

Jornada por Videoconferencia  
**V Encuentro Virtual: Capacitación sobre tecnologías  
digitales de la información, para bibliotecarios y  
documentalistas.**  
16 de mayo de 2012



**Motores de búsqueda y  
circulación de  
conocimiento en la Web y  
redes sociales para  
bibliotecarios y  
documentalistas**

**Mela Bosch**

[melabosch@hotmail.com](mailto:melabosch@hotmail.com)

*Una vez que hemos catalogado y clasificado ...*

*¿Cómo buscan los sistemas informáticos ?*



*Tipos de buscadores*

*bases de datos*

*motores Web*

# El archivo y la búsqueda en bases de datos



- La mayor parte de los sistemas que acumulan información de manera persistente, incluidos muchos de los de referencias bibliográficas y bibliotecas digitales, se basan en datos estructurados con un modelo llamado de **entidad-relación:**

## Modelo de datos de entidad-relación o modelo relacional

Origen: 1970, propuesto por Edgar Frank Codd, un investigador de IBM

revolución muy positiva pues constituyó un soporte lógico matemático fuerte y sencillo

### Antes de esa época:

Archivos de registros y campos, al modo del original sistema ISIS, nacido a fines de los 60

**Es necesario mantener la estructura jerárquica en forma de árbol y cuando se necesita modificar una parte, toda la estructura se afecta**





## ¿Y en el mundo de las bases de datos bibliográficas y bibliotecas digitales?

**Predominio comercial de bases de datos con modelo entidad-relación: lógica diferente de la de clases que se usa para indizar, separa los elementos del objeto que los contiene**

IFLA propone en 1997:

Functional Requirements for Bibliographic Records, FRBR,  
basado en conceptos de lógica relacional

**Según uno de sus desarrolladores es el fin 175 años de pensamiento sobre lo que los catálogos son y cómo deberían trabajar. (Denton, 2007).**

**El gran cambio no es la lógica sino que FRBR es un modelo conceptual en el sentido informático:**

**representación abstracta de los datos físicos, de manera que una misma definición conceptual puede ser usada para diferentes implementaciones físicas**

**Ejemplo: FRBR propone pasar los documentos a las obras, la obra es el referente común de todas las manifestaciones de la misma**

# Buscadores en bases de datos y modelo de datos

## El modelo relacional

- permitió que los usuarios no tuvieran necesidad de saber cómo estaban organizados los datos internamente, permitió independizar la visualización del aspecto físico de los datos.

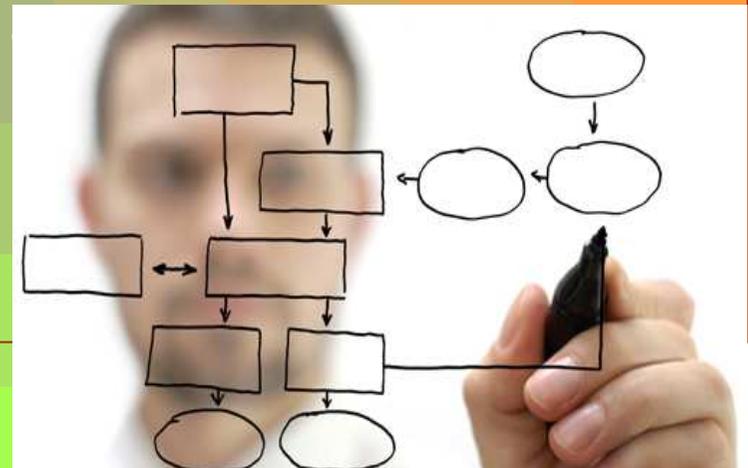
- Esta simplicidad y eficacia tiene precio: es necesario adaptar el dominio que se quiere representar en la base de datos a la lógica del modelo entidad-relación.

Todo debe ser descompuesto en filas y columnas para poder ser manejado en forma de tablas.

- Su eficacia es y ha sido formidable; la vida de todos los días se ha visto modificada por esta forma de modelizar y operar nos rodea.

pero... **El mundo real no está constituido por tablas:**

**Los buscadores deben ser cada vez más sofisticados**



# *El mercado de la búsqueda en bases de datos*

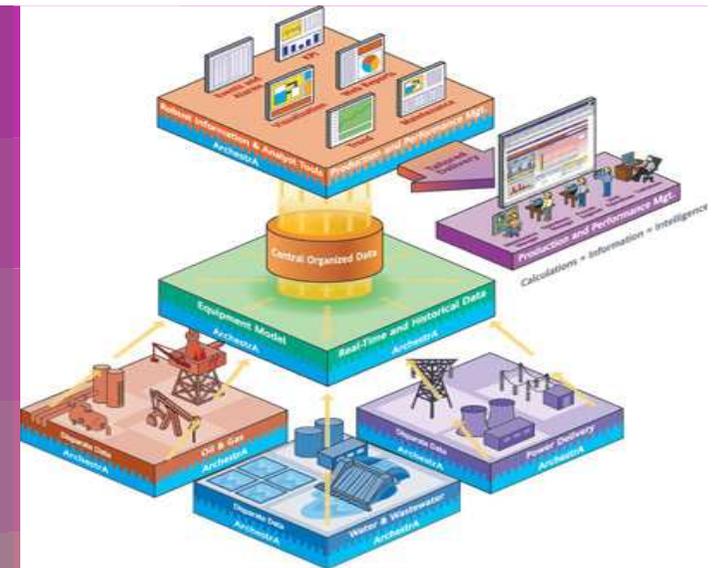
**Buscan en la bases de datos de documentación en las empresas:**

**Lideres del mercado:**

**Autonomy - Oracle - -Microsoft  
Index Server- Hyland / OnBase -  
Open Text- Ask Sam Web Publisher  
- Vertical Search Works-**

**Son costosos y están orientados a las intranet de las empresas, corporaciones, periódicos en línea.**

**También hay sistemas open source:mnGoSearch**



Uno de los líderes es Oracle que está en medio de una batalla judicial con Google por el uso de patentes para sistemas para tablets.

**Entonces el mundo de las intranet e internet confluye...**

# buscadores Web



## 1. Motor de búsqueda textual: (*Text Search Engine*)

### Basado en análisis léxico:

- Procesamiento: divide el texto en párrafos, frases, palabras y también identifica entidades tales como direcciones de correo electrónico y direcciones Web, todos estos elementos para ser procesados son considerados como unidades acumulativas, técnicamente llamadas *tokens*, son sometidos a una serie de parámetros estadísticos con los que se establece un rango de enlaces, esta lista es la que se presenta como respuesta a nuestra pregunta.
- Este tipo de motores son los primeros que aparecieron. Eran de este tipo el Gopher, creado en 1991 por Mark McCahill de la University of Minnesota y Veronica (**V**ery **E**asy **R**odent-**O**riented **N**et-wide **I**ndex to **C**omputerized **A**rchives)



## 2. Motores con indización de la semántica latente (*Latent semantic indexing, LSI*)

- Análisis del significado no explícito utiliza algoritmos con componentes estadísticos y léxicos.
- Latent semantic analysis (LSA) es una técnica de procesamiento del lenguaje natural
- Usa una base de datos de documentos para encontrar términos similares
- puede encontrar un cierto grado de sinonimia y devuelve los enlaces a los sitios Web que mejor se adaptan a la consulta, el LSI no necesita tener el exacto término para poder ofrecerlo como respuesta, puede usar aproximaciones con sinónimos cuasi sinónimos .
- Google utiliza este tipo de análisis, el componente estadístico es más fuerte que el de procesamiento de lenguaje y además tiene otros algoritmos como Page Rank

## 3. Motores de búsqueda semánticos, (*Semantic Web search engines o Search engines of 3rd generation*)



- Toman el sentido de una palabra como factor para los algoritmos de ordenamiento y también pueden ofrecer posibilidades para desambiguar o refinar su consulta.
- Llamados también motores de búsqueda de tercera generación, utilizan las dos tecnologías de búsqueda textual y de búsqueda de semántica latente a las que se suman otras tecnologías de Web semántica.
- Estas son: ontologías, RDF (*Resource Description Format*) OWL (*Ontology Web Language*). Las tecnologías de Web semántica se basan en lógicas de descripción para dar cuenta de manera formal y computable de la semántica de los objetos de un sistema.

### 3. Motores de búsqueda semánticos, (*Semantic Web search engines Search engines of 3rd generation*)

Tres tipos: a)

**Buscadores semánticos orientados al usuario** (*User oriented Semantic Web search engines*)

- devuelven enlaces a páginas Web
- pueden usar internamente tanto tecnologías de Web semántica como de LSI
- Recuperación en diferentes formatos, imagen, sonido, etc para múltiples dispositivos
- Ya hay sistemas comerciales como Hakia, que busca incluso en los Twitters y E-Gains que también trabaja con redes sociales



### 3. Motores de búsqueda semánticos, (*Semantic Web search engines Search engines of 3rd generation*)

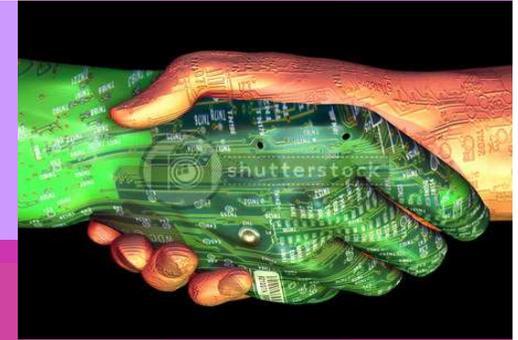
b)

#### **Buscadores semánticos orientados a servicios Web** (*Semantic Web Services oriented*)

- No son para el usuario final
- Dan enlaces a útiles para los especialistas que trabajan en la construcción de recursos de Web semántica, devuelven enlaces a ontologías, archivos en OWL, a instancias de RDF
- Entre ellos tenemos: SOWL, WSE, Watson, Falcons, Sindice y Swoogle



### 3. Motores de búsqueda semánticos, (*Semantic Web search engines Search engines of 3rd generation*)



c)

#### **Motores de búsqueda orientados a la**

#### **Web social semántica *socio-semantic web (s2w)***

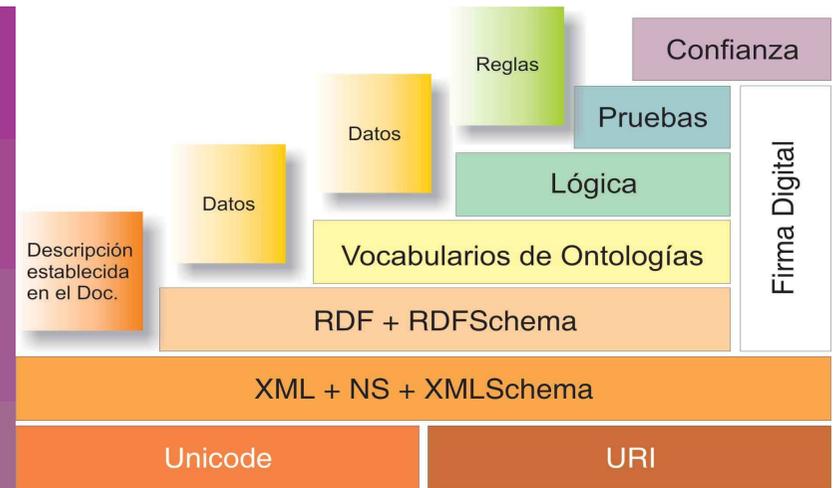
Se proponen complementar la visión formal de la Web semántica con un acercamiento pragmático agregando a los lenguajes controlados creados con fuertes bases lógicas otros aspectos heurísticos basados en experiencias de prueba y error experimentadas por multitudes de usuarios que realizan etiquetado colaborativo (*folksonomy*)

- Entre por ejemplo: <http://www.stumpedia.com/>
- La diferencia los buscadores semánticos orientados al usuario es que utilizan microformatos de Web 2.0 (por ejemplo RSS) para poner etiquetas y usan para el trabajo cooperativo apoyado en computadoras [Computer Supported Cooperative Work \(CSCW\)](#).

¿Por qué tenemos todas estas diferencias?

¿Cuál Web Semántica?

*La Web semántica tiene un **espesor** que va de lo visible, al soporte lógico y a la anotación de datos*



1. Capa superficial de la Web semántica **centrada en los servicios Web: visible en la navegación y recuperación** a través de motores con operadores semánticos más complejos que los de semántica latente (Google) usos desde comercio electrónico a la comunicación con dispositivos móviles, aquí se habla de **Web 3.0**.
2. Se apoya en la Web Semántica **centrada en el conocimiento**, el software es capaz de razonar a partir de la calidad de la descripción del contenido, es decir de la anotación de los datos.
3. El sustento de las otras dos capas es la Web Semántica centrada en la **anotación de los datos** representados en RDF, OWL o en microformatos que se agregan a páginas HTML.



## Aspectos comunes y diferencias entre los servicios documentales tradicionales y los de Web Semántica

### **Bibliotecas:**

**Insumo:** el documento fuera o dentro del sistema.

**Proceso:** su descripción bibliográfica y análisis de contenido:

**Producto:** los registros bibliográficos y su agregación en catálogo (OPAC)

### **Web Semántica:**

**Insumo:** documentos = objetos de información, dentro del sistema.

**Proceso:** asignación de atributos (metadatos) legibles por máquina, embebidos en el propio objeto que o en base de datos de metadatos asociados.

**Producto:** visible en el momento de la recuperación a través de los motores de recuperación semántica

*Ref: Méndez, 2007*

# Aspectos comunes y diferencias entre los servicios documentales tradicionales y los de Web Semántica

Tienen en común la lógica de clases, pero no el objetivo



La indización para catalogación y clasificación en OPACs no es igual a la de la Web Semántica

no es lo mismo indizar un objeto que indizar la referencia a ese objeto





**indiza recursos web:  
objetos, no materiales, sino  
objetos digitales**

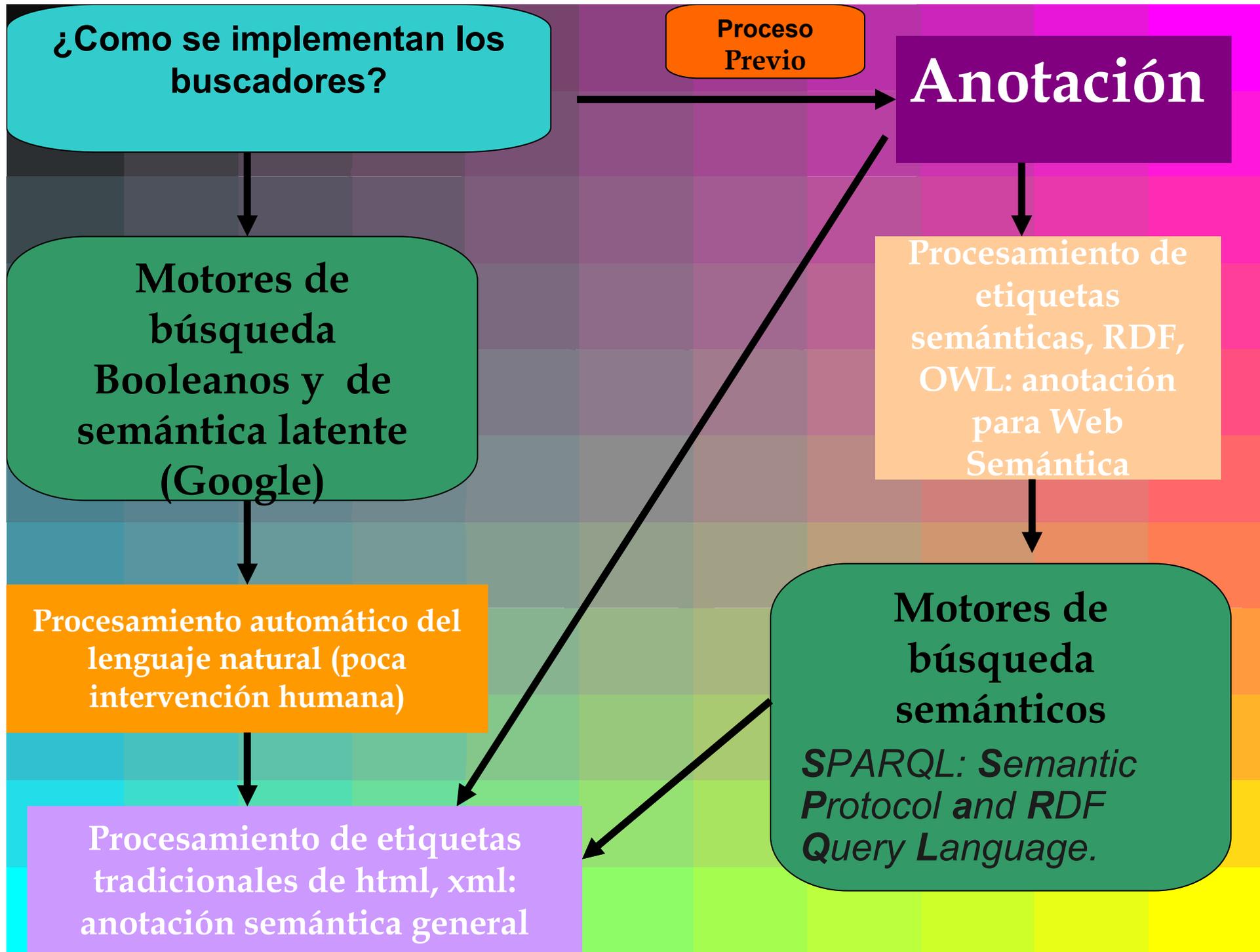
**Esos Objetos son descritos a través de metadatos  
que están especificados en**

- a. una ontología (enfoque top down)**
- b. Esquemas de metadatos (enfoque bottom up)**

**Los items de referencia a objetos digitales: autor, fecha etc:**

**Son similares a los que se usan en la de clasificación y  
catalogación documental**

**Otros aspectos de los objetos digitales como atributos,  
comportamiento, cardinalidad, son expresados con otra  
lógica:**



¿Como se implementan los buscadores?

Proceso Previo

**Anotación**

Motores de búsqueda Booleanos y de semántica latente (Google)

Procesamiento de etiquetas semánticas, RDF, OWL: anotación para Web Semántica

Procesamiento automático del lenguaje natural (poca intervención humana)

Procesamiento de etiquetas tradicionales de html, xml: anotación semántica general

Motores de búsqueda semánticos  
*SPARQL: Semantic Protocol and RDF Query Language.*

## Anotación semántica en general

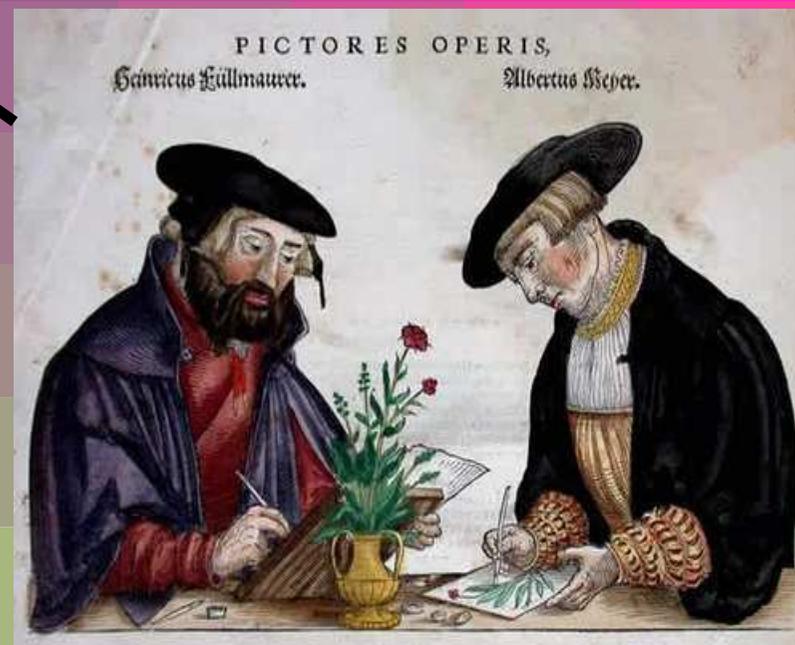
Asociación de una entidad de datos con un elemento de tipo semántico que puede ser: esquema de clasificación, un tesoro, una nota, una glosa: larga tradición bibliotecaria y científica

Anotaciones o etiquetado semántico, una antigua tradición con nueva instrumentación

**Representar y organizar el conocimiento para transmitirlo y conservarlo**

Anotación para Web semántica

Objetivo hacer que las máquinas puedan comprender un dato en uno o varios sentidos y puedan usarlo para tomar decisiones y realizar acciones en determinadas y precisas situaciones





## Anotación proceso y resultado

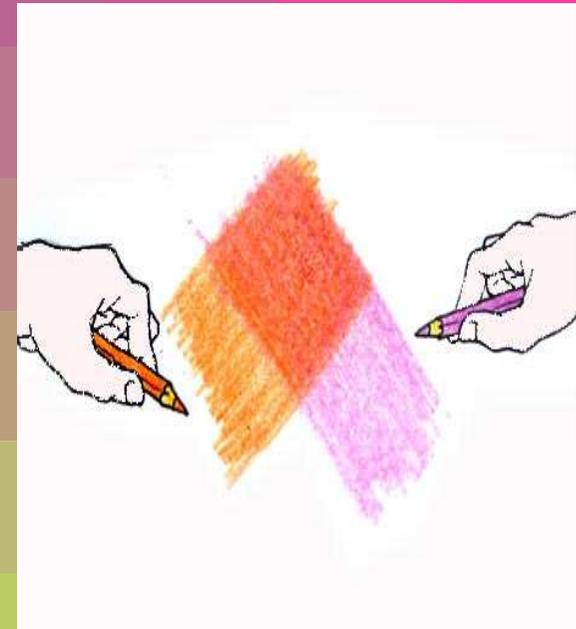
- *El procedimiento para incluir datos comprensibles y procesables en la Web es agregar datos sobre los datos, es decir por medio de la creación de metadata.*
- *Por eso es que usan también los términos de marcado semántico (semantic markup), etiquetado semántico (semantic tagging o semantic labelling*

El concepto de anotación semántica (semantic annotation) conocido también como:

- marcado semántico (semantic markup)
- etiquetado semántico (semantic tagging o semantic labelling)
- El término **anotación** se ha ido imponiendo.

# Anotación proceso y resultado

- Anotación en texto libre  
(*Free-text annotation*)
- pueden ser cualquier tipo de comentarios, notas, explicaciones, referencias, sugerencias, correcciones
- cualquier tipo de indicación externa que puede ser agregada o incluida en un documento Web o en una parte de él, incluso los comentarios en Word o Excell



# Anotación proceso y resultado

## Anotación semántica en general

Significa la asociación de una entidad de datos con un elemento de tipo semántico:

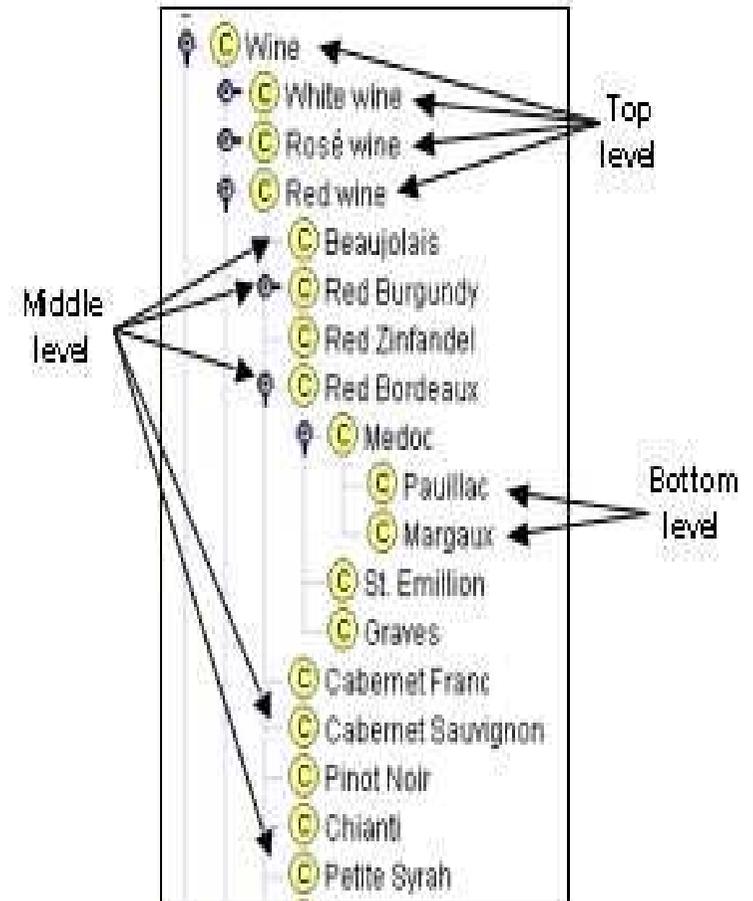
- Un esquema de clasificación, una ontología, un tesoro o cualquier otro instrumento de identificación de conocimiento en un repositorio de información.
- Ejemplo: los descriptores de MeSH de en MEDLINE, una común indización por palabras claves o descriptores, hasta asignaciones de sentido mucho más complejas como los términos de la Gene Ontology a los productos genéticos en UniProt.



## Anotación para Web semántica es más compleja, varios componentes: 1a.

Una ontología que describe el dominio del sistema:

- Términos de una ontología: términos controlados
- El objeto al que se refieren debe responder a una definición ontológica



**Ontología en la Web semántica:** conjunto de términos que refieren a objetos, términos expresados considerando sus propiedades y relaciones en un determinado dominio y sin que eso impida que en el desarrollo de una ontología intervengan diferentes enfoques metodológicos según el campo disciplinario de quienes aportan en la construcción de la ontología o en el dominio de aplicación de la misma



## Bibliotecarios y ontologías: Tenemos mucho que aportar: Aparecen problemas que conocemos en la clasificación documental:

¿es la pizza Margherita una pizza vegetariana?

Errores de construcción lógica

Una clase puede tener más de una superclase y una subclase hereda atributos, aspectos o restricciones de diferentes jerarquías. Y pueden ser resueltos de manera diferentes

Magherita\_Pizza

RDFS:COMMENT:  
A magherita pizza kind of pizza and has has, amongst other things, Mozzarella and Tomato toppings

Asserted Inferred

ASSERTED CONDITIONS:

- NECESSARY
- Pizza
- $\exists$  has\_topping Mozzarella\_topping
- $\exists$  has\_topping Tomato\_topping

Name

Vegetarian\_pizza

rdfs:comment  
A vegetarian pizza is ANY pizza that has no meat topping and no fish topping.

Asserted Inferred

Asserted Conditions

- NECESSARY &
- Pizza
- $\neg(\exists$  has\_topping Fish\_topping)
- $\neg(\exists$  has\_topping Meat\_topping)

# Anotación para Web semántica: componentes

## 1b.

- Proceso de reconocimiento de las instancias de datos, es decir la detección de los objetos que responden a esa ontología, ya sea en un *corpus* de documentos o en un reservorio de objetos multimediales que es manejado por el sistema que hace uso de la ontología.



# Anotación para Web semántica: componentes

1c.

- El tercer componente es la acción que produce la anotación como resultado: se agrega contenido semántico a las instancias que responden a las propiedades y atributos de una ontología u otra estructura de representación de conocimiento
- puede ser manual, automática o semiautomática. Para ello se utilizan las herramientas de anotación semántica (*semantic annotation tools*)



# Tipos de herramientas de anotación semántica (*semantic annotation tools*)

Anotación en línea (*Inline annotation*): la *metadata* está incrustada en el documento

*Embedded metadata*

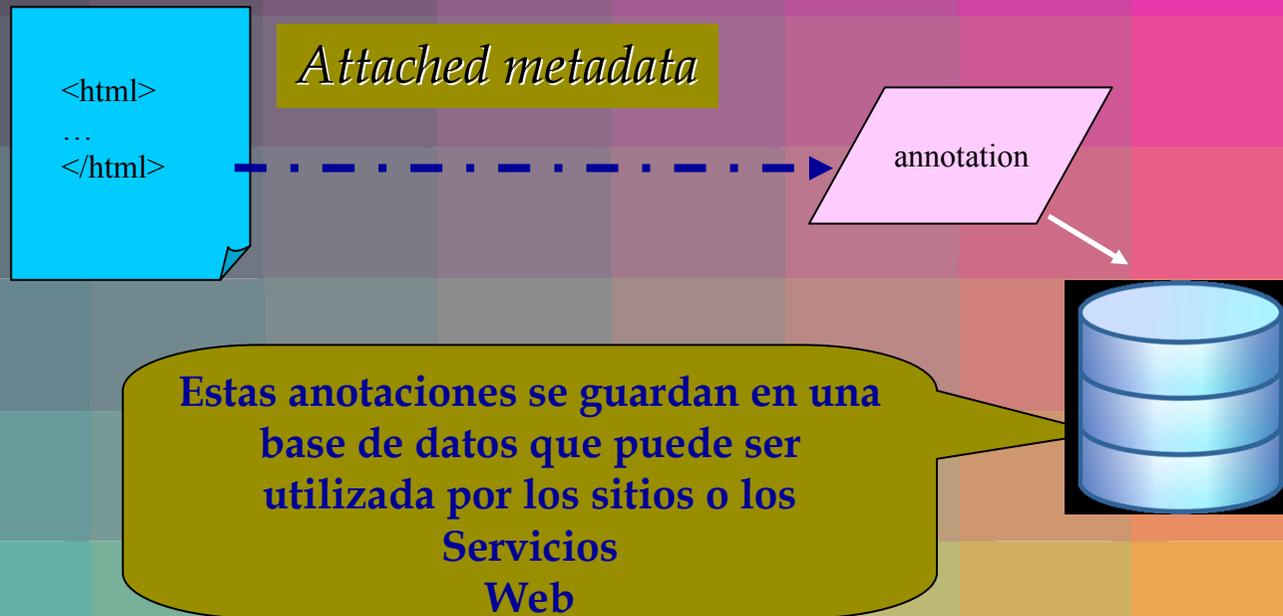
```
<html>  
...  
<annot>  
...  
</html>
```

Se centra en anotar usando RDF, OWL, es decir metadata que pueda ser interpretada por las computadoras

Se llama también:  
Semantic Authoring  
o  
Bottom-up annotation

## Tipos de herramientas de anotación semántica (*semantic annotation tools*)

Standoff annotation: anotación separada, está escrita y archivada fuera del documento o de los objetos Web a los que refiere



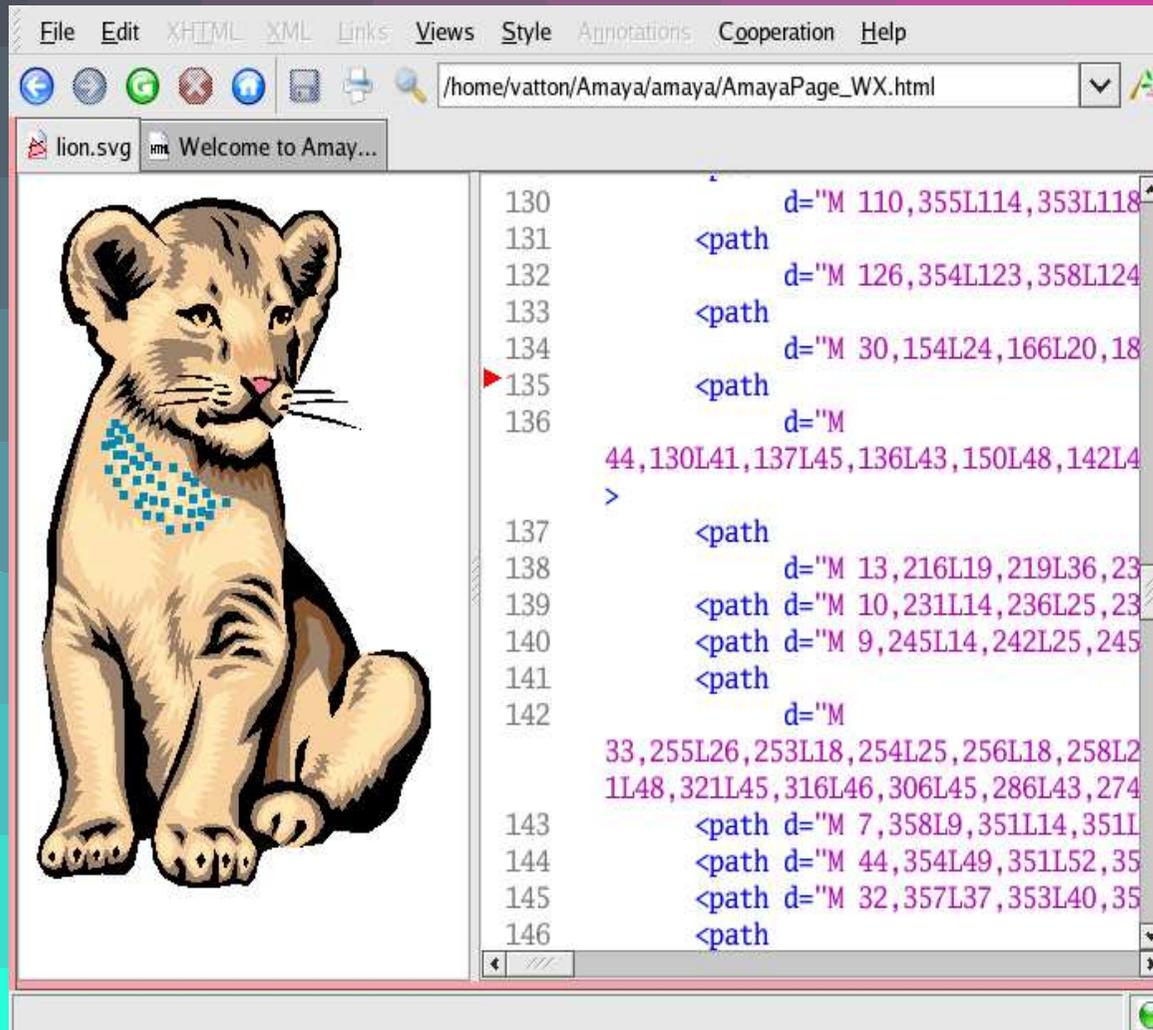
Es preferible desde el punto de vista de la interoperabilidad

Llamada también anotación top-down

Valiosa para relevar información existente

# Ejemplo de herramientas de anotación

- Amaya (embedded annotation)
- Fuente: <http://www.w3.org/Amaya/screenshots/Overview.html>. [



## *Ejemplo de herramientas de anotación*

- Gans, J. Multi-scale, Multi-genome, Multi-platform Visualization and Analysis. Los Alamos National Lab, Bioscience Division.  
[http://public.lanl.gov/jgans/genomorama/genomorama\\_doc.html](http://public.lanl.gov/jgans/genomorama/genomorama_doc.html)

Edit Annotation

Range:

Type:  Strand:

Accession:  GI:  EC:

LOCUS:  LOCUS Tag:

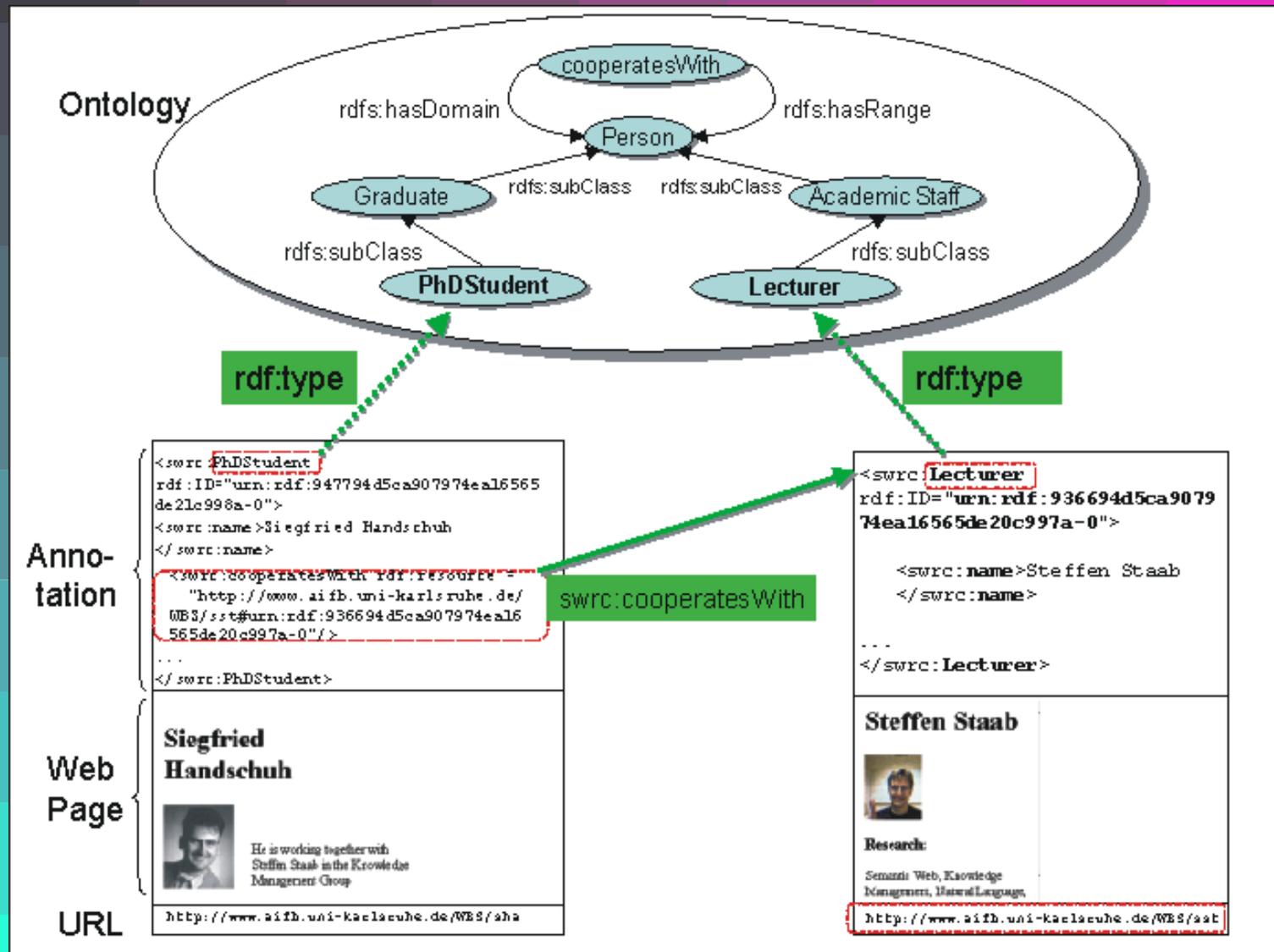
Product:

Note

probable transposase, len: 771 aa; identical to corresponding CDS from Y.pestis KIM5 pCD1 (EMBL:AF053946) (771 aa), fasta scores; opt: 5136 z-score: 6916.0 E(): 0, 100.0% identity in 771 aa overlap. Highly similar to TR:Q00037 (EMBL:X60200), tnpA, E.coli transposon gamma-delta transposase (Tn1000) (1002 aa) (71.2% identity in 1002 aa overlap). The difference in length is accounted for by a large internal deletion from the Y.pestis CDS

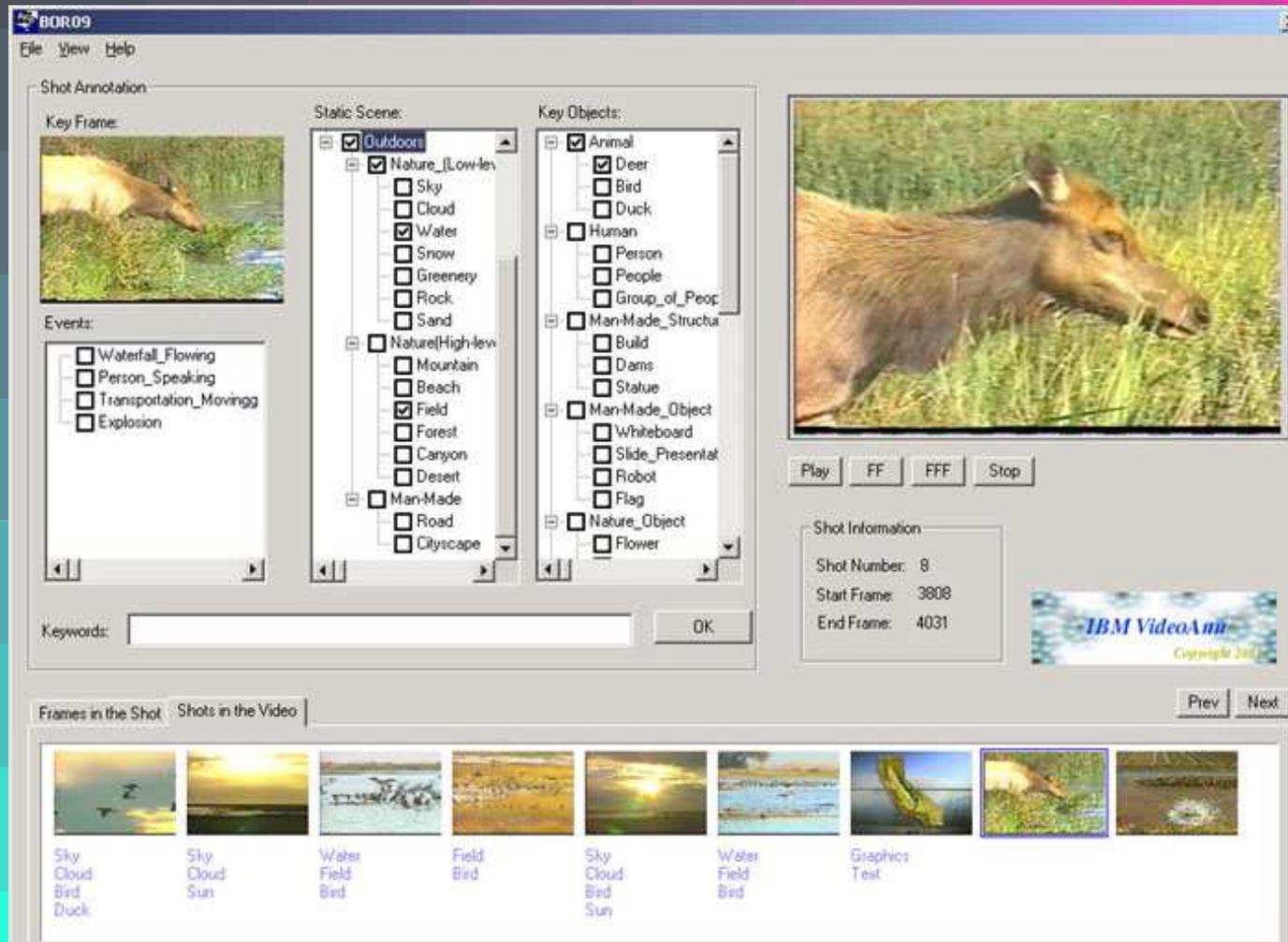
# Ejemplo de herramientas de anotación

- Anotación para Web Semántica. Staab, S. Handschuh, S. Authoring and Annotation of Web Pages in CREAM. <http://www2002.org/CDROM/refereed/506/> [21-07-2009]



# Ejemplo de herramientas de anotación

- Video Semantic Summarization Systems
- <http://www.research.ibm.com/MediaStar/VideoAnn.html>. [21-06-2009]



## Algunas herramientas de anotación

Tool	Source	Type	Language	Origin
<b>Ontomat</b>	Html	Embedded	Formal (DAML+OIL)	Univ. of Karlsruhe, Germany
<b>Mnm</b>	Html	Embedded (file XML), Attached	Formal (DAML+OIL, RDF)	KMI The open Un. - Depart. of CS, Univ. of Sheffield. UK
<b>Smore</b>	Photo, mail, html	Embedded	Formal (RDF, DAML+OIL, OWL)	University of Mariland
<b>Cohse</b>	Doc/ html	Attached	Formal DAML+OIL	Depart. of CS Univ of Manchester UK
<b>Trellis</b>		Attached	Formal (OWL)	USC.Information Science Institute, Depart. of CS,Univ. of Sheffield. UK
<b>Melita</b>	Html	Attached	Formal	
<b>Kim</b>	Txt,Html, xml	Attacched		Ontotext Lab, Sirma AI, Bulgaria
<b>Annotea</b>	Html/xml	Attached	socio-semantic web RDF/XML	W3C INRIA Rhône-Alpes W3C MIT/LCS



# Tendencias en desarrollos de Web Semántica en servicios de bibliotecas

- Línea de horizontalización de recursos a través del enlazado de datos, proceso transformación de catálogos a RDF: Caso Biblioteca Nacional de España
- Línea de verticalización (profundización semántica de contenidos) Ref: OPEN GLAM (<http://openglam.org>)  
Caso local: Acervo Digital Anotado de Literatura Argentina.  
Proyecto de la Dirección de Biblioteca Nacional Argentina
- Para complementar estas tendencias: Registros de esquemas de metadatos: ISO 11179 Compartir diccionarios de datos y el proceso de registro

## Línea de horizontalización de recursos a través del enlazado de datos, proceso transformación de catálogos a RDF: Caso Biblioteca Nacional de España

- **Transformación y publicación de los catálogos bibliográfico y de autoridades en RDF para que los datos sean accesibles en la Web semántica y enlazarlos con otros conjuntos de datos de la nube como el Fichero de Autoridades Virtual Internacional (VIAF), y en un futuro próximo, con Dbpedia.**  
**Se transformaron 4.2 millones de registros almacenados en formato MARC 21 en los catálogos bibliográficos y de autoridades utilizando la herramienta MARiMbA (MArc Rdf and Mappings generAtor).**

- Línea: verticalización (profundización semántica de contenidos)

The screenshot shows the Europeana.eu website homepage. At the top, there is a navigation bar with links for Home, Explore, Help, About Us, Follow Us, My Europeana, and a language selection dropdown. The main header features the Europeana logo (a stylized 'e' with a globe) and the tagline 'think culture'. Below the logo is a search bar with a 'Search' button and a 'Help' link. A social media bar shows 'Like <6.2k', '+1', and '4K'. The main content area is a carousel of featured items:

- World War 1 family history roadshows in 5 countries, spring 2012 (opens in a new window)**: A man in a military uniform holding a flag.
- Exhibitions: Yiddish Theatre in London**: A black and white photograph of a man in a top hat.
- Exhibition: Untold stories of the First World War**: A man in a military uniform playing a trumpet.
- Europeana Remix: First interactive video**: A woman's face in a video frame.

Below the carousel are three columns of content:

- From the blog**:
  - Ministers on Europeana: Géza Szócs**: Article by Géza Szócs, Minister of State for Culture: We have no knowledge of any equestrian statue on public display, or indeed, any statue at all which was...
  - New Virtual Exhibition: Untold Stories of WW1**: In the latest Europeana Virtual Exhibition we showcase some of the most extraordinary and untold real-life stories from the First
- Featured item**: A postcard with handwritten text and a circular stamp.
- Social media activities**:
  - La Joconde - Gioconda & Mona Lisa | Leonardo da Vinci**: 62 personas recomiendan esto.
  - Incursões Monárquicas**: 15 personas recomiendan esto.
  - Bildnis Albert Einstein (1879-1955) | Schmutzer, Ferdinand**: 43 personas recomiendan esto.
  - Molecular model: DNA**

- Línea: verticalización (profundización semántica de contenidos)

The screenshot shows the website [bami.cilea.it](http://bami.cilea.it) in a browser window. The page features the BAMI logo (BIBLIOTECA APERTA DI MILANO) and a navigation link "Visualizza la linea temporale". The main content area contains a description of the project "ICT Milano Città Cablata – Biblioteca Aperta di Milano" and a search section with the following elements:

**BAMI**

Il progetto "ICT Milano Città Cablata – Biblioteca Aperta di Milano", attuato grazie al cofinanziamento dell'Assessorato Ricerche ed Innovazione del Comune di Milano e del CILEA, ha permesso la realizzazione di una **Semantic Web Digital Library**: una infrastruttura tecnologica in grado di far fruire agli utenti di internet una biblioteca di risorse culturali digitalizzate associate tra loro da informazioni che ne specificano il contesto semantico.

Il tema selezionato per la realizzazione di questa prima fase del progetto è stato l'Ottocento musicale milanese. La musica, linguaggio universalmente conosciuto, rappresenta a pieno titolo Milano nel mondo. In particolare l'Ottocento vide Milano affermarsi come centro vitale e propulsore attraverso i suoi teatri, i suoi musicisti e i suoi editori. Il progetto ha consentito la digitalizzazione e la messa on line di documenti di pregio, a stampa e manoscritti appartenenti a questo fulgido periodo della storia di Milano.

Inizia la ricerca

Q Digita il termine da ricercare

Q Ricerca solo tra i documenti digitalizzati

Q Documenti e risorse (220) ~ Documenti e risorse

Q Enti e organizzazioni (29) ~ Enti e organizzazioni

Q Esemplari (199) ~ Esemplari

Q Luoghi (192) ~ Luoghi

Q Opere (448) ~ Opere, Libretti, PublicationWork

Q Performances (29) ~ Performances

Q Persone (366) ~ Persone

- Línea: verticalización (profundización semántica de contenidos)

Para consultar sobre el avance del Acervo de Literatura Argentina:

Tenemos un wiki de trabajo:

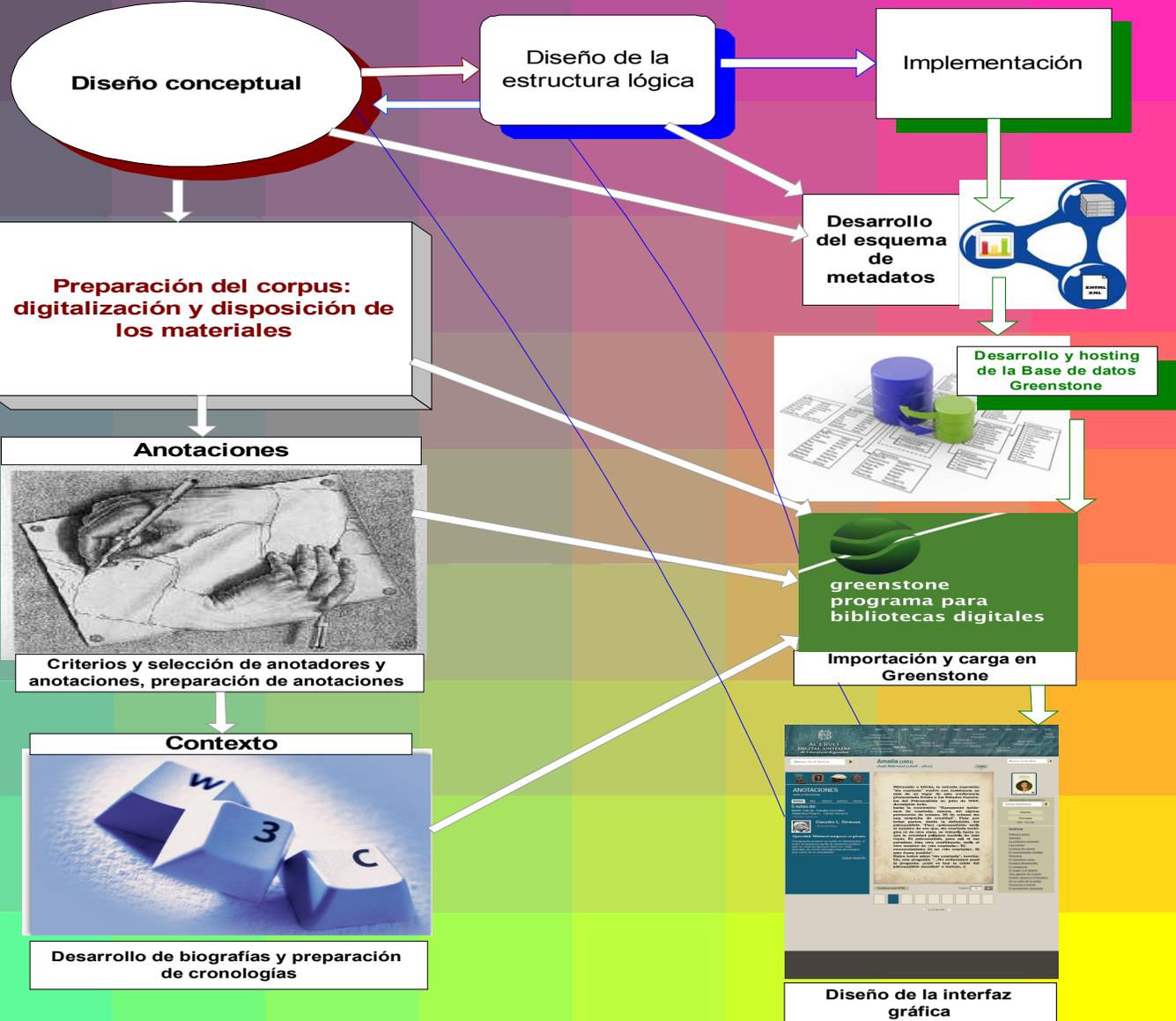
<http://acervodigitalanotadoliteraturaargentina.wikispaces.com/>

Y una maqueta del sitio:

<http://www.acervoliteraturaargentina.net>



**Pasos del desarrollo informático**



## En resumen: Web Semántica en bibliotecas tenemos tres opciones complementarias:

- Línea de horizontalización de recursos a través del enlazado de datos: uso de DC pero cada vez más extendidos
- Línea de verticalización (profundización semántica de contenidos) generan nueva metadata específica para contenidos muy variados

Para complementar estas tendencias: ISO/IEC 19773:2011 basada en ISO/IEC 11179-3, ISO/IEC 19763 y OASIS EBXML: Registros públicos de esquemas de metadatos RDF y SKOS.

**Objetivo:** Compartir diccionarios de datos y el proceso de registro. Se pueden consultar, adoptar y compartir desde una Base de Conocimiento. Aunque el mayor productor de bases de datos Oracle, canceló su apoyo a este estándar para desarrollar su propio sistema comercial, los gobiernos y la comunidad de código abierto le está dando su apoyo.

Ejs: Meteor (Metadata on line register) del Gobierno de Australia, repositorio metadata nacional para salud, también *Statistics Canada's metadata registry*, para comunidad open source: <http://metadataregistry>.

## Registros públicos de esquemas de metadatos RDF y SKOS



### **Simple Knowledge Organization**

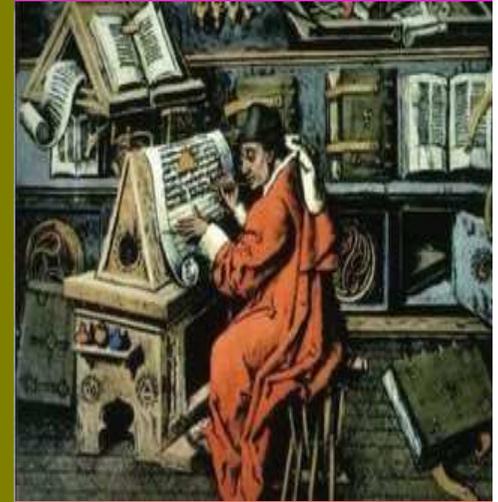
System iniciativa W3C, aplicación de RDF para representar la estructura básica y el contenido de esquemas conceptuales: listas encabezamientos de materia, taxonomías, esquemas de clasificación, tesauros y cualquier tipo de vocabulario controlado.

**Ultimos avances: Grupo Incubadora de Datos Enlazados Bibliotecarios: Conjuntos de Datos, Vocabularios controlados y Conjuntos de Elementos de Metadatos - Informe del grupo incubadora del W3C Incubator 25 de Octubre 2011**

<http://skos.um.es/Incubator/ld/XGR-ld-vocabdataset/>

## En resumen...

Se puede convertir los OPAC en RDF, pero también necesario bibliotecas digitales o referenciaciales con anotación de metadatos muy específicos que mejoran en la medida en que más personas participan anotando, glosando los textos, haciendo comentarios, como hace siglos en el *scriptorium* monástico



```
<!DOCTYPE HTML PUBLIC "-//W3C//DTD HTML 4.01//EN" "http://www.w3.org/TR/html4/strict.dtd">
<html>
<head>
<meta name="TITLE" content="...">
<meta name="KEYWORDS" content="...">
<meta name="DESCRIPTION" content="...">
<link rel="stylesheet" type="text/css" href="...">
<script language="javascript" type="text/javascript">
</script>
</head>
<body bgcolor="#ffffff" width="100%">
```

**Complementariamente es necesario compartir metadata.: permite además obtener indicadores para mejoramiento de servicios e investigación. Web del conocimiento, no solo de la recuperación de información**

**compartir esquemas de metadatos para aprovechar y armonizar las anotaciones: dimensión social del tratamiento de contenido... nos lleva al tema siguiente:**

Además de los motores de búsqueda tenemos otro gran escenario para la circulación del conocimiento que son los medios sociales

Presentamos un cuadro de situación de la circulación del conocimiento, de los saberes y experiencias en los medios sociales.

El objetivo es considerar nuestra propia intervención como usuarios así como los aspectos que debemos tener presentes como bibliotecarios

## El conocimiento en los medios sociales: visiones





**Visión optimista:** inteligencia colectiva que hizo realidad el postulado de que numerosos juicios sencillos producen un conjunto más complejo y preciso que el que daría una sola persona.



**Visión de marketing digital:** se basa en el concepto de SMO o *Social Media Optimization*. Su preocupación es computar el tiempo e interacciones de los usuarios aprovechar al máximo los beneficios que aporta en términos de tráfico y de posicionamiento en buscadores.



**Visión semiológica:** el impacto histórico de Internet es comparable a la aparición de la escritura y a la imprenta, pero hay un aspecto de masificación nunca antes logrado: es la posibilidad de gestionar en gran escala el *metaconocimiento*, es decir los índices, rastros, símbolos.

# Hipótesis sobre el conocimiento los medios sociales

***Están en el sustrato de nuestras interacciones y conviven, no solo en diferentes espacios, sino que muchos de nosotros actuamos suponiendo una o varias de ellas***

**Conspirativa:** Hay una manipulación constante y generalizada por parte de grandes corporaciones económicas y sociopolíticas tendiente a obtener el máximo beneficio de nuestra presencia e interacciones en los medios sociales

**Caótica:** entraña peligros de todo tipo por falta de controles y seguridad: terrorismo, psicópatas, pedofilia, etc.

**Indefensión:** Las técnicas que subyacen no son transparentes y por tanto nos dejan en una posición vulnerable

**Constructivista:** es posible generar espacios de emergencia de recursos respetuosos del planeta y la dignidad humana

## Procedimientos más difundidos en el tráfico de conocimientos en los medio sociales



Tal como suponen las hipótesis más escépticas la fuerza de la Web es que es una red, pero la jerarquía entra de manera subrepticia debido a que:

- Las empresas que actúan en la Web están en competencia permanente por ganar la atención
- La verificación directa de la enorme masa de información no es posible a las personas individuales
- La coincidencia mecánica con lo solicitado, aún separando los enlaces pagados de los que no lo son, no es suficiente

*Entran en juego otras formas de jerarquizar la información para evidenciar lo que cada uno tiene para ofrecer en el tráfico de conocimientos en los medios sociales*

# Procesos de tratamiento y de Social Media Optimization (SMO)

*Los medios sociales funcionan filtrando y confrontando las elecciones individuales con las preferencias colectivas que a su vez también han sido filtradas.*

- agentes de software analizan cada hecho individual (una búsqueda en Google, consulta o escribe una entrada en Wikipedia, hace transacciones en eBay, crea su blog, arma su red Facebook)
- computan la relación de cada interacción individual con la totalidad de interacciones de otra gran masa de personas
- de esa correlación surgen parámetros que se usan para filtrar y procesar la nueva información así como la existente.

**Un proceso circular que tiene como resultado la creación de sesgos en la información que a su vez afectan nuestras interacciones individuales.**

# Las recomendaciones



Son filtros automáticos que analizan las *preferencias* individuales y hacen recomendaciones basadas en cálculos que generalizan las interacciones de multitudes de personas

Múltiples usos: venta electrónica de producto y servicios: compra de libros, ropa vacaciones, parejas sentimentales

*Ventaja: nos sugiere qué han hecho los otros*

*Desventaja: Elitización de los resultados*

**PageRank:** ranquea popularidad de los sitios, pero el valor se calcula con la cantidad de enlaces pero da también puntaje que sean sitios ya marcado con PageRank.

El valor que Google asigna a una página puede aumentar aunque pocos sitios hayan insertado el enlace del mismo, siempre y cuando provengan de páginas consideradas valiosas para Google, es decir que tengan un PageRank alto.

Es útil para detectar sitios atendibles, pero genera una “elitización” de los resultados. Daña a los sitios recientes o a los que no se plantean la política de insertar enlaces a otros sitios en su espacio.

➔ *Peligroso: deja poco espacio para la originalidad y lo nuevo suele quedar aislado*

## Las reputaciones



*El mercado electrónico sigue siendo próspero pero siempre con un halo de duda y sospecha.*

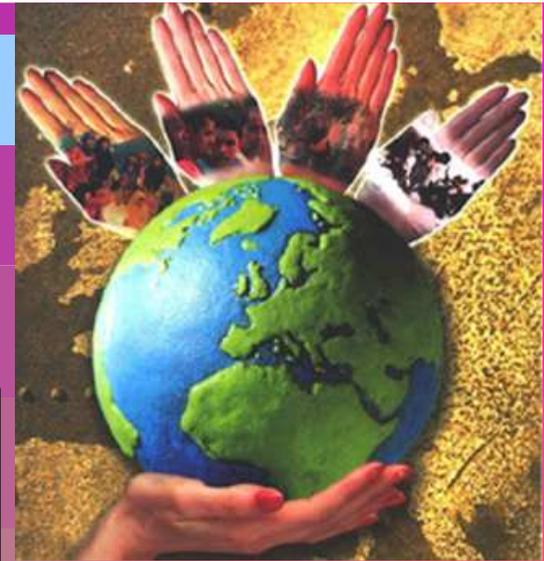
**Son métodos de filtrado que hacen previsiones a partir del *comportamiento* de los usuarios**

Ej. eBay luego de cada transacción tanto el vendedor como el comprador hacen una evaluación que es resumida y vista por los usuarios que se basan en ellas para futuras compras o ventas

**Distorsiones: muchos vendedores venden muy barato algunos productos gancho para ganar reputación**

## La cooperación

Ej: Proyectos de contenido abierto colaborativo como Wikipedia.



*Positivo: Increíble oportunidad de cooperación en la construcción y acceso al conocimiento jamás registrada*

### Algunos problemas

**Unidireccionalidad lingüística y cultural: demasiadas traducciones del inglés, la producción en castellano crece en forma genuina y es también traducida a otras lenguas. No hay formas de estimular la producción en lenguas locales o minoritarias por lo que se produce en esas lenguas no es fácil que sea pasado al inglés o español.**

**Disminución de colaboradores, discusiones sobre las forma de revisión de los aportes**

## Coautoría: Apertura de códigos fuente y derechos de autor



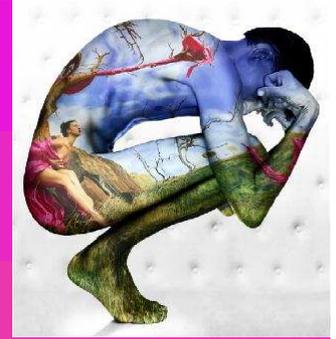
**Ej Comunidades que desarrollan programas *Open Source*, autorías *CC* y *Copy Left***

*Inconveniente del Open Source:* el trabajo de control es manual, el código es accesible a una grupo que puede corregir o borrar según criterios personales o poco compartidos, comercialización de esfuerzos colectivos.

***Conocimiento colectivo que se apoya en esfuerzos individuales, tarea común con normas de cooperación compartidas y explícitas y que ha marcado un punto de no retorno en la producción intelectual en el mundo.***

## En resumen....

**Las redes estimulan la masificación y globalización de interacciones para dar potencia a los algoritmos de recomendación, reputación y ranking, estamos regalando información con cada interacción**



**Opciones para bibliotecas: utilizar marcado semántico para web social, integrado con marcado para contenido**

- FOAF (Friend of a friend) Es un ontología que sirve para identificar personas dentro de las redes sociales para hacer una red de ellas en forma transversal a los recursos: [http://wiki.foaf-project.org/w/Main\\_Page](http://wiki.foaf-project.org/w/Main_Page)
- SIOC initiative (Semantically-Interlinked Online Communities) es otra ontología que sirve para interconectar sitios de discusión, desde blogs, foros y listas de correo. <http://sioc-project.org/>
- SKOS (Simple Knowledge Organization System) aplicación de RDF para representar la estructura básica y el contenido de esquemas conceptuales como listas tesauros, encabezamientos de materia y taxonomías.

Opciones para bibliotecas: utilizar marcado semántico para web social, integrado con marcado para contenido: FOAF

## FOAF : Friend of Friend



FLICKR

```
<foaf:Person rdf:about="http://apassant.net/home/2007/12/flickrdf/people/33669349@N00">
  <foaf:name>Alexandre Passant</foaf:name>
  <foaf:mbox_shalsum>80248cbb1109104d97aae884138a6afcd688bd2</foaf:mbox_shalsum>
  <geonames:locatedIn rdf:resource="http://sws.geonames.org/3038213/" />
  <foaf:holdsAccount rdf:resource="http://apassant.net/home/2007/12/flickrdf/user/33669349@N00" />
  <foaf:knows rdf:resource="http://apassant.net/home/2007/12/flickrdf/people/86846122@N00" />
  <foaf:knows rdf:resource="http://apassant.net/home/2007/12/flickrdf/people/32233977@N00" />
  <foaf:knows rdf:resource="http://apassant.net/home/2007/12/flickrdf/people/43184127@N00" />
  <foaf:knows rdf:resource="http://apassant.net/home/2007/12/flickrdf/people/24266175@N00" />
  <foaf:knows rdf:resource="http://apassant.net/home/2007/12/flickrdf/people/7787294@N06" />
  <sioc:member_of rdf:resource="http://apassant.net/home/2007/12/flickrdf/groups/49656594@N00" />
  <sioc:member_of rdf:resource="http://apassant.net/home/2007/12/flickrdf/groups/23989049@N00" />
</foaf:Person>
```

TRIPLES FROM

TWITTER

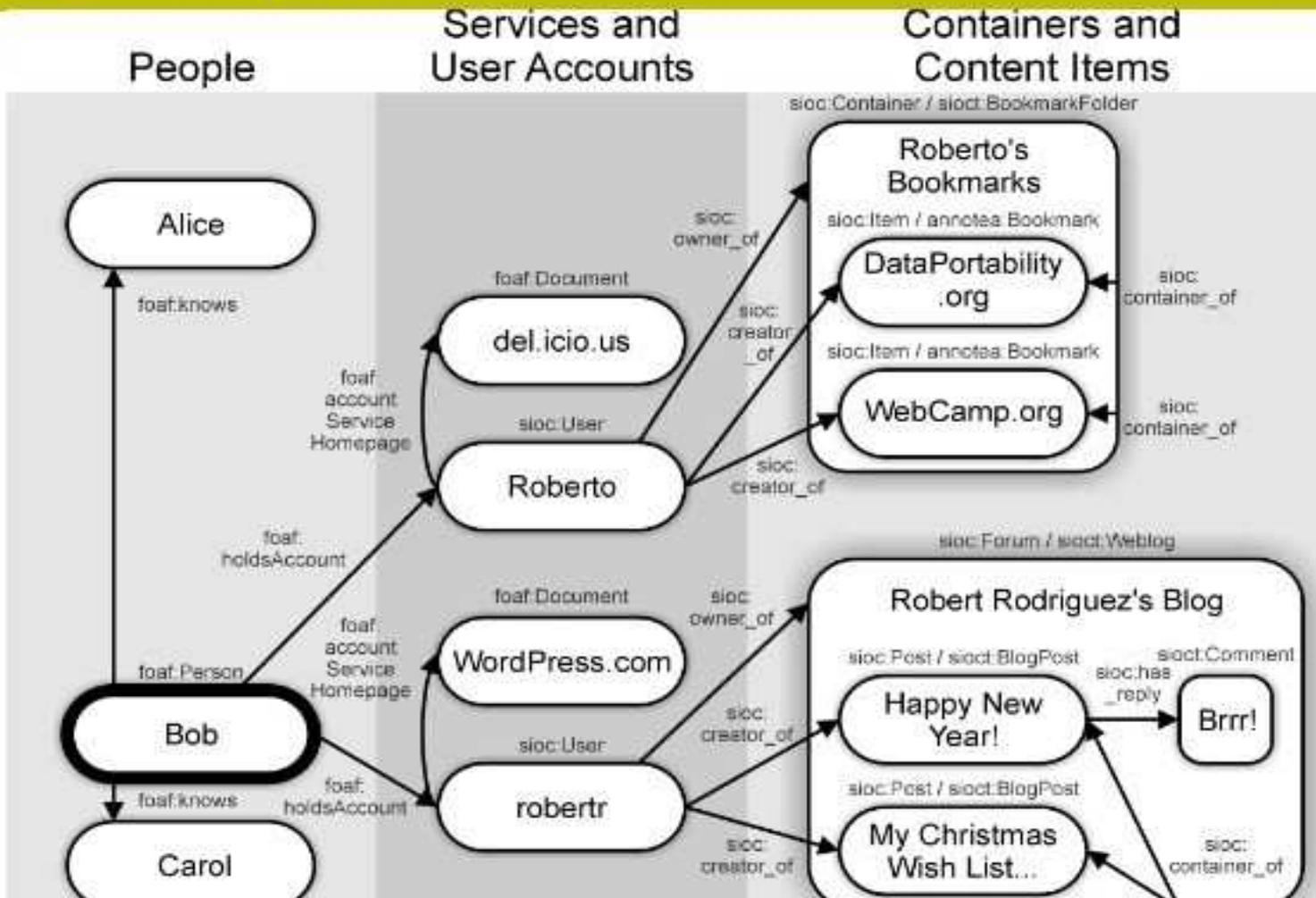
```
<foaf:Agent rdf:about="http://twitter.com/terraces">
  <foaf:knows rdf:resource="http://twitter.com/factoryjoe" />
  <foaf:knows rdf:resource="http://twitter.com/danja" />
  <foaf:knows rdf:resource="http://twitter.com/edd" />
  <foaf:knows rdf:resource="http://twitter.com/dajobe" />
  <foaf:knows rdf:resource="http://twitter.com/iand" />
  <foaf:knows rdf:resource="http://twitter.com/ldodds" />
  <foaf:knows rdf:resource="http://twitter.com/tomorris" />
```

```
<foaf:Person rdf:ID="me">
  <foaf:name>Alexandre Passant</foaf:name>
  <foaf:givenname>Alexandre</foaf:givenname>
  <foaf:family_name>Passant</foaf:family_name>
  <foaf:gender/>
```

Source Alexandre Passant

Opciones para bibliotecas: utilizar marcado semántico para web social, integrado con marcado para contenido: uso de SIOC

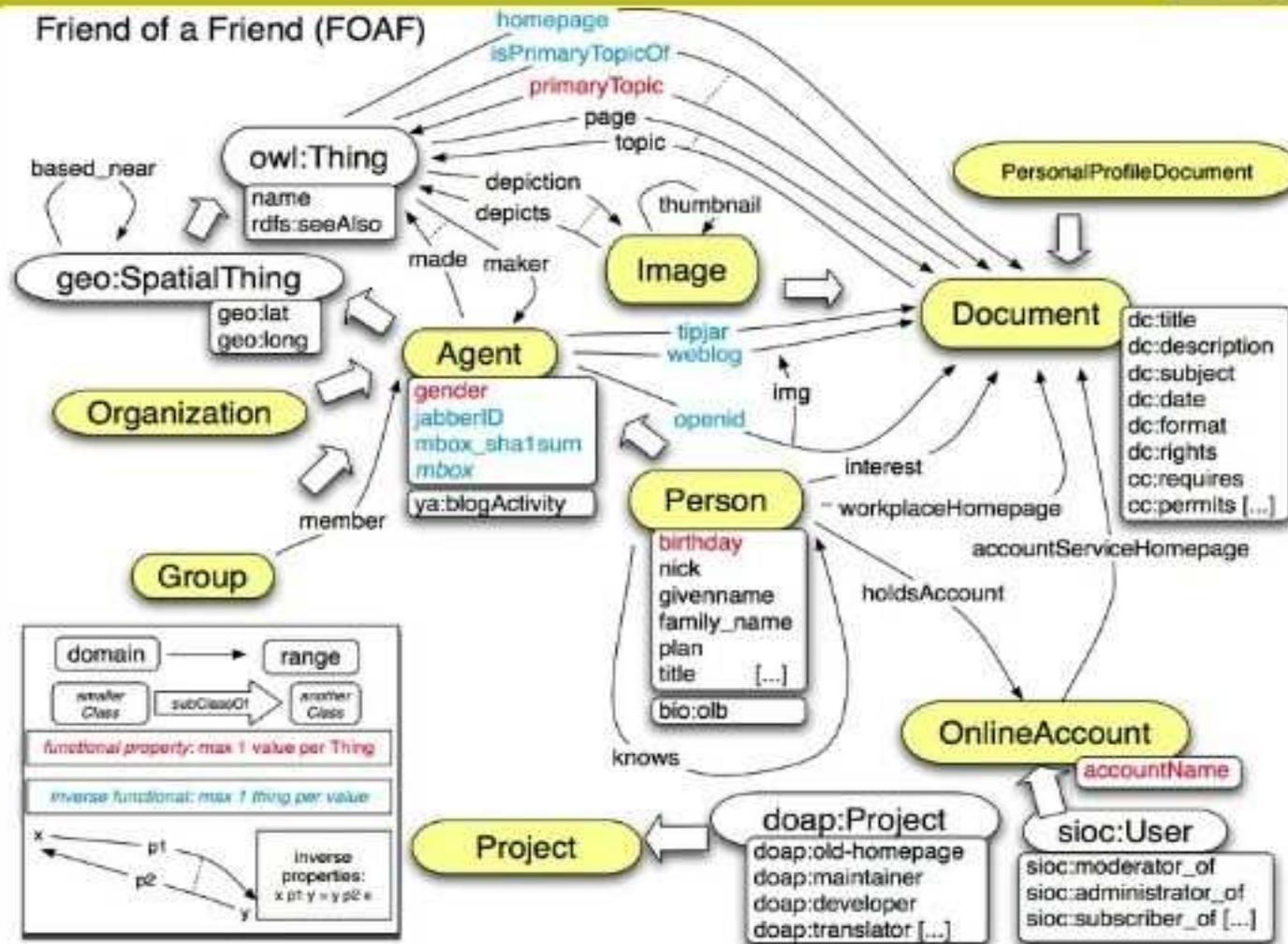
# SIOC: Semantically-Interlinked Online Communities



# Opciones: utilizar marcado semántico para web social, integrado con marcado para contenido: Integración de FOAF , SIOC DOAP en uso en comunidades de software libre

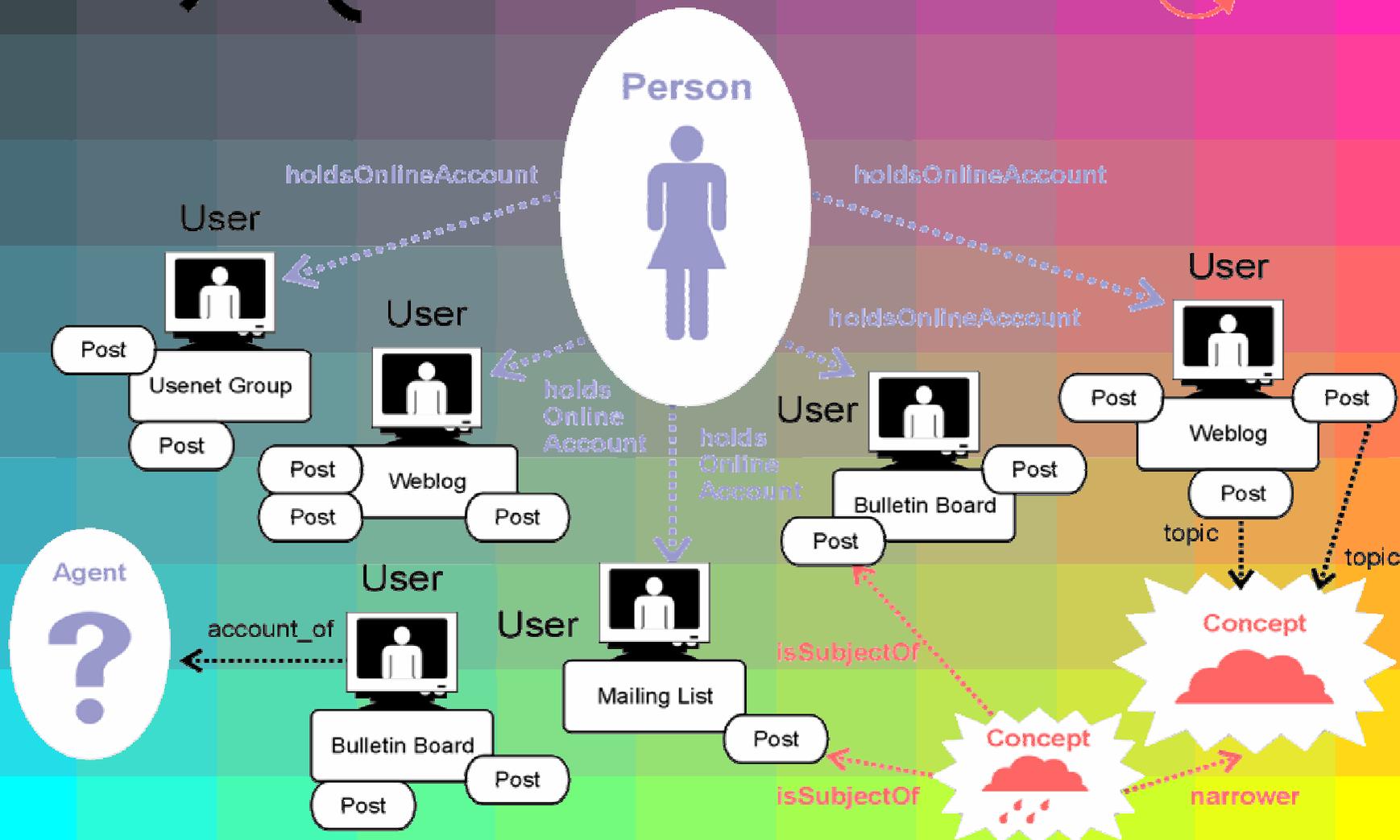


## FOAF : Friend of Friend



# El futuro de bibliotecas en Web 2.0: marcado semántico para web social, integrado con marcado para contenido: uso combinado de metadatos semánticos

## SIOC + FOAF + SKOS



**El futuro de bibliotecas en Web 2. 0:  
marcado semántico para web social,  
integrado con marcado para contenido:  
uso combinado de metadatos semánticos:  
aún pocas posibilidades:**



**WordPress SIOC Exporter: genera metadatos SIOC instalando 2 archivos en la carpeta de plugins en WordPress autoalojado**

**Semantic radar, complemento para Firefox aun en desarrollo permitirá ver los contenidos RDF de cualquier página que los contenga (además de realizar un ping automático a Pingthesemanticweb, servicio que recoge RDF**

**identi.ca red social y microblogging en software libre, permite la libre exportación y el intercambio de personal y de "Amigos", basada en el estándar FOAF.**

**Aún no he localizado widget gratuito para exportar Foaf a Facebook, tiene política Facebook para exportar los contactos de esta red social a otras redes como Google+ y LinkedIn, las herramientas como "Facebook Friend Exporter" no funcionan.**

## ✓ **Conclusión sobre web semántica y bibliotecas**

### **Tres opciones complementarias:**

- **Línea de horizontalización de recursos a través del enlazado de datos: uso de DC pero cada vez más extendidos**
- **Línea de verticalización (profundización semántica de contenidos) generan nueva metadata específica para contenidos muy variados**

**Para complementar estas tendencias: ISO/IEC 19773:2011 basada en ISO/IEC 11179-3, ISO/IEC 19763 y OASIS EBXML: Registros públicos de esquemas de metadatos RDF y SKOS.**

## ✓ **Conclusión sobre circulación de conocimiento en Redes sociales**

**Las redes estimulan la masificación y globalización de interacciones para dar potencia a los algoritmos de recomendación, reputación y ranking, estamos regalando información con cada interacción**

**El futuro de bibliotecas en Web 2. 0: marcado semántico para web social, integrado con marcado para contenido, uso combinado de metadatos semánticos: aún pocas posibilidades**

# Bibliografía

Atanas, K. et al. Semantic Annotation, Indexing, and Retrieval. Ontotext Lab.

[http://www.ontotext.com/publications/SemAIR\\_ISWC169.pdf](http://www.ontotext.com/publications/SemAIR_ISWC169.pdf)

Berrueta, D.; Cantera, J.; Jiménez, M. et al. (2007) Aplicación de las tecnologías de la Web Semántica a la problemática de cumplimentación automática de formularios en la Web Móvil . Fundación CTIC de Asturias, Telefónica I+D, Universidad Politécnica de Madrid. Accesible en:

<http://www.w3c.es/Eventos/2007/MWeb/Comunicaciones/Papers/p4.pdf>

Bosch, M. (2012) La piel de jaguar: La escritura móvil. Heurística y hermenéutica en el tratamiento informático de documentos. Saarbrücken, EAE, Lambert Academic Publishing, ISBN: 978-3-8473-6869-4.

Bosch, Mela. Ontologies, (2006) Different Reasoning Strategies, Different Logics, Different Kinds of Knowledge Representation: Working Together. Knowledge Organization, Ergon Verlag. 33(3) 153-159. ISSN . 0943-7444 Knowl.Org.

Day, Michael (2005) Preservation metadata initiatives: practicality, sustainability, and interoperability1. UKOLN, University of Bath, Accesible en [.http://www.ukoln.ac.uk/preservation/publications/erpanet-marburg/day-paper.pdf](http://www.ukoln.ac.uk/preservation/publications/erpanet-marburg/day-paper.pdf)

Encuentro de Catalogación y Metadatos (4 : 2009 : Mexico, D.F.) IV Encuentro de Catalogación y metadatos : memoria, 4 al 6 de noviembre de 2009 / comp. Filiberto Felipe Martínez, Arellano. México : UNAM, Centro Universitario de Investigaciones Bibliotecológicas, 2010. xi, 429 p. (Sistematización de la Información Documental) ISBN: 978-607-02-1427-

1 [http://132.248.242.3/~publica/archivos/libros/iv\\_encuentro\\_catalogacion.pdf](http://132.248.242.3/~publica/archivos/libros/iv_encuentro_catalogacion.pdf)

Garlatti, S. Telecom Uk. Semantic Web in action. Ontologies for Social Web

<http://www.slideshare.net/garlatti/social-web-ontologies-7714997>

Méndez, E. (2007) Catalogación/organización de documentos digitales: estado de la cuestión, tendencias y perspectivas desde España. Madrid, Dpto. Biblioteconomía y Documentación, Universidad Carlos III de Madrid. Accesible en:

[http://www.oasis-open.org/committees/download.php/6233/c002349\\_ISO\\_IEC\\_11179-1\\_1999%28E%29.pdf](http://www.oasis-open.org/committees/download.php/6233/c002349_ISO_IEC_11179-1_1999%28E%29.pdf)

Pastor-Sánchez, Juan-Antonio.(2011) Tecnologías de la Web Semántica Tecnologías de la web semántica. Madrid, UOC. ISBN: 978-84-9788-474-7