

ESCUELA POLITÉCNICA NACIONAL

**INGENIERÍA EN SISTEMAS INFORMÁTICOS Y DE
COMPUTACIÓN**

**DESARROLLO DE SISTEMA OLAP PARA BIBLIOTECAS
DIGITALES QUE USAN TECNOLOGIAS LINKED DATA**

**PROYECTO PREVIO A LA OBTENCIÓN DEL TÍTULO INGENIERO EN
SISTEMAS INFORMÁTICOS Y DE COMPUTACION**

DARIO JAVIER TUBON MOPOSITA

dario.tbn@gmail.com

DIRECTOR: ING. MARIA HALLO

maria.hallo@epn.edu.ec

Quito, Septiembre de 2016

DECLARACIÓN

Yo, Dario Javier Tubón Moposita, declaro bajo juramento que el trabajo aquí descrito es de mi autoría; que no ha sido previamente presentada para ningún grado o calificación profesional; y, que he consultado las referencias bibliográficas que se incluyen en este documento.

A través de la presente declaración cedo mis de propiedad intelectual correspondientes a este trabajo, a la Escuela Politécnica Nacional, según lo establecido por la Ley de Propiedad Intelectual, por su Reglamento y por la normatividad institucional vigente.

Dario Javier Tubón Moposita

CERTIFICACIÓN

Certifico que el presente trabajo fue desarrollado por Dario Javier Tubón Moposita, bajo mi supervisión.

Ing. María Hallo
DIRECTOR DE PROYECTO

AGRADECIMIENTOS

Agradezco a mis padres Carlos y Elvia, que me han apoyado para conseguir esta meta, enseñándome a no desmayar en el camino y que el esfuerzo junto a la dedicación permiten alcanzar grandes objetivos.

A Silvia y Anthony, mis dos compañeros incondicionales, en las buenas y malas situaciones presentadas en este periodo de mi vida. Compañeros que Dios puso en mi camino.

A mis hermanas, Lorena y Magaly, por sus ánimos y alegría a lo largo de la carrera.

A mis tios, Jaime y Monica, con sus consejos y amistad estuvieron a mi lado y fortalecieron el espíritu de familia.

A mis compañeros y amigos conseguidos a lo largo de estos años, por permitirme conocer excelentes personas.

A todos mis profesores, que gracias a su conocimiento permitieron formarme y crecer personal y profesionalmente.

Y un especial agradecimiento a mi directora Msc. Maria Hallo, por el apoyo brindado a lo largo de la realización del presente proyecto.

Dario T.

DEDICATORIA

A mis padres, por su apoyo incondicional.

A mi compañera, Silvia, por tu apoyo, constancia y amor.

A Anthony, por ser mi faro en los momentos difíciles, indicándome el camino que debo seguir.

A todas las personas, amigos y familiares que me acompañaron en este periodo de mi vida.

Dario T.

CONTENIDO

CAPITULO 1 PLANTEAMIENTO DEL PROBLEMA	1
1.1 SITUACIÓN ACTUAL	1
1.1.1 WEB SEMÁNTICA	2
1.1.2 RDF	4
1.1.3 LINKED DATA	6
1.1.4 OLAP (<i>ONLINE ANALYTICAL PROCESSING</i>)	8
1.1.5 VOCABULARIO PARA CUBO DE DATOS RDF	9
1.2 DESCRIPCIÓN DEL PROBLEMA	10
1.3 JUSTIFICACIÓN	11
1.4 SELECCIÓN Y JUSTIFICACIÓN DE LA METODOLOGÍA	11
1.4.1 JUSTIFICACIÓN DE LA METODOLOGÍA	11
1.4.1.1 Metodología de Bill Inmon	11
1.4.1.2 Metodología Kimball	12
1.4.2 SELECCIÓN DE LA METODOLOGÍA	14
1.5 SELECCIÓN Y JUSTIFICACIÓN DE LAS HERRAMIENTAS DE DESARROLLO	14
1.5.1 MOTORES DE ALMACENAMIENTO RDF	14
1.5.1.1 OpenLink Virtuoso	14
1.5.1.2 Sesame	15
1.5.2 HERRAMIENTAS PARA LIMPIEZA Y TRANSFORMACIÓN DE DATOS	15
1.5.2.1 Google Refine	15
1.5.2.2 Data Wrangler	15
1.5.3 WIKIS SEMÁNTICOS	16
1.5.3.1 OntoWiki	16
1.5.3.2 Semantik MediaWiki	16
1.5.4 VISUALIZADORES DATOS ESTADÍSTICOS	16
1.5.4.1 CubeViz	16
1.5.4.2 Payola	16
1.5.5 SELECCIÓN DE LAS HERRAMIENTAS	16
CAPÍTULO 2 DESARROLLO DEL SISTEMA	20
2.1 ESPECIFICACIÓN DE REQUERIMIENTOS	20
2.1.1 CASO DE ESTUDIO	20

2.1.2	CASOS DE USO	20
2.1.3	OBJETIVOS DEL SISTEMA	21
2.2	ANÁLISIS Y DISEÑO	22
2.2.1	MODELO DIMENSIONAL	22
2.2.2	DISEÑO DE ETL Y DESARROLLO	25
2.2.2.1	Identificación de fuentes de datos	25
2.2.2.2	Extracción de datos	28
2.2.2.3	Diseño de URI.....	29
2.2.2.4	Limpieza de datos.....	31
2.2.2.5	Modelamiento del cubo de datos RDF	36
2.3	PRUEBAS E IMPLANTACIÓN DEL SISTEMA	51
2.3.1	PRUEBAS DEL SISTEMA	51
2.3.1.1	Pruebas de funcionalidad	51
2.3.1.1.1	Pruebas de seguridad	52
2.3.1.1.2	Pruebas de sistema	52
2.3.2	IMPLANTACIÓN DEL SISTEMA.....	56
2.3.2.1	Instalación del servidor Virtuoso	57
2.3.2.2	Instalación de OntoWiki.....	58
2.3.2.3	Instalación de CubeViz	60
CAPÍTULO 3 EVALUACIÓN DEL SISTEMA EN EL CASO DE APLICACIÓN		61
3.1	EJECUCIÓN DEL SISTEMA	61
3.2	RECOPIACIÓN DE INFORMACIÓN DEL CASO DE APLICACIÓN	64
3.2.1	METADATOS DE LA REVISTA POLITÉCNICA.....	64
3.2.2	CONJUNTOS DE DATOS EXTERNOS	65
3.3	ANÁLISIS DE RESULTADOS.....	65
3.3.1	VISUALIZACIÓN DE GRÁFICOS ESTADÍSTICOS	65
3.3.2	CONSULTAS SPARQL	68
CONCLUSIONES Y RECOMENDACIONES		72
CONCLUSIONES.....		72
RECOMENDACIONES		72
BIBLIOGRAFÍA.....		74
GLOSARIO		76
ANEXOS.....		77

INDICE DE FIGURAS

Figura 1 Evolución de la Web.....	2
Figura 2 Arquitectura de la Web Semántica	3
Figura 3 Representación gráfica de una tripleta.....	4
Figura 4 Grafo RDF	5
Figura 5 Representación grafo RDF en RDF/XML.....	5
Figura 6 Conjuntos de Linked Data	7
Figura 7 Vocabulario para cubos de datos RDF	9
Figura 8 Niveles de la arquitectura de Inmon	12
Figura 9 Tareas de la metodología de Kimball.....	13
Figura 10 Arquitectura de Virtuoso Open Link.....	15
Figura 11 Menú principal de Virtuoso	17
Figura 12 Datos en Google Refine.....	17
Figura 13 Presentación de una base de conocimiento en OntoWiki.....	18
Figura 14 Gráfico de pastel en Cube Viz.....	19
Figura 15 Caso de uso del sistema OLAP.....	21
Figura 16 Diagrama dimensional	23
Figura 17 Representación del modelo de datos dimensional con vocabulario de cubo de datos RDF	24
Figura 18 Proceso para la publicación de Linked Data	25
Figura 19 Lista de conjuntos de datos en OJS	26
Figura 20 Visita considerada como valida	28
Figura 21 Consolidación de conjunto de datos	31
Figura 22 Función Replace de Google Refine	32
Figura 23 Separación en varias columnas en Google Refine	32
Figura 24 Unión de diferentes registros en Google Refine.....	33
Figura 25 Agrupación por títulos de los artículos.....	34
Figura 26 Vista de los registros cargados de OAI-PMH	34

Figura 27 Parte de los registros limpiados	35
Figura 28 Limpieza de datos	36
Figura 29 Definición de estructura o esqueleto RDF	37
Figura 30 Importación de vocabulario para cubos de datos RDF	37
Figura 31 Lista de vocabularios y prefijos	38
Figura 32 Definición de dimensiones y medidas	39
Figura 33 Etiquetado de dimensiones y medidas	40
Figura 34 Definición de conjunto y la estructura de datos	41
Figura 35 Componentes	41
Figura 36 Especificación de componentes	42
Figura 37 Propiedades qb:dimension y qb:measure de los componentes.....	43
Figura 38 Metadatos del conjunto de datos	44
Figura 39 Selección de la columna con las observaciones.	45
Figura 40 Observaciones en Google Refine	46
Figura 41 Propiedades de las observaciones.....	47
Figura 42 Observación en formato TTL.	47
Figura 43 Observación con las propiedades: dimensión, medida y conjunto de datos.....	47
Figura 44 Cubo expresado en formato RDF/XML	48
Figura 45 Esquema de proyecto en Silk.....	49
Figura 46 Esquema de tareas para generación de enlaces	50
Figura 47 Enlaces generados.....	51
Figura 48 Dimensión: Titulo.....	53
Figura 49 Dimensión: Año.....	53
Figura 50 Dimensión: Mes.....	54
Figura 51 Dimensión: Titulo.....	54
Figura 52 Medida: Número de visitas.....	54
Figura 53 Definición de la estructura de datos del cubo de datos RDF.....	55
Figura 54 Arquitectura de sistema	56

Figura 55 Configuración de usuario en Virtuoso.....	57
Figura 56 Interfaz de ingreso a Virtuoso	58
Figura 57 Configuración ODBC	59
Figura 58 Pantalla de Inicio de OntoWiki	59
Figura 59 Extensión CubeViz de OntoWiki	60
Figura 60 Creación de grafo RDF	61
Figura 61 Grafos cargados en Virtuoso	62
Figura 62 Creación de nueva base de conocimiento.....	62
Figura 63 Panel de selección de datos	63
Figura 64 Lista de conjuntos de datos.....	64
Figura 65 Gráfico de barras horizontales.....	66
Figura 66 Gráfico de barras verticales	66
Figura 67 Gráfico de dispersión.....	67
Figura 68 Gráfico de dispersión.....	67
Figura 69 Endpoint para consultas SPARQL	68
Figura 70 Consulta: total de visitas.....	69
Figura 71 Consulta: total de visitas a artículos	69
Figura 72 Consulta: países con más visitas.....	70
Figura 73 Consulta: artículo con más visitas	71

INDICE DE TABLAS

Tabla 1 Ejemplo de declaración en RDF	6
Tabla 2 Metadatos de un artículo obtenido mediante OAI-PMH.....	27
Tabla 3 Dimensiones y medidas	29
Tabla 4 Diseño de URIs principales.....	29
Tabla 5 Diseño de URIs para dimensiones y medidas del cubo de datos RDF	30
Tabla 6 Vocabulario adicionales	36
Tabla 7 Metadatos del conjunto de datos.....	44
Tabla 8 Pruebas de seguridad.....	52
Tabla 9 Registros generados	65

RESUMEN

El presente proyecto presenta la elaboración del sistema OLAP para bibliotecas digitales que utilizan Linked Data. Empieza con la descripción de conceptos sobre Web Semántica, Linked Data, OLAP, y el vocabulario para la creación de cubos de datos RDF.

Para la selección de la metodología se tuvo en cuenta las que tradicionalmente se utilizan para el desarrollo de sistemas de inteligencia de negocios, como son Inmon y Kimball, de las cuales se optó por la metodología de Kimball. Adicionalmente se adaptó las tareas de extracción, transformación y carga de la metodología para que en este proceso se pueda utilizar la guía de buenas prácticas para la publicación de datos en la Web Semántica que indica el W3C.

La fuente de datos fue la Revista Politécnica. Para extraer los metadatos de los artículos académicos se utilizó el protocolo OAI-PMH. Las estadísticas de las visitas se obtuvieron desde Google Analytics, estos dos conjuntos de datos fueron consolidados en un conjunto único a partir del cual se obtuvo el cubo de datos RDF.

La visualización se realizó usando CubeViz. Las tripletas del cubo de datos RDF pueden ser consultadas desde un endpoint SPARQL, y lo que es más importante, pueden ser reutilizados y consumidos por cualquier usuario.

Una vez elaborado el cubo de datos RDF se procedió a instalar todos los componentes necesarios para la implementación, servidor Virtuoso, wiki semántico OntoWiki, y CubeViz para la interpretación gráfica del cubo de datos RDF y sus posteriores consultas.

Para las pruebas del sistema se llevaron a cabo las que permiten verificar el correcto funcionamiento de todos los componentes, no centradas en el desarrollo, sino más orientadas hacia la funcionalidad del mismo.

Una vez instalados todos los componentes, se cargó el cubo de datos RDF en el servidor Virtuoso. Las consultas SPARQL se realizaron directamente al grafo RDF mediante el endpoint SPARQL de Virtuoso.

Finalmente se recogen las recomendaciones y conclusiones derivadas de la elaboración de este proyecto.

Palabras clave: Linked Data, metadatos, RDF, cubo de datos, endpoint.

PRESENTACIÓN

La Web ha evolucionado a pasos agigantados desde su concepción por Timothy “Tim” Berners-Lee hasta como la conocemos hoy. Durante esta evolución se ha ido agregando contenido, contenido que ha crecido de forma exponencial. Cada vez es más complicado obtener resultados relevantes en la búsqueda de un recurso específico en la Web.

Con la Web Semántica se busca de alguna forma dar solución a este inconveniente, mediante la publicación de información entendible para máquinas y seres humanos, permitiendo de esta manera que el resultado de una búsqueda sea realmente el esperado.

Las Tecnologías Linked Data, aplicadas a la Web Semántica, buscan enlazar toda esta información, de tal manera que el resultado sea un consolidado de información que brinde conocimiento al consumidor final. De la misma forma que la Web tradicional, los elementos de la Web Semántica generan datos estadísticos, que también deben ser tratados con sus propias herramientas y vocabularios.

Es por esto que en el presente trabajo se realiza el desarrollo de un sistema que permita la manipulación de información estadística dentro de la Web Semántica, orientado exclusivamente a una porción de información generada a partir de las visitas realizadas al sitio web de la Revista Politécnica, utilizando el vocabulario para cubos de datos RDF.

En el primer capítulo se indica una breve reseña de la situación actual de la Web y su evolución a la Web Semántica. Se indican ideas básicas acerca del marco de referencia para describir un recurso en la Web conocido como RDF y el enlazamiento de conjuntos de datos generados. Se describe el problema que se busca dar solución con el desarrollo del presente trabajo, en la parte final de este capítulo se realiza la selección de la metodología y herramientas a utilizarse.

En el segundo capítulo se presentan las etapas de análisis, diseño e implementación del sistema OLAP para bibliotecas que utilizan tecnologías Linked Data.

En el tercer capítulo se realiza la evaluación del sistema con datos reales del caso de estudio, probando que todos los componentes funcionen correctamente. Se realizan consultas SPARQL que permite dar respuestas a interrogantes sobre la información estadística del cubo de datos RDF generado en este trabajo.

Finalmente se recogen las recomendaciones y conclusiones derivadas de la elaboración de este proyecto.

CAPITULO 1 PLANTEAMIENTO DEL PROBLEMA

En la actualidad existen varios proyectos que impulsan la publicación de información apta para ser consumida en la Web Semántica. DBpedia es el proyecto que más conjuntos de datos con este tipo de información posee. La publicación de esta información permite al usuario obtener resultados más próximos a los que realmente está buscando.

Las interconexiones entre estos conjuntos de datos permiten encontrar conocimiento, que puede ser aprovechado por cualquier usuario: humano o máquina. La información aislada es menor a medida que se incrementa la información en la Web Semántica. Todos estos recursos poseen consigo información estadística, por lo que es necesario expresar dicha información para que también sea consumida por los usuarios de la Web Semántica. La publicación de los datos estadísticos debe realizarse en un lenguaje adecuado.

Varias bibliotecas digitales han implementado herramientas que facilitan la publicación de recursos en la Web Semántica, pero la información relacionada a estos aun es limitada, y especialmente la información estadística que generan estos recursos.

Con el presente trabajo se elaborará un conjunto de datos estadísticos, conocido como cubo de datos RDF a partir de la información de los recursos alojados en el repositorio digital de la Revista Politécnica, información que ha sido recolectada dentro de un periodo específico.

En este capítulo se realiza el planteamiento del problema, inicia con una breve revisión de conceptos, que incluyen:

- RDF
- Linked Data
- Vocabulario para cubos de datos RDF

En la sección 1.2 se indica la descripción del problema al que se busca dar solución mediante el sistema OLAP para bibliotecas digitales que utilizan tecnologías Linked Data, la sección 1.3 indica la debida justificación a la solución presentada. Finalmente, las secciones 1.4 y 1.5 indican la selección de la metodología y herramientas necesarias a utilizarse respectivamente.

1.1 SITUACIÓN ACTUAL

Cualquier recurso alojado en la Web tiene sus propios datos, conocidos como metadatos. Los metadatos contienen información relevante para identificar de forma particular a un recurso. Con el auge de nuevas Tecnologías de Información y Comunicación los metadatos de los recursos han tomado mayor importancia.

El análisis de estos metadatos concentrados en conjuntos de datos (*datasets*) permite el descubrimiento de nuevo y valioso conocimiento; que puede ser utilizado en cualquier ámbito. La información aislada se transforma en este valioso

conocimiento al momento de enlazar entre sí todos estos conjuntos de datos. El conocimiento que se pueda obtener de estos conjuntos de datos incrementa mientras más conexiones existan entre sí.

Para llegar a este punto es necesario hacer una breve reseña de la evolución de la Web, la que desde su concepción por Tim Berners-Lee ha evolucionado radicalmente como indica la Figura 1.

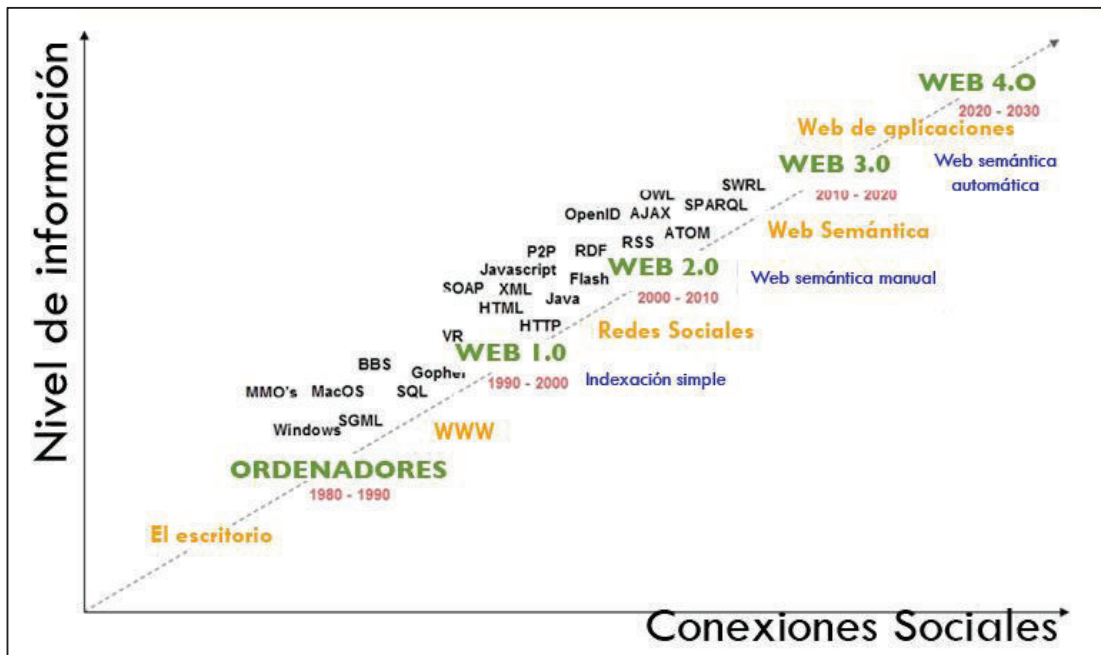


Figura 1 Evolución de la Web [1]

- **Web 1.0,**
La Web más simple, interacción mínima entre productores y consumidores de contenido.
- **Web 2.0**
La Web 2.0 se cimienta en el trabajo colaborativo entre los usuarios de la Web; comunidades digitales. La colaboración permite la difusión del conocimiento de forma sencilla.
- **Web Semántica**
La Web Semántica surge de la necesidad de generar conocimiento a partir de los recursos diseminados en la Web actual.

1.1.1 WEB SEMÁNTICA

Desde que la WWW (*World Wide Web*) -de aquí en adelante llamada Web- fue ideada por Tim Berners-Lee, esta se ha convertido en un gran contenedor de información muy variada, basada principalmente en documentos y enlaces de

hipertexto [2]. Teniendo esta gran cantidad de información disponible, si un usuario realiza una búsqueda no siempre encuentra lo que realmente necesita. Este inconveniente surge principalmente por que la Web ha sido concebida para ser interpretada por humanos; impidiendo que las máquinas puedan entender su contenido.

“La Web Semántica es una Web extendida, dotada de mayor significado en la que cualquier usuario en Internet podrá encontrar respuestas a sus preguntas de forma más rápida y sencilla gracias a una información mejor definida” [3]

Tim Berners-Lee indica que la Web Semántica debe estar cimentada sobre una arquitectura de capas como se muestra en la Figura 2. Cada una indica las herramientas y el lenguaje apropiado a ser utilizados.

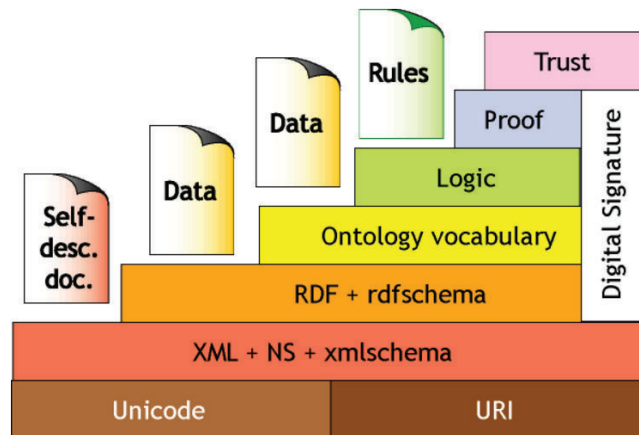


Figura 2 Arquitectura de la Web Semántica. [4]

- **Unicode**
El unicode permite la interpretación de los recursos descritos en la Web Semántica en cualquier idioma.
- **URI (*Uniform Resource Identifier*)**
Una URI es el identificador único para cada recurso alojado en la Web, permitiendo un acceso más fácil, con su uso se evita ambigüedades.
- **XML+NS+xmlschema**
Esta capa es más técnica, permite que los agentes puedan mantener comunicación mediante un formato único de intercambio de documentos (XML).
- **RDF+RDFschema**
En esta capa se define un lenguaje universal (RDF), el mismo que permite expresar claramente diferentes ideas en la Web Semántica.

- **Lenguajes de Ontologías**
Un lenguaje de ontologías facilita la descripción de recursos y las relaciones de estos con otros en la Web.
- **Capa Lógica**
En esta capa se especifica las reglas de inferencia que utilizarán las ontologías para determinar las relaciones entre recursos.
- **Confianza**
La confianza indica a los agentes que realizan las búsquedas en la Web Semántica no dar como válido un resultado hasta que sean comparadas con fuentes de información comprobadas.
- **Firma Digital**
La firma digital es un segmento de datos, encriptado, que verifica la confiabilidad de una fuente de información.

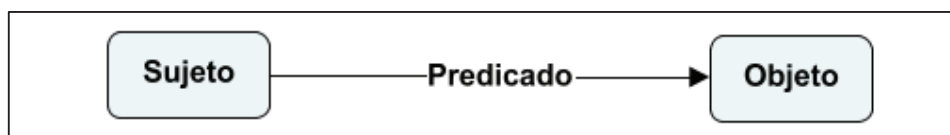
La Web Semántica necesita de tecnologías y lenguajes propios para que la información sea publicada correctamente, los mismos que se describen en los siguientes apartados.

1.1.2 RDF

RDF (*Resource Description Framework*) es un lenguaje que permite la representación de la información de cada recurso alojado en la Web. Especialmente indicado para representar los metadatos que describen e identifican a un recurso, estos metadatos pueden ser tan variados como: el título, autor, las modificaciones efectuadas, derechos de autor de un recurso, etc. [5].

La descripción de estos recursos en RDF son expresados en forma de tripletas. Una tripleta está formada por sujeto, predicado y objeto, elementos que son identificados por URIs, los cuales al unirse forman un grafo dirigido, en el cual los sujetos y los objetos son nodos, mientras que el predicado es el arco que los enlaza.

La Figura 3 muestra un ejemplo gráfico de tripleta.



*Figura 3 Representación gráfica de una tripleta
Realizado por el Autor*

La sintaxis de RDF está basada sobre lenguaje XML (*Extensible Markup Language*) siendo conocida como RDF/XML [5], la utilización de esta sintaxis permite a la información ser intercambiable y entendible por las máquinas. En la

Figura 4 y Figura 5 podemos ver la representación de un grafo RDF en RDF/XML formado por 3 triplas.

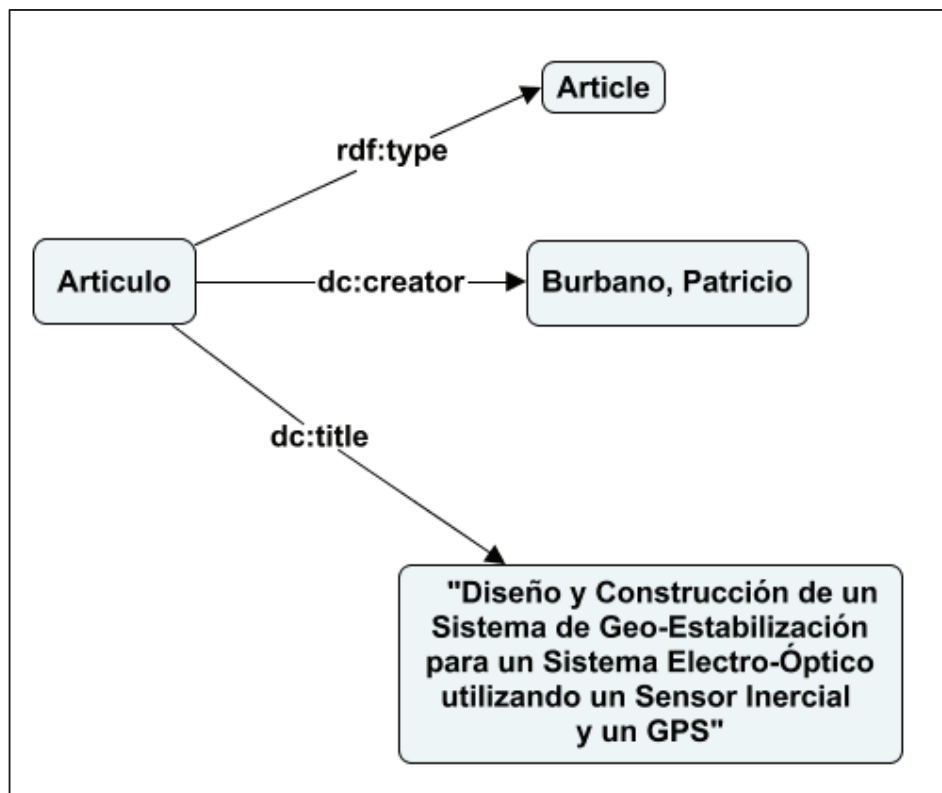


Figura 4 Grafo RDF
Realizado por el Autor

```

<rdf:RDF>
  <rdf:Description rdf:about="http://opendata.epn.edu.ec/article/79">
    <rdf:type rdf:resource="http://purl.org/ontology/bibo/Article"/>
    <dc:creator xml:lang="en">Burbano, Patricio</dc:creator>
    <dc:title>
      Diseño y Construcción de una Plataforma Robótica para Control
      PTZ de un Sistema Electro-Óptico
    </dc:title>
  </rdf:Description>

```

Figura 5 Representación grafo RDF en RDF/XML
Realizado por el Autor

La descripción de un recurso en RDF está fundamentado en la utilización de:

➤ **Recursos**

Un recurso es cualquier objeto que pueda ser descrito mediante RDF, estos pueden ser objetos tan variados como un artículo académico, una

persona o cualquier recurso alojado en la Web. Cada recurso debe ser identificado por su respectiva URI.

➤ **Propiedades**

Las propiedades son las características específicas que permiten la descripción del recurso.

➤ **Declaraciones**

Una declaración es la combinación de un recurso, sus propiedades y los valores que puedan tomar estas propiedades. Esta combinación deriva en una tripleta en la cual el sujeto es el recurso, el predicado es la propiedad y el objeto es el valor de dicha propiedad.

La Tabla 1 ilustra la declaración que permite describir en RDF un artículo académico de la Revista Politécnica.

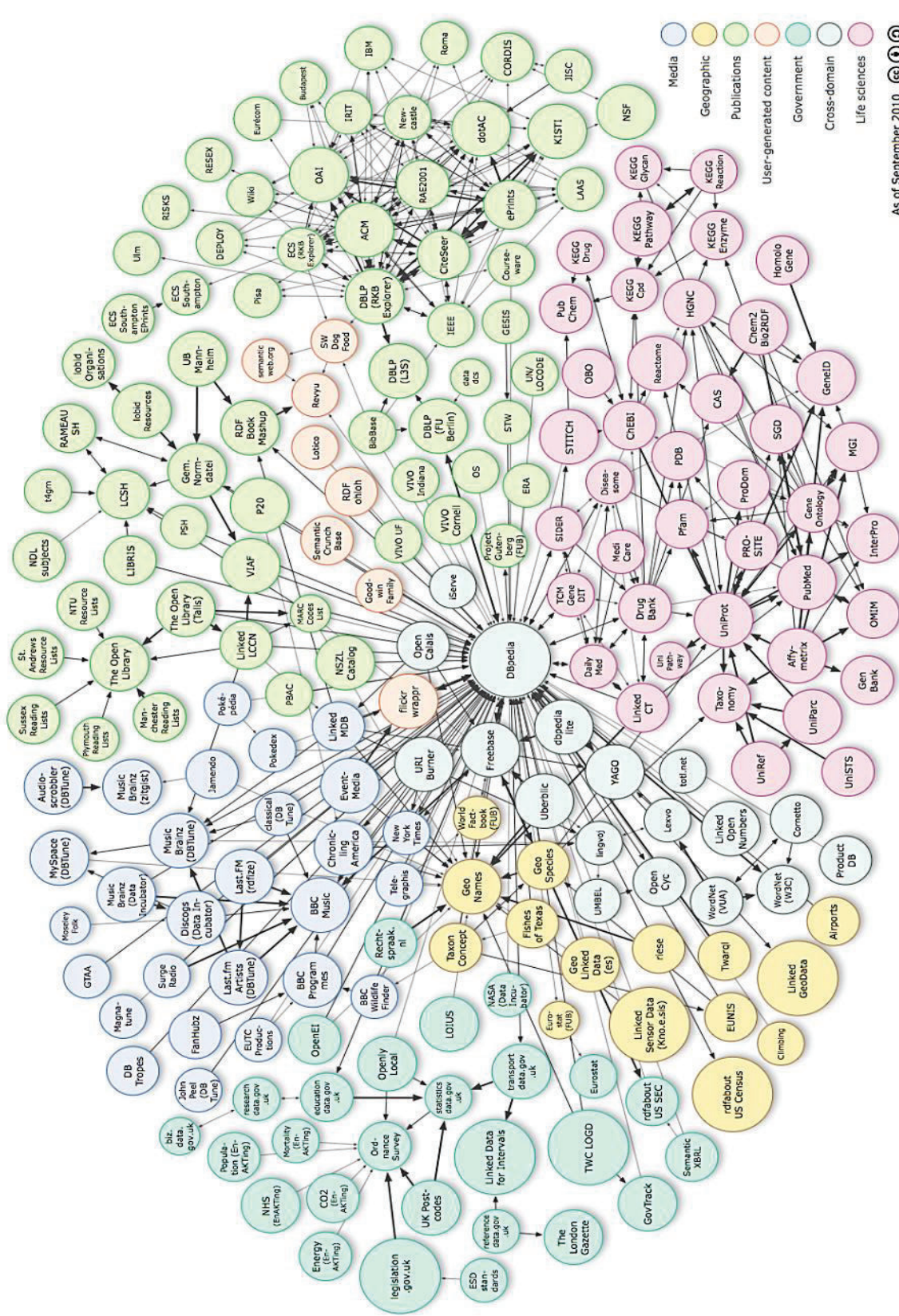
Tabla 1 Ejemplo de declaración en RDF

	Recurso	Propiedad	Valor de la propiedad
Sujeto	Artículo académico de la Revista Politécnica		
Predicado		Título	
Objeto			“Diseño y Construcción de un Sistema de Geo-Estabilización para un Sistema Electro-Óptico utilización de un Sensor Inercial y un GPS”

1.1.3 LINKED DATA

La creación de una Web que entregue conocimiento, con información entendible tanto para máquinas y personas, no sería posible si todos sus recursos no se encuentran enlazados entre sí. Linked Data permite el enlazamiento de cualquier recurso alojado en la Web, similar a los enlaces utilizados dentro de una página web tradicional [6].

Las interconexiones entre conjuntos de datos que contengan información similar, forman una gran nube, cuyo objetivo es tener la menor cantidad de información aislada. La Figura 6 muestra de forma gráfica varios conjuntos de datos publicados en la Web Semántica actualmente interconectados entre sí.



As of September 2010



Figura 6 Conjuntos de Linked Data

La publicación de conjuntos de datos con Linked Data debe tener en cuenta 4 principios básicos:

- **Utilizar URIs para nombrar las cosas**, permitiendo la identificación inequívoca de los recursos en la Web.
- **Utilizar URIs HTTP**, para que un usuario pueda buscar e identificar un recurso en la Web.
- **Cuando un usuario busque un recurso mediante su URI**, toda la información debe ser provista mediante la utilización de estándares como RDF y SPARQL.
- **Incluir enlaces a otras URIs**, de tal forma que el usuario pueda descubrir más cosas.

1.1.4 OLAP (*ONLINE ANALYTICAL PROCESSING*)

Como hemos visto, en la Web cada día se alojan más recursos: documentos, imágenes, audio, etc., lo que hace necesaria la utilización de herramientas BI (*Bussines Intelligence*) para su análisis. El análisis de esta información facilita a una organización la toma de decisiones en beneficio de la continuidad del negocio.

OLAP permite un análisis de datos multidimensional, conocidos como cubos multidimensionales, facilita la consulta de la información por parte del usuario, múltiples vistas de los datos según sean los requerimientos y operaciones intuitivas sobre estos cubos multidimensionales [7].

OLAP es un proceso interactivo de creación, administración, análisis y reporte de datos. Las principales operaciones de un OLAP son:

- **ROLL-UP**
La operación ROLL-UP incrementa el nivel de agregación, disminuyendo la granularidad de la información.
- **DRILL-DOWN**
La operación DRILL-DOWN disminuye el nivel de agregación incrementando la granularidad de la información.
- **SLICE-DICE**
La operación SLICE-DICE permite la manipulación de un subconjunto de datos del cubo, el cual cumple con determinadas restricciones.

Para el análisis de toda esta información publicada en la Web Semántica el OLAP utilizado debe ser creado con las tecnologías adecuadas para ser interpretado por máquinas y personas. El punto de partida es el vocabulario a utilizarse: conocido como vocabulario para cubos de datos RDF (*RDF Data Cube Vocabulary*).

1.1.5 VOCABULARIO PARA CUBO DE DATOS RDF (*RDF DATA CUBE VOCABULARY*)

Cuando son publicados los datos en la Web, especialmente lo relacionado a estadísticas, este vocabulario permite darle un significado semántico mediante la utilización del estándar RDF.

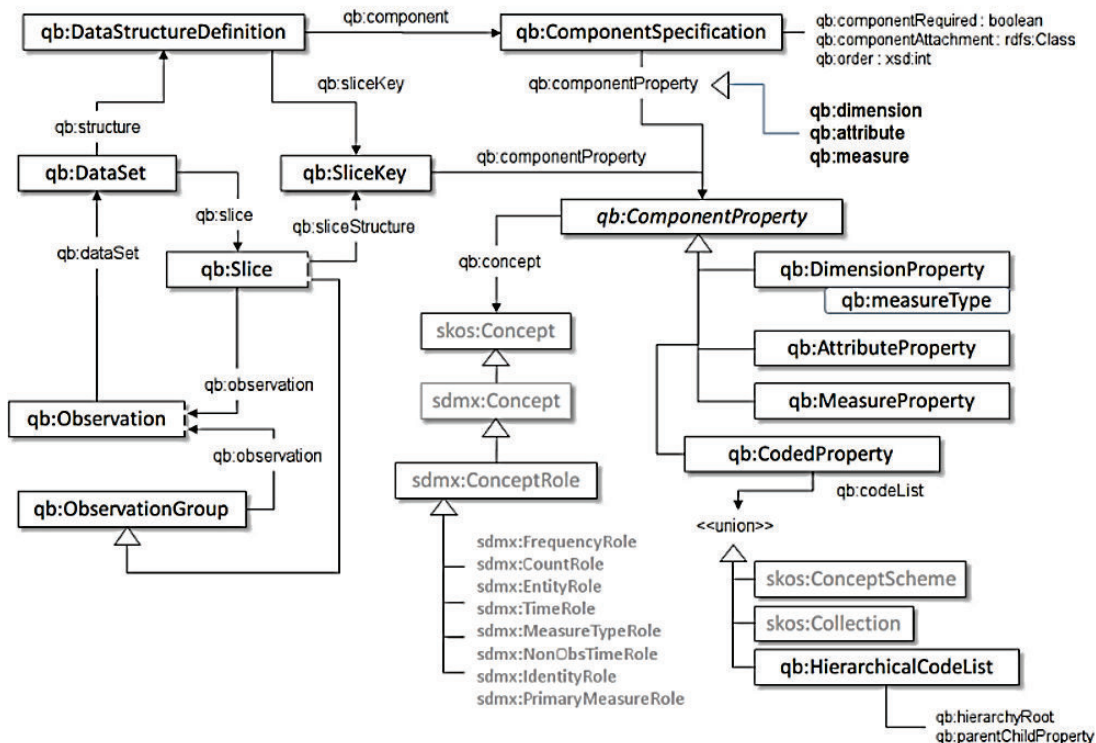


Figura 7 Vocabulario para cubos de datos RDF [8]

Como se puede observar en la Figura 7, en términos generales, este vocabulario se encuentra dividido en clases, propiedades (incluyendo sus dimensiones, atributos y medidas), observaciones y operaciones que pueden ser aplicadas a un conjunto de datos.

Las principales clases y propiedades de este vocabulario son:

- **Conjunto de Datos (*qb:DataSet*)**, es el conjunto de datos estadísticos, los mismos que tienen una estructura definida.
- **Observación (*qb:Observation*)**, es el valor de la medida, si hablamos de una tabla estadística, la observación es el valor que tiene cada celda de esta tabla.
- **Definición de la Estructura de los Datos (*qb:DataSetDefinition, DSD*)**, permite definir la estructura de uno o más conjunto de datos. Específicamente las dimensiones, atributos y medidas que serán utilizados.

- **Dimensiones, atributos y medidas**, el vocabulario para cubo de datos RDF permite representarlos como propiedades RDF [8], que son instancias de la clase abstracta *qb:ComponentProperty*, que contiene a las sub clases:
 - ***qb:DimensionProperty***, es la sub clase que representa los componentes que forman las dimensiones del cubo.
 - ***qb:AttributeProperty***, es la sub clase formada por todos los componentes que representan los atributos de las observaciones dentro del cubo de datos.
 - ***qb:MeasureProperty***, es la sub clase que representa el valor de la medida para el fenómeno observado.
- **Slices**, representa un subconjunto del conjunto de datos que se forma a partir de un subconjunto de valores dimensionales.

La propiedad de un componente encapsula una gran cantidad de información:

- El concepto que va a ser representado (p. ej. tiempo, ubicación geográfica)
- La naturaleza del componente (dimensión, atributo o medida) tal como es representado por la propiedad del tipo del componente.
- El tipo utilizado para representar el valor.

Un mismo concepto puede estar presente en diferentes componentes. Así por ejemplo, el concepto de moneda puede ser utilizado como dimensión o un atributo. En el primero caso el conjunto de datos almacenaría información acerca de los tipos de cambio, en el segundo este mismo concepto describiría la moneda en la que se llevó a cabo una transacción comercial.

1.2 DESCRIPCIÓN DEL PROBLEMA

La publicación de una gran cantidad de recursos académicos en la Web, sin control alguno, permite que valioso potencial conocimiento respecto a este ámbito se encuentre aislado [9], por ejemplo: la detección de nuevas temas relevantes de estudio, investigadores que pueden trabajar conjuntamente sobre una misma temática, etc. Conocimiento que beneficia directamente a la comunidad académica.

El desarrollo del Sistema OLAP mediante la utilización de tecnologías Linked Data, busca procesar la información estadística de los recursos publicados en la Web Semántica con principios de Linked Data, en este caso de estudio con datos relacionados al ámbito académico. Esta información será enriquecida mediante el enlazamiento con conjuntos de datos externos. Los resultados de este trabajo están disponibles para su acceso y reutilización por cualquier usuario.

1.3 JUSTIFICACIÓN

Varias bibliotecas a nivel mundial se encuentran implementando nuevos servicios usando tecnologías Linked Data, que permite el enlazamiento de sus datos a fuentes externas. Enlazamiento que facilita consultas más complejas y resultados que entregan información valiosa al usuario.

Sin embargo existen pocas iniciativas que busquen desarrollar módulos de apoyo a la toma de decisiones, por lo que el desarrollo del sistema OLAP para bibliotecas digitales que utilizan tecnologías Linked Data contribuirá en este sentido. El sistema contribuirá también con sus resultados al apoyo de iniciativas de utilización de Linked Data en bibliotecas digitales, permitiendo la integración y una mejor explotación de la información publicada.

1.4 SELECCIÓN Y JUSTIFICACIÓN DE LA METODOLOGÍA

Para llevar a cabo el presente Proyecto es necesario seguir las indicaciones que dictan las metodologías para elaborar proyectos BI, tareas que serán combinadas con guía de buenas prácticas para la publicación en la Web Semantica del W3C. En los siguientes puntos se realizan la explicación de las metodologías más utilizadas Bill Inmon y Kimball, para posteriormente justificar su elección.

1.4.1 JUSTIFICACIÓN DE LA METODOLOGÍA

1.4.1.1 Metodología de Bill Inmon

Esta metodología se basa en la utilización de un CIF (*Corporate Information Factory*), el CIF concentra toda la información de los demás sistemas transaccionales de la organización; que podrá ser analizada posteriormente [10]. Para Inmon, este CIF debe contar con ciertas características específicas:

- **Orientado a temas.**- La organización del contenido de un almacén de datos debe permitir que todo lo relativo a un hecho u objeto del mundo real queden unidos entre sí.
- **Integrado.**- Los almacenes de datos deben contener la información pertinente a todos los sistemas operacionales de la organización; únicamente información consistente.
- **No volátil.**- Una vez almacenada la información, esta no sufrirá modificación alguna; solo podrá ser consultada.
- **Variante en el tiempo.**- Cualquier cambio que se pueda producir en la información es registrado y presentados mediante informes.

La arquitectura mostrada por Inmon consta de 4 niveles de datos, como se indica en la Figura 8, el nivel operacional, el atómico o nivel de *data warehouse*, el departamental o *data mart* y el nivel individual [11]. El nivel operacional se centra en los datos primitivos. El nivel de *data warehouse* en conjuntos de datos

integrados y ordenados. El nivel departamental contiene conjuntos de datos relacionados de diferentes partes de la organización. Y el nivel individual, donde se realiza el análisis heurístico.

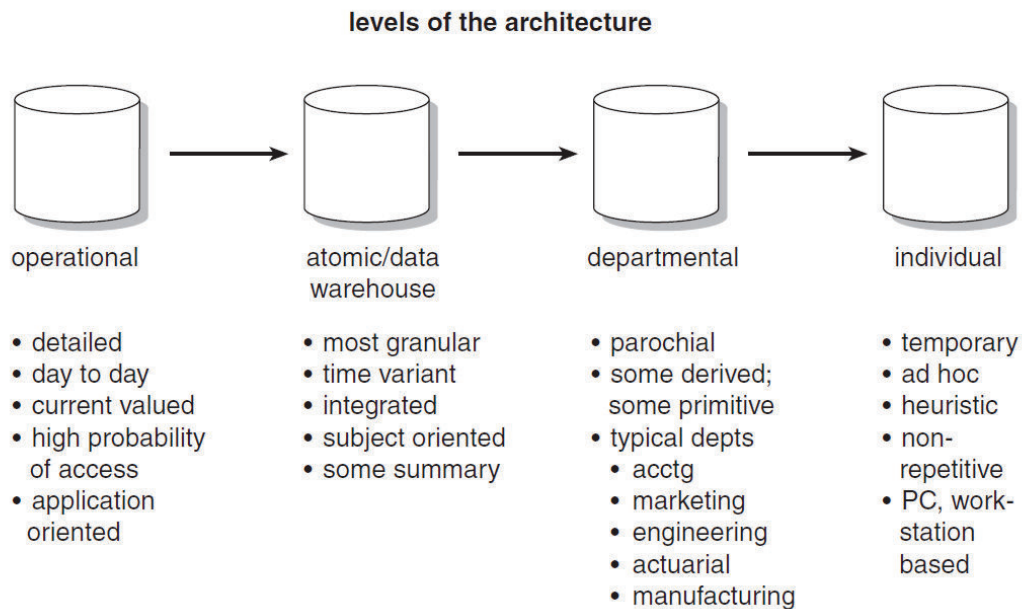


Figura 8 Niveles de la arquitectura de Inmon [11]

El enfoque de Inmon es Top-down, los datos extraídos de los sistemas transaccionales mediante procesos ETL (*Extract, Transform and Load*), pasan a formar parte de un *data warehouse* corporativo. Estos *data warehouse* contienen metadatos que describen exactamente el contenido del mismo [10].

1.4.1.2 Metodología Kimball

Esta metodología está fundamentada en el denominado Ciclo de Vida Dimensional del Negocio (*Business Dimensional Lifecycle*) [12], este ciclo considera a un *data warehouse* como la agrupación de todos los *data mart* departamentales, que contienen los datos de las transacciones dispuestos de forma específica para su análisis.

Este ciclo de vida tiene en cuenta 4 principios fundamentales:

- **Centrado en el negocio**
Este principio permite identificar correctamente los requerimientos del negocio y el valor que estos tienen para el mismo.
- **Construcción de una infraestructura de información adecuada.**
Esta infraestructura es la única base de información, integrada y fácil de utilizar, la misma que poseerá todos los requerimientos del negocio que se hayan identificado.

➤ **Entregables con incrementos significativos**

El *data warehouse* se crea mediante entregables; dentro de plazos determinados. Para determinar la secuencia que deben seguir estos incrementos, se cuantifica el valor de negocio que posee cada elemento identificado.

➤ **Dar una solución completa al problema**

La solución presentada debe proveer todo los elementos necesarios que permiten dar valor a los usuarios del negocio. Lo que se traduce en datos sólidos, diseñados correctamente, de calidad, y accesibles.

Las tareas que forman parte de esta metodología las podemos ver en la Figura 9, las cuales facilitan llevar a cabo un proyecto BI.

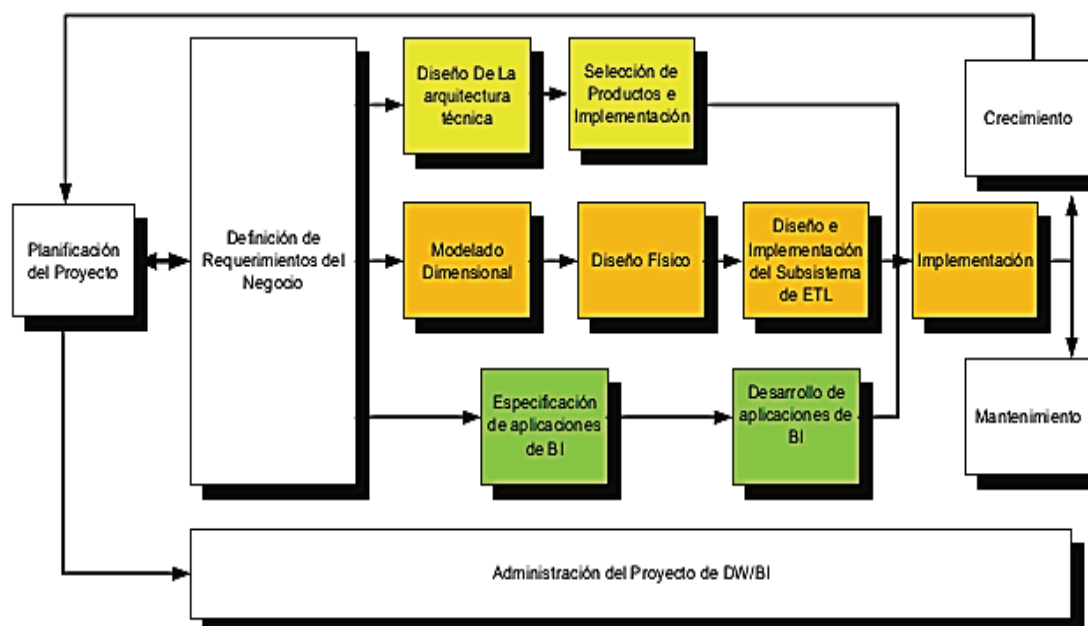


Figura 9 Tareas de la metodología de Kimball [12]

De todas las tareas, la que mayor importancia posee es la definición de los requerimientos del negocio, a partir de esta tarea se desarrollaran las siguientes. Por otro lado, la metodología de Kimball muestra -por decirlo de alguna forma- 3 caminos:

➤ **La tecnología.**

Estas tareas hacen referencia a las herramientas software a utilizar en el proyecto.

➤ **Los datos**

Estas tareas permitirán la extracción, transformación y carga de datos (ETL, por sus siglas en Ingles).

➤ **Las aplicaciones de BI.**

Las tareas que se realizan en este camino permiten el diseño y desarrollo de las aplicaciones de negocio.

1.4.2 SELECCIÓN DE LA METODOLOGÍA

La metodología de Inmon, por su enfoque Top-Down, hace que en un proyecto BI empiece en un nivel superior y vaya hacia el nivel más bajo, con mayor detalle. Esta metodología es ideal para el desarrollo de proyectos complejos de gran envergadura, en los que la información generada perdure y sean consistentes; aun cuando los procesos de la organización vayan cambiando a lo largo del tiempo.

La metodología de Kimball permite el desarrollo de proyectos BI pequeños (generalmente departamentales, los mismos que luego serán acoplados en un *data warehouse* corporativo), este enfoque se centra en la accesibilidad del usuario y además facilita que la solución BI se ajuste a los cambios que aparecen durante su desarrollo.

En este proyecto se utilizó la metodología de Kimball, metodología que permite crear el sistema OLAP a partir de los metadatos de la Revista Politécnica. Teniendo en cuenta el enfoque dimensional se deja la opción de ampliar e ir acoplando nuevos cubos con la información de los demás repositorios académicos de la EPN (Escuela Politécnica Nacional).

1.5 SELECCIÓN Y JUSTIFICACIÓN DE LAS HERRAMIENTAS DE DESARROLLO

Para el desarrollo de este Proyecto son necesarias varias herramientas que permitan la limpieza, el almacenamiento y la publicación de Linked Data, que se han agrupado en:

- Motores de almacenamiento RDF
- Herramientas para limpieza y transformación de datos
- Wikis Semánticos
- Visualizador de Datos Estadísticos

1.5.1 MOTORES DE ALMACENAMIENTO RDF

Un motor de almacenamiento RDF es un software destinado al almacenamiento de grafos RDF, entre los principales tenemos al servidor Virtuoso OpenLink y Sesame.

1.5.1.1 OpenLink Virtuoso

OpenLink Virtuoso es un SQL-ORDBMS y Servidor de Aplicaciones Web híbrido que provee la administración de datos SQL, XML y RDF en un proceso multi hilo

[13]. También funciona como un servidor Web de documentos, servidor de Linked Data y servidor de aplicaciones Web. En lo relacionado a Linked Data incluye el endpoint SPARQL para realizar consultas directamente sobre grafos RDF. En la Figura 10 se muestra la arquitectura de Virtuoso OpenLink.

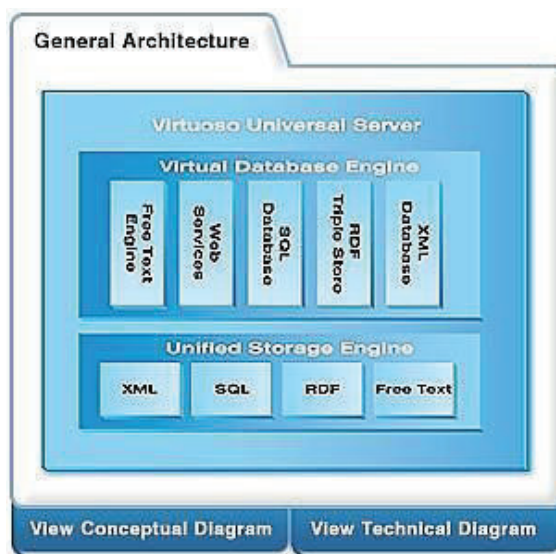


Figura 10 Arquitectura de Virtuoso Open Link [14]

1.5.1.2 Sesame

Sesame es un marco de referencia (*framework*) de Java, de código abierto, para el procesamiento de datos RDF. Sesame permite el análisis, almacenamiento y las consultas sobre estos datos mediante un endpoint SPARQL [14].

1.5.2 HERRAMIENTAS PARA LIMPIEZA Y TRANSFORMACIÓN DE DATOS

Para la elaboración y consolidación de los conjuntos de datos es necesario primero limpiarlos y transformarlos. Las siguientes herramientas facilitan esta tarea.

1.5.2.1 Google Refine

Google Refine es una herramienta que permite la limpieza y transformación de datos. Permite el diseño de la estructura de un grafo RDF que luego puede ser exportado en formato RDF/XML.

1.5.2.2 Data Wrangler

Data Wrangler es una aplicación web que permite la manipulación de datos de forma interactiva. Esta herramienta disminuye considerablemente el tiempo destinado a las tareas de limpieza y transformación de los datos.

1.5.3 WIKIS SEMÁNTICOS

1.5.3.1 OntoWiki

OntoWiki es una aplicación semántica, que permite la administración del conocimiento generado en la Web Semántica. Este conocimiento es organizado mediante una interfaz simple e intuitiva, manipulando clases, propiedades y recursos [15]. La presentación de una base de conocimiento es una especie de mapa de información, con diferentes puntos de vista sobre los datos instanciados.

1.5.3.2 Semantik MediaWiki

Semantik MediaWiki es una extensión de Wikipedia, facilita la búsqueda, organización, etiquetado, evaluación y difusión del contenido de un wiki. Un wiki tradicional contiene texto, no entendible para la máquina, Semantik MediaWiki añade anotaciones semánticas, los enlaces de hipertextos son transformados en atributos RDF, permitiendo al wiki transformarse en parte de una base de datos colaborativa [16].

1.5.4 VISUALIZADORES DATOS ESTADÍSTICOS

1.5.4.1 CubeViz

CubeViz es un navegador por facetas (*faceted browser*), permite visualizar la información de cubos de datos RDF [17]. Este navegador visualiza las observaciones realizadas de forma interactiva, en gráficos de diferente tipo: barras, líneas, columnas y pasteles.

1.5.4.2 Payola

Payola es una aplicación web desarrollada en HTML5, que permite la visualización de datos enlazados que utilizan el vocabulario para cubo de datos RDF. Mediante la utilización de complementos, estos datos pueden ser expresados como tablas, grafos, etc.

1.5.5 SELECCIÓN DE LAS HERRAMIENTAS

Cuando hablamos de herramientas para la publicación de Linked Data, existen varias herramientas que facilitan este proceso. En este trabajo se han seleccionado las que adicionalmente permitan hacerlo con datos estadísticos. Las herramientas seleccionadas son:

- **Virtuoso** como motor de almacenamiento RDF, tiene características similares a las que brinda Sesame, adicional posee métodos de replicación para la información, funciona sobre sistemas operativos: AIX, FreeBSD, HP-UX, Linux, OS X, Solaris y Windows [18]. Permite a la utilización de visualizadores especializados como CubeViz y Payola que interactúan directamente con su almacén de datos. La Figura 11 muestra el panel principal de Virtuoso OpenLink.



Figura 11 Menú principal de Virtuoso
Realizado por el Autor

- **Google Refine**, se usará para la limpieza de datos, a diferencia de Data Wrangler, Google Refine permite la creación de un RDF directamente sobre el resultado de la limpieza de los datos; para finalmente ser exportado en formato RDF/XML. Google Refine también posee la opción de recolectar la información directamente desde la Web, sin necesidad de poseer un archivo en un formato específico. Y en el caso de Linked Data facilita la reconciliación de los datos con fuentes externas, facilitando el enlazamiento de nuestro conjunto de datos.

	País	Mes	Visitas	Nuevos Usuario
1.	Ecuador	2	1.415	707
2.	Ecuador	1	1.293	673
3.	Mexico	2	205	182
4.	Colombia	2	174	161
5.	Mexico	1	107	99
6.	Spain	2	78	62
7.	Colombia	1	59	48
8.	Chile	1	48	31
9.	Peru	1	47	43
10.	Venezuela	2	47	33

Figura 12 Datos en Google Refine
Realizado por el Autor

- **OntoWiki**, se usará como Wiki Semántico, una de las principales características para su elección son sus complementos, dos ejemplos de extensiones son: CubeViz para visualizar cubos de datos RDF y CSVImport

para realizar la transformación de datos estadísticos expresados en CSV a RDF/XML.

OntoWiki permite la edición directa de los conjuntos de datos en RDF. Para el trabajo colaborativo, OntoWiki indica el seguimiento de los usuarios y todas las modificaciones que se hayan realizado en la base de conocimiento. La Figura 13 muestra la presentación de una base de conocimiento en OntoWiki.

The screenshot displays the OntoWiki web interface. At the top, there is a navigation bar with 'User', 'Extras', and 'Help' links, and a search box labeled 'Search for Resources'. The main content area is divided into several panels:

- Data Selection:** Contains options to 'Select a part of the Dataset' (EPN Journal Visits...), 'Select Unit and Measurement' (dcs.sessions), and 'Configure the Dimensions' (City: 10 of 100 Selected, Country: 10 of 38 Selected). It also includes 'Share' and 'Update Chart' buttons.
- Resource List:** Shows a list of instances for the resource 'qb:Observation'. The list is numbered 1 through 8, with each entry showing a URI (e.g., ds.obs183) and the resource name. A 'View' tab is active, and there are 'Instances', 'History', and 'Source' tabs. An 'Add Instance' button is visible.
- Show Properties:** A panel on the right showing the RDF type 'qb:Observation' and its associated properties: 'qb:dataSet', 'Country', 'City', 'Sessions', and 'about:1'.
- Filter:** A panel on the right with an 'Edit' button and a search box. It shows 'Active Filters: Type: qb:Observation'.

At the bottom of the interface, there is a pagination bar with 'First', 'Previous', '1', '2', '3', 'Next', and 'Last' buttons. It also displays 'Search returned 209 results.' and 'Show me: 10 20 30 40 50'. A footer note states 'Query execution took 1209 ms.'

*Figura 13 Presentación de una base de conocimiento en OntoWiki
Realizado por el Autor*

- **CubeViz** se usará para la visualización de datos estadísticos, a más de la visualización y análisis de todos los componentes del Cubo de Datos RDF en CubeViz podemos crear nuestros propios conjuntos de datos estadísticos, los mismos que serán almacenados directamente en Virtuoso. La Figura 14 muestra el gráfico de pastel de CubeViz a partir de un conjunto de datos de ejemplo.

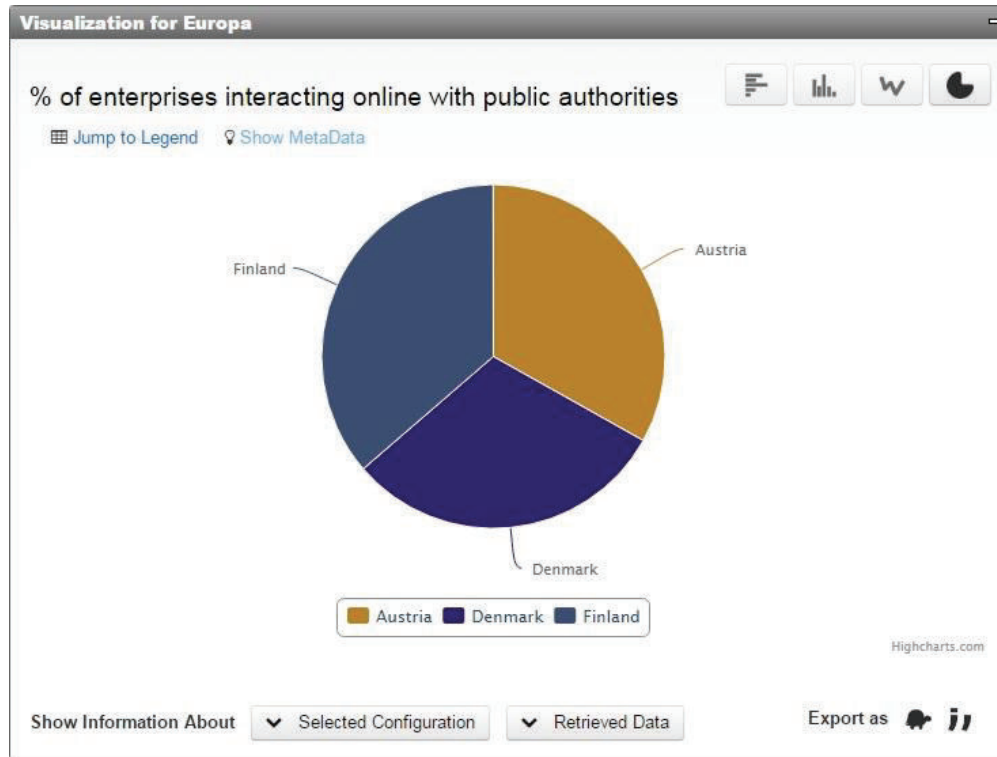


Figura 14 Gráfico de pastel en Cube Viz
Realizado por el Autor

CAPÍTULO 2 DESARROLLO DEL SISTEMA

En este capítulo se realiza el desarrollo del sistema OLAP para bibliotecas digitales que utilizan Linked Data. Este desarrollo seguirá el ciclo de vida de un proyecto BI que indica la metodología de Kimball. De esta metodología se han seleccionado varias etapas teniendo en cuenta las Buenas Prácticas Para la Publicación de Linked Data que indica el W3C (World Wide Web Consortium).

Este capítulo se encuentra distribuido de la siguiente manera: en el apartado 2.1 se realiza la tarea de especificación de requerimientos del sistema OLAP para bibliotecas digitales que utilizan Linked Data, posteriormente se realiza el diseño del mismo, incluido el diseño del subsistema ETL y la elaboración del cubo de datos RDF, finalmente se realiza la implantación del sistema.

2.1 ESPECIFICACIÓN DE REQUERIMIENTOS

2.1.1 CASO DE ESTUDIO

Como se ha dicho anteriormente, el desarrollo de este sistema OLAP es utilizable para cualquier biblioteca digital que utilice tecnologías de Linked Data, que vea necesaria publicar su información estadística en forma de conjuntos de datos en la Web Semántica. El caso de estudio del presente proyecto es la Revista Politécnica.

La Revista Politécnica es el repositorio digital de la Revista Científica publicada por la EPN, la Revista Politécnica alberga todos los artículos publicados por sus investigadores e implementa ciertas tecnologías para la publicación de metadatos, como es el caso del protocolo OAI-PMH.

2.1.2 CASOS DE USO

El diagrama de la Figura 15 representa las tareas que pueden llevar a cabo en el sistema los usuarios. Se han identificado 3 tipos de actores:

- **Usuario Administrador**, es el encargado de manipular los conjuntos de datos estadísticos alojados en el sistema. Desde su creación hasta su eliminación.
- **Usuario**, es la persona que consultará la información contenida en los conjuntos de datos. Esta información es presentada mediante gráficos estadísticos, las consultas también podrán llevarse a cabo desde un endpoint SPARQL.
- **Usuario máquina**, es cualquier dispositivo que pueda utilizar la información publicada en el sistema mediante consultas SPARQL.

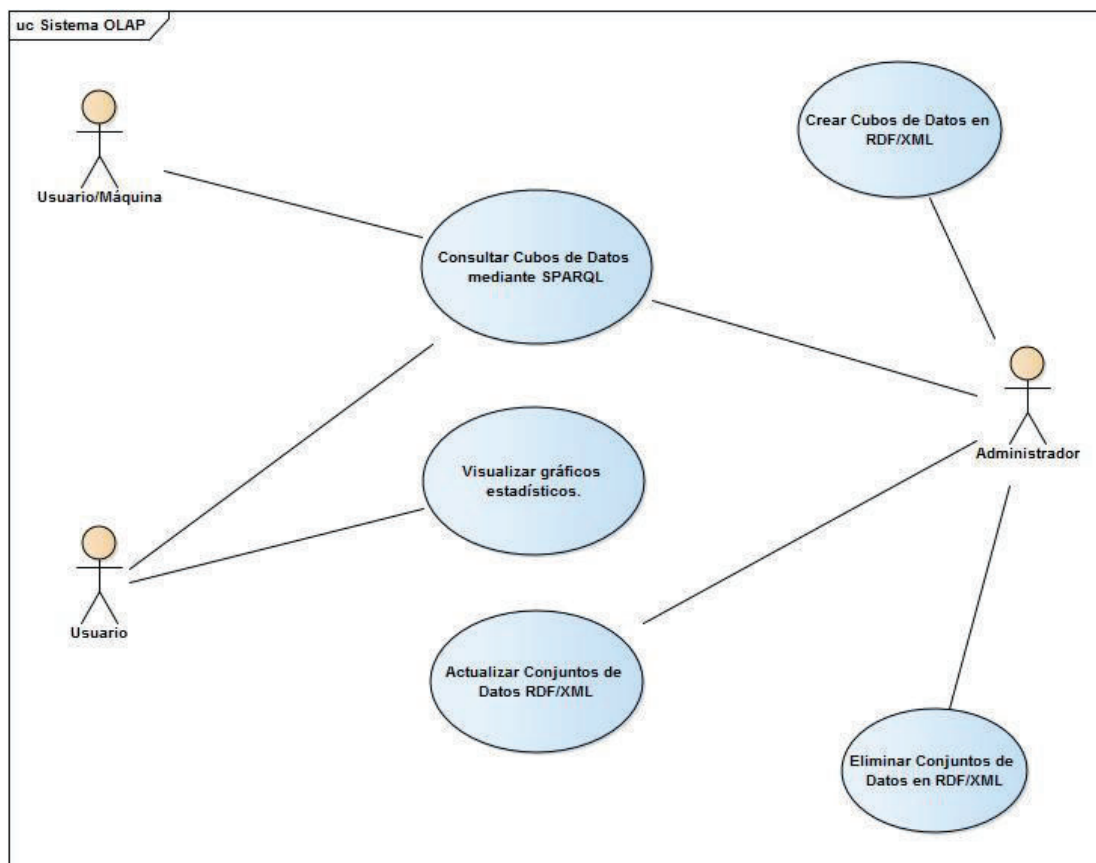


Figura 15 Caso de uso del sistema OLAP
Realizado por el Autor

2.1.3 OBJETIVOS DEL SISTEMA

El presente trabajo tiene como objetivo principal el desarrollo de un sistema OLAP para bibliotecas digitales que usan tecnologías Linked Data.

Los objetivos específicos del desarrollo del sistema OLAP son los siguientes:

- Analizar metodologías y estrategias de desarrollo de sistemas OLAP usando datos enlazados en la Web Semántica.
- Desarrollar un sistema OLAP para bibliotecas digitales en base a datos enlazados.
- Probar el sistema con datos de una biblioteca digital.

El sistema OLAP para bibliotecas digitales que usan tecnologías Linked Data presentará información estadística de estos repositorios, con la utilización de tecnologías de Linked Data y herramientas de código abierto.

Las tecnologías para Linked Data, después de ser implementadas, como resultado brindan conjuntos de datos entendibles tanto para seres humanos como para máquinas. Por lo que es necesario especificar los siguientes objetivos particulares, que se han definido a partir de los principios para la publicación de datos enlazados:

Para ser entendible por maquinas el sistema debe:

- Permitir la manipulación de archivos RDF/XML.
- Utilizar el vocabulario para cubos de Datos RDF.
- Los cubos de datos RDF deben ser creados sobre los principio para publicar en la Web Semántica.

A su vez, el sistema, para un usuario humano debe:

- Presentar información gráfica, en la que se permita identificar dimensiones y medidas.
- La información publicada debe estar enlazada a conjuntos de datos externos, permitiendo el descubrimiento de nueva información.
- Los conjuntos de datos deben ser publicados en repositorios que permitan la reutilización de los mismos.

2.2 ANÁLISIS Y DISEÑO

2.2.1 MODELO DIMENSIONAL

Un modelo dimensional permite organizar la información en un sistema BI, facilitando su comprensión y análisis. Generalmente, por la forma en la que se almacena esta información, puede ser entendida como un cubo con tres dimensiones, que permite tener acceso a la información por cualquiera de estas.

Una de las principales características de un modelo de datos dimensional es que el análisis de su información se la puede realizar mediante cubos OLAP. Adicionalmente, este tipo de modelo de datos es altamente escalable, permitiendo la agregación de nuevas dimensiones y hechos según requerimientos futuros del negocio.

La información se estructura en hechos y dimensiones. Las dimensiones nos dan diferentes puntos de vista de la información a ser analizada, las mismas que están formadas por atributos y poseen una estructura jerárquica. Los hechos a su vez son quienes dotan de sentido al análisis dimensional, son los indicadores resultantes de este análisis.

En la Figura 16 se indica una aproximación del modelo de datos dimensional de nuestro cubo de datos RDF, en el que se han identificado:

- **Artículo (*Article*)**
Los artículos son todos los recursos alojados en la Revista Politécnica, sobre los cuales se contarán las visitas.
- **Creador (*Creator*)**
El creador es el autor del artículo sobre el cual se medirán las visitas.

- **Palabra clave (Keyword)**
Las palabras claves son los metadatos asociados al artículo, facilita la búsqueda de este y permite el enlazamiento a recursos externos.
- **Fecha de visita (ViewDate)**
La fecha de visita indica el año y mes de la visita.
- **País (Country)**
El país indica el nombre del lugar geográfico desde donde se realizaron las visitas a los artículos de la Revista Politécnica.

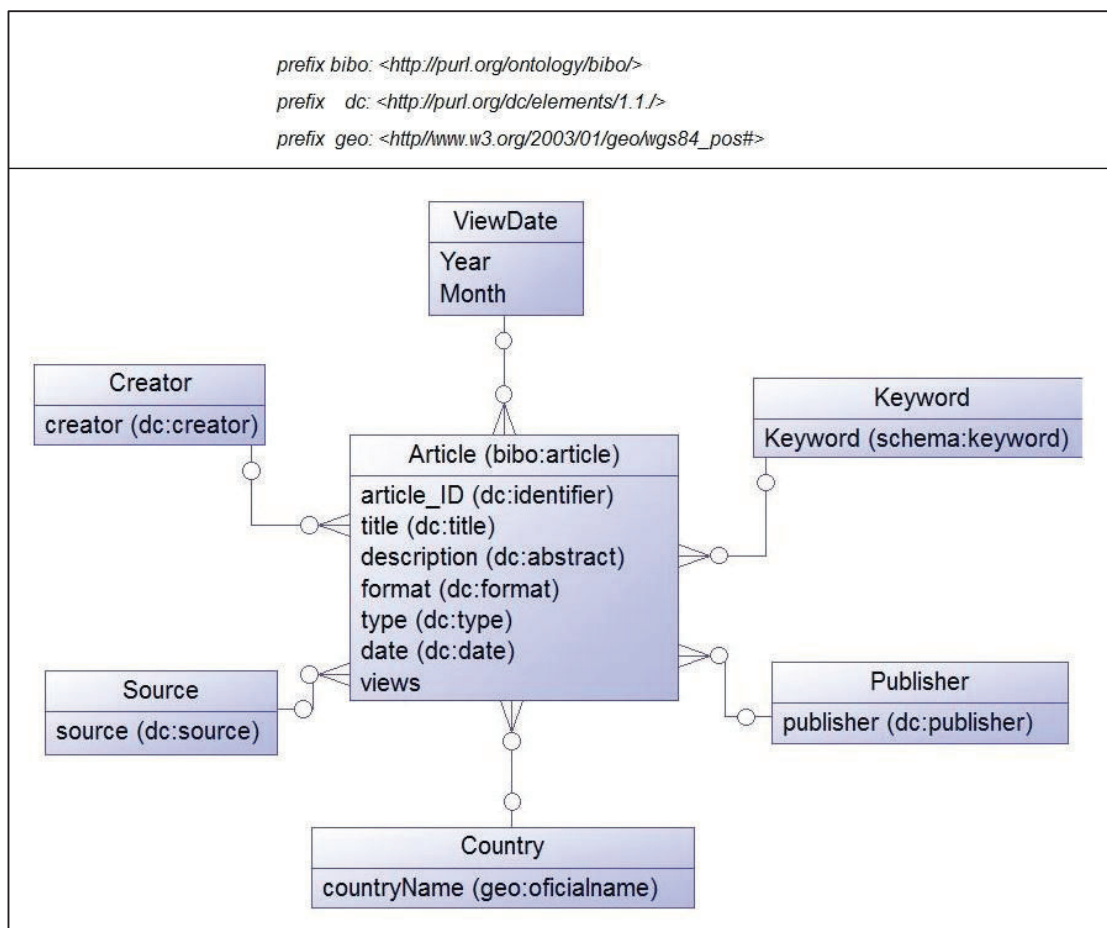


Figura 16 Diagrama dimensional
Realizado por el Autor

La Figura 17 representa la estructura del modelo de datos dimensional, con la utilización del vocabulario para cubos de datos RDF.

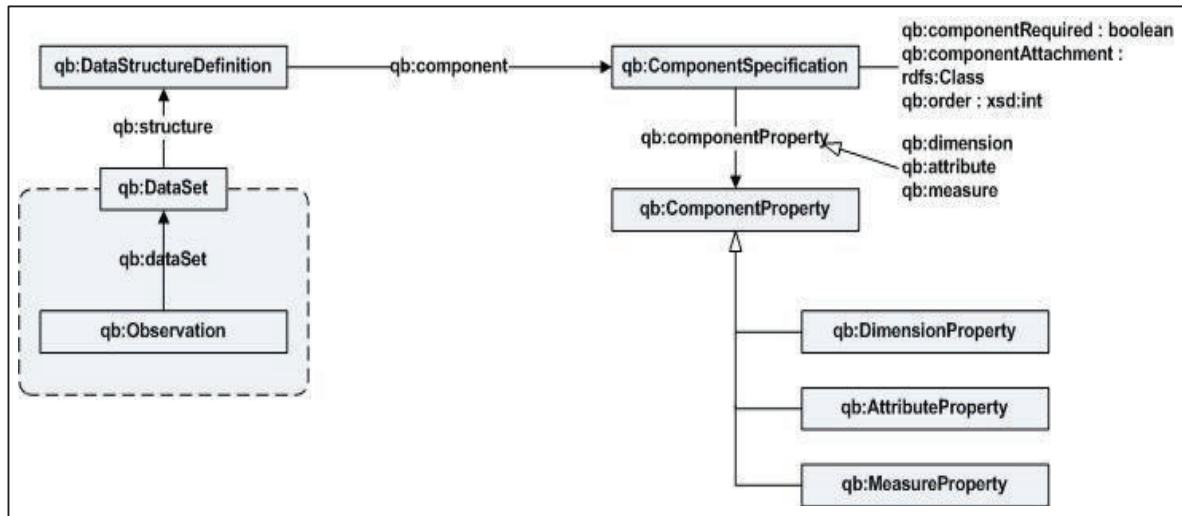


Figura 17 Representación del modelo de datos dimensional con el vocabulario de cubo de datos RDF
Realizado por el Autor

2.2.2 DISEÑO DE ETL Y DESARROLLO

Para el diseño y desarrollo del subsistema ETL ha sido tomada en cuenta la guía de buenas prácticas para la publicación de Linked Data del W3C [19]. El diagrama que recoge la Figura 18 muestra algunas de las tareas de esta guía.

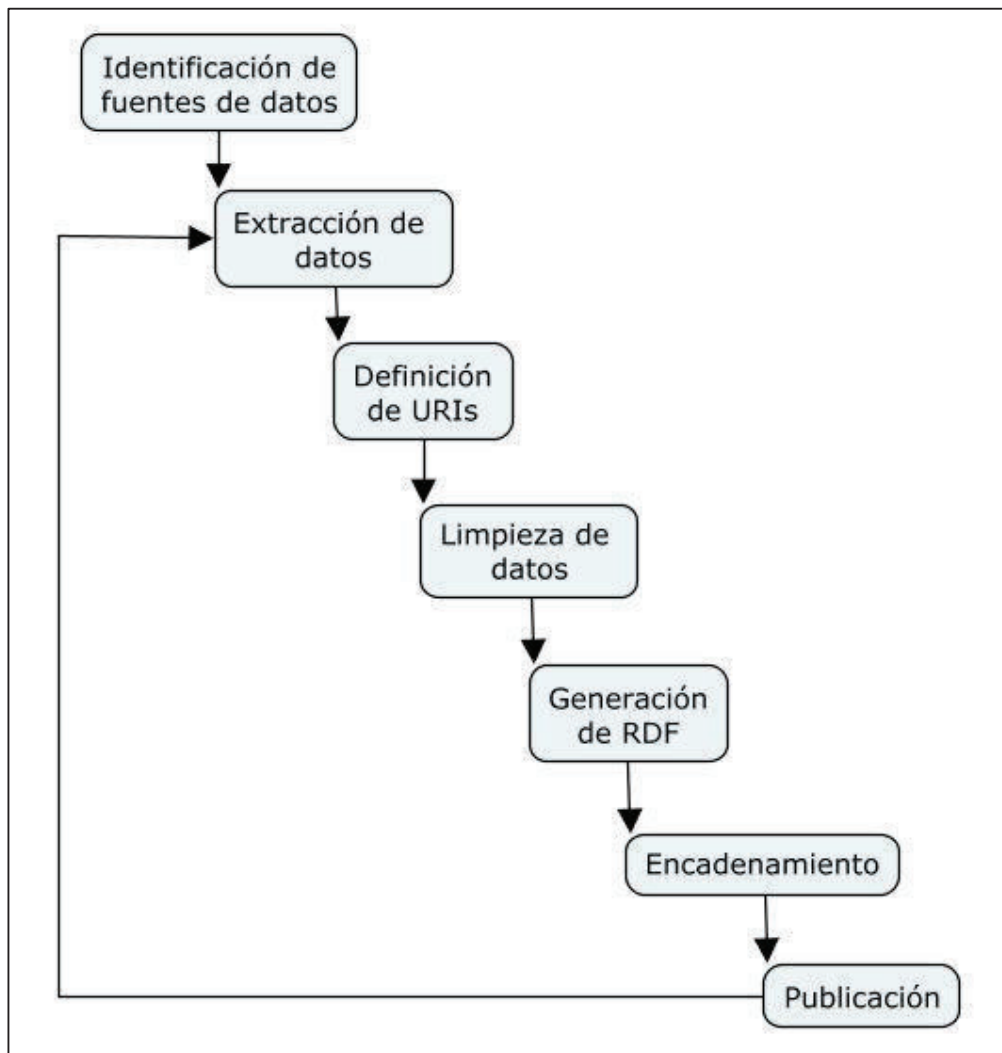


Figura 18 Proceso para la publicación de Linked Data [19]

2.2.2.1 Identificación de fuentes de datos

➤ Metadatos de la Revista Politécnica

Los metadatos son el punto de partida para el desarrollo del subsistema ETL, es importante identificar la fuente de extracción de estos metadatos correctamente.

El acceso a la base de datos del OJS (*Open Journal System*) que aloja a la Revista Politécnica es limitado, pero el protocolo OAI-PMH permite obtener todos los metadatos de los recursos de esta. Esta recolección se la realiza accediendo a la siguiente dirección: <http://revistapolitecnica.epn.edu.ec/ojs2/index.php/index/oai>.

En esta dirección podremos elegir entre los diferentes conjuntos de datos que se muestran en la Figura 19.

Datestamp of response	2016-04-07T16:38:08Z
Request URL	http://revistapolitecnica.epn.edu.ec/ajs2/index.php/index/oai
Request was of type ListSets.	
Set	
setName	Revista Politécnica
setSpec	revista_politecnica2 Identifiers Records
Set	
setName	Articulos
setSpec	revista_politecnica2:ART Identifiers Records

*Figura 19 Lista de conjuntos de datos en OJS
Realizado por el Autor*

El conjunto de datos elegido en este trabajo es “Artículos”. Un ejemplo de la información respecto a un artículo proporcionado por el protocolo OAI-PMH se encuentra en la Tabla 1.

Tabla 2 Metadatos de un artículo obtenido mediante OAI-PMH
Realizado por el Autor

Metadatos Dublin Core ®	Valor
Title	Methodologies and Programming Development of Applications on Modeling, Identification and Simulation
Author or Creator	Burbano, Patricio; Escuela Politécnica Nacional
Author or Creator	Rosales, Andrés
Author or Creator	Cabezas, Lady; Escuela Politécnica Nacional, Departamento de Automatización y Control
Subject and Keywords	Electrónica; Control
Description	Resumen.- La modelación, simulación e identificación de prototipos permiten realizar predicciones acertadas del comportamiento de los sistemas, además de hallar las expresiones matemáticas tales como funciones de transferencia, ecuaciones diferenciales y variables de estado necesarias para la implementación de controladores, sin necesidad de la implementación de los prototipos reales, lo que permite simular en un entorno similar las perturbaciones o simplemente observar su respuesta en condiciones normales. Este trabajo presenta la modelación de varios sistemas, utilizando diversos métodos sean éstos directos, de redes o variacional, además del desarrollo de metodologías enfocadas al desarrollo de casos de estudio, a través de los cuales se implementa cinemática inversa, transformaciones homogéneas, métodos de Newton-Euler, Euler-Lagrange y otros que facilitan la simulación de sistemas a través de Matlab.
Publisher	Escuela Politécnica Nacional
Date	2013-10-15
Resource Type	info:eu-repo/semantics/article
Resource Type	Peer-reviewed Article
Resource Type	info:eu-repo/semantics/publishedVersion
Format	application/pdf
Resource Identifier	http://revistapolitecnica.epn.edu.ec/ojs2/index.php/revista_politecnica2/article/view/60
Source	Revista Politécnica; Vol 32 (2013): Revista Politécnica (Julio)
Source	Revista Politécnica; Vol 32 (2013): Revista Politécnica (Julio)
Relation	http://revistapolitecnica.epn.edu.ec/ojs2/index.php/revista_politecnica2/article/view/60/pdf
Coverage	Robótica, Modelación, Simulación, Identificación, Procesos, Modeling, Identification, Simulation, Robotics, Process.
Rights Management	Los autores que publican en esta revista están de acuerdo con los siguientes términos: Los autores conservan los derechos de autor y garantizan a la revista el derecho de ser la primera publicación del trabajo al igual que licenciado bajo una Creative Commons Attribution License que permite a otros compartir el trabajo con un reconocimiento de la autoría del trabajo y la publicación inicial en esta revista. Los autores pueden establecer por separado acuerdos adicionales para la distribución no exclusiva de la versión de la obra publicada en la revista (por ejemplo, situarlo en un repositorio institucional o publicarlo en un libro), con un reconocimiento de su publicación inicial en esta revista. Se permite y se anima a los autores a difundir sus trabajos electrónicamente (por ejemplo, en repositorios institucionales o en su propio sitio web) antes y durante el proceso de envío, ya que puede dar lugar a intercambios productivos, así como a una citación más temprana y mayor de los trabajos publicados (Véase The Effect of Open Access) (en inglés).

➤ Google Analytics

Las visitas que recibe cada artículo son registradas por la herramienta Google Analytics, los informes que nos presenta pueden ser personalizados a elección del usuario. Pudiendo elegir la fecha, el lugar, tiempo de duración, país, ciudad desde donde se realiza la visita al mismo.

Una visita es considerada como válida para este proyecto, cuando el usuario ingresa al sitio de la Revista Politécnica e ingresa al módulo donde se presenta la información del artículo. Un ejemplo de visita valida se indica en la Figura 20, la misma que será registrada por Google Analytics.

The screenshot shows the homepage of the 'Revista Politécnica' website. At the top, there is a navigation menu with links: INICIO, ACERCA DE..., INGRESAR, REGISTRO ANUNCIOS, BUSCAR, ACTUAL, and ARCHIVOS. The main content area features an article titled 'Análisis Crítico de la Economía Solidaria en el Ecuador' by Steevens Oswaldo Góngora Almeida and Cynthia Johanna Ramos Caizaluisa. The article includes a 'Resumen' (Summary) and a full text section. The website also has a sidebar with 'Notificaciones', 'Volumen 37', and 'AYUDA' sections. On the right side, there are user login fields, language selection (Español), font size adjustment, and a search bar.

*Figura 20 Visita considerada como valida
Elaborada por el Autor*

2.2.2.2 Extracción de datos

El resultado de la consulta mediante el protocolo OAI-PMH es un archivo en formato XML, archivo que contiene los metadatos de todos los artículos de la Revista Politécnica. De Google Analytics el resultado es un archivo en formato CSV (*Comma-Separated Values*), que contiene únicamente los valores de las dimensiones y métricas seleccionadas en un informe personalizado.

Para el informe personalizado de Google Analytics se eligió las dimensiones y métricas que se indican en la Tabla 2. El periodo de análisis seleccionado inicia en el 1 de Octubre de 2015 y finaliza el 31 de Marzo de 2016.

Tabla 3 Dimensiones y medidas
Realizado por el Autor

Dimensiones	Página de destino	Indica la URL de la página que ha sido visitada por el usuario.
	País	País del que procede la visita.
	Año	Año del periodo de análisis seleccionado.
	Mes del año	Mes del año del periodo de análisis seleccionado
Métricas	Sesiones	Número total de sesiones que se han realizado en el periodo de análisis seleccionado. Una sesión es el periodo durante el cual un usuario interactúa con el sitio web, aplicación, etc. Todos los datos de uso (visitas a una pantalla, eventos, comercio electrónico, etc.) están asociados a una sesión.

2.2.2.3 Diseño de URI

En el diseño de las URIs se definieron los patrones que deben seguir las URIs de los componentes que forman parte del cubo de datos RDF.

La Tabla 3 muestra las URIs principales que se utilizaron en este proyecto, a partir de la URI base <http://opendata.epn.edu.ec/>. Esta URI base, seguida de la subcarpeta “cubos” es donde se alojó el cubo de datos RDF y se podrá alojar cualquier otro cubo de datos RDF creado con la información de la Revista Politécnica.

Tabla 4 Diseño de URIs principales
Realizado por el Autor

Estructura de la URI	Patrón del componente	Prefijo	URI para este proyecto
Nombre del cubo	{nombre del cubo}		revista_epn
URI base	{http://ejemplo.org/}		http://opendata.epn.edu.ec/
Cubo de Datos RDF	{URI base}cubos/ {nombre del cubo}		http://opendata.epn.edu.ec/cubos/revista_epn
Conjunto de datos	{URI Base}cubos/ {nombre del cubo}/ dataset/	[ds]	http://opendata.epn.edu.ec/cubos/revista_epn/dataset/
Propiedades	{URI Base}dc/ {nombre del cubo}/ prop/	[prop]	http://opendata.epn.edu.ec/cubos/revista_epn/prop/
Especificación de los componentes de los Cubos de Datos	{URI Base}dc/ {nombre del cubo}/ dccc/	[dccc]	http://opendata.epn.edu.ec/cubos/revista_epn/dccc/

La Tabla 4 muestra los patrones y las URIs resultantes que corresponden a las medidas y dimensiones que forman parte del cubo de datos RDF. Estas medidas y dimensiones son las que se consiguieron identificar en las fuentes de datos.

Dimensiones:

- País desde donde se realiza la visita.
- Año de la visita.
- Mes de la visita.
- Título del artículo visitado.

Medidas:

- Número de visitas a un artículo.

*Tabla 5 Diseño de URIs para dimensiones y medidas del cubo de datos RDF
Realizado por el Autor*

Estructura de la URI	Patrón del componente	URI para este proyecto
Dimensión	{URI base}dc/cubos/ {nombre del cubo}/ prop/{nombre de la dimensión}	http://opendata.epn.edu.ec/cubos/revista_epn/prop/pais http://opendata.epn.edu.ec/cubos/revista_epn/prop/año http://opendata.epn.edu.ec/cubos/revista_epn/prop/mes http://opendata.epn.edu.ec/cubos/revista_epn/prop/titulo
Medidas	{URI base}dc/cubos/ {nombre del cubo}/ prop/{nombre de la medida}	http://opendata.epn.edu.ec/cubos/revista_epn/prop/visita

2.2.2.4 Limpieza de datos

Los dos conjuntos de datos obtenidos son limpiados por separado. Para luego ser consolidados en un único conjunto de datos. La Figura 21 recoge gráficamente las tareas que se llevaron a cabo para obtener un conjunto de datos consolidado de la información de la Revista Politécnica.

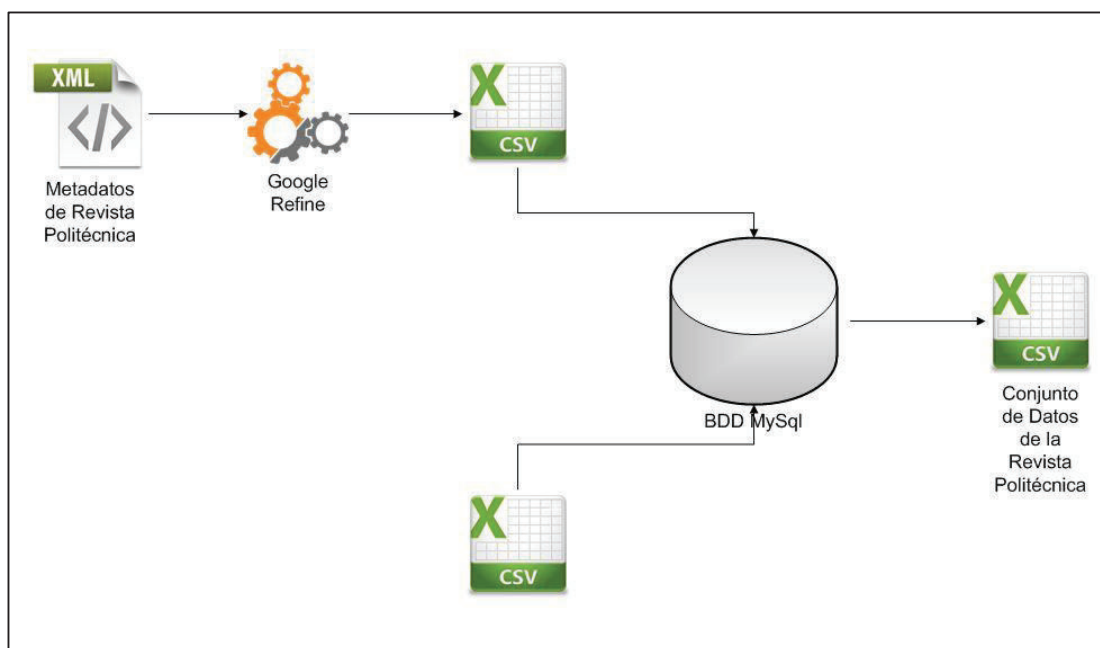


Figura 21 Consolidación de conjunto de datos
Realizado por el Autor

En los siguientes párrafos se describen varias funciones necesarias para cumplir con el proceso de la Figura 21.

La limpieza de datos se consigue mediante la utilización de Google Refine y sus funciones que se describen a continuación:

➤ **Reemplazar valores (*Replace*)**

La función *replace()* de Google Refine permite reemplazar una cadena de caracteres existente por una nueva, permitiendo así la transformación de los registros.

▼ Mes del año		▼ Mes del año
11		11
12		Diciembre
2		2
10		10
1	value.replace("1","Diciembre") →	1
1		1
12		Diciembre
2		2
10		10
10		10

Figura 22 Función Replace de Google Refine
Realizado por el Autor

➤ Separar en columnas (Split into several columns)

Esta función permite dividir una columna en varias, a partir de una cadena de caracteres dentro del registro que hace de punto separador. Otra forma de división es especificando el número de caracteres de la cadena que irán en una columna, quedando el resto de caracteres para la nueva columna. Tal como se puede observar en el ejemplo de la Figura 23.

▼ Página de destino		▼ Página de destino 1	▼ Página de destino
/ojs2/index.php/revista_politecnica2/issue/current		/ojs2/index.php/revista_politecnica2/issue/current	
/ojs2/index.php/revista_politecnica2/article/view/11		/ojs2/index.php/revista_politecnica2/article/	/11
/ojs2/index.php/revista_politecnica2/pages/view/plantillas		/ojs2/index.php/revista_politecnica2/pages/	/plantillas
/ojs2/index.php/revista_politecnica2/announcement/view/6		/ojs2/index.php/revista_politecnica2/announcement/	/6
/ojs2/index.php/revista_politecnica2/article/view/11	«views» →	/ojs2/index.php/revista_politecnica2/article/	/11
/ojs2/index.php/revista_politecnica2/pages/view/plantillas		/ojs2/index.php/revista_politecnica2/pages/	/plantillas
/ojs2/index.php/revista_politecnica2		/ojs2/index.php/revista_politecnica2	
/ojs2/index.php/revista_politecnica2/article/view/11		/ojs2/index.php/revista_politecnica2/article/	/11
/ojs2/index.php/revista_politecnica2/article/view/77		/ojs2/index.php/revista_politecnica2/article/	/77
/ojs2/index.php/revista_politecnica2/pages/view/plantillas		/ojs2/index.php/revista_politecnica2/pages/	/plantillas

Figura 23 Separación en varias columnas en Google Refine
Realizado por el Autor

➤ **Unir valores de celdas (*Join multi-valued cell*)**

Esta opción permite la unión de los registros de diferentes celdas de una misma columna en una sola celda. Para unir estos registros es necesario indicar el caracter con el cual quedarán unidos los registros, como se indica en la Figura 24.

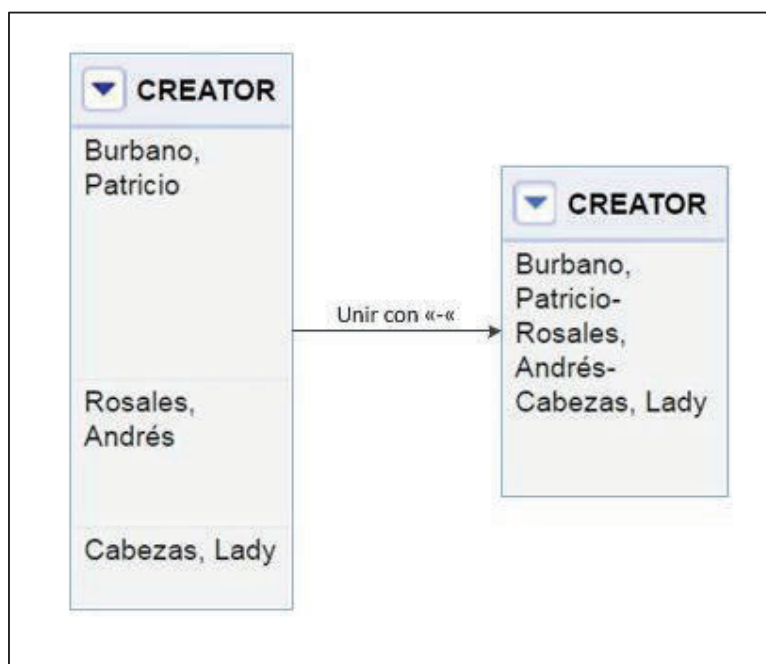


Figura 24 Unión de diferentes registros en Google Refine
Realizado por el Autor

Filtro (Facet)

Google Refine permite agrupar y filtrar el contenido de una columna, una vez agrupado el contenido, los registros resultantes pueden ser manipulados libremente, sin afectar al resto de registros. Los criterios de agrupamiento de Google Refine son:

- Texto
- Numero
- Línea de tiempo
- Celdas en blanco

La Figura 25 muestra la agrupación de la información por el título de los artículos de nuestro conjunto de datos.

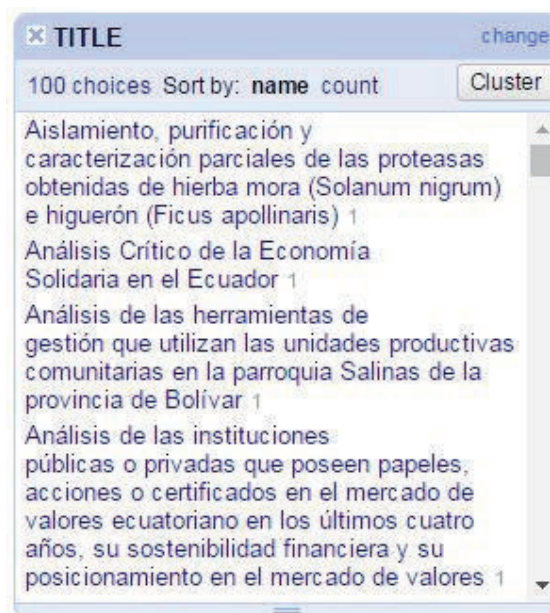


Figura 25 Agrupación por títulos de los artículos.
Realizado por el Autor

El archivo XML resultante de la consulta mediante el protocolo OAI-PMH es cargado a Google Refine, este archivo contiene todos los metadatos que describen a un artículo de la Revista Politécnica. En la Figura 26 se muestra parte de los registros previa a la limpieza y transformación que fueron sometidos

All	ListRecords - recc	ListRecords - record -	ListRecords - recc	ListRecords - recc	ListRecords - recc	
☆	1.	Identificación de la actividad proteolítica en especies vegetales presentes en las provincias de Loja y Pastaza	http://www.openarchives.org/OAI/2.0/oai_dc/ http://www.openarchives.org/OAI/2.0/oai_dc.xsd	info:eu-repo/semantics/article	Castillo, Patricio; Escuela Politécnica Nacional	Revista Politécnica; Vol 31 (2012); Revista Politécnica (Diciembre 2012)
☆				Peer-reviewed Article	Gómez, Christian; Escuela Politécnica Nacional	Revista Politécnica; Vol 31 (2012); Revista Politécnica (Diciembre 2012)
☆				info:eu-repo/semantics/publishedVersion	Sinche, Marco; Escuela Politécnica Nacional	
☆					Duchicela, Jessica; Escuela Politécnica Nacional	
☆					Muñoz, Ricardo; Escuela Politécnica Nacional	

Figura 26 Vista de los registros cargados de OAI-PMH
Realizado por el Autor

De todos estos metadatos, utilizando Google Refine y las funciones anteriormente descritas, mantenemos los que se especificaron en el modelo dimensional:

- Título (Title)
- Resumen (Abstract)
- Autor (Creator)
- Fuente (Source)
- Palabras Clave (Keywords)
- Fecha (Date)
- Identificador del artículo (Article id)
- Asignatura (Subject)

La Figura 27 muestra parte de los registros ya limpiados, manteniendo únicamente los metadatos relevantes para este trabajo.

▼ All		▼ TITLE		▼ CREATOR
☆	🗨️	1.	Análisis Crítico de la Economía Solidaria en el Ecuador	Góngora Almeida, Steevens Oswaldo;Ramos Caizaluisa, Cyntia Johanna
☆	🗨️	2.	Análisis Crítico de la Economía Solidaria en el Ecuador	Góngora Almeida, Steevens Oswaldo;Ramos Caizaluisa, Cyntia Johanna
☆	🗨️	3.	Análisis Crítico de la Economía Solidaria en el Ecuador	Góngora Almeida, Steevens Oswaldo;Ramos Caizaluisa, Cyntia Johanna
☆	🗨️	4.	Proceso de soldadura GMAW, para aceros ordinarios al carbón, aceros inoxidables y aluminio	Granja, Mario;Hidalgo, Victor
☆	🗨️	5.	Aprovechamiento de las propiedades funcionales del jengibre (zingiber officinale) en la elaboración de condimento en polvo, infusión filtrante y aromatizante para quema directa.	Acuña, Oswaldo;Torres, Alejandra

*Figura 27 Parte de los registros limpiados
Realizado por el Autor*

El conjunto de datos obtenido del reporte de Google Analytics es limpiado de la misma forma que el conjunto de datos obtenido con el protocolo OAI-PMH. La Figura 26 muestra parte de los datos resultantes luego de este proceso.

▼ All	▼ COUNTRY	▼ YEAR	▼ MONTH	▼ VIEWS
☆ ↻ 1.	Ecuador	2015	Diciembre	56
☆ ↻ 2.	Ecuador	2016	Enero	40
☆ ↻ 3.	Ecuador	2016	Febrero	32
☆ ↻ 4.	Ecuador	2015	Octubre	29
☆ ↻ 5.	Mexico	2016	Febrero	28
☆ ↻ 6.	Mexico	2016	Marzo	27
☆ ↻ 7.	Ecuador	2016	Enero	27
☆ ↻ 8.	Ecuador	2016	Marzo	22
☆ ↻ 9.	Argentina	2016	Marzo	22
☆ ↻ 10.	Ecuador	2016	Febrero	20

*Figura 28 Limpieza de datos
Realizado por el Autor*

Google Refine no cuenta con la opción de unir varios conjuntos de datos, debido a lo cual fue necesaria la creación de dos tablas temporales en la base de datos MySQL. El resultado de la consulta a estas tablas es el conjunto de datos consolidado en formato CSV.

Este conjunto de datos es el que nuevamente fue cargado a Google Refine, pero esta vez para modelar el cubo de datos RDF.

2.2.2.5 Modelamiento del cubo de datos RDF

El vocabulario para cubos de datos RDF es utilizado para la transformación del conjunto de datos a uno que contenga el cubo de datos RDF. Este conjunto de datos posteriormente fue exportado en formato RDF/XML. El modelado del conjunto de datos RDF adicionalmente necesita otros vocabularios, los mismos que se indican en la Tabla 5.

*Tabla 6 Vocabularios adicionales
Realizado por el Autor*

Prefijo	URI	Descripción
rdfs	http://www.w3.org/2000/01/RDF-schema#	Para describir clases y propiedades de un recurso.
xsd	http://www.w3.org/2001/XMLSchema#	Para describir tipos de datos.
dcat	http://www.w3.org/ns/dcat#	Para la descripción de conjuntos de datos.
dc	http://purl.org/dc/elements/1.1/	Para describir los metadatos de un recurso bibliográfico.
prov	http://www.w3.org/ns/prov#	Indicia de donde proceden los datos.

El conjunto de datos consolidados, obtenidos en el punto anterior, fue cargado a Google Refine, al que previamente se instaló la extensión para RDF (ANEXO A). Esta extensión permite, entre otras cosas, definir la estructura del cubo de datos RDF y exportar el resultado en formato RDF/XML o TTL.

Definir la estructura del Grafo RDF (*RDF Skeleton*)

Google Refine permite la definición del esqueleto o estructura mediante una interfaz gráfica, como muestra la Figura 29, en la cual se especifican como será generado el cubo de datos RDF a partir de nuestro conjunto de datos. Esta estructura está formada por nodos, los que pueden tomar el valor de los registros de nuestro conjunto de datos según la necesidad.

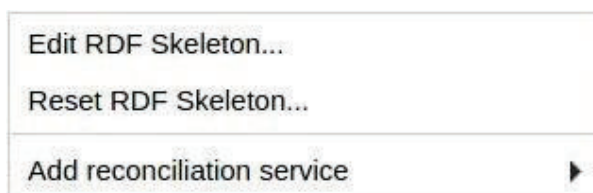


Figura 29 Definición de estructura o esqueleto RDF
Realizado por el Autor

Google Refine permite la exportación de la estructura del grafo RDF, en formato JSON, para ser aplicada a otro conjunto de datos similar, disminuyendo el tiempo en la definición de esta estructura en el modelamiento de conjuntos similares. El ANEXO B muestra la estructura del conjunto de datos generado en este trabajo.

Importación de Vocabularios y agregación de prefijos

Todos los lenguajes utilizados para el modelado del cubo de datos RDF deben ser agregados a Google refine; únicamente ingresando el prefijo del mismo. La Figura 30 muestra la importación del vocabulario para cubos de datos RDF, el prefijo de este vocabulario es qb, después de ingresar este prefijo automáticamente se carga su URI.

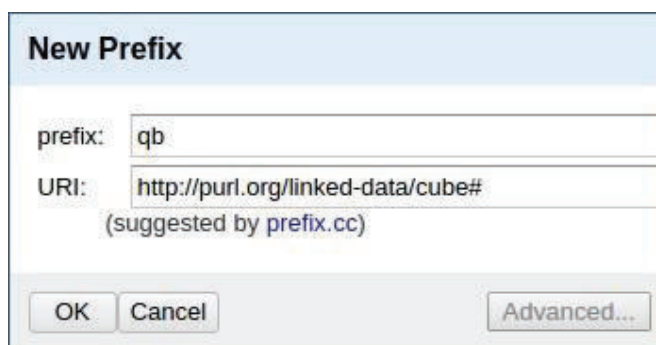


Figura 30 Importación de vocabulario para cubos de datos RDF
Realizado por el Autor

El ingreso de los prefijos que describen los componentes del cubo de datos RDF especificados en la Tabla 3 es similar a los vocabularios, pero adicionalmente habrá que ingresar manualmente la URI de los mismos, hasta finalmente tener todos, vocabularios y componentes, como se indica en la Figura 31.

Prefix	URI	Delete	Refresh
rdfs	http://www.w3.org/2000/01/rdf-schema#	delete	refresh
foaf	http://xmlns.com/foaf/0.1/	delete	refresh
xsd	http://www.w3.org/2001/XMLSchema#	delete	refresh
owl	http://www.w3.org/2002/07/owl#	delete	refresh
rdf	http://www.w3.org/1999/02/22-rdf-syntax-ns#	delete	refresh
qb	http://purl.org/linked-data/cube#	delete	refresh
dcat	http://www.w3.org/ns/dcat#	delete	refresh
dct	http://purl.org/dc/terms/	delete	refresh
prov	http://www.w3.org/ns/prov#	delete	refresh
ds	http://opendata.epn.edu.ec/cubos/revista_epn/dataset/	delete	refresh
prop	http://opendata.epn.edu.ec/cubos/revista_epn/prop/	delete	refresh
dccs	http://opendata.epn.edu.ec/cubos/revista_epn/dccs/	delete	refresh

Figura 31 Lista de vocabularios y prefijos
Realizado por el Autor

La URI principal, igualmente definida en la Tabla 3, debe ser ingresada en el campo Base URI de la interfaz gráfica para el modelamiento del cubo de datos RDF.

Definición de dimensiones y medidas

Una dimensión o una medida se define mediante la propiedad *rdf:type*. La propiedad *rdf:type* indica que un recurso es la instancia de una clase [21]. Los recursos deben poseer las mismas características de la clase a la que instancian.

En Google Refine agregamos un nodo para cada una de las dimensiones y medidas, para su configuración seleccionamos: “el contenido de la celda es utilizado como URI”, para el contenido constante utilizamos las URIs definidas en la Tabla 4. La definición de dimensiones y medidas debe quedar como muestra la Figura 32.



Figura 32 Definición de dimensiones y medidas
Realizado por el Autor

Una dimensión puede ser etiquetada con *rdf:label*, *rdf:label* es una instancia de la clase *rdf:property* [21] que permite indicar el nombre de un recurso. Esta etiqueta es configurada de tal forma que “el contenido de la celda es utilizado como etiqueta de lenguaje”, y en el contenido constante ingresamos palabras que describen a la dimensión o medida. La Figura 33 muestra el etiquetado de las dimensiones y medidas.

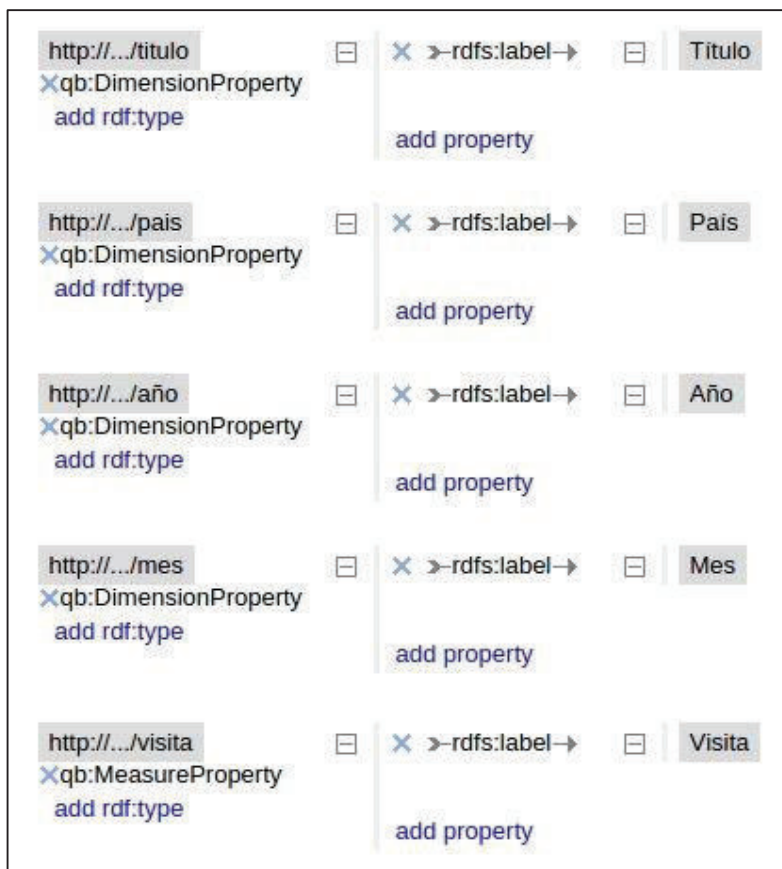


Figura 33 Etiquetado de dimensiones y medidas
Realizado por el Autor

Definición del conjunto de datos, estructura de datos (DSD por sus siglas en inglés) y componentes

La definición del conjunto de datos (*qb:dataset*), y su estructura de datos (*qb:DataStructureDefinition*) están especificados dentro de la misma ubicación: http://opendata.epn.edu.ec/cubos/revista_epn/dataset/. El conjunto de datos contiene todas las observaciones del fenómeno estudiado. Para el presente proyecto, estas observaciones corresponden a las visitas que se han realizado en el periodo de tiempo seleccionado.

La estructura de datos a su vez contiene las especificaciones de los diferentes componentes del cubo de datos RDF o DCCS (por sus siglas en inglés). La ubicación común del conjunto de datos y su estructura hace necesario diferenciarlos mediante prefijos únicos: *dataset-* y *dsd-*, respectivamente, permitiendo obtener URIs únicas para identificar a estos dos elementos.

El conjunto de datos es definido mediante *rdf:type*, sobre un nuevo nodo de la estructura del grafo RDF. El conjunto de datos es configurado de tal forma que “el contenido de la celda es utilizado como URI”, y para el contenido constante utilizaremos la URI definida para el conjunto de datos en la Tabla 4.

La estructura se la agrega mediante una nueva propiedad dentro del mismo nodo en el que está definido el conjunto de datos. Esta estructura es configurada de la siguiente manera: “el contenido de la celda es un nodo en blanco”. A esta estructura le agregamos mediante *rdf:type* la estructura en sí de los datos con *qb:DataSetDefinition*.

Esta Estructura de datos es configurada de la siguiente forma: “el contenido de la celda es utilizado como URI”, en el contenido constante ingresaremos la siguiente URI: http://opendata.epn.edu.ec/cubos/revista_epn/dataset/dsd-revista_epn que identifica a la DSD como indica la Figura 34.

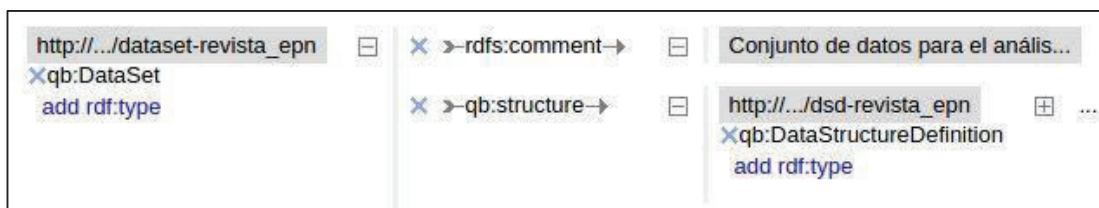


Figura 34 Definición de conjunto y la estructura de datos
Realizado por el Autor.

Definición de los componentes del cubo de datos RDF

Los componentes (*qb:component*) van dentro de la estructura de datos creado previamente, un componente por cada dimensión y medida como se indica en la Figura 35.

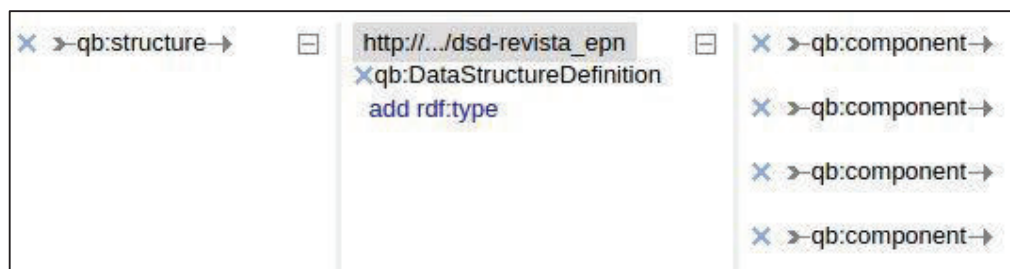


Figura 35 Componentes
Realizado por el Autor

Las especificaciones de cada componente se realiza mediante *qb:ComponentSpecification*, *qb:ComponentSpecification* permite definir las propiedades de un componente que se encuentra especificado dentro de una DSD.

La especificación de componentes es configurada de la siguiente forma: “el contenido de la celda es usado como URI”, en el contenido constante ingresaremos la URI del DCCS, seguido del nombre de la dimensión y medida.

Mediante *rdf:type* agregamos *qb:ComponentSpecification*. Uno por cada dimensión y medida configurada inicialmente como se muestra en la Figura 36.

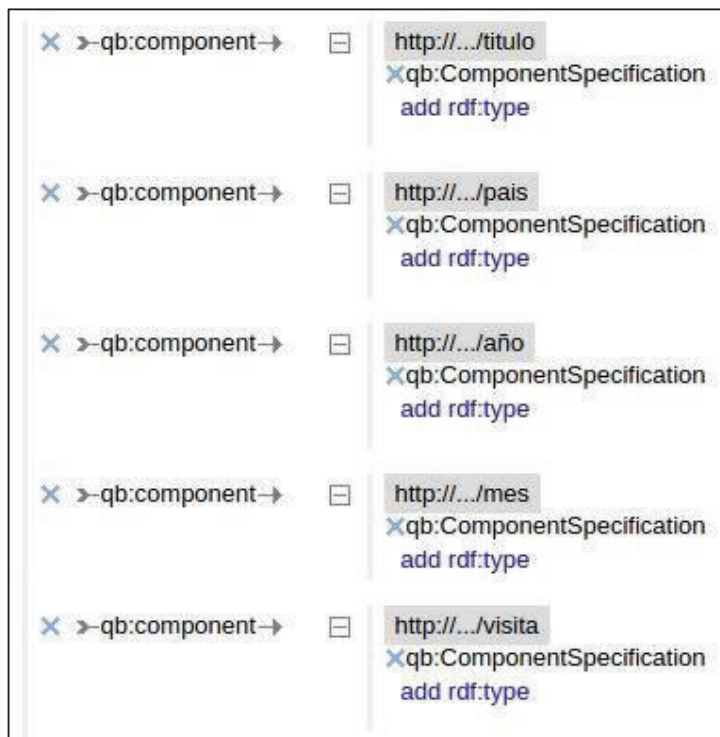


Figura 36 Especificación de componentes
Realizado por el Autor

Propiedades de los componentes

Las propiedades deben ser agregadas para cada uno de los componentes especificados. Las propiedades *qb:dimension* y *qb:measure* permiten indicar explícitamente que un determinado componente es una dimensión o una medida. En la configuración de estas dimensiones y medidas: “el contenido de la celda es usado como URI”. Estas URIs son los valores definidos en la Tabla 4 y que se pueden ver en la Figura 37.



Figura 37 Propiedades qb:dimension y qb:measure de los componentes
Realizado por el Autor.

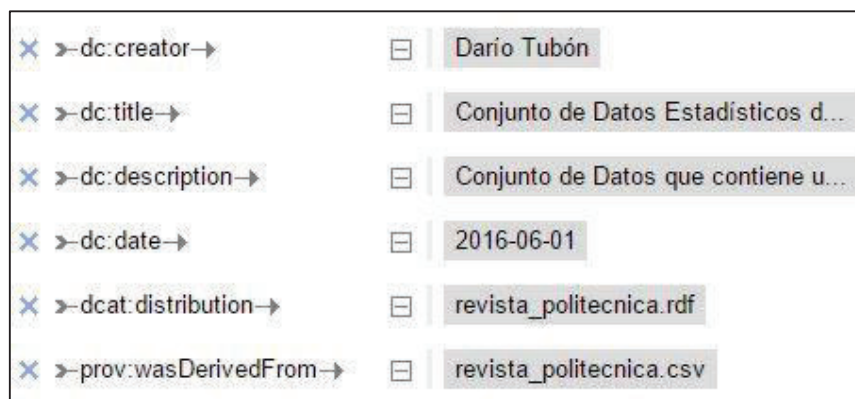
Metadatos para el conjunto de datos creado

Los metadatos que describen el cubo de datos RDF se indican en la Tabla 7. Todas las propiedades de tipo texto deben ser configuradas de la siguiente forma: “el contenido de la celda es utilizado como etiqueta del lenguaje a utilizar”.

Tabla 7 Metadatos del conjunto de datos
Realizado por el Autor

Propiedad	Valor
dc:creator	Dario Tubón
dc:title	Conjunto de Datos Estadísticos de la Revista Politécnica
dc:description	Conjunto de Datos que contiene un cubo de datos RDF, el mismo que representan a las visitas al sitio web de la Revista Politécnica.
dc:date	2016-06-01
dcat:distribution	revista_politecnica.RDF
prov:wasDerivedFrom	revista_politecnica.csv

Los metadatos de la Tabla 7 una vez agregados al cubo de datos RDF se muestran en la Figura 38, estos metadatos permiten la descripción e identificación del conjunto de datos RDF, donde indica su autor, título o nombre, una corta descripción, la fecha de creación, el nombre del archivo con el que será distribuido (archivo en formato RDF) y la fuente a partir del cual se ha creado el conjunto de datos (archivo en formato CSV).



✕ >dc:creator→	☐ Dario Tubón
✕ >dc:title→	☐ Conjunto de Datos Estadísticos d...
✕ >dc:description→	☐ Conjunto de Datos que contiene u...
✕ >dc:date→	☐ 2016-06-01
✕ >dcat:distribution→	☐ revista_politecnica.rdf
✕ >prov:wasDerivedFrom→	☐ revista_politecnica.csv

Figura 38 Metadatos del conjunto de datos
Realizado por el Autor

Definición de las observaciones

En esta sección se indican el proceso para agregar las observaciones al cubo de datos RDF, las observaciones quedan ligadas a cada una de las propiedades de los componentes: dimensiones y medidas definidas. Una observación es representada como una instancia de tipo *qb:Observation* [8]. Representan cualquier simple observación del fenómeno estudiado, la misma que puede tener asociada uno o más valores que permitan medir dicho fenómeno.

Cada observación tiene asociadas un conjunto de propiedades y cuenta con una URI propia y única. Al utilizar Google Refine para la definición de estas

observaciones podemos definir un índice, que facilita la interpretación, nombrando a una observación como obs1, obs2, obs3....obsn.

En un nuevo nodo del grafo RDF en la interfaz de Google Refine, mediante *rdf:type* agregamos la observación, con “el contenido a ser utilizado como URI”. Aquí es importante indicar la columna que contiene nuestras observaciones, que en este caso es la columna VIEWS, el dialogo de configuración para una observación se muestra en la Figura 39.

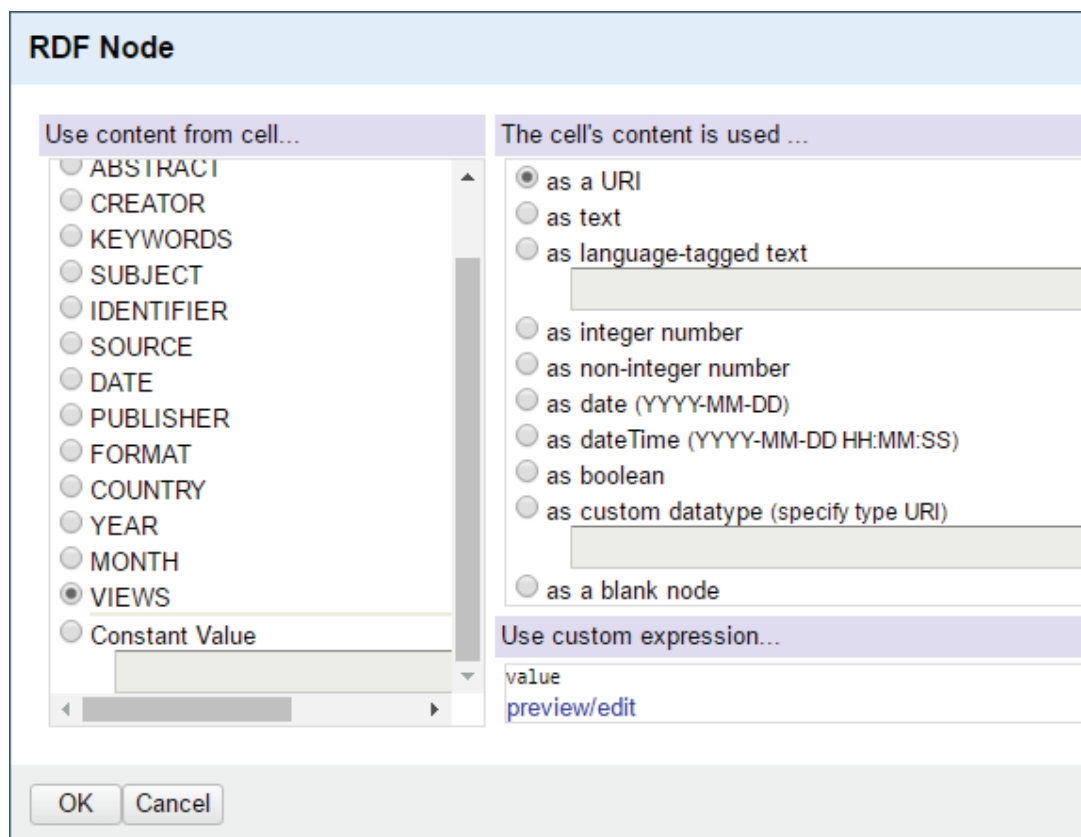
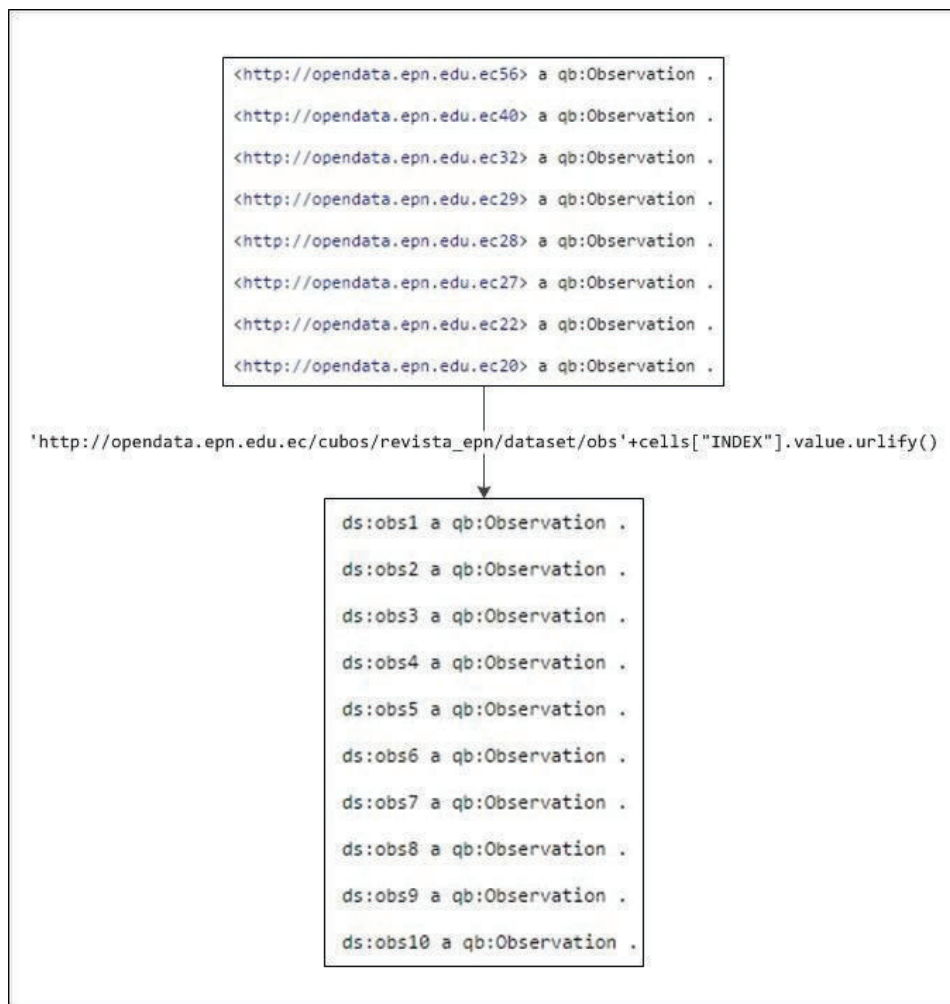


Figura 39 Selección de la columna con las observaciones.
Realizado por el Autor.

Una observación del conjunto de datos RDF debe ser fácilmente interpretada, tanto por máquinas y humanos, lo que se consigue mediante la utilización de una expresión por defecto en Google Refine. Todas estas observaciones se encuentran alojadas bajo /dataset/, cada una con el subíndice que toma de la columna Index. La Figura 40 muestra la definición de una observación, antes y después de agregar su respectivo índice.



*Figura 40 Observaciones en Google Refine
Realizado por el Autor*

A cada una de estas observaciones se las debe agregar sus propiedades: dimensiones y medidas. Cada una de las cuales tiene su propia configuración.

Agregando dimensiones y medidas

Por cada dimensión se agrega una propiedad a la observación, la que es nombrada según la definición de propiedades indicada en la Tabla 4. Estas propiedades, a su vez deben apuntar a las columnas correspondientes que contienen los valores que representan. La diferencia entre la definición de dimensiones y medidas como propiedades de una observación radica en que, la medida obviamente será un número entero que representa el número de visitas, la Figura 41 muestra la estructura de una observación en Google Refine.



Figura 41 Propiedades de las observaciones.
Realizado por el Autor.

En la Figura 42 se observa la definición de una observación en formato TTL con las propiedades hasta el momento agregadas: dimensiones y medidas.

```
ds:obs1 a qb:Observation ;
  prop:pais "Ecuador" ;
  prop:año "2015" ;
  prop:mes "Diciembre" ;
  prop:titulo "Análisis Crítico de la Economía Solidaria en el Ecuador" ;
  prop:visita "56"^^xsd:int .
```

Figura 42 Observación en formato TTL.
Realizado por el Autor.

La definición de una observación está completa cuando se indica el conjunto de datos a la que pertenece. Este conjunto de datos es agregado como propiedad mediante *qb:dataSet*, es utilizada como URI y el valor de esta es el definido en la Tabla 3 para el conjunto de datos, obteniendo el resultado que se muestra en la Figura 43.

```
ds:obs8 a qb:Observation ;
  prop:pais "Ecuador" ;
  prop:año "2016" ;
  prop:mes "Marzo" ;
  prop:titulo "Análisis Crítico de la Economía Solidaria en el Ecuador" ;
  prop:visita "22"^^xsd:int ;
  qb:dataSet ds:dataset-revista_epn .
```

Figura 43 Observación con las propiedades: dimensión, medida y conjunto de datos
Realizado por el Autor

Exportar en formato RDF/XML

El grafo creado en Google Refine se exporta en formato RDF/XML. Este archivo contiene el cubo de datos RDF, con su estructura, dimensiones y medidas. La Figura 44 muestra una parte del mismo, el archivo completo se presenta en el ANEXO D.

```

- <rdf:RDF>
+ <rdf:Description rdf:about="http://opendata.epn.edu.ec/cubos/revista_epn/prop/titulo">
  </rdf:Description>
+ <rdf:Description rdf:about="http://opendata.epn.edu.ec/cubos/revista_epn/prop/pais"></rdf:Description>
+ <rdf:Description rdf:about="http://opendata.epn.edu.ec/cubos/revista_epn/prop/año"></rdf:Description>
+ <rdf:Description rdf:about="http://opendata.epn.edu.ec/cubos/revista_epn/prop/mes"></rdf:Description>
+ <rdf:Description rdf:about="http://opendata.epn.edu.ec/cubos/revista_epn/prop/visita">
  </rdf:Description>
+ <rdf:Description rdf:about="http://opendata.epn.edu.ec/cubos/revista_epn/dataset/dataset-revista_epn">
  </rdf:Description>
+ <rdf:Description rdf:about="http://opendata.epn.edu.ec/cubos/revista_epn/dataset/dsd-revista_epn">
  </rdf:Description>
+ <rdf:Description rdf:about="http://opendata.epn.edu.ec/cubos/revista_epn/dccs/titulo">
  </rdf:Description>
+ <rdf:Description rdf:about="http://opendata.epn.edu.ec/cubos/revista_epn/dataset/dsd-revista_epn">
  </rdf:Description>
+ <rdf:Description rdf:about="http://opendata.epn.edu.ec/cubos/revista_epn/dccs/pais"></rdf:Description>
+ <rdf:Description rdf:about="http://opendata.epn.edu.ec/cubos/revista_epn/dataset/dsd-revista_epn">
  </rdf:Description>
+ <rdf:Description rdf:about="http://opendata.epn.edu.ec/cubos/revista_epn/dccs/año"></rdf:Description>
+ <rdf:Description rdf:about="http://opendata.epn.edu.ec/cubos/revista_epn/dataset/dsd-revista_epn">
  </rdf:Description>
+ <rdf:Description rdf:about="http://opendata.epn.edu.ec/cubos/revista_epn/dccs/mes"></rdf:Description>
+ <rdf:Description rdf:about="http://opendata.epn.edu.ec/cubos/revista_epn/dataset/dsd-revista_epn">
  </rdf:Description>
+ <rdf:Description rdf:about="http://opendata.epn.edu.ec/cubos/revista_epn/dccs/visitas">
  </rdf:Description>
+ <rdf:Description rdf:about="http://opendata.epn.edu.ec/cubos/revista_epn/dataset/dsd-revista_epn">
  </rdf:Description>
- <rdf:Description rdf:about="http://opendata.epn.edu.ec/cubos/revista_epn/dataset/dataset-revista_epn">
  <qb:structure rdf:resource="http://opendata.epn.edu.ec/cubos/revista_epn/dataset/dsd-revista_epn"/>
  <dc:creator xml:lang="es">Darío Tubón</dc:creator>
  - <dc:title xml:lang="es">
    Conjunto de Datos Estadísticos de la Revista Politécnica
  </dc:title>
  - <dc:description>
    Conjunto de Datos que contiene un Cubo de Datos RDF, el mismo que representan a las visitas al sitio
    web de la Revista Politécnica.
  </dc:description>
  <dc:date rdf:datatype="http://www.w3.org/2001/XMLSchema#date">2016-06-01</dc:date>
  </rdf:Description>
- <rdf:Description rdf:about="http://opendata.epn.edu.ec/cubos/revista_epn/dataset/obs1">
  <rdf:type rdf:resource="http://purl.org/linked-data/cube#Observation"/>
  <prop:pais>Ecuador</prop:pais>
  <prop:año>2015</prop:año>
  <prop:mes>Diciembre</prop:mes>
  - <prop:título>
    Análisis Crítico de la Economía Solidaria en el Ecuador
  </prop:título>
  <prop:visita rdf:datatype="http://www.w3.org/2001/XMLSchema#int">56</prop:visita>
  <qb:dataSet rdf:resource="http://opendata.epn.edu.ec/cubos/revista_epn/dataset/dataset-revista_epn"/>
  </rdf:Description>
- <rdf:Description rdf:about="http://opendata.epn.edu.ec/cubos/revista_epn/dataset/obs2">
  <rdf:type rdf:resource="http://purl.org/linked-data/cube#Observation"/>
  <prop:pais>Ecuador</prop:pais>
  <prop:año>2016</prop:año>
  <prop:mes>Enero</prop:mes>
  - <prop:título>
    Análisis Crítico de la Economía Solidaria en el Ecuador
  </prop:título>
  <prop:visita rdf:datatype="http://www.w3.org/2001/XMLSchema#int">40</prop:visita>
  <qb:dataSet rdf:resource="http://opendata.epn.edu.ec/cubos/revista_epn/dataset/dataset-revista_epn"/>
  </rdf:Description>

```

Figura 44 Cubo expresado en formato RDF/XML
Realizado por el Autor

Encadenamiento

El encadenamiento o enlazamiento de datos permite la relación con otros conjuntos de datos que contengan alguna información en común; la esencia misma del concepto de Linked Data.

El encadenamiento se realizó con Silk Workbench, en Silk se creó el proyecto llamado “RevistaPolitecnica”, en el primer caso se enlaza con DBpedia y en el segundo caso con el conjunto de datos creado en el proyecto interno de la EPN: “Plataforma de Integración de Bibliotecas digitales usando Linked Data”.

El enlazamiento con DBpedia tiene los conjuntos de datos origen y objetivo (*source* y *target*). Silk permite realizar el enlazamiento mediante varias opciones, las principales: mediante archivo en diferentes formatos o la conexión directa a un endpoint SPARQL.

La tarea de enlazamiento debe contener estos conjuntos de datos de origen y destino. En la configuración de la tarea de enlazamiento se definen las restricciones con las que se realiza el enlazamiento. En este trabajo la dimensión País fue enlazada con DBpedia. En la Figura 45 se muestra la estructura del proyecto “RevistaEPN” dentro de Silk.

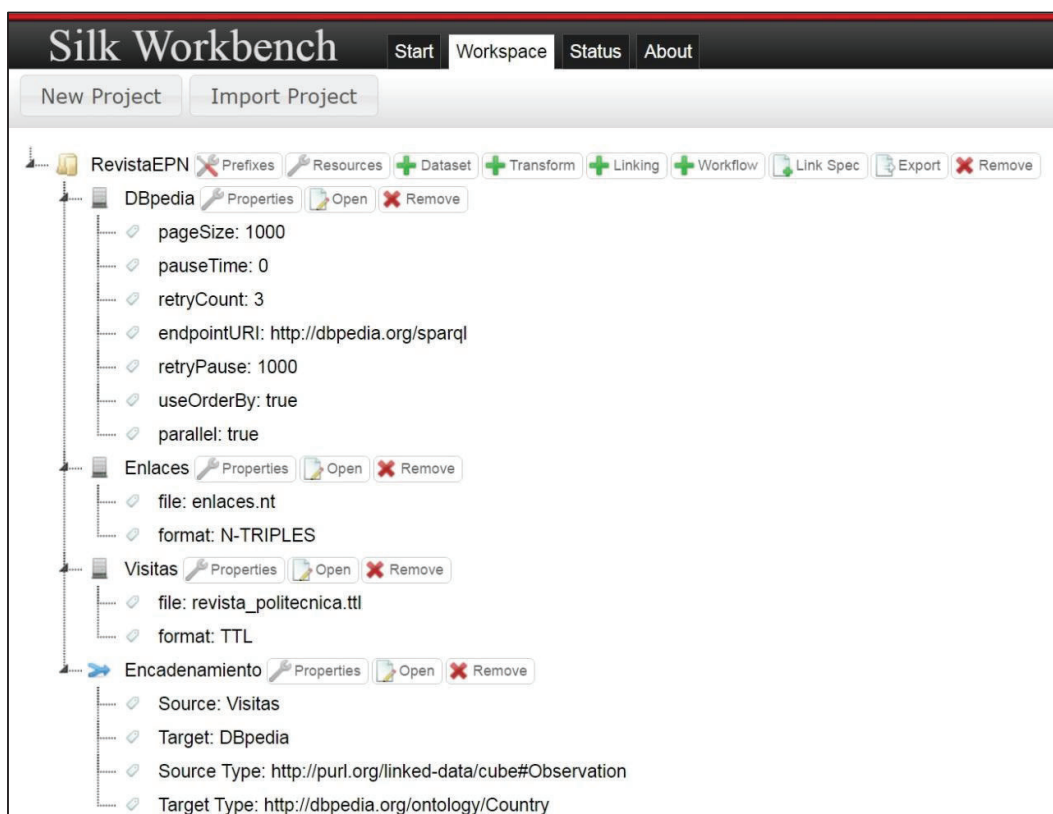


Figura 45 Esquema de proyecto en Silk
Realizado por el Autor

Silk posee una Interfaz gráfica que permite esquematizar subtareas, dentro de estas subtareas podemos seleccionar las propiedades que van a ser comparadas para encontrar sus similitudes. Silk permite esquematizar tareas de transformación, como se muestra en la Figura 46, las propiedades primero son transformadas a una cadena de caracteres minúsculos.

La siguiente tarea consiste en comparar la cadena de caracteres mediante uno de los varios algoritmos para dicho fin. En este proyecto se utilizó el algoritmo “*Levenshtein distance*”, que permite identificar la similitud entre cadenas de caracteres mediante el número de operaciones necesarias para transformar una cadena en la otra.

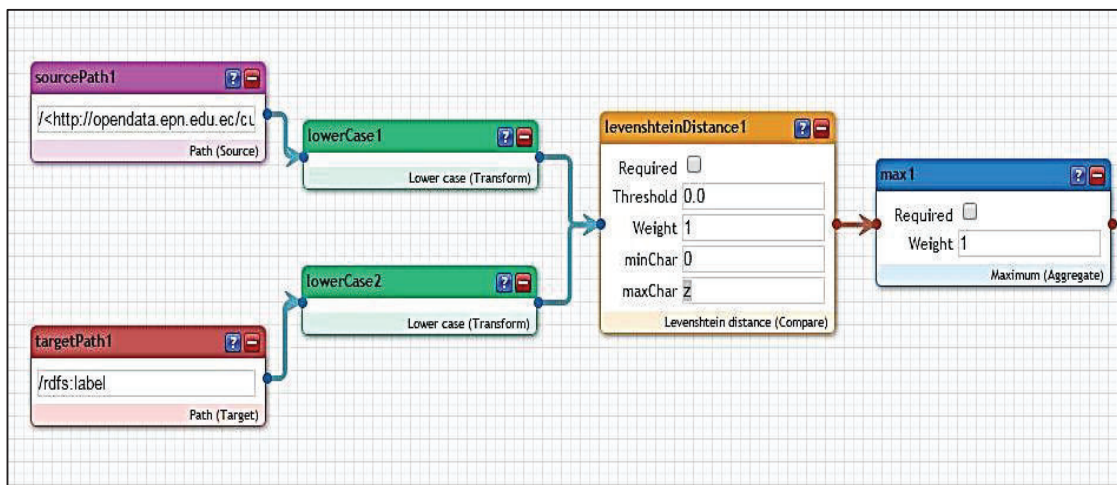


Figura 46 Esquema de tareas para generación de enlaces
Realizado por el Autor

La tarea final es comprobar la generación de los enlaces que se presentan como resultado, todo este procedimiento es manual, el usuario de Silk debe guiarse en los resultados que se presentan y el porcentaje de similitud entre los enlaces origen y destino para dar como valido un enlace. La Figura 47 muestra el resultado de la generación de los enlaces de la dimensión País, contenida en cada una de las observaciones del cubo de datos RDF con DBpedia.

Source: RevistaPolitecnica	Target: DBpedia	Score
1.epn.edu.ec/cubos/revista_epn/dataset/obs980	http://dbpedia.org/resource/Colombia	100,0%
Aggregation: max (max1) 100,0%		
Comparison: levenshteinDistance (levenshteinDistance1) 100,0%		
Transform: lowerCase (lowerCase1) colombia		
Input: <http://opendata.epn.edu.ec/cubos/revista_epn/prop/pais> (sourcePath1) Colombia		
Transform: lowerCase (lowerCase2) colombia كولومبيا kolumbien colombia colombie コロンビア		
Input: rdfs:label (targetPath1) Colombia كولومبيا Kolumbien Colombia Colombie コロンビア		
Colombia Colombia Kolumbia Colómbia Колумбия 哥伦比亚		
1.epn.edu.ec/cubos/revista_epn/dataset/obs932	http://dbpedia.org/resource/Colombia	100,0%
Aggregation: max (max1) 100,0%		
Comparison: levenshteinDistance (levenshteinDistance1) 100,0%		
Transform: lowerCase (lowerCase1) colombia		
Input: <http://opendata.epn.edu.ec/cubos/revista_epn/prop/pais> (sourcePath1) Colombia		
Transform: lowerCase (lowerCase2) colombia كولومبيا kolumbien colombia colombie コロンビア		
Input: rdfs:label (targetPath1) Colombia كولومبيا Kolumbien Colombia Colombie コロンビア		
Colombia Colombia Kolumbia Colómbia Колумбия 哥伦比亚		

Figura 47 Enlaces generados
Realizado por el Autor

Publicación

La publicación de datos consiste en alojarlos en un medio apropiado para que este pueda ser consumido y consultado por cualquier usuario. Este medio es el servidor Virtuoso, punto que es abarcado en el apartado 3.2 del presente trabajo, que indica la carga del cubo de datos RDF en el servidor Virtuoso para ser consumido por OntoWiki, mediante su extensión CubeViz y realizar consultas directamente mediante SPARQL.

2.3 PRUEBAS E IMPLANTACIÓN DEL SISTEMA

En este apartado se detallan las pruebas a llevar acabo al sistema y el despliegue del mismo; etapa indicada por la metodología seleccionada.

2.3.1 PRUEBAS DEL SISTEMA

2.3.1.1 Pruebas de funcionalidad.

Las pruebas de funcionalidad se realizaron sobre el conjunto de datos RDF creado en Google Refine, se verificó que los datos creados a partir del archivo RDF se almacenen correctamente en el servidor Virtuoso y se visualicen de igual forma en OntoWiki.

Las pruebas de seguridad permiten verificar que el acceso al conjunto de datos y manipulación del mismo sea de forma segura y controlada.

2.3.1.1.1 Pruebas de seguridad

El principal objetivo de llevar a cabo pruebas de seguridad es verificar el correcto funcionamiento de los mecanismos de seguridad que implementa el sistema. Lo que permite la protección de la información almacenada en el mismo.

Las pruebas de seguridad se centran en la comprobación de los permisos que cada usuario posee en el sistema, la verificación se la realizó en OntoWiki y en el servidor Virtuoso. De esta forma se verificó el acceso a la creación, eliminación, y actualización de los conjuntos de datos únicamente a usuarios con rol de administrador.

En la Tabla 8 se muestra las pruebas realizadas a los componentes principales del sistema: OntoWiki y servidor Virtuoso.

Tabla 8 Pruebas de seguridad

Objetivo	Verificar que únicamente usuarios con la respectiva autorización puedan ingresar a cada uno de los componentes.	
	Casos de Prueba	Resultados obtenidos
	Ingresar al Servidor Virtuoso sin usuario y contraseña. Ingresar a OntoWiki sin usuario y contraseña.	El Servidor Virtuoso impide el ingreso de cualquier usuario sin sus respectivas credenciales. OntoWiki impide el ingreso de usuarios sin credenciales, los mismos que solo pueden visualizar la información almacenada.
Conclusión	Cada uno de los componentes del Sistema posee sus propios mecanismos de seguridad, lo que hace más robusto al mismo al trabajar todos a la par.	

2.3.1.1.2 Pruebas de sistema

Las pruebas a las que será sometido el sistema son pruebas en general, aplicables a cualquier tipo de sistema; no centradas en el desarrollo. Busca evaluar el correcto funcionamiento de todos los componentes, trabajando como

un único sistema. Después de tener el cubo de datos RDF cargado en el sistema, se verificó que todos los componentes se hayan creado correctamente; con toda su información.

En las siguientes imágenes se muestra la información presentada en OntoWiki de todos los componentes que forman el conjunto de datos RDF diseñado en Google Refine. Se muestra la información de: dimensiones, medidas, definición de la estructura de datos y la especificación de componentes.

➤ **Dimensión: País**

La dimensión País se creó y es interpretada correctamente por OntoWiki como indica la Figura 48.



Figura 48 Dimensión: Título
Realizado por el Autor

➤ **Dimensión: Año**

La dimensión Año se creó y es interpretada correctamente por OntoWiki como indica la Figura 49.



Figura 49 Dimensión: Año
Realizado por el Autor

➤ **Dimensión: Mes**

La dimensión Mes se creó y es interpretada correctamente por OntoWiki como indica la Figura 50.

▣ Mes	qb:DimensionProperty
qb:DimensionProperty	
rdf:type	▣ qb:DimensionProperty
rdfs:label	Mes

*Figura 50 Dimensión: Mes
Realizado por el Autor*

➤ **Dimensión: Titulo**

La dimensión Titulo se creó y es interpretada correctamente por OntoWiki como indica la Figura 51.

▣ Titulo	qb:DimensionProperty
qb:DimensionProperty	
rdf:type	▣ qb:DimensionProperty
rdfs:label	Titulo

*Figura 51 Dimensión: Titulo
Realizado por el Autor*

➤ **Medida: Vistas**

La medida visitas se creó y es interpretada correctamente por OntoWiki como indica la Figura 52.

▣ Visitas	qb:MeasureProperty
qb:MeasureProperty	
rdf:type	▣ qb:MeasureProperty
rdfs:label	Visitas

*Figura 52 Medida: Número de visitas
Realizado por el Autor*

➤ **Definición de la estructura del conjunto de datos RDF (DSD).**

La DSD es la columna vertebral de un cubo de datos RDF, su correcta definición implica una correcta interpretación del cubo de datos RDF en el visualizador CubeViz. Si la información presentada es errónea o no se presenta, el objetivo del sistema OLAP no tendría sentido. La Figura 53 muestra la información relacionada a la DSD mostrada en OntoWiki.

ds:dsd-revista_epn		qb:DataStructureDefinition										
qb:DataStructureDefinition												
qb:component	<table border="1"> <tr> <td colspan="2">Titulo</td> </tr> <tr> <td>qb:dimension</td> <td>Titulo</td> </tr> <tr> <td>rdf:type</td> <td>qb:ComponentSpecification</td> </tr> <tr> <td>rdfs:comment</td> <td>Corresponde al titulo del documento visitado.</td> </tr> <tr> <td>rdfs:label</td> <td>Titulo</td> </tr> </table>		Titulo		qb:dimension	Titulo	rdf:type	qb:ComponentSpecification	rdfs:comment	Corresponde al titulo del documento visitado.	rdfs:label	Titulo
Titulo												
qb:dimension	Titulo											
rdf:type	qb:ComponentSpecification											
rdfs:comment	Corresponde al titulo del documento visitado.											
rdfs:label	Titulo											
	<table border="1"> <tr> <td colspan="2">Pais</td> </tr> <tr> <td>qb:dimension</td> <td>Pais</td> </tr> <tr> <td>rdf:type</td> <td>qb:ComponentSpecification</td> </tr> <tr> <td>rdfs:comment</td> <td>Ubicación geográfica desde donde se realiza la visita al sitio web de la Revista Politécnica.</td> </tr> <tr> <td>rdfs:label</td> <td>Pais</td> </tr> </table>		Pais		qb:dimension	Pais	rdf:type	qb:ComponentSpecification	rdfs:comment	Ubicación geográfica desde donde se realiza la visita al sitio web de la Revista Politécnica.	rdfs:label	Pais
Pais												
qb:dimension	Pais											
rdf:type	qb:ComponentSpecification											
rdfs:comment	Ubicación geográfica desde donde se realiza la visita al sitio web de la Revista Politécnica.											
rdfs:label	Pais											
	<table border="1"> <tr> <td colspan="2">Mes</td> </tr> <tr> <td>qb:dimension</td> <td>Mes</td> </tr> <tr> <td>rdf:type</td> <td>qb:ComponentSpecification</td> </tr> <tr> <td>rdfs:comment</td> <td>Mes en el cual se realiza la visita.</td> </tr> <tr> <td>rdfs:label</td> <td>Mes</td> </tr> </table>		Mes		qb:dimension	Mes	rdf:type	qb:ComponentSpecification	rdfs:comment	Mes en el cual se realiza la visita.	rdfs:label	Mes
Mes												
qb:dimension	Mes											
rdf:type	qb:ComponentSpecification											
rdfs:comment	Mes en el cual se realiza la visita.											
rdfs:label	Mes											
	<table border="1"> <tr> <td colspan="2">Visitas</td> </tr> <tr> <td>qb:measure</td> <td>Visitas</td> </tr> <tr> <td>rdf:type</td> <td>qb:ComponentSpecification</td> </tr> <tr> <td>rdfs:comment</td> <td>Representa el N° de visitas que se realizan a sitio Web de la Revista Politécnica.</td> </tr> <tr> <td>rdfs:label</td> <td>Visitas</td> </tr> </table>		Visitas		qb:measure	Visitas	rdf:type	qb:ComponentSpecification	rdfs:comment	Representa el N° de visitas que se realizan a sitio Web de la Revista Politécnica.	rdfs:label	Visitas
Visitas												
qb:measure	Visitas											
rdf:type	qb:ComponentSpecification											
rdfs:comment	Representa el N° de visitas que se realizan a sitio Web de la Revista Politécnica.											
rdfs:label	Visitas											
	<table border="1"> <tr> <td colspan="2">Año</td> </tr> <tr> <td>qb:dimension</td> <td>Año</td> </tr> <tr> <td>rdf:type</td> <td>qb:ComponentSpecification</td> </tr> <tr> <td>rdfs:comment</td> <td>Año en el que se realiza la visita.</td> </tr> <tr> <td>rdfs:label</td> <td>Año</td> </tr> </table>		Año		qb:dimension	Año	rdf:type	qb:ComponentSpecification	rdfs:comment	Año en el que se realiza la visita.	rdfs:label	Año
Año												
qb:dimension	Año											
rdf:type	qb:ComponentSpecification											
rdfs:comment	Año en el que se realiza la visita.											
rdfs:label	Año											
rdf:type	qb:DataStructureDefinition											

Figura 53 Definición de la estructura de datos del cubo de datos RDF
Realizado por el Autor

2.3.2 IMPLANTACIÓN DEL SISTEMA

Con las herramientas elegidas para el sistema OLAP para bibliotecas digitales que usan tecnologías Linked Data, se procedió a la instalación de cada una de estas, de tal forma que puedan almacenar cualquier tipo cubo de datos RDF; en este caso de estudio el cubo de datos RDF contiene las visitas al sitio web de la Revista Politécnica.

El sistema tiene la arquitectura que se indica en la Figura 54, donde podemos ver que el almacenamiento del cubo de datos es el servidor Virtuoso, que se alimenta de los conjuntos de datos ya transformados y utilizables en la Web Semántica con la herramienta Google Refine.

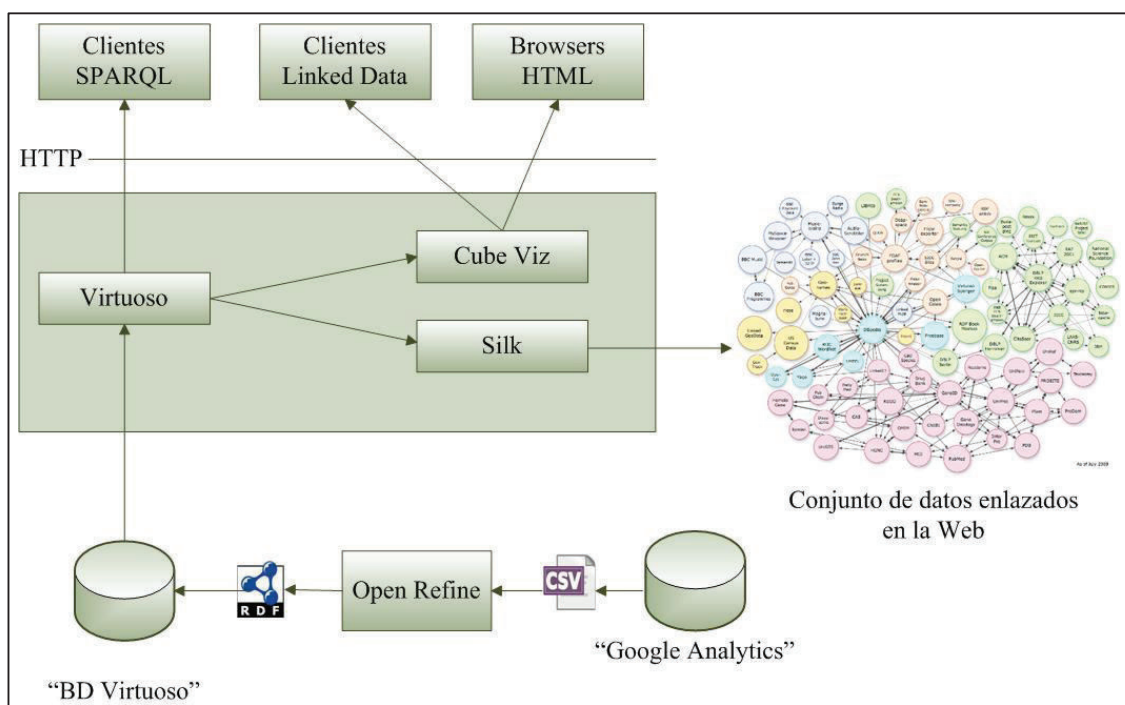


Figura 54 Arquitectura de sistema [22]

La fuente de datos en este caso de estudio son los resultados de Google Analytics y las consultas al protocolo OAI-PMH. Proceso de consolidación indicado en la Figura 21.

Las consultas SPARQL se realizan desde el endpoint SPARQL al servidor Virtuoso. Los usuarios interactúan con la información del cubo de datos RDF accediendo al software OntoWiki. La extensión CubeViz, como se indicó en capítulos anteriores, permite la visualización gráfica de los datos estadísticos del cubo de datos RDF.

La instalación de todas las herramientas se realizó bajo las siguientes condiciones de software y hardware:

Software

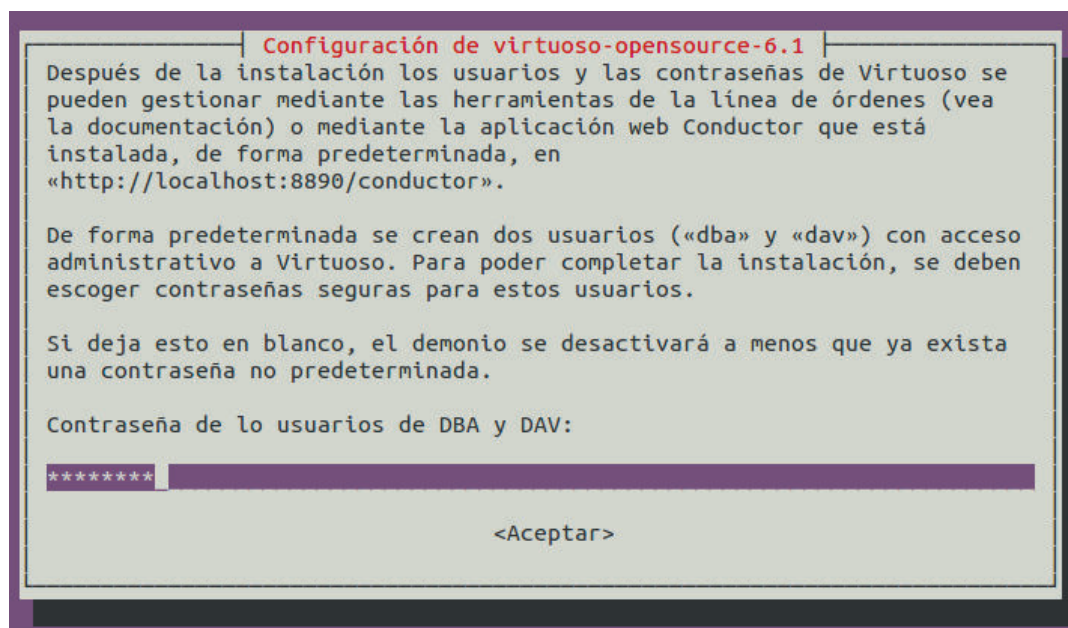
- Sistema Operativo: Ubuntu 12.04.5 LTS
- Máquina virtual: Virtualizador VirtualBox
- Lenguaje de Programación: PHP 5

Hardware

- Procesador: Intel Pentium® Core® i3-3240, 3.40GHz
- RAM: 4GB
- Disco Duro: 20 GB

2.3.2.1 Instalación del servidor Virtuoso

Los siguientes puntos indican a breves rasgos la instalación de Virtuoso (los detalles se encuentran en el ANEXO C), para lo cual se tiene como referencia la documentación oficial que se indica en su página oficial [23] . En la Figura 55 se muestra la pantalla para la configuración de credenciales de ingreso para los usuarios de Virtuoso.



*Figura 55 Configuración de usuario en Virtuoso
Realizado por el Autor*

Una vez que termina la instalación de Virtuoso, el acceso al mismo se lo realiza desde la siguiente dirección: <http://opendata.epn.edu.ec/conductor>. La Figura 56 muestra la interfaz de Virtuoso accedida desde el navegador web.



*Figura 56 Interfaz de ingreso a Virtuoso
Realizado por el Autor*

2.3.2.2 Instalación de OntoWiki

Este wiki semántico se instaló siguiendo la documentación oficial [24], los pasos más importantes, se presentan en las siguientes imágenes (los detalles se encuentran en el ANEXO D).

Apache es el servidor que aloja a la aplicación web OntoWiki, por lo que este es clonado desde su repositorio de GitHub en su carpeta pública por defecto. Que en Linux se encuentra bajo la siguiente ruta: `/var/www/html`. Uno de los componentes principales a tener en cuenta es la configuración del ODBC (*Open DataBase Connectivity*) que permite la comunicación con el servidor Virtuoso.

El archivo `odbcinst.ini` debe ser editado, se agregó las líneas que indica la ruta del driver que permite la comunicación entre la aplicación OntoWiki y el servidor Virtuoso; como se indica en la Figura 57.

```

Archivo odbcinst.ini

[virtuoso-odbc]
Driver = <prefix>/lib/virtodbc.so

Archivo odbc.ini

[ODBC Data Sources]
VOS = Virtuoso

[VOS]
Driver = virtuoso-odbc
Description = Virtuoso Open-Source Edition
Address = localhost:1111

[VOS]
Driver          = /usr/lib/odbc/virtodbc.so
Description     = Virtuoso OpenSource Edition
Address         = localhost:1111

```

Figura 57 Configuración ODBC
Realizado por el Autor.

Las credenciales creadas en la instalación de Virtuoso son agregadas a la configuración de OntoWiki, desde esta aplicación se puede acceder a los grafos almacenados en Virtuoso, toda esta configuración se encuentra en el archivo config.ini de OntoWiki (los detalles se encuentran en el ANEXO E). La Fig. 58 muestra la pantalla de inicio del wiki semántico OntoWiki.

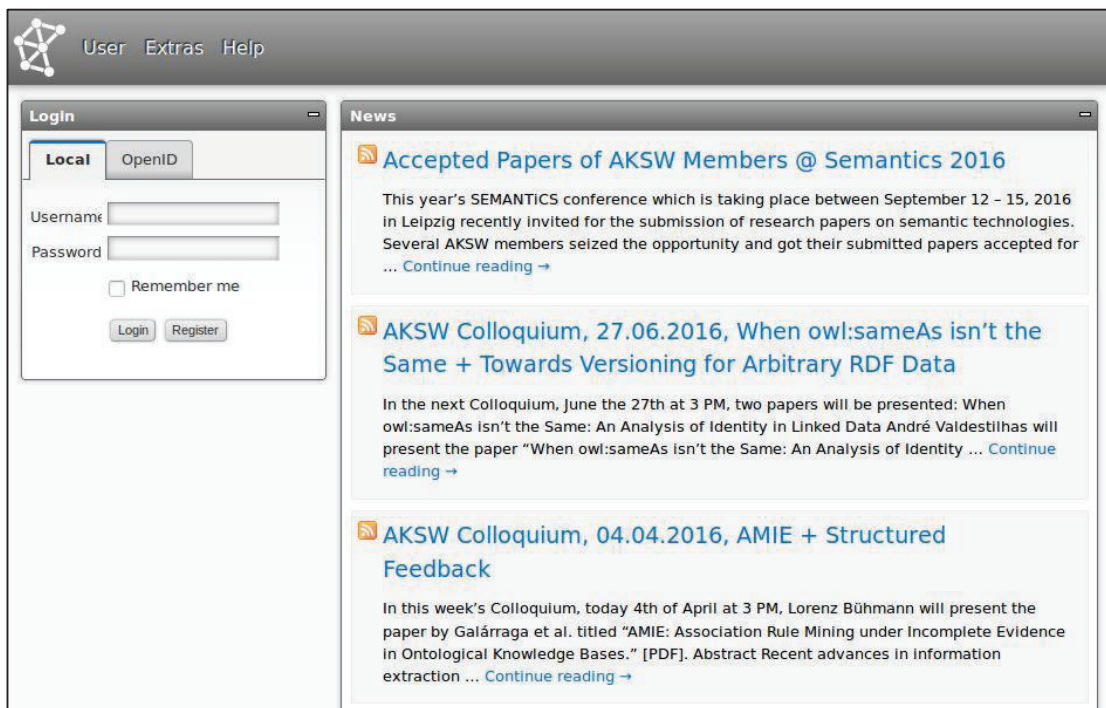


Figura 58 Pantalla de Inicio de OntoWiki
Elaborado por el Autor

2.3.2.3 Instalación de CubeViz

Una de las principales ventajas que se tuvo en cuenta al momento de seleccionar a OntoWiki para la interpretación de cubos de datos RDF, es la existencia de varias extensiones que permiten adaptar esta herramienta a nuevos requerimientos. Es precisamente lo que permite la extensión CubeViz: la interpretación gráfica de metadatos estadísticos elaborados con el vocabulario para cubos de datos RDF.

De la misma forma que las instalaciones previas, también posee su guía instalación oficial [25]. OntoWiki cuenta con un módulo destinado a extensiones, donde se aloja CubeViz, una vez agregada cualquier extensión se debe eliminar todos los registros de la carpeta cache y reiniciar el servidor Apache2 como indica su sitio oficial y puede verse en el ANEXO F. La Fig. 59 indica un ejemplo de visualización de datos con CubeViz.

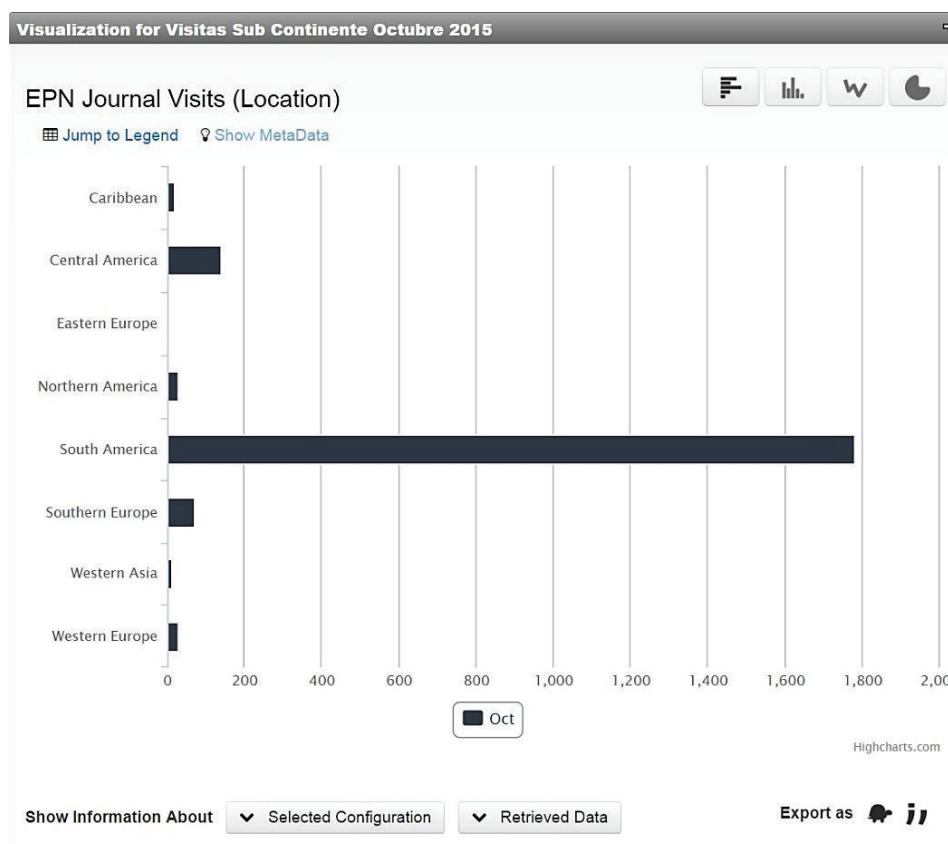


Figura 59 Extensión CubeViz de OntoWiki [22]

CAPÍTULO 3 EVALUACIÓN DEL SISTEMA EN EL CASO DE APLICACIÓN

En este capítulo se realiza la ejecución de todos los componentes como un único sistema. El sistema aloja el cubo de datos RDF creado con un vocabulario propio para formar parte de la Web Semántica. Los resultados se muestran de forma gráfica. También se realizan varias consultas SPARQL con el fin de verificar los datos de las visitas realizadas a la Revista Politécnica.

3.1 EJECUCIÓN DEL SISTEMA

El sistema está compuesto por dos componentes principales: el servidor Virtuoso y el visualizador OntoWiki con su extensión CubeViz, cada uno tiene una dirección de acceso propia.

- Servidor Virtuoso: <http://opendata.epn.edu.ec:8890>
- OntoWiki: <http://opendata.epn.edu.ec/OntoWiki>

Otro componente a tener en cuenta es el endpoint SPARQL, al mismo que se tiene acceso en la siguiente dirección:

- SPARQL: <http://opendata.epn.edu.ec:8890/sparql>

La publicación del cubo de datos RDF se la realizó en el servidor Virtuoso en forma de grafo RDF. En Virtuoso tenemos acceso al menú con todas sus funcionalidades, incluida la dedicada a la publicación de Linked Data. La Figura 60 muestra la carga del grafo directamente al servidor y la Figura 61 muestra todos los grafos que se han creado hasta el momento.

*Figura 60 Creación de grafo RDF
Realizado por el Autor.*

SPARQL	Sponger	Statistics	Graphs	Schemas	Namespaces	Access Control	Views	Quad Store Upload
Graphs								
Graph:								
http://www.openlinksw.com/schemas/virtrdf#				Rename	Delete			
http://localhost:8890/DAV				Rename	Delete			
http://www.w3.org/2002/07/owl#				Rename	Delete			
http://localhost:8890/DAV/Continentes				Rename	Delete			
http://opendata.epn.edu.ec/DAV/revista_politecnica				Rename	Delete			
http://localhost/OntoWiki/index.php/Continentes/				Rename	Delete			
http://192.168.1.7/OntoWiki/index.php/VisitasRevistaPolitecnica/				Rename	Delete			
http://localhost/OntoWiki/Config/				Rename	Delete			
http://ns.ontowiki.net/SysOnt/				Rename	Delete			
http://192.168.1.7/OntoWiki/index.php/Ejemplo/				Rename	Delete			
Copyright © 1998-2016 OpenLink Software								

*Figura 61 Grafos cargados en Virtuoso
Realizado por el Autor*

OntoWiki permite crear bases de conocimiento, el contenido de estas bases de conocimiento es alojado en Virtuoso. OntoWiki mediante consultas SPARQL y la utilización de CubeViz facilita la interpretación gráfica de la información del cubo de datos RDF. La Figura 62 muestra la creación de la base de conocimiento, esta base de conocimiento contiene el cubo de datos RDF.

Create New Knowledge Base

Basic information

Title (leave blank if your import provides the name)

URI (leave blank for auto creation)

Data import actions

Select a Vocabulary

Find vocabularies by searching the [Linked Open Vocabularies \(LOV\)](#) repository.

Search in LOV Repository

- SKOS Vocabulary
- Friend of a Friend (FOAF) vocabulary
- SIOC Core Ontology Namespace
- Dublin Core Metadata Element Set, Version 1.1
- DCMI Metadata Terms - other

Please use one of these options to import data from various sources or with different transformations:

- Create a (nearly) empty knowledge base
- Import an RDF resource from the web
- Upload an RDF Dump
- Paste Source

*Figura 62 Creación de nueva base de conocimiento
Realizado por el Autor*

OntoWiki llama a cada conjunto de datos alojados en Virtuoso como bases de conocimiento (*Knowledge Bases*), si se selecciona una de estas bases de conocimiento se la puede visualizar gráficamente, siempre que haya sido creado con el vocabulario para cubos de datos RDF. La Figura 64 muestra todos los conjuntos de datos alojados en el servidor Virtuoso hasta el momento.

Al seleccionar el conjunto de datos, se muestran también todos los componentes del cubo de datos RDF.

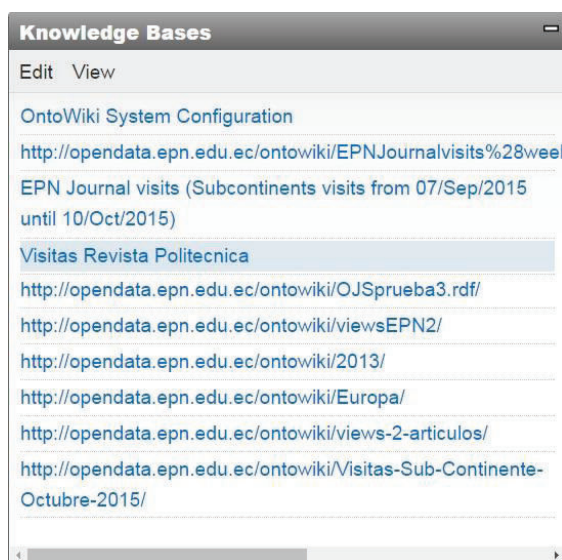


Figura 64 Lista de conjuntos de datos
Realizado por el Autor

3.2 RECOPIACIÓN DE INFORMACIÓN DEL CASO DE APLICACIÓN

Los datos para el caso de estudio fueron recopilados desde fuentes libres, a partir de los cuales con las herramientas utilizadas se los ha consolidado y enriquecido; hasta obtener un conjunto de datos que contiene el cubo de datos RDF.

3.2.1 METADATOS DE LA REVISTA POLITÉCNICA

En el apartado de diseño del cubo de datos RDF se indicó que gracias al protocolo OAI-PMH es posible obtener todos los metadatos ligados a un recurso alojado en la Revista Politécnica. Estos metadatos contienen información fundamental para identificar y describir un artículo académico.

La información de los metadatos no indica una información estadística; necesaria para nuestro caso de estudio. Información que fue obtenida desde Google Analytics, para el periodo seleccionado. La Tabla 9 muestra la cantidad de registros obtenidos de cada fuente y el total de registros al hacer la consolidación de los dos conjuntos de datos.

Tabla 9 Registros generados
Realizado por el Autor

OAI-PMH	Google Analytics	Observaciones del conjunto de datos RDF
456 registros	2911 Registros	2861 Visitas

3.2.2 CONJUNTOS DE DATOS EXTERNOS

Una de las fuentes más grande de datos a la que se pueden acceder de forma libre es DBpedia, la misma que ha sido utilizada para poder realizar el encadamiento y enriquecer los datos que contiene el cubo de datos RDF generado en este trabajo.

Otros conjuntos de datos fueron utilizados, el enlazamiento a estos conjuntos de datos fue nulo debido a que los criterios de búsqueda y enlazamiento no se encontraban. El enlazamiento se realizó, a más de con DBpedia, con los conjuntos de datos RDF elaborados en el proyecto interno de la EPN “Plataforma de Integración de Bibliotecas digitales usando Linked Data”.

3.3 ANÁLISIS DE RESULTADOS

El objetivo de esta sección es mostrar los resultados esperados de este trabajo, como se pudo ver en capítulos anteriores, se empezó con la extracción de los metadatos, elaboración del ETL aplicando la guía de buenas prácticas para la publicación en la web semántica y finalmente la visualización del cubo de datos RDF.

3.3.1 VISUALIZACIÓN DE GRÁFICOS ESTADÍSTICOS.

CubeViz visualiza la información gracias a 3 tipos de gráficos, en las Figura 65, Figura 66, Figura 67 y Figura 68 podemos ver diferentes tipos de representación de la información: barras horizontales, barras verticales y gráficos de dispersión.

El gráfico de la Figura 65 muestra la información estadística en forma de barras horizontales, este gráfico indica que de todos los artículos seleccionados el que más visitas ha recibido es el que tiene como título “*Aplicaciones de Procesamiento de Lenguaje Natural*”. La Figura 66 muestra la misma información pero esta vez en forma de barras verticales.

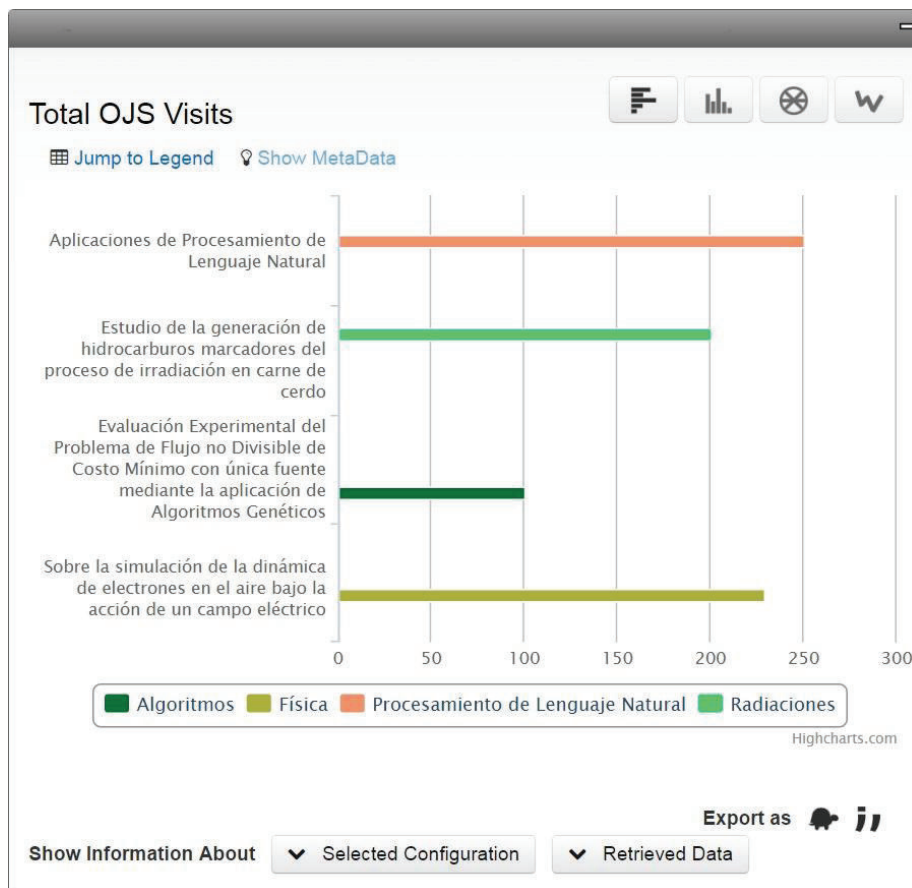


Figura 65 Gráfico de barras horizontales.
Realizado por el Autor

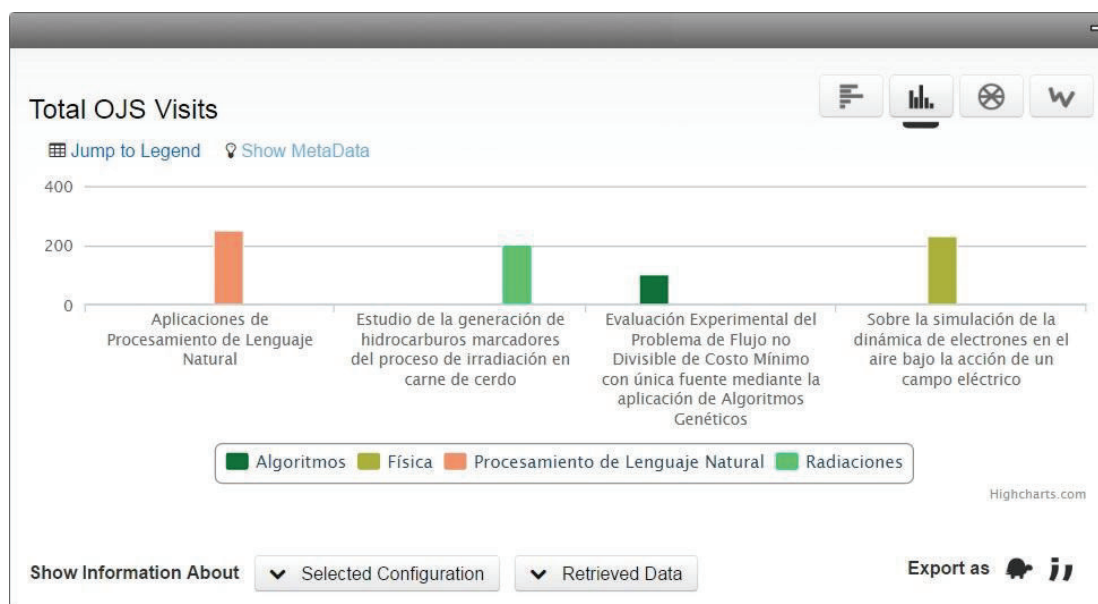


Figura 66 Gráfico de barras verticales.
Realizado por el Autor.



Figura 67 Gráfico de dispersión
Realizado por el Autor

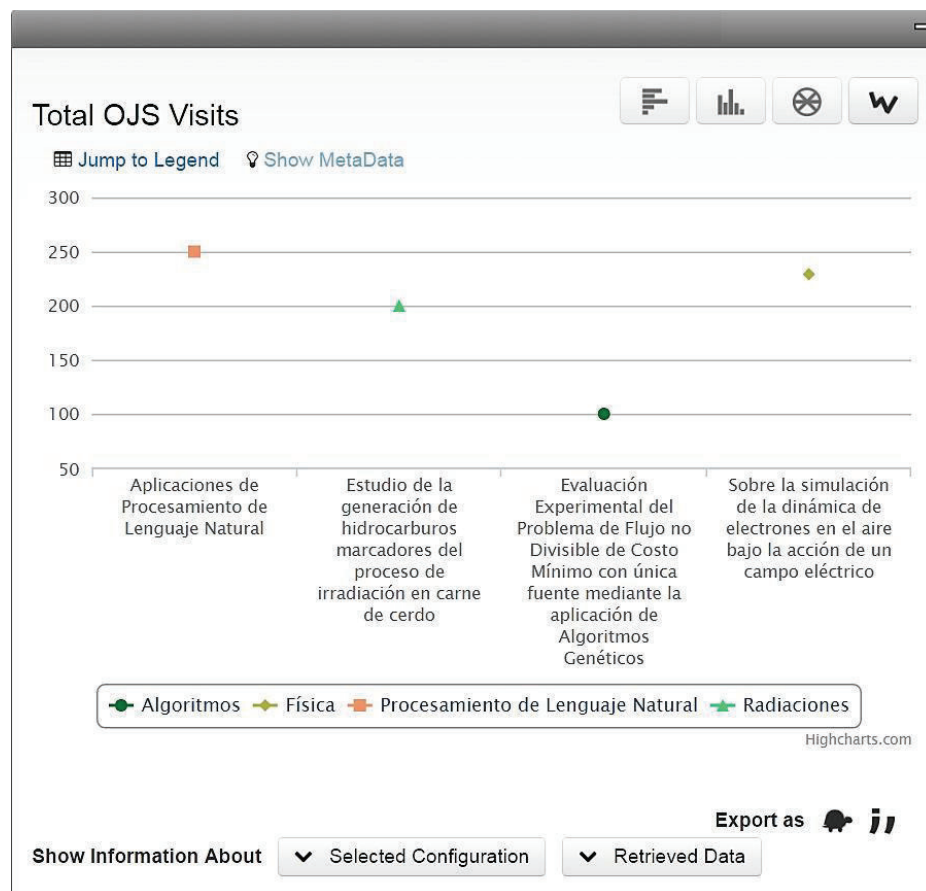


Figura 68 Gráfico de dispersión
Realizado por el Autor

3.3.2 CONSULTAS SPARQL

El sistema permite consultar la información del cubo de datos RDF mediante consultas en lenguaje SPARQL. A continuación se indican los resultados de varias consultas realizadas al cubo de datos RDF generado en este trabajo. Estas consultas se las realizaron desde el endpoint SPARQL de Virtuoso.

Virtuoso SPARQL Query Editor

[About](#) | [Namespace Prefixes](#) | [Inference rules](#)

Default Data Set Name (Graph IRI)

Query Text

(Security restrictions of this server do not allow you to retrieve remote RDF data, see [details](#).)

Results Format:

Execution timeout: milliseconds (values less than 1000 are ignored)

Options: Strict checking of void variables

(The result can only be sent back to browser, not saved on the server, see [details](#).)

Figura 69 Endpoint para consultas SPARQL
 Realizado por el Autor

Virtuoso tiene varios conjuntos de datos almacenados en su almacén, por lo que para realizar una consulta sobre un conjunto de datos específico es necesario ingresar la IRI (*Internationalized Resource Identifier*) del mismo y así evitar resultados erróneos. Las consultas para evaluar los resultados del cubo de datos RDF se realizaron sobre el grafo con la IRI: <http://opendata.epn.edu.ec/OntoWiki/VisitasRevistaPolitecnica/>.

Consulta 1: Total de visitas realizadas al sitio web de la Revista Politécnica.

```

PREFIX dc: <http://purl.org/dc/elements/1.1/>
PREFIX qb: <http://purl.org/linked-data/cube#>
PREFIX prop: <http://opendata.epn.edu.ec/cubos/revista_epn/prop/>
SELECT ?Conjunto_Datos_URI SUM(xsd:int(?Visitas)) as ?Total_Visitas
WHERE{
  ?x prop:visita ?Visitas.
  ?x qb:dataSet ?Conjunto_Datos_URI.
}

```

Conjunto_Datos_URI	Total_Visitas
http://opendata.epn.edu.ec/cubos/revista_epn/dataset/dataset-revista_epn	5466

*Figura 70 Consulta: total de visitas
Realizado por el Autor*

Resultado: En el periodo de análisis del presente trabajo se muestra que la Revista Politécnica ha tenido una total de 5466 visitas, a todos sus artículos, sin distinción.

Consulta 2: Las visitas que se han realizado a un recurso, que contiene una palabra determinada en su Título. En el ejemplo se utiliza la palabra “software”.

```

PREFIX dc: <http://purl.org/dc/elements/1.1/>
PREFIX qb: <http://purl.org/linked-data/cube#>
PREFIX prop: <http://opendata.epn.edu.ec/cubos/revista_epn/prop/>
SELECT ?Titulo SUM(xsd:int(?Visitas)) as ?Visitas
WHERE{
  ?x prop:titulo ?Titulo.
  ?x prop:visita ?Visitas
  FILTER (REGEX(?Titulo, "software", "i"))
}
GROUP BY ?Titulo

```

Titulo	Visitas
Diseño e implementación de un programa de software para el dimensionamiento de un secador de doble tambor	5
Desarrollo de una Aplicación Cliente/Servidor para un Wall View en base a la Plataforma – Cruzada Opensource – FFmpeg (Colección de Software Libre que puede Grabar, Convertir y hacer Streaming de Audio y Video)	9
Desarrollo de una Herramienta de software que implemente Cloud Computing de Modelo Privado para Ofrecer Infraestructura como Servicio	11
Sistema de vigilancia tecnológica de software libre	19
Herramientas de Software Libre para Aplicaciones en Ciencias e Ingeniería	75

*Figura 71 Consulta: total de visitas a artículos
Realizado por el Autor*

Resultado: En este periodo se han consultado 5 artículos, que contienen la palabra “software” en su título.

Consulta 3: Los 10 países desde los cuales se han registrado más visitas.

```

PREFIX dc: <http://purl.org/dc/elements/1.1/>
PREFIX qb: <http://purl.org/linked-data/cube#>
PREFIX prop: <http://opendata.epn.edu.ec/cubos/revista_epn/prop/>
PREFIX ds: <http://opendata.epn.edu.ec/cubos/revista_epn/dataset/>
SELECT ?Pais SUM(xsd:int(?Visitas)) as ?Visitas
WHERE{
  ?dataset a qb:DataSet.
  ?observation qb:dataSet ?dataset;
  a qb:Observation;
  prop:pais ?Pais;
  prop:visita ?Visitas.
}
ORDER BY DESC(SUM(xsd:int(?Visitas)))
LIMIT 10

```

Pais	Visitas
Ecuador	2301
Mexico	821
Colombia	649
Peru	375
Spain	267
Chile	163
Argentina	144
Venezuela	115
United States	77
Bolivia	54

*Figura 72 Consulta: países con más visitas
Realizado por el Autor*

Resultado: En el periodo de análisis, es Ecuador el país que registra más accesos a sitio web de la Revista Politécnica, seguidos por México y Colombia; a considerable distancia respecto del primero.

Consulta 4: Título del artículo más visitado en la Revista Politécnica.

```

PREFIX dc: <http://purl.org/dc/elements/1.1/>
PREFIX qb: <http://purl.org/linked-data/cube#>
PREFIX prop: <http://opendata.epn.edu.ec/cubos/revista_epn/prop/>
PREFIX ds: <http://opendata.epn.edu.ec/cubos/revista_epn/dataset/>
SELECT ?Titulo SUM(xsd:int(?Visitas)) as ?Visitas
WHERE{
  ?dataset a qb:DataSet.
  ?observation qb:dataSet ?dataset;
  a qb:Observation;
  prop:titulo ?Titulo;
  prop:visita ?Visitas.
}
ORDER BY DESC(SUM(xsd:int(?Visitas)))
LIMIT 1

```

Titulo	Visitas
Aprovechamiento de las propiedades funcionales del jengibre (zingiber officinale) en la elaboración de condimento en polvo, infusión filtrante y aromatizante para quema directa.	400

*Figura 73 Consulta: artículo con más visitas
Realizado por el Autor*

Resultado: El artículo más visitado en el periodo seleccionado es el que lleva por título: “Aprovechamiento de las propiedades funcionales del jengibre (zingiber officinale) en la elaboración de condimentos en polvo, infusión filtrante y aromatizante para quema directa”, con 400 visitas del total.

CONCLUSIONES Y RECOMENDACIONES

En los siguientes apartados se ponen de manifiesto varias conclusiones y recomendaciones que surgieron después de culminado el presente Proyecto de Titulación. Los mismos que deberían tener presente para la realización de trabajos similares.

Las conclusiones se generaron a partir de los resultados obtenidos. Las recomendaciones se han elaborado en torno a todas las variantes que se deben tener en cuenta, y que de alguna u otra forma influyeron en el correcto desarrollo de este Proyecto.

CONCLUSIONES

- Para la elaboración de sistemas OLAP en la Web Semántica se deben tener en cuenta las metodologías existentes para sistemas BI, sin dejar de lado las recomendaciones o buenas prácticas para la publicación de información en la Web Semántica.
- El sistema fue probado con la información de la Revista Politécnica, esta información se obtuvo de fuentes libres, accesible a cualquier usuario, como es el protocolo OAI-PMH.
- El sistema puede contener un cubo de datos RDF creado con el vocabulario para cubo de datos RDF de cualquier repositorio académico que vea necesaria la publicación de su información estadística.
- Los conjuntos de datos publicados con tecnologías de la Web Semántica relacionados a la comunidad académica ecuatoriana son limitados, pero se espera que con nuevas iniciativas estos vayan creciendo.
- Todas las herramientas utilizadas son libres, o en su defecto se han utilizado las versiones “*Open Source*”, como es el caso del Servidor Virtuoso, OntoWiki y CubeViz.
- El Sistema es accesible desde cualquier equipo con acceso a Internet.

RECOMENDACIONES

- Adoptar las recomendaciones para la publicación de recursos en la Web Semántica, lo que permitirá tener una Web con más conocimiento disponible para todos los usuarios de la Web.
- Utilizar herramientas libres, para que el costo monetario no sea un limitante en la creación y difusión de información en la Web Semántica.

- Publicar conjuntos de datos reutilizables en cualquier proyecto que utilice tecnologías Linked Data, de temática académica y que pueda ser de relevancia, como puede ser obras publicadas, autores, títulos de sus obras, etc.
- Publicar conjuntamente con la respectiva guía para poder crearlos. Esta guía debe cubrir todas las etapas del desarrollo de un conjunto de datos para ser publicado en la Web Semántica.

BIBLIOGRAFÍA

- [1] Radar Networks & Nova Spivack. The Future of Productivity. [Online]. Disponible en: www.radarnetworks.com. [Ultimo acceso: 25-05-2016]
- [2] María Jesús Lamarca Lapuente. Hacia la Web Semántica. [Online]. Disponible en: http://www.hipertexto.info/documentos/web_semantica.htm. [Ultimo acceso: 01-04-2016]
- [3] W3C. Guía Breve de Web Semántica. [Online]. Disponible: <http://www.w3c.es/Divulgacion/GuiasBreves/WebSemantica> [Ultimo acceso: 26-06-2016]
- [4] Tim Berners-Lee. Sematic Web-XML200. [Online]. Disponible en: <https://www.w3.org/2000/Talks/1206-xml2k-tbl/slide10-0.html>. [Ultimo acceso: 19-07-2016]
- [5] María Jesús Lamarca Lapuente. RDF. [Online]. Disponible en: <http://www.hipertexto.info/documentos/RDF.htm>. [Ultimo acceso: 30-05-2016]
- [6] W3C. Guía Breve de Linked Data.
- [7] Juan Carlos Trujillo, *Diseño y explotación de almacenes de datos: conceptos básicos de modelado multidimensional*. Alicante: ECU, 2013.
- [8] W3C. (2014, Enero) The RDF Data Cube Vocabulary. [Online]. Disponible en: <https://www.w3.org/TR/vocab-data-cube/#intro-RDF>. [Ultimo acceso: 05-09-2016]
- [9] Nelson Piedra, Janeth Chicaiza, Cadme Elizabeth, and Richar Guaya, "Una aproximación basada en Linked Data para la detección de potenciales redes de colaboración científica a partir de la anotación semántica de producción científica: Piloto aplicado con producción científica de investigadores ecuatorianos.," *MASKANA*, 2014.
- [10] Roberto Esinosa. (2010, Abril) El Rincon del BI. [Online]. Disponible en: <https://churriwifi.wordpress.com>. [Ultimo acceso: 17-09-2016]
- [11] W. H. Inmon, *Building the Data Warehouse.*: John Wiley & Sons, 2002.
- [12] Ralph Kimball and Maggy Ross, *The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling*, 3rd ed. Indianapolis: John Wiley & Sons, 2013.
- [13] W3C. (2012, Feb.) OpenLink Virtuoso. [Online]. Disponible: https://www.w3.org/2001/sw/wiki/OpenLink_Virtuoso [Ultimo acceso: 10-05-2016]

- [14] RDF4J. (2014, Oct.) About SESAME. [Online]. Disponible en: <http://RDF4j.org/about.docbook?view> [Ultimo acceso: 21-06-2016]
- [15] User and Developer. (2015, Diciembre) AKS/OntoWiki. [Online]. Disponible en: <https://github.com/AKSW/OntoWiki/wiki>. [Ultimo acceso: 12-06-2016]
- [16] Semantic MediaWiki. (2016, Enero) Introduction to Semantic MediaWiki. [Online]. Disponible en: https://www.semantic-mediawiki.org/wiki/Help:Introduction_to_Semantic_MediaWiki. [Ultimo acceso: 15-06-2016]
- [17] AKSW. CubeViz. [Online]. Disponible en: <http://aksw.org/Projects/CubeViz.html>. [Ultimo acceso: 25-07-2016]
- [18] DB-ENGINES. System Properties Comparison Sesame vs. Virtuoso. [Online]. Disponible en: <http://db-engines.com/en/system/Sesame%3BVirtuoso>. [Ultimo acceso: 25-03-2016]
- [19] W3C. (2014, Enero) Best Practices for Publishing Linked Data. [Online]. Disponible: <https://www.w3.org/TR/ld-bp/>. [Ultimo acceso: 06-02-2016]
- [20] Paul Davidson. (2009, Enero) GOV.UK. [Online]. Disponible en: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/60975/designing-URI-sets-uk-public-sector.pdf. [Ultimo acceso: 31-06-2016]
- [21] Dan Brickley and R.V. Guha. (2014, Febrero) RDF Schema 1.1. [Online]. Disponible en: https://www.w3.org/TR/RDF-schema/#ch_subclassof. [Ultimo acceso: 15-08-2016]
- [22] Maria Hallo, Sergio Lujan, and Tubón Dario, "Cuadro de mando para la evaluación de la Revista Politécnica," Quito, 2015.
- [23] OpenLink Software. (2009) How to Install Virtuoso Open Source (VOS) on Ubuntu Linux. [Online]. Disponible en: http://Virtuoso.openlinksw.com/dataspace/doc/dav/wiki/Main/VOSUbuntuNotes#Installing_Virtuoso. [Ultimo acceso: 23-06-2016]
- [24] Natanael Arndt. (2014, Marzo) Carrying out the OntoWiki Installation. [Online]. Disponible en: <https://github.com/AKSW/OntoWiki/wiki/Carrying-out-the-Installation> [Ultimo acceso: 24-06-2016]
- [25] Konrad Abicht. (2013, Octubre) Install CubeViz. [Online]. Disponible en: <https://github.com/AKSW/CubeViz.OntoWiki/wiki/Install-using-archive>. [Ultimo acceso: 25-06-2016]

GLOSARIO

- **Dataset:** conjunto de datos.
- **Endpoint:** es la interface a través de la cual los sistemas externos pueden consultar, enviando y recibiendo información.
- **Grafo:** representación simbólica de los elementos constituidos de un sistema o conjunto, mediante esquemas gráficos.
- **Metadato:** datos altamente estructurados que describen información, describen el contenido, la calidad, la condición y otras características de los datos. Es "Información sobre información" o "datos sobre los datos".
- **Nodos:** es un punto o espacio en diversas disciplinas en donde confluyen varios otros puntos en interrelación.
- **Tripleta:** grupo de tres personas o cosas.
- **Tuplas:** es una secuencia de valores agrupados.
- **SPARQL:** SPARQL Protocol and RDF Query Language. Lenguaje estandarizado para la consulta de grafos RDF.
- **Unicode:** es un estándar de codificación de caracteres diseñado para facilitar el tratamiento informático, transmisión y visualización de textos de múltiples lenguajes y disciplinas técnicas.

ANEXOS

ANEXO A Instalación de Extensión RDF en Google Analytics.

Descargar la extensión desde la siguiente dirección:

<https://github.com/fadmaa/grefine-rdf-extension/downloads>

Personal Open source Business Explore Pricing Blog Support This repository Search Sign in Sign up

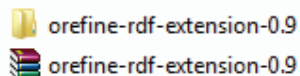
fadmaa / grefine-rdf-extension Watch 18 Star 82 Fork 24

Code Issues 64 Pull requests 0 Projects 0 Wiki Pulse Graphs

Download Packages

- 14,032 downloads
[grefine-rdf-extension-0.8.0.zip](#) — Latest version. Released on 25/07/2012
 10.4MB · Uploaded on 26 Jul 2012
- 118 downloads
[grefine-rdf-extension-0.6.0.zip](#)
 13.5MB · Uploaded on 25 Jul 2012
- 0 downloads
[grefine-rdf-extension-0.7.0.zip](#)
 14.0MB · Uploaded on 25 Jul 2012
- 92 downloads
[grefine-rdf-extension-0.7.5.zip](#)
 14.0MB · Uploaded on 25 Jul 2012

Descomprimir el contenido de la carpeta.



Copiar el contenido de la extensión dentro del directorio google-refine-2.5-r2407/webapp/extensions de Google Refine.



Verificar que se ha agregado la pestaña Edit RDF en Google Refine.

ANEXO B Estructura del Cubo de Datos RDF en formato JSON.

```
[
  {
    "op": "rdf-extension/save-rdf-schema",
    "description": "Save RDF schema skeleton",
    "schema": {
      "baseUri": "http://opendata.epn.edu.ec/",
      "prefixes": [
        {
          "name": "dcss",
          "uri": "http://opendata.epn.edu.ec/cubos/revista_epn/dcss/"
        },
        {
          "name": "qb",
          "uri": "http://purl.org/linked-data/cube#"
        },
        {
          "name": "rdf",
          "uri": "http://www.w3.org/1999/02/22-rdf-syntax-ns#"
        },
        {
          "name": "owl",
          "uri": "http://www.w3.org/2002/07/owl#"
        },
        {
          "name": "prop",
          "uri": "http://opendata.epn.edu.ec/cubos/revista_epn/prop/"
        },
        {
          "name": "xsd",
          "uri": "http://www.w3.org/2001/XMLSchema#"
        },
        {
          "name": "rdfs",
          "uri": "http://www.w3.org/2000/01/rdf-schema#"
        },
        {
          "name": "dcat",
          "uri": "http://www.w3.org/ns/dcat#"
        },
        {
          "name": "prov",
          "uri": "http://www.w3.org/ns/prov#"
        },
        {
          "name": "foaf",
          "uri": "http://xmlns.com/foaf/0.1/"
        },
        {
          "name": "dc",
          "uri": "http://purl.org/dc/elements/1.1/"
        },
        {
          "name": "ds",
          "uri": "http://opendata.epn.edu.ec/cubos/revista_epn/dataset/"
        }
      ]
    },
    "rootNodes": [
      {

```

```

        "nodeType": "resource",
        "value":
"http://opendata.epn.edu.ec/cubos/revista_epn/prop/titulo",
        "rdfTypes": [
            {
                "uri": "http://purl.org/linked-
data/cube#DimensionProperty",
                "curie": "qb:DimensionProperty"
            }
        ],
        "links": [
            {
                "uri": "http://www.w3.org/2000/01/rdf-schema#label",
                "curie": "rdfs:label",
                "target": {
                    "nodeType": "literal",
                    "value": "Titulo"
                }
            }
        ]
    },
    {
        "nodeType": "resource",
        "value":
"http://opendata.epn.edu.ec/cubos/revista_epn/prop/pais",
        "rdfTypes": [
            {
                "uri": "http://purl.org/linked-
data/cube#DimensionProperty",
                "curie": "qb:DimensionProperty"
            }
        ],
        "links": [
            {
                "uri": "http://www.w3.org/2000/01/rdf-schema#label",
                "curie": "rdfs:label",
                "target": {
                    "nodeType": "literal",
                    "value": "Pais"
                }
            }
        ]
    },
    {
        "nodeType": "resource",
        "value":
"http://opendata.epn.edu.ec/cubos/revista_epn/prop/año",
        "rdfTypes": [
            {
                "uri": "http://purl.org/linked-
data/cube#DimensionProperty",
                "curie": "qb:DimensionProperty"
            }
        ],
        "links": [
            {
                "uri": "http://www.w3.org/2000/01/rdf-schema#label",
                "curie": "rdfs:label",
                "target": {
                    "nodeType": "literal",

```

```

        "value": "Año"
      }
    ]
  },
  {
    "nodeType": "resource",
    "value":
"http://opendata.epn.edu.ec/cubos/revista_epn/prop/mes",
    "rdfTypes": [
      {
        "uri": "http://purl.org/linked-
data/cube#DimensionProperty",
        "curie": "qb:DimensionProperty"
      }
    ],
    "links": [
      {
        "uri": "http://www.w3.org/2000/01/rdf-schema#label",
        "curie": "rdfs:label",
        "target": {
          "nodeType": "literal",
          "value": "Mes"
        }
      }
    ]
  },
  {
    "nodeType": "resource",
    "value":
"http://opendata.epn.edu.ec/cubos/revista_epn/prop/visita",
    "rdfTypes": [
      {
        "uri": "http://purl.org/linked-data/cube#MeasureProperty",
        "curie": "qb:MeasureProperty"
      }
    ],
    "links": [
      {
        "uri": "http://www.w3.org/2000/01/rdf-schema#label",
        "curie": "rdfs:label",
        "target": {
          "nodeType": "literal",
          "value": "Visitas"
        }
      }
    ]
  },
  {
    "nodeType": "resource",
    "value":
"http://opendata.epn.edu.ec/cubos/revista_epn/dataset/dataset-
revista_epn",
    "rdfTypes": [
      {
        "uri": "http://purl.org/linked-data/cube#DataSet",
        "curie": "qb:DataSet"
      }
    ],
    "links": [

```

```

    {
      "uri": "http://www.w3.org/2000/01/rdf-schema#comment",
      "curie": "rdfs:comment",
      "target": {
        "nodeType": "literal",
        "value": "Conjunto de datos para el análisis de las
visitas al sitio Web de la Revista Politécnica"
      }
    },
    {
      "uri": "http://purl.org/linked-data/cube#structure",
      "curie": "qb:structure",
      "target": {
        "nodeType": "resource",
        "value":
"http://opendata.epn.edu.ec/cubos/revista_epn/dataset/dsd-revista_epn",
        "rdfTypes": [
          {
            "uri": "http://purl.org/linked-
data/cube#DataStructureDefinition",
            "curie": "qb:DataStructureDefinition"
          }
        ],
        "links": [
          {
            "uri": "http://purl.org/linked-data/cube#component",
            "curie": "qb:component",
            "target": {
              "nodeType": "resource",
              "value":
"http://opendata.epn.edu.ec/cubos/revista_epn/dccs/titulo",
              "rdfTypes": [
                {
                  "uri": "http://purl.org/linked-
data/cube#ComponentSpecification",
                  "curie": "qb:ComponentSpecification"
                }
              ],
              "links": [
                {
                  "uri": "http://purl.org/linked-
data/cube#dimension",
                  "curie": "qb:dimension",
                  "target": {
                    "nodeType": "resource",
                    "value":
"http://opendata.epn.edu.ec/cubos/revista_epn/prop/titulo",
                    "rdfTypes": [],
                    "links": []
                  }
                }
              ],
              "links": []
            }
          }
        ],
        "uri": "http://www.w3.org/2000/01/rdf-
schema#label",
        "curie": "rdfs:label",
        "target": {
          "nodeType": "literal",
          "value": "Titulo"
        }
      }
    },
  ],
}

```

```

        {
            "uri": "http://www.w3.org/2000/01/rdf-
schema#comment",
            "curie": "rdfs:comment",
            "target": {
                "nodeType": "literal",
                "value": "Corresponde al titulo del documento
visitado."
            }
        }
    ],
},
{
    "uri": "http://purl.org/linked-data/cube#component",
    "curie": "qb:component",
    "target": {
        "nodeType": "resource",
        "value":
"http://opendata.epn.edu.ec/cubos/revista_epn/dccs/pais",
        "rdfTypes": [
            {
                "uri": "http://purl.org/linked-
data/cube#ComponentSpecification",
                "curie": "qb:ComponentSpecification"
            }
        ],
        "links": [
            {
                "uri": "http://purl.org/linked-
data/cube#dimension",
                "curie": "qb:dimension",
                "target": {
                    "nodeType": "resource",
                    "value":
"http://opendata.epn.edu.ec/cubos/revista_epn/prop/pais",
                    "rdfTypes": [],
                    "links": []
                }
            }
        ],
    },
    {
        "uri": "http://www.w3.org/2000/01/rdf-
schema#label",
        "curie": "rdfs:label",
        "target": {
            "nodeType": "literal",
            "value": "Pais"
        }
    },
    {
        "uri": "http://www.w3.org/2000/01/rdf-
schema#comment",
        "curie": "rdfs:comment",
        "target": {
            "nodeType": "literal",
            "value": "Ubicación geográfica desde donde se
realiza la visita al sitio web de la Revista Politécnica."
        }
    }
]

```

```

    }
  },
  {
    "uri": "http://purl.org/linked-data/cube#component",
    "curie": "qb:component",
    "target": {
      "nodeType": "resource",
      "value":
"http://opendata.epn.edu.ec/cubos/revista_epn/dccs/año",
      "rdfTypes": [
        {
          "uri": "http://purl.org/linked-
data/cube#ComponentSpecification",
          "curie": "qb:ComponentSpecification"
        }
      ],
      "links": [
        {
          "uri": "http://purl.org/linked-
data/cube#dimension",
          "curie": "qb:dimension",
          "target": {
            "nodeType": "resource",
            "value":
"http://opendata.epn.edu.ec/cubos/revista_epn/prop/año",
            "rdfTypes": [],
            "links": []
          }
        }
      ],
      {
        "uri": "http://www.w3.org/2000/01/rdf-
schema#label",
        "curie": "rdfs:label",
        "target": {
          "nodeType": "literal",
          "value": "Año"
        }
      },
      {
        "uri": "http://www.w3.org/2000/01/rdf-
schema#comment",
        "curie": "rdfs:comment",
        "target": {
          "nodeType": "literal",
          "value": "Año en el que se realiza la
visita."
        }
      }
    ]
  }
},
{
  "uri": "http://purl.org/linked-data/cube#component",
  "curie": "qb:component",
  "target": {
    "nodeType": "resource",
    "value":
"http://opendata.epn.edu.ec/cubos/revista_epn/dccs/mes",
    "rdfTypes": [
      {

```



```

        "uri": "http://purl.org/linked-
data/cube#ComponentSpecification",
        "curie": "qb:ComponentSpecification"
    }
  ],
  "links": [
    {
      "uri": "http://purl.org/linked-
data/cube#dimension",
      "curie": "qb:dimension",
      "target": {
        "nodeType": "resource",
        "value":
"http://opendata.epn.edu.ec/cubos/revista_epn/prop/mes",
        "rdfTypes": [],
        "links": []
      }
    },
    {
      "uri": "http://www.w3.org/2000/01/rdf-
schema#label",
      "curie": "rdfs:label",
      "target": {
        "nodeType": "literal",
        "value": "Mes"
      }
    },
    {
      "uri": "http://www.w3.org/2000/01/rdf-
schema#comment",
      "curie": "rdfs:comment",
      "target": {
        "nodeType": "literal",
        "value": "Mes en el cual se realiza la
visita."
      }
    }
  ]
},
{
  "uri": "http://purl.org/linked-data/cube#component",
  "curie": "qb:component",
  "target": {
    "nodeType": "resource",
    "value":
"http://opendata.epn.edu.ec/cubos/revista_epn/dccs/visitas",
    "rdfTypes": [
      {
        "uri": "http://purl.org/linked-
data/cube#ComponentSpecification",
        "curie": "qb:ComponentSpecification"
      }
    ]
  },
  "links": [
    {
      "uri": "http://purl.org/linked-
data/cube#measure",
      "curie": "qb:measure",
      "target": {

```

```

        "nodeType": "resource",
        "value":
"http://opendata.epn.edu.ec/cubos/revista_epn/prop/visita",
        "rdfTypes": [],
        "links": []
    }
},
{
    "uri": "http://www.w3.org/2000/01/rdf-
schema#label",
    "curie": "rdfs:label",
    "target": {
        "nodeType": "literal",
        "value": "Visitas"
    }
},
{
    "uri": "http://www.w3.org/2000/01/rdf-
schema#comment",
    "curie": "rdfs:comment",
    "target": {
        "nodeType": "literal",
        "value": "Representa el N° de visitas que se
realizan a sitio Web de la Revista Politécnica."
    }
}
]
}
},
{
    "uri": "http://purl.org/dc/elements/1.1/creator",
    "curie": "dc:creator",
    "target": {
        "nodeType": "literal",
        "value": "Darío Tubón",
        "lang": "es"
    }
},
{
    "uri": "http://purl.org/dc/elements/1.1/title",
    "curie": "dc:title",
    "target": {
        "nodeType": "literal",
        "value": "Conjunto de Datos Estadísticos de la Revista
Politécnica",
        "lang": "es"
    }
},
{
    "uri": "http://purl.org/dc/elements/1.1/description",
    "curie": "dc:description",
    "target": {
        "nodeType": "literal",
        "value": "Conjunto de Datos que contiene un Cubo de Datos
RDF, el mismo que representan a las visitas al sitio web de la Revista
Politécnica."
    }
}

```

```

    },
    {
      "uri": "http://purl.org/dc/elements/1.1/date",
      "curie": "dc:date",
      "target": {
        "nodeType": "literal",
        "value": "2016-06-01",
        "valueType": "http://www.w3.org/2001/XMLSchema#date"
      }
    },
    {
      "uri": "http://www.w3.org/2000/01/rdf-schema#label",
      "curie": "rdfs:label",
      "target": {
        "nodeType": "literal",
        "value": "Conjunto de Datos de la Revista Politecnica de
la EPN"
      }
    }
  ],
  {
    "nodeType": "cell-as-resource",
    "expression":
"http://opendata.epn.edu.ec/cubos/revista_epn/dataset/obs'+cells[\"INDEX
\"]].value.urlify()",
    "isRowNumberCell": false,
    "columnName": "VIEWS",
    "rdfTypes": [
      {
        "uri": "http://purl.org/linked-data/cube#Observation",
        "curie": "qb:Observation"
      }
    ],
    "links": [
      {
        "uri":
"http://opendata.epn.edu.ec/cubos/revista_epn/prop/pais",
        "curie":
"http://opendata.epn.edu.ec/cubos/revista_epn/prop/pais",
        "target": {
          "nodeType": "cell-as-literal",
          "expression": "value",
          "isRowNumberCell": false,
          "columnName": "COUNTRY"
        }
      },
      {
        "uri":
"http://opendata.epn.edu.ec/cubos/revista_epn/prop/año",
        "curie":
"http://opendata.epn.edu.ec/cubos/revista_epn/prop/año",
        "target": {
          "nodeType": "cell-as-literal",
          "expression": "value",
          "isRowNumberCell": false,
          "columnName": "YEAR"
        }
      }
    ],
  },
  {

```

```

        "uri":
"http://opendata.epn.edu.ec/cubos/revista_epn/prop/mes",
        "curie":
"http://opendata.epn.edu.ec/cubos/revista_epn/prop/mes",
        "target": {
            "nodeType": "cell-as-literal",
            "expression": "value",
            "isRowNumberCell": false,
            "columnName": "MONTH"
        }
    },
    {
        "uri":
"http://opendata.epn.edu.ec/cubos/revista_epn/prop/titulo",
        "curie":
"http://opendata.epn.edu.ec/cubos/revista_epn/prop/titulo",
        "target": {
            "nodeType": "cell-as-literal",
            "expression": "value",
            "isRowNumberCell": false,
            "columnName": "TITLE"
        }
    },
    {
        "uri":
"http://opendata.epn.edu.ec/cubos/revista_epn/prop/visita",
        "curie":
"http://opendata.epn.edu.ec/cubos/revista_epn/prop/visita",
        "target": {
            "nodeType": "cell-as-literal",
            "expression": "value",
            "isRowNumberCell": false,
            "valueType": "http://www.w3.org/2001/XMLSchema#int",
            "columnName": "VIEWS"
        }
    },
    {
        "uri": "http://purl.org/linked-data/cube#dataSet",
        "curie": "qb:dataSet",
        "target": {
            "nodeType": "resource",
            "value":
"http://opendata.epn.edu.ec/cubos/revista_epn/dataset/dataset-
revista_epn",
            "rdfTypes": [],
            "links": []
        }
    }
]
},
{
    "nodeType": "cell-as-resource",
    "expression": "value",
    "isRowNumberCell": true,
    "rdfTypes": [],
    "links": []
},
{
    "nodeType": "cell-as-resource",
    "expression": "value",

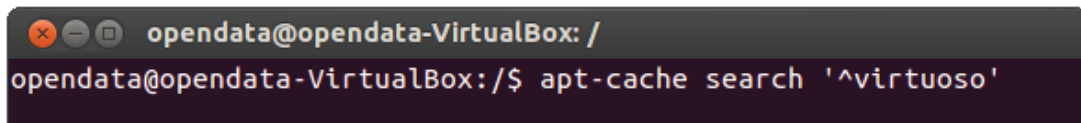
```

```
    "isRowNumberCell": true,  
    "rdfTypes": [],  
    "links": []  
  },  
  {  
    "nodeType": "cell-as-resource",  
    "expression": "value",  
    "isRowNumberCell": true,  
    "rdfTypes": [],  
    "links": []  
  }  
]  
}  
]
```

ANEXO C Instalación de Servidor Virtuoso.

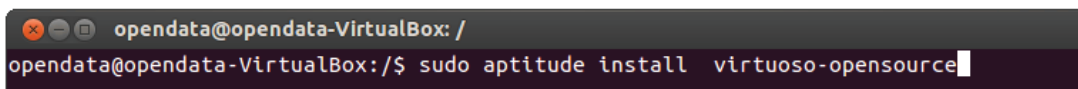
En la consola de Linux ejecutar el siguiente comando.

```
apt-cache search '^virtuoso'
```



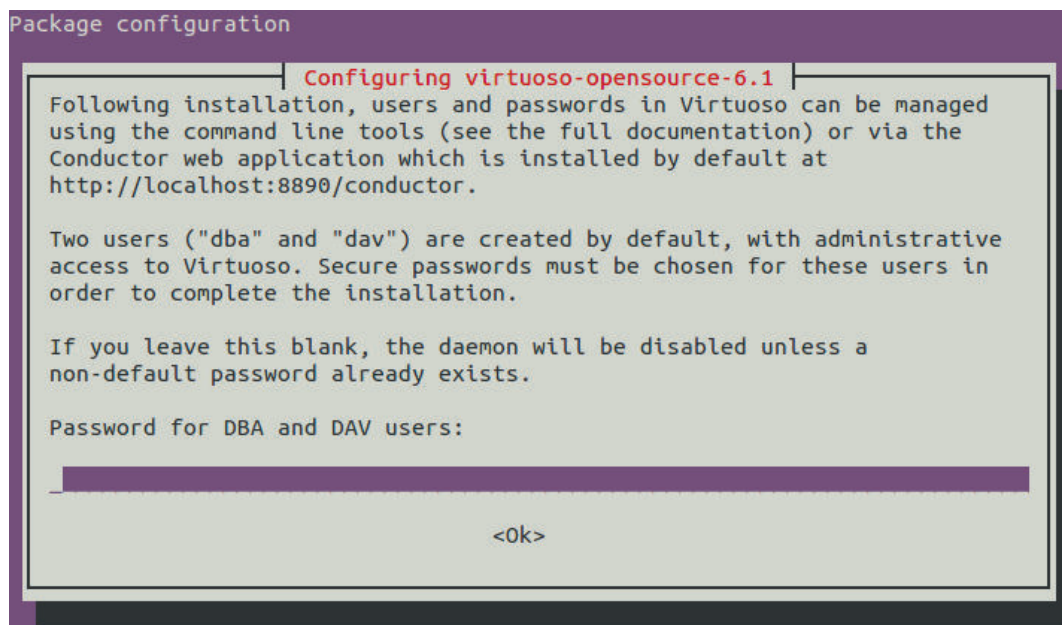
```
opendata@opendata-VirtualBox: /
opendata@opendata-VirtualBox:/$ apt-cache search '^virtuoso'
```

Ejecutar la instalación de Virtuoso Open Source.



```
opendata@opendata-VirtualBox: /
opendata@opendata-VirtualBox:/$ sudo aptitude install virtuoso-opensource
```

Ingresar la contraseña para ingresar al servidor Virtuoso, la misma que será utilizada por Ontowiki.



```
Package configuration
Configuring virtuoso-opensource-6.1
Following installation, users and passwords in Virtuoso can be managed
using the command line tools (see the full documentation) or via the
Conductor web application which is installed by default at
http://localhost:8890/conductor.

Two users ("dba" and "dav") are created by default, with administrative
access to Virtuoso. Secure passwords must be chosen for these users in
order to complete the installation.

If you leave this blank, the daemon will be disabled unless a
non-default password already exists.

Password for DBA and DAV users:
_____

<Ok>
```

Terminado este proceso, el servidor Virtuoso está listo y corriendo.

The screenshot displays the Virtuoso Conductor web interface. At the top left is the logo for VIRTUOSO CONDUCTOR. To the right of the logo, it indicates the user is "not logged in" and provides a "Home" link. Below the logo is a navigation menu with the following items: Home, System Admin, Database, Replication, Web Application Server, XML, Web Services, Linked Data, and NNTP.

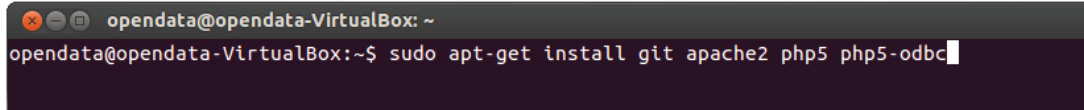
On the left side, there is a sidebar with a "Login into Database" section. This section includes an "Account" label, a text input field containing "dba", a "Password" label, a password input field with masked characters, a "Remember me" checkbox, and a "Login" button. Below the login section are several links: "Interactive SQL (ISQL)", "Virtuoso Start Menu", "Documentation (web)", "Tutorials (web)", "Virtuoso Web Site", and "OpenLink Software". At the bottom of the sidebar, it shows "Version: 06.01.3127" and "Build: Feb 15 2012".

The main content area features a central graphic with a stylized human figure at the base. Above the figure are five blue rectangular boxes representing different capabilities: "Enterprise Data Management", "Enterprise Information Integration", "Web Services Platform", "Business Process Integration", and "Enterprise Collaboration". These boxes are arranged around a central semi-circular arc that contains the text "BUSINESS PROCESS INTEGRATION" and "COLLABORATION".

ANEXO D Instalación de OntoWiki.

Instalar las dependencias de PHP necesarias, ingresar el siguiente comando.

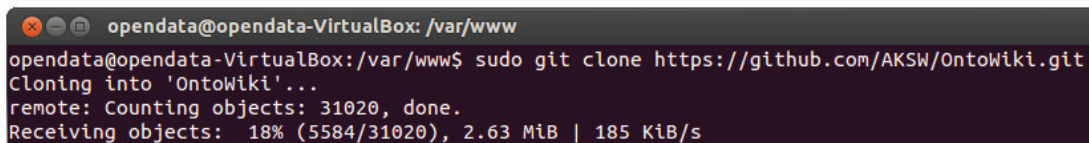
```
sudo apt-get install git apache2 php5 php5-odbc
```

A terminal window with a dark background. The prompt is 'opendata@opendata-VirtualBox: ~'. The command 'sudo apt-get install git apache2 php5 php5-odbc' is entered and followed by a cursor.

```
opendata@opendata-VirtualBox: ~  
opendata@opendata-VirtualBox:~$ sudo apt-get install git apache2 php5 php5-odbc
```

Instalar Ontowiki desde su repositorio en GitHub.

```
cd /var/www/html  
git clone https://github.com/AKSW/OntoWiki.git  
cd OntoWiki
```

A terminal window with a dark background. The prompt is 'opendata@opendata-VirtualBox: /var/www'. The command 'sudo git clone https://github.com/AKSW/OntoWiki.git' is entered. The output shows the cloning process: 'Cloning into 'OntoWiki'...', 'remote: Counting objects: 31020, done.', and 'Receiving objects: 18% (5584/31020), 2.63 MiB | 185 KiB/s'.

```
opendata@opendata-VirtualBox: /var/www  
opendata@opendata-VirtualBox:/var/www$ sudo git clone https://github.com/AKSW/OntoWiki.git  
Cloning into 'OntoWiki'...  
remote: Counting objects: 31020, done.  
Receiving objects: 18% (5584/31020), 2.63 MiB | 185 KiB/s
```

Ejecutar el comando **make deploy** para descargar las librerías Erfurt, RDFAuthor y Zend. Dentro del directorio OntoWiki (**/var/www/html/OntoWiki**)

ANEXO E Configuración de OntoWiki.

Los archivos que se deben configurar son: `odbc.ini` y `odbcinst.ini`. Estos archivos se encuentran en el directorio `/etc`.

Agregar la siguiente línea en el archivo `odbcinst.ini`

```
[virtuoso-odbc]
Driver = <prefix>/lib/virtodbc.so
```

```
[virtuoso-odbc]
Driver = <prefix>/lib/virtodbc.so
```

Agregar las siguientes líneas al archivo `odbc.ini`

```
[ODBC Data Sources]
VOS = Virtuoso
[VOS]
Driver = virtuoso-odbc
Description = Virtuoso Open-Source Edition
Address = localhost:1111
```

```
[ODBC Data Sources]
VOS = Virtuoso

[VOS]
Driver = virtuoso-odbc
Description = Virtuoso Open-Source Edition
Address = localhost:1111
```

Para comprobar el correcto funcionamiento del ODBC ejecutamos el comando `make odbctest`, como resultado se muestran los grafos que tiene alojado Virtuoso.

```
ontowiki@ubuntu:/var/www/html/OntoWiki$ make odbctest
Your connection to Virtuoso seems to work fine:
- http://www.openlinksw.com/schemas/virttrdf#
- http://localhost:8890/sparql
- http://localhost:8890/DAV/
- http://www.w3.org/2002/07/owl#
- http://localhost/OntoWiki/Config/
- http://ns.ontowiki.net/SysOnt/
```

Copiar el contenido de `/var/www/html/OntoWiki/config.ini.dist` a `/var/www/html/OntoWiki/config.ini`, en el cual se cambiaran las credenciales configuradas para el servidor Virtuoso.

```

;;;;
;; Virtuoso backend specific options
;;
store.virtuoso.dsn          = VOS|
store.virtuoso.username    = "dba"
store.virtuoso.password    = "opendata"
;; affect the main search: searches <= 4 will be exact search (instead
of bif:contains)
store.virtuoso.search_max_length_for_bifcontains = "4"
;store.virtuoso.use_persistent_connection = true

```

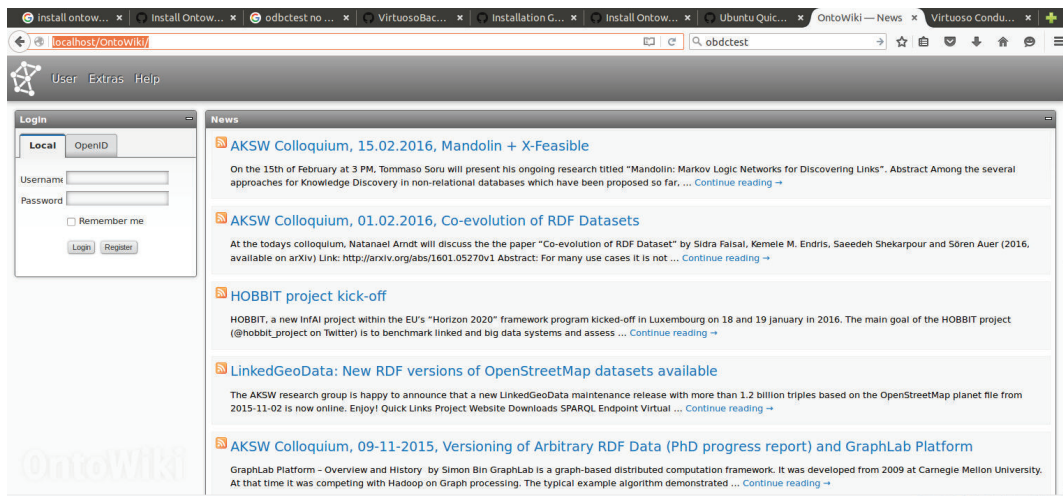
Reiniciar el servidor Apache y el servidor Virtuoso.

```

sudo service virtuoso-opensource-6.1 start
sudo service apache2 start/restart

```

Ingresar a <http://localhost/OntoWiki>, para verificar que OntoWiki se ha instalado correctamente.



Con las credenciales configuradas en la instalación de Virtuoso ingresamos a Ontowiki.

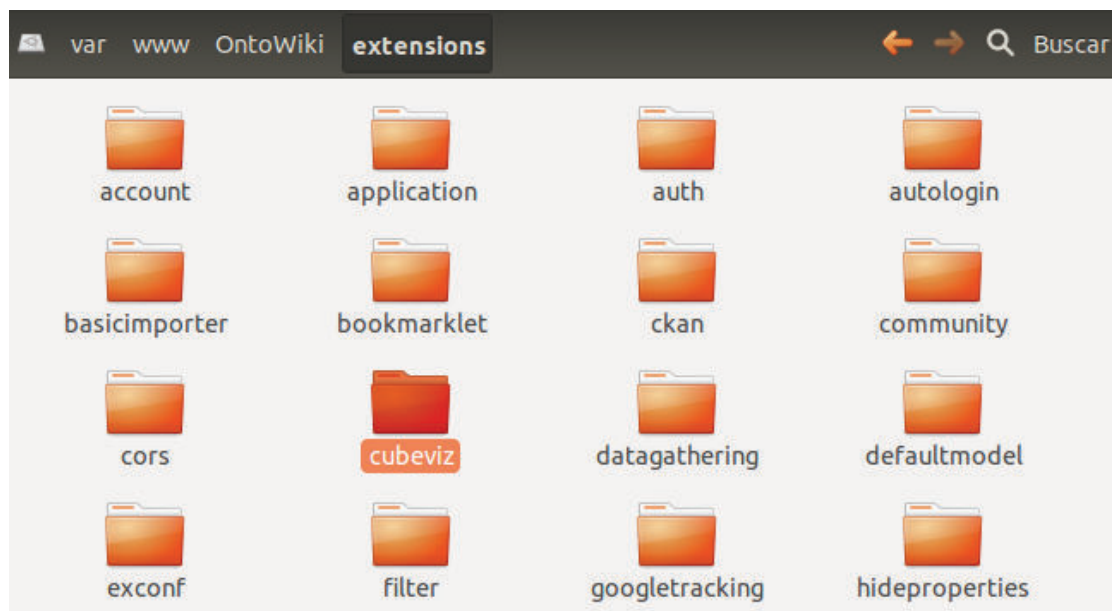
ANEXO F Instalación CubeViz.

Descargar la carpeta que contiene la extensión CubeViz para OntoWiki, en el siguiente enlace se encuentra el enlace de descarga: <https://github.com/AKSW/cubeviz.ontowiki/wiki/Install-using-archive>

Descomprimir la carpeta que contine CubeViz




Copiar la carpeta cubeviz dentro de la carpeta que contiene las extensiones utilizadas por OntoWiki: `/var/www/ontowiki/extensions`




Vaciar el contenido de la carpeta `/var/www/ontowiki/cache`, obligando a OntoWiki a reconocer la nueva extensión instalada.

Data Selection


Select a part of the Dataset ?


 Conjunto de Datos de la..

Select Unit and Measurement ?

 Visitas
Representa el N° de visitas que se realizan a sitio Web...

Configure the Dimensions ?

 Pais Change
10 of 44 Selected
Ubicación geográfica desde donde se..

 Titulo Change
10 of 100 Selected
Corresponde al título del documento..

[Share](#) [Update Chart](#)