



UNIVERSIDAD DE VALLADOLID

ESCUELA TÉCNICA SUPERIOR  
INGENIEROS DE TELECOMUNICACIÓN

TRABAJO FIN DE MASTER

MASTER UNIVERSITARIO EN INVESTIGACIÓN  
EN TECNOLOGÍAS DE LA INFORMACIÓN Y LAS COMUNICACIONES

# **ESTUDIO DEL SOPORTE DE SEMÁNTICA EN GESTORES DE CONTENIDOS**

Autor:

**D. Alfredo Carrión García**

Tutor:

**Dra. Dña. Mercedes Martínez**

Valladolid, 10 de Septiembre de 2014

---

TÍTULO: **ESTUDIO DEL SOPORTE DE SEMÁNTICA EN GESTORES DE CONTENIDOS**

AUTOR: **D. Alfredo Carrión García**

TUTOR: **Dra. Dña. Mercedes Martínez**

DEPARTAMENTO:

---

### **Tribunal**

---

PRESIDENTE: **Dr. D. Pablo de la Fuente**

VOCAL: **Dr. D. Juan Ignacio Asensio**

SECRETARIO: **Dr. D. Eduardo Gómez**

---

FECHA: **10 de Septiembre de 2014**

CALIFICACIÓN:

---

### **Resumen del TFM**

Uno de los aspectos más importantes de cara a progresar hacia la Web Semántica es convertir el contenido web nuevo y el existente, expresándolo en lenguaje natural, en su equivalente semántico.

El marcado semántico de documentos Web es el primer paso hacia la adaptación del contenido a la Web Semántica y por tanto, los formatos de marcado son una de las herramientas consideradas como fundamentales en el diseño de la Web Semántica, ya que no sólo permiten acceder a la información sino que además definen su semántica, facilitando por tanto su procesamiento automático y posterior reutilización para distintas aplicaciones.

Los Sistemas de Gestión de Contenidos Web han tener capacidad para reutilizar los datos que almacenan, automatizando el etiquetado semántico con el fin de hacer el contenido procesable por máquinas y disponer de capacidad para obtener datos RDF de fuentes externas para enriquecer sus contenidos.

Este estudio presenta una propuesta metodológica que permite evaluar y comparar el soporte de marcado semántico en distintos Sistemas de Gestión de Contenidos, protagonistas —sin lugar a duda— en la transición hacia la Web Semántica.

### **Palabras clave**

Sistema de Gestión Contenidos, CMS, RDFa, Microdatos, Microformatos, Web Semántica, metodología de evaluación, Marcado Semántico, Drupal, Wordpress, Joomla.

## **Abstract**

One of the most important aspects for progress to the Semantic Web is to articulate, in natural language, the new and the existing web content, in its semantic equivalent.

Semantic markup of Web documents is the first step to adapting the content to the Semantic Web. The markup formats are one of the most important tools in the design of the Semantic Web, because they provide access information defining the semantics, making possible automatic processing and reusability for different applications.

The Content Management Systems must reuse their stored data, automating the semantic tagging in order to make the content processable by machines and have the capacity for RDF data from external sources to enrich their content.

This project shows a proposal to evaluate and compare the semantic markup support in Content Management Systems, one of the keys in the transition to the Semantic Web.

## **Keywords**

Semantic Web, Content Management Systems, CMS, RDFa, Microdata, Microformats, Evaluation Methodology, Drupal, Joomla, Wordpress, Semantic Markup





# Agradecimientos

En este punto quiero agradecer, en especial, a mi tutora la Dra. Dña. M<sup>a</sup>. Mercedes Martínez, así como al Coordinador del Máster, el Dr. D. Pablo de la Fuente, y a los compañeros de promoción y resto del profesorado, el apoyo en la realización del Máster y todas las aportaciones y buenos momentos compartidos.



# Índice general

<b>1. Introducción</b>	<b>1</b>
1.1. Motivación . . . . .	1
1.2. Estructura de la memoria . . . . .	5
<b>2. Conocimientos Previos</b>	<b>7</b>
2.1. La Web Semántica . . . . .	7
2.1.1. Conceptos Generales . . . . .	7
2.1.2. Mercado Semántico . . . . .	9
2.2. Sistemas de Gestión de Contenidos . . . . .	18
2.2.1. Conceptos Generales . . . . .	18
2.2.2. Estado del Arte . . . . .	20
2.3. Soporte Semántico en CMS . . . . .	31
<b>3. Planteamiento del Estudio</b>	<b>35</b>
3.1. Objetivos de la Evaluación . . . . .	35
3.2. Metodología de la Evaluación . . . . .	35
<b>4. Fases de Planificación y Experimentación</b>	<b>41</b>
4.1. Aplicación de los Criterios de Evaluación . . . . .	41
4.2. Validación del Mercado . . . . .	46
<b>5. Resultados de la Evaluación</b>	<b>51</b>
5.1. Descripción y funcionalidades de las herramientas analizadas . . . . .	51
5.1.1. AlchemyTagger para Wordpress . . . . .	51
5.1.2. RDFaCE para Wordpress . . . . .	51
5.1.3. Schema Creator para Wordpress . . . . .	52
5.1.4. Schema.org para Drupal . . . . .	52
5.1.5. Open Calais para Drupal . . . . .	52
5.1.6. COM J4Schema para Joomla . . . . .	52
5.1.7. Krizalys Breadcrumb para Joomla . . . . .	53
5.2. Tipología de Mercado Semántico . . . . .	54
5.3. Integridad del Software . . . . .	55
5.4. Formatos de Mercado Semántico Admitidos . . . . .	55
5.5. Soporte en diferentes CMS . . . . .	57
<b>6. Conclusiones y Trabajo Futuro</b>	<b>59</b>



# Índice de figuras

1.1.	Los CMS en el global de la red . . . . .	2
1.2.	Estimación del Avance de Wordpress . . . . .	3
2.1.	Propuesta de Web Sintáctica vs. Web Semántica . . . . .	8
2.2.	Mapa conceptual de la Web Semántica. . . . .	10
2.3.	Tabla comparativa Formatos de Marcado Semánticos. . . . .	16
2.4.	Estadísticas de uso de los CMS según BuiltWith 2014 . . . . .	20
2.5.	Tendencia Temporal de los 3 CMS a estudio en Trends . . . . .	21
2.6.	Comparativa de los Requisitos de Sistema en CMS . . . . .	22
2.7.	Comparativa de Seguridad en CMS . . . . .	23
2.8.	Comparativa de Rendimiento en CMS . . . . .	23
2.9.	Comparativa de Soporte en CMS . . . . .	24
2.10.	Comparativa de Facilidad de uso en CMS . . . . .	25
2.11.	Comparativa de Interoperabilidad en CMS . . . . .	26
2.12.	Comparativa de Administración en CMS . . . . .	26
2.13.	Comparativa de Flexibilidad en CMS . . . . .	27
2.14.	Repercusión de Wordpress en 2013. . . . .	28
2.15.	Sistemas de Gestión de Contenidos Semánticos. . . . .	32
3.1.	Metodología de benchmarking de software . . . . .	37
4.1.	Vista del Backend de Alchemy para Wordpress . . . . .	43
4.2.	Vista del editor de Wordpress con RDFaCE integrado . . . . .	43
4.3.	Vista del backend de Schema para Wordpress . . . . .	44
4.4.	Vista de creación de Person mediante Schema.org para Drupal . . . . .	45
4.5.	Backend del plugin Open Calais para Wordpress . . . . .	46
4.6.	Esquema del Funcionamiento de Open Calais . . . . .	47
4.7.	Vista del backend de J4Schema para Joomla! . . . . .	48
4.8.	Vista del backend de Krizalys Breadcrumbs para Joomla! . . . . .	48
4.9.	Vista del validador de marcado de Yandex . . . . .	49
4.10.	Vista de Editor de Texto de Wordpress con RDFaCE integrado . . . . .	49
4.11.	Validación de la entrada de RDFaCE en Wordpress con Yandex . . . . .	50
5.1.	Tabla Comparativa de Funcionalidades esperadas por herramienta . . . . .	54
5.2.	Tabla Comparativa de Tipologías de Marcado por herramienta . . . . .	55
5.3.	Tabla Comparativa de la Integrabilidad de cada herramienta . . . . .	56
5.4.	Tabla Comparativa de Formatos de Marcado admitidos por herramienta. . . . .	56

5.5. Tabla Comparativa del soporte en diferentes CMS . . . . . 57

# Capítulo 1

## Introducción

### 1.1. Motivación

Tim Berners-Lee propuso la Web Semántica como “una extensión de la Web actual en la que la información tiene un significado bien definido, facilitando a los ordenadores y a las personas trabajar en cooperación” [1].

Añadir semántica explícita a la Web —en forma de metadatos—, implica permitir documentos que tienen información en formatos comprensibles por máquinas, además de permitir que los enlaces se creen con valores en las relaciones. Se basa en la idea de la integración y reutilización de los datos en la Web estando definidos y enlazados entre ellos pudiendo ser compartidos y procesados por aplicaciones de manera automática al mismo tiempo que por personas [2]

Según Berners Lee[3], los datos presentes en la web pueden estar distribuidos y ser heterogéneos.

La Web Semántica se considerada también la Web de Datos [4] y se presenta como una extensión de la Web actual en la que a la información disponible se le otorga una semántica bien definida, permitiendo que aplicaciones heterogéneas descubran y utilicen la información presente en la Web.

El mecanismo utilizado para representar el conocimiento en la Web Semántica son las ontologías [5].

Por otro lado, los Sistemas de Gestión de Contenidos (CMS) han alcanzado, durante los últimos años, un alto grado de desarrollo conceptual desde el punto de vista de la arquitectura de la información.

Asimismo, los Sistemas de Gestión de Contenidos se proponen como una solución que ofrece características como la reutilización de contenidos la posibilidad de interconexión con bases de datos muy diversas y un alto grado de modularidad.

La disponibilidad en abierto de muchas de estas aplicaciones y su facilidad de instalación y manejo han propiciado un uso masivo.

Como se aprecia en la Figura 1.1 (realizada a partir de la infografía desarrollada por Moove <sup>1</sup>) un 32 % de los sitios web ya son Sistemas de Gestión de Contenidos y la previsión de crecimiento de éstos en los dos próximos años es elevadísima.

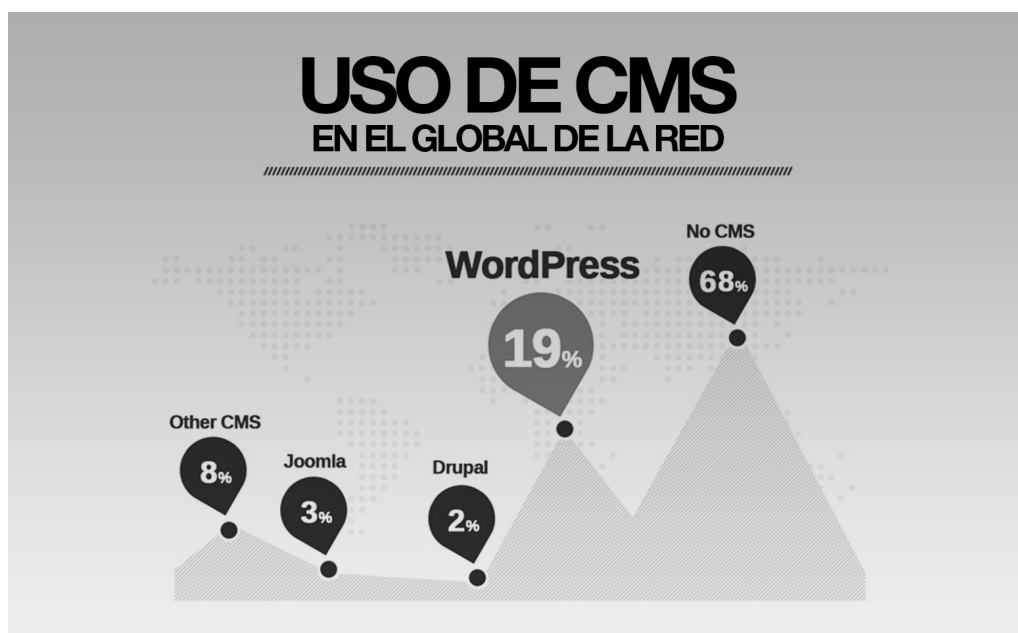


Figura 1.1: Los CMS en el global de la red

En particular, se puede apreciar la tendencia prevista para Wordpress en la figura 1.2 (realizada igualmente a partir de la infografía desarrollada por Moove)

En la Biblia de la Gestión de Contenidos[6] se define, en general, CMS (Content Management System) como una aplicación informática que se usa para editar, administrar, publicar y realizar búsquedas sobre distintos tipos de medios digitales.

Parece necesario, teniendo en cuenta la gran aceptación de este estándar, adaptar las pautas de la Web Semántica a los gestores de contenidos ya que, de manera recíproca, existirá una gran implicación en el éxito de ambos.

Como plantea Pastor-Sánchez[7], se observa una clara tendencia al desarrollo de una infraestructura de metadatos para todo tipo de sistemas presentando una convergencia entre éstos y otros sistemas (como los sistema de gestión de contenidos) en la Web Semántica, dando lugar a nuevas funciones y modos de localización, consulta y reutilización de contenidos a través de estructuras de metadatos dentro de un entorno de estándares y datos

<sup>1</sup><http://mooveagency.com>





Figura 1.2: Estimación del Avance de Wordpress

abiertos.

El presente estudio se centra en la definición de Sistemas de Gestión de Contenidos Web (Web Content Management System) como una aplicación web diseñada para facilitar la gestión de sitios web a usuarios sin necesidad de un nivel alto de conocimientos técnicos, facilitando el proceso de autoría de contenidos, mejorando los tiempos de creación y actualización y la consistencia de las publicaciones, así como otros factores clave como la navegabilidad, flexibilidad, seguridad y dimensionamiento, todos ellos, bajo el paraguas de una reducción de costes muy grande.

Vistos los datos de uso de los Sistemas de Gestión de Contenidos en la totalidad de la web, es posible afirmar que una gran parte de los datos publicados en internet tienen como origen un WCMS —con una clara tendencia al alza— y por tanto se presume como fundamental implementar en ellos este tipo de tecnologías y estándares abiertos.

Los Sistemas de Gestión de Contenidos han tener capacidad para reutilizar los datos que almacenan, automatizando el etiquetado semántico con el fin de hacer el contenido procesable por máquinas y disponer de capacidad para obtener datos RDF de fuentes externas para enriquecer sus contenidos.

La motivación principal de este trabajo es, por tanto, evaluar y comparar el soporte de semántica que presentan los principales Sistemas de Gestión de Contenidos, usando una metodología que guíe la evaluación y comparación futura, centrando el estudio en el soporte dado en el ámbito del marcado semántico.

El marcado semántico de documentos es el primer paso para permitir el procesamiento automático de la Web y, por tanto para la creación de la Web Semántica [8].

Como se verá más adelante, los lenguajes de marcado son consideradas como herramientas fundamentales en el diseño de la web semántica, ya que no sólo permiten acceder a la información sino que además definen su semántica, facilitando por tanto su procesamiento automático y posterior reutilización para distintas aplicaciones.

## 1.2. Estructura de la memoria

La presente memoria se estructura en cinco apartados principales:

1. En el primero se presenta una breve motivación apoyada en el estado actual de los principales actores del estudio para proseguir con el objeto marcado inicialmente.
2. Posteriormente se repasan conceptos esenciales de cada una de las áreas tratadas, indicando el estado del arte del estudio para facilitar la comprensión del resto del trabajo. En este punto el estudio se centra en los dos actores a escena. Por un lado la Web Semántica y por otro los Sistemas de Gestión de Contenidos (CMS).
3. El tercer apartado muestra el planteamiento general del estudio presentando los objetivos de la evaluación, así como la metodología seguida y la definición de los criterios a estudio.
4. En el cuarto punto se presentan la experimentación, donde se define la muestra de CMS a estudio, así como las herramientas disponibles que componen su soporte semántico sobre el que se realizará la evaluación. En este punto se realiza la evaluación sobre los criterios previamente planteados así como la comparativa entre las diferentes herramientas a estudio.
5. A continuación se presentan los resultados de la evaluación sobre las diferentes herramientas a estudio a través de tablas y gráficas.
6. Para finalizar el estudio, previo a la presentación de referencias bibliográficas, se propondrán algunas conclusiones arrojadas por la investigación, así como posibles líneas de trabajo futuro relacionadas con el trabajo actual.



# Capítulo 2

## Conocimientos Previos

### 2.1. La Web Semántica

#### 2.1.1. Conceptos Generales

Según la definición de la W3C<sup>1</sup> “La Web Semántica es una web extendida, dotada de mayor significado, en la que cualquier usuario en Internet podrá encontrar respuestas a sus preguntas de forma más rápida y sencilla gracias a una información mejor definida. Al dotar a la web de más significado y, por tanto, de mayor grado de semántica, se pueden obtener soluciones a problemas habituales en la búsqueda de información gracias a la utilización de una infraestructura común, mediante la cual, es posible compartir, procesar y transferir información de forma sencilla. Esta web extendida y basada en el significado, se apoya en lenguajes universales que resuelven los problemas ocasionados por una web carente de semántica en la que, en ocasiones, el acceso a la información se convierte en una tarea difícil y frustrante”.

Por tanto, la Web Semántica no es una nueva Web segregada de la actual. Es una extensión de la actual en la que la información se ofrece con un significado bien definido, permitiendo a ordenadores y personas trabajar de forma cooperativa [1].

La idea que existe detrás de la Web Semántica es tener datos en la Web definidos y enlazados de manera que puedan ser usados de forma más efectiva para un descubrimiento, una automatización, una integración y una reutilización entre diferentes aplicaciones. Para ello la Web debe evolucionar, ofreciendo una plataforma accesible que permita que los datos se compartan y se procesen por herramientas automatizadas o personas. Añadir semántica explícita (metadatos) a la Web implica permitir documentos que tienen información en formatos comprensibles por máquinas, además de permitir que los enlaces se creen con valores en las relaciones.

El reto de la Web Semántica es ofrecer el lenguaje que exprese tanto datos como reglas para razonar sobre los datos, y además permita que las reglas sobre cualquier sistema de representación del conocimiento sean exportadas a la Web. [11].

---

<sup>1</sup><http://www.w3c.es/Divulgacion/GuiasBreves/WebSemantica>

La web Semántica está situada en la confluencia de la inteligencia artificial y las tecnologías web. Este enfoque propone el enriquecimiento estructural de la información añadiendo componentes semánticos que puedan procesarse automáticamente. Formatos de la web semántica liderados por XML (extensible Markup Language) y RDF (Resource Description Framework), los cuales incluyen ontologías para la especificación de reglas lógicas que permitan a los agentes reconocer y clasificar conceptos. [12].

La figura 2.1 muestra la propuesta de [13] donde la web sintáctica se representa como un grafo conformado por nodos del mismo tipo mientras la Web Semántica se presenta como un grafo con nodos (recursos) de diferentes tipos relacionados explícitamente mediante arcos.

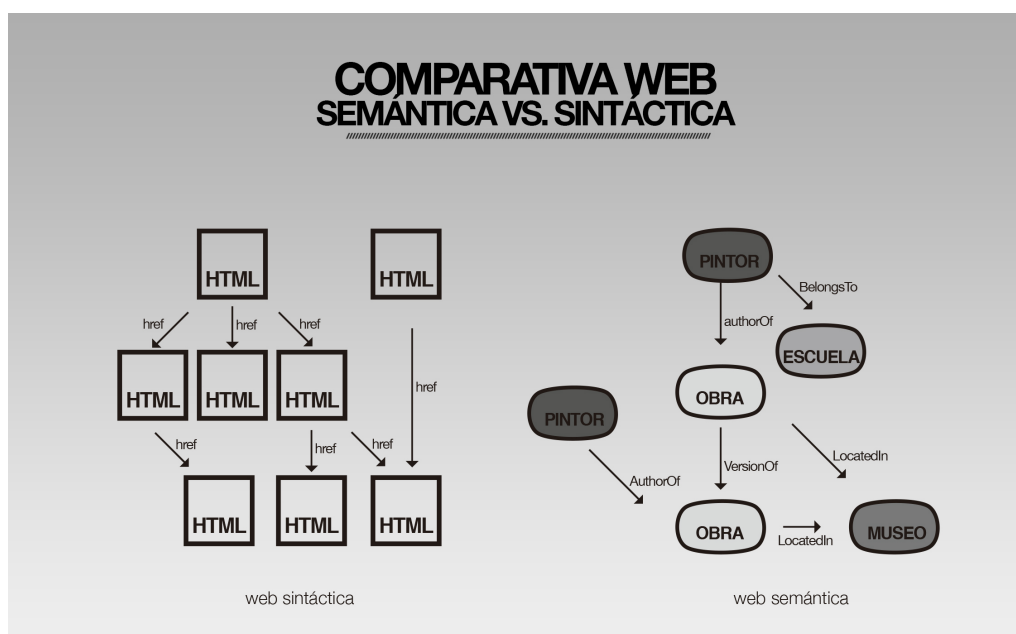


Figura 2.1: Propuesta de Web Sintáctica vs. Web Semántica

Diferentes sistemas web pueden utilizar diferentes identificadores para un mismo concepto. Así, una aplicación que quiera comparar o combinar información entre dichos sistemas debe conocer qué términos son sinónimos. Idealmente, la aplicación deberá tener capacidad de descubrir los significados comunes de cualquier base de datos que encuentre. Una solución a este problema es incluir un nuevo elemento a la Web Semántica, colecciones de información denominadas ontologías.

Las ontologías son herramientas conceptuales que definen un vocabulario común para permitir la compartición de información dentro de un determinado dominio. Esto incluye definiciones de los conceptos básicos del dominio, así como sus relaciones, que tienen que ser interpretables por máquinas. [14].

En la Web se publican grandes conjuntos de datos de una manera estructurada siguiendo dichas pautas, al tiempo que se definen de forma explícita las relaciones entre recursos mediante uno o varios vocabularios, como por ejemplo RDF, OWL (web ontology language ) o SKOS (simple knowledge organization system) entre otros [15].

Datasets y contenidos web deberían dar forma a una realidad en la que los procesos de publicación partieran de la interrelación y la consiguiente integración de grandes cantidades de datos y contenidos informativos [16].

Como puede apreciarse en la figura 2.2, alrededor de la Web Semántica, han surgido innumerables tipos de aplicaciones que permiten crear y procesar el conocimiento, presentando cada una sus propios modelos y técnicas de conocimiento e inferencia.

El éxito de la Web Semántica pasa, entre otros factores, por la interoperabilidad de todas ellas siendo capaces de intercambiar información para poder utilizarla.

### 2.1.2. Mercado Semántico

En 1996, el W3C comenzó el desarrollo del lenguaje de marcado XML (eXtensible Markup Language) [17] que separa la visualización del contenido. XML se diseñó para resolver algunas limitaciones del HTML, permitiendo introducir etiquetas que describen la estructura del contenido de la Web. No obstante, XML no proporciona semántica a dichas etiquetas, con lo cual las aplicaciones informáticas presentan ciertas dificultades en la comprensión del significado de una etiqueta codificada en XML.

La siguiente generación de la Web, conocida como la Web Semántica [18] surge para paliar este problema describiendo los recursos disponibles en la Web, haciéndolos comprensibles por una máquina, con independencia del idioma en el que se encuentren escritos.

Los lenguajes de marcado, permiten la inserción del vocabulario procedente de una ontología dentro de los documentos Web. Estos lenguajes se basan en paradigmas de la representación del conocimiento en inteligencia artificial [5].

Los lenguajes de marcados más representativos son los recomendados por el W3C y son:

- RDF (Resource Description Framework) [19], desarrollado por el W3C como un lenguaje basado en redes semánticas que describe recursos.
- RDF Schema [20], define una extensión semántica de RDF, proporcionando mecanismos para describir grupos de recursos y sus relaciones.
- OWL (Ontology Web Language) [21], creado por el W3C como una extensión de RDF Schema proporcionando vocabulario adicional y una semántica formal.

Se concibe así, como se ha visto previamente, la Web Semántica como una extensión de la Web actual, en la que a la información disponible se le ‘marca’ una semántica bien



Figura 2.2: Mapa conceptual de la Web Semántica.

definida (proporcionada por las ontologías), permitiendo que aplicaciones heterogéneas descubran y utilicen información presente en la Web.

A su vez, según García Jiménez[22], las herramientas para la elaboración de las ontologías se subdividen en:

1. **Herramientas de desarrollo de ontologías:** este grupo incluye las herramientas que sirven para la construcción de nuevas ontologías o bien para la reutilización de las existentes. Destacan entre sus funcionalidades la edición y la consulta, así como la exportación e importación de ontologías, la visualización en diversos formatos gráficos, etcétera.
2. **Herramientas de fusión e integración de ontologías:** pretenden solucionar el problema de la combinación y la integración de diversas ontologías de un mismo do-



minio, lo que ocurre cuando se unen dos organizaciones diferenciadas, o cuando se pretende obtener una ontología de calidad, a partir de las existentes.

3. **Herramientas de evaluación de ontologías:** aparecen como instrumentos de apoyo que deben asegurar que tanto las ontologías como las tecnologías relacionadas tengan un nivel mínimo de calidad. Para el futuro, este esfuerzo puede también conducir a las certificaciones estandarizadas.
4. **Herramientas basadas en la anotación:** estas herramientas se han diseñado para permitir a los usuarios insertar informaciones y datos. La mayoría de estas herramientas han aparecido recientemente, junto con la aparición de la idea de la web semántica.
5. **Herramientas de almacenamiento y preguntas:** son instrumentos que se han creado para permitir utilizar con facilidad las ontologías. La clave está en el intento de que el web se convierta en una auténtica plataforma para transmitir conocimiento.
6. **Herramientas de aprendizaje:** se utilizan semiautomáticamente para construir ontologías a partir del lenguaje natural.

El foco del presente estudio, como se verá más adelante, serán las Herramientas Basadas en la anotación.

Como se ha indicado con anterioridad, uno de los aspectos más importantes de cara a progresar hacia la Web Semántica es cómo convertir el contenido Web nuevo y el existente, expresándolo en lenguaje natural, en su equivalente semántico en el que el contenido se enriquece con metadatos en formato estructurado. El marcado semántico de documentos Web es el primer paso hacia la adaptación del contenido Web a la Web Semántica [8].

El marcado semántico de contenidos supone la incorporación explícita de información a documentos web, de forma que éstos sean inteligibles y procesables también por aplicaciones informáticas y es el primer paso para permitir el procesamiento automático de la Web y, por tanto para la creación de la Web Semántica [8].

Por tanto, los lenguajes de marcado son una de las herramientas consideradas como claves en el diseño de la Web Semántica, ya que no sólo permiten acceder a la información sino que además definen su semántica, facilitando por tanto su procesamiento automático y posterior reutilización para distintas aplicaciones.

El enriquecimiento semántico se hace posible mediante el etiquetado del contenido web con metadatos, los cuales posibilitan describir las entidades que se encuentran en el contenido y las relaciones entre ellas [23].

Como se ha visto previamente, en el contexto de la web semántica, la anotación semántica se lleva a cabo mediante lenguajes de marcado como XML, RDF o OWL, adjuntando ciertos elementos del texto en una etiqueta. [24].

Proporcionar un significado bien definido a los elementos que actualmente forman la Web posibilita, entre otras cosas, mejorar las capacidades de búsqueda contextual e incrementar la interoperabilidad de sistemas en contextos colaborativos [25]. Sin embargo, la mayoría del contenido de la Web permanece desestructurado debido a la dificultad que supone su marcado.

Con independencia del esquema de descripción utilizado, los principales formatos usados en el mercado semántico son:

- Microformatos
- RDFa
- Microdatos

Así se desprende en el estudio cuantitativo de Bizer[26] en el que se demuestra numéricamente que RDFa, microdatos, microformatos tienen una considerable adopción en la web y están siendo utilizados por cientos de miles de sitios web.

Además en estudio presenta una correlación entre los sitios web que proporcionan datos estructurados y su popularidad la lista Alexa, razón suficiente para pensar que el éxito del crecimiento de la Web Semántica vendrá dado por su estrecha relación con el posicionamiento de las webs y las implicaciones económicas que esto conlleva.

### **Microformatos**

Los microformatos son un conjunto de formatos simples de datos abiertos no estándar contruidos sobre estándares. En su diseño se ha optado por la legibilidad de los humanos sobre las máquinas.

Algunos estándares en los que se basan los microformatos son vCard o iCalendar. Son pues, una forma de aplicar nomenclaturas y convenciones simples en atributos HTML ya existentes para definir la semántica de dichos datos [27].

La gran aceptación inicial de este tipo de formato se ha debido a su sencillez de aplicación ya que presenta limitaciones en su modelo de datos (exclusivamente jerárquico) y características como la imposibilidad de definir recursos referenciables mediante URIs y la dificultad para definir nuevos vocabularios o combinar otros ya existentes [28].

A continuación se presenta una representación de código de utilizando este tipo de marcado y podrá compararse con el resto de tipos, en siguientes apartados.

```
<dl class="dublincore">
<dt>Título:</dt>
<dd class="title">TFM MUITIC UVA </dd>
<dt>Dirección:</dt>
<dd><a href="http://wordpress.ochoytres.com"
```

```

class="identifier">http://wordpress.ochoytres.com</a></dd>
<dt>Descripción:</dt>
<dd class="description">Marcado Semántico en CMS.</dd>
<dt>Palabras clave:</dt>
<dd class="subject">Microdatos, Microformatos, RDFa</dd>
<dt>Código de idioma:</dt>
<dd class="language">es</dd>
<dt>Autor/a:</dt>
<dd><a href="http://www.ochoytres.com" class="creator">Alfredo Carrión<
</dl>

```

### Microdatos

Los Microdatos son una propuesta para HTML5 cuyo objeto es crear un esquema de datos estructurados que sea soportado por los principales motores de búsqueda.

Los microdatos estructuran la información de la siguiente forma:

- Identifican nuevas entidades mediante itemscope
- Las propiedades son identificadas con itemprop
- Los tipos con itemtype

En general, los microdatos son considerados como un formato de marcado semántico de complejidad y expresividad media.

Google pese a reconocer las propiedades de microformatos, de RDFa y de microdatos para el marcado de semántico, apuesta por esta última opción como estándar apoyado en su relación con Schema.org <sup>2</sup>, iniciativa respaldada por parte de los principales buscadores.

Schema.org es un diccionario creado en asociación con los principales buscadores, GOOGLE, YAHOO, BING y el buscador ruso Yandex [29], para buscar, entre otras cosas un standard común. Gran parte del vocabulario de Schema.org está basado en otros existentes como FOAF, GoodRelations, OpenCyC, rNews... [30].

Algunos autores dan una definición más amplia de Schema.org y la llaman: "Mega Ontología.<sup>en</sup> forma de esquema de carácter general que intenta representar una amplia variedad de hechos: obras, eventos, intangibles, organizaciones, personas, lugares y productos [7].

```

<div itemscope itemtype="http://schema.org/BlogPosting">
Título <span itemprop="name">TFM MUITIC UVA </span>
Dirección <span itemprop="url">http://wordpress.ochoytres.com</span>
Descripción

```

---

<sup>2</sup>ttp://www.schema.org

```

<span itemprop="Description">Marcado Semántico en CMS.</span>
Keywords
<span itemprop="Keywords">Microdatos, Microformatos, RDFa</span>
Idioma <span itemprop="inLanguage">Sp</span>
Autor <span itemprop="creator">Alfredo Carrión</span>
</div>

```

## RDFa

El presente estudio se centra, principalmente, en RDFa, una de las soluciones del W3C<sup>3</sup> para el marcado semántico, fundamentada en una total compatibilidad con la tecnología clave de la web semántica: RDF. Su última versión (RDFa 1,1) utiliza un conjunto de atributos específicos y la reutilización de otros ya existentes para la inclusión de información semántica.

La especificación contempla su uso en diferentes lenguajes “anfitrión” como xhtml (versiones 1 y 5), html (versiones 4 y 5), xml, SVG (scalable vector graphics), ePUB y Open-Document.

La potencia de RDFa contrasta con cierta complejidad en su aplicación. Su creación ha precisado de la definición de nuevos atributos y la reutilización de otros. Por este motivo, el W3C desarrolló la recomendación RDFa Lite<sup>4</sup>, que indica cómo utilizar RDFa de un modo más simplificado, cubriendo las necesidades más comunes de marcado semántico.

La propuesta de RDFa consiste en codificar tripletas RDF dentro del HTML:

- El sujeto se especifica con el atributo about
- El predicado a través de property, rel o rev
- Las URIs se especifican mediante href, resource o src
- Los literales se especifican mediante content y en algunos casos mediante el atributo datatype que aporta el tipo de dato.

Pastor-Sánchez, J. A [16] propone a RDFa como una alternativa flexible a la par que potente para la inclusión en una página web de cualquier tipo de sentencia RDF.

RDFa fue originalmente diseñado para ser usado únicamente con XHTML. Sin embargo, actualmente, la última versión, RDFa Lite 1,1 funciona para cualquier lenguaje basado en XML y sobre todo con HTML5. RDFa Lite 1,1 usa una mezcla de atributos HTML, como href y src, y nuevos atributos, creados para solucionar algunos problemas específicos. [31].

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).

<sup>3</sup><http://www.w3.org/TR/rdfa-syntax>

<sup>4</sup><http://www.w3.org/TR/rdfa-lite>

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.

A continuación se presenta una representación de código de utilizando este tipo de marcado.

```
<div xmlns:v="http://rdf.data-vocabulary.org/#"
typeof="v:Person">
<span property="v:title">TFM MUITIC UVA </span>, que está en:
<span property="v:url">http://wordpress.ochoytres.com</span>
recoge información donde
<a href="http://www.example.com"
rel="v:description">Marcado Semántico en CMS</a>.
Está escrito en
<span property="v:language">español</span> por
<span property="v:creator">Alfredo Carrión</span>
```

En el estudio comparativo de M. Sporny[32] se propone una comparación exhaustiva de los formatos de marcado semántico anteriormente citados, resumida a modo tabla en la figura 2.3.

Según los principales 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, 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 paginas web enriquecidas con datos estructurados.

Uno de los beneficios más importantes aportado por este proyecto es la mejora de la indexación facilitando el posicionamiento orgánico de paginas web, debido a que se proporciona un sistema único y compartido de marcado.

La aplicación de Schema.org puede conducir, según [7], a una situación de divergencia tecnológica, debido al uso de microdatos en vez de la aplicación de RDFa. De esta forma podría reducirse la capacidad de interoperabilidad entre ambas tecnologías, puesto que la paginas web diseñadas con marcado XHTML no pueden funcionar con los microdatos, mientras que la paginas que incorporan el marcado HTML5 no contemplaban, hasta hace poco, el uso de RDFa.

Feature	RDFa 1.1	Microdata 1.0	Microformats 1.0
Relative Complexity	High	Medium	Low
Data Model	Graph	Tree	Tree
Item optionally identified by IRI	Yes	Yes	No
Item type optionally specified by IRI	Yes	Yes	No
Item properties specified by IRI	Yes	Yes	No
Multiple objects per page	Yes	Yes	Yes
Overlapping objects	Yes	Yes	No
Plain Text properties	Yes	Yes	Yes
IRI properties	Yes	Yes*	No
Typed Literal properties	Yes	No	No
XML Literal properties	Yes	No	No
Language tagging	Yes	Yes	Inconsistent
Override text and IRI content	Yes	No	Text only
Clear mapping to RDF	Yes	Problematic	No
Target Languages	8 (XHTML1, HTML4, HTML5, XHTML5, XML, SVG, ePub, OpenDocument)	2 (HTML5, XHTML5)	4 (XHTML1, HTML4, HTML5, XHTML5)
New Attributes	8 about, datatype, profile, prefix, property, resource, typeof, vocab	5 itemid, itemprop, itemref, itemscope, itemtype	0
Re-used Attributes	5 content, href, rel, rev, src	5 content, src, href, data, datetime	4 class, title, rel, href
Multiple IRI types per object	Yes	RDF only	No
Multiple statements per element	Yes	No	Yes
“Locally scoped” vocabulary terms	Yes, via vocab	Yes, via itemscope	No
Item Chaining	Yes	Basic	No
Transclusion	No	Yes	Yes, via include pattern
Compact IRIs	Yes	No	No
Prefix rebinding	Yes	No	No
Vocabulary Mashups	Yes	No	No
HTML5 time element support	Not yet	Yes	No
Different attributes for different property types	Yes property for text, rel/rev for URLs, resource/content for overrides	No	Yes class for text and rel for URLs
Transform to JSON	Yes (RDFa API)	Yes (Parser and Microdata DOM API)	No
DOM API	Yes	Yes	No
Unified Parser	Yes	Yes	No

Figura 2.3: Tabla comparativa Formatos de Marcado Semánticos.

El mayor inconveniente a la hora de marcar datos estructurados con microdatos es su implantación, ya que la definición del elemento se realiza embebiendo dentro del HTML en diferentes lugares con los problemas que puede conllevar.

```

<div itemscope itemtype="http://schema.org/Event">
  El evento deportivo del año en España, la
  <span itemprop="name">Final del Mundial de Baloncesto</span>
  en directo el
  <span itemprop="startDate"
  content="2014-09-14">14 de Septiembre de 2014</span>
  en
  <span itemprop="location" itemscope
    itemtype="http://schema.org/Place">
    <span itemprop="name">Madrid</span>
  </span>
</div>

```

En la parte final de la realización del estudio se tuvo constancia de una nueva recomendación del W3C <sup>5</sup>, el formato JSON-LD. Publicada por el W3C el 16 de enero de 2014 y adoptada por Google desde hace unas semanas como alternativa a los microdatos para implementar en sitios web. La documentación aportada por Google, recomienda el uso de una tecnología únicamente, recomendando —por el momento— el uso de microdatos en lugar de JSON-LD.

Según la información sobre JSON-LD —escasa en el momento del estudio—, parece que solucionaría el inconveniente dado por la realización del marcado embebiendo dentro del HTML en diferentes lugares, introduciendo un script en un sólo lugar sin mezclarlo con el resto del código HTML, facilitando el marcado.

```

<script type="application/ld+json">// 
{
  "@context" : "http://schema.org",
  "@type" : "Event",
  "name" : "Final del Mundial de Baloncesto",
  "startDate" : "2014-09-14",
  "location" : {
    "@type" : "Place",
    "name" : "Madrid"
  }
}
// ]]&gt;&lt;/script&gt;
</pre>
</div>
<div data-bbox="138 749 862 800" data-label="Text">
<p>JSON-LD es un formato de Linked Data ligero. Es sencillo de escribir y leer a la hora de su implantación. Se basa como su nombre indica en JSON y proporciona una manera de interoperar a escala Web.</p>
</div>
<div data-bbox="161 884 382 900" data-label="Footnote">
<hr/>
<p><sup>5</sup><a href="http://www.w3.org/TR/json-ld/">http://www.w3.org/TR/json-ld/</a></p>
</div>
```

## 2.2. Sistemas de Gestión de Contenidos

### 2.2.1. Conceptos Generales

Este apartado del estudio, se dedica a realizar una introducción a los Sistemas de Gestión de Contenido (CMS).

En el presente estudio se analizan de forma exclusiva los Sistemas de Gestión de Contenidos Web. De aquí en adelante, cuando se utilice el acrónimo CMS se estará haciendo referencia a los Sistemas de Gestión de Contenidos Web (WCMS), pues dicha acepción es la más extendida y aceptada en el mercado.

Los Sistemas de Gestión de Contenidos (CMS) no son sólo un producto o una tecnología. CMS es un término genérico que define una amplia gama de procesos que apuntala la "próxima generación" de medio a sitios web de gran escala [33].

Es un proceso de gestión de contenidos; a través de la creación, almacenado, modificación y recuperación así como la presentación de datos o contenidos. [34].

En la Biblia de la Gestión de Contenidos[6] se define, en general, Sistema de Gestión de Contenidos como una aplicación informática que se usa para editar, administrar, publicar y realizar búsquedas sobre distintos tipos de medios digitales.

Por su parte, Lin Clark <sup>6</sup> define también CMS como una colección de módulos que pueden ser conectados dentro de un entorno principal para otorgar una funcionalidad avanzada al sitio. En Agosto de 2014 existen más de 15000 módulos en Drupal.org<sup>7</sup> apoyando la famosa frase "There's a module for that".

En los últimos años la gestión de contenidos web ha sufrido una importante evolución, superando problemas como la excesiva información obsoleta debida a la complejidad para actualizar información debida, entre otros, a la falta de control sobre el diseño y a la necesidad de un perfil de conocimiento muy alto requerido para administrar la web. [16]

Por tanto, se proponen los Sistemas de Gestión de Contenidos Web o WCMS como una aplicación web diseñada para facilitar la gestión de sitios web a usuarios con escasos conocimientos técnicos, facilitando el proceso de autoría de contenidos, mejorando los tiempos de creación y actualización y la consistencia de las publicaciones, así como otros factores clave como la navegabilidad, flexibilidad, seguridad y dimensionamiento, todos ellos, bajo el paraguas de una reducción de costes muy grande.

Según Navarro-Galindo[8], a nivel general, la funcionalidad que ofrece un Sistema de Gestión de Contenidos (CMS) se puede descomponer en las siguientes categorías principales:

---

<sup>6</sup><http://www.w3.org/2010/Talks/RDFa-Drupal-Tutorial/>

<sup>7</sup><http://drupal.org>



- **Creación de Contenidos** Los Sistemas de Gestión de Contenidos proporcionan herramientas conocidas como WYSIWYG (What You See Is What You Get), clave del éxito de los WCMS ya que proporcionan un mecanismo sencillo para gestionar el sitio web.
- **Gestión de Contenidos** Los contenidos generados, almacenados en Bases de Datos, pueden ser gestionados —control de versiones y cambios e integración con fuentes y tecnologías de información existentes— a través del repositorio central del CMS, ofreciendo la capacidad de usar flujos de trabajo de diferentes roles de usuario.
- **Publicación de Contenidos** Una vez que los contenidos están en el repositorio pueden ser publicados, tanto en el sitio web como en una intranet, aprovechando los poderosos motores de publicación de los WCMS que permiten aplicar una apariencia automáticamente durante este fase.
- **Presentación de Contenidos** Los CMS permiten asegurar una homogeneidad estética en todas las páginas de forma totalmente independiente a los contenidos.

Real Story Group <sup>8</sup> es un referente en análisis de soluciones tecnológicas de contenidos y asesoría a empresas tecnológicas que publica prestigiosos estudios de diferentes áreas, entre las que destacan los de Sistemas de Gestión de Contenidos Web (WCMS). En este punto, Real Story Group clasifica los WCMS diferenciándolos entre productos comerciales según su posicionamiento en mercado y su orientación en función de su target.

La clasificación que propone es la siguiente:

- *Plataformas Empresariales Avanzadas*. Se comercializan a gran escala y por lo general incluyen paquetes multidimensionales multifunción. Un ejemplo de este tipo de gestor es el IBM WCM<sup>9</sup>
- *Plataformas de Gama Alta*. Ocupan un espacio de expansión moderada entre instalaciones de departamentos y vendedores empresariales. Un ejemplo de este tipo es Adobe Business Catalyst<sup>10</sup>
- *Plataformas de Gama Media*. Ofrecen servicio a PYME, centrándose en necesidades con una personalización un poco mayor que la media. Un ejemplo de este tipo puede ser Drupal<sup>11</sup>
- *Productos de Gama Media*. Paquetes de herramientas con característica preestablecidas. Un ejemplo de este tipo de producto es Magnolia<sup>12</sup>.
- *Productos Simples*. Productos para soluciones más sencillas y open source. Dos ejemplos muy populares por su alta penetración en el mercado son Wordpress<sup>13</sup> y Joomla<sup>14</sup>.

<sup>8</sup><http://www.realstorygroup.com/About/>

<sup>9</sup><http://www-03.ibm.com/software/products/es/ibmwebcontmana>

<sup>10</sup><http://businesscatalyst.com/>

<sup>11</sup><https://www.drupal.org/>

<sup>12</sup><http://www.magnolia-cms.com/>

<sup>13</sup><http://www.wordpress.org>

<sup>14</sup><http://www.joomla.org>

### 2.2.2. Estado del Arte

Tal y como se ha presentado previamente, el objeto del presente estudio radica en la presentación de una propuesta metodológica que permite evaluar y comparar el soporte de marcado semántico en distintos Sistemas de Gestión de Contenidos.

Para ello, el primer paso es definir los criterios básicos que permitan seleccionar los principales gestores de contenidos para, posteriormente, evaluar sobre ellos el soporte semántico centrado en el marcado de una manera más extensa.

Existe una extensa variedad de Sistemas de Gestión de Contenidos Web. En la actualidad más de 1200 según CMS Matrix<sup>15</sup>. Por otro lado, como puede apreciarse en la figura 2.4, según Builtwith<sup>16</sup> —a fecha Julio de 2014— los tres CMS más utilizados son WordPress, Joomla! y Drupal, sumando entre los tres casi 14 millones de instalaciones. Los dos primeros pertenecen a los denominados, por Real Story Group, productos simples mientras Drupal es considerado una plataforma de Gama Media.



Figura 2.4: Estadísticas de uso de los CMS según BuiltWith 2014

Se han tenido en cuenta dos criterios clave, alineados con los intereses de un estudio de investigación para la selección de los Sistemas de Gestión de Contenidos, sobre los que realizar la propuesta metodológica que permite evaluar y comparar el soporte de marcado semántico. Los criterios de selección han sido:

<sup>15</sup><http://www.cmsmatrix.org/>

<sup>16</sup><http://trends.builtwith.com/cms>

- **Cuota de uso y relevancia.** El uso de tecnologías con la mayor comunidad de usuarios posible ofrece importantes ventajas: mayor soporte disponible, mayor número de módulos de terceros desarrollados y probados por la comunidad, garantía de evolución de la plataforma. Como se ha visto previamente entre Wordpress, Joomla! y Drupal suman casi 14 millones de instalaciones y aproximadamente el 47 % de los sitios web gestionados mediante CMS utilizan Wordpress, mientras que los sitios que no utilizan ningún CMS se han reducido considerablemente durante los últimos 4 años. Con fecha de realización del presente estudio, se ha comprobado el impacto de los tres mediante Google Trends<sup>17</sup>, al apreciarse que en las búsquedas del término ‘CMS’, aparecen los tres como búsquedas relacionadas, como es posible apreciar en la figura 2.5.



Figura 2.5: Tendencia Temporal de los 3 CMS a estudio en Trends

**OpenSource.** El software libre no impone un coste de licencia, y cuenta con las grandes ventajas del código abierto: flexibilidad, seguridad y rapidez en las tareas de desarrollo y actualización. Este es un punto clave en la elección ya que asegura la posibilidad de reproducción del presente estudio por cualquier perfil debido a su gratuidad.

Como se ha señalado anteriormente, CMS Matrix proporciona un servicio de comparación a través de las características de los más de 1200 CMS disponibles en el mercado. La información recogida en las comparativas ha sido desarrollada mediante una comunidad abierta, lo que aporta una visión con un alto grado de objetividad y de revisión continua.

<sup>17</sup><http://www.google.com/trends>

A continuación se presenta una comparativa de algunas de las características propuestas por CMS Matrix, para la comparativa de Sistemas de Gestión de Contenidos, aplicada a los tres CMS seleccionados previamente. Las versiones utilizadas en el presente trabajo son:

- Wordpress (3.3.1)
- Joomla (2.5)
- Drupal(7.12).

Las valoraciones aplicadas a las tablas pueden presentar los valores:

- Sí
- No
- OPC.(Opcional).
- Valor concreto.

1. **Requisitos del sistema:** Esta comparativa, presentada a modo de tabla en la figura 2.6, presenta los requisitos necesarios para que el CMS funcione correctamente, además de sus características como CMS, entre estos requisitos se encuentra información sobre el sistema operativo, servidor Web, base de datos, licencia, lenguaje de programación,...

	Drupal	Joomla	WordPress
<b>REQUISITOS DEL SISTEMA</b>			
Sistema operativo	Indep.	Indep.	Indep.
Servidor Web	Apache MS IIS	Apache MS IIS	Apache
Base de datos	MySQL SQL S. Oracle	MySQL	MySQL
Licencia	Libre	Libre	Libre

Figura 2.6: Comparativa de los Requisitos de Sistema en CMS

2. **Seguridad:** La tabla en la figura 2.7 recoge aquellas características que posee el CMS, para protegerse frente seguridad, como por ejemplo aprobación del contenido, verificación de email, granularidad de privilegios, autenticación kerberos, LDAP, historial de login, registro de auditoría, compatibilidad ssl, versionado,...

	Drupal	Joomla	WordPress
<b>SEGURIDAD</b>			
Aprobación del contenido	Sí	Sí	Sí
Verificación de email	Sí	Sí	Sí
Granularidad privilegios	Sí	Sí	Sí
Autenticación Kerberos	No	No	No

Figura 2.7: Comparativa de Seguridad en CMS

3. **Rendimiento:** La tabla en la figura 2.8 presenta características relacionadas con el rendimiento del CMS, es decir, es la capacidad de la realización de las tareas con respecto a los medios disponibles, algunas de estas características son, el balanceo de carga, cacheo de páginas, replicación de base de datos, exportación de contenido estático, etc.

	Drupal	Joomla	WordPress
<b>RENDIMIENTO</b>			
Balanceo de carga	Sí	Sí	Sí
Cacheo de páginas	Sí	Sí	OPC.
Cacheo avanzado	Sí	Sí	OPC.
Replicación de base de datos	Sí	No	OPC.
Exportación de contenido estático	No	No	OPC.

Figura 2.8: Comparativa de Rendimiento en CMS

4. **Soporte:** Esta comparativa, presentada a modo de tabla en la figura 2.9, compara el soporte que ofrecen los CMS, mediante programas de certificación, formación comercial, soporte comercial, comunidad de desarrollo, ayuda online, API para desarrollar extensiones, foros públicos, listas de correos, servicios profesionales,...

	Drupal	Joomla	WordPress
<b>SOPORTE</b>			
Programas de certificación	OPC.	No	No
Esqueleto de código para nuevos plugins	Sí	OPC.	No
Formación comercial	Sí	Sí	Sí
Soporte comercial	Sí	Sí	Sí
Formación comercial	Sí	Sí	Sí
Comunidad de desarrollo	Sí	Sí	Sí
Ayuda online	Sí	Sí	Sí
API para desarrollar extensiones	Sí	Sí	Sí
Foro públicos	Sí	Sí	Sí
Listas de correos públicas	Sí	Sí	Sí
Manuales comerciales	Sí	Sí	Sí
Servicios profesionales	Sí	Sí	No
Conferencia para usuarios	Sí	Sí	Sí

Figura 2.9: Comparativa de Soporte en CMS

5. **Facilidad de uso:** La tabla en la figura 2.10 recoge funcionalidades que facilitan realizar ciertas tareas, como redimensionar imágenes, subida de archivos masivo, URL amigables, editor wysiwyg, lenguaje de plantillas, lenguaje para macros, contenido drag-n-drop, prototipado, asistente de configuración, corrector ortográfico...

	Drupal	Joomla	WordPress
<b>FACILIDAD DE USO</b>			
Lenguaje de plantillas	Sí	Sí	No
Lenguaje para macros	OPC.	Sí	OPC.
Contenido drag-n-drop	OPC.	No	Sí
Redimensionado de imágenes	OPC.	OPC.	OPC.
Subida de archivos masivos	OPC.	Sí	Sí
Prototipado	OPC.	Sí	OPC.
Asistente para la configuración del sitio	OPC.	No	No
Corrector ortográfico	OPC.	OPC.	Sí
Asistente para generar plantillas/styles	OPC.	No	No
Subscripciones a secciones	OPC.	Sí	OPC.
Deshacer	OPC.	No	OPC.
Editor WYSIWYG	OPC.	Sí	Sí
URL amigables	No	No	OPC.

Figura 2.10: Comparativa de Facilidad de uso en CMS

6. **Interoperabilidad:** Esta comparativa, presentada a modo de tabla en la figura 2.11, recoge aquellos sistemas/componentes o información, con el que el sistema es capaz de intercambiar información, como RSS, soporte FTP, Cumplimiento WAI, XHTML, ICAL,...

	Drupal	Joomla	WordPress
<b>INTEROPERABILIDAD</b>			
RSS	Sí	Sí	Sí
Soporte FTP	OPC.	Sí	OPC.
Cumplimiento WAI	OPC.	No	OPC.
Cumplimiento XHTML	Sí	Sí	Sí
ICAL	OPC.	OPC.	No
Soporte WebDAV	No	No	No

Figura 2.11: Comparativa de Interoperabilidad en CMS

7. **Administración:** La tabla en la figura 2.12 recoge funcionalidades que nos facilitan la administración del sitio, como por ejemplo, la gestión de registros, portapapeles, programación de contenido, edición de contenido en línea, gestión de temas, estadísticas Web, administración de temas, workflow,...

	Drupal	Joomla	WordPress
<b>ADMINISTRACIÓN</b>			
Gestión de registros	Sí	Sí	Sí
Portapapeles	No	No	No
Programación de contenido	OPC.	Sí	OPC.
Edición de contenido en línea (inline)	Sí	Sí	OPC.
Gestión de skins/temas	Sí	Sí	Sí
Papelera de reciclaje	No	Sí	Sí
Estadísticas Web	Sí	Sí	OPC.
Administración de plantillas	Sí	Sí	Sí
Gestión de traducción	Sí	OPC.	OPC.
Gestión de anuncios	OPC.	Sí	No
Motor de Workflow	OPC.	No	No

Figura 2.12: Comparativa de Administración en CMS



8. **Flexibilidad:** La tabla en la figura 2.13 presenta características que facilitan y hacen flexible la realización y configuración de ciertas tareas, por ejemplo, reutilización de contenidos, traducción del interfaz, contenido multilinguaje, reescritura de URL, despliegue multisite,...

	Drupal	Joomla	WordPress
<b>FLEXIBILIDAD</b>			
Soporte CGI	Sí	Sí	No
Reutilización de contenidos	OPC.	Sí	No
Traducción de interfaz según localización	Sí	Sí	Sí
Metadatos	Sí	Sí	Sí
Contenido multilinguaje	Sí	OPC.	OPC.
Reescritura de URL	Sí	Sí	Sí
Despliegue multisite	Sí	OPC.	No

Figura 2.13: Comparativa de Flexibilidad en CMS

A continuación se describen algunas generalidades destacables de cada uno de los tres Sistemas de Gestión de Contenidos (CMS) seleccionados previamente:

### Wordpress

WordPress<sup>18</sup> comenzó como un Sistema de Gestión de Contenidos (CMS) orientado al desarrollo de Blogs, evolucionando hacia lo que es hoy, uno de los CMS de uso general más completos del mercado. Su origen se encuentra en la herramienta b2/cafelog, diseñada por Michel Valdrighi, y desarrollada inicialmente como un sistema para la gestión de blogs personales. La primera versión de WordPress, la 0.7, surge en 2003, encontrándose actualmente en la versión 3.3.2. El modelo de actualizaciones ha sido, hasta la fecha, bastante dinámico con 3 o 4 versiones base anuales, y versiones menores con una frecuencia mensual. WordPress es uno de los CMS más conocidos, utilizados y descargados del mercado, el motivo se encuentran es la sencillez de uso e implantación.

En el Open Source CMS Market Share Report de 2011[35] destacan algunos datos significativos que reflejan la importancia de este CMS son:

- Es la solución más usada en la actualidad con una cuota del mercado de más del 45 %.

<sup>18</sup><http://wordpress.org/>

- Dispone de más de 600.000 descargas medias mensuales.
- Dispone de más de 80 libros editados, reflejo del nivel de soporte que ofrece el producto.
- La visión de los usuarios con respecto a este CMS es positiva en un 75 %.
- La protección de la la privacidad de los contenidos es otra de las características principales de Wordpress, a través de la definición de niveles de usuario, protección de contenidos por contraseña, filtros antispam o controles de comentarios.
- WordPress dispone de un buen soporte a través de abundante documentación y foros, siendo una de las comunidades más dinámicas en el contexto de los Sistemas de Gestión de Portales Web.

Como resumen gráfico de la repercusión de Wordpress en el avance de los Sistemas de Gestión de Contenidos como estándar de facto en la web, en la figura 2.14 pueden verse algunos datos ofrecidos por Moove Agency<sup>19</sup>



Figura 2.14: Repercusión de Wordpress en 2013.

## Joomla!

Joomla<sup>20</sup> es un potente Sistema de Gestión de Contenidos (CMS) para la creación de sitios web dinámicos. Joomla! surge en 2005 como resultado de una división del proyecto

<sup>19</sup><http://mooveagency.com>

<sup>20</sup><http://www.joomla.org>

Mambo. La primera versión de Joomla integra el núcleo de Mambo, pero con nuevo software libre y muchos cambios importantes en el código. A partir de esta escisión, muchos colaboradores, comunidades y diseñadores, respaldaron el proyecto, qué evolucionó hasta convertirse en lo que es hoy en día, uno de los CMS mas usados y conocidos del mercado. Hasta la fecha ha pasado por cinco versiones estables, aproximadamente una por año.

Joomla es uno de los CMS más utilizados y mejor posicionado del mercado. Para extraer todo el potencial de Joomla, se requiere cierto conocimiento y experiencia, ya que su máxima versatilidad se obtiene de la integración, adaptación y desarrollo de nuevos módulos. A diferencia de Wordpress la curva de aprendizaje es más lenta y la versatilidad para proyectos más complejos es mayor.

En el Open Source CMS Market Share Report de 2011[35] destacan algunos datos significativos que reflejan la importancia de este CMS son:

- En 2011, obtuvo una media semanal de descargas de 86.547, el segundo CMS más descargado, después de WordPress.
- Dispone de una gran documentación, en total 65 libros impresos, de los cuales 13 fueron lanzados en 2011.
- Elevado PageRank de Google. En Julio 2014, joomla.org presenta un valor de 9/10.
- La visión de los usuarios con respecto a este CMS es positiva en casi un 50
- Una de las mayores ventajas de Joomla es su magnífica comunidad que presenta una gran participación de los usuarios, ofreciendo un sistema en continua actualización frente a vulnerabilidad, bugs, nuevas funcionalidades y extensiones. Dispone de un soporte muy completo a través de Web oficiales, foros, y todo tipo de documentación generada.
- Otra de las características destacadas de Joomla en la versatilidad que ofrece el sistema a través de plantillas, extensiones y adaptaciones. Existen cientos de módulos, componentes y plugins que extienden la funcionalidad original del CMS, gestión de archivos, gestión de contactos, sistema de búsqueda, tiendas online, bolsas de trabajo, integración con redes sociales, gestión de noticias y newsletter, sistemas de encuestas, . . .

## Drupal

Drupal <sup>21</sup> es un sistema de administración de contenidos Web especialmente versátil. Inicialmente dirigido a dar soporte a una comunidad de Weblog. Su desarrollo fue iniciado por Dries Buytaert en 1999 y no fue hasta 2001 cuando se publico la primera versión del CMS. Hasta el lanzamiento de la versión 4.0.0, Drupal publicaba una versión anualmente, tras ésta, el lanzamiento de cada nueva versión base, se ha ralentizado a una cada 2 o 3 años, publicando entre 10 y 20 versiones menores sobre cada una de las versiones base.

---

<sup>21</sup><http://www.drupal.org>

Actualmente Drupal se encuentra en la versión 7.3.1. Entre 2008 y 2009

Drupal se sitúa como uno de los CMS referentes del sector empresarial. De los tres CMS a estudio es el que más versatilidad potencial tiene, dispone de un entorno de personalización robusto, tanto el contenido como la presentación pueden ser tratados de forma individual de acuerdo a unas preferencias definidas por el usuario, siendo necesario conocer a fondo el mismo. La curva de aprendizaje es, probablemente, la más lenta de los tres a estudio.

Al igual que en los dos anteriores, el estudio se basa en los datos más significativos que reflejan la importancia de este CMS, en Open Source CMS Market Share Report de 2011 [35]:

- Se encuentra entre en los primeros puestos (top 5, 10 y 20), de la mayor parte de los ranking que aparecen en Internet.
- Elevado PageRank de Google. En Julio 2014, drupal.org presenta un valor de 8/10.
- La visión de los usuarios con respecto a este CMS es positiva en un 70
- Es uno de los CMS más mencionados en los blogs y en redes sociales, sobre un 40
- Dispone de gran documentación, 65 libros impresos, de los cuales 13 fueron lanzados en 2011.

Son muchas las características que sitúan a Drupal entre los CMS más destacados del mercado:

1. Dispone de un entorno de personalización robusto, tanto el contenido como la presentación pueden ser tratados de forma individual de acuerdo a unas preferencias definidas por el usuario. La gestión de contenido se realiza como objetos independientes, de forma que puede realizarse un tratamiento individualizado de la información, facilitando su inclusión en cualquier página o permitiendo comentarios específicos sobre cada uno de ellos.
2. Los mecanismos de actualización de contenidos son realmente sencillos, permite editar la mayor parte de los contenidos tanto desde el *frontend* como desde el *backend*.
3. Ofrece la posibilidad de gestionar las taxonomías y la estructuración de contenidos de forma personalizable, algo indispensable para sitios de complejidad media-alta.
4. Seguridad: Permite la gestión de permisos ofreciendo un sistema muy avanzado y completamente personalizable a nivel de rol y páginas.
5. Escalabilidad y rendimiento: sistema de cache avanzado, replicación de base de datos, balanceo de carga, mecanismos de control de congestión configurable para habilitar módulos, ...

6. La comunidad de Drupal, ofrece un desarrollo dinámico y un soporte amplio basado en foros Web. Además, Drupal presenta un grupo muy activo sobre Web Semántica, que comenzó en 2007 en Barcelona y debate, entre otras cosas sobre como integrar tecnología semántica en Drupal.<sup>22</sup>
7. Dispone de cientos de extensiones agrupadas según funcionalidad en distintas categorías: Administración, Control de Acceso, Eventos, Comercio, Comunidad, contenidos, Gestión de usuarios, Búsquedas, etc.

Con respecto a las características más técnicas, cabe mencionar que Drupal se encuentra liberado bajo licencia GPL y utiliza PHP como lenguaje de programación, MySQL como motor de base de datos, aunque también puede funcionar con PostgreSQL o SQLite, y Apache o Microsoft IIS como servidor Web.

En este Sistema de Gestión de Contenidos (CMS) cabe destacar la web de la Casa Blanca<sup>23</sup> que ha optado por esta solución demostrando entre otras, su alto rendimiento a nivel de seguridad.

## 2.3. Soporte Semántico en CMS

Los sistemas de organización del conocimiento parecen haber encontrado en los Sistemas de Gestión de Contenidos (CMS), una opción de plataforma muy interesante para su publicación y gestión, a través de la incorporación inmediata del marcado semántico en estos.

Como se ha visto previamente, la extracción de información semántica de los propios contenidos web —publicados de forma extremadamente distribuida— [16], mejorará la eficiencia de los motores de búsqueda y será el paso definitivo para la creación automática de conjuntos de datos semánticos. Asimismo, el marcado semántico ofrece nuevas posibilidades para la reutilización de datasets como fuente de información semántica para ser incorporada a los contenidos.

Continuamente aparecen enfoques prácticos para la gestión y apoyo del ciclo de vida del contenido semántico en la web de Datos. Sin embargo, como expone Khalili[30] el aspecto menos desarrollado en este campo es la parte *userfriendly* del marcado semántico manual y semi-automático .

En la figura 2.15 puede verse de manera esquemática una comparativa entre un Sistema de Gestión de Contenidos (CMS) y un sCMS (Semantic Content Management System) basada en un estudio que presentó recientemente el Dr. Tilman Becker realizó, durante los OpenCMS Days de Colonia (Alemania) y que muestra las relaciones a nivel unidad atómica, tipos de búsqueda y gestión entre ambos tipos de gestor.

---

<sup>22</sup><https://groups.drupal.org/semantic-web>

<sup>23</sup><http://www.whitehouse.gov/>



Figura 2.15: Sistemas de Gestión de Contenidos Semánticos.

Los Sistemas de Gestión de Contenidos (CMS) por tanto se presentan como una de las grandes bazas en el proceso de marcado semántico sea cual fuere el formato de marcado elegido, existiendo la opción de combinar formatos dentro de un mismo recurso. El administrador del Sistema de Gestión de Contenidos (CMS) definirá las relaciones entre las estructuras de contenido y los elementos de los vocabularios de descripción elegidos, siendo el Sistema de Gestión de Contenidos (CMS) el encargado de realizar el marcado convenientemente. Los Sistemas de Gestión de Contenidos (CMS) incorporan cada vez más funcionalidades en forma de extensiones, plugins... que permiten implementar el soporte semántico de una forma modular y escalable.

La amplia difusión de los Sistemas de Gestión de Contenidos (CMS) sugiere la integración del marcado semántico en estos sistemas para simplificar su realización y, por tanto, hacerlo de una forma mucho más eficiente. El marcado semántico realizado por un Sistema de Gestión de Contenidos (CMS) debe partir de un mapeado entre los campos descriptivos y los correspondientes elementos de las ontologías y esquemas de metadatos utilizados como modelos descriptivos.

Pastor-Sánchez[16] destaca la potencia de Drupal desde su versión 7, al incorporar el marcado semántico mediante RDFa tanto en su núcleo como mediante extensiones.

En el ámbito del marcado automático, actualmente, existen varios proveedores de ser-

vicios compatibles con CMS a través de APIs (application programming interfaces). Los CMS se conectan con estos servicios mediante extensiones que hacen de puente entre dichas APIs y el motor de marcado de contenido, generalmente basado en un editor WYSIWYG —What You See Is What You Get—. Las APIs devuelven los resultados en formatos como RDF, XML, o JSON.





# Capítulo 3

## Planteamiento del Estudio

### 3.1. Objetivos de la Evaluación

Tras la selección previa de los principales CMS, Wordpress, Joomla y Drupal —de entre los más de 1200 CMS—, atendiendo a criterios como la cuota de uso y la relevancia, se ha realizado una evaluación de las principales características de los CMS según CMS Matrix <sup>1</sup>.

Como se ha visto anteriormente, uno de los aspectos esenciales para el progreso hacia la Web Semántica es cómo convertir el contenido web —tanto el nuevo como el ya existente— expresado en lenguaje natural en su equivalente semántico, en el cual el contenido es enriquecido con metadatos en formato estructurado. [8]

El objetivo de este estudio se dirige a proponer criterios evaluativos que permitan valorar y comparar el soporte de semántica que presentan los principales gestores de contenidos, poniendo el foco en el soporte que ofrecen en el ámbito del marcado semántico.

Para ello se proponen indicadores que atenderán a las funcionalidades exigibles a este tipo de soluciones prestando atención especial a su adecuación a los estándares de web semántica.

### 3.2. Metodología de la Evaluación

Previo a la elaboración del estudio, se han tomado referencias de algunas propuestas evaluativas similares, Moya y Gil[9] y García Castro[10] que han permitido instaurar marcos de evaluación en el entorno de la Web Semántica, sirviendo de guía para la elección y posterior formalización de los indicadores a utilizar en este estudio.

Como propone García Castro[10], para lograr consolidar la tecnología de la Web Semántica, es necesario que alcance un nivel de madurez en el que pueda cumplir con los requisitos de calidad necesarios y para ello necesita ser evaluada exhaustivamente proporcionan-

---

<sup>1</sup><http://www.cmsmatrix.org/>

do resultados objetivos y por otro lado obtener una mejora masiva en su calidad.

De acuerdo con la ISO 14598 estándar (ISO / IEC, 1999)[36], la Evaluación de Software se propone como un examen sistemático que mide hasta qué punto una entidad es capaz de cumplir con los requisitos especificados; esta norma considera el software no sólo como un conjunto de programas, sino también como un conjunto de procedimientos, documentación y datos.

En el estudio "Software inspection benchmarking-a qualitative and quantitative comparative opportunity"[37] se define el benchmarking como un proceso de mejora continua que consiste en ser el mejor de los mejores a través de la comparación de procesos similares en diferentes contextos.

Algunas definiciones vistas en [10], ponen de manifiesto las dos principales características del Benchmarking de Software:

- La búsqueda de las mejores prácticas
- La Mejora Continua.

García Castro y Gómez-Pérez en su estudio "Benchmarking in the Semantic Web"[10], proponen una metodología de Benchmarking de software, basada en iteraciones compuestas por tres fases principales:

- Fase de Planificación
  1. Identificación de Objetivos
  2. Identificación de Herramientas y Métricas
  3. Identificación de Participantes
  4. Redacción de Propuestas
  5. Gestión de la participación
  6. Selección del Partner
  7. Planificación y asignación de Recursos
- Fase de Experimentación
  1. Definición del Experimento
  2. Ejecución del Experimento
  3. Análisis de Resultados del Experimento
- Fase de Mejora
  1. Redacción del Informe de Benchmarking
  2. Comunicación
  3. Planning de Mejoras

4. Mejoras
5. Monitorización

La propuesta finaliza con una tarea de recalibrado como puede verse en la figura 3.1

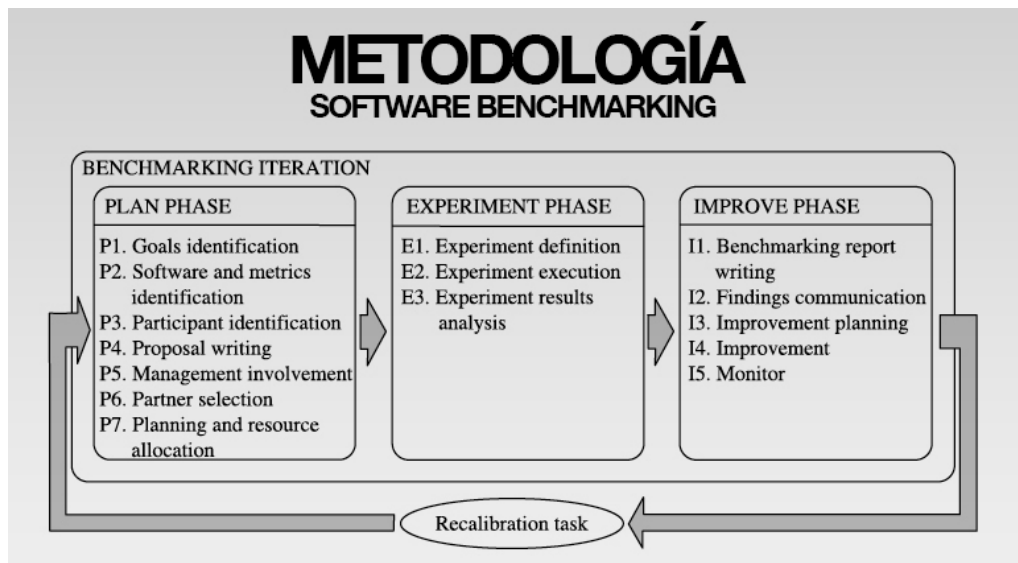


Figura 3.1: Metodología de benchmarking de software

La identificación de objetivos está clara. En el ámbito de la Identificación de Participantes, tendremos en cuenta el perfil de usuario que gestiona contenidos dentro de Sistemas de Gestión de Contenidos (CMS). Un perfil que presenta un conocimiento medio en cuanto a desarrollo web se refiere y, según la literatura consultada, opta generalmente por la integración a través del uso de plugins, extensiones y otras herramientas que se puedan implementar con servicios web.

En el estudio, se proponen criterios evaluativos que permiten valorar y comparar el soporte de marcado semántico que presentan los principales gestores de contenidos, con especial atención a la compatibilidad con los estándares de la web semántica.

No se olvidan aspectos trascendentales como las funcionalidades de la aplicación, Requisitos del Sistema,... determinantes para usuarios finales. Asimismo, se ha de tener en cuenta que un mayor número de formatos de marcado soportados por la herramienta que supone una mayor garantía de éxito en la búsqueda de cubrir las necesidades del usuario.

En este punto, cabe remarcar la especial atención a la metodología propuesta por [38], definiendo los indicadores en los que se centrará el análisis del presente estudio:

- **Propósito de la aplicación** Este parámetro indica la finalidad particular de la aplicación dentro del ámbito general a estudio que es dar soporte semántico al propio

CMS, centrándose en el marcado semántico. En el presente estudio se engloba con el término aplicación a los módulos, plugins y/o extensiones relacionadas o no con APIs así como otras funcionalidades propias del Core del Sistema de Gestión de Contenidos.

- **Requisitos del sistema** Este parámetro de evaluación indica los requisitos necesarios para la utilización del software.
- **Funcionalidades** Este parámetro muestra una selección de funcionalidades comunes a la mayoría de las aplicaciones y cuyo carácter resulta significativo en el presente estudio. En este sentido, la aplicación del marcado semántico en los CMS cubre cinco funciones principales[39]:
  - Integración de contenidos en el ámbito de las redes sociales mediante la aplicación del protocolo Open Graph [43].
  - Inclusión de sistemas de navegación del tipo breadcrumbs que indican la ruta de navegación recorrida por el usuario en el sitio web y además ubica un contenido en la estructura jerárquica del sitio web.
  - Adición de descripciones estructuradas de contenidos para el uso de la tecnología rich snippets de google.
  - Marcado semántico del cuerpo de una página web.
  - Incorporación de metadatos descriptivos para mejorar el Posicionamiento SEO.
- **Tipología de marcado semántico admitido**

Este parámetro muestra el tipo de marcado semántico que utiliza la aplicación, ofreciendo, de esta manera, una información de gran relevancia por las implicaciones que conlleva. Los Tipos de Anotaciones Semánticas, según [8] Aunque se pueden realizar diferentes clasificaciones, nuestro foco de interés se centra en dos criterios:

#### 1. De acuerdo al lugar donde se almacenan

- Internas o embebidas: Almacenadas dentro del mismo documento web donde se realiza la anotación.
- Externas: Almacenadas en ficheros o en servicios distintos del documento web que se anota.

#### 2. Conforme a su nivel de automatización

- Directa o Manual: El usuario realiza las anotaciones directamente en un contenido dado mediante el uso de herramientas específicas.
- Automática: De alguna manera, un proceso automático genera las anotaciones, identificando entidades semánticas y sus relaciones.

- **Integrabilidad del software** La integrabilidad se beneficia de la adopción de estándares abiertos. El software puede ser reutilizable por medio de APIs (API es un conjunto de funciones y procedimientos que ofrecen ciertas bibliotecas para ser utilizado por otro software como una capa de abstracción.) que ofrecen un conjunto de

librerías, provistas de interfaces que describen las definiciones de clase y su comportamiento (en el caso de los lenguajes orientados a objetos). SaaS es un modelo de distribución de software donde el soporte lógico y los datos que se alojan en servidores de una compañía, a los que se accede desde un cliente web. [40].

Por su parte, un servicio web es un "sistema de software diseñado para apoyar la interoperabilidad e interacción entre equipos en una red" [41]. Se trata, por tanto, de tecnología que utiliza un conjunto de protocolos y estándares que sirven para intercambiar datos entre aplicaciones. Permiten que distintas aplicaciones de software, desarrolladas en lenguajes de programación diferentes, y ejecutadas sobre cualquier plataforma, puedan utilizar los servicios web para intercambiar datos, esto es, se conviertan en aplicaciones web.

En el contexto de internet, los servicios web se proponen como una técnica óptima para la reutilización de funcionalidad externa. Los servicios web impulsan la interoperabilidad a través del intercambio de mensajes XML. [38].

- **Formatos de marcado semántico admitidos** Este factor es de vital importancia por su vinculación con los posibles objetivos del usuario. Como se ha visto anteriormente, los formatos de marcado aceptados en la actualidad son microdatos, microformatos o RDFa.

Las implicaciones que conllevan estos formatos serán adquiridas por las propias herramientas a estudio, por lo que la información que subyace en este punto, pese a su simplicidad, es clave, en el contexto del presente estudio.

- **Soporte a otros CMS** En el caso de herramientas basadas en servicios web y/o API, se evalúa si ofrece soporte al resto de CMS a estudio.

El test de la propuesta evaluativa se realiza sobre un conjunto de aplicaciones en forma de Plugins, API con soporte para CMS y el propio Core del CMS, que completan el soporte de marcado semántico de los CMS seleccionados anteriormente.



# Capítulo 4

## Fases de Planificación y Experimentación

### 4.1. Aplicación de los Criterios de Evaluación

Tal y como se ha expuesto con anterioridad, son numerosas las extensiones que presentan los Sistemas de Gestión de Contenidos (CMS) que implementan el marcado semántico de un modo muy sencillo y escalable. En cuanto soporte del core del CMS, cabe destacar Drupal desde su versión 7 (que ofrece el marcado semántico mediante RDFa , tanto en su núcleo como mediante extensiones al que parece sumarse —a partir de su versión 3.3— Joomla.

Una particularidad del tipo de información presente en estos Sistemas de Gestión de Contenido (CMS) es el hecho de ser información altamente estructurada, situación que no se da en todas las páginas web.

Con el fin de validar la propuesta metodológica planteada previamente y como parte de la Fase de Planificación vista anteriormente, se debe proceder a la Selección de Partners y para ello se ha realizado una selección de herramientas para el marcado semántico, dentro del ámbito de Wordpress, Joomla y Drupal que cumplieran ciertos requisitos tales ser aplicaciones libres o contar con una versión de evaluación, permitiendo la reproducción de los experimentos a cualquier perfil interesado y tratarse de aplicaciones para el marcado semántico.

En la análisis se incluyen algunas herramientas que amplían las funciones de los CMS para el marcado semántico manual. Estas se presentan como soluciones de diferentes proveedores de servicios que realizan el marcado de manera automática a través de APIs (application programming interfaces) que permiten su portabilidad mediante extensiones que sirven de nexo entre dichas APIs y el motor de marcado de contenido basado, casi siempre, en editores WYSIWYG (What You See is What You Get). Las APIs devuelven los resultados en formatos como XML, RDF o json (javascript object notation). Asimismo, se ha valorado la obtención de una representación que aborde todos los formatos de marcado semántico vistos con anterioridad, con el fin de contribuir a la posible evaluación

de los mismos en el ámbito de los Sistemas de Gestión de Contenidos.

Este punto de la Fase de Planificación, la selección de partners, debe seleccionar el software empleado para el estudio de referencia, teniendo en cuenta su importancia y uso en la comunidad o en la industria, su disponibilidad pública,...

Al tratarse de herramientas desarrolladas, en muchos casos, por la propia comunidad de cada Sistema de Gestión de Contenidos, se ha consultado en los principales foros de cada una de las plataformas con el fin de seleccionar herramientas cuyo uso está extendido y que posean un soporte de actualizaciones que aseguren su permanencia al máximo.

Este marco de actuaciones nos lleva a fijarnos en las herramientas que permiten en la actualidad la gestión de marcado semántico y proponer una metodología para su evaluación en la que se tenga en cuenta de modo específico su clara adaptación a la web semántica.

Finalmente, las aplicaciones objeto del análisis, que permitirán realizar un test sobre la metodología propuesta anteriormente son:

### 1. **AlchemyTagger para Wordpress**<sup>1</sup>

Es un plugin que funciona de forma automática y transparente y permite aportar marcado semántico al cuerpo de las publicaciones. Está basado en Alchemy<sup>2</sup> SaaS basado en técnicas de procesamiento del lenguaje natural (PLN). Se utiliza mediante una API que interactúa con un servicio web o un SDK (Software Development Kit) disponible en múltiples lenguajes de programación. La API tras suministrar un texto, en alguno de los idiomas soportados, identifica personas, organizaciones, etiquetas, conceptos y relaciones entre distintos tipos de entidades. Como puede apreciarse en la figura 4.1 es necesario el registro para poder conectar con la API. Es gratuito.

### 2. **RDFaCE para Wordpress**<sup>3</sup>

Es el plugin para Wordpress que permite implementar RDFaCE en el editor, como se observa en la figura 4.2, permitiendo realizar de manera automática y/o manual, marcado semántico mediante RDFa o Microdatos y permite el uso de vocabularios de Schema.org. que permite representar una amplia variedad de recursos: personas, trabajos creativos, organizaciones, eventos, etc.

El marcado manual se efectúa seleccionando parte o la totalidad del texto del cuerpo del contenido e identificando el tipo de entidad al que se hace referencia: persona, evento, organización, noticia, lugar, objeto audiovisual, etc. En el texto correspondiente a la entidad identificada, se seleccionan fragmentos más específicos para definir sus propiedades: nombre, URL, fechas, etc. También es posible establecer relaciones entre diferentes entidades.

---

<sup>1</sup><http://wordpress.org/plugins/alchemytagger/>

<sup>2</sup><http://www.alchemyapi.com>

<sup>3</sup><https://wordpress.org/plugins/rdface/>



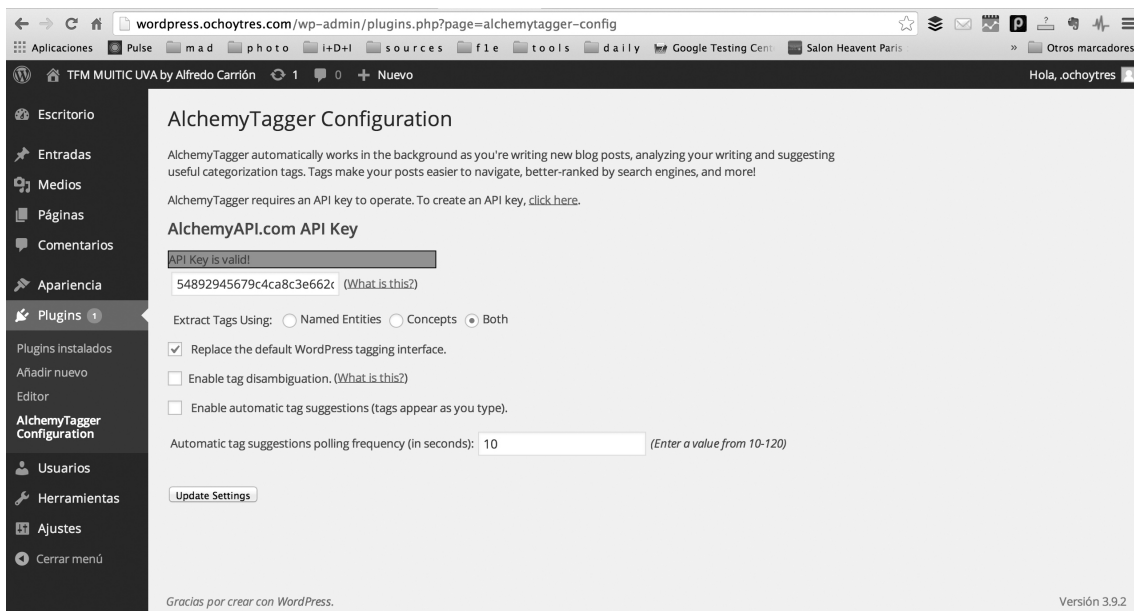


Figura 4.1: Vista del Backend de Alchemy para Wordpress

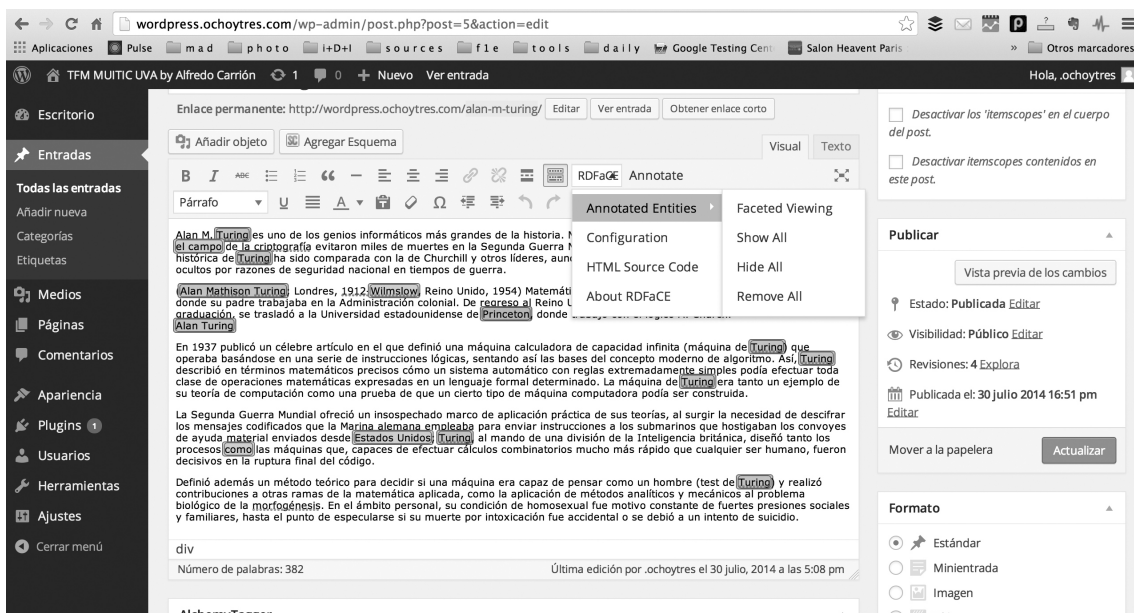


Figura 4.2: Vista del editor de Wordpress con RDFaCE integrado

### 3. Schema Creator para Wordpress<sup>4</sup>

Schema Creator es un plugin para Wordpress que permite simplificar el proceso de marcado semántico mediante microdatos con Schema.org. El marcado se realiza de forma manual, tal y como se observa en la figura 4.3 siendo automática la creación

<sup>4</sup><https://wordpress.org/plugins/schema-creator/>

de la estructura de datos. Es gratuito y no necesita registro.

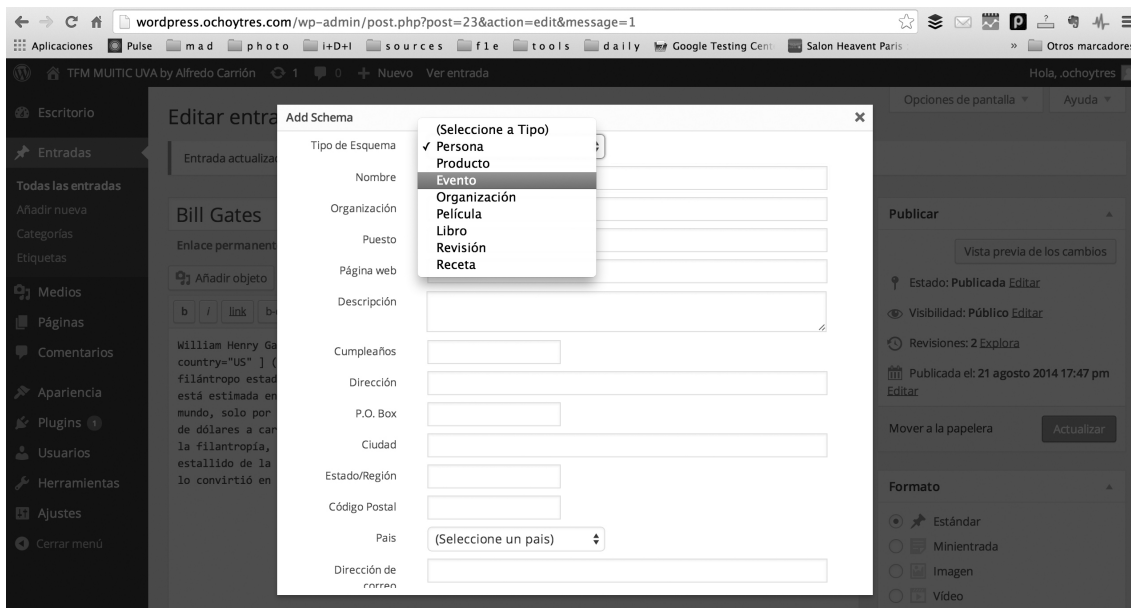


Figura 4.3: Vista del backend de Schema para Wordpress

#### 4. Schema.org para Drupal<sup>5</sup>

Es un módulo de Drupal que soporta marcado semántico en formato RDFa 1.0.(el propio del core de Drupal 7). No es compatible con el formato de microdatos pese a que utiliza como vocabulario Schema.org. La comunidad de Drupal no recomienda su uso simultáneamente con otros módulos de marcado. Este módulo ha sido correctamente validado por sus autores en la herramienta de Google Rich Snippets. Es actualizada con frecuencia y dispone de un gran número de sitios que la usan. Su funcionamiento no es tan inmediato como otras soluciones, pero permite una gran versatilidad a la hora de crear *Content Types*, como puede verse en la figura 4.4. En la propia URL del módulo, se presenta un videotutorial para su uso, realizado por Stéphane Corlosquet<sup>6</sup>, actualmente, uno de los máximos impulsores de la RDFa y Web Semántica mediante Drupal.

#### 5. Open Calais para Wordpress<sup>7</sup> y Drupal<sup>8</sup>

Módulo que sirve de nexo entre Wordpress y/o Drupal y la API de Open Calais, presentando sus panel de configuración en el propio backend del Sistema de Gestión de Contenidos —en este caso Wordpress— como puede apreciarse en la figura 4.5, que presenta el proceso de publicación de una entrada sobre la que el plugin, envía a Open Calais el código del contenido (previa validación de la API KEY) y recibe varias propuestas de marcado.

<sup>5</sup><https://www.drupal.org/project/schemaorg>

<sup>6</sup><https://www.drupal.org/u/scor>

<sup>7</sup><https://wordpress.org/plugins/calais-auto-tagger/>

<sup>8</sup><https://www.drupal.org/project/opencalais>

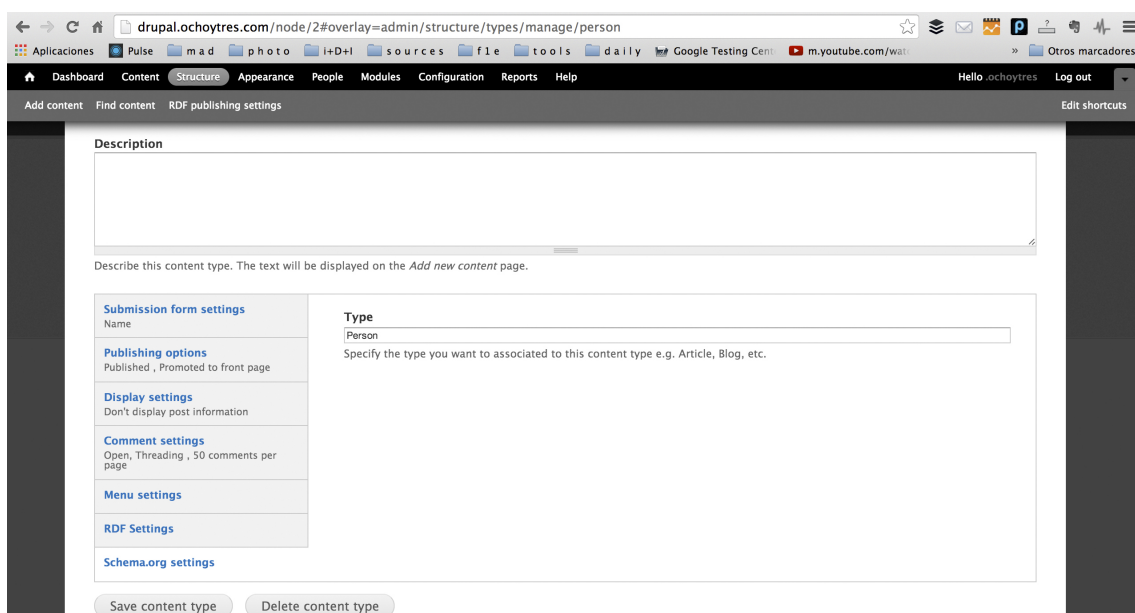


Figura 4.4: Vista de creación de Person mediante Schema.org para Drupal

Está basado en Open Calais [<http://www.opencalais.com>], servicio web basado en técnicas de PLN para el marcado semántico de contenidos web. El funcionamiento, al igual que el de otros SaaS de este tipo, consiste en la misma pauta que en el envío de un texto al servicio web y éste, tras analizarlo, localiza —únicamente— entidades, hechos y eventos, como es posible observar en la figura 4.6 Es un servicio gratuito limitado a 50.000 transacciones diarias. Tras el análisis, OpenCalais devuelve un conjunto de etiquetas que podemos incorporar a nuestro sitio web, devolviendo código html marcado con microformatos o declaraciones RDF en los formatos json, N3 (notation3) o simple format.

6. **COM J4Schema para Joomla**<sup>9</sup> Extensión de Joomla que permite realizar marcado semántico con microdatos utilizando Schema.org. Es un add-on del editor JCE, por lo que su uso está supeditado al cambio del editor, como puede verse en la figura 4.7. Existe una versión PRO, que permite añadir microdatos en otros tipos de contenido más allá de los artículos, y también asocia el nombre de editor con el perfil de Google+ entre otras características.

7. **Krizalys Breadcrumb para Joomla**<sup>10</sup>

Krizalys Breadcrumbs es un módulo para Joomla que permite integrar BreadCrumbs y Rich Snippets en el contenido. El marcado semántico puede realizarse en Microdatos y RDFa como se puede observar en el panel de configuración de la figura 4.8.

<sup>9</sup><http://extensions.joomla.org/extensions/site-management/seo-a-metadata/meta-data/20671>

<sup>10</sup><http://extensions.joomla.org/extensions/structure-a-navigation/site-navigation/23316>

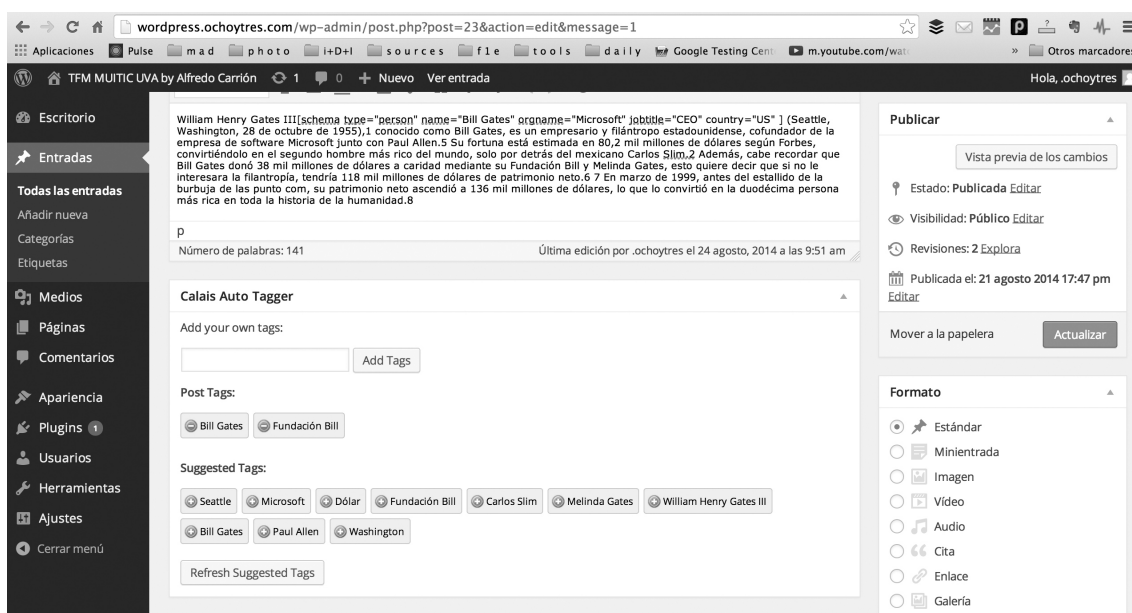


Figura 4.5: Backend del plugin Open Calais para Wordpress

## 4.2. Validación del Marcado

Si bien el objeto del presente estudio no es la validación del marcado, se ha optado por realizarla sobre algunos de los contenidos marcados semánticamente para verificar el formato del marcado y otras funcionalidades evaluadas en las herramientas. Para ello se ha optado por el uso del validador de Yandex<sup>11</sup> (vease figura 4.9 (uno de los buscadores implicados en el proyecto Schema.org)) por su facilidad de uso y capacidad para validar cualquier tipo de formato de marcado semántico, ofreciendo el código completo para su revisión.

Para realizar esta tarea se han dispuesto 3 subdominios, cada uno de ellos con uno de los 3 CMS seleccionados previamente, Wordpress<sup>12</sup>, Joomla<sup>13</sup> y Drupal<sup>14</sup>. En cada uno de ellos se han generado varios artículos sobre los que se han aplicado las diferentes herramientas.

En el caso de RDFaCE de Wordpress se ha realizado una prueba con dicha herramienta sobre una entrada referente a Alan M. Turing como puede verse en la figura 4.10 y al realizar la validación se ha obtenido la siguiente respuesta en el validador de Yandex 4.11.

<sup>11</sup><http://webmaster.yandex.com/microtest.xml>

<sup>12</sup>[wordpress.ochoytres.com](http://wordpress.ochoytres.com)

<sup>13</sup>[joomla.ochoytres.com](http://joomla.ochoytres.com)

<sup>14</sup>[drupal.ochoytres.com](http://drupal.ochoytres.com)



Figura 4.6: Esquema del Funcionamiento de Open Calais

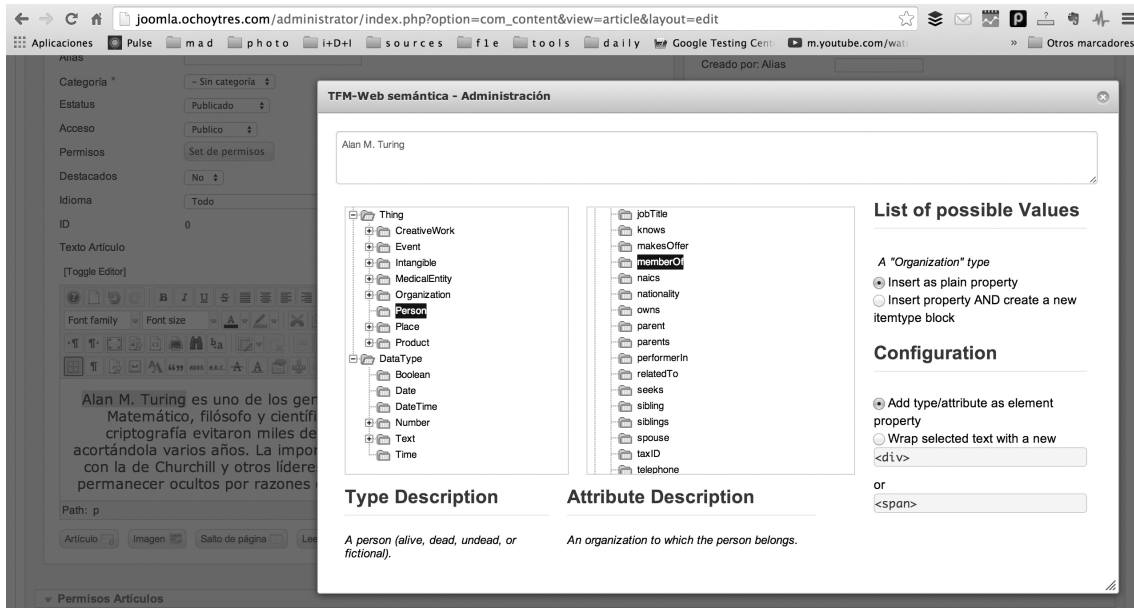


Figura 4.7: Vista del backend de J4Schema para Joomla!

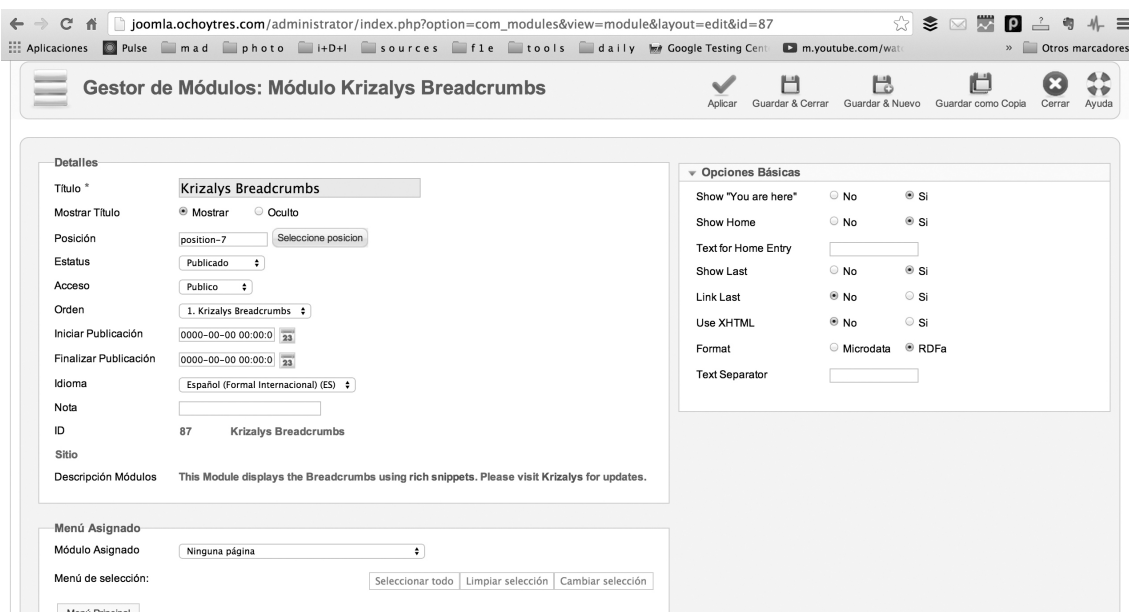
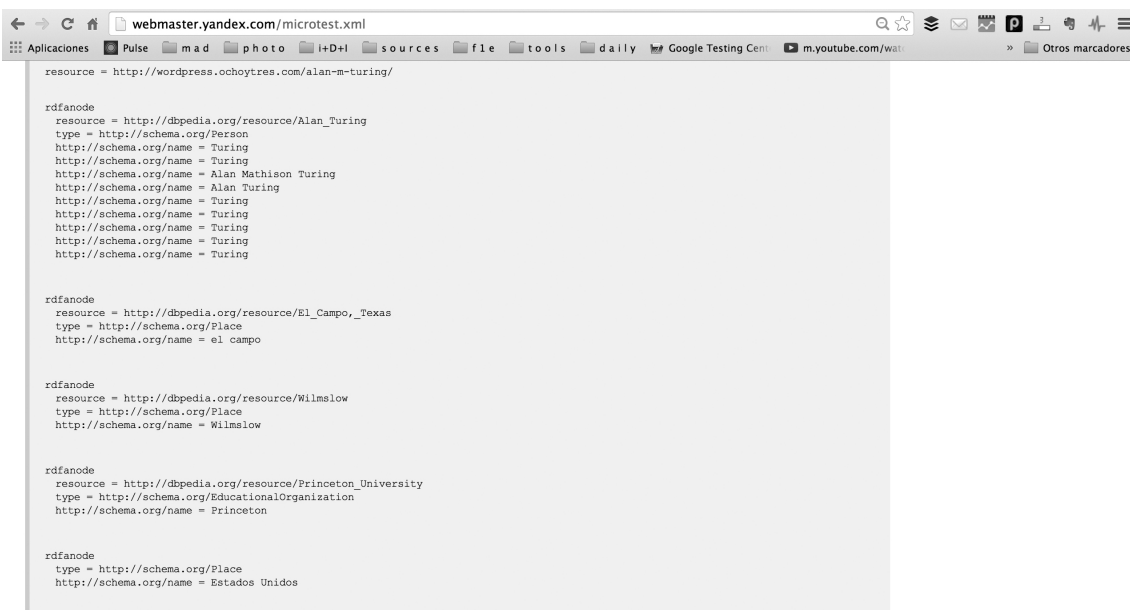


Figura 4.8: Vista del backend de Krizalys Breadcrumbs para Joomla!





```
resource = http://wordpress.ochoytres.com/alan-m-turing/

rdfanode
resource = http://dbpedia.org/resource/Alan_Turing
type = http://schema.org/Person
http://schema.org/name = Turing
http://schema.org/name = Turing
http://schema.org/name = Alan Mathison Turing
http://schema.org/name = Alan Turing
http://schema.org/name = Turing
http://schema.org/name = Turing
http://schema.org/name = Turing
http://schema.org/name = Turing
http://schema.org/name = Turing

rdfanode
resource = http://dbpedia.org/resource/El_Campo,_Texas
type = http://schema.org/Place
http://schema.org/name = el campo

rdfanode
resource = http://dbpedia.org/resource/Wilmslow
type = http://schema.org/Place
http://schema.org/name = Wilmslow

rdfanode
resource = http://dbpedia.org/resource/Princeton_University
type = http://schema.org/EducationalOrganization
http://schema.org/name = Princeton

rdfanode
type = http://schema.org/Place
http://schema.org/name = Estados Unidos
```

Figura 4.11: Validación de la entrada de RDFaCE en Wordpress con Yandex



# Capítulo 5

## Resultados de la Evaluación

A continuación se presentan los resultados tras la aplicación de los criterios evaluativos propuestos anteriormente. Previamente se detallan las versiones utilizadas de cada herramienta en el momento del estudio, con objeto de guiar en futuros estudios.

### 5.1. Descripción y funcionalidades de las herramientas analizadas

Se parte de dos de los criterios definidos previamente, Propósito de la aplicación y Requisitos del sistema evaluados en cada una de las herramientas junto con descripción y funcionalidades, otro de los criterios marcados inicialmente. Respecto a los requisitos, tomaremos como resultados los expuestos previamente en el estudio de Sistemas de Gestión de Contenidos, debido a que este indicador se impone en una capa superior.

#### 5.1.1. AlchemyTagger para Wordpress

Es un plugin de Wordpress que sirve de puente entre la API de Alchemy el Sistema de Gestión de Contenidos. Su uso es gratuito y requiere el registro para la obtención de una clave de la API.

**Especificaciones de la herramienta en la fecha de realización del estudio:**

Ver.	Wordpress	Última Actualización	Descargas
1.1.4	desde 2.5.1	2013-3-18	3891

#### 5.1.2. RDFaCE para Wordpress

RDFaCE es un plugin para Wordpress que sirve de puente entre la API de RDFaCE y el Sistema de Gestión de Contenidos. Combina el marcado manual con el automático en el propio cuerpo del contenido y permite hacerlo en formato RDFa o microdatos.

**Especificaciones de la herramienta en la fecha de realización del estudio:**

Ver.	Wordpress	Última Actualización	Descargas
0.70	desde 3.9.2	2014-6-6	2300

### 5.1.3. Schema Creator para Wordpress

Es un plugin de Wordpress que permite el marcado mediante microdatos de forma manual. Es gratuito.

**Especificaciones de la herramienta en la fecha de realización del estudio:**

Ver.	Wordpress	Última Actualización	Descargas
1.05	desde 3.4	2013-8-8	52101

### 5.1.4. Schema.org para Drupal

Es un módulo de Drupal que soporta marcado semántico en formato RDFa 1.0. No es compatible con el formato de microdatos pese a que utiliza como vocabulario Schema.org.

**Especificaciones de la herramienta en la fecha de realización del estudio:**

Ver.	Drupal	Última Actualización	Descargas
beta4+4	desde desde 7.0	2014-7-20	120618

### 5.1.5. Open Calais para Drupal

Este módulo presenta una API flexible para integrar el Web Service de Open Calais. Es una función basada en una API orientada a objetos.

**Especificaciones de la herramienta en la fecha de realización del estudio:**

Ver.	Drupal	Última Actualización	Descargas
1.05	desde 3.4	2013-8-8	52101

### 5.1.6. COM J4Schema para Joomla

Extensión de Joomla que permite realizar marcado semántico con microdatos utilizando Schema.org. Es un add-on del editor JCE, por lo que su uso está supeditado al cambio del editor. Existe una versión PRO, que permite añadir microdatos en otros tipos de contenido más allá de los artículos, y también asocia el nombre de editor con el perfil de Google+ entre otras características.

**Especificaciones de la herramienta en la fecha de realización del estudio:**

Ver.	Joomla!	Última Actualización	Descargas
beta4+4	desde desde 7.0	2014-7-20	120618

### 5.1.7. Krizalys Breadcrumb para Joomla

Krizalys Breadcrumbs es un módulo para Joomla que permite integrar BreadCrums y Rich Snippets en el contenido. El marcado semántico puede realizarse mediante Microdatos y RDFa.

Krizalys Breadcrumbs cumple, según su desarrollador Krizalys<sup>1</sup> con las normas establecidas por las principales entidades de Internet:

- RDFa Core 1.1, por el W3C
- Rich Snippets (microdatos, microformatos, RDFa y Data Highlighter), por Google
- Microdatos - HTML estándar, por el W3C

**Especificaciones de la herramienta en la fecha de realización del estudio:**

Ver.	Joomla!	Última Actualización	Descargas
1.03	desde desde 7.0	2013-11-14	desconocido

Como se ha visto anteriormente, en el trabajo de Carolina Martínez[39] se propone que la aplicación del marcado semántico en los CMS cubre cinco funciones principales que, como se presentó previamente, son:

- Integración de contenidos en el ámbito de las redes sociales mediante la aplicación del protocolo Open Graph [43]. Como se aprecia en la figura5.1, ninguna de las aplicaciones seleccionadas gestiona directamente esta funcionalidad. No obstante, debido a que OGP es un protocolo basado en RDFa, se posibilita la introducción de este tipo de semántica en cualquier tipo de documento mediante algunos de ellos de una manera sencilla. Gracias al OGP, cualquier página web puede ser integrada en el grafo social, ganando así la misma funcionalidad que otros objetos gráficos. Existen diferentes servicios (no presentes en el estudio), similares a Open Calais, que ofrecen funcionalidades similares.
- Inclusión de sistemas de navegación del tipo breadcrumbs que, indican la ruta de navegación recorrida por el usuario en el sitio web y además ubica un contenido en la estructura jerárquica del sitio web. Es una funcionalidad presente en la herramienta evaluada de Joomla!, Krizalys BreadCrums y opcional, según se ha podido apreciar el varios foros oficiales, para RDFaCE. El resto de herramientas no presentan directamente esta funcionalidad si bien es cierto que, existen en todos los CMS a estudio, opciones compatibles con el resto, que si lo permiten.
- Adición de descripciones estructuradas de contenidos para el uso de la tecnología Rich Snippets de Google. Funcionalidad presente en la gran mayoría de las herramientas a evaluación y que está muy alineada con el gran interés actual con todo lo relacionado con el posicionamiento orgánico web (Search Engine Optimization o SEO).

<sup>1</sup><http://www.krizalys.com/book/references>

- Marcado semántico del cuerpo de una página web. En relación a otra de las características evaluadas, se ha podido concluir que todas las herramientas realizan el marcado en el propio cuerpo de la página.
- Incorporación de metadatos descriptivos para mejorar el Posicionamiento SEO. En todos los casos se aprecia que el marcado semántico en uno u otro formato influye positivamente en el posicionamiento SEO al aportar metadatos descriptivos sobre el contenido publicado.

A continuación, en la figura 5.1, se muestran, a modo sumario, los resultados obtenidos respecto a las funcionalidades esperadas por las herramientas evaluadas.

	Integración con Open Graph	Integración con BreadCrumb	Rich Snippets	Marcado del Cuerpo	Metadatos SEO
<b>FUNCIONALIDADES BÁSICAS</b>					
AlchemyTagger	No	No	No	Sí	Sí
RDFaCE	No	OPC.	OPC.	Sí	Sí
Schema Creator	No	No	OPC.	Sí	Sí
Schema.org	No	No	Sí	Sí	Sí
Open Calais	No	No	Sí	Sí	Sí
COM J4Schema	No	No	Sí	Sí	Sí
Krizalys BreadCrumbs	No	Sí	Sí	Sí	Sí

Figura 5.1: Tabla Comparativa de Funcionalidades esperadas por herramienta

## 5.2. Tipología de Marcado Semántico

Tras el análisis, se aprecia que existen aplicaciones que ofrecen marcado semántico con nivel de anotación tanto manual como automático, permitiendo complementar uno con otro. La tendencia deseada es la obtención de un grado de automatización máxima, respaldada por una fase opcional de validación y complemento por parte del usuario.

En lo que al lugar de almacén se refiere, se presentan dos opciones, el almacén embebido o el almacén externo. Todas las herramientas evaluadas optan por embeber el marcado en el propio documento web, por lo que parece ser la norma más extendida en este tipo de herramientas.

Un ejemplo de aplicación —no evaluada por no considerarse de interés para el presente estudio— es Annotea <sup>2</sup> Proyecto que pretende mejorar el ambiente colaborativo de la W3C a través del uso de anotaciones compartidas que se puedan adjuntar a un documento Web externamente, es decir, sin necesidad de tocar el documento mismo. En la figura 5.2 puede verse a modo de tabla, los resultados obtenidos.

<sup>2</sup><http://www.w3.org/2001/Annotea/>

	Embebido	Externo	Manual	Automático
TIPOLOGÍA DE MARCADO SEMÁNTICO				
AlchemyTagger	Sí	No	OPC.	Sí
RDFaCE	Sí	No	Sí	Sí
Schema Creator	Sí	No	Sí	No
Schema.org	Sí	No	Sí	No
Open Calais	Sí	No	No	Sí
COM J4Schema	Sí	No	Sí	Sí
Krizalys BreadCrumbs	Sí	No	Sí	Sí

Figura 5.2: Tabla Comparativa de Tipologías de Marcado por herramienta

### 5.3. Integrabilidad del Software

Se evalúa esta característica teniendo en cuenta la provisión de paquetes de software o servicios. Algunas de las herramientas que conforman el soporte de marcado semántico de los principales Sistemas de Gestión de Contenidos (CMS), pueden ser reutilizables por medio de APIs que ofrecen un conjunto de librerías, provistas de interfaces que describen las definiciones de clase y su comportamiento (en el caso de los lenguajes orientados a objetos) y/o Servicios Web, pudiendo integrarse en varios de los mismos simultáneamente debido al soporte, en forma de módulos, extensiones, plugins, addons... dados por la propia comunidad.

En la figura 5.3 pueden verse los resultados obtenidos. Más adelante centraremos esta integrabilidad exclusivamente en el ámbito de los Sistemas de Gestión de Contenidos, que nos permitirá ver la aceptación generalizada de determinadas herramientas en el entorno de los CMS a estudio.

En el caso de Open Calais así como en el de Alchemy y RDFaCE presentan el mayor grado de integrabilidad ofreciendo soporte a todo tipo de soluciones a través de API, plugins o mediante su soporte SaaS.

### 5.4. Formatos de Marcado Semántico Admitidos

En general la tendencia en lo que a formatos de marcado semántico en Sistemas de Gestión de Contenidos se refiere, parece ir hacia RDFa y sobre todo hacia el estándar de los Microdatos. Como se ha visto hasta ahora, los Microdatos ganan enteros debido a la gran implicación de Google y otros buscadores en ello.

	SaaS	API	API+Plugin	Plugin
INTEGRABILIDAD DEL SOFTWARE				
AlchemyTagger	Sí	Sí	Sí	Sí
RDFaCE	Sí	Sí	Sí	Sí
Schema Creator	No	No	No	Sí
Schema.org	No	No	No	Sí
Open Calais	Sí	Sí	Sí	Sí
COM J4Schema	No	No	No	Sí
Krizalys BreadCrumbs	No	No	No	Sí

Figura 5.3: Tabla Comparativa de la Integrabilidad de cada herramienta

En la figura 5.4 puede verse a modo de tabla, los resultados obtenidos. Los valores han sido obtenidos de la propia documentación del desarrollador verificando dicha información en la instalación de los mismos sobre los Sistemas de Gestión de Contenidos de prueba y/o a través del validador.

Destaca la versatilidad de la herramienta Krizalys, que acepta todos los formatos de marcado del mercado.

	RDFa	Microdatos	Microformatos
FORMATOS DE MARCADO SEMÁNTICO ADMITIDOS			
AlchemyTagger	No	No	Sí
RDFaCE	Sí	Sí	No
Schema Creator	Sí	Sí	No
Schema.org	Sí	No	No
Open Calais	Sí	No	Sí
COM J4Schema	No	Sí	No
Krizalys BreadCrumbs	Sí	Sí	Sí

Figura 5.4: Tabla Comparativa de Formatos de Marcado admitidos por herramienta.

Como se presentó anteriormente este factor es de vital importancia por su vinculación con los posibles objetivos del usuario. Las implicaciones que conllevan los formatos —vistas previamente a modo resumen en la figura 2.3— serán adquiridas por las propias herramientas a estudio, por lo que la información que subyace en este punto, pese a su simplicidad, es clave.

## 5.5. Soporte en diferentes CMS

En general, las herramientas que se presentan como plugins o extensiones nexos entre el CMS y una API, son las que mayor presencia tienen en los diferentes Sistemas de Gestión de Contenidos.

Es el caso de Alchemy o Open Calais, que son Servicios Web que disponen de API y su soporte está presente a través de plugins y extensiones en los 3 Sistemas de Gestión de Contenidos del estudio.

Esto parece indicar que el futuro del mercado semántico pasará por la gestión de servicios web con un alto grado de integrabilidad que aporte un interés en las comunidades y les invite a dar soporte mediante extensiones o plugins.

En la figura 5.5 puede verse a modo de tabla, los resultados obtenidos, sacados todos ellos de la documentación de cada herramienta y los foros de las comunidades de cada uno de los Sistemas de Gestión de Contenidos.

	Wordpress	Joomla	Drupal
<b>SOPORTE A CMS</b>			
AlchemyTagger	Sí	Sí	Sí
RDFaCE	Sí	Sí	Sí
Schema Creator	Sí	Sí	No
Schema.org	No	No	Sí
Open Calais	Sí	No	Sí
COM J4Schema	No	Sí	No
Krizalys BreadCrums	No	No	Sí

Figura 5.5: Tabla Comparativa del soporte en diferentes CMS





## Capítulo 6

# Conclusiones y Trabajo Futuro

Como se ha presentado al inicio del documento, el marcado semántico de documentos es el primer paso para permitir el procesamiento automático de la Web y, por tanto, para la creación de la Web Semántica. La importancia del marcado semántico en el desarrollo de la Web Semántica se complementa a la perfección con el peso específico de los Sistemas de Gestión de Contenidos en el ámbito de la Web, factor que, como se ha podido ver, va en claro aumento.

En este binomio, se centra el objeto marcado al inicio del presente estudio, el de aportar una propuesta metodológica para evaluar y comparar el soporte de marcado semántico en los Sistemas de Gestión de Contenidos.

Primeramente, se ha realizado una revisión de la bibliografía para conocer el estado del arte en la materia, contextualizando los dos términos del binomio.

A continuación, se ha realizado un análisis comparativo del estado actual de los Sistemas de Gestión de Contenidos Web. A través de una propuesta comparativa detallada, se han seleccionado 3 de entre más de 1200 CMS para realizar una propuesta metodológica que permitiera evaluar y comparar el soporte de marcado semántico en ellos.

Como se ha visto en el presente estudio la inclusión de datos estructurados a través del marcado semántico, pese a presentar grandes opciones a nivel de interoperabilidad semántica y reutilización de la información, no posee aún un estándar claro en cuanto a formato de marcado semántico a utilizar, generando una situación de dispersión que afecta a la eficiencia y desarrollo del mismo y por ende, a la evolución del soporte semántico que ofrecen los Sistemas de Gestión de Contenidos —objeto principal del estudio—.

Posteriormente, se han presentado los diferentes formatos de marcado semántico donde, por un lado, los principales motores de búsqueda apuestan por los por los microdatos y Schema.org. mientras por otro, se defiende el formato del W3C y la mayoría de tecnologías de la Web Semántica, el RDFa, más expresivo desde el punto de vista de formalización, granularidad semántica y potencialidad [16].

Google es, actualmente, el paradigma establecido en lo que a motores de búsqueda se

refiere y por tanto su influencia en el SEO y todo lo que ello conlleva, parecen suficientes argumentos para un establecimiento de los microdatos como estándar de facto. RDFa presenta grandes ventajas para la representación de la semántica de los contenidos con mayor detalle y la generación automática de conjuntos de datos RDF a partir de contenidos web.

Se ha podido comprobar, dada la trayectoria de desarrollo de soporte en los diferentes CMS, que la tendencia de formato de marcado que se logre instaurar, influirá decisivamente en el desarrollo de este soporte semántico por parte de las comunidades de los diferentes Sistemas de Gestión de Contenido. Los servicios web como Open Calais, tienen mucho que decir en el desarrollo del marcado semántico y su integrabilidad a través de API y plugins se presume decisiva en la expansión de la información estructurada que permita cumplir las bases de la Web Semántica.

Asimismo se ha apreciado que, a nivel cuantitativo, el soporte de Wordpress es el mayor de los tres comparados, coincidiendo con la cuota de uso del CMS. A nivel de implementación en el Core y, probablemente, a nivel cualitativo, la opción más completa parece Drupal, que presenta, a priori, una comunidad con un registro técnico mayor que el resto de propuestas.

Por último, es importante tener en cuenta la nueva recomendación del W3C <sup>1</sup>, el formato JSON-LD que fue publicada por el W3C el 16 de enero de 2014 y adoptada por Google desde hace unas semanas como alternativa a los microdatos para implementar en sitios web. Perfiles clave en la literatura del marcado semántico en CMS como Lin Clark<sup>2</sup>, presentan información de interés al respecto.

Como trabajo futuro se plantea la evolución de la presente propuesta adaptándola a los formatos y herramientas más actuales a la vez que se definen nuevos indicadores que aporten mayor nivel de detalle sobre la evolución del soporte de marcado semántico, presente en los Sistemas de Gestión de Contenidos.

Asimismo, parece interesante la actualización del estado del arte tanto en Sistemas de Gestión de Contenidos como en formatos de marcado, con atención especial a la nueva recomendación comentada anteriormente.

Desde el punto de vista de Trabajo Final de Máster, ha servido al alumno para realizar un trabajo de investigación de principio a fin. Se ha pasado por diferentes fases, desde la previa, de Análisis de la bibliografía para conocer el estado del arte en la materia, pasando por las fases de diseño, ejecución y análisis para, finalmente, cerrar el ciclo del Trabajo con la fase de Documentación.

El siguiente paso será identificar algún foro científico para intentar contribuir, en mayor o menor grado, a la comunidad científica y se obtendrá feedback de expertos en este campo.

---

<sup>1</sup><http://www.w3.org/TR/json-ld/>

<sup>2</sup><http://lin-clark.com/>

# Bibliografía

- [1] Berners-Lee, T., Hendler, J., & Lassila, O. (2001). The semantic web. *Scientific american*, 284(5), 28-37.
- [2] J. Hendler, T. Berners-Lee, and E. Miller, "Integrating Applications on the Semantic Web," *Journal of the Institute of Electrical Engineers of Japan*, vol. 122, pp. 676-680, October 2002.
- [3] Shadbolt, N., Hall, W., & Berners-Lee, T. (2006). The semantic web revisited. *Intelligent Systems, IEEE*, 21(3), 96-101.
- [4] Bizer, C., Heath, T., & Berners-Lee, T. (2009). Linked data The story so far. *International journal on semantic web and information systems*, 5(3), 1-22.
- [5] García-Castro, R. (2006). *Tecnologías de la Web Semántica: Cómo funcionan y cómo interoperan*.
- [6] Boiko 2005: Boiko, B. (2005). *Content management bible*. John Wiley & Sons.
- [7] Saorín, T., & Pástor-Sánchez, J. A. (2012). *Gestión documental y de contenidos web: informe de situación*. *Anuario ThinkEPI*, 6, 232-239.
- [8] Navarro-Galindo, J. L., & Samos, J. FLERSA: Soporte a la Definición de Anotaciones y Búsquedas Semánticas en un CMS.
- [9] Moya Martínez, G., & Gil-Leiva, I. (2001). Evaluación de software de gestión de tesauros. *Ciencias de la Información*.
- [10] García-Castro, R., & Gómez-Pérez, A. (2009). *Benchmarking in the Semantic Web*.
- [11] Peñalvo, F. J. G. (2006). *Web Semántica y Ontologías*. Departamento de Informática y Automática. Facultad de Ciencias, Universidad de Salamanca.
- [12] Castells, P. (2003). *La web semántica. Sistemas interactivos y colaborativos en la web*, 195-212.
- [13] Fonseca, J. M. C., Sureda, J. J. H., & Zabala, P. Á. R. (2006). *Web semántica: Tecnologías y arquitectura*. *Comunicaciones de Telefónica I+ D*, (39), 211-221.
- [14] Noy, N. F., McGuinness, D. L. (2001) *Ontology Development 101: A Guide to Creating Your First Ontology*. Stanford Knowledge Systems Laboratory Technical Report KSL-01-05 and Stanford Medical Informatics Technical Report SMI-2001-0880.

- [15] Pastor-Sánchez, Juan-Antonio (2011). *Tecnologías de la web semántica*. Barcelona: Editorial UOC, colección *El profesional de la información*, 1. ISBN: 978 84 9788 474 7.
- [16] Pastor-Sánchez, J. A., Orduña-Malea, E., & Saorín, T. (2013). Mercado semántico automático en gestores de contenidos: integración y cuantificación. *El profesional de la información*, 22(5), 381-391.
- [17] Bray, T., Paoli, J., Sperberg-McQueen, C., Maler, E., Yergeau, F.: *Extensible markup language (xml) 1.0 (fourth edition) w3c recommendation (2006)*
- [18] Berners-Lee, T.: *Weaving the Web: The Original Design and Ultimate Destiny of the World Wide Web*. HarperCollins, New York, NY (1999)
- [19] Primer, R. D. F., Manola, F., & Miller, E. (2007). *W3C Recommendation 10 February 2004*.
- [20] Brickley, D., & Guha, R. V. (2000). *Resource Description Framework (RDF) Schema Specification 1.0: W3C Candidate Recommendation 27 March 2000*
- [21] McGuinness, D. L., & Van Harmelen, F. (2004). *OWL web ontology language overview. W3C recommendation, 10(10), 2004*.
- [22] García Jiménez A. Instrumentos de representación del conocimiento: tesauros versus ontologías. *Anales de la Documentación 2004*;7(7):7.
- [23] Sheth, A., Bertram, C., *Avant: Managing semantic content for the web. IEEE Internet Computing 6, 8087 (July 2002)*.
- [24] De Cea, G. A., Álvarez de Mon, I., & Lora, R. A. P. (2009). Una visión interdisciplinar de la anotación semántica. In *Terminología y sociedad del conocimiento* (pp. 219-254). Peter Lang.
- [25] Tsai, T.M., Yu, H.K., Liao, P.Y., Shih, H.T.: *Semantic modeling among web services interfaces for services integration. In: Proceedings of the 14th International Workshop on Database and Expert Systems Applications. pp. 579 DEXA '03, IEEE Computer Society, Washington, DC, USA (2003)*
- [26] Bizer, C., Eckert, K., Meusel, R., Mühleisen, H., Schuhmacher, M., & Völker, J. (2013). Deployment of rDfA, microdata, and microformats on the web—A quantitative analysis. In *The Semantic Web—ISWC 2013* (pp. 17-32). Springer Berlin Heidelberg.
- [27] Méndez, E., Bravo, A., & López, L. M. (2007). Microformatos: web 2.0 para el Dublin Core. *El profesional de la información*, 16(2), 107-113
- [28] Adida, B., Herman, I., Sporny, M., & Birbeck, M. (2012). *RDFa 1.1 Primer. 20re12-06 07*[2013-07-22]. <http://www.w3.org/TR/xhtml1-rdfa-primer>.
- [29] Marco, F. J. G. (2013). Schema.org: la catalogación revisitada. *Anuario ThinkEPI*, (1), 169-172.

- [30] Khalili, A., & Auer, S. (2013). Wysiwym authoring of structured content based on schema. org. In *Web Information Systems Engineering–WISE 2013* (pp. 425-438). Springer Berlin Heidelberg.
- [31] Bernardis, S. (2014). Estudio del impacto del SEO. semántico en los motores de búsqueda: aplicando microdatos y RDFa Lite 1,1 en el ámbito de Schema. org. *Cuadernos de Gestión de Información.*, 3(1), 85-104.
- [32] Sporny, M. (2011). An uber-comparison of RDFa, microdata and microformats. *The beautiful, tormented machine.*
- [33] Browning, P., & Lowndes, M. (2001). JISC TechWatch Report: Content Management Systems. Techwatch report TSW, 01-02.
- [34] Michelinakis, D. (2004). *Open Source Content Management Systems: An Argumentative Approach.* The University of Warwick: Warwick Manufacturing Group, A report submitted for the award of MSc Electronic Business Management.
- [35] Water & Stone. 2011 Open Source CMS Market Share Report: 4th annual report on the industry. Water & Stone, November 2011. <http://www.waterandstone.com/book/2011-opensource-cms-market-share-report>
- [36] ISO/IEC 14598-1 (1999). *Information Technology – Software Product Evaluation. Part 1: General Overview.* Secretaría General de ISO. Ginebra.
- [37] Wohlin, C., Aurum, A., Petersson, H., Shull, F., & Ciolkowski, M. (2002). Software inspection benchmarking-a qualitative and quantitative comparative opportunity. In *Software Metrics, 2002. Proceedings. Eighth IEEE Symposium on* (pp. 118-127). IEEE.
- [38] González, M. M. M., & Díez, M. L. A. (2014). Propuesta metodológica de evaluación de gestores de tesauros compatibles con la web semántica. En *Anales de documentación: Revista de biblioteconomía y documentación* (Vol. 17, No. 1, pp. 4-18). Servicio de Publicaciones.
- [39] Martínez Martínez, Carolina (2013). *El Mercado Semántico en Wordpress.* Mucia: Facultad de Comunicación y Documentación (Universidad de Murcia), 2013. Trabajo de Fin de Grado.
- [40] Mell, P., & Grance, T. (2011). *The NIST definition of cloud computing.*
- [41] [Haas, H., & Brown, A. (2004). *Web services glossary.* W3C Working Group Note (11 February 2004).
- [42] Khalili, A., & Auer, S. *RDFaCE-Lite: a WYSIWYM editor for user-friendly semantic text authoring.*
- [43] [Ko, M. N., Cheek, G. P., Shehab, M., & Sandhu, R. (2010). *Social-networks connect services.* *Computer*, 43(8), 37-43.].