Wiki semática para recopilar información operativa.

20130108 (revisado 20190916)

En el día a día del trabajo de un departamento cualquiera de una empresa, se genera mucha información interesante sobre los aspectos operativos de lo que se va haciendo. La parte mas formal de esta información suele quedar registrada en los sistemas de información de la empresa: ERP, archivos en servidores, PDM, etc. Pero otra parte suele residir tan solo en la cabeza o en las notas personales de cada persona.

Esa parte informal puede resultar, en no pocas ocasiones, tanto o más útil que la formal. Pero, precisamente por esa falta de registro formal, suele ser difícil de recuperar cuando se necesita. Sobre todo cuando las personas que estuvieron directamente involucradas no están disponibles o no se acuerdan de lo que hicieron.

Por eso, la solución que se propone es ir recopilando, sobre la marcha, lo que se pueda, de manera informal, pero en un repositorio concreto; para conservar al menos parte de la información ("más vale poco que nada."). El sistema utilizado para registrar la información ha de ser sencillo y sin excesivas reglas, para fomentar la escritura. Pero ha de contar con algún tipo de estructura, para facilitar posteriores búsquedas. En ese sentido, una herramienta capaz de satisfacer ambos requisitos podria ser una wiki (o, mejor aún, una wiki semática).



clic aqui para ver la licencia

Juan Murua Olalde

documento derivado del trabajo de fin de carrera "Aplicación de las wikis semáticas como herramienta de gestión del conocimiento operativo en un departamento de TICs" (UOC.edu - repositorio O2)

Índice

1.	Bre	ve resumen de conceptos y tecnologías semánticas	4
	1.1.	Datos almacenados en una BD relacional	4
	1.2.	vs. datos almacenados de forma semántica	5
	1.3.	Principales estándares semánticos	6
		URI (U niform R esource I dentifier) (http://tools.ietf.org/html/rfc3986)	6
		$ \begin{array}{lll} \text{RDF } (R \text{esource } D \text{escription } F \text{ramework}) \\ & (\text{http://www.w3.org/TR/rdf-mt/}) \\ & . \\ \\ & . \\ & . \\ & . \\ & . \\ & . \\ & . \\ & . \\ & . \\ \\ & . \\ & . \\ \\ & . \\ \\ & . \\ \\ & . \\ \\ & . \\ \\ & . \\ \\ & . \\ \\ & . \\ \\ & . \\ \\$	7
		RDFS (RDF Schema) (http://www.w3.org/TR/rdf-schema/)	7
		$\begin{array}{cccccccccccccccccccccccccccccccccccc$	7
			7
	1.4.	Formas de representar información almacenada en un sistema semántico	8
		Grafos	8
		RDF/XML (http://www.w3.org/TR/rdf-syntax-grammar/)	8
		$ \begin{array}{c} {\rm Notation 3 \ (N3)} \\ {\rm (http://www.w3.org/TeamSubmission/n3/)} \\ \end{array} \\ {\rm $	S
		N-Triples	
		$(\text{http://www.w3.org/TR/rdf/testcases/\#ntriples}) \dots \dots \dots \dots \dots \dots \dots \dots \dots $	(
		Turtle (Terse RDF Triple Language) (http://www.w3.org/TeamSubmission/turtle/)	
	1.5.	Formas de embeber información semántica en páginas web	10
		$\label{eq:microformatos} \mbox{(http://microformats.org/wiki/Main_Page)} . \ . \ . \ . \ . \ . \ . \ . \ . \ .$	10
		$\label{linear_model} \begin{tabular}{ll} Microdatos & (schema.org) (http://schema.org/docs/gs.html) & $	10
			11
2.	\mathbf{Bre}	ve descripción de las herramientas software utilizadas	12
	2.1.	Instalación	13
		Prerequisitos:	13
		La wiki propiamente dicha, Media Wiki (MW)	13
		La extensión semántica, Semantic Media Wiki (SMW)	13
	2.2.	Configuración básica	14
3.	\mathbf{Bre}	ve manual de usuario SMW (Semantic Media Wiki)	15
	3.1.	Escribir contenido	15
	3.2.	Introducir anotaciones semánticas	15
		Categorías:	15
		Propiedades:	15
		Formularios:	16
	3.3.	Exportar información en formato semántico	16
	3.4.	Utilizar la información almacenada	17

4.	Una	o posible forma de trabajar para recopilar información operativa	19
	4.1.	Necesidades a cubrir según tipo de usuario	19
	4.2.	Un par de recomendaciones generales respecto a la forma de trabajar $\dots \dots \dots$	19
	4.3.	Recomendaciones respecto a la Ontologia	20
	4.4.	Recomendaciones respecto al modelo de datos	20
	4.5.	Un ejemplo práctico:	21
5.	Glos	sario	27
6.	Bib	liografía	27

1. Breve resumen de conceptos y tecnologías semánticas

La mayor parte de la información almacenada en sistemas digitales está pensada para su consumo por humanos. Su estructura y metadatos están pensados o bien para guiar la lectura/búsqueda por parte de una persona o bien para guiar aplicaciones expresamente diseñadas (por humanos) para explotar esa información.

Sin embargo, si se dotara a cada pieza de información de algún tipo de indicación estandarizada de a qué ente o sujeto se refiere y qué dato aporta sobre ese ente o sujeto, se posibilitaría su tratamiento automatizado por parte de aplicaciones generalistas diseñadas para tratamiento de información en general.

Ilustrándolo con un ejemplo:

1.1. Datos almacenados en una BD relacional

La típica forma de almacenar datos en una BD es en forma de tablas. Para su consulta, se requiere de un conocimiento específico de su estructura interna, las relaciones entre tablas, sus índices, etc.

Usuarios Ubicaciones Departamentos idUsuario idDepartamento idUbicacion Nombre Nombre Nombre Apellidos idUbicacion idDepartamento Equipos TiposDeEquipo idEquipo idTipoEquipo idTipoEquipo Nombre idUsuario Fabricante Modelo Procesador Memoria Disco

Figura 1: Ejemplo de almacenamiento de datos en una BD relacional.

 $SELECT\ Ubicaciones. Nombre,$

Departamentos. Nombre,

Equipos. Fabricante,

Equipos. Modelo

FROM Ubicaciones,

Departamentos,

Usuarios,

Equipos,

TiposDeEquipo

WHERE Ubicaciones.idUbicacion = Departamentos.idUbicacion

AND Departamentos.idDepartamento = Usuarios.idDepartamento

AND Usuarios.idUsuario = Equipos.idUsuario

AND Equipos.idTipoEquipo = Equipos.idTipoEquipo

AND Equipos. Nombre = "Impresora"

GROUP BY Ubicaciones. Nombre, Departamentos. Nombre

1.2. vs. datos almacenados de forma semántica

En un sistema semántico la metainformación para interpretar los datos está almacenada junto con ellos. Es decir tanto los propios datos como los metadatos que los describen están escritos en una misma forma: triples 'sujeto-predicado-objeto'

Figura 2: Ejemplo de datos almacenados en forma semántica

```
ejmpl: TiposDeEquipo rdf: type rdfs: Class.
ejmpl:Impresora rdf:type rdfs:Class;
    rdfs:SubClassOf ejmpl:TiposDeEquipo.
ejmpl:R345WT ejmpl:TieneNumeroDeSerie "R345WT";
    rdf:type ejmpl:Impresora;
    ejmpl: EsFabricadoPor ejmpl: Epson;
    ejmpl: TieneModelo ejmpl: FX 800;
    ejmpl: EstaAsignadoA ejmpl: BenzirpiMirvento.
ejmpl: Epson rdfs: label "Epson".
ejmpl:FX 800 rdfs:label "Epson FX-800: impresora de chorro de tinta";
    ejmpl: EsFabricadoPor ejmpl: Epson.
ejmpl: Benzirpi Mirvento rdf: type ejmpl: Usuario;
    foaf:name "Benzirpi Mirvento";
    foaf: givenName "Benzirpi";
    foaf:familyName "Mirvento";
    foaf:genre "female";
    ejmpl:TrabajaEn ejmpl:TICs central.
ejmpl: TICs central rdf: type ejmpl: Departamento;
    rdfs: label "Dpto. TICs, Oficinas Centrales";
    ejmpl:EstaUbicadoEn ejmpl:Bilbao.
ejmpl: Bilbao rdf: type ejmpl: Ubicacion;
    rdfs:label "Bilbao".
```

De esta manera:

■ Es mas sencillo definir nuevas categorías o nuevas relaciones. Tan solo hay que añadir nuevos triples al almacén.

```
ejmpl: Subordinado rdfs: SubClassOf ejmpl: Usuario.
ejmpl: Jefe rdfs: SubClassOf ejmpl: Usuario.
ejmpl: MandaSobre rdf: type rdfs: Property;
rdfs: domain ejmpl: Jefe;
rdfs: range ejmpl: Subordinado.
```

■ Las consultas son mas naturales.

■ Es posible inferir automáticamente nuevos datos a partir de los datos almacenados. En base a las reglas de inferencia asociadas con algunas de las relaciones indicadas.

Por ejemplo, una aserción como:

```
ejmpl: Benzirpi Mirvento \ ejmpl: Manda Sobre \ ejmpl: Lued by Hert \,.
```

Permite al sistema deducir automáticamente que:

```
ejmpl: BenzirpiMirvento rdf: type ejmpl: Jefe; rdf: Type ejmpl: Usuario.

ejmpl: LuedbyHert rdf: type ejmpl: Subordinado; rdf: Type ejmpl: Usuario.
```

• Se simplifica la utilización de datos provenientes de fuentes distintas. Tan solo se han de considerar conjuntamente todos los triples de todas las fuentes, y se han de ir añadiendo al sistema los triples necesarios para describir las equivalencias entre los datos que se están mezclando.

Por ejemplo, datos sobre impresoras de este estilo:

```
otro: GK4234L otro: TieneEtiquetaCodigo "GK4234L"; rdf: type otro: ImpresoraDeTinta; otro: TieneFabricante otro: Epson; otro: TieneModelo otro: FX_1200; otro: TieneUbicacion otro: Piso2.
```

Se pueden compatibilizar fácilmente con parte de los ya existentes:

```
otro: TieneEtiquetaCodigo owl: equivalentProperty ejmpl: TieneNumeroDeSerie. otro: TieneFabricante owl: equivalentProperty ejmpl: EsFabricadoPor. otro: TieneModelo owl: equivalentProperty ejmpl: TieneModelo. otro: ImpresoraDeTinta rdfs: SubClassOf ejmpl: Impresora.
```

1.3. Principales estándares semánticos

```
egin{aligned} 	ext{URI } & (\textit{U} \, 	ext{niform} \, \textit{R} \, 	ext{esource} \, \textit{I} \, 	ext{dentifier}) \ & (	ext{http:}//	ext{tools.ietf.org/html/rfc3986}) \end{aligned}
```

Es una cadena de caracteres, un nombre propio, que identifica de forma única un ente. Idealmente a nivel mundial, en todo Internet. Es decir, el URI de un ente es el identificador que debiera utilizar todo aquel que quiera referirse a dicho ente para dar cualquier tipo de información acerca del mismo.

Puede ser todo lo corto o todo lo largo que se desee. Normalmente suele venir precedido por un identificador del espacio de nombres donde se ha definido, para evitar confusiones en caso de ser utilizado un mismo nombre para dos entes distintos (por ejemplo: en nuestros anteriores ejemplos "ejmpl:R345WT" se refiere a la impresora con número de serie R345WT; mientras que "film:R345WT" podría referirse a la película de título R345WT en una base de datos sobre filmografía).

Nota: A lo largo de este documento se utilizan siempre nombres cortos para los identificadores de espacios de nombres. Así los ejemplos son más sencillos de leer. Pero en realidad, para evitar que coincidan dos iguales en Internet, los identificadores suelen ser bastante más largos y específicos. Aunque es posible sustituirlos por equivalentes cortos, siempre y cuando se indique expresamente en la cabecera del documento; por ejemplo con sentencias como: "xmlns:ejmpl=http://miejemplo.org" o "@prefix film: <http://rdf.freebase.com/ns/films".

Si en Internet existe alguna página que defina bien el ente en cuestión, lo habitual es utilizar directamente su URL (Uniform Resource Locator), tal cual, como URI de ese ente. También suele ser práctica habitual expresar todos los URIs como si fueran URLs, aunque no lo sean realmente (por ejemplo: http://miejemplo.org/equipos/impresoras/R345WT). Esto puede llevar a confusión; y para evitarlas es conveniente no perder de vista que, por mucho que el URI parezca una dirección web, no hay ninguna obligación de que exista una página web real tras esa dirección.

En la práctica, el URI ideal no suele existir. Siempre cabe la posibilidad de que un mismo ente sea identificado de distintas maneras según las diversas fuentes de la información. En esos casos, hay medios para indicar que dos URIs se refieren en realidad a una misma cosa; por ejemplo con aserciones como:

<http://miejemplo.org/usuarios/BenzirpiMirvento><owl:sameAs><http://www.ieee.org/members/bmirvento>>

```
RDF (Resource Description Framework) (http://www.w3.org/TR/rdf-mt/)
```

Define los elementos básicos para codificar información:

- Las propiedades (rdf:Property) que relacionan datos de unos ciertos tipos (rdf:type).
- Las sentencias (rdf:Statement) que expresan esas relaciones en forma de triples: (rdf:subject) (rdf:predicate) (rdf:object).
- Colecciones de datos. (rdf:Bag) (rdf:Seq) (rdf:Alt) (rdf:value)
- Listas de datos. (rdf:List) (rdf:nil) (rdf:first) (rdf:rest)

```
\begin{array}{l} {\rm RDFS} \,\, (RDF \,\, S{\rm chema}) \\ ({\rm http://www.w3.org/TR/rdf\text{-}schema/}) \end{array}
```

Define los elementos básicos para codificar información semántica, permitiendo expresar aspectos tales como:

- Jerarquía entre tipos de datos. (rdfs:Class) (rdfs:SubClassOf)
- Jerarquía entre propiedades que los relacionan. (rdfs:SubPropertyOf)
- Tipos de datos relacionados por una cierta propiedad. (rdfs:domain) (rdfs:range)
- Mas información acerca de un dato. (rdfs:label) (rdfs:seeAlso)

```
egin{aligned} 	ext{OWL ($W$ eb $O$ ntology $L$ anguage)} \ & (	ext{http:}//	ext{www.w3.org}/	ext{TR/owl-overview}/) \end{aligned}
```

Es un lenguaje que permite modelar completos modelos semánticos acerca de cualquier campo del conocimiento.

Entre las muchas propiedades establecidas por el estándar OWL, algunas de las más utilizadas son:

- Para definir características de las propiedades. (owl:InverseOf) (owl:SymmetricProperty) (owl:TransitiveProperty)
- Para definir qué se corresponde con qué cuando se mezcla información procedente de diversas fuentes. (owl:equivalentClass) (owl:equivalentProperty) (owl:sameAs) (owl:FunctionalProperty) (owl:InverseFunctionalProperty)
- Para definir caracteristicas de entidades y propiedades. (owl:Class) (owl:ObjectProperty) (owl:DatatypeProperty) (owl:XMLLiteral) (owl:AnnotationProperty)
- Para definir restricciones.
 (owl:Restriction) (onProperty) (hasValue) (someValuesFrom) (allValuesFrom)
- Para operaciones de conjunto.
 (owl:unionOf) (owl:intersectionOf) (owl:complementOf) (owl:oneOf) (owl:differentFrom) (owl:disjointWith)
- Para limitar número.
 (owl:Cardinality) (owl:minCardinality) (owl:maxCardinality)

```
\begin{array}{l} {\rm SPARQL} \ (S{\rm PARQL} \ P{\rm rotocol} \ a{\rm nd} \ R{\rm DF} \ Q{\rm uery} \ L{\rm anguage}) \\ ({\rm http://www.w3.org/TR/rdf-sparql-query/}) \end{array}
```

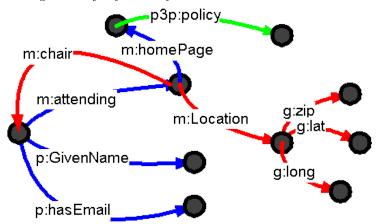
Es un lenguaje que permite realizar consultas sobre almacenes de datos en forma de triples RDF.

1.4. Formas de representar información almacenada en un sistema semántico

Grafos

Son una representación visual que encaja naturalmente con la forma de registrar datos como triples sujetopredicado-objeto. Sujetos y objetos pasan a ser representados como nodos; mientras que los predicados se representan como líneas que los relacionan entre sí.

Figura 3: Ejemplo de representación en forma de Grafo.



RDF/XML (http://www.w3.org/TR/rdf-syntax-grammar/)

Es una representación textual, basada en XML

Figura 4: Ejemplo de representación en formato RDF/XML

```
< rdf:RDF \ xmlns: p="http://www.example.org/personal_details#"
  xmlns: rdf = "http://www.w3.org/1999/02/22 - rdf - syntax - ns #"
  xmlns="http://www.example.org/meeting_organization#">
  <rdf: Description about="http://meetings.example.com/cal#m1">
    <\!Location\ parseType = "Resource"\!>
      <zip xmlns="http://www.another.example.org/geographical#">02139</zip>
      <lat xmlns="http://www.another.example.org/geographical#">14.124425</lat>
      <long xmlns="http://www.another.example.org/geographical#">14.245 < /long>
    </Location>
        <chair resource="http://www.example.org/people#fred"/>
        <homePage resource="http://meetings.example.com/m1/hp"/>
  </rdf: Description>
  <rdf:Description about="http://meetings.example.com/m1/hp">
    <policy resource="http://meetings.example.com/privacyPolicy"/>
  </rdf: Description>
  <rdf: Description about="http://www.example.org/people#fred">
    <attending resource="http://meetings.example.com/cal#m1"/>
    <p:GivenName>Fred</p:GivenName>
    <\!p: has Email\ resource = "mailto: fred@example.com"/>
  </rdf: Description>
</rdf:RDF>
```

```
Notation3 (N3) (http://www.w3.org/TeamSubmission/n3/)
```

Es un lenguaje de propósito general para representar triples RDF en un formato legible por humanos.

Figura 5: Ejemplo de representación en formato N3

```
@prefix g: <http://www.another.example.org/geographical#> .
@prefix : <http://www.example.org/meeting organization#> .
@prefix p: <http://www.example.org/personal details#>..
<http://meetings.example.com/cal#m1>
         : Location
             g:zip "02139";
             g:lat "14.124425";
             g:long "14.245" ];
         : chair <http://www.example.org/people#fred>;
         : homePage < http://meetings.example.com/m1/hp> .
<http://meetings.example.com/m1/hp>
         : policy < http://meetings.example.com/privacyPolicy> .
<http://www.example.org/people#fred>
         : attending <http://meetings.example.com/cal#m1>;
         p:GivenName "Fred";
         p: hasEmail < mailto: fred@example.com> .
```

N-Triples

```
(http://www.w3.org/TR/rdf/testcases/#ntriples)
```

Es un lenguaje simple, basado en líneas de texto plano. Cada línea representando un triple RDF.

Figura 6: Ejemplo de representación en formato N-Triples

```
<http://www.example.org/people#fred> <http://www.example.org/meeting_organization#atte
<http://www.example.org/people#fred> <http://www.example.org/personal_details#GivenNar
<http://www.example.org/people#fred> <http://www.example.org/personal_details#hasEmai</pre>
```

```
Turtle (Terse RDF Triple Language) (http://www.w3.org/TeamSubmission/turtle/)
```

Es un subconjunto de Notation3, pero sin llegar a la simplicidad de N-Triples.

1.5. Formas de embeber información semántica en páginas web

Microformatos

(http://microformats.org/wiki/Main Page)

Son extensiones de HTLM que permiten etiquetar los datos básicos de determinados tipos de información, tales como personas y organizaciones (hCard), eventos (hCalendar), entradas de blogs (hAtom), noticias (hNews), productos de consumo (hProduct), reseñas(hReview), curricula vitae (hResume), recetas (hRecipe), etc.

De tal manera que, sin mucho esfuerzo adicional, las páginas web así etiquetadas ofrecen parte de sus datos en un formato apto para ser procesado por buscadores y otras herramientas automatizadas que las visiten.

Figura 7: Ejemplo de información semántica embebida en una página web, empleando Microformats.



```
 \\  < strong \ class = "fn \ org" > Waikiki \ Parc \ Hotel < / strong > \\ < br / > \\ < span \ class = "adr" > < span \ class = "street-address" > 2233 \ Helumoa \ Rd < / span > \\ < br / > \\ < span \ class = "locality" > Honolulu < / span > , < span \ class = "region" > Hawaii < / span > , \\ < br / > \\ < span \ class = "postal-code" > 96815 < / span > \\ < br / > \\ < span \ class = "country-name" > United \ States < / span > \\
```

Microdatos

(schema.org) (http://schema.org/docs/gs.html)

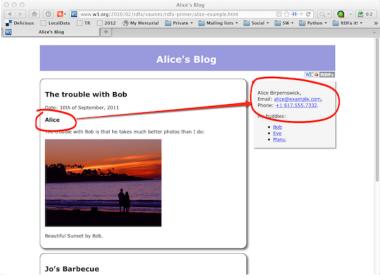
Es una variante parecida a los microformatos, pero con una sintaxis y un vocabulario alternativo. Está patrocinado por los principales buscadores: Google, Bing, Yandex y Yahoo.

Figura 8: Ejemplo de información semántica embebida en una página web, empleando Microdata.

RDFa (Resource Description Framework -in- attributes) (http://www.w3.org/TR/xhtml-rdfa-primer/)

Va camino de convertirse en una recomendación oficial del W3C. Para ser el mecanismo estándar de etiquetado de páginas HTML, dotándolas así de información semántica útil para buscadores y otras herramientas automatizadas que las visiten.

Figura 9: Ejemplo de información semántica embebida en una página web, empleando RDFa.



2. Breve descripción de las herramientas software utilizadas

La herramienta principal:

 Semantic mediawiki (SMW) (http://semantic-mediawiki.org)

En la web, tienen una sandbox online para probar a trabajar con ella.

Para ser instalado en un equipo, necesita que estén previamente instalados en él:

- Un entorno AMP: Apache web server, MySQL y PHP. (Nota: para Windows, Linux, Mac OS X o Solaris, puede obtenerse un paquete completo, Xampp, por ejemplo, desde http://www.apachefriends.org/en/xampp.html)
- $\bullet \ \ MediaWiki\ pueden\ obtenerse\ desde\ http://www.mediawiki.org/wiki/MediaWiki.$

(Nota: Para Mac OS X recomiendan utilizar el paquete "todo en uno", MAMPstack+MediaWiki, que puede obtenerse desde http://bitnami.org/stack/mediawiki.)

Algunas extensiones que podrían resultar útiles:

- Halo (https://www.semantic-mediawiki.org/wiki/Extension:Halo): Una serie de interfaces gráficos intuitivos para facilitar la edición, búsqueda, navegación y organización de datos semánticos:
 - Barra de herramientas semántica: con comandos para revisar, crear y modificar rápidamente las anotaciones realizadas en una página.
 - Modo avanzado de anotación: para realizar estas directamente en forma visual (WYSIWYG), sin necesidad de hacerlo en el texto fuente de la página.
 - Auto-completado: sugerencias automáticas de entes o datos ya existentes en otras anotaciones similares.
 - Interfaz gráfico de búsqueda: para facilitar la composición de las sentencias de búsqueda y la visualización de los resultados en diversos formatos de salida.
 - Editor de ontologías: para facilitar la visualización y mantenimiento de metainformación acerca de instancias y propiedades.

Nota: La extensión Halo mas algunas otras tales como TreeView, Enhanced Retrieval, Acces Control List, Semantic Notificatios, Rich Media,... estaban incluidas dentro del paquete SMW+ de la extinta empresa Ontoprise, cuyos principales activos fueron adquiridos por la empresa Semafora en Julio de 2012 para incorporarlos dentro de su gama de productos propietaria. No obstante, la extensión Halo continua estando disponible desde SourceForge (Halo extension).

- Semantic MediaWiki Plus, SMW+ (https://www.semantic-mediawiki.org/wiki/Semantic_MediaWiki_Plus)

 Nota (Septiembre 2019): Solo sobreviven algunas de esas extensiones. Pero algunas de ellas han dado
 lugar a un producto comercial para recopilar conocimiento (https://classic.diqa-pm.de/en/DataWiki).
- Semantic Result Formats (http://www.semantic-mediawiki.org/wiki/Semantic_Result_Formats): Añade multitud de nuevos formatos de salida para mostrar los resultados de consultas inline. Formatos tales como: calendarios, líneas temporales, gráficos de datos, funciones matemáticas, etc.
- Semantic Watchlist (http://www.mediawiki.org/wiki/Extension:Semantic_Watchlist): Complemento que extiende la funcionalidad de las "listas de seguimiento"; permitiendo realizar también recopilaciones automáticas de cambios en función de determinadas anotaciones semánticas.
- Semantic Forms (http://www.mediawiki.org/wiki/Extension:Semantic_Forms): Permite la creación de formularios para la recogida sistematizada de información semántica a incluir en un determinado tipo de páginas.

Nota (Septiembre 2019): Existen algunos productos comerciales adaptados expresamente para el fin que se persigue en este documento. Por ejemplo: Data Wiki - The Knowledge Portal

2.1. Instalación

Prerequisitos:

Necesita que estén previamente instalados: un 'servidor web' (Apache o ISS), PHP, una 'base de datos' (MySQL, PostgreSQL o SQLite) y 'MediaWiki' con la extensión 'Validator'.

Todos esos elementos se pueden ir descargando e instalando uno a uno.

(http://www.mediawiki.org/wiki/Manual:Installation_guide)

O se puede emplear alguno de los paquetes preconfigurados existentes, instalando todos los elementos de un solo golpe. (http://www.mediawiki.org/wiki/Software bundles)

En mi caso, he optado por el paquete BitNami (http://www.mediawiki.org/wiki/MediaWiki_with_BitNami), porque dispone de soporte nativo para Mac OS X. Simplificandose así sobremanera la instalación en mi equipo (un Mac Mini).

La wiki propiamente dicha, Media Wiki (MW)

La instalación de los prerequisitos ha sido muy sencilla, gracias a la preconfiguración del paquete BitNami. Tan solo ha requerido descargar el archivo .dmg desde la web, abrirlo y volcar su contenido sobre la carpeta 'Aplicaciones' del equipo.

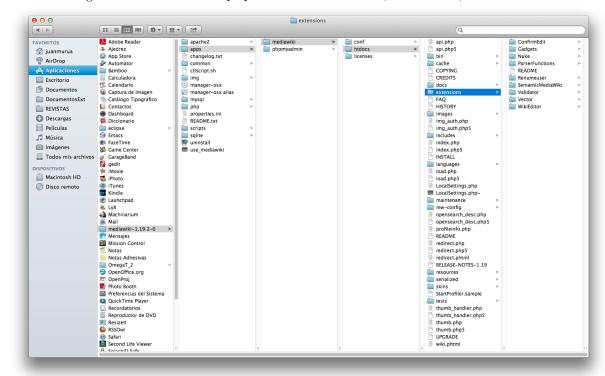


Figura 10: Instalación del paquete 'MediaWiki Stack', de BitNami, sobre Mac OS X

La extensión semántica, Semantic Media Wiki (SMW)

La instalación de la extensión semántica consiste en tres sencillos pasos:

- Copiar el contenido del archivo descargado a la carpeta "extensions" de la instalación de Media Wiki.
- Añadir un par de líneas en el archivo de configuración de Media Wiki (LocalSettings.php).
- Iniciar sesión en Media Wiki con poderes administrativos. Ir a las páginas especiales (Special Pages), concretamente al grupo de páginas específicas de administración de 'Semantic MediaWiki', a la página "Admin functions for Semantic MediaWiki". Lanzar los scripts de configuración de la base de datos ("Database installation and upgrade" y "Data repair and upgrade").

2.2. Configuración básica

Algunos aspectos que es conveniente ajustar desde el principio:

■ El idioma por defecto de la navegación en la wiki. (En el archivo LocalSettings.php)

```
\# Site language code, should be one of ./languages/Language(.*).php \ SwgLanguageCode = "en";
```

• El nombre del site y el identificador del espacio de nombres para la información en él contenida. Así como el logotipo a mostrar en esquina superior izquierda de la página principal. (En el archivo LocalSettings.php)

- Las páginas de "Política de protección de datos", "Acerca de..." y "Aviso Legal" ("Privacy policiy", "About..." y "Disclaimers"), que aparecen siempre en la parte inferior de cada página. (Nota: Para tener acceso a la edición de estas páginas, es necesario iniciar sesión con un usuario con permisos de administrador.)
- Cuando se está editando una página, el mensaje de aviso al fondo es el típico de wikipedia. ("Por favor, ten en cuenta que todas las contribuciones a Benzirpi's experimentation site pueden ser editadas, modificadas o eliminadas por otros colaboradores. Si no deseas que las modifiquen sin limitaciones y las distribuyan libremente, entonces no las pongas aquí. También nos aseguras que tú escribiste esto y te pertenecen de los derechos de autor, o lo copiaste desde el dominio público u otra fuente libre. (véase Benzirpi:Derechos de autor para más detalles). ¡No uses escritos con copyright sin permiso!") ("Please note that all contributions to Benzirpi's experimentation site may be edited, altered, or removed by other contributors. If you do not want your writing to be edited mercilessly, then do not submit it here. You are also promising us that you wrote this yourself, or copied it from a public domain or similar free resource (see Benzirpi:Copyrights for details). Do not submit copyrighted work without permission!"). Es conveniente poner algo mas acorde a nuestro uso.
- La página de "Ayuda" ("Help"). Donde es conveniente poner, como mínimo: por un lado, algo que guíe a los usuarios novatos en sus primeros pasos; y, por otro lado, algo de referencia rápida para el día a día de los usuarios habituales.
- A no ser que el gran volumen de información o la gran complejidad de las consultas lo desaconsejen, es conveniente desactivar el mecanismo de "concept caching" de la wiki. De esa manera, todas las consultas inline de las páginas visitadas se ejecutan en el mismo momento de cada visita, mostrando así siempre la información actualizada.

Para desactivar todas las cachés, se han de añadir dos variables en el archivo Localsettings.php:

```
$wgEnableParserCache = false;
$wgCachePages = false;
```

Si no se desean desactivar completamente, en el archivo SMW_Settings.php se pueden ajustar ciertos aspectos del funcionamiento de las cachés. Entre otros, su tiempo de vigencia; que, por defecto, está establecido en un día (es decir, solo se recalculan consultas cuya caché de resultados tiene más de un día de antigüedad).

```
// Cache life time in minutes. If a concept cache exists but is older than // this, SMW tries to recompute it, and will only use the cache if this is not // allowed due to settings above: \$smwgQConceptCacheLifetime = 24 * 60;
```

3. Breve manual de usuario SMW (Semantic Media Wiki)

En SMW, cada tópico tiene su 'página', (recodar la estructura de 'artículos' de la Wikipedia). Y, a todos los efectos, se identifica por el título de la misma. Incluso la definición de categorías, propiedades, etc. se realiza creando páginas especiales para cada una de ellas con un determinado formato de título.

SMW tiene un lenguaje propio de modelado: SMW-ML y de interrogación: SMW-QL.

3.1. Escribir contenido

Para abrir una nueva página, simplemente se ha de intentar ir a ella:

- bien sea sobre la marcha, escribiendo un enlace en otra página ya existente, guardando y cliclando sobre dicho enlace (que estará en rojo),
- bien tecleando un título en el cuadro de búsqueda,
- o bien escribiendo directamente un título en la URL del navegador;

cuando la wiki detecta que no existe la página a la que se pretende ir, propone automáticamente su creación.

Para poner un enlace a otra página de la wiki, se escribe el título de esta entre corchetes dobles [[título]]

Es conveniente familiarizarse con las formas básicas de indicar ciertos tipos de formato tales como "italica", "'negrita"', <s>tachado<s/>, === Título ===, listas con bolo: *item1 *item2 *item3, listas numeradas: #item1 #item2 #item3, etc.

Para que el historial de ediciones tenga sentido, suele ser conveniente resumir lo editado en la línea de "sumario" (al fondo de la página que se está editando), sobre todo cuando se realizan ediciones de cierto calado.

Para adjuntar archivos externos a una página, la carga de estos ha de estar configurada expresamente por los administradores de la wiki (https://www.mediawiki.org/wiki/Manual:Configuring file uploads)

3.2. Introducir anotaciones semánticas

Categorías:

Son el mecanismo básico de Media Wiki para clasificar páginas. Permiten hacer búsquedas y confeccionar listas automáticas, pero sus capacidades son bastante limitadas en comparación con las capacidades derivadas de las anotaciones semánticas.

Propiedades:

Son el mecanismo básico de cualquier herramienta semántica para formar triples: sujeto-propiedad-objeto

Se recomienda utilizar frases con verbo como nombres de propiedad. Así se gana en precisión y se evitan ambigüedades. Por ejemplo, es mas claro un nombre tal como "esta ubicado en" o "se guarda en" en lugar de un simple "ubicación".

Para introducir una propiedad en una página, se ha de escribir entre corchetes dobles; separando la propiedad de su valor con dos dobles puntos :: [[nombredelapropiedad::valorqueseleasignaenestapagina]]

Si se desea introducir una propiedad de modo "silencioso" (sin que su valor aparezca visible en la página), se puede emplear la notación alternativa $\{\{\#set:nombredelapropiedad=valorqueseleasignaenestapagina\}\}$

Para añadir una propiedad a un enlace, se ha de poner esta tras el enlace, dentro de los corchetes de este [enlace[[propiedad::valor]]]

Se pueden definir "plantillas semánticas", para agilizar con ellas la introducción de determinados tipos de anotación habitual.

Para recoger información más compleja o más estructurada, se pueden definir y utilizar los clásicos formularios de recogida de datos.

Extensiones tales como Halo permiten ayudas a la anotación y proporcionan herramientas adicionales muy útiles: autocompletar según información y anotaciones ya presentes en la wiki, consultar ontologías sobre la marcha, barras con botones específicos para anotación semántica, etc.

El 'factbox' es una ayuda interesante. Aparece al fondo de la página, cuando se clica sobre [Mostrar previsualización] estando editando una página. Resume toda la información semántica que se ha reconocido en la página, da avisos sobre información mal introducida y permite realizar consultas rápidas a partir de la información para verificar la pertinencia o no de las anotaciones semánticas realizadas. (Nota: La aparición o no del factbox se regula desde la variable \$smwgShowFactbox, en el archivo SMW_Settings.php; o a través de las palabras mágicas __NOFACTBOX__ o __SHOWFACTBOX__ puestas expresamente en páginas concretas.)

Es muy recomendable crear una página descriptiva para todas y cada una de las propiedades que se definan. Esto se hace en el espacio de nombres 'Property', poniendo como título de la página: Property:nombredelapropiedad. En esta página descriptiva se suele explicar el significado y del uso correcto de la propiedad, para evitar la dispersión de significados; además de indicar las caracteristicas especiales que la propiedad pueda tener:

- Por defecto, todas las propiedades se consideran del tipo 'page' y se les pone el correspondiente enlace a la página de su valor. Para propiedades representativas de entes tales como cantidades, fechas, etc. no tiene sentido una página por cada valor. De ahí que sea interesante definir expresamente cual es el tipo de esas propiedades. Y eso se realiza con la etiqueta [[Has type::xxxx]], por ejemplo [[Has type::number]], [[Has type::date]], [[Has type::telephone number]], etc.
- Si se desea restringir los valores que la propiedad puede tomar, se puede dar una lista de valores válidos indicándolos con una serie de etiquetas [[Allows value::xxx]]: [[Allows value::unvalorqueesadmitido]], [[Allows value::otrovalorqueesadmitido]],...
- etc. http://semantic-mediawiki.org/wiki/Help:Special properties

Nota: SMW realiza inferencias con relaciones tales como SubTypeOf y SubPropertyOf. Pero no las hace con relaciones más complejas, tales como TransitiveProperty, domain, range, cardinality, FunctionalProperty,...; ni aún en el caso de explicitar dichas relaciones directamente en las páginas o vía ontologías externas. Si se precisaran ese tipo de inferencias avanzadas, no hay más remedio que recurrir a extensiones de terceros: por ejemplo, a http://www.mediawiki.org/wiki/Extension:SemanticQueryPlus

Formularios:

Cuando se trata de recoger muchos datos, suele ser más práctico preparar un formulario que guíe dicha recogida. En SMW se pueden definir estos formularios a través de los recursos aportados por la extensión 'Semantic Forms': https://www.mediawiki.org/wiki/Extension:Semantic Forms

3.3. Exportar información en formato semántico

Aunque internamente todas las propiedades se definen localmente. Y al exportarse, van con el espacio de nombres del sistema local. También es posible relacionar una propiedad interna con un vocabulario externo (por ejemplo, FOAF o Dublin Core). De tal manera que, cuando se exporte el contenido de la wiki, la propiedad interna sea sustituida por la externa. Permitiendo así que usuarios externos la vean en esa terminología externa. (Nota: Esto se puede hacer en cualquier momento, con cualquier propiedad que se esté ya usando dentro de la wiki. Lo cual nos da libertad para organizar nuestra ontología interna como nos parezca más adecuado para nuestros usuarios. Sin perder por ello la interoperabilidad con otras ontologías externas estándares.)

3.4. Utilizar la información almacenada

'Special pages', 'Páginas especiales': No son solo para administración, sino que también tienen enlaces a herramientas para consulta y para búsqueda de determinadas páginas o determinados tipos de información.

Listas de páginas

- Categorías
- Lista de redirecciones
- Propiedades
- Páginas que enlazan con páginas de desambiguación
- Tipos de datos
- Todas las páginas
- Todas las páginas con prefijo

Semantic MediaWiki

- Buscar por atributo
- Búsqueda semántica
- Explorar artículos

- Exportar el artículo como RDF
- Query Creator
- 'My watchlist', 'Lista de seguimiento': Permite a cada cual hacerse una página de seguimiento de aquellas páginas que le interesen. => se da una lista de páginas (sus títulos), y la watchlist muestra los últimos cambios realizados en dichas páginas desde la última vez que se visitó la watchlist.
- 'Inline queries' (http://semantic-mediawiki.org/wiki/Help:Inline_queries), por medio de consultas semánticas (lenguaje SMW-QL), es posible:
 - Generar automáticamente páginas recopilatorias (páginas de noticias/eventos/alertas, calendarios, tablas, listas, etc). La selección de páginas http://semantic-mediawiki.org/wiki/Help:Selecting_pages (función #ask), se puede realizar tanto en base a categorías wiki como a propiedades semánticas, o hasta incluso en base a algunas otras caracteristicas de Media Wiki tales como los espacios de nombres.

Figura 11: Ejemplo de consulta inline para generar listas automáticamente

```
== Están ubicados en esta sede ==
{{#ask: [[Esta ubicado en :: Barcelona]]
|? Es un
| ? Es una
| format=ul
| sort=Es un
| order=ascending }}

== Páginas relacionadas con esta ==
{{#ask: [[Esta relacionado con :: Barcelona]]
OR [[Esta relacionado con :: <q>[[Esta ubicado en :: Barcelona]] </q>]]
|? Modification date
| format=ul
| sort=Modification date
| order=descending }}
```

Están ubicados en esta sede

[editar]

- EsbcnLJ2500 (Es una Impresora)
- Benzirpi Mirvento (Es una Persona)
- Esbcnfs1 (Es un Servidor)
- Esbartr1 (Es un Servidor)
- Esbcnfs3 (Es un Servidor)

Páginas relacionadas con esta

[editar]

- Despliegue de Autocad 2013 (Modification date 16 diciembre 2012 11:57:23)
- Instalacion 20121205T0900 (Modification date 9 diciembre 2012 16:57:18)
- Incidente 20121128T2306 (Modification date 2 diciembre 2012 20:32:35)

• Hacer aflorar un cierto dato mediante la búsqueda de un determinado valor para una determinada propiedad (función #show). De esta forma, se contribuye a mantener la consistencia de los datos a lo largo y ancho de toda la wiki, (la cuestión es tener claro en cada caso cual es la página maestra "dueña" del dato y de la que dependerán todas las demás que simplemente lo muestran en diferentes contextos).

Figura 12: Ejemplo de consulta inline para hacer aflorar un dato de una página en otra *[[Relacionado con :: Benzirpi Mirvento]] ({{#show: Benzirpi Mirvento | ?Esta ubicado en}})

Benzirpi Mirvento (Barcelona)

- Los propios usuarios pueden definirse cada cual las consultas que necesite => a través de las funcionalidades aportadas por extensiones tales como Halo o Semantic Drilldown.
- La información se puede exportar a formatos tales como CSV, JSON, RDF => permitiendo a la wiki actuar como repositorio de datos para otras aplicaciones externas.

4. Una posible forma de trabajar para recopilar información operativa

En este capítulo resumo los frutos de las pruebas que he realizado a medida que iba leyendo los manuales de SMW e intentando construir un sistema de recogida de información lo más sencillo posible. Las recomendaciones recogidas han de considerarse como unas meras ideas base, sobre las que es posible comenzar a construir un sistema práctico.

4.1. Necesidades a cubrir según tipo de usuario

- Usuarios normales: escriben y consultan.
 - Ayudas para saber las opciones de formato de que disponen.
 - Ayudas para saber qué interesa que etiqueten semánticamente y qué no, dependiendo del tipo de información que estén introduciendo => ha de quedar claro qué es lo mínimo de información semántica que seria deseable.
 - Ayudas para no inventarse nuevas propiedades ni nuevos entes allá donde ya hay algo definido para el significado que pretenden etiquetar => hay que evitar la dispersión de sinónimos para una misma cosa
- Jardineros semánticos: revisan los contenidos y velan para que estén adecuadamente etiquetados.
 Hay varias páginas especiales que pueden serles útiles:
 - Special: Unused Properties (propiedades abandonadas, que no se usan; tienen página descriptiva, pero no hay ninguna anotación en ninguna página de la wiki)
 - Special: Wanted Properties (propiedades que se han utilizado, hay anotaciones en la wiki, pero que no disponen de su propia página descriptiva 'Property: nombredela propiedad')
- Administradores: se encargan de que la infraestructura de la wiki funcione y esté bien engrasada.

4.2. Un par de recomendaciones generales respecto a la forma de trabajar

1. Al comenzar a escribir sobre algo, abrir una página nueva cada vez. Si se ha de buscar la página adecuada a la que añadir la información, a no ser que se tenga muy clara la estructura de la wiki, el esfuerzo extra de intentar "acertar" puede hacer que la gente se retraiga y deje "para más tarde" la escritura.

Nota: Para aquellos contenidos que no tengan un título único evidente, se podría usar un título genérico seguido de la fecha en formato ISO comprimido, más que nada para asegurarnos fácilmente de que el título es único (seria extraño que otra persona se ponga a escribir algo similar justo a la misma hora y en el mismo minuto). (Por ejemplo, si comenzamos a relatar algo sobre un problema que hemos resuelto, y lo comenzamos a escribir el 28 de Noviembre de 2012 a las 15:06, podríamos poner algo como "incidente 201211281506" como título para la página.)

2. Según se va escribiendo, acordarse de ir "dejando caer" las etiquetas (propiedades) que se crean convenientes

De forma natural, según van apareciendo las palabras clave en nuestro pensamiento o en lo que estemos escribiendo. La parte de la página donde se sitúen las etiquetas no tiene importancia alguna para luego encontrar la información etiquetada.

Sin obsesionarnos por etiquetarlo todo. Mas vale información no etiquetada, pero escrita en la wiki; que información exhaustiva, pero solo "en intención" de escribirla.

Es importante también tener presente la función de "jardinería" (refactorización continua): la labor de ir incorporando y arreglando etiquetas sobre la marcha, según se utilizan las páginas o según se detectan inconsistencias o faltas de etiquetado.

4.3. Recomendaciones respecto a la Ontologia

Algunas propiedades que sería interesante definir para esta aplicación:

- EsUn o EsUna => Es una propiedad equivalente a la propiedad rdf:type, permitirá hacer listas automáticas por tipo de recurso.
- EstaRelacionado Con o EstaRelacionada Con => Es una propiedad equivalente, en cierta medida, a la propiedad rdfs:seeAlso, permitirá hacer recopilaciones automáticas sobre todo aquello "que tiene algo que ver" con determinados entes o recursos.
- Esta Ubicado En o Esta Ubicada En => Esta propiedad complementa a las propiedades EsUn y a Relacionado Con, para permitir la realización de inventarios automáticos de lo que hay (cosas, personas, servicios,...) en cada sede o ubicación de la empresa.
- HablaSobreUn => incidente, mantenimiento, despliegue,... (permitirán hacer estadísticas sobre determinados tipos de actuaciones); procedimiento (instrucciones formales), truco (instrucciones informales), comentario, etc.
 - Será una propiedad especial, con lista limitada de valores posibles: [[Allows value :: xxxx]]
- *TieneFecha* => Será una propiedad especial: [[Has type :: date], para luego permitir búsquedas por periodos temporales.

Nota: El hecho de definir dos formas, masculina y femenina, en algunas propiedades se debe al intento de facilitar la escritura de las mismas de forma fluida según se está redactando el contenido de la página. Luego, con la asignación cruzada de [[SubPropertyOf::xxxx]] en la definición de ambas, se consigue tener el equivalente a una sola propiedad (lo que seria owl:equivalentProperty); de tal forma que se puedan utilizar indistintamente y se localicen los resultados correspondientes a ambas cuando alguna de ellas se utiliza en cualquier búsqueda.

Nota: Al escribir relaciones, el código de colores de MediaWiki será muy útil como indicador de algún error de escritura en el título de la página relacionada: si el valor aparece en rojo, es señal de que la página relacionada no existe (y un ente o recurso común con el que se enlace habitualmente, debería existir).

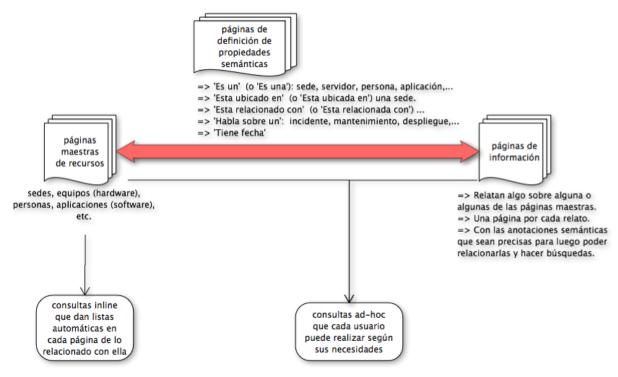
Nota: Es posible utilizar las categorías de MW [[Category:xxxx]] en lugar de la propiedad [[EsUn:xxxx]], y los enlaces de MW [[xxxx]] en lugar de la propiedad [[EstaRelacionadoCon::xxxx]]; para luego utilizar las herramientas estándares MW de confección de listas en base a categorías y de búsqueda de páginas relacionadas, "Lo que enlaza aquí" ("What links here"). Pero el uso de propiedades semánticas SMW permite ir más allá y utilizar búsquedas inline, búsquedas con las que se mejora en gran medida la presentación de información de uso habitual y la forma de navegar a través de esta.

4.4. Recomendaciones respecto al modelo de datos

Prácticamente se basa en un mismo patrón repetido continuamente (cuanto más simple, mayores posibilidades de que los usuarios comprendan y sigan las directrices marcadas):

- Se definen 'páginas maestras' para cada uno de los recursos tales como sedes, equipos (hardware), personas, aplicaciones (software),... En estas páginas maestras, además de los datos prácticos básicos sobre el recurso, se definirán consultas inline para tener listas automáticas del estilo de "todo lo relacionado con esto" o "todo lo que se ubica aquí", enlazando con el resto de páginas que están relacionadas con ellas.
- Por cada información a registrar, se abre una 'página de relato', que se enlazará con los recursos involucrados mediante las correspondientes anotaciones semánticas.

Figura 13: Esquema general del modelo de datos propuesto para la wiki

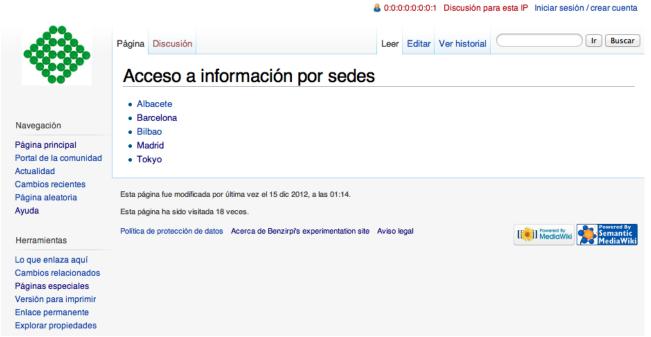


4.5. Un ejemplo práctico:

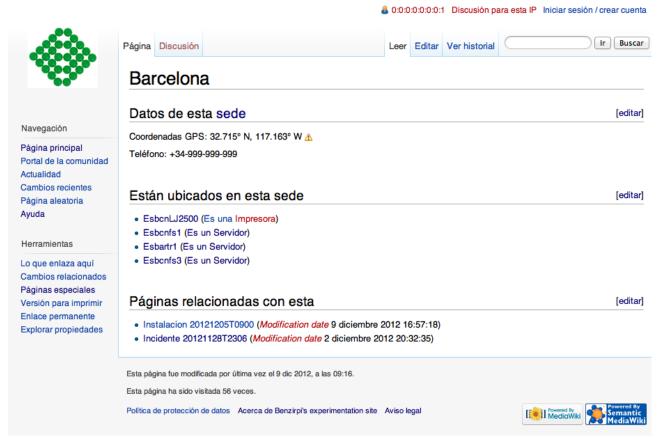
Un ejemplo de información recogida en la wiki de pruebas.





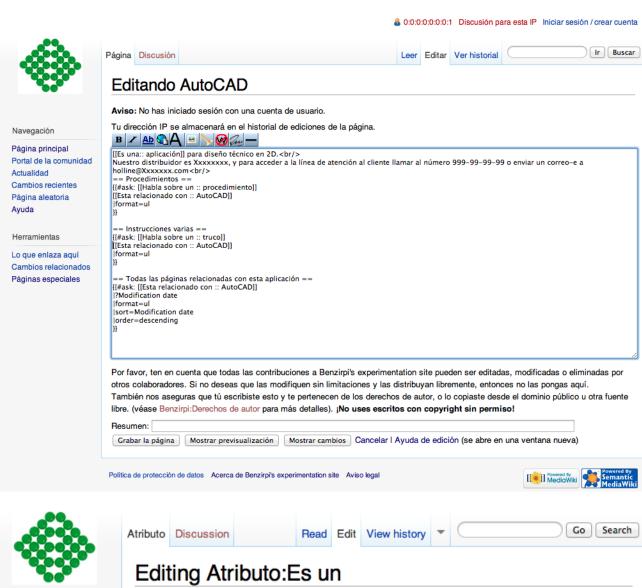


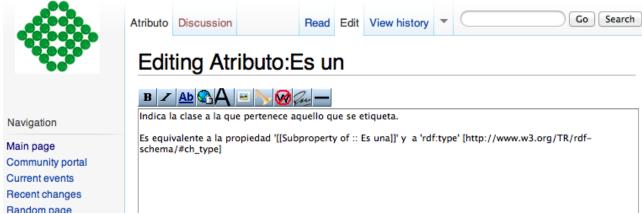


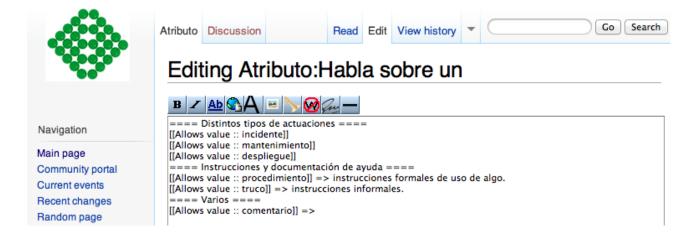


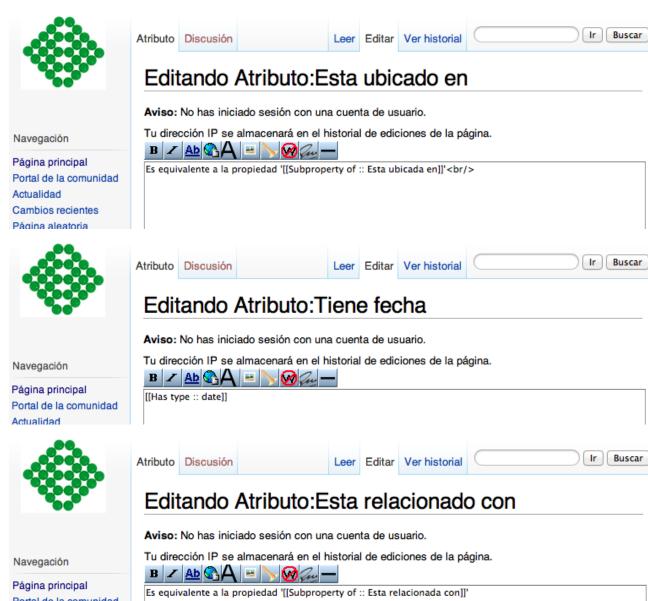












Página principal
Portal de la comunidad
Actualidad

5. Glosario

Grafo Representación compuesta por nodos, y aristas que los relacionan entre sí.

Microformat Un formato para embeber información semántica en una página web escrita en HTML.

MW abreviatura de 'Media Wiki'

N-Triples Uno de los formatos estandarizados para escribir/leer información RDF

N3 (Notation3) Uno de los formatos estandarizados para escribir/leer información RDF.

Ontología Modelo semántico (metadatos), que describe de manera formal las entidades (recursos) y las relaciones (propiedades) necesarias para codificar coherentemente información (datos) relativa a un determinado campo del saber.

OWL (Web Ontology Language) Un sistema estandarizado para escribir ontologías. A su vez, OWL está escrito en RDF.

RDF (Resource Description Framework) Un sistema estandarizado en que se pueden escribir tanto información (datos) como información semántica acerca de ella (metadatos).

RDFa Un formato para embeber información semántica en una página web escrita en HTML.

RDFS (RDF Schema) Versión reducida de OWL, contemplando tan solo las relaciones básicas.

RDF/XML Uno de los formatos estandarizados para escribir/leer información RDF.

SMW abreviatura de 'Semantic Media Wiki'

SPARQL Un lenguaje de interrogación para realizar consultas sobre almacenes de triples.

Triple Sentencia del tipo "Sujeto-Predicado-Objeto". Es la forma habitual de expresar una pieza de información en un sistema semántico.

Turtle (Terse RDF Triple Language) Uno de los formatos estandarizados para escribir/leer información RDF

URI (Uniform Resource Identifier) Cadena de caracteres que identifica unívocamente una determinada entidad.

URL (Uniform Resource Locator) Cadena de caracteres que representa la dirección donde se puede encontrar una determinada página web.

6. Bibliografía

Referencias

- [1] Allemang, Dean. Hendler, James. "Semantic Web for the Working Ontologist, Effective Modeling in RDFS and OWL". Morgan Kauffmann Publishers (2008).
- [2] Segaran, Toby. Evans, Colin. Taylor, Jamie. "Programming the Semantic Web, Building Flexible Applications with Graph Data". O'Reilly (2009).
- [3] Fogel, Karl. "Producing Open Source Software". O'Reilly (2009)
- [4] Raymond, Eric S. "The Cathedral & the Bazaar". O'Reilly (2008)
- [5] http://www.w3.org/2001/sw/BestPractices/Tutorials . Visitada por última vez el 20130102T2214
- [5] http://www.mediawiki.org/wiki/Help:Contents . Visitada por última vez el 20130102T2214
- [6] http://semantic-mediawiki.org/wiki/Help:User manual . Visitada por última vez el 20130102T2214
- [7] $http://semantic-mediawiki.org/wiki/Help:Administrator_manual . Visitada por última vez el 20130102T2214$
- [8] http://videolectures.net/eswc08 schaffert sw/. Visitada por última vez el 20130102T2214