

# Seminario Web Semántica y Ontologías

Inteligencia Artificial  
5º Informática

IA curso 2000-2010

CCIA

Diciembre 2009

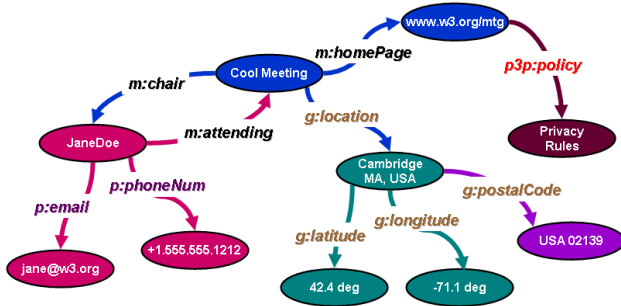
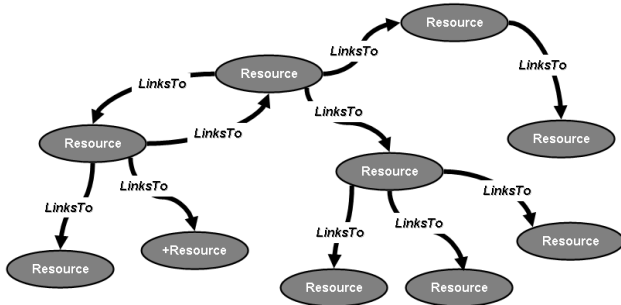
# Web Semántica vs. Web Actual

Evolución de la web actual en la que el contenido es procesable de forma automática

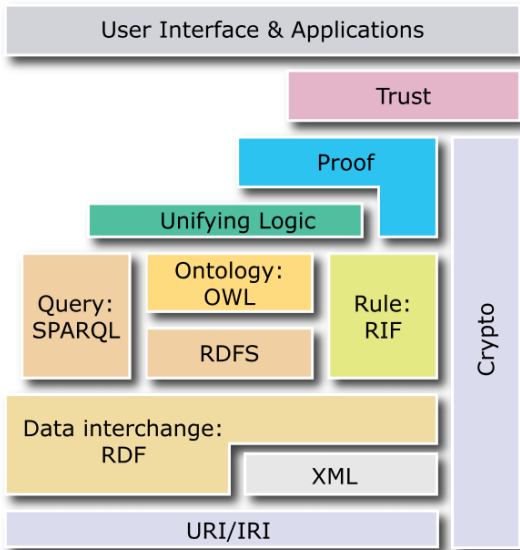
- Web actual representa la información utilizando documentos en lenguaje natural con poca estructura
  - ▶ Fácil comprensión por humanos (`html` sólo define presentación)
  - ▶ Difícil soportar procesado automático
- Alternativas
  - ▶ Máquinas más inteligentes → comprender el significado de la información que hay en la web
    - ★ Procesamiento de lenguaje natural, reconocimiento de imágenes, etc
  - ▶ Información más inteligente → representar la información de modo que sea sencilla de comprender a las máquinas
    - ★ Expresar contenidos en un formato procesable automáticamente.
    - ★ Uso de **metainformación** (metadatos = datos sobre los datos)

## Definición Web Semantica [Tim Berners-Lee(2002)]

- *"La Web Semántica es una extensión de la Web actual en la que a la información disponible se le otorga (anota o marca) una semántica bien definida."*
- *"Está basada en la idea de proporcionar en la Web datos definidos y enlazados, permitiendo que aplicaciones heterogéneas descubran, integren, razonen y utilicen la información presente en la web"*
  
- Web convencional: enlaces "tontos" entre documentos HTML
- Web Semántica: relaciones con *significado* entre conceptos



# Estándares de la Web Semántica



# Estándares básicos

**UNICODE** estándar que proporciona el medio por el cual codificar un texto en cualquier forma e idioma

**URI** *uniform resource identifier* cadena caracteres que identifica inequívocamente un recurso (servicio, página, documento, etc) físico o abstracto

- identifica el recurso, pero no tiene por que localizar su ubicación(URL)

**XML** meta-lenguaje extensible de etiquetas usado para el intercambio de datos en la web

- uso de etiquetas con significado intuitivo para humanos, pero no para las máquinas
- XML estandariza formato no significados
  - ▶ nombre de las etiquetas XML no ofrece semántica por si mismo

# Representación de conceptos y relaciones (RDF)

**RDF** (*Resource Definition Format*) estándar W3C para describir recursos (cualquier concepto que tenga una URI) en la web

- Formato común para describir información que pueda ser leída y entendida por una aplicación informática
- Permite representar conceptos y relaciones mediante un conjunto de tripletas.
  - ▶ **tripleta**: describe propiedades de un recurso identific. por una URI
    - ★ Recurso (o nodo): representa un documento, parte de él, una colección de documentos, un objeto, etc
    - ★ Propiedad: es siempre una URI (predefinida y con significado preestablecido)
  - ▶ cada tripleta combina  $\left\{ \begin{array}{l} \text{un recurso (SUJETO)} \\ \text{una propiedad (PREDICADO)} \\ \text{un valor para la propiedad (OBJETO)} \end{array} \right.$



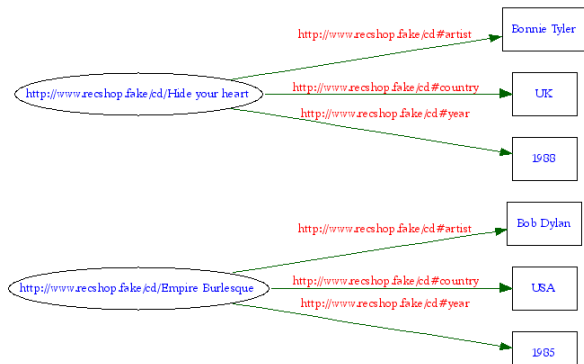
- RDF permite usar vocabularios semánticos definidos por expertos para describir recursos
  - ▶ *Dublin Core*: descripción de recursos digitales (páginas HTML, libros, etc) [<http://dublincore.org/>]
  - ▶ FOAF (*friend of a friend*): ontología para descripción de personas [<http://www.foaf-project.org/>]
- Representable en forma de documentos XML [serialización RDF/XML]
- Posibilidad de usar lenguajes de consulta sobre triples RDF
  - ▶ SPARQL (<http://www.w3.org/TR/rdf-sparql-query/>): sintaxis tipo SQL sobre bases de datos de triples RDF
  - ▶ DBPEDIA (<http://wiki.dbpedia.org/>): versión estructurada (tuplas RDF) de la Wikipedia

# Ejemplo RFD

```
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:cd="http://www.recshop.fake/cd#">
  <rdf:Description rdf:about="http://www.recshop.fake/cd/Empire Burlesque">
    <cd:artist>Bob Dylan</cd:artist>
    <cd:country>USA</cd:country>
    <cd:year>1985</cd:year>
  </rdf:Description>

  <rdf:Description rdf:about="http://www.recshop.fake/cd/Hide your heart">
    <cd:artist>Bonnie Tyler</cd:artist>
    <cd:country>UK</cd:country>
    <cd:year>1988</cd:year>
  </rdf:Description>
</rdf:RDF>
```

<http://www.w3.org/RDF/Validator/>



# Ontologías

**Ontología:** Conjunto de términos y relaciones entre ellos que describen un dominio de aplicación concreto

- Esquema conceptual de un dominio dado, que facilita la comunicación y compartición de información entre diferentes sistemas
  - ▶ Son quienes realmente aportan el significado
  - ▶ Definen  $\left\{ \begin{array}{l} \text{vocabulario para expresar conceptos, propiedades, relaciones} \\ \text{una jerarquía de clases para crear instancias} \end{array} \right.$
- Posibilidad de inferencias implícitas (herencia)
- Ofrece las bases para integrar el uso de herramientas de razonamiento y/o reglas para inferir nuevo conocimiento

Definición (*Studer, Benjamins en Knowledge Engineering: Principles and Methods*)

*"A formal explicit specification of a shared conceptualization"*

- *Conceptualization*: una forma de entender/describir un dominio
- *Shared*: compartida/consensuada por un grupo
- *Explicit*: descrita en un lenguaje
- *Formal*: comprensible para una máquina → (XML, RDF, RDFS, OWL)

# Ontologías (RDFS)

**RDFS** (*RDF Schema*) lenguaje extensible que proporciona los elementos básicos para para crear ontologías (vocabularios semánticos RDF)

- Permite definir clases, relaciones entre clases, restricciones sobre propiedades, etc
  - ▶ `rdfs:Class` declarar recursos como clases para otros recursos
  - ▶ `rdfs:subClassOf` definir jerarquías (relaciona clase con superclases)
  - ▶ `rdf:property` definir subconjunto de recursos RDF que son propiedades
  - ▶ `rdfs:subPropertyOf` definir jerarquías de propiedades
  - ▶ `rdfs:domain` dominio de una propiedad (clase de recursos que aparecen como sujetos en las tripletas de ese predicado)
  - ▶ `rdfs:range` rango de una propiedad (clase de recursos que aparecen como objetos en las tripletas de ese predicado)
- RDF Schema define el significado de los términos usados en los tripletas RDF

# Ejemplo RDFS

```
<rdfs:Class
  rdf:ID="Person"
  rdfs:label="Person"
  rdfs:comment="A person">
</rdfs:Class>

<rdf:Property
  rdf:about="http://xmlns.com/foaf/0.1/knows"
  rdfs:label="knows"
  rdfs:comment="A person known by this person.">
  <rdfs:domain rdf:resource="http://xmlns.com/foaf/0.1/Person"/>
  <rdfs:range rdf:resource="http://xmlns.com/foaf/0.1/Person"/>
</rdf:Property>

<rdf:Description rdf:ID="Person">
  <rdf:type rdf:resource=http://www.w3.org/2000/01/rdf-schema#Class>
  <rdfs:label>Person</rdfs:label>
  <rdfs:comment>A person</rdfs:comment>
</rdf:Description>
```

# Ontologías OWL

**OWL** (*ontology web language*) extiende RDFS para permitir la expresión de relaciones complejas entre clases RDFS, y mayor precisión en las restricciones de clases y de propiedades.

- Derivado de la fusión de los lenguajes de ontologías DAML y OIL
- Permite  $\left\{ \begin{array}{l} \text{expresar relaciones entre clases} \\ \text{expresar y restringir clases (rango, dominio)} \\ \text{expresar y restringir propiedades (cardinalidad)} \end{array} \right.$
- Tres variantes/sublenguajes (menor a mayor potencia expresiva)
  - ▶ *OWL-lite*: versión simplificada (representación de jerarquías simples)
  - ▶ *OWL-DL*: incluye constructores tomados de *Description Logics (DL)*  
Busca compromiso entre máxima expresividad y eficiencia computacional (sólo constructores decidibles de DL)
  - ▶ *OWL-full*: soporte completo de constructores DL
- Usos de OWL
  - ▶ Definición de Clases: *intersectionOf*, *unionOf*, *complementOf*
  - ▶ Características de Propiedades: *inverseOf*, *FunctionalProperty*, *InverseFunctionalProperty*, *TransitiveProperty*, *SymmetricProperty*
  - ▶ Igualdad/desigualdad: *equivalentClass*, *equivalentProperty*, *sameIndividualAs*, *differentFrom*, *allDifferent*
  - ▶ Restricciones de Cardinalidad: *minCardinality*, *maxCardinality*

# Temas a desarrollar

- Usos prácticos de la Web semántica y/o sus conceptos relacionados (ontologías, etc)
- Revisión de herramientas
  - ▶ editores de ontologías
  - ▶ lenguajes/sistemas de consulta sobre la Web semántica
- Metodologías para desarrollo de ontologías
- Descripción en profundidad/ejemplos de los estándares protocolos básicos (RDF, RDF Schema, OWL) o de otros relacionados (serialización RDF)
- Ontologías públicas, repositorios de ontologías
- etc,...

## Referencias

- W3C (*World Wide Web Consortium*): <http://www.w3.org/2001/sw/>
  - ▶ Especificaciones: Web Semántica, RDF, RDF Schema, OWL