



UNIVERSIDAD POLITÉCNICA DE PUEBLA

PROGRAMA ACADÉMICO DE POSGRADO

**Implementación de un servicio web para la  
recuperación de información semántica del  
repositorio institucional de la UPPue**

TESIS PARA OBTENER EL GRADO DE

MAESTRÍA EN INGENIERÍA EN SISTEMAS Y CÓMPUTO  
INTELIGENTE

PRESENTA:

**ING. PAULO DANIEL VÁZQUEZ MORA**

**Directora:** Dra. María Auxilio Medina Nieto

Juan C. Bonilla, Puebla, México. Diciembre 2019.

---



*El presente trabajo fue realizado en el Departamento de Posgrado de la Universidad  
Politécnica de Puebla, ubicada en Tercer Carril del Ejido "Serrano" S/N, San Mateo  
Cuanalá, Municipio Juan C. Bonilla, Puebla CP 72640. Beca de posgrado número 863914.  
Consejo Nacional de Ciencia y Tecnología.*



UNIVERSIDAD POLITÉCNICA DE PUEBLA  
MAESTRÍA EN INGENIERÍA EN SISTEMAS Y CÓMPUTO INTELIGENTE

# **Implementación de un servicio web para la recuperación de información semántica del repositorio institucional de la UPPue**

TESIS REALIZADA POR:

**ING. PAULO DANIEL VÁZQUEZ MORA**

12 de Diciembre del 2019

**Comité evaluador**

**(Firma)**

Dra. María Auxilio Medina Nieto .....

Dra. Mireya Vidal Tovar .....

M.C. Rebeca Rodríguez Huesca .....

M. C. Antonio Felipe Razo Rodríguez .....





UNIVERSIDAD POLITÉCNICA DE PUEBLA  
MAESTRÍA EN INGENERÍA EN SISTEMAS Y CÓMPUTO INTELIGENTE

Derechos de autor: Universidad Politécnica de Puebla 2019

El contenido se distribuye bajo los términos de la Licencia Abierta  
*Creative Commons* (CC BY-NC-ND 2.5 MX)

*(Firma)*

.....  
Ing. Paulo Daniel Vázquez  
Mora

<b>1 Planteamiento del problema</b>	<b>11</b>
1.1 Introducción . . . . .	11
1.2 Objetivo general . . . . .	13
1.3 Objetivos específicos . . . . .	13
1.4 Justificación . . . . .	13
<b>2 Marco teórico</b>	<b>15</b>
2.1 La importancia de los metadatos . . . . .	15
2.2 RDF en el contexto de la web semántica . . . . .	15
2.3 SPARQL: lenguaje de consulta para datos en RDF . . . . .	17
2.4 Servicios REST . . . . .	18
2.5 Características de los repositorios según COAR . . . . .	21
2.6 Trabajos relacionados . . . . .	23
2.6.1 Servicios de búsqueda simple, avanzada y semántica . . . . .	25
2.6.2 Plataformas que procesan consultas semánticas . . . . .	25
<b>3 Metodología</b>	<b>30</b>
3.1 Modelo de desarrollo . . . . .	30
3.2 Análisis de requerimientos . . . . .	30
3.3 Diseño . . . . .	31
3.3.1 Casos de uso . . . . .	31
3.3.2 Diagrama de clases . . . . .	31
3.3.3 Diseño de alto nivel . . . . .	32
3.4 Implementación . . . . .	33
3.4.1 Etapa 1: exportación de la colección de tesis . . . . .	34
3.4.2 Etapa 2: instalación de Apache Jena Fuseki . . . . .	34
3.4.3 Etapa 3: ejecución del servidor Jena Fuseki . . . . .	36
3.4.4 Etapa 4: ejecución del servicio RDFizer . . . . .	36
3.4.5 Etapa 5: creación del almacén de ternas . . . . .	37
3.4.6 Etapa 6: obtención de ternas . . . . .	37
3.4.7 Herramientas software . . . . .	38
3.4.7.1 Herramientas de la interfaz gráfica de usuario . . . . .	38
3.4.7.2 Servicio especializado en lenguaje Python 3.6 . . . . .	38
3.4.7.3 Librería RDFLib . . . . .	38
3.4.7.4 Librería OWLReady2 . . . . .	39
3.4.7.5 Librería xml.etree.ElementTree . . . . .	40
3.4.7.6 Framework Flask . . . . .	40
3.4.7.7 AdminLTE . . . . .	40
3.5 Evaluación . . . . .	40
3.6 Mantenimiento . . . . .	41

---

<b>4 Resultados</b>	<b>42</b>
4.1 Resultados . . . . .	42
4.1.1 Fase 1. Captura de metadatos en DSpace 6.2 . . . . .	42
4.1.2 Fase 2. Creación de almacén de ternas . . . . .	43
4.1.3 Fase 3. Generación de documentos en RDF . . . . .	44
4.1.4 Fase 4. Extracción y procesamiento de metadatos . . . . .	45
4.1.5 Fase 5. Lectura e inserción de instancias en la ontología Onto4UPPue	46
4.1.6 Fase 6. Implementación del servicio web . . . . .	46
4.1.7 Fases 7 y 8: evaluación y mantenimiento . . . . .	50
4.1.7.1 Verificación del proceso de integración de instancias . . . . .	50
4.1.8 Evaluación de la experiencia del usuario . . . . .	52
4.1.9 Mantenimiento del servicio web . . . . .	55
<b>Referencias</b>	<b>57</b>

## Lista de figuras

2.1	Modelo de capas de Berners-Lee para la web semántica[18] . . . . .	16
2.2	Ejemplo de representación gráfica provenientes de objetos vCard en RDF / XML . . . . .	18
2.3	Ejemplo de conjunto de ternas en RDF . . . . .	19
2.4	Arquitectura cliente servidor de servicio REST . . . . .	19
2.5	Servicios sin estado de la arquitectura cliente servidor REST . . . . .	19
2.6	Servicios REST apoyados en caché . . . . .	20
2.7	Servicios REST uniformes . . . . .	20
2.8	Arquitectura de capas de los servicios REST . . . . .	21
2.9	Resultado en formato JSON del servicio áreas de conocimiento del RN . . . . .	22
2.10	Arquitectura de servicio web de búsqueda semántica SW001. . . . .	23
2.11	Estadísticas generales reportadas en LA Referencia . . . . .	24
2.12	Repositorios por país en OpenDOAR . . . . .	24
2.13	Repositorios por lenguaje y tipo de software en OpenDOAR . . . . .	25
2.14	Repositorios con mayor producción en Latinoamérica . . . . .	25
2.15	Vista del portal del Sistema Nacional de Repositorios Digitales . . . . .	26
2.16	Vista del portal brasileño de publicaciones científicas y acceso abierto - oasisbr . . . . .	26
2.17	Vista del panel de búsquedas en RI2.0 . . . . .	27
2.18	Vista del panel de búsquedas avanzadas en RI2.0 . . . . .	27
2.19	Vista del Sistema Nacional de Acceso Abierto al Conocimiento . . . . .	27
2.20	Vista del portal de acceso libre a información científica para la innovación . . . . .	28
2.21	Vista del panel de búsqueda en el portal Sistema Nacional de Repositorios Digitales en Argentina . . . . .	28
2.22	Vista del panel de búsqueda en el portal instituto Brasileño de Información en Ciencia y Tecnología en Brasil . . . . .	29
2.23	Servicios de búsqueda en RDs de AA . . . . .	29
2.24	Ejemplo de conjunto de ternas en RDF . . . . .	29
3.1	Modelo de desarrollo del servicio web SW001 . . . . .	30
3.2	Diagrama de casos de uso . . . . .	31
3.3	Diagrama de clases. Fuente: [19] . . . . .	32
3.4	Diseño de consumo del servicio web. Fuente: [19] . . . . .	32
3.5	Diseño de alto nivel del servicio web. Fuente: [19] . . . . .	33
3.6	Módulos adicionales de DSpace para la versión 6.2 . . . . .	33
3.7	Etapas del proceso de migración de metadatos del RI-UPPue a la web semántica . . . . .	34
3.8	Exportación de la colección tesis a archivo CSV . . . . .	35
3.9	Opciones de descarga para el almacén de ternas Jena Fuseki . . . . .	35
3.10	Contenido de la carpeta del servidor <i>Jena Fuseki</i> . . . . .	36

3.11	Ejecución del servidor Jena Fuseki en terminal de Windows . . . . .	36
3.12	Extracción, transformación y carga . . . . .	37
3.13	Acceso a metadatos de tesis en RDF . . . . .	37
3.14	Metadatos de RDF para una tesis del almacén de ternas . . . . .	38
3.15	Arquitectura del servicio web . . . . .	39
3.16	Tecnologías empleadas en el servicio web . . . . .	39
3.17	Reactivos para evaluar la heurística generalidades [40] . . . . .	41
4.1	Esquema general de fases para la obtención de resultados . . . . .	42
4.2	Creación de la comunidad de <i>tesis</i> . . . . .	43
4.3	Creación de la colección para tesis de <i>maestría</i> . . . . .	43
4.4	Ítems agregados a la colección de <i>tesis</i> . . . . .	44
4.5	Implementación del almacén de ternas empleando el servidor Apache Jena Fuseki versión 1.6 . . . . .	44
4.6	Ejecución local del servidor Apache Jena Fuseki . . . . .	45
4.7	Ejecución del serializador <i>RDFizer</i> . . . . .	45
4.8	Descripción de la clase para la extracción de metadatos desde archivo RDF . . . . .	46
4.9	Metadatos recuperados para una tesis del almacén de ternas . . . . .	47
4.10	Método de inserción de instancias . . . . .	47
4.11	Estructura de una instancia agregada por el servicio . . . . .	48
4.12	Estructura de una instancia agregada por el servicio . . . . .	48
4.13	Sección de inicio del servicio web SW001 . . . . .	49
4.14	Servicio de búsqueda semántica . . . . .	49
4.15	Resultados de una de búsqueda semántica . . . . .	50
4.16	Descarga de la ontología ONTO4-UPPUE en formato JSON u OWL . . . . .	50
4.17	Sección de preguntas frecuentes . . . . .	51
4.18	Sección de contacto . . . . .	51
4.19	Interfaz del servicio web SW001 desde un dispositivo móvil . . . . .	51
4.20	Vista de la instancia T24 agregada por el servicio web SW001. . . . .	52
4.21	Pantalla principal del TestUX . . . . .	53
4.22	Resultados de la evaluación del servicio web . . . . .	54
4.23	Resultado general de la evaluación del servicio web . . . . .	54
4.24	Usuarios realizando la evaluación del servicio web . . . . .	55
4.25	Iconos para ayuda contextual . . . . .	55
4.26	Mensaje para la descripción detallada de elementos y ejemplos de uso . . . . .	56
4.27	Mensajes contextuales sobre los servicios ofrecidos en la barra de navegación . . . . .	56

## Lista de tablas

2.1	Catálogo de servicios REST ofrecidos por el RN . . . . .	21
2.2	Características técnicas entre DSpace, Virtuoso y VIVO . . . . .	28
3.1	Tabla de requerimientos funcionales del servicio web . . . . .	31
3.2	Escala de Likert empleada para evaluar la interfaz gráfica del servicio web	41
4.1	Actividades de verificación con instancias migradas al almacén de ternas .	45
4.2	Actividades de verificación con archivos en RDF . . . . .	46
4.3	Actividades de verificación con archivos CSV . . . . .	47
4.4	Verificación de instancias agregadas a Onto4UPPue . . . . .	52

## 1.1 Introducción

El acceso abierto (en adelante AA) es digital, en línea, libre de cargo y libre de la mayoría de restricciones de derechos de autor y licencias; elimina las barreras de precios por suscripciones, cuotas o pago de licencias [42]. Existen diversos tratados y convenciones internacionales que promueven el AA del público en general a la literatura científica y académica de manera digital representada en revistas, tesis, artículos, memorias de eventos, libros, entre otros recursos digitales. En inglés, las siglas para el AA son OA de *Open Access*.

La propuesta del AA se planteó en la Iniciativa de Budapest en 2002, durante un evento organizado por el Instituto de la Sociedad Abierta (*Open Society Institute*); en 2003, se llevó a cabo la Declaración de Berlín sobre el AA al conocimiento en las ciencias y las humanidades [48]. Para [3], el AA ofrece a las Instituciones de Educación Superior (IES) y Centros de Investigación (CIs) beneficios como los siguientes:

- Rendición de cuentas transparente ante la sociedad con respecto a la inversión pública
- Incremento en la difusión e impacto de la producción científica
- Fomento de la producción científica y académica, aumenta las posibilidades de acceso
- Facilita el intercambio de información entre IES, CIs, comunidades locales, nacionales e internacionales
- Garantiza la preservación electrónica de recursos documentales digitales

Entre los beneficiarios del AA están los autores, las comunidades científicas, académicas y usuarios de *repositorios institucionales* (RIs). Los RIs son plataformas tecnológicas diseñadas para almacenar, preservar y difundir documentos digitales, son un medio de divulgación (*ruta verde*) o de publicación (*ruta dorada*) de los contenidos producidos por una institución o comunidad [46].

Según [2], un repositorio institucional (RI) es un conjunto de servicios prestados por las universidades y organismos de investigación a la comunidad para recopilar, administrar, difundir y preservar la producción documental digital, cualquiera que sea su tipología, a través de la creación de una colección digital organizada, abierta e interoperable que emplea el protocolo de la Iniciativa de Archivos Abiertos para la disseminación de metadatos, *Open Archives Initiative Protocol for Metadata Harvesting*, (OAI-PMH) [43], con el fin de garantizar un aumento en la visibilidad e impacto de la propia institución.

En Diciembre de 2018, según el sitio OpenDOAR [4], directorio de repositorios de AA de la Universidad Nottingham, existían 3,779 repositorios alrededor del mundo, de los cuales (46%) se encuentran en Europa, el 27% en Norteamérica y sólo el 0.9% en México. En México, al 9 de Noviembre del 2019, la página web del Repositorio Nacional (RN) [13]

reporta la existencia de 105 repositorios de Ciencia Abierta (CA) e INDEXE, Sistema de Búsqueda de la Red Mexicana de Repositorios Institucionales (REMERI) [12] indica 100 RIs comprometidos con la divulgación de sus contenidos institucionales y temáticos. De acuerdo con datos de la red REMERI, más del 45% de RIs pertenecen a instituciones públicas, siendo la Universidad Nacional Autónoma de México (UNAM), la institución con mayor número de repositorios y documentos publicados [12].

Según datos de OpenDOAR [4], la distribución del *software* empleado para la implementación de los 3,779 RIs es como sigue: *DSpace* [11] (44.2%), *EPrints* (13.4%), *Digital Commons* (4.7%) y *WEKO* (2.7%). A diferencia de *Digital Commons* que es un *software* licenciado por la empresa *Bepress*, *WEKO*, *EPrints* y *DSpace* cuentan con licencia libre, por lo que su uso se ha extendido en múltiples IES, CIs y otras organizaciones.

Por un lado, *EPrints* [14] se desarrolló de la Universidad de Southampton en el Reino Unido del 2010, esta plataforma soporta la preservación, diseminación y generación de reportes para instituciones que requieran que servicios de AA, también permite la construcción de repositorios de educación abierta (en inglés, *open education*) y bancos de datos para investigación (*research data*). *EPrints* se emplea como medio de integración con redes sociales. Por otro lado, *DSpace* [11] surgió como proyecto desarrollado en sus inicios por el Instituto Tecnológico de Massachusetts (MIT) en el año 2002 en conjunto con los Laboratorios HP. Actualmente, se mantiene en la fundación *DuraSpace* que entre sus objetivos se encuentran la innovación en tecnologías de AA y basadas en nube, principalmente para bibliotecas, universidades, CIs y organizaciones de patrimonio cultural. *DSpace* soporta el almacenamiento de tesis, administración de registros electrónicos, preservación digital y publicación. Los RIs que interoperan con el RN [13] emplean en su mayoría (más del 90%) una versión de *DSpace*. Una comparación que considera aspectos de funcionalidad, elementos técnicos y de administración entre las plataformas *EPrints* y *DSpace* se presenta en [49].

En *EPrints* y *DSpace* la interoperabilidad por omisión se implementa al utilizar el protocolo OAI-PMH [43] y el estándar de metadatos Dublin Core [44]. En el RI de la Universidad Politécnica de Puebla, en adelante RI-UPPue, está soportado en la versión 6.2 de *DSpace*. En esta tesis se identifica como problemática para la comunidad universitaria lo siguiente:

- Los mecanismos de recuperación de información disponibles desde la interfaz del RI-UPPue sólo distinguen entre el autor principal y los coautores de los documentos; a la fecha, no se puede determinar un rol de los coautores como asesor o sinodal si se trata de una tesis, o autor principal, segundo o tercer autor si se trata de un artículo
- Existe *ambigüedad* en la interpretación de los datos descriptivos o *metadatos* cuando se depositan documentos en el RI-UPPue, por ejemplo, qué se debe colocar dentro del elemento *contributor* al depositar una tesis de maestría
- A pesar de que otras IES que interoperan con el RN tal como el RI-UPPue y de que emplean la plataforma *DSpace*, la interpretación de los datos exportados está

sujeta a los usuarios finales, dado que ésta se realiza por omisión únicamente en el formato CSV<sup>1</sup>.

Previamente, en la UPPue se desarrolló un modelo semántico u ontología denominada *Onto4AIR*, es un producto de software que representa formalmente conocimiento de dominio y operativo de los RIs, hace énfasis en los documentos, los tipos de usuario y sus relaciones [6]. La tesis propone utilizar una instancia de esta ontología, identificada como *onto4UPPue* y tecnologías semánticas como alternativas para atender la problemática anterior, de manera que los objetivos son:

## 1.2 Objetivo general

Extender los mecanismos de búsqueda de tesis del RI-UPPue utilizando tecnologías semánticas para atender necesidades de información de usuarios del RI-UPPue

## 1.3 Objetivos específicos

- Describir la funcionalidad del componente RDF de la plataforma DSpace 6.2
- Integrar datos de tesis de maestría del RI-UPPue a instancias de la ontología *Onto4UPPue* validando su consistencia lógica de forma automática mediante razonadores
- Diseñar e implementar un servicio web tipo REST que permita recuperar datos utilizando instancias de la ontología *Onto4UPPue*

## 1.4 Justificación

En los últimos años, diferentes IES, CIs y comunidades se han sumado el esfuerzo de promover el AA a sus contenidos científicos mediante RIs que en su mayoría emplea modelos de datos relacionales para almacenar información. De acuerdo con [16] y [17], el uso de tecnologías semánticas como las ontologías se caracteriza por lo siguiente:

- Intercambio de datos entre aplicaciones, programas y/o plataformas mediante el lenguaje XML<sup>2</sup> o lenguajes derivados de éste
- Las ontologías permiten gestionar datos incompletos y reutilizar el conocimiento
- Los datos en las ontologías forman conjuntos de datos que son procesables por computadora

Los beneficios esperados del desarrollo de la tesis son:

- Recopilación y divulgación de datos de tesis del RI-UPPue en formatos de la web semántica

<sup>1</sup>CSV corresponde a las siglas de *Comma Separated Values*

<sup>2</sup>XML son las siglas de *eXtensible Markup Language*

- Uso de la ontología Onto4UPPue como alternativa de integración de información semántica a datos provenientes del RI-UPPue
- Extensión de los mecanismos de búsqueda y recuperación del RI-UPPue, recuperación desde un punto de vista conceptual y no sólo textual

El capítulo 2 se organiza como sigue: en primer lugar trata la importancia de los metadatos, considerados como datos descriptivos para los documentos de los repositorios. Posteriormente, se presenta el modelo de datos RDF y su contexto en la web semántica, seguido de una breve descripción del lenguaje de consulta estándar para datos en RDF, SPARQL. En la tesis, el acceso a los datos se implementa en servicios REST, por lo que se describen sus características principales junto con los requerimientos para repositorios de acuerdo con la *Confederation of Open Access Repositories*, Confederación de Repositorios de Acceso Abierto (COAR). Finalmente, se presentan los trabajos relacionados.

## 2.1 La importancia de los metadatos

Actualmente, en la web existe gran cantidad de información sobre cualquier tema, tanta que se requiere del desarrollo de servicios que consideren su pertinencia, veracidad y calidad para satisfacer necesidades de información específicas de los usuarios tomando en cuenta aspectos técnicos de accesibilidad y disponibilidad, esto porque el acceso se realiza desde cualquier lugar y mediante una gama amplia de dispositivos.

El etiquetado y descripción de los contenidos son cruciales en el desarrollo de ese tipo de servicios, el primero permite categorizar o clasificar, el segundo se refiere al uso de las descripciones, elementos descriptores o *metadatos* para que los recursos digitales se localicen y procesen adecuadamente por agentes tales como computadoras, aplicaciones móviles y usuarios específicos.

Según [20], una prioridad particular del *World Wide Web Consortium*, Consorcio W3C, es usar la web para documentar el significado de los metadatos. La importancia de la gestión, uso y representación de los metadatos en modelos semánticos u ontologías, se relaciona directamente con el modelo de datos conocido como Marco de Descripción de Recursos, en inglés *Resource Description Framework*, (RDF). La sección 2.2 describe características de este modelo junto con otras tecnologías de la web semántica.

## 2.2 RDF en el contexto de la web semántica

La web es una plataforma tecnológica que constituye la mayor base de datos existente, se conforma por todo tipo de recursos, donde las personas o usuarios realizan tareas como publicar, explorar, consultar, almacenar datos e información. En sus inicios, en la web se consideraba que los recursos debieran ser entendidos sólo por los usuarios, con el paso del tiempo y la gran cantidad de datos que están en ella, es de suma importancia su gestión y procesamiento mediado por las computadoras.

La web semántica, extensión de la web tradicional, promueve el modelado, etiquetado y representación de la información de manera que tanto los humanos como las computadoras sean capaces de “comprender” el contenido y la descripción de los recursos; como menciona [24], esta web propone una alternativa para agregar significado a los contenidos

almacenados con el propósito de propiciar una interacción más fluida entre aplicaciones, servicios, computadoras y el ser humano en comparación con la web tradicional.

Las tecnologías de la web semántica se rigen bajo ciertas normas y lenguajes estándar organizados en capas o niveles como muestra la Figura 2.1.

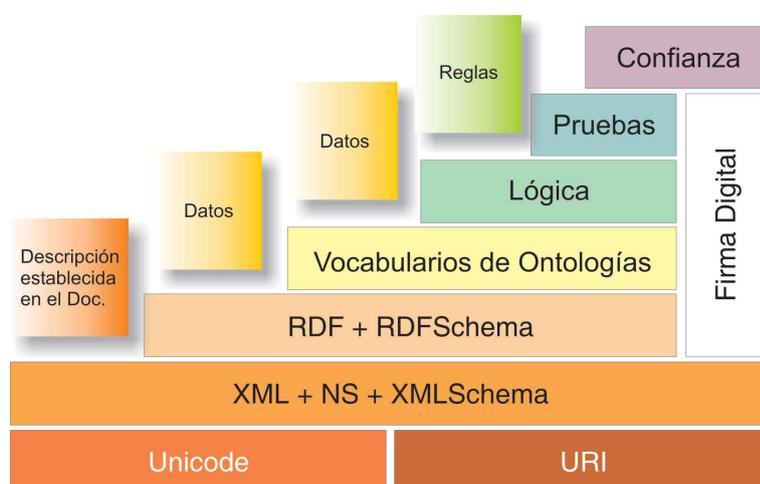


Figura 2.1: Modelo de capas de Berners-Lee para la web semántica[18]

Una descripción amplia y detallada de los primeros niveles está disponible en [20] y [24]. En forma breve, los niveles inferiores de la Figura 2.1 tratan de lo siguiente:

- *Unicode*. Estándar para documentos de texto que permite codificar la mayoría de los sistemas de escritura del mundo.
- *Uniform Resource Identifier*, (URI), identificador uniforme de recursos. Estándar para crear identificadores de recursos web a través de cadenas compactas de caracteres, identifican a los recursos de forma unívoca, se emplean para localizarlos de forma automática. Un *Uniform Resource Locator*, (URL), identificador de recursos uniforme, es un tipo de URI, por ejemplo, <http://informatica.up Puebla.edu.mx/> identifica específicamente a la página inicial de un servidor.
- *XML*<sup>1</sup>. Es un lenguaje de marcado similar a HTML, es una especificación del W3C de propósito general en la que cada usuario define sus propias etiquetas.
- *XML namespaces*. Los espacios de nombre de XML se emplean para atender el problema de ambigüedad en los documentos XML, de manera que cada elemento está identificado por una URI que lo hace único y universal. Los espacios de nombre son recomendación del W3C.

De acuerdo con [22] y [35], una descripción para los niveles medios y superiores de la Figura 2.1 es la siguiente:

- *RDF*. Modelo estándar para el intercambio de datos en la web, con características que facilitan la fusión de datos incluso en diferentes esquemas. *RDF* amplía la

<sup>1</sup>*eXtensible Markup language* (Lenguaje de Marcado Extensible o Lenguaje de Marcas Extensible)

estructura de enlaces web (*links*) para nombrar a dos recursos denominados *sujeto* y *objeto*, así como a la relación entre ellos (*predicado*). Estos enunciados se conocen como *ternas* o *trípletas*. En el modelo de datos RDF se representan datos estructurados y semiestructurados, los cuales se comparten y utilizan por diferentes aplicaciones.

- *RDF Schema*. Proporciona un vocabulario para el modelo de datos RDF, describe los conceptos básicos y la sintaxis abstracta de RDF; es una extensión semántica de RDF, se utiliza para determinar las características de otros recursos como dominio y rango de propiedades.
- *Ontología*. Según [45], una ontología es un marco común o una estructura conceptual sistematizada y de consenso no sólo para almacenar la información, sino también para buscarla y recuperarla. Una ontología define los términos y las relaciones básicas para la comprensión de un área del conocimiento o dominio, así como las reglas para combinar los términos que definen las extensiones de su vocabulario
- *Vocabularios de ontologías*. El lenguaje de ontologías web, (*Ontologies Web Language*), está diseñado para ser usado en aplicaciones que necesitan procesar el contenido de la información, cuenta con un vocabulario más extenso al de RDF y RDF Schema, junto con una semántica formal. OWL tiene tres sublenguajes que varían por su nivel de expresividad: OWL Lite, OWL DL y OWL Full
- *Capa lógica*. Permite determinar si la estructura de los razonamientos es válida a través del estudio de las reglas formales. En esta capa se infiere conocimiento, requiere de la interacción entre las ontologías y agentes de software, (programas o aplicaciones)
- *Capa de pruebas*. A través de demostraciones matemáticas se comprueba que el procesamiento del agente de software alcance la máxima confiabilidad en sus razonamientos
- *Capa de confianza*. Establece las políticas de seguridad que permitan asignar niveles de fiabilidad a determinados recursos, de forma comprobable por agentes. Esta capa usa firmas digitales y redes de confianza

Los lenguajes utilizados ampliamente para representar los datos en la web semántica son XML, RDF y OWL. El lenguaje de consulta para datos en RDF se describe en la sección 2.3.

## 2.3 SPARQL: lenguaje de consulta para datos en RDF

RDF es un modelo de datos que se asocia con diferentes representaciones, una de ellas es como un modelo de grafos dirigidos etiquetados que representan información en la web. Se emplea para representar información personal, datos de redes sociales,

metadatos sobre objetos digitales o como medio para la integración de fuentes de información heterogéneas.

Los datos en RDF se recuperan utilizando SPARQL, *SPARQL protocol and RDF query language*, protocolo y lenguaje de consulta para RDF, sirve para extraer la información contenida en una ontología RDF, devuelve resultados en forma de enlaces o esquemas RDF. Entre sus especificaciones, de acuerdo con [32], se encuentran las siguientes:

- La especificación del protocolo SPARQL para RDF *SPROT* que define el protocolo remoto para enviar consultas SPARQL y recibir los resultados
- La especificación del formato XML de los resultados de consultas SPARQL *RESULTS*, define un formato de documento XML para representar los resultados de las consultas SELECT y ASK

La Figura 2.2 muestra una representación gráfica simple de *vc-db-1.rdf*, archivo que contiene RDF para varias descripciones de vCard de personas, descritas en las notas del W3C “Representación de objetos vCard en RDF / XML”.

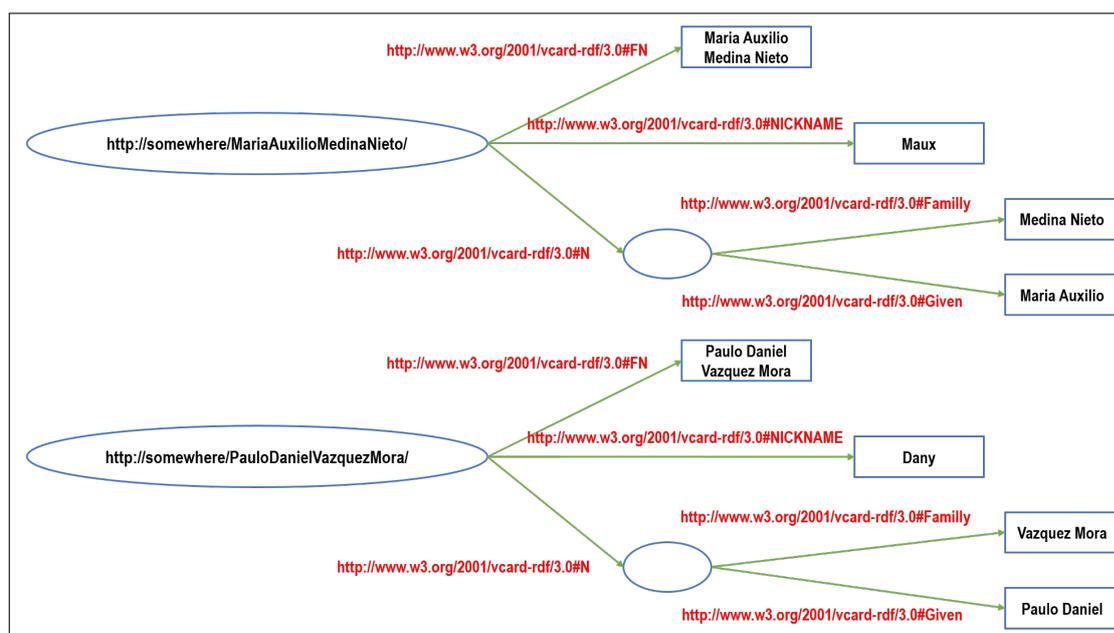


Figura 2.2: Ejemplo de representación gráfica provenientes de objetos vCard en RDF / XML

La Figura 2.3 muestra el esquema RDF expresado como un conjunto de ternas.

## 2.4 Servicios REST

Los servicios *Representational State Transfer*, (REST), se emplean para explotar recursos a partir de la migración a la web 2.0, sustituyen a los servicios que hacían uso de los protocolos SOAP y WSDL. La arquitectura de los servicios REST es sencilla y está orientada a recursos, hacen uso del protocolo HTTP y cumplen con las premisas siguientes [30]:

```

@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix xsd: <http://www.w3.org/2001/XMLSchema#> .
@prefix dcterms: <http://purl.org/dc/terms/> .
@prefix bibo: <http://purl.org/ontology/bibo/> .
@prefix foaf: <http://xmlns.com/foaf/0.1/> .
@prefix dspace: <http://digital-repositories.org/ontologies/dspace/0.1.0#> .
@prefix dc: <http://purl.org/dc/elements/1.1/> .

<http://34.236.228.85/rdf/resource/123456789/8400>
  dspace:isPartOfCollection <http://34.236.228.85/rdf/resource/123456789/7882> ;
  dc:contributor
    "Gutiérrez Álvarez, Brenda Alma" ;
  dc:creator
    "Gutiérrez Álvarez, Brenda Alma" ;
  dc:date
    "2018-05-08T04:54:00Z"^^xsd:dateTime ;
  dc:language
    "spa" ;
  dc:publisher
    "Universidad de las Américas Puebla" ;
  dc:rights
    "Acceso Abierto" ;
  dcterms:available
    "2018-05-08T04:54:00Z"^^xsd:dateTime ;
  dcterms:title
    "Proyecto de Inversión de una Planta de Extracción y Empaque de Aceite de
Aguacate para el Rancho Villa Reguero" ;

```

Figura 2.3: Ejemplo de conjunto de temas en RDF

- Definen una *interfaz* de comunicación *cliente-servidor* que separa las responsabilidades entre ambas partes, ver la Figura 2.4.



Figura 2.4: Arquitectura cliente servidor de servicio REST

- Trabajan con peticiones *sin estado*, esto es, cada petición que se realiza es completamente independiente de cualquier otra, sin embargo, todas las solicitudes al mismo servicio son idénticas, ver la Figura 2.5.

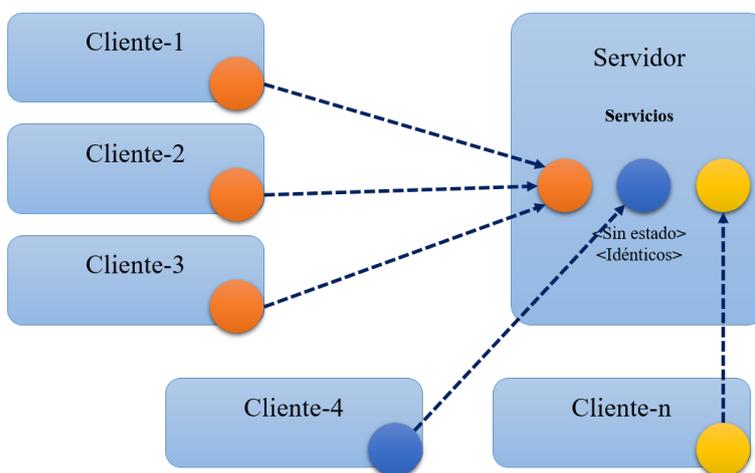


Figura 2.5: Servicios sin estado de la arquitectura cliente servidor REST

- Guardan en caché su contenido de tal manera que una vez realizada la primera petición, el resto de peticiones puedan apoyarse en la caché si fuera necesario, ver la Figura 2.6.
- Son *uniformes*, es decir, comparten una forma de invocación y métodos GET, POST, PUT, DELETE, ver la Figura 2.7.

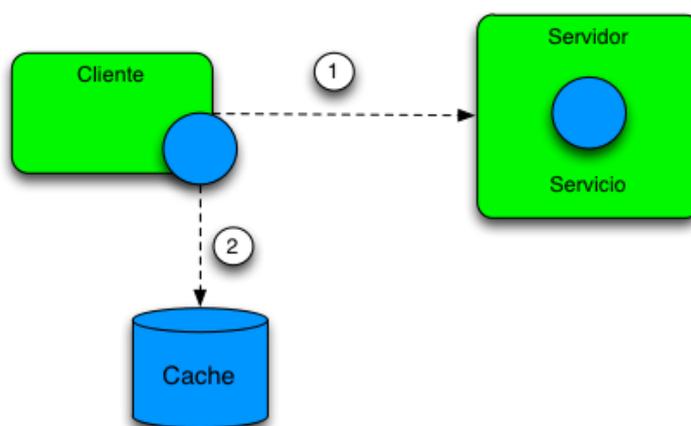


Figura 2.6: Servicios REST apoyados en caché

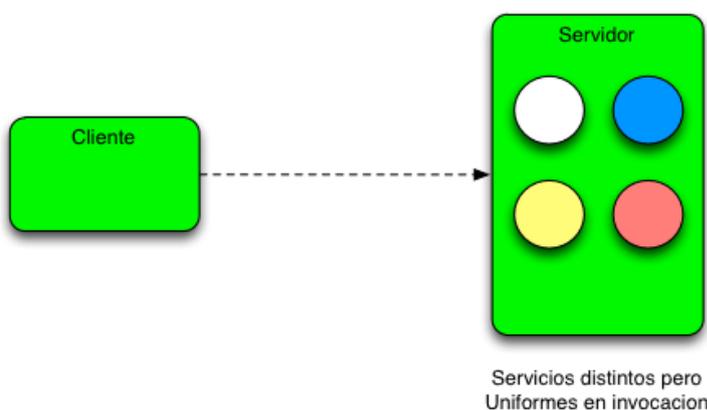


Figura 2.7: Servicios REST uniformes

- Son escalables y para el cliente transparentes, es decir, el cliente no puede distinguir si la petición es atendida directamente por el servidor o por el sistema de cachés debido al servicio de balanceo de cargas, ver la Figura 2.8.

Si bien algunos repositorios institucionales (RIs) cuentan con servicios tipo REST, el grado de complejidad en las consultas que hasta el momento están implementadas emplean sólo un atributo o características de los documentos y que las búsquedas se basan en palabras clave, lo cual restringe el tipo de información que puede recuperarse. Por ejemplo, el RN tiene un catálogo de más de doscientos servicios REST que regresan los resultados en formato *JavaScript Object Notation* (JSON) [50], los principales se incluyen en la Tabla 2.1. Dependiendo del catálogo, los servicios se consideran básicos o específicos. El resultado de los servicios se presenta en formato JSON, como muestra la Figura 2.9.

En general, el uso del protocolo REST permite el intercambio y manipulación de datos a través de Internet, lo cual contribuye al desarrollo de aplicaciones y servicios que gestionan datos provenientes de diversos orígenes. Por lo tanto, los servicios REST funcionan como interfaces entre los RIs que use HTTP para obtener datos o generar operaciones

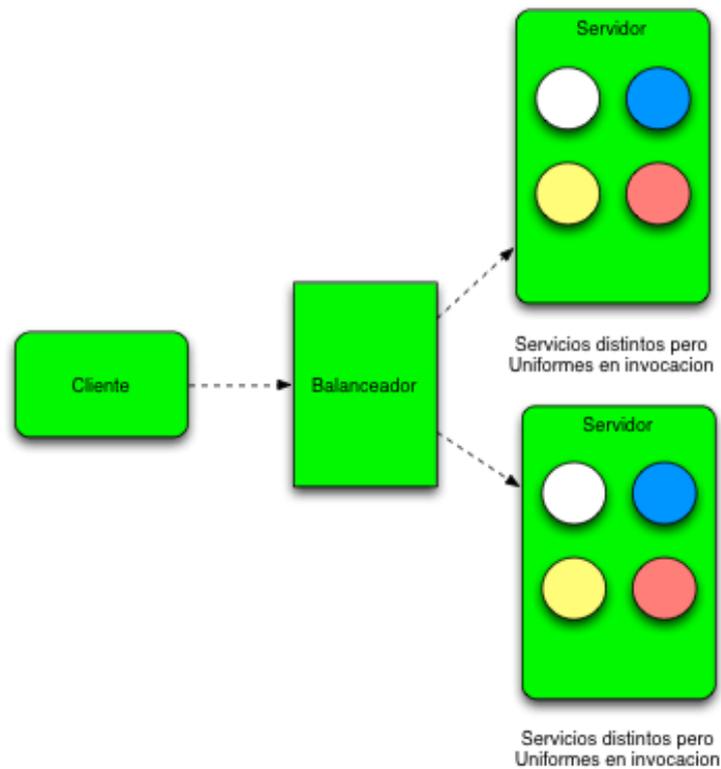


Figura 2.8: *Arquitectura de capas de los servicios REST*

sobre esos datos.

## 2.5 Características de los repositorios según COAR

Junto con el desarrollo de servicios especializados, para las IES y CI que distribuyen documentos en AA, diversos medios han apoyado la difusión de la investigación y de la producción académica que se caracterizan por recuperar documentos a partir de palabras clave. En 2016, la Confederación de Repositorios de Acceso Abierto, *Confederation of Open Access Repositories* (COAR) [34], integró el equipo de trabajo *Next Generation Repository Working Group* para caracterizar a la siguiente generación de repositorios, cuyo propósito

Tabla 2.1: *Catálogo de servicios REST ofrecidos por el RN*

Áreas de conocimiento	Licencia
Campos de conocimiento	Localidad
Disciplinas de conocimiento	Municipio
Subdisciplinas de conocimiento	Nivel de acceso
Audiencia	País
Estado	Persona
Institución	Formato
Idioma	Programa
Financiador-Programa	Proyecto

```

JSON  Datos sin procesar  Encabezados
Guardar Copiar Contraer todo Expandir todo Filtro JSON
0: { ... }
1: { ... }
2: { ... }
3: { ... }
4:
  cveArea: "4"
  descripcion: "HUMANIDADES Y CIENCIAS DE LA CONDUCTA"
  description: "[Por definir CONACyT]"
  idArea: 4
5:
  cveArea: "7"
  descripcion: "INGENIERÍA Y TECNOLOGÍA"
  description: "[Por definir CONACyT]"
  idArea: 7
6:
  cveArea: "3"
  descripcion: "MEDICINA Y CIENCIAS DE LA SALUD"
  description: "[Por definir CONACyT]"
  idArea: 3
    
```

Figura 2.9: Resultado en formato JSON del servicio áreas de conocimiento del RN

fue definir las funcionalidades y tecnologías a desarrollar en los RIs y RDs en los próximos años.

El informe *Behaviours and Technical Recommendations of the COAR Next Generation Repositories* [34] indica que será necesario adoptar nuevas tecnologías, estándares y protocolos que permitan una mejor integración de los repositorios en los entornos web, de esta manera, los RDs jugarán un papel importante en el amplio campo de la comunicación académica. El informe plantea las siguientes características para que los RDs sean fuentes sólidas de publicación, confiables y de AA [34]:

- Exponer *identificadores*
- Declaración de *licencias* a nivel de recursos
- Descubrimientos a través de la *navegación*
- Interactuar con *recursos*
- Descubrimiento de *lotes*
- *Transferencia* de recursos
- *Metadatos* de la actividad de recopilación y exposición de la información
- *Identificación* del usuario
- *Autenticación* del usuario
- Exponer *métricas* de uso estandarizadas

- *Conservación de los recursos*

La visión general del informe señala que se deben establecer comunidades de investigadores y académicos que, por un lado, aporten al acervo de los repositorios, por otro lado, que al mismo tiempo incentiven la producción en su comunidad o en otras comunidades interesadas en participar, de tal manera que los repositorios brinden una visión global de la investigación científica y académica.

En esta tesis, se propone la implementación de un servicio web tipo REST conforme a la arquitectura que muestra la Figura 2.10 como alternativa tecnológica acorde con algunos de los elementos del informe y que extienda los mecanismos de búsqueda de tesis del RI-UPPue utilizando tecnologías semánticas para atender necesidades de información de usuarios del RI-UPPue. El servicio se identifica con la etiqueta SW001, siglas de *servicio web 001*.

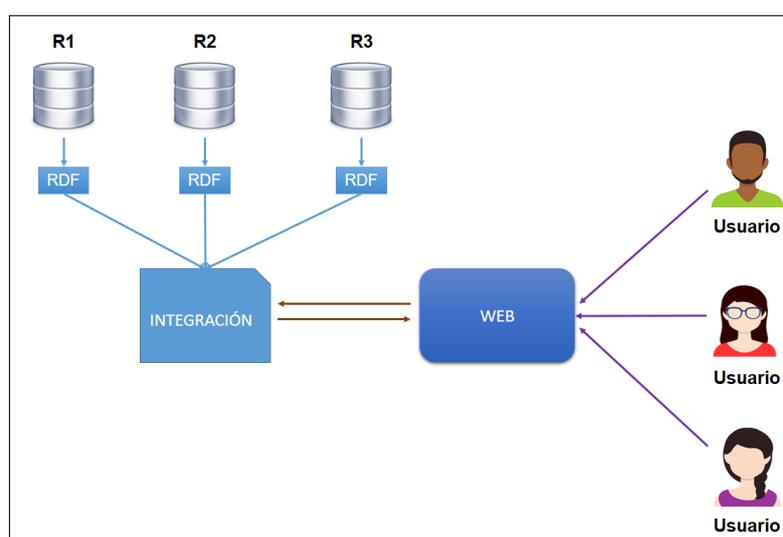


Figura 2.10: *Arquitectura de servicio web de búsqueda semántica SW001.*

## 2.6 Trabajos relacionados

La sección de trabajos relacionados se integra de la revisión documental y de sitios web que abordan temáticas como repositorios, acceso abierto y tecnologías semánticas. Por ejemplo, [21] plantea la transferencia de los resultados de investigación entre universidades o instituciones que, aunque los preservan, generan y transmiten, la transferencia se requiere sea lo más eficiente y efectiva posible, produciendo resultados favorables para los involucrados. Al mismo tiempo, se señala la importancia de la representación de un dominio adecuado de la gestión de la información científica definido mediante una ontología sobre un software licenciado para la gestión de contenidos, *Content Management System*, sistema de administración de contenidos (CMS) y contenidos enriquecidos semánticamente.

LA-Referencia<sup>2</sup>, Red de Repositorios de Acceso Abierto a la Ciencia [28], brinda un es-

<sup>2</sup>Disponible en: <http://www.lareferencia.info/es/>

pacio para la divulgación de la ciencia en Latinoamérica, incluye en su catálogo repositorios de países como Argentina, Brasil, Chile, Colombia, Costa Rica, Ecuador, El Salvador, México y Perú. En los contenidos, los usuarios pueden consultar nodos nacionales, documentos varios, artículos, reportes, tesis de maestría y doctorado. La Figura 2.11 muestra estadísticas generales de LA-Referencia.



Figura 2.11: Estadísticas generales reportadas en LA Referencia

OpenDOAR<sup>3</sup> es un directorio global de repositorios de acceso abierto académico, permite la identificación, navegación y búsqueda en función de una serie de características de los repositorios como la ubicación, el software o el tipo de material que se posee. Al mes de Octubre de 2019, OpenDOAR reporta las estadísticas de crecimiento por tipo de plataformas tecnológicas empleadas por los repositorios de acceso abierto (RAAs) que se muestran en la Figura 2.12.

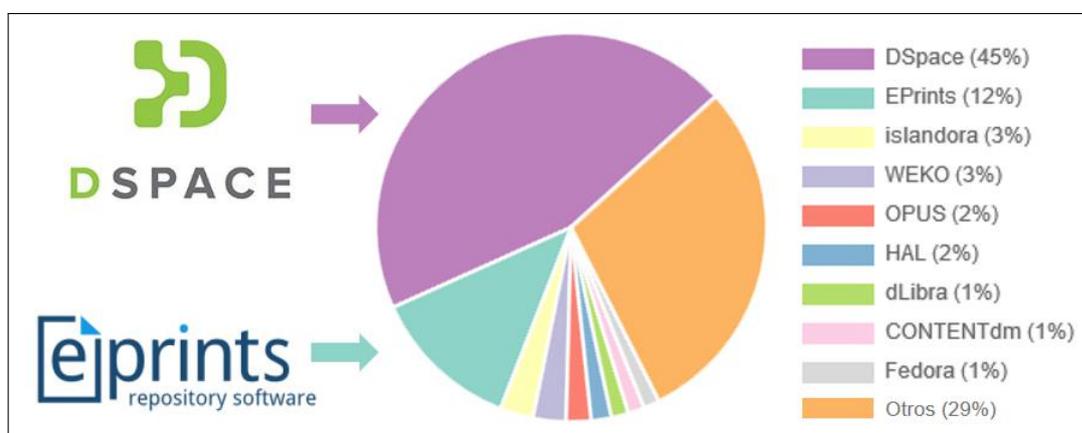


Figura 2.12: Repositorios por país en OpenDOAR

En Latinoamérica, LA Referencia reporta la producción de recursos de AA que muestra la Figura 2.14 y el tipo de servicios de búsqueda disponibles para los usuarios, notar que los atributos recurrentes para las búsquedas básicas son: título y autor, mientras que en el caso de las avanzadas, el número de atributos varía entre cuatro y doce. A la fecha de elaboración de esta tesis, los repositorios enumerados en la Figura 2.14 y que pertenecen a LA-Referencia, no reportan la implementación de mecanismos de búsqueda semántica.

Las figuras 2.15 a 2.20 muestran las páginas iniciales de los repositorios consultados y secciones de la interfaz correspondientes a sus módulos de búsqueda.

<sup>3</sup>Disponible en: <http://v2.sherpa.ac.uk/opensoar/>

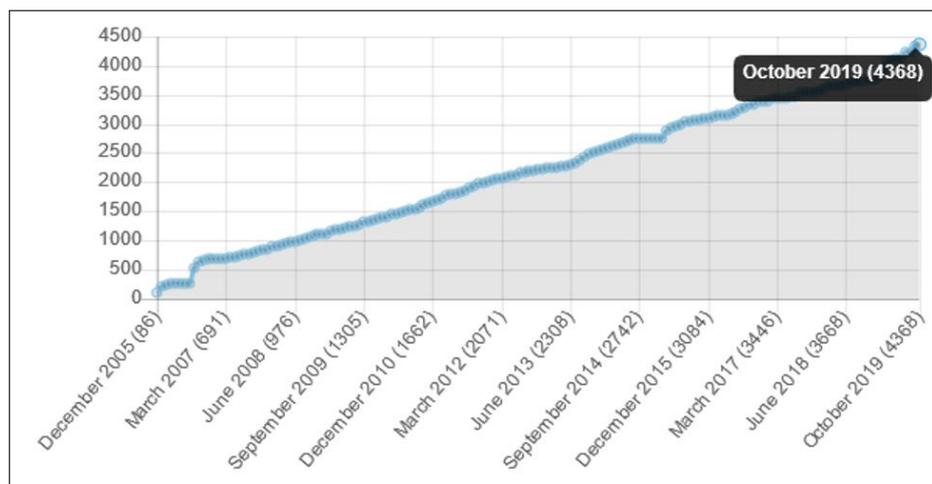


Figura 2.13: Repositorios por lenguaje y tipo de software en OpenDOAR

Organización / Red	I	R	D
Sistema Nacional de Repositorios Digitales (ARG)	30	31	176,927
Repositorio CONICYT (CHI)	5,077	-	99,026
Portal Brasileño de Publicaciones Científicas y A. A. (BRA)	1,000	-	2,405,456
Sistema Nacional de Acceso Abierto al Conocimiento (COL)	45	28	116,041
Acceso Libre a Información Científica para la Innovación (PER)	233	0	836,582
Repositorio Nacional (MX)	80	87	59,792
Red Mexicana de Repositorios Institucionales (MX)	67	98	483,603

I - Instituciones  
R- Repositorios  
D - Documentos

Fecha de consulta: 5-febrero-2019

Figura 2.14: Repositorios con mayor producción en Latinoamérica

### 2.6.1 Servicios de búsqueda simple, avanzada y semántica

Las figuras 2.21 y 2.22 muestran los servicios de búsqueda que ofrecen diferentes RDs en AA a sus usuarios. Los criterios de búsqueda comunes son por autor, título, editor, lenguaje, fecha de publicación, entre otros. La figura 2.23 concentra información de los portales con mayor producción en Latinoamérica y los servicios de búsqueda que ofrecen, notar que sólo uno de los repositorios ofrece servicios basados en tecnologías semánticas.

### 2.6.2 Plataformas que procesan consultas semánticas

Los RAAs que cuentan con características técnicas que permiten implementar servicios semánticos se caracterizan por lo siguiente:

- Cuentan con un servicio o módulo de procesamiento de datos en RDF
- Mantienen un almacén de ternas en RDF



Figura 2.15: Vista del portal del Sistema Nacional de Repositorios Digitales



Figura 2.16: Vista del portal brasileño de publicaciones científicas y acceso abierto - oasisbr

- Emplean un lenguaje de consulta para acceder al almacén

La Tabla 2.2 muestra características técnicas entre las plataformas DSpace, Virtuoso y VIVO, las cuales emplean tecnologías de la web semántica. Estas plataformas trabajan con datos en RDF como los que muestra la Figura 2.24 y emplean como lenguaje de consulta al estándar SPARQL.



Figura 2.17: Vista del panel de búsquedas en RI2.0

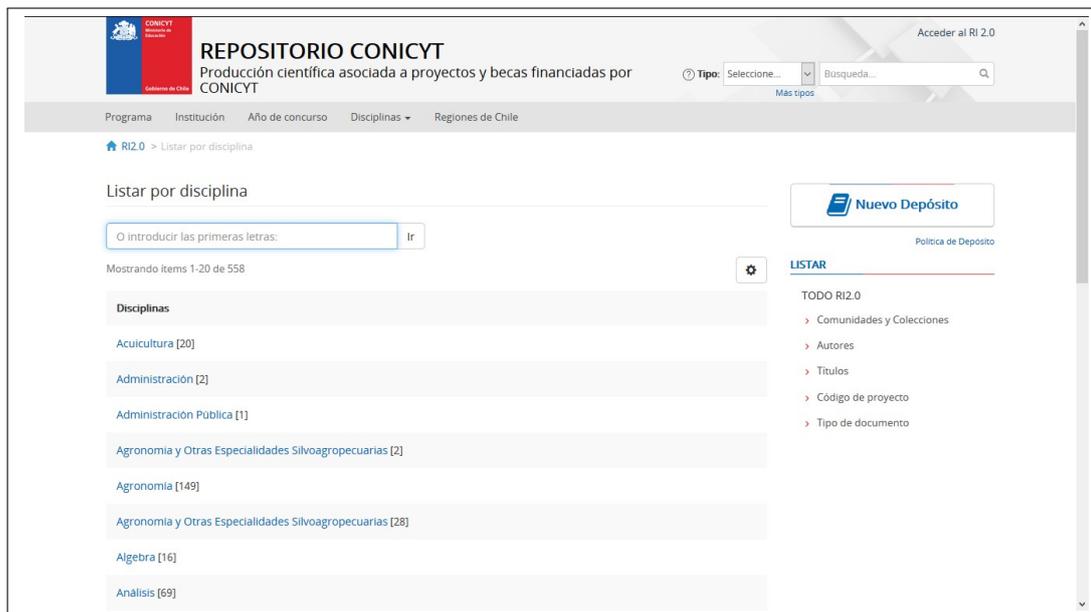


Figura 2.18: Vista del panel de búsquedas avanzadas en RI2.0



Figura 2.19: Vista del Sistema Nacional de Acceso Abierto al Conocimiento



Figura 2.20: Vista del portal de acceso libre a información científica para la innovación



Figura 2.21: Vista del panel de búsqueda en el portal Sistema Nacional de Repositorios Digitales en Argentina

Tabla 2.2: Características técnicas entre DSpace, Virtuoso y VIVO

Plataforma	Servicio RDF	Proveedor de almacenamiento
DSpace 6.2	Si	Apache Jena
Virtuoso	Si	Sesame, Redland, Apache Jena
VIVO	Si	-

Figura 2.22: Vista del panel de búsqueda en el portal instituto Brasileño de Información en Ciencia y Tecnología en Brasil

Repositorio	Básica	Avanzada	Semántica
Sistema Nacional de Repositorios Digitales (ARG)	Si	Si	No
Repositorio CONICYT (CHI)	Si	Si	No
Portal Brasileño de Publicaciones Científicas y A. A. (BRA)	Si	Si	No
Sistema Nacional de Acceso Abierto al Conocimiento (COL)	Si	Si	No
Acceso Libre a Información Científica para la Innovación (PER)	Si	Si	No
Repositorio Nacional (MX)	Si	Si	No
Red Mexicana de Repositorios Institucionales (MX)	Si	No	No
Enciclopedia de la Literatura en México	Si	Si	Si
Red CEDIA	Si	No	No

Fecha de consulta: 5-febrero-2019

Figura 2.23: Servicios de búsqueda en RDs de AA

```

@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix xsd: <http://www.w3.org/2001/XMLSchema#> .
@prefix dcterms: <http://purl.org/dc/terms/> .
@prefix bibo: <http://purl.org/ontology/bibo/> .
@prefix foaf: <http://xmlns.com/foaf/0.1/> .
@prefix dspace: <http://digital-repositories.org/ontologies/dspace/0.1.0#> .
@prefix dc: <http://purl.org/dc/elements/1.1/> .

<http://34.236.228.85/rdf/resource/123456789/8400>
  dspace:isPartOfCollection <http://34.236.228.85/rdf/resource/123456789/7882> ;
  dc:contributor "Gutiérrez Álvarez, Brenda Alma" ;
  dc:creator "Gutiérrez Álvarez, Brenda Alma" ;
  dc:date "2018-05-08T04:54:00Z"^^xsd:dateTime ;
  dc:language "spa" ;
  dc:publisher "Universidad de las Américas Puebla" ;
  dc:rights "Acceso Abierto" ;
  dcterms:available "2018-05-08T04:54:00Z"^^xsd:dateTime ;
  dcterms:title "Proyecto de Inversión de una Planta de Extracción y Empaque de Aceite de
Aguacate para el Rancho Villa Reguero" ;

```

Figura 2.24: Ejemplo de conjunto de ternas en RDF

El capítulo 3 se organiza como sigue: al inicio se presentan los modelos empleados para el desarrollo del servicio web, después se describe el *análisis de requerimientos* que determina las actividades clave. Posteriormente, en la etapa de *diseño* se explican los modelos de referencia y en la *implementación* se trabajan con etapas. Finalmente, se presenta una herramienta de *evaluación* para estimar la usabilidad del servicio web SW001 y se enumeran las tareas de *mantenimiento*.

### 3.1 Modelo de desarrollo

El desarrollo del servicio web SW001 se muestra la Figura 3.1. Se emplearon principalmente dos modelos en distintos momentos; como modelo base, se emplea el de cascada [1] para las etapas de análisis de requerimientos y diseño, para las etapas subsecuentes, se adaptó el modelo incremental que consiste en la codificación de módulos reducidos, pruebas rápidas de funcionamiento, realimentación, corrección, liberación o escalamiento hasta concluir la implementación; finalmente se considera la etapa de mantenimiento. Las secciones siguientes exponen este modelo de desarrollo.

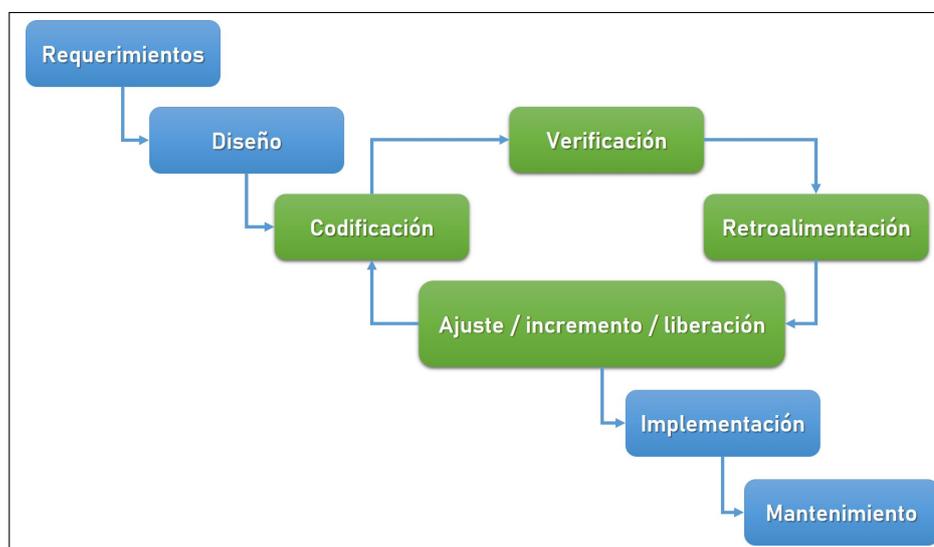


Figura 3.1: Modelo de desarrollo del servicio web SW001

### 3.2 Análisis de requerimientos

El servicio web SW001 forma parte del *back end* del RI-UPPue, el usuario final no tendrá una interacción con éste, sino que está diseñado para usuarios de tipo administrador. Junto con la administradora del RI-UPPue, se identificaron los requerimientos funcionales de la Tabla 3.1.

Tabla 3.1: Tabla de requerimientos funcionales del servicio web

No.	Descripción	Prioridad
1	Consulta de información semántica	Alta
2	Capacidad de almacenamiento de ternas en RDF	Alta
3	Integración de datos provenientes de un RI	Alta
4	Exportación de datos en formatos no propietarios como JSON y OWL	Mediana

### 3.3 Diseño

En la tesis se usan herramientas de modelado UML y los requerimientos planteados en la etapa de análisis para elaborar los modelos de casos de uso, diagrama de clases y diseño de alto nivel del servicio web SW001, los cuales se describen a continuación.

#### 3.3.1 Casos de uso

La Figura 3.2 muestra el diagrama de casos de uso que modela la funcionalidad general del RI-UPPue.

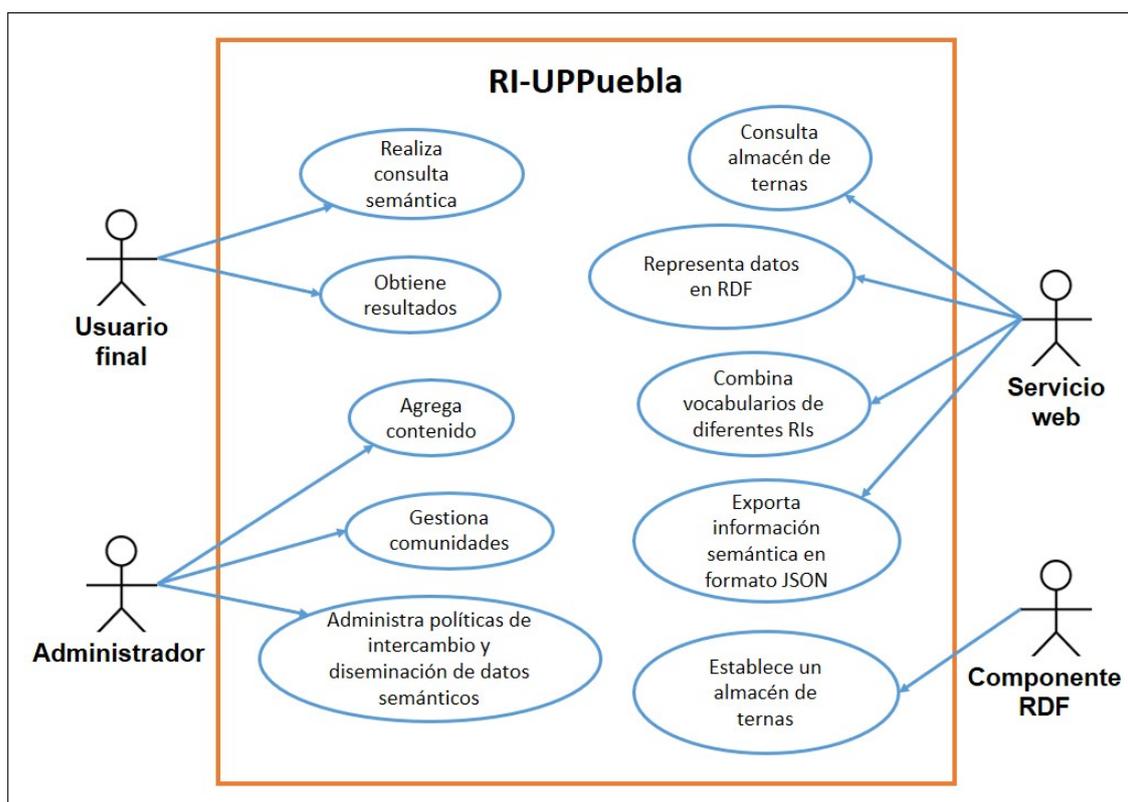


Figura 3.2: Diagrama de casos de uso

#### 3.3.2 Diagrama de clases

La Figura 3.3 muestra el diagrama de clases del servicio web, los rectángulos en color verde representan los nombres de las clases. Este diseño se propuso en el artículo [19],

en el cual se encuentra la descripción detallada de los atributos y métodos definidos para cada clase.

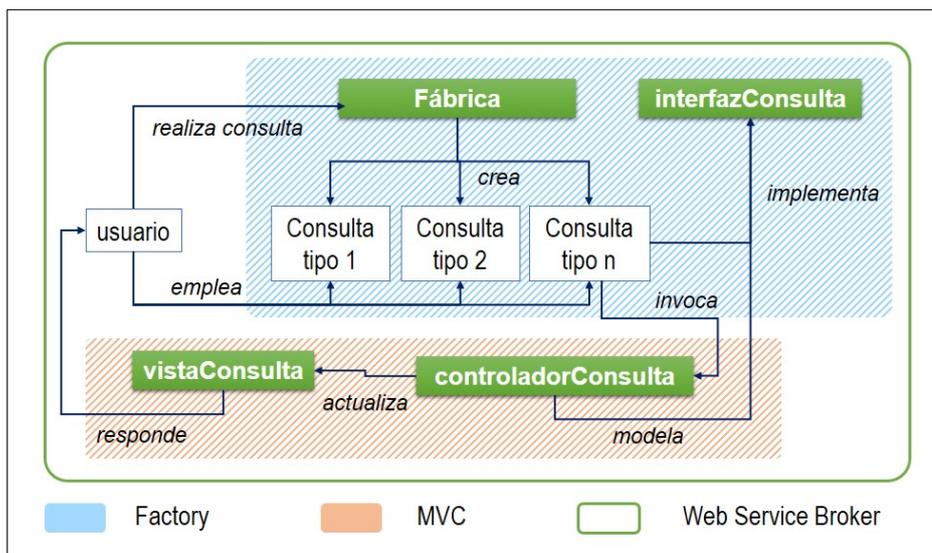


Figura 3.3: Diagrama de clases. Fuente: [19]

### 3.3.3 Diseño de alto nivel

Las Figuras 3.4 y 3.5 muestran los diseños elaborados para representar, por un lado, la manera en que se puede emplear el servicio web, por otro lado, un modelo general del comportamiento en un contexto general, donde a corto plazo, se propone que opere en el servidor del RI-UPPue y a mediano plazo, se buscará su instalación en diferentes RIs que compartan la plataforma tecnológica DSpace.

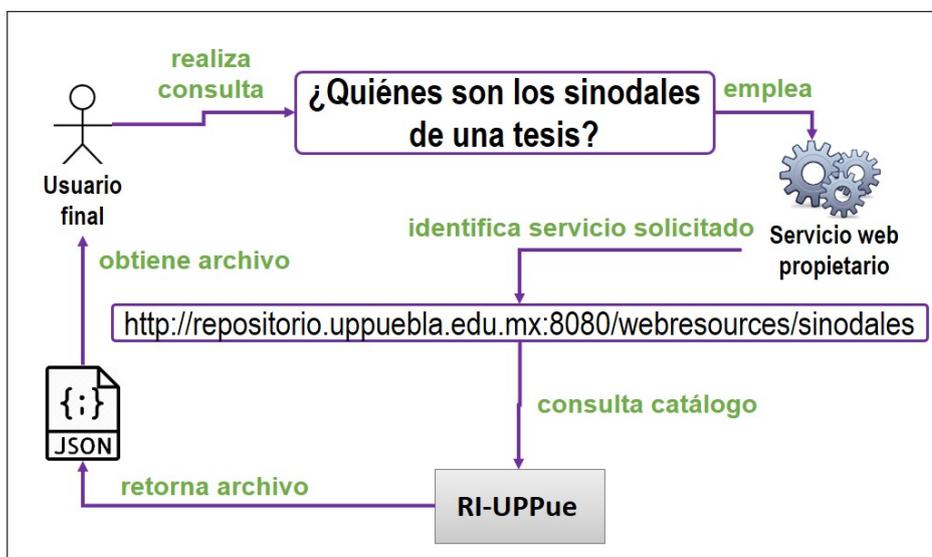


Figura 3.4: Diseño de consumo del servicio web. Fuente: [19]

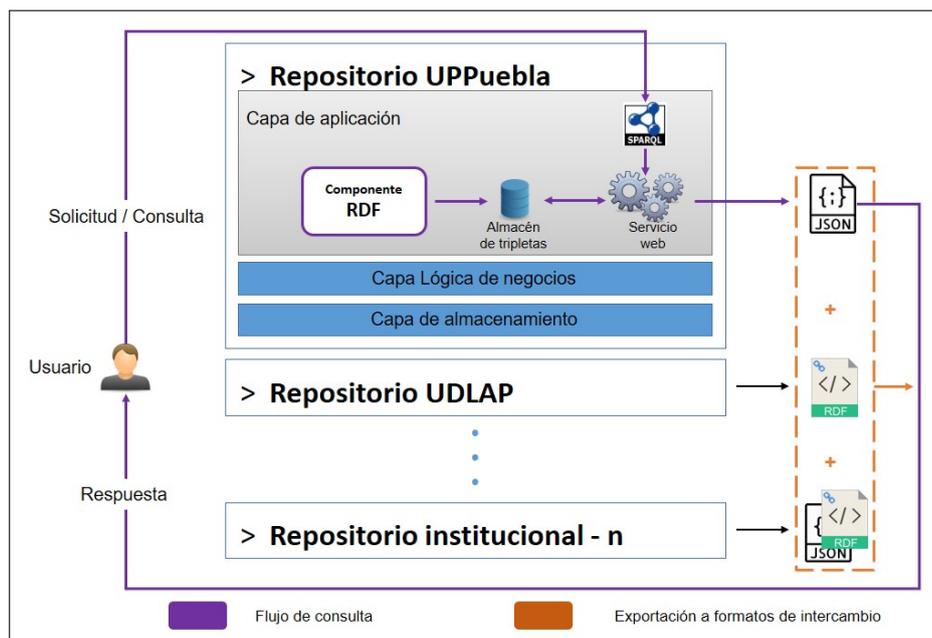


Figura 3.5: Diseño de alto nivel del servicio web. Fuente: [19]

### 3.4 Implementación

Previo a la implementación de los módulos, se asume que está instalada la plataforma DSpace 6.2. Los pre-requisitos de operación para el servicio se relacionan con los elementos que muestra la Figura 3.6, la cual contiene la arquitectura de la plataforma DSpace junto con los servicios adicionales. Cabe mencionar que la documentación de referencia de [11] indica que DSpace cuenta con un módulo RDF, sin embargo, éste no se habilita durante la instalación estándar.

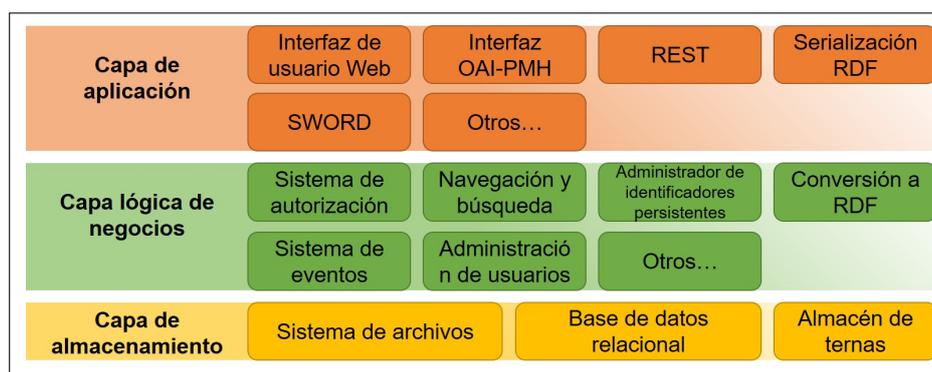


Figura 3.6: Módulos adicionales de DSpace para la versión 6.2

La habilitación del módulo de serialización RDF requiere de tres elementos de software: 1) un almacén de ternas, 2) un mecanismo de serialización y 3) un convertor a formato RDF. La instalación de estos elementos forman parte de las etapas del proceso técnico que muestra la Figura 3.7, estas etapas se describen en las secciones siguientes.

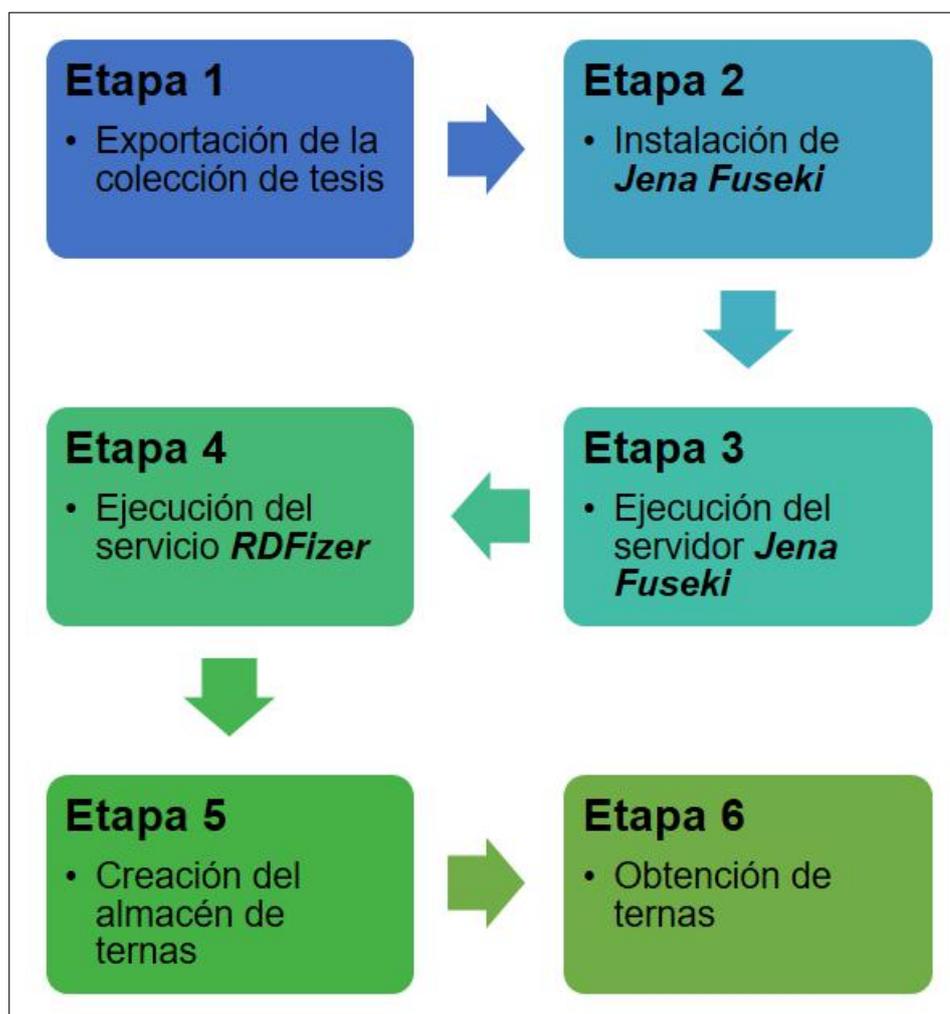


Figura 3.7: Etapas del proceso de migración de metadatos del RI-UPPue a la web semántica

### 3.4.1 Etapa 1: exportación de la colección de tesis

DSpace permite exportar cualquier colección, en particular, se emplea la colección de tesis del RI-UPPue, formada por más de 50 tesis del Departamento de Posgrado. La Figura 3.8 muestra la interfaz para esta tarea, los metadatos se almacenan a un archivo con extensión CSV<sup>1</sup>. El número 1 muestra el botón para exportar y el 2 el archivo CSV con los metadatos exportados. La tarea de exportación se llevó a cabo al utilizar la versión 75.0.3770.142 del navegador Chrome y la interfaz en JSP de DSpace.

### 3.4.2 Etapa 2: instalación de Apache Jena Fuseki

El almacén de ternas es Jena-Fuseki versión 1.6.0, este servidor provee también de una interfaz web para realizar consultas en lenguaje SPARQL. Fuseki se integra con TDB<sup>2</sup> para proporcionar una capa de almacenamiento persistente transaccional y robusta, permite consultas de texto y consultas espaciales [36]. La página oficial de Apache Jena

<sup>1</sup>Siglas del inglés de *Comma Separated Values*, formato separado por comas

<sup>2</sup>Componente de Jena Fuseki para el almacenamiento y consulta de datos en RDF

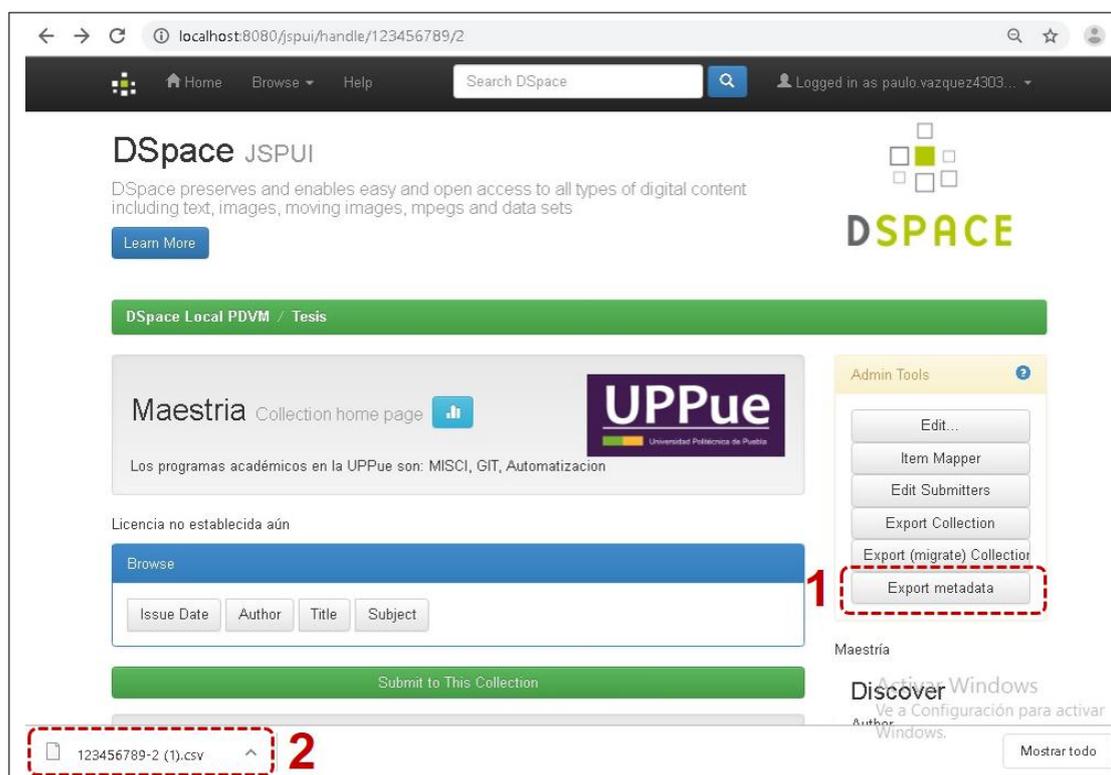


Figura 3.8: Exportación de la colección tesis a archivo CSV

Fuseki<sup>3</sup> provee diversas opciones para la descarga dependiendo del sistema operativo: Linux (.tar.gz) o Windows(.zip).

## Download Fuseki

Releases of Apache Jena Fuseki can be downloaded from one of the mirror sites:

[Jena Downloads](#)

and previous releases are available from [the archive](#). We strongly recommend that users use the latest official Apache releases of Jena Fuseki in preference to any older versions.

**Fuseki download files**

Filename	Description
<code>fuseki-*VER*.distribution.zip</code>	Fuseki download, includes everything.
<code>fuseki-*VER*-server.jar</code>	Fuseki server, as an executable jar.
<code>fuseki-*VER*-server.war</code>	Fuseki server, as a web application archive (.war) file.

Figura 3.9: Opciones de descarga para el almacén de ternas Jena Fuseki

Para terminar la instalación, los archivos descargados se descomprimen y se guardan en una carpeta que se nombra como *Fuseki*.

<sup>3</sup>Disponible en <https://jena.apache.org/download/index.cgi>

Nombre	Fecha de modifica...	Tipo	Tamaño
bin	03/04/2019 09:53 a...	Carpeta de archivos	
webapp	03/04/2019 09:53 a...	Carpeta de archivos	
fuseki	03/04/2019 09:52 a...	Archivo	13 KB
fuseki.service	03/04/2019 09:53 a...	Archivo SERVICE	3 KB
fuseki.war	03/04/2019 09:53 a...	Archivo WAR	23,315 KB
fuseki-backup	03/04/2019 09:53 a...	Archivo	2 KB
fuseki-server	03/04/2019 09:52 a...	Archivo	3 KB
fuseki-server	03/04/2019 09:53 a...	Archivo por lotes ...	2 KB
fuseki-server	03/04/2019 09:53 a...	Executable Jar File	26,045 KB
LICENSE	03/04/2019 09:52 a...	Archivo	30 KB
NOTICE	03/04/2019 09:52 a...	Archivo	10 KB
README	03/04/2019 09:52 a...	Archivo	3 KB

Figura 3.10: Contenido de la carpeta del servidor Jena Fuseki

### 3.4.3 Etapa 3: ejecución del servidor Jena Fuseki

Jena Fuseki se ejecuta como servicio de sistema operativo, como aplicación web Java (WAR) o como servidor independiente. El componente Apache Shiro se utiliza para implementar características de seguridad, tiene una interfaz de usuario para el monitoreo y la administración. Jena Fuseki se inicializa ejecutando el archivo `fuseki-server.bat`, (.jar en Linux) como muestra la Figura 3.11.

```

C:\Windows\system32\cmd.exe
[2019-04-03 10:12:56] Server      INFO  Apache Jena Fuseki 3.10.0
[2019-04-03 10:12:57] Config    INFO  FUSEKI_HOME=C:\Fuseki\
[2019-04-03 10:12:57] Config    INFO  FUSEKI_BASE=C:\Fuseki\run
[2019-04-03 10:12:57] Config    INFO  Shiro file: file://C:\Fuseki\run\shiro.ini
[2019-04-03 10:13:03] Server      INFO  Started 2019/04/03 10:13:03 CST on port 3030
[2019-04-03 10:14:01] Admin      INFO  [1] GET http://localhost:3030/$/server
[2019-04-03 10:14:01] Admin      INFO  [1] 200 OK (38 ms)
    
```

Figura 3.11: Ejecución del servidor Jena Fuseki en terminal de Windows

### 3.4.4 Etapa 4: ejecución del servicio RDFizer

Un componente RDFizer [37] o *spongers*, transforma datos de DSpace en una o más de las serializaciones al modelo de datos RDF. La Figura 3.12 representa esta transformación. Los RDFizers implementan las tareas básicas siguientes: extraer, transformar y cargar, en inglés, *extract*, *transform* y *load* (ETL), mismas que se describen como sigue:

- *Extraer*. Los datos se obtienen de una base de datos o se recopilan de múltiples y diversas fuentes
- *Transformar*. Los datos obtenidos se transforman en los requeridos mediante formularios; la transformación utiliza reglas o tablas de búsqueda o combina los datos con otros.

- *Cargar*. Los datos se escriben en la fuente o base de datos destino.

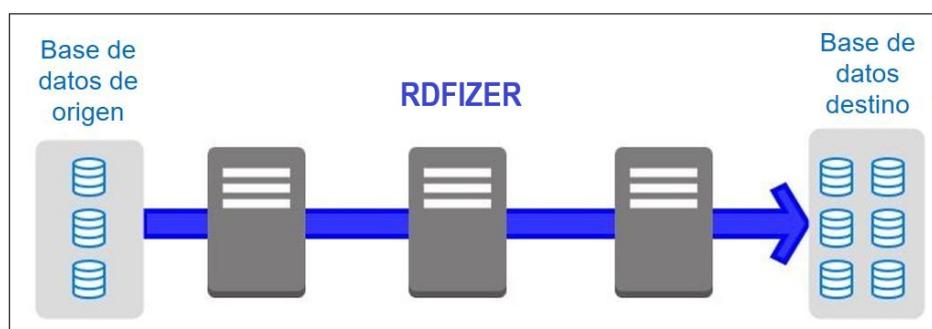


Figura 3.12: *Extracción, transformación y carga*

Los RDFizers soportan la migración de los metadatos a formatos de intercambio como XML, RDF o JSON.

### 3.4.5 Etapa 5: creación del almacén de ternas

Una vez que los metadatos del repositorio se han procesado por los RDFizes o ETLs, los archivos con datos en RDF se almacenan en la ubicación indicada en el archivo de configuración de la plataforma DSpace fuseki-assembler-ttl, forman así el almacén de ternas o servicio persistente.

### 3.4.6 Etapa 6: obtención de ternas

Los metadatos en RDF para cada tesis se acceden empleando una dirección de internet o URL<sup>4</sup> como muestra la Figura 3.13.

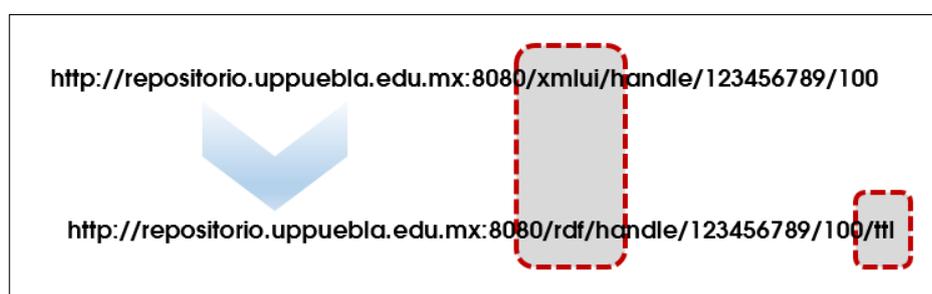


Figura 3.13: *Acceso a metadatos de tesis en RDF*

La Figura 3.14 muestra algunos metadatos de una tesis en RDF. Note que el archivo RDF contiene los metadatos capturados en la plataforma DSpace cuando se agrega un documento a la colección. En el ejemplo, se muestra un error lógico, notar que los elementos creator y contributor contienen el mismo valor.

<sup>4</sup>Localizador Uniforme de recursos, en inglés *Uniform Resource Locator*

```

dSPACE:isPartOfCollection <http://localhost:8080/rdf/resource/123456789/2> ;
dc:contributor "Bravo Hernández, Paola" ;
dc:creator "Bravo Hernández, Paola" ;
dc:date "2019-03-28T11:20:23Z"^^xsd:dateTime ;
dc:language "es" ;
dc:publisher "Universidad Politécnica de Puebla" ;
dcterms:abstract "-"@en-US ;
dcterms:available "2019-03-28T11:20:23Z"^^xsd:dateTime ;

```

Figura 3.14: Metadatos de RDF para una tesis del almacén de ternas

### 3.4.7 Herramientas software

Esta sección contiene una breve descripción de las herramientas de software utilizadas en la implementación del servicio web SW001.

#### 3.4.7.1 Herramientas de la interfaz gráfica de usuario

La interfaz gráfica de usuario (*Graphical User Interface*, GUI) o *front end* del servicio web SW001 integra tres tecnologías web: *HTML 5*, lenguaje de modelado de hipertexto, en inglés *HyperText Markup Language*, que es un lenguaje de marcado que sirve para modelar contenidos; *Java Script*, que es un lenguaje de programación del lado del cliente para manipular elementos HTML y hojas de estilo en cascada, en inglés *Cascading Style Sheets* (CSS), que sirven para definir reglas de diseño para la presentación del contenido. Estas tecnologías son nativas a la web por lo que para ejecutar el servicio, no se requiere de ningún tipo de instalación adicional, sino usar navegadores como Chrome, Mozilla, Opera, Edge o Safari.

#### 3.4.7.2 Servicio especializado en lenguaje Python 3.6

Se utiliza la versión 3.6 del lenguaje de programación Python para extraer elementos (*items*) del almacén de ternas a través de la librería RDFLib versión 4.2.2 [38] y así integrar los datos provenientes del RI-UPPue a una instancia de la ontología Onto4UPPue a través de la librería OWLReady2 [39] versión 0.19. La Figura 3.15 muestra la arquitectura propuesta para el servicio web y sus componentes.

La Figura 3.16 muestra las tecnologías empleadas para el desarrollo del servicio web desde el punto de vista del *front end* (parte gráfica) y del *back end* (parte lógica).

#### 3.4.7.3 Librería RDFLib

La librería RDFLib<sup>5</sup> de Python sirve para trabajar con datos en RDF en diferentes sintaxis como RDF/XML, N3, NTriples, Turtle, TriX, RDFa y Microdata; cuenta con la interfaz *Graph*, el núcleo *rdflib* incluye implementaciones de Store para almacenamiento en memoria, almacenamiento persistente y un contenedor para puntos finales remotos de SPARQL 1.1.

La instalación de RDFLib se realiza a través del servicio pip<sup>6</sup> en la instrucción `pip install rdflib` como se indica en la documentación.

<sup>5</sup>Disponible en: <https://pypi.org/project/rdflib/>

<sup>6</sup>Repositorio de software para el lenguaje Python, en inglés *Python Index Package*

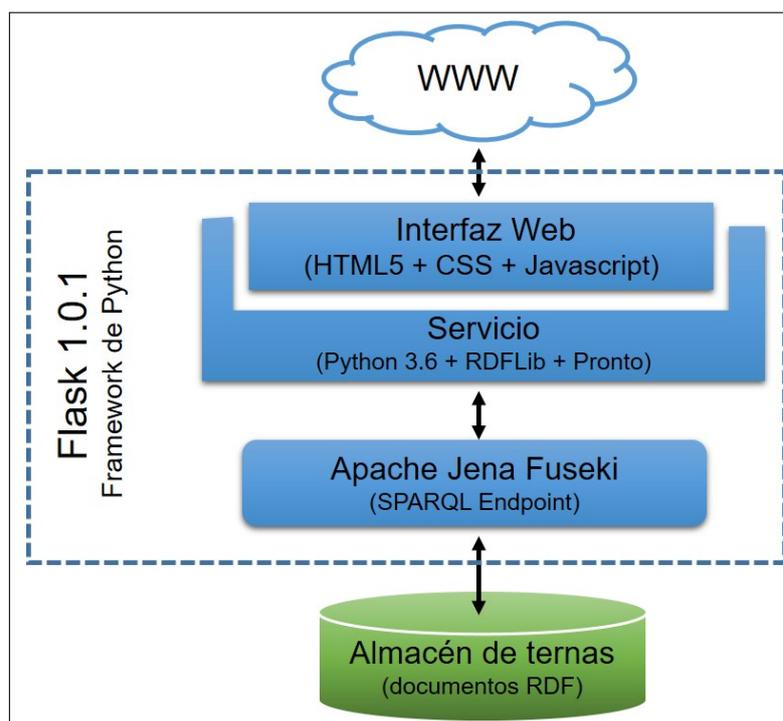


Figura 3.15: Arquitectura del servicio web

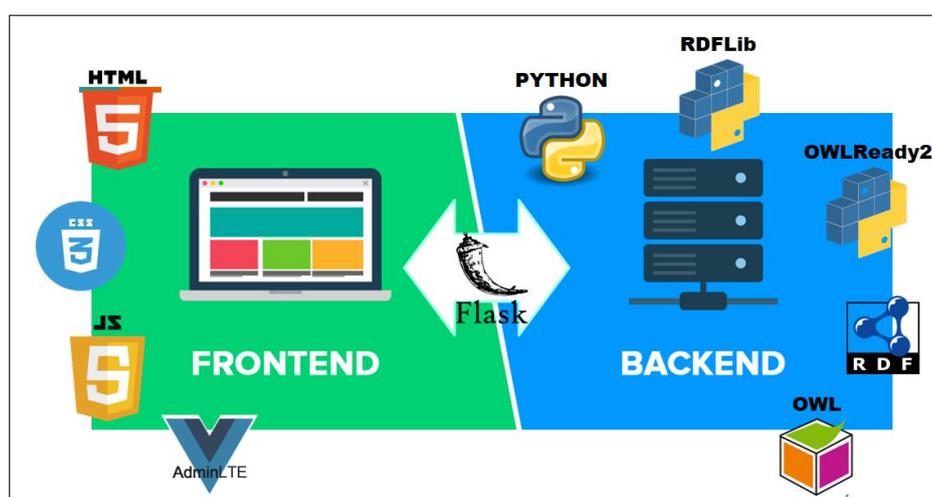


Figura 3.16: Tecnologías empleadas en el servicio web

#### 3.4.7.4 Librería OWLReady2

La librería OWLReady2<sup>7</sup> procesa ontologías escritas en la versión 2.0 de OWL, las cuales se gestionan como objetos Python, se conecta con el razonador Hermit y soporta tareas como las siguientes:

- Importación de ontologías OWL 2.0 en ternas-N, RDF/XML o formato OWL/XML.
- Exportación de ontologías OWL 2.0 a ternas-N o RDF/XML

<sup>7</sup>Disponible en <https://pypi.org/project/Owlready2/>

- Manipulación de clases, instancias y propiedades de las ontologías de forma transparente (como si fueran objetos de Python)
- Agregación de métodos de Python a las clases de una ontología
- Clasificación automática de clases e instancias utilizando los razonadores HermitT o Pellet
- Compatibilidad con la librería RDFlib

#### 3.4.7.5 Librería `xml.etree.ElementTree`

La librería `xml.etree.ElementTree`<sup>8</sup> se usa para procesar documentos XML, cuenta con diversos métodos para acceder a elementos específicos de un documento, insertar, modificar y eliminar nodos. En la tesis, esta librería se utiliza para obtener información de la ontología *Onto4UPPue* escrita en OWL.

#### 3.4.7.6 Framework Flask

El desarrollo del servicio web emplea la versión 1.1.1 de Flask<sup>9</sup>, el cual es un *framework* ligero para aplicaciones web. Flask implementa el patrón de diseño Modelo-Vista-Controlador (MVC), en inglés *Model-View-Controller*, ampliamente usado en el desarrollo de proyectos de software.

#### 3.4.7.7 AdminLTE

Finalmente, en el desarrollo del *front end* se integra el panel de control o *dashboard* que conforma un tipo de interfaz gráfica de usuario con *AdminLTE*, versión 2.4.15.

### 3.5 Evaluación

La evaluación de la usabilidad de la interfaz gráfica del servicio web se orienta a obtener información de la experiencia del usuario<sup>10</sup>, su propósito es identificar problemas de uso. En la tesis se adapta la plantilla diseñada para hacer análisis heurísticos descrita en [40], disponible en <https://www.torresburriel.com/weblog/2008/11/28/plantilla-para-hacer-analisis-heuristicos-de-usabilidad/>.

La plantilla consta de un cuestionario que explora las once heurísticas siguientes:

1. Generalidades
2. Identidad e información
3. Lenguaje y redacción
4. Rotulado

---

<sup>8</sup>Disponible en <https://pypi.org/project/elementtree/>

<sup>9</sup>Disponible en <https://pypi.org/project/Flask/>

<sup>10</sup>En inglés, las siglas *UX* abrevian la expresión *User eXperience*

5. Estructura y navegación
6. Estructura de la página
7. Búsqueda
8. Elementos multimedia
9. Ayuda
10. Accesibilidad
11. Control y retroalimentación

Cada usuario indica su grado de satisfacción para cada heurística al utilizar una escala de Likert de cinco grados con los valores mostrados en la Tabla 3.2.

Tabla 3.2: Escala de Likert empleada para evaluar la interfaz gráfica del servicio web

Valor	Interpretación
1	Valor mínimo
2	Valor bajo
3	Valor medio
4	Valor alto
5	Valor máximo

A manera de ejemplo, la Figura 3.17 muestra un extracto de los elementos utilizados en [40] para evaluar la heurística *generalidades* (heurísticos generales).

Heurísticos generales <sup>2</sup>	
Generales	Puntos
¿Cuáles son los objetivos del sitio web? ¿Son concretos y bien definidos?	
¿Los contenidos y servicios que ofrece se corresponden con esos objetivos?	
¿Tiene una URL correcta, clara y fácil de recordar? ¿Y las URL de sus páginas internas? ¿Son claras y permanentes?	
¿Muestra de forma precisa y completa qué contenidos o servicios ofrece realmente el sitio web?	
¿La estructura general del sitio web está orientada al usuario?	
¿El look & feel general se corresponde con los objetivos, características, contenidos y servicios del sitio web?	

Figura 3.17: Reactivos para evaluar la heurística generalidades [40]

### 3.6 Mantenimiento

La fase final del desarrollo del servicio web consta de realizar las modificaciones o adecuaciones identificadas en la etapa de evaluación.

## 4.1 Resultados

Este capítulo presenta los resultados obtenidos de las fases propuestas en la metodología, las cuales se muestran en la Figura 4.1.



Figura 4.1: Esquema general de fases para la obtención de resultados

Se validó el funcionamiento de los servicios tecnológicos instalados, configurados y desarrollados a través de la verificación de los procesos siguientes:

- Exportación de metadatos en formato CSV
- Recuperación de metadatos del almacén de ternas en formato RDF
- Exportación de datos a formato JSON
- Integración de instancias a la ontología Onto4UPPue en formato OWL
- Verificación de la funcionalidad del servicio web para búsqueda semántica SW001

### 4.1.1 Fase 1. Captura de metadatos en DSpace 6.2

Para la implementación de la fase 1, captura de metadatos en la versión 6.2 de la plataforma DSpace, se realizaron las siguientes actividades:

- Instalación de una instancia local de DSpace que emula las condiciones técnicas del RI-UPPue
- Creación de la comunidad *Tesis*, Figura 4.2
- Creación de la colección *Maestría*, Figura 4.3

- Inserción de un conjunto de once tesis (o ítems) en la colección *Maestría*, se introdujeron archivos y metadatos, Figura 4.4

DSpace Local PDVM / Administer

### Create Community ?

Community's metadata

Name: Tesis

Short Description: Tesis de maestría UPPue

Introductory text (HTML):  
 Las tesis de maestría pertenecientes a los programas educativos:  
 \* Maestría en Automatización  
 \* Maestría en Ingeniería en Sistemas y Cómputo Inteligente  
 \* Maestría en Biotecnología

Copyright text (plain text):

Figura 4.2: Creación de la comunidad de tesis

DSpace Local PDVM / Administer

### Edit Collection 123456789/2 ?

Collection's Metadata

Name: Maestría

Short Description: Tesis de maestría UPPue

Introductory text (HTML):  
 Los programas académicos en la UPPue son: MISCI, GIT, Automatización

Copyright text (plain text):  
 Licencia no establecida aún

Side bar text:

Delete this Collection...

Submission Workflow

Submitters: Edit... Delete

Accept/Reject Step: Create...

Accept/Reject/Edit Metadata Step: Create...

Edit Metadata Step: Create...

Collection Administrators: Create...

Collection's settings

Item template: Create...

Figura 4.3: Creación de la colección para tesis de maestría

#### 4.1.2 Fase 2. Creación de almacén de ternas

La Figura 4.5 muestra las actividades principales para crear el almacén de ternas, servicio persistente indispensable para la implementación de las fases posteriores.



Figura 4.4: Ítems agregados a la colección de tesis

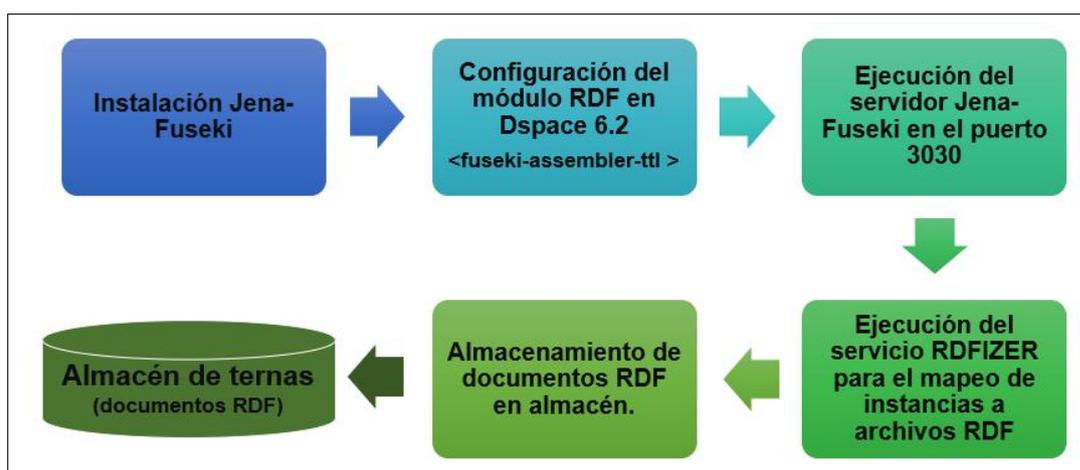


Figura 4.5: Implementación del almacén de ternas empleando el servidor Apache Jena Fuseki versión 1.6

El Anexo A contiene el manual que describe la habilitación del componente RDF de la plataforma DSpace versión 6.2 a través de la instalación del servidor Apache Jena Fuseki versión 1.6. La Figura 4.6 muestra la interfaz desde un navegador web posterior al iniciar la ejecución.

### 4.1.3 Fase 3. Generación de documentos en RDF

La generación de documentos en RDF requiere de ejecutar el serializador *RDFizer* como muestra la Figura 4.7, esta acción permite extraer los metadatos almacenados en DSpace y migrarlos al almacén de ternas. La verificación de la migración se realizó a través de la revisión de los documentos RDF generados conforme a las actividades de la Tabla 4.1.

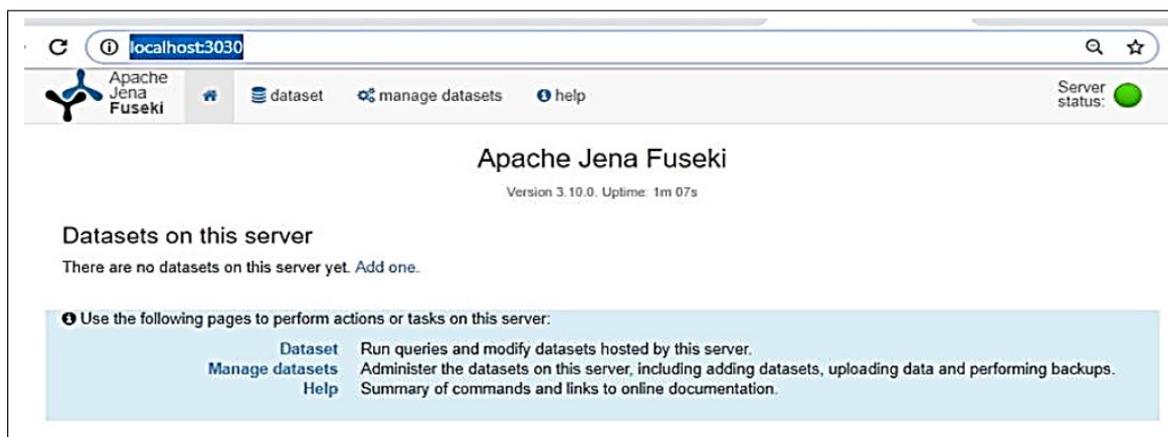


Figura 4.6: Ejecución local del servidor Apache Jena Fuseki

```

C:\Windows\system32\cmd.exe - dspace rdfizer -convert-all
28/03/2019 05:19 a. m. <DIR> upload
18/08/2018 07:58 a. m. <DIR> var
18/08/2018 08:01 a. m. <DIR> webapps
5 archivos 254,158,680 bytes
22 dirs 11,351,777,280 bytes libres

C:\DSpace>cd bin
C:\DSpace\bin>dspace rdfizer -convert-all
Using DSpace installation in: C:\DSpace
  
```

Figura 4.7: Ejecución del serializador RDFizer

#### 4.1.4 Fase 4. Extracción y procesamiento de metadatos

Se implementó un servicio (ver Figura 4.8) utilizando la versión 3.6 del lenguaje Python y la biblioteca `RDFlib`<sup>1</sup> para verificar automáticamente el número de documentos migrados. La Tabla 4.2 muestra los resultados. La Figura 4.9 muestra el resultado de ejecutar dicho servicio para una tesis.

Para verificar los resultados del proceso de migración, se llevó a cabo la revisión de los datos exportados al formato CSV y RDF al conjunto de prueba, (ver la Tabla 4.3). Los metadatos exportados en CSV son: id, collection, author, accessioned, available, issued, abstract, provenance, sponsorship, description, citation, uri, iso, publisher,

<sup>1</sup> biblioteca de Python para trabajar con archivos XLM y RDF

Tabla 4.1: Actividades de verificación con instancias migradas al almacén de ternas

No.	Actividad	Resultado esperado	Resultado obtenido
1	Serialización de ítems almacenados en DSpace	Once instancias migradas a formato RDF	Once instancias migradas a formato RDF
2	Acceso a documento RDF mediante URL	Consulta de once documentos RDF	Consulta de once documentos RDF

Tabla 4.2: Actividades de verificación con archivos en RDF

No.	Acción	Resultado esperado	Resultado obtenido
1	Recuperación de archivos RDF integrados en el almacén	Extracción de once tesis	Extracción de once tesis
2	Recuperación de los metadatos de un archivo RDF	Recuperación de quince metadatos DCMI	Recuperación de doce metadatos DCMI, dos metadatos propios de DSpace y otros tres

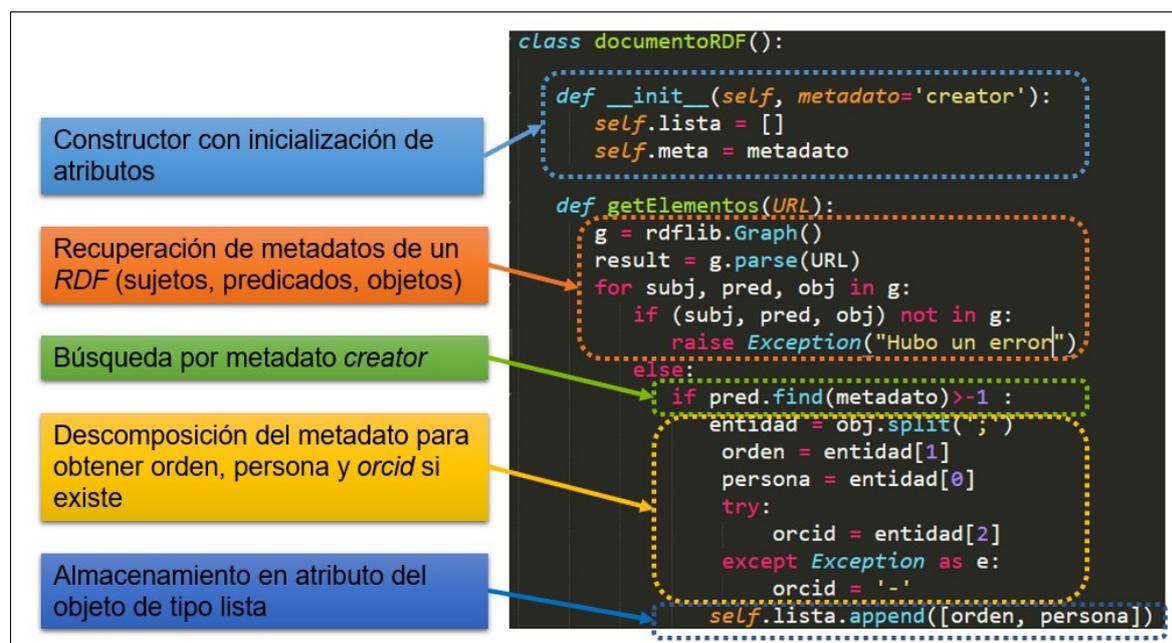


Figura 4.8: Descripción de la clase para la extracción de metadatos desde archivo RDF

subject, alternative title, title y type.

#### 4.1.5 Fase 5. Lectura e inserción de instancias en la ontología Onto4UPPue

Para implementar la lectura de datos en RDF y la inserción de instancias en la ontología Onto4UPPue, el método `getroot` permite ubicarse en el nodo raíz de árbol en XML y los métodos `set` y `dump` se emplean para la inserción de nuevos nodos en el cuerpo del documento, conservando la estructura de las instancias existentes en la ontología. La Figura 4.10 muestra un extracto de la implementación de la inserción de una instancia en la ontología; la Figura 4.11 muestra la estructura que deben guardar las instancias insertadas mediante el servicio de integración.

#### 4.1.6 Fase 6. Implementación del servicio web

La Figura 4.12 muestra las tecnologías empleadas para el desarrollo del servicio web SW001, éstas lo dotan de características como las siguientes: adaptable, validación HTML, dinámico, accesible e intuitivo.

```
> Identificando metadatos en http://localhost:8080/rdf/handle/123456789/39/ttl

DCMI metadatos:
1.- http://purl.org/dc/elements/1.1/contributor
2.- http://purl.org/dc/terms/bibliographicCitation
3.- http://purl.org/dc/elements/1.1/publisher
4.- http://purl.org/dc/elements/1.1/creator
5.- http://purl.org/dc/terms/hasPart
6.- http://purl.org/dc/elements/1.1/language
7.- http://purl.org/dc/elements/1.1/date
8.- http://purl.org/dc/terms/issued
9.- http://purl.org/dc/terms/isPartOf
10.- http://purl.org/dc/terms/abstract
11.- http://purl.org/dc/terms/available
12.- http://purl.org/dc/terms/title

Otros metadatos:
1.- http://digital-repositories.org/ontologies/dspace/0.1.0#isPartOfCollection
2.- http://purl.org/ontology/bibo/uri
3.- http://xmlns.com/foaf/0.1/homepage
4.- http://digital-repositories.org/ontologies/dspace/0.1.0#hasBitstream
5.- http://rdfs.org/ns/void#sparqlEndpoint
[Finished in 1.1s]
```

Figura 4.9: Metadatos recuperados para una tesis del almacén de ternas

Tabla 4.3: Actividades de verificación con archivos CSV

No.	Actividad	Resultado esperado	Resultado obtenido
1	Exportación de instancias en formato CSV	Archivo con once instancias	Archivo con once instancias
2	Exportación de metadatos en formato CSV	Archivo con quince metadatos DCMI	Archivo con dieciseis metadatos DCMI y dos metadatos DSpace

```
import xml.etree.ElementTree as ET
ontologia = ET.parse('ontologias/onto4RI-UPPue.owl')
raiz = ontologia.getroot()

... # MetadatosInstancias -> diccionario de instancias con sus metadatos
for x in len(MetadatosInstancias):
    dpa_Autor = raiz.Element('DataPropertyAssertion')
    dpi_Autor = raiz.SubElement(dpa_Autor, 'DataProperty').set('IRI', '#Autor')
    nii_Autor = raiz.SubElement(dpa_Autor, 'NamedIndividual').set('IRI', '#T24')
    ldi_Autor = raiz.SubElement(dpa_Autor, 'Literal').set('datatypeIRI',
        'http://www.w3.org/1999/02/22-rdf-syntax-ns#PlainLiteral')
    ldi_Autor.text = nombreAutor
    raiz.dump(dpa_Autor)

... # Mas metadatos
```

Figura 4.10: Método de inserción de instancias

El servicio web SW001 cuenta con las siguientes secciones:

- *Inicio*. Representa la página inicial del servicio
- *Búsqueda semántica*. Implementa la búsqueda semántica al consultar las relaciones almacenadas en la ontología Onto4UPPue

```

2631 <DataPropertyAssertion>
2632 <DataProperty IRI="#Autor"/>
2633 <NamedIndividual IRI="#T24"/>
2634 <Literal datatypeIRI="http://www.w3.org/1999/02/22-rdf-syntax-ns#PlainLiteral">Pérez Juárez, Irianelly
Ivonne</Literal>
2635 </DataPropertyAssertion>
2636 <DataPropertyAssertion>
2637 <DataProperty IRI="#Descripcion"/>
2638 <NamedIndividual IRI="#T24"/>
2639 <Literal datatypeIRI="http://www.w3.org/1999/02/22-rdf-syntax-ns#PlainLiteral"></Literal>
2640 </DataPropertyAssertion>
2641 <DataPropertyAssertion>
2642 <DataProperty IRI="#Fecha"/>
2643 <NamedIndividual IRI="#T24"/>
2644 <Literal datatypeIRI="http://www.w3.org/1999/02/22-rdf-syntax-ns#PlainLiteral">2019-03-24T03:33:23Z</
Literal>
2645 </DataPropertyAssertion>
2646 <DataPropertyAssertion>
2647 <DataProperty IRI="#Lenguaje"/>
2648 <NamedIndividual IRI="#T24"/>
2649 <Literal datatypeIRI="http://www.w3.org/1999/02/22-rdf-syntax-ns#PlainLiteral">Español</Literal>
2650 </DataPropertyAssertion>
2651 <DataPropertyAssertion>
2652 <DataProperty IRI="#Tema"/>
2653 <NamedIndividual IRI="#T24"/>
    
```

Figura 4.11: Estructura de una instancia agregada por el servicio



Figura 4.12: Estructura de una instancia agregada por el servicio

- *Test UX.* Accede al instrumento de evaluación para estimar la usabilidad del servicio web
- *Exportar.* Exporta la ontología, incluye las instancias de la colección de tesis
- *Preguntas frecuentes.* Muestra preguntas frecuentes y respuestas del dominio de repositorios institucionales
- *Contacto.* Contiene información de los desarrolladores

Las Figuras de la 4.13 a la 4.18 ilustran algunas de secciones. Los elementos principales de sección de *inicio* son: un mensaje de bienvenida y el objetivo del servicio.



Figura 4.13: Sección de inicio del servicio web SW001

La sección *búsqueda semántica* muestra un panel para seleccionar las relaciones *es autor* y *es sinodal*, las cuales modelan la participación de los autores en una tesis de maestría. Actualmente, la interfaz incluye únicamente estas relaciones, sin embargo, se requeriría realizar pequeñas modificaciones en el código de Python para integrar otras.

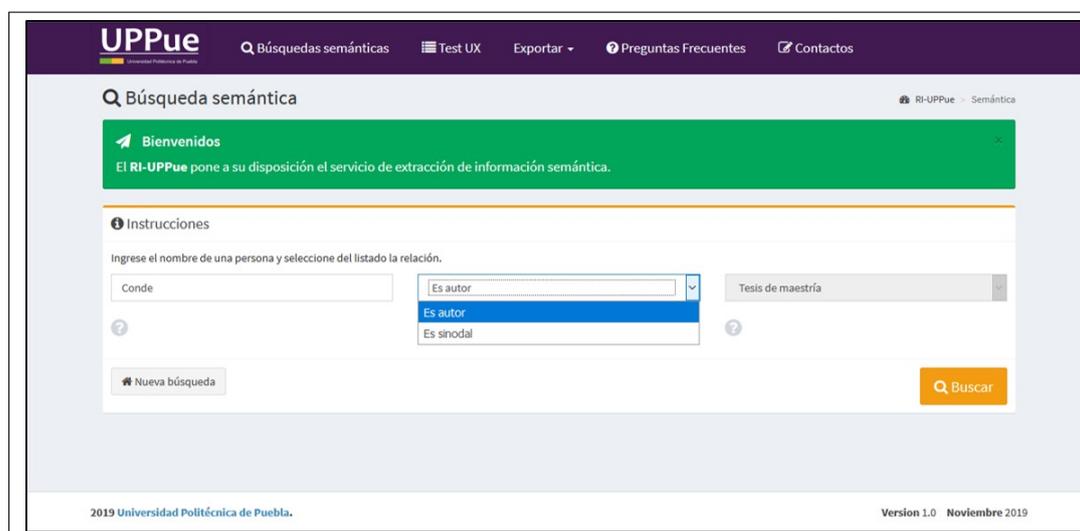


Figura 4.14: Servicio de búsqueda semántica

Una vez seleccionada la relación o criterio de búsqueda, el usuario presiona el botón *Buscar* en la parte inferior del panel y posteriormente se muestran los resultados en forma de tabla.

El servicio web SW001 implementa una tarea que permite descargar la ontología en dos formatos de intercambio: JSON y OWL, ambos pueden procesarse por cualquier lenguaje de programación de alto nivel y ser utilizados con otros propósitos.

La sección de *preguntas frecuentes* tiene el propósito de apoyar a los usuarios en la

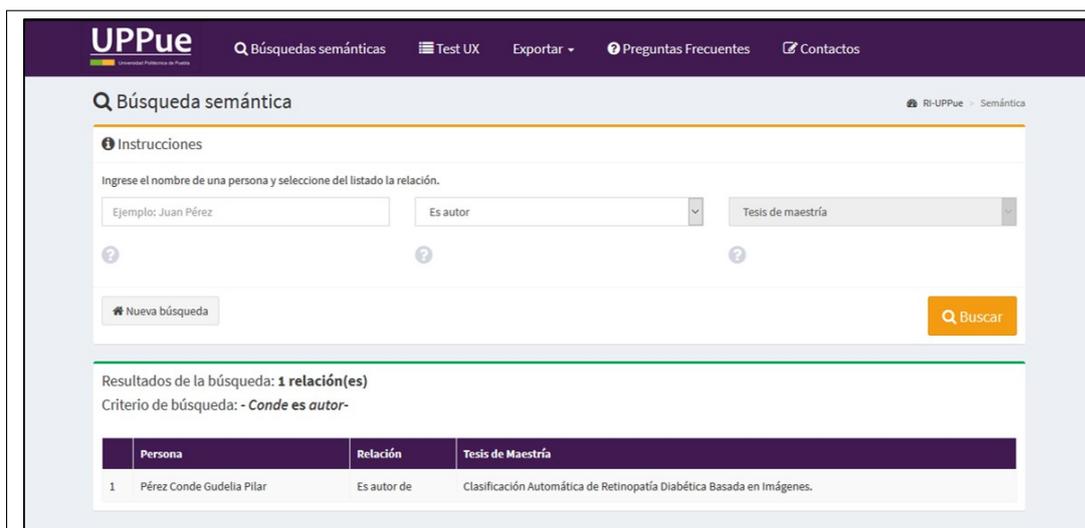


Figura 4.15: Resultados de una de búsqueda semántica

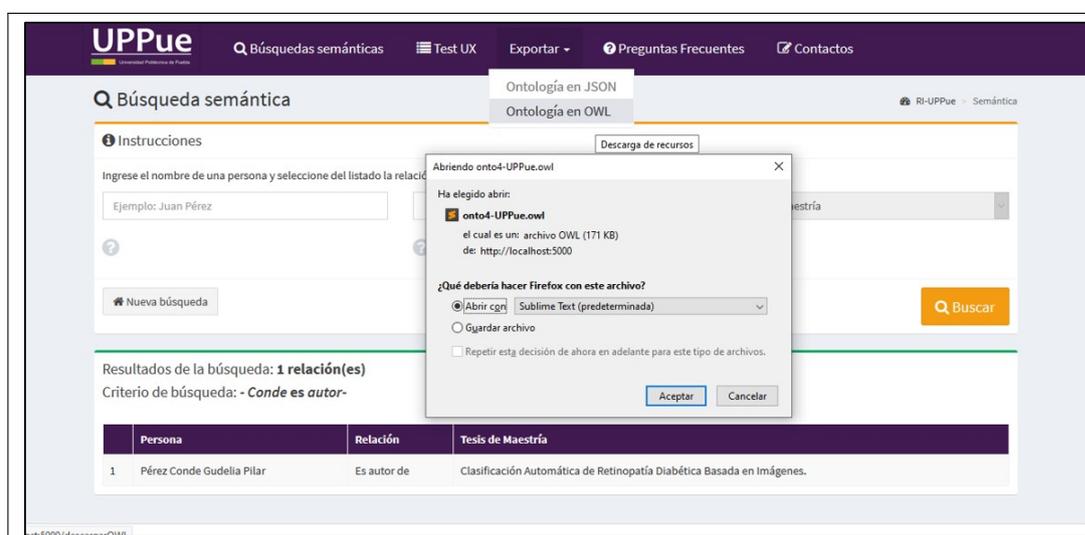


Figura 4.16: Descarga de la ontología ONTO4-UPPUE en formato JSON u OWL

adquisición de conceptos del dominio de los repositorios institucionales. La Figura 4.17 muestra la interfaz de esta sección.

Finalmente, la sección de *Contacto* muestra datos de la directora de tesis como del desarrollador del servicio web SW001.

El servicio web es accesible desde equipos de cómputo de escritorio o cualquier dispositivo móvil que cuente con conexión a internet. La Figura ?? muestra las interfaces de algunas secciones cuando se accede desde un dispositivo móvil.

#### 4.1.7 Fases 7 y 8: evaluación y mantenimiento

##### 4.1.7.1 Verificación del proceso de integración de instancias

Se realizaron observaciones en el contenido del archivo OWL para identificar las instancias integradas a la ontología ONTO4UPPue mediante el servicio web SW001. La

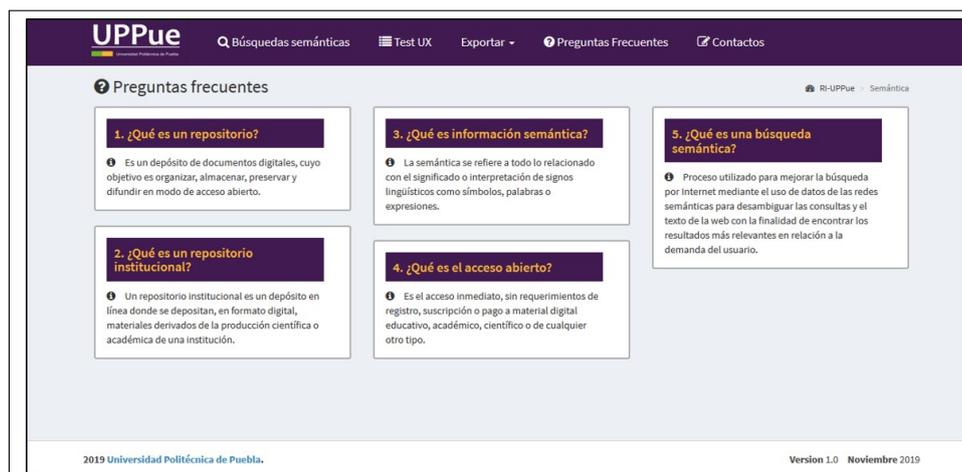


Figura 4.17: Sección de preguntas frecuentes

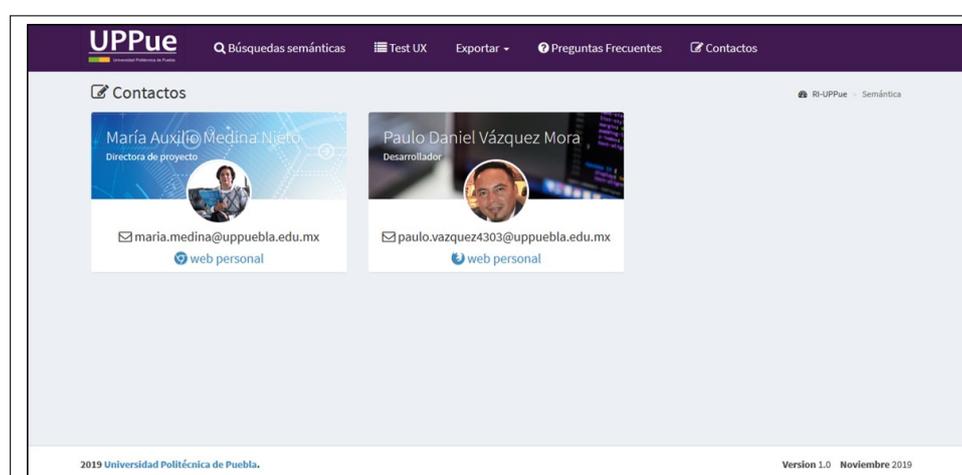


Figura 4.18: Sección de contacto



Figura 4.19: Interfaz del servicio web SW001 desde un dispositivo móvil

Tabla 4.4 muestra los resultados.

La Figura 4.20 muestra la integración de instancias dentro de la ontología Onto4UPPue. El panel Individuals del editor Protégé muestra las instancias agregadas mediante el servicio web SW001, en la denominación de T24 hasta T32. Las propiedades de datos de

Tabla 4.4: Verificación de instancias agregadas a Onto4UPPue

No.	Actividad	Resultado esperado	Resultado obtenido
1	Integración de nuevas instancias de tesis en la ontología	Integración de nuevas instancias de tesis en formato OWL	Integración de nuevas instancias de tesis en formato OWL
2	Validación de consistencia lógica utilizando el editor Protégé <sup>2</sup> y los razonadores Hermit y Pellet	Inserción de nuevas instancias (individuals) en la ontología Onto4UPPue	Inserción de nuevas instancias (individuals) en la ontología Onto4UPPue

cada instancia contienen los metadatos.

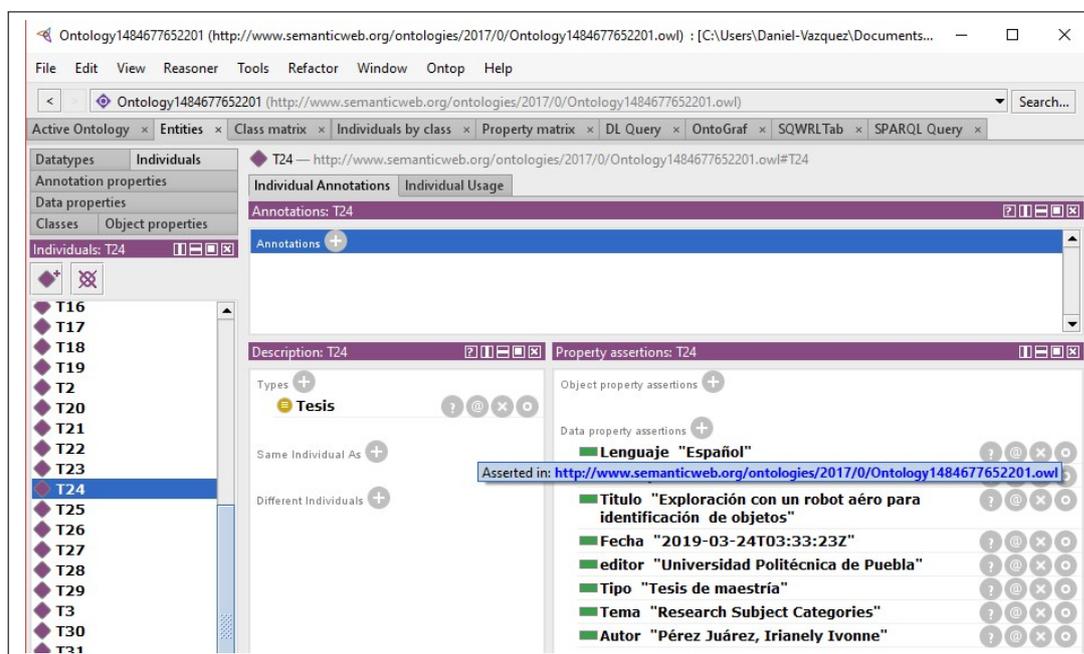


Figura 4.20: Vista de la instancia T24 agregada por el servicio web SW001.

#### 4.1.8 Evaluación de la experiencia del usuario

Como un servicio complementario, se desarrolló una aplicación web que evalúa la experiencia del usuario mediante un cuestionario que considera las once heurísticas propuestas por [40]. La figura 4.21 muestra la pantalla principal del test.

Para la evaluación del servicio web participaron dieciséis usuarios con las siguientes características generales:

- 6 mujeres y 10 hombres
- Las edades de los usuarios están en el rango de los 21 y 24 años
- Todos alumnos de la carrera de ingeniería en informática, próximos a egresar.

Figura 4.21: Pantalla principal del TestUX

Cabe resaltar que según [?] y [40], sólo se requiere de cinco personas para poder realizar la evaluación de la experiencia del usuario, con lo que se excede del número sugerido por los autores antes mencionados.

La evaluación consistió en una serie de afirmaciones, en las cuales los usuarios indicaron a través de una escala Likert de cinco grados, el nivel de satisfacción con las características, elementos o funcionalidades que el servicio web debería contar. Las escalas consideradas son las siguientes:

- No aplica (0)
- Totalmente insatisfecho (1)
- Muy insatisfecho (2)
- Satisfecho (3)
- Muy satisfecho (4)
- Totalmente satisfecho (5)

Las figuras 4.22 y 4.23 muestran los resultados obtenidos en la evaluación por cada una de las heurísticas y de manera general respectivamente.

Derivado de los resultados mostrados en figuras 4.22 y 4.23 se concluyó que la heurística mejor evaluada fue la de accesibilidad y la peor fue ayuda.

El promedio de las calificaciones obtenidas fue de **4** por lo que se concluye que el usuario se siente **muy satisfecho** con el servicio web.

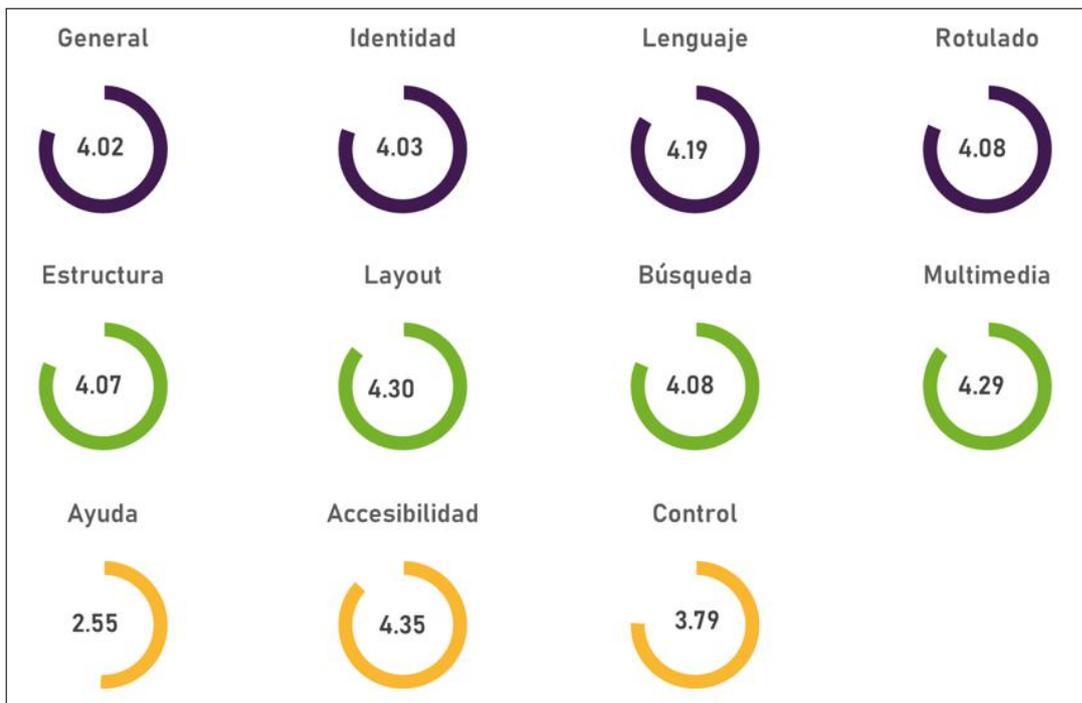


Figura 4.22: Resultados de la evaluación del servicio web

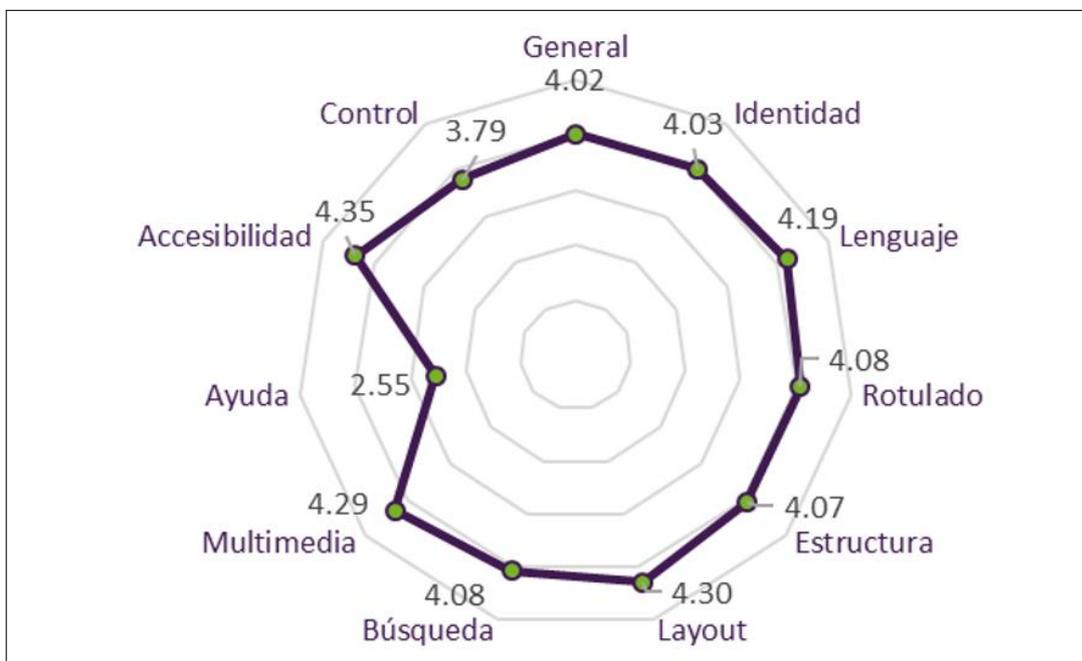


Figura 4.23: Resultado general de la evaluación del servicio web

La figura 4.24 muestra evidencia fotográfica de los usuarios que realizaron la evaluación del servicio web.

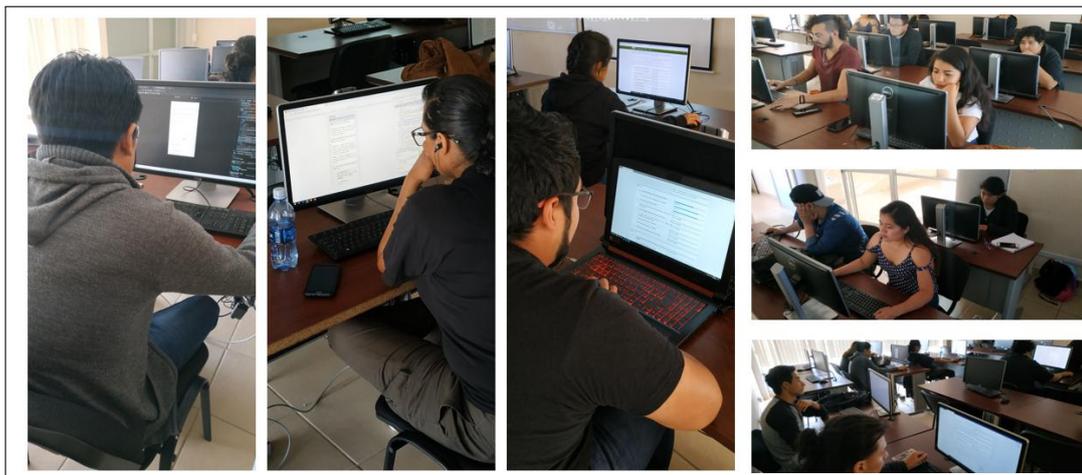


Figura 4.24: Usuarios realizando la evaluación del servicio web

#### 4.1.9 Mantenimiento del servicio web

A través de la aplicación de la evaluación de experiencia del usuario se identificaron las siguientes áreas de oportunidad:

- El usuario requiere mayor número de elementos de ayuda.
- El usuario requiere una mejor retroalimentación sobre las cosas que están sucediendo con los servicios ofrecidos.

Por lo anterior, se implementaron diferentes elementos de ayuda contextual y mensajes de retroalimentación que permitieran al usuario a identificar correctante el tipo de acciones que debería llevar a cabo en cada sección. las figuras 4.25, 4.26 y 4.27 muestran los elementos implementados para atender las necesidades identificadas en la evaluación.

**i Instrucciones**

Ingrese el nombre de una persona y seleccione del listado la relación.

Ejemplo: Juan Pérez Es autor

? ?

¿Tiene duda en como ingresar el nombre?

Nueva búsqueda

Figura 4.25: Iconos para ayuda contextual

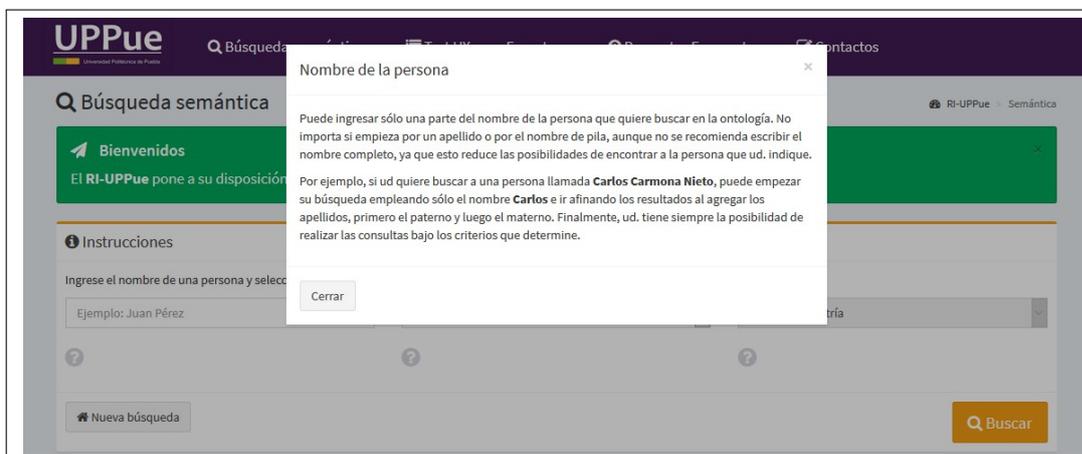


Figura 4.26: Mensaje para la descripción detallada de elementos y ejemplos de uso



Figura 4.27: Mensajes contextuales sobre los servicios ofrecidos en la barra de navegación

- [1] I. Sommerville, "Ingeniería de Software", p.p. 1-63, 2005
- [2] J. A. Merlo, "Ecosistemas del acceso abierto", p.p. 43-44, *Ediciones Universidad de Salamanca*. 2018
- [3] REBUIN, "Beneficios del acceso abierto en tu universidad", *Universidad de Navarra, España*. Disponible en: [http://www2.unavarra.es/gesadj/servicioBiblioteca/acceso\\_abierto/BeneficiosAccesoAbierto.pdf](http://www2.unavarra.es/gesadj/servicioBiblioteca/acceso_abierto/BeneficiosAccesoAbierto.pdf). Fecha de consulta: 2018-10-01
- [4] JISK, "OpenDOAR - The Directory of Open Access Repositories", *University of Nottingham, UK*. Disponible en: <http://www.opendoar.org/>. Fecha de consulta: 2019-06-01
- [5] M. Hagemann, "Iniciativa de Budapest para el acceso abierto". Disponible en: <https://www.budapestopenaccessinitiative.org/translations/spanish-translation>. 2007
- [6] M. A. Medina, J. A. Sánchez, O. Cervantes, R. C. Medina , J. De la Calleja, A. Benítez. "Representción semántica de conocimiento operativo y de dominio para repositorios institucionales". 2017. Registro Público No. 03-2017-042511235500-01
- [7] RAE, "Real Academia de la Lengua Española". Disponible en: <http://www.rae.es/>. Fecha de consulta: 2018-03-07
- [8] J. Polanco-Cortés, "Repositorios digitales. Definición y pautas para su creación". *Universidad de Costa Rica*, Disponible en: <https://ucrindex.ucr.ac.cr/docs/repositorios-digitales-definicion-y-pautas-para-su-creacion.pdf>. Fecha de consulta: 2018-03-07
- [9] Biblioteca Universitaria, "¿Qué es un repositorio institucional?", *Universidad de León*. Disponible en: <https://biblioteca.unileon.es/ayuda-formacion/repositorio-institucional>. Fecha de consulta: 2018-03-12
- [10] REMERI, "Red Mexicana de Repositorios Institucionales". Disponible en: [www.remeri.org.mx/](http://www.remeri.org.mx/). Fecha de consulta: 2018-01-12
- [11] DuraSpace, "DSpace is a turnkey institutional repository application". Disponible en: <http://www.dspace.org/>. 2018
- [12] REMERI, "Directorio de Red Mexicana de Repositorios Institucionales". Disponible en: <http://www.remeri.org.mx/repositorios/>. Fecha de consulta: 2018-01-12
- [13] CONACyT, "Repositorio Nacional". Disponible en: <https://www.repositorionacionalcti.mx/>. Fecha de consulta: 2018-01-25
- [14] Electronics in Computer Science, "EPrints for Open Access", *University of Southampton*. Disponible en: <http://www.eprints.org/uk/>. Fecha de consulta: 2018-01-25
- [15] LYRASIS, "VIVO open source platform". Disponible en: <http://www.vivoweb.org/>. 2018

- [16] S. Blanco. "Anotaciones semánticas en WebQuest". *Universidad de Valladolid*. Disponible en: <https://www2.infor.uva.es/~sblanco/>. Fecha de consulta: 2018-02-21
- [17] N. F. Noy, D. L. McGuinness, "Ontology development 101: a guide to creating your first ontology". *Stanford University*. Disponible en: [stanford.edu/publications/ontologydevelopment/ontology101noymcguinness.html](http://stanford.edu/publications/ontologydevelopment/ontology101noymcguinness.html). Fecha de consulta: 2018-02-21
- [18] K. Rodríguez, R. Ronda, "Web semántica: un nuevo enfoque para la organización y recuperación de información en el web", *SciELO Cuba*, Disponible en: [http://scielo.sld.cu/scielo.php?script=sci\\_arttext&pid=S1024\discretionary{-}{-}{-}94352005000600003](http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S1024\discretionary{-}{-}{-}94352005000600003), Fecha de consulta: 2018-12-11
- [19] P. D. Vázquez, M. A. Medina, "Diseño de un servicio web para la recuperación de información semántica del repositorio institucional de la Universidad Politécnica de Puebla". *Universidad Politécnica de Puebla*. En *Congreso de Instrumentación SOMI*, 2018
- [20] MIT, ERCIM, Keio, Beihang, "W3C Standards", Disponible en: <https://www.w3.org/standards/>. Fecha de consulta: 2018-11-13
- [21] D. J. Soler, "Implementación de un servicio web para la transferencia de los resultados de la investigación de la región de Murcia mediante tecnologías de la web semántica y de gestión de contenidos". *Universidad de Murcia, Esp.*, Disponible en: <https://digitum.um.es/xmlui/handle/10201/54364>, Fecha de consulta: 2018-10-09
- [22] RDF Working Group, "Resource Description Framework", Disponible en: <https://www.w3.org/2001/sw/wiki/RDF>, 2018-11-13
- [23] D. Brickley, "RDF Schema 1.1", Disponible en: <https://www.w3.org/TR/rdf-schema/>, 2014
- [24] R. Angles, "Introducción a las bases de datos RDF", Disponible en: <http://renzoangles.net/files/libro.pdf>, 2015
- [25] OpenLink Software, "About OpenLink Virtuoso", Disponible en: <https://virtuoso.openlinksw.com/>, 2015
- [26] F. Bauer, M. Kaltenböc, "Linked Open Data: The Essentials", Disponible en: *Graphik-service GmbH*, 2012
- [27] OntoText, "What is a semantic repository", Disponible en: <https://ontotext.com/knowledgehub/fundamentals/semantic-repository/>, 2018
- [28] RedCLARA, "La Referencia, Red de Repositorios de Acceso Abierto a la Ciencia", Disponible en: <http://www.lareferencia.info/es/>, 2019
- [29] Infotec, "Descripción de servicios REST, Catálogos del repositorio nacional", Disponible en: <http://catalogs.repositorionacionalcti.mx/webresources/>, 2018

- [30] C. Á. Caules, "Introducción a Servicios REST", Disponible en: <https://www.arquitecturajava.com/servicios-rest/>, 2013
- [31] J. Texier, "Repositorio Semántico para la Universidad Nacional de Chilecito", Disponible en: [http://sedici.unlp.edu.ar/bitstream/handle/10915/62792/Documento\\_completo.pdf-PDFA.pdf?sequence=1](http://sedici.unlp.edu.ar/bitstream/handle/10915/62792/Documento_completo.pdf-PDFA.pdf?sequence=1), 2017
- [32] W3C, "SPARQL Query Language for RDF", Disponible en: <https://www.w3.org/TR/rdf-sparql-query/>, 2008
- [33] D. Robinson, "The Incredible Growth of Python", Disponible en: <https://stackoverflow.blog/2017/09/06/incredible-growth-python/>, 2017
- [34] Next Generation Repository Working Group, "Next Generation Repositories: Behaviours and Technical Recommendations of the COAR", *Confederation of Open Access Repositories*, Disponible en: <https://www.coar-repositories.org/activities/advocacy-leadership/working-group-next-generation-repositories/>, 2017
- [35] Ll. Codina, C. Rovira, "La Web semántica", Disponible en: [http://eprints.rclis.org/8899/1/web\\_semantica.pdf/](http://eprints.rclis.org/8899/1/web_semantica.pdf/), 2006
- [36] The Apache Software Foundation, "Apache Jena Fuseki", Disponible en: <https://jena.apache.org/documentation/fuseki2/>, 2019
- [37] OpenSemanticFramework.org, "RDFizer Concept", <http://wiki.opensemanticframework.org/index.php/Concept>, 2019
- [38] Eikeon, Gromgull, Joernhees, "RDFLib Library version 4.2.2", Disponible en: <https://pypi.org/project/rdflib/>, 2019
- [39] Lamy, Jean-Baptiste, "OWLReady2 Library version 0.19", Disponible en: <https://pypi.org/project/owllready2/>, 2019
- [40] D . Torres-Burriel, "Pruebas de usabilidad con usuarios", Disponible en: <https://www.torresburriel.com/>, 2008
- [41] F. Lundh, "Elementtree 1.2.6", Disponible en: <https://pypi.org/project/elementtree/>, 2005
- [42] P. Suber, "Open Access Overview", Disponible en: <http://legacy.earlham.edu/~peters/fos/overview.htm>, 2015
- [43] C. Lagoze, H. Van de Sompel, M. Nelson, S. Warner, "Implementation Guidelines for the Open Archives Initiative Protocol for Metadata Harvesting", Disponible en: <http://www.openarchives.org/OAI/2.0/guidelines.htm>, 2005
- [44] DCMI, "Dublin Core Metadata Initiative", Disponible en: <https://www.dublincore.org/>, 2019

- [45] M. J. Lamarca, "Hipertexto, el nuevo concepto de documento en la cultura de la imagen". Disponible en: <http://www.hipertexto.info>, 2013
- [46] Elsevier Connect, "Tipos de Open Access: diferencias entre la vía verde y la vía dorada", *Elsevier*. Disponible en: <https://www.elsevier.com/es-es/connect/actualidad-sanitaria/tipos-de-open-access-via-verde-y-la-via-dorada>, 2019
- [47] M.A. Medina, J.A. Sánchez, O. Cervantes, J. De la Calleja, A. Benitez, "Representación semántica de conocimiento operativo y de dominio para repositorios institucionales", *Universidad Politécnica de Puebla*, Reg. Pub. 03-2017-042511235500-01, 2017
- [48] . Iniciativa de Budapest para el Acceso Abierto. Disponible en: <http://www.budapestopenaccessinitiative.org/translations/spanish-translation>. Vol 1. 2010.
- [49] M. Medina, G. De la calleja, C. Zepeda, A. Benitez y R. Medina, "Evaluación de usabilidad de las interfaces para depositar objetos de aprendizaje en EPrints y DSpace", *Revista Visión e Innovación Politécnica*, 2018
- [50] . INFOTEC. Descripción de servicios REST. Catálogos del repositorio nacional. Disponible en: <http://catalogs.repositorionacionalcti.mx/webresources/>. 2018.

