



UNIVERSIDAD DE CUENCA

FACULTAD DE INGENIERÍA

ESCUELA DE INFORMÁTICA

***“Estudio, Diseño e Implementación de un
Buscador Semántico”***

***Tesis previa a la obtención del
Título Ingeniero en Sistemas***

AUTORES:

María Augusta Escandón Cueva

Freddy Orlando Ñauta Ñauta

DIRECTOR:

Jorge Mauricio Espinoza Mejía

Cuenca-Ecuador

2011



RESUMEN

El presente trabajo de investigación, “Estudio, Diseño e Implementación de un Buscador Semántico”, elabora un acercamiento a los conceptos tecnológicos que implica una Web basada en Semántica. También demuestra estos conceptos a través de su aplicación en un **prototipo de buscador semántico**, mismo que; aunque está restringido al dominio turístico, permitirá que todos sus conceptos y tecnologías sean extrapolados a otros dominios de interés.

PALABRAS CLAVE: Web Semántica, Buscador, Inteligencia Artificial, Ontología.



CONTENIDO

CAPÍTULO I, INTRODUCCION	- 12 -
1.1. Presentación	- 12 -
1.2. Justificación.....	- 15 -
1.3. Objetivos:	- 19 -
1.3.1. Objetivos Generales	- 19 -
1.3.2. Objetivos específicos.....	- 20 -
1.4. Contexto.....	- 20 -
1.5. Problemas	- 21 -
1.6. Suposiciones.....	- 23 -
1.7. Hipótesis del Trabajo.....	- 24 -
1.8. Restricciones.....	- 25 -
1.9. Estructura.....	- 27 -
 CAPÍTULO 2, ESTADO DEL ARTE	 - 31 -
2.1. Evolución de la Web.....	- 31 -
2.1.1. La Web 1.0	- 32 -
2.1.2. Web 2.0.	- 34 -
2.1.3. Web 3.0.	- 36 -



2.2. Buscadores Actuales.....	- 39 -
2.2.1. Definición.....	- 39 -
2.2.2. Técnicas de extracción de la Información.....	- 39 -
2.2.3. Problemas	- 45 -
2.3. Ontologías.....	- 46 -
2.3.1. Definición.....	- 46 -
2.3.2. Elementos.....	- 48 -
2.3.3. Tipos de Ontologías.....	- 55 -
2.3.4. Herramientas para soportar Ontologías.....	- 59 -
2.4. Técnicas de utilización del Conocimiento.....	- 64 -
2.4.1. Representación del Conocimiento	- 65 -
2.4.2. Adquisición de Conocimiento	- 67 -
2.5. Buscadores Semánticos.....	- 71 -
2.5.1. Buscadores Semánticos: ¿Para qué sirven?.....	- 71 -
2.5.2. Buscadores Semánticos: ¿Cómo funcionan?.....	- 74 -
2.5.3. Mecanismos utilizados en el proceso de búsqueda Web	- 78 -
2.5.4. Tipos de Buscadores Semánticos	- 80 -
2.5.5. Buscadores Semánticos: Limitaciones para su aplicación	- 83 -
2.6. Transición a la Web Semántica.....	- 85 -
 CAPÍTULO 3, PLATAFORMA TECNOLÓGICA DEL BUSCADOR.....	 - 89 -
3.1. Desarrollo de Ontologías.....	- 90 -



3.2. Metodologías para construir una ontología	- 94 -
3.2.1. Metodologías de desarrollo más destacadas.....	- 95 -
3.3. Lenguajes de representación Semántica	- 99 -
3.3.1. XML	- 99 -
3.3.2.RDF	- 100 -
3.3.3.RDFS	- 104 -
3.3.4.OWL.....	- 104 -
3.4. Lenguajes de Consulta.....	- 107 -
3.4.1.RDQL	- 108 -
3.4.2.SeRQL Query Language.....	- 112 -
3.4.3.SPARQL	- 112 -
3.5. Procesamiento del Lenguaje Natural	- 115 -
3.6. Interfaces y APIS de Programación	- 118 -
CAPÍTULO IV, METODOLOGIA NEON	- 126 -
4.1. Introducción.....	- 126 -
4.2. PASO 1: Especificación de la Ontología	- 128 -
4.3. PASO 2: Reutilización de Recursos No-Ontológicos	- 147 -
4.4. PASO 3: Reingeniería de recursos No-Ontológicos.....	- 155 -
4.5. PASO 4: Reutilización de Recursos Ontológicos	- 172 -
4.6. PASO 5: Reutilización de Plantillas de diseño para Ontologías	- 198 -
4.7. Conclusiones.....	- 200 -



CAPÍTULO V, DISEÑO E IMPLEMENTACION DEL BUSCADOR	- 204 -
5.1. Fase de Análisis	- 205 -
5.2. Fase de Diseño	- 206 -
5.3. Fase de Implementación	- 209 -
5.3.1. Capa de Datos.....	- 209 -
5.3.2. Capa de Negocios	- 219 -
5.3.3. Interfaz.....	- 233 -
5.4. Fase de Pruebas	- 239 -
 CAPÍTULO VI, CONCLUSIONES.....	 - 246 -
6.1. Conclusiones.....	- 246 -
6.2. Recomendaciones.....	- 249 -
6.3. Trabajos Futuros	- 251 -
 ANEXOS	 -236-
 FUENTES DE INFORMACIÓN	 - 263 -



*“Todos los grandes descubrimientos
han surgido de la voluntad, y la
capacidad, de no estar en lo cierto.”*

Neale Donald Walsch



CERTIFICO

Que, bajo mi dirección se desarrolló la Tesis “Estudio, Diseño e Implementación de un Buscador Semántico”, de los alumnos Freddy Ñauta y María Augusta Escandón, como requisito previo a su incorporación como Ingenieros de Sistemas.

Ing. Mauricio Espinoza



Que, los conceptos desarrollados, análisis realizados y las conclusiones del presente trabajo, son de exclusiva responsabilidad de los autores.

Freddy Orlando Ñauta
010414514-9

María Augusta Escandón
010432360-5

Cuenca, 05 de Enero de 2011

DEDICATORIA Y AGRADECIMIENTO



Este trabajo de grado va dedicado a mis padres, que gracias a su dedicación y esfuerzo han sabido apoyarme en cada paso de mi vida, y guiarme con su ejemplo.

Ma. Augusta

Para mi papá, prometo seguir siempre tu ejemplo y nunca defraudarte, agradezco toda la confianza que siempre has tenido en mí.

A mi mami, pues no lo habría podido lograr sin ti, lo que soy y seré siempre será gracias a ti, se que podré contar contigo para siempre.

Para mi hermano Paul, mis hermanitas Mayra, Andrea y María, y de una forma especial para Angie; este trabajo se los dedico de todo corazón, por brindarme su cariño y comprensión.



Freddy

CAPÍTULO I

INTRODUCCIÓN



CAPÍTULO I

INTRODUCCIÓN

1.1. Presentación

La *Web Semántica* es una extensión de la World Wide Web como la conocemos en la actualidad. Permite dotar a la Web de un mayor significado y contenido semántico, organizando la información de forma definida y estructurada, de modo que cualquier usuario en Internet pueda encontrar respuestas a sus preguntas de forma rápida y eficiente.

Esta Web extendida y basada en el significado, se apoya en lenguajes formales para describir la semántica de los términos; y, por lo tanto puede resolver los problemas de una Web privada de semántica en la que, muchas veces, el acceso a la información se convierte en una tarea difícil y larga.

Una Web dotada de mayor significado, es capaz de obtener soluciones a problemas comunes en cuanto a la búsqueda de información, ya que usa una infraestructura común al usuario final, que es el contenido semántico al cual los



usuarios estamos acostumbrados por ser la base del lenguaje natural. De esta manera se vuelve posible compartir, procesar y transferir información de forma sencilla, al mismo tiempo que el usuario puede delegar tareas de extracción y clasificación de información a las herramientas de la Web Semántica.

Los buscadores actuales¹ nos muestran sus resultados basados principalmente en palabras claves, arrojando muchas veces respuestas que no se relacionan con el tema buscado y sin ningún tipo de información en cuanto a sus relaciones con otros temas.

Como ejemplo de esta desventaja; supongamos que contamos con un artículo publicado de cierto Autor. En un buscador semántico podemos preguntar cuáles artículos ha publicado ese autor mediante las relaciones que existen; por ejemplo el concepto “Artículo” estará relacionado con un “Autor”, una “Editorial”, un “Año de Publicación”, etc.; y estos a su vez con otros “Artículos” que podrían ser de relevancia para el usuario. Por el contrario, un buscador sintáctico nos devolvería como resultado cualquier documento que contenga el nombre del autor, y la palabra artículo, sin ningún tipo de clasificación ni organización previa.

Esta manera de extraer información es posible ya que la Web Semántica está basada en la idea de añadir *metadatos* semánticos a la Web. Los metadatos

¹ como por ejemplo Google, Yahoo, etc.



constituyen información adicional que describe el contenido, el significado y la relación entre los datos, y están agregados de manera formal, para que sea posible evaluarlos automáticamente por máquinas de procesamiento.

De esta manera, los contenidos de la Web pueden ser expresados en un lenguaje natural, y también en un formato que pueda ser entendido, interpretado y usado por diferentes tipos de software, permitiéndoles buscar, compartir e integrar información de manera más fácil.

El objetivo de la Web Semántica es mejorar el Internet ampliando la interoperabilidad entre los sistemas informáticos y usando "agentes inteligentes"; programas que buscan información sin necesitar la intervención de operadores humanos.

Con el presente trabajo de investigación se aspira elaborar un acercamiento a los conceptos tecnológicos que implica una Web basada en Semántica. También se pretende demostrar estos conceptos a través de su aplicación en un **prototipo de buscador semántico**, mismo que; aunque está restringido al dominio turístico, permitirá que todos sus conceptos y tecnologías sean extrapolados a otros dominios de interés.



En la siguiente sección señalaremos la motivación de este trabajo a través de la justificación del mismo; en segundo lugar pasaremos a detallar los objetivos, contexto, hipótesis y restricciones del presente trabajo de investigación. Y finalmente se describirá el contenido del Trabajo capítulo a capítulo junto con los principales puntos a tratar en cada uno de estos.

1.2. Justificación

El internet cambió abismalmente la forma en la que nos comunicamos, hacemos negocios y realizamos nuestro trabajo. Hoy en día es posible comunicarse prácticamente con todo el mundo en cualquier momento y a bajo coste. A través de la Web podemos realizar transacciones económicas, y acceder a millones de recursos, independientemente de nuestra situación geográfica e idioma. Todos estos factores han contribuido al éxito de la Web. Sin embargo, y al mismo tiempo, estos factores también han originado sus principales problemas: la sobrecarga de información y la heterogeneidad de fuentes de información con el consiguiente problema de interoperabilidad.

La Web Semántica es un área nueva que está cobrando importancia en la interrelación entre la Inteligencia Artificial y las tecnologías Web. Propone introducir descripciones explícitas sobre el significado de los recursos, es decir



añadir metadatos semánticos a la Web, que describan la relación que existe entre los datos, haciendo posible evaluar la información que se encuentra en el ciberespacio automáticamente.

En el pasado, cuando una persona necesitaba obtener algún tipo de información acerca de algún tema en específico, la biblioteca era el lugar adecuado para consultar los documentos existentes y; mediante catálogos de autor, título, etc., se podía saber la ubicación de estos documentos en la biblioteca, esto obviamente involucraba una inversión importante de tiempo y esfuerzo.

En la actualidad, el internet es considerado como una gran biblioteca cerca ya de convertirse en una gran enciclopedia universal de conocimiento humano. Casi todo en la actualidad está representado de una u otra forma en la Web, y con la ayuda de un buen buscador, podemos encontrar información sobre casi cualquier tema que necesitemos.

Desde el uso masivo del internet, a finales de los 90, los buscadores Web han sido una herramienta más que indispensable en la búsqueda de información en la Web; la definición de un buscador, según un diccionario informático², es la siguiente:

² <http://www.alegsa.com.ar/Dic/buscador.php>



“Sitios especializados que poseen herramientas para facilitar la búsqueda de información de entre las millones de páginas Web existentes”, ejemplos de los buscadores más comunes hoy en día son Google, Yahoo!, Ask, entre otros.

Pero a la vez que el internet evoluciona, las demandas de los usuarios para acceder a la misma también son cada vez más exigentes. Si bien los buscadores facilitan la navegación, existen problemas que van incrementando la ineficacia de los mismos; uno de estos problemas es el acceso a la información, ya que el tiempo que los navegantes requieren para encontrar información relevante se considera, en algunos casos, demasiado elevado. Además, la clasificación y categorización de la información requieren una intervención manual, y muchas veces, los buscadores basados en palabras clave suelen devolver información irrelevante.

En los motores de búsqueda actuales, por lo general se encuentran multitud de respuestas a una misma consulta; algunas pueden contener la respuesta que se busca y otras no guardar ninguna relación con lo que se está buscando. Un ejemplo práctico para reflejar esto puede ser el siguiente, si en un motor de búsqueda cualquiera introducimos el siguiente texto “artículo sobre García Márquez”, encontraremos cientos de artículos de García Márquez, y tal vez muy pocos que traten sobre este autor. Obviamente teniendo en cuenta los continuos



cambios de la Web y los algoritmos de los buscadores, los resultados de esta prueba pueden variar de un día para otro, pero lo que se quiere dejar en constancia con este ejemplo es el síntoma de una causa común: la falta de capacidad de las representaciones en que se basa la Web actual para expresar significados.

La Web Semántica ofrece una alternativa a este tipo de inconveniente, etiquetando la información con metadatos semánticos basados en *ontologías*. La posibilidad de etiquetar la información semánticamente abre nuevas y sofisticadas formas de responder a las consultas de los usuarios. Esto no solo potencia la precisión y calidad comparado con los motores de búsqueda actuales, sino que también ofrecer funcionalidades adicionales como: i) ofrecer información adicional acerca de una respuesta, ii) proveer medidas de confiabilidad y certeza, y/o iii) justificar una respuesta.

La Web Semántica fue creada por Tim Berners-Lee, inventor de la WWW, URIs, HTTP y HTML. El objetivo de la Web Semántica es mejorar el Internet ampliando la interoperabilidad entre los sistemas informáticos y utilizando agentes inteligentes. Agentes inteligentes son programas en las computadoras que buscan información sin operadores humanos.



El precursor de la idea, Tim Berners-Lee, intentó desde el principio incluir información semántica en la Web, pero por diferentes causas no fue posible. Por ese motivo introdujo el concepto de semántica con la intención de recuperar dicha omisión.

Una Web extendida, dotada de mayor significado permite que las propias máquinas tengan un nivel de comprensión suficiente como para hacerse cargo de una parte, quizá la más rutinaria, o físicamente inabarcable del trabajo que actualmente realizan manualmente los usuarios de la Web: la búsqueda, abstracción y clasificación de la información.

1.3. Objetivos:

1.3.1. Objetivos Generales

- Investigación, planteamiento y aplicación de los conceptos y metodologías de la Web Semántica.
- Implementación de un prototipo que ejecute búsquedas semánticas, en el que se apliquen técnicas de búsqueda de conocimiento en lugar de las técnicas clásicas de búsqueda sintáctica.



1.3.2. Objetivos específicos

- Estudiar e identificar las principales diferencias entre información y conocimiento y las técnicas asociadas para su tratamiento.
- Estudiar y analizar las metodologías para la construcción de ontologías de dominio.
- Estudiar, analizar e identificar las técnicas más adecuadas para la ejecución de búsquedas de conocimiento.
- Diseñar e implementar un programa prototipo de búsqueda que permita responder ciertas preguntas en lenguaje natural, contando con una ontología definida previamente y enfocada al dominio turístico.

1.4. Contexto

El contexto del presente trabajo es el marco de los buscadores, concretamente en el análisis, diseño e implementación de un prototipo de buscador semántico.



Nos enfocamos en estudiar las técnicas y herramientas para el desarrollo de un buscador de información, en el que se apliquen técnicas de búsqueda semántica.

El conocimiento estará descrito mediante una ontología en concreto que permita representar el dominio de la aplicación; para el presente trabajo, el dominio estará limitado al conjunto de conceptos y relaciones que se manejan dentro del área Turística de la ciudad de Cuenca (iglesias, hoteles, lugares turísticos, etc.).

Hemos escogido el dominio turístico debido a la importancia que se le está brindando al turismo como una estrategia fundamental de desarrollo para el país y de forma especial en la ciudad de Cuenca, teniendo presente que al finalizar este proyecto, el mismo pudiera ayudar en cierta parte a facilitar la búsqueda de lugares e información turística dentro de la Ciudad.

1.5. Problemas

Los problemas que surgen al momento de plantearnos los objetivos específicos para este proyecto son los siguientes:



- Falta de difusión y estandarización de las tecnologías y metodologías para la construcción de ontologías:

El problema nace en el momento en que comenzó la concepción de la Web tal como la conocemos. El desarrollo y difusión masivos de la Web Semántica tienen algunas dificultades que no ha podido ser superadas todavía: una de ellas es tecnológica y la otra está relacionada con la falta de interés de los propietarios de las páginas Web en realizar las anotaciones semánticas sobre sus páginas. Esta resulta ser la variable del trabajo, ya que a medida que avance la investigación, esto puede mejorar. Estudiar y analizar las metodologías para la construcción de las ontologías de dominio, nos permitirá tener un mejor entendimiento de las mismas y contribuir con su difusión.

- Estudiar, analizar e identificar las técnicas más adecuadas para la ejecución de búsquedas de conocimiento.

Conocemos que los buscadores actuales realizan búsquedas textuales en su mayoría, es decir no toman en cuenta el contexto de una consulta, sino únicamente indexan los resultados en base a estadísticas de acceso y coincidencias de palabras exactas. Esto ocurre debido a que la mayoría de



páginas en la Web no comparten un formato común ni anotaciones que permitan dotarles de mayor significado.

De igual manera no existe en la actualidad un regulador de los contenidos que los autores colocan en sus páginas Web, por esto, la anotación semántica debe permitir la intervención del usuario.

Además cabe recalcar que los buscadores sintácticos actuales arrojan un conjunto de resultados que no es priorizado correctamente, siendo este despliegue de resultados insuficiente para que un usuario encuentre lo que busca en realidad.

1.6. Suposiciones.

- Teniendo en cuenta que la Ontología se ocupa de la definición del ser; y de establecer las categorías fundamentales para un dominio particular en este caso, el dominio Turístico, partimos del punto en el cual será necesario desarrollar la Ontología para adaptarle a nuestro caso en particular y que las relaciones entre sus conceptos están plenamente definidas.



1.7. Hipótesis del Trabajo

La información contenida en el trabajo de investigación ayudará al lector a conocer el marco conceptual y tecnológico en el cual se basa la Web Semántica.

Las siguientes hipótesis plantean los objetivos que tiene la Web Semántica, y que se pretenden exponer y desarrollar en el presente proyecto de Tesis.

Hipótesis 1.

La tecnología aplicada en la Web Semántica puede mejorar el acceso a la información publicada en la Web ampliando la interoperabilidad entre los sistemas informáticos mediante el uso de metadatos semánticos y ontologías.

Hipótesis 2.

Un buscador que aplique comprensión semántica será capaz de interactuar con el lenguaje natural, establecer relaciones entre elementos ontológicos, y sintetizar la información utilizando sus propias herramientas. De esta



manera proporcionará un mejor resultado que el resultado de una búsqueda sintáctica o tradicional.

1.8. Restricciones

Este proyecto tiene como objetivo estudiar y analizar las tecnologías más importantes desarrolladas para llevar a cabo las propuestas de la Web Semántica. Esto se realizará mediante la investigación de los métodos, técnicas y herramientas para la implementación de un prototipo que permita extraer conocimiento de un dominio particular. El buscador semántico que planteamos puede explotar la disponibilidad de sentencias semánticas descritas en ontologías, y así proveer respuestas precisas a consultas complejas. Pero, al mismo tiempo estará sujeto a las siguientes restricciones:

- Tomando en consideración que la implementación de un buscador semántico que abarque múltiples dominios de aplicación sería un proyecto demasiado ambicioso, hemos enfocado nuestra propuesta sobre un dominio concreto. Particularmente, hemos seleccionado el **Dominio Turístico**, delimitando la implementación a un área más específica, como la actividad turística en la ciudad de Cuenca. Cabe resaltar que la descripción



de la actividad turística, con información de lugares, hoteles, iglesias, rutas, etc. será descrito mediante el uso de la ontología de dominio a desarrollar.

- La implementación de este prototipo será realizada en un entorno Web, y no incluirá más temas de ontología que el mencionado anteriormente, tampoco incluirá búsqueda ni presentación de imágenes o ningún tipo de material *multimedia*. Su propósito es sacar el máximo provecho al buscador semántico y poder efectuar las preguntas usando expresiones en lenguaje natural y obtener respuestas.
- El prototipo de buscador semántico a desarrollar se basa en la extracción de información desde la ontología y en la respuesta a preguntas concretas. No trabajará con información de páginas Web ya que, al momento, no existen sitios Web sobre información turística de Cuenca que contengan anotaciones semánticas.
- El trabajo de investigación, y su prototipo demostrativo, no incluyen el análisis ni la aplicación del concepto de 'Agentes Inteligentes' u otra herramienta de 'Inteligencia Artificial'.



1.9. Estructura

El proyecto consta de dos grandes partes; la primera es el proceso teórico de INVESTIGACIÓN, y la segunda parte es la aplicación mediante la creación de la ontología y la implementación de un PROTOTIPO DE BUSCADOR SEMÁNTICO.

Los primeros capítulos conforman a una compilación teórica de los conceptos que serán de utilidad para sustentar la aplicación, para luego abordar la parte práctica en los capítulos finales.

En la parte de investigación del proyecto, se aplicará el método de Investigación, el cual requiere un profundo proceso de estudio y aprendizaje al tratarse en gran parte de una temática nueva y actual. Esta metodología de investigación abarcará en primera parte el entendimiento de las técnicas concerniente a los buscadores actuales, para luego enfocarnos en el estudio de la Web Semántica, y las técnicas y métodos más apropiados para la ejecución de este proyecto.

Seguidamente se sintetizará la información de manera clara y explícita en los capítulos 2 y 3, esto facilitará el proceso de desarrollo de la ontología y el prototipo en los capítulos 4 y 5 respectivamente.

El capítulo segundo, ESTADO DEL ARTE contiene información sobre el estado actual de la Web, aquí trataremos de la evolución de la Web, características de los



buscadores actuales, el concepto de "Ontologías", sus elementos y algunas herramientas de software que permiten trabajar con las mismas. Posteriormente se tratarán algunas técnicas de la Web Semántica, para terminar con un análisis de las características de los buscadores semánticos.

En el tercer capítulo, PLATAFORMA TECNOLÓGICA DEL BUSCADOR SEMÁNTICO, se explica el proceso de desarrollo de una ontología, las metodologías existentes para construir una ontología, los lenguajes de representación semántica y de consultas, y finalmente las APIs e interfaces de programación con soporte semántico.

En el Capítulo Cuarto, NEON METHODOLOGY; TEORÍA Y APLICACIÓN, se presentarán los pasos y actividades que utiliza la metodología NeOn para la construcción de ontologías, detallando cada uno de estos. Seguidamente se aplicará esta metodología para el dominio de Turismo, creando y modificando nuestra ontología dentro de la plataforma NeOn Tool Kit.

El Capítulo Quinto, DISEÑO E IMPLEMENTACIÓN DEL BUSCADOR SEMÁNTICO, trata de la planeación y realización del prototipo de búsqueda semántica en sí. Aquí se explicará el desarrollo del proyecto siguiendo las fases de Ingeniería de Software: Análisis, Diseño, Implementación y Pruebas.



El Capítulo Sexto, CONCLUSIONES Y RECOMENDACIONES, redacta las conclusiones obtenidas al final del proyecto de tesis, y las recomendaciones y sugerencias para trabajos futuros.

CAPÍTULO II

ESTADO DEL ARTE



El capítulo segundo, ESTADO DEL ARTE contiene información sobre el estado actual de la Web, y algunos aspectos tecnológicos relacionados con la Web Semántica.

Cómo primer punto se estudiarán las características más importantes de la Web mediante un análisis de su evolución. La segunda parte trata de las técnicas de búsqueda sintáctica y su situación actual. En el apartado tercero, definiremos el concepto "Ontología"; detallaremos sus elementos y explicaremos algunas herramientas de software que permiten trabajar con Ontologías. A continuación entraremos en el campo de la Web Semántica



explicando las técnicas de administración del conocimiento, anotaciones semánticas y buscadores semánticos. Finalmente, se concluirá el capítulo con un análisis de la problemática que representa la Transición hacia la Web Semántica.

CAPÍTULO 2

ESTADO DEL ARTE

2.1. Evolución de la Web

La palabra ‘Web’ hace referencia a una red informática que abarca un sinnúmero de dominios de cualquier índole alojados en servidores en cada lugar del planeta.

Se atribuye la creación de la Web al físico Británico Tim Berners Lee.

Berners Lee, durante su trabajo en el CERN³ propuso un proyecto basado en hipertexto, cuya finalidad era compartir información entre los investigadores. Así nació su proyecto “Enquire Within”. Este programa podía apuntar a cualquier tipo de documento, estuviera o no en un ordenador del CERN y sin importar el formato del documento, todo esto gracias a su diseño extensible.

³ Organización Europea para la Investigación Nuclear - Suiza



En 1989, el CERN era el nodo de Internet más grande de Europa y Berners-Lee vio la oportunidad de unir el Internet y el hipertexto (HTTP y HTML). En 1990 diseñó y construyó el primer navegador Web llamado WorldWideWeb, y el primer servidor Web al que llamó HTTPD (HyperText Transfer Protocol daemon); que se encontraba en el CERN y fue puesto en línea el 6 de agosto de 1991.

Hoy en día, la World Wide Web (WWW) es una compleja red de páginas relacionadas y enlazadas entre sí que forman un medio de comunicación basado en texto, gráficos, y otros objetos *multimedia* a través del Internet. En otras palabras, la Web es un sistema de hipertexto que utiliza Internet como su mecanismo de transporte.

A continuación analizaremos cada una de las etapas de la Web a través de su evolución.

2.1.1. La Web 1.0

Se conoce como WEB 1.0 a la primera red de páginas Web publicadas en el Internet. Comprende el período de los años 1991 a 2003 y es la forma básica de la Web. En sus inicios presentó navegadores de sólo texto como ELISA^[1]. Después, con el surgimiento del lenguaje HTML, surgieron los navegadores visuales que



permitirían mostrar contenido *multimedia* y hacer las páginas Web más agradables para el usuario. Ejemplos de estos son Internet Explorer y Netscape.

Durante esta etapa, la Web fue un repositorio de páginas Web estáticas que ofrecían información de solo lectura a sus usuarios, y estaba totalmente limitada a las publicaciones del Webmaster o Administrador del sitio Web. Los usuarios no podían interactuar con este contenido de ninguna manera.

Algunos elementos de diseño típicos de un sitio Web 1.0 incluyen:

- Páginas estáticas.
- El uso de *framesets* o Marcos.
- Libros de visitas online o Guestbooks.
- Botones GIF, promocionando navegadores Web u otros productos.
- Formularios HTML enviados vía e-mail.

Durante esta etapa, el Internet llegó a tener aproximadamente 45 millones de usuarios en todo el mundo, para un total de 250,000 sitios Web que eran actualizados de manera periódica por sus Administradores^[2]. Esta baja popularidad y expansión (en comparación con la expansión actual de la Web), se



debían a la falta de acceso al Internet en muchos países, a su elevado costo, y a su falta de difusión por parte de las instituciones educativas.

2.1.2. Web 2.0.

En el año 2001 el Internet dio un giro al introducir la Web 2.0. Aquí empezó la interacción de los usuarios hacia la Web, y la Web pasó a actuar como un punto de encuentro, volviéndose, en muchos casos, dependiente de los usuarios.

La Web 2.0 está orientada a la interacción y a las redes sociales en donde se puede servir contenido que explota los efectos de las redes. Se puede entender a la Web 2.0 como la orientación que ha tomado el Internet a través de la utilización de tecnologías como *Ajax*, *DHTML*, *XML*, y *Soap*, y no como un tipo de tecnología o una aplicación en sí.

Algunas características de la Web 2.0 incluyen:

- La usabilidad simplificada de los Sitios Web,
- El ahorro de tiempo para los usuarios,
- La estandarización de los lenguajes para una mejor re-utilización del código,



- La interoperabilidad entre aplicaciones y entre las aplicaciones y las máquinas (software-hardware),
- La facilidad de convergencia entre los medios de comunicación y los contenidos, que facilita la publicación, la investigación y la consulta de contenidos Web; y,
- La posibilidad de aprovechar y estimular la inteligencia colectiva en beneficio de la Web, ayudándose de la *folcsonomía*.

Sin negar todos los beneficios que trae consigo la Web 2.0; esta ‘democratización’ de la Web involucra nuevos inconvenientes: Imaginemos una consulta cotidiana en un buscador Web; si un turista necesita información GPS del lugar en donde se encuentra, es probable que el buscador, entre muchas páginas que tengan la información correcta, despliegue cientos de páginas que contengan el término ‘GPS’, pero que se encuentren en Blogs, Redes Sociales o Sitios Web de Opinión, y, una vez que el usuario se dirija hasta estos sitios solamente encontrará la opinión de usuarios Web sobre la tecnología GPS, sobre los nuevos dispositivos GPS que existen o simplemente encontrará referencias vagas e inútiles sobre este tema, ya que la información que se puede hallar en redes sociales o wikis, no siempre es exacta.



En este sentido, la Web 1.0 era mucho más exacta ya que los sitios Web de información no permitían entradas de comentarios por parte de los usuarios y el proceso de búsqueda era mucho más sencillo.

Algunos de estos problemas de búsqueda de información son tratados en la siguiente evolución de la Web, la cual describimos a continuación.

2.1.3. Web 3.0.

La Web 3.0 se encuentra en etapa de desarrollo, es decir, todavía no se puede hablar de la Web 3.0 como una realidad, pero su propuesta pretende detallar cómo será el próximo paso en la evolución de la Web.

El término “Web 3.0” apareció por primera vez en el año 2006 en un artículo de Jeffrey Zeldman^[3], crítico de la Web 2.0. Actualmente se habla del inicio de la era Web 3.0 y se estima su vigencia entre los años 2010 y 2020.

El principal objetivo de la Web 3.0 es el de “dotar significado a la Web”, lo cual será posible al pasar de la Gestión de la Información a la Gestión del Conocimiento.



En términos informáticos, se puede entender al “Conocimiento” como la relación que existe entre la información. Es decir, el conocimiento es la convergencia lógica entre distintas fuentes de información, basado en la experiencia y en el significado de un concepto dentro de su contexto.

Es por este motivo que Tim Berners-Lee, la llama “Web Semántica”. Sin embargo hay que saber diferenciar entre estos dos conceptos ya que la Web 3.0 es un estado que será alcanzado en mayor o menor plazo, mientras la Web Semántica es un proceso evolutivo de la Web que esta en construcción permanente; proceso en el que ya estamos inmersos y que trasciende el concepto de Web 3.0.

Principales Características

- La “Web Semántica” se refiere a la visión del World Wide Web Consortium, o W3C, de una Web de información enlazada. Las tecnologías de esta Web permitirán que los usuarios creen almacenes de datos en la Web, construyan vocabularios y escriban reglas para el manejo de los datos. Todo esto gracias a la introducción de los llamados “Datos semánticos” y a la información enlazada, potenciados por tecnologías como RDF, SPARQL, y OWL; mismas que serán tratadas en los capítulos siguientes.



- La Web 3.0 utilizará herramientas tecnológicas conocidas como programas Inteligentes, o programas inmersos en la Web que permiten realizar tareas de clasificación de la información gracias a los datos semánticos dentro de las páginas Web. Esta clasificación es posible solamente si los Sitios Web poseen etiquetas semánticas, es decir que; estos sitios Web, además de contener palabras clave como las que contienen en la actualidad, deberán contener información semántica sobre estas palabras clave.
- La información semántica implica incluir una relación entre un artículo Web y su contexto, esto significa que los sitios Web incluirán un enlace de relación hacia una ontología general a la cual pertenece la información publicada en ellos.
- Básicamente, la idea se refiere a una Web capaz de interpretar e interconectar un número mayor de datos, lo que permitiría un avance importante en el campo del conocimiento. Del mismo modo, los *motores de búsqueda* podrán evolucionar y hacerse compatibles con la *Web Semántica*, en la medida en que sus usuarios exijan dicha compatibilidad.



En la actualidad, se utilizan los programas inteligentes y los datos semánticos a pequeña escala en algunas compañías para conseguir una manipulación de datos más eficiente. En los últimos años, sin embargo, ha habido un mayor enfoque dirigido a trasladar estas tecnologías de inteligencia semántica al público en general.

2.2. Buscadores Actuales

2.2.1. Definición

Un Buscador Web es un sistema informático que ofrece consultar una base de datos en la cual se relaciona el contenido de millones de páginas Web, y presentar como respuesta un listado de referencias a páginas que contienen datos sobre las *palabras clave* del tema buscado.

2.2.2. Técnicas de extracción de la Información



Básicamente existen tres tipos de buscadores, los que se clasifican según la forma en que obtienen la información; de la siguiente manera:

- **Índices temáticos:** Son sistemas de búsqueda por temas o categorías cuyas bases de datos son creadas manualmente, hay personas que se encargan de asignar cada página Web a una categoría o tema determinado. Un ejemplo de este tipo, es el primer buscador Yahoo!.
- **Motor de Búsqueda:** Son sistemas de búsqueda por palabras clave y tienen bases de datos que incorporan automáticamente páginas Web mediante "robots" de búsqueda en la red.
- **Metabuscadore:** Utilizan las bases de otros buscadores ya que no tienen una propia.

Los buscadores más utilizados hoy en día, son los motores de búsqueda. Están basados en páginas de hipertexto, y en ellos se introducen las palabras o conjunto de palabras clave a buscar. Estos programas hacen el papel de un intermediario entre los usuarios y las bases de datos que están almacenadas en servidores y a las cuales pueden acceder.

Funcionamiento Básico



El funcionamiento básico de todos los motores de búsqueda es el mismo; se basa en tres componentes que interactúan entre sí para obtener resultados de manera rápida. La diferencia está en la tecnología que se utiliza para cada uno de estos componentes y los algoritmos de búsqueda e indexación, propios de cada buscador. Estos componentes son:

1. La “Araña”

Cuando hacemos una petición de información a un buscador, la búsqueda se realiza con palabras claves a través de árboles jerárquicos clasificados por temas, como resultado, estas páginas muestran una lista de direcciones Web en las cuales se mencionan algunos temas relacionados con las palabras clave buscadas. Los sistemas de búsqueda indexan parte de la información en su propia base de datos y ya no es necesario recorrer toda la Web al realizar una consulta.

Para indexar esta información; los buscadores actuales incluyen *índices* automáticos de páginas Web y utilizan programas autómatas conocidos como “Arañas” o Robots que están diseñadas para recorrer el ciberespacio siguiendo enlaces entre páginas; durante este trayecto se busca páginas nuevas o actualizaciones en las páginas antiguas contenidas en la base de datos; se analiza el código fuente (HTML) y el contenido de cada página visitada, y se lleva esta información a una central en donde es procesada y almacenada.



Existen muchas Arañas navegando constantemente en la Web, la mayoría de estas pertenecen a los buscadores. Las 6 principales arañas son Google Bot (Google), Slurp (Yahoo), MSNBot (Microsoft), Scooter (Altavista), Fast (Webs privadas), y Teoma (Ask Jeeves).

Se puede analizar qué Arañas visitan una página Web utilizando programas de análisis estadístico, esto es posible ya que las Arañas dejan un rastro en los logs (registros) de las páginas Web que visitan.

2. La Base de Datos.

En segundo lugar tenemos la Base de Datos; esta almacena información sobre un gran número de páginas que se extraen de la Web. Las Bases de Datos de los buscadores deben permanecer en constante actualización y estructurar su información para optimizar las búsquedas. Como resultado, estos sistemas son capaces de procesar una consulta de manera casi instantánea aunque contengan millones de registros.

La base de datos debe guardar la información de cada una de las páginas Web que la araña le ha enviado. Para esto, cada página se analiza para determinar cómo clasificarla, por ejemplo a partir de su título, del texto que contiene o de unas



etiquetas invisibles denominadas metatags, que proporcionan datos sobre su contenido.

Después, el sistema crea una estructura de datos conocida como “Índice” que contiene las palabras más utilizadas en cada página Web, y ordena estas palabras por relevancia. Los *índices* de palabras de una página Web son listas en las que se tiene en cuenta las veces que aparece cada palabra en una página, así como el número total de palabras que contiene esa página y la densidad que tiene las palabra en el conjunto de palabras de la página (las veces que se repite una palabra, dividido por el número total de palabras).

Estos índices serán utilizados para detectar coincidencias con palabras clave al momento de una búsqueda, y para mostrar la información más relevante conforme a las mismas.

Existe una técnica conocida como “Smart Representacion”^[4] o búsqueda inteligente, que pretende elaborar el Índice con una estructura tal que permita reducir el tiempo de búsqueda al mínimo. Bajo la Smart Representation, los datos se ordenan en forma de árbol y las búsquedas comienzan desde arriba en el nodo raíz, este nodo contiene la “palabra nudo”; si el término a buscar, comienza por una letra del alfabeto anterior a la letra inicial de la palabra nudo, el buscador parte de la ramificación izquierda; y si la letra es posterior en el alfabeto, el buscador parte de la ramificación derecha. Cada nudo sucesivo ofrece más ramificaciones



hasta que se localiza el término en cuestión, o hasta concluir que no se encuentra en el árbol.

3. El Algoritmo de relevancia

Cuando se realiza una búsqueda, el motor busca en el índice y recupera un listado de las páginas que, según su criterio, se adecúan más a la consulta. Como resultado ofrece numerosos enlaces *URL* (unified resource locator). Pero todos los enlaces resultantes podrían tener la misma relevancia, es aquí en donde el algoritmo de relevancia ordena los resultados basándose en una serie de variables que tienen en cuenta la importancia de la página Web y la importancia del término buscado, dentro de esta página Web.

Un método habitual, es el “peso según frecuencia”, este método asigna un peso relativo a las palabras para indicar su importancia en cada página Web; la asignación de relevancia se basa en la distribución de las palabras y en la frecuencia con que aparecen. Para mostrar los resultados de una búsqueda, se leen los criterios de relevancia y se ordena las páginas de acuerdo a estos. Por último, se compone la página de respuesta al usuario, en la que se muestran los resultados ordenados de mayor a menor relevancia.



2.2.3. Problemas

Los motores de búsqueda Web nos permiten localizar cualquier tipo de información albergada en el Internet de manera rápida, económica y eficiente. Sin embargo, estos sistemas presentan mejores resultados para personas con una mayor experiencia en su utilización, como por ejemplo, una persona que sabe exactamente qué tipo de palabras clave y qué operadores lógicos darán mejores resultados para su búsqueda, y sobretodo qué palabras o términos estarían demás al momento de acudir a un buscador Web.

El gran crecimiento de la Web agrava este inconveniente, ya que si un usuario ingresa palabras o frases ambiguas, el número de páginas Web que el buscador dará como resultado será muy grande y, lo que es peor, la mayoría de estos resultados no serán capaces de satisfacer las necesidades del usuario. En consecuencia, sería necesario invertir mucho tiempo en clasificar y filtrar toda esta información de forma manual.

Al parecer, la gran popularidad de los motores de búsqueda basados en palabras clave se debe a la paciencia de sus usuarios, y sobre todo, a que nos hemos acostumbrado a clasificar los resultados manualmente.

La solución a todos estos inconvenientes es la implementación de buscadores Semánticos; buscadores que tomen en cuenta el contexto y el significado de la



consulta, y no solamente la sintaxis exacta de las palabras que la conforman. De esta manera los usuarios podrían formular búsquedas más expresivas y obtener resultados precisos sin necesidad de una mayor intervención de su parte.

2.3. Ontologías

La utilización de ontologías en la Web puede hacer posible la organización de la información de manera lógica y puede facilitar su tratamiento a la hora de encontrar información y sacar deducciones a partir de ella.

2.3.1. Definición

En el ámbito informático, una *Ontología* es una descripción explícita y formal de los conceptos de un dominio dado. Cada concepto perteneciente a la ontología es representado a través de entidades llamadas CLASES (CLASS); y las propiedades y métodos de las Clases se conocen como ATRIBUTOS (SLOTS) y representan sus características. Una ontología puede utilizarse como una herramienta para definir un dominio y razonar sobre sus características.



El científico americano Thomas Gruber, precursor de la *Ingeniería Ontológica*, describe este concepto de la siguiente manera: “Una ontología es una especificación explícita de una conceptualización” [5].

Para entender este concepto podemos decir que una conceptualización es una abstracción o simplificación de un dominio que se quiere representar. La especificación que aporta la ontología permite describir, entender y analizar la naturaleza de los elementos de dicho dominio. Sin embargo, no se debe confundir a una ontología con una base de datos ni una tabla de contenidos, ya que no contiene registros ni datos en sí, si no que permite organizarlos y relacionarlos, intercambiar estos datos entre programas, simplificar la traducción de los datos desde distintas representaciones y facilitar la comunicación entre las personas que las utilizan.

Actualmente las ontologías se utilizan en importantes áreas científicas como la Inteligencia Artificial, Informática Biomédica^[6], Tecnología de Dotación Lógica^[7], y la *Arquitectura de la Información*, entre otras.

La Figura 2.1 muestra un ejemplo de ontología para el concepto “ARBOL”; como se puede observar en la figura, la ontología parte de un concepto base o padre, luego se desglosa por categorías o jerarquías, las cuales se detallan cada vez más en sus niveles inferiores.

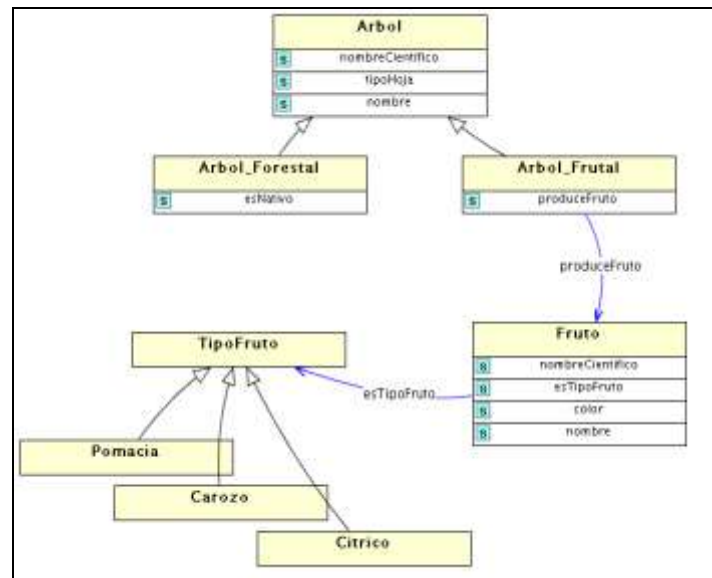


Figura 2.1. Ejemplo de Ontología

2.3.2. Elementos

Todo tipo de ontología, sin importar la naturaleza del dominio que representa, ni el lenguaje en el cual se expresa, utiliza una misma estructura para representar el conocimiento. Esta estructura comprende los siguientes elementos:

1. Clases

Las clases son el centro de una ontología. Representan conceptos genéricos que pueden abarcar otros conceptos o instancias más concretos. También determinan



el tipo o categoría de un cierto grupo de elementos que se pueden agrupar gracias a sus características en común.

Una clase es un concepto abstracto que describe la naturaleza de una entidad del dominio que se quiere representar y puede contener a otras subclases que son más específicas que la superclase. Es así, que se puede dividir a las clases según su jerarquía siendo las Clases Universales la jerarquía más alta, ya que son las clases que contienen un mayor número de atributos comunes para muchas otras clases.

Algunos ejemplos de clases son:

- Persona, la clase de toda la gente.
- Empleado, subclase de la clase persona.
- País, la clase de todos los países.

Una clase abstracta es una clase que no admite instancias de ella, y se utiliza solamente para agrupar conceptos e introducir cierto orden en la jerarquía.

2. Atributos

También conocidos como Slots, son las propiedades que permiten describir detalladamente la Clase y sus instancias a través de las características asociadas



a dicha Clase. Un atributo establece una propiedad que posee la Clase y que se concreta mediante un valor particular al momento de instanciar la Clase.

Los atributos pueden ser de distintos tipos dependiendo de los valores que contengan. Un atributo nominal contiene valores textuales o descriptivos como el tipo de suelo, nombres, etc. Los atributos numéricos contienen valores reales, números enteros, o valores de colores RGB en el caso de una composición de colores. También se pueden utilizar otros tipos de atributos como fechas, valores booleanos, otras Clases, etc.

Algunos ejemplos de atributos, para la clase Persona son:

- Nombre.
- Cedula de Identidad.
- E-mail

3. Instancias

Son objetos específicos pertenecientes a una Clase y constituyen el componente más básico de una ontología. Las instancias son entidades concretas cuya naturaleza esta descrita por una Clase definida y pueden ser objetos concretos (personas, animales, automóviles) u objetos abstractos (números, palabras).



Las instancias no son explícitamente parte de la ontología, si no que ésta proporciona medios de describirlos y clasificarlos.

Cómo ejemplo, una “Instancia” de la clase “PERSONA” es el Sr. “José Calle”, cuyos atributos serían:

- Nombre: José Calle
- Cédula de Identidad: 0105734456
- E-mail: jose_calle@hotmail.com

Este es un ejemplo de una instancia Directa, ya que pertenece a una sola Clase y no existen clases intermedias. También existen las Instancias Indirectas, que son las instancias de todas las Clases que contengan a la primera Clase Directa, en este caso existen Clases derivadas que actúan como intermediarias entre la Clase y su instancia Indirecta.

4. Relaciones

Simbolizan la manera de relacionarse que tienen los conceptos de un dominio y permiten organizar las Clases de manera jerárquica. Los tres grandes tipos de relaciones entre dos Clases son:



- Generalización / Herencia (es-un): Las Clases relacionadas (heredadas) cuentan con los atributos de la Clase con la cual se relacionan. Se puede dar una Herencia múltiple cuando una clase hereda las propiedades de dos clases padre con las que establece dos relaciones del tipo 'es_un'.

Ejemplo: “Empleado” es-una “Persona”. En este caso la clase Empleado hereda de la clase Persona y correspondientemente, la clase Persona es una generalización de la clase Empleado.

- Agregación (todo-parte//tiene-un)

Ejemplo: “Empresa” tiene-un “Presidente” (Instancia de la clase persona)

- Asociación (tiene y esta dentro de)

Ejemplo: La “Empresa” esta presente en varias “Ciudades”; una ciudad tiene varias “Empresas”.

Las relaciones representan una propiedad de una clase cuyo valor es una instancia de otro concepto. Por ejemplo; para la clase “Empresa”, podríamos tener el atributo “Presidente”, que sería de tipo “Persona”; es decir, la clase “Empresa” necesita una instancia de la clase “Persona” como valor de su atributo “Presidente”.

Otro ejemplo para la misma ontología, podría ser el atributo “Cargo” de la Clase “Empleado”; el cargo de un empleado puede ser una clase independiente ya que



muchos empleados pueden tener el mismo cargo; y sí, hipotéticamente, un empleado también podría tener varios cargos dentro de la Empresa, nos encontraríamos ante una relación de cardinalidad múltiple, y por lo tanto debe crearse una tercera clase que represente la misma; en este caso se podría tener la clase “Rol”, con los atributos “Empleado” y “Cargo”.

5. Funciones o Métodos

Las funciones son un tipo especial de relación en donde se identifica un elemento mediante la realización de uno o varios cálculos.

Están constituidos por una Función que puede recibir parámetros de entrada y siempre devuelve una acción u objeto como resultado de su ejecución. Esta función puede cambiar el estado de los atributos de una instancia dada o acceder a sus datos, es por esto que la función debe ser llamada solamente desde una instancia particular y solo realizará la acción sobre sus atributos sin poder afectar a otras instancias. Una función, por ejemplo, podría retornar el número de elementos asociados que posee un concepto o bien realizar una categorización de las clases, asignaciones de fechas, o cualquier otro tipo de dato importante que se relacione con un concepto. Por ejemplo, pueden aparecer funciones como categorizarClase, asignarFecha, etc.



6. Restricciones

También conocidas como “Facetas”; son las descripciones formales que indican las propiedades que se deben cumplir dentro de la ontología. Se utilizan frecuentemente para describir las propiedades de los Atributos de una Clase, por ejemplo cardinalidad, obligatoriedad, etc.

Si no se cumplen las restricciones, las afirmaciones no serán aceptadas como entradas para la Ontología.

7. Axiomas

Son teoremas que determinan el estado de la Ontología. Al introducir reglas en la Ontología estamos definiendo las relaciones que deben cumplir los elementos de la misma. Por ejemplo: “Si A y B son de la clase C, entonces A no es subclase de B”, “Para todo A que cumpla la condición C1, A es B”, etc.

Un ejemplo para la Clase Empresa, es la regla:

- El presidente es una “persona”;

En este caso estamos afirmando la naturaleza del objeto “presidente” como una instancia de la clase “persona”.



Las reglas también pueden tener la forma de una sentencia “si-entonces”, en este caso se establece un antecedente y su respectivo consecuente que describe inferencias lógicas para una afirmación en particular. Un ejemplo para este caso; en la Ontología “Familia” sería la siguiente:

- Si Juan, es “padre” de Diego; entonces Diego es “hijo” de Juan.

Aquí se establece la relación bidireccional entre las instancias “Juan” y “Diego” a través de una afirmación lógica, pero que deber ser introducida a la Ontología a manera de Regla, para que esta pueda procesar los casos similares en el futuro.

El conjunto de Axiomas abarcan la teoría total que describe la Ontología e incluyen las declaraciones afirmadas como conocimiento a priori y la teoría derivada de declaraciones axiomáticas anteriores.

2.3.3. Tipos de Ontologías

Las ontologías pueden clasificarse de acuerdo a tres criterios:

1. Cantidad y Tipo de Estructura

Según Van Heist^[8] las ontologías pueden clasificarse de acuerdo a la cantidad y tipo de estructura de la conceptualización en:



- Ontologías terminológicas: Especifican los términos que representan el conocimiento en el dominio del discurso. Se utilizan para unificar el vocabulario en un campo determinado.
- Ontologías de información: Especifican la estructura de almacenamiento de las bases de datos y ofrecen un marco estandarizado para el almacenamiento de la información.
- Ontologías del modelado del conocimiento: Especifican conceptualizaciones del conocimiento. Contienen una rica estructura interna y suelen estar ajustadas al uso particular del conocimiento que describen. Incluyen todas las primitivas de representación utilizadas para formalizar el conocimiento en paradigmas de representación del conocimiento como Clases, Instancias, Slots, Facetas, etc.

2. Dependencia y Relación de la Ontología

Según su dependencia y relación con una tarea específica, las ontologías se clasifican en:

- Ontologías de Alto nivel o Genéricas: Permiten modelar los niveles más altos de una realidad describiendo los conceptos generales y las categorías



de alto nivel como Tiempo y Espacio. Al ser genéricas, pueden ser reutilizadas a través de dominios. Algunos ejemplos de ontologías de este tipo son CyC^[9], WordNet^[10], y SUMO.

- Ontologías de Dominio: Describen un vocabulario relacionado con un dominio específico. Se pueden utilizar para representar un tipo de artefacto y especificaciones sobre tareas determinadas, por ejemplo, ontología del proceso de producción de hidrocarburos, de la red eléctrica, de barcos, etc.
- Ontologías de Tareas o Técnicas básicas: Describen una tarea, actividad o artefacto, por medio de sus componentes, procesos o funciones. Este tipo de ontología describe las actividades, esta característica puede resultar útil en las organizaciones.
- Ontologías de Aplicación: Describen los conceptos conforme a un dominio determinado y a tareas concretas; generalmente son una especialización de ambas. Son utilizadas por la aplicación para procesos de producción, diagnóstico de fallas, de diseño intermedio, etc.

3. Granularidad conceptual de las categorías representadas

Esta clasificación se refiere a la profundidad del detalle o especificidad de los conceptos representados, y encontramos dos tipos de ontologías:



- Ontologías Livianas o Ligeras: Incluyen los conceptos, las taxonomías de los conceptos, las relaciones y las propiedades que los describen. Constituyen una descripción limitada ya que no incluyen axiomas ni restricciones, por esto son inconvenientes para la reutilización de conocimiento.

Una ontología ligera permite integrar fuentes de información heterogéneas ya que sus conceptos pueden asociarse a diferentes términos y responder a diferentes intereses al mismo tiempo: clínicos, de investigación, epidemiológicos, económicos, formativos, etc.

- Ontologías de Peso o Pesadas: Se caracterizan por su alta axiomatización. Buscan capturar una perspectiva clara de la realidad. Incluyen la definición de las propiedades de los conceptos, axiomas y restricciones explícitas; también incluyen subconceptos y conceptos hermanos. Las restricciones entre los conceptos relacionados incrementan el control de la ambigüedad. Las ontologías pesadas son de mucha utilidad para Sistemas de Toma de Decisiones y como recurso para la PLN ya que pueden ser utilizadas en los procesos de razonamiento (inferencia y descubrimiento de nuevo conocimiento) definidos en la base de conocimiento.

La figura 2.2 muestra algunas de las propiedades de las Ontologías Ligeras frente a las Ontologías Pesadas.



Figura 2.2. Ontologías Ligeras Vs Ontologías Pesadas

2.3.4. Herramientas para soportar Ontologías

Los editores de ontologías son programas que permiten crear, diseñar y modificar ontologías, incluyendo conceptos y relaciones pertenecientes a dominios específicos, así como propiedades e instancias de las Clases. Muchos de estos editores también permiten analizar la ontología a través de un entorno gráfico que representa las relaciones existentes entre sus conceptos. Estas herramientas pueden ser organizadas dependiendo de las características que presenten. Se puede clasificar a los editores según las siguientes categorías:

- Desarrollo de ontologías
- Evaluación de ontologías



- Combinación e Integración de ontologías
- Herramientas de anotación basadas en ontologías
- Almacenamiento y consulta de ontologías
- Aprendizaje sobre ontologías
- Librerías o *APIs* de programación
- Herramientas de desarrollo Web basado en ontologías
- Herramientas de desarrollo de Servicios Web Semánticos

Es evidente, que la mayoría de las herramientas de gestión de ontologías presentan varias de estas categorías disponibles. A continuación describimos las principales características de los editores de ontologías más utilizados:

- **Protégé**

Esta herramienta, desarrollada en la Universidad de Stanford, es un software libre y de código abierto. Su plataforma soporta dos formas de modelar ontologías; una es a través de los Marcos de Protégé (Protégé-Frames), y la otra utilizando los editores OWL del programa. Una ontología creada en Protégé puede ser exportada en varios formatos, incluyendo RDF, OWL y XML Schema.



Protégé está basado en Java y cuenta con varias extensiones disponibles para descargar desde su página principal⁴, estas extensiones convierten al programa en un entorno flexible para el desarrollo de aplicaciones.

En la actualidad, Protégé tiene más de 100,000 usuarios registrados, y cuenta con una fuerte comunidad de desarrolladores y usuarios académicos, corporativos y de gobierno; mismos que lo utilizan para soluciones de conocimiento en diversas áreas cómo la biomedicina, integración de inteligencia y modelos corporativos.

- **Ontolingua**

Proporciona un entorno distribuido y colaborativo para la creación, edición, modificación, navegación y utilización de ontologías mediante la Web. También incluye una API de integración de ontologías del servidor. ^[11]

Ontolingua está disponible en la Web y se accede desde cualquier navegador estándar, para poder utilizar sus servicios solo es necesario registrarse como usuario e iniciar sesión en el servidor.

⁴ <http://protege.stanford.edu/>



- **OntoEdit**

Forma parte del software OntoStudio, basado en la tecnología OntoBroker de la Universidad de Karlsruhe, Alemania. Esta herramienta de edición de ontologías apoya el desarrollo mediante un entorno gráfico y un entorno Web; y permite la representación semántica de lenguajes conceptuales y estructurales mediante conceptos, jerarquías, relaciones y axiomas.

Existen varias versiones del programa, desde la 2.0 hasta la 2.6. Cada una de estas con una versión libre y una profesional. Algunas de las características más relevantes de la última versión de OntoEdit son: Importar las estructuras del directorio, importar tablas de Excel, construir reglas gráficas, y; visualizar y editar ontologías en un gráfico.

- **KAON**

El software KAON o KARlsruhe Ontology fue desarrollado en la Universidad de Karlsruhe, Alemania y es una herramienta de libre acceso que puede descargarse desde su sitio Web⁵.

⁵ <http://kaon.semanticWeb.org/download>



KAON ofrece una infraestructura de manejo de ontologías enfocada a las aplicaciones de negocios. Incluye herramientas amigables al usuario y que permiten la creación y manipulación intuitiva de ontologías y de aplicaciones basadas en ontologías.

Un importante aspecto de KAON es su enfoque en la integración de las tecnologías tradicionales para la gestión y aplicaciones de ontologías que se usan en las aplicaciones comerciales, como son las bases de datos de relaciones.

- **WebOnto**

Es una herramienta de software libre desarrollada por el Knowledge Media Institute del Reino Unido. WebOnto es un componente Java que cuenta con un servidor Web personalizado y que permite navegar en modelos de conocimiento en la Web. Está construida como parte de los proyectos PatMan, HCREMA y Enrich. Además de estos proyectos, WebOnto también se usa en PlanetOnto y ScholOnto.

- **Neon ToolKit**

Es un editor de ontologías gratuito y de código abierto. Cuenta con varias extensiones y módulos y esta basado en el editor de ontologías OntoStudio, en el



entorno ECLIPSE. Se puede descargar de la página <http://neon-toolkit.org/wiki/Download>.

Neon ToolKit permite modelar ontologías y trabajar con dos tipos de lenguajes ontológicos: OWL y *F-Logic*. El entorno dispone de diferentes perspectivas, vistas y plugins que amplían su funcionalidad, por ejemplo la Perspectiva OWL, OWL Ontology Visualization y utilidades varias para FLOGIC.

La ontología a utilizar para el desarrollo del prototipo de búsqueda semántica, se encuentra realizada y modificada en esta herramienta.

2.4. Técnicas de utilización del Conocimiento

La Web semántica mantiene los principios que han hecho un éxito de la Web actual, como son los principios de descentralización, compartición, compatibilidad, máxima facilidad de acceso, contribución, y la apertura al crecimiento; todo esto relacionado con la información que contiene la Web. A esto se incluyen las nuevas tecnologías para lograr el objetivo de incorporar metadatos en la misma, y convertir esta información en conocimiento.



En esta sección detallaremos algunas de las técnicas para representación y adquisición del conocimiento, presentes en la implementación de la Web Semántica.

2.4.1. Representación del Conocimiento

Estas tecnologías incluyen lenguajes para la representación de Ontologías como los que ya se mencionaron en el capítulo anterior, librerías, parsers, lenguajes de consultas sobre Ontologías, entornos de desarrollo, conversión de Ontologías, y módulos de visualización, entre los más relevantes.

Retrocediendo un poco en el tiempo, nos encontrarnos con el primer lenguaje para la construcción de la Web Semántica: SHOE (Simple HTML Ontology Extension), desarrollado en 1997 por Jim Hendler en la Universidad de Maryland. Se trataba de un lenguaje de representación del conocimiento diseñado para la Web, y permitía asociar el contenido de páginas Web con una o varias ontologías.

Desde entonces, y hasta la actualidad, se han desarrollado muchas tecnologías como XML^[12], XMLS^[13], DAML + OIL^[14] y recientemente OWL^[15]. Todos incluyen mecanismos que ayudan a transformar la Web en una infraestructura global en la que es posible compartir, y reutilizar datos y documentos entre diferentes tipos de usuarios.



Tal como podemos apreciar en la figura 2.3, los lenguajes de marcado sobre los que se asienta la Web Semántica están basados en XML.

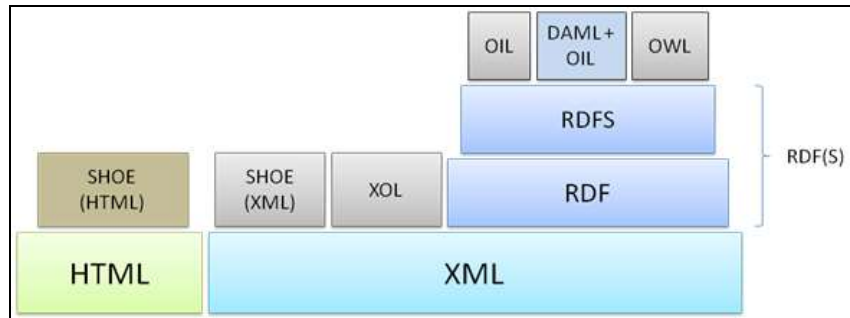


Figura 2.3. Lenguajes de marcado de Ontologías

Es evidente que el desarrollo de la Web Semántica requiere la utilización de estos lenguajes, de manera que puedan dotar de lógica y significado al contenido de cada página, archivo y recurso de la Web.

Las tecnologías de la Web semántica están organizadas de acuerdo a la manera en que se identifica, clasifica y visualiza la información, comenzando desde el enfoque básico en el que se concibió la Web (objetos), hasta lo que se pretende implantar hoy en día que tiene que ver con la adquisición de conocimiento. La organización es la siguiente:

- Web de datos / objetos: HTML / XML
- Web de información / relaciones: RDF
- Web de conocimiento / razonamiento: OWL



En la Figura 2.4, se muestra la evolución de la Web, desde el punto de vista de las tecnologías que se le han incorporado para incrementar su potencialidad.

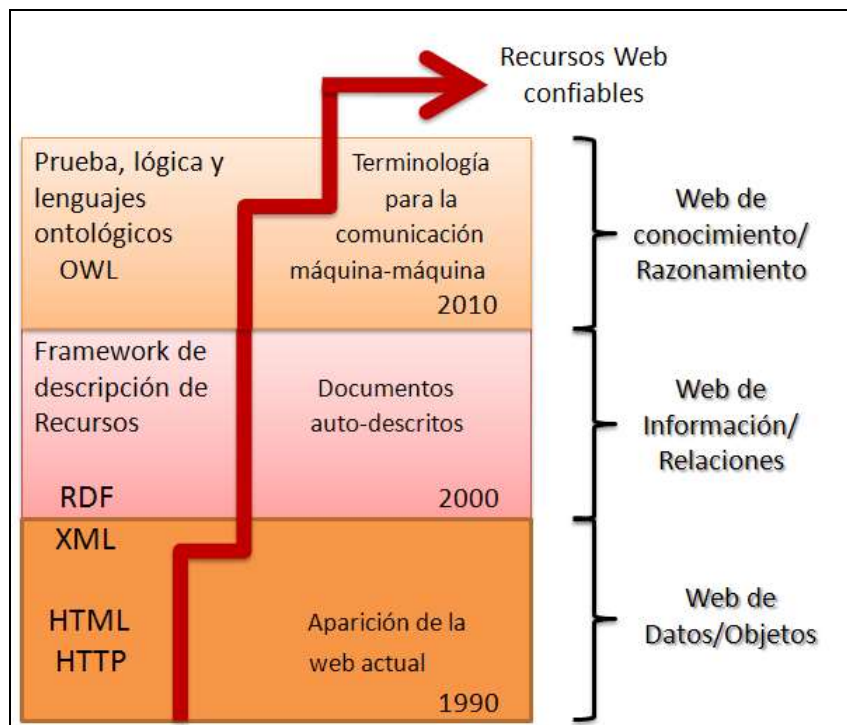


Figura 2.4. Evolución de la Web y tecnologías involucradas⁶

2.4.2. Adquisición de Conocimiento

La adquisición del conocimiento en los Agentes Inteligentes de la Web, tiene su base en el concepto filosófico de la “Epistemología”.

⁶ Berners-Lee y Handler, 2001



La epistemología busca entender las definiciones y relaciones entre lo que se conoce y el objeto conocido. Las diferentes teorías filosóficas de la adquisición de conocimiento se construyen en base a referentes epistemológicos.

Estos conceptos filosóficos se traspasan a los sistemas Inteligentes, mismos que necesitan una estructura categórica de la información que pretenden utilizar y analizar. Esta estructura de conocimiento debería reflejar hechos y objetos del mundo real. Todo esto es posible gracias a la aplicación de las técnicas de la Ingeniería Ontológica, y de métodos como la inferencia, razonamiento y aprendizaje, mismos que describiremos a continuación.

Ingeniería Ontológica

La ingeniería Ontológica abarca el conjunto de actividades referentes al proceso de desarrollo de una ontología, el ciclo de vida de una ontología, metodologías, herramientas y lenguajes necesarios para la construcción de ontologías.

La Ingeniería Ontológica nos proporciona factores claves como son:

- Nos proporciona un lenguaje consensuado al considerar polisemias y sinónimos, reduciendo la ambigüedad terminológica.



- Reutilización de conocimiento. Nos permite el aprovechamiento de ontologías realizadas sobre cualquier área, reflejando cualquier dominio del mundo a través de la utilización de las mismas.
- Traducción e intersección semántica. Esto logrado a partir de los *mappings*, es decir las correspondencias entre diferentes ontologías.

Estos factores han hecho posible que la Ontología sea presentada como una solución factible para la interoperabilidad semántica, originando una mejora notable en la búsqueda y presentación de la información.

La manera en la que hoy se puede representar el conocimiento depende en gran medida de los componentes que se puedan reutilizar, pudiendo enfocarse específicamente en construir conocimientos especializados y nuevos razonadores de acuerdo a las tareas específicas de las aplicaciones. Permitiendo que de esta manera el conocimiento declarativo, las técnicas de solución de problemas y los servicios de razonamiento sean todos compartidos entre sistemas.

Inferencia, Razonamiento y Aprendizaje

Un proceso de razonamiento implica la realización de inferencias a partir de un conjunto de premisas. Y, a través del razonamiento, un sistema es capaz de aprender y adquirir nuevos conocimientos. Ya que la naturaleza de un axioma no



puede cambiar o ser modificado; el conjunto de axiomas crece en tamaño de manera continua. Todo esto es posible gracias al motor de inferencia de los SBC⁷. Algunos de los principales métodos de inferencia son la deducción, la abducción y la inducción.

- La **deducción** parte de un principio conocido hacia uno desconocido y de reglas generales a situaciones específicas, es decir que va desde la premisa hasta la conclusión lógica; de esta manera garantiza el resultado de conclusiones verdaderas a partir de premisas verdaderas.
- El método de **abducción** trata de encontrar afirmaciones que soporten una conclusión verdadera con la que se cuenta previamente. Es decir que va de lo general a lo particular. Este método se utiliza para generar explicaciones pero no garantiza conclusiones verdaderas.
- La **inducción** es el razonamiento a partir de hechos particulares hacia una conclusión general; es la base de la investigación científica pero no es un método garantizado de inferencia.

Una ontología puede utilizar Inferencia durante el proceso de instanciación en el cual se puede generar una instancia llenando automáticamente sus slots, gracias a los métodos automáticos de adquisición del conocimiento.

⁷ Sistemas Basados en Conocimiento



2.5. Buscadores Semánticos

Un buscador semántico es un programa que entiende las búsquedas del usuario y los textos de la Web utilizando algoritmos de simulación de comprensión del lenguaje, y que proporciona resultados correctos sin que el usuario tenga que inspeccionar los documentos por sí mismo.

2.5.1. Buscadores Semánticos: ¿Para qué sirven?

El objetivo de utilizar Buscadores Semánticos es clasificar, estructurar y registrar los recursos Web, procesando automáticamente el contenido semántico de la Web, y de esta manera compartir y facilitar el acceso al conocimiento.

La tecnología de un Buscador Semántico permite clasificar la información de manera formal de acuerdo al significado de los conceptos dentro de un dominio, y permite también las siguientes aplicaciones:

- **Recuperación de información:** Las búsquedas semánticas trabajan con el significado de las palabras asegurando una alta precisión en las búsquedas. Un Buscador Semántico puede relacionar la información contenida en páginas Web, procesarla, separar la información más fiable, e



incluso deducir información no registrada tomando decisiones con un cierto grado de autonomía. El resultado es la información solicitada en forma de conceptos del dominio.

- **Respuesta a preguntas desde una Ontología:** Un buscador semántico puede recibir una consulta por parte de un usuario y buscar la respuesta dentro de una ontología, para esto relacionará los términos de la consulta con los conceptos que contiene dicha ontología. Esta es la aplicación en la cual estará basado el prototipo de buscador semántico del presente trabajo.
- **Publicación de la información:** Permite la navegación y la presentación de la información de acuerdo a su contenido, de manera que el usuario pueda visualizar los conceptos del modelo y aprovecharlos. También incorpora Interfaces inteligentes basadas en lenguaje natural y la posibilidad de formular consultas en un lenguaje cercano al natural.
- **Sistema de inferencia y agregación de información:** Un buscador Semántico hace posible validar y aumentar la información mediante sistemas de inferencia automáticos, mismos que podrán realizar inferencias automáticas al buscar información relacionada con la que se encuentra situada en las páginas.



- **Intercambio de información:** Se puede intercambiar información en diferentes formatos para aplicaciones específicas. Esta posibilidad de traducir la información entre formatos, permite aumentar la rentabilidad de la codificación de la misma. Actualmente el gasto de hacer compatibles a sistemas heterogéneos representa un 30% del gasto de toda la industria de tecnologías de la información.

Los agentes Web están siendo diseñados para tratar la información situada en las páginas Web de manera semiautomática con la finalidad de que los datos puedan ser utilizados y “comprendidos” sin necesidad de la intervención humana. Con esto, se mejorará la búsqueda de información y se potenciará el desarrollo de aplicaciones de comercio electrónico, ya que las anotaciones de información seguirán un esquema común, y los buscadores Web compartirán los mismos esquemas con estas anotaciones. Así por ejemplo, diferentes empresas que traten con clientes y proveedores, podrán intercambiar sus datos de productos siguiendo esquemas afines.



2.5.2. Buscadores Semánticos: ¿Cómo funcionan?

Para que un Buscador Semántico pueda trabajar adecuadamente; la información publicada en la Web, debe contener información adicional que agregue un sentido lógico a los datos; indicando la semántica, la explicación de otros datos a los que hacen referencia, y el contexto en donde se encuentran.

Como ya se mencionó anteriormente, ésta información adicional, se conoce como “*metadatos*”, y representa una explicación al contenido que se desea publicar; también permite que un programa pueda interpretar los contenidos de forma rápida, sencilla y sin supervisión humana. Los programas capaces de interactuar con los metadatos se conocen como “Agentes” y pueden ser enviados hacia la Web en búsqueda de páginas que contengan la información exacta de lo buscado. De esta manera se pretende obtener una concordancia semántica del tema buscado y obtener una búsqueda refinada y acertada.

El punto fundamental de la Web Semántica es convertir la estructura de la Web, del actual almacén de datos en una estructura de almacenamiento de información. Para transformar la Web 2.0 en Web Semántica, es necesario apoyarse en dos conceptos fundamentales:

- La descripción del significado que tienen los contenidos en la Web.



- La manipulación automática de estos significados.

La descripción del significado de los contenidos Web consta de los siguientes elementos:

- La Semántica, en este caso, el significado de los datos procesable por máquinas.
- Los Metadatos, que permiten convertir la información en conocimiento, referenciando datos dentro de las páginas Web a metadatos con un esquema común consensuado sobre algún dominio. Los metadatos no sólo especificarán el esquema de datos que debe aparecer en cada instancia, sino que además podrán obtener información adicional de cómo hacer deducciones con ellos.
- Las Ontologías, que son el conjunto de conceptos y relaciones entre ellos que describen un dominio de aplicación concreto. El propósito de una ontología es la descripción de un dominio mediante un vocabulario común de representación.

La manipulación automática de los contenidos se hace a través de las siguientes herramientas:

- La lógica matemática, que permite establecer reglas para tratar el contenido semántico.



- Los motores de inferencia, que permiten combinar información conocida para elaborar nuevos conocimientos.

La Web y la información contenida en ella, deben contener una estructura lógica; esta estructura se conoce como Red Semántica, y consiste en un conjunto de redes similares de conceptos interdependientes. Es decir, es un diagrama en el cual se relacionan y se van descomponiendo los conceptos en otros. A continuación se presenta un ejemplo de una red semántica para un Vehículo.

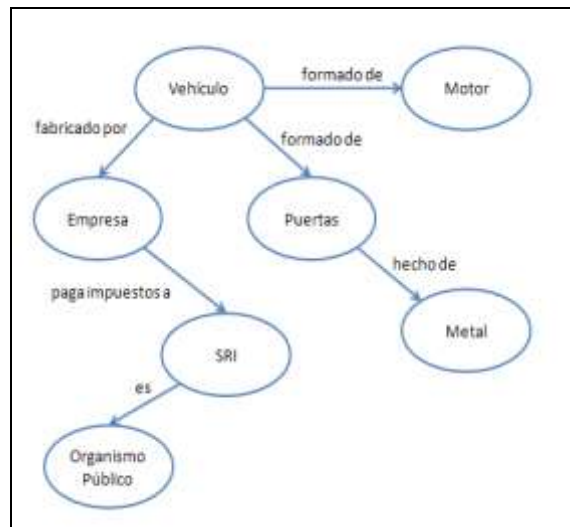


Figura 2.5: Red Semántica del dominio Vehículo

En el ejemplo de la Figura 2.5, partimos del concepto “Vehículo”, el cual está formado de motor y puertas, fabricadas de metal. A su vez el vehículo es fabricado por una Empresa, la cual debe de pagar impuestos a la SAT, una entidad del



gobierno. Gracias a este ejemplo, se puede observar claramente cómo a partir de un concepto se van creando las relaciones con otros conceptos y poco a poco se va profundizando en el tema de un dominio dado. Así se hace posible interpretar una red semántica. El propósito de la Web Semántica es almacenar toda la información de esta manera para crear una Red de Conocimiento general que pueda ser recorrida e interpretada automáticamente.

Las búsquedas semánticas son muy superiores a las búsquedas basadas en palabras clave, pero la complejidad de la búsqueda semántica hace que de momento esté restringida a grupos de investigación y empresas con una alta inversión.

Un Buscador Semántico se basa en algoritmos que simulan la comprensión de las palabras, utilizando sistemas PLN⁸ y estableciendo relaciones entre las palabras para realizar búsquedas de interés. Es posible, que en varias ocasiones los documentos devueltos no contengan las palabras o expresiones de búsqueda si no uno o varios sinónimos de las mismas.

En general el proceso de búsqueda semántica es el siguiente:

⁸ Procesamiento de Lenguaje Natural



1. Interpretar la pregunta del usuario extrayendo los conceptos más relevantes de la frase.
2. Utilizar este grupo de conceptos para crear una consulta y utilizarla contra la ontología del sistema.
3. Presentar los resultados al usuario.

Siguiendo estos pasos, la búsqueda semántica determinará cuáles son las páginas más relevantes para una búsqueda en base los conceptos que contienen, en lugar de aplicar métodos estadísticos sobre el texto.

2.5.3. Mecanismos utilizados en el proceso de búsqueda Web

El proceso de búsqueda, abstracción y clasificación de la información esta a cargo de aplicaciones Web conocidas como Agentes Inteligentes. Estas aplicaciones son capaces de interpretar las anotaciones y esquemas ontológicos en las páginas Web, además de reconocer los dominios a los que pertenece la información. También son capaces de realizar búsquedas con inferencias utilizando los axiomas que están situados en los esquemas, y de realizar ligaduras entre los árboles taxonómicos de varias ontologías.



Todo esto es posible gracias a la utilización de Ontologías; cuyos elementos describen e interrelacionan los conceptos de un dominio y por lo tanto son esenciales para poder realizar búsquedas. Las Ontologías permiten que los Agentes comprendan la estructura lógica que representa la información. Así, hacen posible realizar búsquedas un poco más refinadas que toman en cuenta las estructuras jerárquicas de los conceptos, y sobretodo, que son capaces de inferir conocimiento a partir de los axiomas y de acceder a una mayor cantidad de información.

La figura 2.6 nos muestra la ubicación de las ontologías dentro de la pila de componentes de la Web Semántica; entre los cuales también encontramos XML, XML Schema, RDF, RDF Schema y OWL (mismos que se tratan en el capítulo 3); todos estos elementos interactúan para potenciar las búsquedas de información en la Web, ya sean estas realizadas por un usuario o por un Agente Inteligente.

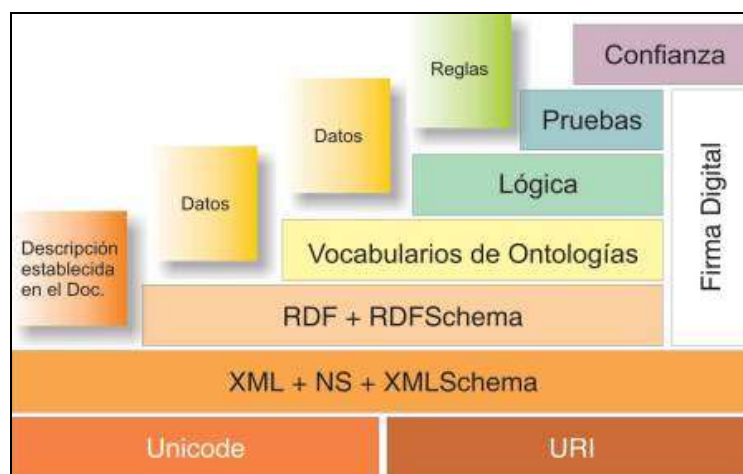




Figura 2.6. Pila de la Web Semántica

En el contexto de la Web Semántica, las ontologías suponen una capa por encima de RDF y los RDFSchemas ya que dependen de estos. XML y el lenguaje RDF pueden agregar una lógica de significado en cada página, archivo, recurso o contenido de la Web; esto permite que los buscadores conozcan el significado de la información que manejan, todo con el fin de que ésta información pueda ser presentada en pantalla, y sobretodo que pueda ser integrada y reutilizada.

La capa de “Lógica” abarca todo proceso de búsqueda y extracción de información desde las capas inferiores; esta capa puede utilizar técnicas distintas dependiendo del tipo de información a extraer y de la finalidad de la búsqueda; entre las técnicas más conocidas podemos mencionar la Representación del conocimiento, Extracción del Conocimiento, Gestión de Conocimiento y Minería de Datos.

Todos estos conceptos se aplican en las búsquedas Semánticas en la Web, pero también se pueden plasmar para búsquedas Semánticas en aplicaciones de escritorio como se hace en nuestro trabajo.

2.5.4. Tipos de Buscadores Semánticos

Los Buscadores Semánticos, también conocidos como buscadores de tercera generación, no son más que buscadores de ontologías que buscan ficheros RDF, micro formatos y OWL en el entorno Web. Los Buscadores Semánticos se



encuentran clasificados en dos categorías de acuerdo a su utilización: Buscadores Semánticos no orientados a usuarios y Buscadores Semánticos orientados a usuarios.

Buscadores Semánticos no orientados a usuarios

Los Buscadores Semánticos no orientados a usuarios son de gran utilidad para usuarios expertos pero no para usuarios finales, ya que buscan información representada de manera formal o técnica en la Web, es decir que busca ontologías, ficheros rdf, entre otros. Entre este tipo de buscadores destacan los siguientes:

- NaturalFinder: Es un complemento que puede agregarse a cualquier buscador para Internet e intranets y permite realizar consultas en lenguaje natural (español o inglés). ^[16]
- Blinkx: Su función es buscar vídeos mediante un proceso de reconocimiento de imagen y voz que se encuentran en su base de datos, así como información oculta y escrita de los mismos.
- Watson: Desarrollado por Knowledge Media Institute (KMi) de la Open University (UK), es un buscador de ontologías y documentos semánticos, que muestra un listado de las ontologías existentes en la Web de acuerdo al dominio buscado.^[17]



- Swoogle: Desarrollado en la Universidad de Maryland, Baltimore County. Este buscador de ontologías proporciona documentos en RDF, OWL, etc. Pero no está orientado a proporcionar documentos HTML que son los documentos que requieren los usuarios finales. ^[18]

Buscadores Semánticos orientados a usuarios

Entre los Buscadores Semánticos orientados a usuarios más importantes tenemos:

- Naveganza: Es un buscador español basado en la comprensión de las preguntas tecleadas por los usuarios y en los documentos indexados. Naveganza va dirigido a sectores concretos, empresas o administraciones que tengan buena parte de sus datos organizados bajo los criterios de Web Semántica ^[19]
- Hakia: Es un Buscador Semántico de propósito general. La calidad de los resultados cumple con tres criterios: fuentes dignas de confianza, información reciente e información relevante para consulta. ^[20]
- Swotti: Utiliza tecnologías de la Web Semántica para extraer las opiniones que realizan los usuarios en blogs y foros sobre empresas o productos. ^[21]



- Reiver: Es el buscador original de imágenes en *Flickr*. La búsqueda se puede hacer de dos formas: dibujando en una pantalla la forma y los colores que se desea o subiendo una fotografía que sirva de modelo. [22]
- True Knowledge: No solo proporciona respuestas a una pregunta concreta si no que ofrece enlaces con información relacionada a la misma. [23]
- Bing: El nuevo buscador Bing utiliza ontologías para identificar conceptos y asociaciones entre ellos, y de esta manera relacionarlos con la consulta del usuario. Bing combina tecnología de búsqueda semántica y minería de datos. [24]

2.5.5. Buscadores Semánticos: Limitaciones para su aplicación

El desarrollo y difusión masivos de los Buscadores Semánticos, dentro del marco de la Web Semántica, tienen algunas dificultades que aparecen con respecto a tecnología o relacionadas con la falta de interés de los propietarios de sitios Web. A continuación presentamos un resumen de las limitaciones más notables:

- Actualmente la Web está basada, en su mayoría, en el lenguaje HTML, que no ofrece herramientas para categorizar los elementos que forman un texto.



Es por esto que la información no se encuentra relacionada entre sí, ni esta almacenada en una estructura de conceptos lógicos.

- Por la falta de estructura y anotaciones semánticas en la Web actual, los Buscadores Semánticos se ven obligados a analizar los recursos palabra por palabra y oración por oración mediante algoritmos cognitivos para asignar las palabras y oraciones a conceptos ontológicos.
- Se necesita un lenguaje común basado en Web, con suficiente capacidad expresiva y de razonamiento para representar la semántica de las ontologías; este lenguaje debería permitir que las reglas de cualquier sistema existente de conocimiento, sean exportadas sobre la Web.
- Hace falta un componente tecnológico que permita convertir de forma automática el contenido de texto de las páginas Web en marcas OWL. Actualmente existen herramientas y páginas Web que realizan este tipo de traducción pero se encuentran en etapa de desarrollo y todavía no equiparan la eficiencia de una anotación manual.
- El modelo de negocio de muchas páginas Web consiste en obtener ingresos de la publicidad. Estos ingresos son posibles únicamente si sus



páginas son visitadas por una persona, y se pierden si los datos quedan disponibles para que los interprete y extraiga un Agente Inteligente.

La Web Semántica representará un importante adelanto para el potencial de la Web, desarrollando aplicaciones con esquemas de datos comunes, fomento de las transacciones entre empresas y búsqueda de información con inferencias. Pero para poder lograr estos objetivos se necesita unificar los contenidos Web y convertirlos en contenidos Semánticos, formalizando así el conocimiento de manera consensuada y reutilizable.

Aún queda camino por recorrer en el campo de la Web Semántica; el Buscador Semántico ideal, capaz de responder preguntas de toda índole en lenguaje natural, está muy lejos; pero la investigación e implementación de nuevas tecnologías, que se acercan a este ideal, avanzan día con día.

2.6. Transición a la Web Semántica

La Web Semántica busca enriquecer la comunicación mediante metadatos semánticos que aporten un valor añadido a la información, la diferencien y la hagan más inteligente. No obstante, los avances de esta tecnología son lentos y



difíciles. La solución podría estar en la combinación de las técnicas de inteligencia artificial con el acceso a la capacidad humana de realizar tareas muy complejas para un ordenador como la de relacionar, seleccionar, y sintetizar información.

El paso de la Web 2.0 hacia la Web 3.0 no es responsabilidad única de los administradores de grandes motores de búsqueda o eminentes investigadores tecnológicos encargados de implementar la inteligencia en la Web; también es una tarea de todos sus usuarios y todos y cada uno de los propietarios de sitios Web.

Para que la transición hacia la Web Semántica se concrete, es necesario que cada página Web contenga sus etiquetas semánticas respectivas que la relacionen con una o varias ontologías de dominio. Si un Sitio Web no cumple con este requerimiento, no solo se vería excluido de la Web 3.0, si no que podría llegar a entorpecer su funcionamiento y su finalidad, que es la de entrelazar la información, organizando y compartiendo conocimiento.

Hemos visto que las herramientas de anotación ayudan a mantener anotaciones sobre sitios Web pero, incluso con la ayuda de estos programas, la responsabilidad de transformar los sitios Web en sitios con etiquetas semánticas, seguirá siendo de los propietarios de las páginas Web. Por esta razón es necesario concientizar a la comunidad acerca de la importancia de la dotación de



significado a la Web, y a la vez, se puede recompensar la publicación de contenido semántico, para que los sitios Web puedan posicionarse en buscadores y otros servicios de búsqueda.

Se puede decir que la Web llegará al estado de Web 3.0 cuando la generación de contenido semántico sea automática, y exista el procesamiento automático de lenguaje natural. Actualmente existen programas etiquetadores que permiten añadir contenido semántico a sitios Web enteros, pero estos programas y páginas Web se encuentran en etapa de prueba. Se espera que en un futuro cercano se pueda llevar a cabo la tarea de relacionar información semántica en los sitios Web de manera automática; facilitando así, la transición hacia la Web 3.0.

CAPÍTULO III



PLATAFORMA TECNOLÓGICA DEL BUSCADOR SEMÁNTICO



En el tercer capítulo, PLATAFORMA TECNOLÓGICA DEL BUSCADOR SEMÁNTICO, se exponen las tecnologías que dan soporte al prototipo de Buscador Semántico implementado en este trabajo. En primer lugar se explica el proceso de desarrollo de una ontología, inmediatamente se tratan las metodologías existentes para construir una ontología y las diferencias entre las mismas. Como tercer punto se explican los lenguajes de representación semántica existentes. El cuarto punto trata de los Lenguajes de consultas, para continuar con un estudio del Procesamiento de Lenguaje Natural en quinto lugar; y terminar con una exposición general de las APIs e interfaces de soporte semántico en el sexto punto.

CAPÍTULO 3

PLATAFORMA TECNOLÓGICA DEL BUSCADOR SEMÁNTICO



3.1. Desarrollo de Ontologías

Un punto clave para el avance en cuanto a las búsquedas Semánticas, es el rol que cumplen las Ontologías, qué, como vimos en el capítulo 2, permiten representar el conocimiento de forma consensuada, legible por los ordenadores, y reutilizable.

Tales aspectos deben ser el punto de partida en el diseño de una Ontología, de tal manera que se proporcione un vocabulario común que no solamente sea de utilidad para una aplicación en concreto si no que permita compartir el conocimiento en un sin número de ámbitos y aplicaciones.

Por estas razones es indispensable contar con una metodología que nos brinde soporte durante todo el proceso de representación del dominio.

Las ontologías aplicadas en la Web van desde grandes taxonomías que se utilizan en la categorización de sitios Web tales como Yahoo!, hasta categorizaciones de productos para ventas, tales como Amazon.com.

Algunas de las razones para crear una ontología se encuentran en la Guía para Desarrollar Ontologías de la Universidad de Stanford^[25], y son las siguientes:



- Compartir el entendimiento común de la estructura de información entre personas o agentes de software.
- Permitir la reutilización de conocimiento de un dominio.
- Explicitar suposiciones de un dominio.
- Analizar el conocimiento de un dominio; en especial esta es la razón por la cual desarrollaremos nuestra ontología.

El primer tópico, compartir el entendimiento común de la estructura de información entre personas o agentes de software, es quizás el más importante. Por ejemplo, supongamos que distintos sitios web contengan información turística de la ciudad de Cuenca; si estos sitios web comparten y publican la misma Ontología subyacente de los términos que usan, entonces, los agentes de software podrían extraer y agregar información de esos sitios diferentes, y a la vez podrían usar esta información para responder diversas solicitudes de los usuarios. Sin embargo, y por motivos de restricciones⁹, esta funcionalidad no está incluida en el prototipo de buscador semántico del presente trabajo.

En muchas aplicaciones semánticas, no es necesario desarrollar una Ontología en concreto, puesto que se pueden utilizar ontologías existentes. La reutilización de conocimiento de un dominio reduce significativamente el trabajo en la

⁹ Restricciones, Capítulo 1, Sección 1.8



investigación, ya que permite integrar y compartir esa información. Por ejemplo, si un grupo de investigadores desarrolla una ontología detallada, otros investigadores y desarrolladores podrían reusarla para dominios incluidos dentro de esta. Además, si se presenta la necesidad de construir una ontología grande, simplemente sería posible integrar varias ontologías existentes que describen porciones del dominio más grande. También es posible reutilizar una Ontología general y extenderla hacia nuestro dominio de interés.

La separación del conocimiento del dominio, del conocimiento operacional es otro uso común de las ontologías. Podemos describir la tarea de configuración de un producto a partir de sus componentes de acuerdo a especificaciones requeridas e implementar un programa que hace que esta configuración sea independiente de los productos y los componentes en sí.

En todos estos casos es recomendable describir explícitamente las suposiciones de un dominio al momento de su implementación, ya que esto facilitará las modificaciones sobre las mismas y el entendimiento de los términos de la ontología por parte de nuevos usuarios.

Por último, es posible analizar el conocimiento de un dominio una vez que una especificación declarativa de los términos se encuentra disponible. El análisis



formal de los términos es el cimiento fundamental al reusar ontologías existentes y al extenderlas.

Implementar una ontología en sí, no es un tema complicado si se definen claramente todos los aspectos que regirán a lo largo de su desarrollo. Básicamente es necesario tener claro el dominio en que se va a desarrollar la ontología, y seguir las reglas que guían el desarrollo de una Ontología.

3.1.1. Reglas Fundamentales en el desarrollo de Ontologías

Antes de pasar al proceso de desarrollo de una ontología, es indispensable conocer las reglas o conceptos fundamentales que regirán a lo largo del proceso de diseño y desarrollo. Las siguientes reglas son de gran utilidad para tomar decisiones de diseño en cualquier momento del proceso de desarrollo:

- No existe una forma correcta de modelar un dominio, siempre existen varias alternativas viables. La mejor solución siempre dependerá de la aplicación específica que se tenga en mente.
- El desarrollo de una Ontología es necesariamente un proceso iterativo.



- Se debe tomar en cuenta que los conceptos de la Ontología deben ser cercanos a los objetos (físicos y lógicos) que se van a definir, probablemente estos serán sustantivos o verbos en oraciones que describirán el dominio de uso.

Estas reglas, a más de ser una ayuda para el desarrollador, nos servirán para conocer el uso específico que tendrá a la Ontología y para saber cuan detallada podrá llegar a ser, ya que; al tratarse de un proceso iterativo, es posible perfeccionar la ontología desde una versión inicial hasta el nivel de detalle que se quiera conseguir. La creación de ontologías sigue un conjunto de principios de creación, mismos que varían de acuerdo a la metodología aplicada.

3.2. Metodologías para construir una ontología

Dada la importancia de las ontologías en la Web Semántica, se ha visto necesario contar con metodologías para el desarrollo colaborativo y contextual de las redes de ontologías.

Existen varias metodologías para el desarrollo de ontologías, dependiendo del grado de dificultad y especificación en su aplicación. Estas metodologías describen los procedimientos de desarrollo y validación de ontologías, brindando



soporte durante todo el proceso de representación del dominio. Toda metodología, sin importar su naturaleza específica, cuenta con los siguientes elementos:

- **Ciclo de vida:** Comprende todos los procesos que se realizan desde la concepción, construcción y mantenimiento de la ontología.
- **Herramientas de soporte:** Todas estas metodologías deben ser soportadas por herramientas Ontológicas, teniendo claro el nivel de expresividad al pasar de lo informal a lo formal.
- **Nivel de abstracción:** Este nivel permite la correcta extracción de conceptos, permitiendo que la Ontología se desarrolle sin exceso de información en ciertos aspectos del dominio.
- **El uso de ontologías base:** Permite analizar otras ontologías existentes y añadirlas a la Ontología con la que se esta trabajando.

3.2.1. Metodologías de desarrollo más destacadas

Una vez analizadas las características de todas las metodologías en general, vamos a definir a grandes rasgos las metodologías más destacadas:

Methontology



Incluye un ciclo de vida de construcción para la ontología que consiste en un conjunto de prototipos evolutivos. Permite agregar, cambiar y renovar términos en cada nueva versión de los prototipos. ^[26]

Metodología CyC

Los pasos para la construcción de la ontología usando esta metodología son: la extracción manual del conocimiento de diversas fuentes y la aplicación de herramientas de procesamiento de lenguaje natural para la adquisición de nuevo conocimiento en la ontología. ^[27]

Metodología Uschold y King

Permite la creación de ontologías en base a otras ontologías ya existentes. Esta metodología se basa en la experiencia de construir el “Enterprise Ontology”, una ontología que incluye un conjunto de conceptos para modelado empresarial. ^[28]

On-To-Knowledge

Se enfoca en las aplicaciones futuras de la ontología. Los pasos esenciales de esta metodología involucran el estudio de viabilidad, la delimitación del



dominio, el objetivo de la ontología y el mantenimiento de la misma a través de todo el proceso de desarrollo. ^[29]

NeOn Methodology

La Metodología NeOn permite la construcción de ontologías basándose en un conjunto de posibles escenarios. En total se trabaja con 9 escenarios distintos que combinan las posibilidades de construir una ontología desde cero o de reutilizar recursos existentes para crearla, a su vez estos recursos podrían ser ontológicos o no ontológicos. También permite combinar cualquiera de estos escenarios entre sí para adecuarse a las necesidades del desarrollador. ^[30]

La figura 3.1 muestra un cuadro de análisis de estas características estas metodologías.



Características Metodologías	Ciclo de vida propuesto	Herramientas de soporte	Modelado	Nivel de abstracción	Uso de Ontología de base
Methontology	Secuencial con prototipos de desarrollo y actividades de administración	ODE Web ODE OntoEdit Protégé 2000	Identificación directa de conceptos y organización taxonómica	Elevado	No
CyC	Secuencial	No especificada	Extracción manual	Alto	No
Uschold and King	No especificada	No especificada	Identificación directa de conceptos	Normal	No
OTK	Incremental y cíclica con prototipos de desarrollo	OntoEdit Corporum	Casos de Uso semilla	Bajo	No
Sensus	No propuesto	No especificada	Semilla	Normal	Si
Neon Methodology	Secuencia e Incremental	No especificada	Preguntas de Competencia	Elevado	Con o sin ontologías base

Figura 3.1. Metodologías de desarrollo de Ontologías ^[31]

Las metodologías existentes: Methontology, Dilligent y On-To-Knowledge, representan un avance en el proceso de construcción de ontologías; pero carecen de directrices para la creación de ontologías dentro de un contexto de redes de ontologías. Además, no se describen con un enfoque orientado al usuario, y no abordan el desarrollo de aplicaciones para la Web Semántica.



Methontology, Dilligent y On-To-Knowledge no presentan una guía para tratar la dinámica y la evolución de una ontología y tampoco presentan una guía detallada para cada uno de los procesos o actividades concernientes a la creación de ontologías.

Por otra parte, la Metodología de NeOn, cuenta con guías más detalladas en cuanto a su aplicación; y es por estas razones que ha sido seleccionada como la metodología para construir la ontología de Turismo en el presente trabajo.

3.3. Lenguajes de representación Semántica

A continuación analizaremos a fondo cada uno de los lenguajes que hacen posible la implementación de nuestra ontología y del buscador semántico.

3.3.1. XML

XML, o Lenguaje de Marcado Extensible (Extensible Markup Lenguaje), es un lenguaje de marcado que se puede utilizar para crear etiquetas propias. Fue creado por el consorcio W3C con la finalidad de superar las limitaciones del lenguaje HTML.



Su principal característica es que las páginas XML pueden ubicar metadatos, esquemas XML, y esquemas RDF, aportando un vocabulario descriptivo para interpretar y comprender los documentos en forma de árboles de etiquetas con atributos. Esta característica representa una primera aproximación a la Web Semántica.

Debido a la gran operatividad que posee, XML ha logrado convertirse en un lenguaje estándar actualmente. Se trata de una evolución del sofisticado SGML¹⁰ que aporta datos estructurados a la Web; además, por la fiabilidad en cuanto a portabilidad se ha convertido en la infraestructura preferida para el intercambio de datos.

Existe una extensión de XML denominado XML Schema (XMLS) y la única diferencia es que se acuerdan de antemano las estructuras a utilizarse.

3.3.2. RDF

La Especificación RDF (Marco para descripción de recursos) se trata de un modelo estándar para el intercambio de información en la Web. RDF tiene características que facilitan la compartición de información, aún bajo esquemas de

¹⁰ Standard Generalized Markup Language o "Lenguaje de Marcado Generalizado"



representación diferentes. El modelo RDF permite la descripción y el procesamiento de los metadatos, pudiendo además describir metadatos de cualquier dominio.

La información manejada es tratada mediante grafos que permiten modelar las relaciones entre los objetos. Una similitud del modelo de RDF podría darse con los lenguajes de programación orientados a objetos, ya que manejan clases y subclases dispuestas en una jerarquía y con un nivel de herencia múltiple.

RDF maneja el modelo de datos en base a tres tipos de objetos:

- Recursos, que son expresiones RDF descritas mediante URI's.
- Propiedades, que describen las características, atributos y relaciones de los recursos.
- Instrucciones, que asignan un valor a una propiedad de un recurso específico.

Una instrucción RDF se basa en la idea de convertir las declaraciones de los recursos en expresiones con la forma sujeto-predicado-objeto, conocidos como tripletas, tal como se puede apreciar en la figura 3.2.

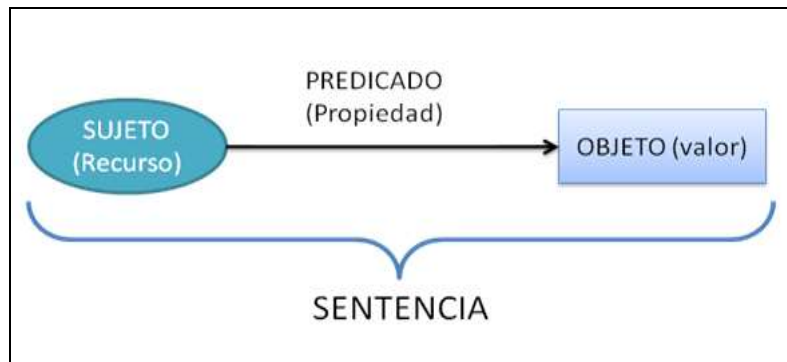


Figura 3.2. Una instrucción RDF

El sujeto es el recurso, aquello que se describe en la instrucción. El predicado es la propiedad o relación que se desea establecer acerca del recurso. Y el objeto es el valor de la propiedad o el otro recurso con el que se establece la relación.

Un ejemplo de instrucciones RDF podría ser el siguiente:

- Juan escribe un artículo;

Donde *Juan* es el sujeto, *escribe* es la relación, y *artículo* es el objeto.

En el ejemplo “*Juan*” y “*artículo*” son recursos denotados por nodos y “*escribe*” es una propiedad denotada gráficamente por un arco, tal como se puede apreciar en la figura 3.3.

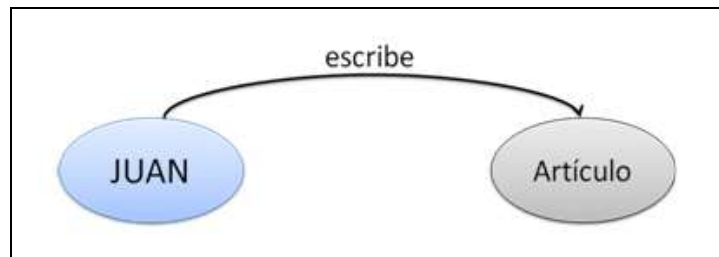


Figura 3.3. Ejemplo de un grafo en RDF

XML aporta una primera aproximación a los metadatos potenciando su poder sintáctico para soportar el desarrollo de RDF y RDFs como herramientas iniciales para compartir datos en la Web. RDF utiliza XML como lenguaje base, de ahí la importancia del mismo.

La combinación de RDF con otras herramientas como RDF Schema y OWL permite añadir significado a las páginas, y es una de las tecnologías esenciales de la Web Semántica.

El modelo de datos RDF no toma conclusiones previas sobre la estructura de un documento que contiene información RDF, permitiendo que las instrucciones puedan aparecer en cualquier orden dentro de una ontología.



3.3.3. RDFS

RDF Schema es un vocabulario para describir las propiedades y clases de los recursos RDF.

Teniendo en cuenta que un esquema hace referencia a un conjunto de clases que definen un dominio o una aplicación, RDFS se encarga de proporcionar el mecanismo para declarar propiedades y definir relaciones entre estas propiedades y recursos, es decir proporciona una jerarquización de las clases.

RDFS provee las primitivas básicas, necesarias para modelar Ontologías, las mismas que son agrupadas en clases y propiedades.

3.3.4. OWL

OWL, Lenguaje de Ontologías Web por sus siglas en español, es el resultado de la estandarización de lenguajes como DAML + OIL, mismo que fue desarrollado sobre RDF y RDFS, es decir que toda especificación de OWL está basada en RDF.

OWL añade más vocabulario para describir clases y propiedades, con la finalidad de definir Ontologías estructuradas que puedan ser usadas a través de diferentes



sistemas. Este vocabulario más extenso, puede contener relaciones entre clases como por ejemplo: desunión, cardinalidad, igualdad, más tipos de propiedades, características de propiedades (por ejemplo, simetría), y clases enumeradas.

El principal objetivo de OWL es facilitar un modelo de marcado sobre RDF y codificado en XML. OWL provee un mejor mecanismo de interpretación del contenido Web que los mecanismos admitidos por XML, RDF, y esquema RDF (RDFS); proporcionando de esta manera, un vocabulario adicional junto con una semántica formal. Proporciona además un lenguaje para definir ontologías estructuradas que pueden ser usadas a través de diferentes sistemas. En el caso de una Ontología descrita en OWL, siempre se inicia con la declaración del nodo raíz de RDF, en donde se puede incluir el espacio de nombres para RDF, RDFS y OWL.

Actualmente la especificación de OWL tiene tres variantes. Cada una de estas variantes cuenta con una mayor capacidad expresiva y con diferentes funcionalidades, y a la vez cada uno de estos sublenguajes es una extensión de su predecesor más simple.

- OWL Lite: Se recomienda para cubrir necesidades de clasificación jerárquica y restricciones simples. Utiliza únicamente algunas



características del lenguaje OWL, y presenta una noción limitada de cardinalidad que solamente permite valores de 0 y 1.

- OWL DL: Esta variante de OWL requiere separación de tipos, por ejemplo una clase no puede ser un individuo o una propiedad, una propiedad no puede ser tampoco un individuo o una clase. Esto implica que no se pueden aplicar restricciones a elementos del lenguaje de OWL; algo que se permite en OWL Full.
- OWL Full: Proporciona soporte a usuarios que requieren el máximo de expresividad y la libertad sintáctica de RDF.

Nos queda ahora la interrogante de definir cuál será el estándar apropiado para cumplir los objetivos antes planteados. Como se ha venido recalcando, RDF/RDFs nos proporciona un modelo de datos para cualquier recurso y las relaciones que se puedan establecer entre ellos, incorporando definiciones de clases y propiedades. Pero no facilita completamente la interoperabilidad semántica.

La variante de OWL que más se presta para cumplir tal meta es OWL DL, ya que al utilizarlo contamos con las siguientes ventajas:

- Se pueden declarar restricciones de rango por propiedad, perteneciente a cada clase.



- Se pueden tratar las propiedades transitivas, de relación de orden (menor que, mayor que, igual), de simetría, inversa y relación única.
- Se puede indicar que determinadas clases son disjuntas. Por ejemplo, supongamos que en OWL declaramos la clase Persona, y las subclases Mujer y Hombre, entonces se puede especificar que una persona no puede pertenecer a las clases Mujer y Hombre al mismo tiempo.
- Se permite expresar restricciones de cardinalidad.

3.4. Lenguajes de Consulta

La idea de *Berneers Lee*, de establecer una serie de estándares que permitieran añadir datos a los datos existentes en la web, expresando semántica a dicha información, paso de ser una utopía a una necesidad, pero la idea de poblar la web de “metadatos” no es el único objetivo a cumplir, sino también definir estándares que sean capaces de abarcar tareas como la búsqueda y clasificación, pero teniendo siempre en cuenta la semántica en los resultados.

Hasta ahora nos hemos estacionando en las tecnologías que hacen posible tener una web más dinámica, por ejemplo hemos visto como el uso de RDF y OWL permiten realizar inferencias; pero, para completar la funcionalidad de una



ontología, también son necesarias las herramientas que permiten extraer la información que contienen.

A continuación analizaremos algunas de las tecnologías que permiten consultar y extraer el conocimiento de los metadatos de la Web Semántica.

3.4.1. RDQL

RDQL¹¹, es un lenguaje de consultas con un enfoque totalmente declarativo, y al igual que con el lenguaje RDF, considera un modelo RDF como un conjunto de tripletas (Objeto, Propiedad, Valor).

Es un modelo de búsqueda orientada a datos en la que solo se consulta información contenida en los modelos, es decir no se realizan inferencias. Sin embargo, el modelo RDQL que implementa *Jena* puede proveer cierta “inteligencia”, ya que proporciona la impresión de que ciertas tripletas existen, mediante la creación de tripletas bajo-demanda.

¹¹ *RDF Data Query Language*, por sus siglas en inglés ó Lenguaje de Consultas de RDF



El modelo de recuperación RDQL no hace más que tomar la descripción que necesita la aplicación en forma de consulta, y devuelve la información en forma de un conjunto de enlaces.

Cada enlace que se devuelve es un conjunto de pares *nombre-valor* para los valores de las variables. Todas las variables son obligadas, es decir que no existe disyuntiva en la consulta.

Para aclarar de mejor manera el funcionamiento de RDQL, ilustraremos un ejemplo mediante un modelo en RDF, mismo que se muestra en el Figura 3.4.



```

- <rdf:RDF>
  - <rdf:Description rdf:about="http://somewhere/JohnSmith">
    <vCard:FN>John Smith</vCard:FN>
    - <vCard:N rdf:parseType="Resource">
      <vCard:Family>Smith</vCard:Family>
      <vCard:Given>John</vCard:Given>
    </vCard:N>
  </rdf:Description>
  - <rdf:Description rdf:about="http://somewhere/RebeccaSmith">
    <vCard:FN>Becky Smith</vCard:FN>
    - <vCard:N rdf:parseType="Resource">
      <vCard:Family>Smith</vCard:Family>
      <vCard:Given>Rebecca</vCard:Given>
    </vCard:N>
  </rdf:Description>
  - <rdf:Description rdf:about="http://somewhere/SarahJones">
    <vCard:FN>Sarah Jones</vCard:FN>
    - <vCard:N rdf:parseType="Resource">
      <vCard:Family>Jones</vCard:Family>
      <vCard:Given>Sarah</vCard:Given>
    </vCard:N>
  </rdf:Description>
  - <rdf:Description rdf:about="http://somewhere/MattJones">
    <vCard:FN>Matt Jones</vCard:FN>
    <vCard:N vCard:Family="Jones" vCard:Given="Matthew"/>
  </rdf:Description>
</rdf:RDF>

```

Figura 3.4. Modelo de ejemplo en RDF.

Este modelo contiene una serie de descripciones para la clase PERSONA. Podemos entonces realizar la consulta que se muestra en la Figura 3.5, con la sintaxis en RDQL.



```
SELECT? X
WHERE (? X <http://www.w3.org/2001/vcard-rdf/3.0#FN> "John Smith")
```

Figura 3.5. Consulta en RDQL

La figura 3.5 indica la estructura de una consulta RDQL, en la primera línea de la misma se indica la extracción de cierto atributo X de la ontología; la segunda línea contiene la condición que dicho atributo deberá cumplir utilizando la cláusula WHERE; en este caso la condición será coincidir con la instancia "John Smith".

La ejecución de esta consulta, muestra una respuesta con el formato mostrado en la figura 3.6.

```
x
=====
<http://somewhere/JohnSmith/>
```

Figura 3.6. Resultado de la consulta en RDQL.

En este caso X representa el atributo extraído de la ontología de acuerdo a la consulta.



3.4.2. SeRQL Query Language

SeRQL¹², es un lenguaje de consulta en RDF o RDF Schema. Este lenguaje combina las características fundamentales de otros lenguajes de consulta como RQL, RDQL, N3, etc., con algunos aspectos nuevos.

Algunas de las funcionalidades más importantes que ofrece SeRQL son:

- Transformación gráfica.
- Soporte para RDF Schema.
- Soporte para tipos de datos de XML Schema.

3.4.3. SPARQL

SPARQL^[32] es un lenguaje que se puede utilizar para expresar consultas a través de diversas fuentes de datos, siempre y cuando estos datos estén almacenados de forma nativa como RDF o sean vistos como RDF a través de un middleware.

El lenguaje de consulta SPARQL para RDF está diseñado para satisfacer los casos de uso y las necesidades identificadas por el acceso a datos RDF.

¹² Sesame RDF Query Language



SPARQL está basado en el lenguaje de base de datos SQL, y al igual que sucede con el mismo, es necesario distinguir entre el lenguaje de consulta y el motor para el almacenamiento y la recuperación de los datos. Por tal razón existen algunas implementaciones de SPARQL, ligados generalmente a entornos de desarrollo.

La mayoría de tipos de consultas SPARQL contienen un conjunto de patrones de tripletas llamadas Patrones Gráficos de Base. Estos patrones de tripletas son como las tripletas RDF, con la excepción de que cada uno de los sujetos, predicados y objetos pueden ser una variable.

Un patrón gráfico de base coincide con la información de un sub-grafo RDF cuando los términos RDF de ese sub-grafo pueden ser sustituidos por las variables, y el resultado es un grafo RDF equivalente al sub-grafo.

Para entender de mejor manera el funcionamiento de este lenguaje vamos a ilustrar una consulta sencilla, tomada de la web oficial de SPARQL de la W3C.

El siguiente ejemplo muestra una consulta SPARQL para encontrar el título de un libro:

Datos:

```
<http://example.org/book/book1> <http://purl.org/dc/elements/1.1/title>  
"" Tutorial SPARQL.
```

Consulta:

```
SELECT? Título  
DÓNDE  
(  
  <http://example.org/book/book1> <http://purl.org/dc/elements/1.1/title>? título.  
)
```

La consulta consta de dos partes: la cláusula Select, que identifica las variables que aparecen en los resultados de la consulta; y la cláusula Where, que establece el patrón gráfico de base, que se comparara con el gráfico de datos.

La respuesta a la consulta anterior es la siguiente:

título
"SPARQL Tutorial"



El resultado de una consulta es una secuencia de solución correspondiente a la forma en la cual el patrón gráfico de la consulta coincide con los datos. Puede haber cero, una, o múltiples soluciones a una consulta.

3.5. Procesamiento del Lenguaje Natural

Más que un lenguaje de soporte para la infraestructura de la Web Semántica; el Procesamiento de Lenguaje Natural o PLN es la técnica que permite comunicarse con el usuario final, utilizando el lenguaje natural. Es decir, nos permite pasar de la porción lógica de la inteligencia semántica (OWL, RDF, RDFs, etc.) a la parte de interacción con el usuario en la Interfaz de un sistema.

El procesamiento de una consulta es una tarea compleja ya que trata con el lenguaje natural; es por esto que se utilizan las técnicas de PLN en el ámbito de búsquedas y anotaciones semánticas. Estas técnicas facilitan la descripción del contenido de una frase o consulta, con la finalidad de clasificarla según su significado y contexto; y así obtener resultados que satisfagan las necesidades del usuario en mayor grado.



El procesamiento del Lenguaje Natural trata de crear mecanismos de comunicación que sean eficaces computacionalmente, enfocándose en la comprensión del lenguaje natural; para esto se basa en la comprensión de los aspectos cognitivos y la organización de la memoria humana. Para lograr este objetivo, el PLN tiene que hacer frente a grandes problemas del lenguaje natural como son la ambigüedad; la detección de separación entre palabras, y la recepción imperfecta de datos.

A continuación exponemos las técnicas que permite llevar a cabo un proceso de PLN:

- **Análisis morfológico:** Es el análisis de las palabras para extraer raíces, rasgos flexivos, unidades léxicas compuestas y otros fenómenos gramaticales. Es ejecutado por los programas etiquetadores que asignan una categoría gramatical a cada palabra a partir de los rasgos morfológicos identificados en las mismas.
- **Análisis sintáctico:** Es el análisis de la estructura sintáctica de la frase mediante una gramática del idioma; permite ver cómo se relacionan y combinan las palabras entre sí para formar unidades superiores y frases. Aquí se identifican únicamente las estructuras más significativas: frases



nominales, sintagmas verbales y preposicionales, entidades, etc. En este punto se aplican parsers o formalismos descriptivos del lenguaje, que tienen como objetivo fijar la estructura sintáctica del texto.

- **Análisis semántico:** Es la extracción del significado de una frase, y la resolución de ambigüedades léxicas y estructurales. Se obtiene el significado de las frases que componen un texto cuando se consigue la representación semántica de la misma a partir de los elementos que la forman.
- **Planificación de la frase.** Implica estructurar cada frase del texto con el fin de expresar su significado adecuado.
- **Generación de la frase.** Es la generación de la cadena lineal de palabras a partir de la estructura general de la frase, con sus correspondientes flexiones, concordancias y restantes fenómenos sintácticos y morfológicos.

Existen varios métodos de PLN disponibles en la Web (como por ejemplo la integración de recursos semánticos basados en WordNet^[33]), pero, con la finalidad de entender de mejor manera este proceso, y cumplir con los objetivos del prototipo de búsqueda semántica, implementaremos un procedimiento de interpretación del lenguaje natural propio. Este proceso será explicado en el



capítulo 5, “Diseño e Implementación del Buscador Semántico”. No se realizará un proceso PLN como tal, ya que dicha tarea se encuentra fuera del alcance de este trabajo.

3.6. Interfaces y APIS de Programación

Con la definición de las tecnologías mencionadas en la sección anterior, es importante ahora conocer la manera de plasmar el estándar a utilizar (OWL DL), mediante la ayuda de alguna API de programación para la manipulación y tratamiento de una Ontología.

En la actualidad existen varias interfaces de programación que permiten cumplir con las reglas de estándares como RDF, RDFS, OWL; la gran mayoría se encuentran incluidos como *plugins* para distintos lenguajes de programación, otros en cambio se presentan de manera independiente.

A continuación analizaremos las APIs “OWL API” y “Jena”; que se basan en el lenguaje de programación Java, por tratarse de un lenguaje universal, completo y que permite la fácil incorporación de *plugins*.



3.6.1. OWL API

La OWL API^[34] es una API Java y permite crear, manipular y serializar Ontologías OWL. La API de OWL es de código abierto y está disponible bajo la licencia LGPL.

Específicamente la API OWL está enfocada hacia OWL Lite y OWL DL, y también implementa un motor de inferencia.

La API de OWL incluye los siguientes componentes:

- Una API para OWL2
- Parser de lectura y escritura para RDF/XML
- Parser de lectura y escritura para PPP/XML, y
- Analizador KRSS

3.6.2. Jena

De igual manera, Jena^[35] es una es un API para Java, es decir, un entorno para el desarrollo de aplicaciones para la Web Semántica. Proporciona un ambiente de programación para RDF, RDFS, OWL, y SPARQL; e incluye un motor de inferencia basado en reglas. Además, Jena permite gestionar todo tipo de ontologías, almacenarlas y realizar consultas sobre ellas.

El Framework de Jena incluye:



- Un API para RDF.
- Lectura y Escritura de RDF en RDF/XML, N3 y N-Triples.
- Una API para OWL.
- Almacenamiento persistente en memoria, y
- Motor de consultas SPARQL.

Una característica importante de Jena, es la incorporación de un Reasoner, es decir un motor de razonamiento que implementa la capacidad de inferir conocimiento nuevo.

Todas estas características, añadido a la facilidad de implementación e integración a la plataforma Eclipse para Java; constituyen la razón por la cual hemos seleccionado esta tecnología en particular para aplicarla en nuestro prototipo de buscador semántico; además existe una gran cantidad de repositorios de información y ejemplos para JENA.

A continuación estudiaremos más a fondo el funcionamiento de Jena, mediante la definición de las funciones más relevantes que posee, aclarando los conceptos de estándares como XML, RDF y OWL.

En primer lugar, un recurso para Jena es cualquier cosa que se pueda identificar; una persona puede ser un recurso, así como también una página web, o este



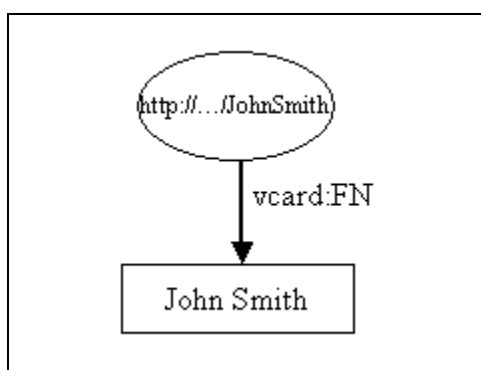
documento. A lo largo de esta explicación los ejemplos tratarán sobre personas, utilizando la representación RDF de VCARDS¹³ tomados de la documentación oficial de Jena^[36].

Como primer paso se debe definir un modelo vacío, como se indica en la Figura 3.7.

```
//Create a empty model
Model model= ModelFactory.createDefaultModel();
```

Figura 3.7. Creación de un modelo

Una manera alternativa de representación de RDF son los grafos, tal y como se aprecia en la figura 3.8, en donde se describe el recurso “John Smith”. El cual deberá ser declarado en el modelo que se creó anteriormente.



¹³ VCards es un lenguaje embebido en RDF al igual que OWL



Figura 3.8. Representación RDF mediante grafos

En este caso, contamos únicamente con un recurso y una propiedad asociada, entonces, una vez creado el modelo habrá que crear el recurso y asignarle la propiedad tal como se indica en la Figura 3.9.

```
String personURI= "http://www.johnsmith.com";
String FullName= "John Smith";

//create the resource
Resource JhonSmith = model.createResource(personURI);
JhonSmith.addProperty(VCARD.FN, FullName);

//write the model
model.write(System.out);
```

Figura 3.9. Ejemplo de modelo con JENA

La ejecución de estas líneas de código da lugar al siguiente fichero en RDF:

```
<rdf:RDF
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:vcard="http://www.w3.org/2001/vcard-rdf/3.0#" >
  <rdf:Description rdf:about="http://www.johnsmith.com">
    <vcard:FN>John Smith</vcard:FN>
  </rdf:Description>
</rdf:RDF>
```

Figura 3.10. Resultado RDF obtenido con Jena



Jena se puede utilizar para crear y manipular gráficos RDF. En el primer ejemplo el valor de la propiedad era un literal, pero las propiedades RDF pueden tomar también otros recursos como sus valores; esto se puede apreciar en la figura 3.11.

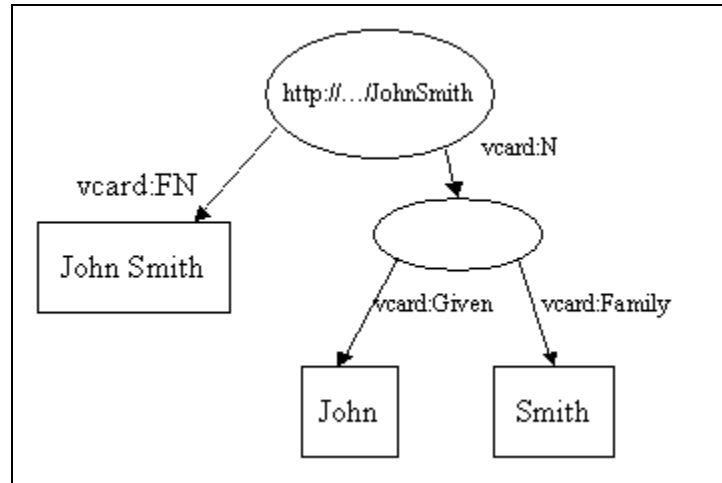


Figura 3.11. Modelo de Propiedades con valores

En el modelo de la figura 3.11, se ha añadido la propiedad “Vcard:N”, para representar la estructura del nombre “John Smith”. Se puede notar que la propiedad Vcard:N toma un recurso como su valor, pero al no estar representado con una URI se le conoce como un “Nodo en blanco”.

Finalmente, podemos decir que Jena provee un entorno de desarrollo para RDF, OWL, SPARQL y RDFS, permitiendo cualquier tarea de lectura o escritura, como navegar entre los nodos, añadir información, etc.



CAPÍTULO IV

METODOLOGÍA NEON,



TEORÍA Y APLICACIÓN



En el Capítulo Cuarto, METODOLOGÍA NEON; TEORÍA Y APLICACIÓN, se presentarán los pasos y actividades que utiliza la metodología NeOn para la construcción de ontologías, detallando cada uno de estos, implementándolos para el dominio de Turismo, y creando y modificando nuestra ontología dentro de la plataforma NeOn Tool Kit.

CAPÍTULO IV

METODOLOGÍA NEON; TEORÍA Y APLICACIÓN

4.1. Introducción



El objetivo de la metodología NeOn es construir redes de ontologías y cubrir todos los inconvenientes de sus metodologías antecedentes, beneficiándose de sus fortalezas. Concretamente, la metodología NeOn incluye los beneficios de colaboración de Dilligent, la propuesta del análisis del alcance, dada por Methontology y el uso de preguntas de competencia de On-To-Knowledge.

Características de la Metodología NeOn:

- En las diferentes etapas del proceso de desarrollo de la ontología, la información contextual puede ser introducida por expertos y profesionales del dominio.
- La metodología incluye el proceso de desarrollo y diferentes modelos de ciclo de vida, así como los métodos, técnicas, y herramientas que pueden ser utilizados durante la construcción de redes de ontologías.
- La metodología se puede aplicar en conjunto con la herramienta NeOn Tool Kit, aunque esto no es obligatorio.

Actividades de la Metodología NeOn:

La metodología NeOn utiliza el siguiente conjunto de pasos para desarrollar una ontología:

1. Especificación de la Ontología.



2. Reutilización de recursos No-Ontológicos.
3. Reingeniería de recursos No-Ontológicos.
4. Reutilización de recursos Ontológicos.
5. Reutilización de Plantillas de diseño para Ontologías.

A continuación explicaremos cada uno de estos pasos y presentaremos la manera en que se aplican para la creación de nuestra Ontología de Turismo:

4.2. PASO 1: Especificación de la Ontología

El proceso de especificación de la ontología tiene como finalidad explicar la razón para la cual se esta construyendo la ontología, cuales son sus usos deseados, quienes serán sus usuarios finales y cuales son los requerimientos que la ontología debe cumplir.

En este paso debemos definir el propósito de la Ontología en términos generales, pudiendo ayudarnos de preguntas como las siguientes:

- ¿Cuál es el dominio que la ontología cubrirá?
- ¿Para qué usaremos la ontología?



- ¿Para qué tipos de preguntas la información en la ontología deberá proveer respuestas?
- ¿Quién usará y mantendrá la ontología?

Las respuestas a estas preguntas ayudaran a delimitar el alcance de la Ontología en cualquier momento del diseño.

Para especificar los requerimientos de la ontología utilizaremos las preguntas de competencia. Pero, antes de esto, es necesario identificar el propósito y alcance de la ontología, su nivel de formalidad, sus usuarios y sus aplicaciones deseadas.

Para este primer paso; la metodología NeOn propone las tarjetas de llenado; presentadas en la Tabla 4.1. Estas incluyen el objetivo de la ontología, entradas y salidas, el responsable de la actividad y cuando debería llevarse a cabo esta actividad.



ESPECIFICACIÓN DE LA ONTOLOGÍA	
DEFINICION La especificación de la ontología es la actividad de reunir los requerimientos que la ontología debe cumplir. Por ejemplo las razones para construir la ontología, el grupo al que va dirigida, los usos deseados para la misma, todo esto conseguido a través de un proceso de consenso.	
OBJETIVO La actividad de especificación de la ontología presenta la razón por la cual la ontología esta construida, sus usos deseados, quienes serán los usuarios finales y los requerimientos que la ontología debería cumplir.	
ENTRADA Un conjunto de necesidades ontológicas	SALIDA Documentos de especificación de los Requerimientos de la Ontología. (ORSO)
QUIEN Los desarrolladores de software y los profesionales de ontologías que forman parte del equipo de desarrollo de la ontología (ODT), en colaboración con los usuarios y expertos del dominio.	
CUANDO Esta actividad debe ser llevada a cabo en paralelo con la actividad de adquisición del conocimiento.	

Tabla 4.1. Tarjeta de llenado para la Especificación de la Ontología

La actividad de Especificación de la Ontología contiene las tareas mostradas en la Figura 4.1:

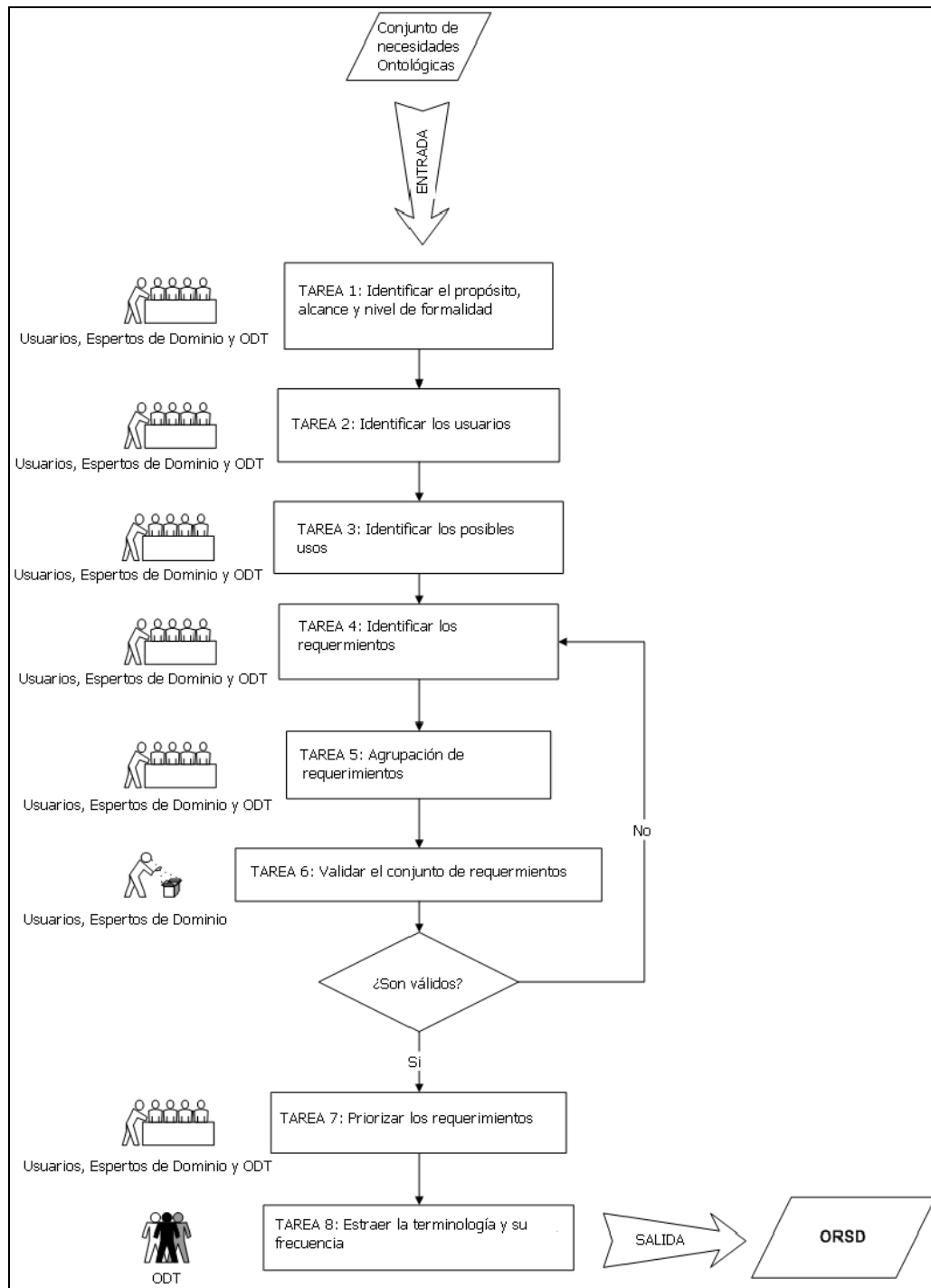


Figura 4.1: Tareas para la especificación de la Ontología



El resultado de esta actividad es el ORSD, o Documento de Especificación de los Requerimientos de la Ontología (Ontology Requirements Specification Document). La metodología NeOn propone una plantilla para escribir el ORSD con los slots, que se muestran en la Tabla 4.2:

Documento plantilla para la especificación de los requerimientos de la Ontología	
1	Propósito
	Incluye la intención de la ontología y sus objetivos.
2	Alcance
	Incluye la cobertura y la granularidad de la ontología.
3	Nivel de Formalidad
	Incluye el nivel de formalidad de la ontología.
4	Usuarios
	Incluye los principales usuarios que se desea que tenga la ontología.
5	Usos
	Incluye los escenarios principales en los cuales se desea usar la ontología.
6	Grupos de Competencia
	Grupos de Preguntas de Competencia o CQs, incluyendo sus respuestas y prioridades
7	Pre-Glosario de Términos
	Términos
	Lista de términos incluida en el CQs y sus frecuencias.
	Objetos
	Lista de objetos y sus frecuencias.



Tabla 4.2: Plantilla para el OSDR

Especificación de la Ontología para la Ontología de Turismo

El objetivo principal del prototipo de búsqueda semántica es el de demostrar las ventajas de la extracción de información por significado en la Web. Para esto se utilizará una Ontología de referencia que represente el dominio turístico para la Ciudad de Cuenca.

A continuación presentaremos la especificación de la Ontología de Turismo a utilizar en el prototipo de buscador semántico, siguiendo las guías propuestas por la Metodología NeOn.

Tarea 1: Identificar el propósito, alcance y nivel de formalidad.

PROPÓSITO	El propósito de la Ontología es definir una estructura formal de conceptos referentes al dominio turístico, de manera que permita responder a preguntas simples realizadas en lenguaje natural.
	Nuestra Ontología estará restringida al dominio turístico,



RESTRICCIÓN	teniendo como instancias lugares turísticos de la ciudad de Cuenca, tales como hoteles, parques, iglesias, museos, lugares arqueológicos y sitios naturales.
NIVEL DE FORMALIDAD	Nivel Técnico. La Ontología esta representada en lenguaje XML, y OWL.

Tarea 2: Identificar Usuarios

Dado el alcance y el propósito de nuestra ontología, se concluye que la misma estará al servicio de los siguientes usuarios:

Usuario 1: Docentes universitarios que necesiten la Ontología con fines de pruebas y experimentación.



Usuario 2: Estudiantes o investigadores que pretendan poner en práctica el concepto de buscadores semánticos o utilizar la ontología descrita como extensión de una ontología más grande.

Usuario 3: Usuarios en general que deseen encontrar lugares turísticos dentro de la ciudad de Cuenca.

La razón por la cual no se especifican más usuarios, es por que se trata de un prototipo con fines académicos, restringiendo por el momento el acceso al mismo.

Tarea 3. Identificar usos de la Ontología.

Uso 1: Buscar información de lugares. Proveer información como nombre, dirección, tipo de lugar turístico de las instancias de la Ontología.

Uso 2: Información sobre la naturaleza de los lugares. Podemos conocer acerca de un lugar específico en particular, conociendo el nombre del mismo.

Tarea 4. Identificar los requerimientos



Para especificar los requerimientos de la ontología usaremos la técnica de preguntas de competencia (CQ). Utilizamos la práctica *bottom up* para identificar los mismos. Como resultado de este análisis, se encontraron 32 preguntas de competencia para la Ontología de Turismo.

Las preguntas de competencia se almacenan en un archivo de Excel y se reescriben en una herramienta de mapa mental, como aparecen en las Figura 4.2 y 4.3, respectivamente.

	A	B	C
1	Nº	Pregunta de competencia	Respuesta
2	CQ1	¿Cuál es la dirección del Hotel?	Calle Benigno Malo y Sucre
3	CQ2	¿Cuál es la dirección de la Iglesia?	Calle Luis Cordero y Sucre
4	CQ3	¿Cuál es la dirección del Restaurante?	Calle Bolívar y Benigno Malo
5	CQ4	¿Cuál es la dirección del Bar?	Paseo 3 de Noviembre
6	CQ5	¿Cuál es la dirección del Museo?	Calle Larga y Av. Huaynacapac
7	CQ6	¿Cuál es la dirección del Parque?	Calle Bolívar y Tomás Ordoñez
8	CQ7	¿Qué tipo de lugar es el Hotel Carvallo?	Es un Hotel
9	CQ8	¿Qué tipo de lugar es San Alfonso?	Es una Iglesia
10	CQ9	¿Qué tipo de lugar es Raymipamba?	Es un Restaurante
11	CQ10	¿Qué tipo de lugar es Dos Dos?	Es un Bar
12	CQ11	¿Qué tipo de lugar es el Banco Central?	Es un Museo
13	CQ12	¿Qué tipo de lugar es el Paraíso?	Es un Parque
14	CQ13	¿Cuál es el año de Fundación del Hotel?	1997
15	CQ14	¿Cuál es el año de Fundación de la Iglesia?	1977
16	CQ15	¿Cuál es el año de Fundación del Museo?	1952
17	CQ16	¿Cuál es el año de Fundación del Parque?	1955
18	CQ17	¿Cuál es el horario de atención de Restaurante?	Lun-Vier: 12h00-22h00
19	CQ18	¿Cuál es el horario de atención de Bar?	Mier-Jue: 21h00-00h00, Vier-Sab: 21h00-2h30
20	CQ19	¿Cuál es el horario de atención de Museo?	Lun-Vier: 9h00-16h00
21	CQ20	¿Cuál es el horario de atención de Hotel?	Lun-Vier: 9h00-16h00, Sab: 9h00-12h30
22	CQ21	¿Qué tipo de museo es el Banco Central?	Es un museo arqueológico.
23	CQ22	¿San Alfonso es una iglesia ?	Si, San Alfonso es una iglesia.
24	CQ23	¿San José es un museo?	No, San José es una iglesia.

Figura 4.2. Preguntas de Competencia en Excel

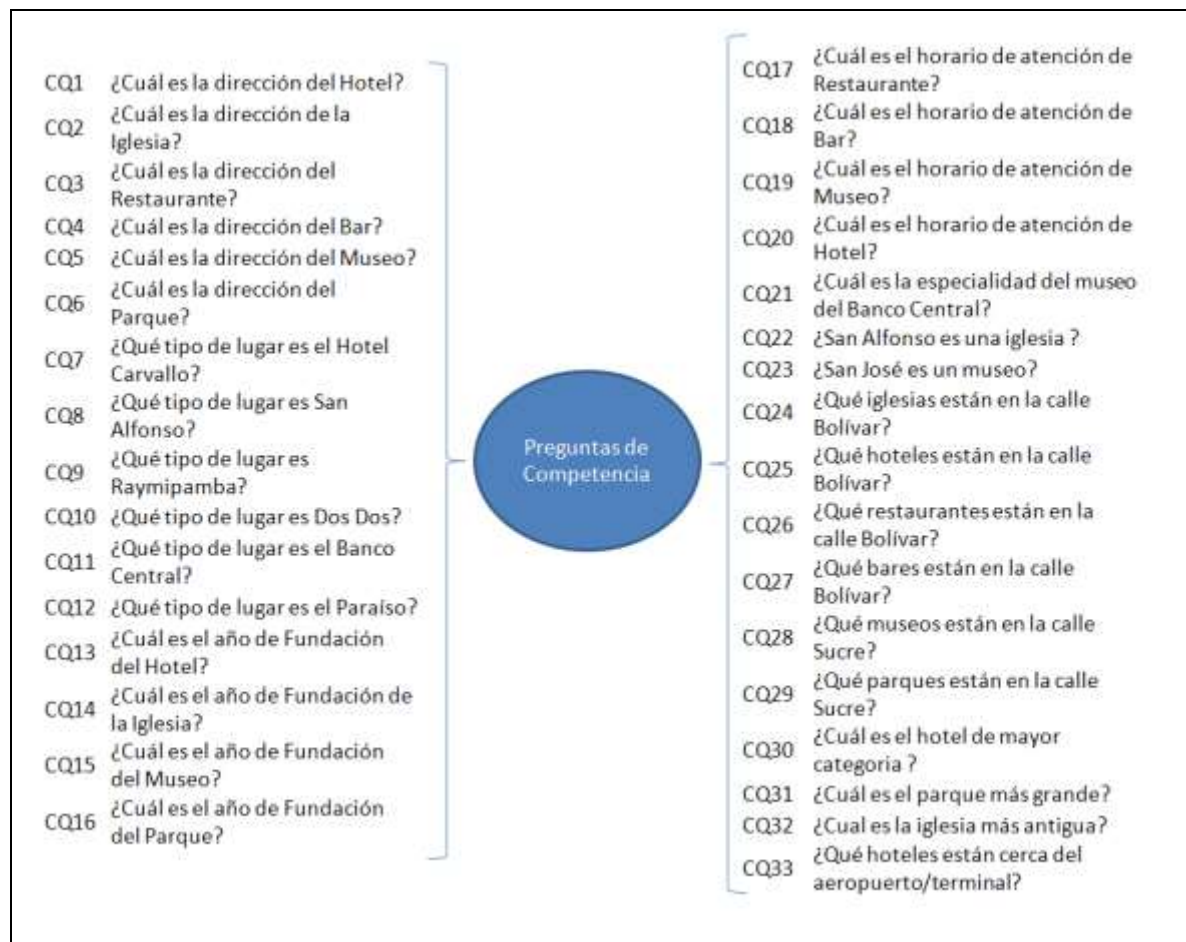


Figura 4.3. Preguntas de Competencia, Mapa Mental

Tarea 5: Agrupación de los Requerimientos

Las 33 preguntas de competencia descritas en la sección anterior han sido agrupadas manualmente en 7 grupos según las entidades del mundo real a las cuales hacen referencia.

La Figura 4.4 muestra estos grupos: Lugares Turísticos, Restaurante, Bar, Iglesia, Museo, Hotel, y Parque.



Figura 4.4. Grupos de Preguntas de Competencia

La Figura 4.5 muestra cada una de las preguntas de competencia dentro de sus respectivos grupos:

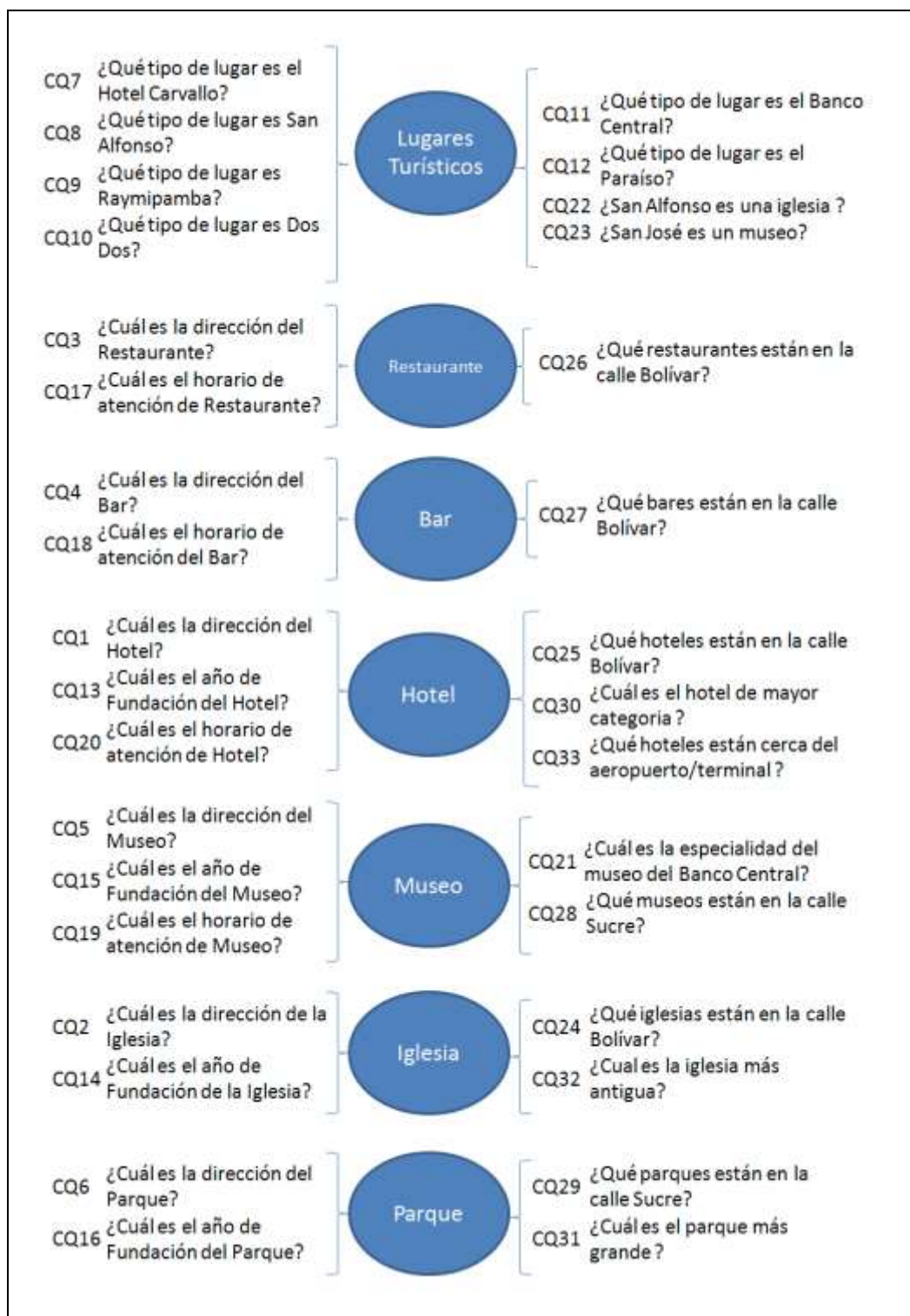




Figura 4.5. Preguntas de Competencia para cada grupo

Tarea 6: Validar el conjunto de requerimientos

Para validar el conjunto de preguntas de competencia se utilizaron los siguientes criterios:

- Exactitud: Se verificó que cada una de las preguntas de competencia sean correctas y sus respuestas estén de acuerdo a los objetivos de la ontología, también se verificó que las CQs permitan la consecución de los usos que se encontraron en la tarea 3.
- Completitud: Consideramos que el conjunto de CQs es completo ya que no hay necesidad de agregar más requerimientos, esto de acuerdo al uso que tendrá la ontología.
- Consistencia y realista: Ninguna de las CQs se contradicen entre sí. Y cada uno de los grupos de CQs hace referencia a entidades independientes del dominio.
- Verificable y Entendible: Cada una de las CQs es entendible por los usuarios y no presenta ambigüedad en sus términos.



- Adaptabilidad y Trazabilidad: El conjunto de CQs para la ontología de Turismo esta estructurado de manera que puede ser modificado y seguido durante todo el desarrollo de la ontología.

Tarea 7: Priorización de los requerimientos

La priorización de requerimientos para la Ontología de Turismo es una tarea que se pasará por alto, ya que no es necesario representar el conocimiento de ninguna de las CQs de manera prioritaria para la consecución de los objetivos de nuestra aplicación final.

Tarea 8: Extracción de la terminología y su frecuencia

Extraemos la terminología más utilizada de la lista de requerimientos. Estos términos nos permitirán representar formalmente a la ontología.

Se realiza la extracción por cada uno de los grupos de preguntas de competencia encontrados en la Tarea 5. Los términos más importantes, encontrados en las CQ



se encuentran en la Tabla 4.3 junto con su frecuencia y clasificados según el grupo de requerimientos al cual pertenecen:

Término	Frecuencia
LUGAR TURÍSTICO	6
Tipo	8
Dirección	6
Nombre	7
Hotel	6
Año de Fundación	1
Horario	1
Categoría	1
Aeropuerto	1
IGLESIA	4
Año de Fundación	1
RESTAURANTE	3
Horario	1
BAR	3
Horario	1
MUSEO	5
Año de Fundación	1
Horario	1
Especialidad	1
PARQUE	4
Año de Fundación	1
Tamaño	1

Tabla 4.3. Terminología y Frecuencia



La Tabla 4.4 muestra algunos ejemplos de objetos o instancias de los atributos NOMBRE, DIRECCIÓN, AÑO DE FUNDACIÓN, y HORARIO, mismos que han sido extraídos de las respuestas a las CQs obtenidas en la tarea 5.

NOMBRE	DIRECCIÓN	AÑO DE FUNDACIÓN	HORARIO
San Blas	Calle Benigno Malo y Sucre	1997	Lun-Vier: 12h00-22h00
Hotel Carvallo	Calle Luis Cordero y Sucre	1977	Mier-Jue: 21h00-00h00
El Cántaro	Calle Bolivar y Benigno Malo	1952	Lun-Vier: 9h00-16h00
Bar el Ángeluz	Paseo 3 de Noviembre	1955	Lun-Vier: 9h00-16h00
Museo de Arte Moderno	Calle Larga y Av. Huaynacapac		
Parque Calderón	Calle Bolivar y Tomás Ordoñez		

Tabla 4.4. Ejemplos de Objetos

Una vez terminadas estas tareas, obtenemos como salida el Documento de Especificación de los Requerimientos de la Ontología u ORSD; mismo que se muestra en la Tabla 4.5:



Documento plantilla para la especificación de los requerimientos de la Ontología	
1	Propósito
	Definir una estructura formal de conceptos referentes al dominio turístico, de manera que permita responder a preguntas simples realizadas en lenguaje natural, además de permitir la inclusión de anotaciones semánticas.
2	Alcance
	Restringida al dominio turístico, teniendo como instancias lugares turísticos de la ciudad de Cuenca. La ontología no realizará inferencias automáticas ni procesos de razonamiento complejos.
3	Nivel de Formalidad
	La ontología utilizará lenguaje natural en su totalidad ya que va a interactuar con un buscador dirigido a los usuarios.
4	Usuarios
	Usuario 1: Docentes universitarios que necesiten la Ontología con fines de pruebas y experimentación. Usuario 2: Estudiantes o investigadores que pretendan poner en práctica el concepto de buscadores semánticos o utilizar la ontología descrita como extensión de una ontología más grande. Usuario 3: Usuarios en general que deseen encontrar algún turístico dentro de la ciudad de Cuenca.
5	Usos
	Uso 1: Buscar información de lugares. Proveer información como nombre, dirección, tipo de lugar turístico de las instancias de la Ontología. Uso 2: Encontrar relaciones geográficas. Podemos relacionar un lugar turístico con uno cercano, en base a la dirección. Uso 3: Listar lugares turísticos. Obtener lugares específicos dentro de un área geográfica; por ejemplo la lista de hoteles en la Calle Larga. Uso 4: Información sobre la naturaleza de los lugares. Podemos conocer acerca de un lugar específico en particular, conociendo



6 Grupos de Competencia	
G1: Lugares Turísticos	CQ7 ¿Qué tipo de lugar es el Hotel Carvallo? CQ8 ¿Qué tipo de lugar es San Alfonso? CQ9 ¿Qué tipo de lugar es Raymipamba? CQ10 ¿Qué tipo de lugar es Dos Dos? Lugares Turísticos CQ11 ¿Qué tipo de lugar es el Banco Central? CQ12 ¿Qué tipo de lugar es el Paraíso? CQ22 ¿San Alfonso es una iglesia? CQ23 ¿San José es un museo?
G2: Hotel	CQ1 ¿Cuál es la dirección del Hotel? CQ13 ¿Cuál es el año de Fundación del Hotel? CQ20 ¿Cuál es el horario de atención de Hotel? Hotel CQ25 ¿Qué hoteles están en la calle Bolívar? CQ30 ¿Cuál es el hotel de mayor categoría? CQ33 ¿Qué hoteles están cerca del aeropuerto/terminal?
G3: Iglesia	CQ2 ¿Cuál es la dirección de la Iglesia? CQ14 ¿Cuál es el año de Fundación de la Iglesia? Iglesia CQ24 ¿Qué iglesias están en la calle Bolívar? CQ32 ¿Cuál es la iglesia más antigua?
G4: Bar	CQ4 ¿Cuál es la dirección del Bar? CQ18 ¿Cuál es el horario de atención del Bar? Bar CQ27 ¿Qué bares están en la calle Bolívar?
G5: Restaurante	CQ3 ¿Cuál es la dirección del Restaurante? CQ17 ¿Cuál es el horario de atención del Restaurante? Restaurante CQ26 ¿Qué restaurantes están en la calle Bolívar?
G6: Parque	CQ6 ¿Cuál es la dirección del Parque? CQ16 ¿Cuál es el año de Fundación del Parque? Parque CQ29 ¿Qué parques están en la calle Sucre? CQ31 ¿Cuál es el parque más grande?
G7: Museo	CQ5 ¿Cuál es la dirección del Museo? CQ15 ¿Cuál es el año de Fundación del Museo? CQ19 ¿Cuál es el horario de atención de Museo? Museo CQ21 ¿Cuál es la especialidad del museo del Banco Central? CQ28 ¿Qué museos están en la calle Sucre?



7 Pre-Glosario de Términos		
Términos		
	Término	Frecuencia
a.	Lugar Turístico	6
b.	Tipo	8
c.	Dirección	6
d.	Nombre	7
e.	Hotel	6
f.	Año de Fundación	4
g.	Horario	4
h.	Categoría	1
i.	Aeropuerto	1
j.	Iglesia	4
k.	Restaurante	3
l.	Bar	3
m.	Museo	5
n.	Especialidad	1
o.	Parque	4
p.	Tamaño	1
Objetos		
	- Tipo - O1. Hotel - O2. Iglesia - O3. Bar – Restaurante - O4. Museo - O5. Parque - Dirección - O6. Calle Benigno Malo y Sucre - O7. Calle Luis Cordero y Sucre - O8. Calle Bolívar y Benigno Malo - O9. Paseo 3 de Noviembre - O10. Calle Larga y Av. Huaynacapac - O11. Calle Bolívar y Tomás Ordoñez - Nombre - O12. San Blas - O13. Hotel Carvallo - O14. Restaurante el Cántaro - O15. Bar el Ángeluz - O16. Museo de Arte Moderno - O17. Parque Calderón - Año de Fundación - O18. 1997 - O19. 1977 - O20. 1952 - O21. 1955	- Horario - O22. Lun-Vier: 12h00-22h00 - O23. Mier-Jue: 21h00-00h00, Vier-Sab: 21h00-2h30 - O24. Lun-Vier: 9h00-16h00 - O25. Lun-Vier: 9h00-16h00, Sab: 9h00-12h30 - Categoría - O26. 1 estrella - O27. 2 estrellas - O28. 3 estrellas - O29. 4 estrellas - O30. 5 estrellas - Aeropuerto - O31. Cerca del Aeropuerto - O32. Lejos del Aeropuerto - Especialidad - O33. Arqueológico - O34. Religioso - O35. Barroco - O36. Natural - Tamaño - O37. 10 m2 - O38. 20 m2 - O39. 22 m2



Tabla 4.5. ORDS de la Ontología de Turismo

4.3. PASO 2: Reutilización de Recursos No-Ontológicos

La segunda actividad de la metodología NeOn permite analizar recursos no-ontológicos con la finalidad de reutilizarlos en la construcción de la ontología, para esto, dichos recursos deben contener terminología consensuada con el dominio a modelar.

Si uno o más recursos pueden ser reutilizados, entonces se llevará a cabo el proceso de reingeniería de recursos no-ontológicos, extendiendo los recursos con el conocimiento específico del dominio en coincidencia con la información de la nueva ontología en construcción.

La reutilización de terminología existente y consensuada permite ahorrar tiempo y dinero en el proceso de desarrollo de ontologías y promueve la aplicación de buenas prácticas.

Los Recursos No-Ontológicos son recursos de conocimiento existente, cuya semántica no ha sido formalizada por medio de una ontología.

Existe una gran cantidad de Recursos No-Ontológicos que representan conocimiento sobre dominios particulares y que representan algún grado de



consenso para una comunidad de usuarios. Estos recursos se presentan en forma de textos libres, cuerpos de textos, páginas web, estándares, catálogos, directorios web, clasificaciones, tesauros, y léxicos; entre otras.

Los recursos No-Ontológicos deben tener una semántica relacionada al dominio que interese, de forma que permita interpretar el conocimiento que representan. Algunas veces esta semántica está especificada explícitamente en documentos o en lenguaje natural; en otros casos, la semántica no estará disponible de manera explícita pero podrá ser extraída de fuentes de información adicionales como la comunidad de usuarios propietarios de este recurso. Sea cual fuere el caso, el mayor problema de este tipo de recursos, es su falta de formalización, lo cual impide su uso como ontologías.

El objetivo de la reutilización de recursos no-ontológicos de la Metodología NeOn es escoger los recursos más adecuados para la construcción de ontologías. La Tabla 4.6 muestra la platilla de llenado para el proceso de reutilización de recursos no-ontológicos.

REUTILIZACIÓN DE RECURSOS NO-ONTOLÓGICOS	
DEFINICION La reutilización de recursos no ontológicos hace referencia al proceso de escoger los recursos más adecuados para el desarrollo de ontologías.	
OBJETIVO Escoger los recursos mas adecuados para la construcción de ontologías.	
ENTRADA Documento ORSD.	SALIDA Conjunto de recursos no-ontológicos que cubren alguna extensión del dominio
QUIEN Expertos de dominio, desarrolladores de software y profesionales de ontologías.	
CUANDO Después de la actividad de especificación y antes de la reingeniería de recursos no-ontológicos.	

Tabla 4.6. Plantilla de Llenado para la Reutilización de Recursos No-Ontológicos.

Este proceso incluye las actividades y tareas presentadas en la Figura 4.6:

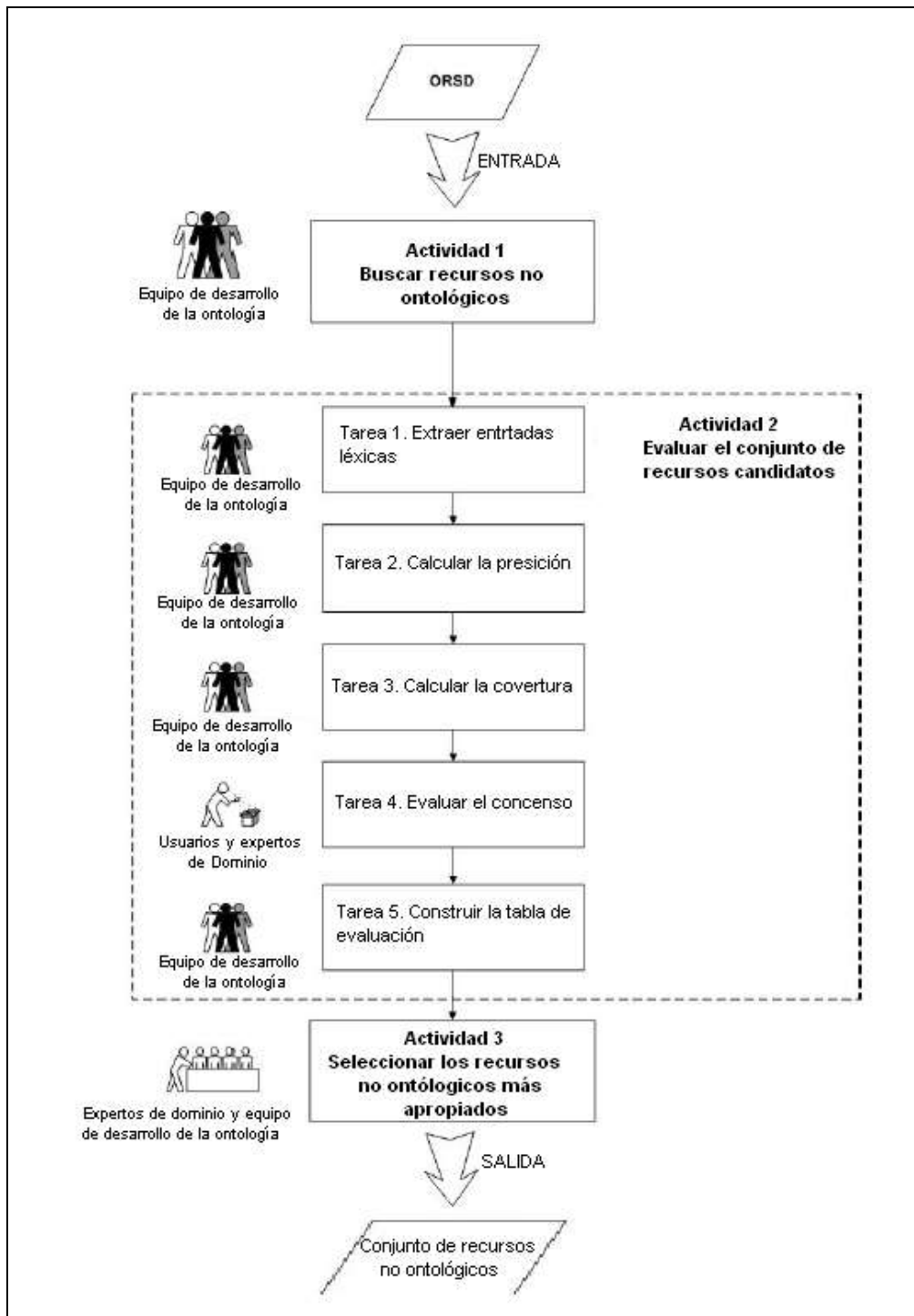




Figura 4.6. Actividades de la Metodología NeOn para el proceso de reutilización de recursos no-ontológicos

Reutilización de Recursos No-Ontológicos para la Ontología de Turismo

A continuación mostraremos el proceso que seguimos para seleccionar los recursos No-Ontológicos a reutilizar en nuestro dominio turístico.

Actividad 1. Buscar recursos no-ontológicos

Se buscó una clasificación existente en cuanto al dominio turístico de la ciudad de Cuenca; y se encontró un sitio Web confiable, que contiene un gran número de sitios de interés, organizados de acuerdo a su naturaleza. Este es el sitio <http://www.cuenca.com.ec/> , mostrado en la Figura 4.7, de la fundación municipal Turismo para Cuenca. Aquí encontramos la siguiente clasificación de los lugares de interés turísticos, en cuanto a nuestro dominio:

- Bar
- Cafetería
- Hostal
- Hostal Residencial
- Hostería



- Hotel
- Iglesias
- Parque
- Restaurante

Dentro de cada una de estas categorías se encuentra un listado de sitios concernientes e información sobre cada uno de estos.



Figura 4.7. Página Principal www.cuenca.com



Actividad 2. Validar el conjunto de Recursos No-Ontológicos candidatos

Dado que este proyecto no cuenta con el apoyo de expertos de dominio; y, que el alcance de la ontología a desarrollar es limitado tratándose de una aplicación tipo prototipo; se analizó el contenido de está página de manera subjetiva y sin tomar en cuenta las medidas de precisión y cobertura.

El resultado es la jerarquía de lugares turísticos para la ciudad de Cuenca mostrada en la figura 4.8:

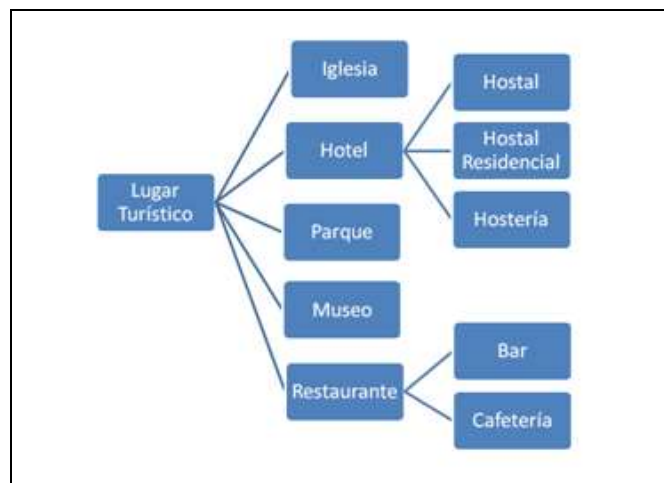


Figura 4.8. Jerarquía de Lugares Turísticos desde Recursos No Ontológicos



Actividad 3. Seleccionar los recursos No-Ontológicos más apropiados.

En conclusión; se ha decidido utilizar la jerarquía de Lugares Turísticos encontrada en la página Web www.cuenca.com como recursos No-Ontológicos para nuestra ontología. La misma, una vez resumida, y adaptada a nuestro interés; se muestra en la figura 4.9:

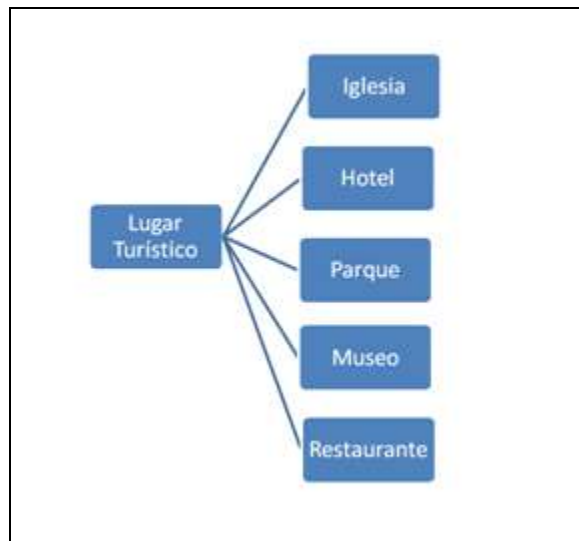


Figura 4.9. Jerarquía de la Ontología de Turismo para el Paso 2: Reutilización de Recursos No Ontológicos



4.4. PASO 3: Reingeniería de recursos No-Ontológicos

El proceso de Reingeniería de Recursos No-Ontológicos de la Metodología NeOn considera como entrada al conjunto de recursos no-ontológicos y las plantillas de reingeniería de recursos no-ontológicos que se encuentran incluidas en el repositorio de plantillas de NeOn.

El proceso de Reingeniería de Recursos No-Ontológicos consta de las siguientes actividades:

1. Ingeniería inversa de recursos no ontológicos
2. Transformación de recursos no ontológicos
3. Ingeniería progresiva de ontologías

La Ingeniería inversa de Recursos No-Ontológicos tiene como objetivo analizar los recursos para identificar sus componentes fundamentales y crear una representación de los mismos en un nivel más alto de abstracción.

La transformación de Recursos No-Ontológicos tiene como objetivo generar un modelo conceptual a partir de los recursos.

La tercera actividad, la ingeniería progresiva de ontologías, pretende obtener una nueva implementación de la ontología en base al nuevo modelo conceptual. Para



esto se utilizan los niveles de abstracción de la ontología ya que están directamente relacionados con el proceso de desarrollo de la ontología.

Actividades y Tareas el Proceso de Reingeniería de Recursos No-Ontológicos

El objetivo de la Reingeniería de Recursos No-Ontológicos es transformar los recursos en ontologías. Como salida de este proceso se obtiene una ontología. La Tabla 4.7 muestra la plantilla de llenado para esta actividad, incluyendo su definición, objetivo, entrada, salida, quien lleva a cabo este proceso, y cuando debería llevarse a cabo el mismo.



REINGENIERÍA DE RECURSOS NO-ONTOLÓGICOS	
DEFINICION Es el proceso de tomar recursos no-ontológicos existentes y transformarlos en una ontología.	
OBJETIVO Crear una ontología a partir de un recurso no-ontológico.	
ENTRADA Uno o más recursos no-ontológicos.	SALIDA Una ontología.
QUIEN Expertos de dominio, desarrolladores de software y profesionales de software.	
CUANDO Después del proceso de reutilización de recursos no-ontológicos y antes de la actividad de conceptualización.	

Tabla 4.7. Plantilla de llenado para la Reingeniería de Recursos No-Ontológicos

Las tareas propuestas para la reingeniería de recursos no-ontológicos se presentan en la Figura 4.10.

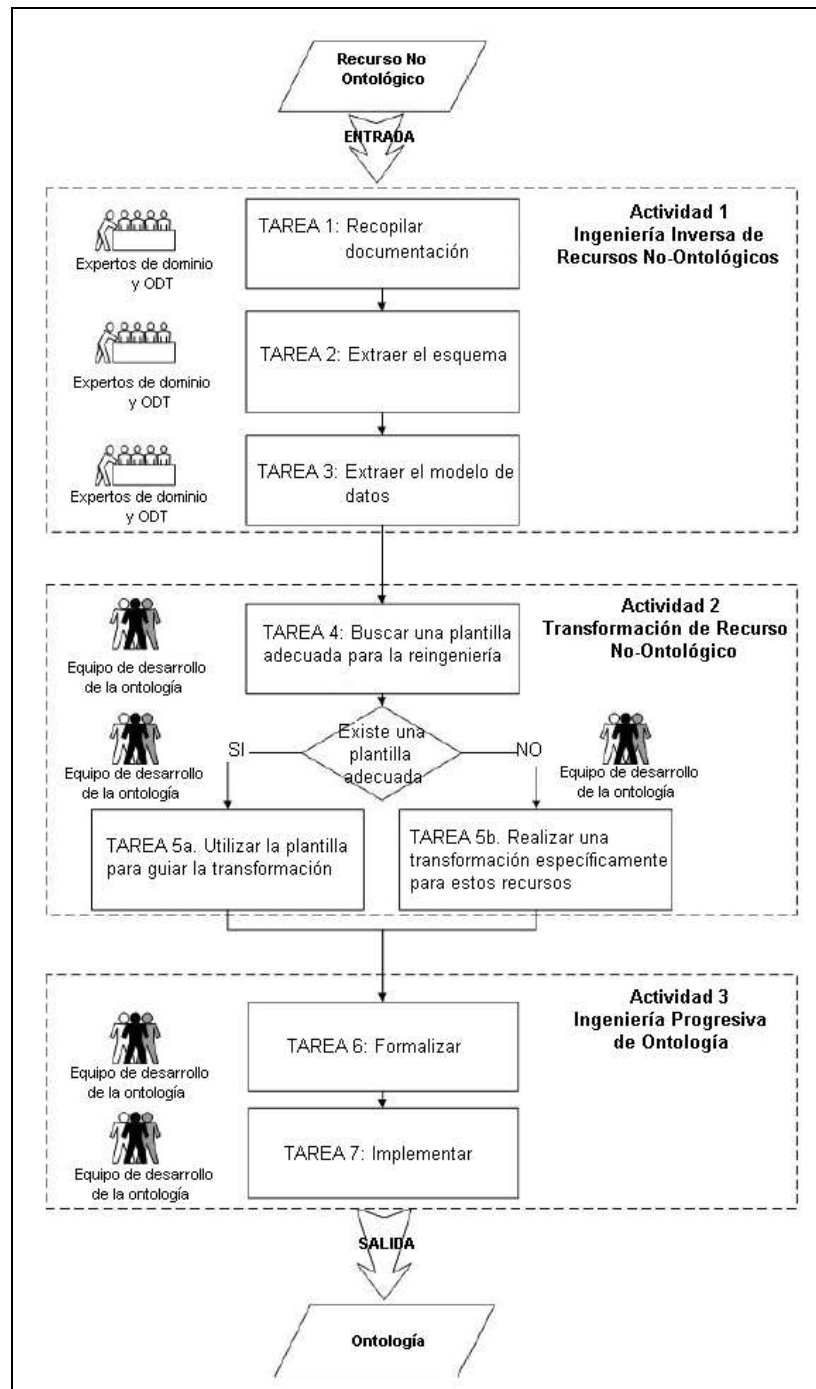


Figura 4.10. Actividades NeOn para el proceso de



Reingeniería de Recursos No-Ontológicos

Reingeniería de Recursos No-Ontológicos para la Ontología de Turismo

Actividad 1. Ingeniería Inversa de Recursos.

El esquema de la Ontología y el modelo de Datos se obtienen del proceso de Reutilización de Recursos No-Ontológicos; y, de acuerdo a la figura 4.9, se concluye que la ontología contará con las siguientes Clases y Propiedades:

CLASES:

- Lugar Turístico: Clase de alto nivel; representa a todas las subclases que contiene y sus propiedades en común.
- Hotel: Representa una clasificación de hoteles según su categoría.
- Iglesia: Representa a las Iglesias de la ciudad de Cuenca.
- Museo: Representa a los museos de la ciudad de Cuenca.
- Parque: Representa a los parques de la ciudad de Cuenca.
- Restaurante: Representa a los restaurantes de la ciudad.

PROPIEDADES:

- nombre (string): Nombre comercial o cultural del sitio de Interés Turístico; está presente en la clase Lugar Turístico, y por lo tanto en todas sus subclases.



- dirección (string): Dirección del sitio de Interés Turístico; esta presente en la clase Lugar Turístico, y por lo tanto en todas sus subclases.
- añoFundación (integer): Año de Fundación del sitio de Interés Turístico; está presente en las clases Iglesia, Museo y Parque.
- horario (string): Horario de atención del sitio de Interés Turístico; esta presente en las clases Hotel, Museo y Restaurante.
- categoria (Integer): Categoría de un Hotel, sus valores pueden ser únicamente: “1 estrella”, “2 estrellas”, “3 estrellas”, “4 estrellas”, y “5 estrellas”. Esta presente solamente en la clase Hotel.
- cercaAeropuerto (boolean): Bandera que indica si un hotel esta cerca (true) o lejos (false) del aeropuerto y/o terminal. En este caso se entiende al aeropuerto y al terminal cómo un solo objeto de referencia ya que se encuentran en una misma zona geográfica de la ciudad de Cuenca.

Esta presente solamente en la clase Aeropuerto.
- tamaño (integer): número entero que indica el tamaño en metros cuadrados de un parque. Esta presente solamente en la clase Parque.
- especialidad (String): Especialidad de un museo. Sus valores pueden ser únicamente: “Arqueológico”, “Religioso”, “Moderno”, “Artesanías”, y “Ciencia”. Esta presente solamente en la clase Museo.



Actividad 2: Transformación de Recursos No-Ontológicos

Para el caso de nuestra ontología de Turismo no se utilizarán plantillas existentes, si no que se creará una a partir de los recursos encontrados. El modelo conceptual para los componentes listados en la Actividad 1; es el siguiente:

- CLASE: Lugar Turístico.

COMPONENTES:

- Nombre
 - Dirección
-
- Hotel.
 - Año de Fundación
 - Categoría
 - Cerca del Aeropuerto
 - Horario
 - Iglesia.
 - Año de Fundación



- Parque.
 - Año de Fundación
 - Tamaño
- Museo.
 - Año de Fundación
 - Especialidad
 - Horario
- Restaurante.
 - Horario

La Jerarquía de estas Clases se puede observar en la figura 4.11:

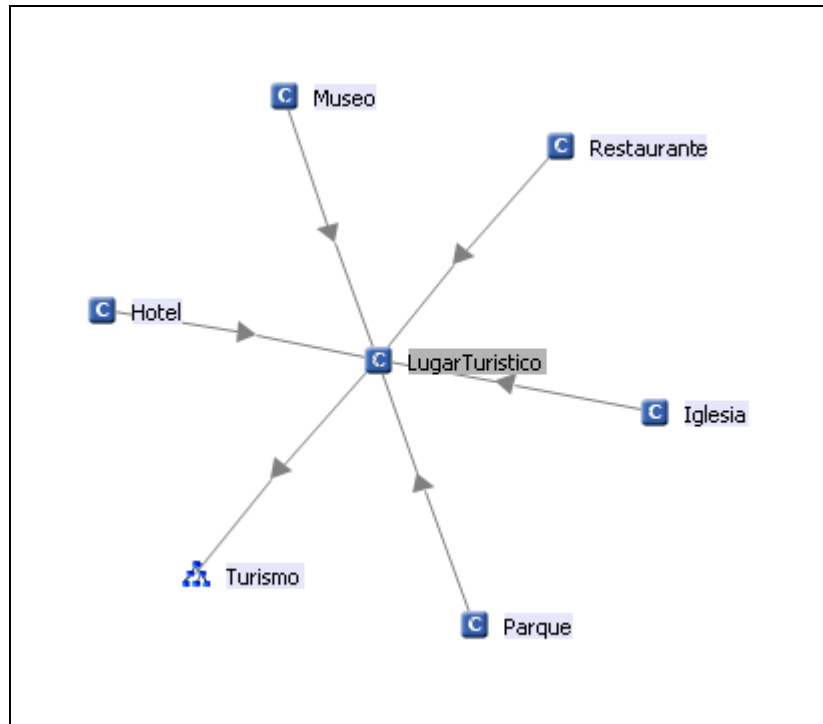


Figura 4.11. Jerarquía de la Ontología de Turismo, para la Actividad 3;
Reingeniería de Recursos No Ontológicos

Actividad 3: Ingeniería Progresiva de Ontologías

Para obtener la representación formal de los conceptos obtenidos para nuestra ontología, se procede a implementarla en lenguaje OWL utilizando la plataforma



NeOn Toolkit versión 2.3.2, descargada de la página principal de la comunidad oficial NeOn¹⁴.

Para la creación de Ontologías; se siguen los siguientes pasos:

1. Creación de CLASES

En el proceso de definición y jerarquización de clases, debemos analizar la lista de términos del punto anterior, clasificando y considerando únicamente los de relevancia alta.

NeOn Tool Kit provee una interfaz intuitiva en la cual se pueden crear las clases para una ontología y agregar propiedades a las mismas; esto se puede apreciar en la figura 4.12:

¹⁴ <http://neon-toolkit.org/wiki/Download>,

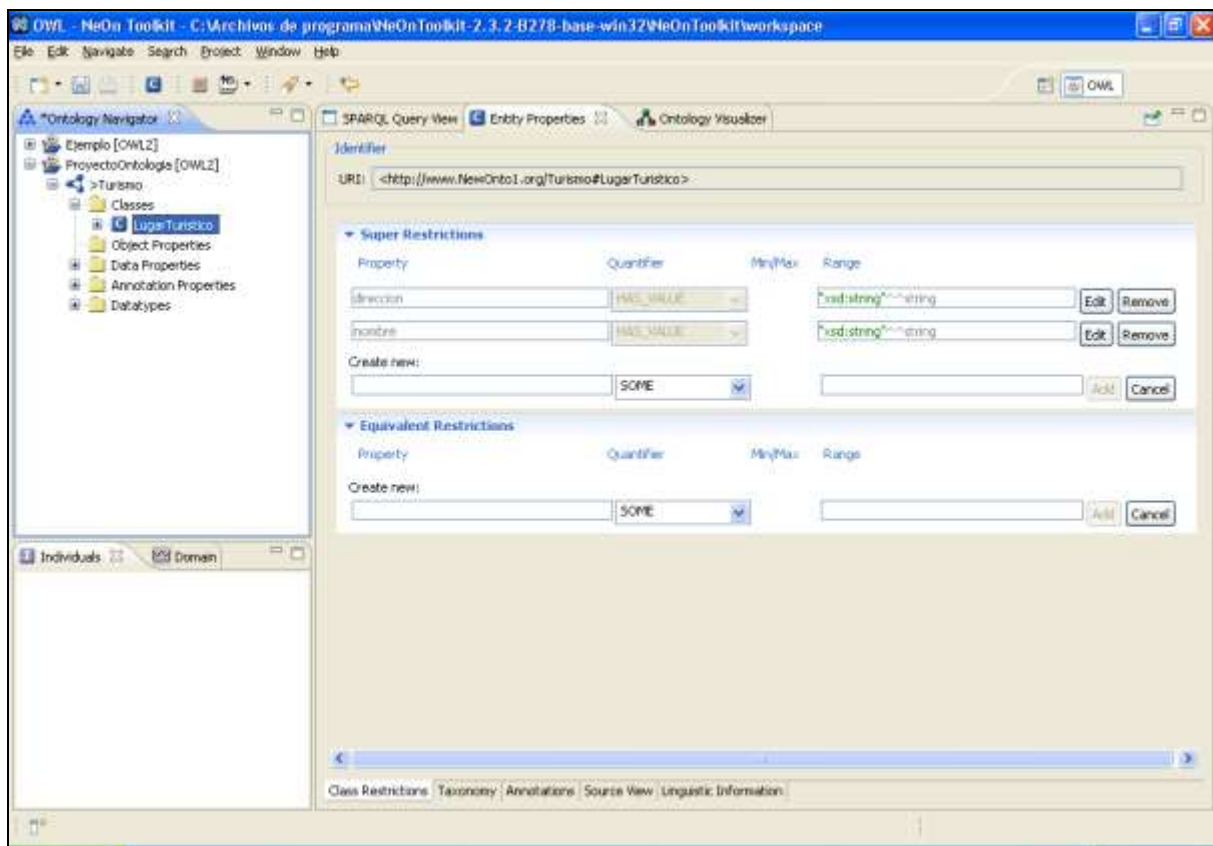


Figura 4.12. Interfaz de NeOn Tool Kit para la creación de Clases de una Ontología

Una vez creadas todas las clases en NeOn Tool kit, se puede observar que constan dentro de la ontología en orden jerárquico, esto se puede observar en la figura 4.13:

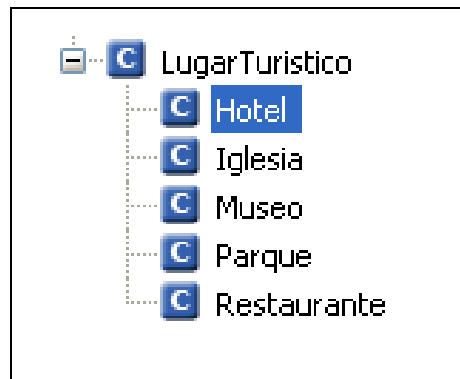


Figura 4.13. Clases de la Ontología de Turismo, para la Actividad 3 del Paso 3:
Reingeniería de Recursos No-Ontológicos

2. AGREGAR PROPIEDADES

Definidas las clases, debemos describir la estructura interna de los conceptos, es decir, crear propiedades para las clases; esto se realiza considerando que un slot (propiedad) deberá estar adosado a la clase más general que pueda tener esa propiedad, para luego poder ingresar las instancias particulares.

Generalmente las instancias pueden tener diferentes facetas para cada slot, así como también diferente cardinalidad, pudiendo esta ser simple (admite un solo valor por slot) o múltiple (admite cualquier cantidad de valores).



De igual manera que con las clases, NeOn Tool kit permite crear propiedades y adjuntarlas a las clases de la ontología, esto se ve en la figura 4.14:

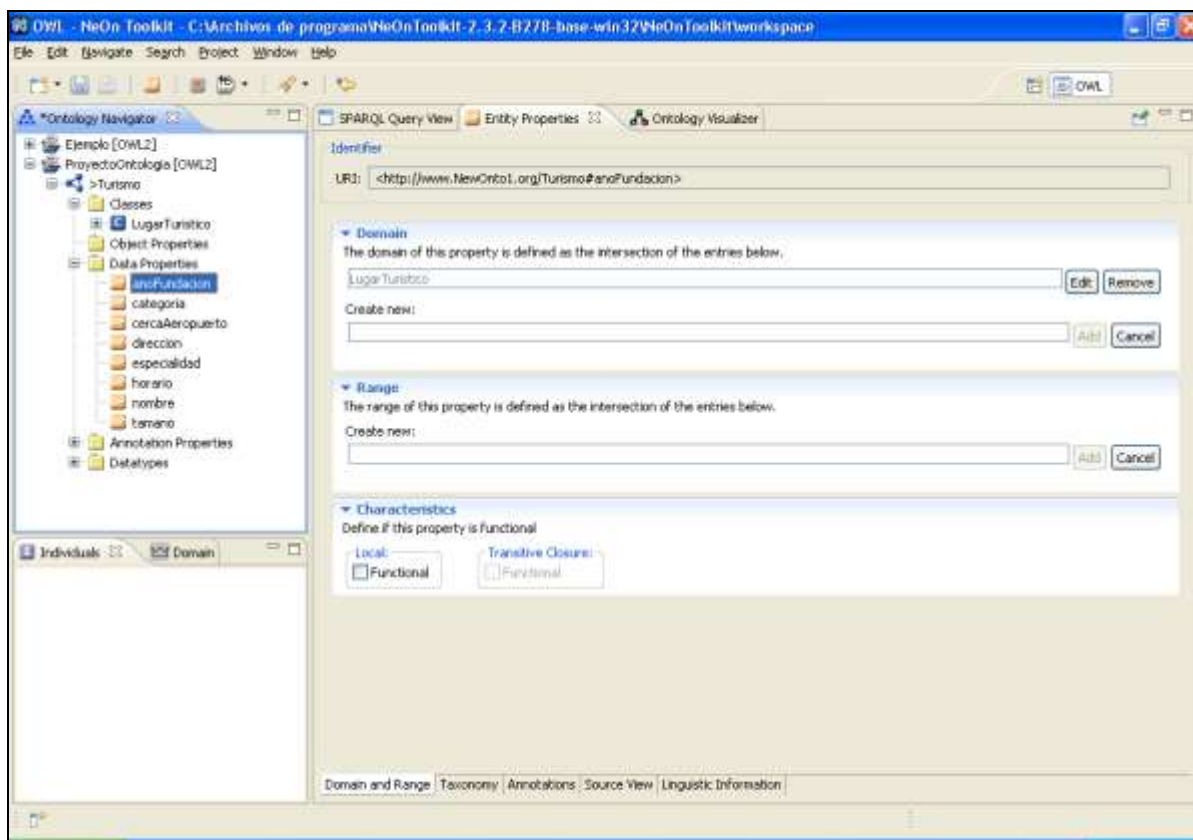


Figura 4.14. Interfaz de NeOn Tool Kit para la agregación de propiedades

Una vez agregadas todas las propiedades, podemos ver que la ontología cuenta con las siguientes propiedades de datos (Data Properties), que permiten describir las clases:

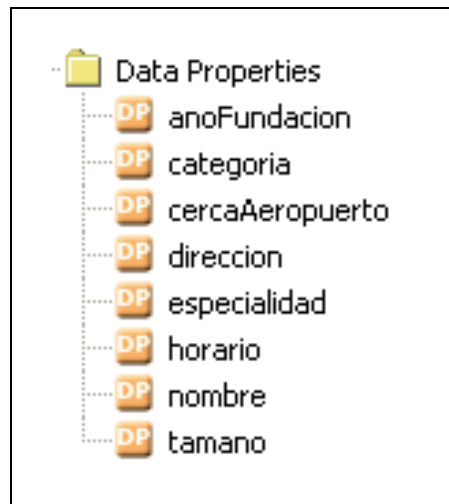


Figura 4.15. Propiedades de Datos de la Ontología de Turismo

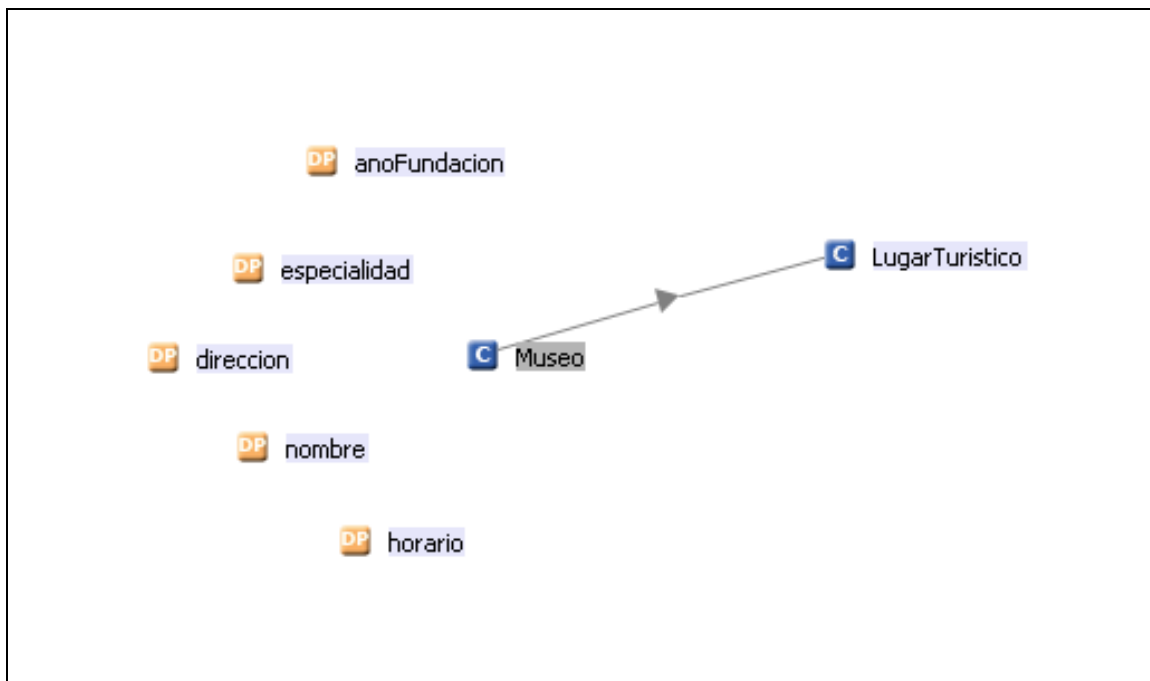


Figura 4.16. Propiedades de Datos de la Clase Museo



3. AGREGAR INSTANCIAS

Una instancia particular u objeto es la representación de algo real, un ejemplo claro podría darse al definir la Clase “Iglesia”, que tiene como una propiedad a la “ubicación”, para este caso una instancia específica sería la “Catedral de la Inmaculada”, cuya ubicación es “Calle Benigno Malo y Sucre”.

Para propósito de responder a las CQs resultantes de la aplicación de la Metodología NeOn; y, para verificar la funcionalidad del prototipo de buscador semántico; se han agregado las siguientes Instancias en la ontología de turismo de la ciudad de Cuenca:

HOTEL

- hotelCrespo
- hotelCarvallo
- hotelItalia

IGLESIA

- catedral
- catedralVieja



- santoDomingo

MUSEO

- arteModerno
- bancoCentral
- museoConceptas

PARQUE

- parqueCalderón
- parqueCandelas
- parqueDeLaMadre

RESTAURANTE

- mondayBlue
- raymipampa
- zoe

La figura 4.16 muestra la creación de la Instancia “CatedralVieja” para la clase “Iglesia”.

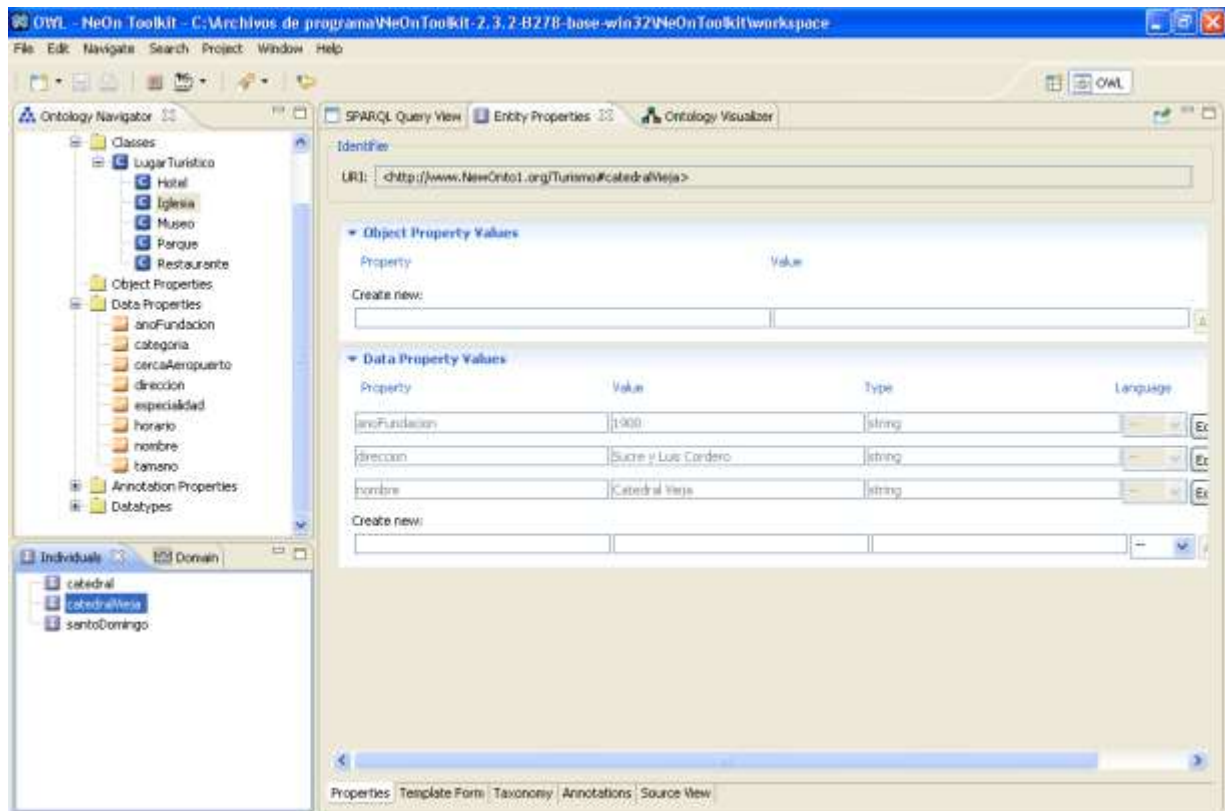


Figura 4.16. Instancia “catedralVieja”

De esta manera queda completa la transformación de los recursos no-ontológicos; y queda como resultado de esta actividad la ontología Turismo.owl; misma que será modificada en la siguiente actividad de la Metodología NeOn.



4.5. PASO 4: Reutilización de Recursos Ontológicos

La reutilización de Recursos Ontológicos es el proceso de utilizar recursos ontológicos disponibles, tales como ontologías, módulos, sentencias o plantillas de diseño de ontologías, en la solución de diferentes problemas de desarrollo de aplicaciones basadas en ontologías.

La reutilización de recursos ontológicos reduce significativamente el tiempo y los costos asociados al desarrollo de ontologías. En esta actividad, la Metodología NeOn, nos permite analizar si es posible reutilizar recursos ontológicos para construir una ontología o adaptarla a una red de ontologías. Este proceso se ve reforzado por el creciente número de ontologías, librerías de ontologías y repositorios ontológicos disponibles en la Web.

El proceso de reutilización de ontologías depende del tipo de ontología a reutilizar, tal como se presenta en la Figura 4.17. Se puede considerar la reutilización de ontologías generales o de ontologías de dominio basándose en los términos extraídos del documento de especificación de requerimientos de la ontología (ORSO).



Es aconsejable consultar una ontología existente, ya que, tomando como base una ontología ya desarrollada, se pueden tomar ideas o fragmentos de ésta y reutilizarlos en la ontología a desarrollar. Esta tarea de reutilización cobra importancia cuando se desea interactuar con aplicaciones que ya definen Ontologías en concreto.

También es útil integrar dos ontologías que describen una conceptualización de un dominio común; es decir, que están constituidas por un conjunto importante de conceptos comunes, tales como clases y restricciones que puedan ser utilizados entre ellas y con otras Ontologías del mismo dominio. Para esto se puede reutilizar ya sea Ontologías Generales, Ontologías de Dominio, o solamente Sentencias de Ontologías.

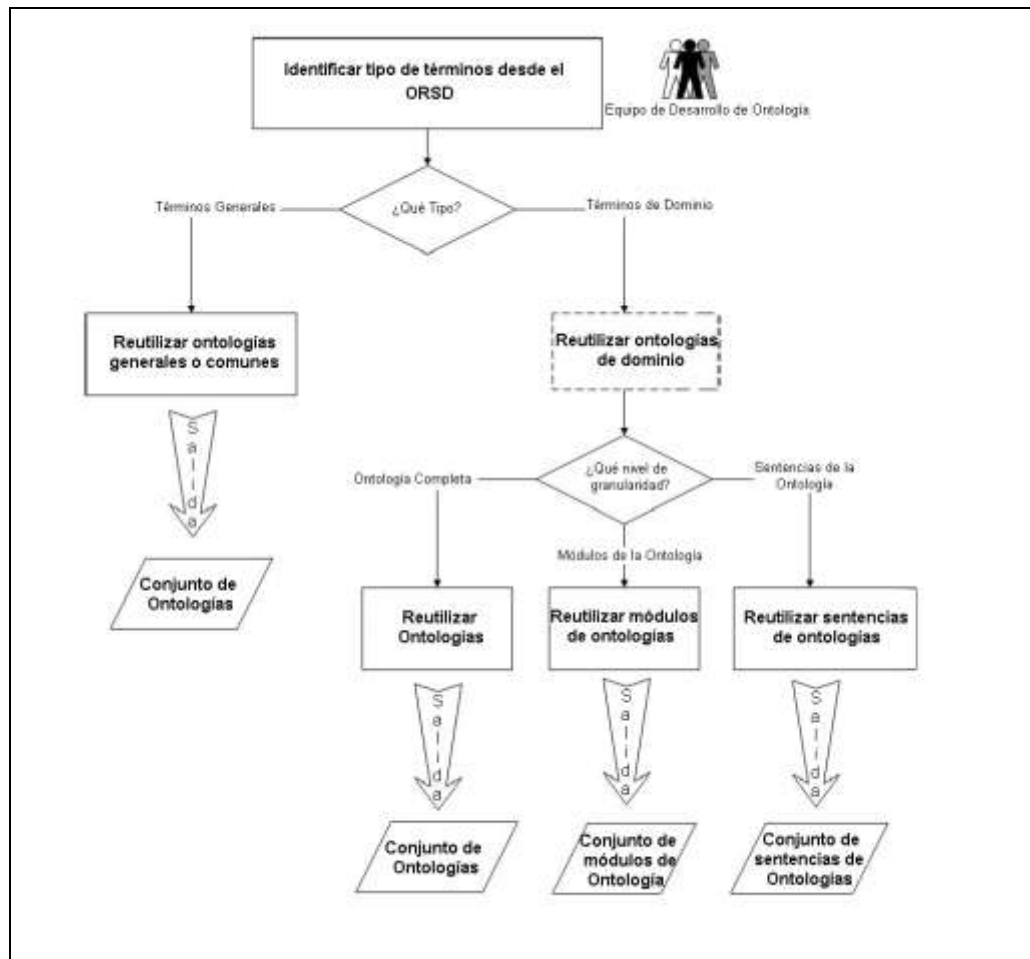


Figura 4.17. Diferentes tipos de Reutilización de Recursos Ontológicos

Las ontologías generales o comunes, proveen una conceptualización de un tema genérico, como por ejemplo tiempo o espacio. Dada la generalidad del tema descrito, es común tener muchas ontologías que traten del mismo tema, cada una de ellas con un punto de partida diferente en cuanto a la especificación del tema. Cuando se reutiliza este tipo de ontologías, es necesario tomar en cuenta los



diferentes puntos de vista y suposiciones que tiene la ontología. Y dado que, por lo general, estas ontologías son bien formadas y contienen sus propias teorías; es común reutilizarlas completamente.

Por otra parte, las ontologías de dominio presentan un conocimiento concreto sobre un dominio amplio pero no general; por ejemplo Medicina, Farmacia, Turismo, etc. Estas ontologías pueden ser útiles cuando se va a construir una ontología del mismo dominio. Al contrario de las ontologías generales, estas ontologías pueden ser reutilizadas a diferentes niveles, tal como se indica en la Figura 4.17.

Es importante tomar una decisión en cuanto al nivel de reutilización; esto es, si se van a reutilizar muchos recursos ontológicos, pocos recursos, o ninguno. En el caso de reutilizar muchos recursos, se puede llegar a un buen resultado, pero se podría estar utilizando los recursos de manera inapropiada e ineficiente. Mientras que si no se reutiliza ningún recurso, se puede llegar a una solución exitosa y creativa pero con un alto costo de tiempo y esfuerzo.

Basados en el criterio anterior, la metodología NeOn clasifica los tipos de reutilización de Ontologías de acuerdo a la Figura 4.18:

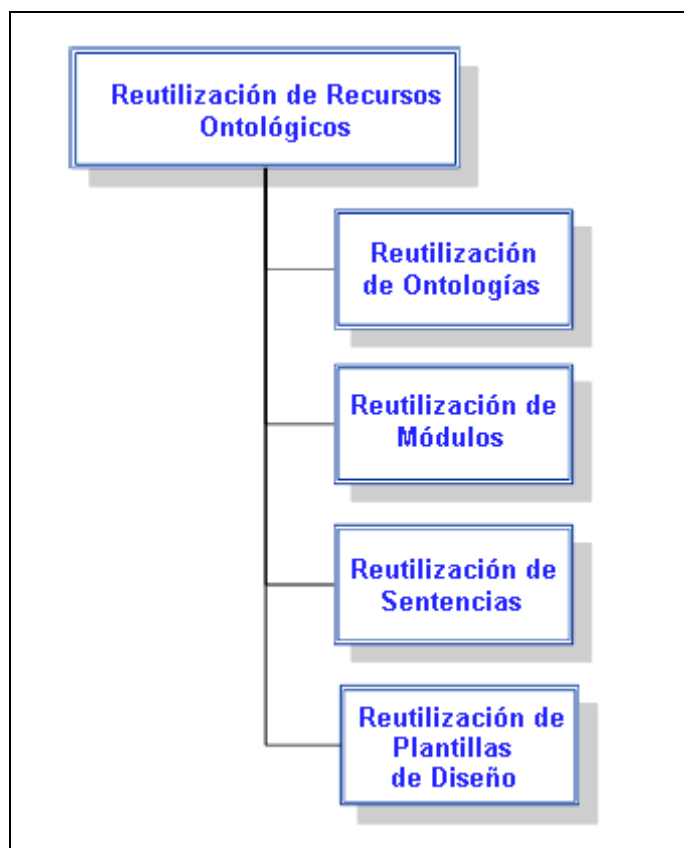


Figura 4.18. Definiciones de la Reutilización de Recursos Ontológicos

Criterios Generales para la Reutilización de Recursos Ontológicos

La Metodología NeOn presenta un conjunto de criterios generales para utilizar en este proceso. Estos criterios han sido obtenidos de la experiencia en la reutilización de recursos ontológicos en la construcción de ontologías para diferentes proyectos; y son presentados a continuación:



- **Cobertura de Dominio del Recurso Ontológico:** Significa analizar si el recurso cubre total o parcialmente el conjunto de CQs del ORSD.
- **Esfuerzo de la Reutilización del Recurso Ontológico:** Se refiere a la estimación del esfuerzo que sería necesario para utilizar el recurso; en este caso se toma en cuenta el costo económico, el tiempo, y la accesibilidad al recurso.
- **Esfuerzo del Entendimiento del Recurso Ontológico:** Se refiere al esfuerzo necesario para analizar y entender el recurso. En este caso se toma en cuenta la documentación del recurso, disponibilidad de sus desarrolladores, y claridad en la terminología del recurso.
- **Esfuerzo de la Modularización del Recurso Ontológico:** Es el esfuerzo necesario para extraer una parte de la ontología en el caso de no reutilizar la ontología completa. En este caso se debe considerar si la ontología ya esta modularizada, y si las herramientas que permitan extraer e identificar los módulos de la ontología están disponibles.
- **Esfuerzo de la Integración del Recurso Ontológico:** Se refiere a la estimación del esfuerzo que se necesita para integrar el recurso en la ontología a desarrollar; en este caso se consideran criterios como la similaridad entre las convenciones de nombres del recurso y de la ontología



a desarrollar; la similaridad en el lenguaje de implementación, las contradicciones en el conocimiento, y la adaptación de las definiciones y axiomas que pueden satisfacer las restricciones existentes de la ontología a desarrollar.

- **Confiabilidad del Recurso Ontológico:** Es el análisis de la confiabilidad del recurso a reutilizar. En este caso se toman en cuenta las pruebas realizadas con este recurso, la teoría que lo soporta, si el equipo de desarrollo es confiable, y si existen proyectos conocidos que hayan reutilizado este recurso anteriormente.

Reutilización de ontologías comunes o generales

La reutilización de ontologías comunes consiste en encontrar y seleccionar ontologías existentes y reutilizarlas en el desarrollo de nuevas ontologías. En esta actividad, se pueden considerar dos situaciones diferentes:

Situación 1. Cuando existe un estudio comparativo previo de la teoría que soporta el tipo de ontologías comunes a ser reutilizadas, por ejemplo un estudio de modelamiento.



Situación 2. Cuando no existe un estudio comparativo sobre la teoría.

La Tabla 4.8 muestra la plantilla para el proceso de reutilización de ontologías generales:

General or Common Ontology Reuse	
<i>Definición</i> Es el proceso de utilizar ontologías comunes o generales en la solución de problemas diferentes.	
<i>Objetivo</i> Encontrar y seleccionar ontologías comunes o generales para integrarlas en las redes de ontologías a desarrollar.	
<i>Entrada</i> Preguntas de Competencia incluidas en el ORSD de la ontología a desarrollar, y el lenguaje de implementación de dicha ontología.	<i>Salida</i> Una ontología general o común integrada en la ontología a ser desarrollada.
<i>Quien</i> Desarrolladores de software y profesionales de ontologías involucrados en el desarrollo de la ontología.	
<i>Cuando</i> Este proceso debería llevarse a cabo después de la especificación de la ontología.	



Tabla 4.8. Plantilla para la Reutilización de Ontologías Generales

Las Figuras 4.19 y 4.20 muestran los pasos a seguir para la Reutilización de Ontologías Generales en las situaciones 1 y 2, respectivamente.

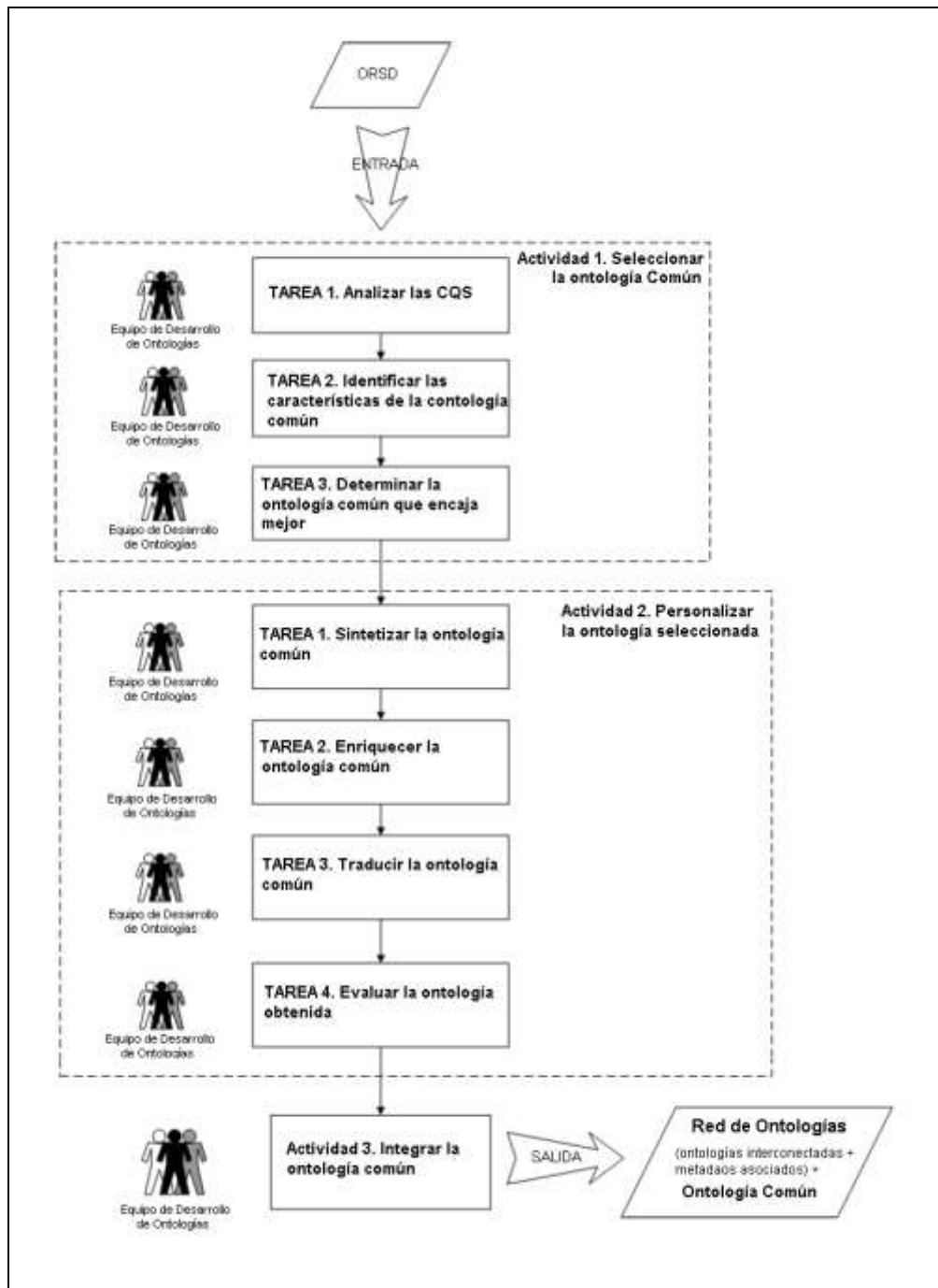




Figura 4.19. Actividades para la reutilización de Ontologías Generales en la
Situación 1

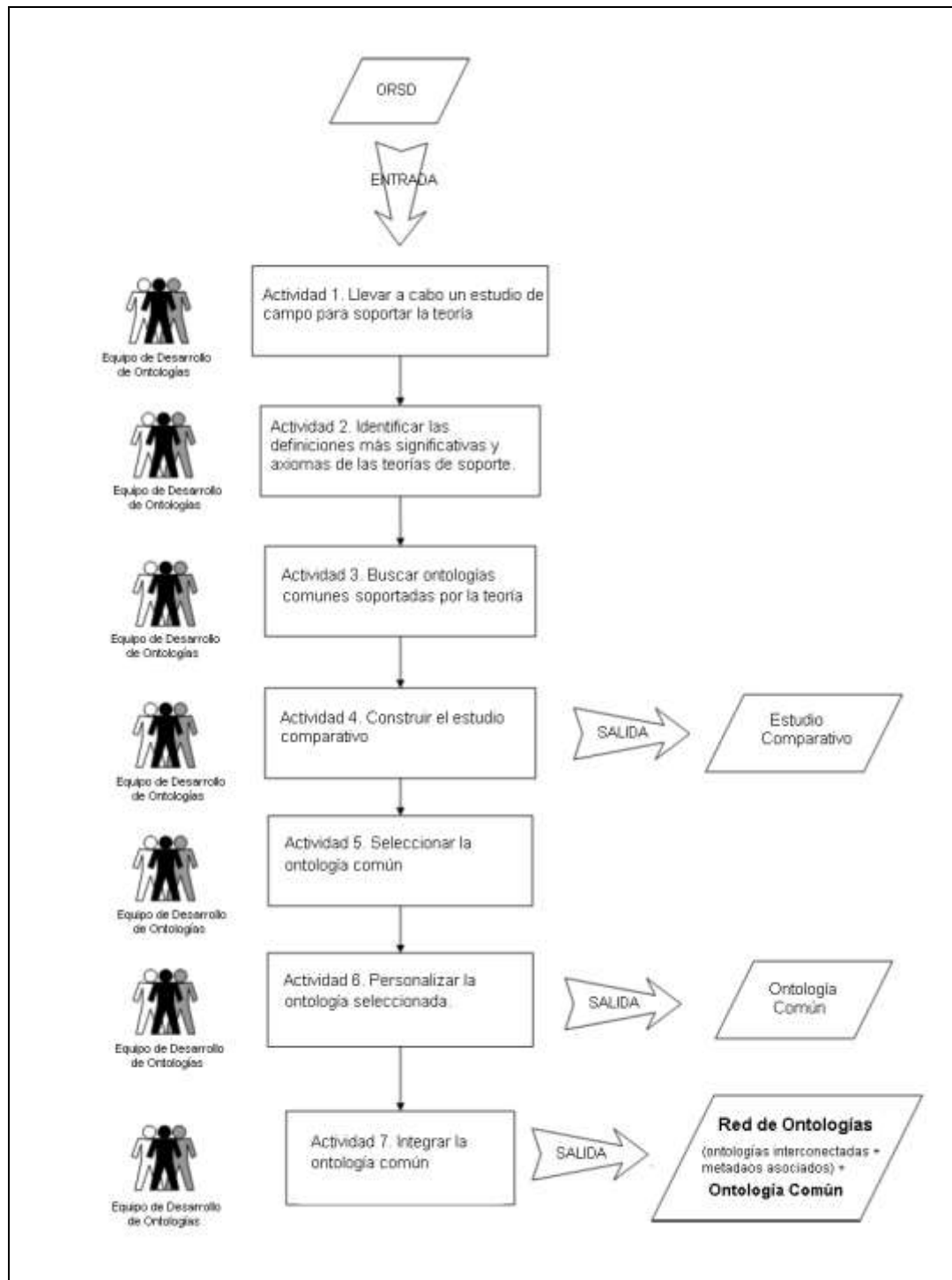




Figura 4.20. Actividades para la reutilización de Ontologías Generales en la Situación 2

Reutilización de ontologías de Dominio

El objetivo de la Reutilización de ontologías de Dominio es seleccionar una o más ontologías relacionadas con el dominio de la ontología a desarrollar y utilizarlas en la construcción de dicha ontología.

La Tabla 4.9 muestra la plantilla para el proceso de Reutilización de ontologías de Dominio:



Reutilización de Ontologías de Dominio	
<i>Definición</i> Es el proceso de utilizar ontologías de dominio en la solución de diferentes problemas.	
<i>Objetivo</i> Encontrar y seleccionar una o varias ontologías de dominio relacionadas con el dominio de la ontología a ser desarrollada.	
<i>Entrada</i> EI OSRD.	<i>Salida</i> Una red de ontologías extendida con la ontología de dominio reutilizada.
<i>Quien</i> Desarrolladores de software y profesionales de ontología.	
<i>Cuando</i> Se lleva a cabo después de la actividad de especificación de la ontología.	

La Tabla 4.9. Plantilla para la reutilización de ontologías de Dominio

Las actividades para llevar a cabo el proceso de reutilización de ontologías de Dominio se encuentran en la Figura 4.21:

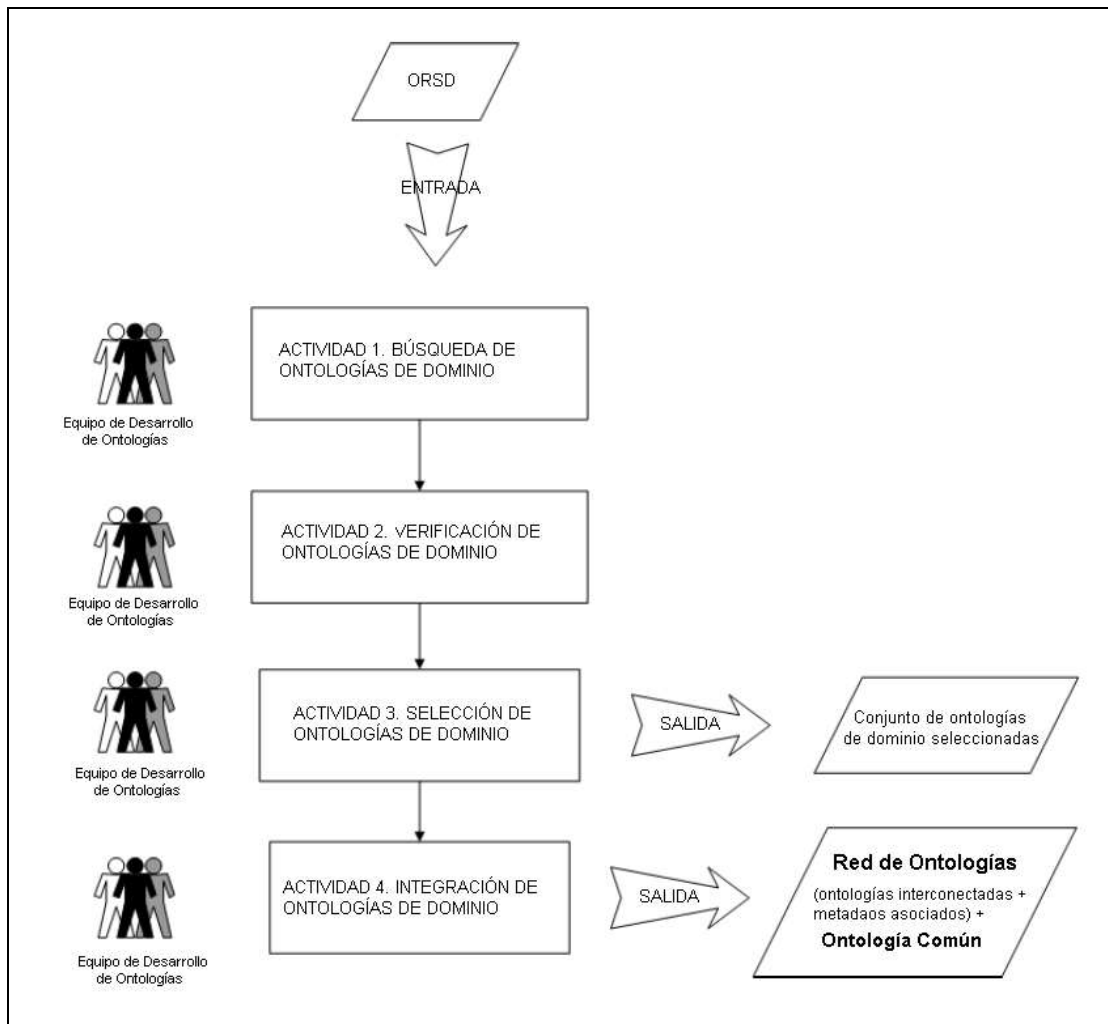


Figura 4.21. Actividades para Reutilizar Ontologías de Dominio

Reutilización de Sentencias de Ontologías



La Tabla 4.10 muestra la plantilla para el proceso de Reutilización de sentencias de ontología incluyendo su definición, objetivo, entradas y salidas, quien debe llevar a cabo este proceso y cuando debe ser ejecutado.

Reutilización de Sentencias de Ontología	
<i>Definición</i> Es el proceso de utilizar sentencias de ontologías de dominio, en la solución de diferentes problemas.	
<i>Objetivo</i> Utilizar sentencias de ontologías de una ontología que no fue desarrollada originalmente para el problema actual.	
<i>Entrada</i> El OSRD y sentencias de ontologías disponibles, en el mismo dominio de la ontología a desarrollar.	<i>Salida</i> Red de ontologías extendida con la reutilización de sentencias de ontologías.
<i>Quien</i> Desarrolladores de software y profesionales de software.	
<i>Cuando</i> Este proceso se puede llevar a cabo en varios momentos del ciclo de vida de la ontología. Por lo general se realiza al momento de construir la ontología y puede ser útil en varias situaciones. También se puede llevar a cabo en las últimas etapas del ciclo de vida cuando se actualiza o extiende la ontología.	



Tabla 4.10. Plantilla del Proceso de Reutilización de Sentencias Ontológicas

La reutilización de sentencias ontológicas se puede aplicar en dos situaciones diferentes:

Situación 1: Construir redes de ontologías desde cero. En este caso es de gran utilidad si ya existe un modelo preliminar de la ontología establecido, y tener una idea clara de los requerimientos de la ontología, su dominio y sus objetivos.

Situación 2: Extender o mejorar una red de ontologías existente.

Las actividades que permiten llevar a cabo el proceso de reutilización de sentencias ontológicas se muestran en la Figura 4.22.

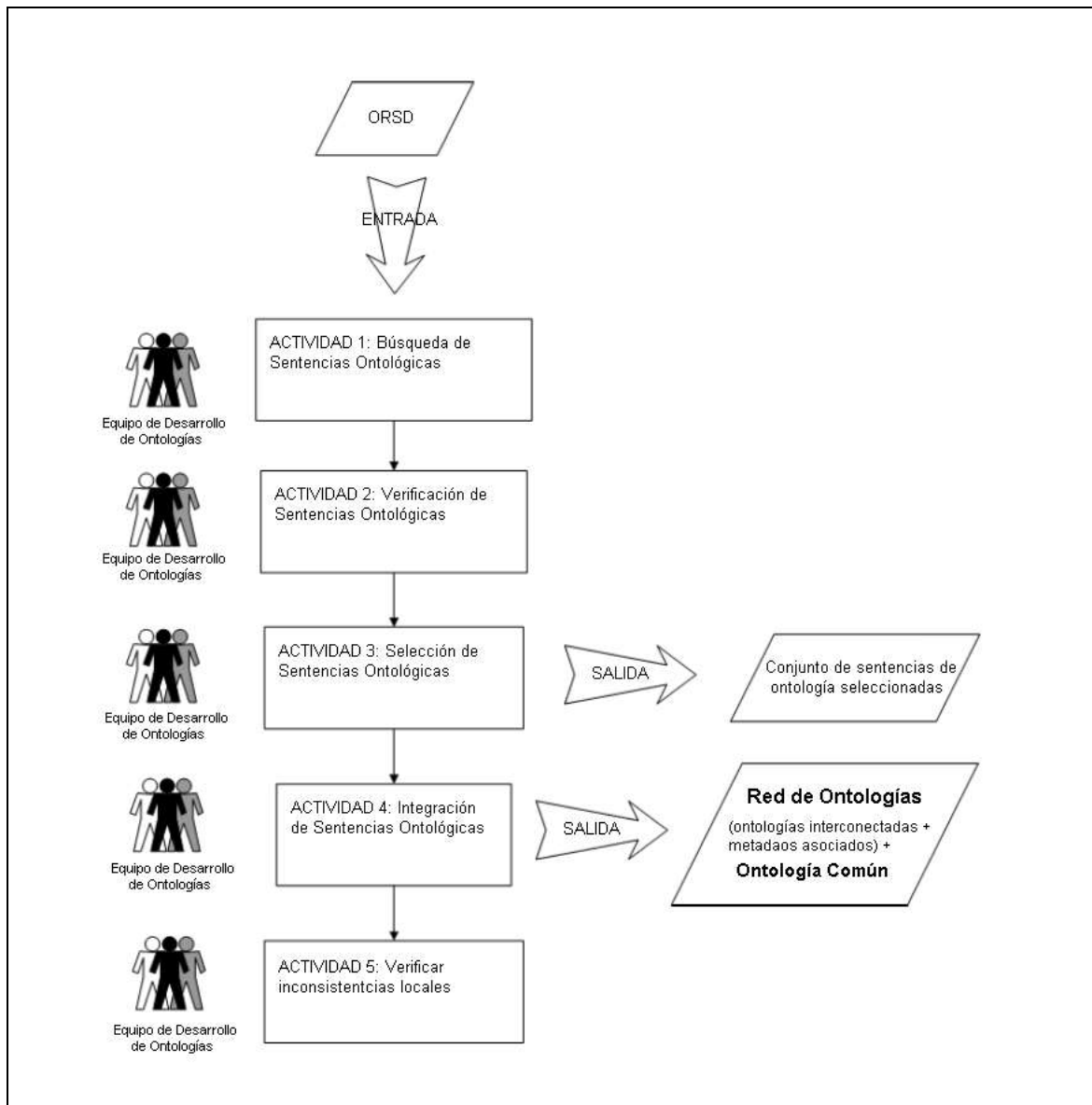


Figura 4.22. Actividades para la reutilización de sentencias de Ontologías



Reutilización de Recursos Ontológicos para la Ontología de Turismo

Para el caso de nuestra Ontología de Turismo, se reutilizarán Ontologías de Dominio, es por esto que utilizamos las actividades de la figura 4.21.

Actividad 1: Búsqueda de Ontologías de Dominio

Hemos consultado las librerías de ontologías Swoogle y Watson, para encontrar ontologías de dominio de Turismo publicadas en la Web; además se revisaron algunos artículos Web sobre el estado del arte de las Ontologías de Turismo^[37].

Actividad 2: Verificación de Ontologías de Dominio

Las ontologías analizadas presentan características diferentes de acuerdo al área de aplicación para la que fueron creadas, y muchas de estas no están disponibles para su descarga y utilización de manera libre. Sin embargo todas contienen características relacionadas al dominio Turístico.

Actividad 3: Selección de Ontologías de Dominio



Del conjunto de Ontologías analizadas se seleccionaron la ontología Qall-Me^[38] y la ontología Harmonise^[39]. Estas ontologías fueron seleccionadas debido a su relación con nuestra Ontología de Turismo, y la similaridad en la clasificación de sus clases y subclases.

Actividad 4: Integración de Ontologías de Dominio

Puesto que las ontologías analizadas están descritas en idioma inglés, no es posible integrarlas directamente a nuestra ontología. Es por esto que se reutiliza solamente algunos de sus conceptos, mismos que serán reescritos para nuestro caso.

De acuerdo a las Clases y sub clases encontradas en las ontologías a reutilizar; hemos agregado las siguientes clases a nuestra ontología:

Atracción Turística: Contiene todos los lugares Turísticos que son objetos de visita por su atractivo turístico. Esta es una sub clase de *LugarTurístico*, y contiene, por lo tanto sus características heredadas de *Nombre* y *Dirección*, además contiene las siguientes características propias:

- Año de Fundación

Las subclases de la clase *AtraccionTuristica* son:



- Exposición
- Parque
- Templo

Exposición: Son los lugares turísticos que presentan exposiciones o exhibiciones de arte para los visitantes. Esta es una sub clase de *AtraccionTurística*, y contiene, por lo tanto sus características heredadas de *Nombre, Dirección y Año de Fundación*, además contiene las siguientes características propias:

- Especialidad
- Horario

Las subclases de la clase *Exposición* son:

- Galería
- Museo
- Plaza de Arte

Parque: Contiene todos los tipos de parques de la ciudad. Esta es una sub clase de *AtraccionTurística*, y contiene, por lo tanto sus características



heredadas de *Nombre*, *Dirección* y *Año de Fundación*, además contiene las siguientes características propias:

- Tamaño

Las subclases de la clase *Parque* son:

- Parque arqueológico
- Parque natural
- Parque Público
- Plazoleta

Templo: Contiene todos los lugares Turísticos en donde se llevan a cabo ceremonias religiosas. Esta es una sub clase de *AtraccionTurística*, y contiene, por lo tanto sus características heredadas de *Nombre*, *Dirección* y *Año de Fundación*, además contiene las siguientes características propias:

- Categoría; en este caso la categoría hace referencia al tipo de ceremonia religiosa; pudiendo ser católica, cristiana, evangélica, etc.

Las subclases de la clase *Templo* son:

- Iglesia



- Capilla

Comedor: Contiene todos los lugares Turísticos en donde se ofrecen alimentos. Esta es una sub clase de *LugarTurístico*, y contiene, por lo tanto sus características heredadas de *Nombre y Dirección*, además contiene las siguientes características propias:

- Horario

Las subclases de la clase *Comedor* son:

- Bar
- Fonda
- Restaurante

Hospedaje: Contiene todos los lugares Turísticos en donde se ofrece hospedaje temporal. Esta es una sub clase de *LugarTurístico*, y contiene, por lo tanto sus características heredadas de *Nombre y Dirección*, además contiene las siguientes características propias:

- Categoría
- Cerca del Aeropuerto



- Horario

Las subclases de la clase *Hospedaje* son:

- Hostería
- Hotel
- Restaurante

Transporte: Contiene todos los medios de Transporte disponibles en la ciudad de Cuenca. Contiene las siguientes características:

- Nombre
- Dirección
- Categoría

Las subclases de la clase *Transporte* son:

- Transporte Aéreo
- Transporte Terrestre



Las clases agregadas a la ontología contienen solamente propiedades que ya existían en nuestra ontología de Turismo y agregan una organización jerárquica a la misma.

La estructura final de las Clases de la Ontología de Turismo, y su jerarquía se muestra en la figura 4.23.

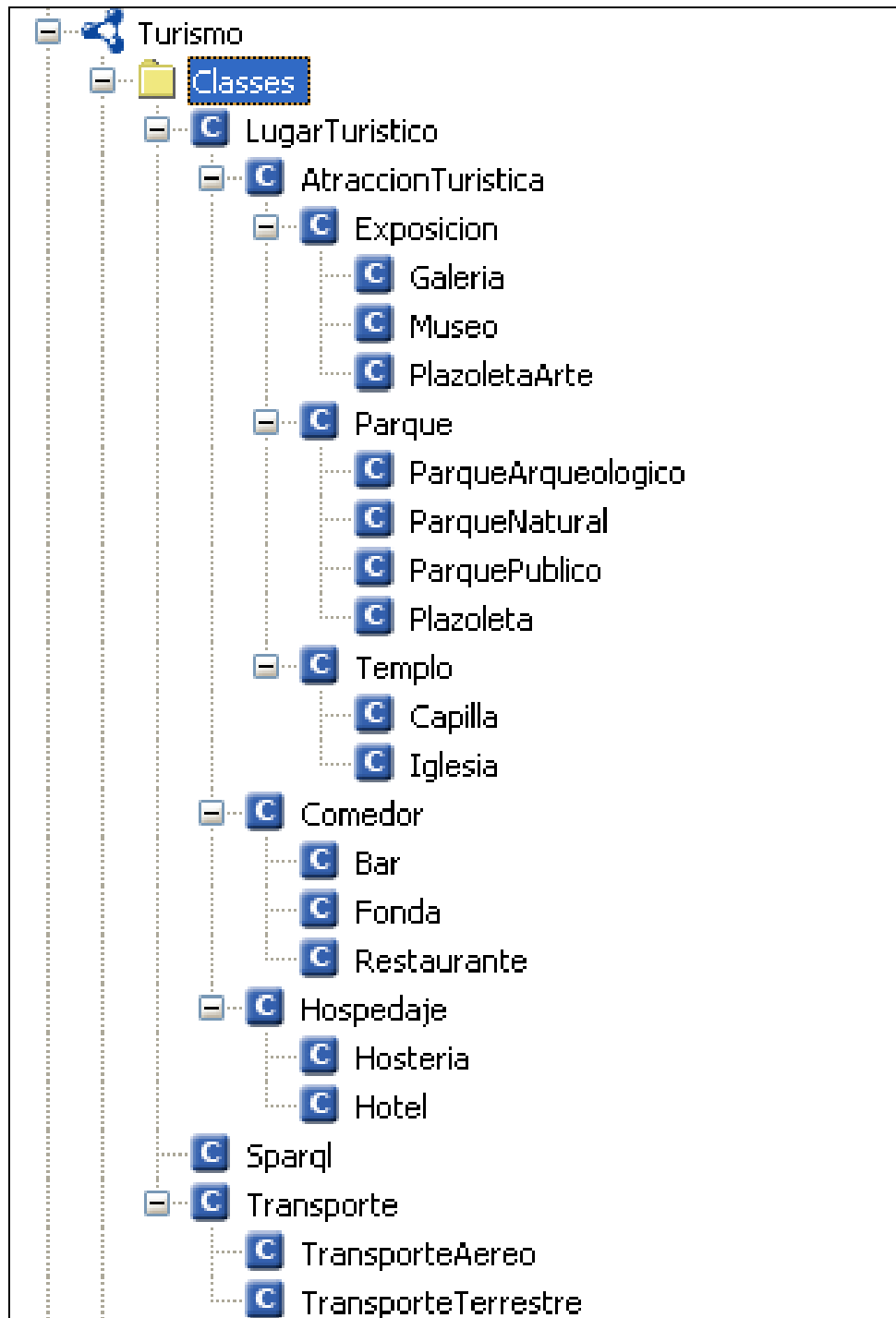




Figura 4.23. Jerarquía de Clases de la Ontología de Turismo, para la Actividad 4 del Paso 4: Reutilización de Recursos Ontológicos

4.6. PASO 5: Reutilización de Plantillas de diseño para Ontologías

La reutilización de plantillas de diseño, permite designar soluciones de modelado, que ya han sido utilizadas recurrentemente, para resolver problemas de diseño. La reutilización de plantillas de diseño en la programación orientada a objetos es una práctica común y existen repositorios de plantillas y manuales tales como el Gamma^[40]. Actualmente, estos repositorios se encuentran integrados en herramientas de software que permiten un acceso y una integración de las plantillas de manera rápida. Sin embargo, todos estos requieren de un conocimiento de diseño previo por parte de sus usuarios.

Otra limitación para su aplicación, es la falta de metodologías generales o estándares para la reutilización de diferentes repositorios de plantillas; además las plantillas contienen diferentes estilos, de tal manera que las recomendaciones dadas por ciertos autores no pueden ser extrapoladas o reutilizadas para la aplicación de otros repositorios.

La reutilización de plantillas de Diseño en la Ingeniería Ontológica, es una práctica que no se encuentra muy difundida ya que las ODPs (ontology design patterns) no



cuentan con repositorios actualizados ni amplios; y al igual que en la programación orientada a objetos, no existen metodologías guías que faciliten la reutilización de las plantillas. Sin embargo existen iniciativas que ayudan a los usuarios en el proceso de adaptar o implementar ODPs por medio de asistentes (Wizards).

Para permitir que el usuario promedio tenga mayor acceso a las ontologías, sería necesario que se lleven a cabo los siguientes avances:

- Crear plantillas estandarizadas para la descripción de los ODPs y que sean entendibles para diferentes tipos de usuarios;
- Crear métodos generalizados o guías de usuario sin experiencia previa en la reutilización de ODPs.
- Creación de técnicas y herramientas para transformar la selección de plantillas automática o semiautomáticamente.

Al momento; los métodos que se siguen para la reutilización de repositorios de plantillas se limitan a algunos consejos y recomendaciones. Los usuarios expertos tienen acceso a estos repositorios gracias a medios de criterio analógico, basados en su experiencia y en las descripciones incluidas en los catálogos. De esta manera seleccionan las plantillas mas adecuadas a sus necesidades.



La reutilización de plantillas de diseño, es un paso opcional en la Metodología NeOn, puesto que los pasos anteriores son suficientes para la construcción de una ontología completa.

No se llevará a cabo este paso de la Metodología NeOn para el caso particular de nuestra ontología de Turismo ya que no contamos con plantillas apropiadas que permitan su uso. Además la ontología creada es correcta y suficiente para cumplir con los objetivos del presente trabajo, y se encuentra dentro de las restricciones del mismo.

4.7. Conclusiones

La Metodología NeOn para la creación de ontologías, nos permite crear una ontología desde cero o bien reutilizando ontologías similares que ya existan. Esta metodología provee plantillas y herramientas para dicha tarea; y es de gran utilidad para los profesionales de Ontologías y desarrolladores de Software que pretenden incursionar en el mundo de la Web Semántica, ya que las ontologías son la base de la misma.

La Metodología NeOn se encuentra en constante desarrollo y espera publicar su segunda edición para principios del año 2011. Esta edición contará con un mayor



número de plantillas guía y herramientas para el desarrollo y reutilización de ontologías.

Al momento; se cuenta con un número bastante limitado de recursos para aplicar la metodología NeOn, además de estar al alcance de pocos expertos debido a su complejidad.

Pero sin duda, la metodología NeOn ha sido una herramienta invaluable para la creación de la Ontología de Turismo para el prototipo de Buscador Semántico, del presente trabajo.



CAPÍTULO V

DISEÑO E IMPLEMENTACIÓN

DEL BUSCADOR SEMÁNTICO



El Capítulo Quinto, DISEÑO E IMPLEMENTACIÓN DEL PROTOTIPO DE BÚSCADOR SEMÁNTICO, trata de la planeación y realización del prototipo de buscador semántico en sí; en el transcurso de este capítulo se detallarán las técnicas utilizadas para llevar a cabo el proyecto en sus fases de Análisis, Diseño, Implementación y Pruebas, explicando la tecnología implicada en sus capas: de



Datos, de Negocios e Interfaz. También se presentará la arquitectura del buscador en general, y la manera en que interactúan sus capas y tecnologías entre sí; esto nos permitirá tener una visión general del prototipo, su estructura, y su funcionamiento.

CAPÍTULO V

DISEÑO E IMPLEMENTACIÓN DEL BUSCADOR SEMÁNTICO

El nombre seleccionado para nuestro buscador semántico es “BUTUCU” o Buscador Turístico de Cuenca; ya que trabaja únicamente con información de lugares turísticos de la Ciudad de Cuenca.

El prototipo de buscador semántico implementado, tiene como objetivo aplicar y demostrar algunas de las tecnologías que dan soporte a la Web Semántica. A continuación se presentará todo el proceso de creación del prototipo mediante la explicación de sus fases de Análisis, Diseño, Implementación, y Pruebas.



5.1. Fase de Análisis

En esta fase se van a especificar los requisitos establecidos para la realización del prototipo y se va a analizar el sistema a desarrollar.

Requerimientos del prototipo de Buscador Semántico:

- El buscador debe ser un sistema que permita realizar búsquedas semánticas sobre la Ontología de Turismo creada para este fin.
- El buscador debe ser capaz de interactuar con el usuario a través de la Interfaz, y utilizando lenguaje natural en idioma español.
- El buscador permite contestar preguntas acerca de la información más relevante de Lugares Turísticos de la ciudad de Cuenca, Ecuador. Estas preguntas corresponden a las Preguntas de Competencia detalladas en el capítulo 4, para la creación de la ontología.
- El buscador trabajará solamente con la ontología de Turismo creada en este trabajo de manera estática; pero dicha ontología podría ser modificada dinámicamente durante la ejecución del programa, y éste debe ser capaz de detectar las modificaciones y trabajar con ellas.
- El sistema funciona en una interfaz Web, es intuitivo para el usuario y de fácil uso.



Funcionalidades del Sistema:

- El sistema cuenta con tres capas de programación que interactúan entre sí; capa de Datos, capa de Negocios e Interfaz; y que pueden ser modificadas de manera independiente.
- El sistema recibe una pregunta por parte del usuario en lenguaje natural y devuelve una respuesta a la interfaz; en caso de no encontrar una respuesta, el sistema deberá indicar el motivo del inconveniente.
- El sistema realiza un procesamiento de la pregunta en lenguaje natural para transformarla en una consulta Sparql¹⁵ y enviarla a la capa de Datos.

5.2. Fase de Diseño

La creación de la Ontología de Turismo, descrita en el Capítulo 4; es parte de la fase de Diseño del Buscador semántico; además de esto, la fase de Diseño; basándose en las conclusiones de la fase de Análisis; nos permite concluir que; para la consecución de los requerimientos del programa; el prototipo deberá realizar las siguientes funciones:

¹⁵ Sparql: Lenguaje de consultas introducido en la sección 3.4 del Capítulo 3: Plataforma Tecnológica del Buscador Semántico.



1. La Interfaz recibe la consulta de usuario y, una vez validada, la envía a la Capa de Negocios. En este paso, el usuario ingresa la consulta en lenguaje natural directamente en la interfaz.
2. La Capa de Negocios recibe la consulta y la clasifica de acuerdo a sus términos. Aquí será necesario agregar un módulo de Identificación de términos para clasificar la consulta y que el buscador pueda interactuar con ella.
3. La Capa de Negocios genera el código para la consulta SPARQL y la envía a la Capa de Datos. Será necesario hacer la transformación de la consulta de usuario a una consulta SPARQL para que pueda ser procesada por la capa de Datos.
4. La Capa de Datos ejecuta la consulta SPARQL y envía la respuesta a la Capa de Negocios.
5. La Capa de Negocios genera la respuesta en lenguaje natural, a partir de la respuesta enviada por la Capa de Datos, y la envía a la Interfaz.
6. La Interfaz muestra la respuesta al usuario.

Todos estos pasos deberán ser ejecutados por el prototipo en ese preciso orden, y la interacción de sus capas de programación será una característica clave para el buen funcionamiento del mismo. La figura 5.1 nos muestra el diagrama flujo para

este conjunto de pasos, y la Capa del sistema en que se ejecuta cada uno de estos.

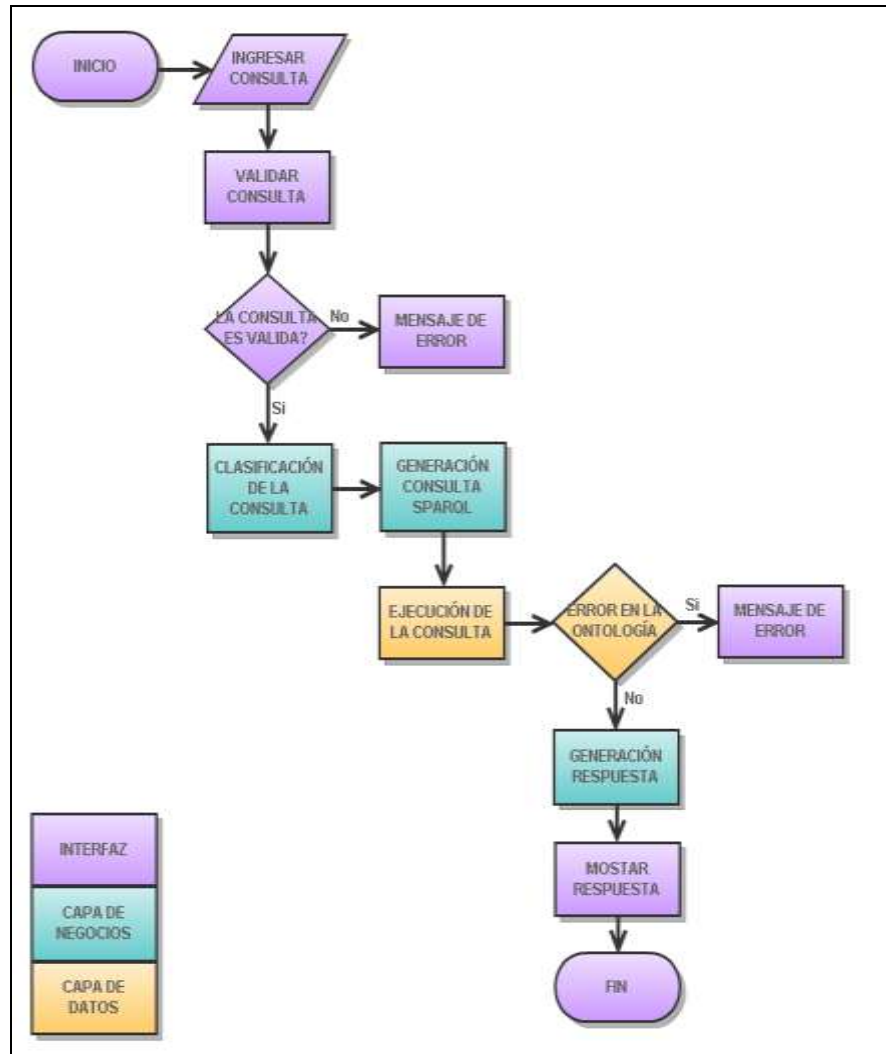


Figura 5.1. Diagrama de Flujo de las Funciones del Buscador Semántico



5.3. Fase de Implementación

A continuación explicaremos la implementación de las Funciones del Buscador Semántico dentro de cada una de sus capas de Programación.

5.3.1. Capa de Datos

La Capa de Datos, en el caso de nuestro proyecto, está constituida por la ontología 'Turismo.owl', que fue desarrollada en la plataforma Neon Toolkit y siguiendo la Metodología Neon, como se encuentra detallado en el Capítulo 4.

Terminado el proceso de desarrollo de la ontología, se procedió a comprobar las consultas utilizando el lenguaje Sparql en Protégé. Este paso es necesario, ya que la función principal de la Capa de Datos es la resolución de consultas Sparql enviadas desde la capa de Negocios, y el envío de las respuestas a la misma.

Se seleccionaron las consultas a realizar de acuerdo a los distintos tipos de consultas que serán necesarias para responder a las CQs¹⁶, obtenidas tras la aplicación de la Metodología NeOn. Cada una de estas consultas nos ayuda a extraer la información de la ontología en varias formas que serán necesarias dentro de la aplicación del buscador semántico.

¹⁶ Preguntas de Competencia



A continuación se explicarán algunos ejemplos de las consultas implementadas:

1. Como un primer ejemplo, explicaremos una consulta en la que intervengan solamente propiedades sobre los datos. Como se muestra en la figura 5.2, la consulta listará todos los lugares turísticos de la ontología junto con su dirección y año de fundación.

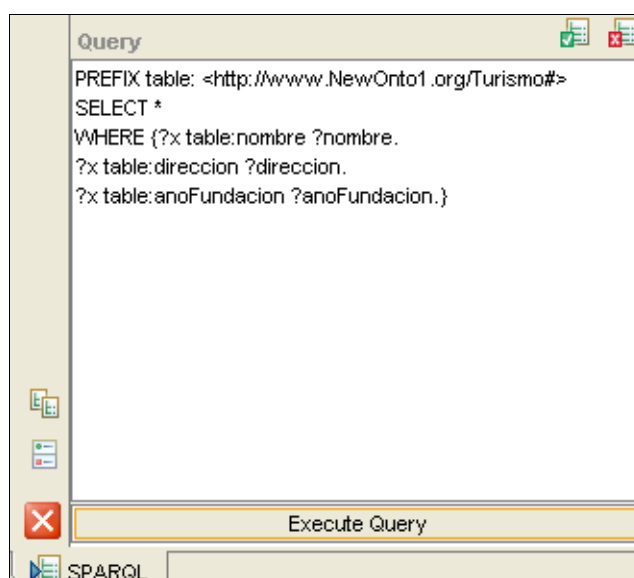


Figura 5.2. Consulta Sparql (1) sobre la Ontología de Turismo

Para la ejecución de esta consulta, Sparql extraerá todas las instancias de la ontología que contengan las propiedades de *Dirección* y *Año de Fundación*, es decir las que corresponden a las clases *Iglesia*, *Museo*, *ParquePublico*, y *Hotel*, pasando por sus clases padre *Templo*, *Exposición*, *Parque* y *Hospedaje* respectivamente. Como se puede ver por la



simplicidad de la consulta en la figura 5.2, este tipo de extracción representa una gran ventaja en comparación con una consulta Sql para una Base de Datos, en la cual sería necesario agregar una cláusula JOIN para agrupar las clases y buscar las coincidencias.

El resultado de la ejecución de esta consulta en Protégé, se muestra en la Figura 5.3. Analizando estos resultados podemos apreciar que en base a la consulta de la figura 5.2, obtenemos el nombre de las instancias de los lugares Turísticos en la columna “nombre”.

Results			
x	nombre	direccion	anoFundacion
◆ santoDomingo	Santo Domingo	Gran Colombia y Padre Aguirre	1917
◆ arteModerno	Museo de Arte Moderno	Sucre y Coronel Talbot	1959
◆ parqueCandelas	Parque de las Candelas	Gonzalo Cordero	1973
◆ hotelCarvallo	Hotel Carvallo	bolivar y padre aguirre	1977
◆ museoConceptas	Museo de las Conceptas	Hermano Miguel y Juan Jaramillo	1599
◆ catedralVieja	Catedral Vieja	Sucre y Luis Cordero	1900
◆ hoteCrespo	Hotel Crespo	Calle Larga y Luis Cordero	1979
◆ catedral	Catedral de la Inmaculada	Benigno Malo y Sucre	1933
◆ hotelItalia	Hotel Italia	av. españa y huaynacapac	1959
◆ bancoCentral	Museo del Banco Central	Calle Larga y Huaynacapac	1992
◆ parqueCalderon	Parque Abdon Calderon	Sucre y Benigno Malo	1900
◆ parqueDeLaMadre	Parque de la Madre	Av. 12 de abril	1977

Figura 5.3. Resultado de la Consulta Sparql (1).

2. Una segunda consulta de ejemplo extraerá la dirección de un Lugar Turístico concreto, en este caso el *Museo del Banco Central*; para esto se utiliza la siguiente sentencia en Sparql:

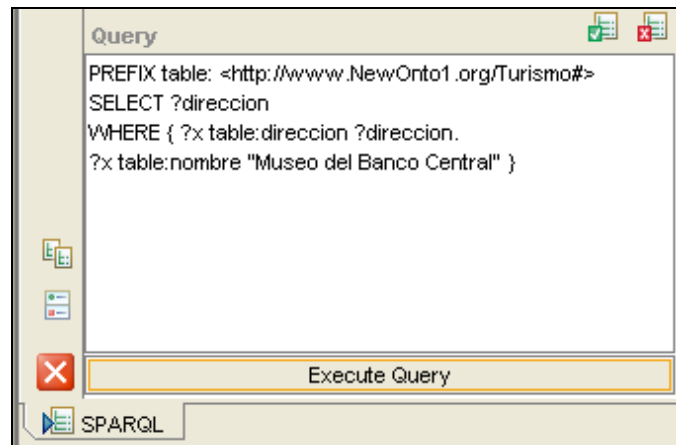


Figura 5.4. Consulta Sparql (2) sobre la Ontología de Turismo

Podemos apreciar que para tal fin, la propiedad de la cual se desea obtener el valor, debe ir acompañada por el nombre de la instancia en particular.

Lógicamente esta es una consulta de prueba, y se puede realizar el mismo tipo de consulta para cualquier instancia y extrayendo cualquiera de sus propiedades, para esto se utiliza el envío de parámetros en el código, como se explicará en la sección 5.3, fase de Implementación, descrita más abajo.

El resultado de la ejecución de esta consulta en Protégé, se muestra en la Figura 5.5:



Results	
	direccion
	Calle Larga y Huaynacapac

Figura 5.5. Resultado de la Consulta Sparql (2).

3. En la siguiente consulta de ejemplo (3), obtendremos la clase a la cual pertenece una determinada instancia, para esto se utiliza la sentencia en Sparql mostrada en la figura 5.6:

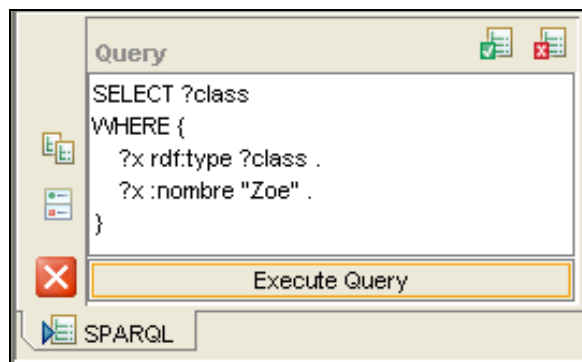


Figura 5.6. Consulta Sparql (3) sobre la Ontología de Turismo

En esta consulta se utiliza la sentencia type que esta definida como parte del lenguaje RDF y por tanto se requiere especificar donde encontrarlo.



En la estructura de la consulta, mostrada en la figura 5.6, podemos apreciar que utilizamos el prefijo *rdf* para obtener la clase a la cual pertenece la instancia particular “Zoe”, que está presente en la ontología de Turismo. Se puede decir que es como importar la librería para poder usar una función, lo mismo que ocurre en cualquier lenguaje de programación.

Mediante este tipo de consultas, Sparql nos permite conocer la clase a la cual pertenece una instancia de manera inmediata. Ya que ésta funcionalidad es ejecutada automáticamente por el motor de inferencia de la ontología. Esto sería bastante más complicado en una base de datos, en donde sería necesario extraer todas las instancias de una clase para comprobar si se encuentra dentro de esta, y, de no ser así, probar con otras clases hasta encontrar la instancia.

El resultado de la ejecución de esta consulta en Protégé, se muestra en la Figura 5.7.

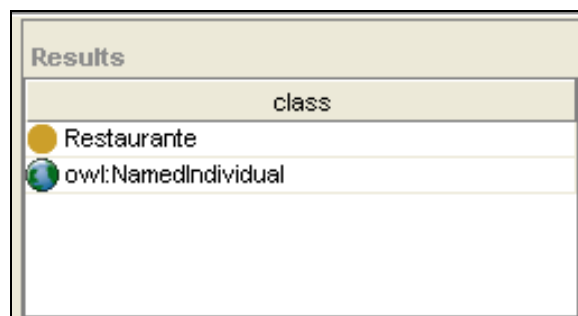


Figura 5.7. Resultado de la Consulta Sparql (3).



4. Una última consulta de ejemplo (4) nos permitirá extraer las clases equivalentes de una Clase particular en base a una de sus instancias. Como se puede apreciar en la Fig. 5.8, obtendremos la clase a la cual pertenece la instancia “Catedral de la Inmaculada”, y las clases equivalentes de la misma, en caso de que existan.

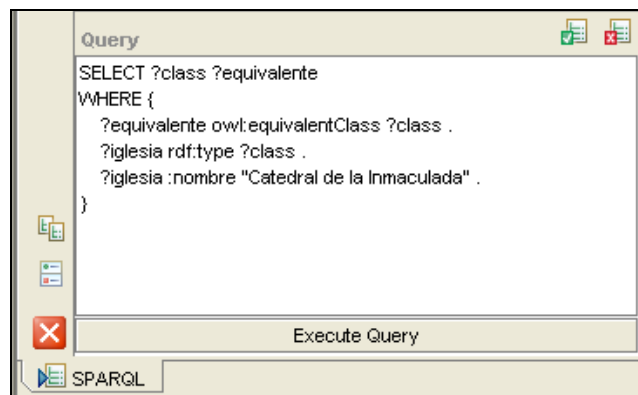


Figura 5.8. Consulta Sparql (4) sobre la Ontología de Turismo

Para realizar esta consulta, Sparql primero extrae la clase a la cual pertenece la instancia y posteriormente busca las clases equivalentes de la clase encontrada. En caso de que una instancia no pertenezca a una clase equivalente simplemente el valor de la columna “equivalente” aparecerá vacío.

En este caso Sparql trabaja con las subclases de la clase *Templo*; cuya relación se puede apreciar en la figura 5.9.

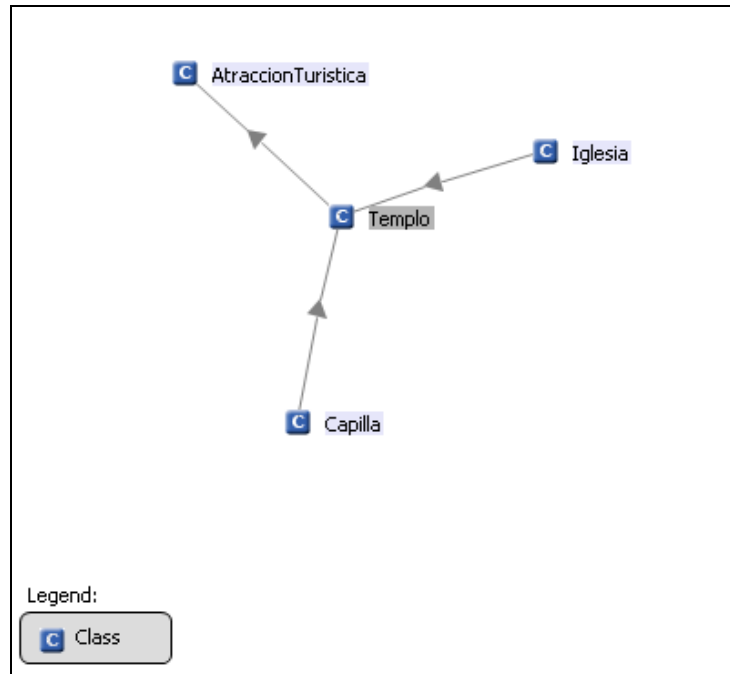


Figura 5.9. Clase Templo y Subclases

El resultado de la ejecución de esta consulta en Protégé, se muestra en la Figura 5.10.



Results	
class	equivalente
 Iglesia	 Capilla

Figura 5.10. Resultado de la Consulta Sparql (4).

Estas consultas son solo algunas de las consultas que permite ejecutar el sistema; y han sido tomadas como casos de prueba, para probar la eficiencia de la resolución de consultas SPARQL dentro de la capa de Datos.

Conexión de la Capa de Datos con la Capa de Negocios

Después de probar la resolución de consultas en nuestra Ontología, procederemos a conectarla con la Capa de Negocios. Este procedimiento está implementado en el lenguaje de programación Java, usando el entorno de programación Eclipse, versión “Helios Service Release 1”. El siguiente código, en la figura 5.11, nos permite cargar la ontología en Eclipse utilizando el Api de Jena:

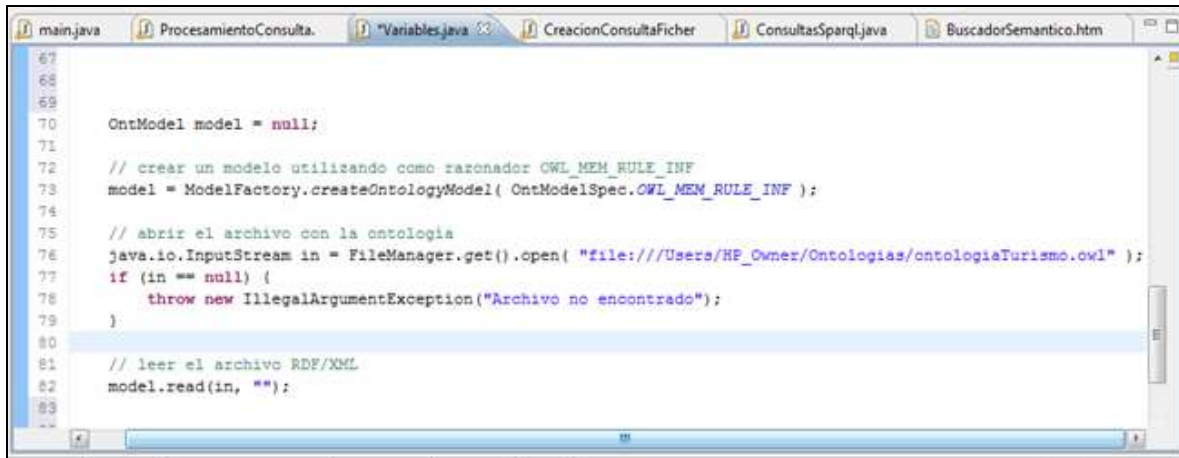


Figura 5.11. Conexión de la Ontología con la Capa de Negocios

El segmento de código mostrado en la figura 5.11, permite crear un objeto de tipo Modelo (línea 70), propio de la librería Jena para Java, y cargar nuestra ontología a la plataforma de programación (línea 76).

Jena nos permite trabajar con ontologías desde la plataforma de programación, como que se explicó en el capítulo 2.

De esta manera, queda enlazada la Capa de Datos con la Capa de Negocios, y estamos listos para empezar a trabajar con la ontología.



5.3.2. Capa de Negocios

La Capa de Negocios es la encargada de las funciones lógicas y operativas del prototipo. Proporciona la inteligencia necesaria para la comunicación con el usuario mediante lenguaje natural a través de la interfaz y para la comunicación con la Capa de Datos y la ontología.

Para el funcionamiento del buscador, la Capa de Negocios realiza las siguientes tareas:

- a) Clasificación de la Consulta
- b) Generación de Código SPARQL
- c) Generación de la Respuesta para el usuario

A continuación explicaremos las funciones que son realizadas en cada una de las tareas de la Capa de Negocios.

a) Clasificación de la Consulta

En nuestro prototipo, la interpretación del lenguaje natural se realiza mediante un proceso de identificación de términos. Este proceso permite que el programa reconozca si las palabras ingresadas por el usuario coinciden con CLASES, ATRIBUTOS o INSTANCIAS de la ontología de Turismo.



Tras el ingreso de una pregunta en Lenguaje Natural, existe una clasificación, procesamiento y ejecución de una consulta en SPARQL, para obtener una salida.

Hemos creído conveniente clasificar las consultas que se realicen sobre este prototipo en base al sentido de la pregunta. La clasificación es la siguiente:

- En primer lugar tenemos las preguntas que desean conocer los detalles referentes a un lugar determinado, es decir preguntas del tipo: ¿Cuál?

Dentro de este grupo podemos tener todas las posibles combinaciones entre propiedades e Instancias del modelo, como por ejemplo:

- ¿Cuál es la dirección del Parque de las Candelas?
- ¿Cuál es la categoría del Hotel Italia?
- ¿Cuál es el año de fundación del Museo Remigio Crespo?

- Como segunda clasificación tenemos las preguntas que desean conocer información acerca de una instancia, como el tipo de clase al que corresponde un sitio en particular, denominamos a este segundo grupo como las preguntas de tipo: ¿Qué?

Dentro esta clasificación podemos tener preguntas que involucren clases e instancias, algunos ejemplos son:

- ¿Qué es Zoe?



- ¿Qué tipo de lugar es Raymipamba?
- ¿Qué tipo de lugar turístico es Santo Domingo?
- El tercer tipo de clasificación hace referencia a aquellas preguntas abiertas, tales como:
 - ¿Santo Domingo es una Iglesia?; o
 - ¿Raymipamba es un Parque?

El conjunto completo de preguntas, está especificado en la sección 4.2¹⁷ del capítulo 4. Ahora, para cada clasificación, se sigue el siguiente proceso de identificación:

1. El primer paso es almacenar la pregunta ingresada. En nuestro caso, creemos conveniente hacer uso de un contenedor para la pregunta. De esta manera, al ingresar la pregunta se puede tener acceso a cada palabra de la pregunta en una posición del contenedor, lo cual es de ayuda para conocer a qué tipo de clasificación pertenece, esto se puede ver en la Figura 5.12.

¹⁷ Especificación de la Ontología, Metodología NeOn



Figura 5.12. Clasificación de una Pregunta

2. En segundo lugar, se analiza la primera posición del contenedor, misma que corresponde a la primera palabra la pregunta. De acuerdo al contenido de esta palabra, clasificamos la consulta en uno de los tres tipos de consultas expuestos anteriormente. Como ya se mencionó, no nos hemos basado en ningún modelo en particular para realizar esta acción; a nuestro criterio es más fácil identificar el tipo de pregunta conociendo cada palabra que la compone.

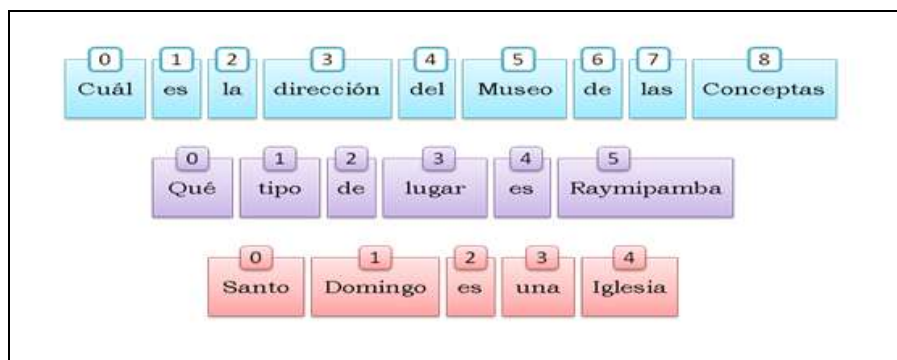


Figura 5.13. Tipos de preguntas, de acuerdo a al primera posición del contenedor de consulta



Como se aprecia en la Figura 5.13, analizar el primer elemento del Contenedor nos permite saber si es que se trata de una pregunta de tipo ¿Cuál?, ¿Qué?, o de tipo abierto.

Revisando las CQs, hemos podido identificar algunos patrones en cuanto a la estructura de las preguntas. Por citar un ejemplo podemos decir que gran parte de las preguntas de tipo ¿Cuál? por lo general van acompañadas de una propiedad en la posición 3 o 4 de la consulta como se puede apreciar en la figura 5.13.

3. Una vez que se conoce el tipo de la pregunta, el siguiente paso es identificar una propiedad y una instancia, esto para el caso de la pregunta ¿Cuál?, mientras que en el caso de las preguntas ¿Qué? se debe identificar al menos una instancia.

En cada una de las clasificaciones es importante identificar claramente los conceptos claves en la pregunta, tal y como se aprecia en la figura 5.14.





Figura 5.14. Términos Clave en una consulta

Para lograr identificar los conceptos claves dentro de la consulta, utilizamos tres contenedores de comparación que son cargados desde la ontología al momento de ejecutar el programa; un contenedor de CLASES, otro de PROPIEDADES y un tercero de INSTANCIAS.

El programa recorre el contenedor de la consulta y compara cada uno de sus elementos con los elementos de estos tres contenedores hasta encontrar una coincidencia. Dependiendo del contenedor en el que se encuentre un término de la consulta, este será clasificado como CLASE, PROPIEDAD o INSTANCIA, y almacenado en una variable que posteriormente será enviada como parámetro para generar el código Sparql de la consulta.

En caso de que los términos de la consulta no se encuentren en ninguno de estos contenedores, el programa descartará el término suponiendo que se trata de un artículo, un verbo o palabras auxiliares que no serán de utilidad para la Capa de Datos. En la PLN, esta eliminación de términos, se conoce como la normalización de un término.

Cabe mencionar que el proceso de comparación de contenedores esta sujeto a excepciones de ortografía, esto se debe al hecho de que los nombres de



Clases, Propiedades e Instancias en la ontología no están escritos necesariamente en lenguaje natural o con el acento adecuado. Un ejemplo de esto es la propiedad “dirección”; que se encuentra como “direccion”, en la ontología. Otro ejemplo sería el caso de la clase “parquePublico”, al cual el usuario podría referirse bien como “parque” o “Parque”. Estas excepciones están incluidas en el código del programa al momento de comparar las palabras en los contenedores.

b) Generación de código SPARQL

Los términos clave de la consulta (clases, propiedades e instancias), extraídos en el procedimiento anterior; son enviados como parámetros al proceso que genera el código de la consulta Sparql.

La figura 5.15 muestra un esquema conceptual de lo mencionado:

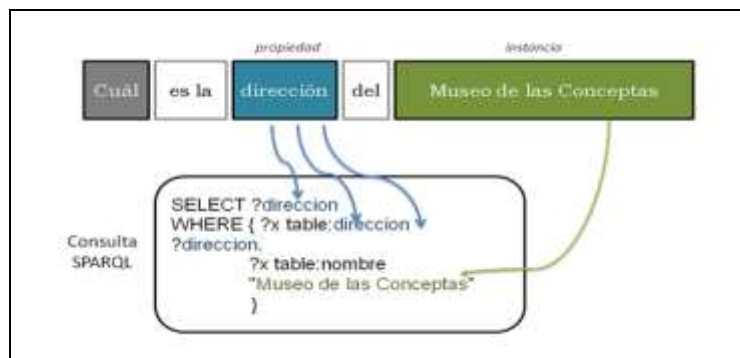


Figura 5.15. Armado de la Consulta Sparql



Sin embargo, el proceso de armado de la consulta, no depende solamente de la identificación de estas palabras clave. Quizá una de las partes más complejas de todo el proyecto es tratar con las consultas en SPARQL, ya que involucra un conjunto de prefijos para ser utilizados dentro de la consulta. Podemos ver algunos de los más utilizados en la Figura 5.16.

Prefix	Namespace
cicero	http://isweb.uni-koblenz.de/cicero#
owl	http://www.w3.org/2002/07/owl#
owl2	http://www.w3.org/2006/12/owl2#
protege	http://protege.stanford.edu/plugins/owl/protege#
rdf	http://www.w3.org/1999/02/22-rdf-syntax-ns#
rdfs	http://www.w3.org/2000/01/rdf-schema#

Figura 5.16. Prefijos para consultas SPARQL

El prefijo define el estándar que se va a utilizar dentro de la consulta, y al declararlo previamente en la consulta nos permite utilizar todas las funcionalidades de dicho estándar.

Un ejemplo claro de esto se encuentra en la figura 5.17, y se da al momento de generar la consulta SPARQL que devuelve las propiedades sobre los datos:

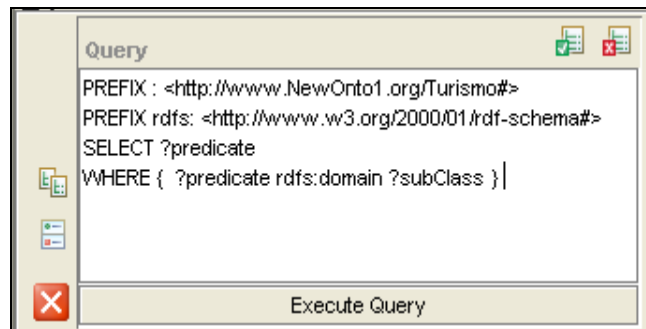


Figura 5.17. Uso de los prefijos en una Consulta.

El uso del prefijo `rdfs` nos permite utilizar la palabra clave “domain”, que pertenece a la sintaxis del mencionado estándar.

Anteriormente se detalló el proceso completo de generación y ejecución de una consulta SPARQL para una pregunta del tipo ¿Cuál?, podemos aseverar que, en base a nuestra experiencia, el proceso es idéntico para las preguntas pertenecientes a otro tipo, por ejemplo del tipo ¿Qué?

Para aclarar esto, podemos escoger una pregunta abierta perteneciente a la última clasificación y seguir su proceso de clasificación y generación de la consulta.

Como se indicó anteriormente, el prototipo está en la facultad de responder preguntas sobre clases Equivalentes (Figura 5.18) agregando un nivel de inferencia al modelo. A continuación trataremos una pregunta de este tipo, siguiendo el proceso para responder a la pregunta: ¿Santo Domingo es una



Capilla?, en este caso no se trata de una pregunta de tipo ¿Qué? ó ¿Cuál?, si no de una pregunta de naturaleza de la instancia, en donde será necesario identificar primero si “Santo Domingo” es una instancia en la ontología y después identificar la clase a la cual pertenece. A continuación se explica como se procesan este tipo de preguntas.

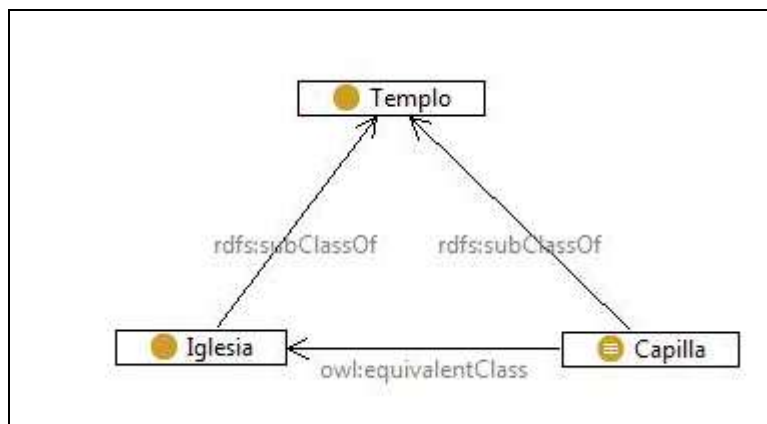


Figura 5.18. Ejemplo de Clase Equivalente

Dentro del modelo, la clase Templo define dos subclases: Iglesia y Capilla, en la cual Capilla es una clase Equivalente de Iglesia, como se aprecia en la figura 5.18; de tal manera responder a la pregunta ¿Santo Domingo es una Iglesia?, no traería mayores complicaciones.



Para responder a esta pregunta, es necesario identificar en primer lugar la clase a la cual pertenece la instancia, (Figura 5.19) y luego obtener las clases Equivalentes de esta clase.

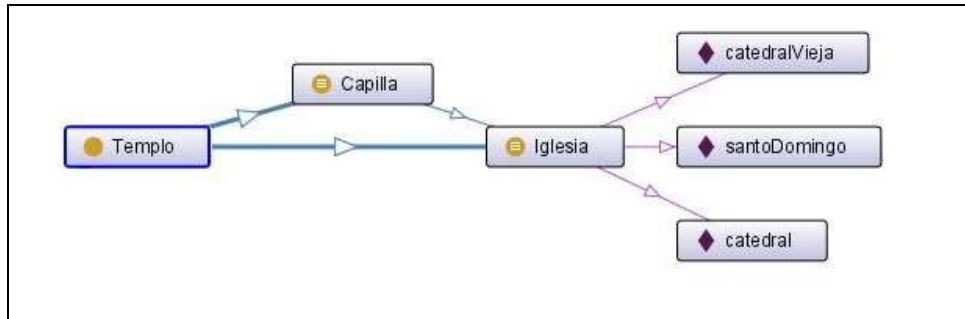


Figura 5.19. Instancias de Clases Equivalentes

Las Clases Equivalentes son una mejora que incorpora el estándar owl, por tal motivo para hacer uso del mismo, debemos referenciar el prefijo de owl, tal como se indica en la figura 5.20.

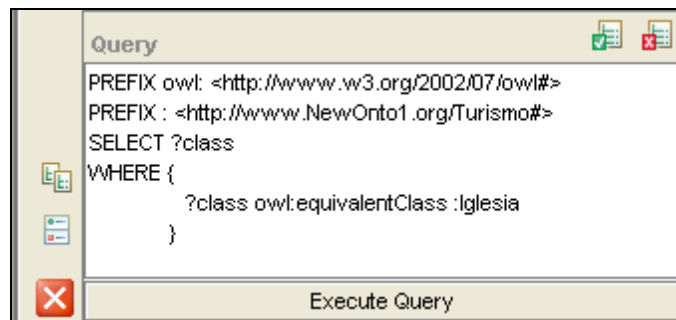


Figura 5.20. Consulta SPARQL de Clases Equivalentes



De esta manera, podemos obtener solución a preguntas que involucren clases Equivalentes.

Después de determinar la naturaleza de las palabras que conforman la consulta, la Capa de Negocios genera la consulta en Sparql utilizando del código mostrado en la figura 5.21:

```

40      try {
41          BufferedWriter bw =
42              new BufferedWriter(new FileWriter(sFichero));
43
44              bw.write("PREFIX table: <http://www.NewOnto1.org/Turismo#> ");
45              bw.newLine();
46              bw.write("PREFIX rdf: <http://www.w3.org/2000/01/rdf-schema#> ");
47              bw.newLine();
48              bw.write("PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> ");
49              bw.newLine();
50              //bw.write("SELECT ?especialidad ");
51              bw.write("SELECT ");bw.write("?");bw.write(auxiliar);
52              bw.newLine();
53              //bw.write("WHERE { ?x table:especialidad ?especialidad.");
54              bw.write("WHERE { ?x table:");bw.write(auxiliar+" ");bw.write("?");bw.write(auxiliar);bw.write(".");
55              bw.newLine();
56              bw.write("?x table:nowhere ");
57              bw.newLine();
58              bw.write("");bw.write(variable);bw.write("");
59              //bw.write("");bw.write(preguntaC[6]);bw.write(preguntaC[7]);bw.write("");
60              bw.newLine();
61              bw.write(" ");
62
63              // Hay que cerrar el fichero
64              bw.close();

```

Figura 5.21. Generación del código SPARQL

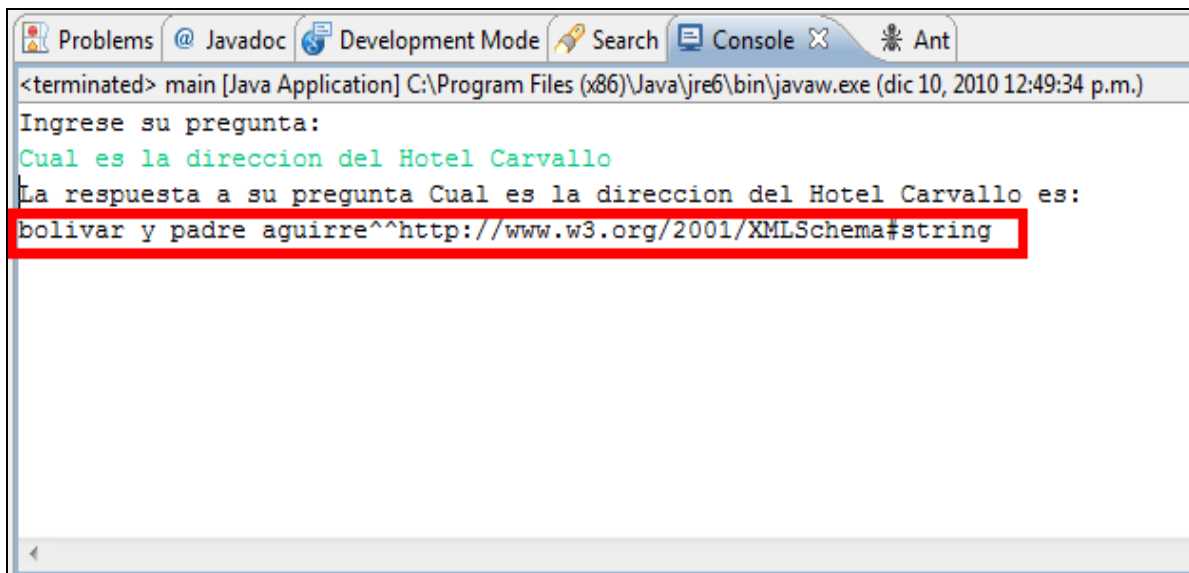
Este código genera automáticamente la consulta con los términos ingresados por el usuario y la almacena en un archivo de texto. Posteriormente la capa de



Negocios pasa el archivo que contiene la consulta hacia la capa de Datos para su ejecución.

c) Generación de la respuesta para el usuario

La Capa de Datos resuelve la consulta y devuelve el resultado a la Capa de Negocios, con el formato mostrado en la figura 5.22; es aquí en donde se debe completar el proceso de lenguaje natural, creando una respuesta entendible para el usuario.



```
<terminated> main [Java Application] C:\Program Files (x86)\Java\jre6\bin\javaw.exe (dic 10, 2010 12:49:34 p.m.)
Ingrese su pregunta:
Cual es la direccion del Hotel Carvallo
La respuesta a su pregunta Cual es la direccion del Hotel Carvallo es:
bolivar y padre aguirre^^http://www.w3.org/2001/XMLSchema#string
```

Figura 5.22. Respuesta a la Consulta SPARQL

La Capa de Negocios recibe la respuesta a la consulta y genera una frase o enunciado con la misma. Para esto es necesario abstraer los términos clave de la



consulta, como son los nombres de los atributos de las instancias que constan en la pregunta del usuario, y los términos que utilizó el usuario para formular la pregunta.

Este proceso devuelve una respuesta como se muestra en el siguiente ejemplo; figura 5.23:

```
<terminated> main [Java Application] C:\Program Files (x86)\Java\jre6\bin\javaw.exe (dic 10, 2010 12:49:34 p.m.)
Ingrese su pregunta:
Cual es la direccion del Hotel Carvallo
La respuesta a su pregunta Cual es la direccion del Hotel Carvallo es:
bolivar y padre aguirre
```

Figura 5.23. Respuesta generada en la Capa de Negocios

Esta respuesta se pasa a la Interfaz, misma que se encarga de presentarla ante el usuario.



5.3.3. Interfaz

La última capa, la Interfaz, es la encargada de comunicarse directamente con el usuario. Está construida con lenguaje HTML, y utilizando tecnología GWT.

GWT o Google Web Toolkit es un framework creado por Google que permite trabajar con la tecnología AJAX. Este framework permite crear código Java utilizando cualquier entorno de desarrollo, en nuestro caso Eclipse; y traducir este código automáticamente a HTML y JavaScript para la aplicación Web.

La interfaz del Buscador Semántico BUTUCU se muestra en la figura 5.24.

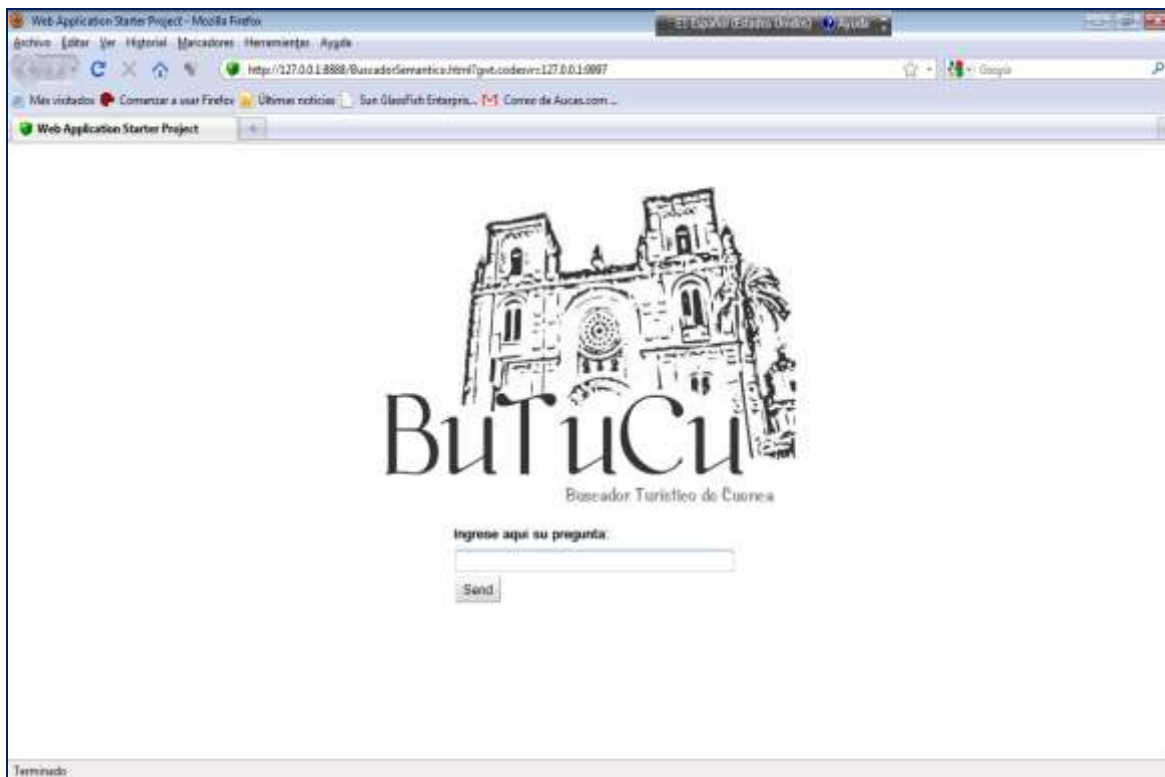




Figura 5.24. Interfaz del Buscador Semántico

Para el caso de nuestro prototipo de búsqueda semántica, la Capa de Interfaz cumple con dos funciones principales: Recibir la pregunta del usuario, y mostrar la respuesta a la consulta; ambos en lenguaje natural.

Al recibir la información de usuario, la interfaz es capaz de mostrar sugerencias sobre las preguntas de competencia para nuestra ontología, esto es posible gracias a una función de Java Script para añadir lo que se conoce como “Suggestion Text”; la imagen 5.25 muestra esta función, que es programada directamente en la interfaz para el textbox correspondiente:

```

58      //suggest box
59      MultiWordSuggestOracle oracle = new MultiWordSuggestOracle();
60      String[] words = new String[36]; //agregamos la lista de sugerencias
61      words=Variables.getOracle();
62
63      for (int i = 0; i < words.length; ++i) {
64          oracle.add(words[i]);
65      }
66      // Create the suggest box
67      final SuggestBox suggestBox = new SuggestBox(oracle);
68      suggestBox.setWidth("400px");
69      suggestBox.ensureDebugId("cwSuggestBox");
70      suggestBox.setText("");

```

Figura 5.25. Código para Suggestion Text

Así, se facilita el ingreso de la consulta por parte del usuario. Posteriormente la Interfaz envía esta consulta a la Capa de Negocios para su tratamiento y, una vez



ejecutada la consulta, recibe la respuesta desde la Capa de Negocios y la muestra directamente al usuario. Ya que la respuesta llega en Lenguaje Natural, la interfaz no necesita realizar más operaciones.



En resumen, el prototipo de buscador semántico debe combinar las tecnologías que se encuentran dentro de la Capa de Datos, de Negocios, y la interfaz para otorgar los resultados esperados. La combinación de estas tecnologías se puede entender mejor en la figura 5.26:

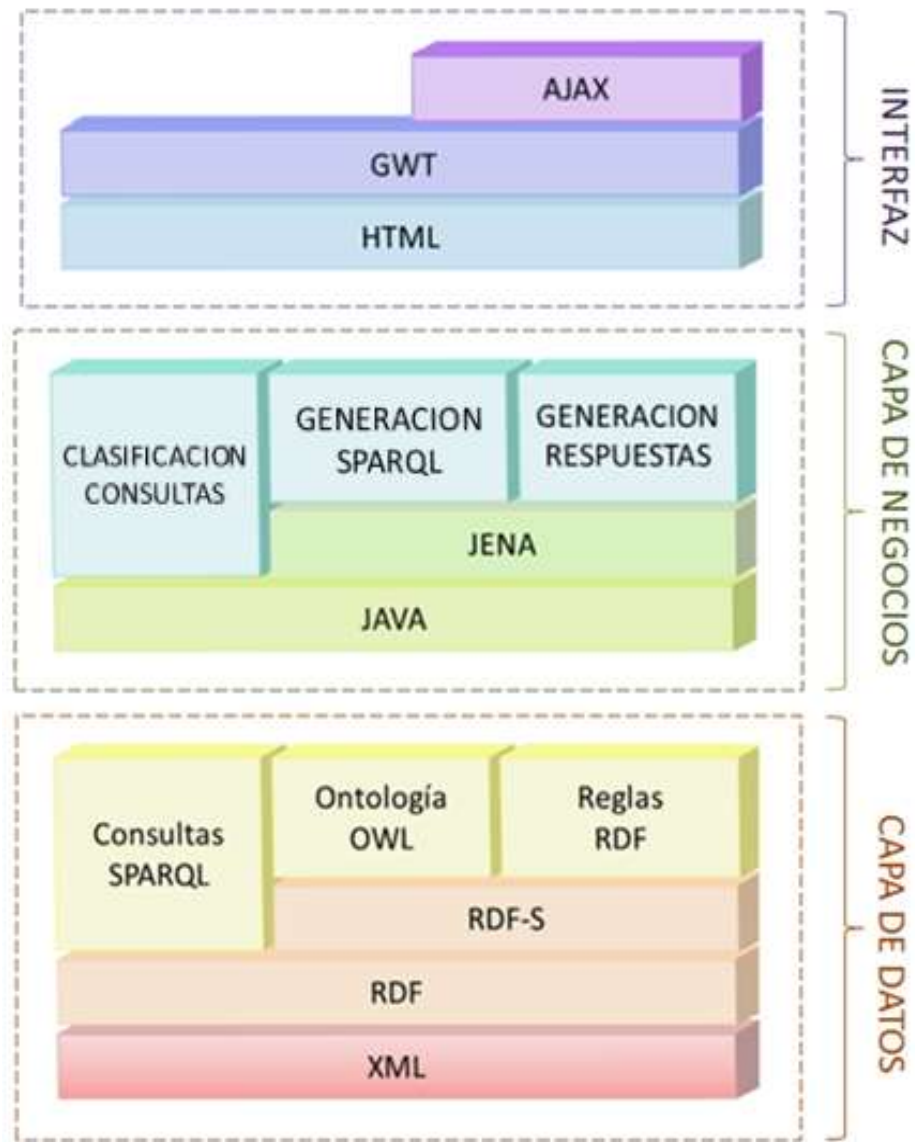


Figura 5.26. Arquitectura del Buscador Semántico

Cómo se puede ver en la figura 5.26; la Capa de Datos se sostiene de la tecnología XML, que permite trabajar con la ontología OWL utilizando RDF, RDFS, Reglas del RDF, y SPARQL para las consultas sobre la Ontología de Turismo.



Esta capa se encuentra en directa comunicación con la Capa de Negocios a través de su código JENA soportado desde JAVA. Además, la Capa de Negocios también contiene su propia lógica en cuanto a la Clasificación de Consultas, y la Generación de Respuestas, misma que le permiten, a su vez, estar en contacto directo con la Interfaz gracias a la interpretación del Lenguaje Natural

La Capa de Interfaz se basa en HTML y GWT que le permiten trabajar con AJAX y JavaScript para soportar Suggestion Text e interactuar con el usuario.



5.4. Fase de Pruebas

En esta sección se evaluará el prototipo de búsqueda Semántica, para esto probaremos su funcionamiento en cada una de las tareas que realiza:

1. La Interfaz recibe la consulta de usuario y, una vez validada, la envía a la Capa de Negocios.

La Figura 5.27 muestra el ingreso de una consulta por parte del usuario; aquí se puede observar el funcionamiento del Suggestion Text. Y, la Figura 5.28 muestra la consulta ingresada.

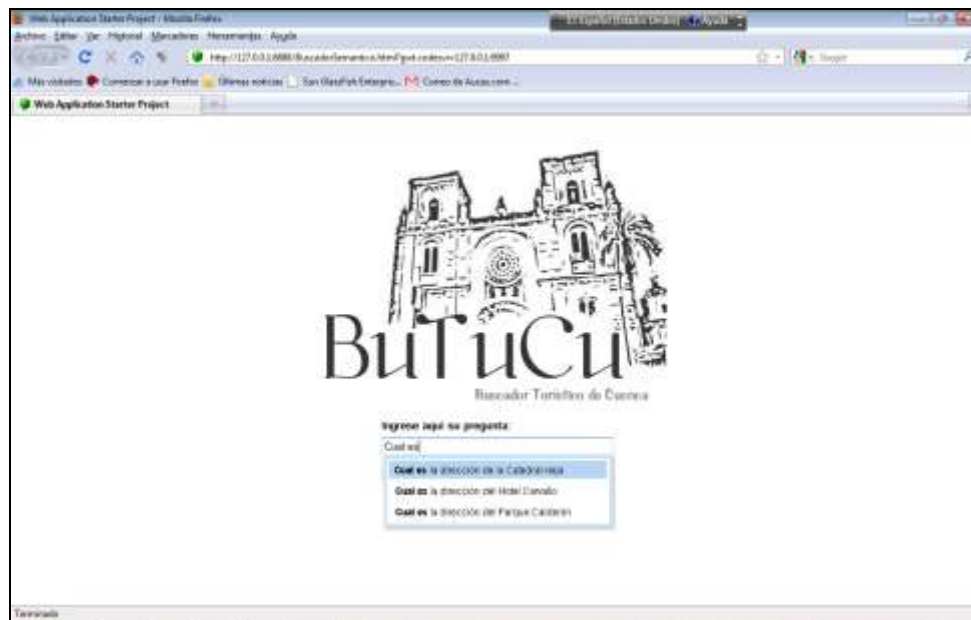




Figura 5.27. Suggestion Text



Figura 5.28. Ingreso de la Consulta en la Interfaz

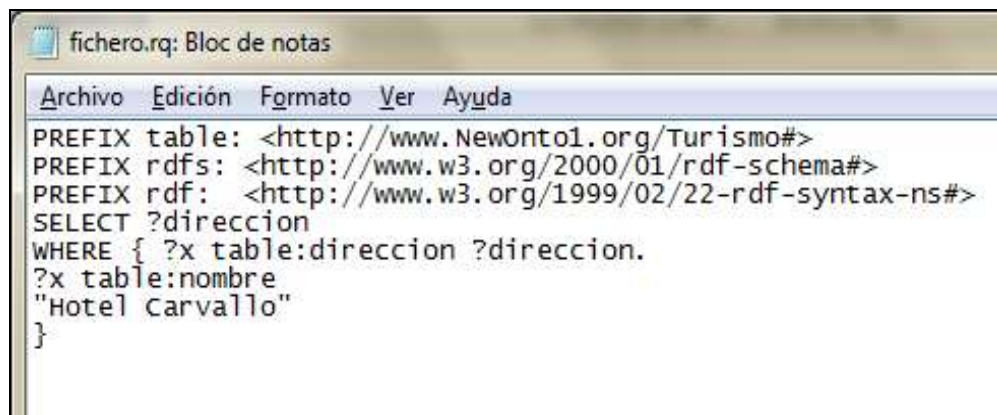
2. La Capa de Negocios recibe la consulta y la clasifica de acuerdo a sus términos.

En la capa de negocios, la Consulta es analizada y separada de acuerdo a cada uno sus términos; para probar el funcionamiento de esta tarea, pasamos a la siguiente etapa del programa: la generación de la consulta Sparql.



3. La Capa de Negocios genera el código para la consulta SPARQL y la envía a la Capa de Datos.

Una vez clasificada la consulta, la Capa de Negocios genera la consulta SPARQL como se explicó en el punto anterior. La Figura 5.29 muestra el archivo de texto que contiene la consulta SPARQL:



```
fichero.rq: Bloc de notas
Archivo Edición Formato Ver Ayuda
PREFIX table: <http://www.NewOnto1.org/Turismo#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
SELECT ?direccion
WHERE { ?x table:direccion ?direccion.
?x table:nombre
"Hotel Carvallo"
}
```

Figura 5.29. Consulta SPARQL

Cómo se puede observar en la figura, la capa de Negocios ha generado correctamente la consulta SPARQL a partir de los datos encontrados en la consulta ingresada por el usuario.

4. La Capa de Datos ejecuta la consulta y envía la respuesta a la Capa de Negocios.



La Figura 5.30 nos muestra la ejecución de la consulta desde la Capa de Datos; como se puede ver, la consulta es ejecutada con éxito y se devuelven los datos correspondientes.

```

<terminated> main [Java Application] C:\Program Files (x86)\Java\jre6\bin\javaw.exe (dic 10, 2010 12:49:34 p.m.)
Ingrese su pregunta:
Cual es la direccion del Hotel Carvallo
La respuesta a su pregunta Cual es la direccion del Hotel Carvallo es:
bolivar y padre aguirre^http://www.w3.org/2001/XMLSchema#string
    
```

Figura 5.30. Ejecución de la Consulta

5. La Capa de Negocios genera la respuesta en lenguaje natural, a partir de la respuesta enviada por la Capa de Datos, y la envía a la Interfaz.

La respuesta enviada por la Capa de Datos, es analizada y transformada en una respuesta entendible por el usuario; la Figura 5.31 nos muestra cómo este procedimiento se ejecutó con éxito en la Capa de Negocios.

```

<terminated> main [Java Application] C:\Program Files (x86)\Java\jre6\bin\javaw.exe (dic 10, 2010 12:49:34 p.m.)
Ingrese su pregunta:
Cual es la direccion del Hotel Carvallo
La direccion del Hotel Carvallo es bolivar y padre aguirre.
    
```



Figura 5.31. Respuesta generada en Lenguaje Natural

6. La Interfaz muestra la respuesta al usuario.

La respuesta en Lenguaje Natural se pasa a la Interfaz, misma que solo tiene la función de mostrarla al usuario. La Figura 5.32 muestra la realización exitosa de esta tarea:

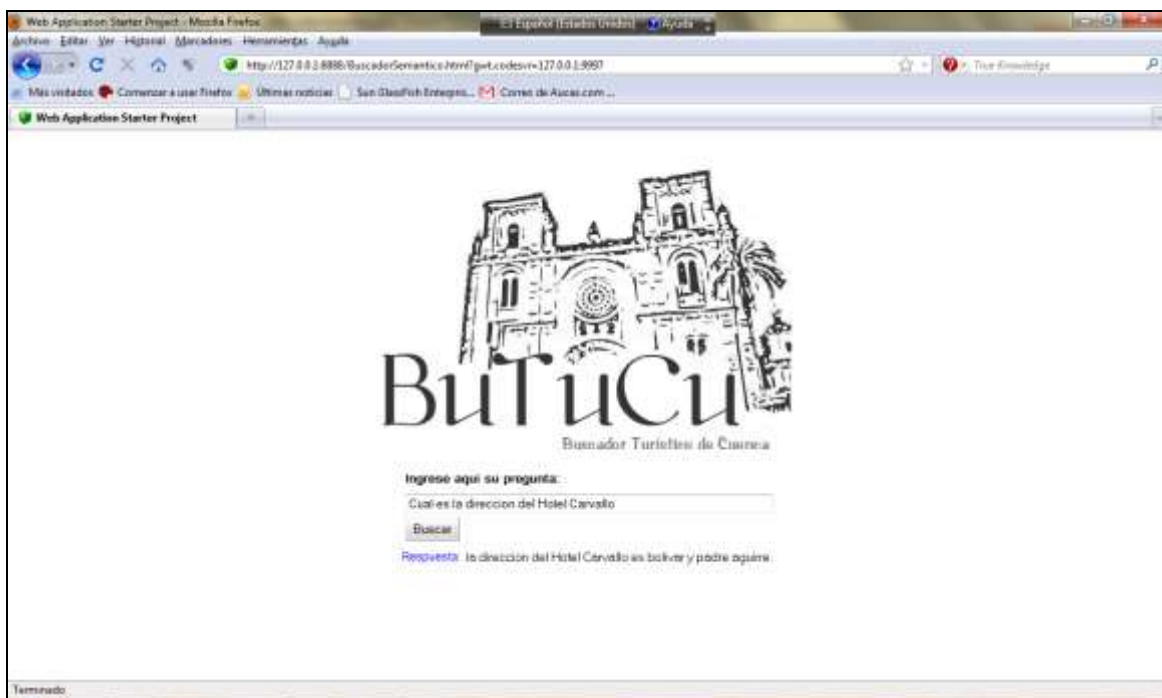


Figura 5.32. Respuesta en la Interfaz



De esta manera queda demostrado el funcionamiento del prototipo de Buscador Semántico, el mismo que esta limitado al conjunto de preguntas de competencia obtenido en la Metodología NeOn¹⁸.

CAPÍTULO VI

CONCLUSIONES

¹⁸ Capítulo 4. Metodología NeOn; Teoría y Aplicación



El Capítulo Sexto redacta las conclusiones obtenidas al final del proyecto, presentando un enfoque global de los temas analizados y las inferencias obtenidas a partir de los mismos. También se exponen algunas recomendaciones en cuanto a la experiencia adquirida en el transcurso de la realización del presente trabajo. Finalmente se lista un conjunto de trabajos futuros, sugeridos para los estudiantes, investigadores y desarrolladores interesados en la continuación de nuestro proyecto, y en



*general en la aplicación de proyectos
dentro del tema de la “Web Semántica”.*

CAPÍTULO VI

CONCLUSIONES

6.1. Conclusiones

Del trabajo de investigación del tema “Web Semántica”, se concluye que:

- Una vez finalizada la parte teórica-investigativa sobre las tecnologías involucradas en la Web Semántica y cómo interactúan entre ellas; se ha obtenido un importante marco teórico, mismo que servirá como un conocimiento básico para todas las personas que deseen conocer acerca de este tema.
- La Web Semántica propone un esquema nuevo de la Web pero altamente necesario para la organización de su información. En los últimos años el volumen de la información en la Web ha crecido de manera exponencial pero sin una organización significativa. De no ser por la organización



semántica y ontológica, esta sobredosis de información correría el riesgo de convertirse en un caos.

- La Web Semántica, y su infraestructura basada en metadatos, aporta una herramienta de razonamiento en la Web, extendiendo así sus capacidades.
- El presente trabajo de investigación contribuye con un estudio de las tecnologías que involucran la Web Semántica, de tal manera que será de utilidad para entender esta tendencia de la Web de manera general pero bastante clara.
- Aún queda un largo camino por recorrer en el campo de investigación y aplicación de las tecnologías de la Web Semántica, y esta llegará a ser una realidad en muy poco tiempo, siempre y cuando la comunidad de usuarios y desarrolladores Web colaboren con la misma. Es por esto que es importante que los investigadores, desarrolladores e ingenieros, estén al tanto de esta tecnología y sepan guiar a los usuarios a través de la misma. Así, se podrá llegar a formar una comunidad Web comprometida cuya información se encuentre integrada y sea de fácil acceso e interpretación tanto para los usuarios como para los programas y agentes inteligentes.



Del trabajo de implementación del prototipo de Buscador Semántico, se concluye que:

- La implementación del prototipo de Buscador Semántico ha hecho posible aplicar algunos de los conceptos y tecnologías de la Web Semántica, y por lo tanto entenderlos más a fondo.
- El estudio de las metodologías de desarrollo de ontologías nos permitió conocer las distintas maneras que existen para este proceso, sus ventajas y desventajas, y posteriormente a seleccionar la Metodología NeOn para la construcción de nuestra ontología de ejemplo, considerando la documentación disponible de la misma, y sus ventajas frente a las demás metodologías.
- El prototipo realiza búsquedas semánticas, a diferencia de los buscadores actuales que trabajan con búsquedas sintácticas. Esto es posible gracias a la clasificación de consultas realizada desde la interfaz y en donde se categorizan a las consultas por su contenido sintáctico y su dominio.
- El prototipo de Buscador Semántico es capaz de interactuar con el lenguaje natural utilizado en las consultas gracias al proceso de interpretación de lenguaje de la capa de Negocios.



6.2. Recomendaciones

- Una aplicación semántica puede diferir de una aplicación común, sea esta hacia la web o de escritorio, en la manera en que trabaja con la información. Por ello es necesario conocer las técnicas de extracción del conocimiento así como los lenguajes que permiten representarlo y extraerlo mediante una ontología.
- Una recomendación al momento de comenzar la construcción del proyecto es dividir el trabajo, en el caso específico de un prototipo de un buscador podemos enfocarnos por un lado en trabajar con el modelo de la ontología con la cual vamos a tratar, y, de esta manera podemos comprobar la legibilidad de la estructura del modelo.
- En cuanto al uso de Ontologías, se recomienda manejar directamente los datos desde Jena, al momento de realizar consultas. También es posible trabajar paralelamente con un modelador de Ontologías como Protégé o NeOn Toolkit, que proveen una interfaz gráfica para trabajar con los modelos de Ontologías



- En la actualidad existen muchos frameworks para diseñar interfaces de usuario, mismas que nos permiten mejorar la interacción de cualquier aplicación. Además de proveer componentes innovadores y con funcionalidades distintas, la elección del mismo dependerá de la aplicación en concreto a realizar.
- La mayor parte de los problemas que se suscitaron en el transcurso de este proyecto tuvieron que ver con la falta de fuentes de información o documentación en español. Debido a que la semántica en la web se encuentra todavía en proceso, no existen aún muchas fuentes en la web.
- En el caso de las herramientas semánticas que usamos en este proyecto podemos decir que, la principal dificultad con la que nos encontramos fue en el modelado de consultas, ya que no siempre se obtienen los mismos resultados o del mismo tamaño y forma desde un editor de Ontologías como Protégé, que incorpora una extensión de un razonador de Sparql, que ejecutando la misma consulta desde un API semántico como lo es Jena, por lo general desde este último se debe dar un tratamiento a las respuestas para su posterior presentación.



6.3. Trabajos Futuros

- Se recomienda estudiar más a fondo cada una de las tecnologías presentadas en este trabajo para su mejor entendimiento y aplicación.
- La ontología de Turismo realizada en este trabajo de investigación puede ser ampliada ya sea con ontologías existentes o agregando conceptos de manera manual. Es evidente que los temas que abarca son limitados por tratarse de una ontología de demostración para el prototipo, pero una ontología de turismo podría contener mucha más información del dominio como por ejemplo reservas en hoteles y vuelos; información geo-referencial en cuanto a la ubicación de los lugares turísticos, y por supuesto más instancias, entre otras ampliaciones recomendadas.
- Una manera de ampliar la información en la ontología es a través de Anotaciones Semánticas, para esto sería necesario una investigación previa y la búsqueda de herramientas que permitan la anotación automática de metadatos en la Web.



- Con el objetivo de ampliar la funcionalidad del Buscador Semántico, se recomienda agregar un proceso de inclusión automática de ontologías al motor del buscador, esto con la finalidad de ampliar el rango de respuestas del mismo y también para que no sea necesario incluir las ontologías de manera manual como se hizo en el prototipo.
- Un posible trabajo futuro, en cuanto a Buscadores Semánticos, es la implementación de técnicas de Procesamiento de Lenguaje Natural, mismas que no se incluyeron en el presente documento por tratarse de un tema tan amplio que podría ser implementado en una sola tesis por si mismo.
- Por último, es recomendable publicar el buscador semántico en la Web para que pueda ser utilizado y probado por cualquier usuario de la Web. De esta manera se contribuirá al entendimiento de la Web Semántica y sus tecnologías de apoyo.

ANEXOS



ANEXO I

GLOSARIO DE TÉRMINOS

El Anexo I; Glosario de Términos, contiene una recopilación de términos o expresiones técnicas utilizadas en el presente trabajo de grado; su objetivo es clarificar el significado de las mismas y evitar la ambigüedad en su interpretación dentro del contexto actual.

Índices:

Son buscadores organizados por categorías o temas, por ejemplo: Terra, Yahoo, Todoenlaces.

Motores de Búsqueda:



Son sistemas de búsqueda por palabras clave. Como operan en forma automática, los motores de búsqueda contienen generalmente más información que los directorios, por ejemplo: Google, Altavista. Recopilan las páginas web en un robot que constantemente recorre Internet buscando nuevo material.

Web Semántica:

Es una Web extendida, dotada de mayor significado en la que cualquier usuario en Internet podrá encontrar respuestas a sus preguntas de forma más rápida y sencilla gracias a una información mejor definida.

Lenguaje de recuperación:

Son lenguajes informáticos utilizados para recuperar información de sus almacenes. Puede definirse un lenguaje de recuperación como un conjunto de órdenes, operadores y estructuras que, organizados según unas normas lógicas, permiten la consulta de fuentes y recursos de información electrónica.

El World Wide Web Consortium:



Abreviado W3C, es un consorcio internacional que produce recomendaciones para la World Wide Web. Está dirigida por Tim Berners-Lee, el creador original de URL (Uniform Resource Locator, Localizador Uniforme de Recursos), HTTP (HyperText Transfer Protocol, Protocolo de Transferencia de HiperTexto) y HTML (Lenguaje de Marcado de HiperTexto) que son las principales tecnologías sobre las que se basa la Web.

Metadatos:

Representan la informaciones sobre los recursos. Los metadatos significan datos sobre datos. Estos permiten describir, identificar y localizar contenidos en los documentos de la Web.

Los metadatos deben contener información estructurada para que los ordenadores los entiendan y deben contener patrones iguales. Es decir, los metadatos son datos altamente estructurados que describen información, el contenido, la calidad, la condición y otras características de los datos.

Proxies:

Un proxy es un programa o dispositivo que realiza una acción en representación de otro. Un servidor proxy sirve para permitir el acceso a Internet a todos los



equipos de una organización cuando sólo se puede disponer de un único equipo conectado, es decir disponiendo de una única dirección IP.

Ontología:

Una ontología es una descripción formal de los conceptos y las relaciones entre conceptos, son teorías que especifican un vocabulario relativo a un cierto dominio.

Web centrada en multimedia:

Se refiere a una web que ofrezca, no solamente búsquedas basadas en metadatos, sino por similitudes en la multimedia.

Web permanente:

También llamada Web omnipresente o 'pervasive web', es una web que está en todas partes, no sólo en el PC o celular sino en la ropa, joyas, automóviles, etc. Es un concepto que se ha extendido a partir del pasaje de los PC a los teléfonos celulares y PDA. El laboratorio de multimedios del MIT, ya está trabajando en estas ideas a partir de la creación de espejos de baños y ventanas de casas conectados a internet.



Araña:

Programa informático diseñado para recorrer la Web siguiendo enlaces entre páginas, en búsqueda de nuevas páginas. Esta es la forma habitual empleada por los principales buscadores para encontrar las páginas que posteriormente forman parte de sus bases de datos. También se les conoce como Robots o Agentes Robots.

Flickr

Red social basada en administrar y compartir fotografías. Permite publicar fotografías y videos y compartirlas con el resto de usuarios en la Red Social.

Parser

Programa de computación de análisis sintáctico que permite analizar el texto conformado por una secuencia de términos, y determinar su estructura gramatical.

Folcsonomía



Hace referencia al creciente sistema de clasificación colectiva que permite aproximarse a una organización, basándose en la colaboración de las personas que clasifican la información por medio de etiquetas o tags.

URL

La URL es el 'nombre' de la página Web, y es la forma en que un usuario podrá identificarla y acceder a la misma. Cada página Web tiene una dirección propia, o una URL (Localizador Universal de Recursos) única.

Como ejemplo de URL tenemos: 'http://www.altavista.com/index.htm', que especifica claramente el protocolo de Internet que utiliza la Web (http), el nombre del servidor Web solicitado (www), el nombre del servidor específico que contiene la página Web solicitada (altavista), y en este caso, que dicho servidor está gestionado por una organización sin fines de lucro (com). Además indica el documento que se está solicitando a dicho servidor Web (index.htm).

Framework



Un framework, en el desarrollo de software, es una estructura de soporte definida en la cual otro proyecto de software puede ser organizado y desarrollado. Típicamente, puede incluir soporte de programas, bibliotecas y un lenguaje interpretado entre otros software para ayudar a desarrollar y unir los diferentes componentes de un proyecto.¹⁹

API

Una interfaz de programación de aplicaciones o API (del inglés Application Programming Interface) es el conjunto de funciones y procedimientos que ofrece cierta biblioteca para ser utilizado por otro software como una interfaz de comunicación entre componentes de software. Se trata del conjunto de llamadas a ciertas bibliotecas que ofrecen acceso a ciertos servicios desde los procesos.

AJAX

Ajax; o JavaScript asíncrono y XML, es una tecnología de desarrollo web que permite crear aplicaciones interactivas que se ejecutan en el cliente, y de esta manera realizar cambios sobre las páginas sin necesidad de recargarlas.

¹⁹ <http://www.codebox.es/glosario>



BOTTOM UP

Estrategia de procesamiento de información que permite diseñar componentes generales a partir de sus partes individuales.

FUENTES DE



INFORMACIÓN



FUENTES DE INFORMACIÓN

CAPÍTULO 1

- <http://www.w3c.es/divulgacion/guiasbreves/websemantica>
- www.bvs.sld.cu/revistas/aci/vol13_6_05/aci030605.htm
- <http://www.slideshare.net/lanegris/web-semantica-2do-comunicacion-a-presentation>
- http://es.wikipedia.org/wiki/Web_semántica
- <http://www.slideshare.net/Dmignz/web-semantica-2972370>

CAPÍTULO 2

REFERENCIAS BIBLIOGRÁFICAS

- [1] Navegador Web de Texto ELISA, <http://es.wikipedia.org/wiki/ELISA>
- [2] Expansión del Internet, <http://coevolucion.net/index.php/component/content/article/131-cerebro-global-2>
- [3] Jeffrey Zeldman, http://en.wikipedia.org/wiki/Jeffrey_Zeldman



- [4] Funcionamiento de Buscadores Actuales, <http://idem.wordpress.com/2007/06/14/%C2%BFcomo-funcionan-los-buscadores-de-internet/>
- [5] Thomas Gruber, concepto de ontología, <http://www.eumed.net/rev/cccss/02/rcb.htm>.
- [6] Informática Biomédica, <http://www.gib.fi.upm.es/es/vocabulario>.
- [7] Dotación Lógica, http://www.worldlingo.com/ma/enwiki/es/Software_Engineering_Body_of_Knowledge
- [8] HONRUBIA LOPEZ, Francisco. "Introducción a las Ontologías". Escuela Politécnica de Albacete, <http://www.dsi.uclm.es/asignaturas/42551/trabajosAnteriores/Trabajo-Ontologias.pdf>
- [9] Ontología CyC, <http://www.cyc.com/>
- [10] Ontología WORDNET, <http://wordnet.princeton.edu/>
- [11] Ontolingua, <http://ksl.stanford.edu/software/ontolingua/>
- [12] XML, www.w3.org/XML/
- [13] XMLS, <http://www.clikei.net/XMLS>



- [14] DAML + OIL, www.w3.org/TR/daml+oil-reference
- [15] OWL, www.w3.org/TR/owl-features/
- [16] Natural Finder, <http://www.bitext.com/es/naturalfinder.html>
- [17] Watson, <http://kmi-web05.open.ac.uk/WatsonWUI/>
- [18] Swoogle, <http://swoogle.umbc.edu/>
- [19] Naveganza, <http://www.isoco.com/naveganza.htm>
- [20] Hakia, <http://www.hakia.com/>
- [21] Swotti, <http://www.swotti.com/>
- [22] Retriver, <http://labs.systemone.at/retrievr/about>
- [23] True Knowledge, <http://www.trueknowledge.com/>
- [24] Bing, <http://www.bing.com/>

FUENTES DE INFORMACIÓN

- VARIOS AUTORRES. "INTRODUCCION AL INTERNET". Seminario Misiones. <http://www.un.org/spanish/Depts/dpi/seminariomisiones/intro-internet.pdf>



- CONTRERAS, Jesús. "TUTORIAL ONTOLOGÍAS". ISOCO.
http://deki.uca.es/@api/deki/files/327/=gt_normalizacion_tutorial_ontologias.pdf
- GOMEZ LOPEZ Alberto. "DESCRIPCIÓN DE LA ONTOLOGÍA".
"<http://axel.deri.ie/teaching/ri2007/alberto.pdf>
- GOMEZ-PEREZ Asunción. "La Web Semántica". Universidad
Politécnica de Madrid, 2006. <http://www.ibercajalav.net/img/websemantica.pdf>
- VAN REES, Reinout. "CLARITY IN THE USAGE OF THE TERMS
ONTOLOGY, TAXONOMY AND CLASSIFICATION". Delft University
of Technology. [http://itc.scix.net/data/works/att/w78-2003-432.
content.pdf](http://itc.scix.net/data/works/att/w78-2003-432.content.pdf)
- BORGO Stefano, LESMO Leonardo. "The Attractiveness of
Foundational Ontologies in Industry". Laboratory for Applied
Ontology, Università di Torino. "[http://www.loa-cnr.it/Files/Fomi08
BorgoLesmo.pdf](http://www.loa-cnr.it/Files/Fomi08BorgoLesmo.pdf)

CAPÍTULO 3

REFERENCIAS BIBLIOGRÁFICAS



- [25] Guía para el desarrollo de la Ontología de la Universidad de Standford, http://protege.stanford.edu/publications/ontology_development/ontology101-es.pdf
- [26] Methontology, <http://semanticweb.org/wiki/METHONTOLOGY>
- [27] Metodología CyC, <http://semanticweb.org/wiki/Cyc>
- [28] Metodología Uschold y King, http://iwayan.info/Research/Ontology/Tutor_Workshop/Tutorial_2_Descriptions.pdf
- [29] Metodología OntoKnowledge, <http://www.ontotext.com/otk/index.html>
- [30] NeonMethodology, http://www.neon-project.org/web-ontent/index.php?option=com_content&task=view&id=24&Itemid=43
- [31] Cuadro comparativo de las metodologías para el desarrollo de ontologías, <http://www.utpl.edu.ec/semanticlab/http://www.utpl.edu.ec/semanticlab/mvagila/metodologias-para-el-desarrollo-de-ontologias/98>
- [32] SPARQL, <http://www.w3.org/TR/rdf-sparql-query/>



- [33] Integración de recursos semánticos basados en WordNet, <http://www.sepln.org/ojs/ojs-2.2/index.php/pln/article/viewArticle/780>
- [34] OWL Api, <http://owlapi.sourceforge.net/>
- [35] Jena, <http://jena.sourceforge.net/>
- [36] Documentación oficial de Jena, http://jena.sourceforge.net/tutorial/RDF_API/

FUENTES DE INFORMACIÓN

CHAVEZ María, CÁRDENAS Oscar, BENITO Oscar. "LA WEB SEMÁNTICA". Universidad Nacional Mayor de San Marcos.

http://sisbib.unmsm.edu.pe/bibvirtualdata/publicaciones/risi/n3_2005/a06.pdf

- VARIOS AUTORES. "EVOLUCIÓN DE LOS LENGUAJES UTILIZADOS EN LA CONSTRUCCIÓN DE LA WEB SEMÁNTICA". Universidad Tecnológica de Pereira, Colombia.

<http://redalyc.uaemex.mx/redalyc/pdf/849/84911652067.pdf>



- KRSULOVIC Ernesto, GUITIERREZ Claudio. "Infraestructura para la Incorporación de Metadatos en Sitios Web de Departamentos Universitarios Chilenos". Universidad de Chile.
<http://www.dcc.uchile.cl/~cguiterr/articulos/wis03.pdf>
- ISOCO. "La Web Semántica". Noviembre 2004.
http://www.w3c.es/gira/paradas/presentaciones/Isoco_SemanticWeb.pdf
- BYLANDER, T. & CHANDRASEKARAN, B. (1988). "Generic tasks in knowledge-based reasoning". London, Academic Press.

CAPÍTULO 4

REFERENCIAS BIBLIOGRÁFICAS

- [37] GUTIÉRREZ Ignacio, CONESA Jordi, GEVA Felipe. "Ontologías Turísticas Geográficas", Sección 2: Estado del Arte de las Ontologías Turísticas. Universidad Abierta de Cataluña, Barcelona, España.
http://openaccess.uoc.edu/webapps/o2/bitstream/10609/2284/1/igutierrezl_articulo.pdf



- [38] Ontología Qall-Me. <http://www.slideshare.net/dinel/qallme-ontology-and-semantic-web>
- [39] Ontología Harmonise. <http://www.etourism-austria.at/harmonet/images/documentation/d2.3%20final%20ontology%20report.pdf>
- [40] Plantillas de diseño de Ontologías.
http://ontologydesignpatterns.org/wiki/Main_Page

FUENTES DE INFORMACIÓN

- http://www.NeOn-project.org/nw/NeOn_Methodology_for_Ontology_Lifecycle

DESARROLLO DEL PROTOTIPO

Ejemplos de SPARQL

- <http://ontoguate.wordpress.com/2009/01/25/utilizacion-de-sparql-para-consultas-a-ontologias/>

Consultas SPARQL desde Jena



- <http://www.eslomas.com/index.php/archives/2007/05/31/algunos-experimentos-basicos-con-protege-jena/>

Buscador semántico de Luis Criado

- <http://vissem.criado.org/>

Panel de Elementos de GWT

- <http://gwt.google.com/samples/Showcase/Showcase.html#!CwCheckBox>

Showcase de GWT

- <http://gwt-ext.com/demo/>

Creación de proyectos para GWT, como crear un nuevo proyecto

- <http://code.google.com/intl/es-ES/webtoolkit/gettingstarted.html#New>
- <http://smi-protege.stanford.edu/repos/protege/owl/trunk/examples/>

Ejemplos de Ontologías de Turismo

- <http://protege.cim3.net/file/pub/ontologies/travel/travel.owl>

Tutorial de un ejemplo básico en el GWT

- <http://sites.google.com/site/julioverne/paulojara Fuentes2>



Consultas desde Protege

- <http://protege.stanford.edu/doc/sparql/>

Showcase de SmartGWT

- <http://www.smartclient.com/smartgwt/showcase/#main>

Sintaxis de consultas SPARQL en Protégé

- http://www.co-ode.org/resources/reference/manchester_syntax/

Ejecutando consultas SPARQL con Jena Api

- <http://yuhanaresarch.wordpress.com/2008/04/20/executing-sparql-queries-with-the-jena-api/>

Soluciones individuales a las consultas

- http://jena.sourceforge.net/ARQ/app_api.html