

Tecnologías XML y Web Semántica



Departamento de Informática
Universidad de Oviedo

Fundamentos de la Web Semántica

Justificación
Esquema General
Principales Vocabularios



Departamento de Informática
Universidad de Oviedo



Reflexiones sobre la WWW

Web actual = mayor almacén de información recopilado por personas humanas

Grandes cantidades de información sobre cualquier asunto

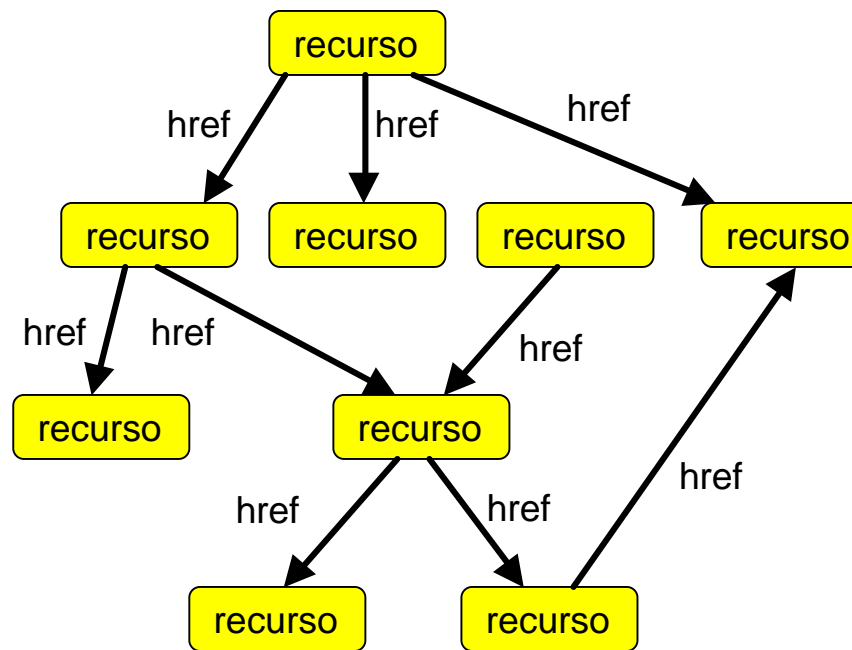
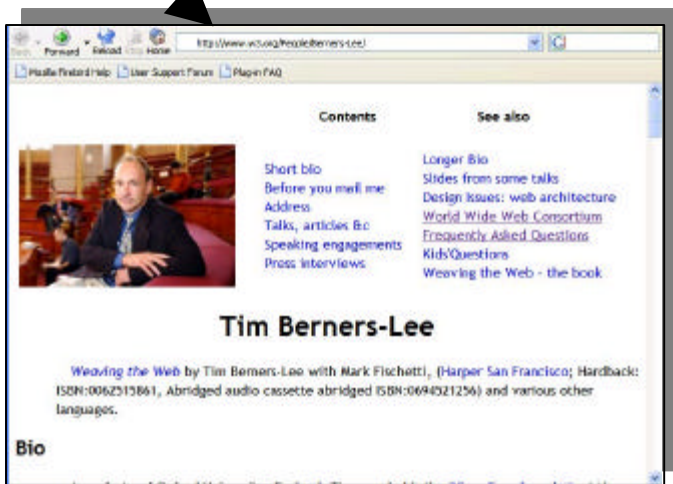
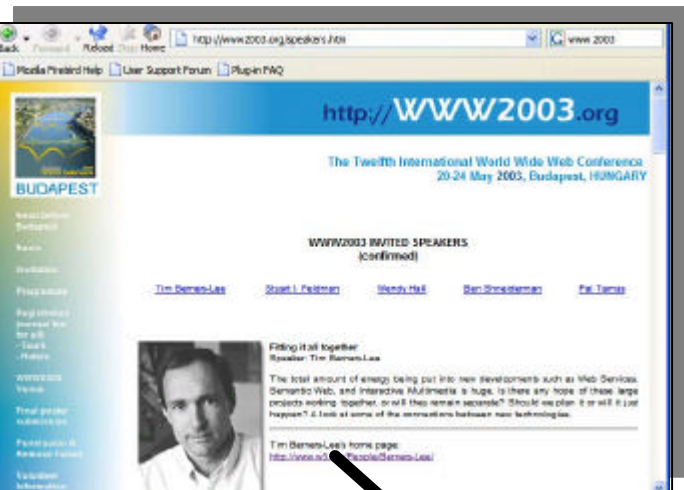
Acceso casi instantáneo desde cualquier lugar con conexión a Internet

Sistema no centralizado \Rightarrow Cualquier persona puede añadir más información



Reflexiones sobre la Web

Ahora estamos en la **Web Sintáctica**





La Web Sintáctica

Características actuales de la Web sintáctica

Biblioteca Digital con hipertexto

Enorme biblioteca con documentos (llamados *páginas Web*) conectados entre sí mediante enlaces

Una base de datos (o plataforma común de aplicaciones)

Un portal común de aplicaciones accesibles a través de páginas Web y que muestran sus resultados como páginas Web

Una plataforma para multimedia

Una nueva forma de transmitir programas de radio, TV y vídeos

Un esquema de nombres

Identidad única para los documentos

Los ordenadores realizan la presentación visual (tarea fácil) y las personas navegan e interpretan el contenido (tarea difícil)

¿Sería posible que los ordenadores hiciesen algo más?



Tareas difíciles en la Web sintáctica

Buscar fotos del profesor de este curso (Labra)



Buscar fotos del Director del Depto. donde trabaja el profesor de este curso....



Tareas difíciles en la Web Sintáctica

Buscar información sobre la Universidad de **Beihang** en China...

The screenshot shows a Mozilla browser window displaying the website of Beihang University (北京航空航天大学). The browser's address bar shows the URL <http://www.buaa.edu.cn/>. The website header features the university's logo and name in Chinese and English, along with the text "AERONAUTICS & ASTRONAUTICS". The main content area is divided into several sections:

- Left Sidebar:** A vertical menu with links for "学校概况" (School Overview), "组织机构" (Organizational Structure), "院系设置" (Departmental Settings), "科学研究" (Scientific Research), "人才培养" (Talent Cultivation), "招生就业" (Admission and Employment), "合作交流" (Cooperation and Exchange), "师资队伍" (Faculty and Staff), "招聘指南" (Recruitment Guide), "校办产业" (University-owned Enterprises), "校园文化" (Campus Culture), and "校友总会" (Alumni Association).
- News Section (新闻 News):** A list of recent news items, including:
 - 动力学院为廊坊学生指导科技制作 [2004-03-16]
 - 北京市教委、市卫成区司令部来校检查评估军训工作 [2004-03-16]
 - 后勤集团举办庆“三八”座谈会 [2004-03-16]
 - 2004年能源与动力工程学院师生足球联赛开赛 [2004-03-16]
 - 仪器科学与光电工程学院召开教职工大会总结布置学院工作 [2004-03-15]
 - 仪器科学与光电工程学院首届工会委员会成立 [2004-03-15]
 - 宇航学院研究生体验“拓展”活动 [2004-03-15]
 - 机械学院党支部开展文明礼貌月活动 [2004-03-15]
- Search Section (搜索):** A search box with a "搜索" (Search) button and a dropdown menu for "校内办公电话" (Campus Office Phone).
- Mail Login Section (邮箱登录):** Fields for "帐号:" (Account) and "密码:" (Password), with a "登录" (Login) button.
- Quick Links Section (快速链接 Quick link):** Links for "新闻中心" (News Center) and "校园地图" (Campus Map).

¿Hacer un viaje de la Universidad de **Beihang** a la Universidad de **Hainan**?



Tareas difíciles en la Web Sintáctica

Buscar un ave que utilice el oído para orientarse y que no sea un murciélago

y si os digo que en inglés es *barn owl* ?





Tareas difíciles en la Web Sintáctica

Búsquedas complejas

Localizar información en almacenes de datos

Búsqueda de viajes

Comparar Precios de productos

Encontrar y utilizar “*servicios web*”

Delegar tareas complejas a agentes de la Web

Organizar un viaje en algún lugar con playa no demasiado caro en el que hablen inglés

Buscar y comparar noticias que hablen de las últimas elecciones

Encargar una comida en el restaurante que no tenga demasiadas calorías



El problema de la Web Sintáctica



Las marcas de HTML incluyen información sobre la visualización (tipo de letra, color, etc.)

El significado es accesible a las personas pero no es (fácilmente) accesible a los ordenadores...

```
<html><head><title>Pizzeria Al Capone</title></head>
<body bgcolor="blue" text="yellow">
<h1>Pizzas del Restaurante Al Capone</h1>
<table>
<caption>Tipos de Pizzas</caption>
<tr>
<td>Barbacoa</td><td>Salsa barbacoa, mozzarella, Pollo
Bacon, Ternera </td><td>8&euro;</td>
</tr>
...
</body>
</html>
```



El problema de la Web Sintáctica

La información que vemos nosotros...

Pizzas del Restaurante Al Capone

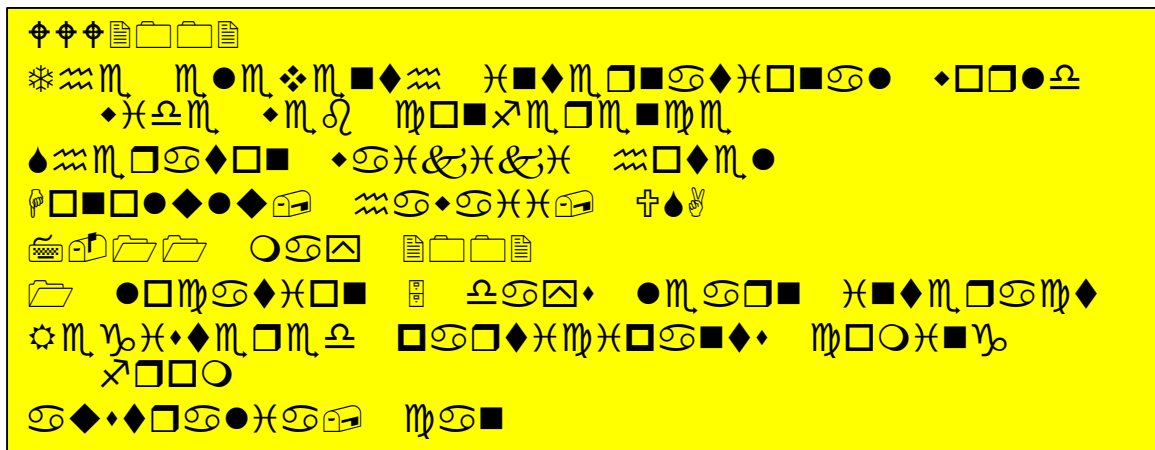
Tipos de Pizzas

Barbacoa

Salsa barbacoa, mozzarella, Pollo, Bacon,
Ternera 8€

...

Lo que ve la máquina...





XML como posible solución?

Incluir etiquetas con un significado determinado...

```
<título>Pizzas del Restaurante Al Capone</título>  
<pizza>  
  <nombre>Barbacoa</nombre><ingredientes>Salsa barbacoa,  
    mozzarella, Pollo, Bacon, Ternera  
  </ingredientes><precio>8€ </precio></pizza>  
...
```

La máquina vería...

```
<título>☼ ℓ ʏ ⋆ ⋄ ℓ □ ℓ ㊦ ㊦ ㊦ ⋆ ㊦ ⋄ </título>  
<pizza><nombre> □ ㊦ □ ⋄ ⋆ ℓ ⋆ □ ㊦ ⋄ ⋄  
  </nombre><ingredientes> ☒ ㊦ ③ ⑩ ㊦ ㊦ ㊦ ㊦ ㊦ ㊦ ㊦ ㊦ ㊦  
  ④ ⑥ ⑥ ⑥ ㊦ ㊦ ㊦ ㊦ ㊦ ㊦ ㊦ ㊦ ㊦ ㊦ ㊦ ㊦ ㊦ ㊦ ㊦ ㊦ ㊦  
  ☒ ㊦ ㊦ ㊦ ㊦ ㊦ ㊦ </ingredientes>  
<precio> ⋆ ℓ </precio></pizza>  
...
```



XML como posible solución?

Pero si vamos a otro contexto

```

<ave><nombreComún>Lechuza Común</nombreComún>
<nombre>Tyto alba</nombre>
<clasificación>Orden Estrigiformes, familia Titónidos.
  </clasificación>
<características>33-39 cm. 300-380 g. </características>
...

```

La máquina vería...

```

⤴⚡②⚡⇒⤴⑤⑥④⚡⑨⚡⇒⑥④□⑤⇒⤴⚡⚡⚡①⑥⚡
⤴⑥④□⑤⤴□⑤⑥④⚡⑨⚡⇒⑥④□⑤⇒
⤴⚡⚡④①③①⇒☒⑤①⑥
⚡③⚡⚡⤴□⚡⚡④①③①⚡⇒⤴□⚡②⚡⇒
⤴⚡③⚡⑩①⚡①⚡⚡⚡①★⑤⇒×⑨⚡⚡⑤
⤴⑩①⑨①⚡①⚡⑥⑨④⚡⑩☐⚡⚡⚡④①③①⚡
☒①①★⑤①⚡⑥⑩☐⚡⤴□⚡⚡③⚡⑩①⚡①⚡

```

El problema es que las etiquetas no tienen un significado compartido



Posibilidad: **Acuerdo global** sobre el significado de las anotaciones

Ejemplo: Dublin Core

Especifica etiquetas globales de anotación de recursos. Por ejemplo: dc:creator

Problemas:

Inflexibilidad

Limita el número de cosas que pueden expresarse

Otra posibilidad: Utilización de **Ontologías**

Definen el significado de las anotaciones

Los términos nuevos pueden formarse a partir de otros anteriores

El significado se define formalmente

Pueden especificarse relaciones entre términos de varias ontologías...



La Web semántica (**Tim Berners-Lee**) pretende desarrollar lenguajes que faciliten la inclusión en la Web de contenido legible por las máquinas

Características de la Web que deben tenerse en cuenta...

No centralizada: problemas para garantizar integridad de la información)

Información Dinámica: puede cambiar la información e incluso el conocimiento sobre esa información

Mucha información: El sistema no puede pretender acaparar toda la información

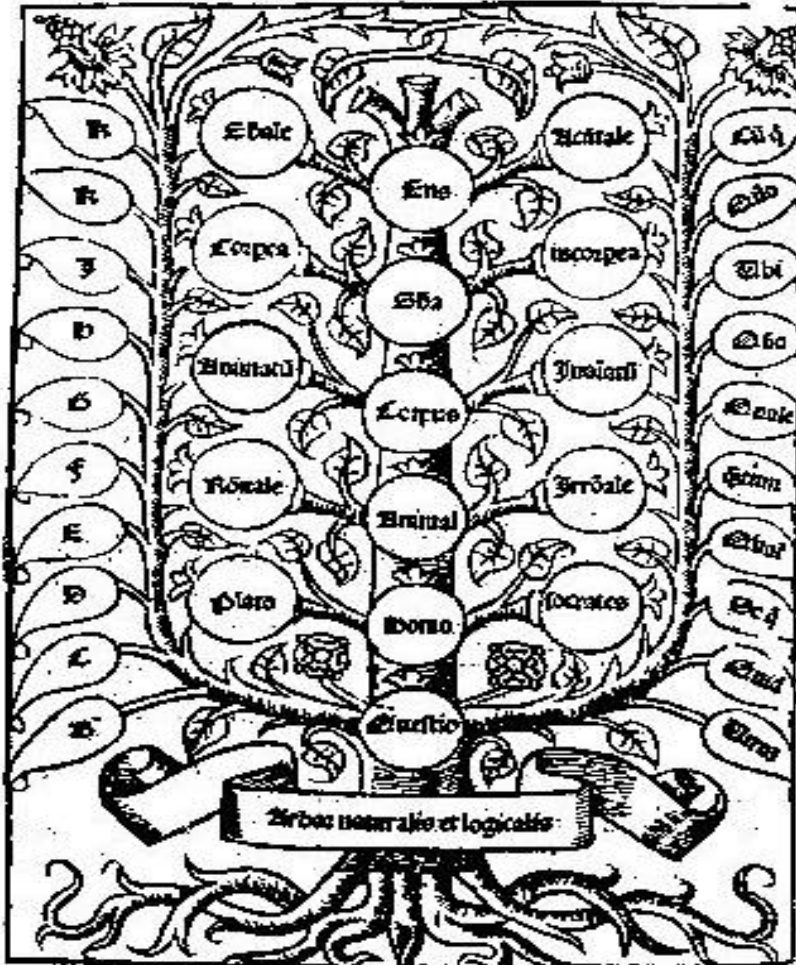
Es abierta: Muchos sistemas anteriores usaban la *Closed World Assumption*



Web Semántica

Antecedentes

Sistemas de representación del conocimiento



Árbol de la naturaleza y de la lógica
Ramón Llull (1235-1316)



Web Semántica

Antecedentes

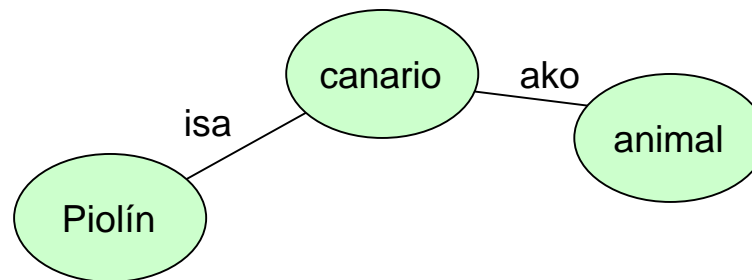
Sistemas de representación del conocimiento

Redes semánticas (R. Quillian, 1968)

Redes de conceptos enlazados entre sí.

is-a = pertenece

ako (a kind of) = incluido





Web Semántica

Antecedentes

Lógica = Estudio de los razonamientos

Lenguaje de la lógica permite representar conocimiento e inferencias

Niveles:

Lógica proposicional

Lógica de predicados de primer orden

Órdenes superiores

$$\frac{\forall x(\text{canario}(x) \rightarrow \text{animal}(x)) \quad \text{canario}(\text{piolín})}{\text{animal}(\text{piolín})}$$

Sistemas de inferencia:

Consistencia: Todo lo que se deduce es correcto

Compleitud: Todo lo que es correcto es deducible

Decidibilidad: El algoritmo de deducción finaliza

Tratabilidad: El algoritmo termina en un tiempo razonable

Problema P - NP

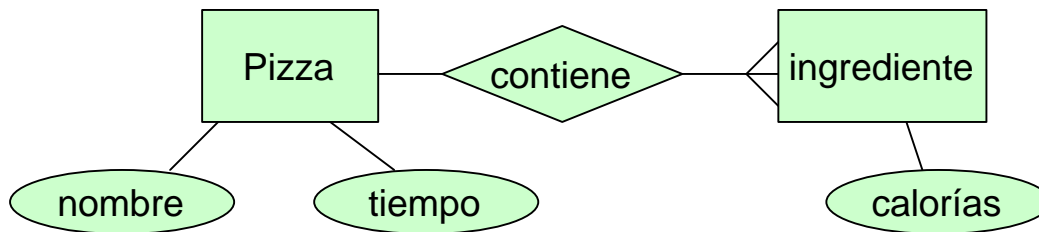


Web Semántica

Antecedentes

Diagramas entidad-relación

Utilizados en la modelización de Bases de Datos



Otras notaciones de modelado en Ingeniería del Software

Ejemplo: UML



Web Semántica

Antecedentes

Mapas de tópicos

(<http://www.topicmaps.org/>)

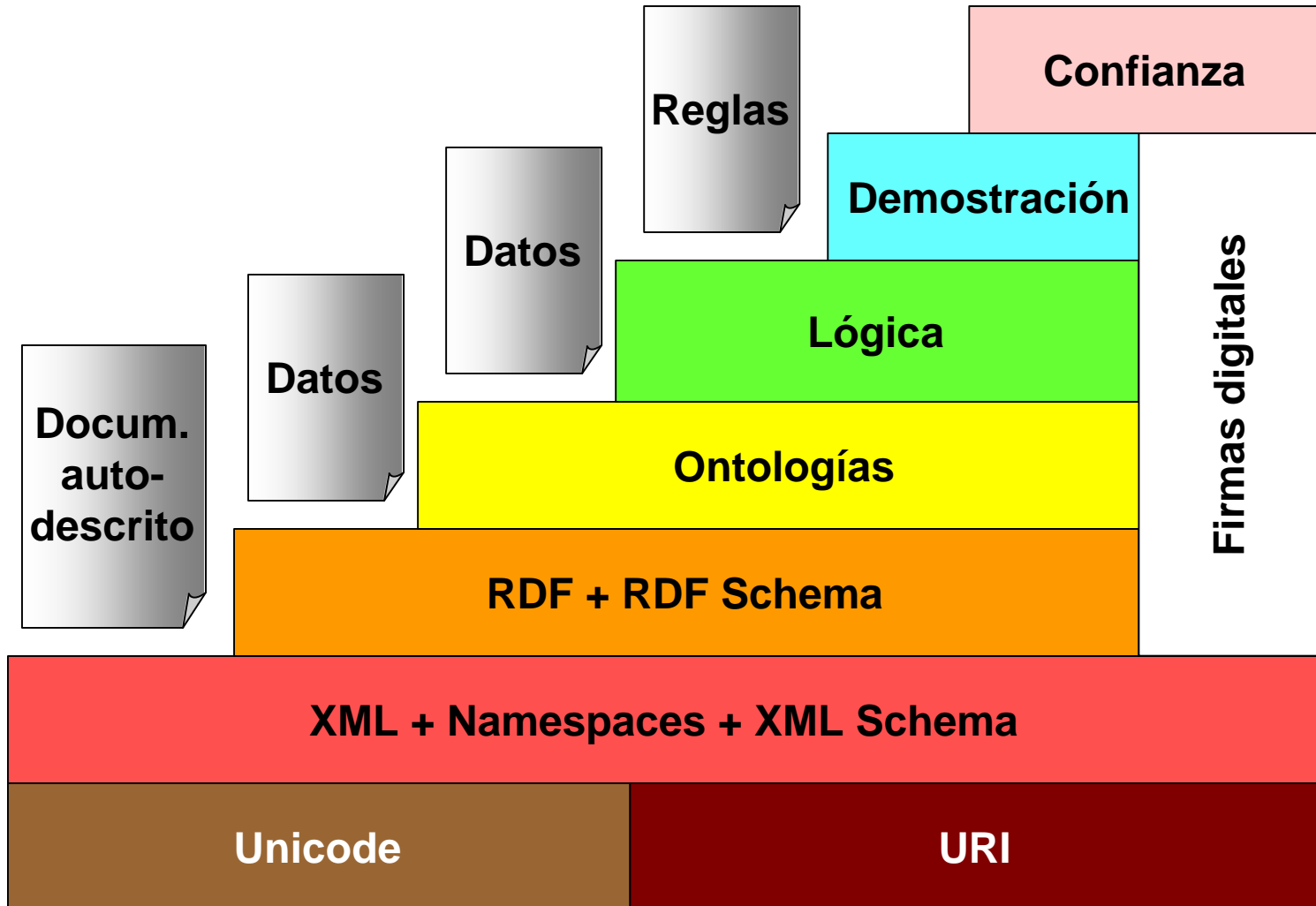
Estándar de definición de índices

XTM es un vocabulario para mapas de tópicos basado en XML

```
<topic id="pizzas"/> ...  
  <occurrence>  
    <instanceOf>  
      <topicRef xlink:href="#barbacoa"/>  
    </instanceOf>  
    <scope>  
      <topicRef xlink:href="#pizza"/>  
    </scope>  
    <resourceRef xlink:href="barbacoa.jpg"/>  
  </occurrence>  
  ...  
</topic>
```



Web Semántica





RDF (Resource Description Framework)

Origen: PICS (Platform for Internet Content Selection)

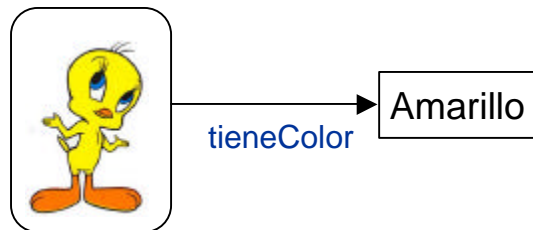
Descripción de contenidos en Internet (meta-información)

RDF define un modelo basado en tripletas

Dicho modelo tiene una sintaxis XML asociada

Las tripletas son de la forma:

Sujeto --- Predicado --- Objeto



Incluye algunas propiedades predefinidas (type)

Mecanismos de *reificación*



Permite definir:

Clases y propiedades

Ej. *"juan" es un individuo*

"profesores" es una clase

"daClaseDe" es una propiedad

Jerarquías de clases y herencia

Ej. La clase *"profesores" está incluida en "personas"*

Jerarquías de propiedades

Ej. *"esPadre" es una subpropiedad de "esProgenitor"*



OWL (Web Ontology Language)

Permite definir Ontologías

Se basa en lógica descriptiva (*description logics*)

Aumenta la expresividad de RDF(S)

Compromiso entre expresividad/computabilidad (3 niveles)

Se pueden indicar:

Restricciones de rango (las vacas sólo comen hierba)

Clases disjuntas (ej. hombres y mujeres)

Combinaciones booleanas (unión, intersección, etc.)

Personas=Hombres \cup Mujeres

Restricciones de Cardinalidad (una persona tiene 2 progenitores)

Características de propiedades (transitividad, unicidad, inversa, ...)



Lógica y Demostración

Orígenes = Representación del conocimiento

Lógica = Fundamento de Representación del Conocimiento

Existen técnicas de prueba para subconjuntos de lógica de primer orden

Propiedades: Consistencia, Completud, tratabilidad

RDF(S) y OWL son subconjuntos de lógica descriptiva

Otros formalismos: Cláusulas Horn

Lenguajes de reglas en desarrollo. Ej. RuleML

Propiedades:

Reglas Monótonas vs. No monótonas

Mundo cerrado vs. Abierto

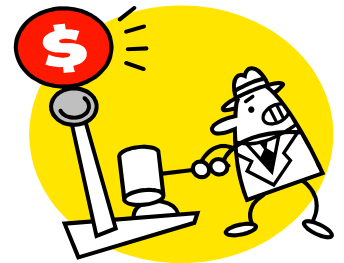


Ejercicio para mañana

Plantear una posible aplicación para la Web Semántica
Se puede consultar...

<http://challenge.semanticweb.org/>

Premio: 1000€ para viaje y 250€ para libros...





Selección de Enlaces



Página del consorcio: <http://www.w3c.org/RDF>

SemanticWeb: <http://www.semanticweb.org>

Directorio de Ontologías: <http://www.schemaweb.info>

Ontologías: <http://www.ontology.org>

Dublin Core: <http://www.dcmi.org>

Darpa Markup Language: <http://www.daml.org>

Open Directory Project: <http://www.dmoz.org>

OntoWeb: <http://www.ontoweb.org>

Topic Maps: <http://easytopicmaps.com>



Fin de la Presentación

