

Introducción a la Web Semántica

Web Semántica
Universidad de Valladolid
Curso 2016-2017

M. Mercedes Martínez
Dep. Informática (U. Valladolid, España)

Contenidos

- **Introducción al problema de la información en la web**
 - Problemas
 - Web de datos
 - Datos abiertos
- **Tecnologías para la representación, manipulación y compartición de datos en la Web**
 - Formatos para datos abiertos
 - Modelos de datos para representación de semántica
 - Acceso a la información semántica: consulta e integración
 - Casos de estudio: datos enlazados, datos abiertos enlazados, wikidata, mashup semánticos
 - Semántica y big data

Modelos de datos, tecnologías y estándares

- **RDF**
 - Serialización de RDF: RDF/XML, Notación N3, Turtle, N-Triples
 - RDFa
- **Vocabularios y esquemas**
 - RDFS
 - SKOS
 - FOAF y Schema.org
- **Lenguajes de consulta**
 - SPARQL

¿Por qué?

- La Web está hecha para las personas
- En los buscadores se busca por términos
- Hay silos de datos aislados

Web Semántica

“The Semantic Web has, in a way, become almost a celebrity (...) although most people don't know what it is, and although there really isn't a Semantic Web yet”

[Passin04, pag. 2]

¿Qué es la Web Semántica?

- *“The first step is putting data on the Web in a form that machines can naturally understand, or converting it to that form. This creates what I call a Semantic Web – a web of data that can be processed directly or indirectly by machines”* [BernersLee99]
- *“an extended web of machine-readable information and automated services that amplify the Web far beyond current capabilities”* (Tim Berners-Lee, según [Daconta03, p. xxi])
- *“The trend toward the Semantic Web is based on the goal of semantic interoperability of data, which enables application independence, improved search facilities, and improved machine inference”* [Daconta03, p. xiv]
- *“So, I think, data integration is the name of the game”* [BernersLee08]

¿Qué es Web Semántica?

La Web actual

Conocimiento compartido

+ Extracción de conocimiento

Web Semántica

Web de Datos = Web Semántica

- *La iniciativa de Web Semántica del W3C se ha integrado en la iniciativa de Web de Datos*
- Principios de diseño:
 - Hacer disponibles datos estructurados en formatos estándar en la web
 - Dejar accesibles en la web no sólo los conjuntos de datos, sino también elementos de datos individuales y sus relaciones
 - Describir la semántica de esos datos en un formalismo tal que pueda ser procesado por las máquinas

Historia: cómo se llega a la idea de Web semántica

- **La evolución de la informática**
 - Los datos son secundarios respecto a las aplicaciones
 - Los datos se transforman en elementos de primer nivel, pero siguen estando en formatos propietarios, que no se pueden intercambiar entre aplicaciones
 - Los datos crecen en importancia frente a las aplicaciones con la llegada de la Web

[Daconta03]

Historia: cómo se llega a la idea de Web semántica

- **La evolución de los datos**
 - Registros de bases de datos y documentos de texto
 - Documentos XML con vocabularios sencillos
 - Taxonomías XML y documentos con varios vocabularios
 - Ontologías XML y razonamiento automático

[Daconta03]

Fundamentos de la Web Semántica

La Web actual proporciona las siguientes bases para la Web Semántica:

- **Recursos**
 - Identificados por URI (Uniform Resource Identifier)
 - Un documento o trabajo (por ejemplo, el libro [Passin04]) es un recurso, con una URI, aunque puede haber múltiples versiones o copias de dicha obra
 - Una ley, una obra artística, una obra intelectual, ... son recursos
 - Las URI se ajustan a *estándares* (RFC 2396)
- **Pocos comandos**
 - HTTP es un protocolo con pocos comandos, y sencillos
- **Escalabilidad**
- **Es abierta, incompleta e inconsistente**
 - **Incompleta:** puede haber enlaces que no funcionan
 - **Inconsistente:** El contenido de dos páginas puede ser contradictorio

Problemas de la Web actual

- Sobrecarga de información
- Buscar y extraer información
- Presentación de la información
- Relaciones
 - Las relaciones entre los items de información como “include”, “describe” no se modelan
- Metadatos

Problemas actuales

- **Sistemas cerrados**
 - Sistemas cuyos componentes están diseñados para trabajar conjuntamente. La información fluye dentro del sistema pero no se puede compartir entre sistemas u organizaciones.
- **Integración de información**

El problema de la sobrecarga de información

- **La Web**

- La Web ha supuesto un impulso enorme en la facilidad para acceder a grandes masas de información
- La masificación de datos dificulta encontrar la información que necesitamos y también su manipulación

El problema de la sobrecarga de información

- La Web

↑ Datos ≡ ↓ Calidad

- ¿Por qué?
 - La información está en lenguaje natural
 - El lenguaje natural es adecuado para los humanos, pero no para las máquinas ⇒
 - No se puede procesar la información automáticamente

Problemas para buscar y extraer información

- Encontramos páginas que no nos interesan
- Tenemos que leer la página que nos devuelve el buscador para encontrar la información que realmente nos interesa. Las herramientas automáticas no son capaces de extraerlo de texto.
- Podemos buscar *términos*, pero no *conceptos*

Problemas con la presentación de la información

- Mantener los sitios Web puede ser una tarea costosa
- Controlar la redundancia y consistencia (sobre todo cuando hay actualizaciones) es muy difícil

Problemas con la manipulación de conocimiento

Manipulación de conocimiento: adquisición, mantenimiento y acceso al conocimiento de una organización

- **Problemas:**
 - Buscar información
 - Extraer información
 - Mantenimiento
 - Generación automática de documentos
 - Documentos (datos semiestructurados) adaptados al perfil del usuario

Conocimiento tácito

- Las personas usamos conocimiento tácito. Somos inconscientes de que lo hacemos.
 - Para representarlo antes debemos tomar consciencia de él.
- Los humanos nos expresamos de forma laxa. Los lenguajes formales son rígidos.
 - En conocimiento que evoluciona esto es especialmente relevante.

Conocimiento dependiente del entorno

- Lo mismo se puede describir de distintos modos y con diferentes niveles de detalles en entornos distintos.
- Interesar ser capaz de convertir/transformar unas descripciones en otras de modo automático.

Conocimiento dependiente del entorno

Library of Congress (LoC)

British Library (BL)

Author

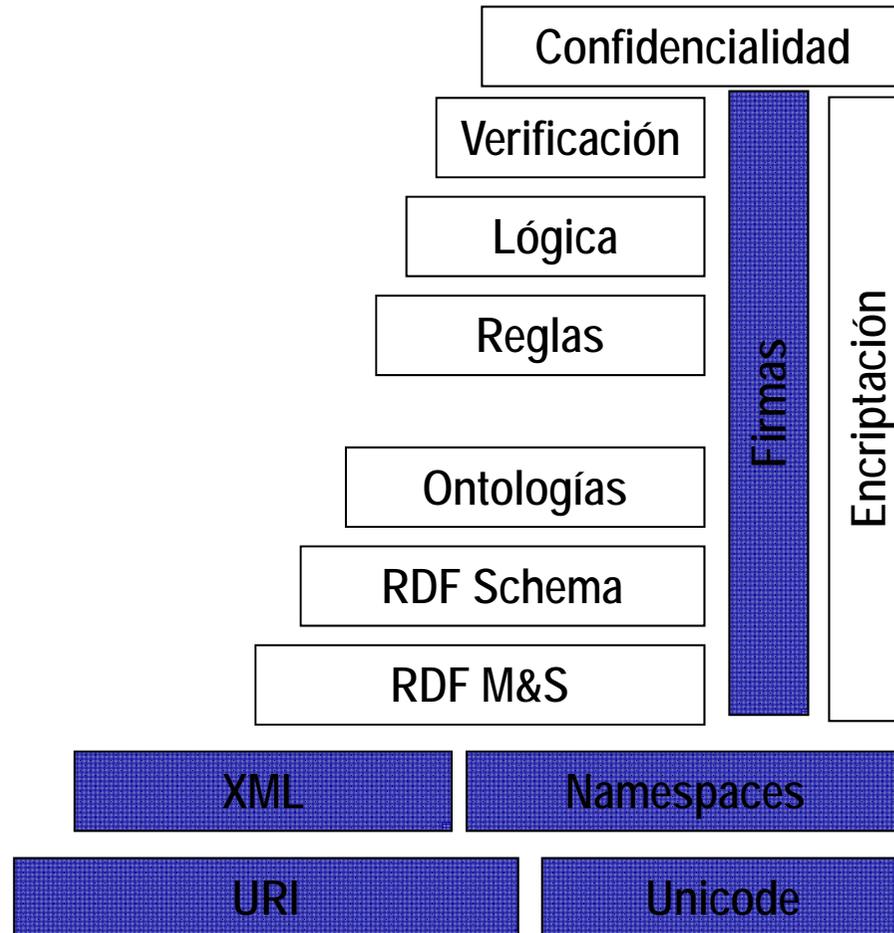
Creator

∀ persona X

∀ recurso Y

Si X es (LoC) 'Author' de Y
entonces X es (BL) 'Creator' de Y

Los componentes de la Web semántica



The Semantic Web Stack.
Tim Berners-Lee.
Copyright [2002] World Wide Web Consortium

Las piezas del puzzle



Argumentos en contra de la Web Semántica

- **El mal precedente de la Inteligencia Artificial**
 - Respuesta: *“A Semantic Web is not Artificial Intelligence. The concept of machine-understandable documents does not imply some magical artificial intelligence which allows machines to comprehend human mumblings. It only indicates machine’s ability to solve a well-defined problem by performing well-defined operations on existing well-defined data. Instead of asking machines to understand people’s language, it involves asking people to make the extra effort.”* [Tim Berners-Lee. “What the Semantic Web can Represent”, <http://www.w3.org/DesignIssues/RDFnot.html>]
- **Los costes de implementación**
- **La postura conservadora**
 - “¿Para qué cambiar lo que tenemos?”

Otras consideraciones

- Viabilidad en el campo de las iniciativas relacionadas con metadatos y estándares
- Adopción de estándares por el mercado
- Eficacia y utilidad de la Web Semántica frente a herramientas como Google y Amazon
 - ¿O integración en estas herramientas?

Referencias

- **Libros básicos:**
 - [Antoniou12] “A Semantic Web Primer”. Antoniou, Groth, Harmelen and Hoekstra. 3^a ed. MIT Press. 2012.
 - [Allemang11] “Semantic web for the working ontologist – effective modeling in RDFs and OWL”, 2^a ed. Allemang and Hendler. Morgan Kaufmann. 2011.
 - [Yu14] “A Developer’s Guide to the Semantic Web”. Yu. 2^a ed. Springer. 2014.

Referencias

- URLs:
 - W3C Semantic Web Activity. <http://www.w3.org/2001/sw/>
 - Semantic Web wiki. <http://www.semanticweb.org/>
 - RDF wiki. <http://www.w3.org/RDF/>
 - Linked Data. <http://www.linkeddata.org/>