

INTRODUCCIÓN A LA WEB SEMÁNTICA Y LINKED DATA



La Web actual

- Representa la información con:
 - Texto
 - Gráficos
 - Multimedia
- Publicación
 - Información
 - Maquetación
- Consumo humano



Un ejemplo



Rastreator.com

Lo rastrea torrrrrrr

Hacia la Web Semántica

- ¿Y si se quiere combinar los datos de la Web?
 - v.g. Información de viajes y hoteles
 - v.g. Búsquedas en varias bibliotecas digitales
- Los humanos la combinan fácilmente
 - Incluso si se usan terminologías diferentes
 - Pero lo hacen lentamente

Sin embargo...

- Las máquinas son ignorantes
 - No pueden hallar sentido en, por ejemplo, una imagen
 - Trazar analogías automáticamente es difícil
 - Difícil combinar información automáticamente
 - `<foo:creator>` es lo mismo que `<bar:author>`?

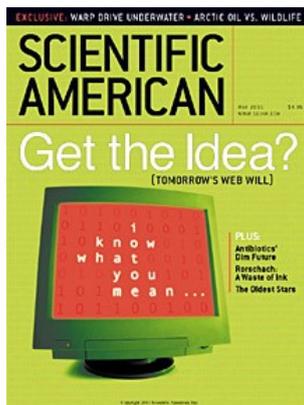
Ejemplos

- Sistemas de reserva de vuelos
 - Conoce tus preferencias y hábitos
 - Combina conocimiento local con servicios remotos
 - Disponibilidad de líneas aéreas
 - Calendario disponible
 - Diseño ad-hoc: difícil integrar nuevos servicios
- Integración de [bases de] datos
 - Datos de la web 'oculta'
- Redes sociales
 - ¿Cuántas veces has tenido que volver a añadir todos tus contactos?
- Bibliotecas digitales

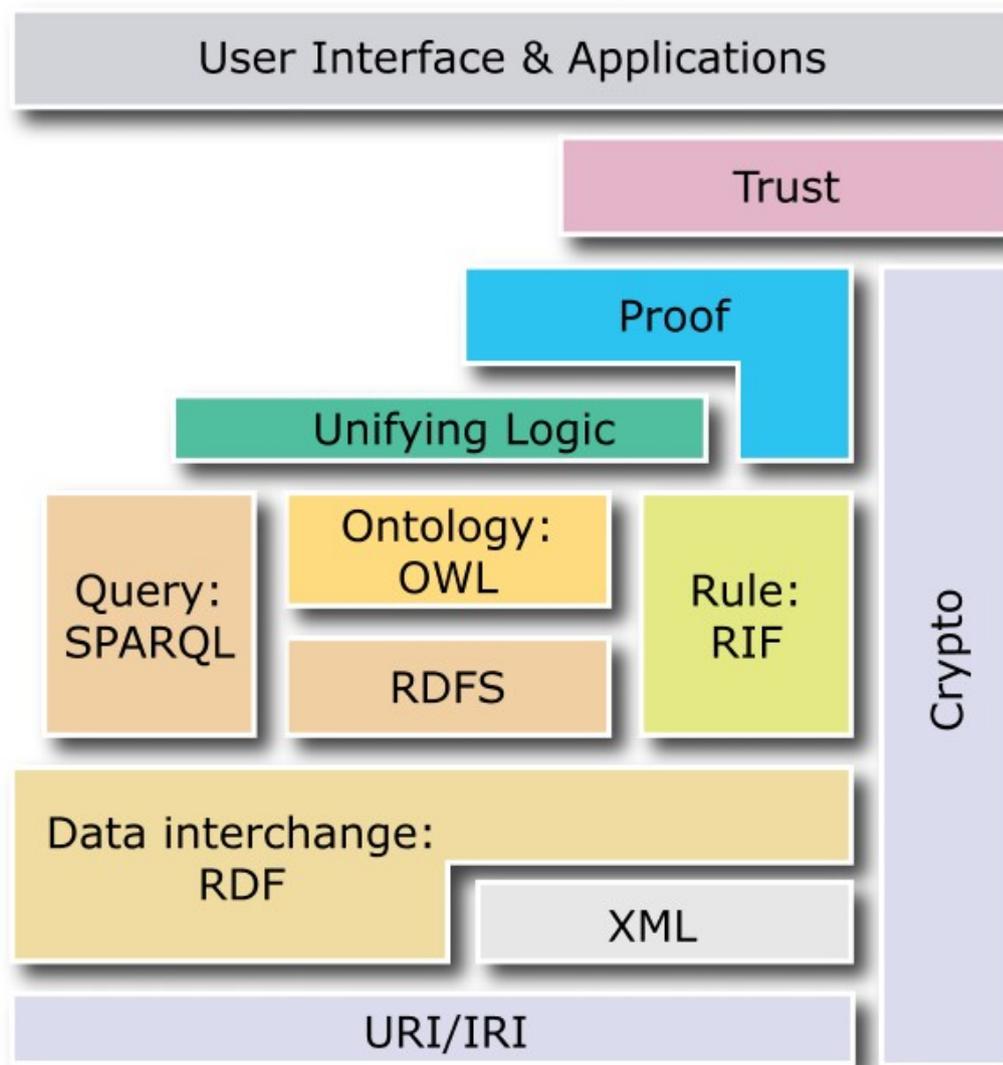
¿Qué hace falta?

- Datos preparados para su consumo por las máquinas
- Poder combinar y cruzar datos a escala Web
- Algunos datos describen otros datos (v.g. Metadatos en bibliotecas digitales)
- Las máquinas deberían poder *razonar* con todos esos datos

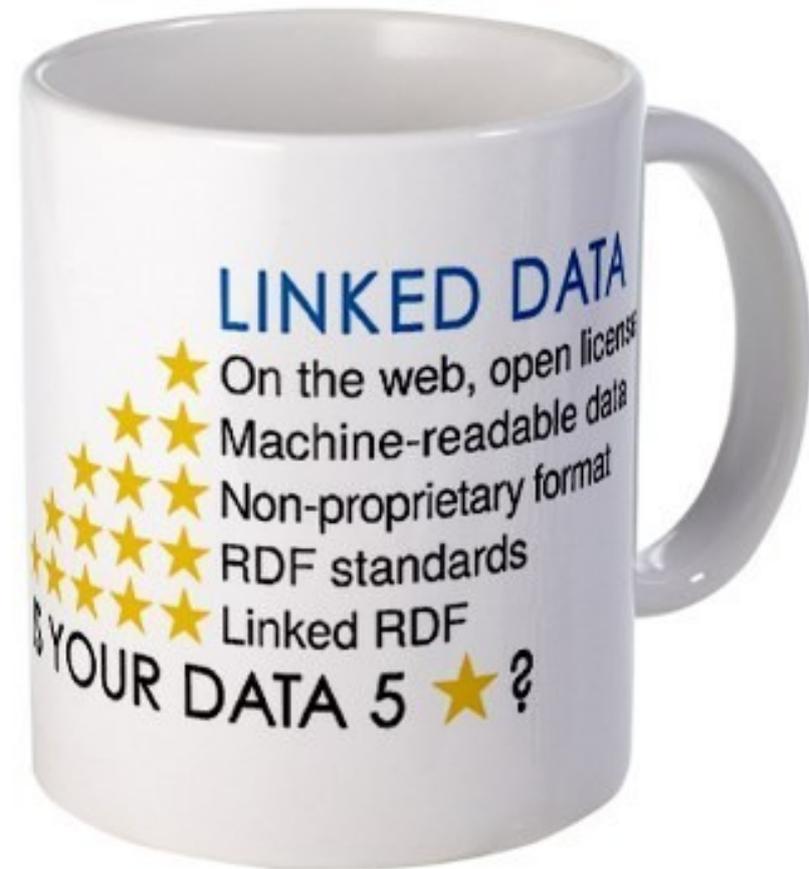
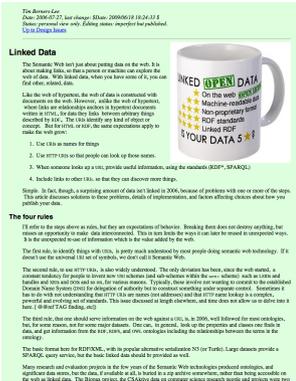
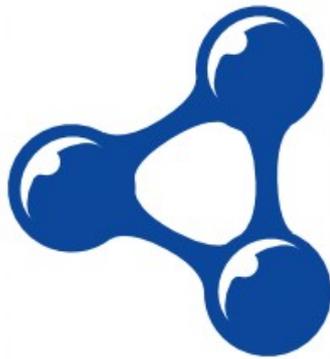
Web Semántica - 2001



Tim Berners-Lee: "The Semantic Web", *Scientific American Magazine*, May 2001



Linked Data - 2006



Tim Berners-Lee: "Linked Data", Design Issues note, 2006

¿Esto es algo nuevo?

- En realidad no, pero...
- Lo que se ha hecho automáticamente es lo que hacemos todo el tiempo los usuarios de la Web
- La diferencia: un poco de rigor (v.g. dar nombre a las relaciones) necesario para que las máquinas puedan hacerlo

¿Qué hemos hecho?

- Hemos combinado datasets diferentes que
 - están en algún sitio de la web
 - están en formatos distintos (MySQL, Excel sheet, XHTML, etc.)
 - tienen nombres diferentes para las relaciones
- La combinación fue posible porque algunas URI eran idénticas (v.g. El ISBN-s)
- Añadimos alguna información adicional sencilla (“pegamento”) empleando terminologías comunes acordadas por la comunidad
- Como resultado, se pueden encontrar y recuperar nuevas relaciones entre los datos

Podría haberse hecho mejor

- Podríamos añadir conocimiento extra a los datasets combinados
 - v.g. una clasificación completa de varios tipos de datos de biblioteconomía
 - Información geográfica
 - etc.
- Aquí intervienen las *ontologías, reglas, etc.*
- Como resultado podrían hacerse consultas más complejas



Open Data



OpenData

DATOS

- Formatos
- Microformatos
- Esquemas
- Metadatos
- Microdata



microformats

schema.org



PROTOCOLOS

- CKAN Data API
- Microsoft OData
- Facebook OGP
- SODA/ReST
- JSON-Stat



OData

Ejemplo: microformatos

The image shows two screenshots. The top one is a Google search for 'chicken pasta' with the 'Recipes' tab selected. It displays search filters like 'Ingredients', 'Any cook time', and 'Any calories'. A result for 'Cajun Chicken Pasta' is visible. The bottom screenshot shows a Facebook event page for 'Great British Booze-up' on March 14, 2011, from 17:00 to 20:00 at Shakespeare's Pub. The event description mentions SXSW and lists participants like CitySocialising and HeadScape. A 'Shopping' overlay is present over the Facebook page, listing ingredients like gorgonzola, pine nuts, lemon, tomatoes, linguine, penne, and grated parmesan with 'Yes/No' checkboxes. Below the overlay, recipe snippets for 'Chicken Piccata Pasta Toss' and 'Angel Chicken Pasta' are visible.

```
<div>Monday, March 14 ·  
<span class="dtstart">  
  <span class="value-title" title="2011-03-14T18:00:00"> </span>6:00pm</span> -  
  <span class="dtend">  
    <span class="value-title" title="2011-03-14T21:00:00"> </span>9:00pm</span>  
</div>
```

Ejemplo: schema.org

Data-Vocabulary.org

Event

See our [Event webmaster docs](#) for full explanation of the Event type, and how to encode Event data in RDFa, microformats or microdata. This page exists to contain a machine-readable microdata description of the Event type itself (at the URI <http://data-vocabulary.org/Event>).

Looking for Google's [webmaster docs](#)?
Looking for [RDF](#)?

An item with the [item type](#) <http://data-vocabulary.org/Event> represents an event such as a concert, lecture, class, etc. The following are the type's [defined property names](#):

Property	Description
summary	<i>Required.</i> The name of the event.
url	Link to the event details page.
location	The location or venue of the event. Can optionally be represented by nested Organization data, or nested Address data. Recommended (unless the page containing the markup is a page about the venue, and the location is the same for every event.) More information about nested entities.
description	A description of the event.
startDate (dtstart)	<i>Required.</i> The starting date and time of the event in ISO date format .
endDate (dtend)	The ending date and time of the event in ISO date format .

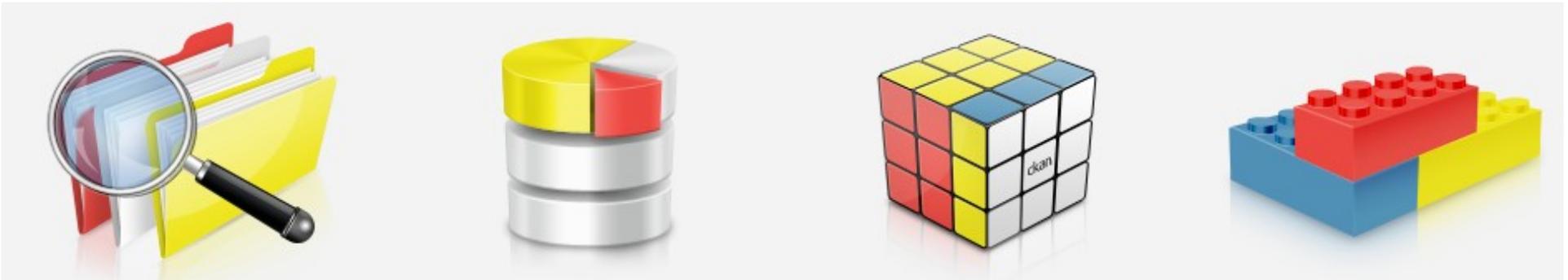
Ejemplo: CKAN

Publicar

Almacenar

Federar

Customizar

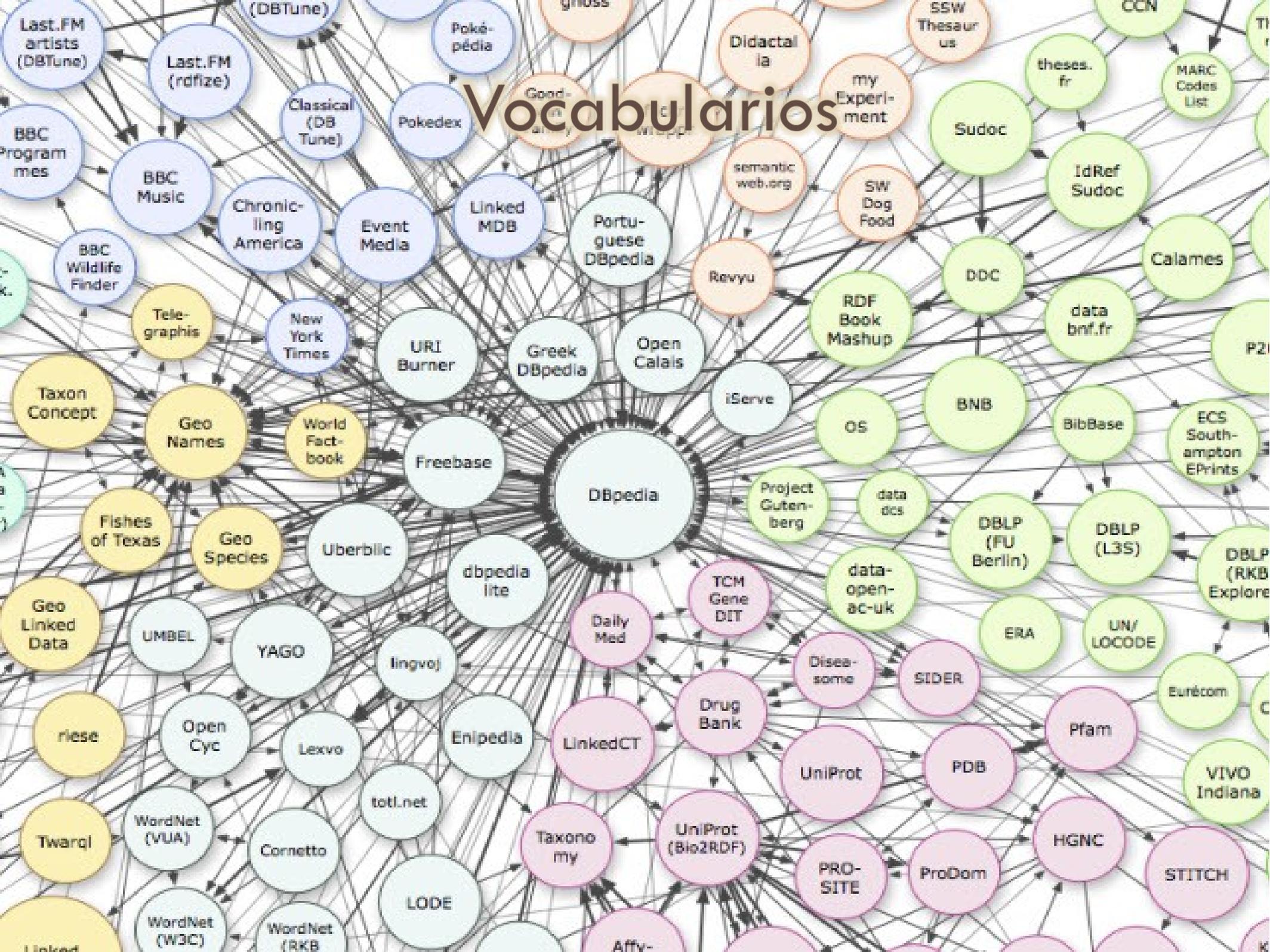


- Búsqueda
- Metadatos
- Versionado
- Comunidad
- Analítica
- Visualización
- API JSON RESTful



Open Knowledge
Foundation

Vocabularios



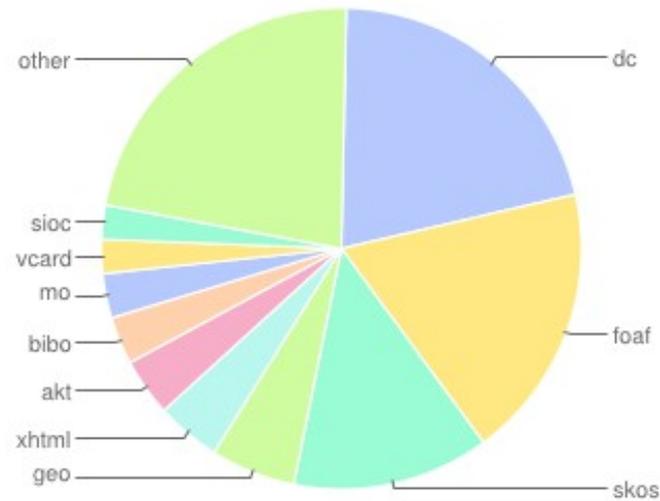
Vocabularios básicos

□ RDF(S)

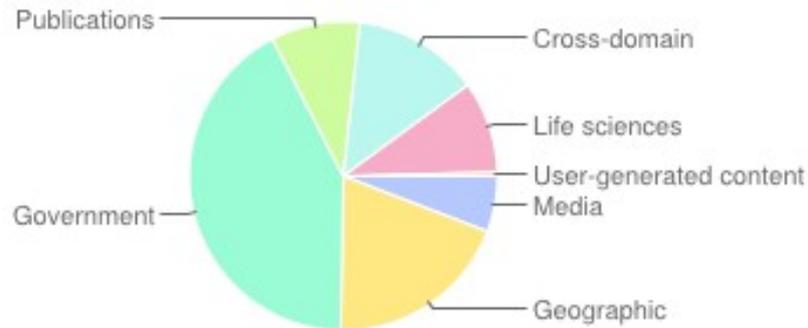
- Dublin Core: recursos de información, bibliotecas digitales, derechos, permisos, derechos de digitales
- FOAF: sobre personas y sus organizaciones
- DOAP: descripción de proyectos de software
- SIOC: Comunidades online
- vCard en RDF
- Schema.org & schema.rdfs.org
- ...

□ Ontologías/vocabularios deben compartirse y reutilizarse!

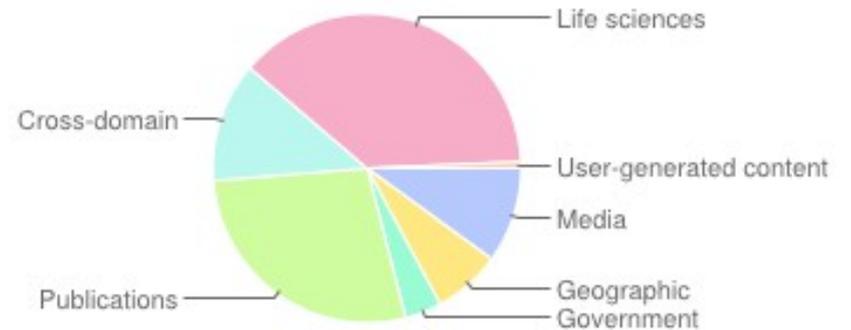
Linked Data Cloud



Datasets por dominio



Enlaces por dominio



Foo Fighters & Al Green?

Foo Fighters

Formed 1994.



SCOTT MILLS | BBC RADIO 1

Foo Fighters - Preview Footage

A very lucky member of the audience at Big Weekend recounts her experience with the Foo Fighters.

FOO FIGHTERS 1-4 OF 12

Featured on BBC MUSIC SHOWCASE

Al Green

Born 13 abril 1946.



Tony Russell/Redferns

Biography

Albert Greene (born April 13, 1946), better known as Al Green or Reverend Al Green, is an American gospel and soul music singer. He reached the peak of his popularity in the 1970s, with hit singles such as "Tired of Being Alone", "Let's Stay

Ejemplo SPARQL

```
PREFIX po: <http://purl.org/ontology/po/>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX mo: <http://purl.org/ontology/mo/>
PREFIX foaf: <http://xmlns.com/foaf/0.1/>
PREFIX event: <http://purl.org/NET/c4dm/event.owl#>
PREFIX tl: <http://purl.org/NET/c4dm/timeline.owl#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
SELECT DISTINCT ?programme ?label
WHERE {
  ?event1 po:track ?track1 .
  ?track1 foaf:maker ?maker1 . ?maker1 owl:sameAs <http://www.bbc.co.uk/music/artists/67f6
  ?event2 po:track ?track2 .
  ?track2 foaf:maker ?maker2 . ?maker2 owl:sameAs <http://www.bbc.co.uk/music/artists/fb72
  ?event1 event:time ?t1 .
  ?event2 event:time ?t2 .
  ?t1 tl:timeline ?tl .
  ?t2 tl:timeline ?tl .
  ?version po:time ?t .
  ?t tl:timeline ?tl .
  ?programme po:version ?version .
  ?programme rdfs:label ?label .
}
```

Ontologías y OWL

¿Son lo mismo?

DBpedia

```
<http://dbpedia.org/resource/Kolkata>  
...
```

Geonames

```
<http://sws.geonames.org/1275004/>  
wgs84_pos:lat "22.5697222";  
wgs84_pos:long "88.3697222";  
sws:population "4631392"  
...
```

Ontologías

- RDFS es útil, pero no resuelve todos los posibles requisitos
- Posibilidades de aplicaciones más complejas:
 - Caracterizar propiedades (no sólo rango+dominio)
 - Identificar objetos con URIs diferentes
 - Clases disjuntas o equivalentes
 - Construir clases y no sólo nombrarlas
 - Esquemas de clasificación más complejos
 - ¿Puede un programa razonar sobre ciertos términos?
v.g.:
 - “si los recursos «A» y «B» de tipo «Person» tienen la misma propiedad «foaf:email», entonces «A» y «B» son idénticos”
 - Etc.

Ontologías (cont.)

- Una **ontología** (*lógica de descripciones*) define los conceptos y relaciones empleados para describir y representar un área de conocimiento
- Lenguajes de ontologías para la Web:
 - RDFS: lenguaje sencillo de ontologías
 - DAML+OIL: antiguo lenguaje para agentes software
 - OWL: Web Ontology Language (W3C)
 - Etc.
- Los lenguajes deben ser un compromiso entre
 - Semántica rica para aplicaciones “con sentido”
 - Factibilidad, posibilidad de implementación

Limitaciones de RDF(S)

- RDF(S) no resuelve todos los posibles requisitos
- Posibilidades más complejas:
 - Caracterizar propiedades (no sólo rango+dominio)
 - Identificar objetos con URIs diferentes
 - Clases disjuntas o equivalentes
 - Construir clases y no sólo nombrarlas
 - Esquemas de clasificación más complejos
 - ¿Puede un programa razonar sobre ciertos términos?
v.g.:
 - “si los recursos «A» y «B» de tipo «Person» tienen la misma propiedad «foaf:email», entonces «A» y «B» son idénticos”

Ontologías

- Una ontología (**lógica de descripciones**) define los conceptos y relaciones empleados para describir y representar un área de conocimiento
- Lenguajes de ontologías para la Web:
 - RDFS: lenguaje sencillo
 - DAML+OIL: antiguo lenguaje para agentes software
 - OWL: Web Ontology Language (W3C)
 - Etc.
- **OWL** es una capa extra sobre RDF(S)
 - Espacio de nombres y términos propios
 - Basado en esquemas RDF

Perfiles OWL

- OWL 1
 - OWL Full: sin restricciones, no decidible
 - OWL DL (Description Logics): restricciones para asegurar la implementabilidad
 - OWL Lite
- OWL 2
 - OWL 2 EL: razonamiento $O(n^k)$ – para ontologías muy grandes
 - OWL 2 QL: para ontologías ligeras con muchos individuos en RDBs accesibles vía SQL
 - OWL 2 RL: para ontologías ligeras con muchos individuos en RDBs accesibles vía RDF

Enlazar datasets

DBpedia

```
<http://dbpedia.org/resource/Kolkata>  
  owl:sameAs <http://sws.geonames.org/1275004/>;  
  ...
```

Geonames

```
<http://sws.geonames.org/1275004/>  
  owl:sameAs <http://DBpedia.org/resource/Kolkata>  
  wgs84_pos:lat "22.5697222";  
  wgs84_pos:long "88.3697222";  
  sws:population "4631392"  
  ...
```

Perfiles OWL

- OWL 1
 - OWL Full: sin restricciones, no decidible
 - OWL DL (Description Logics): restricciones para asegurar la implementabilidad
 - OWL Lite
- OWL 2
 - OWL 2 EL: razonamiento $O(n^k)$ – para ontologías muy grandes
 - OWL 2 QL: para ontologías ligeras con muchos individuos en RDBs accesibles vía SQL
 - OWL 2 RL: para ontologías ligeras con muchos individuos en RDBs accesibles vía RDF

Conclusiones

La Web Semántica sirve para integrar automáticamente datos/información/conocimiento en la Web

El objetivo es crear una Web de datos preparada para su procesamiento automático por agentes software

Lecturas

Nigel Shadbolt, William H. Dutton, Wendy Hall and James A. Hendler (eds.), “Web science: a new frontier” Special issue, *Philosophical Transactions of the Royal Society A*, vol. 371, March 2013

