UNIVERSIDAD DE CUENCA

FACULTAD DE INGENIERÍA
CENTRO DE POSGRADOS

MAESTRÍA EN GESTIÓN ESTRATÉGICA DE TECNOLOGÍAS DE LA INFORMACIÓN

Integración de los Sistemas de Información de la Universidad de Cuenca utilizando tecnología semántica para la utilización de buscadores inteligentes

TRABAJO DE GRADUACIÓN PREVIO A LA OBTENCION DEL GRADO DE MAGISTER EN GESTIÓN ESTRATÉGICA DE TECNOLOGÍAS DE LA INFORMACIÓN

Autor: Ing. José Rolando Zumba Gómez C.I.: 0102492972

Director: Ing. Víctor Hugo Saquicela Galarza, PhD. C.I.: 0103599577

CUENCA – ECUADOR
Octubre – 2017

RESUMEN

El presente trabajo consiste en realizar la integración de los datos que se generan al operar los sistemas informáticos de la Universidad de Cuenca. Los sistemas están desarrollados con distintas tecnologías y almacenan sus datos en fuentes de datos heterogéneas. Los Sistemas informáticos no están integrados, esto genera complicaciones a la hora de consultar datos de los sistemas. Por esta razón se propone un modelo de integración utilizando tecnologías de la Web semántica, que permita a los usuarios realizar búsquedas de información más precisas en un contexto global. Para la integración de los datos se generó una arquitectura que servirá de guía para la integración de los datos generados en los distintos ámbitos de acción de la Universidad utilizando tecnología semántica que por sus grandes avances es una alternativa que da muchos beneficios al momento de integrar los datos y que servirá para la utilización de un buscador semántico que es una vista al usuario final como un único punto de acceso a los datos de la Universidad. Se pudo concluir que mediante el presente trabajo se estableció una arquitectura que permitió integrar los datos de la estructura organizacional de la Universidad con el área académica de grado y que también permitirá integrar con los datos de las otras áreas que la Institución crea necesaria hacerlo en un futuro.

Palabras Clave: Web semántica, Ontologías, Datos enlazados, Linked Data, NeOn

ABSTRACT

The present work is based on the integration of the data that is generated by the operation of the information systems of the University of Cuenca. The systems where developed using different technologies and they store their data in heterogeneous data sources. The information systems are not integrated which generates complications for the data consumption. It is because of this that a model of integrations is proposed using Semantic Web technologies, which allows more precise searches in a global context. For the data integration a new architecture has been generated that will be used as a guide for the integration of the data generated by deferent fields of action of the University of Cuenca using semantic technology that - because of its great advances - is an alternative that brings a lot of benefits for data integration and the use of a semantic search engine which will be the only point of access to the users of the University of Cuenca. It has been concluded that the present work has established an architecture that allowed the integration of the data of the organizational structure of the University including the third degree academic area that will also allow the integration with data from other areas that the institution will believe necessary.

Keywords: Semantic Web, Ontologies, Linked Data, NeOn

DEDICATORIA

Dedico especialmente este trabajo a DIOS por ser mi guía, y mi fuerza en todo momento y el Ser que nunca me abandona. A mi esposa por ser mi apoyo incondicional y ser la persona que con su amor me dio el aliento que necesité para culminar con éxito este camino que juntos nos propusimos seguir. A mi hija que es la personita que hace mi vida feliz y por quien doy mucho más de lo que puedo. A mi director de tesis Ing. Víctor Saquicela, que con su conocimiento y capacidad me ayudó a plasmar las ideas en este trabajo. A mi mamá que fue un gran apoyo en los momentos que más necesitaba. A mi papá y hermanas que con su amor son un pilar para seguir adelante cada día. A la familia que siempre está conmigo.

AGRADECIMIENTOS

Agradezco a DIOS por darme la vida y haber permitido llegar a este momento importante de mi formación profesional. Agradezco infinitamente a mi esposa, por ser el complemento que me ayudó a plasmar de mejor manera el presente trabajo.
Un agradecimiento especial y muy afectuoso a mi director de tesis Ing. Víctor Saquicela, por la confianza que me brindó, y que, con su conocimiento, experiencia, y tiempo, fue el pilar principal para la realización de este trabajo. A Carlos Plaza y Yolanda Aucapiña, que con su apoyo y colaboración me ayudaron a cumplir el objetivo planteado. A mi mamá, papá, hermanas y familia que de una u otra manera me ayudaron y estuvieron siempre junto a mí.

Tabla de Contenidos

Lista de figuras **7**
Lista de Tablas **8**
Cláusulas de derecho de autor **9**
Cláusulas de propiedad intelectual **10**
Capítulo 1 INTRODUCCIÓN **11**
1.1. Antecedentes 11
1.2. Problemática 11
1.3. Solución 12
1.4. Objetivos 12
Objetivo general 12
Objetivos específicos 12
1.5. Alcance 13
1.6. Contenido del documento 13
Capítulo 2 MARCO TEORICO **14**
2.1. Integración de Datos 14
Virtual vs. Materializado 15
2.2. Metadatos 16
2.3. Ontologías 16
2.3.1. Componentes de una ontología 16
2.3.2. Tipos de ontologías 17
2.4. Metodologías para la construcción de ontologías 18
2.4.1. CYC 18
2.4.2. USCHOLD Y KING 19
2.4.3. METHONTOLOGY 19
2.4.4. NeOn 20
2.5. Web Semántica 21
2.5.1. Arquitectura de la Web Semántica 22
2.5.2. Ventajas de la Web Semántica 24
2.6. Datos enlazados (Linked Data) 25
2.6.1. Principios de los datos enlazados 25
2.6.2. Ciclo de vida de los datos enlazados 25

2.6.3. Almacenamiento de datos en la Web Semántica 26
2.7. Lenguajes de consulta para RDF 27
SPARQL 28
2.8. Buscadores 28
Capítulo 3 TRABAJOS RELACIONADOS 30
Capítulo 4 PROCESO DE INTEGRACIÓN DE DATOS PARA LOS SISTEMAS DE INFORMACIÓN DE LA UNIVERSIDAD DE CUENCA 34
4.1. Análisis de las alternativas para la integración de los datos: Virtual vs. Materializado 34
4.2. Proceso de integración de datos 35
4.2.1. Especificación 37
4.2.2. Modelamiento 44
4.2.3. Generación 51
4.2.4. Enlaces 56
4.2.5. Publicación 59
4.2.6. Validación 60
4.2.7. Explotación 61
4.3. Resumen del proceso de integración 62
Capítulo 5 EXPLOTACIÓN DE LOS DATOS 64
5.1. Definición del prototipo 65
5.2. Buscador semántico 66
5.3. Búsqueda 67
5.4. Pantalla de resultado de las búsquedas 68
Capítulo 6 CONCLUSIONES Y TRABAJOS FUTUROS 71
Glosario de términos 74
ANEXO 1 Herramientas utilizadas para la publicación de datos enlazados 76
Especificación 76
Modelamiento 76
Generación 77
Enlaces 77
Publicación 77
Explotación 77
Referencias Bibliográficas 79

Lista de figuras

Figura 1:Arquitectura de la Web Semántica 22
Figura 2: Ciclo de vida de los datos enlazados 26
Figura 3: Representación de una tripleta 27
Figura 4: Ciclo de vida para integrar los datos de la Universidad de Cuenca 37
Figura 5:Arquitectura para la integración de datos en la Universidad de Cuenca 39
Figura 6:Análisis de las fuentes y tipos de datos 40
Figura 7: Organigrama de la Universidad de Cuenca 41
Figura 8: Especificación de requerimientos del SGA 42
Figura 9: Mapa conceptual de alto nivel de la red odontológica de la Universidad de Cuenca 50
Figura 10: Vista de la ontología generada con la herramienta Protégé 51
Figura 11: Mapeo de datos y recursos ontológicos 54
Figura 12: Herramientas utilizadas en la fase de generación 55
Figura 13: Visualización de las tripletas en forma de grafos enlazados 57
Figura 14: Visualización de las tripletas en formato XML 58
Figura 15: Publicación de datos enlazados 59
Figura 16:Ejemplo de consulta SPARQL vs. SQL 61
Figura 17:Definición del prototipo 66
Figura 18:Pantalla principal del buscador semántico 67
Figura 19:Ejemplo de consulta en el buscador semántico 68
Figura 20:Página de resultados 69
Figura 21:Página de descripción del recurso RDF 69
Figura 22:Información consultada vista en forma de grafo 70
Figura 23:Componentes en Pentaho para el ciclo de vida de los datos enlazados 78

Lista de Tablas

Tabla 1: Dominios analizados 35
Tabla 2: Detalle de datos publicados 43
Tabla 3: Ejemplo del documento de especificaciones 45
Tabla 4: Ontologías candidatas según la categoría 48
Tabla 5: Criterios para seleccionar ontologías para reutilizar 48
Tabla 6: Calificaciones de las ontologías candidatas 49
Tabla 7: Ontologías seleccionadas según la categoría 49
Tabla 8: Ejemplos de preguntas planteadas para la validación 60

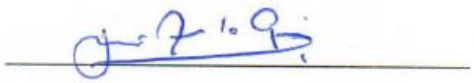
Cláusulas de derecho de autor

Cláusula de licencia y autorización para publicación en el Repositorio Institucional

José Rolando Zumba Gómez en calidad de autor y titular de los derechos morales y patrimoniales del trabajo de titulación **"Integración de los Sistemas de Información de la Universidad de Cuenca utilizando tecnología semántica para la utilización de buscadores inteligentes"**, de conformidad con el Art. 114 del CÓDIGO ORGÁNICO DE LA ECONOMÍA SOCIAL DE LOS CONOCIMIENTOS, CREATIVIDAD E INNOVACIÓN reconozco a favor de la Universidad de Cuenca una licencia gratuita, intransferible y no exclusiva para el uso no comercial de la obra, con fines estrictamente académicos.

Asimismo, autorizo a la Universidad de Cuenca para que realice la publicación de este trabajo de titulación en el repositorio institucional, de conformidad a lo dispuesto en el Art. 144 de la Ley Orgánica de Educación Superior.

Cuenca, 8 de octubre de 2017



José Rolando Zumba Gómez
C.I: 0102492972

Cláusulas de propiedad intelectual

Cláusula de Propiedad Intelectual

José Rolando Zumba Gómez, autor del trabajo de titulación **"Integración de los Sistemas de Información de la Universidad de Cuenca utilizando tecnología semántica para la utilización de buscadores inteligentes"**, certifico que todas las ideas, opiniones y contenidos expuestos en la presente investigación son de exclusiva responsabilidad de su autor.

Cuenca, 8 de octubre de 2017

José Rolando Zumba Gómez
C.I: 0102492972

Capítulo 1
INTRODUCCIÓN

1.1. Antecedentes

En la actualidad, la creciente e inmensa cantidad de información que se encuentra disponible en la WEB, hace que las búsquedas de información generen resultados irrelevantes, y numerosos, que pueden o no contener lo que el usuario busca, por lo que los resultados obtenidos deben ser procesados por la persona que realiza la búsqueda para encontrar lo que realmente necesita, esta es una tarea compleja pues requiere gastar tiempo en realizar nuevamente el filtrado de información. Por lo anteriormente mencionado, es necesario realizar búsquedas con la mayor precisión y dando sentido a los resultados según los intereses del usuario. En la Universidad de Cuenca existe gran cantidad de datos que día a día se recolectan en las bases de datos a través de los sistemas operacionales de la Institución, cuya complejidad llega a ser mayor, porque se encuentran en distintas formatos y gestores de base de datos, adicionalmente a esto los datos se encuentra repetida y dispersa, haciendo más difícil la recuperación y combinación de la misma para satisfacer las necesidades de los usuarios.

1.2. Problemática

La información es uno de los activos intangibles más importantes en toda empresa y por lo tanto su correcto manejo es de vital importancia, puesto que es necesaria para la toma de decisiones. Los sistemas informáticos de la Universidad de Cuenca almacenan sus datos en bases de datos relacionales, sin embargo, cada sistema de información los almacena en distintos gestores de bases de datos incluso de distinta tecnología por lo que los datos no se encuentran integrados. La falta de integración de los datos ha ocasionado que los usuarios no cuenten con datos confiables para obtener las respuestas a sus consultas.

1.3. Solución

Los datos además de ser integrados deben ser manejados estructuradamente para que se pueda obtener las respuestas a las consultas de los usuarios (Tapia & Fuertes, 2014). Integrar datos de distintas fuentes es un reto, es por esto que a través de este trabajo se definirá una arquitectura para la integración mediante redes ontológicas, con el fin de que los datos se encuentren estructurados adecuadamente para que los usuarios puedan accederla y encontrarla más apropiadamente y con menos esfuerzo. Cabe indicar que los usuarios contarán con un rol específico que define el nivel de acceso a la información de cada uno de los usuarios. Para realizar la integración de las distintas fuentes de datos es necesario un modelo de datos común, para esto se utilizará el marco de descripción de recursos RDF (Codina & Rovira, 2006) para representar este modelo, puesto que es una tecnología flexible que permite representar la información de distintas fuentes.

1.4. Objetivos

A continuación, se describen los objetivos general y específicos de este trabajo.

Objetivo general

Definir una arquitectura que permita realizar la integración de los datos de los sistemas de información de la Universidad de Cuenca utilizando tecnología semántica.

Objetivos específicos

- Analizar las alternativas para la integración de los datos, ya sea una integración Materializada, Virtual o Híbrida, considerando la realidad de la Universidad de Cuenca.
- Definir una ontología organizacional (Ontología base) que servirá de base para la integración de las demás ontologías que se planteen para la Universidad de Cuenca.
- Definir un proceso de integración de datos.

- Definir un proceso de transformación o acceso de datos utilizando la arquitectura de integración planteada.
- Generar un prototipo de buscador semántico sobre los datos semánticos integrados.

1.5. Alcance

Definir una arquitectura para modelar la información de las bases de datos de la Universidad de Cuenca, de manera que a través de tecnología Semántica pueda ser consultada. Para lo cual, considerando la realidad de la Universidad de Cuenca, se analizará la mejor alternativa para la integración de datos, sea a través de integración Materializada (Anguita, 2012), Virtual (Anguita, 2012) o pudiendo aplicar una combinación entre las dos. Luego se definirá una ontología base, la misma que se crea siguiendo una metodología específica que se establecerá en este trabajo. Esta ontología base servirá para integrarse con otras ontologías que se pueden crear en un futuro para soportar la integración de los sistemas que se operan en la Universidad. Es importante indicar que el escenario de uso para este trabajo es la estructura organizacional de la Instrucción integrada con el área académica de grado.

1.6. Contenido del documento

El presente documento está estructurado de la siguiente manera: como se puede revisar en este primer capítulo se realiza una breve introducción al objeto de estudio del trabajo; indicando sus antecedes, la problemática, solución, los objetivos y alcance del proyecto. En el capítulo dos se realiza un repaso de los principales conceptos relacionados con la integración de datos enlazados (Linked Data) y la Web Semántica. El capítulo tres analiza los dominios en las que se ha aplicado la Web Semántica. En el capítulo cuatro se realiza un análisis de las alternativas de integración y se explica paso a paso el proceso realizado para publicar los datos enlazados. El capítulo cinco realiza una explicación detallada de la fase de explotación de los datos enlazados, indicando el proceso de creación del prototipo del buscador semántico. En el capítulo seis se presentan las conclusiones, recomendaciones y trabajos futuros relacionados al trabajo realizado.

Capítulo 2
MARCO TEORICO

En este capítulo se describe brevemente las temáticas referentes a la Web Semántica y Linked Data, que permitirá un mejor entendimiento de este trabajo. El capítulo está estructurado de la siguiente manera: El primer tema tratado es la integración de datos y las formas de hacerla para que se pueda obtener el conocimiento sobre los dominios de interés de la Universidad. Luego se habla sobre los metadatos porque es necesario conocer cómo se agrega la semántica a los documentos. Un tercer tema abordado es lo que se refiere a las ontologías, los componentes, los tipos y las metodologías para su construcción, de la cuales se ha seleccionado una que permitió construir la ontología base para este trabajo. A continuación, se describe brevemente los conceptos de la Web Semántica y su arquitectura que permite soportar el modelamiento de los datos en base a la ontología creada. Otro tema y bastante importante tratado en este capítulo es lo referente a los datos enlazados, sus principios y el ciclo de vida que fue utilizado y mejorado en este trabajo para la publicación de los datos en la Web. Se aborda también el tema de los lenguajes de consulta de RDF en particular SPARQL que se utilizó para las consultas de los datos. Por último, se habla de los buscadores tomando en cuenta también los buscadores semánticos.

2.1. Integración de Datos

La integración de datos permite la explotación transparente de los datos para obtener conocimiento sobre algún dominio de interés específico. Se define como el problema de combinar los datos existentes en diferentes fuentes en la web, para proveer al usuario una visión unificada de estos datos (Cali, Calvanese, de Giacomo, & Lenzerini, 2002). Por lo que, la meta de un sistema de integración de datos es ofrecer un acceso uniforme a un conjunto de fuentes de datos autónomos y heterogéneos (Doan & Halevy, 2012). Por lo tanto, se puede concluir que los

usuarios pueden acceder a la información proveniente de distintas fuentes, sin que tengan que enterarse de aspectos como su localización o estructura.

Virtual vs. Materializado

Para Doan & Halevy (2012) la integración es el proceso de combinar los datos que están dispersos en distintas fuentes y presentarlos de forma unificada. Para lograr la integración de datos surge el concepto de esquema global que consiste en una visión unificada de los datos y que brinda la posibilidad de expresar las consultas necesarias para extraer la información del sistema integrado de forma transparente para el usuario, es decir, que éste no requiere conocer sobre la localización de los datos, estructura y la manera de unificarlos. Para construir sistemas que ofrecen una interfaz homogénea e integrada se toma muy en cuenta el factor que hace referencia a la localización de los datos unificados, este factor da lugar a dos modelos bien definidos que son: 1) Materializado y 2) Virtualizado.

El modelo materializado consiste en: abstraer los datos de las distintas fuentes y bases de datos y almacenarlos en un repositorio central, este repositorio es consultado para obtener los resultados de las búsquedas, sin tener que hacer consultas específicas a cada fuente de datos. Este tipo de integración es muy utilizado cuando se conoce las fuentes y se tiene el control total de los cambios en los datos, facilitando de esta manera la evaluación de las consultas (Anguita, 2012).

Por otro lado, se tiene el modelo virtualizado o distribuido, en este modelo los datos permanecen en las fuentes originales y son accedidas de forma dinámica según sea requerida. Este modelo es adecuado en entornos donde las fuentes de datos aparecen sin un control específico, y las actualizaciones son impredecibles. Como por ejemplo en ambientes colaborativos, donde no se tiene un control adecuado de los datos y las actualizaciones son generados en forma desordenada. Por esta razón, se requieren sistemas más flexibles ante los cambios y más fáciles de actualizar, haciendo que este tipo de sistemas sean mucho más costoso al desarrollarlos y menos eficientes debido a la complejidad en los procesos a ejecutar. Hay que

tener en cuenta que este modelo no asegura: la disponibilidad de los datos, el rendimiento en cuanto a consultas y problemas de fiabilidad de la información (Anguita, 2012).

2.2. Metadatos

Los metadatos son datos sobre los datos (Codina & Rovira, 2006). Los metadatos se encargan de agregar semántica a los documentos, facilitando la indexación de los datos para que los motores de búsqueda puedan encontrar la información buscada por los usuarios. Los lenguajes de visualización de contenidos Web carecen de semántica, por esa razón los lenguajes estandarizados como XML permiten la descripción de los contenidos a través de metadatos para facilitar la comunicación entre los humanos y las computadoras (Piñeres & Bonilla, 2008).

2.3. Ontologías

Según Gruber (1993) "Una ontología es una especificación explícita de una conceptualización" p.2. Por otro lado, Studer et al. (1998) afirman que "Una ontología es una especificación formal y explícita de una conceptualización compartida" p. 184. De estas dos definiciones se puede indicar que las ontologías definen el conocimiento, a través de conceptos, relaciones y axiomas o reglas de inferencia de un cierto dominio, permitiendo inferir nuevo conocimiento (Bustamante & Sequeda, 2006).

2.3.1. Componentes de una ontología

Gruber (1993) indica también que, los componentes de las ontologías que permiten representar el conocimiento son las siguientes:

- ***Conceptos***

Los conceptos son las ideas básicas que se desea formalizar, pueden ser clases de objetos, métodos, estrategias, etc.

- ***Relaciones***

Representan la interacción entre los conceptos de la ontología y definen la taxonomía del dominio.

- ***Funciones***

Las funciones son un tipo específico de una relación. Un elemento puedo ser identificado mediante el cálculo de una función en la que intervienen varios elementos de la ontología.

- ***Instancias***

Las instancias son utilizadas para representar los objetos de los conceptos.

- ***Axiomas***

Los axiomas son los teoremas o enunciados que se declaran sobre las relaciones que deben cumplir los conceptos de la ontología.

2.3.2. Tipos de ontologías

Dependiendo del enfoque, las ontologías tienen diferentes clasificaciones, por ejemplo, pueden ser clasificadas de acuerdo al nivel de generalidad. Otra posible clasificación se da de acuerdo al destino que se da a la ontología, o también según su nivel de abstracción, estas entre otras formas de clasificarlas (Reuco, 2008). Según Guarino (1998), de acuerdo al nivel de generalidad se clasifican de la siguiente manera:

- ***Ontologías de alto nivel***

Estas ontologías describen conceptos muy generales, sin importar un dominio o problema en particular.

- ***Ontologías de dominio***

Este tipo de ontologías representan el conocimiento de un dominio o subdominio genérico, especializando los conceptos descritos en la ontología de nivel superior.

- ***Ontologías de tareas***

Este tipo de ontologías se centran en la representación del conocimiento de una determinada actividad de un dominio o subdominio, al igual que las ontologías de dominio, éstas también especializan los conceptos descritos en la ontología de nivel superior.

- ***Ontologías de aplicación***

Estas ontologías describen los conceptos que pertenecen a un determinado dominio y a una tarea en particular, eso se hace a través de la especialización de los conceptos de las ontologías de dominio y de tareas.

2.4. Metodologías para la construcción de ontologías

Para construir una ontología se debe seguir una metodología. Existen varias metodologías que facilitan la construcción de ontologías como por ejemplo CYC, USCHOLD Y KING, GRÜNINGER Y FOX, KACTUS, METHONTOLOGY, NeOn, entre otras. A continuación, se realiza una breve descripción de algunas de estas metodologías.

2.4.1. CYC

Esta metodología nace en 1984 en la Corporación de Tecnología en Computación y Microelectrónica, surge como un proyecto de Inteligencia Artificial. Esta metodología consiste en dos pasos, el primer paso es para extraer manualmente el conocimiento común de diversas fuentes. El segundo paso consiste en utilizar herramientas de procesamiento de lenguaje natural con el objetivo de obtener nuevo conocimiento para la ontología (Lenat & Guha, 1990).

2.4.2. USCHOLD Y KING

Esta metodología permite usar otras ontologías para crear una nueva. Recomienda los siguientes pasos para la construcción de la ontología (Guzmán, López, & Durley, 2012):

1. Identificar el propósito de la ontología
2. Capturar los conceptos y las relaciones entre ellos
3. Codificar la ontología
4. Evaluar la ontología
5. Documentar la ontología.

2.4.3. METHONTOLOGY

Esta metodología trata la construcción de una ontología como un proyecto informático, es decir que además de los pasos que se deben ejecutar para la creación de una ontología, también se realizan actividades de planificación del proyecto. Además esta metodología permite la construcción de nuevas metodologías o reutilizar las existentes (Guzmán, López, & Durley, 2012). Es un método estructurado para construir ontologías que está basado en la experiencia adquirida para el desarrollo de ontologías de dominio de productos químicos. Los pasos a seguir en esta metodología son (Fernández, Gómez-Pérez, & Juristo, 1997):

1. Especificación: Define el alcance y granularidad
2. Adquisición del conocimiento: Es una actividad independiente, pero se realiza de forma simultánea con la especificación
3. Conceptualización: Organiza y estructura el conocimiento adquirido
4. Integración: Considera reutilizar definiciones de otras ontologías
5. Implementación: Formaliza la ontología, utilizando un lenguaje para su construcción
6. Evaluación: Comprueba su funcionamiento

2.4.4. NeOn

Según Suárez-Figueroa, Gómez-Pérez, Motta, & Gangemi (2012) la metodología NeOn está basada en un conjunto de nueve escenarios que orientan el desarrollo de ontologías colaborativas y la reutilización de recursos ontológicos. Cada ontología es independiente, sin embargo, se interconectan entre ellas a través de relaciones como son:

- Importaciones y dependencias
- Control de versiones
- Alineaciones
- Modularización

A continuación, se explica brevemente cada uno de los nueve escenarios de esta metodología:

Escenario 1. ***Desde la especificación a su implementación***: Se realiza el proceso de especificación de requisitos de la ontología, a continuación, se analiza si existen recursos que pueden ser reutilizados y por último se construye la ontología utilizando el lenguaje para su construcción.

Escenario 2. ***Reutilización y reingeniería de los recursos no-ontológicos***: Los desarrolladores analizan que recursos no-ontológicos se van a reutilizar para la elaboración de la ontología. El objetivo es transformar los recursos no-ontológicos en una ontología.

Escenario 3. ***Reutilización de recursos ontológicos***: Los desarrolladores buscan recursos ontológicos existentes en la web para ser reutilizados para el desarrollo de la nueva ontología.

Escenario 4. ***Reutilización y reingeniería de recursos ontológicos***: Los desarrolladores buscan los recursos ontológicos presentes en la web para proceder a una reingeniería según las necesidades de la nueva ontología.

Escenario 5. ***Reutilización y fusión de recursos ontológicos***: Varios recursos ontológicos que pertenecen al dominio de la nueva ontología son seleccionados para crear la nueva ontología a partir de éstos.

Escenario 6. ***Reutilización, fusión y reingeniería de recursos ontológicos***: En este escenario los desarrolladores seleccionan los recursos ontológicos pertenecientes al dominio, pero la diferencia con el escenario anterior es que se realiza una reingeniería de los recursos seleccionados combinando uno o más recursos.

Escenario 7. ***Reutilización de patrones de diseño***: Los patrones de diseño de ontologías (ODPs) son una guía para ayudar a los desarrolladores a modelar la ontología y se los puede conseguir en librerías online, estos pueden ser utilizados por los desarrolladores para superar problemas recurrentes que se presentan durante el diseño de las ontologías.

Escenario 8. ***Reestructuración de recursos ontológicos***: Los desarrolladores pueden reestructurar: esto es modularizar, extender o especializar los recursos ontológicos modificando así la red de ontologías que está siendo construida.

Escenario 9. ***Localización de recursos ontológicos***: En este escenario los desarrolladores pueden adaptar una ontología a otros lenguajes (español, francés, alemán, etc) y culturas, el resultado será una ontología multilingüe.

La metodología NeOn para la construcción de ontologías no es rígida, sino que permite combinar los escenarios obteniendo así varias opciones para la construcción de nuevas ontologías (Barrera, Nuñez, & Ramos, 2012).

2.5. Web Semántica

La Web Semántica es una extensión de la Web actual, está dotada de más significado, de tal manera que las personas y las computadoras trabajan colaborativamente. A través de la Web Semántica se solucionan problemas comunes y habituales de las búsquedas de información que realizan los usuarios, gracias a que utiliza una infraestructura común que permite compartir, procesar y transferir información de forma sencilla. La Web Semántica utiliza lenguajes universales para resolver problemas de la Web actual que carece de significado. (W3C - Web Semántica).

2.5.1. Arquitectura de la Web Semántica

La Web Semántica está estructura en varias capas (Berners-Lee, 1998). Como se puede observar en la Figura 2 la Web Semántica reposa sobre varias tecnologías, protocolos y lenguajes. Se puede ver que en el nivel inferior están los conceptos de la Web actual como son el protocolo URI y el estándar Unicode, además se puede ver una capa para la construcción de las ontologías, sobre la cual está la capa lógica y por último están las capas de pruebas y confianza.

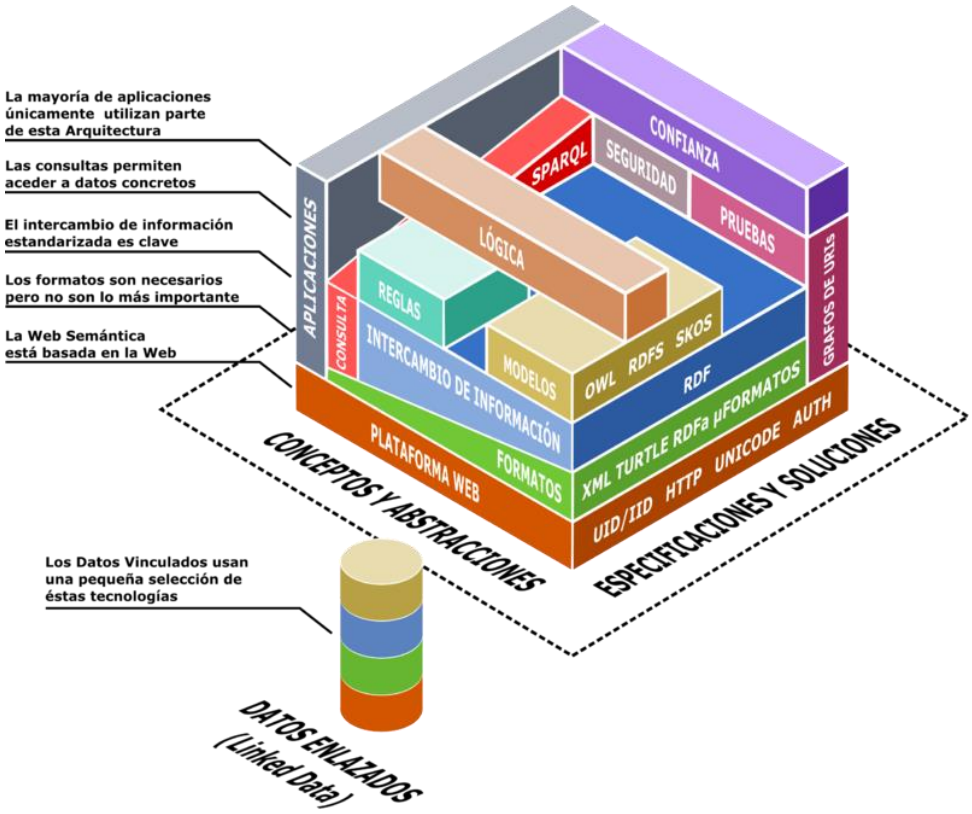


***Figura 1**:Arquitectura de la Web Semántica*

Fuente: (Wikipedia). Arquitectura tecnológica de la Web Semántica. Recuperado de https://es.wikipedia.org/wiki/Web_semántica

Capa URI – Unicode

URI (**U**niforme **R**esource **I**dentifier), es una cadena de caracteres que sirve para identificar a los recursos en Internet. Por otro lado, Unicode es un estándar que sirve para codificar el texto, para visualizar información de cualquier idioma y no permite símbolos extraños. Este nivel se encuentra en la parte inferior de la arquitectura, puesto que son la base de la Web en general (Villalba, 2007).

Capa XML+NS+XML Schema

En esta capa de la Web Semántica se agrupan las tecnologías que permiten la comunicación entre agentes. XML (E**X**tended **M**arkup **L**anguage), este es un lenguaje de etiquetado que permite el intercambio de datos y permite la lectura de datos a través de diferentes aplicaciones (W3C - XML). XML es la base para la estructuración del contenido en la Web. Por otro lado, NS (Name Spaces), permite combinar los marcados creados con XML, por último, XML Schema, que sirve para definir tipos de documentos complejos porque establece la estructura, tipos de datos y restricciones de los contenidos de un documento XML (Codina & Rovira, 2006).

Capa RDF+RDF Schema

Esta capa se basa en la anterior, define el lenguaje para expresar las ideas de la Web Semántica. RDF (**R**esource **D**escription **F**ramework) es un modelo de datos que permite representar los recursos y sus propiedades, está concebido para representar cualquier clase de recurso. Su sintaxis está basada en XML. Por otro lado, RDF Schema, es el modelo que permite definir las relaciones entre los recursos mediante clases y objetos (Codina & Rovira, 2006).

Capa Ontology vocabulary

Este nivel permite clasificar la información y extender la funcionalidad de la Web Semántica a través nuevas clases y propiedades para la definición de los recursos (Pascual,

Valdés, & Gómez). El uso de ontologías permite representar a los objetos y las relaciones que mantienen con otros objetos.

Capa Lógica

La capa lógica permite realizar las consultas e inferir el conocimiento para lograr interoperabilidad entre aplicaciones y sistemas heterogéneos (Rodriguez & Ronda, 2005). En esta capa se define los pasos que los agentes deben seguir para llegar a inferir el conocimiento como respuesta a un proceso de búsqueda de información (Peis, Herrera-Viedma, & Morales, 2007).

Capa de pruebas

En la capa de pruebas se evalúan y validan las reglas y sentencias de la capa lógica, con el fin de determinar la confiabilidad de los recursos. Esto se hace a través de tres elementos: primero a través de reglas de inferencia que fueron definidas en la capa lógica, segundo con la capacidad que tienen los agentes para probar el origen de una secuencia lógica y tercero a través de firmas digitales para verificar que la información proviene de una fuente fiable (Peis, Herrera-Viedma, & Morales, 2007).

Capa de confianza

La capa de confianza permite determinar el grado de confiabilidad de los datos, servicios y agentes. Siendo esta capa la de más alto nivel, es crucial, puesto que el éxito de la Web Semántica se alcanza cuando los usuarios confían en las operaciones realizadas y en la calidad de la información que han encontrado (Villalba, 2007).

2.5.2. Ventajas de la Web Semántica

La principal ventaja de la Web Semántica es que al estar dotada de significado incorpora contenido semántico a los documentos de la Web, esto permite que la información sea organizada por conceptos, permitiendo así que las consultas realizadas devuelvan información relevante de manera más fácil y más útil para los usuarios.

2.6. Datos enlazados (Linked Data)

La Web Semántica vincula los datos que se encuentran distribuidos en la Web a través de los datos enlazados, esto lo realiza de manera similar a como se referencian los enlaces en las páginas web, con la diferencia de que los datos se encuentran interconectados. El objetivo es que las personas y las computadoras puedan explorar y entender la información presente en la web. Gracias a los datos enlazados se puede construir la Web de los datos, esto es que los datos se encuentran distribuidos en la web formando una gran base de datos (W3C - Datos Enlazados).

2.6.1. Principios de los datos enlazados

Para lograr los objetivos de los datos enlazados se han establecido los siguientes principios (Heath & Bizer, 2011):

1. Usar URIs como identificadores
2. Usar URIs HTTP, esto es el uso de URIs sobre HTTP de manera que los recursos puedan ser buscados y encontrados en la web.
3. Usar RDF para ofrecer información sobre los recursos
4. Incluir enlaces a otros URIs

2.6.2. Ciclo de vida de los datos enlazados

Para realizar la publicación de los datos enlazados Villazón-Terrazas et al. (2012) propone un método simple y unificado para lograr publicar los datos enlazados.

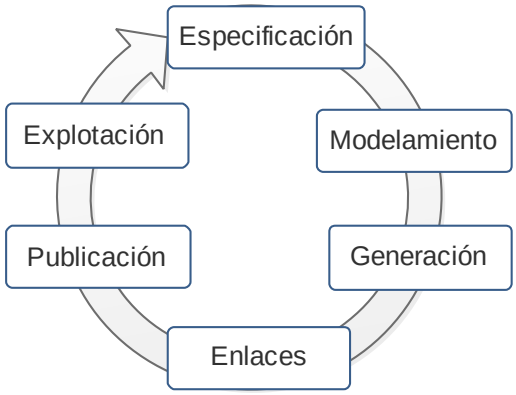


***Figura 2**: Ciclo de vida de los datos enlazados*
Fuente: Villazón-Terrazas, B et al. (2012). Main Activities for Publishing Linked Data [Figura]. En Publishing Linked Data - There is no One-Size-Fits-All Formula (p. 2)

Como se puede visualizar en la Figura 2, este método consiste de seis fases que se describen a continuación:

a) Especificación: Para analizar y seleccionar las fuentes de datos.
b) Modelamiento: Para desarrollar el modelo que representa el dominio de información de las fuentes de datos**.**
c) Generación: Para transformar las fuentes de datos en RDFs.
d) Enlaces: Para crear enlaces entre los recursos RDF desarrollados con otros recursos externos.
e) Publicación: Para publicar los recursos RDF y los enlaces generados.
f) Explotación: Para consumir los datos que serán explotados a través de aplicaciones construidas para esta finalidad.

2.6.3. Almacenamiento de datos en la Web Semántica

El almacenamiento de datos en la Web Semántica se lo realiza en forma de tripletas de tipo sujeto-predicado-objeto (Figura 3). El sujeto y el objeto son los nodos de la tripleta, por otro

lado, el predicado es la conexión dirigida que expresa la relación entre los dos nodos, esta dirección siempre apunta hacia el objeto (Castelló, 2006). .

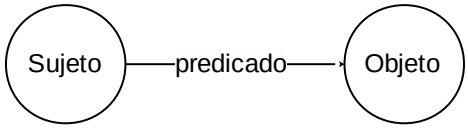


***Figura 3**: Representación de una tripleta*

Las tripletas se almacenan en una base de datos llamadas: "triplestore". Algunas de estas han sido construidas basadas en tecnologías de grafos, es decir no se han basado en ningún otro motor de bases de datos relacionales existentes. Como ejemplo de estas triplestore se puede citar: Jena[1], Sesame[2], entre otras. Por otro lado, también están las triplestore que para su construcción han usado como base los motores de bases de datos relacionales existentes, con el fin de beneficiarse de sus funcionalidades; como ejemplo de estas están: 3store[3], Oracle[4], Virtuoso[5], entre otras más. Finalmente cabe indicar que las bases de datos triplestore deben ser consultadas mediante un lenguaje de consultas semántico (Priyatna, 2015).

2.7. Lenguajes de consulta para RDF

Según Llamas (2006), los documentos RDF pueden ser consultados en tres niveles: sintáctico, estructural y semántico. La explotación a nivel sintáctico se refiere a las consultas sobre la sintaxis del documento. A nivel estructural los documentos son el conjunto de tripletas, las consultas explotan el modelo de datos RDF es decir atacan a los grafos. Por último, sabiendo que el conocimiento se encuentra almacenado en las tripletas RDF, se necesita entonces un

[1] https://jena.apache.org/
[2] http://rdf4j.org/
[3] http://threestore.sourceforge.net/
[4] http://www.oracle.com/technetwork/database-options/spatialandgraph/overview/rdfsemantic-graph-1902016.html
[5] https://virtuoso.openlinksw.com/

lenguaje de consultas que permita realizar la explotación a nivel semántico. Ejemplos de estos lenguajes son RQL, RDQL, SPARQL, entre otros.

SPARQL

SPARQL (W3C - SPARQL), es un lenguaje estandarizado para realizar consultas sobre diversas fuentes de datos almacenadas en formato RDF que permite a los desarrolladores y usuarios finales la manera de escribir y obtener resultados de búsquedas realizadas sobre una información extensa y variada. Utiliza un protocolo común que permite que las aplicaciones puedan acceder y combinar información de la Web.

2.8. Buscadores

Un buscador es un software diseñado para realizar consultas de información digital de diferentes formatos como, por ejemplo: páginas web, documentos de texto, imágenes, videos, etc. almacenada en distintos servidores en la red. Los buscadores presentan a los usuarios los resultados de las consultas a través de enlaces que se vinculan con la información solicitada (quees.info).

A continuación, se listan algunos tipos de buscadores conocidos hoy en día (Moreno & Sánchez, 2012):

a) **Buscadores por palabras clave:** El usuario ingresa en el buscador una palabra clave y este examina la información hasta encontrar datos que coincidan para dar el resultado al usuario.
b) **Buscadores por categorías:** La información está organizada por categorías temáticas organizadas jerárquicamente, de manera que la información puede ser visualizada desde los temas más generales hasta llegar a los temas más específicos.
c) **Metabuscadores:** Los usuarios realizan su consulta y los metabuscadores consultan en otros buscadores y la respuesta es lanzada de otro buscador que encontró la respuesta a la consulta.

d) Buscadores específicos: Son aquellos que dan información sobre temas específicos y darán mejores resultados que los buscadores genéricos.

e) Buscadores semánticos: Son aquellos buscadores que utilizan las ontologías y agentes inteligentes para inferir el conocimiento y presentar los resultados al usuario sin que tenga que analizarlos para obtener la información más adecuada como respuesta a su búsqueda.

Capítulo 3
TRABAJOS RELACIONADOS

En la WEB existe gran cantidad de información sobre la Web semántica y las ontologías vistas como una herramienta útil para integrar información registrada en diferentes fuentes de datos. La Web Semántica permite superar las limitaciones de la Web actual, puesto que permite que la información sea interpretada tanto por las personas como por las computadoras, facilitando así la obtención del conocimiento. El uso de tecnología semántica se ha hecho presente en distintos dominios como, por ejemplo: el área financiera, la educación, gestión documental, entre otras. A continuación, se describen algunos trabajos que hacen uso de la Web Semántica:

Un dominio en la que se ha hecho uso de la tecnología semántica es en la gestión de referencias bibliográficas (Galey, 2010). El autor de este trabajo en su proyecto de fin de carrera: "Aplicación web semántica para la gestión de referencias bibliográficas", desarrolla un gestor bibliográfico de publicaciones científicas, utilizando tecnologías de la web semántica. El autor de este trabajo afirma que el gestor facilita el tratamiento de referencias y permite transformar los formatos de las publicaciones. Un resultado más de este trabajo ha sido una ontología que describe publicaciones científicas.

Las artes plásticas es otro dominio en el que se ha visto que se ha aplicado la tecnología semántica (Guzmán, López, & Durley, 2012) . En el trabajo llamado "Metodologías y métodos para la construcción de ontologías" los autores hablan sobre la necesidad de utilizar una metodología para el diseño e implementación de una ontología, por esta razón realiza una comparación de diferentes metodologías y métodos para el diseño e implementación de las ontologías; e indica que aunque las metodologías tienen características comunes, éstas también se diferencian por la naturaleza de su aplicación y la selección de la metodología adecuada es muy subjetiva, pues depende del alcance de su aplicación. Finalmente, en este trabajo sus

autores seleccionan una de las metodologías estudiadas y la aplican en un caso de uso en el dominio de artes plásticas.

La Web Semántica también ha sido aplicada en contextos gubernamentales. En el trabajo llamado "Methodological Guidelines for Publishing Government Linked Data" (Villazón-Terrazas, Vilches, Corcho, & Gómez, 2011), sus autores indican que para realizar la integración de datos y publicarlos se requiere de varios pasos, por esta razón proponen un conjunto de lineamientos metodológicos para realizar las actividades que permitan llevar a cabo todo el proceso de datos enlazados y lo aplican en varios contextos gubernamentales como por ejemplo GeoLinkedData y AEMETLinkedData.

El dominio de los sistemas de información geográfica también ha incursionado en el mundo de la tecnología semántica, es así como en la tesis doctoral "Modelo de integración de datos, metadatos y conocimientos geográficos", su autor elabora un prototipo del modelo de integración de las abstracciones geográficas MIGEO, concluyendo que el prototipo demuestra que es viable y puede ser aplicado a las infraestructuras de datos espaciales, con el propósito de tomar decisiones basadas en la semántica asociada a las abstracciones geográficas y sus vínculos (Oliva, 2013).

Con respecto al dominio de la educación, se ha visto que es bastante beneficiado con los avances tecnológicos. Según: Arroyo, Castro, & Rosario (2008) en su trabajo llamado "La Educación y la Web Semántica" proponen que el paradigma de la Web Semántica ofrece al área educativa y en particular a la educación a distancia un gran apoyo, puesto que permite que la Web no sea utilizada únicamente como un repositorio sino que se pueda tratar la información para obtener conocimiento. Los autores de este trabajo también proponen que se debe continuar con el estudio de la tecnología semántica de manera que se puedan desarrollar sistemas de colaboración que permite realizar búsquedas y manejar de mejor manera el contenido para lograr mejores ambientes de aprendizaje.

Un trabajo más relacionado con la educación se llama: "Generación de datos semánticos a partir de una base de datos relacional de una institución de educación superior" (Tapia & Fuertes, 2014), en este sus autores explican cómo se utilizó la tecnología semántica para generar una ontología a partir de los datos de una universidad pública que están almacenados en bases de datos relacionales y también cuentan que los resultados de este trabajo fueron publicados exitosamente en la Web. Otro ejemplo es el artículo "Diseño de una ontología para la gestión de datos heterogéneos en universidades: marco metodológico", en este trabajo los autores proponen la creación de una ontología para el manejo de datos registrados en diferentes fuentes de la Universidad de la Habana, a través de una metodología para construir un sistema que permite la anotación semántica para realizar búsquedas federadas y semánticas. Además, hablan de las ventajas que brinda a la gestión de la información institucional y de cómo a través de esta aplicación, la Institución puede llegar a ser visible en las redes (Rosell, Senso, & Leiva, 2016).

Otro trabajo que habla sobre la aplicación tecnología semántica en la educación se llama "Modelo ontológico para la representación de datos académicos y su publicación con tecnología semántica" (Mora & Segarra, 2016). En este trabajo sus autores explican cómo se construye un modelo ontológico de información académica (planes de curso) aplicando la metodología NeOn para la creación de una ontología y su publicación a través del ciclo de vida de los datos enlazado. En este trabajo se realiza la descripción de los datos con un significado claro y preciso, para su posterior publicación con el objetivo de compartir conocimiento y vincular con datos de otras instituciones.

Como se ha visto la web semántica puede ser aplicada en diferentes dominios y existen varias aplicaciones en la educación, por lo se concluye que es completamente viable la construcción de la ontología para los datos académicos de la Universidad de Cuenca y también la construcción del prototipo para el buscador semántico de la Institución. Luego de analizar los trabajos relacionados, para el desarrollo de este proyecto, la generación, publicación y

explotación de los datos enlazados se realizó siguiendo los seis pasos: especificación de requerimientos, modelamiento, generación, establecer enlaces, publicación y explotación propuestos en "Publishing Linked Data - There is no One-Size-Fits-All Formula" (Villazón-Terrazas, y otros, 2012). Por otro lado, para la construcción de la ontología se siguió la metodología NeOn, debido a que permite la combinación de escenarios y no es rígida, puesto que sugiere una gran variedad de opciones para construir la ontología (Ramos, Barrera, & Núñez, 2012). Así también cuando se realiza el modelamiento de una ontología es recomendable reutilizar, tanto como sea posible, las ontologías y vocabularios existentes, puesto que esta reutilización aumentará la probabilidad de que los datos sean consumidos por aplicaciones que pertenezcan al mismo dominio (Heath & Bizer, 2011).

Capítulo 4
PROCESO DE INTEGRACIÓN DE DATOS PARA LOS SISTEMAS DE INFORMACIÓN DE LA UNIVERSIDAD DE CUENCA

En este capítulo se explica cómo fue realizado el proceso de integración de datos de la Universidad de Cuenca. Cabe indicar que como se anotó en el alcance del presente trabajo, a modo de ejemplo de caso de uso, se construyó una ontología base que consiste en una ontología organizacional que cubre toda la estructura de la Universidad. Esta ontología está preparada para luego integrar los datos con los datos de otros dominios. Al momento a esta ontología organizacional se ha integrado los datos del dominio académico de carreras, puesto que uno de los ejes principales de la Universidad de Cuenca es el eje académico y es por esta razón que a partir de septiembre del 2009 se inició la operación del sistema académico y hoy en día es el sistema más robusto de todos los sistemas que se operan en la Universidad.

4.1. Análisis de las alternativas para la integración de los datos: Virtual vs. Materializado

Hoy en día, la Universidad de Cuenca opera grandes cantidades de datos provenientes de los distintos sistemas informáticos de la Institución. Cabe indicar que existen diferentes tipos de fuentes de datos, así como también diferentes motores de base de datos, haciendo más difícil el acceso de forma rápida y confiable. Actualmente, para obtener los datos de los sistemas, se realizan consultas específicas a las diferentes bases, posteriormente se consolida los datos de forma manual y de esta forma se obtienen la información requerida. Por lo antes expuesto, es evidente que la información no puede ser obtenida desde una única fuente debido a la heterogeneidad y dispersión de los datos. Para dar solución a este inconveniente, es necesario homogeneizar e integrar los datos con el propósito de explotarlos fácilmente y de forma transparente para el usuario. Según lo expuesto en el marco teórico en la sección 2.1, en donde se habla sobre las formas de integración de los datos, y en vista de que en la Universidad de Cuenca tiene control total de sus fuentes de datos que serán consultadas, en el presente trabajo se opta por un modelo centralizado para almacenar la información y atacar directamente a la

información almacenada. Con esto se logra mejorar la disponibilidad, orden y aseguramiento de la información.

4.2. Proceso de integración de datos

En el campo de la integración y publicación de datos de datos, se han propuesto algunas alternativas, una de ellas es la integración de datos utilizando tecnología semántica. Esta tecnología ha logrado grandes alcances y ha tomado fuerza durante los últimos años, dando lugar a varios proyectos relacionados con Linked Open Data. La información integrada y unificada ha sido un reto para varios sectores, y como se vio en el capítulo 3, el sector de la educación también ha puesto gran énfasis en lograr la integración de sus datos para mejores resultados en su quehacer diario. Por esta razón a través de este trabajo se permitió que la Universidad de Cuenca integre y publique su información utilizando Linked Open Data, y las tecnologías que aportan de semántica a los datos y por ende dan mayor significado a las búsquedas para obtener información concreta, siendo éstas menos complicadas y mucho más exactas. Tener la información integrada, abierta y disponible es de gran ayuda para una empresa, puesto que los datos integrados permiten brindar resultados óptimos el momento de proporcionar respuestas a las consultas realizadas por los usuarios.

En la Tabla 1., se puede observar los dominios sobre los cuales la Universidad de Cuenca puede aplicar el proceso de integración de datos enlazados. Sin embargo, debido al alcance de este trabajo, la publicación de datos enlazados está orientada a la estructura organizacional de la Universidad de Cuenca y a los datos académicos y curriculares de sus carreras. A partir de este dominio se construirá una ontología base, para que en un futuro se puedan implementar las ontologías pertenecientes a los demás dominios identificados y por tanto lograr la publicación de los datos enlazados de toda la Universidad.

***Tabla 1**: Dominios analizados*

Dominio	Fuentes de Datos	Formato	Descripción

Académico	SGA,SGAP	Oracle 11 g, DB2	La ontología debe contestar preguntas con respecto a las carreras ofertadas por la Universidad de Cuenca, así como también los estudiantes matriculados y sus calificaciones en las asignaturas y los docentes que dictan las materias
Estructura Organizacional	Estatuto de la Universidad	PDF	Contesta preguntas con respecto a toda la estructura organizacional y quienes son las autoridades
Académico de Posgrados	SGAP	Oracle 11 g	La ontología debe contestar preguntas sobre los programas de posgrados ofertados por la Universidad de Cuenca, así como también los estudiantes matriculados y sus calificaciones en las asignaturas o módulos
Evaluación del desempeño del docente	SGE	Oracle 11 g	La ontología debe responder a consultas sobre el modelo de evaluación del docente, como por ejemplo: funciones y ámbitos que se evalúan y debe también responder los resultados obtenidos por los docentes
Investigación	SGI	Oracle 11 g	Se debe responder a búsquedas sobre el proceso de investigación que se desarrolla en la Universidad, incluyendo la producción científica de los docentes
Financiero	ERP	Postgres	Debe responder a preguntas de la situación financiera de la Universidad
Proyectos	Sistema de proyectos Redmine	MySql	Esta ontología debe responder a las preguntas de los proyectos que se desarrollan en la Universidad, y dar a conocer el estado de cada uno de ellos, sus integrantes, director, instituciones involucradas, entre otras consultas.
Talento Humano	SGP	DB2	A través de esta ontología se responderán a preguntas sobre todo el personal que labora en la Institución y de las condiciones en las que laboran como por ejemplo su tipo de relación, el estado, tipo de servidor, etc.

Como se vio anteriormente, hasta el momento se han desarrollado algunas guías para la publicación de datos enlazados, sin embargo para realizar la integración y publicación de las fuentes de datos de la Universidad de Cuenca se utilizó como base la metodología propuesta por Villazón – Terrazas (2012). Esta metodología fue seleccionada para ser aplicada en este trabajo, puesto que en comparación con otras metodologías presenta un método general organizado que consiste en 6 fases bien definidas que se debe seguir para llegar a la publicación de los datos enlazados.

Para realizar la integración de los datos de los sistemas de información de la Universidad de Cuenca, en este trabajo antes de la fase de explotación, se incluyó una fase más para realizar

la validación de los datos generados (Figura 4), esta fase se incluyó con el objeto de asegurar que la explotación que se realiza a través de un buscador semántico tenga una mayor precisión en los resultados que el usuario busca.

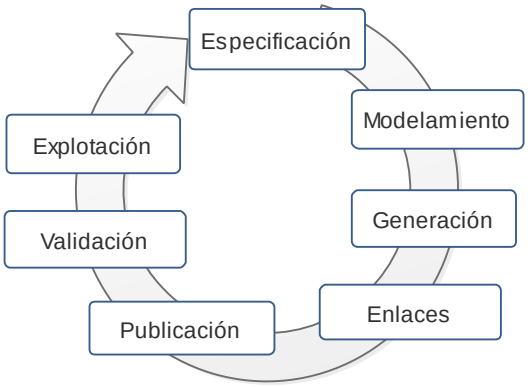


***Figura 4**: Ciclo de vida para integrar los datos de la Universidad de Cuenca*

Fuente: Elaboración propia basada en Villazón-Terrazas, B et al. (2012). Main Activities for Publishing Linked Data [Figura]. En Publishing Linked Data - There is no One-Size-Fits-All Formula (p. 2)

En este trabajo también se desarrolló una arquitectura que describen las actividades que se deben realizar en cada fase del ciclo de vida de la publicación de los datos enlazados. Esta arquitectura se puede visualizar en la Figura 5, en donde se puede ver los sistemas que se operan en la Universidad, así como también los gestores que almacenan los datos generados por cada uno de los sistemas. Además, se puede también observar la arquitectura de software tecnológica implementada, es decir se puede ver las herramientas que se utilizaron en cada una de las fases de la arquitectura. A continuación, se explica cada uno de los componentes de la arquitectura definida.

4.2.1. Especificación

Como en cualquier tipo de proyecto de tecnología en este trabajo también se consideró la etapa de especificación como la más crítica, puesto que el éxito de los proyectos depende de un buen análisis para que luego no se presenten retrasos en el avance de cada etapa. Para este trabajo la fase de especificación consiste en el análisis de los datos que reposan en las distintas

fuentes de datos operadas en la Universidad de Cuenca. El análisis permite examinar y seleccionar los datos que serán necesarios para obtener información con respecto al dominio seleccionado. Para este presente trabajo el dominio definido fue la estructura organizacional y el área académica de las carreras de la Universidad.

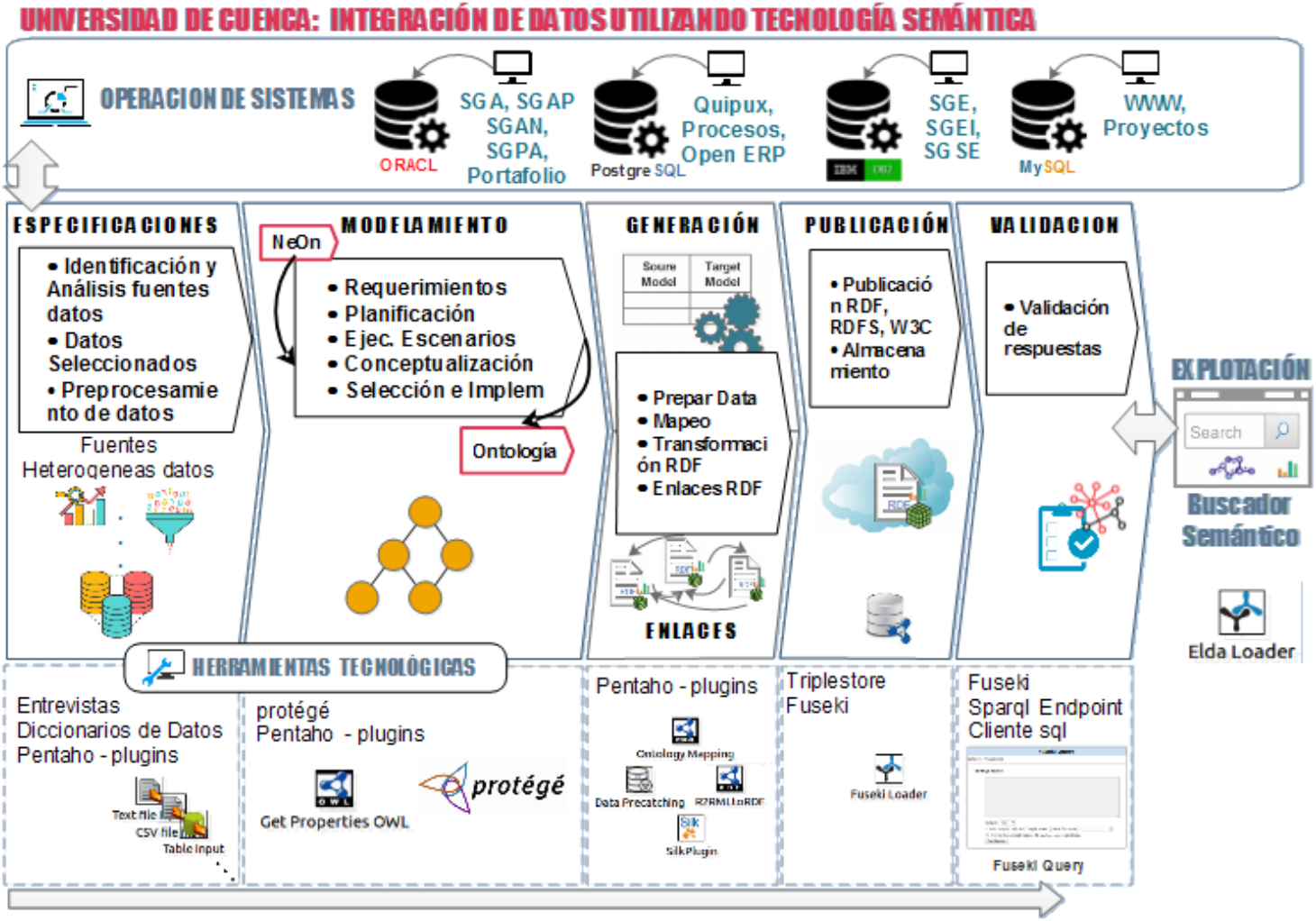


***Figura 5**:Arquitectura para la integración de datos en la Universidad de Cuenca*

La fuente de datos utilizada para obtener la información sobre la estructura organizacional, fue el documento del estatuto de la Universidad de Cuenca, aprobado en el año 2013; en este documento se establece la normativa y el funcionamiento de la Institución. Por otro lado, la información para el modelamiento del área académica se obtuvo de la documentación del actual sistema de gestión académica (SGA) y de los datos que se almacena en la base de datos Oracle. La Figura 6 presenta gráficamente una breve ilustración de las fuentes de datos analizadas, en esta se puede ver que en lo que respecta a la estructura organizacional se pudo analizar de un archivo PDF y de la página Web institucional. Por otro lado, la información académica de carrera se pudo analizar de los archivos de documentación del proyecto del Sistema de Gestión Académica y del esquema de la base de datos que se encuentra en el gestor Oracle 11g.

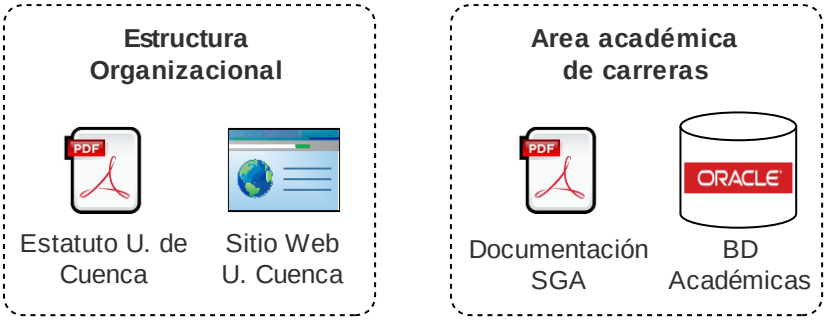


***Figura 6**:Análisis de las fuentes y tipos de datos*

El organigrama de la Universidad se puede visualizar en la Figura 7, y en la Figura 8 se puede ver el modelo conceptual del sistema de gestión académica, que se generó luego de analizar la documentación del SGA, el modelo entidad-relación y el diccionario de datos del esquema académico de Oracle 11g. A partir de este modelo y de algunas entrevistas realizadas a autoridades académicas, se realizó el análisis de los requerimientos del área académica de carreras de la Universidad.

Art. 7 de la Ley Orgánica de Transparencia y Acceso a la Información Pública - LOTAIP

Literal a1) Estructura orgánica funcional

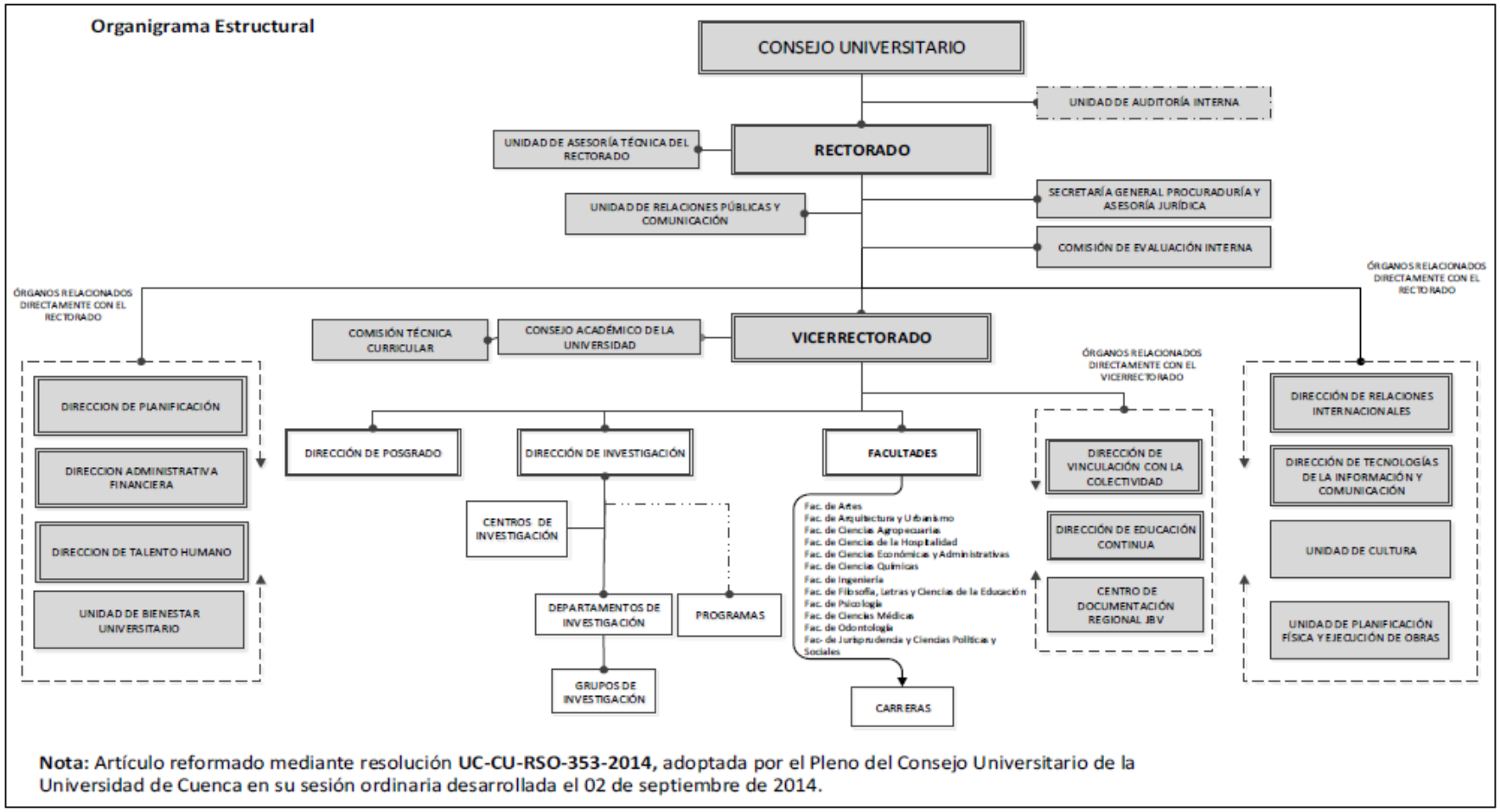


Figura 7: Organigrama de la Universidad de Cuenca
***Fuente**: Organigrama Estructural. [Figura]. Recuperado de https://ucuenca.edu.ec*

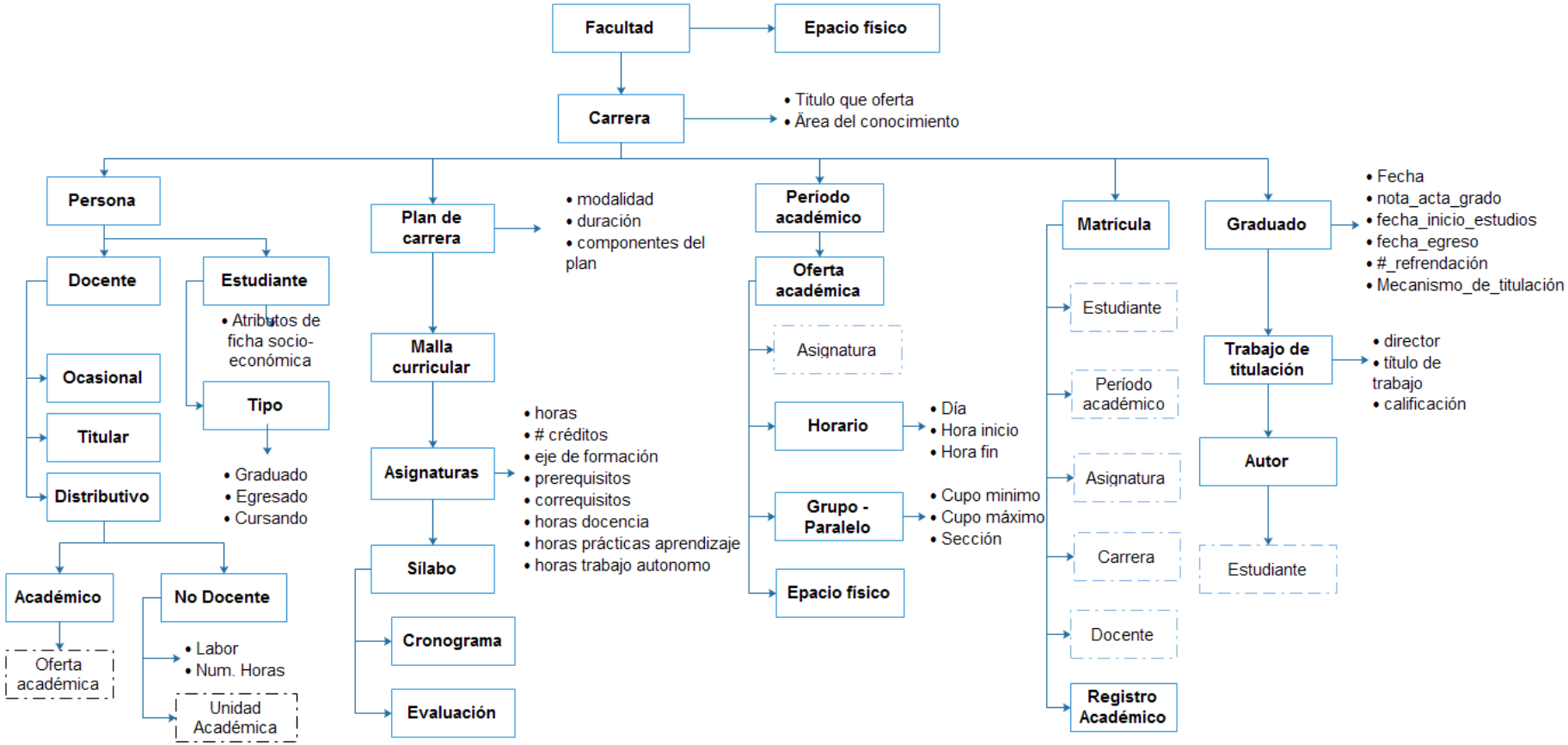


Figura 8: Especificación de requerimientos del SGA

Para Villazón-Terrazas (2011), la identificación precisa y eficiente de los requisitos para la vinculación de datos de un dominio es de gran importancia, por esta razón propone algunos lineamientos a seguir durante la especificación:

i. Identificación y análisis de las fuentes:

En esta tarea se definieron los datos del área académica y la estructura organizacional de la Universidad que fueron publicados, cuyo detalle se lo puede ver en la Tabla 2, en la que se explica cuáles y de qué tipo son las fuentes de datos. Por ejemplo, se puede ver que los datos correspondientes a las carreras, estudiantes, graduados, etc. se obtuvieron de una base de datos relacional que está en el gestor de bases de datos Oracle.

***Tabla 2**: Detalle de datos publicados*

Fuente de Datos	Datos publicados	Tipo
Estatuto Universidad de Cuenca	Organización Administrativa Organización Académica	PDF Página Web
Base de Datos Relacional	Facultades Espacios Físicos Títulos Ofertados Personas Dependencias	ORACLE (ADMINUC)
Base de Datos Relacional	Carreras Áreas de Conocimiento Estudiantes Graduados Egresados Docentes Planes de Carreras Mallas de Carreras Asignaturas Oferta Académica (Grupos, docentes, horarios, cupos) Matrículas Registros Académicos Períodos Académicos	ORACLE (ACADEMICO)
Base de Datos Relacional	Trabajos de Titulación Tribunal	ORACLE (PRERREQUISITOS)

Base de Datos Relacional	Sílabos Cronograma	ORACLE (SILABOS)

ii. Diseño de URIs

Linked data permite construir una web de datos enlazadas, como se vio en la Figura 1, la base de la Web Semántica y de los datos enlazados es la tecnología de la web actual, por lo que son accedidos a través del protocolo HTTP. Por otro lado Linked Data define 4 principios básicos, uno de los cuales es utilizar URIs simples, estables y manejables, para identificar y nombrar de forma única los recursos (Berners-Lee, 1998). La actividad de especificación cubre la tarea sobre la definición de URIs, que son utilizadas en los grafos RDF. Para la definición de la URIs se utilizaron las pautas sugeridas por Villazón-Terrazas (2011), una de las cuales indica que se deben definir URIs significativas, incluyendo la mayor cantidad de información de manera que sea de fácil entendimiento. Como ejemplo de URIs se nombran las siguientes:

http://ucuenca.edu.ec/resource/{resourcetype}/{resourcename}
http://ucuenca.edu.ec/resource/estudiante/0102492972
http://ucuenca.edu.ec/resource/facultad/Filosofia

4.2.2. Modelamiento

El siguiente paso del ciclo de vida de los datos enlazados es el modelamiento, esta fase modela la ontología que permite compartir los datos que son procesados por humanos y por computadoras. El buen análisis de las fuentes de datos facilita la construcción de la ontología que da como resultado el vocabulario que permite la conversión de los datos en las tripletas RDF. Cuando se realiza el modelamiento de la ontología, es recomendable reutilizar tanto como sea posible las ontologías y vocabularios existentes, puesto que esta reutilización aumenta la probabilidad de que los datos sean consumidos por aplicaciones que pertenezcan al mismo dominio y acelera el desarrollo de la ontología ahorrando tiempo, recursos y esfuerzo (Heath & Bizer, 2011). Para realizar la construcción de la ontología se aplicó la metodología NeOn,

puesto que de acuerdo a lo descrito en el marco teórico, es una metodología flexible que sugiere varias alternativas para el desarrollo de ontologías. A continuación, se explican brevemente las principales actividades realizadas para la creación del modelo ontológico, en base a la metodología de NeOn.

i. Especificación de requerimientos ontológicos:

Para realizar esta actividad, la metodología de NeOn propone la creación de documento de especificación de requerimientos ontológicos (DERO), Este documento permite identificar el conocimiento almacenado en la ontología, facilita la reutilización de recursos, así como también la validación de resultados. Además, este documento permite principalmente conocer los usos de la ontología y los usuarios a quienes está dirigida. En la Tabla 3 se puede ver un extracto de este documento perteneciente a la ontología creada, en este se puede conocer el propósito, el alcance, el lenguaje utilizado para la construcción de la ontología, entre otros.

Tabla 3: *Ejemplo del documento de especificaciones*

Documento de especificación de requerimientos ontológicos	
1.	**Propósito**
El propósito de construir una ontología universitaria es proveer un modelo genérico de conocimiento que permita describir y presentar conceptos y relaciones existentes en el dominio universitario. El modelo debe ser capaz de responder a consultas referentes a la estructura organizacional y académica de la Universidad de Cuenca. Además, el modelo obtenido debe almacenar la información necesaria para ser utilizado posteriormente en un sistema de búsqueda semántica.	
2.	**Alcance**
La ontología plantea modelar la estructura organizacional y académica de la Universidad de Cuenca, para lo cual es necesario el almacenamiento de información de diversa índole dentro del contexto universitario. Esta información es planteada en base a: • Estructura organizacional: Organigrama organizacional aprobado en el estatuto de la Universidad • Estructura académica: Organigrama planteado luego de realizar el estudio de los sistemas y fuentes de datos utilizados en la Universidad	
3.	**Lenguaje de implementación**
El lenguaje utilizado para la construcción de la ontología es OWL/XML	
4.	**Usuarios finales previstos**
1. Autoridades 2. Personal docentes y administrativo 3. Estudiantes	

5.	**Usos previstos**
	1. Consultar información relacionada a la estructura organizacional 2. Consultar información relacionada a la estructura académica 3. Permitir la conexión con otras ontologías

La especificación de los requerimientos funcionales se realizó a través de preguntas de competencia, este método consiste en especificar un conjunto de preguntas que la ontología debe contestar para satisfacción del usuario. Para este trabajo las preguntas planteadas se basan en el contexto del modelo organizacional y académico de la Universidad.

Se plantearon 121 preguntas clasificadas en cinco categorías: Estructura organizacional, Estructura académica, Personas, Espacios físicos, Recursos bibliográficos, además se definió el pre-glosario de términos, que es el resultado de estas preguntas y se refiere a la frecuencia de aparición de un término tanto en las preguntas como en las respuestas. Cabe indicar que todo se registra en el DERO de la ontología. (DTIC, 2017). Los elementos que contiene la ontología son definidos de acuerdo a la terminología que se extrae de un pre-glosario de términos que se establece a partir de las preguntas de competencia.

A continuación, se nombran algunos ejemplos de las preguntas de competencia que están planteadas en el DERO:

- ¿Qué organismo es la máxima autoridad de la Universidad de Cuenca?
- ¿Cuántas facultades tiene la Universidad de Cuenca?
- ¿Cuál es el período académico vigente en la carrera X?
- ¿Cuál es la capacidad del aula de clases X?
- ¿Qué título de bachiller tiene el estudiante X?
- ¿Cuál es la calificación del trabajo de titulación X?
- ¿El estudiante tiene algún tipo de discapacidad X?
- ¿Cuántos miembros del grupo familiar del estudiante X estudian en niveles diferentes al superior?

ii. Modelamiento de la ontología:

Como se mencionó anteriormente, la metodología de NeOn propone nueve escenarios que pueden ser utilizados para el modelamiento de la ontología. Una vez elaborado el documento DERO, la siguiente actividad fue establecer el escenario que se adaptó al dominio de estudio de este trabajo que se refiere al modelamiento de la estructura organizacional y académica de la Universidad de Cuenca. Se seleccionó el Escenario 6: "La reutilización, la fusión y re-ingeniería de recursos ontológicos". Puesto que existen recursos ontológicos en la WEB que pertenecen al mismo dominio que este estudio, y que fueron modificados para que sean de utilidad para la nueva ontología.

iii. Búsqueda de recursos ontológicos:

En base a las preguntas establecidas y al pre-glosario de términos definidos en la especificación de requerimientos ontológicos, se realizó la búsqueda de recursos ontológicos pertenecientes al dominio de este trabajo. Para realizar la búsqueda de los recursos que cumplen con los requerimientos identificados se utilizaron algunas herramientas online como por ejemplo Swoogle[6], Watson[7], y Sindice[8]. Además se realizaron búsquedas manuales en librerías de ontologías como LOV[9] (**L**inked **O**pen **V**ocabularies). El resultado de esta actividad fue un conjunto de ontologías candidatas para cada una de las categorías planteadas, tal y como se puede ver en la Tabla 4.

[6] http://swoogle.umbc.edu/2006/
[7] http://watson.kmi.open.ac.uk/
[8] http://sindice.com
[9] http://lov.okfn.org/dataset/lov/

Tabla 4: *Ontologías candidatas según la categoría*

Categoría definida	Ontologías candidatas
Estructura organizacional	Univ-Bench Academic Institution Internal Structure Ontology (AIISO) Core Organization Ontology (ORG) Higher Education Reference Ontology (HERO)
Estructura académica	Learning Object Metadata Ontology (LOM) Teaching Core Vocabulary Specification (TEACH) Curriculum Ontology HERO
Personas	FOAF VCARD
Espacios físicos	ROOMS COBRA-ONT Space Ontology WGS84 Basic Geo
Recursos bibliográficos	VIVO-ISTF Bibliographic Ontology (BIBO) Semantic Web for Research Communities (SWRC)

iv. Comparación de ontologías:

El siguiente paso dentro de esta fase fue realizar una comparación entre las ontologías candidatas para seleccionar la que mejor se adapte a las necesidades del dominio de este trabajo. Para realizar esta comparación se definieron cuatro criterios que son recomendados por: Chimbo, Contreras, & Espinoza (2017). Esto se puede visualizar en la Tabla 5.

Tabla 5: *Criterios para seleccionar ontologías para reutilizar*

Criterio	Actividad analizada
Propósito similar	Se revisó que el propósito de la ontología candidata concuerde con el propósito de la nueva ontología
Alcance similar	Se verificó que el alcance general de la ontología existente o el grupo al cual pertenece sea similar al de la nueva ontología
Cobertura de requisitos no funcionales	Se definió que el idioma será el único requisito no funcional a ser considerado
Cobertura de requisitos no funcionales	Se comparó el nivel de coincidencia igual o semejante de los términos de la ontología

Una vez definidos los criterios de comparación, las ontologías fueron calificados con valores cualitativos (Chimbo, Contreras, & Espinoza, 2017), a continuación se puede ver el resultado de la calificación en la Tabla 6.

Tabla 6: *Calificaciones de las ontologías candidatas*

Criterio	Calificación
Si-Totalmente (Si-T)	La ontología cumple de manera exacta con el criterio calificado
Si-Parcialmente (Si-P)	La ontología candidata cumple de manera parcial con el criterio calificado.
No (N)	La ontología candidata no cumple con el criterio calificado
Desconocido (D)	La ontología candidata no proporciona documentación suficiente para determinar si es válida o no para ser reutilizada dentro del criterio calificado

v. Selección de ontologías e implementación:

Para facilitar el proceso de selección de ontologías a ser reutilizadas, se analizaron las ontologías candidatas siguiendo las guías propuestas en la metodología de NeOn, que consisten en dar una puntuación a cada una de ellas, en base a ciertas características no funcionales como son: costo de reutilización, esfuerzo de comprensión, esfuerzo de integración y fiabilidad. Luego de realizar este análisis se seleccionaron las ontologías que cumplieron con los criterios establecidos. En la Tabla 7, se nombran las ontologías seleccionadas según las categorías definidas en el DERO.

Tabla 7: *Ontologías seleccionadas según la categoría*

Categoría definida	Ontologías seleccionadas
Estructura organizacional	Univ-Bench
Estructura académica	HERO
Personas	FOAF
Espacios físicos	ROOMS WGS84 Basic Geo (para la latitud y longitud)
Recursos bibliográficos	Bibliographic Ontology (BIBO)

Luego de analizar los recursos y definir las ontologías reutilizables se modeló la red ontológica para el caso de estudio del presente trabajo. En la Figura 9 se describen las relaciones existentes en la red ontológica. En primer lugar, está la ontología base denominada ***Ontología Organizacional*** facilita la integración con diferentes dominios ontológicos. La **estructura organizacional** se vinculó con el dominio ***académico***, mediante las unidades académicas que son parte de la estructura organizacional de la Universidad y que ofertan las carreras, es decir: la universidad oferta carreras y las carreras son ofertadas por las unidades académicas. Por otro lado, la universidad es gestionada por ***Personas*** y las Personas gestionan la Universidad. A las carreras se vinculan estudiantes o docentes que son personas y los estudiantes o docentes cursan o dictan clases en los cursos ofertados. Los cursos se ofertan en ***espacios físicos***. Los cursos se dictan en ciertos ***periodos de tiempo***. Los cursos generan ***recursos bibliográficos***, como por ejemplo trabajos de titulación, publicaciones, etc. Cabe indicar que siendo esta la ontología base, ésta podrá ser utilizada para realizar la publicación de los datos enlazados de los demás dominios descritos en la Tabla 1 de la sección 4.2.

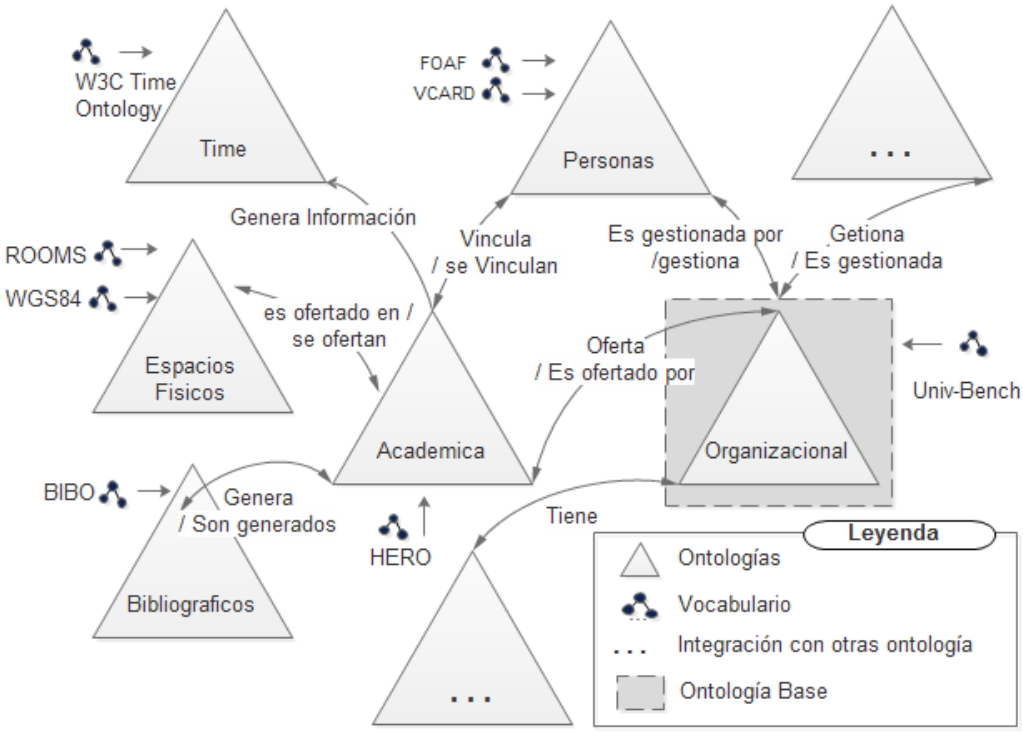


Figura 9*: Mapa conceptual de alto nivel de la red odontológica de la Universidad de Cuenca*

En esta etapa se implementó la ontología del dominio de estudio en el lenguaje formal OWL, a través del editor de ontologías Protégé[10]. Para la implementación se realizaron actividades como: la generación de nuevos conceptos, la poda de ontologías seleccionadas, adición de anotaciones que describen la ontología, utilización de vocabularios y términos específicos, y la generación de patrones de diseño ontológicos. En la Figura 10 se puede apreciar una vista de la ontología vista generada con Protégé.

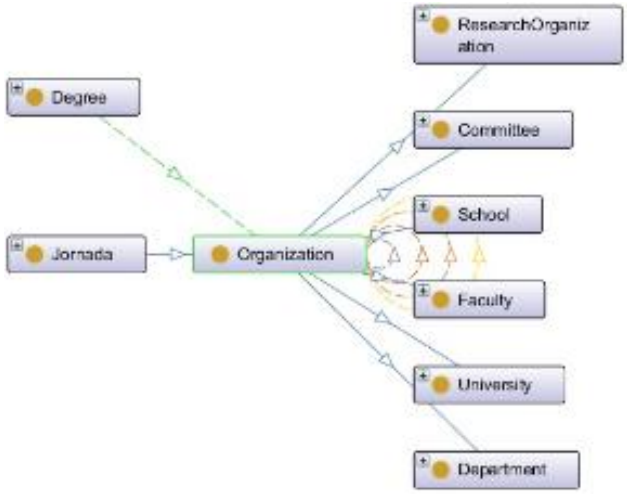


***Figura 10**: Vista de la ontología generada con la herramienta Protégé*

4.2.3. Generación

El siguiente paso del ciclo de vida de los datos enlazados, es la conversión de los datos a formato RDF. Para realizar esta transformación se utiliza la ontología creada en la etapa de modelamiento, y se aplica a las fuentes de datos seleccionadas, dando como resultado un conjunto de tripletas RDF. Durante esta etapa se realizaron las siguientes tareas:

[10] https://protege.stanford.edu/

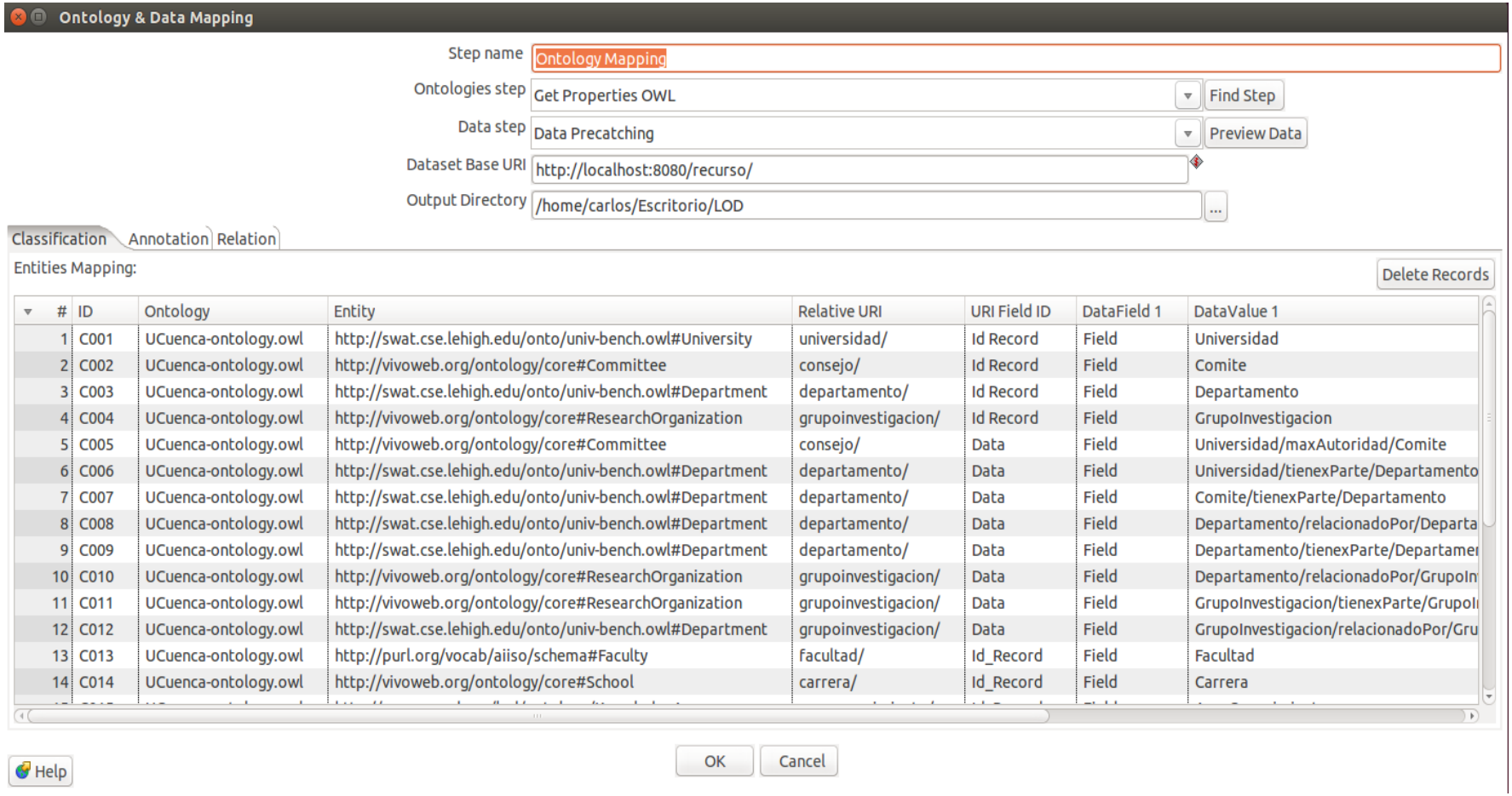
i. Preparación de datos.

Esta actividad consiste en el descubrimiento y corrección o eliminación de datos erróneos, asegurando la calidad de los datos, para su posterior utilización. Por lo que para la generación de los grafos RDF en el presente trabajo se realizó primero un proceso de preparación de los datos de manera que estos se encuentren consistentes y libre de errores. Posteriormente se procedió con la limpieza de los datos referentes a la estructura organizacional y las carreras ofertadas por la Universidad. Con respecto a la estructura de la Universidad se vio que algunas de las dependencias están duplicadas, por esta razón se realizó un proceso para eliminar inconsistencias y redundancias para recuperar las dependencias sin duplicidades y así evitar futuros errores en la ontología. Por otro lado, en la revisión de la información correspondiente a las carreras, se pudo notar que en las bases de datos se almacena información sobre carreras auxiliares utilizadas por el SGA (Sistema de Gestión Académica) para cumplir con ciertas funcionalidades. Se seleccionaron entonces únicamente las carreras válidas. Se realiza también un mapeo en caso de que una carrera tenga más de un registro con el mismo nombre debido a historiales de carreras. Estos procesos y otros más se realizaron con la finalidad de llegar a una publicación exitosa para que los usuarios puedan explotar la información con un alto grado de confianza. Para realizar la limpieza y preparación de los datos, se modelaron y ejecutaron procesos de extracción, transformación y carga de datos (proceso ETL) (Bernabeu, 2010), los resultados de estos procesos se almacenan temporalmente en memoria utilizando un plugin de Pentaho, llamado “Data Pre catching”. La función de este plugin es permitir que los datos almacenados temporalmente sean manipulados para la generación de los RDF.

ii. Mapeo de datos

Luego del proceso de limpieza y transformación de los datos, se realizó el mapeo entre las fuentes de datos y los recursos ontológicos que sirvieron de base para la generación automática del grafo RDF. Los datos antes de ser almacenados en el plugin de Pentaho[11] se preparan en forma de tripletas que se obtienen a través de una consulta a la base de datos. En la Figura 11 se puede ver cada campo de las entidades mapeadas. ID es un código generado por el plugin, Ontology se refiere a la ontología de la cual se selecciona un determinado término. Entity es la entidad o clase procedente de la base de datos que se mapea con el registro procedente de la consulta de la base de datos. Relative URI es el texto para diferenciar el recurso generado. URI Field es el campo clave que permite identificar de manera única el recurso generado. DataField es el campo que con el que se filtran los registros de la base de datos y por ultimo DataValue es el valor que es mapeado con el campo Entity.

[11] http://www.pentaho.com/



#	ID	Ontology	Entity	Relative URI	URI Field ID	DataField 1	DataValue 1
1	C001	UCuenca-ontology.owl	http://swat.cse.lehigh.edu/onto/univ-bench.owl#University	universidad/	Id Record	Field	Universidad
2	C002	UCuenca-ontology.owl	http://vivoweb.org/ontology/core#Committee	consejo/	Id Record	Field	Comite
3	C003	UCuenca-ontology.owl	http://swat.cse.lehigh.edu/onto/univ-bench.owl#Department	departamento/	Id Record	Field	Departamento
4	C004	UCuenca-ontology.owl	http://vivoweb.org/ontology/core#ResearchOrganization	grupoinvestigacion/	Id Record	Field	GrupoInvestigacion
5	C005	UCuenca-ontology.owl	http://vivoweb.org/ontology/core#Committee	consejo/	Data	Field	Universidad/maxAutoridad/Comite
6	C006	UCuenca-ontology.owl	http://swat.cse.lehigh.edu/onto/univ-bench.owl#Department	departamento/	Data	Field	Universidad/tienexParte/Departamento
7	C007	UCuenca-ontology.owl	http://swat.cse.lehigh.edu/onto/univ-bench.owl#Department	departamento/	Data	Field	Comite/tienexParte/Departamento
8	C008	UCuenca-ontology.owl	http://swat.cse.lehigh.edu/onto/univ-bench.owl#Department	departamento/	Data	Field	Departamento/relacionadoPor/Departa
9	C009	UCuenca-ontology.owl	http://swat.cse.lehigh.edu/onto/univ-bench.owl#Department	departamento/	Data	Field	Departamento/tienexParte/Departamer
10	C010	UCuenca-ontology.owl	http://vivoweb.org/ontology/core#ResearchOrganization	grupoinvestigacion/	Data	Field	Departamento/relacionadoPor/GrupoIn
11	C011	UCuenca-ontology.owl	http://vivoweb.org/ontology/core#ResearchOrganization	grupoinvestigacion/	Data	Field	GrupoInvestigacion/tienexParte/GrupoI
12	C012	UCuenca-ontology.owl	http://swat.cse.lehigh.edu/onto/univ-bench.owl#Department	grupoinvestigacion/	Data	Field	GrupoInvestigacion/relacionadoPor/Gru
13	C013	UCuenca-ontology.owl	http://purl.org/vocab/aiiso/schema#Faculty	facultad/	Id_Record	Field	Facultad
14	C014	UCuenca-ontology.owl	http://vivoweb.org/ontology/core#School	carrera/	Id_Record	Field	Carrera

***Figura 11**: Mapeo de datos y recursos ontológicos*

iii. Transformación

El siguiente paso fue realizar la descripción semántica utilizando los vocabularios de las ontologías que fueron modeladas en la fase anterior. Se procedió entonces a realizar los mapeos necesarios de los datos registrados temporalmente con el plugin de Pentaho nombrado en la sección de preparación de datos, con las ontologías definidas. La última actividad realizada en esta fase fue la transformación de los datos al formato RDF, los grafos generados fueron almacenados en una base de datos de tripletas, las cuales son consultadas en el momento de la explotación.

A continuación, en la Figura 12, se puede ver todas las herramientas que se utilizaron para la generación de los datos enlazados. Para más detalle de las herramientas que se utilizaron en este trabajo se puede consultar el Anexo 1.

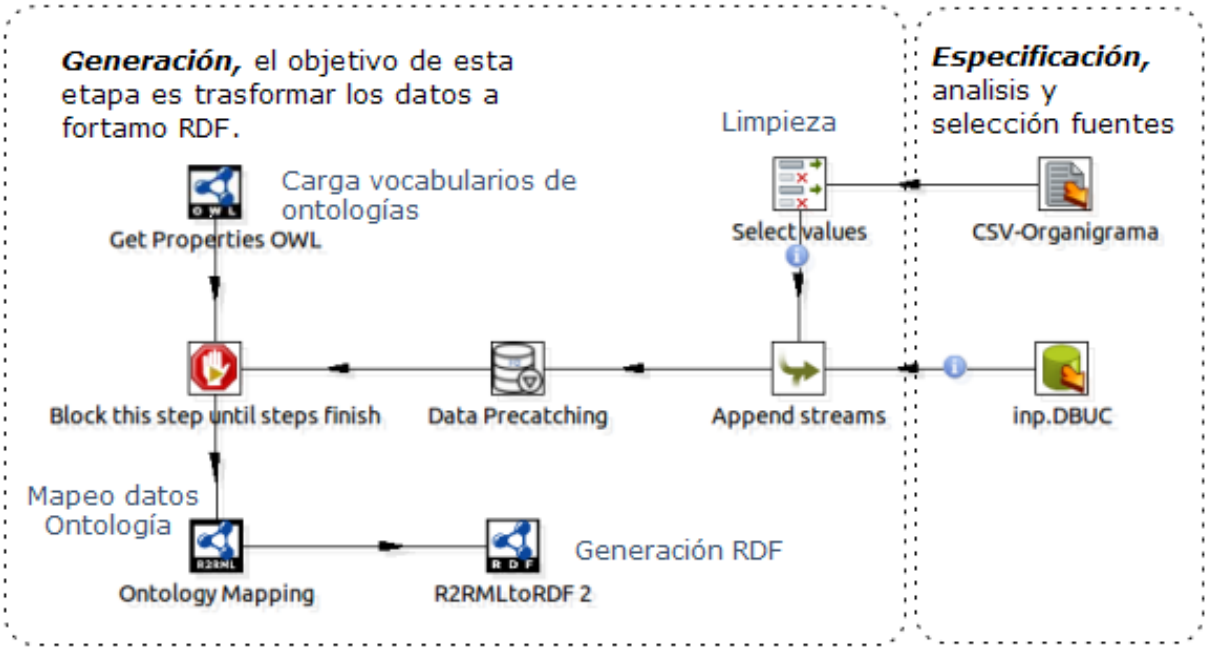


***Figura 12**: Herramientas utilizadas en la fase de generación*

4.2.4. Enlaces

Los enlaces se realizan con el propósito de que la nueva ontología enriquezca a las fuentes externas que pertenezcan al mismo dominio. La definición de los enlaces entre los RDFs que han resultado de la fase de generación con grafos similares existentes en la Web, es decir con fuentes de datos externas, es una tarea bastante importante, puesto que pueden aportar de manera significativa a los recursos publicados como datos enlazados en la Web enriqueciendo los recursos de los que se dispone y permitiendo que los resultados obtenidos tengan mayor alcance. El establecimiento de enlaces con las fuentes externas se propone como un trabajo futuro a este trabajo y para realizar esta actividad primero se deberán definir los recursos que podrán ser enlazados e identificar las posibles fuentes de datos externas que pertenezcan al mismo dominio.

Para cumplir con esta actividad se investigó que puede ser realizada utilizando la herramienta web llamada ***Silk Workbench***[12]**.** El propósito de esta aplicación es buscar recursos relacionados en fuentes de datos internas o externas, y generar enlaces RDF con dichos recursos en la web.

Luego de la fase de generación y la fase de definición de enlaces se obtiene como resultado la base de datos de tripletas. En la Figura 13 se visualizan algunas de las tripletas en forma de grafo, en donde se puede ver el sujeto, propiedad y objeto. Por otro lado, en la Figura 14 se pueden ver las tripletas escritas en lenguaje RDF.

[12] http://silkframework.org/

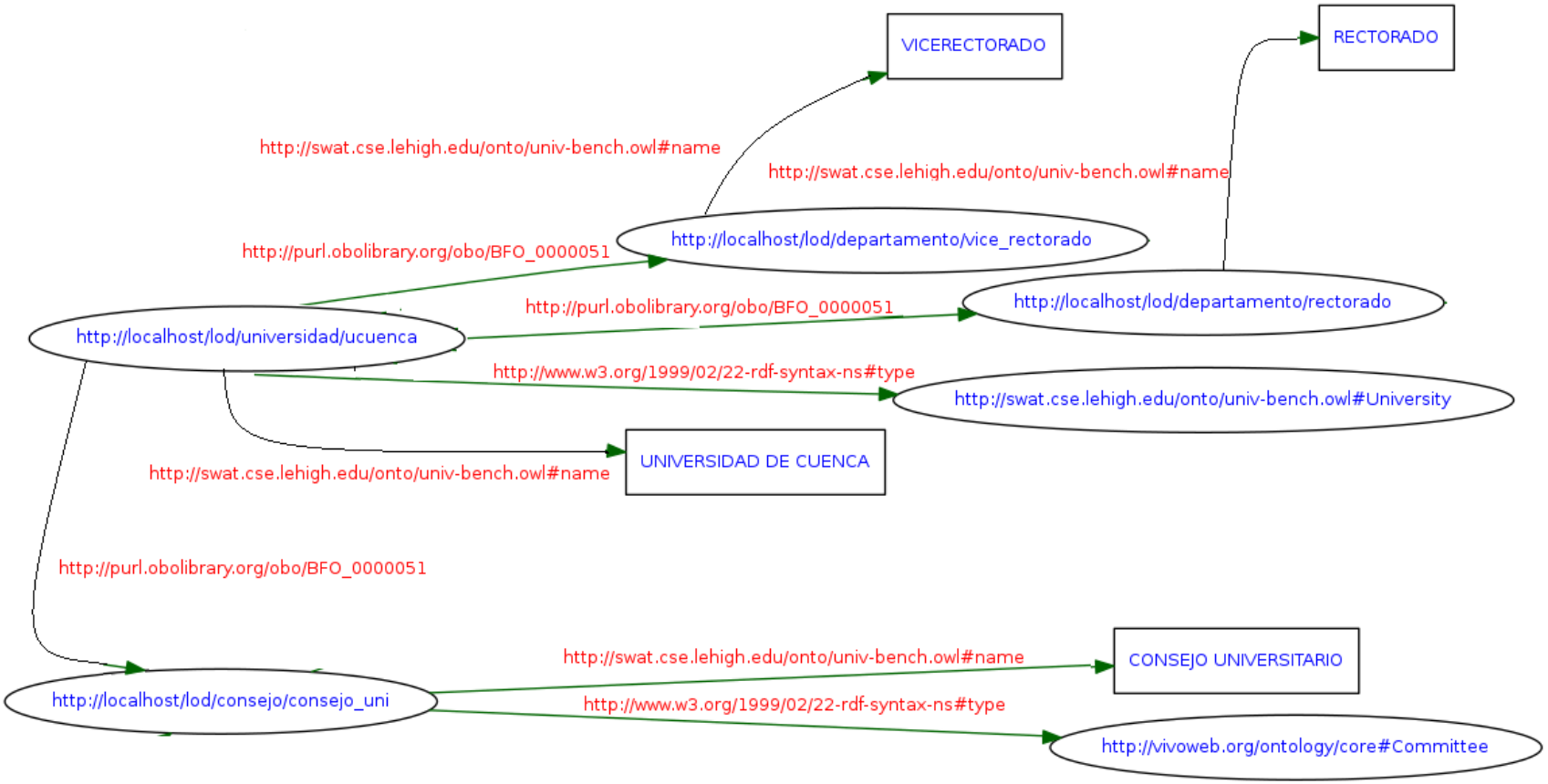


***Figura 13**: Visualización de las tripletas en forma de grafos enlazados*

Organizacion.rdf

```
<?xml version="1.0" encoding="UTF-8"?>
<rdf:RDF
    xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#">

<rdf:Description rdf:about="http://localhost/lod/consejo/consejo_uni">
    <name xmlns="http://swat.cse.lehigh.edu/onto/univ-bench.owl#">CONSEJO UNIVERSITARIO</name>
</rdf:Description>

<rdf:Description rdf:about="http://localhost/lod/departamento/vice_rectorado">
    <BFO_0000051 xmlns="http://purl.obolibrary.org/obo/" rdf:resource="http://localhost/lod/grupoinvestigacion/direccion_investigacion"/>
</rdf:Description>

<rdf:Description rdf:about="http://localhost/lod/grupoinvestigacion/direccion_investigacion">
    <relatedBy xmlns="http://vivoweb.org/ontology/core#" rdf:resource="http://localhost/lod/departamento/programas"/>
</rdf:Description>

<rdf:Description rdf:about="http://localhost/lod/departamento/rectorado">
    <relatedBy xmlns="http://vivoweb.org/ontology/core#" rdf:resource="http://localhost/lod/departamento/direccion_planificacion"/>
    <relatedBy xmlns="http://vivoweb.org/ontology/core#" rdf:resource="http://localhost/lod/departamento/direccion_adm_financiera"/>
    <relatedBy xmlns="http://vivoweb.org/ontology/core#" rdf:resource="http://localhost/lod/departamento/direccion_talento_humano"/>
    <relatedBy xmlns="http://vivoweb.org/ontology/core#" rdf:resource="http://localhost/lod/departamento/unidad_bie_universitario"/>
    <relatedBy xmlns="http://vivoweb.org/ontology/core#" rdf:resource="http://localhost/lod/departamento/direccion_rel_internacionales"/>
    <relatedBy xmlns="http://vivoweb.org/ontology/core#" rdf:resource="http://localhost/lod/departamento/direccion_tec_informacion"/>
    <relatedBy xmlns="http://vivoweb.org/ontology/core#" rdf:resource="http://localhost/lod/departamento/unidad_cultura"/>
    <relatedBy xmlns="http://vivoweb.org/ontology/core#" rdf:resource="http://localhost/lod/departamento/unidad_pla_fisica"/>
</rdf:Description>

<rdf:Description rdf:about="http://localhost/lod/departamento/vice_rectorado">
    <relatedBy xmlns="http://vivoweb.org/ontology/core#" rdf:resource="http://localhost/lod/departamento/direccion_vin_colectividad"/>
    <relatedBy xmlns="http://vivoweb.org/ontology/core#" rdf:resource="http://localhost/lod/departamento/direccion_edu_continua"/>
```

***Figura 14**: Visualización de las tripletas en formato XML*

4.2.5. Publicación

En esta etapa del ciclo de vida de los datos enlazados, se publica la información obtenida en la fase de generación, para que los grafos en RDF sean accedidos en la Web a través de consultas. Según Villazón-Terrazas (2011), en esta fase se realizan las siguientes tareas:

a) Publicación y almacenamiento de las fuentes de datos transformadas en tripletas RDF en un servidor de base de datos de tripletas
b) Publicación de los metadatos de las fuentes de datos antes mencionadas
c) Mantener los datos actualizados para su posterior recuperación

Para el caso de estudio de este trabajo, se publicó en un triple store FUSEKI[13], el cual permite publicar los archivos de información RDF en la web y ser consultada mediante el lenguaje de consultas SPARQL. La Figura 15 visualiza gráficamente como se realiza la publicación de los datos enlazados.

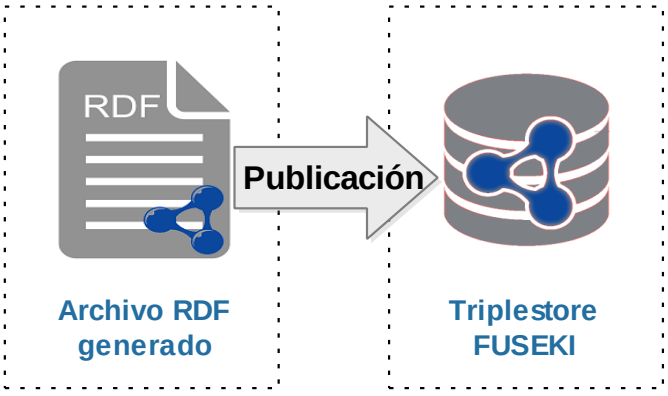


***Figura 15**: Publicación de datos enlazados*

[13] https://jena.apache.org/documentation/fuseki2/index.html

4.2.6. Validación

Una vez que se realizó la publicación de los datos enlazados, se vio la necesidad de incluir la fase de validación en el ciclo de vida, para asegurar que la ontología que fue modelada puede contestar las preguntas que el usuario consulta sobre un dominio específico, en este caso sobre la estructura organizacional de la Universidad y el área académica de carreras. Para realizar la validación se solicitó la colaboración de autoridades como directores de carreras y miembros de la comisión técnico-curricular quienes comprobaron que las tripletas almacenadas pueden contestar las preguntas que fueron definidas durante la fase de modelamiento. En la Tabla 8 se puede ver algunas de las preguntas planteadas para realizar la validación.

Tabla 8: *Ejemplos de preguntas planteadas para la validación*

	Preguntas
1.	Cuántas Facultades hay en la Universidad
2.	Cuáles son las Facultades de la Universidad
3.	Cuáles son los títulos que oferta la carrera de Comunicación Social

Para obtener las respuestas se realizaron consultas SPARQL sobre las tripletas almacenadas. Por otro lado, se realizaron consultas SQL a las bases de datos de los sistemas informáticos para obtener la información sobre las mimas preguntas planteadas. Se comprobó entonces que a pesar de que las respuestas obtenidas a través de consultas a las bases de datos fueron las mismas, para obtener el resultado final se realizó un proceso más largo y desgastante a diferencia de la consulta a la ontología que resulta más liviana y con mejor velocidad en dar la respuesta, puesto que no se necesita de procesos manuales para consolidar la información. Por todo esto las autoridades antes nombradas pudieron concluir que la ontología fue modelada correctamente.

A continuación, en la Figura 16 se presenta gráficamente la respuesta a la consulta de la segunda pregunta visualizando el resultado de la consulta SPARQL y de la consulta directa a la

base de datos. En la parte superior de la Figura se puede visualizar la consulta SPARQL del modelo RDF publicado con Fuseki y el resultado de la misma y en la parte inferior se puede ver la consulta realizada en el cliente de Oracle SQL developer.

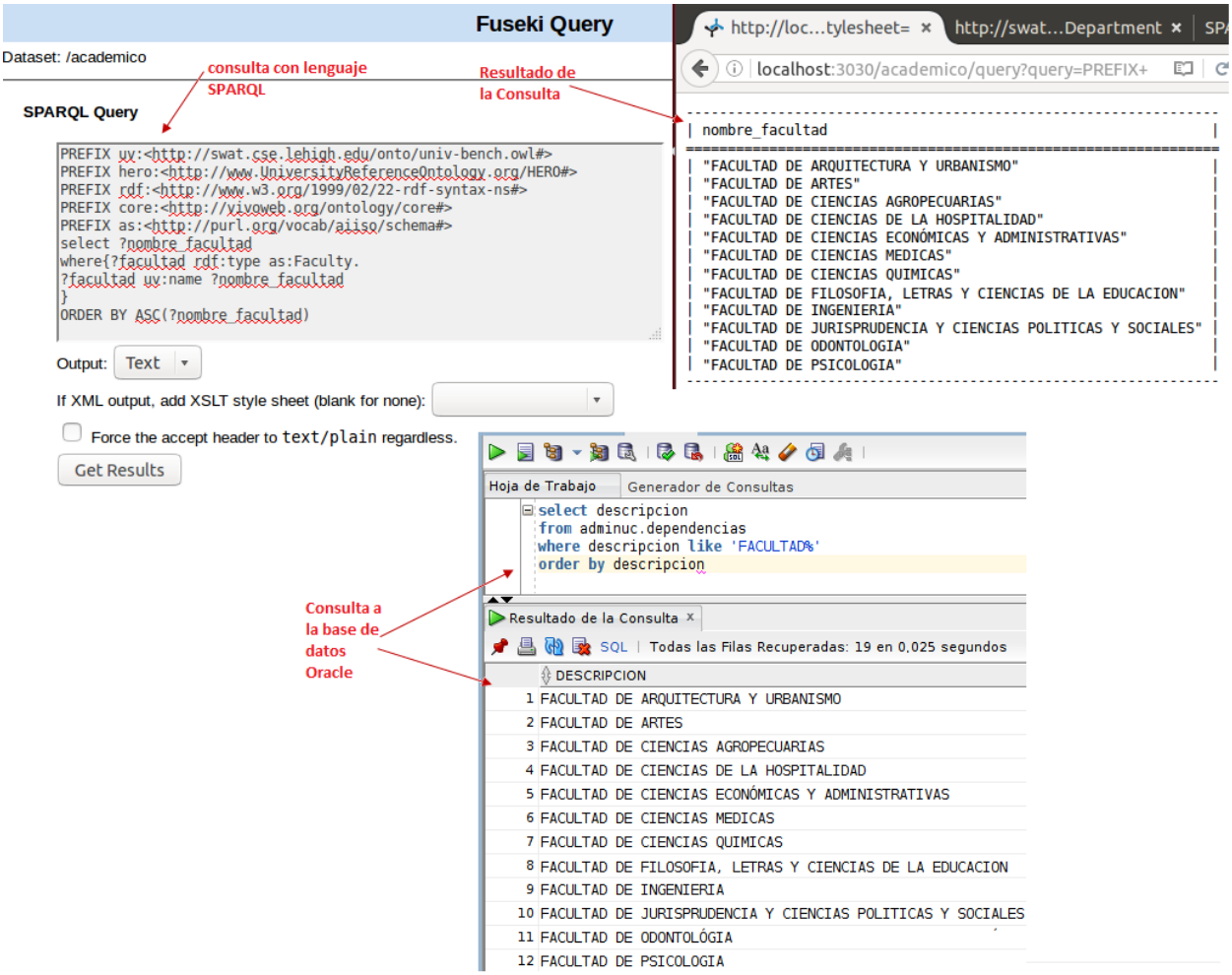


***Figura 16**:Ejemplo de consulta SPARQL vs. SQL*

4.2.7. Explotación

En el ciclo de vida de los datos enlazados la última fase es la explotación. No es suficiente que los datos estén publicados en la Web, por esta razón, como parte de este trabajo se construyó un prototipo de un buscador semántico para que los datos publicados puedan ser consumidos. Todas las consultas que se realizan a través de este buscador semántico fueron

implementadas a través de consultas escritas en lenguaje SPARQL. En el capítulo 5 se puede revisar a detalle cómo se realizó el prototipo del buscador.

4.3. Resumen del proceso de integración

Debido al alcance de este trabajo, como se ha indicado anteriormente se ha realizado la integración de datos correspondientes a la estructura organizacional de la Universidad de Cuenca y de sus Carreras, sin embargo, es importante indicar que el proceso utilizado para la integración de los datos, puede ser aplicado a cualquiera de los dominios que fueron identificados en la sección 4.1. A continuación, se nombra rápidamente cada una de las tareas en cada fase del ciclo de vida:

1. Especificación
    a. Identificación y análisis de las fuentes de datos, con el objetivo de definir los requisitos para la vinculación de datos
    b. Diseño de URIs para identificar y acceder de manera única a los recursos y atributos de los datos enlazados.
2. Modelamiento de la Ontología, tratando de reutilizar de elementos ontológicos existentes
    a. Para la construcción de ontologías se utiliza la metodología NeOn.
    b. Especificar los requerimientos ontológicos que debe responder la ontología (DERO).
    c. Modelar la ontología.
    d. Buscar recursos ontológicos compatibles para reutilizarlos
    e. Comparar ontologías que permitirá determinar aquellas que más se ajusten a los criterios de selección preestablecidos.
    f. Utilizar protégé para la construir la ontología en lenguaje OWL.
3. Generación: convertir los datos a formato RDF utilizando la ontología antes desarrollada.
    a. Realizar proceso de preparación y limpieza de errores en los datos.
    b. Mapear las fuentes de datos y los recursos ontológicos.
    c. Transformar los datos a formato RDF.

4. Enlaces de los datos con dominios internos y externos.
5. Publicación
    a. Publicar y almacenar el RDF generado en un servidor de base de tripletas.
    b. Publicar los metadatos
    c. Mantener los datos actualizados
6. Validación: Verificar que el conjunto de tripletas RDF sean correctas y contesten las preguntas definidas.
7. Explotación: Para consumir los datos publicados.

Para integrar nuevas fuentes de datos pertenecientes a los dominios identificados en la Tabla 1 del capítulo 4, el paso inicial es realizar a conciencia la fase de especificación, luego en el modelamiento se debe definir y completar de manera correcta el documento de requerimientos ontológicos (DERO). Por tanto se integrarán nuevos vocabularios u ontologías que cumplan los requerimientos identificados, dando como resultado una nueva versión de la ontología base que incluirá el nuevo dominio que se está integrando. A continuación, se actualizará la base de datos de tripletas y se establecerán los enlaces entre los dominios. Se procederá entonces a publicar los datos para que deben ser validados para que finalmente puedan ser explotados por los usuarios.

Capítulo 5
EXPLOTACIÓN DE LOS DATOS

Este capítulo relata más detalladamente el desarrollo del prototipo del buscador semántico para realizar la explotación de los datos de la Universidad de Cuenca.

Como se ha visto a lo largo del desarrollo de este proyecto, se ha definido un proceso a seguir para la publicación de datos enlazados, que da como resultado la homogenización e integración de los datos con los que cuenta la Universidad de Cuenca. Para el caso de estudio se utilizó las fuentes de información sobre la estructura organizacional y académica de la Institución. Cabe anotar que este proceso servirá como punto de partida para la integración y publicación de las demás fuentes de datos heterogéneas con las que cuenta la institución.

Como objetivo de este trabajo se planteó la construcción de un prototipo de buscador semántico para realizar la explotación de los datos, la cual consiste en el desarrollo de una herramienta de software con capacidad de búsqueda semántica por medio de consultas SPARQL. Es por esta razón, que en este capítulo se detalla la tecnología utilizada para la generación del prototipo, que cabe indicar que se ha basado en trabajos anteriores (Peñaloza & Santacruz, 2015) y no se ha iniciado desde cero.

Para explotar los datos es importante definir la periodicidad con la cual serán actualizadas las tripletas en la base de datos. Esta depende que tan cambiantes son los datos. Para el caso de la estructura organizacional se deben actualizar cuando por ejemplo se cambian los jefes de dependencia, entonces se definió que a través de triggers estos cambios serán almacenados en una tabla de la base de datos relacional, de manera que cada mes se haga un barrido de esta tabla y se proceda a realizar la actualización de los RDFs. Para el caso del área académica los datos son más cambiantes, por ejemplo, cuando se registran matrículas o calificaciones. Por esta razón para el área académica se establecieron distintos periodos de actualización. El primero se da

cuando se termina el proceso de matrículas. Con respecto a las calificaciones, las tripletas serán actualizadas luego de culminar los registros de cada aporte.

La ventaja de contar con un buscador semántico que ataque a una base de datos de tripletas RDF es sin duda que este tipo de buscadores pueden realizar interpretaciones más precisas de las búsquedas de los usuarios. Es decir, el buscador puede lograr inferencias sobre las consultas. Como un ejemplo simple se puede inferir si una persona es un estudiante, solo con saber si tiene matrícula y tiene calificaciones.

Otra ventaja de tener una base de datos RDF que será atacada por el buscador es que: el momento que se deba realizar una integración con otro modelo de datos RDF, no es importante conocer las plataformas relacionales que almacenan los datos originales, puesto que lo único que será necesario es crear el enlace y agregar nuevas tuplas para RDF.

Es importante anotar que al término de este trabajo el usuario a través del buscador, obtendrá información únicamente de las áreas indicadas. Las respuestas a consultas de datos de otras áreas una vez se conseguirán una vez que se integren los demás dominios.

5.1. Definición del prototipo

El prototipo pretende satisfacer las siguientes necesidades:

a) Búsquedas sencillas de información en la base de datos de tripletas generadas para el dominio específico. Para este caso de estudio el repositorio contiene datos de la estructura organizacional y académica.
b) Reutilizar herramientas que ya han sido desarrolladas en proyectos anteriores
c) Generar el componente de búsqueda que acceda a la base de datos RDF.
d) Los procesos de búsqueda se basan en modelos ontológicos y herramientas de consulta SPARQL, que gestione las relaciones de los recursos ontológicos existentes en la base de

datos triplestore, de tal forma que los resultados generados puedan tener un detalle amplio y enlazado.

e) Los usuarios tienen la posibilidad de definir criterios de búsquedas con respecto al dominio específico requerido

f) El buscador permite el autocompletado de lo que el usuario está ingresando para consultar, así también se puede definir el ámbito de búsqueda, seleccionando los íconos que se encuentran debajo del cuadro de texto de búsqueda.

g) El prototipo está implementado en la Web

En la Figura 17 se puede ver la arquitectura definida para el prototipo del buscador

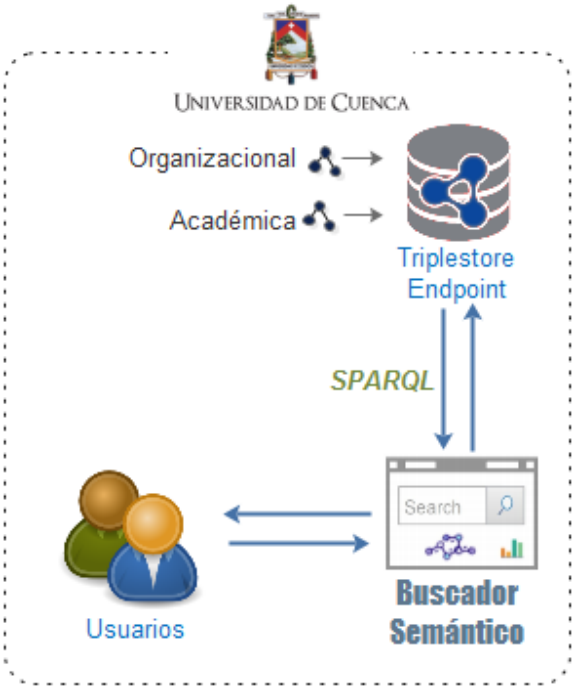


***Figura 17**:Definición del prototipo*

5.2. Buscador semántico

En esta sección se describe el funcionamiento del prototipo del buscador semántico. El acceso al buscador se realiza a través de un navegador web. La Figura 18 visualiza la pantalla principal del buscador semántico. En el centro de la pantalla se presenta un cuadro de búsqueda y

en la parte superior está la barra de menús que permite acceso a diferentes opciones del buscador. Así también el prototipo tiene la posibilidad de gestionar los niveles de acceso y cuentas de usuario.

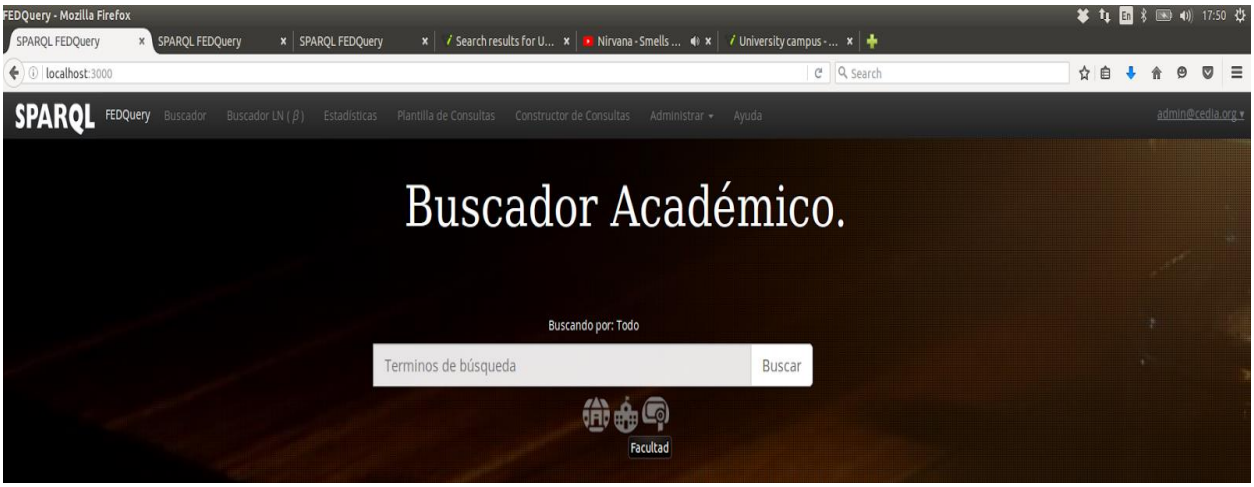


***Figura 18**:Pantalla principal del buscador semántico*

Como se puede ver en el gráfico en la parte superior está la barra de menús que permite acceder a diferentes opciones del buscador, las cuales serán visibles de acuerdo al nivel de acceso que tiene el usuario:

- FEDQuery: Página principal para búsquedas.
- Buscador: Interfaz de búsqueda y obtención de resultados.
- Buscador LN (Beta): Interfaz para la generación de consultas en lenguaje natural.
- Estadísticas: Resumen de las fuentes en forma gráfica.
- Plantilla de Consultas: Constructor
- Constructor de Consultas: Herramienta gráfica para la creación de consultas.
- Administrar: Gestión de usuarios.

5.3. Búsqueda

La búsqueda consiste en ingresar un texto referente a una pregunta sobre la cual se busca una respuesta. Este texto puede ser una palabra clave o algo descriptivo del recurso solicitado.

En la parte inferior del cuadro de búsqueda existen íconos que permiten realizar búsquedas específicas por recurso. Esto se puede ver en la Figura 19.

***Figura 19**:Ejemplo de consulta en el buscador semántico*

El buscador también permite el autocompletado por el criterio de búsqueda, haciendo sugerencias en tiempo de ejecución.

5.4. Pantalla de resultado de las búsquedas

La interfaz de la figura 20 es utilizada para presentar los resultados, con otras funcionalidades, como por ejemplo el cuadro de búsqueda, fuente del recurso, nombre y descripción del recurso en RDF, página de descripción del recurso, grafo, etc. En la Figura 21 se puede observar el resultado del recurso RDF.

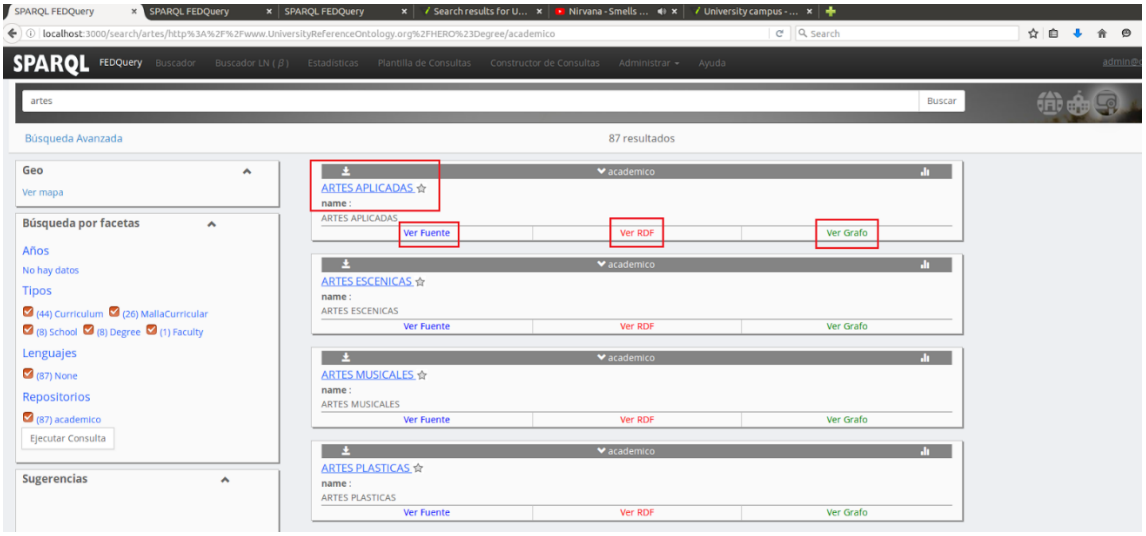


***Figura 20**:Página de resultados*

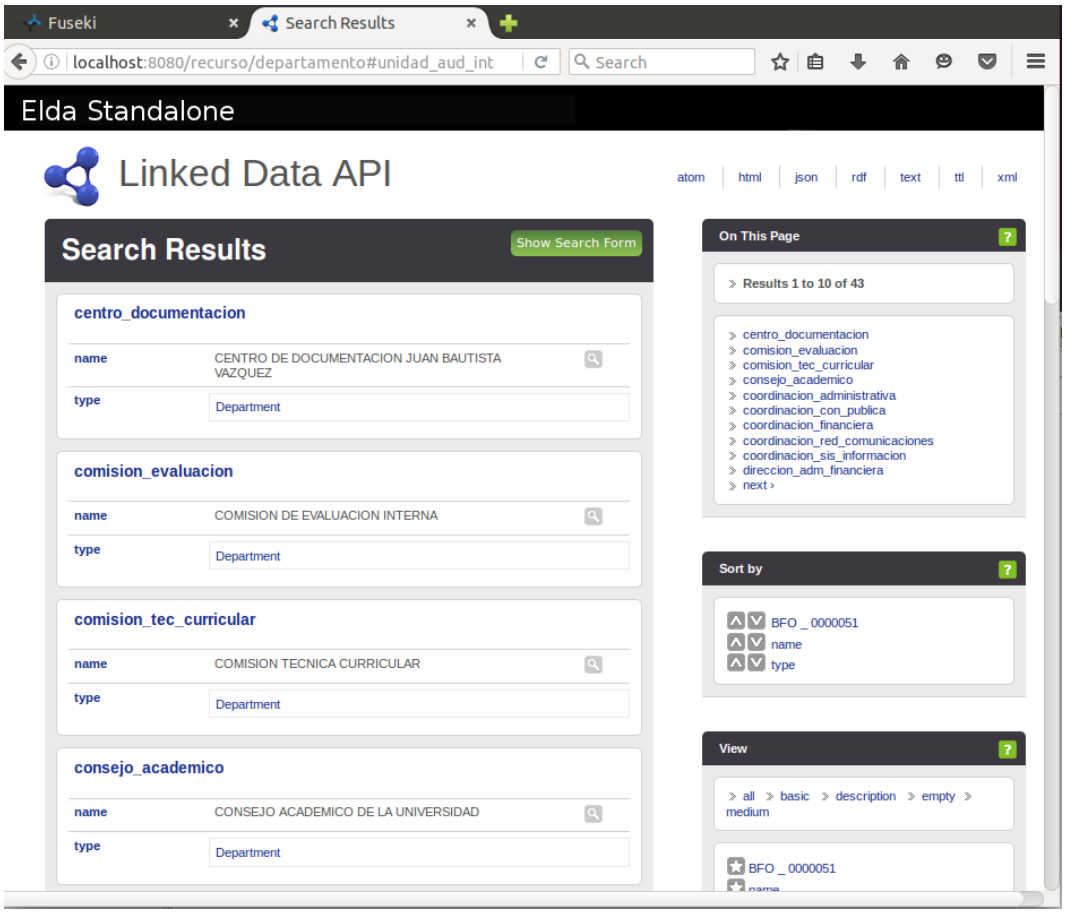


***Figura 21**:Página de descripción del recurso RDF*

En la Figura 22 se puede ver el resultado de la información en forma de grafo.

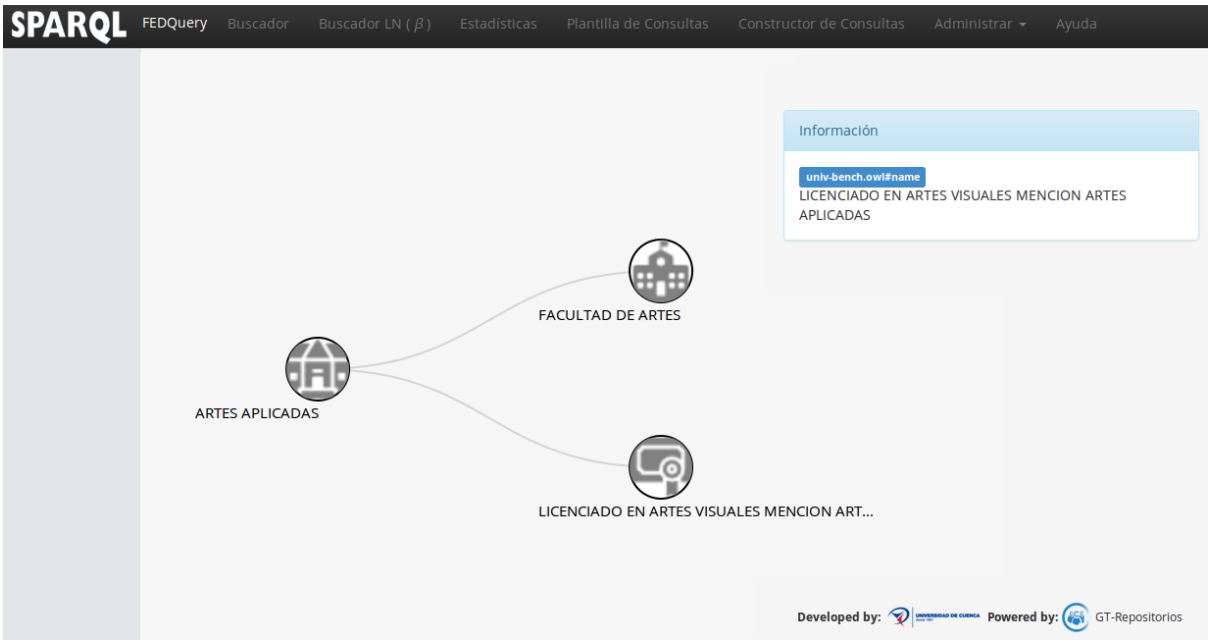


***Figura 22**:Información consultada vista en forma de grafo*

Capítulo 6
CONCLUSIONES Y TRABAJOS FUTUROS.

Luego de culminado este trabajo se pudo constatar que el problema de la Web actual es que su significado no es accesible por las computadoras, este problema ha sido superado a través de la Web Semántica, puesto que a través de un medio universal de intercambio de información aporta semántica, es decir significado, a los documentos en la Web.

Al analizar las alternativas para realizar la integración de los datos, se optó por implementar un modelo de integración materializado, puesto que todos los datos pueden ser obtenidos de bases de datos y de archivos propios de la Universidad, lo cual permite que la actualización de las tripletas RDF sin depender de terceros y los usuarios pueden encontrar la respuesta a sus consultas de manera fiable y a tiempo.

La ontología definida representa de manera sencilla y entendible el dominio de la estructura organizacional y el área de las carreras de la Universidad. Esta ontología ha sido construida siguiendo el proceso definido en este trabajo en la sección 4.2, permitiendo que la información organizacional y académica este integrada a partir no solamente de fuentes como las bases de datos sino también documentos tanto digitales como físicos.

A pesar de que para el desarrollo de la ontología en este trabajo se utilizó el escenario 6 de la metodología NeOn, se recomienda que para la creación de nuevas ontologías para la Universidad se debe seleccionar el escenario que más se adapte al dominio a generar.

Uno de los retos que se tuvo que enfrentar fue la búsqueda de vocabularios y ontologías que pudieron ser reutilizadas, puesto que se tuvo que analizar de manera profunda la descripción de la ontología y sus términos para entenderla y conocer si puede ser reutilizada. Hubo casos en

los que por los términos parecían ser útiles, sin embargo, al no tener la descripción clara y entendible en algunos casos estas fueron descartadas.

El proceso de integración que se definió en este trabajo es fácil de seguir, puesto que, para cada fase, todas las actividades están muy bien establecidas y las herramientas utilizadas ya han sido probadas, dando muy buenos resultados, por esta razón se recomienda el uso de estas. Más detalles sobre las herramientas se puede ver en el Anexo I. Sin embargo, un proceso importante a tomar en cuenta es lo referente a la generación de las tripletas, puesto que, para llegar a tener datos consistentes y libres de errores, fue necesario realizar varios procesos de limpieza, como por ejemplo eliminar las duplicidades de dependencias u omitir la carga de datos de carreras que en el sistema académico son consideradas como auxiliares para las cuales no existen estudiantes matriculados en dichas carreras. Así también para la estructura organizacional al no tener los datos en una base de datos relacional ni en un documento estructurado se procedió a generar un archivo .csv con todos los datos requeridos de manera que pudo luego ser leído para generar el RDF.

La implementación de un prototipo de buscador semántico permitió comprobar que los resultados de las búsquedas pueden satisfacer las expectativas de los usuarios pues podrán obtener resultados más precisos.

Como trabajos futuros se plantea la posibilidad de profundizar en la fase de enlaces perteneciente al ciclo de vida definido en este trabajo para la publicación de datos. El enlazar con recursos externos es una oportunidad para que la ontología definida e implementada para uso de la Universidad pueda ser útil y reutilizada en otros proyectos pertenecientes al mismo dominio.

Es importante también indicar que el modelo y arquitectura definida en este trabajo, el cual fue utilizado para implementar la publicación de los datos enlazados de la estructura organizacional y del área académica es la base para la construcción de las nuevas ontologías

correspondientes a los diferentes dominios de interés para la Universidad identificados en la Tabla 1 de la sección 4.2.

Se recomienda también que a partir de este trabajo el prototipo propuesto para el buscador semántico sea implementado para uso de toda la comunidad universitaria, de manera que los usuarios puedan verse beneficiados de este trabajo, puesto que los resultados en las búsquedas son más precisos y de mejor calidad en comparación con los resultados que hoy en día obtienen en los buscadores actuales.

Glosario de términos

3tore	Base de datos de tripletas
BD	Base de datos
Dataset	Conjunto de datos
DERO	Documento de especificación de requerimientos ontológicos
FOAF	Ontologia Friend of a Friend
Fuseki	Servidor de SPARQL
HTTP	Protocolo de Transferencia de Hipertextos. En inglés: Hypertext Transfer Protocol
Jena	Framework para construir aplicaciones de Web Semántica y Linked Data
Linked data	Datos enlazados
LOD	Linked Open Data
LOV	Buscador de vocabularios ontológicos
Metadato	dato para describir más datos
MySQL	Base de datos relacional
NS	Name Spaces
Ontología	Vocabulario que describe conocimiento
Open ERP	Sistema financiero
Oracle	Base de datos relacional
OWL	Web Ontology Language. Lenguaje para desarrollar temas o vocabularios específicos
Pentaho	Es una herramienta de **Business Intelligence** desarrollada bajo la filosofía del software libre para la gestión y toma de decisiones empresariales
Plugin	Aplicación que permite añadir funcionalides adicionales a un programa informático
Portafolio	Portal del docente
PostgresSQL	Base de datos relacional
Proceso ETL	Proceso de extracción, transformación y carga de datos

Protégé	Herramienta para desarrollo de ontologías
Quipux	Sistema documental
RDF	Marco de Descripción de Recursos. Lenguaje para la definición de ontologías y metadatos en la web. En inglés: Resource Description Framework.
RDF Schema	lenguaje de definicion de ontologias basado en RDF
Sesame	Framework open source para consultar y analizar datos RDF
SGA	Sistema de gestión académica
SGAP	Sistema de gestión académica de posgrados
SGE	Sistema de gestión de la evaluaión
SGEI	Sistema de Gestión de Evaluación Institucional
SGPA	Sistema de gestión de para-académicos
Silk Workbench	Framework de software libre para integrar fuentes de datos heterogéneas
SPARQL	lenguaje de consulta sobre RDF
Swoogle	Buscador de vocabularios ontológicos
triple store	Base de datos para almacenar tripletas
URI	**U**niforme **R**esource **I**dentifier. Cadena de caracteres para identificar recursos
URL	Uniform Resource Locator
Virtuoso	Servidor que permite trabajar con datos enlazados
W3C	World Wide Web Consortium
Watson	Buscador de vocabularios ontológicos
XML	E**X**tended **M**arkup **L**anguage, Lenguaje de etiquetado

ANEXO 1

Herramientas utilizadas para la publicación de datos enlazados

Para lograr la publicación de los datos enlazados se utilizó la herramienta Pentaho Data Integration, la misma que a pesar de que es un software libre para la toma de decisiones relacionado con la inteligencia de negocios, provee herramientas a través de pluggins que permiten la publicación de datos enlazados. Por otro lado, la tesis de grado desarrollada en la Universidad, "Creación de componentes para el framework de generación de resource description framework (RDF)", proporciona un entorno unificado para la generación de datos enlazados que abarca las fases de la publicación de datos enlazados. Los autores de esta tesis crean los componentes utilizando la herramienta Pentaho Data Integration, estos componentes permiten realizar las actividades de cada una de las fases como por ejemplo extracción de datos, carga de ontologías, mapeos, generación de RDFs, publicar y explotar. A continuación, se realiza una breve descripción de los componentes utilizados (Peñaloza & Santacruz, 2015).

Especificación

Para realizar las tareas de la especificación, Pentaho ofrece componentes propios que permiten la lectura de las distintas fuentes como por ejemplo archivos, bases de datos, servicios web, etc. Estos componentes permiten modelar los procesos ETL (extracción, tranformación y carga). Además de estos componentes propios de Pentaho, existen otros que son compatibles como por ejemplo *OAILoader*, que permite extraer información de repositorios digitales como Dspace. Otro componente es *MARC21* utilizado para la lectura de metadatos de recursos bibliográficos (LOD-GF).

Modelamiento

Para la fase de modelamiento, la plataforma desarrollada provee de un plugin que permite cargar los vocabularios de nuevas ontologías o las presentes en la web. Este plugin se denomina Get Projperties OWL y permite dos tipos de carga de archivo o de la web.

Generación

Para lograr la generación se realizan algunos pasos, cada uno de los cuales usa algunos componentes. Para realizar la **limpieza de los datos** se utiliza componentes propios para el procesamiento de datos. El siguiente componte que no es propio, sino más bien un plugin llamado *Data Pre catching* que permite que los datos procesados puedan ser manipulados en pasos posteriores. **La conversion de datos** en la fase de generación utiliza el plugin *Ontology & Data Mapping*, este plugin es el que permite vincular los recursos con los vocabularios.

Enlaces

Para realizar los enlaces de datos con fuentes externas se utiliza el componente ***SilkPlugin***, que permite utilizar los beneficios de ***Silk Workbench*** para de esta forma lograr el enriquecimiento de la información porque permite encontrar recursos similares entre dos fuentes.

Publicación

Para realizar la publicación se requiere guardar la base de datos de tripletas en un triplestore que dispone de un cliente de acceso a los datos (SPARQL Endpoint). Para esta fase se utiliza el componente ***Fuseki Loader***, que se utiliza para obtener la configuración de los parámetros básicos para el despliegue de un triplestore ***Fuseki.***

Explotación

Componente **ELDA Loader,** que genera una página de descripción de los recursos, que permite navegar en la información que está en formato RDF.

En la Figura 23 se puede ver gráficamente un resumen de los componentes que soportan cada una de las etapas de Linked Data.

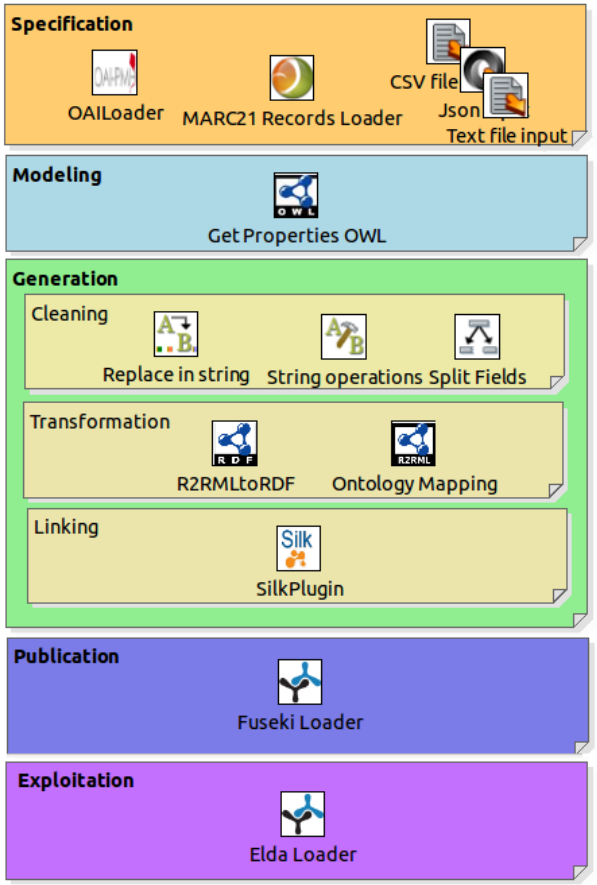


***Figura 23**:Componentes en Pentaho para el ciclo de vida de los datos enlazados*

Fuente: [Figura]. Recuperado de https://ucuenca.github.io/lodplatform/#introducci%C3%B3n

Referencias Bibliográficas

Anguita, A. (2012). *Modelo de mediación semántica para la integración de fuentes de datos heterogéneas.* Tesis doctoral, Universidad Politécnica de Madrid, Facultad de Informática, Departamento de Inteligencia Artificial.

Antiñaco, M. (2013). *Bases de Datos NoSQL: Escalabilidad y alta disponibilidad a través de patrones de diseño.* Universidad Nacional de la Plata, Facultad de Informática.

Arroyo, E., Castro, E., & Rosario, P. (2008). La Educación y la Web Semántica. *Telematique , 7* (1), 117-126.

Barrera, M., Nuñez, H., & Ramos, E. (2012). Ingeniería Ontológica . *Lecturas en Ciencias de la Computación* .

Bernabeu, R. (2010). *HEFESTO*. Córdoba, Argentina.

Berners-Lee, T. (1998). *W3C*. Obtenido de Semantic Web - XML2000: https://www.w3.org/2000/Talks/1206-xml2k-tbl/

Bravo, J., Carranza, C., Castells, Pablo, F. J., Rico, M., Alonso, J., y otros. (2004). Aplicación de tecnologías de la Web Semántica a la gestión de información financiera y económica. *V Congreso en Interacción Persona-Ordenador (Interacción 2004)*, (págs. 326-329).

Bustamante, D., & Sequeda, J. (2006). *Propuesta del Uso de Ontologías para la Búsqueda Semántica en Laboratorios de Investigación y Desarrollo: OLID.* Universidad del Valle.

Cali, A., Calvanese, D., de Giacomo, G., & Lenzerini, M. (2002). Data Integration under Integrity Constraints. *International Conference on Advanced Information Systems Engineering*, (págs. 335-352).

Castelló, A. (2006). *WEB SEMÁNTICA: RDF Y SGBD QUE LO SOPORTAN*. Trabajo final de carrera.

Castro, R. (2008). *Representación del Conocimiento. Web Semántica.* Universidad Carlos III de Madrid, Madrid, España.

Chimbo, D., Contreras, P., & Espinoza, M. (2017). Aplicación de tecnologías semánticas y realidad aumentada paa realizar búsquedas de personas, puntos de interés y actividades dentro del campus centra de la Universidad de Cuenca. *MASKANA , 5* (2).

Codina, L., & Rovira, C. (2006). La Web Semántica. En J. Tramullas, *Tendencias en documentación digital* (págs. 9-54). Trea.

Criado, L. (2009). *PROCEDIMIENTO SEMI-AUTOMÁTICO PARA TRANSFORMAR LA WEB EN WEB SEMÁNTICA.* Tesis doctoral, Universidad Nacional de Educación a Distancia (UNED), Departamento de Inteligencia Artificial, Madrid, España.

del Busto, G., & Yanez, O. (2012). Bases de datos NoSQL. *Revista Telem@tica , 11* (3), 21-33.

Doan, A., & Halevy, A. I. (2012). *Principles of Data Integration.* Elsevier.

DTIC. (2017). *Docuentación de proyectos de la DTIC* . Cuenca, Ecuador.

Duque, N., Chavarro, J., & Moreno, R. (2006). EVOLUCIÓN DE LOS LENGUAJES UTILIZADOS EN LA CONSTRUCCIÓN DE LA WEB. *Revista Scientia Et Technica , XII* (32), 381-386.

Escribano, R., García, A., Alcañiz, M., & Marcos, A. B. (2004). *Sistemas de Representación y Procesamiento Automático del Conocimiento. ONTOLOGIAS EN LA WEB SEMANTICA.* Grado de Licenciatura, Universidad Politécnica de Valencia, Facultad de Informática, Valencia.

Fernández, M., Gómez-Pérez, A., & Juristo, N. (1997). METHONTOLOGY: From Ontological Art Towards Ontological Engineering. *AAAI Technical Report SS-97-06* , 33-40.

Fernandez, N., & Sánchez, L. (2005). La Web Semántica: fundamentos y breve "estado del arte". *Novática: Revista de la Asociación de Técnicos de Informática* (178), 6-11.

Fierros, I., Menéndez, V., & Castellanos, M. (2016). Bases de Datos Semánticas. *Revista Latinoamericana de Ingenieria de Software* , 209-215.

Galey, J. M. (2010). *Aplicación web semántica para la gestión de referencias bibliográficas.* Proyecto de fin de carrera, Universidad Rey Juan Carlos, Escuela Superior de Ingeniería Informática.

Gruber, T. R. (1993). A Translation Approach to Portable Ontology Specifications. *Knowledge Acquisition , 5*, 119-220.

Guarino, N. (1998). Formal Ontology and Information Systems. *Proceedings of FOIS'98* , 3-15.

Guzmán, J., López, M., & Durley, I. (2012). Metodologías y métodos para la construcción de ontologías. *Scientia et Technica* (50), 133-140.

Heath, T., & Bizer, C. (2011). *Linked Data Envolving the Web into a Global Data Space.* Morgan & Claypool Publishers.

Lenat, D., & Guha, R. (1990). Building Large Knowledge-Based Systems: Representation and Inference in the Cyc Project. *Artificial Intelligence , 61* (1993), 95-104.

Llamas, M. I. (2006). *Lenguajes de consulta para documentos RDF*. Trabajo final de carrera, Universitat Oberta de Catalunya.

LOD-GF. (s.f.). Obtenido de LOD-GF: An Integral Open Data Generation Framework: https://ucuenca.github.io/lodplatform/#introducci%C3%B3n

Lozano, A. (2001). *Ontologías en la Web Semántica.* Universidad de Extremadura, España.

Mora, M. B., & Segarra, V. (2016). Modelo ontológico para la representación de datos académicos y su publicación con tecnología semántica. *Opción , 32* (10), 267-282.

Moreno, C., & Sánchez, Y. (2012). *Prototipo de buscador semántico aplicado a la búsqueda de libros de ingeniería de sistemas y computación en la biblioteca Jorge Roa Martínez de la Universidad Tecnológica de Pereira.* Proyecto de grado, Universidad Tecnológica de Pereira, Pereira.

Oliva, R. (2013). *Modelo de integración de datos, metadatos y conocimiento geográficos.* Tesis doctoral, Universidad de Alicante, Departamento de Tecnología Informática y Computación.

Pascual, I., Valdés, O., & Gómez, E. (s.f.). *Web Semántica*. Obtenido de Componentes de la Web Semántica: http://lawebsemantica.weebly.com/componentes-de-la-web-semaacutentica.html

Pedraza-Jimenez, R., Codina, L., & Rovira, C. (2007). Web semántica y ontologías en el procesamiento de la. *El profesional de la información , 16* (6), 569-578.

Peis, E., Herrera-Viedma, E., & Morales, J. (2007). Aproximación a la web semántica desde la perspectiva de la Documentación. *Investigación bibliotecológica , 21* (43), 47-71.

Peñaloza, F., & Santacruz, F. (2015). *Creación de componentes para el framework de generación de resource description framework (RDF).* Tesis de grado, Universidad de Cuenca, Facultad de Ingeniería.

Piñeres, M., & Bonilla, I. (2008). De la web actual a la web semántica. *Prospectiva , 6* (2), 65-70.

Priyatna, F. (2015). *Methods and Techniques for the Generation and Efficient Exploitation of RDB2RDF Mapping.* Tesis doctoral, Universidad Politécnica de Madrid, Escuela Técnica Superior de Ingenieros Informáticos, Departamento de Inteligencia Artificial , Madrid, España.

quees.info. (s.f.). Obtenido de Buscador de Internet - Explicación y definición de buscador: https://www.quees.info/que-es-un-buscador.html

Reuco, R. (2008). *arco de Trabajo para el desarrollo de aplicaciones con ontologías de dominio.* Tesis de master, Universidad Central "Marta Abreu" de Las Villas, Santa Clara, Cuba.

Rodriguez, K., & Ronda, R. (2005). Web semántica: un nuevo enfoque para la organización y recuperación de información en el web. *ACIMED , 13* (6).

Rosell, Y., Senso, J., & Leiva, A. (2016). Diseño de una ontología para la gestión de datos heterogéneos en universidades: marco metodológico. *Revista Cubana de Información en Ciencias de la Salud* , 545-567.

Studer, R., Benjamins, V. R., & Fensel, D. (1998). Knowledge Engineering: Principles and methods. *Data & Knowledge Engineering* , 161-197.

Suárez-Figueroa, M. C., Gómez-Pérez, A., Motta, E., & Gangemi, A. (2012). *Ontology Engineering in a Networked World.* Springer.

Tapia, F., & Fuertes, W. (2014). Generación de Datos Semánticos a partir de una Base de Datos Relacional de una Institución de Educación Superior. *Twelfth LACCEI Latin American*

and Caribbean Conference for Engineering and Technology (LACCEI'2014). Guayaquil, Ecuador.

Villalba, R. (2007). *La Web semántica.* Trabajo Práctico, Universidad Católica "Nuestra Señora de la Asunción, Ingeniería Informática, Asunción, Paraguay.

Villazón-Terrazas, B., Vila-Suero, D., Garijo, D., Vilches, L., Poveda, M., Mora, J., y otros. (2012). Publishing Linked Data - There is no One-Size-Fits-All Formula.

Villazón-Terrazas, B., Vilches, L., Corcho, O., & Gómez, A. (2011). Methodological Guidelines for Publishing Goverment Linked Data.

W3C - Datos Enlazados. (s.f.). Obtenido de Guía Brede de Linked Data: http://www.w3c.es/Divulgacion/GuiasBreves/LinkedData

W3C - SPARQL. (s.f.). Obtenido de SPARQL Query Language for RDF: https://www.w3.org/TR/2006/CR-rdf-sparql-query-20060406/

W3C - Web Semántica. (s.f.). Obtenido de Guía Breve de Web Semántica: http://www.w3c.es/Divulgacion/GuiasBreves/WebSemantica

W3C - XML. (s.f.). Obtenido de Guía Breve de Tecnologías XML: http://www.w3c.es/Divulgacion/GuiasBreves/TecnologiasXML

Wikipedia. (s.f.). Obtenido de Web semántica: https://es.wikipedia.org/wiki/Web_semántica