados mediante la sintaxis compuesta de pares *nombre-valor* con el contenido existente. Esto conlleva la posibilidad de extraer los conceptos más relevantes de una frase, y luego crear una consulta que se realiza contra la ontología del sistema (Méndez, 2007), devolviendo finalmente los resultados al usuario.

Schema.org aplica el mismo mecanismo, basándose en la utilización de metadatos de un vocabulario compartido que describe detalladamente y de forma estructurada el contenido de

la página, y los cuales han de ser incluidos en el código HTML de las páginas web. Para anotar la información en los documentos HTML *Schema.org* utiliza los microdatos, una sintaxis semántica que permite añadir información entendible por las máquinas.

Según los motores de búsqueda, el uso de un vocabulario de marcado compartido facilitaría, en primer lugar, la recuperación de información y datos por parte de aplicaciones informáticas, puesto que se trata de generar contenidos semánticamente relevantes y estructurados (Schema.org, 2012b), y en segundo lugar, permitiría la unificación de la Web, proporcionando un único sistema compartido de criterios, esquemas y ayudas para que los desarrolladores pueden diseñar páginas web enriquecidas con datos estructurados. Uno de los beneficios más importantes aportado por este proyecto es, pues, la mejora de la indexación para facilitar el posicionamiento de páginas web, SEO, puesto que proporcionando un sistema único y compartido de marcado conlleva una mayor visibilidad de las páginas web (Bradley, 2011).

Su objetivo es, pues, ofrecer un mecanismo común de estructuración de las páginas web, de manera que la recuperación de la información en la web es más relevante y exhaustiva. Algunos autores dan una definición más amplia de *Schema.org* y la llaman:

mega-ontología en forma de esquema de carácter general que intenta representar una amplia variedad de hechos: obras, eventos, intangibles, organizaciones, personas, lugares y productos (Pástor-Sánchez, 2012)

En este caso se define *Schema.org* como una pseudo mega-ontología, porque efectivamente *Schema.org* no puede ser ceñido entre la restrictiva etiqueta de vocabulario, como algunos autores hacen (Sánchez-Cuadrado, Colmenero-Ruiz y Moreira, 2012), puesto que hace referencia a un riguroso y exhaustivo esquema conceptual. Aunque no es una “verdadera” ontología, puesto que no permite hacer inferencia entre los hechos, es decir los objetos representados.

A pesar de eso, hace falta resaltar algunos problemas que surgen con el planteamiento de este nuevo estándar. En primer lugar, aún no sabemos si el añadido de marcado semántico al código HTML puede efectivamente mejorar la indexación y el posicionamiento de páginas web. Los buscadores aconsejan la utilización de marcado semántico, pero efectivamente hace falta

averiguar con casos prácticos si se producen las ventajas declaradas.

Por otro lado, la aplicación de *Schema.org* puede conducir a “una situación de divergencia tecnológica, debido al uso de microdatos en vez de la aplicación de RDFa” (Saorín y Pástor-Sánchez, 2012).

Al afirmar esto, Saorín y Pástor-Sánchez (2012) parten desde el punto de vista que esta divergencia tecnológica puede reducir la capacidad de interoperabilidad entre las dos tecnologías, puesto que las páginas web diseñadas con marcado XHTML no pueden funcionar con los microdatos, mientras que las páginas que incorporan el marcado HTML5 no contemplaban, hasta hace unos meses, el uso de RDFa (3). Además, la comunidad de los linked data está impulsando el uso de RDFa Lite 1.1 en lugar de los microdatos a través del portal <http://schema.rdfs.org/> proporcionando ejemplos para aplicar RDFa Lite 1.1 al *Schema.org*

Aunque el proyecto *Schema.org* fue realizado juntamente por los tres principales motores de búsqueda (Google, Yahoo y Bing), el motor de búsqueda elegido para este estudio es Google, primero porque proporciona consejos para la mejora del posicionamiento de una web a través de una guía sobre optimización en los motores de búsqueda, secundariamente porque en el año 2012 la cuota de mercado de Google (ver figura 2) era la más alta entre los motores de búsqueda (Martínez-Méndez, 2013). Se ha ceñido el experimento a este motor de búsqueda además porque le da más importancia al contenido generado rápidamente y aquel generado en las redes sociales, sobre todo en *Google+*. Esta última red social, al usar datos estructurados, permite enlazar la identidad del propietario de la cuenta a una página web que incluye marcado semántico.

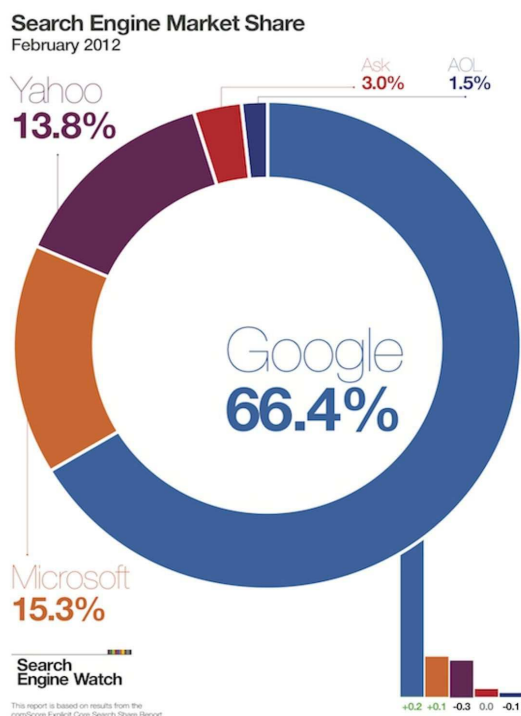


Figura 2: Cuota de mercado de motores de búsqueda a nivel mundial (Martínez-Méndez, 2013). Fecha de consulta: 17 Junio 2013

1.1. Objetivos del estudio.

En el apartado anterior se ha visto que el uso de marcado semántico y de *Schema.org* debería mejorar la visibilidad de una página web y su encontrabilidad, y, según Google, crear fragmentos enriquecidos en los resultados de búsqueda (Herramientas para webmasters de Google, 2012). Además, se observa una divergencia tecnológica importante debida a la elección por parte de los tres motores de búsqueda de una única sintaxis, los microdatos, prefiriéndola a otras sintaxis (RDFa Lite 1.1, por ejemplo) mucho más difundida y usada por los webmasters. Sin embargo, RDFa Lite 1.1 podría tener las características para ser aplicada a *Schema.org* y encima, antes de recomendar el uso de una sintaxis o de otra, haría falta analizar con un experimento cuales ventajas aportan las dos sintaxis. A partir de estas reflexiones se plantean algunas cuestiones:

- ¿El añadido de marcado semántico mediante *Schema.org* incide efectivamente en la mejora de la visibilidad de una página web y en su encontrabilidad? ¿Y cuál es la mayor ventaja que aporta?

- Al usar indistintamente el vocabulario de *Schema.org*, ¿cuál sintaxis, entre los microdatos y RDFa Lite 1.1, aporta mayores ventajas al posicionamiento de un sitio web?
- Puesto que los resultados de Google son cambiantes con el tiempo, ¿en que manera el actualización del contenido puede favorecer el posicionamiento de un sitio web enriquecido con datos estructurados?

Para contestar dichas preguntas se ha realizado un experimento, aplicando la metodología que explicaremos en el siguiente apartado.

2. Metodología.

En primer lugar se lleva a cabo una revisión bibliográfica con respecto al análisis de los microdatos y RDFa Lite 1.1. (los formatos de marcado semántico elegido para este estudio), para estudiar las motivaciones de la elección de los microdatos como formato de marcado recomendado y averiguar si RDFa Lite 1.1 tenga las características para ser aplicado al *Schema.org*. Todo ello mediante el análisis de las especificaciones de los Microdatos y de RDFa Lite 1.1 y de la producción científica encontrada sobre el tema en las principales bases de datos y directorios digitales de revistas especializadas en el campo de las ciencias de la información: ISI Web of Knowledge (WOK), Scopus, Google Académico, Dialnet, así como en las bases de datos de tesis doctorales Teseo, TDR, Tesinalia, Redial y NDLTD y catálogos de bibliotecas públicas y universitarias españolas: BPE y Rebiun. También se extiende la revisión al motor de búsqueda Google.

Sucesivamente se ha realizado un experimento que consiste en la puesta en marcha y desarrollo de tres sistemas – es decir tres páginas web – realizadas usando el CMS Joomla y optimizada con técnicas SEO tradicionales. En el código HTML de dos de ellas se añaden datos estructurados (en una los microdatos y en la otra RDFa Lite 1.1) junto con el *Schema.org*.

Finalmente, se realiza un estudio longitudinal mediante el lanzamiento de tres series de cadenas de búsqueda en tres periodos distintos (1 en el mes de Abril y 2 en el mes de Junio), con el fin de analizar como cambian los resultados a lo largo del tiempo y si la actualización del contenido de las tres web puede incidir en los resultados. En consecuencia, con el análisis de los resultados extraídos de las tres tomas de datos se evalúa:

- Cual de las tres sitios se posiciona mejor;
- Qué ventajas conlleva el uso de *Schema.org*;
- Qué formato de marcado semántico produce los mejores resultados;
- Como incide la actualización del contenido en el posicionamiento de una web con datos estructurados.

3. Formatos de marcado semánticos.

Los tres formatos más utilizados para el marcado semántico y la inclusión de datos estructurados en el código HTML / XHTML son: microformatos (4) microdatos y RDFa.

En este trabajo no se tratarán los microformatos, porque nos interesa centrar la atención en los microdatos, como sintaxis recomendada para *Schema.org* por parte de los motores de búsqueda y RDFa Lite 1.1 como la recomendación elaborada por el W3C.

Los microdatos son una especificación que fue introducida con HTML5 y que se podría definir como una "meta-sintaxis semántica", porque utiliza los atributos propios del código HTML, en lugar de ficheros XML separados. Proporciona un significado a las paginas para que sus contenidos puedan ser procesados automáticamente por aplicaciones informáticas, como los motores de búsqueda.

Se utilizan ciertos atributos para describir entidades, denominadas *ítems*. Entre ellos, los más utilizados son:

- *ítemscope* que indican que los descendentes del elemento marcado contienen microdatos (información semántica) sobre el contenido de dicho elemento.
- *ítemtype* que une el elemento usando una URL al vocabulario de *Schema.org* e indica el tipo de objeto (Persona, evento, dirección...).
- *ítemprop* que indica las propiedades de un *ítem*.

Para etiquetar el código HTML deben incluirse los ítems de microdatos, junto con enlaces al vocabulario de *Schema.org*. En la pagina web Getting Started with *Schema.org* (Schema.org 2012a) se proporciona un ejemplo practico de como etiquetar el marcado web (Figura 3).

En primer lugar se declara que se está utilizando un ítem de microdatos (*ítemscope*) con el correspondiente tipo utilizado (en este caso *Mo-*

vie) y a continuación el tipo utilizado.

En segundo lugar, se aplican las diferentes propiedades, como *name*, *director*, *birthDate* y *genre*. De esta manera, los motores de búsqueda reconocerán "Avatar" como la película de James Cameron y no como la encarnación terrestre de un dios o la imagen virtual de una persona.

Como se puede ver *ítemtype* añade un enlace a una etiqueta de *Schema.org*, <http://schema.org/Movie>, que permite identificar de manera univoca que estamos hablando de una película. Como se puede observar, una de las mayores ventajas de *Schema.org* es el uso de las direcciones URLs para eliminar las ambigüedades en la comprensión de los términos.

```
<div itemscope itemtype="http://schema.org/Movie">
<h1 itemprop="name">Avatar</h1>
<div itemscope itemtype="http://schema.org/Person">
<span itemprop="director">Director:
<span itemprop="name">James Cameron</span>
(born <span itemprop="birthDate">August 16,1954</span>
</div>
<span itemprop="genre">Science fiction</span><br />
<a href="/movies/avatar-theatrical-trailer.html" itemprop="trailer">
Trailer</a></div>
```

Figura 3: *Uso de los microdatos con Schema.org.*
Fuente: (Schema.org, 2012)

RDFa es otra tecnología para marcar datos HTML incluyendo indicaciones legibles por maquina. En esencia se trata de un conjunto de extensiones de XHTML, el cual viene propuesta directamente por el W3C como estándar específico para introducir datos semánticamente estructurados en el código de las páginas web.

RDFa fue originalmente diseñado para ser usado únicamente con XHTML. Sin embargo, ahora mismo la ultima versión, RDFa Lite 1.1 funciona para cualquier lenguaje basado en XML y sobre todo con HTML5 (W3C, 2011). RDFa Lite 1.1 usa una mezcla de atributos HTML, como href y src, y nuevos atributos, creados para solucionar algunos problemas específicos.

```
<html>
<head>...</head>
<body vocab="http://purl.org/dc/terms/">
<h1 property="title">La Huella</h1>
<p>Date: <span property="created">1972-12-10</span></p>
</body>
</html>
```

Figura 4: *Uso de RDFa Lite 1.1 con Dublin Core.*
Fuente: elaboración propia

Una característica que acerca RDFa Lite 1.1 a

los microdatos es el uso de enlaces para identificar muchos datos, que en lenguaje más técnico se llaman *recursos* (*resources*).

Esto es así porque el uso de direcciones URLs unívocas permite eliminar ambigüedades en la interpretación de un término (W3C, 2012, 13 de Diciembre), puesto que una URL se asocia a un elemento concreto de un vocabulario con un significado semántico específico.

Otra importante característica de RDFa es que permite definir un vocabulario por defecto, introduciendo el atributo *vocab*. Esto resulta muy útil cuando se utiliza un único vocabulario (ver Figura 4).

3.1. Microdatos y Rdfa Lite 1.1: dos metodologías de trabajo.

Básicamente se podría elegir entre dos metodología de trabajo de cara al añadido de las dos sintaxis:

- la aplicación de dos diferentes vocabularios de marcado (*Schema.org* con los microdatos y RDFa Lite 1.1 junto con *Dublin Core*);
- la aplicación del mismo vocabulario de marcado (*Schema.org* con ambas sintaxis).

Nuestro objetivo es aplicar *Schema.org* con ambas sintaxis. Para establecer si también RDFa Lite 1.1 tiene las características para ser aplicada a *Schema.org* fue realizada una revisión de la literatura existente sobre dicha cuestión. En la búsqueda de informaciones para sustentar la elección de una metodología o de otra hemos tenido en consideración algunos aspectos fundamentales:

- la motivación propuesta por el equipo de *Schema.org* de cara a la recomendación de usar los microdatos;
- la posibilidad de usar RDFa Lite 1.1 con diferentes vocabularios, incluso usar múltiples vocabularios en el mismo fichero;
- algunas características técnicas de los microdatos y RDFa Lite 1.1 que pueden afectar la elección de una o de otra opción

En consecuencia, vemos cuales son estos elementos que nos ayudan en nuestra elección.

El equipo de *Schema.org* ha elegido los microdatos por su sencillez de uso, y encima para proveer un sistema de marcado único con el fin de evitar la mezcla de sintaxis y el riesgo que los webmaster realicen marcados complejos que afecten a la consistencia y funcionamiento

de los motores de búsqueda (Herramientas para webmasters de Google, 2012). De todas formas, la específica RDFa Lite 1.1 resulta ser bastante sencilla de usar cuanto los microdatos, y, diversamente de los microdatos, tiene la ventaja de poder ser usada tanto con HTML5 como con XHTML.

RDFa Lite 1.1 puede ser usado contemporáneamente con diferentes vocabularios, introduciendo un vocabulario por defecto y añadiendo otros vocabularios para que estos sean aplicados juntos al primero con el fin de enriquecer la definición de hechos/objetos que no se podrían definir con el otro vocabulario. Por ejemplo, RDFa Lite 1.1 se usa por defecto con el vocabulario Dublin Core y, para representar los hechos que no son representados en Dublin Core se usa el vocabulario FOAF.

Al revés, algunos autores (Sporny 2011) afirman que se podría aplicar a *Schema.org* indistintamente ambas tecnologías, fundamentalmente porque ambas sirven para describir hechos/objetos y siguen un modelo de datos muy parecido (W3C, 2012, 8 de Marzo).

A pesar de eso, los microdatos y RDFa muestran algunas relevantes diferencias que pueden afectar la elección de una en lugar de la otra. Por ejemplo, según el HTML Data Guide del W3C (2012, 8 de Marzo) a la hora de elegir una sintaxis hay que tener en cuenta que en sitios multilingües sería mejor el uso de RDFa, porque en los microdatos, el vocabulario debería proporcionar un mecanismo separado para indicar el idioma, mientras que en RDFa se aplica automáticamente el idioma usado en la pagina web mediante el atributo *@lang*, así como en microdatos no se puede introducir del marcado como valor de una propiedad, hecho que se puede hacer fácilmente con RDFa.

Además, parece que, al analizar los datos recogidos en <http://www.commoncrawl.org>, el formato (Mika y Potter, 2012) RDFa es más utilizado por los webmasters que los microdatos y que este último se está difundiendo sobre todo gracias al éxito de *Schema.org*. Finalmente, la comunidad de los linked data ha realizado un portal (<http://schema.rdfs.org/>) donde encontrar informaciones para utilizar el vocabulario de *Schema.org* junto con RDFa Lite 1.1, con el fin de impulsar el uso de RDFa Lite 1.1 en lugar de los microdatos.

En consecuencia, por todas estas motivaciones (la sencillez de uso de RDFa Lite 1.1, su flexibilidad de uso con diferentes vocabularios, su di-

fusión en la web y dentro de la comunidad de los linked data) para desarrollar dicho proyecto de investigación se ha utilizado un único vocabulario, con el fin de comprobar cual sintaxis aporta mayores ventajas al posicionamiento de un sitio web.

4. Diseño del experimento.

En la consecución de los objetivos que marcan las pautas del presente trabajo y con el experimento que hemos realizado se pretende averiguar:

- el efecto que a nivel de posicionamiento orgánico de las páginas web produce el añadido de datos estructurados, es decir, comprobar si efectivamente mejoran la indexación y el posicionamiento de paginas web;
- en caso de una mejora del posicionamiento, evaluar cual de las dos sintaxis produce los efectos más relevantes al combinarla con *Schema.org*.

Antes de nada, se ha construido una primera página web donde se lo optimizó todo lo posible, siguiendo ciertos criterios SEO de interés para los motores de búsqueda y que explicaremos en los siguientes apartados. En las dos siguientes instalaciones se ha incluido los datos estructurados y el vocabulario compartido *Schema.org*, usando en una página los microdatos y en la otra RDFa Lite 1.1 como sintaxis para añadir datos estructurados semánticamente enriquecido en el código HTML.

Al principio las tres URLs eran iguales y se había introducido el mismo contenido en las tres instalaciones. Sucesivamente, fueron aportadas algunas actualizaciones al contenido por ciertas razones que explicaremos en el apartado de los resultados.

Las URLs de cada instalación son:

- Optimizada sólo SEO: <http://www.esebedesign.org/> (Esebedesign);
- Optimizada SEO añadiendo los microdatos y el vocabulario *Schema.org*: <http://www.esebedesign1.org/> (Esebedesign1);
- Optimizada SEO añadiendo RDFa Lite 1.1 y el vocabulario *Schema.org*: <http://www.esebedesign2.org/> (Esebedesign2).

Las tres web fueron creadas y lanzadas en el mes de Febrero 2013, así que se trata de pági-

nas web con pocos meses de vida:

- las web <http://www.esebedesign1.org/> y <http://www.esebedesign.org/> han sido lanzadas el 19 de Febrero 2013;
- la web <http://www.esebedesign2.org/> ha sido lanzado el 23 de Febrero 2013

La elección de investigar páginas web creada desde hace poco tiempo, es relacionada con el objetivo de investigar que pasa si utilizamos una estrategia SEO que desde el principio abarca también el añadido de los datos estructurados en el código HTML.

Por cuanto concierne la realización del *template* se ha elegido la opción de realizar la estructura desde cero, instalando la plantilla en el CMS Joomla! y usando HTML5 para definir la estructura y el contenido de una página web y las hojas de estilo en cascada CSS3 para definir la presentación de dicho contenido, puesto que ambas tecnologías están ligadas con el posicionamiento SEO y son elementos favorecedores de un buen posicionamiento.

A nivel de SEO, SMO y Web 2.0 se han introducido y analizado algunos elementos de interés para los motores de búsqueda y capaces de mejorar el posicionamiento.

Finalmente se ha definido una serie de cadenas de búsqueda para introducir en el motor de búsqueda, usando las palabras clave previamente elegidas para optimizar las páginas web, con el fin de evaluar cual de las tres resulta ser la versión mejor posicionada.

4.1 Creación de la plantilla.

Para la creación del *template* de Joomla! se ha preferido desarrollar una plantilla desde cero, para tener bajo control todo los aspectos que pueden ser incluidos en la realización de una página web, desde el código HTML hasta los diferentes elementos SEO a utilizar.

Se ha preferido evitar la inclusión de elementos *flashs* y *frames* que podrían generar problemas de accesibilidad, escogiendo librerías jquery para dinamizar las páginas y se ha desarrollado la versión adaptada para móvil usando la técnica del Responsive Design (5).

Además, a nivel de código se ha preferido el empleo de HTML5 en lugar de, por ejemplo, XHTML. Como CMS se ha elegido el CMS Joomla! para construir las web contando en algunos aspectos característicos de dicho CMS y

se han incorporado algunas extensiones que contestasen a requisitos específicos.

4.1.1 HTML5, CSS3 y Responsive Design, elementos favorecedores para el SEO semántico.

La aplicación de la técnica Responsive Design es bastante intuitiva y permite dividir una plantilla en bloques independientes que se colocan o que se ordenan según el dispositivo y consecuentemente la resolución utilizada, puesto que cada dispositivo usa una resolución diferente. Normalmente para aplicar la técnica del Responsive Design se usan las hojas de estilo CSS3, que permiten juntos con las media queries de crear efectos y adaptar la plantilla a cualquier dispositivos.

A nivel SEO el Responsive Design puede aportar diferentes beneficios (Developers Google, 2013):

1. usando dicha técnica no se producen múltiples URLs, sino todas las versiones de la página web comparten la misma URL;
2. no surgen problemas de información duplicada en la web, porque existe un único contenido, sin duplicados y los crawlers adquirirán una sola página;
3. habrá una sola web a optimizar y una sola analítica a estudiar.

Además, la web adaptada permite una navegación mucho más satisfactoria para los usuarios (Gardner, 2011), permitiendo al sitio web de ser usable y accesible desde cualquier dispositivo. Con respecto a HTML5, su razón de ser y su mayor beneficio se encuentra en la creación de nuevas etiquetas semánticas que han permitido mejorar enormemente la indexación del contenido y su clasificación semántica, y también la presentación de documentos (Franganillo, 2010):

- <header> para identificar las cabeceras en una página HTML5;
- <footer> para identificar los pies de una página web;
- La etiqueta <nav> para identificar la zona de navegación, formada por enlaces o un menú de navegación.

Por tanto, la aplicación de HTML5 puede ser útil a nivel de aplicación del SEO semántico.

Además, HTML5 facilita la interpretación de los

documentos en el caso que no sean validos o no están bien formados. Dicha relación de HTML5 con la semántica resulta útil a la ahora de mejorar la accesibilidad de las paginas web, porque permite a los lectores de pantallas de desplazarse rápidamente por la pagina web (Lawson y Faulkner, 2011).

4.1.2 Elección de Joomla!

Las páginas web han sido construidas con el gestor de contenidos Joomla!, porque se trata de un CMS que, diversamente de otros (Drupal, por ejemplo), posee una curva de aprendizaje bastante baja y resulta ser más usado en la construcción de sitios web “escaparate”, como en el caso de las páginas web creadas para el nuestro estudio. Otro CMS que se hubiera podido elegir era Wordpress, que posee algunos plugins favorecedores al posicionamiento en motores de búsqueda (Soler del Pozo, 2012)

Sin embargo la elección de Joomla! está relacionada con el objetivo de averiguar si también Joomla! tiene elementos favorecedores en este ámbito, ya que sería interesante profundizar esta relación con un trabajo a nivel académico.

4.1.3 Elección de las extensiones.

Las extensiones de Joomla! Son agrupadas en diferentes categorías. Dentro de las extensiones catalogadas como “Site Management” se han tenido en cuenta para la elección las extensiones incluidas en las sub-categorías SEF y SEO & Metadata (restringiendo la evaluación sólo a las incluidas en la sub-categoría “Metadata”). Las siguientes son extensiones imprescindibles y incluidas en los dos sistemas con el marcado semántico:

- *J4schema*: añade los microdatos y el vocabulario de *Schema.org*;
- *Htm5 breadcrumbs*: permite añadir automáticamente a los moldes de pan fragmentos enriquecidos incluyendo microdatos o RDFa Lite 1.1;

No existe en Joomla! una extensión parecida a *J4schema* que permita añadir código en RDFa Lite 1.1 a las páginas web. Por ello, el añadido de RDFa Lite 1.1 se ha realizado modificando manualmente el marcado de los microdatos, siguiendo las indicaciones incluidas en la página web <http://schema.org/docs/datamodel.html> para aplicar la sintaxis RDFa Lite 1.1 al vocabula-

rio *Schema.org*. En consecuencia se ha substituido *itemprop* con *property* y *itemtype* con *typeof*, eliminando *itemscope*.

En los tres sitios se ha usado un componente para rescribir las URLs en forma amigables, y actualizar automáticamente las metaetiquetas *description*, *keywords* y otras metaetiquetas. Para elegir el componente SEF management más adecuado para la experimentación, es decir la ARTIOJoomSEF, se han comparado dos componentes, seleccionando el más apto basándose en algunos criterios indicados en la Tabla 1 del apéndice.

4.2 Elementos SEO a estudiar.

Durante la fase de implementación de una página web hay que tener en cuenta la realización de una optimización *on-page* de algunos elementos, que resulta ser imprescindible para conseguir la indexación de una página web y la mejora de su posicionamiento según Google y su *Guía para principiantes sobre la optimización para motores de búsqueda* (Google, 2006). Una buena optimización para motores de búsqueda afecta sólo a los resultados de búsqueda orgánicas, que representan nuestro reto, es decir posicionar aquí y en la mejores posiciones nuestras páginas web.

Se ha elaborado una tabla (Tabla 2 del apéndice) con todos los elementos SEO favorecedores de una buena optimización y aplicado en los tres sitios web. Aunque dos de ellas incluyen datos estructurados, se supone que se deberían también optimizar las páginas que incluyen datos estructurados. Con la ayuda de algunas extensiones Joomla! (en particular ARTIOJoomSEF) se han optimizado los elementos más importantes y que tienen la prioridad alta. Entre estos elementos, llamados también *factores SEO* (Gil Leiva, 2008), deseables para una página web hay algunos que destacan y que resultan ser imprescindible a la hora de optimizar una página web.

El elemento más importante a la hora de construir una página web es crear títulos de páginas únicos y relevantes, es decir, que estén relacionados con el contenido de la página web, contener las palabras claves y además ser una combinación única en cada página (Vasta, 2010). Los títulos tienen que colocarse dentro de la etiqueta `<title>`. Por lo tanto, es un elemento importante porque su contenido se muestra en los resultados de la búsqueda.

Otro elemento básico a la hora de optimizar una página es la inclusión de la metaetiqueta *description* que viene siendo usada por Google para mostrar fragmentos de descripción, *snippet* (Google, 2006) y que permite incluir un resumen del contenido de la página.

Finalmente, se realizó un *sítemap*, es decir, un mapa web con todos los vínculos del sitio que permite a Google de rastrear sitios, puesto que ayudan a los *spiders* de Google a comprender cuando una página ha sido modificada por última vez y cuál es la frecuencia con que cambia dicha página.

Para conseguir un mejor posicionamiento en los motores de búsqueda hay que usar URLs amigables que contengan palabras claves y usar una etiqueta de cabecera `<h1>` única en cada página con las palabras clave y que repita el contenido de la etiqueta `<title>`.

En el caso de los sitios web desarrollados para este trabajo se ha utilizado el comando `echo` de php para crear dinámicamente una etiqueta `<h1>` a partir del contenido de la etiqueta `<title>`. Esto permite ahorrar tiempo, porque evita de incluir manualmente todas las etiquetas de cabecera.

A nivel de validación de las res páginas web se han usado sobre todo el *W3C Markup Validation Service* para validar el código HTML5 (Validator W3C, s.f.a) y a nivel de marcado CSS3 (Validator W3C, s.f.b) se ha aplicado el test de validación *W3C CSS Validation Service* (Tabla 3 en Apéndice). En ambos casos se ha conseguido un listado de errores que se han corregido en las tres versiones del sitio, dejando sólo un error a nivel de código HTML5: la metaetiqueta `title`.

Aunque se trata de una metaetiqueta desaprobada en HTML5 esta ha siempre sido bastante importante para la mejora del posicionamiento de una página (Vasta, 2010) y dejarla no significa afectar negativamente a los resultados de búsqueda. Aplicando el test de Accesibilidad Web WAI (eXaminator, s.f.) se han encontrado sólo algunos errores a nivel de conformidad AAA que se han corregido.

Finalmente, se ha aplicado el Test de Google, Herramienta de prueba de datos estructurados (Google, s.f.), para evaluar si se han aplicado correctamente los datos estructurados tanto en la versión *Schema.org* y los microdatos, como en la versión *Schema.org* con RDFa Lite 1.1. Este caso no se han encontrado errores.

4.3 Elementos SMO a estudiar.

Se han incluido en las tres páginas web unos botones sociales para compartir contenido y para ver si conseguimos algún resultado significativo capaz de mejorar el posicionamiento.

En todas tres las páginas hemos puesto los botones a las cuentas de Facebook y Twitter, mientras que sólo en las dos páginas con datos estructurados se ha incluido el botón a *Google+*, creando dos páginas diferentes para cada sitio web. En dichas páginas se han publicado el contenido de los sitios web, y se han enviado enlaces a los dos sitios.

Sin embargo, a nivel de análisis de las mejoras aportadas por las redes sociales, la red más interesante a estudiar es la de *Google+*, que tiene más relación con el SEO semántico. Teóricamente crear una página diferente para cada sitio web sirve para obtener una prueba de autoría que permita a Google enlazar el contenido de una página web con datos estructurados con su verdadero autor. En este caso las dos páginas son enlazadas con el perfil de *Google+* de Samantha Bernardis, comprobando que existe un autor real del contenido de los sitios web.

4.4 Cadenas de búsqueda.

Las cadenas de búsqueda se han obtenido de los contenidos de las tres páginas. En primer lugar se ha utilizado la herramienta de sugerencia de Google Adwords (s.f.) para seleccionar las mejores palabras clave conectadas con el tema tratado en los tres sistemas creados.

Las palabras clave elegidas se encuentran en la tabla 4 del Apéndice, e incluye diferentes categorías de términos que abarcan el mundo del diseño web, de la gráfica y del marketing online. En la elección de las palabras clave se ha dado prioridad a grupos de palabras de dos términos para evaluar el nivel de competencia que tienen en el entorno web. Sin embargo, en la fase de búsqueda se ha pensado de lanzar también cadenas de términos que incluyen 3 o 4 palabras. No obstante, en la mayoría de los casos uno de los términos se refiere a la ubicación geográfica, Alicante, donde ejerce la empresa ficticia de la que se habla en los tres sitios web. La decisión de usar grupos de palabras de 3 o 4 términos está relacionada con la nueva tendencia en el ámbito del comportamiento informacional en la búsqueda web. Según una investigación realizada por Statowl.com (s.f., c.p. Martínez-Méndez, 2013) se observa que la mayoría de los usuarios utiliza 3 términos (23,73%) o 4 (27%) en la fase de la búsqueda (ver figura 5).

A nivel metodológico se ha establecido de realizar tres series de búsqueda distintas con un número creciente de cadenas de búsqueda (9, 19 y 30). Según algunos autores (Vasta, 2010) en caso de sitios recientemente creados en los resultados de búsqueda en los primeros meses se encuentran sólo resultados conectados con las palabras clave. Vasta (2010) aconseja de aumentar cada vez el número de cadenas de búsquedas y usar en estas unas de las palabras para evaluar qué clase de resultados salen. Por ello, las dos primeras tomas de datos han sido realizadas con el fin de averiguar si aparecían ya resultados.

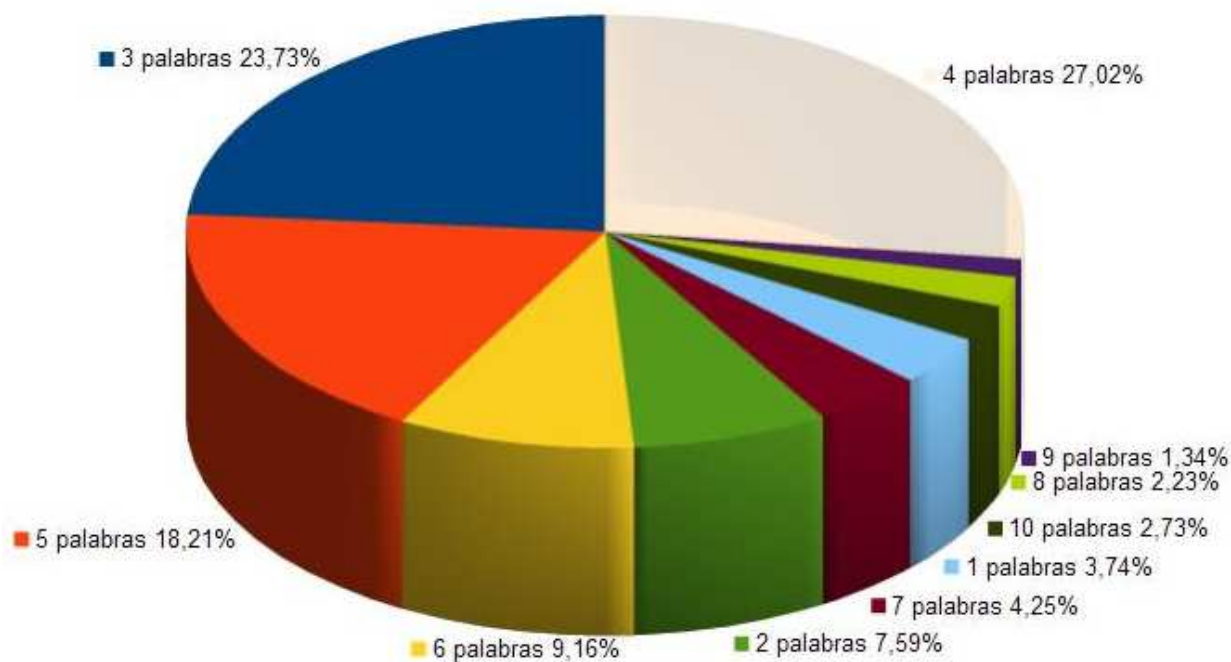


Figura 5: Comportamiento informacional en la búsqueda (Martínez Méndez, 2013). Fecha de consulta: 15 de mayo 2013

5. Resultados.

En un primer lugar se ha realizado una toma de datos en el mes de Abril 2013 (entre el 13 de abril de 2013 y el 19 de abril 2013), al cabo de una semana de la primera introducción de contenido en las tres páginas. Para efectuar dicha primera toma de datos se han incluido sólo las 8 cadenas de palabras clave previamente seleccionadas para posicionar el sitio web más la cadena "Samantha Bernardis", que suman 9 cadenas de búsqueda distintas, con el fin de comprobar si ya aparecían resultados relevantes.

Las tres páginas han sido creadas sólo en el mes de febrero 2013 y los resultados obtenidos de esta primera toma de datos fueron bastante escasos: sólo 4 resultados (tabla 5 del apéndice), dos de ellos conectados con una de las palabras clave, "proyectos gráficos Alicante" y los otros dos conectados con el nombre del autor del contenido de las tres web, "Samantha Bernardis". Dicha cadena de búsqueda fue diseñada para comprobar si funciona la autoría establecida con la red social de Google+, vinculando el nombre del propietario de la cuenta a los sitios web enriquecidos con los microdatos y RDFa lite 1.1.

Entre el 15 y el 19 de Junio de 2013 se realizó una segunda toma de datos. En este segundo caso se ha modificado una variable: los sitios

web no han sido actualizados, es decir, el contenido no ha cambiado con respecto a la última actualización efectuada en el mes de abril. Se realizaron 19 búsquedas (tabla 5 del apéndice). De estas cadenas, 8 se han sido palabras aisladas y a la vez de palabras entrecorridas, lo que suma 16 cadenas de búsquedas distintas. Las otras tres son el nombre "Samantha Bernardis", y dos grupo de palabras sueltas que incluyen más de 4 términos en la búsqueda. Las primeras 16 cadenas de búsquedas son las palabras clave elegidas para posicionar los sitios web más la palabra "Alicante" usada para restringir la área de búsqueda. En este caso se ha obtenido 10 resultados (Tabla 5 en apéndice).

En la última semana de Junio 2013, del 25 al 30 de Junio, se ha realizado la tercera toma de datos, tras actualizar el contenido de algunas páginas de los tres sitios web. Se han lanzado 30 búsquedas (tabla 6 del apéndice), 19 son las mismas cadenas de búsquedas lanzadas previamente y las otras 11 son grupos sintácticos de palabras individuales que incluyen 3 o más términos (también unas palabras clave) en la búsqueda. En este caso se han obtenido 36 resultados.

6. Discusión.

En los resultados de la primera toma de datos, encontramos dos enlaces procedentes del sitio web marcado con microdatos, mientras que

aparece una del sitio web optimizado con RDFa Lite 1.1 y una del sitio optimizado sólo SEO. La web con los microdatos resulta ser la mejor posicionada por la palabra clave “*proyectos gráficos Alicante*” y la única a alcanzar, ya al cabo de una semana, la primera página, con respecto a la página sólo SEO que se coloca en la tercera página.

Respecto a la segunda clase de resultados obtenidos, se observa que salen las dos webs con la cuales hemos establecido la autoría en el perfil de *Google+*, *Esebedesign1* y *Esebedesign2*, aunque se recupera antes la página web optimizada con RDFa Lite 1.1 que la URL de la instalación optimizada con los microdatos.

Si se observa la tabla 5, en la segunda toma de datos se detectan algunas mejoras en el posicionamientos, aunque sólo para *Esebedesign*.

De hecho, emerge claramente una preponderancia de URLs procedentes de las páginas optimizadas únicamente con SEO tradicional y se observa como la palabra clave que previamente devolvía la página optimizada con los microdatos (*proyectos gráficos Alicante*), ahora no devuelve más el mismo enlace. Se podría suponer que las páginas web que incluyen datos estructurados son mucho más valoradas por los crawlers a la hora de actualizar el contenido y que, entonces, necesitan actualizarse más frecuentemente para mantener el posicionamiento adquirido. Por su parte, las páginas optimizadas sólo con técnicas SEO tradicionales se posicionan mucho más lentamente pero mantienen el posicionamiento aunque no sean actualizadas.

La tercera toma de datos confirma esta última afirmación. A partir de los 36 resultado válidos, observamos que de ellos, 15 corresponden a una URL proveniente del sitio web marcado con los microdatos, otros 15 del sitio optimizado con SEO tradicional y 6 de la web con RDFa Lite 1.1. Los enlaces procedentes de las páginas con microdatos aumentan considerablemente, y se anotan 13 URLs más que en la primera toma de datos, contra el aumento de sólo 5 enlaces de las otras dos web. Además, la instalación con los microdatos substituye la web sólo optimizada SEO en el posicionamiento por la palabra clave “*proyectos gráficos Alicante*” en poco más que una semana.

Si observamos el orden de aparición de los resultados, en la primera y segunda posición, vemos que las URLs de la instalación sólo optimizada SEO aparece antes (1ª posición) que una URL de las otras instalaciones en 9 ocasiones,

ocurriendo la situación contraria para la página con microdatos únicamente en 6 ocasiones. Además, en las primeras páginas aparece la URL del sitio con SEO tradicional en 14 ocasiones contra los 9 resultados de la web con microdatos y los 3 de la instalación con RDFa Lite 1.1. Estos datos dejan claro que la web optimizada sólo con SEO es la mejor posicionada.

7. Conclusiones.

Tras analizar los resultados obtenidos en el apartado anterior obtenemos algunas conclusiones que nos sirven para contestar a las preguntas hechas en el apartado introducción.

Puede afirmarse que el añadido de marcado semántico produce una efectiva mejora al posicionamiento de una página web sólo a la hora de actualizar su contenido, y sin la presencia de este elemento no se producen significativas mejoras. Además, la presencia en la segunda toma de datos de enlaces procedentes sólo de *Esebedesign* nos lleva a concluir que los beneficios que *Schema.org* (una mayor visibilidad y “*encontrabilidad*” de las páginas web) debería aportar no se producen a largo plazo.

De la lectura de estos datos se pueden extrapolar algunas suposiciones. De estas la más interesante es que los motores de búsqueda dan relevancia al contenido actual y actualizado, y además que, en caso de actualización, las web enriquecidas con datos estructurados alcanzan en tiempos breves un mejor posicionamiento que las que son optimizadas únicamente con SEO. En consecuencia, una recomendación que se podría hacer es la actualización frecuente del contenido de una página para aprovechar las ventajas del marcado semántico de datos estructurados.

Por cuanto concierne a las otras dos cuestiones: la ventaja principal que puede aportar el uso de *Schema.org* se relaciona con la ventaja de alcanzar las primeras posiciones de una S.E.R.P. (Search Engine Result Page, Página de Resultados de Búsqueda) en un plazo relativamente breve de tiempo y al actualizar el contenido de una página.

Respecto a la segunda cuestión cabe concluir que el uso de RDFa Lite 1.1 no ayuda a mejorar el posicionamiento de una página web.

Sobre la implicación que tienen las Redes Sociales en el posicionamiento (Optimización SMO) tan sólo se ha conseguido comprobar el funcionamiento de la autoría con *Google+* a la

hora de actualizar las páginas. Sin embargo, dicha autoría en la segunda y tercera tomas de datos no ha aparecido más.

Debido a esto puede ser que influyan en las otras redes factores como el número de seguidores, implicación de los usuarios (“menciones”, “retwitts”, “+1”) etc, algo que no ha podido conseguirse en nuestro estudio. Se podrían haber obtenido otro tipo de resultados teniendo en cuenta y mejorando algunos de los siguientes aspectos que añadimos como limitaciones al estudio planteado:

- Poco contenido generado en la web. Se podría haber añadido mucho más contenido para tener mucha más páginas indexadas;
- Poco tiempo para la generación de resultados. Las páginas han sido creadas en febrero de 2013;
- Pocos enlaces entrantes. Las páginas han sido enlazadas sólo con las redes sociales y de un par de otros sitios web;
- Bajo Pagerank. En este estudio no se ha tenido en cuenta el PageRank de las páginas, lo que afecta también al posicionamiento de las páginas;
- Utilizar una mayor diversidad de cadenas de búsqueda en cada una de las búsquedas. Para conseguir datos uniformes y coherentes con este estudio preliminar haría falta realizar otras tomas de datos con el mismo número de cadenas (30 cadenas);
- usar otras palabras clave. La presencia de muchas palabras clave con media y alta competencia puede haber influido negativamente en el posicionamiento de las tres páginas web. Por ello, en un estudio futuro se debe tener en cuenta la posibilidad de utilizar más palabras clave con baja competencia;
- Realización del experimento con Joomla!. Se habría podido utilizar otro CMS (Wordpress o Drupal) para analizar si el gestor de contenidos utilizado influye en los resultados.

A partir del estudio realizado en esta investigación se podría seguir las siguientes líneas o vías futuras de investigación:

- Repetir el experimento lanzando el mismo número de cadenas de búsqueda (30) para conseguir datos más uniformes;
- Repetir el experimento usando otro gestor de contenidos, con el fin de averiguar si el CMS

incide en el posicionamiento de sitios enriquecidos con datos estructurados.

- Como la actualización del contenido es un elemento que favorece el buen posicionamiento de un sitio enriquecido con datos estructurados, averiguar que sucede al aplicar la metodología seguida en este trabajo, con sitios web de comercio electrónico donde se actualiza frecuentemente el contenido.

Notas

- (1) La composicionalidad es un principio según el cual el significado de las expresiones complejas está completamente determinado por su estructura y el significado de sus expresiones componentes (Compositionality, 2004).
- (2) La SEO es un acrónimo inglés que significa *Search Engine Optimization*, y se refiere a la *Optimización de Motores de Búsqueda*, un conjunto de técnicas que ayudan a una página web a conseguir la mejor posición en las SERP ofrecidas por los motores de búsqueda al hacer una búsqueda determinada (Gil Leiva, 2008).
- (3) RDFa Lite 1.1 es una especificación que hace referencia a RDF. Que es otro modelo de inclusión de datos semánticamente relevantes basado en RDF y propuesto directamente como especifica estándar por W3C (2011).
- (4) Junto con los microdatos y RDFa hay también otra especificación estándar para extraer datos de las páginas HTML bajo la denominación de microformatos (W3C, s.f.)
- (5) La traducción del término Responsive Design en Castellano sería “diseño web adaptativo”. Se refiere a un conjunto de técnicas de diseño web que usando hojas de estilo CSS3 permite adaptar una misma página para diferentes dispositivos y en particular para los dispositivos móviles, como la tableta y el smartphone (Wisniewski, 2013).

Referencias

- Bradley, A. (2011, 3 de Junio). *What Schema.org Means for SEO and Beyond*. *SEOSkeptic*. Recuperado 07-12-2012, de <http://www.seoskeptic.com/what-schema-org-means-forseo-and-beyond/>
- Compositionality. (2004, 8 de Abril). En *Stanford Encyclopedia of Philosophy*. Recuperado 02-07-2013, de <http://plato.stanford.edu/entries/compositionality/#1.4>
- Developers Google. (2013). *Building Smartphone-Optimized Websites*. Recuperado 01-06-2013, de <https://developers.google.com/webmasters/smartphone-sites/details>
- Franganillo, J. (2010). HTML5: el nuevo estándar básico de la Web. *Anuario, ThinkEPI*. 5 (Septiembre 2010) 261-265.
- eXaminator. (s.f.). *Evaluación automática de la accesibilidad*. Recuperado 05-05-2013, de <http://examinator.ws/>

- Gardner, B.S. (2011). Responsive Web Design: Enriching the User Experience. *SigmaΣ, Inside the digital ecosystem*, 11:11 (2011) 13-17.
- García Marco, F. J. (2012) Schema.org: la catalogación revisitada". *Notas ThinkEPI*. Recuperado 26-12-2012, de <http://www.thinkepi.net/schema-org-catalogacion-srevisitada>
- Gil Leiva, I. (2008). *Manual de Indización, teoría y práctica*. Asturias (España): Ediciones Trea. 431 p.
- Google (2006). *Guía para principiantes sobre optimización para motores de búsqueda*. Recuperado 17-06-2013, de http://static.googleusercontent.com/external_content/untrusted_dlcp/www.google.es/es/es/webmasters/docs/guia_optimizacion_motores_busqueda.pdf
- Google (s.f.). *Herramientas de pruebas de datos estructurados*. Recuperado 05-05-2013, de <http://www.google.es/webmasters/tools/richsnippets>
- Google Adwords (s.f.). *Herramienta para palabras clave*. Recuperado 01-03-2013, de https://adwords.google.com/o/Targeting/Explorer?__c=9026451409&__u=8897273449&__o=cues&ideaRequestType=KEYWORD_IDEAS
- Herramientas para webmasters de Google. (2012). *Preguntas frecuentes sobre Schema.org*. Recuperado 29-12-2012, de <http://support.google.com/webmasters/bin/answer.py?hl=es&answer=1211158>
- Lawson, B. y Faulkner, S. (2011, 25 de Mayo). HTML5 and Accessibility. *Scriptjunkie* β . Recuperado 29-05-2013, de <http://msdn.microsoft.com/en-us/magazine/hh204741.aspx>
- Martínez Méndez, F.J. (2013). *Apuntes de clase. Asignatura: Evaluación y medida de la economía del conocimiento en las organizaciones*. Facultad de Comunicación y Documentación, Universidad de Murcia
- Méndez, E. (2007). Cuando la búsqueda se vuelve semántica: SWoogle. *Anuario ThinkEPI*, 1 (febrero 2006) 65-68.
- Mika, P.; Potter, T. (2012). Metadata Statistics for a Large Web Corpus. En: *Actas de Congreso Linked Data Workshop (LDOW) at the International World Wide Web Conference*. Recuperado 07-12-2012, de <http://events.linkedata.org/ldow2012/papers/ldow2012-inv-paper-1.pdf>
- Pástor Sánchez, J. A. (2012). Prospectiva de la Web semántica: divergencia tecnológica y creación de mercados Linked Data. *Notas ThinkEPI*, 6 (diciembre 2011) 269-275.
- Page, L., Brin, S., Motwani, R. y Winograd, T. (1999). *The PageRank Citation Ranking: Bringing Order to the Web*. Recuperado 02-07-2013, de <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.31.1768>
- Sánchez Cuadrado, S.; Colmenero Ruiz, M. J.; Moreira, J. A. (2012) Tesoros: estándares y recomendaciones. *El Profesional de la Información*. 21(3) (Mayo-Junio 2012) 229-235.
- Saorín, T.; Pástor Sánchez, J. A.. (2012). Gestión documental y de contenidos web: informe de situación. *Anuario ThinkEPI*. 6 (Junio 2012) 232 – 239.
- Schema.org (2012a). *Getting Started with Schema.org*. Recuperado el 7 de Diciembre de 2012, de <http://schema.org/docs/gs.html>
- Schema.org (2012b). *What is Schema.org*. Recuperado el 7 de Diciembre de 2012, de <http://schema.org/>
- Soler del Pozo, J. (2012). Aproximación al estudio del posicionamiento en buscadores con WordPress: elementos internos favorecedores. *Cuadernos de Gestión de Información* 2. 85-104. Recuperado el 7 de Diciembre de 2012 de <http://fcdmurcia.es/ojs/index.php?journal=cuadernos&page=article&op=view&path%5B%5D=111>
- Sporny, M.. (2011). The false choice of Schema. Org. *The Beautiful, Tormented Machine*. Recuperado 07-12-2012, de <http://manu.sporny.org/2011/falsechoice/>
- Validator W3C. (s.f.a). *Markup Validation Service*. Recuperado 17-06-2013, de <http://validator.w3.org/>
- Validator W3C. (s.f.b). *CSS Validation Service*. Recuperado 05-05-2013, de <http://jigsaw.w3.org/css-validator/>
- Vasta, D. (2010). *SEO Ottimizzazione web per motori di ricerca*. Milano: Apogeo. 166p.
- Wisniewski, J. (2013). Responsive Design, Control-shift. *Online Searcher*. 1, 74-76. Recuperado 14-04-2013, de <http://pqasb.pqarchiver.com/infotoday/access/2882128061.html?FMT=ABS&FMTS=ABS:FT:PAGE&type=current&date=Jan%2FFeb+2013&author=Jeff+Wisniewski&pub=Online&edition=&startpage=74&desc=Responsive+Design>
- W3C (2011, 19 de Abril). *RDFa 1.1 Primer*. Recuperado el 28-12-2012, de <http://www.w3.org/TR/2011/WD-rdfa-primer-20110419/>
- W3C. (2012, 8 de Marzo). *HTML Data Guide*. Recuperado el 27-12-2012, de <http://www.w3.org/TR/html-data-guide/>
- W3C (2012, 13 de Diciembre). *HTML+RDFa 1.1*. Recuperado el 28-12-2012, de <http://www.w3.org/TR/rdfa-in-html/>
- W3C. (s.f.). *El W3C establece un puente entre HTML/Microformatos y la Web Semántica*. Recuperado el 24-04-2013, de http://www.w3c.es/Prensa/2007/nota070911_grddl.html

Apéndice.

Característica	MijoSEF (versión free)	ArtioJoomSEF (versión free)
	Valoración	Valoración
Compatible con versión 1.5 de Joomla!	3	4
SEF (amigables) urls	4	4
Metadata generación	4	4
Orientado a SEO	4	4
Orientado a S.M.O.		
Urls redirect 301		4
Compatibilidad con JoomFish para la gestión SEO de paginas traducida a otro idioma		4
Total	15	24

Tabla 1: Criterios de valoración de las extensiones SEF a instalar. Fuente: elaboración persona

Criterio	Recomendación	P	OK	Observación
URL Amigable	Estructura semántica de la URL	SI	SI	
Acceso www.- sin www	Mismo destino	SI	SI	
Titulo	15-75 caracteres	SI	SI	65 caracteres
Titulo no repetido	Distinto en cada pagina y con las keywords	SI	SI	
Metadescription	70-160 caracteres	SI	SI	159 caracteres
Metakeywords	SI	SI	SI	
Cabeceras	H1-H6. Sólo un H1 con keywords y que repite el titulo	SI	SI	Creado dinamicamente con PHP a partir del titulo
Nombre fichero imagen	Amigable y descriptivo	SI	SI	
ALT en imágenes	SI. Obligatorio	SI	SI	
Frames	NO	NO	SI	
Flash	NO	NO	SI	
Links de Navegación	Migas de pan y menú	NO	SI	
Textos ancla	Descriptivos	NO	SI	
rel="nofollow"		NO	SI	Enlaces externos
Fichero robots.txt	SI	SI	SI	
Sitemap XML	SI	SI	SI	
Validación W3C	SI	SI	NO	1 error (presencia de meta-title. No afecta el posicionamiento)
Optimización CSS3		NO	NO	Manual

P: Prioridad alta OK: Indica si se cumple el criterio en los sitios web del experimento

Tabla 2: Criterios SEO deseables en la instalación sólo optimizada SEO (y incluidos también en las otras dos páginas).
Leyenda: "P" Fuente: Elaboración Propia adaptando la propuesta de Soler del Pozo (2012)

PRUEBAS EFECTUADAS		ESEBEDESIGN (SEO)	ESEBEDESIGN1 (MICRODATOS)	ESEBEDESIGN2 (RDFa LITE 1,1)
W3C	HTML5	1	1	1
	CSS3	0	0	0
WAI	A	0	0	0
	AA	0	0	0
	AAA	0	0	0
HPE*	MICRODATOS		0	
	RDFa LITE 1.1			0

Tabla 3: Resultados de validación de código. Fuente: elaboración personal

* = Herramienta de pruebas de datos estructurados

Competencia	Búsquedas globales mensuales	Búsquedas locales mensuales	
páginas web Alicante	Media	720	720
proyectos gráficos	Baja	2900	590
marketing online	Media	7.480.000	33.100
marketing digital	Media	450.000	12.100
Posicionamiento seo	Alta	22.200	14.800
diseño web profesional	Alta	3.600	1.600
diseñador web	Alta	27.100	9,900
diseño web	Alta	368.000	135.000
diseñador gráfico	Media	450.000	60.500
Diseño gráfico Alicante	Media	480	480

Leyenda: competencia, nivel de popularidad de una combinación de palabras clave; búsquedas globales mensuales, indica con qué frecuencia los usuarios de cualquier ubicación buscan las palabras claves; búsquedas locales mensuales, muestran la frecuencia con la que los usuarios en el ámbito local buscan las palabras clave.

Tabla 4: identificación de las palabras clave para las tres páginas web y evaluación de su posible impacto a nivel de posicionamiento en los motores de búsqueda. Fuente: elaboración propia a partir de los datos recogidos mediante Google Adwords (s.f.).

1ª toma de datos	2ª toma de datos	3ª toma de datos	
Cadenas de búsqueda	9	19	30
Resultados	4	10	36
Resultados agotados	14	28	24
Resultados Esebedesign SEO	1	10	15
Resultados Esebedesign1 MICRODATOS	2	0	15
Resultados Esebedesign2 RDFa LITE 1.1	1	0	6
Resultados en la 1º página de la web SEO	0	7	14
Resultados en la 1º página de la web con los microdatos	2	0	9
Resultados en la 1º página de la web con RDFa Lite 1.1	1	0	3
Resultados SEO antes de microdatos o RDFa Lite 1.1	0	0	9
Resultados microdatos antes SEO o RDFa Lite 1.1	1	0	6
Resultados RDFa Lite 1.1 antes microdatos o SEO	1	0	0

Tabla 5: Aspectos analizados en las tres primeras pruebas en las búsquedas.
Primera, segunda y tercera tomas de datos (Abril de 2013 y Junio de 2013). Fuente: Elaboración Propia

P	#	UR	L
CADENA 1: Samantha Bernardis			
1º RESULTADO	1	3	http://www.esebedesign.altervista.org/sobre-esebedesign.html
2º RESULTADO	-	-	Resultados agotados
CADENA 2: "proyectos gráficos Alicante"			
1º RESULTADO	1	3	http://www.esebedesign1.altervista.org/
2º RESULTADO	1	5	http://esebedesign2.altervista.org/
CADENA 3: proyectos gráficos Alicante			
1º RESULTADO	1	4	http://www.esebedesign1.altervista.org/
2º RESULTADO	1	6	http://www.esebedesign.altervista.org/
CADENA 4: "marketing online Alicante"			
1º RESULTADO	3	7	http://www.esebedesign1.altervista.org/
2º RESULTADO	3	9	http://esebedesign2.altervista.org/
CADENA 5: marketing online Alicante			
1º RESULTADO	9	6	http://www.esebedesign1.altervista.org/
2º RESULTADO	9	8	http://esebedesign2.altervista.org/
CADENA 6: "diseño web profesional Alicante"			
1º RESULTADO	1	2	http://www.esebedesign.altervista.org/
2º RESULTADO	-	-	Resultados agotados
CADENA 7: "diseñador web Alicante"			
1º RESULTADO	1	2	http://www.esebedesign.altervista.org/
2º RESULTADO	1	4	http://www.esebedesign1.altervista.org/
CADENA 8: "diseñador gráfico Alicante"			
1º RESULTADO	1	3	http://www.esebedesign.altervista.org/
2º RESULTADO	2	1	http://www.esebedesign1.altervista.org/
CADENA 9: "páginas web Alicante"			
1º RESULTADO	9	4	http://www.esebedesign.altervista.org/
2º RESULTADO	-	-	Resultados agotados
CADENA 10: Solicitar presupuesto página web Alicante			
1º RESULTADO	1	9	http://www.esebedesign.altervista.org/contacto/contatti/pedir-presupuestos.html
2º RESULTADO	-	-	Resultados agotados
CADENA 11: Solicitar presupuesto página web y proyectos gráficos Alicante			
1º RESULTADO	1	1	http://www.esebedesign.altervista.org/contacto/contatti/pedir-presupuestos.html
2º RESULTADO	-	-	Resultados agotados
CADENA 12: realización páginas web Alicante			
1º RESULTADO	1	2	http://www.esebedesign.altervista.org/
2º RESULTADO	1	4	http://www.esebedesign1.altervista.org/
CADENA 13: realización campañas SEM Alicante			
1º RESULTADO	1	1	http://www.esebedesign.altervista.org/
2º RESULTADO	1	2	http://www.esebedesign1.altervista.org/
CADENA 14: realización campañas SEO Alicante			
1º RESULTADO	1	1	http://www.esebedesign.altervista.org/
2º RESULTADO	1	2	http://www.esebedesign1.altervista.org/
CADENA 15: campañas SEO Alicante			
1º RESULTADO	1	3	http://www.esebedesign.altervista.org/
2º RESULTADO	8	6	http://esebedesign2.altervista.org/
CADENA 16: campañas SEM Alicante			
1º RESULTADO	1	3	http://www.esebedesign.altervista.org/
2º RESULTADO	14	8	http://www.esebedesign1.altervista.org/
CADENA 17: "campañas SEM Alicante"			
1º RESULTADO	1	3	http://www.esebedesign.altervista.org/
2º RESULTADO	14	8	http://www.esebedesign1.altervista.org/
CADENA 18: Creación proyectos gráficos alicante			
1º RESULTADO	1	2	http://www.esebedesign1.altervista.org/
2º RESULTADO	1	3	http://esebedesign2.altervista.org/
CADENA 19: campañas SMO Alicante			
1º RESULTADO	1	2	http://www.esebedesign.altervista.org/
2º RESULTADO	2	7	http://www.esebedesign1.altervista.org/
CADENA 20: realización aplicaciones web Alicante			
1º RESULTADO	1	2	http://www.esebedesign1.altervista.org/

2°RESULTADO	-	-	Resultados agotados
CADENA 21: servicios proyectos gráficos y paginas web			
1°RESULTADO	1	2	http://www.esebedesign1.altervista.org/
2°RESULTADO	1	8	http://www.esebedesign2.altervista.org/servicios-web-graficos-marketing.html
CADENAS 22-23-24-25-26-27-28-29-30: posicionamiento seo Alicante; "posicionamiento seo Alicante"; diseño web profesional Alicante; diseñador web Alicante; páginas web Alicante; "diseño web Alicante"; diseño web Alicante; "campanas SEO Alicante"; diseñador gráfico Alicante			
1°RESULTADO	-	-	Resultados agotados
2°RESULTADO	-	-	Resultados agotados

Leyenda: "P", Página de los resultados de búsqueda. "#", posición dentro de la página P.

Tabla 6: Resultados de posicionamiento de palabras clave, tercera toma de datos. Elaboración Propia.
Fecha de la tercera toma de datos: de 25 de Junio de 2013 a 30 de Junio de 2013.