

Tipo de artículo: Artículo original  
Temática: Desarrollo de aplicaciones informáticas  
Recibido: 4/05/2016 | Aceptado: 1/07/2016

## Indización de grafos RDF desde un SPARQL Endpoint

### *Indexing RDF graphs from a SPARQL Endpoint*

Alejandro Jesús Mariño Molerio<sup>1\*</sup>, Juan Carlos Moreira De Lara<sup>1</sup>, Yusniel Hidalgo Delgado<sup>2</sup>, Ernesto Ortiz Muñoz<sup>2</sup>

<sup>1</sup>Universidad de las Ciencias Informáticas, Carretera a San Antonio de los Baños, km 2 ½, Boyeros, La Habana, Cuba. {[jcmoreira](mailto:jcmoreira@estudiantes.uci.cu), [ajmarino@estudiantes.uci.cu](mailto:ajmarino@estudiantes.uci.cu)}

<sup>2</sup>Universidad de las Ciencias Informáticas, Carretera a San Antonio de los Baños, km 2 ½, Boyeros, La Habana, Cuba. {[yhdelgado](mailto:yhdelgado@uci.cu), [ernesto@uci.cu](mailto:ernesto@uci.cu)}

\*Autor para la correspondencia: [yhdelgado@uci.cu](mailto:yhdelgado@uci.cu)

---

#### Resumen

Las bibliotecas han tenido un rol importante en la sociedad. Actualmente, las bibliotecas se agrupan en tres grandes categorías: convencionales, digitales e híbridas. En el caso de las bibliotecas digitales, éstas coleccionan, almacenan y distribuyen la información en soporte digital. Por otro lado, la aplicación de los principios para la publicación de contenido en la Web, ha permitido una mejora en el procesamiento y gestión de la información. En este contexto, los datos enlazados se han convertido en un área de investigación activa en los últimos años. En el grupo de investigación de web semántica de la Universidad de las Ciencias Informáticas se desarrolla un proyecto de investigación que genera grafos RDF con metadatos bibliográficos que son almacenados para su posterior uso en un almacén de tripletas (del inglés, *triplestore*). Estos almacenes no se encuentran optimizados para la realización en tiempo cercanos al real de consultas formuladas por los usuarios, por lo que se hace necesario contar con estructuras de datos optimizadas (índices) para disminuir dicho tiempo y resolver los problemas de búsqueda y recuperación de datos en grandes bases de datos. En este artículo se propone un componente de software que integra un servidor de indización con tripletas RDF existentes en un almacén de tripletas.

**Palabras clave:** biblioteca digital; índice; grafo RDF; metadato bibliográfico; web semántica.

## **Abstract**

*Libraries have played an important role in the society. Nowadays, the libraries are grouped into three broad categories: conventional, digital and hybrid. In the case of digital libraries, they collect, store and distribute digital information. The application of the principles for publishing content on the Web, has allowed an improvement in processing and information management. In this sense, the linked data have become an active research area in recent years. The semantic web research group at University of Informatics Sciences, is developing a research project that generates RDF graphs with bibliographic metadata that is stored for later use in a triplestore. However, these triplestores are not optimized for executing real time queries submitted by users, so it is necessary to have data structures optimized (index) to decrease the query time and solve the issues related to search and retrieval in large databases. In this paper, we propose a software component that integrates an indexing server with existing RDF triples in a triplestore.*

**Keywords:** *digital library; index; RDF graph; bibliographic metadata; semantic web.*

---

## **Introducción**

Las bibliotecas siempre han tenido un valor preponderante en la sociedad, en unas culturas más que en otras, lo cual se debe a que éstas atesoran el conocimiento científico de la humanidad. Sus funciones, según indica (Shanhong 2000) son reunir, procesar, difundir, almacenar y usar la información documental para dar servicio a la sociedad. El surgimiento de Internet y el desarrollo de las *Tecnologías de la Información y las Comunicaciones* (TIC, por sus siglas en español) revolucionaron la forma en que las bibliotecas llevaban a cabo su actividad fundamental; transformando su organización en profundidad como expone (López y Perelló 2005) y articulándose con nuevas perspectivas de servicio como es el caso de: la teledocumentación, la difusión selectiva, el acceso en línea a bases de datos documentales y científicas, los servicios de Internet y correo electrónico, así como el incremento de documentos en soportes no bibliotecarios con sus correspondientes servicios (fonotecas, fototecas, cartotecas, videotecas, entre otros). Según refieren (Singh y Sharma 2015) en el escenario actual las bibliotecas se agrupan en tres grandes categorías: convencionales, digitales e híbridas. En el caso particular de las bibliotecas digitales, estas coleccionan, almacenan y distribuyen la información en soporte digital.

En este contexto, se ha incrementado la atención de la sociedad por la información y el conocimiento, específicamente el conocimiento científico, debido a que ambos, se han convertido en la fuerza conductora para el

desarrollo social. Como tendencia, los científicos y los académicos publican los resultados de sus investigaciones en revistas científicas. En muchos casos estos resultados son almacenados en los repositorios de las instituciones a las que pertenecen o en bases de datos relacionales, donde resulta complejo el acceso al conocimiento científico y tecnológico.

Debido a la gran diversidad y volumen de las fuentes y recursos en Internet, se hizo necesario establecer un mecanismo para etiquetar, catalogar, describir y clasificar el gran volumen de recursos presentes en la World Wide Web con el objetivo de facilitar la búsqueda y recuperación de la información. Este mecanismo lo constituyen los llamados metadatos.

La aplicación de los principios de los datos enlazados para la publicación de contenido en la Web, indican (Rizo y García 2013), ha permitido una considerable mejora en el procesamiento y gestión de la información tanto para humanos como para las máquinas (agentes software). Los datos enlazados se han convertido en un área de investigación activa en los últimos años, específicamente ha cobrado importancia la publicación de metadatos bibliográficos, según alguna de las serializaciones del formato RDF. Esto se debe a la necesidad de publicar y consumir datos estructurados en la Web, potenciando de esta manera el desarrollo de la Web Semántica. Sin embargo, está vigente una barrera técnica en relación con el uso de datos enlazados para los usuarios que no están familiarizados con las tecnologías de la Web Semántica, principalmente por la necesidad de tener que realizar consultas semánticas complejas sobre grafos RDF.

El grupo de investigación de web semántica de la *Universidad de las Ciencias Informáticas* (UCI, por sus siglas en español) está ejecutando el proyecto de investigación: “Extracción, publicación y consumo de metadatos bibliográficos como datos enlazados”. En este proyecto se generan grafos RDF con metadatos bibliográficos que son almacenados en un *triplestore* para su uso posterior. Los *triplestores* son bases de datos que actualmente no se encuentran optimizadas para la realización en tiempo cercanos al real de consultas formuladas por los usuarios.

En el contexto del proyecto de investigación fue implementada una aplicación web para el consumo de datos enlazados a partir de un triplestore (RACien), sin embargo, los resultados en cuanto al tiempo de respuesta no fueron los esperados. En este sentido, se hace necesario contar con índices para disminuir el tiempo de respuesta del *triplestore* a consultas formuladas por los usuarios.

Los índices son estructuras de datos optimizadas que permiten transformar el texto en un formato donde la búsqueda sea más rápida, eliminando el proceso de exploración lento a consultas formuladas por los usuarios. Este proceso de conversión es llamado indización y es necesario para un adecuado almacenamiento y recuperación de la

información. Para lograr una correcta indización es importante utilizar técnicas de indización existentes. Los servidores de indización, por su parte, constituyen una solución eficiente para resolver los problemas de búsqueda y recuperación de datos en grandes bases de datos; problemas para los que las bibliotecas digitales trabajan con el objetivo de elevar su eficiencia.

La *World Wide Web* (WWW, por sus siglas en inglés), se ha convertido en un instrumento de uso cotidiano en nuestra sociedad; es hoy un medio flexible y económico para la comunicación, el comercio y los negocios, ocio y entretenimiento, acceso a información y servicios, difusión de cultura, entre muchos otros escenarios, según refieren (Chávez, Cárdenas y Benito 2005). Paralelamente al crecimiento de la web y las tecnologías que la hacen posible han experimentado una rápida evolución.

Según (Tello 2006), la información se encuentra almacenada en la Web mediante lenguajes de etiquetado, tales como el lenguaje HTML (que está orientado el texto y no al dato) y que únicamente describe la forma en que dicha información debe ser presentada al usuario por el navegador. Sin embargo, este lenguaje carece de etiquetas que permitan representar explícitamente la semántica o significado subyacente de la información publicada (Dadzie y Rowe 2011). Por otra parte, la información en la Web se encuentra dispersa y no existe relación explícita entre los diferentes recursos, lo que provoca ambigüedad en la información, exponen (Dadzie y Rowe 2011), mientras que imposibilita su descubrimiento y utilización por sistemas informáticos. Lo anterior influye en que los motores de búsquedas más utilizados en la actualidad como Google y Yahoo resultan imprecisos y, en muchos casos, no satisfacen las necesidades de búsqueda de los usuarios al responder consultas basadas en palabras claves, no siendo capaces de recuperar la información a partir de consultas expresadas en lenguaje natural (Sudeepthi, Anuradha y Babu 2012).

La Web 3.0 o web de los datos, como también es reconocida, intenta resolver las limitaciones antes expuestas que impiden aprovechar todas sus potencialidades, dígame: formato, integración y recuperación de la información, como explican (Delgado y Puente 2013). Su objetivo es catalogar la información de los recursos web, páginas HTML, documentos *Portable Document Format* (PDF, por sus siglas en inglés), videos, archivos de sonido y otros, mediante ontologías.

Para (Chávez, Cárdenas y Benito 2005) y (Tello 2006), la Web Semántica se basa dos conceptos fundamentales: (1) la descripción del significado que tiene los contenidos en la web y (2) la manipulación automática de estos significados. En el primer caso intervienen los conceptos como: *la Semántica*, que es el estudio y significado de los términos lingüísticos procesables por máquinas; *los Metadatos* como contenedores de información semántica sobre

los datos; y *las Ontologías* para definir conceptos y relaciones de un dominio específico. En el segundo caso, entonces se estaría hablando de: *la Lógica Matemática* (permite establecer reglas para tratar el contenido semántico) y *los Motores de Inferencia* (permiten combinar conocimientos conocidos para elaborar otros nuevos conocimientos). En general para que la comunicación entre las personas y las máquinas funcione a nivel semántico se necesitan tres tecnologías claves: un modelo de datos común (RDF) para leer y escribir en el mismo idioma; un lenguaje de consultas para ese modelo de datos (SPARQL) que permita consultar la información; y una lógica que opera sobre esos mismos datos (OWL) para poder razonar sobre ellos.

## **Materiales y métodos o Metodología computacional**

Se explora, aplicando el método *analítico sintético* las principales aproximaciones existentes en la literatura para resolver el problema de disminución del tiempo de respuesta a las consultas formuladas por los usuarios durante el consumo de metadatos bibliográficos publicados como datos enlazados, de manera que sea posible estudiar en profundidad cada componente, técnica o tecnología involucrada y realizar además una síntesis de los elementos comunes y más importantes de cada teoría y/o aproximación.

Se aplica el método *histórico lógico* para caracterizar la evolución del área del conocimiento durante los últimos años, haciendo particular énfasis en los principales conceptos que forman parte de esta investigación. Para la fundamentación y desarrollo de la propuesta de solución se aplican los métodos de modelación e inductivo deductivo. En el primer caso, se formula una generalización de propuesta arquitectónica utilizando técnicas de visualización de la información, así como algunos artefactos de ingeniería que ayuden a comprender sus componentes y sus interrelaciones.

Como metodología de desarrollo se propone el uso de AUP-UCI, en su escenario cuatro; puesto que se ajusta para proyectos que no modelan un negocio sino modelan el sistema con las historias de usuario, que es justamente el caso que ocupa y el que adopta el grupo de investigación. Por otra parte, se decide el uso de Groovy (v.2.5) por ser lenguaje dinámico para el desarrollo ágil de aplicaciones web basadas en la plataforma Java y ofrecer soporte nativo a lenguajes de marcado como XML y HTML, mientras que es empleado Grails (v.2.5.3) como marco de trabajo altamente productivo desarrollado sobre el lenguaje de programación Groovy; que reutiliza tecnologías Java ya probadas como Hibernate y Spring bajo una interfaz simple y consistente, siendo capaz de extender sus funcionalidades al ser su arquitectura basada en *plugins*.

Se emplean además: SPARQL (v.1.1) como lenguaje de consultas para RDF, Apache Jena Fuseki (v.2.0) como almacén de tripletas RDF, Git como sistema de control de versiones, IntelliJ IDEA (v.14.0.1) como entorno de trabajo y Elasticsearch (v.2.1) como motor de búsqueda.

## Trabajos relacionados

Entre las principales aproximaciones similares existentes en la literatura fueron estudiadas las técnicas y herramientas de indización. Dentro de las técnicas de indización analizadas se encuentran:

**Inverted Index:** un índice invertido (también denominado archivo de publicaciones o archivo invertido) es una estructura de datos de índice de almacenamiento de un mapeo de contenidos, tales como palabras o números, a sus ubicaciones en un archivo de base de datos o en un documento o un conjunto de documentos. El propósito de éste es permitir búsquedas rápidas de texto completo, a un costo de aumento del procesamiento cuando se añade un documento a la base de datos. Los índices invertidos son la estructura de datos más popular utilizada en sistemas de recuperación de documentos. Existen dos variantes principales de índices invertidos: (1) un índice de nivel de grabación invertida, que contiene una lista de referencias a los documentos para cada palabra y (2) un índice invertido a nivel de palabra, que contiene además, las posiciones de cada palabra dentro de un documento. La estructura de datos de índice invertido es un componente central de un algoritmo típico de indexación de los motores de búsqueda.

**Suffix tree:** un árbol de sufijos (también llamado árbol de PAT o, en una forma anterior, árbol de posición) es un trie (estructura de datos en tipo de árbol) comprimido que contiene todos los sufijos del texto dado como sus claves y posiciones en el texto como sus valores. Éstos implementan eficientemente muchas operaciones de cadenas importantes y proporcionan una de las primeras soluciones en tiempo lineal para el problema de la subcadena común más larga. Estas aceleraciones tienen un costo: el almacenamiento de un árbol así normalmente requiere mucho más espacio que el almacenamiento de la propia cadena. Los árboles de sufijos se pueden utilizar para resolver un gran número de problemas de la secuencia que se producen en edición de textos, búsqueda de texto libre, biología computacional y otras áreas de aplicación.

**Document-term matrix:** es una matriz matemática que describe la frecuencia de los términos que se producen en una colección de documentos, siendo útiles en el campo del procesamiento del lenguaje natural. Al crear una base de datos de términos que aparecen en un conjunto de documentos, la matriz contiene filas correspondientes a los documentos y columnas correspondientes a los términos. El análisis semántico latente puede mejorar los resultados de búsqueda por palabras polisémicas, eliminar la ambigüedad y la búsqueda de sinónimos de la consulta. Sin embargo,

buscando en el espacio continuo de alta dimensión es mucho más lenta que la búsqueda de trie estándar de los motores de búsqueda.

Como parte de las herramientas de indización examinadas destacan:

**Solr:** el plugin Solr Grails integra un modelo de dominio Grails con el motor de búsqueda de Apache Solr a través de la API SolrJ. Un índice en Solr posibilita llevar a cabo de manera óptima: la búsqueda de texto completo, agregados y filtrado. Solr acepta documentos JSON, pudiendo transformar su documento RDF en un documento JSON-LD (que es un formato de serialización RDF). Por otra parte SolrRDF (entiéndase Solr + RDF) es un conjunto de extensiones para la gestión de Solr (índice y búsqueda) de datos RDF siendo posible indizar triples al clúster y realizar consultas SPARQL (ASK, CONSTRUCT, SELECT y DESCRIBE) y actualizaciones (como INSERT y DELETE) a cualquier nodo del clúster obteniendo como respuesta un XML, siendo compatible con SPARQL Endpoint 1.1.

**ElasticSeach:** cuenta con un plugin (ElasticSearch Grails Plugin 0.0.4.6) que pretende implementar una integración simple de Grails con el motor de búsqueda de código abierto Elasticsearch, que está basado en Lucene y proporciona capacidades distribuidas. El *plugin* indexa automáticamente todos los dominios de búsqueda cuando Grails Object Relational Mapping (GORM, por sus siglas en inglés) / Hibernate hace una reserva o una actualización en la base de datos. También elimina automáticamente del índice cualquier documento que corresponde a un dominio que se elimina de la base de datos. Por otra parte, EEA RDF River Plugin es también un plugin para ElasticSearch (que utiliza como técnica la de índice invertido) que permite cosechar los metadatos de los criterios de valoración de SPARQL o archivos RDF en Elasticsearch. Entre sus ventajas: indiza grafos RDF proporcionando sus URIs, así como triples recuperados de un SPARQL endpoint; facilidad de un índice personalizable y tipos de nombres; normalización de las propiedades de diferentes espacios de nombres, así como la sincronización con un SPARQL Endpoint.

**Jena ARQ:** es un motor de búsqueda de Jena que soporta el lenguaje de consultas SPARQL. En Jena, toda la información de estado proporcionada por un conjunto de tripletas RDF está contenida en una estructura de datos llamada modelo. El modelo representa un grafo RDF, llamado así porque contiene una colección de nodos RDF, unidos entre sí por relaciones marcadas. Entre las capacidades de Jena ARQ destacan: búsqueda de texto libre a través de Lucene; actualización, acceso y la extensión del álgebra de SPARQL; apoyo a las funciones de filtro personalizados; funciones de propiedad para un tratamiento personalizado de relaciones semánticas; y apoyo al cliente para el acceso remoto a cualquier SPARQL Endpoint.

**Apache Lucene:** es el principal motor de búsqueda de código abierto y se utiliza en muchas empresas, proyectos y productos, originalmente implementada en Java. Es útil para cualquier aplicación que requiera indexado y búsqueda a

texto completo. En esencia, el índice se compone de documentos que se componen de campos. Las consultas de Lucene tienen que pasar a través de los mismos analizadores que se utilizaron durante la indexación - de lo contrario términos idénticos podría no coincidir. Se puede usar Lucene en Grails, como es el caso de los plugins para la integración de Solr y Elasticsearch con Grails.

## **Resultados y discusión**

El proyecto “Extracción, publicación y consumo de metadatos bibliográficos como datos enlazados” tiene como objetivo construir una Biblioteca Digital Semántica. Esta iniciativa es única de su tipo en Cuba, no habiendo dónde coleccionar las contribuciones del quehacer académico y científico. El proyecto cuenta con cuatro actividades fundamentales, que según (Delgado 2015) quedarían definidas de la siguiente manera: (1) Extracción de metadatos, (2) Pre-procesamiento, (3) Publicación y finalmente (4) Consumo. Es en esta última etapa donde se enmarca la presente investigación.

Primeramente, se extraen y almacenan los metadatos provenientes de fuentes de datos heterogéneas en un repositorio de metadatos para luego ser analizados con el fin de limpiar y normalizar los mismos, logrando de esta forma mayor calidad en los datos. Luego, se determinan las ontologías a utilizar para el modelado de datos de la biblioteca. Con el modelo ontológico resultante y los datos limpios y normalizados se realizan tres tareas fundamentales: transformación, enlazado y publicación. Como resultado de estas actividades se generan grafos RDF que son almacenados en un almacén de tripletas RDF para ser manipulados en la etapa de consumo empleando además un servidor de indexación. El consumo de datos incluye: la realización de búsquedas textual y facetada y la realización análisis estadísticos.

La propuesta de solución implementa el componente `sdl-index` con el propósito de llevar a cabo tres tareas: (1) obtener el grafo RDF desde un SPARQL Endpoint; (2) sincronizar el grafo RDF desde un SPARQL Endpoint con el índice en el motor de búsqueda Elasticsearch y (3) realizar búsquedas textuales y facetadas en SPARQL sobre el índice de Elasticsearch. Se pretende implementar un prototipo funcional donde tiene lugar el componente `sdl-consumer` con el fin de proveer una interfaz web para la búsqueda textual y facetada sobre los datos en el índice. Ver Figura 1.

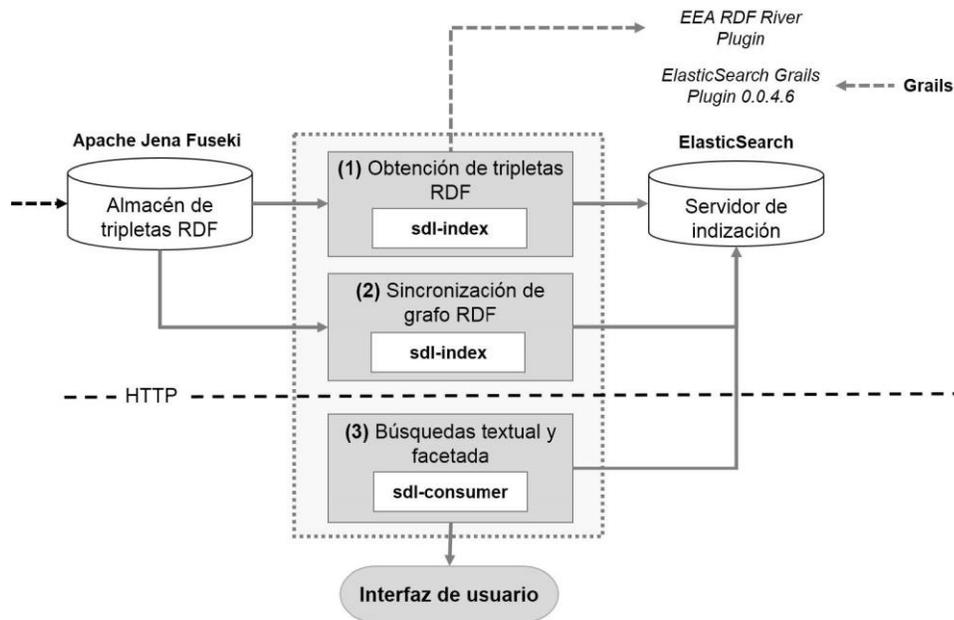


Figura 1. Estilo y arquitectura del componente propuesto (Flujo de datos y Tuberías – filtros respectivamente).

El componente *sdl-index* como propuesta de solución sigue un estilo arquitectónico de flujo de datos, que se adecua para sistemas que implementan transformaciones de datos en pasos sucesivos. La arquitectura utilizada es de tuberías y filtros, que consiste en ir transformando un flujo de datos en un proceso comprendido por varias fases secuenciales, siendo la entrada de cada una la salida de la anterior. Una tubería (del inglés, pipeline) conecta componentes computacionales (filtros) a través de conectores (del inglés, pipes), de modo que las computaciones se ejecutan a manera de un flujo.

La arquitectura presentada, contempla dos elementos fundamentales: el almacén de tripletes Apache Jena Fuseki (como punto de acceso a consultas SPARQL) y el servidor de indización ElasticSearch (como motor de búsqueda). La interacción entre ambos se debe a la necesidad de encontrar una alternativa que resuelva el inconveniente del tiempo de respuesta que tardan las consultas realizadas por los usuarios.

Apache Jena Fuseki almacena los grafos RDF generados *a priori*. Gracias a las facilidades de ElasticSearch para su integración con *plugins*, es posible mediante EEA RDF River Plugin, obtener dichos grafos RDF para indizarlos en el servidor de indización. Estos grafos una vez incorporados a ElasticSearch se encuentran en formato JSON. La tarea de obtención de grafos RDF se acomete en primera instancia, ya que seguidamente basta con sincronizar los grafos RDF del almacén de tripletes con los del índice en el motor de búsqueda. El propósito de esta actividad es consultar el almacén de tripletes para verificar los últimos cambios e indizar sólo esos grafos RDF al índice del motor de

búsqueda de ElasticSearch. Este proceso es ventajoso en el sentido de que no se precisa el almacenamiento repetido de los grafos RDF existente en el almacén de tripletas.

La última de las tres actividades es el núcleo del componente, puesto que la realización de búsquedas textuales y facetadas como técnicas para acceder a la información organizada permite a los usuarios explorar una colección de información mediante la aplicación de varios filtros, lo cual contribuye a la disminución del tiempo para consultar dicha información (ver figura 2). Desde el motor de búsqueda es permitido realizar consultas SPARQL al almacén de tripletas, lo cual se debe a que el formato JSON respeta la estructura del grafo RDF. ElasticSearch Grails Plugin 0.0.4.6 permite una integración simple del marco de trabajo Grails con motor de búsqueda de código abierto ElasticSearch.

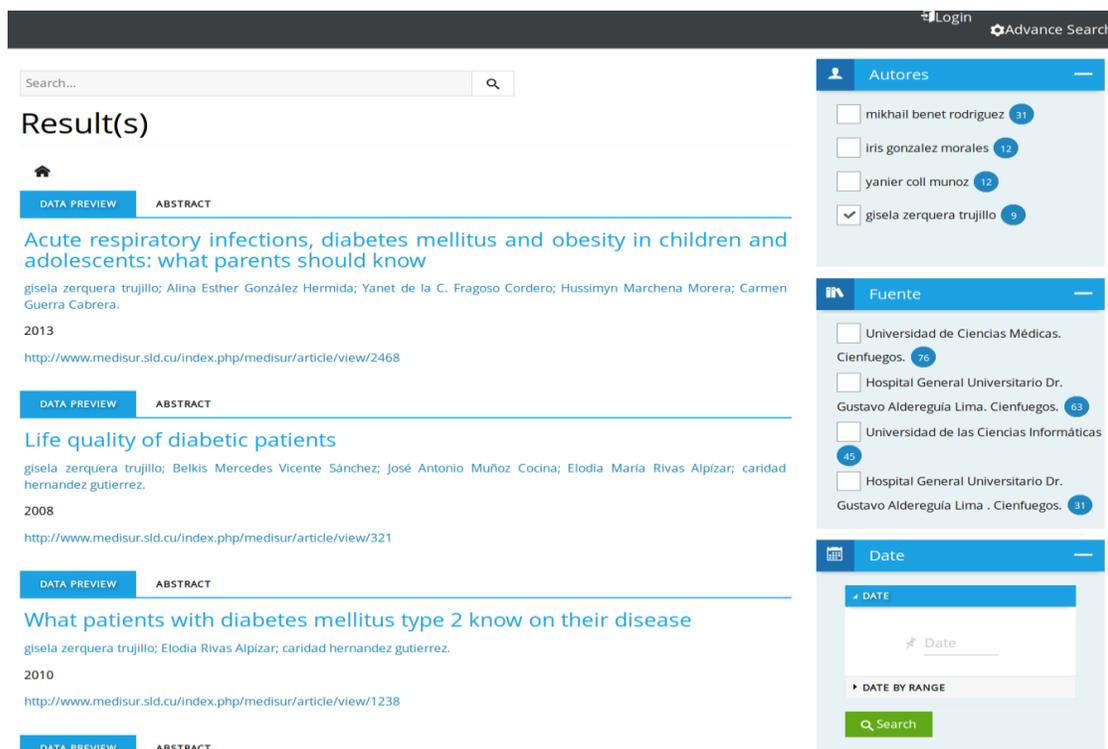


Figura 2. Interfaz gráfica con componentes visuales para realizar la búsqueda textual y facetada

## Conclusiones

Con el análisis de los principales conceptos relacionados con el objeto de estudio y su interrelación, se logró profundizar en la comprensión del problema planteado por la investigación. El estudio de las técnicas y herramientas a partir de la literatura científica consultada, permitió seleccionar la herramienta ElasticSearch por ser la que más se ajusta a los requerimientos de la investigación. Ésta herramienta proporciona un índice invertido como estructura de datos optimizada para el almacenamiento de grafos RDF. El componente *sdl-index* como propuesta de solución para la indización de grafos RDF a partir de un *triplestore*, está basado en tres fases fundamentales. Se implementa un prototipo funcional con el fin de proveer una interfaz web para la búsqueda textual y facetada sobre los datos en el índice.

## Referencias

- CHÁVEZ, M.E., CÁRDENAS, O. y BENITO, O., 2005. La web semántica. Revista de investigación de Sistemas e Informática, vol. 2, no. 3, pp. 43–54.
- DADZIE, A.-S. y ROWE, M., 2011. Approaches to visualising linked data: A survey. Semantic Web, vol. 2, no. 2, pp. 89–124.
- DELGADO, Y.H., 2015. Marco de trabajo basado en los datos enlazados para la interoperabilidad semántica en el protocolo OAI-PMH. Tesis de Maestría, Universidad de las Ciencias Informáticas, Cuba. [en línea], [Consulta: 29 marzo 2016]. Disponible en: <http://eprints.rclis.org/28755/>.
- DELGADO, Y.H. y PUENTE, R.R., 2013. La web semántica: una breve revisión. Revista Cubana de Ciencias Informáticas, vol. 7, no. 1, pp. 76-85. ISSN 2227-1899.
- LÓPEZ, P.L. y PERELLÓ, J.G., 2005. Información, conocimiento y bibliotecas en el marco de la globalización neoliberal [en línea]. S.l.: Trea. [Consulta: 4 febrero 2016]. Disponible en: [https://www.researchgate.net/profile/Rosa\\_San\\_Segundo/publication/33401520\\_Informacin\\_conocimiento\\_y\\_bibliotecas\\_en\\_el\\_marco\\_de\\_la\\_globalizacin\\_neoliberal\\_book\\_review/links/0912f50db41a9e11ec000000.pdf](https://www.researchgate.net/profile/Rosa_San_Segundo/publication/33401520_Informacin_conocimiento_y_bibliotecas_en_el_marco_de_la_globalizacin_neoliberal_book_review/links/0912f50db41a9e11ec000000.pdf).
- RIZO, C.H. y GARCÍA, C.H.C., 2013. Consumo de datos enlazados mediante búsqueda textual y facetada. Tesis de Grado, Universidad de las Ciencias Informáticas, Cuba.
- SHANHONG, T., 2000. Gestión del conocimiento en las bibliotecas del siglo XXI. 66th IFLA Council and General Conference [en línea]. S.l.: s.n., pp. 13–18. [Consulta: 4 febrero 2016]. Disponible en:

<http://www.hacienda.go.cr/centro/datos/Articulo/Gesti%C3%B3n%20del%20conocimiento%20en%20las%20Bibliotecas%20del%20siglo%20XXI.doc>.

- SINGH, T. y SHARMA, A., 2015. Research work and changing dimensions of digital library: A review. *Emerging Trends and Technologies in Libraries and Information Services (ETTLIS)*, 2015 4th International Symposium on [en línea]. S.l.: IEEE, pp. 39–42. [Consulta: 4 febrero 2016]. Disponible en: [http://ieeexplore.ieee.org/xpls/abs\\_all.jsp?arnumber=7048169](http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=7048169).
- SUDEEPTHI, G., ANURADHA, G. y BABU, M.S.P., 2012. A survey on semantic web search engine. *IJCSI International Journal of Computer Science Issues* [en línea], vol. 9, no. 2. [Consulta: 4 febrero 2016]. Disponible en: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.403.482&rep=rep1&type=pdf>.
- TELLO, J.C., 2006. La Web Semántica y el lenguaje RDF. [en línea], [Consulta: 4 febrero 2016]. Disponible en: [https://portal.uah.es/portal/page/portal/GP\\_EPD/PG-MA-PROF/OLD\\_PG-PROF-138886%202008-07-14%2010-07-31/TAB4348465/TAB4348469/TAB4348477/Articulo\\_WebSemantica\\_Jesus\\_Caceres\\_CISTI\\_06.pdf](https://portal.uah.es/portal/page/portal/GP_EPD/PG-MA-PROF/OLD_PG-PROF-138886%202008-07-14%2010-07-31/TAB4348465/TAB4348469/TAB4348477/Articulo_WebSemantica_Jesus_Caceres_CISTI_06.pdf).