

UNIVERSIDAD DE CUENCA



FACULTAD DE INGENIERÍA

ESCUELA DE INGENIERÍA DE SISTEMAS

“Implementación de Procesos de Anotación Semántica de Servicios WFS con
Ontologías de Dominio.”

Tesis previa a la obtención del
título de Ingeniero de Sistemas

Autor:

Coronel Samaniego Holger Marcelo CI: 0105986731
RodasZamora Marcos Javier CI: 0103636593

Director:

Ing Víctor Hugo Saquicela Galarza.
CI: 0103599577

Cuenca - Ecuador
2016

Resumen.

Actualmente los servicios WFS son una fuente de información desordenada y no estructurada. Es por esto que se hace necesaria la transformación de información para estos servicios. El internet está migrando hacia la web semántica, es por esto, que es oportuno si no necesario migrar hacia un repositorio semántico. Las ventajas que se derivan de tener un repositorio semántico son múltiples, las más importantes son: acceso a la información de los servicios WFS a través de consultas semánticas, estructurar la información en tripletas y quizá la ventaja más importante sería la posibilidad de enlazar los datos con otros repositorios. Al intentar un proceso de enlace de datos de los servicios WFS aparecen varios problemas entre los cuales están: la falta de herramientas para implementar el proceso ya mencionado, la falta de información descriptiva de los servicios WFS, vital para el proceso de anotación semántica, y la correlación de las ontologías y servicios WFS.

Los sistemas de información geográfica (SIG) son un conjunto de herramientas que permiten el manejo de información relacionada a cierto lugar geográfico. Los servicios definidos por la OGC son descritos sintácticamente, es por esto, que para mejorar la búsqueda de estos servicios deben ser anotados semánticamente para lo cual se debe describir un proceso.

Para anotar los servicios WFS es recomendado dividir los servicios en tres niveles: meta data, capas e instancias.

Palabras Claves: Web Feature Service (WFS), Web Semántica, Datos Entrelazados (LOD), RDF, SPARQL, GEOSPARQL, Sistemas de Información Geográfica (SIG).

Abstract.

Currently WFS services are a source of information that is messy and unstructured. That is why there is a need for a information transformation for these services. Since the Internet is migrating towards the semantic web it is timely if not necessary for the migration of WFS services to be made to a semantic repository. The advantages to be derived from this semantic repository would be multiple, among which the most important are: access to information in WFS through semantic queries, structuring the information in triplets and perhaps most important of all would be semantic enrichment with other semantic information repositories. When attempting a Linked Open Data(LOD) process of WFS services several problems rise, among which are the lack of tools that implement the methodology chosen for the afore mentioned process, the lack of descriptive information WFS services, vital for the process of semantic annotation, and the correlation between ontologies and WFS services.

Geographic Information Systems (GIS) are a set of tools that allow the management of information related to a certain geographic place.

Another important feature of GIS is the way that the data is published. Usually geographic information is published through web services. WFS services are a collection of features that can be accessed through an URL.

Web services defined by OGC are described syntactically, this is why to improve the search of these web services need to be semantically anotated but a process needs to be described for this annotation.

To annotate WFS services it is recommended to divide the web service in three tiers metadata, layers and instances.

Keywords: Web Feature Service (WFS), Semantic Web, Linked Open Data (LOD), RDF, SPARQL, GEOSPARQL, Geographic Information System (GIS)

Índice General

Resumen	1
Abstract	2
Índice General	3
Índice de figuras.	5
1. Introducción	7
1.1. Introducción	7
1.2. Panorama General	7
1.3. Planteamiento del problema y justificación	8
1.4. Objetivo General.	8
1.5. Objetivos Específicos	9
1.6. Estructura del Documento	9
2. Marco Teórico.	11
2.1. Introducción	11
2.2. Sistemas de Información Geográfica	11
2.2.1. Sistemas de Coordenadas	12
2.2.2. Estándares OGC (Open Geospatial Consortium)	12
2.2.3. WFS (Web Feature Service)	12
2.2.4. GML(Geography Markup Language)	15
2.3. Servicios Web Geoespaciales	16
2.4. ¿Que es la Web Semántica?	17
2.4.1. RDF (Resource Description Framework)	17
2.4.2. OWL (Web Ontology Language)	17
2.4.3. SPARQL	17
2.5. Anotación Semántica	17
2.5.1. ¿Qué es una anotación?	18
2.5.2. Anotación manual	18
2.5.3. Anotación semi-automática	18
2.5.4. Anotación automática	18
2.5.5. Anotación de servicios OGC	19
2.6. Linked Data (Datos Enlazados)	19

3. Anotación semántica de servicios WFS	20
3.1. Introducción	20
3.2. Búsqueda servicios WFS	20
3.3. Anotación de los meta datos del servicio WFS	21
3.4. Anotación de modelo de datos	21
3.5. Análisis de la información de los servicios WFS	22
3.6. Selección de ontologías	24
3.7. Invocación de los servicios WFS	27
3.8. Mapeo R2RML	28
3.9. Técnicas de Anotación	30
3.9.1. Anotación Manual	30
3.10. Publicación de Información	33
3.11. Enlace de datos	37
3.11.1. Creación de un proyecto en la plataforma Silk Workbench	42
3.11.2. Carga de fuentes al proyecto.	42
3.11.3. Asignación de formato de entrada y salida a las fuentes del	
proyecto.	43
3.11.4. Relación de las fuentes para su análisis	44
3.11.5. Plugins para uso de la generación de enlaces entre las fuentes	44
3.11.6. Generación de enlaces entre las fuentes	45
3.11.7. Verificación de resultados en forma gráfica	46
4. Evaluación y resultados	47
4.1. Método	47
4.2. Proceso de evaluación y resultados	48
4.3. Registro de capas	49
4.4. Anotación semántica	50
4.5. Evaluación de datos de entrada	50
4.6. Evaluación de datos de salida	51
4.7. Repositorio WEB	52
4.8. Recomendaciones	53
5. Conclusiones	54
A. INSTALACIÓN DE LA PLATAFORMA SILK WORKBENCH	56
B. INSTALACIÓN DEL ENDPOINT MARMOTTA	59
C. FUENTES DE INFORMACIÓN	63
Bibliografía.	64
Acrónimos	64

Índice de figuras.

1.1. Esquema proceso anotación 3 capas.	10
2.1. getCapabilities Request	13
2.2. getCapabilities Response	13
2.3. describeFeatureType Request	14
2.4. getFeature Request	14
2.5. getFeature Response	15
2.6. Ejemplo de codificación GML	16
2.7. Ejemplo de formato RDF	17
2.8. Anotación a tres niveles (Maué)	19
3.1. Fuentes de información	21
3.2. Descripción de información de un Feature (Ejemplo 1)	23
3.3. Descripción de información de un Feature (Ejemplo 2)	23
3.4. Búsqueda de Ontología	25
3.5. Proceso de selección de Ontología	27
3.6. Proceso de anotación manual y automático.	30
3.7. Proceso de anotación entre servicios WFS y ontología WFS	32
3.8. Proceso de anotación entre servicios WFS y ontología DCAT	32
3.9. Proceso de anotación entre servicios WFS y ontología de dominio	33
3.10. Propiedades de la clase describeFeatureType	34
3.11. Listado de instancias de la clase describeFeatureType	35
3.12. Visualización de todos los datos de una Feature	36
3.13. Relaciones entre la información anotada en dos fuentes distintas	37
3.14. Búsqueda de Información en base a criterios específicos	38
3.15. Operadores silk	39
3.16. Pantalla inicial Silk Workbench	39
3.17. Creación de proyecto Silk Workbench	39
3.18. Agregar Resources en proyecto Silk Workbench	40
3.19. Formato de las fuentes en proyecto Silk Workbench	40
3.20. Configuración de datos para la verificación Silk Workbench	41
3.21. Proceso de verificación Silk Workbench	41
3.22. Resultado del proceso de verificación Silk Workbench	41
3.23. Creación de un proyecto en la plataforma Silk Workbench	42
3.24. Carga de fuentes al proyecto.	43

3.25. Asignación de formato de entrada fuente IGM	43
3.26. Asignación de formato de entrada fuente UDA	44
3.27. Relación de las fuentes para su análisis	44
3.28. Plugin utilizado para el análisis de las geometrías	45
3.29. Proceso para generación de linking	45
3.30. Generación de Linking	45
3.31. Resultados encontrados	46
3.32. Resultados encontrados validados	46
4.1. Esquema de verificación de información.	48
4.2. WFS anotados con respecto a las ontologías WFS y DCAT	49
4.3. Servicios WFS y capas usadas.	50
4.4. Capas mapeadas	50
4.5. Relación campos no válidos.	51
4.6. Opciones Triple Store marmotta	52
A.1. Selección de descarga de plataforma silk-workbench	56
A.2. Contenido carpeta silk-workbench	57
A.3. Ejecución de plataforma silk-workbench	57
A.4. Plataforma silk-workbench	57
A.5. Plugins en silk-workbench	58
A.6. Carga de plugins a silk-workbench	58
B.1. Descarga apache marmotta	59
B.2. Apache marmotta guía de instalación	60
B.3. Términos de licencia de uso	60
B.4. Selección de paquetes de instalación de apache marmotta	61
B.5. Directorio de instalación	61
B.6. Servicios de apache marmotta	61
B.7. Apache marmotta	62
C.1. Fuentes de información	63



Holger Marcelo Coronel Samaniego, autor/a de la tesis "Implementación de Procesos de Anotación Semántica de Servicios WFS con Ontologías de Dominio", reconozco y acepto el derecho de la Universidad de Cuenca, en base al Art. 5 literal c) de su Reglamento de Propiedad Intelectual, de publicar este trabajo por cualquier medio conocido o por conocer, al ser este requisito para la obtención de mi título de Ingeniero de Sistemas. El uso que la Universidad de Cuenca hiciera de este trabajo, no implicará afección alguna de mis derechos morales o patrimoniales como autor/a

Cuenca, 09/11/2016

A handwritten signature in blue ink, appearing to read "Holger Marcelo Coronel Samaniego", written over a horizontal line.

Holger Marcelo Coronel Samaniego

C.I: 0105986731



Marcos Javier Rodas Zamora, autor/a de la tesis "Implementación de Procesos de Anotación Semántica de Servicios WFS con Ontologías de Dominio", reconozco y acepto el derecho de la Universidad de Cuenca, en base al Art. 5 literal c) de su Reglamento de Propiedad Intelectual, de publicar este trabajo por cualquier medio conocido o por conocer, al ser este requisito para la obtención de mi título de Ingeniero de Sistemas. El uso que la Universidad de Cuenca hiciere de este trabajo, no implicará afección alguna de mis derechos morales o patrimoniales como autor/a

Cuenca, 09/11/2016

A handwritten signature in blue ink that reads "Marcos Rodas".

Marcos Javier Rodas Zamora

C.I: 0103636593



Universidad de Cuenca
Clausula de propiedad intelectual

Holger Marcelo Coronel Samaniego, autor/a de la tesis "Implementación de Procesos de Anotación Semántica de Servicios WFS con Ontologías de Dominio", certifico que todas las ideas, opiniones y contenidos expuestos en la presente investigación son de exclusiva responsabilidad de su autor/a.

Cuenca, 09/11/2016

A handwritten signature in blue ink, written over a horizontal line.

Holger Marcelo Coronel Samaniego

C.I: 0105986731



Universidad de Cuenca
Clausula de propiedad intelectual

Marcos Javier Rodas Zamora, autor/a de la tesis "Implementación de Procesos de Anotación Semántica de Servicios WFS con Ontologías de Dominio", certifico que todas las ideas, opiniones y contenidos expuestos en la presente investigación son de exclusiva responsabilidad de su autor/a.

Cuenca, 09/11/2016

A handwritten signature in blue ink that reads "Marcos Rodas".

Marcos Javier Rodas Zamora

C.I: 0103636593

Agradecimientos

Agradezco por realizar esta tesis a una persona genial a un ser humano espectacular cuya ética de trabajo dedicación y empeño han hecho posible culminar mis estudios con gran sacrificio y trabajo duro. Esa persona es mi abuela Esthela Coronel quien siempre inculco los mejores valores en mi.

UNIVERSIDAD DE CUENCA
desde 1867

Agradecimientos

Agradezco primero a Dios por permitirme cumplir una meta importante en mi vida, a mis padres y hermanos que han sido una fuente de inspiración y apoyo, ya que ellos han estado siempre guiándome y formándome como una persona de bien.

A mi director de tesis. Ingeniero Víctor Saquicela quien a sabido guiarnos durante todo este proceso de desarrollo, compartiendo sus conocimientos que han sido de mucha ayuda para la culminación de esta etapa.

A todos mis amigos y compañeros que han estado ahí como apoyo moral a lo largo de toda esta formación profesional.

Holger Coronel

Dedicatoria

Dedico esta tesis a mi abuela Esthela Coronel quien siempre me ha apoyado durante mis estudios desde la primaria hasta la Universidad.

Marcos Rodas.



UNIVERSIDAD DE CUENCA
desde 1867

Dedicatoria

A mis padres Galo y Olga por haber estado a mi lado en todo momento apoyándome con las decisiones que e tomado durante mis estudios, a mis hermanos por la confianza y el apoyo brindado que a servido de inspiración para seguir adelante y cumplir mis metas.

Holger Coronel

Capítulo 1

Introducción

1.1. Introducción

La información geográfica se encuentra publicada en servicios WFS que se encuentran representados sintácticamente. Para poder representar la información geográfica de manera semántica es necesario anotar los servicios WFS. La anotación de servicios necesita llevarse a cabo con un procedimiento específico, para describir este proceso es necesario conocer el estado del arte de la representación semántica de servicios WFS, los problemas que trae la situación actual y los objetivos que se esperan alcanzar con la descripción e implementación del proceso de anotación semántica.

1.2. Panorama General

La anotación de servicios WFS se encuentra descrita en el documento de MAUÉ[8], las herramientas para anotar semánticamente se centran principalmente en la anotación de bases entidad relación. Los servicios web actualmente se encuentran representados sintácticamente con estándares conocidos como WSDL o el propio WFS y la información proporcionada por estos servicios debe ser interpretada completamente por los usuarios; para representar estos estándares sintácticos en forma semántica, la literatura recomienda un proceso de anotación semántica el cual puede ser manual, automático o semiautomático [13]. El proceso de anotación manual es completamente una tarea humana en la cual una persona define la relación que existe entre la información proporcionada por el servicio web y la ontología o dentro del cual se desea representar la información. La anotación semiautomática utiliza herramientas automáticas pero deben ser operadas por un usuario. Finalmente la anotación automática no requiere la intervención de un usuario.

En la actualidad, del análisis del estado del arte se ha podido definir que no existen herramientas para una anotación automática de servicios WFS o cualquier tipo de servicio web, de hecho existen metodologías definidas para el proceso de anotación que ayudan a sentar las bases de un proceso automático, el cual se pue-

de nutrir de herramientas existentes y de principios para el desarrollo de nuevas herramientas de manera que se pueda lograr un proceso automático para la anotación de servicios WFS y de servicios web en general. Para un proceso detallado de anotación de los servicios WFS, la OGC se centra en tres niveles de anotación, estas son: la anotación del servicio como tal, la anotación de la *metadata* y finalmente la anotación de las instancias.

La anotación del primer nivel se puede realizar automáticamente con una ontología que describa los servicios WFS. El segundo nivel también se puede anotar con una ontología que defina los servicios o un catalogo de datos.

Un último nivel del proceso definido por la OGC trata de anotar conceptos dentro de la ontología a cada atributo. Este paso se puede definir como una interiorización del proceso anterior en un nuevo nivel de la estructura del servicio WFS.

En cuanto al estado del arte en herramientas de anotación de servicios WFS tenemos las siguientes:

- SWING herramienta con interfaz visual
- SAPR que utiliza una aplicación de google para la inyección de anotaciones semánticas.

1.3. Planteamiento del problema y justificación

Las fuentes en la web que proporcionan información geográfica a los usuarios en la actualidad se encuentran dispersas y no estructuradas. Las consultas sobre temas necesarios en estos sitios son complejas, debido a que se necesita recorrer cada sitio buscando de capa en capa para encontrar información de utilidad, lo que conlleva que los valores obtenidos dependan de la sensatez y dedicación que el usuario usa para la búsqueda de información. Por tal motivo se considera una transformación de la información con fuentes geográficas a otra representación de datos que permite tener una fuente de datos estructurada y relacionada con diferentes fuentes, ya sean estos geográficos o de cualquier índole. Dado que las fuentes de información geográfica son varias (WFS, WMS, etc.), este trabajo se centra en fuentes de datos de tipo servicio WFS, que son los de mayor uso en la actualidad y como fuente destino en un repositorio semántico, dados los beneficios que tiene la web semántica para la búsqueda y el enriquecimiento de la información.

1.4. Objetivo General

Implementar un proceso para anotar semánticamente los servicios WFS relacionando sus distintos niveles de abstracción con diversas ontologías de dominio.

1.5. Objetivos Específicos

- Usar ontología en base a un tema puntual de dominio con la cual se anotara la estructura de datos de las capas del servicio WFS.
- Anotar metadatos de servicio WFS con ontología específica para catalogar servicios web.
- Anotar la información geométrica de los servicios WFS con la ontología geosparql.
- Implementar método de invocación de servicios WFS.
- Creación de un repositorio con información semántica sobre los servicios WFS catalogados.
- Verificar la información generada en el proceso de anotación mediante consultas al repositorio local.

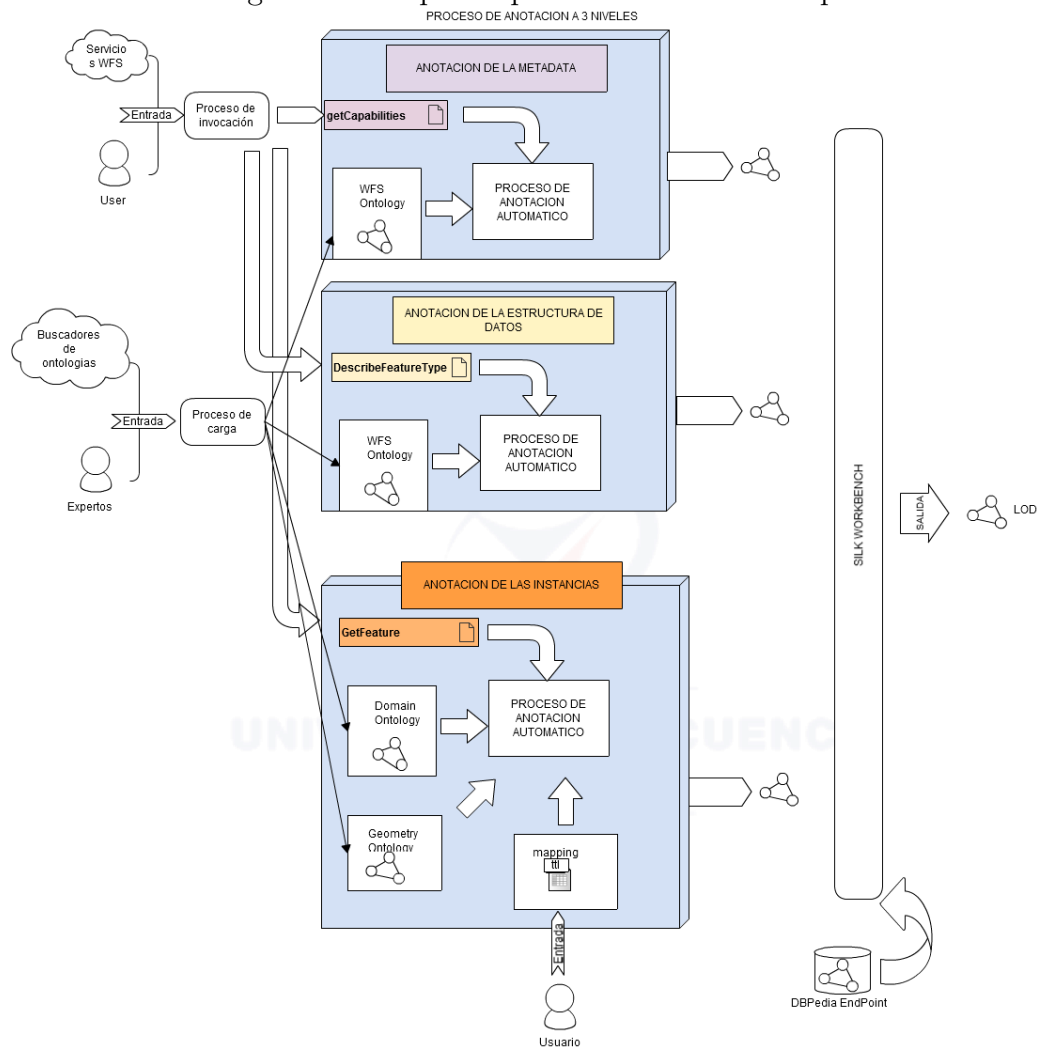
El presente proyecto de tesis tiene como meta final establecer un proceso para pasar de información representada mediante servicios WFS a información representada en un repositorio semántico, el cual permitirá realizar enriquecimiento semántico de la información transformada gracias a las diferentes herramientas dentro de la web semántica. El proceso que se desea establecer deberá hacer uso de las tres operaciones principales dentro de los servicios WFS y definir las restricciones al proceso.

1.6. Estructura del Documento

El proceso que se desea establecer deberá hacer uso de las tres etapas principales dentro de los servicios WFS y definir las restricciones al proceso. A continuación se define la estructura que seguirá el presente documento detallando el proceso realizado. En el segundo capítulo se explica de manera detallada todas las tecnologías y metodologías involucradas en el proyecto dentro de las cuales las más importantes son las tecnologías RDF, RDF schema, OWL, Jena, SPARQL, GEOSPARQL y las metodologías de LOD y de anotación semántica definidas por la OGC para servicios WFS. El capítulo 3 define de manera detallada el proceso de anotación de los servicios WFS utilizando las tecnologías y metodologías descritas. Además se describe el proceso de selección de la ontología de dominio que es de especial importancia para el proceso de anotación. Todo esto se puede resumir de mejor manera en la figura 1.1.

Como se puede observar el proceso de anotación tiene tres etapas, la primera consiste en la búsqueda y selección de servicios WFS y ontologías de dominio para la anotación semántica al igual que la selección de ontologías para la anotación de la metadata de los servicios WFS. La última etapa del proceso como se puede

Figura 1.1: Esquema proceso anotación 3 capas.



ver es el proceso de enriquecimiento semántico con los servicios WFS anotados en los tres niveles definidos por la OGC.

En el capítulo 4 se definen métricas para la medición de los resultados obtenidos en el proceso descrito en el capítulo 3, además la forma de interpretar estos resultados.

Capítulo 2

Marco Teórico.

2.1. Introducción

Dentro del presente marco teórico se definirán diferentes conceptos de suma importancia para el desarrollo del presente trabajo de investigación. Los tópicos involucrados son numerosos por lo que se los ha catalogado dentro de diferentes subtemas los cuales son: Web Semántica, los estándares OGC, y los sistemas de información geográfica.

2.2. Sistemas de Información Geográfica

Los Sistemas de Información Geográfica están formados por los resultados de lo que se conoce como Tecnologías de la Información (TI), por lo que podemos definir de la siguiente manera:

«Conjunto integrado de medios y métodos informáticos, capaz de recoger, verificar, almacenar, gestionar, actualizar, manipular, recuperar, transformar, analizar, mostrar y transferir datos espacialmente referidos a la Tierra»[7].

Las aplicaciones GIS permiten a los usuarios relacionar temáticas como transporte y logística con el análisis de mapas. La captura de información geográfica es digitalizada gracias a la ayuda de sistemas de diseño asistido por computadora y a las orto fotos obtenidas gracias a satélites y drones especializados. La principal manera en la que los sistemas de información geográfica relacionan la información es mediante la localización espacial a la que hace referencia. La veracidad de la información depende de la fuente y el medio tecnológico para obtener dicha información.

La representación de la información en sistemas de información geográfica como zonas de lluvia o zonas de producción agrícola, etc, son representados geométricamente por puntos líneas y polígonos dentro de los diferentes formatos en un sistema de coordenadas geográficas.

La información geográfica por lo general debe de ser procesada luego de ser obtenida, principalmente cuando está en un formato raster y debe pasar a un formato vectorial, también es importante la representación que se escoja para mostrar la

información, pues dependiendo de el sistema de coordenadas la interpretación de la información puede variar enormemente.

Del campo de los sistemas de información geográfica se derivan varios otros tópicos como el análisis de la información y la publicación de la misma en diferentes medios como el web, al igual que la explotación mediante diferentes técnicas.

2.2.1. Sistemas de Coordenadas

Un sistema de coordenadas sirve para ubicar cualquier punto de la superficie terrestre mediante una secuencia de números los cuales indican la latitud (Norte y Sur) y longitud (Este y Oeste), Este tipo de numeración son coordenadas angulares medidas desde el centro de la tierra, conocidas también como coordenadas esféricas[6].

2.2.2. Estándares OGC (Open Geospatial Consortium)

El estándar OGC se centra en el desarrollo de normas de aplicación mediante la adaptación o la ampliación de las normas de servicio Web comunes. La OGC a estado dedicada a la estandarización de interfaces de Servicios Web Geoespaciales para permitir la interoperabilidad.

Todos los Servicios Web OGC están basados en un modelo común y comparten ciertas características.

- Solicitud de operación y respuesta de contenidos.
- Parámetros unificados para las operaciones de petición y respuesta.
- Adopción exclusiva de XML y codificación en valores claves pares.

Bajo dichas especificaciones se han desarrollado ciertos métodos para la publicación y el análisis de los datos geoespaciales en la web(WMS, WFS, WMC).

2.2.3. WFS (Web Feature Service)

El Web Feature Service (WFS) es un estándar utilizado como una interfaz de conexión que interactúa con datos geoespaciales permitiendo a un cliente actualizar o recuperar información, dicha información se encuentra codificada en Geography Markup Language (GML), un estándar que deriva del lenguaje XML mediante el cual se transmiten las ordenes WFS. El WFS usa como plataforma de computación distribuida el protocolo de transferencia HTTP, además posee 3 operaciones para la administración de los elementos geográficos (Features)[4], estas son:

- Operación getCapabilities.
- Operación describeFeatureType.
- Operación GetFeature.

Operación getCapabilities

La operación *getCapabilities* hereda los elementos *ows:ServiceIdentification*, *ows:ServiceProvider*, y *ows:OperationMetadata* de la especificación para la implementación OWS. El WSDL describe las operaciones del servicio WFS describe las operaciones, la sección *FeatureType* define la lista de tipos de *feature* proporcionados por el servicio. La sección *Filter capabilities* muestra las expresiones que pueden ser usadas como predicados de consultas. Por último la sección *FeatureTypeList* contiene una lista de *feature types*[4][14].(ver figura 2.1, 2.2).

Figura 2.1: getCapabilities Request

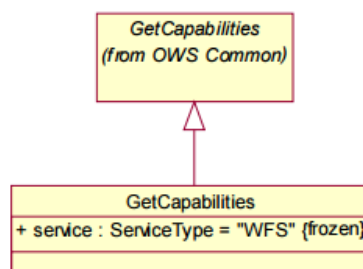
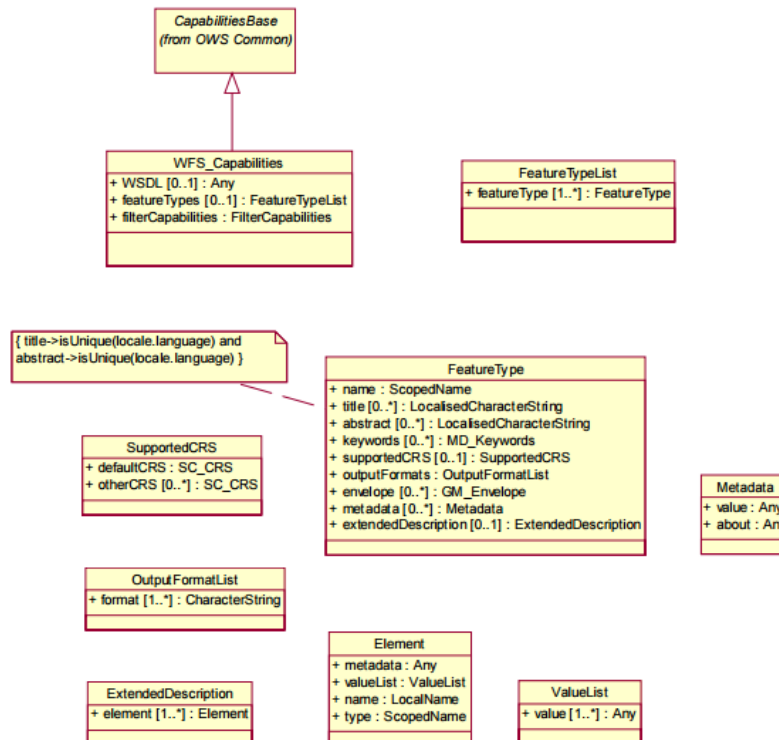


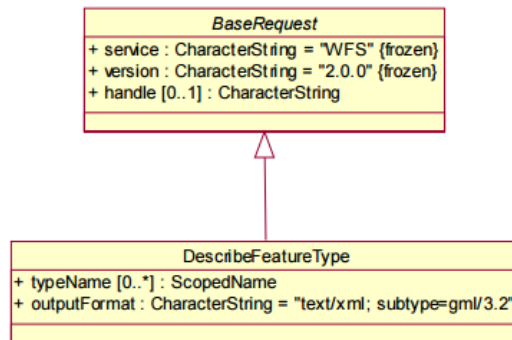
Figura 2.2: getCapabilities Response



Operación describeFeatureType

Esta operación proporciona un esquema de descripción de los tipos de *features* ofrecidos en una instancia WFS, Específicamente, la operación DescribeFeatureType recupera las propiedades y atributos de los *features* disponibles[4][14].(ver figura 2.3)

Figura 2.3: describeFeatureType Request



Operación getFeature

Esta operación recupera todas las instancias que se encuentran en el servicio WFS, es decir los datos con la información como geometrías y el valor de los atributos. la representación canónica de los Features utiliza GML.

Un requerimiento de la operación *GetFeature* contiene una o mas expresiones de consulta la cual identifica las instancias a ser representadas en la respuesta de la operación[4][14]. (ver figura 2.4,2.5)

Figura 2.4: getFeature Request

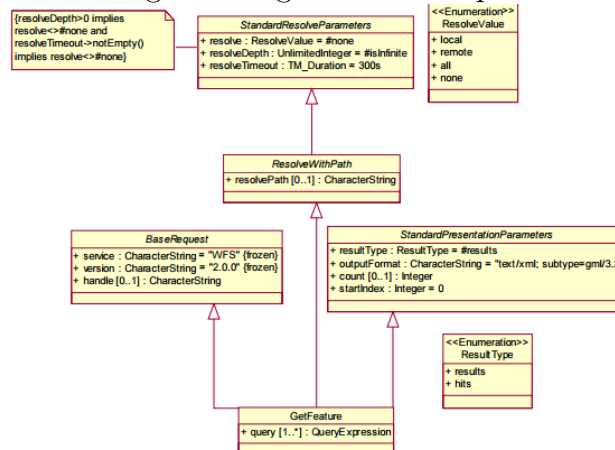
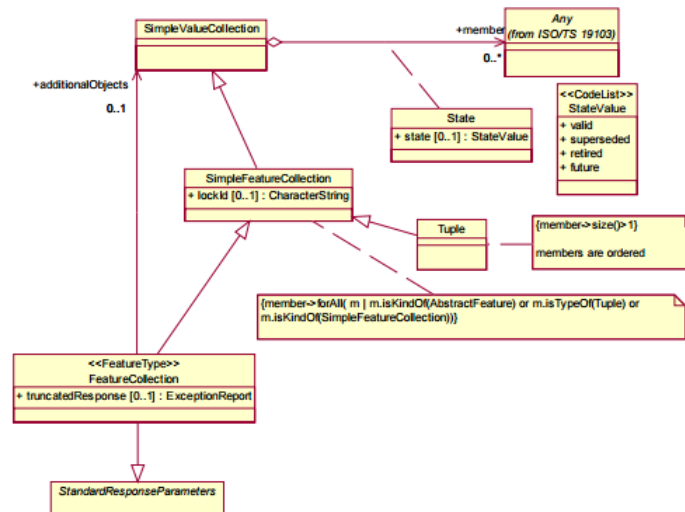


Figura 2.5: getFeature Response



2.2.4. GML(Geography Markup Language)

El GML (Geographic markup language) está basado en un esquema XML el cual describe el almacenamiento y transporte de la información geográfica. Se considera característica geográfica si está asociada a cualquier ubicación en algún lugar de la tierra. Este GML puede ser construido de dos maneras:

- Por la adherencia a las reglas especificadas en el ISO 19109 para esquema de aplicación en UML y conformando ambas restricciones en tales esquemas y reglas para mapearlos a esquema de aplicaciones de GML aplicadas al estándar internacional.
- Por adherencia a las reglas para esquemas de aplicación de GML especificadas en el estándar internacional para las creaciones de esquemas de aplicaciones de GML usando directamente el esquema de XML[3]. (ver figura 2.6)(gml ogc standard)

Figura 2.6: Ejemplo de codificación GML

```

<Landsat7Coverage dimension="2">
  <gml:metaDataProperty>
    <ImageMetadata>
      <SUN_AZIMUTH uom="urn:x-
IHS DM:v2.05a:uom:angle">114.4699122</SUN_AZIMUTH>
      <SUN_ELEVATION uom="urn:x-
IHS DM:v2.05a:uom:angle">66.6625502</SUN_ELEVATION>
      <QA_PERCENT_MISSING_DATA
uom="urn:x-IHS DM:v2.05a:uom:percent">66</QA_PERCENT_MISSING_DATA>
      <CLOUD_COVER uom="urn:x-IHS DM:v2.05a:uom:percent">80</CLOUD_COVER>
    </ImageMetadata>
  </gml:metaDataProperty>
  <gml:rectifiedGridDomain>
    <gml:RectifiedGrid dimension="2"
srsName="urn:ogc:def:crs:EPSG:6.6:32612">
      <gml:limits>
        <gml:GridEnvelope>
          <gml:low>0 0</gml:low>
          <gml:high>17436 15624</gml:high>
        </gml:GridEnvelope>
      </gml:limits>
      <gml:axisName>x</gml:axisName>
      <gml:axisName>y</gml:axisName>
      <gml:origin>
        <gml:Point gml:id="Pt0001"
srsName="urn:ogc:def:crs:EPSG:6.6:32612">
          <gml:pos>270372.375 270372.375</gml:pos>
        </gml:Point>
      </gml:origin>
      <gml:offsetVector srsName="urn:ogc:def:crs:EPSG:6.6:32612">1
0</gml:offsetVector>
      <gml:offsetVector srsName="urn:ogc:def:crs:EPSG:6.6:32612">0 -

```

2.3. Servicios Web Geoespaciales

Los Servicios Web Geoespaciales (SWG) están creciendo día a día con el fin de acceder, integrar y reutilizar datos geoespaciales. Además, los SWG están cambiando la manera en que los sistemas de información y aplicaciones espaciales son diseñadas, desarrolladas y desplegadas. Sin embargo, se ha notado que la búsqueda y recuperación de los datos espaciales ha llegado a tener gran dificultad para los usuarios debido a la variedad en el significado de los datos.

A pesar de que existen ciertos catálogos que proporcionan repositorios de búsqueda de información y servicios, los medios que apoyan el descubrimiento de la información son insuficientes. Por tal motivo es necesario generar procesos que ayuden a encontrar los datos adecuados en el entorno deseado por el usuario.

Es necesario que los servicios cumplan ciertos estándares para facilitar su publicación, y la explotación de la información que contienen. Existen diversas instituciones encargadas de establecer estándares para los servicios geoespaciales entre las que destacan ISO, TC21, y OGC[10].

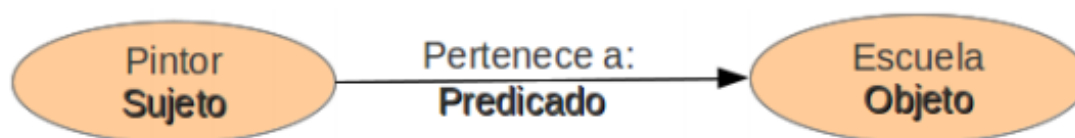
2.4. ¿Que es la Web Semántica?

La web semántica es un conjunto de actividades desarrolladas con la finalidad de crear tecnologías que permitan publicar datos legibles por aplicaciones informáticas añadiendo datos semánticos a la información que se encuentra ya en la web. En la actualidad existen diferentes tecnologías consideradas estándares dentro de la web semántica dentro de las cuales se pueden citar las siguientes: RDF, RDFSchema, OWL, SPARQL[15].

2.4.1. RDF (Resource Description Framework)

Es un conjunto de especificaciones cuyo objetivo inicial es crear un modelo de datos para metadatos lo cual lo ha convertido en el estándar para la definición de recursos web de forma semántica, El modelo de datos que se tiene en el formato RDF tiene por idea el modelo sujeto – predicado – objeto y esta terminología es conocida como triplete.(ver figura 2.7).[12]

Figura 2.7: Ejemplo de formato RDF



2.4.2. OWL (Web Ontology Language)

Es un lenguaje semántico diseñado para representar conocimiento complejo acerca de diferentes tópicos, esta basado en la lógica de manera que esta información puede ser explotada por los computadores. Uno de sus principales objetivos es facilitar un modelo de marcado el cual es construido sobre RDF(Resource Description Framework)[5].

2.4.3. SPARQL

Es un lenguaje estandarizado para la consulta de grafos el cual es una parte esencial de la web semántica como tal pues permite hacer inferencias acerca de la información definida en los repositorios RDF

2.5. Anotación Semántica

La anotación semántica es conocida actualmente como el estudio del lenguaje natural (lenguaje humano), a su vez el procesamiento del lenguaje utilizando

un modelo que ha sido extraído del lenguaje humano, este modelo es conocido como ontología. Esta Ontología extrae y representa el significado de textos en lenguaje natural, haciendo uso de las representaciones de su significado. Con esto se pretende que la información obtenida mediante ciertos recursos sea inteligente y a su vez facilite la búsqueda y navegación a través de sus relaciones[1].

2.5.1. ¿Qué es una anotación?

Por anotación se considera notas, explicaciones, o algún tipo de reseña que se asocie a un documento o una pagina Web, o como datos sobre las entidades o conceptos de una ontología, también indica las referencias que se encuentran en un texto sobre un repositorio semántico. Las anotaciones semánticas generan una relación entre la información contenida en un documento o parte de ella, y un identificador conceptual URI (*Universal Resource Identificador*) tomado de una ontología o de otra fuente de conocimiento[1].

2.5.2. Anotación manual

En un proceso de anotación manual el responsable de analizar la información y encontrar las características principales de los datos y algún tipo de vocabulario es el humano, por tal motivo se considera este tipo de anotación una tarea completamente humana. Este tipo de anotación es considerado como el mas preciso hasta el momento debido a que el mapeo se lo hace mediante un alto nivel de extracción de la información. Puesto que el proceso manual es lento y existen miles de WFS este proceso queda descartado[8][9].

2.5.3. Anotación semi-automática

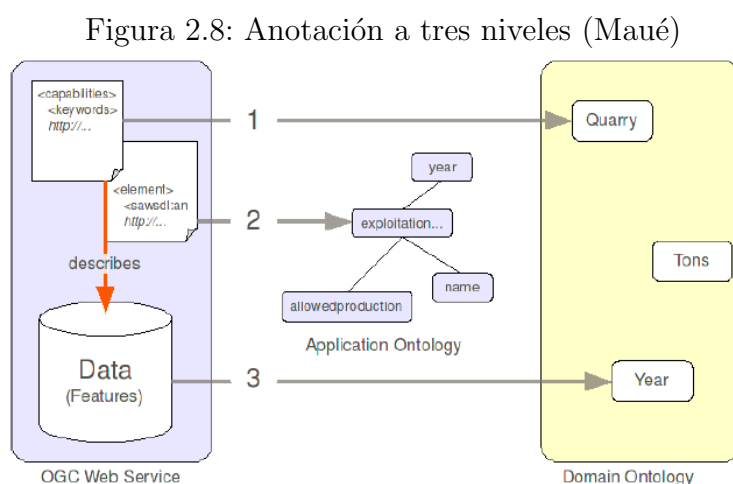
Este tipo de anotación esta formado por una anotación manual y con procesos de extracción de información automáticas, debido a que ciertas partes de este proceso esta automatizado no es tan preciso como el método de anotación manual, Este método posee una intervención humana mucho más baja que la anterior donde su intervención era completa[8][9].

2.5.4. Anotación automática

Este proceso es un método completamente automático, no depende de la intervención humana, esta enfocado más en buscar formas de crear anotaciones de acuerdo al dominio de una ontología especifica, basándose en algoritmos tanto de alineación o de reconocimiento[8][9].

2.5.5. Anotación de servicios OGC

La OGC ha definido las mejores practicas para la anotación semántica de servicios WFS. En la definición de las mejores practicas se desarrolla el concepto de anotación en tres niveles: los meta datos del servicio que pueden ser anotados por una ontología que describe servicios WFS como WSMO, los modelos de datos y descripción de procesos que pueden ser anotados por una ontología que describa catálogos de datos, y finalmente las instancias del servicio que pueden ser anotadas por una ontología de dominio, en la figura 2.8 podemos observar la anotación en tres niveles por parte de Maué[8][13].



2.6. Linked Data (Datos Enlazados)

Los Datos Enlazados o también llamados Linked Data hoy en día vienen siendo una gran extensión a la web que se tiene en la actualidad, y se la utiliza para la publicación de datos estructurados en la web. Es decir, toda la información que se encuentra en base a un documento se encuentra relacionada o interconectada, lo que permite tener una mayor integración y mas accesibilidad permitiendo el entendimiento tanto por los humanos como por las maquinas[2]. Para la publicación en la web Linked Data se basa en 4 puntos:

- Utilizar URIs para nombrar todas las cosas en la Web.
- Utilizar HTTP para que todas las personas puedan encontrar la información en la Web.
- Información útil al ingresar por HTTP y esta debe estar e formato estándar RDF y SPARQL.
- Incluir links a otras URIs para poder navegar en la información que se encuentra enlazada.

Capítulo 3

Anotación semántica de servicios WFS

3.1. Introducción

La motivación para definir un proceso de anotación semántica que tenga como fuente servicios WFS (Web Feature Service) es clasificar la información contenida en estos servicios, debido a que los servicios WFS están almacenados con estructuras diferentes en cuanto a datos y la gran mayoría no describe de manera adecuada la información que contienen. Otra falencia que tiene la información de los servicios WFS es que la información está aislada, es decir, cualquier relación de datos entre una capa y otra es determinada de forma empírica. El objetivo final de la anotación semántica es realizar consultas semánticas a la información de los WFS.

En este capítulo se detalla el proceso seleccionado para anotar servicios WFS. Los estándares seguidos para la anotación de servicios WFS son los recomendados por la OGC dentro de el documento técnico escrito por Patrick Maué [8]. Los estándares de la OGC definen lineamientos pero no procesos descritos de manera estructurada. También se describe el proceso de selección de las ontologías para la anotación de los servicios y las relaciones que se pudieron obtener partiendo de la anotación de los servicios.

3.2. Búsqueda servicios WFS

El proceso de búsqueda de servicios WFS se realizó de forma manual, debido a que no se cuenta con procesos definidos de búsqueda automática, o con herramientas desarrolladas por terceros para este proceso. El primer paso para la búsqueda de servicios WFS es identificar el tipo de información que se necesita para el proceso de anotación, junto con las fuentes de los servicios. Para encontrar fuentes de servicios el primer paso es utilizar buscadores web, en este caso dió resultados mixtos. La mayoría de los resultados fueron definiciones de lo que es un WFS. Posteriormente se intento con palabras como geoportal con mejores resultados pues es en los geoportales es donde se encuentran publicados los servicios

WFS, en la figura 3.1 se puede observar algunos de los servicios WFS encontrados.

Figura 3.1: Fuentes de información

SERVICIOS WFS	LUGAR
http://gis.uazuay.edu.ec:8080/geoserver/ows?service=wfs&version=2.0.0&request=GetCapabilities	Universidad del Azuay
http://mapainteractivo.ambiente.gob.ec/geoserver/ows?service=wfs&version=2.0.0&request=GetCapabilities	Mapa Interactivo Ambiental
http://www.geoportaligm.gob.ec/regional/ows?service=wfs&version=2.0.0&request=GetCapabilities	Instituto Geográfico Militar
http://www.geoportaligm.gob.ec/nacional/ows?service=wfs&version=2.0.0&request=GetCapabilities	Instituto Geográfico Militar
http://www.geoportaligm.gob.ec/galapagos25/wfs?service=WFS&request=GetCapabilities&version=2.0.0	Instituto Geográfico Militar
http://geoportal.magap.gob.ec/geomagap/ows?service=wfs&version=2.0.0&request=GetCapabilities	MAGAP
http://www.ign.es/wfs-inspire/services/transportes?version=2.0.0&request=GetCapabilities&service=WFS	Transporte España
http://geocatmin.ingemmet.gob.pe/arcgis/services/PSAD56_18/SERV_CATASTRO_MINERO_18/MapServer/WFSServer?request=GetCapabilities&service=WFS	Catastro minero de Perú
http://geoservidor.minam.gob.pe/arcgis/services/ServicioCCNN/MapServer/WFSServer?request=GetCapabilities&service=WFS	Comunidades nativas Perú
http://mtcgeo2.mtc.gob.pe:8080/geoserver/MTC_gis/wfs?service=WFS&version=2.0.0&request=getCapabilities	Red vial nacional Perú
http://siga.conagua.gob.mx/wfsconnector/com.esri.wfs.Esrimap/CuerposAgua_Features?request=getcapabilities&service=wfs&version=2.0.0	Cuerpos principales de agua México
http://siga.conagua.gob.mx/wfsconnector/com.esri.wfs.Esrimap/RHA_Features?request=getcapabilities&service=wfs&version=2.0.0	Regiones hidrológicas administrativas
http://siga.conagua.gob.mx/wfsconnector/com.esri.wfs.Esrimap/RiosPrincipales_Feature?request=getcapabilities&service=wfs&version=1.0.0	Ríos principales
http://siga.conagua.gob.mx/wfsconnector/com.esri.wfs.Esrimap/Acuiferos_Features?request=getcapabilities&service=wfs&version=2.0.0	Acuíferos
http://ontorisk.cedia.org.ec/geoserver/ows?service=wfs&version=2.0.0&request=GetCapabilities	CEDIA
http://geoinec.inec.gob.ec/geoinec/ows?service=wfs&version=2.0.0&request=GetCapabilities	INEC

desde 1867

3.3. Anotación de los meta datos del servicio WFS

Los estándares definidos por la OGC para anotar servicios WFS, se enfoca primero en anotar los meta datos del servicio; mismos que se encuentran en la operación *GetCapabilities* dentro de las etiquetas *Service Provider*, *Title*, *Abstract*, *IndividualName*, y *PositionName*.

Los meta datos se anotaron en una ontología que describe los servicios WFS tal como lo sugiere el estándar de la OGC.[8]. La ontología escogida para anotar el servicio WFS en un inicio solo describía las operaciones por lo que fue necesario agregar propiedades para anotar los atributos mencionados anteriormente.

3.4. Anotación de modelo de datos

El estándar de la OGC describe un segundo nivel de anotación semántica que se enfoca en los geo procesos y los modelos de datos que tiene un servicio WFS. Los geo procesos se refieren a las operaciones que se pueden anotar dentro del servicio mientras que los modelos de datos se refiere a la estructura de cada

capa. En esta implementación se anoto solamente los geoprosesos, para esto se recorrió la etiqueta *FeatureTypeList* y se anotaron las operaciones *GetFeature* y *DescribeFeatureType*, pero no se anoto la respuesta del *DescribeFeatureType* pues es un proceso demorado debido al numero de capas que contienen los servicios.[8]

3.5. Análisis de la información de los servicios WFS

Con el conjunto de servicios encontrados se debe analizar la estructura de los datos del servicio WFS, para lo cual, se utiliza la operación *DescribeFeatureType* descrita en el capítulo 2. El análisis de los meta datos de los servicios encontrados sirvió para establecer las siguientes conclusiones:

- El análisis de los meta datos de un servicio WFS consume una gran cantidad de tiempo debido a que es un proceso manual.
- La mayoría de las capas dentro de un servicio WFS contienen campos con nombres poco descriptivos o inentendibles en algunos casos.
- La calidad de los meta datos de los servicios WFS depende de la fuente que los publica y de el proceso para subir esta capa.

Es difícil saber la calidad de los datos de un servicio, pero se puede esperar homogeneidad en todos los servicios publicados por una misma fuente; por ejemplo, los datos publicados por el IGM (Instituto Geográfico Militar) contienen atributos en la mayoría de sus capas que no están bien descritos, en la capa *igm:antena parabólica* se puede apreciar accediendo a la operación *DescribeFeatureType* descrita en el capítulo 2 la lista de los atributos que tiene esta capa. Es por esto que se accede a estos datos con la siguiente URL:

```
http://www.geoportaligm.gob.ec/galapagos25/wfs?service=wfs&request=DescribeFeatureType&typeName=igm:antena_parabolica
```

que brinda los siguientes atributos: *f code*, *descripcion*, *acc*, *acc desc*, *use*, *use desc*, *txt*, *isla*, y *the geom*

Como se observa la mayoría de los nombres de los campos no describen de manera adecuada a que hacen referencia, por lo que se vuelve necesario analizar las instancias de esta capa para explorar el contenido proporcionado por la operación *GetFeature* descrita en el capítulo 2 mediante la URL http://www.geoportaligm.gob.ec/galapagos25/wfs?service=wfs&request=GetFeature&typeName=igm:antena_parabolica. Al analizar los *features* es notable que el contenido de sus campos no aporta información útil para la anotación semántico de los mismos. Los *features* antes mencionados se pueden ver en la figura 3.2

Figura 3.2: Descripción de información de un Feature (Ejemplo 1)

```

--<wfs:FeatureCollection numberOffFeatures="21" timeStamp="2016-01-07T23:16:54.218-05:00"
version=1.1.0&request=DescribeFeatureType&typeName=igm:antena_parabolica http://www.openg
--<gml:featureMembers>
--<igm:antena_parabolica gml:id="antena_parabolica.15">
  <igm:fcode>AT010</igm:fcode>
  --<igm:descripcio>
    OBJETO CÓNCAVO UTILIZADO PARA LA TRANSMISIÓN O RECEPCIÓN DE SEÑ
  </igm:descripcio>
  <igm:acc>1</igm:acc>
  <igm:acc_desc>EXACTA</igm:acc_desc>
  <igm:use>0</igm:use>
  <igm:use_desc>DESCONOCIDO</igm:use_desc>
  <igm:txt>Antena</igm:txt>
  <igm:isla>ISLA SAN CRISTOBAL</igm:isla>
--<igm:the_geom>

```

Campos no descritos correctamente.

La única información útil de este servicio es el campo *igm:isla*, y su posición exacta con el campo *the_geom*, pero el resto de campos es imposible de descifrar pues no permiten diferenciar al usuario a que hace referencia ya sea por el nombre del campo como por el contenido del mismo. Por ejemplo, en el servicio se puede apreciar que el campo *acc_desc* siempre contiene el valor EXACTA pero ni el nombre del campo ni su contenido ayudan a determinar a que hace referencia este campo. Para poder descifrar el valor de la información que este campo representa se tendría que recurrir a personas expertas en el tema o a su vez las personas encargadas de recolectar o publicar dichas capas, debido a que si se busca un experto en el tema es poco probable que entienda el valor de la información solo leyendo el nombre del campo.

Con el análisis previo se da a conocer que la información que posee la capa del servicio es de mala calidad, para aseverar que la información está en mal estado o incomprensible, es necesario analizar por lo menos una capa, para este ejemplo se utiliza la capa *igm:anteplaya*, la cual posee los siguientes campos:

f code, descripción, mcc, mcc desc, txt, isla y the_geom.

Nuevamente en esta capa los campos tienen nombres no descriptivos, por lo que se procede a analizar la operación *GetFeature* para ver si se puede obtener información que describa a que hace referencia esta capa pero como se puede observar en la figura 3.3 se tiene el mismo problema que en la primera capa.

Figura 3.3: Descripción de información de un Feature (Ejemplo 2)

```

--<wfs:FeatureCollection numberOffFeatures="131" timeStamp="2016-01-07T23:05:18
version=1.1.0&request=DescribeFeatureType&typeName=igm:anteplaya http://www.ope
--<gml:featureMembers>
--<igm:anteplaya gml:id="anteplaya.0">
  <igm:fcode>BA020</igm:fcode>
  --<igm:descripcio>
    PARTE DE LA ORILLA O PLAYA LA CUAL CAE ENTRE LA MARCA DE
  </igm:descripcio>
  <igm:mcc>0</igm:mcc>
  <igm:mcc_desc>DESCONOCIDO</igm:mcc_desc>
  <igm:isla>SAN CRISTOBAL</igm:isla>
+<igm:the_geom></igm:the_geom>

```

Campos no entendibles

Después de haber analizado varias capas de este servicio del IGM de la manera descrita en los dos ejemplos anteriores, y habiendo obtenido resultados similares;

se llega a la conclusión de que este servicio prácticamente no proporciona información útil para la anotación semántica. La única información útil es la geométrica, debido a que esta pertenece al estándar GML descrito en el capítulo 2. En conclusión, se puede decir, que este servicio sería difícil de anotar pero su información geométrica puede aportar al procesos de enlace de datos.

El caso de estudio anterior sirve para ilustrar como identificar un servicio que no brinda información de calidad, tanto en su descripción de datos o metadatos como en las instancias de una capa. A continuación se detalla lo contrario, un servicio donde la información esta bien detallada. En la operación *GetCapabilities* cada capa consta con un *abstract*, que define a que se hace referencia, el servicio al que se hace referencia se puede acceder mediante la URL <https://rsg.pml.ac.uk/geoserver/ows?service=wfs&request=GetCapabilities> y pertenece al *remote sensing group*. Al igual que en el caso anterior se procede a analizar la operación *DescribeFeatureType* de la capa *topp:tasmania water bodies* la cual brinda los siguientes campos.

- THE GEOM
- AREA
- PERIMETER
- COUNTRY NAME
- CONTINENT

Con el acceso a la operación *DescribeFeaturetype* de esta capa se puede ver una diferencia con las capas del anterior servicio del IGM. Los nombres de los campos de esta capa describen claramente a que se hace referencia en ellos; el área, el perímetro y el tipo de cuerpo de agua a las que hace referencia el polígono descrito en la geometría. Es un hecho que aunque los meta datos de la capa tenga nombres claros de campos es posible que exista información que no este correcta dentro de los mismos en la operación *GetFeature*. Al acceder a la operación *DescribeFeaturetype* es notorio que la información proporcionada por este servicio facilitará el proceso de anotación debido a que los campos están bien descritos y la información contenida por estos es coherente.

3.6. Selección de ontologías

La selección de ontologías también se realiza siguiendo un proceso que puede ser resumida en los siguientes pasos:

1. Listar las capas que se quieren anotar.
2. Verificar la validez de los datos proporcionados por la capa.
3. Búsqueda de ontologías de diferentes maneras.

4. Verificar la viabilidad para mapear la ontología seleccionada con las capas seleccionadas.

Lista de capas para el proceso de anotación

Con los servicios escogidos en el proceso de búsqueda de servicios WFS se a generado una lista de servicios los cuales ayudan en el proceso de anotación, se procede a recorrer el servicio mediante la operación *DescribeFeatureType*, con la finalidad de saber que clases y propiedades debe contener la ontología que se va a utilizar. Para la ilustración del procedimiento que se siguió se utiliza como ejemplo la capa *water bodies* del servicio <https://rsg.pml.ac.uk/geoserver/ows?service=wfs&request=GetCapabilities>.

Al acceder a esta capa se obtiene meta datos bien definidos, por lo que es de utilidad para el proceso de búsqueda de ontologías. Recopilados los campos que se van a utilizar para el mapeo, se procede a definir el campo dentro de la ontología de dominio al cual se mapearán.

Verificar la validez de los datos proporcionados por la capa

En la sección anterior 3.5 se validaron ciertos datos de forma manual en los servicios WFS datos que se consideran no aptos para el uso, debido a que la información que estos presentan no está clara o fácil de entender. Para efectos de buscar una ontología de la manera mas rápida posible se decidió que la capa se mapearía a la ontología.(ver figura 3.4).

Figura 3.4: Búsqueda de Ontología



Búsqueda de ontologías por varias maneras

Primero se observó la capa de la cual se puede sacar las palabras claves que van a ayudar en la búsqueda de las ontologías, puesto que se ha decidido que las clases o al menos una clase de la ontología debe tener una relación con la capa. En el caso antes descrito el nombre de la capa de ontología es *topp:tasmania water bodies*, se puede obviar el prefijo *topp* pues no aporta un significado ni semántico ni sintáctico. El siguiente paso es eliminar la palabra *tasmania* pues hace referencia a un lugar y las ontologías hacen referencias a conceptos no a lugares. Finalmente se procede a realizar la búsqueda con las palabras *water bodies* junto a la palabra *ontology*.

Los resultados que se obtienen con esta búsqueda en el motor de búsqueda *Google* hacen referencia sobre todo a *papers*. Dentro de los *papers* en ocasiones

se hace referencia a ontologías pero no se especifica en donde se puede acceder a esta ontología como se puede observar en el presente *paper* <http://ceur-ws.org/Vol-1063/paper3.pdf>. En este, se describe la ontología con la que se va a trabajar pero no hay ningún enlace hacia la ontología, como este son la mayoría de los casos de los *papers* que se obtienen en la búsqueda con buscadores como *Google*. Por tal motivo se decidió no seguir con este proceso para buscar las ontologías, posterior a esto se procedió a realizar la búsqueda en sitios web diseñados específicamente para buscar ontologías como lo son <http://swoogle.umbc.edu/> y <http://watson.kmi.open.ac.uk/WatsonWUI/> con las mismas palabras que se intento en el motor de búsqueda google.

Los resultados obtenidos en estos motores de búsqueda especializados son mucho mejores que los anteriores. Estos hacen referencia directa a las URLs donde se puede acceder a la ontología. De esto se puede sacar las clases que estarían relacionadas con la capa, el próximo paso sería ver que propiedades son las que se podrían relacionar con los campos de la capa a la que se hace referencia. En este punto se genera uno de los inconvenientes grandes del proceso de la selección de la ontología, pues al estar los servicios WFS con estructuras diferentes en base a su información como se dijo anteriormente en este capítulo. Existen ocasiones en las que los nombres de los campos no tienen ningún sentido, por ejemplo, en una de las ontologías que se recupero del buscador *watson* con referencia a *water bodies* o cuerpos de agua la cual puede ser descargada desde la siguiente URL http://lstdis.cs.uga.edu/proj/traks/ontologies/ontology_9.owl. En esta ontología se encontró una clase que tiene concordancia con la capa *tazmania water bodies*. Después de un recorrido de las clases de la ontología se encontró que en primera instancia podría tener relación con la capa *water bodies* y la clase *water structure*. Al analizar los campos de la capa se encuentra un campo *water type* que en todas las capas dice Lake. En la ontología que se había mencionado se encuentra una subclase de *water structure* que tiene por nombre *Lake*, de manera que los capas se deberían relacionar a la clase *Lake* de la ontología.

Verificar la viabilidad para mapear la ontología seleccionada con las capas seleccionadas

El próximo paso es ver las propiedades de clase en la ontología y los campos en la capa para ver como se pueden relacionar. Dentro de la capa los nombres de los campos son: AREA, PERIMETER, COUNTRY NAME y CONTINENT. Las propiedades de la clase *Lake* se deriva de la clase *water feature* que hereda de la clase *water estructura* que hereda de *place* que tiene las siguientes propiedades: *bounding box*, *geographic description*, *elevation*, *est population*. Como se puede ver ninguna de estas propiedades concuerdan por lo que, si se quiere realizar una anotación se necesita crear la propiedad que haga referencia en este campo por ejemplo en este caso el atributo área se convertiría en la propiedad **tieneArea**.

El proceso para la selección de ontologías previamente descritos. (ver figura 3.5).

Figura 3.5: Proceso de selección de Ontología



3.7. Invocación de los servicios WFS

Al invocar el servicio con una conexión java surgen cierto tipo de inconvenientes, como el hecho de que la conexión hacia el servicio web se puede perder, por lo cual fue necesario formar archivos temporales donde se almacenaron los servicios en formato GML. También se intento usar la librería *geotools* pero esta no soportaba la version 2.0.0 del estándar WFS, también se pudo constatar que agilizaba el desarrollo de la aplicación acceder al GML directamente. Se pudo constatar que el proceso de lectura de un servicio WFS con las tecnologías antes descritas es variable dependiendo del tamaño de los *features* que existan en las capas de un servicio, es decir, es mucho mas rápido leer una capa que está enteramente compuesta de puntos que una con polígonos bastante detallados.

Otro inconveniente que se encontró en la lectura de los servicios es la heterogeneidad de las etiquetas GML, a pesar que se decidió leer solo servicios WFS con el estándar 2.0.0 se encontraron ciertas divergencias entre los distintos servicios en algunas etiquetas de suma importancia, por ejemplo, en la gran mayoría de los servicios se podía acceder a los *features* dentro de la operación *GetFeature* leyendo los nodos que están dentro de la etiqueta *wfs:member* pero se encontraron casos en los que para acceder a los *features* de ciertos servicios se vio la necesidad de explorar la etiqueta *gml:featuremembers* pues discrepaba con la versión 2.0.0

estándar en la etiqueta principal.

Por último al momento de leer los servicios WFS se debió identificar cual era el tipo de coordenadas que se manejaba en los servicios WFS. Puesto que todos utilizan la proyección ESPG:4326. Por lo que se realizó la conversión de coordenadas cuando era necesario para poder explotar los servicios anotados.

3.8. Mapeo R2RML

Los dos niveles de la anotación del servicio WFS (metadata y data) son automáticos pero el mapeo es manual, para atacar al último nivel del servicio se ha utilizado un mapeo semiautomático que utiliza el lenguaje R2RML para plasmar los mapeos entre las diferentes capas y las clases de las ontologías de dominio. Para iniciar con el proceso de mapeo de las ontologías con los servicios WFS se toma como insumos los resultados obtenidos de la selección de servicio al igual que la selección de la ontología de dominio.

De la selección de servicios se puede obtener los nombres de las capas y los campos dentro de esa capa que se vayan a anotar, mientras que, de la selección de la ontología de dominio se selecciona la clase y las propiedades, como se explicó anteriormente al encontrar relación entre capas y clases en una ontología no significa que también exista una relación entre los campos de la capa y las propiedades de la clase. De todas maneras en caso de existir esa relación se debería realizar el mapeo correspondiente.

Lo descrito en el párrafo anterior se puede comprender de mejor manera en el siguiente ejemplo, si se tiene una capa de cuerpos de agua más concretamente lagos y dentro de la ontología existe una clase lago con correspondencia perfecta entre clase y capa, por lo que se debe realizar un análisis de concordancia entre las campos de la capa y las propiedades de la clase. En este caso las propiedades a ser anotadas serían ÁREA, PERÍMETRO y CONTINENTE; ninguna de las cuales corresponde a alguna de las propiedades de la clase Lago. Al tener esta disyuntiva existen tres opciones: la primera crear propiedades adecuadas dentro de la ontología de dominio para realizar el mapeo, la segunda no mapear los atributos que no existan de manera que no se anote nada de esa capa y por último mapear los campos existentes con las propiedades existentes aunque no exista concordancia entre los mismos. Solo la primera opción devolvería resultados significativos de manera que es la que se ha escogido para realizar el mapeo de el último nivel del servicio [11].

```
<#TriplesMap1>
  rr:logicalTable [ rr:tableName "prefijo:nombre capa" ];
  rr:subjectMap [
    rr:template "IRI de la clase";
    rr:class ex:nombre de la clase;
```

```

];
rr:predicateObjectMap [
  rr:predicate ex:nombre de la propiedad;
  rr:objectMap [ rr:column "nombre del campo" ];
];

```

Dentro de la aplicación una vez se leen todos los insumos para el mapeo de la ontología de dominio se procede a realizar una lectura del archivo *mapping*. Este archivo contiene las diferentes relaciones entre las ontologías y las capas del servicio WFS. Para ilustrar como funciona este proceso se procede a detallar un ejemplo con la capa *tazmania water bodies* a partir del cual se procede a definir las relaciones con la ontología *hidro_ontology*.

El primer paso es añadir el IRI mediante el cual se identifica la ontología con la cual se quiere realizar el mapeo. En el presente caso es el siguiente:

```
@prefix ex: <http://geo.linkeddata.es/ontology/>.
```

Una vez agregado la IRI mediante la cual se identifica la ontología se agrega el mapeo en el formato R2RML en el cual se puede ver que primeramente se mapea la capa con la clase dentro de la ontología. Posteriormente se mapea los campos de la capa con las propiedades pertenecientes a las clase que se mapeo en un principio.

El mapeo en R2RML se complementa en la aplicación con la búsqueda mediante SPARQL de las clases y propiedades que concuerden con la información leída desde el servicio WFS para lo cual se utiliza la librería *Jena* mediante el siguiente pseudocódigo.

```

ncapas=x;
ncampos=0;
PARA i=1:ncapas HAGA
  SI nombrecapa = lista_capas_mapeo ENTONCES
    ncampos = numero_campos_servicio;
    PARA j=1:ncampos HAGA
      SI campos = campos_mapeo ENTONCES
        setear campo a ontología;
      FIN SI
    FIN PARA
  FIN SI
FIN PARA

```

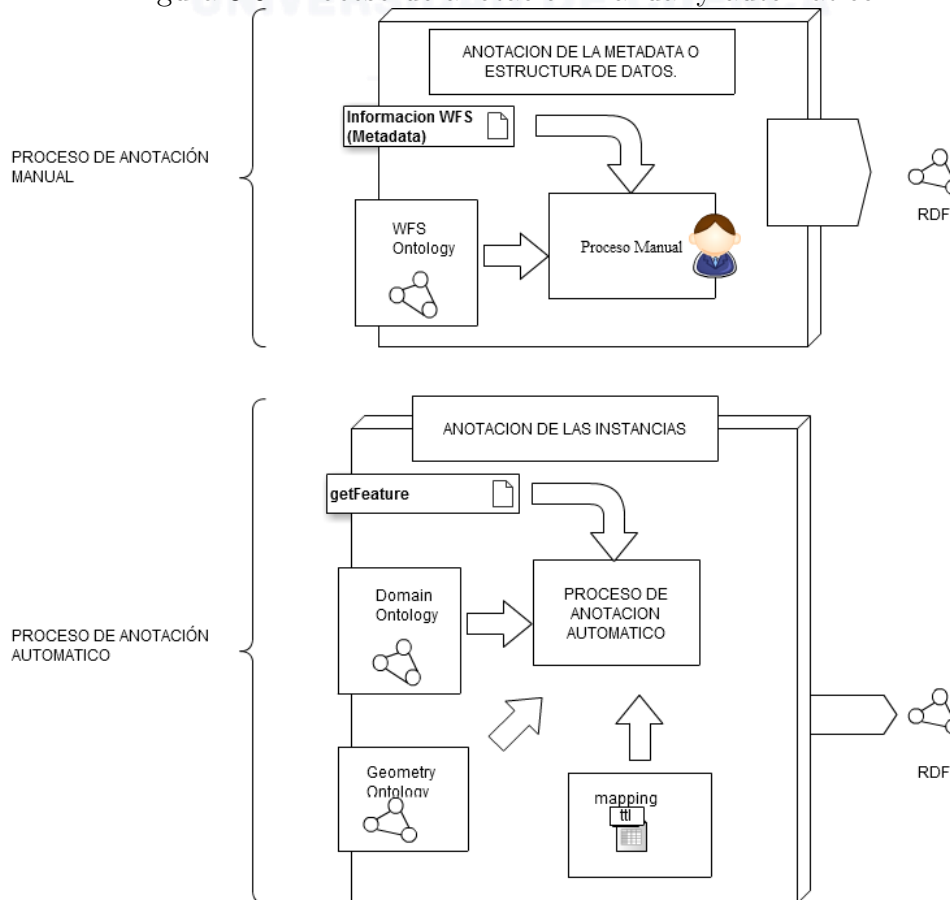
3.9. Técnicas de Anotación

Los procesos de anotación semántica que se han utilizado en este proyecto son: Anotación Manual y Anotación Automática.

3.9.1. Anotación Manual

En la anotación manual los usuarios son los responsables de identificar y analizar los servicios WFS, con la finalidad de relacionarlos con algún vocabulario o base de conocimiento. La anotación manual tiene como desventaja que al ser una tarea humana es propensa a los errores dependiendo de la complejidad de la ontología con la cual se vaya a anotar el servicio, por el contrario el nivel de certeza al extraer la información semántica es mucho mayor. En el proceso se anotaron las operaciones *GetCapabilities* y *DescribeFeatureType* de manera automática. Dentro de la operación *GetCapabilities* se extrajo la información necesaria para poder anotar las URLs mediante las cuales se podía llegar a las operaciones *describeFeatureType* y *GetFeature* para que se pueda catalogar la información contenida dentro del servicio web.(ver figura3.6)

Figura 3.6: Proceso de anotación manual y automático.



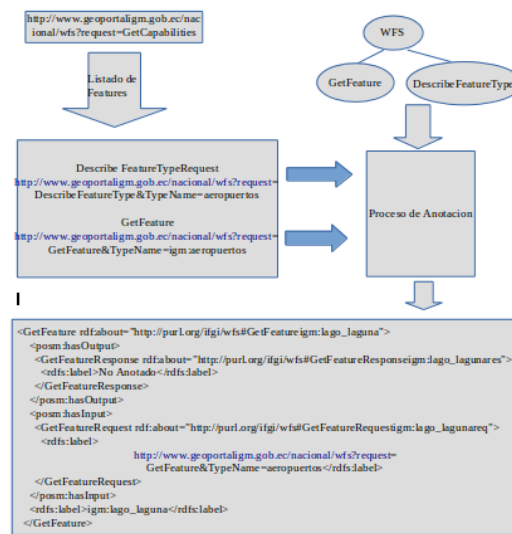
Descripción de anotación de servicio WFS con respecto a la ontología WFS

Las direcciones de los servicios WFS se encuentran almacenados en un archivo de texto plano, en el cual cada línea pertenece a un servicio WFS específico. El objetivo de esto es tener una serie de Servicios, los cuales serán leídos de manera automática por la aplicación. La lectura de los WFS se realiza navegando las diferentes operaciones que contiene el WFS. La primera operación que se utiliza en la aplicación es la de *GetCapabilities*, como por ejemplo: `http://www.geoportaligm.gob.ec/nacional/wfs?request=GetCapabilities`

De esta operación se puede obtener información acerca del servicio, es decir, los meta datos del servicio WFS. La información obtenida en el servicio WFS se anota con una ontología que describe la metadata, en la actual aplicación se implementaron procesos para anotar los meta datos proporcionados por el *GetCapabilities* con la ontología WFS la cual detalla específicamente propiedades de un servicio WFS. Luego se realizó un proceso de anotación para la ontología DCAT que define servicios web en general. La anotación de estas ontologías se hizo de manera semiautomática pues se trabaja con estándares.

El primer paso para anotar la operación *GetCapabilities* es obtener una lista de los componentes dentro del WFS que compartan la misma representación espacial lo cual es posible leyendo los elementos que están dentro del elemento *FeatureTypeList*. Se accede a esta mediante la librería *URLConnection* que permite un flujo de comunicación con la URL de la operación *GetCapabilities*. Se optó por esta técnica de lectura de las diversas operaciones del servicio WFS en vez de librerías implementadas pues con la primera técnica no se puede recorrer un servicio WFS en su última versión la cual a la fecha es la 2.0.0. Una vez obtenida la lista de los atributos dentro del servicio WFS se procedió a anotar las URL de los *requests* de las operaciones *GetFeature* y *DescribeFeatureType* en las clases del mismo nombre dentro de la ontología WFS. Por último se anotaron las propiedades *hasInput* con la URL del *request GetFeauture*, *DescribeFeatureType* y las propiedades *hasOutput* se anotaron con el valor No Anotado[8].(ver figura3.7).

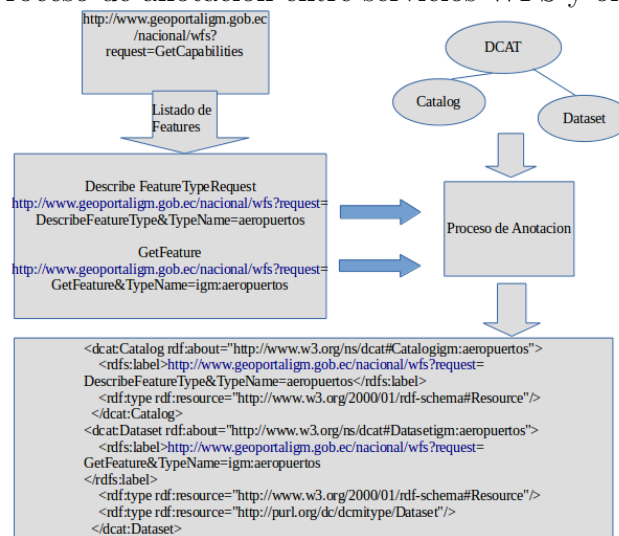
Figura 3.7: Proceso de anotación entre servicios WFS y ontología WFS



Descripción de la anotación de servicio WFS con respecto a la ontología DCAT

Posteriormente se procede a anotar el servicio WFS con la ontología DCAT para lo cual se lee la ontología con la librería *Jena*. Después se anotaron las clases *Dataset* y *Catalog*. Dentro de la clase *Dataset* se anotó la URL del *request* de la operación *GetFeature* pues con esta operación se puede tener acceso a los arreglos de datos que se encuentran dentro de el servicio WFS. En la clase *Catalog* se anotaron las URLs de de la operación *describeFeatureType* de las diferentes capas del servicio WFS. Esta operación brinda un catalogo de los datos de las capas dentro del servicio web. (ver figura 3.8).

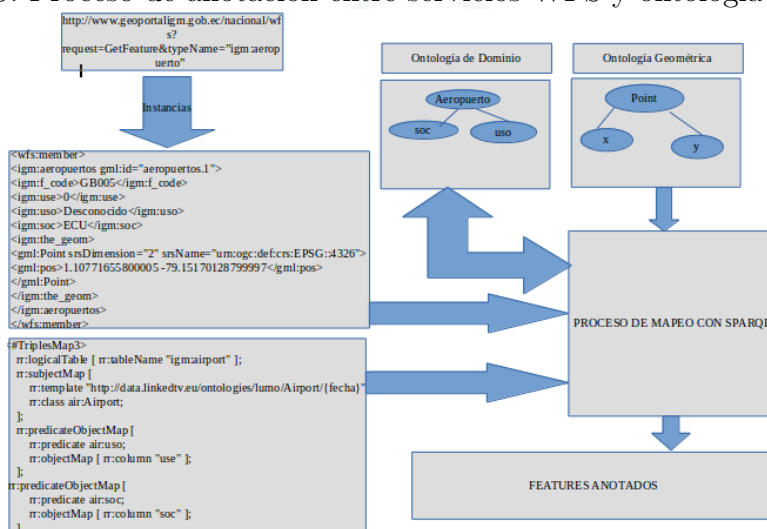
Figura 3.8: Proceso de anotación entre servicios WFS y ontología DCAT



Descripción de la anotación de servicios WFS con respecto a la ontología de dominio

En el proceso de anotación de los servicios se anota la información obtenida dentro de la operación *GetFeature* con respecto a una ontología de dominio ingresada por el usuario. A diferencia de las anteriores el mapeo se realiza de una manera semiautomática, utilizando un archivo R2RML que permite mediante consultas SPARQL ver cual es la clase y el campo dentro de la ontología que pertenece a la capa y campo en el servicio WFS. En cuanto a la anotación de la parte geométrica de los *features* en el servicio WFS se la realiza con un archivo R2RML que mapea las clases de la ontología geométrica con los valores en el atributo *the_geom*. (ver figura 3.9).

Figura 3.9: Proceso de anotación entre servicios WFS y ontología de dominio



3.10. Publicación de Información

El objetivo de la publicación de los servicios WFS es habilitar el acceso a través de la web a la información de los servicios: Direcciones de servicio, relaciones existentes, a que dominios pertenecen, etc. Según Tim Berners Lee los pasos para la correcta publicación de ontologías son:

- Nombrar cosas utilizando URIs
- Utilizar HTTP para construir URIs a través de la web
- Exponer información útil a través de las URIs
- Enriquecer las URIs añadiendo links a otra información a través de la web

Después del proceso de anotación se logra los cuatro puntos expuestos anteriormente, en todos los niveles en los que se atacó lo cual se puede mostrar de mejor manera analizando los resultados obtenidos de dicha anotación. Se empezará con el análisis de los resultados de la anotación de la ontología de WFS.

La información que se anota en la ontología WFS son las clases *DescribeFeatureType* y *GetFeature* que están dentro de la URI <http://purl.org/ifgi/wfs>. Para ver las propiedades que se tiene dentro de la clase *DescribeFeatureType* se utilizara la siguiente consulta.

```
SELECT DISTINCT ?propiedad ?tieneValor WHERE {
  { <http://purl.org/ifgi/wfs#DefaultDescribeFeatureType> ?propiedad
    ?tieneValor }
UNION
  { ?isValueOf ?propiedad
    <http://purl.org/ifgi/wfs#DefaultDescribeFeatureType> }
}
ORDER BY ?propiedad ?tieneValor ?isValueOf
LIMIT 1000
```

Esta consulta generada muestra como resultado las siguientes propiedades. ver figura(3.10)

Figura 3.10: Propiedades de la clase describeFeatureType

propiedad	tiene Valor
http://purl.org/ifgi/pos m#has Input	http://purl.org/ifgi/wfs#DefaultDescribeFeatureType Request
http://purl.org/ifgi/pos m#has Output	http://purl.org/ifgi/wfs#DefaultDescribeFeatureType Response

Si desea ver las instancias de la clase *DescribeFeatureType* se lo puede hacer con la consulta que se muestra a continuación.

```
SELECT DISTINCT ?instance WHERE {
  ?instance a <http://purl.org/ifgi/wfs#DescribeFeatureType>
}
ORDER BY ?instance
LIMIT 1000
```

Esta consulta nos genera un conjunto de salidas que indican todas las instancias encontradas las cuales se listan a continuación. (ver figura 3.11)

Figura 3.11: Listado de instancias de la clase describeFeatureType

instance
http://purl.org/ifgi/wfs#DefaultDescribeFeatureType
http://purl.org/ifgi/wfs#DescribeFeatureTypersg:50m_coastline
http://purl.org/ifgi/wfs#DescribeFeatureTypersg:ais_report
http://purl.org/ifgi/wfs#DescribeFeatureTypersg:amtctd
http://purl.org/ifgi/wfs#DescribeFeatureTypersg:amtctd_days
http://purl.org/ifgi/wfs#DescribeFeatureTypersg:areas
http://purl.org/ifgi/wfs#DescribeFeatureTypersg:areas_new
http://purl.org/ifgi/wfs#DescribeFeatureTypersg:areas_pw
http://purl.org/ifgi/wfs#DescribeFeatureTypersg:discovery_latest
http://purl.org/ifgi/wfs#DescribeFeatureTypersg:full_10m_borders

Las entradas del *DescribeFeatureType* se lo puede sacar con la siguiente consulta:

```
SELECT DISTINCT ?instance ?inobj ?label WHERE {
  ?instance a <http://purl.org/ifgi/wfs#DescribeFeatureType>.
  ?instance <http://purl.org/ifgi/posm#hasInput> ?inobj.
  ?inobj rdfs:label ?label
}
ORDER BY ?instance
```

Las salidas con la siguiente consulta:

```
SELECT DISTINCT ?instance ?inobj ?label WHERE {
  ?instance a <http://purl.org/ifgi/wfs#DescribeFeatureType>.
  ?instance <http://purl.org/ifgi/posm#hasOutput> ?inobj.
  ?inobj rdfs:label ?label
}
ORDER BY ?instance
LIMIT 1000
```

Después de analizar la información anotada de los servicios con la ontología WFS, se procede a realizar el respectivo análisis con la ontología de dominio y la ontología geométrica. Para realizar el análisis de la publicación de la ontología de dominio anotada se utilizara el ejemplo de la capa ya antes mencionada en los puntos previos que es la capa de *topp:tazmania waterbodies* lo cual se puede obtener con la siguiente consulta.

```
SELECT ?a ?url ?area ?perimeter ?continente WHERE {
  ?a <http://webprotege.stanford.edu/tieneUrl> ?url.
  ?a <http://webprotege.stanford.edu/hasArea> ?area.
  ?a <http://webprotege.stanford.edu/hasPerimeter> ?perimeter.
  ?a <http://webprotege.stanford.edu/fromContinent> ?continente
```

```
} ORDER BY ?type
```

con el siguiente resultado. (ver figura 3.12)

Figura 3.12: Visualización de todos los datos de una Feature

a	url	area	perimeter	continente
http://webprotege.stanford.edu/1	https://rsg.pml.ac.uk/geoserver/topp/ows?service=WFS&version=2.0.0&request=GetFeature&typeName=topp:tasmania_water_bodies	1067509088	1073140989	Australia
http://webprotege.stanford.edu/5	https://rsg.pml.ac.uk/geoserver/topp/ows?service=WFS&version=2.0.0&request=GetFeature&typeName=topp:tasmania_water_bodies	1065646817	1071606923	Australia
http://webprotege.stanford.edu/3	https://rsg.pml.ac.uk/geoserver/topp/ows?service=WFS&version=2.0.0&request=GetFeature&typeName=topp:tasmania_water_bodies	1066494066	1071999090	Australia
http://webprotege.stanford.edu/2	https://rsg.pml.ac.uk/geoserver/topp/ows?service=WFS&version=2.0.0&request=GetFeature&typeName=topp:tasmania_water_bodies	1064598241	1071187492	Australia
http://webprotege.stanford.edu/0	https://rsg.pml.ac.uk/geoserver/topp/ows?service=WFS&version=2.0.0&request=GetFeature&typeName=topp:tasmania_water_bodies	1064866676	1071221047	Australia
http://webprotege.stanford.edu/6	https://rsg.pml.ac.uk/geoserver/topp/ows?service=WFS&version=2.0.0&request=GetFeature&typeName=topp:tasmania_water_bodies	1067743969	1073212293	Australia
http://webprotege.stanford.edu/4	https://rsg.pml.ac.uk/geoserver/topp/ows?service=WFS&version=2.0.0&request=GetFeature&typeName=topp:tasmania_water_bodies	1065512599	1071304933	Australia

Como se puede ver en la figura 3.12 al publicar los servicios WFS anotados se pueden mostrar los datos del servicio de una entendible, a diferencia de la forma en la que se visualiza la información del servicio en el formato GML esta manera de visualizar los datos del servicio permite ver de manera clara las relaciones que existen entre servicio, capas e instancias.

En definitiva el proceso de anotación hizo posible pasar de información en un formato sintáctico y aislado a un conjunto de información relacionada y ordenada de mayor utilidad para el usuario, para que este la pueda explotar mejor. La información dentro del servicio no tenía ningún sentido semántico mientras que ahora se puede obtener información con preguntas semánticas.

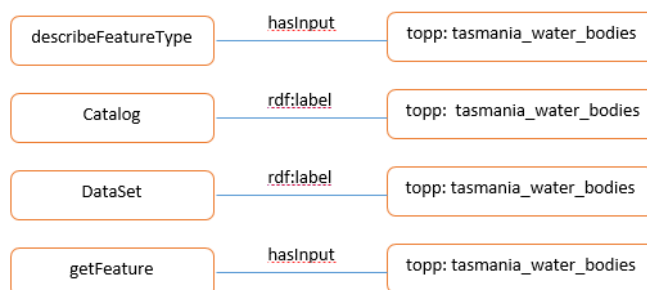
Previamente para saber como se puede obtener información geográfica acerca de lagos en Australia se debía realizar una búsqueda en la web de los diferentes servicios de una manera empírica con resultados inciertos y poco estructurados e imposibles de relacionar con otras fuentes de información de manera automática.

En definitiva se ha realizado un proceso de análisis de información proporcionada por servicios WFS que era completamente dependiente del conocimiento que tenía el usuario de las fuentes de información y los temas a los que estas hacen referencia a un proceso en el que un usuario sin mayor conocimiento podría buscar y acceder a los datos de manera mucho más sencilla.

3.11. Enlace de datos

Una vez anotados los servicios en base a la ontología de dominio, meta datos, geometrías y su publicación, se procede al enlace de datos, este proceso es un método para encontrar información en el repositorio *Triple Store* que tenga relación sea en un tema en específico, o ubicaciones cercanas o coincidentes y que sean de diferentes fuentes de información, con el afán de validar datos o de obtener más información que tengan relaciones en diferentes fuentes. Por ejemplo una vez cargado en el SPARQL ENDPOINT la anotación con respecto a la ontología DCAT se puede obtener un listado de los catálogos y grupos de datos contenidos en el servicio web mientras que en la clase WFS se anotan las URLs de las operaciones *getFeature* y *DescribeFeatureType*, los catálogos de un servicio web están contenidos en la operación *DescribeFeatureType*.(ver figura 3.13.)

Figura 3.13: Relaciones entre la información anotada en dos fuentes distintas

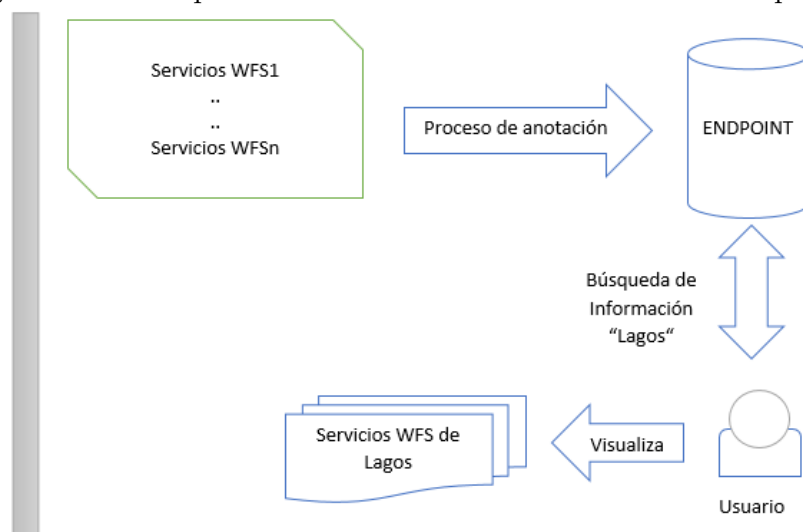


Luego se analiza las relaciones que se obtiene con la anotación del servicio con respecto a la ontología de dominio. Este tipo de anotaciones es la que brinda más información al usuario. Por ejemplo, al anotar una capa de lagos que tiene como uno de sus campos área se podría averiguar todos los lagos que tengan una superficie dentro de un rango determinado y a que servicios pertenecen estos lagos. Suponiendo que se tiene dos servicios cada uno con una capa que hace referencia a lagos y que tengan el campo área y estos se anotan con respecto a una ontología de dominio al mismo tiempo que contra la ontología WFS y DCAT. Se podrían obtener las direcciones de todos los servicios WFS que contengan lagos de cierta área.

A continuación el esquema de búsqueda de información en la figura 3.14

Hasta el momento se ha analizado las oportunidades que brinda la anotación de los servicios WFS por si solos, es decir, sin crear relaciones con otros *sets* de datos. Para analizar en un contexto más amplio las oportunidades que viene con

Figura 3.14: Búsqueda de Información en base a criterios específicos



la anotación semántica de servicios WFS se debe generar relaciones con otros *end-points* como por ejemplo DBPedia. Por lo tanto, la creación de relaciones entre dos grafos representados mediante el formato RDF se puede lograr con la herramienta SILK. Para entender cómo funciona esta herramienta se necesita analizar los siguientes conceptos: fuentes de datos, reglas de relacionamiento, salida y finalmente enlaces de referencia.

Existen dos tipos de *sets* de datos dentro de la herramienta, la primera es un SPARQL ENDPOINT que permite acceder a diferentes ENDPOINTS dentro de la web, tal como dbpedia o geonames, geonames es especialmente útil para el análisis de datos geográficos anotados semánticamente mientras que los archivos RDF pueden ser cargados a la aplicación para realizar la operación de enlace de tripletas. Las reglas de relacionamiento especifica como dos ítems son comparados por similitud y consiste de cuatro componentes: La ruta de entrada que devuelve los valores de una entidad dada para una entidad RDF, una transformación que se aplica a los valores de la entrada RDF, una comparación que evalúa la similitud de dos entradas y finalmente funciones de agregación que devuelven valores de confianza en los datos obtenidos.

Los directorios de entrada pueden servir para hacer referencia a subconjuntos de datos de un RDF el cual puede ser navegado con los operadores que se muestran a continuación. (ver figura 3.15)

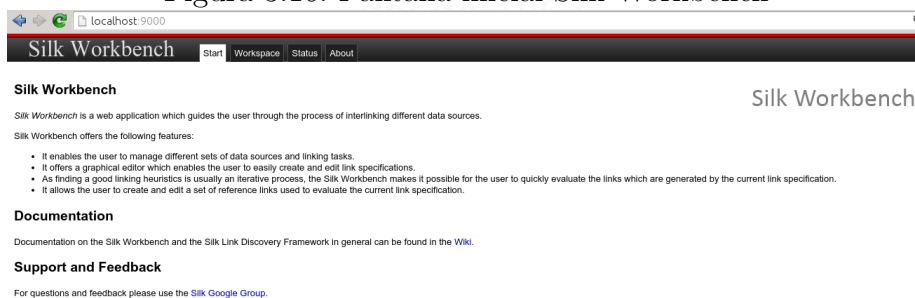
Figura 3.15: Operadores silk

Operator	Name	Use	Description
/	Operador adelante	<code><path_segment>/<property></code>	Se mueve hacia adelante desde un recurso sujeto a través de una propiedad hacia un objeto.
\	Operador reverso	<code><path_segment>*<property></code>	Se mueve inversamente al anterior
[]	Operador filtro	<code><path_segment> [<property> <comp_operator> <value>] <path_segment> [!lang <comp_operator> <value>]</code>	Reduce los recursos de la selección.

Una manera de comprobar la información generada es mediante consultas SPARQL y GEOSPARQL o a su vez ver la relación que tienen dichos archivos RDF mediante la plataforma SILK Workbench.

La instalación completa de la plataforma SILK se puede ver en los anexos (Apéndice A) el cual se encuentra como archivo anexo al presente trabajo. La siguiente URL `http://localhost:9000` se habilitará una vez instalada e iniciada los scripts de silk-workbench tal como muestra en la figura 3.16.

Figura 3.16: Pantalla inicial Silk Workbench



Dentro del cual se creará un proyecto en el que se va a subir los archivos para el proceso. ver figura 3.17.

Figura 3.17: Creación de proyecto Silk Workbench



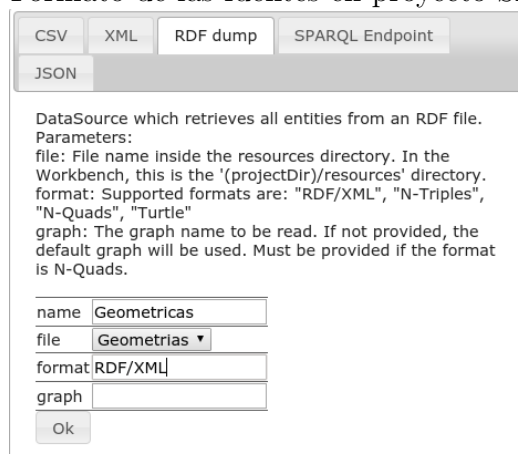
Luego de crear el proyecto se procede a cargar los archivos al sistema, tanto los archivos anotados con las ontologías de metadatos, geométricos y *geatures* detallados en la figura 3.18.

Figura 3.18: Agregar Recursos en proyecto Silk Workbench



A esto se agrega el formato de lectura de las fuentes de información para que el sistema reconozca y genere el proceso necesario, esto esta expresado en la figura 3.19.

Figura 3.19: Formato de las fuentes en proyecto Silk Workbench



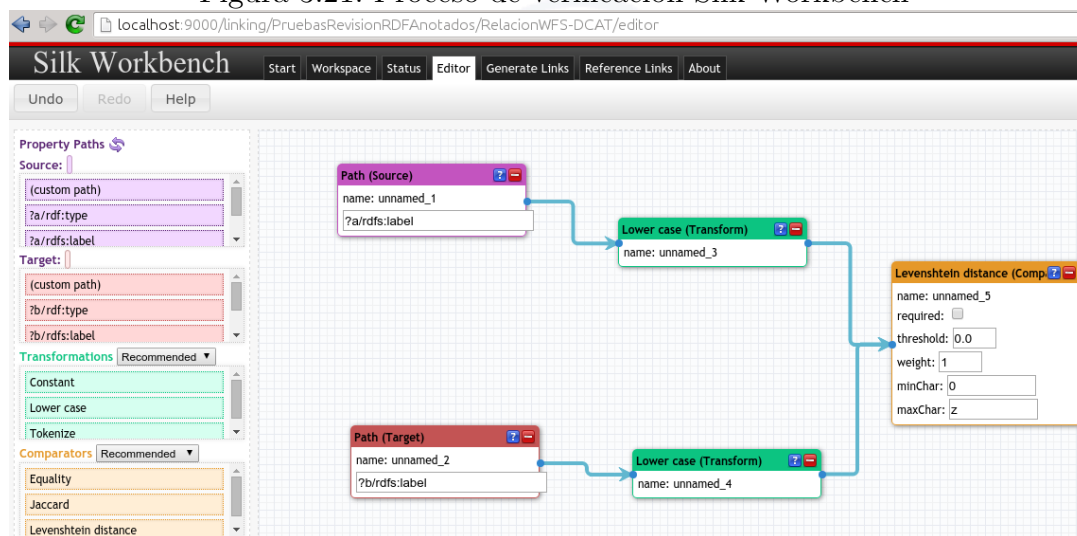
Posteriormente se procede a realizar el proceso de enlace dentro de la plataforma, para ello se indica los archivos RDF y las relaciones que se encuentran entre ellos para que la plataforma inicie el proceso de verificación, además, es necesario configurar ciertos criterios de los cuales ayudarán a la aplicación en su búsqueda como se indica en la figura 3.20.

Figura 3.20: Configuración de datos para la verificación Silk Workbench

Linking Task	
Name	RelacionWFS-DCAT
Source	WFS
Source restriction	?a rdf:type ?obj1
Target	DCAT
Target restriction	?b rdf:type ?obj2
Output	
Ok	

dejando así el proceso de generación de enlaces completo. (ver figura 3.21)

Figura 3.21: Proceso de verificación Silk Workbench



Una vez definido el proceso de manera gráfica se procede a la generación de los enlaces entre la unión de los dos archivos RDF y se detallan los resultados de este proceso en la figura 3.22.

Figura 3.22: Resultado del proceso de verificación Silk Workbench

Source	Target	Score	Correct
http://purl.org/fgi/wfs#GetFeaturetiger_poi	dcatalogtiger_poi	100,0%	✓ ? ✕
http://purl.org/fgi/wfs#GetFeaturetiger_poi	dcatalogtiger_poi	100,0%	✓ ? ✕
http://purl.org/fgi/wfs#DescribeFeatureTypetiger_poly_landmarks	dcatalogtiger_poly_landmarks	100,0%	✓ ? ✕
http://purl.org/fgi/wfs#DescribeFeatureTypetiger_poly_landmarks	dcatalogtiger_poly_landmarks	100,0%	✓ ? ✕
http://purl.org/fgi/wfs#DescribeFeatureTypetiger_poi	dcatalogtiger_poi	100,0%	✓ ? ✕
http://purl.org/fgi/wfs#DescribeFeatureTypetiger_poi	dcatalogtiger_poi	100,0%	✓ ? ✕
http://purl.org/fgi/wfs#DescribeFeatureTypetiger_roads	dcatalogtiger_roads	100,0%	✓ ? ✕
http://purl.org/fgi/wfs#DescribeFeatureTypetiger_roads	dcatalogtiger_roads	100,0%	✓ ? ✕

Con este proceso se puede encontrar relaciones entre dos o más fuentes diferentes de datos y obtener un archivo único el cual contiene las relaciones existentes entre ellos, además, si se agrega a *Triple Store* los resultados obtenidos en este proceso y las fuentes utilizadas se puede realizar consultas tanto en las relaciones y en las fuentes a la vez logrando los resultados esperados y necesarios dentro de

nuestro repositorio de datos RDF.

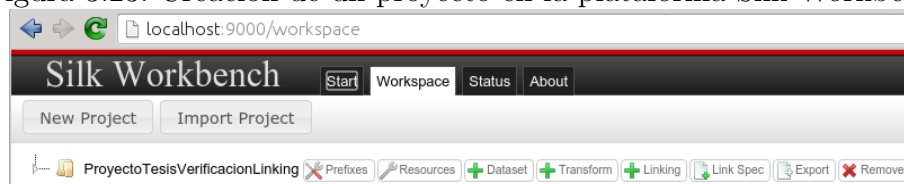
A continuación se muestra un ejemplo completo que se realizó en el presente trabajo. El cual se detalla de la siguiente manera:

- Creación de un proyecto en la plataforma *silk workbench*.
- Carga de fuentes al proyecto.
- Asignación de formato de entrada y salida a las fuentes del proyecto.
- Relacionar las fuentes para su análisis.
- Generar proceso para su análisis respectivo.
- Generación de enlaces entre las fuentes.
- Verificación de resultados en forma gráfica.

3.11.1. Creación de un proyecto en la plataforma Silk Workbench

Dentro de la plataforma *silk workbench* se procede a la creación de un proyecto llamado *ProyectoTesisVerificacionLinking* como se presenta en la figura 3.23.

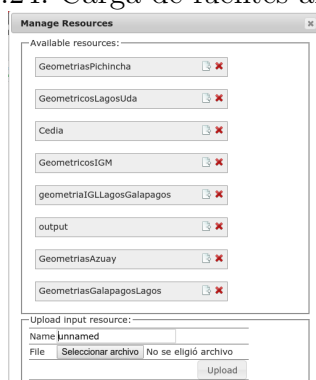
Figura 3.23: Creación de un proyecto en la plataforma Silk Workbench



3.11.2. Carga de fuentes al proyecto.

Con el proyecto creado en la plataforma se procede a la carga de las fuentes para el proceso correspondiente, para ello se da click en el botón *Resources* y aparecerá la siguiente pantalla, en las cuales se procede a cargar las fuentes desde el botón seleccionar archivo, se ingresa un nombre válido y se cargara automáticamente a la parte superior como se muestra en la figura 3.24.

Figura 3.24: Carga de fuentes al proyecto.

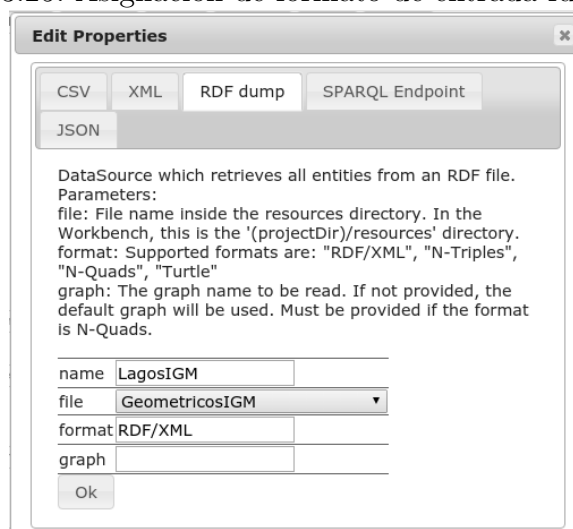


En el ejemplo se usara las fuentes de datos *LagosUDA* y *LagosIGM* servicios que fueron encontrados en el servidor de mapas de la Universidad del Azuay y en el servidor de mapas del Instituto Geográfico Militar.

3.11.3. Asignación de formato de entrada y salida a las fuentes del proyecto.

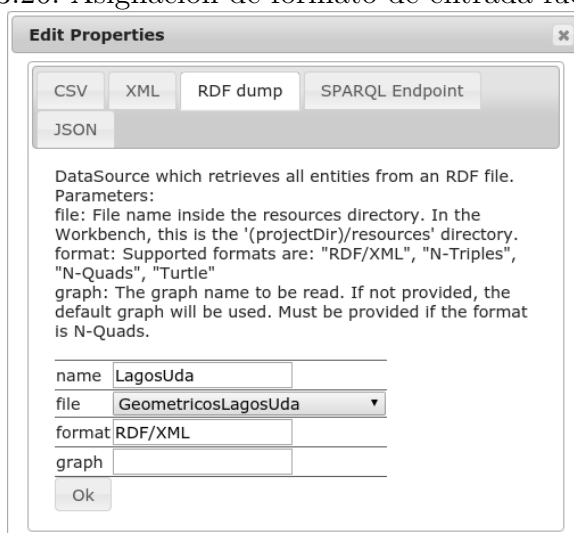
Para realizar el respectivo proceso de enlace de datos de la fuente es necesario indicarle el formato que dichas fuentes poseen, por tal motivo dentro del proyecto creado en el botón *Dataset* el cual presentará la figura 3.25 con los datos requeridos, se ingresa la información de las dos fuentes. Primero se asignará el formato a la capa de IGM.

Figura 3.25: Asignación de formato de entrada fuente IGM



A continuación se agregara el formato a la capa de la UDA (fig. 3.26).

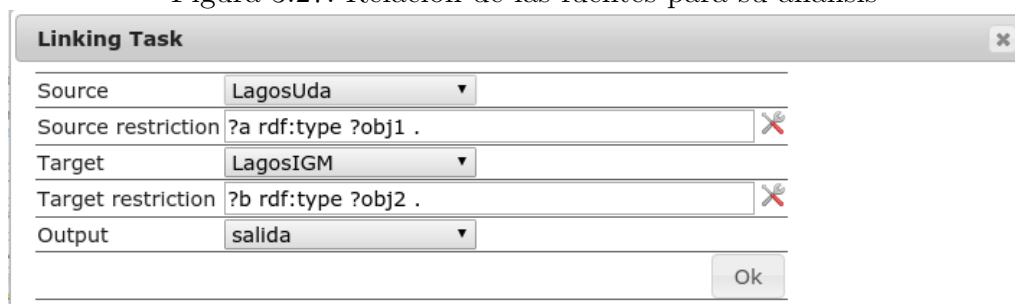
Figura 3.26: Asignación de formato de entrada fuente UDA



3.11.4. Relación de las fuentes para su análisis

Una vez cargadas las fuentes y especificadas el formato en el que estas se encuentran es necesario brindar las relaciones necesarias para que la plataforma proceda a realizar la búsqueda, para ello, en el proyecto presionamos el botón *Linking* el cual presentará la siguiente pantalla. (ver figura A.1).

Figura 3.27: Relación de las fuentes para su análisis

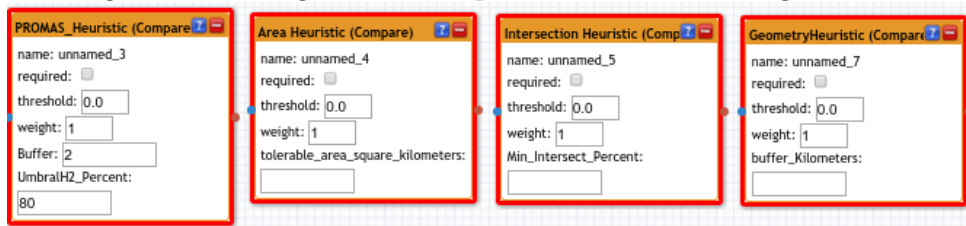


3.11.5. Plugins para uso de la generación de enlaces entre las fuentes

Para el proceso de linking entre las fuentes seleccionadas se a utilizado un plugin construido propiamente para el análisis de geometrías por la Universidad de Cuenca, Proyecto que actualmente dirige el Ing Víctor Saquicela docente universitario.

El plugin consta de varios análisis los cuales son por Áreas, Geometrías, Intersecciones y todas estas opciones juntas en *PROMAS Heuristic*, ver figura 3.28.

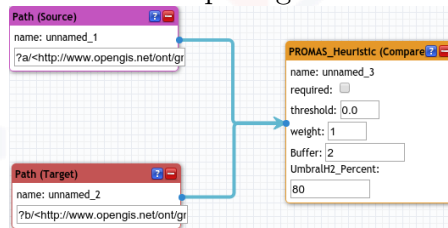
Figura 3.28: Plugin utilizado para el análisis de las geometrías



Cada *plugin* necesita de parámetros para sus respectivos análisis, en este proyecto se utiliza el *plugin PROMAS Heuristic* el cual solicita un porcentaje de aceptación de coincidencia y un *buffer* que servirán para el correcto análisis.

A continuación se procede a modelar el proceso que ayudará en la selección de enlaces, este proceso se detalla en la imagen 3.29.

Figura 3.29: Proceso para generación de linking



3.11.6. Generación de enlaces entre las fuentes

Luego de que el proceso se encuentra listo para su funcionamiento se accede a la opción *Generate Links* la cual se encuentra dentro del editor, este llevará a la pantalla en la que muestra las relaciones existentes entre las dos fuentes (ver figura 3.30).

Figura 3.30: Generación de Linking

The screenshot shows the Silk Workbench interface with the 'Generate Links' window open. The window title is 'Generate Links finished in 4s'. Below the title bar, there are navigation buttons: 'Expand All', 'Collapse All', 'Prev', '1', and 'Next'. A 'Filter:' input field is also present. The main area displays a table of generated links between two sources: 'LagosUda' and 'LagosiGM'.

Source	Target	Score	Correct
▶ http://www.opengis.net/ont/gmi#63	http://www.opengis.net/ont/gmi#31	100,0%	<input checked="" type="checkbox"/> ? ✕
▶ http://www.opengis.net/ont/gmi#158	http://www.opengis.net/ont/gmi#23	100,0%	<input checked="" type="checkbox"/> ? ✕
▶ http://www.opengis.net/ont/gmi#62	http://www.opengis.net/ont/gmi#31	100,0%	<input checked="" type="checkbox"/> ? ✕

At the bottom of the table, there are 'Prev', '1', and 'Next' navigation buttons.

y si se procede a verificar las geometrías se encuentra la siguiente información (figura 3.31).

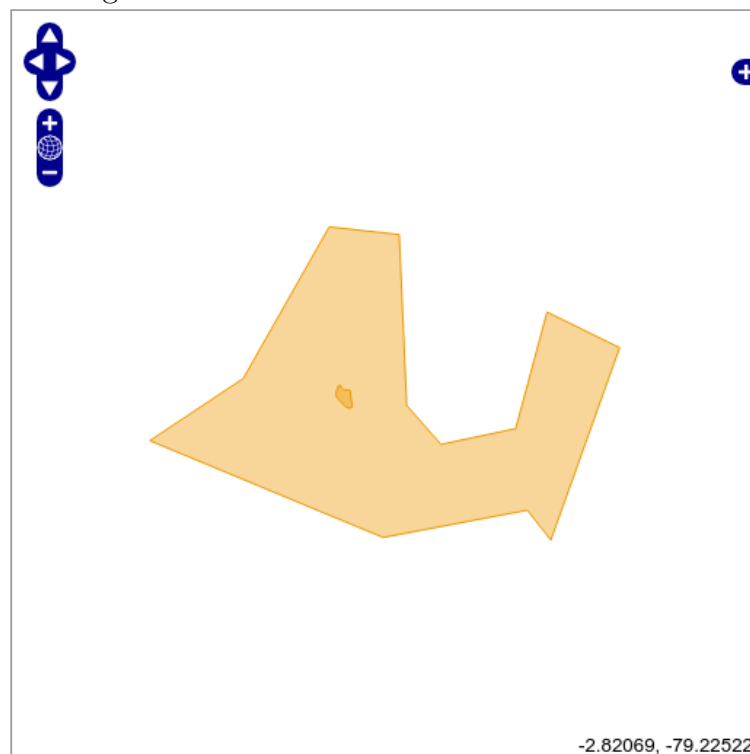
Figura 3.31: Resultados encontrados

Source: LagosUda	Target: LagosIGM	Score	Correct
http://www.opengis.net/ont/gmi#63	http://www.opengis.net/ont/gmi#31	100,0%	<input checked="" type="checkbox"/> <input checked="" type="checkbox"/>
Comparison: PromasHeuristic (unnamed_3) <input type="text" value="100,0%"/>			
Input: ?a<http://www.opengis.net/ont/gmi#asWKT> (unnamed_1) MULTIPOLYGON (((-2.831305346972184 -79.22415717193314, -2.8313054705971554 -79.22423755013541, -2.8312874619816824 -79.22428760324331, -2.831052759467871 -79.22455438930108, -2.8309353061893376 -79.22462145622863, -2.83085390735394 -79.22461258662662, -2.83082673667354 -79.22458508570406, -2.8308266321063136 -79.22451707339384, -2.8308714778139974 -79.22427755634641, -2.830871230601944 -79.22411680000165, -2.830898320031317 -79.22409146490637, -2.831088188936776 -79.22407206513587, -2.831142366920199 -79.22402083284229, -2.831169378530053 -79.22394491006862, -2.8312236386406577 -79.22394707586132, -2.831305346972184 -79.22415717193314)))			
Input: ?b<http://www.opengis.net/ont/gmi#asWKT> (unnamed_2) MULTIPOLYGON (((-2.8340376107750602 -79.22374381239358, -2.836794107197901 -79.2255788301297, -2.833570736021583 -79.22691654018324, -2.8299158452109276 -79.22843331915014, -2.825669795728571 -79.22762703755315, -2.8249727807834266 -79.22850924915515, -2.822945312649911 -79.22283797416083, -2.8250860836933236 -79.22178728637398, -2.826009821967541 -79.22522308467305, -2.828208263659731 -79.22568730516515, -2.8292208167008535 -79.2245397593336, -2.8294414821934084 -79.21950240949063, -2.8315061086726954 -79.21927568960734, -2.8340376107750602 -79.22374381239358)))			

3.11.7. Verificación de resultados en forma gráfica

Una vez encontradas las relaciones entre las dos fuentes, se puede revisar que la información que se encuentra en ellas esta correcta de manera gráfica en cualquier visor de datos WKT el cual se presenta en la imagen 3.32.

Figura 3.32: Resultados encontrados validados



para este ejemplo se uso la dirección <http://dev.openlayers.org/sandbox/docs/examples/wkt.html>.

Capítulo 4

Evaluación y resultados

En este capítulo se describe los resultados del proceso de anotación a varios servicios WFS, con el fin de descubrir la confiabilidad de este método para la anotación y clasificación de servicios. A continuación se detalla los pasos que se siguieron para el proceso de evaluación.

4.1. Método

El análisis de los resultados obtenidos del proceso de anotación de los servicios WFS se divide en 2 fases. El registro del servicio WFS y la evaluación de la anotación semántica. Dentro del proceso de registro de los servicios WFS se escogió un conjunto de 6 servicios de los cuales la totalidad fueron registrados en el sistema, como dato importante hay que tener en cuenta que para el análisis de la aplicación se tomará en cuenta capas y no servicios, pues el proceso de mapeo utilizando R2RML se enfoca en eso y no distingue entre diferentes servicios pues ese nivel de anotación se lo realiza de manera semiautomática. Los pasos que se definieron para la evaluación de los resultados son los siguientes: Calificación de servicios WFS que ha sido ya explicado previamente el proceso que se utilizó para clasificar los servicios WFS, pero en breves palabras se realizó lo siguiente, analizar la fuente, analizar los nombres de las capas y finalmente analizar los campos de las capas.

Las fuentes utilizadas en el proceso de anotación fueron extraídas de las siguientes entidades:

- Instituto Geográfico Militar.
- Universidad de Cuenca.
- Universidad del Azuay.
- Remote Sensing Group.

Para análisis de las capas se ha visto la necesidad de distinguir ciertos parámetros en objetivos y subjetivos, objetivos porque se da por hecho que la información

que indica está de acuerdo a la capa y no existe error, mientras que, subjetivos solo se puede indagar sobre la información, ya que no hay certeza de que el nombre del campo sea en verdad lo que significa, además, existen campos que no poseen información fácil de entender es decir la descripción o nombre del campo no indica nada. Por lo tanto, esta información debe ser verificada de forma manual debido a que no existen herramientas para realizar un proceso automático. Posteriormente se procede a analizar la metadata de las capas y las instancias de la misma, lamentablemente este análisis debe ser subjetivo forzosamente.

El siguiente paso es la evaluación de las ontologías obtenidas con el proceso de búsqueda manual, el cual se explicó en el capítulo 3. Esto incluye métricas para ver el nivel de validez de una ontología, se calcula viendo el nivel de capas que pueden ser anotadas con esta ontología y el nivel de similaridad que hay entre los campos de la capa y la propiedad de la clase. Finalmente se evalúa el proceso de anotación de los servicios WFS tomando en cuenta las métricas que se detallarán más adelante en el documento. A continuación se detalla un esquema de verificación de información en la figura (4.1) a continuación.

Figura 4.1: Esquema de verificación de información.



4.2. Proceso de evaluación y resultados

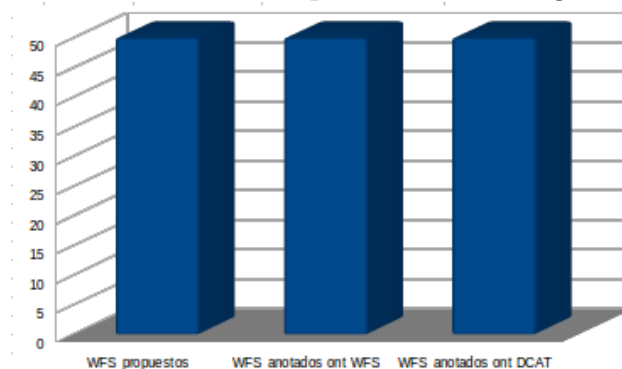
El proceso de evaluación y resultados se lo realizo mediante 3 puntos:

- Análisis y anotación de la operación `getcapabilities` del servicio WFS.
- Análisis y anotación de la operación `DescribeFeatureType` del servicio WFS.
- Análisis y anotación de la operación `getFeature` del servicio WFS.

El primer punto de análisis se realizó en la operación `getCapabilities` del servicio WFS el cual es la operación encargada de mostrar la metadata del servicio. Además, las ontologías usadas para este proceso de anotación son específicas para metadatos absolutamente todos los servicios WFS han sido anotación sin problema alguno.

En la figura 4.2 se detalla el nivel de anotación de la operación `getCapabilities`.

Figura 4.2: WFS anotados con respecto a las ontologías WFS y DCAT



El mapeo que se realizó es manual, por ello, la información anotada con respecto a las ontologías es correcta, es decir, los resultados son los esperados en el proceso de anotación y validación de datos. Además se desarrolló una aplicación que es capaz de leer los servicios WFS y anotarlos mediante lecturas del archivo mapeado con de forma manual explicado en la sección 3.8 de este documento, este procedimiento genera archivos en formato RDF quienes son ingresados en un repositorio de datos *geosparql* para su almacenamiento, en este punto una persona con conocimientos en consultas para este tipo de repositorios podrá ingresar y filtrar información referente a un dominio específico, además, se realizó el proceso de enlace de datos el cual ayuda a encontrar más información referente al dominio que el usuario busca en los repositorios. El problema más grande que se ha encontrado, es que los servicios WFS no poseen información detallada de tal manera que sea capaz de adaptarse fácilmente a ontologías de dominios específicos creados.

4.3. Registro de capas

El sistema invoco a 6 servicios WFS, dentro de los cuales existían 700 capas de las cuales se accedieron a 10 capas, dentro de estas 10 capas la aplicación tuvo un porcentaje del 100% (ver figura 4.3) de éxito al leer estas capas. Durante el proceso de desarrollo de la aplicación se a probado ciertas capas en las cuales se encontraron errores de invocación, la mayoría de los errores en la lectura se suscitaron por las siguientes razones:

- El tamaño de la capa es demasiado grande
- La conexión del servicio es intermitente
- El servicio no tiene la version 2.0.0 de WFS

Figura 4.3: Servicios WFS y capas usadas.

SERVICIOS WFS	CAPAS	¿CONECTO?
http://ide.ucuenca.edu.ec/geoserver/ows?service=wfs&version=2.0.0&request=GetCapabilities	ecu_esca_nacional:BH080_LAGO_LAGUNA	SI
http://ide.ucuenca.edu.ec/geoserver/ows?service=wfs&version=2.0.0&request=GetCapabilities	fonapabase:BH080_LAGUNAS	SI
http://www.geoportaligm.gob.ec/nacional/wfs?request=GetCapabilities	igm:lago_laguna	SI
http://gis.uazuay.edu.ec:8080/geoserver/ows?service=wfs&version=2.0.0&request=GetCapabilities	azuay_50k:Azuay_lagunas	SI
http://gis.uazuay.edu.ec:8080/geoserver/ows?service=wfs&version=2.0.0&request=GetCapabilities	ic_base:lagunas	SI
http://www.geoportaligm.gob.ec/galapagos25/wfs?service=wfs&request=GetCapabilities	igm:lago_laguna	SI
http://ide.ucuenca.edu.ec/geoserver/ows?service=wfs&version=2.0.0&request=GetCapabilities	PRUEBA:Inu2014	SI
http://ontorisk.cedia.org.ec/geoserver/ows?service=wfs&version=2.0.0&request=GetCapabilities	vulnerabilidad:INUNDACION	SI
http://ontorisk.cedia.org.ec/geoserver/ows?service=wfs&version=2.0.0&request=GetCapabilities	wwf_galapagos:GE000_UNIDAD_AMBIENTAL_LAGUNAS_COSTERAS	SI
http://ontorisk.cedia.org.ec/geoserver/ows?service=wfs&version=2.0.0&request=GetCapabilities	wwf_galapagos:GE000_UNIDAD_AMBIENTAL_LAGUNAS_INTERIOR	SI

4.4. Anotación semántica

Este análisis sigue la anotación de los diferentes niveles mediante los cuales se ataca a los servicios WFS. En el primer nivel de los servicios se logro una anotación del 60 % (ver figura 4.4).

Figura 4.4: Capas mapeadas

CAPAS	¿MAPPEO?	OBSERVACIONES
ecu_esca_nacional:BH080_LAGO_LAGUNA	NO	f_code no tiene sentido para un usuario común
fonapabase:BH080_LAGUNAS	NO	No tiene campos descriptivos
igm:lago_laguna	SI	
azuay_50k:Azuay_lagunas	SI	
ic_base:lagunas	SI	
igm:lago_laguna	SI	
PRUEBA:Inu2014	SI	
vulnerabilidad:INUNDACION	NO	No tiene campos descriptivos
wwf_galapagos:GE000_UNIDAD_AMBIENTAL_LAGUNAS_COSTERAS	SI	
wwf_galapagos:GE000_UNIDAD_AMBIENTAL_LAGUNAS_INTERIOR	NO	No se acoplo a ninguna ontología de las encontradas



Se puede interpretar que el proceso de anotación de las capas registradas a tenido un gran valor de aceptación, pero este valor depende de la información con la que es publicada cada capa en el servidor de mapas *geoserver*.

4.5. Evaluación de datos de entrada

Al hablar de la validación de parámetros de entrada el sistema no los evalúa pues el servicio WFS no los tiene, es por esto que el sistema solo toma la URL de los servicios y de ahí verifica la información de sus campos (metadata), estos pueden clasificarse en campos válidos y descriptivos, campos geoespaciales y campos desconocidos. Los campos válidos y descriptivos son aquellos que indica con certeza lo que significa dicho campo, los campos geoespaciales muestran la información

geométrica del *feature* que se está analizando, y los campos desconocidos son los que no tienen suficiente información descriptiva para poder asociar la información con algún tema o valor específico. Los campos de entrada son en su gran mayoría desconocidos por lo menos en los servicios locales por lo cual no sirven de mucho al momento de realizar la anotación semántica, estos campos muchas veces aportan más información por su contenido que por su nomenclatura, de esta manera se obtuvo la información necesaria para el mapeo y posteriormente la anotación. Por lo tanto, para la evaluación de estos campos primero se definió el número de campos desconocidos versus el número de campos conocidos. Los campos conocidos fueron mapeados manualmente y anotados semánticamente con un porcentaje de éxito del 100 % por lo cual no serán analizados, lo que se analizará es el porcentaje de campos desconocido versus campos conocidos dentro de los WFS. La información de campos desconocidos frente a campos desconocidos se encuentra en un 60 % del total de todos los campos de las capas analizadas, mientras que el 40 % de los campos poseen información válida o entendible. (ver figura 4.5)

Figura 4.5: Relación campos no válidos.

CAPAS	Campos	Campos Inentendibles
ecu_esca_nacional:BH080_LAGO_LAGUNA	4	2
fonapabase:BH080_LAGUNAS	5	3
igm:lago_laguna	8	5
azuay_50k:Azuay_lagunas	3	1
ic_base:lagunas	9	4
igm:lago_laguna	8	5
PRUEBA:Inu2014	52	39
vulnerabilidad:INUNDACION	6	2
wwf_galapagos:GE000_UNIDAD_AMBIENTAL_LAGUNAS_COSTERAS	6	1
wwf_galapagos:GE000_UNIDAD_AMBIENTAL_LAGUNAS_INTERIOR	4	1
Total	105	63
Porcentaje de campos	100%	60%

4.6. Evaluación de datos de salida

La validación de los datos de salida se la realiza mediante métricas de precisión y respuesta, la precisión es calculada en base al número de datos anotados correctamente dividido entre el número de datos anotados, debido a la falta de procesos que ayuda a verificar si un dato fue anotado correctamente haciendo referencia al valor que tiene, se puede concluir que estos datos dependen netamente de la persona que realiza el mapeo antes de la anotación, si la persona relaciona bien la información del servicio con la ontología la anotación tendrá éxito, caso contrario, la información saldrá errónea, es decir, se podrá catalogar como buena o mala la anotación dependiendo del conocimiento que tenga la persona sobre la información de las capas que se encuentren en el servicio WFS. Dado que el mapeo fue realizado solo por una persona en este proceso de pruebas los resultados no son comparables. La conclusión que se puede sacar de esto es que se necesita diseñar un método largo para la validación de los datos de salida sea este realizado por un experto o de manera automática.

4.7. Repositorio WEB

Todos los archivos RDF generados mediante el proceso de anotación, sea en *getCapabilities*, *DescribeFeatureType* o *getFeature* son almacenados en un *Triple Store* al cual se podrán conectar para realizar consultas SPARQL y GEOSPARQL, para poder realizar este tipo de consultas sobre la información generada se ha instalado un servidor de apache “MARMOTTA“ el cual permite realizar este tipo de operaciones, para acceder al servidor de manera local ingresa a <http://localhost:8080> el cual mostrará una página con el menú que muestra la siguiente imagen 4.6:

Figura 4.6: Opciones Triple Store marmotta



En el menú existe una opción SPARQL en el cual se puede realizar las consultas deseadas para obtener la información necesaria. A continuación se presenta unos ejemplos de query generado durante este proceso.

Ejemplo 1. Query para obtener la información del servicio y sus capas.

```
SELECT DISTINCT ?property ?hasValue ?isValueOf WHERE {
  {
    <http://purl.org/ifgi/wfs#WebFeatureServicehttp://ide.ucuenca.edu.ec
/geoserver/> ?property ?hasValue }
UNION
  { ?isValueOf ?property
    <http://purl.org/ifgi/wfs#WebFeatureServicehttp://ide.ucuenca.edu.ec
/geoserver/> }
```

```
}  
ORDER BY ?property ?hasValue ?isValueOf  
LIMIT 1000
```

Ejemplo 2. Query para verificar todos los servicios que han sido anotados.

```
SELECT DISTINCT ?instance WHERE {  
  ?instance a <http://purl.org/ifgi/wfs#WebFeatureService>  
}  
ORDER BY ?instance  
LIMIT 1000
```

Ejemplo 3. Query para visualizar los campos de una capa dentro del servicio.

```
SELECT DISTINCT ?property ?hasValue ?isValueOf WHERE {  
  { <http://webprotege.stanford.edu/0> ?property ?hasValue }  
UNION  
  { ?isValueOf ?property <http://webprotege.stanford.edu/0> }  
}  
ORDER BY ?property ?hasValue ?isValueOf  
LIMIT 1000
```

4.8. Recomendaciones

En cuanto a la evaluación del proceso de enlace de datos basándose en los resultados obtenidos se puede llegar a la conclusión de que se necesita un procedimiento semiautomático para realizar un mapeo de las capas. Este proceso debería tener un conjunto de ontologías ingresadas en el sistema al igual que un diccionario de palabras que puedan determinar de manera cuantificable el número de campos desconocidos y ver la correspondencia que existe entre estos para que el sistema sugiera las posibles correspondencias que existan entre los servicios y las ontologías. Las ventajas de esta mejora en el proceso sería en el ahorro de tiempo pues la búsqueda de ontologías y mapeo de las mismas es consumidor de tiempo, en este primer intento para definir un proceso de generar LOD con servicios WFS se ha podido constatar que la anotación del primer nivel de los servicios según la metodología de MAUE se realiza de una manera correcta en un 100% de los casos estudiados a diferencia del último nivel por la heterogeneidad de parámetros y lo empírico del procedimiento.

Capítulo 5

Conclusiones

En el presente proyecto se a desarrollado un proceso en el cual se puede anotar los servicios WFS (Web Feature service) a tres niveles, usando ontologías para catalogar sus datos, debido a que la información de dichos servicios se encuentran disponibles en la red pero de una manera organizada también se a creado un repositorio de datos (Triple Store) para almacenar la información de los servicios WFS anotados, en el cual se almacena la información tanto de los servicios como la información que se a obtenido mediante un proceso de enlace de datos a los mismos, esto indicará cual de los servicios se relacionan o pertenecen a la misma información proporcionada por fuentes distintas.

Al dar por terminado el proyecto se a precisado verificar si los objetivos planteados para el mismo han sido verificados con éxito y en que porcentaje se a podido determinar, para ellos se detallará todos los objetivos planteados y describir con mas detalle el porque si/no se a cumplido, cual fue la clave de éxito o no.

- Usar ontología en base a un tema puntual de dominio con la cual se anotara la estructura de datos de las capas del servicio WFS. Este Objetivo a sido uno de los mayores retos ya que si bien existen ontologías para cualquier dominio de estudio la información de los servicios WFS no poseen la información con la cual relacionar directamente sus campos a las clases o atributos de la ontología, por lo que se a tomado una ontología base y de ahí cada atributo del servicio que no este directamente relacionado a la ontología a sido creado como una instancia en la misma, de esta manera se a podido almacenar toda la información del servicio dentro de la ontología. Sin embargo a pesar de los inconvenientes encontrados se a detallado un proceso en el cual se indica como proceder para la selección de la ontología, por lo que se da como cumplido el objetivo.
- Anotar metadatos de servicio WFS con ontología específica para catalogar servicios web. Se a encontrado varias ontologías que han sido diseñadas para almacenamiento de metadatos de servicios, estos pueden o no ser propios para servicios WFS, se a realizado el proceso de anotación para dos ontologías seleccionadas.
 - Ontología WFS, Esta diseñada propiamente para la anotación de ser-

vicios WFS, haciendo relación directamente a sus atributos y propiedades.

- Ontología DCAT, Esta diseñada para cualquier tipo de servicio web.

dando por entendido que se a cumplido el objetivo planteado.

- Anotar la información geométrica de los servicios WFS con la ontología geosparql. Se han analizado varias ontologías para la anotación de las geometrías y se decidió usar geosparql ya que es la que mas de apega a la información que brindan los servicios WFS, El proceso de anotación se desarrollo sin inconvenientes por lo que se puede decir que este objetivo a sido cumplido.
- Implementar un proceso de invocación de servicios WFS. Este proceso se a realizado mediante la lectura de los servicios WFS atacando directamente al GML que están soportando actualmente los servidores de mapas como geoserver, la invocación del servicio para la obtención del GML se realizo mediante operaciones *URLConnection*, se a analizado otras opciones como el uso de *geotools* para la invocación de servicios pero existían graves problemas al invocar servicios con la versión 2.0.0 (versión por GML) por lo que se opto la primera opción para el desarrollo de este objetivo.
- Creación de un repositorio con información semántica sobre los servicios WFS catalogados. Luego de generar las anotaciones correspondientes con los servicios WFS y las respectivas ontologías se logro tener la información necesaria para el almacenamiento de datos, usando el *Triple Store marmotta* por lo que se cumple el objetivo.
- Verificar la información generada en el proceso de anotación mediante consultas al repositorio local. Se han definido ciertas consultas para la verificar que la información obtenida durante el proceso de anotación esta correcto, las consultas se encuentran en el capítulo 4 para la muestra de resultados, de esta manera se da por completado el objetivo.

Luego de almacenar la información anotada de los servicios se a realizado un proceso para encontrar enlaces hacia otras fuentes de información mediante la plataforma silk, en el cual se pudo constatar que la información de los *features* anotada es información válida, de este modo se logro cumplir el objetivo general el cual es: Implementar un proceso para anotar semánticamente los servicios WFS relacionando sus distintos niveles de abstracción con diversas ontologías de dominio.

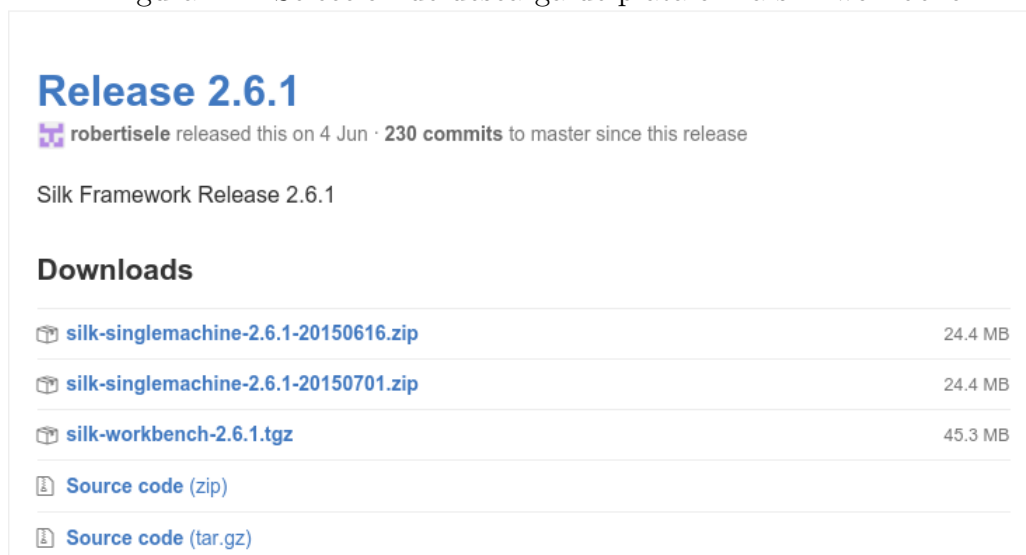
Apéndice A

INSTALACIÓN DE LA PLATAFORMA SILK WORKBENCH

A continuación se describe el proceso de instalación de la plataforma silk workbench, este programa a sido instalado y probado sobre las plataformas de Linux en las versiones de Ubuntu y OpenSuse.

Para descargarnos el archivo de instalación de silk accedemos a la siguiente url <https://github.com/silk-framework/silk/releases> y seleccionamos la versión para Linux (silk-workbench-2.6.1.tgz) que nos presenta la siguiente figura A.1 .

Figura A.1: Selección de descarga de plataforma silk-workbench

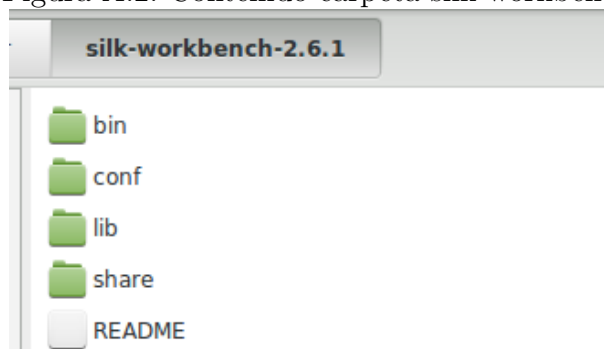


The screenshot shows the GitHub release page for Silk Framework 2.6.1. The title is "Release 2.6.1" and it was released by robertisele on 4 Jun. There are 230 commits to master since this release. The release is titled "Silk Framework Release 2.6.1". Under the "Downloads" section, there are four items:

Download	Size
silk-singlemachine-2.6.1-20150616.zip	24.4 MB
silk-singlemachine-2.6.1-20150701.zip	24.4 MB
silk-workbench-2.6.1.tgz	45.3 MB
Source code (zip)	
Source code (tar.gz)	

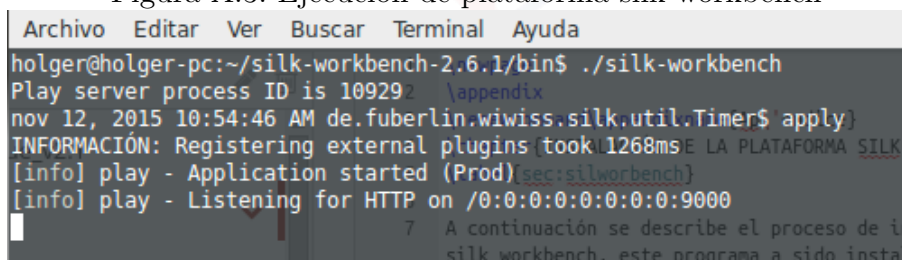
Una vez descargado el archivo, lo copiamos a un directorio donde se vaya a trabajar sin tener problemas por ejemplo: `/home/usuario/Documentos/silk` y lo descomprimos mediante la terminal con el siguiente comando `tar -xvzf silk-workbench-2.6.1.tgz` en el cual obtendremos la carpeta con la siguiente información. (ver figura A.2)

Figura A.2: Contenido carpeta silk-workbench



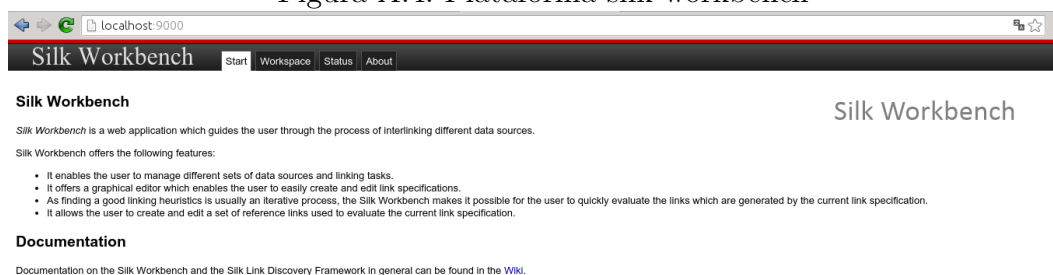
ingresamos a la carpeta **bin** en la cual encontraremos un archivo ejecutable, mediante la consola de Linux accedemos a la ruta en que se encuentra dicho archivo y ejecutamos el siguiente comando.(figura A.3)

Figura A.3: Ejecución de plataforma silk-workbench



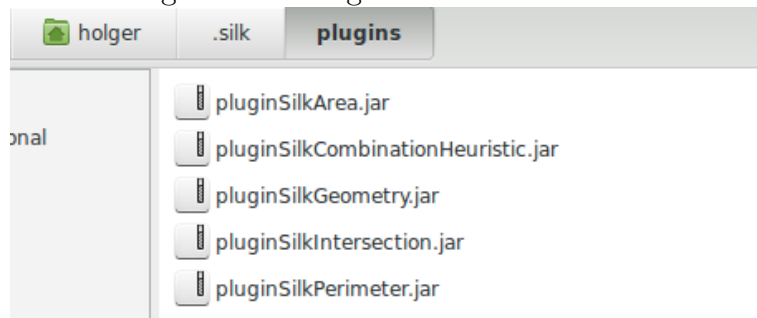
por defecto la plataforma silk usa el puerto 9000 por tal motivo a la hora de ingresar lo podemos hacer por la siguiente url `localhost:9000`. (figura A.4)

Figura A.4: Plataforma silk-workbench



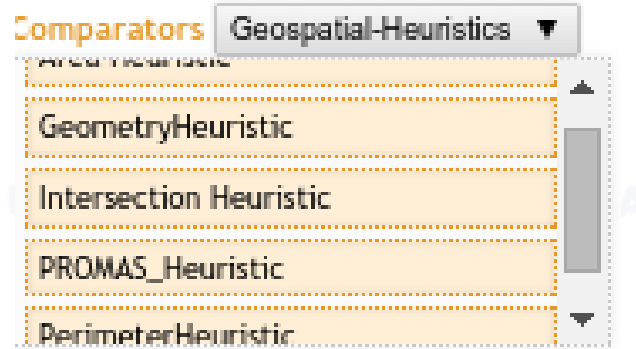
Para el uso de los plugins generados por la universidad de cuenca para el análisis de geometrías es necesario cargar dichos plugins en nuestra plataforma para ello silk crea una carpeta oculta en el directorio `home` de nuestro usuario, dentro de este directorio es necesario crear una carpeta llamada `plugins` y pegar los archivos jar necesarios. (figura A.5)

Figura A.5: Plugins en silk-workbench



los cuales se ven de la siguiente manera en la plataforma silk-workbench. (figura A.6)

Figura A.6: Carga de plugins a silk-workbench



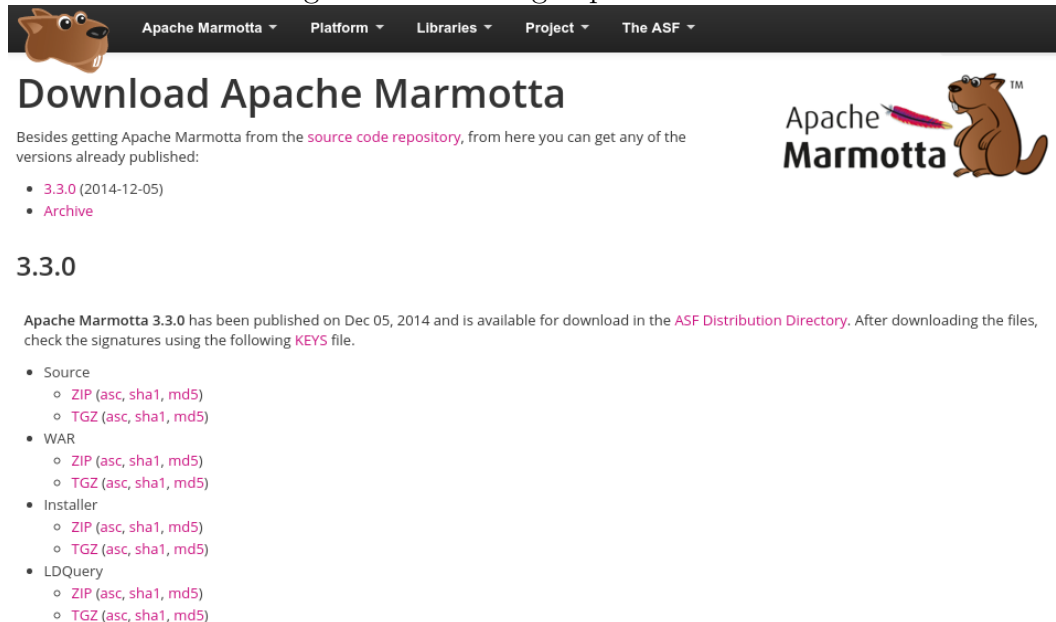
Apéndice B

INSTALACIÓN DEL ENDPOINT MARMOTTA

A continuación se describe el proceso de instalación del endpoint marmotta, este programa a sido instalado y probado sobre las plataformas de Linux en las versiones de Ubuntu y OpenSuse.

Para descargarnos el archivo de instalación de marmotta accedemos a la siguiente url <http://marmotta.apache.org/download.html> y seleccionamos la versión installer ZIP que nos presenta en la siguiente figura.

Figura B.1: Descarga apache marmotta



Download Apache Marmotta

Besides getting Apache Marmotta from the [source code repository](#), from here you can get any of the versions already published:

- [3.3.0 \(2014-12-05\)](#)
- [Archive](#)

3.3.0

Apache Marmotta 3.3.0 has been published on Dec 05, 2014 and is available for download in the [ASF Distribution Directory](#). After downloading the files, check the signatures using the following [KEYS](#) file.

- Source
 - [ZIP \(asc, sha1, md5\)](#)
 - [TGZ \(asc, sha1, md5\)](#)
- WAR
 - [ZIP \(asc, sha1, md5\)](#)
 - [TGZ \(asc, sha1, md5\)](#)
- Installer
 - [ZIP \(asc, sha1, md5\)](#)
 - [TGZ \(asc, sha1, md5\)](#)
- LDQuery
 - [ZIP \(asc, sha1, md5\)](#)
 - [TGZ \(asc, sha1, md5\)](#)

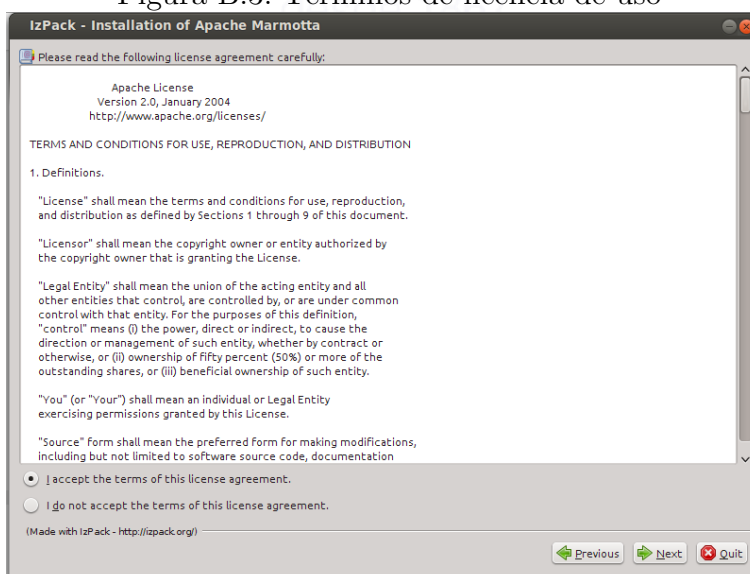
una vez descargado extraemos la información del archivo ZIP y procedemos a la ejecución del instalador mediante el comando `java -jar marmotta-installer-3.3.0.jar` el cual nos presentara la siguiente guía de instalación. (figura B.2)

Figura B.2: Apache marmotta guía de instalación



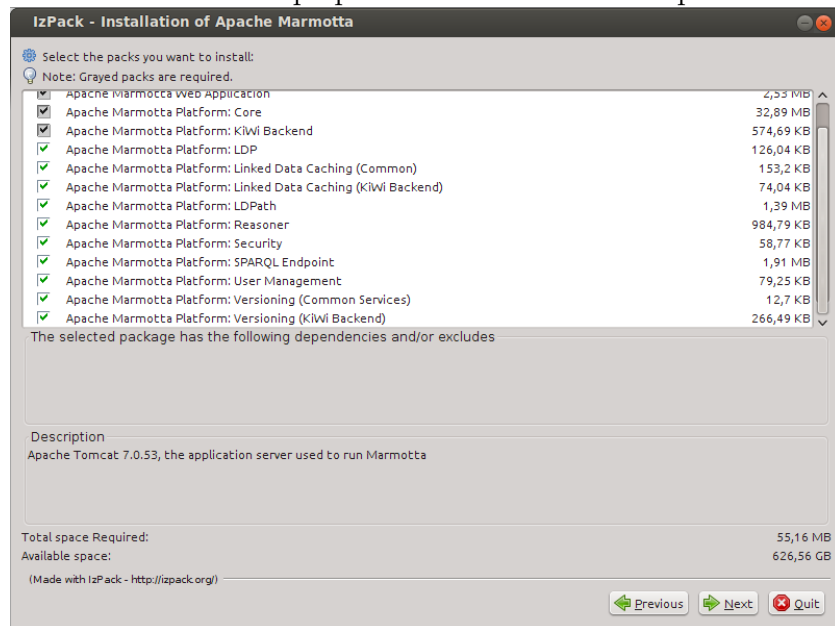
Una vez aceptado nos presenta la licencia de uso de apache, se la aprueba y continua. (figura B.3)

Figura B.3: Términos de licencia de uso



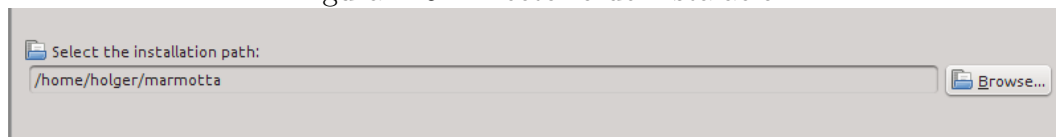
Posterior a esto, se procede a la selección de todos los paquetes que trae consigo apache marmotta para la instalación y su correcto funcionamiento, por defecto se a dejado todos los proyectos que ya están marcados de manera automática, luego presionamos next para continuar con el proceso de instalación. (figura B.4)

Figura B.4: Selección de paquetes de instalación de apache marmotta



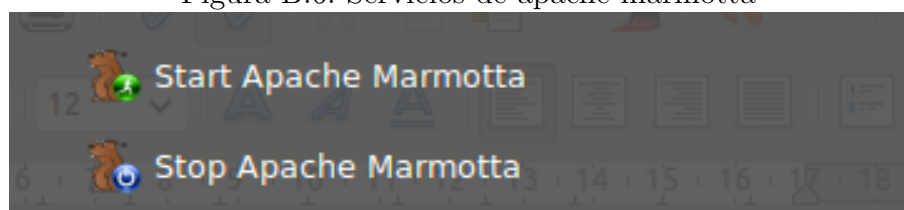
Se selecciona el directorio donde vamos a instalar apache marmotta y presionamos next. (figura B.5)

Figura B.5: Directorio de instalación



De esta manera comienza el proceso de instalación, finalmente al terminar este proceso apache crea dos servicios que nos ayudan a iniciar/detener los servicios en nuestro sistema. (figura B.6)

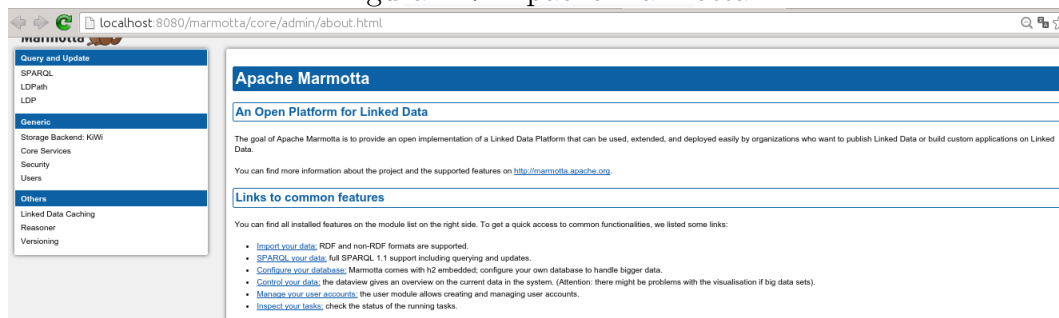
Figura B.6: Servicios de apache marmotta



Al iniciar el servicio con **Start Apache Marmotta** por defecto podemos ingresar a la aplicación por el puerto 8080 quedando así la url establecida `http://localhost:8080`. (figura B.7)



Figura B.7: Apache marmotta



Apéndice C

FUENTES DE INFORMACIÓN

En esta sección presentamos los sitios de donde se sacó la información de los servicios WFS que fueron analizados y anotados respecto a las ontologías en el proceso de trabajo realizado.

Figura C.1: Fuentes de información

SERVICIOS WFS	LUGAR
http://gis.uazuay.edu.ec:8080/geoserver/ows?service=wfs&version=2.0.0&request=GetCapabilities	Universidad del Azuay
http://mapainteractivo.ambiente.gob.ec/geoserver/ows?service=wfs&version=2.0.0&request=GetCapabilities	Mapa Interactivo Ambiental
http://www.geoportaligm.gob.ec/regional/ows?service=wfs&version=2.0.0&request=GetCapabilities	Instituto Geografico Militar
http://www.geoportaligm.gob.ec/nacional/ows?service=wfs&version=2.0.0&request=GetCapabilities	Instituto Geografico Militar
http://www.geoportaligm.gob.ec/galapagos25/wfs?service=WFS&request=GetCapabilities&version=2.0.0	Instituto Geografico Militar
http://geoportal.magap.gob.ec/geomagap/ows?service=wfs&version=2.0.0&request=GetCapabilities	MAGAP
http://www.ign.es/wfs-inspire/services/transportes?version=2.0.0&request=GetCapabilities&service=WFS	transporte España
http://geocatmin.ingemmet.gob.pe/arcgis/services/PSAD56_18/SERV_CATASTRO_MINERO_18/MapServer/WFSServer?request=GetCapabilities&service=WFS	Catastro minero de Peru
http://geoservidor.minam.gob.pe/arcgis/services/ServicioCCNN/MapServer/WFSServer?request=GetCapabilities&service=WFS	Comunidades nativas Peru
http://mtcgeo2.mtc.gob.pe:8080/geoserver/MTC_gis/wfs?service=WFS&version=2.0.0&request=getCapabilities	Red vial nacional Peru
http://siga.conagua.gob.mx/wfsconnector/com.esri.wfs.Esrimap/CuerposAgua_Features?request=getcapabilities&service=wfs&version=2.0.0	Cuerpos principales de agua Mexico
http://siga.conagua.gob.mx/wfsconnector/com.esri.wfs.Esrimap/RHA_Features?request=getcapabilities&service=wfs&version=2.0.0	Regiones hydrologicas administrativas
http://siga.conagua.gob.mx/wfsconnector/com.esri.wfs.Esrimap/RiosPrincipales_Feature?request=getcapabilities&service=wfs&version=1.0.0	Rios principales
http://siga.conagua.gob.mx/wfsconnector/com.esri.wfs.Esrimap/Acuiferos_Features?request=getcapabilities&service=wfs&version=2.0.0	Acuiferos
http://ontorisk.cedia.org.ec/geoserver/ows?service=wfs&version=2.0.0&request=GetCapabilities	CEDIA
http://geoinec.inec.gob.ec/geoinec/ows?service=wfs&version=2.0.0&request=GetCapabilities	INEC



UNIVERSIDAD DE CUENCA
desde 1867

Bibliografía

- [1] Anónimo. <http://www.infor.uva.es/sblanco/tesis/ anotaciones semánticas.pdf>.
- [2] Linked Data. <http://linkeddata.org/>.
- [3] Open Geospatial Consortium Inc. Geography markup language (gml) encoding standard. 2007.
- [4] Open Geospatial Consortium Inc. Web feature service 2.0 interface standard. 2010.
- [5] Web Ontology Language. <https://www.w3.org/2001/sw/wiki/OWL>.
- [6] Robert Lee. Sistema de coordenadas.
- [7] D J Maguire. An overview and definition of gis.
- [8] Patrick Maué. Semantic annotation in ogc standards. 2009.
- [9] Meenaksi Nagarajan. Semantic annotations in web services.
- [10] Estándares OGC. <http://live.osgeo.org/es/standards/standards.html>.
- [11] Maria Pontes, Marco Casanova, Luis Eufrazio, and José Monteiro. A semi-automatic approach for generating customized r2rml mappings.
- [12] Resource Description Framework (RDF). <https://www.w3.org/RDF/>.
- [13] Victor Saquicela. Semantic annotation of restful and wfs ogc services. 2015.
- [14] Victor Saquicela, Luis Vilches, and Oscar Corcho. Anotación semántica de web feature services.
- [15] Web Semántica. <https://www.w3.org/standards/semanticweb/>.