

Acuerdo de Bibliotecas Universitarias de Córdoba

Seminario

27 y 28 de septiembre de 2012

Web semántica, Web 3.0 y entornos Cloud Computing, nuevos horizontes para bibliotecarios, documentalistas y archivistas
Guía de temas a desarrollar



Mela Bosch

melabosch@hotmail.com

Acuerdo de Bibliotecas Universitarias de Córdoba

Seminario

27 y 28 de septiembre de 2012



Octavo encuentro: Bibliotecas y web semántica

Síntesis de conceptos, estado actual, metodologías, tendencias en desarrollos de Web Semántica y herramientas en uso

Síntesis: Los sistemas informáticos dan preponderancia a la sintaxis, pero....la semántica siempre está

Sistemas informáticos → expresan el contenido través de la **forma con:**

✓ **estructuras de datos**

✓ **metainformación,
los metadatos**



Síntesis

W3C:

La Web Semántica es una Web **extendida**, dotada de mayor **significado** a través de **datos mejor definidos**.

Esto permite *razonar* los sistemas informáticos, realizar *inferencias* y tomar *decisiones*.

Se pueden obtener **soluciones a problemas habituales en la búsqueda de información**

(Traducción propia, fuente: <http://www.w3.org/RDF/FAQ/>)

Síntesis

¿Como implementa?

XML: lenguaje de marcado que expresa:

- ✓ ***RDF***: formato básico de metadatos
- ✓ ***RDF/S*** : lenguaje básico de ontologías usando ***RDF***
- ✓ ***OWL***: lenguaje común para expresar ontologías

Agentes de usuario: para implementar servicios web semánticos

Síntesis ¿Como se concreta?

Se articula una **infraestructura** de servidores y sitios web que implementan agentes de usuario que llaman a aplicaciones necesarias para que:

- ✓ navegadores interpretan metadatos de marcado semántico
- ✓ interpretan ontologías asociadas



Síntesis: Metodologías para tratar el conocimiento en los sistemas

Procedimental → el conocimiento está integrado en el programa

Ventajas: *gran especificidad: algoritmos adaptados a cada caso*

Desventajas: *poca versatilidad dificultad para modificar*

Declarativa → La representación del conocimiento es independiente del proceso computacional

Ventajas : *flexible y con fuerte base lógica*

Desventajas: *alto nivel de abstracción, dificultad para mantener una lógica consistente*

Web Semántica: Representación declarativa utilizando metadata

La transición entre las diferentes Web

Síntesis:

Una sintaxis de datos sobre datos para dar semántica

Para que los servidores y navegadores pueden mostrar información al usuario basada en su contenido y no solo en su estructura.



Tiene como fundamento:

✓ Una Lógica descriptiva y una Lógica de objetos (veremos enseguida)

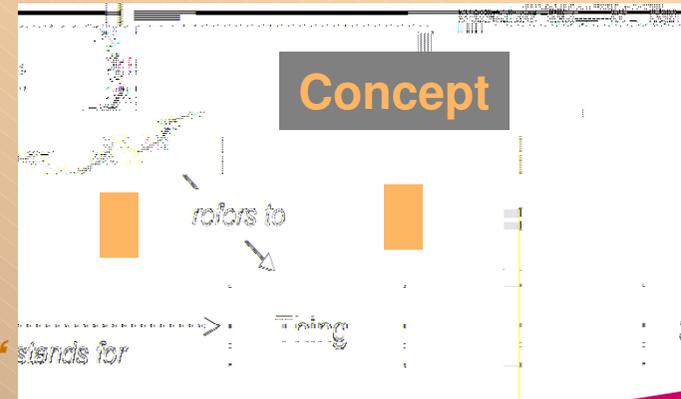
Una sola herramienta:
la anotación semántica:

Formas de etiquetado que describen el tipo de contenido que encapsulan

Síntesis. Orientación a objetos y ontologías

Aspectos en común

- ✓ Proceso iterativo
- ✓ Definición de conceptos del dominio (classes)
- ✓ Ordenamiento de los conceptos en jerarquías de clases
- ✓ Definición de los atributos y propiedades de las clases y sus valores



"Jaguar" stands for



Abstract Class: animal

Abstract Class: Car

Class: feline

Subclass: Jaguar, luxe

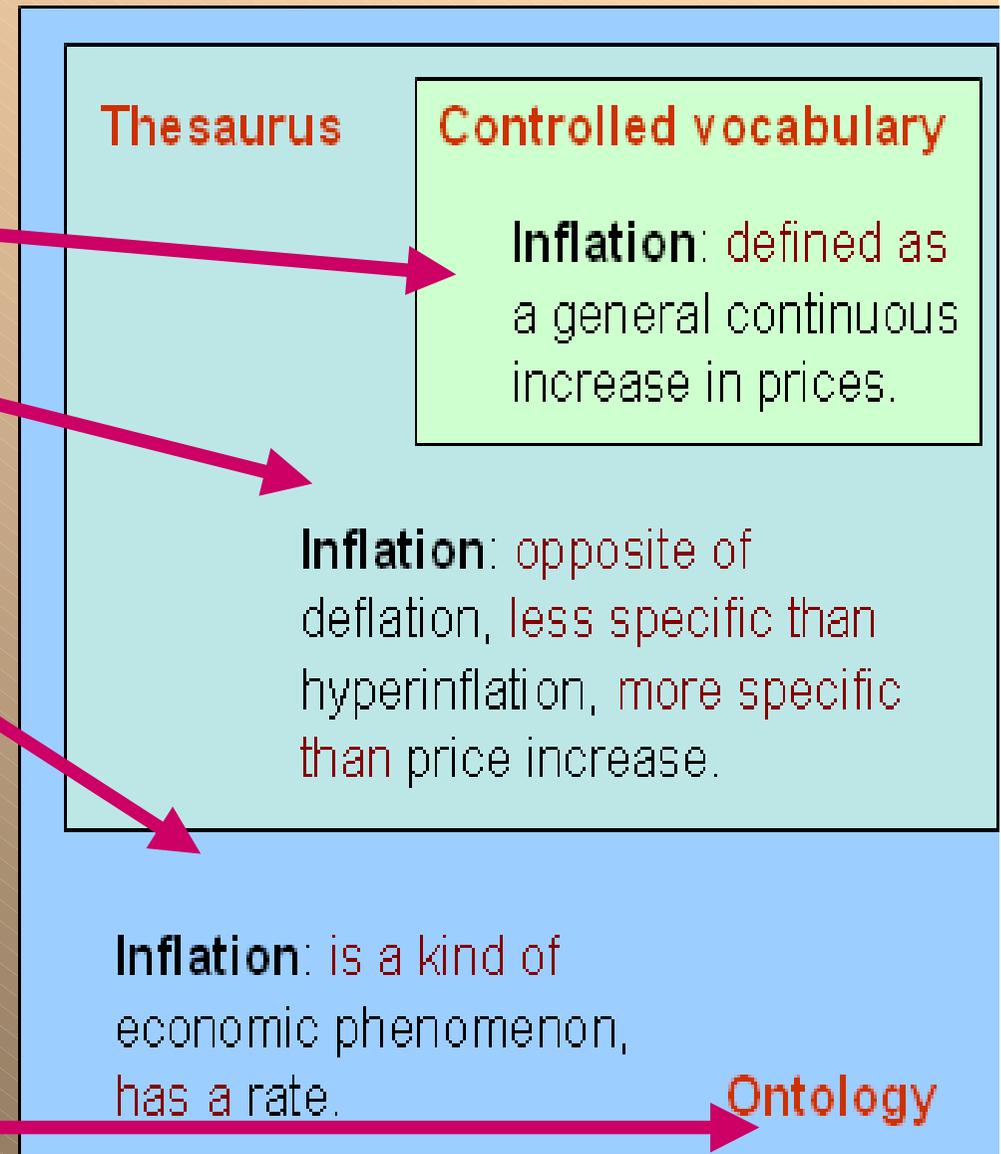
subclass: Jaguar, spots

Different: attributes, properties and values

Síntesis: Lenguajes documentales: toman partes del lenguaje natural para representar un dominio, se pueden usar como base para la primera estructura de clases

- ✓ **conceptos**
- ✓ **Partes de conceptos**
- ✓ **Como se relacionan con otros**
- ✓ **Puede apuntar a otros conceptos**

Puede direccionar a una clase de una ontología se utilizan sus principios para determinar categorías de clases



Síntesis Web 3.0

web

2.0

to

3.0

Web 2.0

+ dispositivos móviles

+ Web Semántica

+ Cloud computing

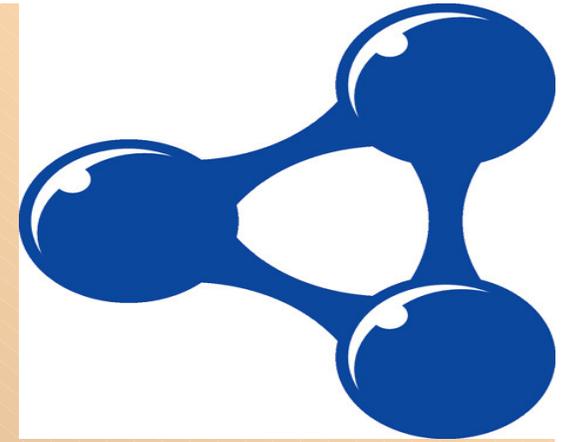


Síntesis: Tecnologías de Web Semántica usando lenguajes de marcado

XML: doble carácter

- ✓ **Metalinguaje de descripción que permite crear etiquetas propias y ocultas**
- ✓ **Lenguaje de especificación de puntos de ejecución de porciones de software y programas completos que pueden hacer uso etiquetas en maneras sofisticadas.**

Síntesis: Tecnologías de Web Semántica usan lenguajes de marcado



Resource Description Framework (RDF)

Es una EXTENSION de XML

Tripletas con verbo, sujeto y predicado: Cada sujeto y cada predicado son identificados por el Universal Resource Identifier (URI), como un enlace en cualquier página web

Verbos son identificados también por URIs, que permite a cualquiera definir un concepto nuevo, un verbo nuevo, simplemente definiendo un URI para ellos en algún lugar de la Web

RDF está siendo desarrollado y promovido por el Consorcio 3WC y varias ISO lo están normalizando

RDF/Schema: permite expresar un conjunto de clases con sus propiedades. Es algo así como una ontología básica

**Grafo de
declaración
RDF**

**Predicado
propiedad**

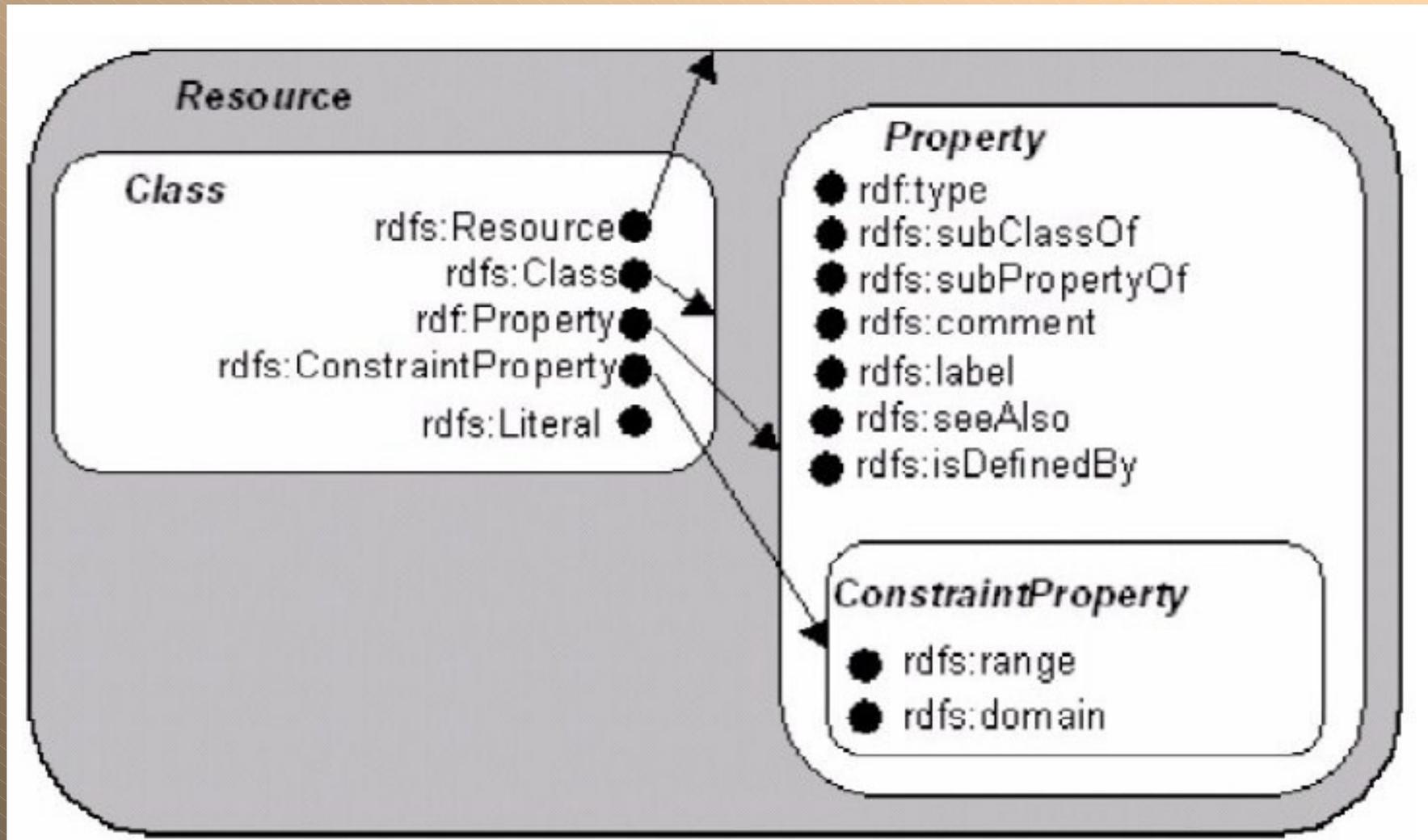
**Sujeto
Recurso
lugares, personas,
objetos
Identificados con URI**

**Objeto, valor
asigando a la
propiedad, puede
ser otro URI**

**Sentencia
(declaración)**

**Reificación:
sentencias
sobre
sentencias**

Ejemplo de una jerarquía de clases en RDF Schema



RDF→ No evita superposiciones, dos recursos pueden usar diferentes identificadores para el mismo concepto.

- Para evitarlo → **Ontologías:**

Origen en la filosofía→ En los sistemas informáticos: especificación de una conceptualización

- Conceptos son definidos por sus atributos y comportamiento
- Reglas que permiten que la ontología deduzca, o por lo menos proponga, a que clase o categoría puede pertenecer cada nuevo concepto que ingresa

La ontología: taxonomía + conjunto de reglas de inferencia.

La taxonomía define clases de objetos y relaciones entre ellos.

Retomando la definición lenguajes naturles, artificiales y controlados, la ontologías se sitúan dentro de los lenguajes controlados

**Síntesis: para llegar Web semántica:
una sucesión de pasos para avanzar:
Desarrollo de servicios en XML →
metadatos semánticos → Esquemas
→ ontologías**

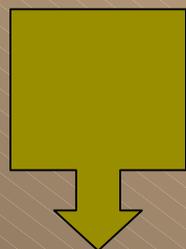
**Si es tan
complicado →
¿por qué
hacerlo?**



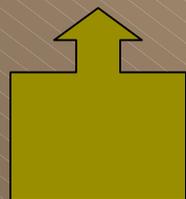
¿Cuál es la situación actual?



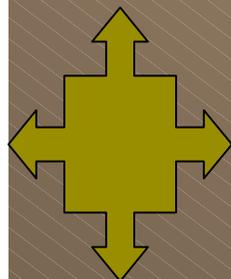
Enfoques metodológicos



Top down: ontologías, se definen los conceptos más generales y luego se especializan



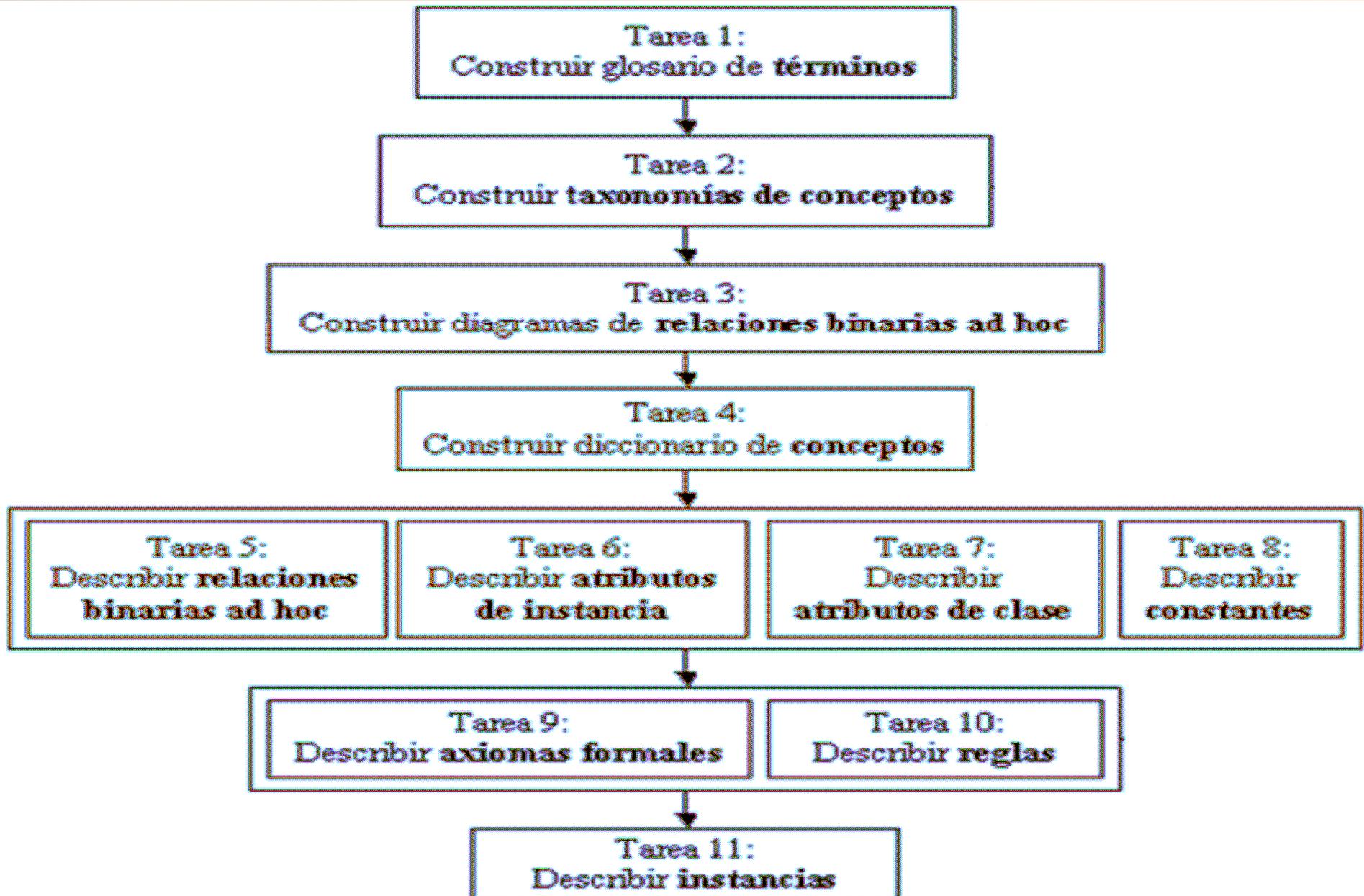
Bottom up: esquemas de metadatos, se definen los conceptos más específicos y luego se generaliza



Combinación o híbrido: se definen los conceptos más importante a usando a veces estructuras de lenguajes documentales o esquemas de metadatos y se generaliza o especializa cada conceptos según el dominio de aplicación

¿Cuál es la situación actual?

Top down ontologías



¿Cuál es la situación actual?

Top down: ontologías

**Sistemas ex novo, content intensive,
dominios acotados, ámbito privado,
corporaciones**

Sistemas Financieros



Medios de comunicación



Diseño y producción industrial



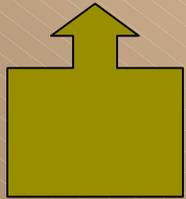
E- commerce



Médicos y farmacéuticos

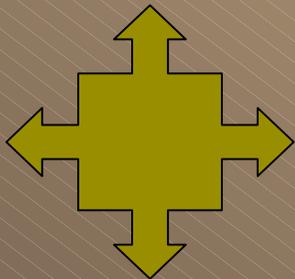


¿Cuál es la situación actual?



Bottom up

Datos enlazados (Linked data) predominantemente recursos de servicios web navegables, web social, medios, e-government, e-learning



Híbrido

Data warehousing con base semántica: datos estructurados en organizaciones, reutilización de datos de bases relacionales

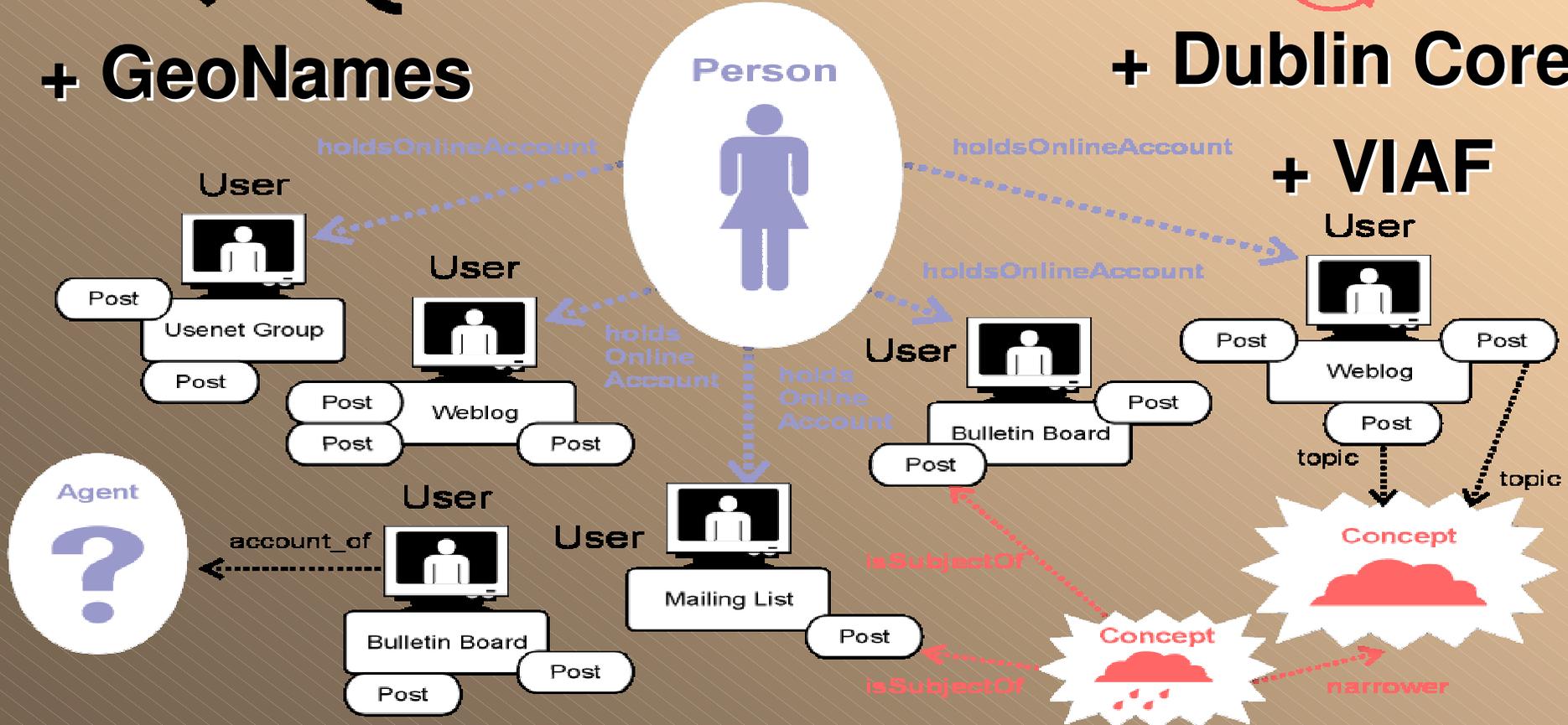
¿Cuál es la situación actual?



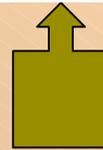
Bottom up

Datos enlazados (Linked data) Se convierten a RDF conjuntos de datos combinando metadatos semánticos

SIOC + FOAF + SKOS + Dublin Core + GeoNames + VIAF



¿Cuál es la situación actual?



Bottom up

Datos enlazados (Linked data) Se convierten a RDF conjuntos de datos combinando metadatos semánticos

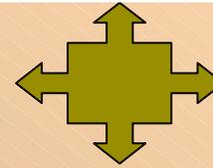
Algunos de ellos: FOAF (Friend of a friend) Es un ontología que sirve para identificar personas dentro de las redes sociales para hacer una red de ellas en forma transversal a los recursos: http://wiki.foaf-project.org/w/Main_Page

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.

Dublin Core: 15 metadatos para obras, VIAF Virtual International Authority File , GEOREF nombres geográficos, DBPedia (remite a Wikipedia... muchos más en desarrollo

¿Cuál es la situación actual?



Híbrido

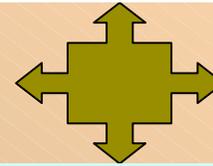
Data warehousing semántico

Se basan en el desarrollo de una “**capa semántica**” que separa los datos operativos y reordena transversalmente los mismos. En algunos sistemas como SAP se la llama “universo”

Sirve para:

- ✓ Ocultar aspectos técnicos irrelevantes para el usuario, como las tablas agregadas, los alias, los loops, los particionamientos, el propio modelo de datos, o los tipos de relación.
- ✓ Reorganiza de acuerdo con el uso y no con el proceso.
- ✓ Se pueden realizar búsqueda flexibles a través de la integración de un motor de búsqueda semántico

¿Cuál es la situación actual?



Híbrido

Data warehousing semántico

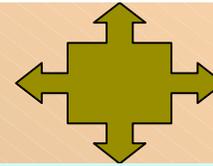
DW tienen como objetivo integrar datos de diferentes bases de datos con diferente análisis de esos datos

'80 se centraron algoritmos automáticos de **clustering** y estadísticos. Paralelamente → Web Semántico: el razonamiento automático partir de metadata

DWS → Extraction/Transformation/Load , ETL+ integración de metadata.

Ontologías orientadas a Dws (ontology-driven Dws) es para el procesamiento semántico automático, para la consulta y su visualización para el usuario final.

¿Cuál es la situación actual?



Híbrido

Data warehousing semántico

Descripción conceptual del contenido DW
independiente de la plataforma utilizando Web
Ontology Language (OWL)

Muchos pasos previos y complementarios

- ✓ análisis de dominio para generar las jerarquías de clases orientadas al ETL.
- ✓ diseño de ontologías que más se adapten para la integración de fuentes, para el macheo de esquemas.
- ✓ análisis de datos orientados a la reformulación del query.
- ✓ Estudio de las interfaces de usuarios necesarias para la mejor explicación de los componentes

¿Cuál es la situación actual?

Herramientas más usadas

Anotación:

Annozilla, Annotea, (*attached metadata*) Smore, Amaya (*embedded metadata*)

- **Jena** – A Semantic Web Framework for Java

- Entorno de programación para RDF, RDFS, OWL, SPARQL

- <http://jena.sourceforge.net/>

- **SESAME**

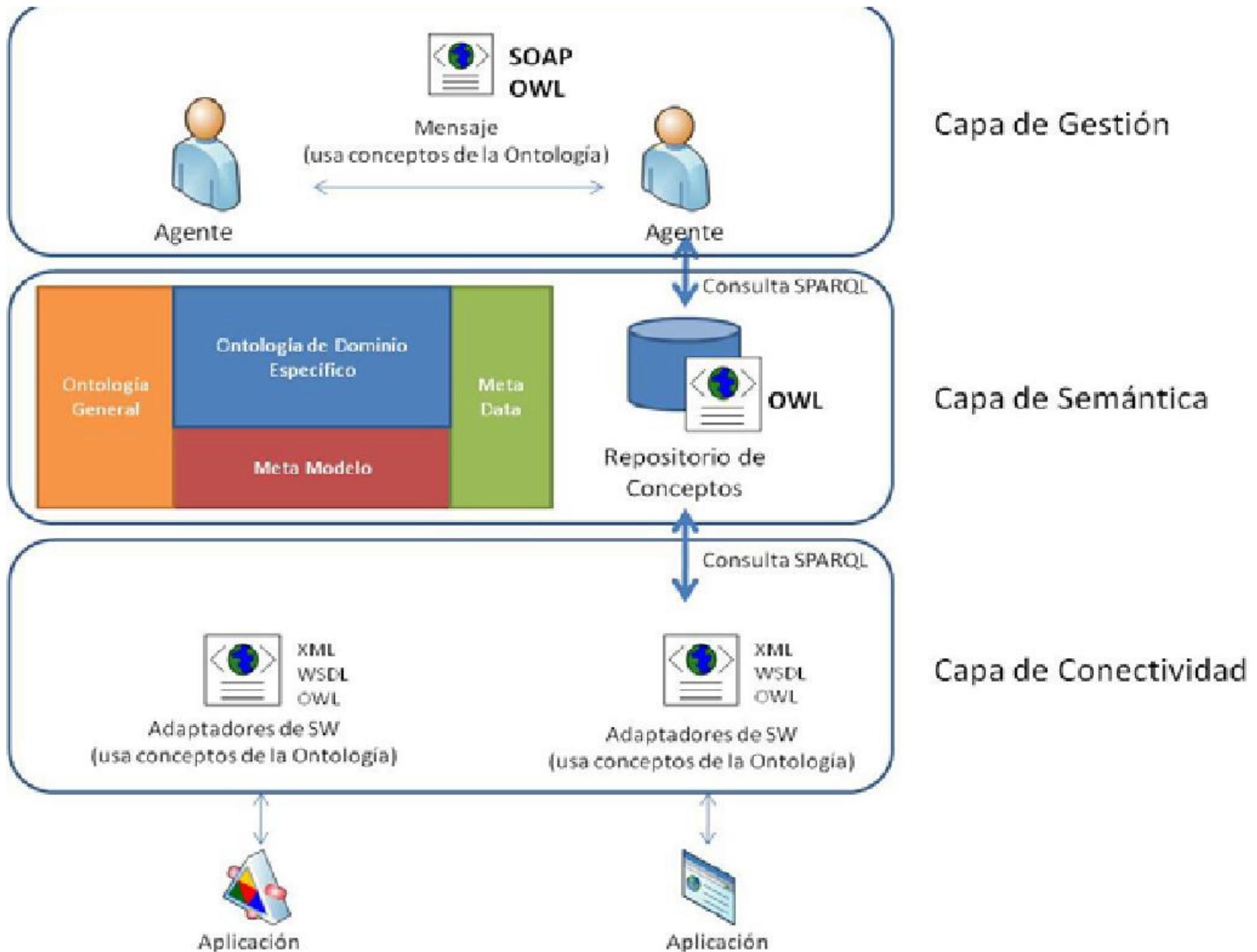
- Entorno de programación para RDF y RDFS

- <http://www.openrdf.org/>

- **Protégé**: editor de ontologías

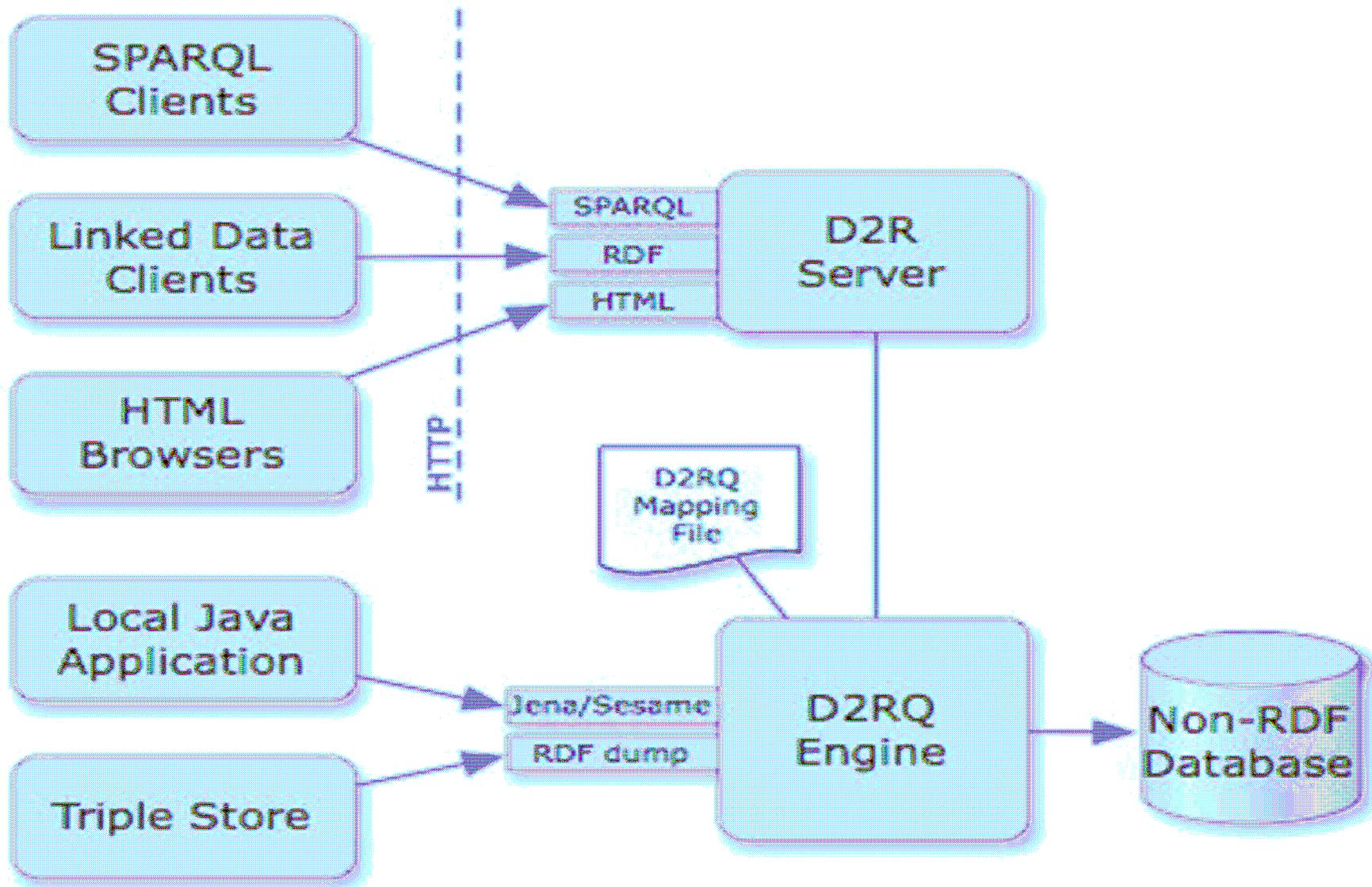
- RDF, RDFS, OWL

- Panel SPARQL



¿Cuál es la situación actual?

Herramientas





Aplicaciones comerciales e institucionales Web 3.0 con Ws integrada

- <http://www.cambridgesemantics.com/semantic-university/example-semantic-web-applications>
- <http://www.adlnet.org/>
- <http://openbiblio.net/2012/02/02/linked-data-at-the-biblioteca-nacional-de-espana/>
- <https://w3.bcn.es/V42/altres/index-en.html>
- <http://www.realinstitutoelcano.org/wps/portal/rielcano/Mapa>
- <http://company.hakia.com/whatis.html>
- <http://www.evri.com/>
- <https://www.tripit.com/>
- <https://www.thesavingsbankcircleville.com/>

Bibliografía, referencias, redes, organizaciones

Grupo: <http://www.mail-archive.com/web-semantica-ayuda@es.tldp.org/info.html>

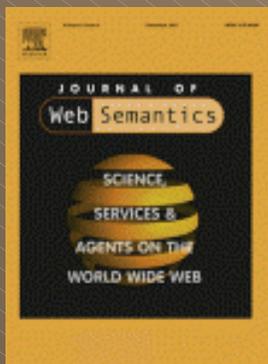
ISKO (International Society for Knowledge Organization) <http://www.isko.org/>

<http://www.gnoss.com/en/about-gnoss/2D>

Coyle, Karen. Linked Data tools: Connecting on the Web. Lbrary Technology reports, 2012.

Corcho, Oscar; Fernández-López, Mariano; Gómez-Pérez, Asunción. Ontological Engineering: Principles, Methods, Tools and Languages. Springer, 2006

Moreda, Paloma; Llorens, Hector; Saquete, Estela; Palomar, Manuel. Combining semantic information in question answering systems Original Research Article Information Processing & Management, Volume 47, Issue 6, November 2011, Pages 870-885



Journal of Web Semantics. Science, Services and Agents on the World Wide Web. Editors-in-Chief: T. Finin, S. Staab, I. Horrocks