

HUMANIDADES, CIENCIA, TECNOLOGÍA E INNOVACIÓN EN PUEBLA

ACADEMIA JOURNALS



OPUS PRO SCIENTIA ET STUDIUM

ISSN 2644-0903 online

VOL. 2, NO. 1, 2020

WWW.ACADEMIAJOURNALS.COM

TRABAJO DE INVESTIGACIÓN AUSPICIADO POR EL CONVENIO CONCYTEP-ACADEMIA JOURNALS



PAULO DANIEL VÁZQUEZ MORA

SERVICIO WEB PARA LA EXTRACCIÓN DE INFORMACIÓN SEMÁNTICA DEL RI-UPPUE

UNIVERSIDAD POLITÉCNICA DE PUEBLA

COMITÉ EVALUADOR:

DRA. MARÍA AUXILIO MEDINA NIETO

DRA. MIREYA TOVAR VIDAL

M.C. REBECA RODRÍGUEZ HUESCA

M.C. ANTONIO FELIPE RAZO RODRÍGUEZ

NÚMERO DE SECUENCIA 2-58



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

Servicio web para la extracción de información semántica del RI-UPPue

TESIS REALIZADA POR:

ING. PAULO DANIEL VÁZQUEZ MORA

12 de Diciembre del 2019

Comité evaluador

(Firma)

Dra. María Auxilio Medina Nieto

Dra. Mireya Tovar Vidal

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

Servicio web para la extracción de información semántica del RI-UPPue

ING. PAULO DANIEL VÁZQUEZ MORA

La tesis describe diferentes interfaces de búsqueda básica y avanzada de repositorios abiertos que distribuyen documentos en español. Se enfoca en la interfaz del Repositorio de la Universidad Politécnica de Puebla, soportado en la plataforma DSpace, la cual permite recuperar una tesis buscando por autor pero se desea ampliar la búsqueda considerando a los profesores que la dirigieron o revisaron. Para atender esta problemática, la tesis describe el diseño y la implementación de un servicio web denominado SW001, éste recupera información semántica utilizando un modelo semántico u ontología que representa formalmente a personas, documentos y sus relaciones. El servicio es tipo REST, emplea consultas en SPARQL y otras herramientas tecnológicas como el módulo de serialización RDF. El servicio establece un vínculo entre el repositorio y la ontología, dado que en la medida en que se incrementa el número de tesis en el repositorio y se exporten sus metadatos, éstos se insertan y gestionan como instancias. Se presenta un sitio web para acceder al servicio y la aplicación web TestUX diseñada para evaluar su usabilidad mediante un conjunto de heurísticas; los resultados experimentales muestran un nivel de satisfacción alto. Desde el sitio se puede descargar la ontología con las instancias en los formatos JSON y OWL. La tesis provee las bases para el desarrollo de otros servicios, contribuye a difundir los beneficios del acceso abierto y en la generación de datos semánticamente enriquecidos.

Agradecimientos

A mis maestros, amigos, compañeros de trabajo, alumnos, a mi Universidad, pero sobre todo a mis hijos y mi amada esposa.

Sin su apoyo nada de esto hubiera sido posible.

Ing. Paulo Daniel Vázquez Mora

Esta investigación fue realizada gracias al apoyo del Consejo de Ciencia y Tecnología del
Estado de Puebla

1 Planteamiento del problema	12
1.1 Introducción	12
1.2 Objetivo general	14
1.3 Objetivos específicos	14
1.4 Justificación	14
2 Marco teórico	16
2.1 La importancia de los metadatos	16
2.2 RDF en el contexto de la web semántica	16
2.3 SPARQL: lenguaje de consulta para datos en RDF	19
2.4 Servicios REST	19
2.5 Características de los repositorios según COAR	22
2.6 Trabajos relacionados	23
2.6.1 Servicios de búsqueda simple, avanzada y semántica	25
2.6.2 Plataformas que procesan consultas semánticas	26
3 Metodología	31
3.1 Modelo de desarrollo del servicio web SW001	31
3.2 Análisis de requerimientos	31
3.3 Diseño	32
3.3.1 Casos de uso	32
3.3.2 Diagrama de clases	32
3.3.3 Diseño de alto nivel	33
3.4 Implementación	34
3.4.1 Etapa 1: exportación de la colección de tesis	35
3.4.2 Etapa 2: instalación de Apache Jena Fuseki	35
3.4.3 Etapa 3: ejecución del servidor Jena Fuseki	35
3.4.4 Etapa 4: ejecución del servicio RDFizer	36
3.4.5 Etapa 5: creación del almacén de ternas	38
3.4.6 Etapa 6: obtención de ternas	38
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	39
3.4.7.3 Librería RDFLib	39
3.4.7.4 Librería OWLReady2	40
3.4.7.5 Librería xml.etree.ElementTree	40
3.4.7.6 Marco de trabajo Flask	41
3.4.7.7 AdminLTE	41
3.5 Evaluación	41
3.6 Mantenimiento	42

4 Resultados	43
4.1 Captura de metadatos en DSpace 6.2	43
4.2 Creación de almacén de ternas	44
4.3 Generación de documentos en RDF	45
4.4 Extracción y procesamiento de metadatos	46
4.5 Lectura e inserción de instancias en la ontología Onto4UPPue	48
4.6 Implementación del servicio web	49
4.7 Evaluación y mantenimiento	50
4.7.1 Evaluación de la usabilidad del sitio web para acceder al servicio SW001	51
5 Conclusiones	57
Referencias	59

2.1	Modelo de capas para la web semántica propuesto por Berners-Lee [17] . . .	17
2.2	Grafo dirigido que representa a una persona en RDF	19
2.3	Conjunto de ternas que representan un recurso digital en RDF	19
2.4	Arquitectura cliente servidor de servicio REST [29]	20
2.5	Servicios sin estado de la arquitectura cliente servidor REST [29]	20
2.6	Servicios REST apoyados en caché [29]	20
2.7	Servicios REST uniformes [29]	21
2.8	Arquitectura de capas de los servicios REST [29]	21
2.9	Resultado del servicio áreas de conocimiento del RN	22
2.10	Arquitectura de servicio web de búsqueda semántica SW001.	24
2.11	Estadísticas generales de LA Referencia. Fecha de consulta: 9 de Noviembre del 2019	24
2.12	Repositorios por país [4]	25
2.13	Repositorios por lenguaje y tipo de software [4]	25
2.14	Repositorios con mayor producción en Latinoamérica [27]	26
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	27
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	28
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	29
2.22	Vista del panel de búsqueda en el portal instituto Brasileño de Información en Ciencia y Tecnología	29
2.23	Servicios de búsqueda en RDs de AA	30
3.1	Modelo de desarrollo del servicio web SW001	31
3.2	Diagrama de casos de uso	32
3.3	Diagrama de clases. Fuente: [18]	33
3.4	Diseño de consumo del servicio web. Fuente: [18]	33
3.5	Diseño de alto nivel del servicio web. Fuente: [18]	34
3.6	Módulos adicionales de DSpace para la versión 6.2	34
3.7	Etapas del proceso de migración de metadatos del RI-UPPue a la web semántica	35
3.8	Exportación de la colección tesis a archivo CSV	36
3.9	Opciones de descarga para el almacén de ternas Jena Fuseki	36
3.10	Contenido de la carpeta del servidor <i>Jena Fuseki</i>	37
3.11	Ejecución del servidor Jena Fuseki en terminal de Windows	37

3.12	Extracción, transformación y carga	37
3.13	Acceso a metadatos de tesis en RDF	38
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	40
3.17	Reactivos para evaluar la heurística generalidades [39]	42
4.1	Tareas implementadas para el desarrollo del servicio web SW001.	43
4.2	Creación de la comunidad <i>Tesis</i>	44
4.3	Creación de la colección <i>Maestría</i>	44
4.4	Elementos de la colección <i>Maestría</i>	45
4.5	Actividades para crear el almacén de ternas	45
4.6	Ejecución local del servidor Apache Jena Fuseki	46
4.7	Ejecución del serializador <i>RDFizer</i>	46
4.8	Código fuente del servicio para verificar metadatos de documentos en RDF	47
4.9	Metadatos recuperados de un documento en RDF para una tesis	47
4.10	Sección de código fuente para insertar instancias en la ontología	48
4.11	Resultado de la inserción de una instancia	49
4.12	Tecnologías para el desarrollo del servicio web SW001	49
4.13	Sección de inicio del servicio web SW001	50
4.14	Sección de búsquedas semánticas, acceso al servicio web SW001	51
4.15	Tabla de resultados de la búsqueda semántica de la Figura 4.14.	51
4.16	Descarga de la ontología <i>Onto4UPPue</i> en formato JSON u OWL	52
4.17	Sección de preguntas frecuentes	52
4.18	Sección de contacto	52
4.19	Acceso al servicio web SW001 desde un dispositivo móvil	53
4.20	Descripción de la instancia T24 en la ontología <i>Onto4UPPue</i>	53
4.21	Pantalla principal de <i>TestUX</i>	54
4.22	Resultados de la evaluación por heurística del sitio web	54
4.23	Integración de resultados por heurística	55
4.24	Estudiantes utilizando la aplicación web <i>TestUX</i>	55
4.25	Íconos para ayuda contextual	56
4.26	Mensaje para la descripción detallada de elementos y ejemplos de uso	56
4.27	Mensaje descriptivo del servicio web SW001 en la barra de navegación	56

2.1	Servicios REST ofrecidos por el RN	22
2.2	Características técnicas entre DSpace, Virtuoso y VIVO	28
3.1	Tabla de requerimientos funcionales del servicio web	32
3.2	Escala de Likert empleada para evaluar la interfaz gráfica del servicio web	42
4.1	Actividades para verificar la generación de documentos en RDF	46
4.2	Actividades de verificación del servicio de la Figura 4.8	48
4.3	Actividades de verificación con archivos CSV	48

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 [41]. Existen diversos tratados y convenciones internacionales que promueven el AA del público en general a la literatura científica y académica 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 [5]. 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, se consideran un medio de divulgación (*ruta verde*) o de publicación (*ruta dorada*) de los contenidos producidos por una institución o comunidad [45].

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 diseminación de metadatos, *Open Archives Initiative Protocol for Metadata Harvesting*, (OAI-PMH) [42], 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. Al 9 de Noviembre del 2019, la página web del Repositorio Nacional (RN) de este país reporta la existencia de 105 repositorios de Ciencia Abierta (CA) [12] e INDEXE, Sistema

de Búsqueda de la Red Mexicana de Repositorios Institucionales (REMERI) [11] 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 que cuenta con el mayor número de repositorios y documentos publicados [11].

Según datos de OpenDOAR [4], la distribución del *software* empleado para la implementación de los 3,779 RIs es como sigue: *DSpace* [10] (44.2%), *EPrints* (13.4%), *Digital Commons* (4.7%) y *WEKO* (2.7%), el porcentaje restante pertenece a plataformas propias. 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* [13] 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* [10] surgió como proyecto del 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; más del 90% de los RIs que interoperan con el RN [12] emplean una versión de *DSpace*. Una comparación que considera aspectos de funcionalidad, elementos técnicos y de administración entre *EPrints* y *DSpace* se presenta en [46].

En *EPrints* y *DSpace* la interoperabilidad por omisión se implementa al utilizar el protocolo OAI-PMH [42] y el estándar de metadatos Dublin Core [43]. En el RI de la Universidad Politécnica de Puebla, en adelante RI-UPPue, está soportado en la versión 6.2 de *DSpace*.

Esta tesis identifica como problemática 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 en el caso de un artículo
- Existe *ambigüedad* en la interpretación de los datos descriptivos o *metadatos* cuando se depositan documentos, por ejemplo, qué se debe colocar dentro del elemento *contributor* al almacenar una tesis de maestría, cartel (póster) o artículo
- La interpretación de los datos exportados de *DSpace* está sujeta a los usuarios finales y es dependiente de cada IES. Por omisión, los datos se exportan en archivos

con extensión CSV¹.

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 [?]. La tesis propone utilizar una instancia de esta ontología identificada como *Onto4UPPue*, la cual está escrita en el estándar marco de descripción de recursos, *Resource Description Framework* (RDF) [21], uso del protocolo y lenguaje de consulta para RDF, SPARQL [31] y otras 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 la ontología *Onto4UPPue* validando su consistencia lógica de forma automática mediante razonadores
- Diseñar e implementar un servicio web tipo REST que permita recuperar información semántica utilizando la ontología *Onto4UPPue* y consultas en SPARQL

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 [15] y [16], el uso de tecnologías semánticas como las ontologías ofrece beneficios como los siguientes:

- Intercambio de datos entre aplicaciones, programas y/o plataformas mediante el lenguaje XML² 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

¹CSV corresponde a las siglas del formato *Comma Separated Values*, valores separados por comas

²XML son las siglas de *eXtensible Markup Language*

- Uso de la ontología Onto4UPPue como alternativa de representación de información semántica de 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. Después presenta el modelo de datos RDF y su contexto en la web semántica, seguido de una breve descripción de SPARQL. Posteriormente, se enumeran las características principales de los servicios REST¹ junto con los requerimientos para repositorios de acuerdo con la *Confederation of Open Access Repositories*, Confederación de Repositorios de Acceso Abierto (COAR). Finalmente, se describen 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 usuarios específicos, computadoras y aplicaciones móviles.

Según [19], una prioridad particular del *World Wide Web Consortium* (W3C) o consorcio web, es usar la web para documentar el significado de los metadatos. La relevancia de la gestión, uso y representación de los metadatos en modelos semánticos u ontologías, se relaciona directamente con RDF. La sección 2.2 describe algunas de sus características principales 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 los 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. [23] afirma que 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.

¹Las siglas REST corresponden a la expresión en inglés *Representational State Transfer*

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.

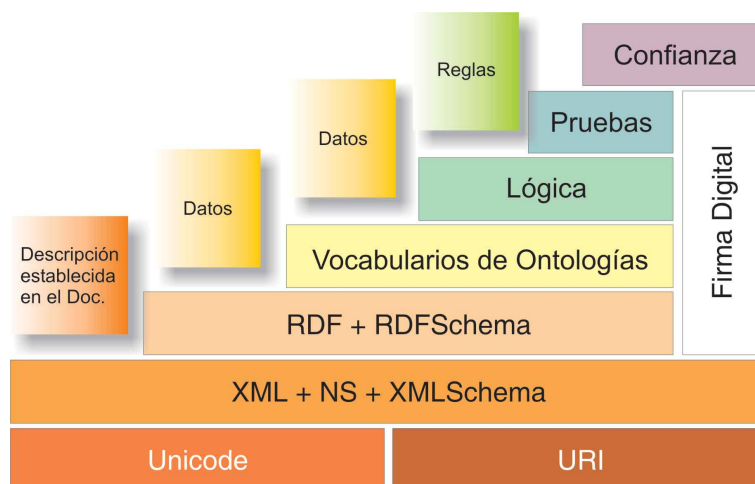


Figura 2.1: Modelo de capas para la web semántica propuesto por Berners-Lee [17]

Una descripción amplia y detallada de los primeros niveles está disponible en [19] y [23]. 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, los recursos se identifican de forma unívoca. Un *Uniform Resource Locator (URL)*, localizador de recursos único, es un tipo de URI
- *eXtensible Markup Language (XML)*. Lenguaje de marcado extensible o lenguaje de marcas extensible, es una especificación del W3C de propósito general en la que cada usuario define sus propias etiquetas.
- *XML namespaces*. Los espacios de nombre se emplean para atender el problema de ambigüedad en los documentos XML, de manera que cada elemento está identificado por un URI que lo hace único y universal. Los espacios de nombre son recomendación del W3C.

De acuerdo con [21] y [34], 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 diferentes esquemas. RDF amplía la estructura de enlaces web (*links*) para nombrar a dos recursos denominados *sujeto* y *objeto*, así como a la relación entre ellos (*predicado*). Los tres elementos forman enunciados conocidos también como *termas* o *triplezas*. RDF representa datos estructurados y semi-estructurados que se comparten entre aplicaciones de software.

- *RDF Schema*. Proporciona un vocabulario para el modelo de datos RDF, describe los conceptos básicos, la sintaxis abstracta y determina las características de los recursos como dominio y rango de propiedades
- *Ontología*. Define los términos y las relaciones básicas para la comprensión de un área del conocimiento o dominio, así como las reglas para combinar estos términos
- *Vocabularios de ontologías*. RDF, RDF-Schema y el lenguaje de ontologías web OWL (*Ontologies Web Language*) son ejemplos de vocabularios establecidos por el W3C que representan ontologías
- *Capa lógica*. Permite determinar si la estructura de los razonamientos es válida a través del estudio formal de reglas. 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 confiabilidad a determinados recursos, de forma comprobable por agentes. Esta capa usa firmas digitales y redes de confianza

Las ontologías se refieren a modelos de representación de conocimiento, se componen de conceptos, individuos, relaciones entre sí y reglas que al integrarse, adquieren una interpretación única para un conjunto de personas o computadoras. Según [44], 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.

La ontología Onto4UPPue está escrita en RDF y RDF-Schema, sin embargo, existen otros lenguajes como OWL, el cual 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*. En síntesis, los lenguajes utilizados ampliamente para representar los datos en la web semántica son XML, RDF y OWL.

RDF es un modelo de datos empleado para representar información personal, datos de redes sociales, metadatos sobre objetos digitales o como medio para la integración de fuentes de información heterogéneas. RDF se asocia con diferentes representaciones, una de ellas es como un modelo de grafos dirigidos etiquetados que representan información en la web [21]. La Figura 2.2 muestra un ejemplo de esta representación, se utiliza el vocabulario vCard disponible en <https://www.w3.org/TR/2001/NOTE-vcard-rdf-20010222/>. Una representación como conjunto de ternas se muestra en la Figura 2.3, (sintaxis Turtle).

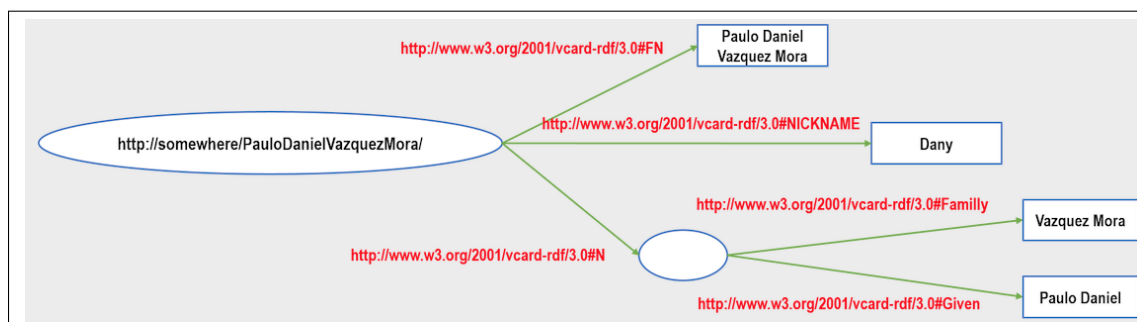


Figura 2.2: Grafo dirigido que representa a una persona en RDF

```

@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: Conjunto de ternas que representan un recurso digital en RDF

2.3 SPARQL: lenguaje de consulta para datos en RDF

Los datos en RDF se recuperan utilizando el lenguaje de consulta SPARQL, éste devuelve resultados en forma de enlaces o esquemas RDF. Entre sus especificaciones, de acuerdo con [31], se encuentran las siguientes:

- 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 especificación del protocolo SPARQL para RDF *SPROT* define el protocolo remoto para enviar consultas SPARQL y recibir los resultados

El capítulo 4 muestra ejemplos de consultas en este lenguaje que recuperan datos de la ontología Onto4UPPUE.

2.4 Servicios REST

[29] establece que a partir de la migración a la web 2.0, los servicios REST sustituyen a los servicios que hacían uso de SOAP² y WSDL³, su arquitectura está orientada a recursos, hacen uso del protocolo HTTP y se caracterizan por lo siguiente [29]:

²Siglas de *Simple Object Access Protocol*, protocolo simple de acceso a objetos

³Abreviación de *Web Services Description Language*, lenguaje de descripción de servicios web

- Definen una interfaz de comunicación *cliente-servidor* que separa sus responsabilidades, (ver la Figura 2.4).



Figura 2.4: *Arquitectura cliente servidor de servicio REST [29]*

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

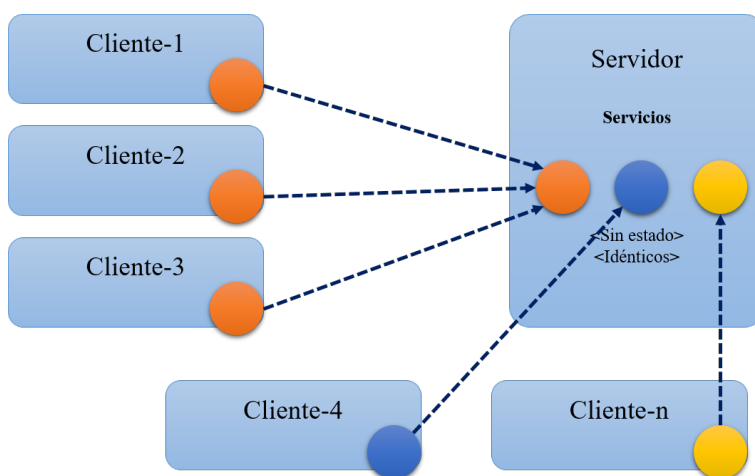


Figura 2.5: *Servicios sin estado de la arquitectura cliente servidor REST [29]*

- Guardan en caché su contenido de manera que una vez realizada la primera petición (1), el resto de peticiones se apoya en caché si fuera necesario (2), (ver la Figura 2.6).

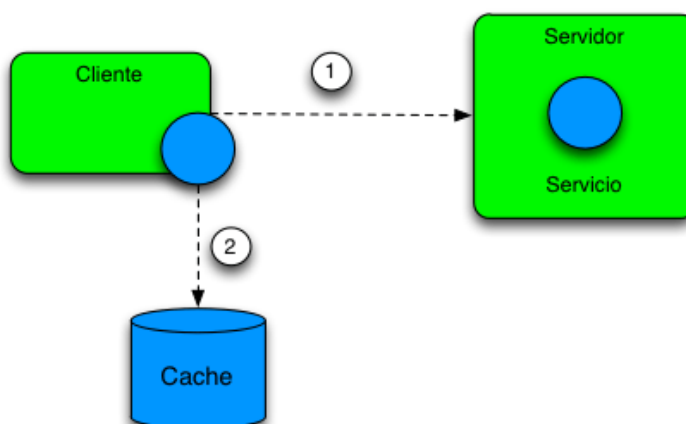


Figura 2.6: *Servicios REST apoyados en caché [29]*

- Son *uniformes*, esto es, comparten una forma de invocación y métodos GET, POST, PUT, DELETE, (ver la Figura 2.7).

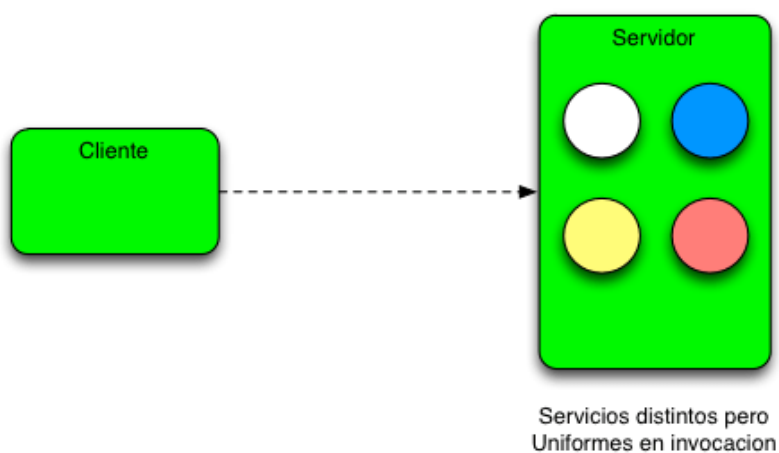


Figura 2.7: *Servicios REST uniformes [29]*

- Son escalables y 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 (balanceador), (ver la Figura 2.8).

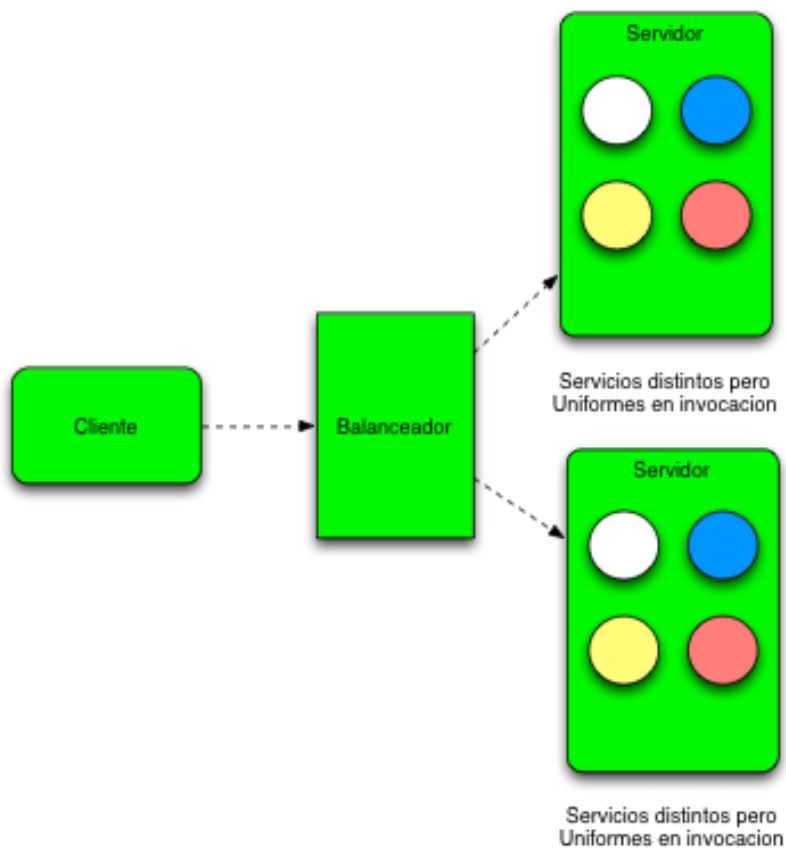


Figura 2.8: *Arquitectura de capas de los servicios REST [29]*

El RN tiene un catálogo de más de dos cientos servicios REST que regresan los resultados en formato *JavaScript Object Notation* (JSON) [28], la Tabla 2.1 contiene algunos de éstos; la Figura 2.9 muestra un ejemplo del resultado del primero, (áreas de

conocimiento).

Tabla 2.1: Servicios REST ofrecidos por el RN

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

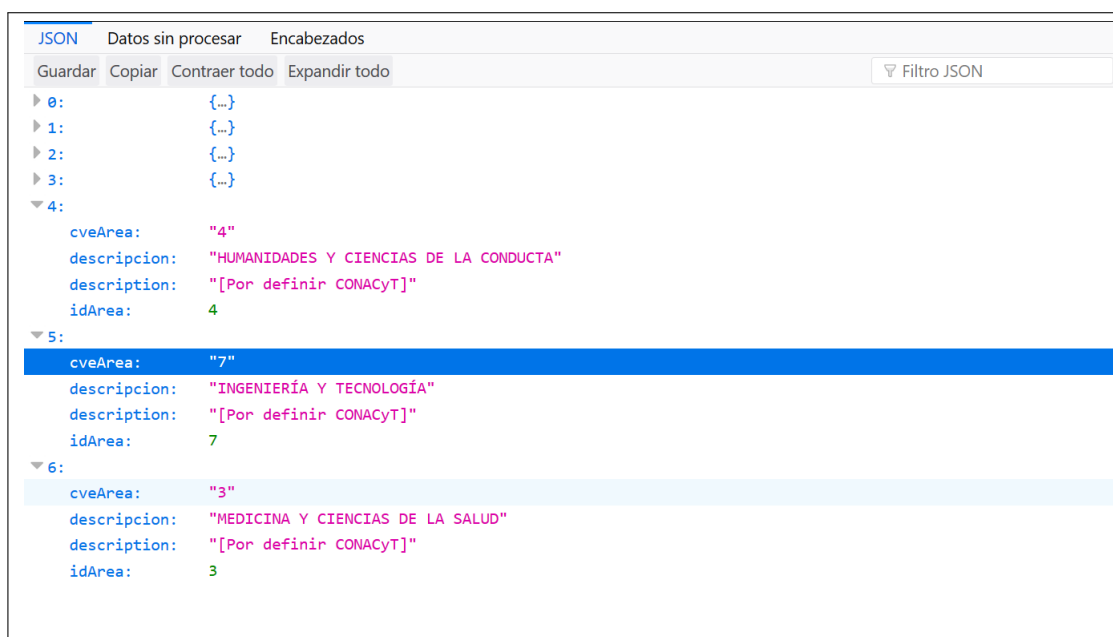


Figura 2.9: Resultado del servicio áreas de conocimiento del RN

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 usan HTTP para obtener datos o generar operaciones con esos datos.

A la fecha de elaboración de la tesis, los repositorios institucionales (RIs) que interoperan con el RN y que cuentan con servicios REST, implementan consultas que recuperan un sólo un atributo o característica de los documentos que coinciden con un conjunto de palabras clave.

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. En 2016, la Confederación de Repositorios de Acceso Abierto, *Confederation of Open Access Repositories* (COAR) [33], integró el equipo de trabajo *Next*

Generation Repository Working Group con el propósito definir las funcionalidades y tecnologías a desarrollar en los RIs y repositorios digitales (RDs) para los próximos años. El informe *Behaviours and Technical Recommendations of the COAR Next Generation Repositories* [33] 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 la comunicación académica y serán fuentes sólidas de publicación, confiables y de AA. Las características propuestas en el informe son las siguientes [33]:

- Exponer *identificadores*
- Declaración de *licencias* a nivel de recursos
- Descubrimientos a través de la *navegación*
- Interacción 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 extiende los mecanismos de búsqueda de tesis del RI-UPPue utilizando tecnologías semánticas. El servicio se identifica con la etiqueta SW001, siglas de *servicio web 001*.

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, [20] 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

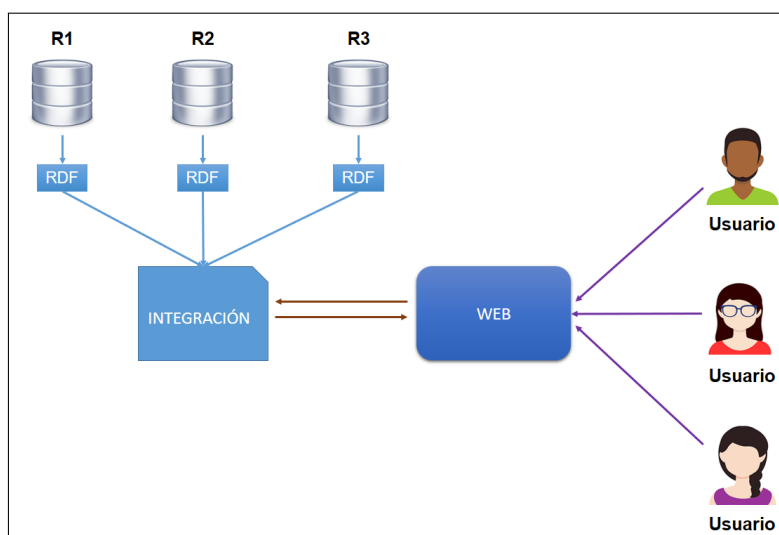


Figura 2.10: *Arquitectura de servicio web de búsqueda semántica SW001.*

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-Referencia⁴, Red de Repositorios de Acceso Abierto a la Ciencia [27], brinda un espacio 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.

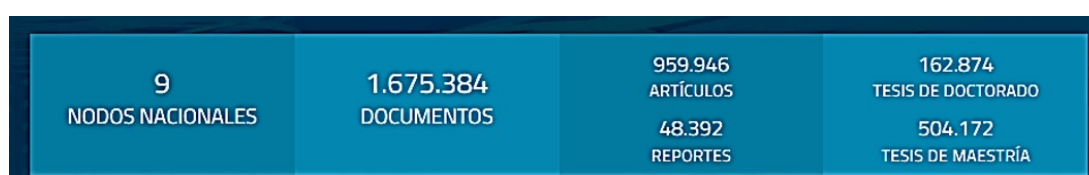


Figura 2.11: *Estadísticas generales de LA Referencia. Fecha de consulta: 9 de Noviembre del 2019*

OpenDOAR⁵ 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 documentos que almacenan. Al mes de Octubre de 2019, OpenDOAR reporta la estadística por país (Figura 2.12) y tipo de plataformas tecnológicas empleadas por los repositorios de acceso abierto (RAAs) (Figura 2.13).

En Latinoamérica, LA Referencia indica la producción de recursos de AA que muestra

⁴Disponible en: <http://www.lareferencia.info/es/>

⁵Disponible en: <http://v2.sherpa.ac.uk/pendoar/>

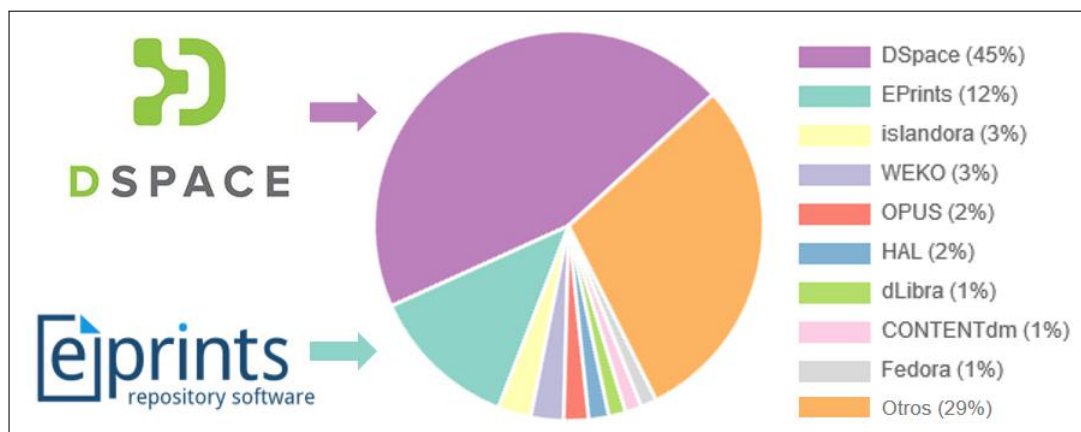


Figura 2.12: Repositorios por país [4]

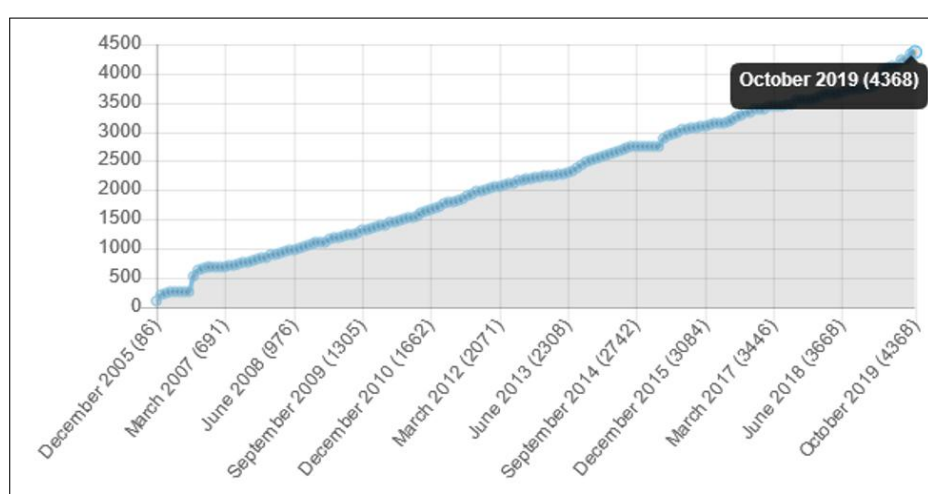


Figura 2.13: Repositorios por lenguaje y tipo de software [4]

la Figura 2.14 y el tipo de servicios de búsqueda disponibles para los usuarios. A la fecha de elaboración de esta tesis, los repositorios enumerados en esta figura pertenecen a LA-Referencia, sus sitios web no indican la implementación de mecanismos de recuperación de información semántica. Las figuras 2.15 a 2.20 muestran las páginas iniciales de algunos de estos repositorios y secciones de la interfaz correspondientes a sus módulos de búsqueda; 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.

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 y fecha de publicación. 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.

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 [27]

Argentina.gov.ar

Inicio / Ministerio de Educación, Cultura, Ciencia y Tecnología / Ciencia, Tecnología e Innovación Productiva / Sistema Nacional de Repositorios Digitales

Sistema Nacional de Repositorios Digitales

La producción científica argentina al alcance de todos

¿Qué estás buscando?

Todos los Campos

176.927 publicaciones
 31 repositorios
 30 instituciones

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

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
- 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 2.3



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



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

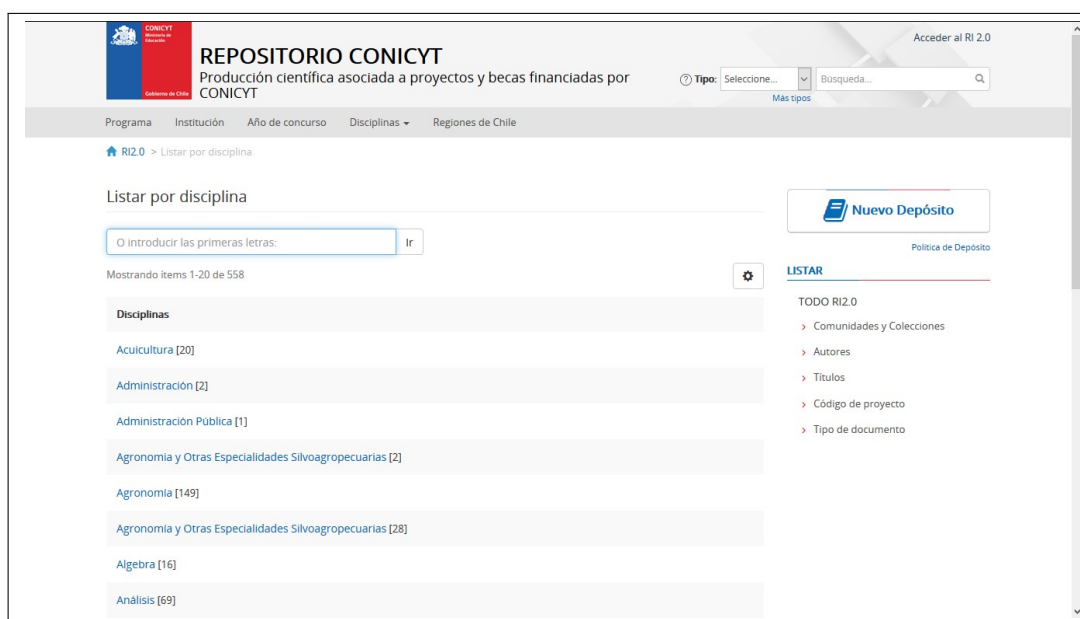


Figura 2.18: Vista del panel de búsquedas avanzadas en RI2.0



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

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



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

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

El capítulo 3 se organiza como sigue: al inicio presenta el modelo de desarrollo para el 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 describe una herramienta de *evaluación* para estimar la usabilidad del servicio web SW001 y se enumeran las tareas de *mantenimiento*.

3.1 Modelo de desarrollo del servicio web SW001

La Figura 3.1 muestra el modelo de desarrollo del servicio web SW001, éste utiliza el modelo de cascada como base [1], en particular en 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.

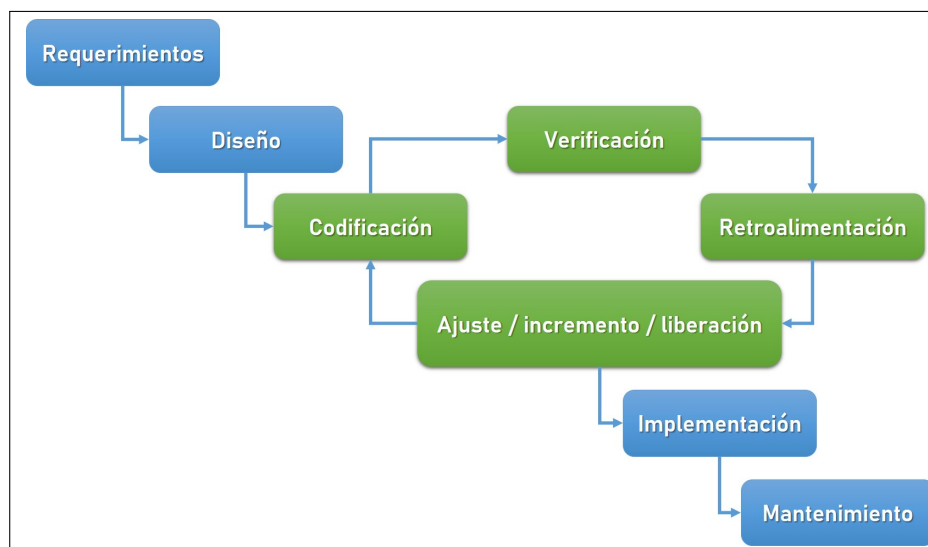


Figura 3.1: Modelo de desarrollo del servicio web SW001

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 de este RI, 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.

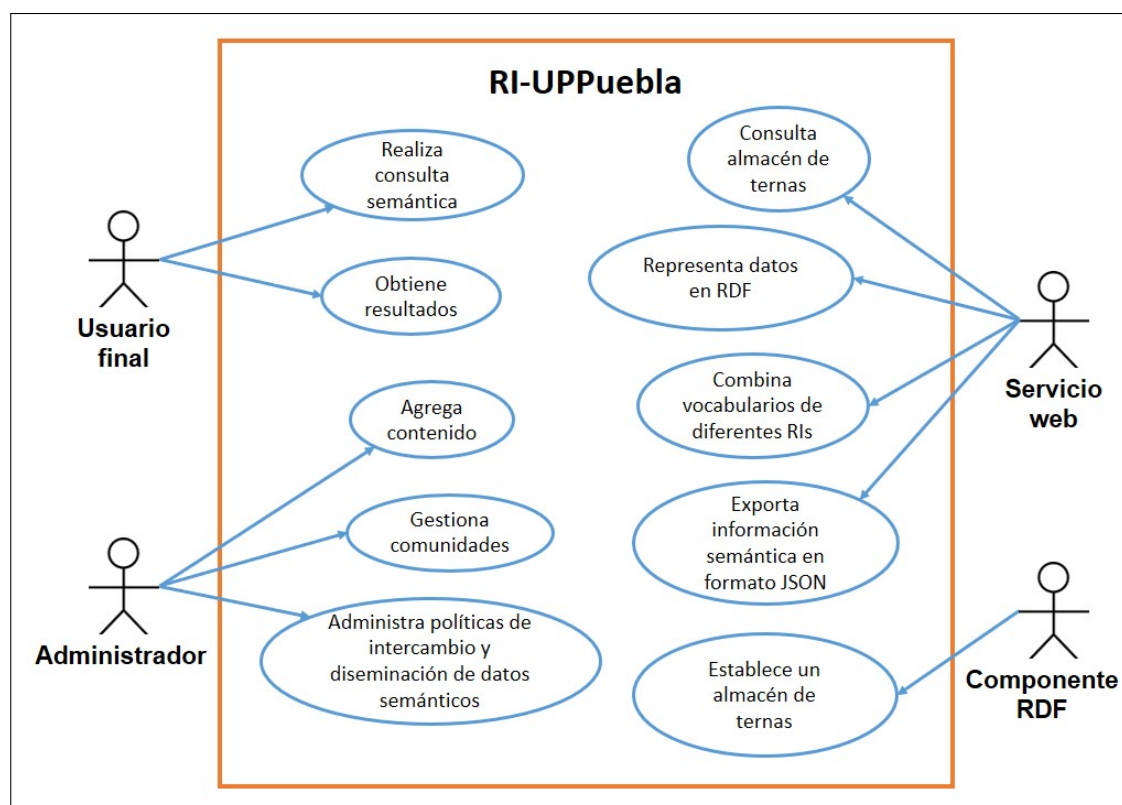


Figura 3.2: *Diagrama de casos de uso*

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 [18],

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

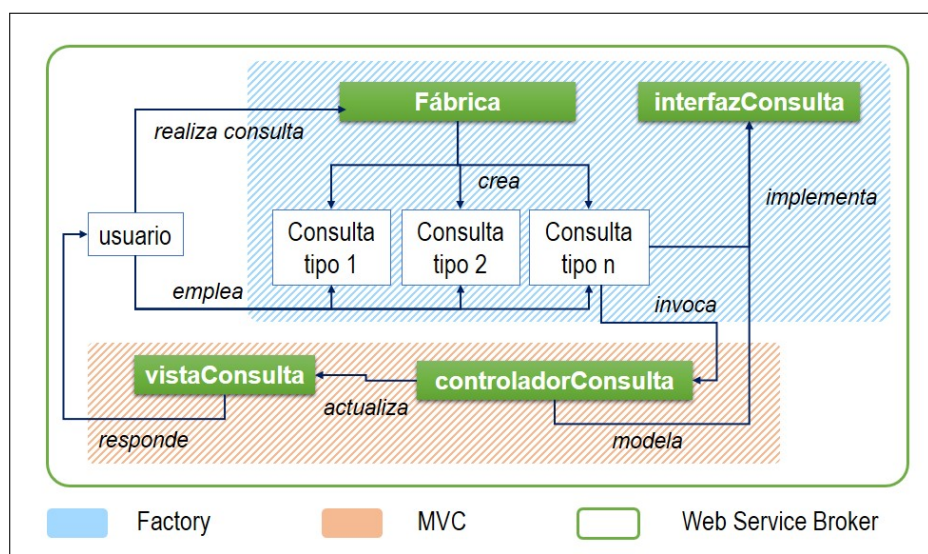


Figura 3.3: Diagrama de clases. Fuente: [18]

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 con datos del RI-UPPue y a mediano plazo, se buscará su instalación en diferentes RIs que compartan la plataforma tecnológica DSpace.

3.4 Implementación

Previo a la implementación de 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 [10] indica que DSpace cuenta con un módulo RDF, sin embargo, éste no se habilita durante la instalación estándar.

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 conversor a formato RDF. La instalación de estos elementos forman parte de las etapas del proceso técnico que muestra la Figura 3.7.

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 50 tesis aproximadamente del Departamento de Posgrado. La Figura 3.8 muestra la interfaz para esta tarea, los metadatos se almacenan



Figura 3.4: *Diseño de consumo del servicio web. Fuente: [18]*

a un archivo con extensión CSV¹. El número 1 muestra el botón para exportar y el 2 el archivo CSV con los metadatos exportados. La 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 TDB² para proporcionar una capa de almacenamiento persistente transaccional y robusta, permite consultas de texto y consultas espaciales [35]. La página oficial de Apache Jena Fuseki³ provee diversas opciones para la descarga dependiendo del sistema operativo: Linux (.tar.gz) o Windows(.zip). Para terminar la instalación, los archivos descargados se descomprimen y se guardan en una carpeta que se nombra como *Fuseki*.

¹Siglas del inglés de *Comma Separated Values*, formato separado por comas

²Componente de Jena Fuseki para el almacenamiento y consulta de datos en RDF

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



Figura 3.5: Diseño de alto nivel del servicio web. Fuente: [18]

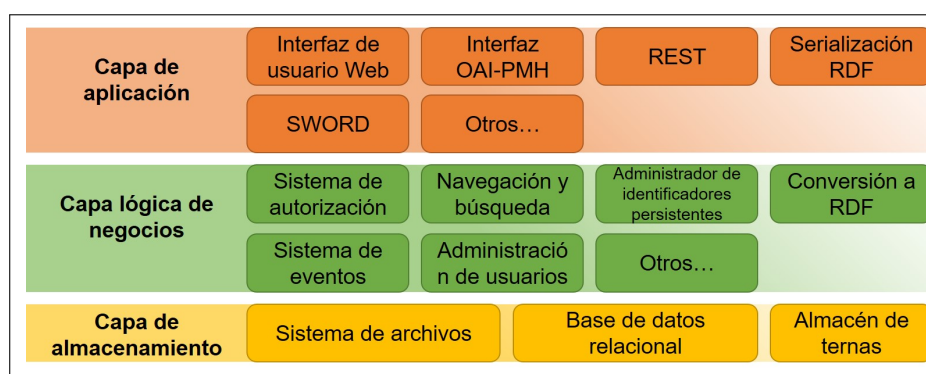


Figura 3.6: Módulos adicionales de DSpace para la versión 6.2

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.

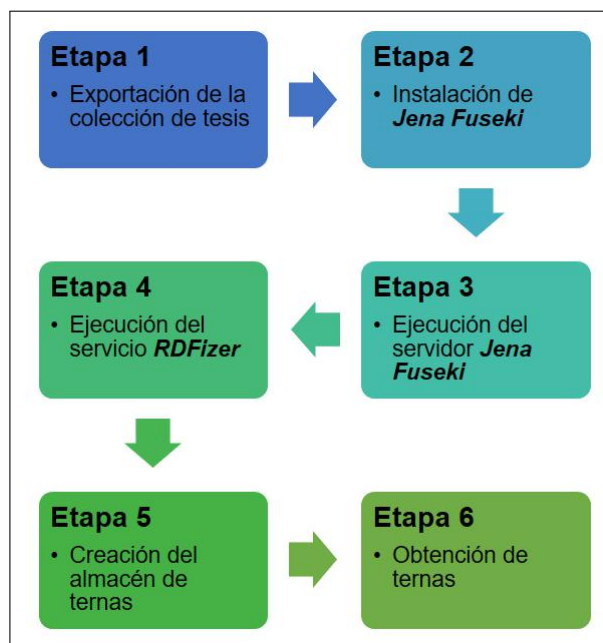


Figura 3.7: Etapas del proceso de migración de metadatos del RI-UPPue a la web semántica

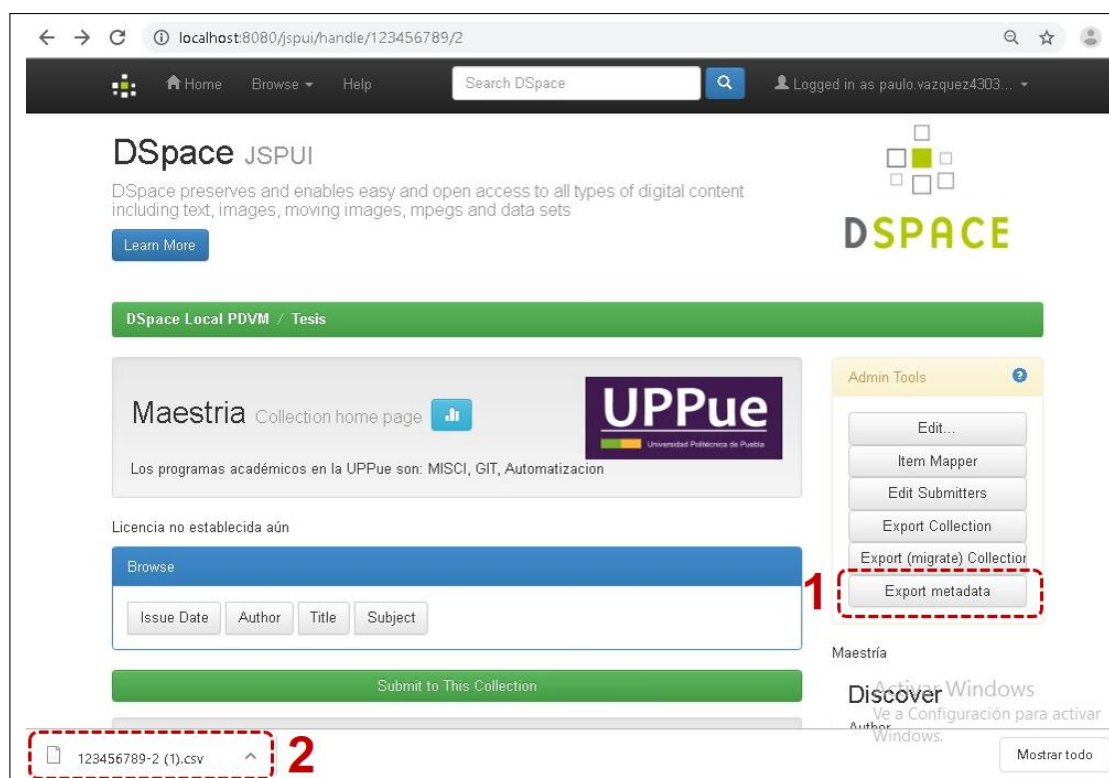


Figura 3.8: Exportación de la colección tesis a archivo CSV

3.4.4 Etapa 4: ejecución del servicio RDFizer

Un componente RDFizer [36] o *sponger*, 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:

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

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

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

```

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

- **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

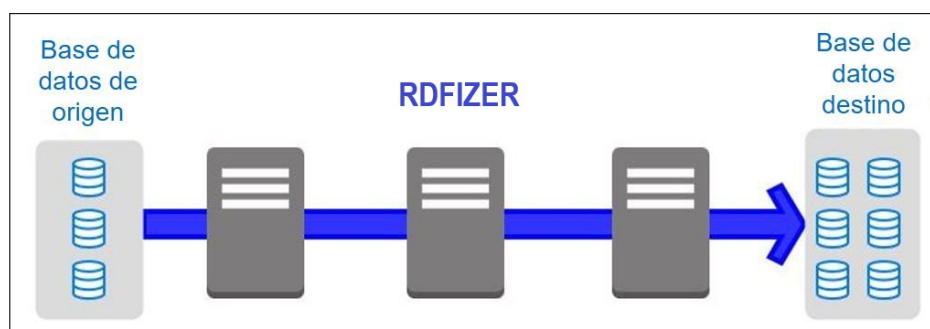


Figura 3.12: *Extracción, transformación y carga*

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 URL⁴ como muestra la Figura 3.13.

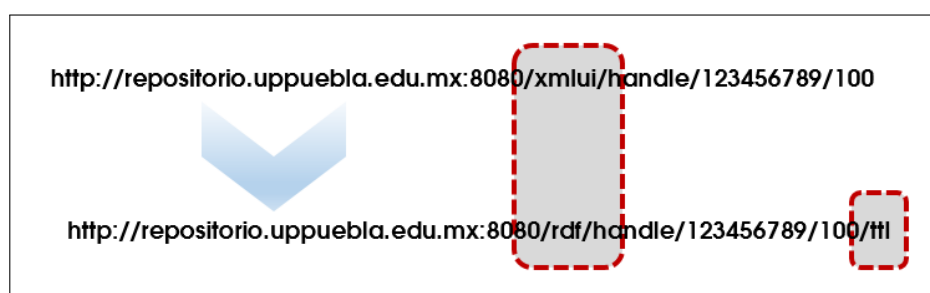


Figura 3.13: *Acceso a metadatos de tesis en RDF*

La Figura 3.14 muestra algunos metadatos de una tesis en RDF, estos provienen de los introducidos en DSpace cuando se agrega un documento a la colección.

3.4.7 Herramientas software

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

⁴Localizador Uniforme de recursos, en inglés *Uniform Resource Locator*



```

1 dspace:isPartOfCollection <http://localhos:8080/rdf/resource/123456789/2> ;
2 dc:contributor "Benitez Ruiz, Antonio" ;
3 dc:creator "Bravo Hernández, Paola" ;
4 dc:date "2019-03-28T11:20:23Z"^^xsd:dateTime ;
5 dc:language "es" ;
6 dc:publisher "Univesidad Politécnica de Puebla" ;
7 dcterms:abstract "-"@en-US ;
8 dcterms:available "2019-03-28T11:20:23Z"^^xsd:dateTime ;

```

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

3.4.7.1 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: 1) en inglés *HyperText Markup Language* (HTML 5), lenguaje de modelado de hipertexto, 2) *Java Script*, que es un lenguaje de programación del lado del cliente para manipular elementos HTML y 3) 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 (*ítems*) del almacén de ternas a través de la librería RDFLib versión 4.2.2 [37] y así integrar los datos provenientes del RI-UPPue a la ontología Onto4UPPue a través de la librería OWLReady2 [38] 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 RDFLib⁵ 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 pip⁶ en la instrucción `pip install rdflib` como se indica en la documentación.

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

⁶Repositorio de software para el lenguaje Python, en inglés *Python Index Package*

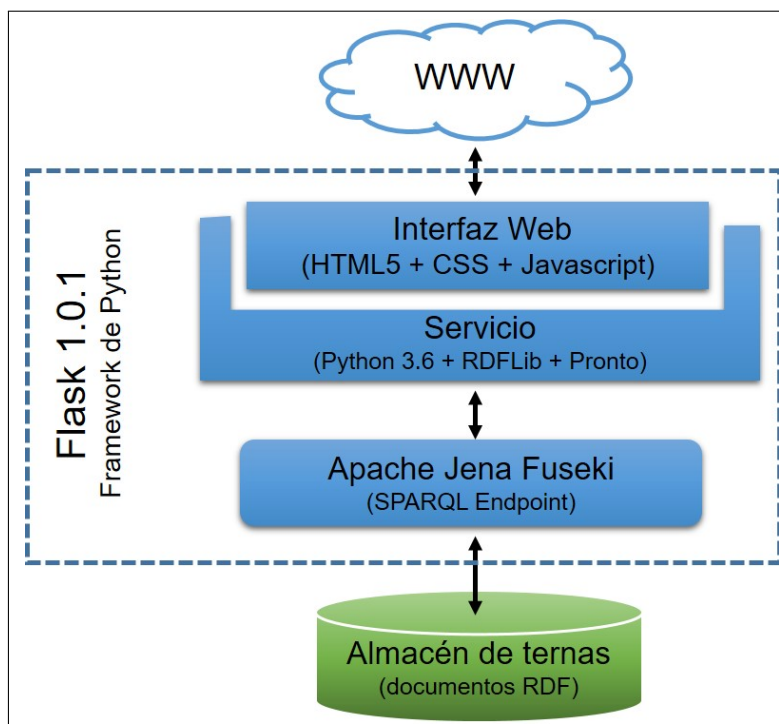


Figura 3.15: Arquitectura del servicio web

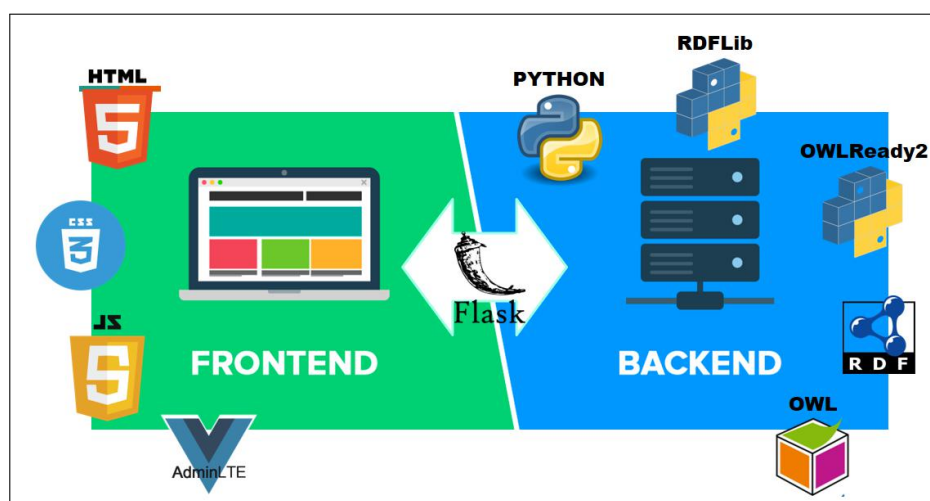


Figura 3.16: Tecnologías empleadas en el servicio web

3.4.7.4 Librería OWLReady2

La librería OWLReady2⁷ 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

⁷Disponible 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 HermiT o Pellet
- Compatibilidad con la librería RDFSlib

3.4.7.5 Librería `xml.etree.ElementTree`

La librería `xml.etree.ElementTree`⁸ 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 Marco de trabajo Flask

El desarrollo del servicio web emplea la versión 1.1.1 de Flask⁹, 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 usuario¹⁰, su propósito es identificar problemas de uso. En la tesis se adapta la plantilla diseñada para hacer análisis heurísticos descrita en [39], 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

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

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

¹⁰En 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 [39] para evaluar la heurística *generalidades* (heurísticos generales).

Heurísticos generales ²	
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 [39]

3.6 Mantenimiento

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

De acuerdo con la metodología planteada para el desarrollo del servicio web SW001, este capítulo presenta los resultados obtenidos por tarea como muestra la Figura 4.1.

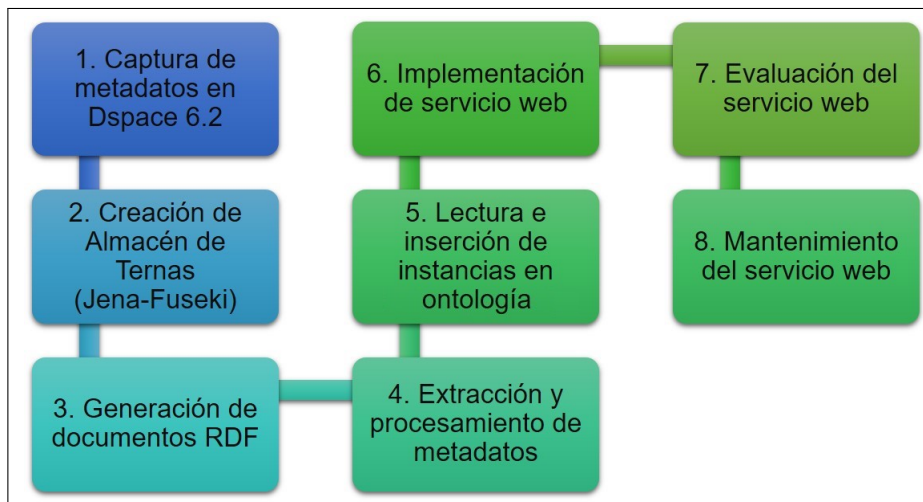


Figura 4.1: Tareas implementadas para el desarrollo del servicio web SW001.

Una vez que se instalaron y configuraron las herramientas tecnológicas, se llevó a cabo lo siguiente:

- 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
- Verificación de la funcionalidad del servicio web SW001

4.1 Captura de metadatos en DSpace 6.2

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

- Instalación de una instancia local de DSpace que emula las condiciones técnicas del RI-UPPue
- Creación de la comunidad *Tesis*, (ver la Figura 4.2)
- Creación de la colección *Maestría*, (ver la Figura 4.3)
- Inserción de once tesis (o *items*) en la colección *Maestría*

La Figura 4.4 muestra elementos de la colección *Maestría*, es decir, algunas de las once tesis insertadas, éstas representan una muestra de las que almacena el RI-UPPue. Esta colección contiene los archivos en formato PDF así como los metadatos correspondientes.

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 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: -

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...

Delete this Collection...

Figura 4.3: Creación de la colección Maestría

4.2 Creación de almacén de ternas

La Figura 4.5 muestra las actividades para crear el almacén de ternas utilizando la versión 1.6 del servidor Apache Jena Fuseki. El almacén está formado por un conjunto de documentos RDF, es un un servicio persistente indispensable para la implementación de tareas posteriores.

La Figura 4.6 muestra la interfaz desde un navegador web posterior al iniciar la ejecución de Fuseki.

The screenshot shows the 'Repositorio Dspace' interface. At the top, there is a navigation bar with 'DSpace Principal' and 'Envíos recientes'. The main content area is titled 'DSpace Local PDVM: Envíos recientes' and displays a list of recent uploads. Each item includes a title, author, and date. For example, the first item is 'Gestión tecnológica aplicada al desarrollo y registro de ISE-LabVIEW' by Castañeda Espinoza, Angélica (Universidad Politécnica de Puebla, 2017-12-31). To the right, there is a search bar labeled 'Buscar en DSpace' and a 'Listar' section with links for 'Comunidades & Colecciones', 'Por fecha de publicación', 'Autores', 'Títulos', and 'Materias'. Below the search bar is a 'Mi cuenta' section with links for 'Acceder' and 'Registro'.

Figura 4.4: Elementos de la colección Maestría

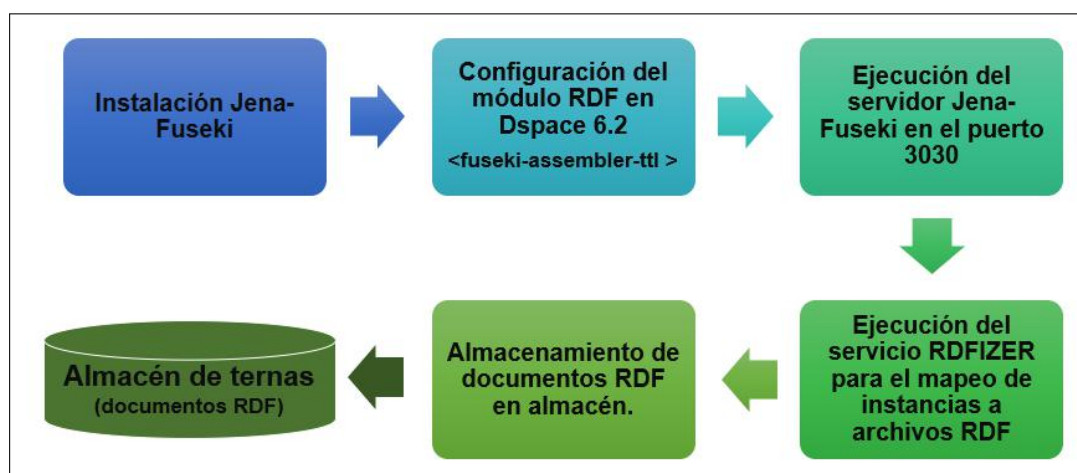


Figura 4.5: Actividades para crear el almacén de ternas

The screenshot shows the Apache Jena Fuseki web interface. The browser address bar displays 'localhost:3030'. The interface includes a navigation bar with 'Apache Jena Fuseki', 'dataset', 'manage datasets', and 'help'. The main content area displays 'Apache Jena Fuseki' and 'Version 3.10.0. Uptime: 1m 07s'. Below this, it states 'Datasets on this server' and 'There are no datasets on this server yet. Add one.' A section titled 'Use the following pages to perform actions or tasks on this server:' provides links for 'Dataset' (Run queries and modify datasets hosted by this server.), 'Manage datasets' (Administer the datasets on this server, including adding datasets, uploading data and performing backups.), and 'Help' (Summary of commands and links to online documentation.).

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

Tabla 4.1: Actividades para verificar la generación de documentos en RDF

Actividad	Resultado esperado	Resultado obtenido
Serialización de elementos almacenados en DSpace	Once instancias descritas en RDF	Once instancias descritas en RDF
Acceso a documentos RDF mediante URL	Consulta de once documentos RDF	Consulta de once documentos RDF

4.3 Generación de documentos en RDF

La generación de documentos en RDF se realiza al ejecutar el serializador *RDFizer*, (ver la Figura 4.7), éste extrae los metadatos almacenados en DSpace y los migra al almacén de ternas en documentos RDF. La verificación de esta tarea se realizó a través de la revisión de los documentos generados conforme a las actividades de la Tabla 4.1.

4.4 Extracción y procesamiento de metadatos

Se diseñó un servicio para verificar que los metadatos provenientes de DSpace estuvieran representados en los documentos RDF. Por omisión, el formato de metadatos es Dublin Core¹ (DC). El servicio usó la biblioteca *RDFlib*² y se implementó en la versión 3.6 del lenguaje Python. La Figura 4.8 muestra parte del código fuente, el resultado de su ejecución para una tesis se muestra en la Figura 4.9, en tanto, la Tabla 4.2 contiene las actividades de verificación.

Los metadatos para el conjunto de prueba que se exportaron en un archivo con extensión CSV son: id, collection, author, accessioned, available, issued, abstract, provenance, sponsorship, description, citation, uri, iso, publisher, subject, alternative title, title y type. Se verificó que los valores para estos metadatos estuvieran presentes en los documentos generados en RDF al implementar las actividades de la Tabla 4.3.

¹Los elementos de la Iniciativa de Metadatos Dublin Core (DCMI) se describen en: <http://purl.org/dc/elements/1.1/>

²biblioteca de Python para trabajar con archivos XML y RDF

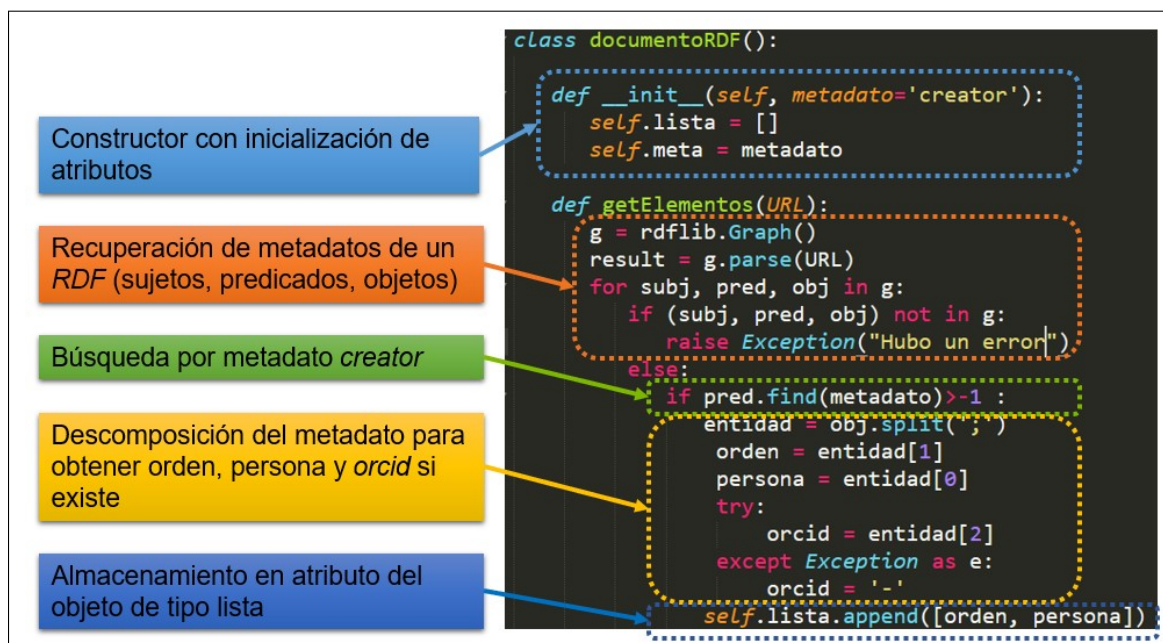


Figura 4.8: Código fuente del servicio para verificar metadatos de documentos en RDF

```
> 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 de un documento en RDF para una tesis

4.5 Lectura e inserción de instancias en la ontología Onto4UPPue

Para implementar la lectura de datos de los documentos en RDF y transformarlos en instancias de la ontología Onto4UPPue, se utilizaron los métodos `getroot`, `set` y `dump`; el primero permite ubicarse en el nodo raíz de árbol en XML de la ontología, los otros insertan y conservan la estructura de las instancias existentes. La Figura 4.10 despliega una sección del código fuente para insertar de una instancia.

La Figura 4.11 muestra el resultado de insertar una instancia en la ontología, cada

Tabla 4.2: Actividades de verificación del servicio de la Figura 4.8

Acción	Resultado esperado	Resultado obtenido
Recuperación de archivos RDF integrados en el almacén	Once documentos recuperados	Once documentos recuperados
Recuperación de los metadatos de un documento RDF	Recuperación de quince metadatos DC	Recuperación de doce metadatos DC, dos metadatos propios de DSpace y tres adicionales

Tabla 4.3: Actividades de verificación con archivos CSV

Actividad	Resultado esperado	Resultado obtenido
Exportación de metadatos de once tesis en archivo CSV	Archivo con metadatos de once tesis	Archivo con metadatos de once tesis
Exportación de metadatos en formato CSV de una tesis	Archivo con quince metadatos DC	Archivo con dieciseis metadatos DC y dos metadatos de 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: Sección de código fuente para insertar instancias en la ontología

tesis emplea bloques de aserciones de propiedades de datos (DataPropertyAssertion) para agrupar a un metadato (DataProperty), con el identificador de una tesis (NamedIndividual) y su valor representado en una literal.

4.6 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 adaptable, validación HTML, dinámico, accesible e intuitivo.

Se diseñó un sitio web para acceder al servicio web SW001, éste cuenta con las siguientes secciones:

- *Inicio*. Representa la página inicial del servicio


```

<DataPropertyAssertion>
  <DataProperty IRI="#Autor"/>
  <NamedIndividual IRI="#T24"/>
  <Literal datatypeIRI="http://www.w3.org/1999/02/22-rdf-syntax-ns#PlainLiteral">Pérez Juárez, Irianely Ivonne</Literal>
</DataPropertyAssertion>
<DataPropertyAssertion>
  <DataProperty IRI="#Descripcion"/>
  <NamedIndividual IRI="#T24"/>
  <Literal datatypeIRI="http://www.w3.org/1999/02/22-rdf-syntax-ns#PlainLiteral"></Literal>
</DataPropertyAssertion>
<DataPropertyAssertion>
  <DataProperty IRI="#Fecha"/>
  <NamedIndividual IRI="#T24"/>
  <Literal datatypeIRI="http://www.w3.org/1999/02/22-rdf-syntax-ns#PlainLiteral">2019-03-24T03:33:23Z</Literal>
</DataPropertyAssertion>
<DataPropertyAssertion>
  <DataProperty IRI="#Lenguaje"/>
  <NamedIndividual IRI="#T24"/>
  <Literal datatypeIRI="http://www.w3.org/1999/02/22-rdf-syntax-ns#PlainLiteral">Español</Literal>
</DataPropertyAssertion>

```

Figura 4.11: Resultado de la inserción de una instancia



Figura 4.12: Tecnologías para el desarrollo del servicio web SW001

- *Búsquedas semánticas.* Es la sección principal, desde ahí se accede al servicio que emplea las relaciones de la ontología Onto4UPPue para buscar instancias
- *Test UX.* Accede al instrumento de evaluación para estimar la usabilidad del servicio
- *Exportación de datos.* 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

La Figura 4.13 muestra el mensaje de bienvenida y el objetivo del servicio, ambos forman parte de la sección *Inicio*.



Figura 4.13: Sección de inicio del servicio web SW001

La Figura 4.14 muestra la sección *búsquedas semánticas*, la caja de selección despliega las relaciones *es autor* y *es sinodal*, las cuales modelan la participación de los autores en una tesis de maestría dentro de la ontología Onto4UPPue. Estas relaciones se utilizan en consultas SPARQL que recuperan los resultados, éstos se despliegan en tablas como muestra la Figura 4.15.

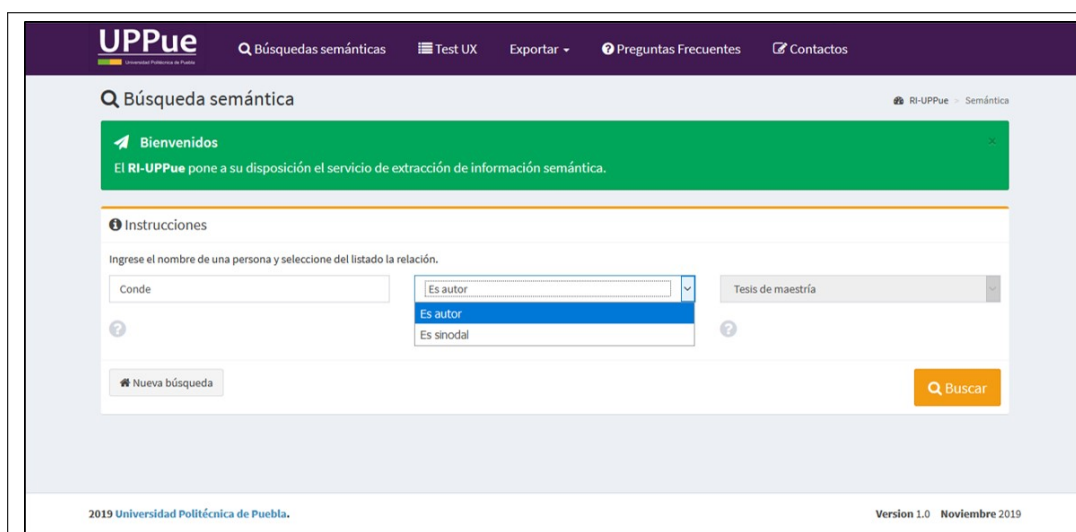


Figura 4.14: Sección de búsquedas semánticas, acceso al servicio web SW001

El servicio web SW001 está implementado de manera que agregar otras relaciones sólo requiere realizar modificaciones sencillas en el código fuente. La ontología se puede exportar en archivos con formato JSON y OWL como muestra la Figura 4.16.

La Figura 4.17 despliega la sección de *preguntas frecuentes* cuyo propósito es apoyar a los usuarios en la adquisición de conceptos del dominio sobre los repositorios institucionales. Los datos de contacto del desarrollador y revisora del servicio se incluyen en la sección *Contacto* como muestra la Figura 4.18.

The screenshot shows the UPPue semantic search interface. At the top, there is a navigation bar with the UPPue logo and links for 'Búsquedas semánticas', 'Test UX', 'Exportar', 'Preguntas Frecuentes', and 'Contactos'. Below this is the 'Búsqueda semántica' section with instructions: 'Ingrese el nombre de una persona y seleccione del listado la relación.' There are input fields for 'Ejemplo: Juan Pérez', 'Es autor', and 'Tesis de maestría'. A 'Nueva búsqueda' button and a 'Buscar' button are also present. The search results section shows 'Resultados de la búsqueda: 1 relación(es)' and 'Criterio de búsqueda: - Conde es autor-'. A table displays the results:

	Persona	Relación	Tesis de Maestría
1	Pérez Conde Gudella Pilar	Es autor de	Clasificación Automática de Retinopatía Diabética Basada en Imágenes.

Figura 4.15: Tabla de resultados de la búsqueda semántica de la Figura 4.14.

This screenshot shows the same UPPue semantic search interface as Figure 4.15, but with a 'Descarga de recursos' dialog box open. The dialog box has tabs for 'Ontología en JSON' and 'Ontología en OWL'. It displays the file 'onto4-UPPue.owl' (171 KB) from 'http://localhost:5000'. The user is prompted to choose an application to open the file, with 'Sublime Text (predeterminada)' selected. There are 'Aceptar' and 'Cancelar' buttons at the bottom of the dialog.

Figura 4.16: Descarga de la ontología Onto4UPPue en formato JSON u OWL

The screenshot shows the 'Preguntas frecuentes' section of the UPPue website. It features five question cards with the following content:

- ¿Qué es un repositorio?**
Es un depósito de documentos digitales, cuyo objetivo es organizar, almacenar, preservar y difundir en modo de acceso abierto.
- ¿Qué es un repositorio institucional?**
Un repositorio institucional es un depósito en línea donde se depositan, en formato digital, materiales derivados de la producción científica o académica de una institución.
- ¿Qué es información semántica?**
La semántica se refiere a todo lo relacionado con el significado o interpretación de signos lingüísticos como símbolos, palabras o expresiones.
- ¿Qué es el acceso abierto?**
Es el acceso inmediato, sin requerimientos de registro, suscripción o pago a material digital educativo, académico, científico o de cualquier otro tipo.
- ¿Qué es una búsqueda semántica?**
Proceso utilizado para mejorar la búsqueda por Internet mediante el uso de datos de las redes semánticas para desambiguar las consultas y el texto de la web con la finalidad de encontrar los resultados más relevantes en relación a la demanda del usuario.

At the bottom of the page, it says '2019 Universidad Politécnica de Puebla.' and 'Versión 1.0 Noviembre 2019'.

Figura 4.17: Sección de preguntas frecuentes

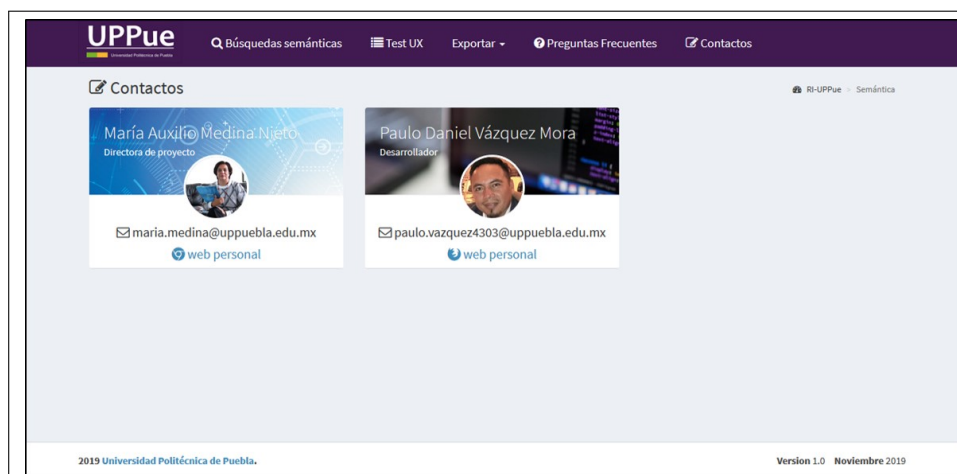


Figura 4.18: Sección de contacto

El servicio web es accesible desde equipos de cómputo de escritorio o cualquier dispositivo móvil que cuente con conexión a internet tal como muestra la Figura ??.



Figura 4.19: Acceso al servicio web SW001 desde un dispositivo móvil

4.7 Evaluación y mantenimiento

En la evaluación del servicio web SW001 se utilizó la versión 5.2 del editor de ontologías Protégé para revisar que la inserción de instancias del conjunto de prueba fuera correcta, lo cual ocurrió en los once elementos del conjunto de prueba. A manera de ejemplo, la Figura 4.20 muestra las propiedades de datos para un elemento identificado como instancia T24, observar que estas propiedades modelan a los metadatos provenientes de DSpace. La verificación automática de la consistencia lógica de la ontología posterior a la inserción de todas las instancias se llevó a cabo utilizando los razonadores Hermit y Pellet, ninguno de ellos desplegó mensajes de error o alarma, lo cual indica que la ontología con instancias resultó ser lógicamente consistente.

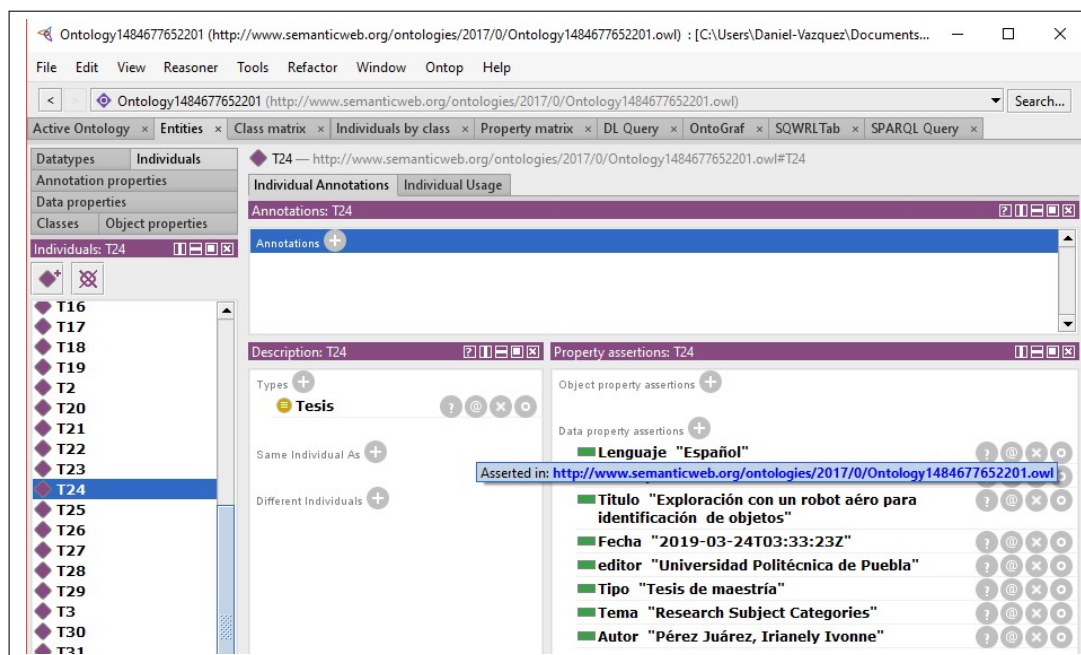


Figura 4.20: Descripción de la instancia T24 en la ontología Onto4UPPue.

4.7.1 Evaluación de la usabilidad del sitio web para acceder al servicio SW001

Durante la tesis, se desarrolló una aplicación web denominada *TestUX* que emplea las once heurísticas propuestas por [39] y las integra en un cuestionario, el propósito es estimar la usabilidad del sitio web que accede al servicio SW001 de acuerdo a la opinión de usuarios potenciales. La Figura 4.21 muestra la pantalla principal de esta aplicación.

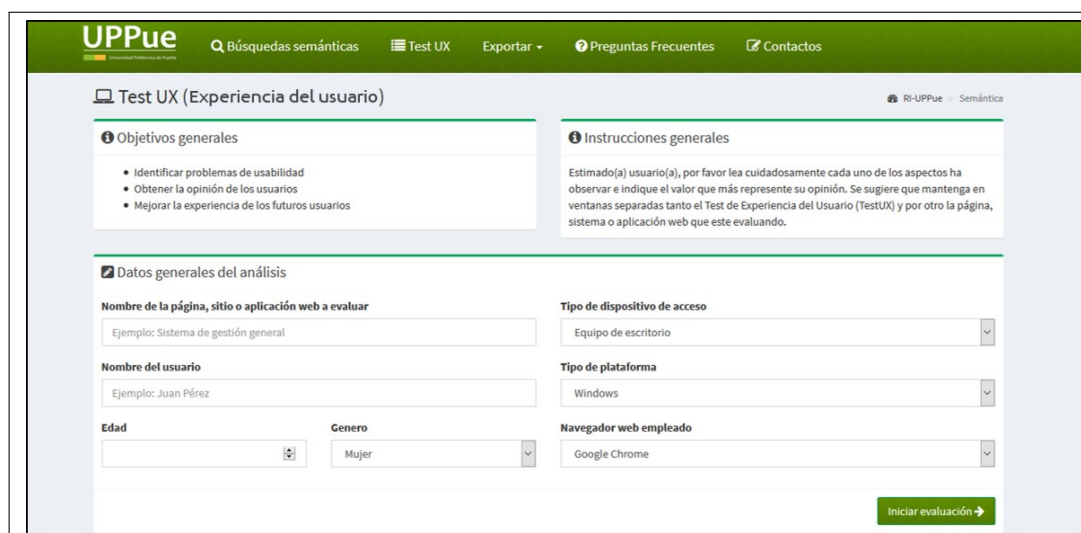


Figura 4.21: Pantalla principal de TestUX

De acuerdo con la opinión de [39], sólo se requiere de cinco personas para poder realizar una evaluación de la experiencia del usuario. En la evaluación reportada en esta tesis, se seleccionaron a 16 estudiantes de un grupo de 30 de Ingeniería en Informática,

quienes indicaron ser usuarios frecuentes del repositorio institucional (RI-UPPue). El género de diez (10) de los estudiantes es masculino y seis (6) femenino, su rango de edad varía entre 21 y 24 años.

Los estudiantes utilizaron la escala Likert siguiente para indicar su nivel de satisfacción con las heurísticas, éstas y sus resultados se muestran en la Figura 4.22. La integración de los resultados así como el promedio general de las heurísticas se despliega en la Figura 4.23.

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

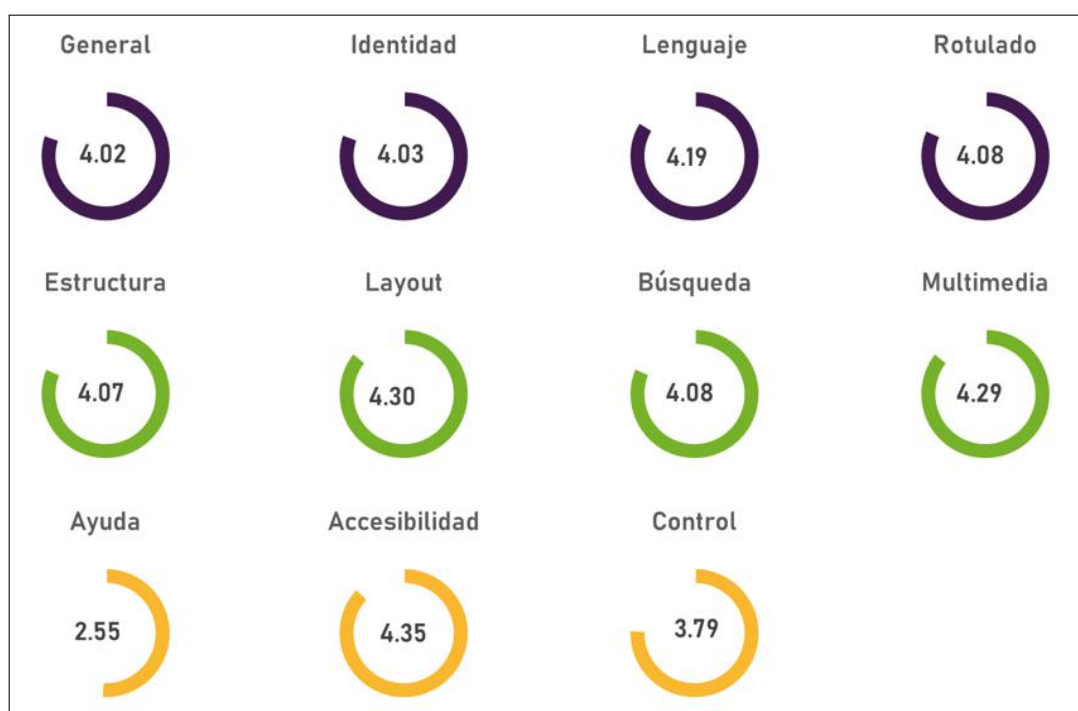


Figura 4.22: Resultados de la evaluación por heurística del sitio web

Las Figuras 4.22 y 4.23 muestran que la heurística con el valor máximo (4.35) fue para la accesibilidad y el mínimo (2.55) se relaciona con la ayuda. El promedio de las heurísticas fue de **4.02**, el cual es muy cercano al valor 4 de la escala de Likert, por lo que se concluye que los estudiantes indicaron estar muy satisfechos con el sitio web. La Figura 4.24 muestra evidencia fotográfica del uso de TestUX por parte de los estudiantes.

Después de la evaluación, se implementaron elementos de ayuda contextual y mensajes descriptivos como muestran las Figuras 4.25, 4.26 y 4.27.

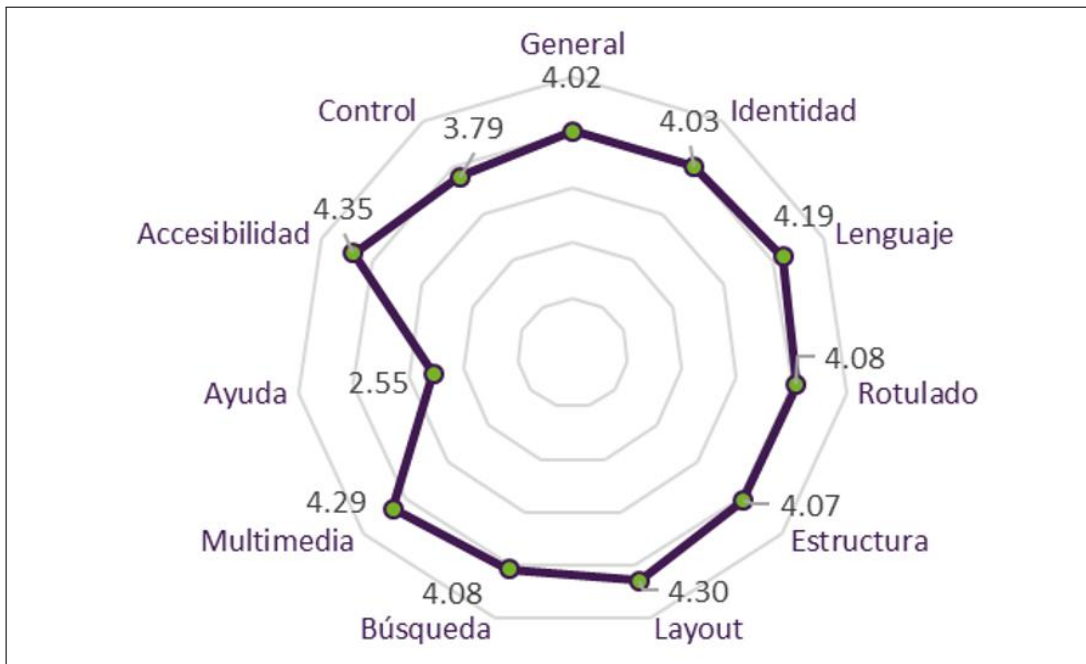


Figura 4.23: Integración de resultados por heurística

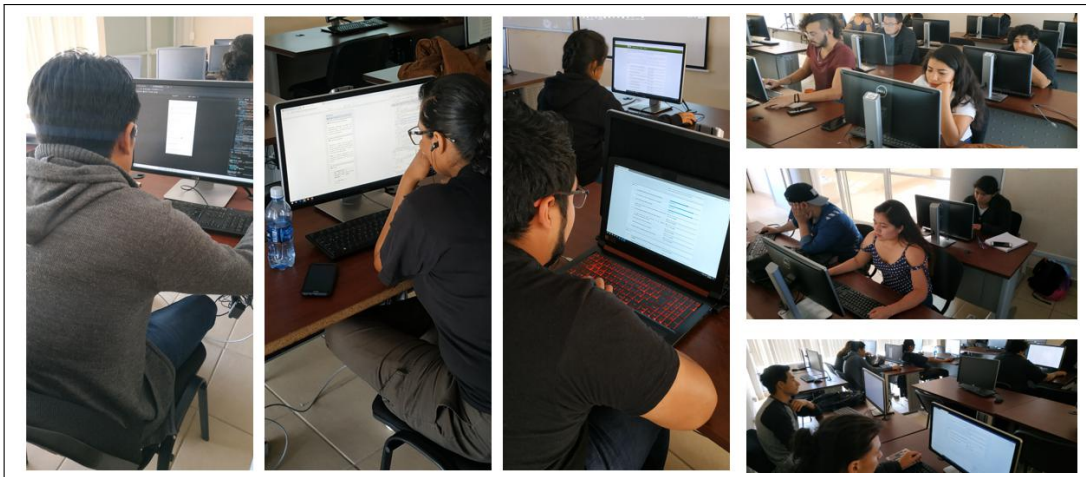


Figura 4.24: Estudiantes utilizando la aplicación web TestUX.

Figura 4.25: Íconos para ayuda contextual

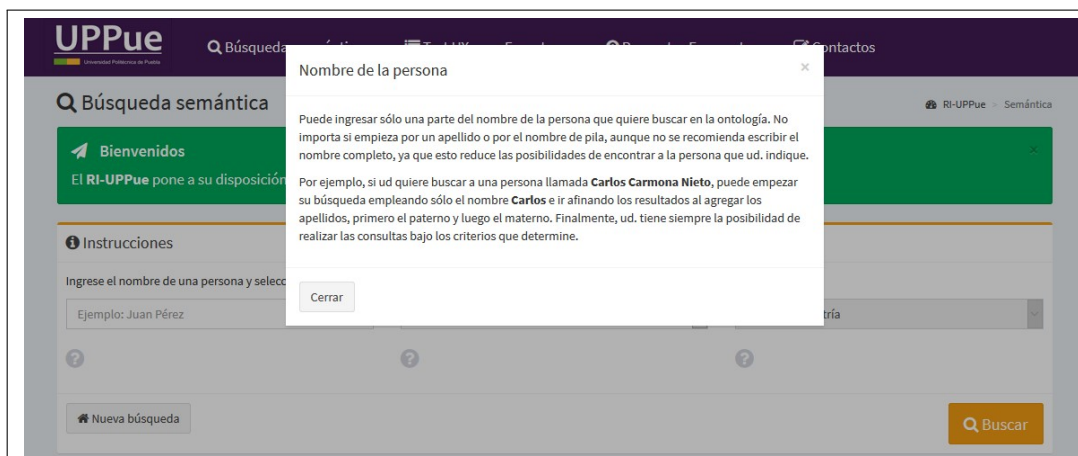


Figura 4.26: Mensaje para la descripción detallada de elementos y ejemplos de uso

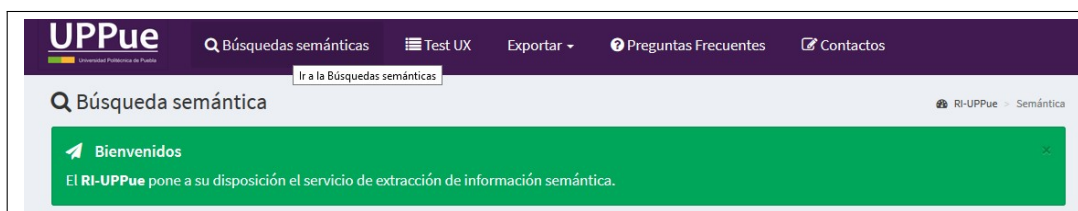


Figura 4.27: Mensaje descriptivo del servicio web SW001 en la barra de navegación

La tesis expuso conceptos relacionados con los repositorios institucionales, los cuales almacenan documentos digitales validados que se producen generalmente como resultado de actividades académicas y científicas dentro de organizaciones educativas. Se hizo una investigación de tipo exploratoria en la que se analizaron interfaces de búsqueda básica y avanzada de diferentes repositorios que distribuyen documentos en español y se observó un uso limitado de tecnologías semánticas, mismas que se utilizaron para atender la problemática expuesta en la sección 1.1.

En la tesis, se revisaron los requerimientos para repositorios propuestos por la Confederación de Repositorios de Acceso Abierto (COAR) y se presentó el diseño y la implementación de un servicio web denominado SW001, éste recupera información semántica utilizando la ontología Onto4UPPue, la cual representa formalmente a personas, documentos y sus relaciones de acuerdo al contexto de operación del RI-UPPue. El servicio usa también consultas en SPARQL y otras herramientas tecnológicas como el módulo de serialización RDF para la plataforma DSpace. Este servicio es tipo REST, por lo que se caracteriza por ser escalable y transparente.

La ontología reduce la ambigüedad en la interpretación de los datos descriptivos o metadatos exportados de la plataforma DSpace que soporta al repositorio RI-UPPue, su vocabulario se puede compartir entre usuarios y computadoras. Por tanto, se considera que el servicio implementado establece un vínculo entre el repositorio y la ontología, dado que en la medida en que se incremente el número de tesis de maestría en el repositorio y se exporten sus metadatos, éstos se insertan y gestionan como instancias. Cabe hacer mención que la ontología en sí representa un medio para encontrar inconsistencias de manera automática en los datos del RI-UPPue.

Por parte del desarrollador, el autor de esta tesis, se verificó la funcionalidad del servicio SW001 a través de la revisión de los metadatos exportados de DSpace y su representación como instancias en la ontología utilizando el editor Protégé; se comprobó la consistencia lógica de ésta utilizando razonadores.

Durante la tesis se diseñó un sitio web para acceder al servicio y se construyó la aplicación web TestUX para evaluar su usabilidad. A la fecha, el servicio recupera datos de dos relaciones: *es sinodal* y *es autor*, éstas permiten determinar el rol de los autores de una tesis. Dado que la información sobre estos roles no podía recuperarse desde la interfaz de búsqueda del RI-UPPue, se demuestra que a través de la implementación del sitio y servicio SW001, se extendieron los mecanismos de búsqueda para las tesis.

La metodología de desarrollo utilizada en el servicio SW001 permite que se puedan agregar otras relaciones a partir de modificaciones relativamente sencillas en el código fuente. Desde el sitio web, los usuarios pueden descargar la ontología con las instancias insertadas por el servicio en los formatos JSON y OWL, se espera que de esta manera sea reutilizada por otras aplicaciones de la web semántica.

En la evaluación del sitio web, se recolectó información de once heurísticas y se contó con la participación de 16 estudiantes de Ingeniería en Informática de la UPPue, quienes reportaron utilizar con frecuencia el RI-UPPue. Los resultados se mostraron

por heurística y en promedio se obtuvo un nivel de satisfacción alto. Posterior a la evaluación, se agregaron elementos de soporte debido a que la heurística de ayuda obtuvo la calificación mínima.

A través de la tesis, se proveen las bases para el desarrollo de otros servicios y se contribuye en la difusión de los beneficios del acceso abierto y en la generación de datos semánticamente enriquecidos. Actualmente, se está trabajando en la inserción de instancias para cualquier tipo de documento almacenado en el RI-UPPue. Como trabajo a futuro, se planea la implementación de servicios relacionados con la obtención de indicadores de producción académica para los autores.

- [1] Sommerville I. Ingeniería de Software. 2005. pp. 1-63
- [2] Merlo J. A. Ecosistemas del acceso abierto. 2018. Ediciones Universidad de Salamanca. pp. 43-44.
- [3] REBUIN. Beneficios del acceso abierto en tu universidad. Red de Bibliotecas Universitarias REBIUN. Universidad de Navarra. España. Disponible en: 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. Disponible en: <https://www.jisc.ac.uk/opendoar>. Fecha de consulta: 2019-06-01
- [5] BOAI15. Iniciativa de Budapest para el acceso abierto. 2007. Disponible en: <https://www.budapestopenaccessinitiative.org/translations/spanish-translation>. Fecha de consulta: 2018-09-01
- [6] Medina M. A., Sánchez J. A., Cervantes O., Medina R. C., De la Calleja J., Benítez A. Representación semántica de conocimiento operativo y de dominio para repositorios institucionales. 2017. Instituto Nacional del Derecho de Autor. Registro Público 03-2017-042511235500-01.
- [7] J. Polanco-Cortés. Repositorios digitales. Definición y pautas para su creación. Universidad de Costa Rica. Vicerrectoría de Investigación. Disponible en: <https://ucrindex.ucr.ac.cr/docs/repositorios-digitales-definicion-y-pautas-para-su-creacion.pdf>. Fecha de consulta: 2018-03-07
- [8] Universidad de León. ¿Qué es un repositorio institucional?. Biblioteca universitaria. Disponible en: <https://biblioteca.unileon.es/ayuda-formacion/repositorio-institucional>. Fecha de consulta: 2018-03-12
- [9] REMERI. 2018. Red Mexicana de Repositorios Institucionales. Disponible en: <http://www.remeri.org.mx/>. Fecha de consulta: 2018-01-12
- [10] DuraSpace. 2019. DSpace is a turnkey institutional repository application. 2020. Copyright ©DSpace 2020. Disponible en: <http://www.dspace.org/>. Fecha de consulta: 2018-02-05
- [11] REMERI. Directorio de la Red Mexicana de Repositorios Institucionales. Disponible en: <http://www.remeri.org.mx/repositorios/>. Fecha de consulta: 2018-01-12
- [12] CONACYT. Consejo Nacional de Ciencia y Tecnología. Repositorio Nacional. Gobierno de México. Disponible en: <https://www.repositorionacionalcti.mx/>. Fecha de consulta: 2018-01-25

- [13] Electronics in Computer Science. 2019. EPrints for REF2021 – EPrints Services. *Electronics and Computer Science*. Universidad de Southampton. Disponible en: <http://www.eprints.org/uk/>. Fecha de consulta: 2019-01-25
- [14] DuraSpace. VIVO - Connect. Share. Discover. 2019. Copyright ©Vivo 2020. Disponible en: <https://duraspace.org/vivo/>. Fecha de consulta: 2019-10-11.
- [15] Blanco S., De la Fuente P., Anguita R. 2007. WebQuest y anotaciones semánticas. *Comunicar, revista científica de comunicación y educación*. Vol. 28. ISSN:1134-3478. pp. 191-196. Disponible en: <https://www.revistacomunicar.com/ojs/index.php/comunicar/article/view/C28-2007-23>. Fecha de consulta: 2018-02-21
- [16] N. F. Noy, D. L. McGuinness. 2001. Ontology development 101: a guide to creating your first ontology. Stanford Knowledge Systems Laboratory Technical Report KSL-01-05 and Stanford Medical Informatics Technical Report SMI-2001-0880. Universidad de Stanford. Disponible en: https://protege.stanford.edu/publications/ontology_development/ontology101.pdf. Fecha de consulta: 2018-02-21
- [17] K. Rodríguez, R. Ronda. 2005. Web semántica: un nuevo enfoque para la organización y recuperación de información en el web. *ACIMED en línea*. Vol 13. No. 6. ISSN: 1024-9435. Disponible en: <http://eprints.rclis.org/7961/>. Fecha de consulta: 2018-12-11
- [18] Vázquez P. D., Medina M. A. 2018. Diseño de un servicio web para la recuperación de información semántica del repositorio institucional de la Universidad Politécnica de Puebla. Congreso de Instrumentación SOMI XXXIII. Año 5. No. 1. Octubre 2018. ISSN: 2395-8499.
- [19] MIT, ERCIM, Keio, Beihang. 2019. W3C Standards. Copyright ©2019. Disponible en: <https://www.w3.org/standards/>. Fecha de consulta: 2018-11-13
- [20] Soler D. J. 2017. 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. Facultad de Comunicación y Documentación. Universidad de Murcia, España. Tesis de doctorado. Disponible en: <http://hdl.handle.net/10201/54364>, Fecha de consulta: 2018-10-09
- [21] RDF Working Group. 2014. Resource Description Framework. Disponible en: <https://www.w3.org/2001/sw/wiki/RDF>. Fecha de consulta: 2018-11-13
- [22] D. Brickley, Guha R. V. 2014. RDF Schema 1.1. 2014. Disponible en: <https://www.w3.org/TR/rdf-schema/>. Fecha de consulta: 2018-11-13
- [23] R. Angles. 2015. Introducción a las bases de datos RDF. Disponible en: <http://renzoangles.net/files/libro.pdf>
- [24] Open Link Software. 2019. About OpenLink Virtuoso. Copyright ©2019. Disponible en: <https://virtuoso.openlinksw.com/>. Fecha de consulta: 2019-04-21.

- [25] F. Bauer, M. Kaltenböc. 2012. Linked open data: the essentials. A quick start guide for decision makers. DGS-Druck-u.Graphikservice GmbH. Disponible en: <http://www.semantic-web.at/LOD-TheEssentials.pdf>
- [26] OntoText LMI. 2018. Semantic technology accelerates document search: how LMI implements semantic search with OpenPolicy. Disponible en: https://cdn2.hubspot.net/hubfs/5619976/White%20Papers/Ontotext%20White%20Paper_Semantic%20Technology%20Accelerates%20Document%20Search.pdf?__hssc=95638467.1.1586626505135&__hstc=95638467.6453e6a92dfc925ef818f5df16c143a8.1586626505135.1586626505135.1586626505135.1&__hsfp=2831686981&hsCtaTracking=15a48ae7-033c-4b27-9a0a-7d3ce19fbee%7C2be5edb1-1de7-4a55-a26a-06afa8b288ac. Fecha de consulta: 2019-05-21.
- [27] RedCLARA. 2019. La Referencia, Red de Repositorios de Acceso Abierto a la Ciencia. 2019. Disponible en: <http://www.lareferencia.info/es/>. Fecha de consulta: 2019-11-13
- [28] CONACYT. Repositorio Nacional. 2019. Catálogos: servicios REST. Disponible en: <https://catalogs.repositorionacionalcti.mx/>. Fecha de consulta: 2019-11-13
- [29] C. Á. Caules. 2013. Introducción a Servicios REST - Arquitectura Java. Copyright ©Arquitectura Java. 2019. Disponible en: <https://www.arquitecturajava.com/servicios-rest/>.
- [30] J. Texier. De Giutsi M., Villareal G. Lira A. 2017. Repositorio Semántico para la Universidad Nacional de Chilecito. Conferencia Latinoamericana de Informática y 46 Jornadas Argentinas de Informática (CLEI XLIII y 46 JAIIO). Disponible en: http://sedici.unlp.edu.ar/bitstream/handle/10915/62792/Documento_completo.pdf-PDFA.pdf?sequence=1. Fecha de consulta: 2018-06-13.
- [31] Prud'hommeaux E., Seaborne A. 2008. SPARQL Query Language for RDF. 2008. Recomendación de W3C desde Enero del 2008. Copyright ©2006-2007 W3C. Disponible en: <https://www.w3.org/TR/rdf-sparql-query/>. Fecha de consulta: 2018-06-13.
- [32] D. Robinson. 2017. The Incredible Growth of Python. Copyright ©2020 Stack Overflow. Disponible en: <https://stackoverflow.blog/2017/09/06/incredible-growth-python/>. Fecha de consulta: 2018-06-15.
- [33] COAR 2017. Next generation repositories: behaviours and technical recommendations of the COAR next generation repositories working group. Copyright ©2017 COAR: Building a Global Knowledge Commons. Disponible en: <https://www.coar-repositories.org/files/NGR-Final-Formatted-Report-cc.pdf>. Fecha de consulta: 2019-03-15.
- [34] Codina L., Rovira C. 2006. La web semántica. Universidad Pompeu Fabra. Departamento de Periodismo y de Comunicación Audiovisual. Instituto Universitario de Lingüística Aplicada. Disponible en: http://eprints.rclis.org/8899/1/web_semantica_.pdf.

- [35] The Apache Software Foundation. 2019. Apache Jena Fuseki. Copyright ©2011-2020. Apache Software Foundation. Disponible en: <https://jena.apache.org/documentation/fuseki2/>. Fecha de consulta: 2019-03-15.
- [36] Open Semantic Framework. 2013. RDFizer Concept. Disponible en: https://wiki.opensemanticframework.org/index.php/RDFizer_Concept. Fecha de consulta: 2019-03-13.
- [37] RDFLib. 2013. RDFLib 4.2.2. documentation. Copyright ©2009-2013. RDFLib Team. Disponible en: <https://rdflib.readthedocs.io/en/stable/>. Fecha de consulta: 2019-03-13.
- [38] Jean-Baptiste L. 2019. OWLReady2 Library version 0.19. Disponible en: <https://pypi.org/project/owlready2/>. Fecha de consulta: 2019-03-13.
- [39] Torres-Burriel D. 2008. Plantilla para hacer análisis heurísticos de usabilidad. Disponible en: <https://www.torresburriel.com/weblog/2008/11/28/plantilla-para-hacer-analisis-heuristicos-de-usabilidad/>. Fecha de consulta: 2019-07-13.
- [40] Lundh F. 2005. Elementtree 1.2.6. Python (MIT style). Disponible en: <https://pypi.org/project/elementtree/>. Fecha de consulta: 2018-07-14.
- [41] P. Suber. 2015. Open Access Overview: focusing on open access to peer-reviewed research articles and their preprints. Disponible en: <http://legacy.earlham.edu/~peters/fos/overview.htm>. Fecha de consulta: 2019-08-23.
- [42] Lagoze C., Van de Sompel H. Nelson M., S. Warner S. 2005. Implementation guidelines for the Open Archives Initiative Protocol for Metadata Harvesting. Disponible en: <http://www.openarchives.org/OAI/2.0/guidelines.htm>. Fecha de consulta: 2019-08-23.
- [43] DCMI. 2001. DCMI Metadata Terms. Dublin Core Metadata Initiative. DCMI Usage Board. Disponible en: <https://www.dublincore.org/specifications/dublin-core/dcmi-terms/>. Fecha de consulta: 2019-07-23.
- [44] Lamarca L. M. J. 2007. Hipertexto, el nuevo concepto de documento en la cultura de la imagen. Universidad Complutense de Madrid. Facultad de Ciencias de la Información. Departamento de Biblioteconomía y Documentación. Tesis de Doctorado. Disponible en: <http://www.hipertexto.info>. Fecha de consulta: 2019-05-21.
- [45] Elsevier Connect. 2019. Tipos de Open Access: diferencias entre la “vía verde” y la “vía dorada”. Copyright ©2020 Elsevier. Disponible en: <https://www.elsevier.com/es-es/connect/actualidad-sanitaria/tipos-de-open-access-via-verde-y-la-via-dorada>. Fecha de consulta: 2019-05-22.
- [46] M. Medina, G. De la Calleja, C. Zepeda, A. Benitez y R. Medina. 2018. Evaluación de usabilidad de las interfaces para depositar objetos de aprendizaje en EPrints y DSpace. *Revista Visión e Innovación Politécnica*. ISSN: 2007-221X