

Consumo y visualización de datos enlazados mediante aplicaciones web de código libre

Consumption and visualization of linked data through of Open Source Applications

Ramiro L. Ramírez-Coronel¹, Pablo A. Quezada-Sarmiento^{1,2}

¹ Universidad Técnica Particular de Loja, Ecuador

² Universidad Internacional del Ecuador, Ecuador

rlramirez@utpl.edu.ec , paquezadasa@uide.edu.ec

RESUMEN. La Web se ha convertido en un recurso necesario de uso diario, los beneficios que ofrece, al ser una fuente de conocimiento y colaboración, dando lugar a nuevas iniciativas como los datos enlazados, cuyo propósito es vincular los datos esparcidos por la Web por medio de relaciones semánticas entre estos. El propósito de este artículo, es mostrar las mejoras del consumo y visualización de datos enlazados en la Web, de tal manera que cualquier usuario, con o sin ningún tipo de conocimiento sobre el tema, pueda explorar bases de datos semánticas, mediante la construcción de un software capaz de proporcionar un buscador para obtener información relacionada a un término de búsqueda en lenguaje natural que es convertido a consulta SPARQL, así como también la visualización de resultados, por medio de un grafo, para lo cual se realiza una investigación y análisis de aplicaciones de código abierto existentes; además de considerar los repositorios semánticos más utilizados que aportaron a la propuesta y desarrollo del software.

ABSTRACT. The Web has become a necessary resource of daily use, the benefits it offers, being a source of knowledge, and collaboration giving rise to new initiatives such as linked data, whose purpose is link the data scattered through the Web through of semantic relationships between them. The propose of this article is show the improving, consumption, and visualization of linked data in the Web, in such a way that any user with or without any knowledge on the subject, can explore semantic databases, through the construction of software capable of providing a search engine for information related to a natural language search term that is converted to SPARQL query, in the same context the visualization of results by means of a graph, for which a research, and analysis of existing open sources applications is carried out. In addition, semantic repositories contributed to the proposal, and for the develop of software.

PALABRAS CLAVE: Datos enlazados, Semántica, Visualización de datos enlazados, Sparql, RUP.

KEYWORDS: Linked data, Semantics, Visualization of linked data, Sparql, RUP.

1. Introducción

Actualmente se crean nuevos dispositivos y servicios sociales que generan datos en la web, este crecimiento sirvió como impulso para proponer la forma de enlazar los datos actualmente esparcidos por la Web; con el objetivo de crear una Web inteligente no solo entendible por los seres humanos sino también por las máquinas. La capacidad y beneficios que ofrece este tipo de tecnología es enorme, lo que ha dado paso a nuevos proyectos que buscan aprovechar el uso de la misma, sin embargo, esta sigue siendo desconocida para un gran número de usuarios en la actualidad lo que provoca que no se explote el máximo potencial de la misma.

Los datos enlazados facilitan la generación de conocimiento (Quezada-Sarmiento, Enciso & Garbajosa, 2016; Barba-Guamán et al., 2016); a través de la explotación de los enlaces semánticos entre los datos, esto trae consigo beneficios como buscadores inteligentes, navegar entre recursos y conseguir información de alto valor (Berners-Lee, Hendler & Lassila, 2001).

De lo expuesto anteriormente se investigó y analizó las aplicaciones ya existentes con el objetivo de proponer nuevas alternativas para mejorar el consumo y visualización de datos enlazados en la Web. La importancia de este artículo, se basa en permitir a los usuarios explorar e interpretar datos enlazados de forma sencilla.

2. Estado del arte

A. Web Semántica. La Web 3.0 o Web Semántica, apareció del ingenio de Tim Berners-Lee, creador de la World Wide Web (Berners-Lee, Hendler & Lassila, 2001), su idea puede describirse como una manera de enlazar todos los datos disponibles en la Web, esto de tal manera de que el computador sea capaz de entender la información, así como una persona lo que permitirá realizar búsquedas más inteligentes en la Web y complementar los servicios que nacieron en la Web 2.0.

Unicode. - Este es un estándar universal, para la codificar caracteres de texto multilingüe para facilitar el intercambio de textos internacionales (Berners-Lee, Hendler & Lassila, 2001; Berners-Lee et al., 2007; Bizer, Heath & Berners-Lee, 2009).

URI.- Es un identificador único que se asigna a algún recurso disponible en la Web como páginas Web, documentos, correos electrónicos, etc. (Berners-Lee, Hendler & Lassila, 2001).

RDF+rdfschema.- RDF es un lenguaje creado para representar los numerosos recursos disponibles en la Web, este es vital para poder enlazar datos y así crear la Web Semántica (Quezada, Garbajosa & Enciso, 2016). RDFSschema le proporciona a RDF la capacidad especificar que los contenidos son parte de un vocabulario y una estructura definida (Quezada, Garbajosa & Enciso, 2016).

B. Datos enlazados. El concepto de Datos Enlazados está encaminado a la creación de vínculos para enlazar los datos en la Web que se encuentran estructurados por lenguaje de hipertexto o HTML (Berners-Lee et al., 2007).

En W3C (2009) se explica que los datos enlazados se basan en archivos de tipo RDF (Resource Description Framework). El fundador de la Web, Tim Berners-Lee (W3C, 2008), hace referencia de un conjunto de reglas necesarias para publicar datos libres y que puedan ser enlazados en la Web, estas son:

- El uso de URIs (Unifor Resource Identifier) para nombrar las cosas, esto proporciona a la información un enlace único y entendible por lenguaje natural, además de ayudar a prevenir ambigüedades al referirnos a información (Duerst & Suignard, 2005).
- El uso del protocolo HTTP junto con las URI de manera que el usuario pueda acceder con facilidad a la información, ya que las URI no son direcciones sino sólo identificadores de la información es



necesario el uso de HTTP para poder acceder a los datos.

- Proporcionar facilidad en la búsqueda de información, mediante el uso de estándares como RDF para representar la información mediante una estructura que se basa en grafos para la conexión de los datos y SPARQL (Protocol and RDF Query Language) como lenguaje de consulta para acceder a los datos contenidos en bases de datos semánticas (W3C, 2008).
- Enlazar URIs unas con otras de manera que se pueda acceder con facilidad a información relacionada, esto permite que los datos no se queden aislados, sino que puedan ser accedidos mediante referencias (Dbpedia Español, 2017; W3C, 2009).

C. RDF (Resource Description Framework). Es un estándar de la W3C (World Wide Web Consortium), utilizado para estructurar y representar recursos disponibles en la Web, su objetivo es el de brindar un formato en el que se pueda describir información que pueda ser procesada y entendida por el computador (W3C, 2009).

De acuerdo a Lapuente (2015) se describe el lenguaje RDF, en donde se puede destacar los múltiples beneficios que este proporciona a los datos enlazados, entre estos tenemos:

- Permite la representación de relaciones y sus conceptos por medio de tripletas.
- Permite el uso de distintos vocabularios para describir recursos.
- Puede ser representado en documentos XML a través de serialización.
- Permite el uso de diferentes lenguajes para la consulta de los datos almacenados en documentos RDF.

D. SPARQL. Protocol and RDF Query Language (SPARQL), es un lenguaje recomendado por la W3C en el año 2008, la creación del mismo nació con la idea de los Datos Enlazados ya que se necesitaba un método por el cual se tenga acceso hacia los datos contenidos en diferentes repositorios, al igual que se hace con Structured Query Language (SQL), para la extracción de datos contenidos en bases de datos relacionales, “SPARQL se puede utilizar para expresar consultas que permiten interrogar diversas fuentes de datos” (W3C, 2008).

E. SPARQL Endpoints. Definidos como servicios Web destinados a la recepción de peticiones para realizar una posterior devolución de datos, de esto nace la iniciativa para la creación de SPARQL Endpoints cuya funcionalidad es otorgar a los usuarios la facilidad de consumir los recursos existentes en los diferentes TripleStores disponibles en la Web. La manera de acceder a estos servicios es por medio de una URI otorgada por los administradores de los mismos.

EndPoint	Enlace
DBpedia	http://dbpedia.org/sparql
DBpedia español	http://es.dbpedia.org/sparql
DBpedia Latinoamérica	http://es-la.dbpedia.org/sparql
EEA (European Environment Agency) Semantic Data Service	http://semantic.eea.europa.es/sparql
OpenLink Virtuoso	http://demo.openlinksw.com/sparql/

Tabla 1. SPARQL Endpoints. Fuente: Elaboración propia.

3. Análisis de la problemática y propuesta de la solución

Con la llegada de los datos enlazados surgieron nuevas herramientas para el uso de la misma, de aquí surgen los repositorios de datos enlazados cuyo propósito es poner a disposición de los usuarios un sinnúmero

de datos de forma libre, los usuarios pueden acceder a estos datos por medio de servicios Web también llamados Endpoints, sin embargo, esta manera de explorar datos está diseñada para usuarios con conocimientos sobre este tipo de tecnología limitando su uso.

El propósito de este artículo es tratar la problemática actual e identificar las razones de esta, así mismo se realiza un análisis de trabajos relacionados que permitan brindar una solución óptima para resolver los problemas encontrados y como último punto se propone una metodología de desarrollo con la que se construirá el proyecto.

F. Problema actual

La Web Semántica y los datos enlazados trajeron consigo cambios significativos los cuales tienen por objetivo mejorar la forma en que accedemos a la información que nos ofrece la Web. Los datos enlazados a diferencia de la Web 2.0 nos permite navegar entre recursos, términos y conceptos; convirtiéndose en una tecnología innovadora que produce grandes cambios en la forma tradicional para acceder a recursos en la Internet, sin embargo, el uso de esta tecnología aún no está al alcance de cualquier usuario, limitando el crecimiento de la misma, es por esto que es necesaria la implementación de soluciones tecnológicas que permitan a los usuarios aprovechar estos beneficios. Las técnicas de visualización de datos permiten una mejor interpretación de los mismos, esto facilita que cualquier usuario pueda captar de múltiples formas la información que se encuentra contenida en un conjunto incontable de datos actualmente disponibles en la Web, al enfocar el problema sobre los datos enlazados permite identificar los beneficios que traería consigo aprovechar eficientemente esta tecnología, en donde el usuario pase de navegar entre páginas Web a recursos enlazados semánticamente.

G. Solución propuesta

Existen proyectos avanzados en esta área de consumo de datos semánticos, funcionalidades ya desarrolladas y de gran utilidad para los usuarios, sin embargo, aún es necesario darle una mayor facilidad para que puedan acceder a estos datos y con ello a información de calidad e interés, por esta razón se propone el desarrollo de una aplicación capaz de proporcionar un buscador que le permita al usuario tener acceso de forma sencilla a la información almacenada en cualquier repositorio de datos enlazados disponibles en la Web, además de poder percibir a través de gráficos que resulten comprensibles e intuitivos.

Entre las características fundamentales que aporta la aplicación:

Buscador: La función de este es ofrecer un buscador con múltiples herramientas para mejorar la precisión de la información que se desea obtener.

- Consumo de datos enlazados: Se consumirá los recursos disponibles en una base de datos semántica.
- Visualización de datos: El usuario tendrá a su disposición una manera fácil e intuitiva de navegar entre datos enlazados.

4. Desarrollo de la solución

Las metodologías de desarrollo de software aportan una estructura ordenada para la construcción del mismo, RUP establece pautas a seguir contenidas en fases, en donde cada etapa define un entregable del producto con el objetivo de producir una línea base a seguir en todo el proceso de elaboración del sistema.

H. Arquitectura de la aplicación

La aplicación se ha desarrollado bajo una arquitectura en tres capas.



- Capa de Presentación: Esta capa alberga la interfaz de usuario, usa tecnologías como HTML5, JQuery y CSS3.
- Capa de Negocios: Contiene la lógica de la aplicación recibiendo los parámetros de la capa y trabaja con el lenguaje de PHP.
- Capa de Datos: Alberga las conexiones con la base de datos y funciones para extraer datos.

I. Requerimientos de interesados

Entre los requerimientos para mejorar la visualización de datos enlazados se encontraron los siguientes:

Requerimiento	Prioridad	Situación actual
Diseñar un buscador de términos sobre los diferentes Endpoints disponibles.	Alta	Realizar consultas SPARQL directamente sobre Endpoints
Disponibilidad de filtros para mejorar los resultados de búsqueda	Alta	N/A
Visualizar los datos por medio de gráficos	Alta	Las consultas arrojan tablas con los datos o en otros formatos RDF
Administrar las configuraciones básicas del sistema	Media	N/A

Tabla 2. Requerimientos. Fuente: Elaboración propia.

Como resultado se identificaron 4 componentes generales dentro del desarrollo del software.

- Diseño de buscador.
- Filtros de búsqueda.
- Visualización de resultados.
- Administración del sistema.

5. Construcción

El objetivo es desarrollar el aplicativo Web que busca satisfacer la problemática antes mencionada en el apartado 4, así como también cubrir los requerimientos funcionales y no funcionales identificados. Por esta razón es fundamental detallar el proceso de codificación y las diferentes herramientas que se han utilizado (Ver tabla 3), asociadas a cada capa de la arquitectura y obtener como resultado el buscador semántico tal como lo muestra la figura 1, donde se visualiza el resultado de una búsqueda como es los grafos, los filtros, categorías, etc.

Herramienta/tecnología	Descripción
Html5, CSS3, Javascript	Para el frontend
PHP	Como lenguaje de programación
MYSQL	Como motor de base de datos para almacenamiento de datos de configuración de la aplicación
Apache	Servidor web
Dbpedia	Endpoints donde consultar los datos semánticos
Json	Como medio de comunicación de las consultas Sparql y el frontend

Tabla 3. Herramientas utilizadas. Fuente: Elaboración propia.



Figura 1. Interfaz del buscador semántico. Fuente: Elaboración propia.

J. Filtros, paginación y resultados

Después de la recepción de resultados de la búsqueda, la aplicación arroja parámetros como lo muestra la figura 2, que serán de gran utilidad para el usuario como:

- Filtros: Agrupación de resultados según su tipo y categoría a la que pertenecen en torno al parámetro de la búsqueda (Barba-Guamán et al., 2016).
- Paginación: Visualización de 5 resultados por página.
- Resultados: Lista de resultados obtenidos al realizar una búsqueda.

En la figura 2 se muestra la interfaz del buscador semántico con filtros así como resultados generados, después de realizar diferentes pruebas al software (Diez, 2010).



Figura 2. Interfaz del buscador semántico con filtros y resultados. Fuente: Elaboración propia.

K. Configuración de parámetros de búsqueda

Dentro de la solución implementada se establece diferentes parámetros de configuración (Ver figura 3), esto con el fin garantizar la mayor accesibilidad de la aplicación, las cuales se listan a continuación:

- Lenguaje: El usuario podrá elegir el idioma en el que desea obtener resultados.
- Buscar sobre: Al ser un buscador por coincidencia de palabras el usuario puede decidir sobre qué etiquetas orientar sus búsquedas.

- Endpoints: La fuente de datos podrá ser cambiada para realizar búsquedas sobre una fuente en particular, así como también la opción de agregar una propia fuente de datos por medio de la dirección URL de un Endpoint.
- Búsqueda Exacta: Esta opción le permite al usuario realizar búsquedas que contengan como parámetro el país de donde se conectan.

Figura 3. Parámetros de configuración, idioma, buscar sobre.

La interfaz de administración del sistema está diseñada para que el usuario tenga control total sobre los datos contenidos en las diferentes tablas que componen el sistema (Liu et al., 2015; Aguilar, et al., 2017).

La información almacenada en este módulo del sistema; se ve reflejada en la configuración del buscador (ver figura 4), en donde, estos datos resultan de utilidad para el usuario, brindándole la opción de establecer los parámetros necesarios, para realizar una búsqueda como son:

- La fuente de datos.
- La etiqueta sobre la cual se realizará la búsqueda.
- El idioma de búsqueda.

id	prefijo	url	Borrar	Editar
1	rdfs:type	http://www.w3.org/2000/01/rdf-schema#type	X	/
2	rdfs:comment	http://www.w3.org/2000/01/rdf-schema#comment	X	/
3	foaf:name	http://xmlns.com/foaf/0.1/name	X	/

Figure 4. Parámetros de configuración de prefijos.

Además se debe incluir la gestión de usuarios del sistema, aquellas personas que podrán acceder a este componente para la gestión de la base de datos del aplicativo.

6. Conclusiones

La disponibilidad de los puntos de acceso hacia los diferentes repositorios de datos enlazados encontrados en la Web, ofrece la oportunidad de crear nuevas propuestas en beneficio del crecimiento de la Web Semántica y los datos abiertos, para que estos a su vez puedan ser aprovechados por los usuarios.

El análisis e investigación de trabajos relacionados permitió conocer varias propuestas basadas en mejorar el consumo de datos enlazados, sin embargo, estas requieren cierto nivel de conocimientos técnicos, lo que brindó la oportunidad de proponer una solución innovadora para mejorar el consumo y visualización de datos

enlazados en la Web.

La representación visual (grafos de navegación) de los resultados que arrojan las consultas SPARQL de los diferentes endpoints ayudan a entender de mejor manera y de forma clara la interconexión de los datos y el contexto de lo que representa linked data, esto con el fin de obtener resultados claros y precisos de lo que el internauta está buscando.

Cómo citar este artículo / How to cite this paper

Ramirez-Coronel, R. L.; Quezada-Sarmiento, P. A. (2017). Consumo y visualización de datos enlazados mediante aplicaciones web de código libre. *International Journal of Information Systems and Software Engineering for Big Companies (IJISEBC)*, 4(2), 15-22. (www.ijisebc.com)

Referencias

- Aguilar, J.; Sánchez, M.; Cordero, J.; Valdiviezo-Díaz, P.; Barba-Guamán, L.; Chamba-Eras, L. (2017). Learning analytics tasks as services in smart classrooms. *Universal Access in the Information Society*.
- Barba-Guamán, L.; Quezada-Sarmiento, P.; Calderon-Cordova, C.; López, J. (2016). Detection of the characters from the license plates by cascade classifiers method. *Future Technologies Conference (FTC)*.
- Berners-Lee, T.; Hendler, J.; Lassila, O. (2001). *The Semantic Web*. Scientific American.
- Berners-Lee, T.; Hollenbach, J.; Lu, K.; Presbrey, J.; Pru d'ommeaux, E.; Schraefel, M. (2007). *Tabulator Redux: Writing Into the Semantic Web*. (<http://eprints.soton.ac.uk/264773/>)
- Bizer, C.; Heath, T.; Berners-Lee, T. (2009). Linked data-the story so far. *International journal on Semantic Web and Information Systems*. (<https://doi.org/10.4018/jswis.2009081901>)
- Brunetti, J.; Auer, S.; García, R. (2012). The Linked Data Visualization Model. *International Semantic Web*. (http://svn.aksw.org/papers/2013/WWW_LDVM/public.pdf)
- DBpedia (2015). ES dbpedia: III Jornadas esDBpedia. (http://es.dbpedia.org/wiki/Wiki.jsp?page=III_Jornadas_esDBpedia)
- Dbpedia Español. (<http://dbpedia.org/sparql>)
- Diez, E. (2010). Tipos De Pruebas Dinámicas. (<http://sistemas.unla.edu.ar/sistemas/sls/ls-4-optativa-algoritmos-y-lenguajes-prueba-del-software/pdf/Pruebas-de-Software-C07-Tipos-de-pruebas-dinamicas.pdf>)
- Duerst, M.; Suignard, M. (2005). Internationalized Resource Identifiers (IRIs). (<http://www.ietf.org/rfc/rfc3987.txt>)
- Heather, T.; Bizer, C. (2011). Linked Data - Evolving the Web into Global Data Space. *Www-Kasm.Nii.Ac.Jp* (Vol. 5). (<https://doi.org/10.2200/S00334ED1V01Y201102WBE001>)
- Hildebrand, M.; Ossenbruggen, J.; Hardman, L. (2006). /facet: A browser for heterogeneous semantic web repositories. *ISWC 2006 - International Semantic Web Conference*. (https://doi.org/10.1007/11926078_20)
- Ian, D.; Steiner, T.; Arnaud, J. (2013). RDF 1.1 JSON Alternate Serialization (RDF/JSON). (<http://www.w3.org/TR/rdf-json/>)
- INEC. (2010). Fascículo provincial loja. (<http://www.ecuadorencifras.gob.ec/wp-content/descargas/Manu-lateral/Resultados-provinciales/loja.pdf>)
- Lapuente, M. (2015). RDF. (<http://www.hipertexto.info/documentos/rdf.htm>)
- Liu, D.; Valdiviezo-Díaz, P.; Riofrio, G.; Sun, Y.; Barba, R. (2015). Integration of Virtual Labs into Science E-learning. In *Procedia Computer Science*.
- OpenLink Virtuoso. (<http://demo.openlinksw.com/sparql/>)
- Quezada, P.; Garbajosa, J.; Enciso, L. (2016). Use of standard and model based on BOK to evaluate professional and occupational profiles *Advances in Intelligent Systems and Computing*.
- Quezada-Sarmiento, P.; Enciso, L.; Garbajosa, J. (2016). Use of body knowledge and cloud computing tools to develop software projects based in innovation. *IEEE Global Engineering Education Conference (EDUCON)*.
- Quezada-Sarmiento, P.; Enciso-Quispe, L.; Garbajosa, J.; Washizaki, H. (2016). Curricular design based in bodies of knowledge: Engineering education for the innovation and the industry. *SAI Computing Conference (SAI)*.
- W3C (2008). SPARQL Query Language for RDF. (<http://www.w3.org/TR/rdf-sparql-query/#introduction>)
- W3C (2009). Guía Breve de Linked Data. (<http://www.w3c.es/Divulgacion/GuiasBreves/LinkedData>)

