

Tema 5. Publicación de Datos abiertos

J. Javier Samper Zapater
José Fco. García Calderaro
Departamento de Informática

VNIVERSITAT  ID VALÈNCIA

Web semántica y los Repositorios Institucionales

Asistimos a diario a iniciativas importantes que se desarrollan en paralelo.

- **Open Data:** Las organizaciones (principalmente gubernamentales) se hallan en el proceso de publicar los datos que generan con la finalidad de posibilitar nuevos servicios y aplicaciones a partir de ellos. Un ejemplo: Periodismo basado en datos.
- Muchas de estas iniciativas liberan sus datos en formatos tales como CSV, texto plano, XML.
- En estos portales comienza a tomar relevancia la utilidad / necesidad de publicar sus instancias de datasets bajo los 4 principios de Linked Open Data (URIs para dataset propio, http, RDF, linkear con otras URIs de datasets externos).
- Se comienza a trabajar con nuevas tecnologías tales como endpoints SPARQL en implementaciones reales de la Web Of Data o Web Semántica.

Breve Introducción a la Web Semántica o Linked Data

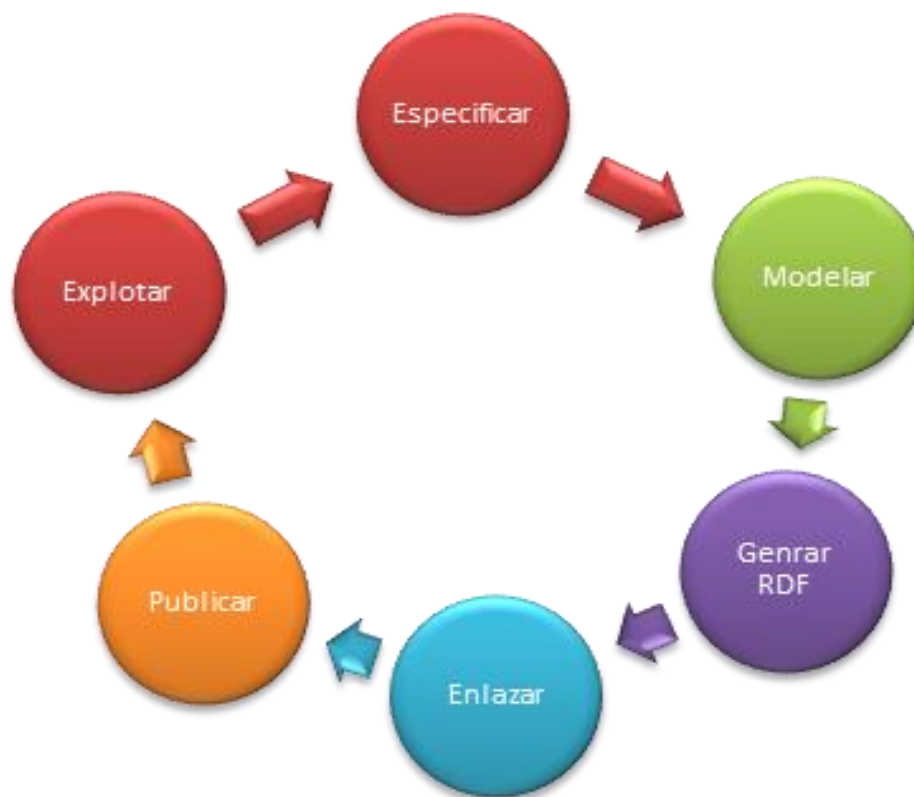
- Si dibujamos los conceptos como nodos y las relaciones como arcos , obtenemos una estructura que sobretodo los Ingenieros Informáticos conocemos bien: el grafo.
- Tim Berners Lee, director del W3C e inventor de la Web, pretende que este grafo sea consultado como si la Web se tratara de una base de datos global, sobre la que se pueden aplicar nuevos algoritmos para adquirir nuevos conocimientos y sirva de base para crear inteligencias artificiales capaces de entender el lenguaje humano.
- Esta idea es todavía utópica. Las Tecnologías Linked Data tienen muchos problemas que resolver antes de llegar a cumplir ese y otros objetivos como, por ejemplo, que la tecnología se fácil de usar para todo el mundo.

Breve Introducción a la Web Semántica o Linked Data

- A pesar de estos problemas , la Web Semántica ha tenido bastante éxito.
- En 2007 surgió en W3C el proyecto **Linked Open Data**.
- En su Web han publicado un grafo que muestra los enlaces creados entre todas las Web semánticas.
- Se va actualizando continuamente. El ritmo de crecimiento del grafo es exponencial.
- En el centro del grafo tenemos a la Dbpedia, la versión semántica de la Wikipedia y muy cerca Freebase que es actualmente propiedad de Google y ha sido base para su Knowledge graph.
- Knowledge graph sirve para complementar las búsquedas con información semántica.

Proceso de publicación de datos para portales de datos abiertos

- El proceso de publicación en Linked Data tiene un modelo de ciclo de vida incremental iterativo:



Especificación

Especificación:

- Análisis de fuentes de datos
- Diseño de URIs
- Definición de licencias

Diseño de URIs

- Usar URIs “cool”, en lugar de URIs opacas, siempre que sea posible
 - URI “cool”: <http://dbpedia.org/resource/Azuaga>
 - URI opaca: <http://www.geonames.org/2521436/>
- Existen varias propuestas para la generación de URIs
 - Reino Unido:
 - <http://data.gov.uk/resources/uris>
- España: Secretaría de Estado para la función pública (Norma Técnica de Interoperabilidad de reutilización de recursos de información)

Definición de licencia

Múltiples posibilidades

- The UK Open Government License
- Open Database License
- Public Domain Dedication and License
- Open Data Commons Attribution License
- The Creative Commons Licenses

También es posible reutilizar y aplicar una licencia existente de las fuentes de datos del gobierno.

Generación de RDF

- Tomar las fuentes de datos seleccionadas en la actividad de especificación y transformarlas a RDF de acuerdo con el vocabulario creado en la actividad de modelado
- Algunas herramientas
 - CSV y hojas de cálculo
 - Apache Any23, RDF Refine, XLWrap, RDF123, NOR2O, Tabela
 - Bases de datos relacionales
 - D2R Server, ODEMapster, Morph, W3C RDB2RDF WG – R2RML
 - XML
 - GRDDL, ReDeFer
- <http://www.w3.org/wiki/ConverterToRdf>

Generación de RDF

- Tomar las fuentes de datos seleccionadas en la actividad de especificación y transformarlas a RDF de acuerdo con el vocabulario creado en la actividad de modelado
- Algunas herramientas
 - CSV y hojas de cálculo
 - Apache Any23, RDF Refine, XLWrap, RDF123, NOR2O, Tabela
 - Bases de datos relacionales
 - D2R Server, ODEMapster, Morph, W3C RDB2RDF WG – R2RML
 - XML
 - GRDDL, ReDeFer
- <http://www.w3.org/wiki/ConverterToRdf>

Enlazado: Fases

1. Identificar conjunto de datos con los que se puede enlazar
2. Descubrir relaciones entre conjuntos de datos
3. Validar las relaciones descubiertas

Registro de conjuntos de datos

- Registrar el conjunto de datos en el CKAN Registry
- Generar ficheros *sitemap* para el conjunto de datos, usando *sitemap4rdf*
- Enviar el fichero *sitemap* a Google y Sindice

Algunos ejemplos:

- El proyecto Innpronta Ciudad2020 se centra en tres pilares de la Smart City, que son transporte, energía y medioambiente, y propone una Plataforma Linked Data para la gestión e integración de datos provenientes de Smart Cities en esos tres dominios. El Portal de Linked Data de Ciudad2020 cuenta con conjuntos de datos cuyos contenidos abarcan sistemas de compartición de bicicletas, museos y restaurantes. Para la administración y los desarrolladores provee también un punto de acceso sparql y posibilidad de consultas en streaming, con resultados en tiempo real.
- El Portal de Datos Abiertos de Zaragoza incluye más de 100 conjuntos de datos abiertos, en distintos ámbitos. Estos conjuntos de datos están disponibles mediante el catálogo, y se puede acceder libremente a ellos bajo términos de uso basados en datos.gob.es. Asimismo, el portal cuenta con un punto de acceso SPARQL asociado a varios de estos conjuntos de datos, de tal manera que puedan ser utilizados por desarrolladores.
- El Portal de Open Data de Euskadi . Su catálogo lo conforman 17 conjuntos de datos sobre diversos dominios: Geografía, Meteorología, Estadística, Turismo y Tramitación administrativa.
- El Gobierno del Principado de Asturias sigue la tendencia RISP (Reutilización de Información del Sector Público). El Catálogo de Datos de Asturias contiene 26 conjuntos de datos disponibles en múltiples formatos, tanto para el ciudadano, como para desarrolladores.
- El Portal de Datos Abiertos del Ayuntamiento de Madrid está dedicado a promover el acceso a los datos del gobierno municipal e impulsar el desarrollo de herramientas creativas para atraer y servir a los ciudadanos de la ciudad de Madrid .
- El portal de Transparencia y Datos Abiertos del Ayuntamiento de Valencia, pone a disposición del ciudadano, datos sobre [Medio Ambiente](#), [Sociedad y Bienestar](#), [Transporte](#), [Urbanismo e Infraestructuras](#), [Salud](#) y [Turismo](#) .

Referencias

- El Imprescindible futuro de la Web Semántica. Ramón Saquete en Google+ - Desarrollo Web
- **Beginner's guide to RDF: 6. Querying with SPARQL. Disponible en <https://code.google.com/p/tdwg-rdf/wiki/Beginners6SPARQL>**
- Fuente Web: <http://www.w3.org/2009/03/xbri/report.html>
- Fuente Web: [Grafo Interactivo LOD: http://inkdroid.org/lod-graph/](http://inkdroid.org/lod-graph/)
- Fuente Web: <http://www.w3.org/wiki/ConverterToRdf>
- Herramientas de publicación de datos para portales de datos abiertos. O. Corcho Tecnologías Semánticas en la Web de Datos. Red.es. 13 de diciembre de 2012.
- Collaborative Project. LOD2 - Creating Knowledge out of Interlinked Data. Deliverable 3.2.2. DBpedia-Live Extraction
- Collaborative Project. LOD2 - Creating Knowledge out of Interlinked Data. Deliverable 3.1.1. Report on Knowledge Extraction from Structured Sources.

Referencias 2

- WhitePaper Ciudad2020_Iniciativas de Linked Data para Smart Cities (www.innprontaciudad2020.es) .
- Ciudad 2020, Transporte linkeddata. <http://transporte.linkeddata.es> , 2013
- Datos abiertos de Zaragoza <http://www.zaragoza.es/ciudad/risp/>
- Open Data Euskadi, <http://opendata.euskadi.net/w79-home/eu/> [4]. Datos de Asturias
- <http://www.asturias.es/portal/site/webasturias/menuitem.6e0e3b8f454aecb96c0be183ec12b2a0/?vgnextoid=9ccb5d0cfd524410VgnVCM100000ce212b0aRCRD>
- Portal de datos abiertos del ayuntamiento de Madrid. <http://datos.madrid.es/>
- Portal de transparencia y datos abiertos del Ayuntamiento de Valencia. <http://gobiernoabierto.valencia.es/es/data/>