

Ontologías y Web Semántica

Aplicaciones RDF

Pablo R. Fillottrani

Depto. Ciencias e Ingeniería de la Computación
Universidad Nacional del Sur

marzo 2019



Librerías

- ▶ **PerlXMLParser** un conjunto de módulos CPAN que implementan un parser SAX para RDF, en PERL
www.w3c.org/1999/02/26-modules/
- ▶ **ICS-FORTH VRP** un parser RDF validador para RDF Schema, en Java
www.ics.forth.gr/proj/isst/RDF
- ▶ **DATAx** una biblioteca de módulos para simplificar el intercambio de datos estructurados
www.megginson.com/DATAX/
- ▶ **Libwww** un parser RDF, de la biblioteca de demostración de W3C
www.w3c.org/Library/



Usos RDF

Herramientas para procesamiento RDF: estado actual

Vocabularios RDF

Proyecto Linked Data



- ▶ **SIRPAC** un conjunto de clases Java que puede convertir un documento RDF/XML en el conjunto de triplas de su modelo de datos
www.w3c.org/RDF/Implementations/SiRPAC
- ▶ **SWI-Prolog Parser** un parser RDF compacto y eficiente escrito en Prolog
swi.psy.uva.nl/projects/SWI-Prolog/packages/sgml/online.html
- ▶ **Jena** una biblioteca completa de código abierto para procesar datos RDF, RDF Schema y OWL
jena.sourceforge.net



Módulos para bases de datos

- ▶ **rdfDB** un sistema de base de datos RDF, con su propio lenguaje de consulta alla SQL
www.guha.com/rdfdb/
- ▶ **Redland** una biblioteca para manipular grafos y triplas RDF, RDF Schema
librdf.org/
- ▶ **Algae** un sistema que implementa un lenguaje de consulta para RDF, que puede tener interfaces con bases de datos relacionales
www.w3.org/2004/05/06-Algae/
- ▶ **Oracle Spatial 10g** una plataforma comercial para el manejo de datos RDF
www.oracle.com/technology/tech/semantic_technologies



Aplicaciones

- ▶ **Reggie: the metadata editor** un applet que permite la creación de metadatos en varios formatos, HTML, RDF, etc
metadata.net/dstc/
- ▶ **DC Dot** a partir de una página web genera información RDF/XML con los metadatos Dublin Core de la misma
www.ukoln.ac.uk/metadata/dcdot/
- ▶ **Mozilla** usa RDF
www.mozilla.org/rdf/doc/



Otras librerías

- ▶ **Virtuoso RDF** un motor de base de datos multi-modelo
virtuoso.openlinksw.com
- ▶ **RDF Crawler** una herramienta en Java que descarga e interconecta fragmentos de RDF, basado en SiRPAC
ontobroker.semanticweb.org/rdfcrawl
- ▶ **RDFStore** un API en PERL que soporta almacenamiento y recuperación de triplas y generación automática de vocabulario
xml.jrc.it/RDFStore



- ▶ **Metalog** un sistema de consultas basado en RDF
www.w3.org/RDF/Metalog/
- ▶ **Protege** una herramienta que permite construir una ontología, e ingresar conocimiento basado en ella. Puede ser extendida en diversas formas. Los resultados se pueden exportar en RDF, RDF Schema, OWL
protege.stanford.edu/



Objetivos

- ▶ FOAF proporciona vocabulario básico para ayudar a publicar en la Web información sobre la conexión entre personas
- ▶ miles de personas tienen hoy información personal en sus páginas web. Mediante FOAF se pueden crear aplicaciones que aprendan sobre las distintas formas de relaciones entre las personas, lugares y objetos descriptos en la Web.
- ▶ FOAF usa RDF para integrar esta información



Términos más comunes

- ▶ **términos FOAF básicos:** Agent, Person, name, nick, title, homepage, mbox, img, surname, family_name, firstName, etc.
- ▶ **información personal:** weblog, knows, interest, currentProject, pastProject, schoolHomepage, publications, etc.
- ▶ **conexión online:** OnlineAccount, OnlineChatAccount, msnChatID, accountName, icqChatID, accountServiceHomepage, etc.
- ▶ **proyectos y grupos:** Project, Organization, Group, member, fundedBy, theme, etc.
- ▶ **documentos e imágenes:** Document, Image, PersonalProfileDocument, topic, made, etc.



Objetivos

- ▶ mediante el uso de este vocabulario en la información disponible en páginas web, es mucho más fácil responder preguntas como
 - ▶ “encontrar las páginas web recomendadas por personas que trabajan para Greenpeace”
 - ▶ “encontrar las publicaciones de las personas con las que he trabajado”
 - ▶ “encontrar críticas a una página web, y las páginas web de los autores de las críticas”
- ▶ es decir, se quiere poder consultar en la Web preguntas comunes sobre documentos, organizaciones y personas



Ejemplo

```
<foaf:Person rdf:about="#me" xmlns:foaf="http://xmlns.com/foaf/0.1/">
<foaf:name>Pablo Fillottrani</foaf:name>
<foaf:mbox>prf@cs.uns.edu.ar</foaf:mbox>
<foaf:mbox_sha1sum>241021fb0e6289f92815fc210f9e9137262c252e</foaf:mbox_sha1sum>
<foaf:homepage rdf:resource="http://www.cs.uns.edu.ar/~prf/" />
<foaf:Organization rdf:resource="http://www.cs.uns.edu.ar/" />
<foaf:img rdf:resource="http://www.cs.uns.edu.ar/~prf/images/yo.jpg" />
</foaf:Person>
```



Objetivos

- ▶ la [Iniciativa para Metadatos de Dublin Core](http://dublincore.org/) (DCMI) dublincore.org/ es una organización dedicada a la promoción de la adopción amplia de estándares interoperables de metadatos, y el desarrollo de vocabularios especializados para describir recursos
- ▶ de esta manera se favorece la implementación de sistemas inteligentes de descubrimiento de la información
- ▶ para esto DCMI realiza
 - ▶ desarrollo y mantenimiento de estándares internacionales para describir recursos
 - ▶ soporte para la comunidad de usuarios y desarrolladores
 - ▶ promoción del uso de las soluciones de Dublin Core



- ▶ existen varias herramientas construidas sobre el vocabulario del Dublin Core
- ▶ en general son buscadores inteligentes de información, editores de metadatos, y extractores automáticos de los metadatos
- ▶ por ejemplo DC-dot
www.ukoln.ac.uk/metadata/dcdot/



Términos más comunes

- ▶ en general los [términos](#) que formaliza el Dublin Core sirven para especificar metadatos de documentos
- ▶ por ejemplo
 - ▶ creator
 - ▶ date
 - ▶ language
 - ▶ description
- ▶ para cada término, el Dublin Core especifica el URI con el cual identificarlo, la etiqueta, una pequeña descripción, y la fecha en que fue aceptado como recomendación



RSS

- ▶ [RSS](#) es una familia de formatos para contenidos web, usados para publicar información cuyo contenido se actualiza frecuentemente
- ▶ ejemplos de uso: contenido blog, titulares de noticias, podcasts
- ▶ existen varias versiones propietarias, pero RSS 1.0 (RDF Site Summary) es un estándar W3C y está basado en RDF
 - ▶ Really Simple Syndication (RSS 2.0)
 - ▶ RDF Site Summary (RSS 1.0 and RSS 0.90)
 - ▶ Rich Site Summary (RSS 0.91)



- ▶ un documento RSS, llamado **feed**, **web feed** o **channel**, contiene ya sea un resumen del contenido asociado a un sitio web, o el texto completo
- ▶ RSS permite mantener actualizado el contenido de sitios web en forma automática (sin necesidad de recargarlo manualmente)
- ▶ el contenido puede ser procesado por distintos tipos de software
 - ▶ un **feed reader** es un programa o un plugin para descargar los *feeds* y visualizarlos de acuerdo a las preferencias del usuario
 - ▶ un **Web aggregator** es un programa que lee los *feeds*, posiblemente de diversos orígenes, y los incorpora al contenido de una página web



Ejemplo

```
<rdf:RDF
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns="http://purl.org/rss/1.0/">
  <channel>
    <title>Example Dot Org</title>
    <link>http://www.example.org</link>
    <description>the Example Organization web site</description>
    <items>
      <rdf:Seq>
        <rdf:li resource="http://www.example.org/status/">
          </rdf:li>
        </rdf:Seq>
      </items>
    </channel>
    <image rdf:about="http://www.example.org/images/logo.gif"/>
    <image rdf:about="http://www.example.org/images/logo.gif"/>
      <title>Example</title>
      <url>http://www.example.org/images/logo.gif</url>
      <link>http://www.example.org</link>
    </image>
    <item rdf:about="http://www.example.org/status/">
      <title>New Status Updates</title>
      <link>http://www.example.org/status/</link>
      <description>News about the Example project</description>
    </item>
  </rdf:RDF>
```



Objetivos

- ▶ en la Biología se pierde actualmente mucho tiempo y esfuerzo en la búsqueda de la información disponible
- ▶ debido a una amplia variedad en la terminología usada en diferentes áreas de investigación
- ▶ el proyecto **Gene Ontology** (GO) www.geneontology.org es un esfuerzo colaborativo para desarrollar descripciones consistentes de productos genéticos existentes en bases de datos distintas
- ▶ nació en 1998 como una colaboración entre bases de datos genéticas específicas de tres organismos, pero actualmente tiene un objetivo mayor



- ▶ en GO se desarrollan vocabularios estructurados y controlados (ontologías) para describir los genes y sus productos: procesos biológicos, componentes celulares y funciones moleculares asociadas
- ▶ existen tres actividades principales
 1. desarrollo y mantenimiento de las ontologías
 2. establecer las asociaciones entre los términos de la ontologías y las bases de datos
 3. desarrollar herramientas que se favorezcan por el uso de la ontología



Herramientas

- ▶ editores, visualizadores y buscadores en la ontología
- ▶ buscadores y visualizadores de información que usan la información de la ontología
- ▶ herramientas de anotación (descripción, comentarios)
- ▶ herramientas de análisis de expresiones genéticas



Cómo se ubica en la Web Semántica

- ▶ en contraposición con la visión de la Web Semántica, Linked Data se focaliza en publicar información estructurada en RDF usando URIs en lugar del razonamiento ontológico
- ▶ esta simplificación permite bajar la barrera de entrada para los proveedores de datos, y favorece una aceptación más amplia



Objetivos

- ▶ **Web de Datos** = linked data + vocabularios + metadatos embebidos (RDFa, microformats, etc.)
- ▶ cuando se publican linked data se provee una forma estandarizada y uniforme para
 - ▶ descubrir
 - ▶ integrar
 - ▶ consultar en forma distribuida
 - ▶ acceder datos y metadatos uniformemente
- ▶ es como una API para acceder a los datos a través de HTTP



Principios

- ▶ presentados también por Tim Berners-Lee en 2006
 - ▶ uso de URIs para identificar objetos (cualquiera, no solamente documentos)
 - ▶ el uso de HTTP URIs – nombres únicos globales, propiedad distribuida- permite buscar objetos
 - ▶ proveer información útil en RDF cuando alguien busca un URI
 - ▶ incluir enlaces RDF a otros URIs para buscar más información relacionada



Problemas

- ▶ son sólo principios, no consejos de implementación
- ▶ muchas dudas (a propósito?)
- ▶ recursos informativos vs recursos no-informativos
- ▶ formato: HTML + RDF/XML vs. RDFa



Negociación de Contenido en HTTP

- ▶ la **Negociación de Contenido** (CN) es el proceso de seleccionar la mejor representación para una dada respuesta cuando hay múltiples representaciones disponibles
- ▶ tres tipos de CN: server-driven, agent-driven CN, transparent CN
- ▶ ejemplo

```
HTTP/1.1 303 See Other
Content-Type: application/rdf+xml
Location: http://www.cs.uns.edu.ar/~prf
```



Repaso HTTP

- ▶ los mensajes HTTP consisten de un requerimiento de un cliente hacia un servidor, y su respuesta del servidor al cliente
- ▶ consisten de un conjunto de métodos predefinidos (GET, POST, etc), que puede ser expandido
- ▶ y un conjunto de códigos de estado definidos
 - ▶ informacionales 1xx: respuesta provisoria (100 Continue)
 - ▶ exitosos 2xx: pedido recibido exitosamente, entendido y aceptado (201 Created)
 - ▶ redirección 3xx: son necesarias acciones extra por parte del cliente para completar el pedido (301 Moved Permanently)
 - ▶ error en el cliente 4xx: método no permitido (405 Method Not Allowed)
 - ▶ error en el servidor 5xx: el servidor detectó una condición inesperada (501 Not Implemented)

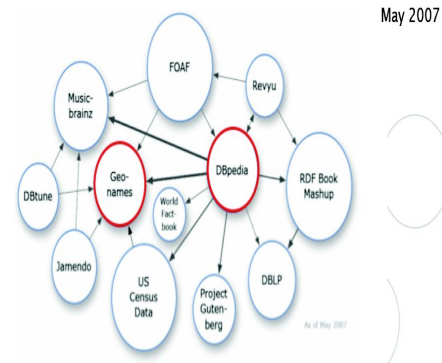


Proyecto

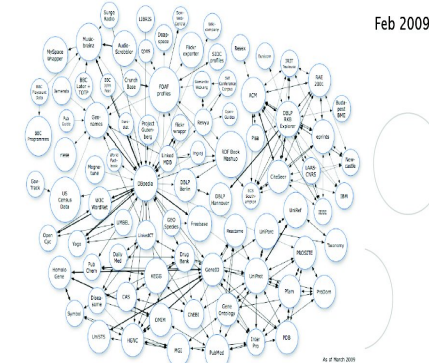
- ▶ la comunidad comenzó con el apoyo de la W3C a principios de 2007
- ▶ la idea es tomar conjuntos de datos existentes y ponerlos disponibles en RDF
- ▶ establecer interconexión con otros conjuntos de datos



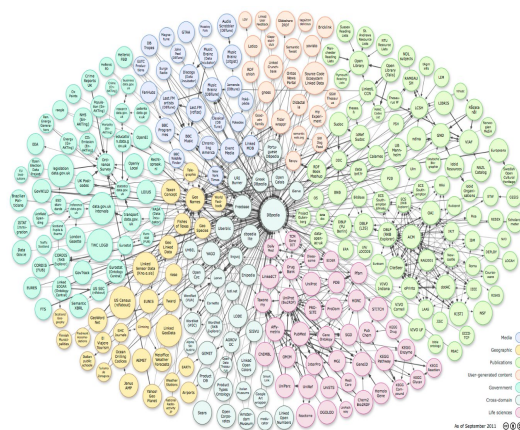
Nodos en 2007



Nodos en 2009



Nodos en 2011



Herramientas y Aplicaciones

- ▶ browsing Tabulator, VisiNav, DBpedia Mobile, etc.
- ▶ búsqueda Sindice, SWSE, Falcons, etc.
- ▶ integración Revyu, BBC Music, DERI Pipes
- ▶ ver

<http://esw.w3.org/topic/SweoIG/TaskForces/CommunityProjects/LinkingOpenData/Applications>

