



INSTITUTO POLITECNICO NACIONAL

Centro de Investigación en Computación

Servicio Web para la Recuperación Semántica de Información Geográfica

TESIS

Que para obtener el grado de
Maestría en Ciencias de la Computación

PRESENTA

Eliezer Alcázar Silva

ASESORES:

Dr. Miguel Jesús Torres Ruíz

Dr. Marco Antonio Moreno Ibarra



Noviembre 2013



INSTITUTO POLITÉCNICO NACIONAL SECRETARÍA DE INVESTIGACIÓN Y POSGRADO

ACTA DE REVISIÓN DE TESIS

En la Ciudad de México, D.F. siendo las 12:00 horas del día 13 del mes de noviembre de 2013 se reunieron los miembros de la Comisión Revisora de la Tesis, designada por el Colegio de Profesores de Estudios de Posgrado e Investigación del:

Centro de Investigación en Computación

para examinar la tesis titulada:

"Servicio web para la recuperación semántica de información geográfica"

Presentada por el alumno:

ALCÁZAR

Apellido paterno

SILVA

Apellido materno

ELIEZER

Nombre(s)

Con registro:


| | | | | | | |
|---|---|---|---|---|---|---|
| B | 1 | 1 | 0 | 8 | 4 | 8 |
|---|---|---|---|---|---|---|

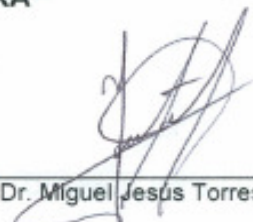
aspirante de: **MAESTRÍA EN CIENCIAS DE LA COMPUTACIÓN**

Después de intercambiar opiniones los miembros de la Comisión manifestaron **APROBAR LA TESIS**, en virtud de que satisface los requisitos señalados por las disposiciones reglamentarias vigentes.

LA COMISIÓN REVISORA


Directores de Tesis


Dr. Marco Antonio Moreno Ibarra


Dr. Miguel Jesús Torres Ruiz


Dr. Grigori Sidorov


Dr. Cornelio Yáñez Márquez


Dr. Rolando Quintero Téllez


Dr. José Giovanni Guzmán Lugo

PRESIDENTE DEL COLEGIO DE PROFESORES


Dr. Luis Alfonso Villa Vargas


INSTITUTO POLITÉCNICO NACIONAL
CENTRO DE INVESTIGACION
EN COMPUTACION
DIRECCION



INSTITUTO POLITÉCNICO NACIONAL
SECRETARÍA DE INVESTIGACIÓN Y POSGRADO

CARTA CESIÓN DE DERECHOS

En la Ciudad de México el día 22 del mes noviembre del año 2013, el (la) que suscribe Eliezer Alcázar Silva alumno (a) del Programa de Maestría en Ciencias de la Computación con número de registro B110848, adscrito a Centro de Investigación en Computación, manifiesta que es autor (a) intelectual del presente trabajo de Tesis bajo la dirección de Dr. Miguel J. Torres Ruiz y cede los derechos del trabajo intitulado Servicio web para la recuperación semántica de información geográfica al Instituto Politécnico Nacional para su difusión, con fines académicos y de investigación.

Los usuarios de la información no deben reproducir el contenido textual, gráficas o datos del trabajo sin el permiso expreso del autor y/o director del trabajo. Este puede ser obtenido escribiendo a la siguiente dirección ealkazar@gmail.com. Si el permiso se otorga, el usuario deberá dar el agradecimiento correspondiente y citar la fuente del mismo.

Eliezer Alcázar Silva

Nombre y firma

Resumen

Los grandes volúmenes de información disponibles en la web mediante documentos, artículos, libros, reportes, mapas, entre muchos otros, y la carencia de una estructura estándar entre los mismos dificultan el acceso y explotación de toda esta información. El acceso también está limitado en parte por los motores de búsqueda de información por palabras clave ya que según algunas estimaciones, en promedio, la posibilidad de que dos personas usen los mismos términos para describir un mismo objeto es entre el 7% y 18%. Se estima que entre el 13% y 15% de las consultas realizadas en los motores de búsqueda contienen el nombre de algún lugar, esto ha motivado a realizar más accesible la información geográfica en la Web, además de incentivar el crecimiento de la recuperación de información geográfica como área académica.

En esta tesis se propone una metodología para permitir el acceso a la información geográfica mediante un servicio web y facilitar la interoperabilidad entre aplicaciones. La recuperación no se limita a una búsqueda por palabras clave, por tanto, no solo trabaja con coincidencias exactas de la información buscada, sino permite explorar información relacionada con el concepto buscado. Se proponen tres etapas: a) *Conceptualización* que consiste en definir el dominio de interés a través de la descripción de los objetos, relaciones y propiedades presentes. b) *Recuperación*, se define un esquema básico para recuperar información geográfica mediante una medida de similitud semántica. Esta recuperación queda disponible por un servicio web que para ser utilizado por cualquier otra aplicación. c) *Visualización*, presenta una manera de explotar la recuperación de información disponible a través del servicio web; además de implementar un análisis espacial, al calcular rutas, distancias y duración. El sistema implementado muestra como resultado un conjunto de entidades geográficas encontradas a partir de un concepto de búsqueda, las cuales se muestran agrupadas de acuerdo con el tipo de entidad y ordenadas de acuerdo con el costo. Asimismo, se muestra mediante un mapa digital una posible ruta entre la ubicación del usuario y cada una de estas entidades, además de su distancia y duración de recorrido.

Abstract

The large volumes of information available on the web through documents, articles, books, reports, maps and more, and the lack of a standard structure between them make difficult the access and exploit this information. In addition, access is also limited in part by the keywords based search engines. Some estimates, on average, the possibility that two people use the same terms to describe the same object is between 7% and 18 %, which is a low rate to exploit the keyword search.

It's estimated that between 13% and 15% of the consultations in the search engines contain the name of a place, this has motivated to make more accessible geographic information on the Web, besides encouraging the growth of geographical information retrieval like academic area.

In this paper we propose a methodology to enable access to geographic information through a web service, without being limited to a keyword search, allowing not only to see if we have exact matches of the information sought in our information repository, but also allowing us to explore information related to the search term. We propose three stages: a) Conceptualization, which is to define the interesting domain, as well as to describe the hierarchy of objects, their relationships and characteristics, with the aim to model the domain. b) Retrieval, at this stage we propose a basic framework for geographic information retrieval using a similarity measure semantic distance between concepts, this recovery is available through a web service that can be used by any other application. Finally c) Visualization, at this stage we propose a way to exploit the information retrieval available through the web service, in addition to implementing a spatial analysis when calculating routes, distances and duration using the Google Maps API .

The implemented system shows results in a set of geographical entities found from a search term, which is grouped according to the type of entity, and ranked according to cost. Likewise, using a digital map shows a possible route from the user's location to each of these entities besides their distance and travel time.

Agradecimientos

Es tanto lo que quiero decir y tan poco lo que me puedo expresar, que en unas pocas palabras espero poder externarles mis más sinceros agradecimientos.

A ustedes Miguel, Marco, por cada corrección u observación a este trabajo, que más bien siempre fue como un consejo de amigos

A los profesores y compañeros del PIIG LAB que gracias a ustedes cada día en el CIC se volvía más grato que el anterior

A mis amigos Ime, Vlad, Luis, Magda, Isela, Víctor, Anabel, Carmen, Mario Alfonso, Rodolfo, Glendy por esos momentos de risas, juegos, estudio, comida, rondas por el pasillo, etc. momentos que quedarán grabados en mi memoria

A mi familia (abuelos, tíos y primos), por el apoyo incondicional y sus palabras de aliento

Papá, Mamá a ustedes no encuentro la manera de agradecerles tanto... El apoyo incondicional, el compartir conmigo cada acierto o desatino, y más aún el apoyo económico, pero sobre todo por ser siempre esa palabra de fe y de aliento para continuar mi camino

Isaias, Max, qué más puedo pedir a la vida con dos amigos, cómplices, y hermanos como ustedes.

Gracias a todos esos mensajes y palabras de apoyo Ene, Nancy, Mirian, Edith, Claris, Ahuactzi, Selene, Lalis.

A todos ustedes.... GRACIAS!!!!

1 Contenido

CAPÍTULO 1. INTRODUCCIÓN..... 1

1. Introducción..... 1

 1.1 Antecedentes 2

 1.2 Planteamiento del problema 3

 1.3 Objetivos 5

 1.3.1 Objetivo General..... 5

 1.3.2 Objetivos específicos..... 5

 1.4 Motivación y justificación..... 6

 1.5 Hipótesis 7

 1.6 Organización de la tesis 7

CAPÍTULO 2. ESTADO DEL ARTE 9

2. Estado del Arte..... 9

 2.1 Información 9

 2.1.1 Información web 9

 2.1.2 Información geográfica 9

 2.2 Recuperación de Información 10

 2.2.1 Búsqueda y recuperación de información 10

 2.2.2 Búsqueda y recuperación semántica de información 11

 2.2.3 Recuperación de Información geográfica 13

 2.2.4 Aspectos de la recuperación semántica de información geográfica..... 14

 2.3 Web semántica 15

 2.3.1 Web semántica espacial..... 17

 2.4 Similitud semántica 18

 2.4.1 Clasificación de medidas de similitud 20

 2.4.2 Distancia semántica 23

 2.5 Servicios WEB 28

 2.5.1 Servicios Web semánticos 30

 2.5.2 Servicios Web Geoespaciales 31

 2.6 Sistemas de información geográfica 31

| | |
|---|----|
| 2.6.1 Componentes y características de un GIS | 32 |
| 2.7 Ontologías..... | 33 |
| 2.7.1 Construcción de una ontología..... | 35 |
| 2.7.2 Tipos de ontologías | 36 |
| 2.7.3 Metodologías para construir una ontología | 37 |
| 2.7.4 Lenguajes de Ontologías..... | 39 |
| 2.8 Trabajos relacionados..... | 41 |
| 2.9 Aplicaciones relacionadas..... | 41 |
| 2.9.1 Servicios web de recuperación de información | 41 |
| 2.9.2 Recuperación semántica de información..... | 42 |
| 2.9.3 Sistemas orientados a la recuperación de información geográfica | 43 |
| CAPÍTULO 3. METODOLOGÍA..... | 45 |
| 3. Metodología | 45 |
| 3.1 Suposiciones Iniciales..... | 45 |
| 3.2 Marco de trabajo | 46 |
| 3.3 Etapa de Conceptualización..... | 46 |
| 3.3.1 Datos semi-estructurados..... | 47 |
| 3.3.2 Criterios de Búsqueda | 48 |
| 3.3.3 Atributos de la información | 48 |
| 3.3.4 Ontología de aplicación | 48 |
| 3.3.5 Poblado de la ontología..... | 59 |
| 3.3.6 Modelo persistente de datos | 61 |
| 3.4 Etapa de Recuperación..... | 64 |
| 3.4.1 Cliente | 64 |
| 3.4.2 Consulta Semántica | 65 |
| 3.4.3 Consultas SPARQL..... | 69 |
| 3.4.4 Servicio WEB | 71 |
| 3.5 Etapa de Visualización | 75 |
| 3.5.1 Entidades + Propiedades | 75 |
| 3.5.2 Análisis Espacial..... | 75 |
| 3.5.3 Ranking de resultados..... | 80 |

| | |
|--|-----|
| 3.5.4 Presentación | 81 |
| CAPÍTULO 4. RESULTADOS | 83 |
| 4. Resultados | 83 |
| 4.1 Escenarios de consulta | 83 |
| 4.1.1 Consulta una entidad | 83 |
| 4.1.2 No se selecciona ningún concepto | 86 |
| 4.1.3 Consulta un concepto | 89 |
| 4.1.4 Consulta en jerarquía..... | 91 |
| 4.1.5 Consulta de un concepto sin instancias..... | 94 |
| 4.1.6 Consulta de un concepto sin instancias en Jerarquía..... | 95 |
| CAPÍTULO 5. CONCLUSIONES Y TRABAJOS FUTUROS | 99 |
| 5. Conclusiones y trabajos futuros | 99 |
| 1.1 Conclusiones..... | 99 |
| 1.1 Aportaciones | 100 |
| 1.2 Limitaciones | 100 |
| 1.3 Trabajos futuros | 101 |
| Referencias | 103 |

Índice de Figuras

| | |
|--|----|
| Figura 2.1 Modelo simple de recuperación de información | 10 |
| Figura 2.2 Modelo de recuperación de información basado en ontologías | 12 |
| Figura 2.3 Arquitectura básica del sistema GIR GeoUJA..... | 14 |
| Figura 2.4 Recuperación de información geográfica basada en ontologías..... | 15 |
| Figura 2.5 Diferentes nociones de similitud..... | 20 |
| Figura 2.6 Relación de los componentes de un servicio web..... | 29 |
| Figura 3.1 Marco de trabajo propuesto | 46 |
| Figura 3.2 Etapa de Conceptualización..... | 47 |
| Figura 3.3 Tareas de la metodología METHONTOLOGY | 49 |
| Figura 3.4 Fragmento de la taxonomía de conceptos | 53 |
| Figura 3.5 Fragmento del diagrama de relaciones binarias | 54 |
| Figura 3.6 Vista de un fragmento de la ontología en Protégé | 59 |
| Figura 3.7 Segmento de individuos de la ontología de universidades del D.F. | 60 |
| Figura 3.8 Una instancia de la ontología | 60 |
| Figura 3.9 Esquema general de almacenamiento persistente | 62 |
| Figura 3.10 Tareas de la etapa de Recuperación | 64 |
| Figura 3.11 Resultado del ejemplo de consulta SPARQL | 71 |
| Figura 3.12 Esquema básico de un servicio Web | 72 |
| Figura 3.13 Estructura general de SOAP | 73 |
| Figura 3.14 WSDL del servicio web de recuperación semántica..... | 74 |
| Figura 3.15 Tareas de la etapa de Visualización..... | 75 |
| Figura 3.16 Esquema básico de rutas mediante el API de Google Maps | 76 |
| Figura 3.17 Indicar la posición a través de un mapa de google | 79 |
| Figura 3.18 Distancia entre dos puntos mediante el API de google maps | 80 |
| Figura 3.19 Ranking de resultados mediante el costo y tipo de entidad | 81 |
| Figura 3.20 Esquema básico de la presentación de resultados al usuario..... | 82 |
| Figura 4.1 Consulta de una entidad en específico. | 84 |
| Figura 4.2 Tabla con los atributos y conceptos de la entidad consultada. | 85 |
| Figura 4.3 Mapa para el cálculo de rutas, distancia y duración..... | 85 |

| | |
|---|----|
| Figura 4.4 Esquema de una consulta sin concepto de búsqueda. | 86 |
| Figura 4.5 Entidades que cumplen con los criterios de búsqueda. | 87 |
| Figura 4.6 Conceptos de la ontología más cercanos al concepto raíz. | 88 |
| Figura 4.7 Consulta las entidades de un concepto en específico..... | 89 |
| Figura 4.8 Resultado de la consulta de un concepto en específico..... | 90 |
| Figura 4.9 Fragmento de la ontología | 90 |
| Figura 4.10 Consulta de una subclase | 91 |
| Figura 4.11 Instancias Recuperadas de la consulta de un subconcepto..... | 92 |
| Figura 4.12 Consulta de un subconcepto que no tiene instancias | 93 |
| Figura 4.13 Fragmento de la ontología | 93 |
| Figura 4.14 Conceptos derivados de la misma clase que el concepto buscado ... | 93 |
| Figura 4.15 Instancias de los Conceptos relacionados con el concepto buscado. | 94 |
| Figura 4.16 Consulta de un concepto sin subconceptos | 94 |
| Figura 4.17 Resultado de la consulta de un concepto sin instancias | 95 |
| Figura 4.18 Ejemplo de un subconcepto sin subconceptos y sin instancias. | 95 |
| Figura 4.19 Conceptos más cercanos al concepto buscado | 96 |
| Figura 4.20 Resultado de la consulta de un concepto sin instancias | 97 |

CAPÍTULO 1. INTRODUCCIÓN

1. Introducción

El constante crecimiento de la información electrónica en la web desde su origen, ha supuesto un cambio radical en cuanto al acceso y la difusión de la información, en cambio, el gigantesco número de documentos aunado a la falta de una estructura lógica entre ellos da lugar a problemas en la recuperación de información en la web (Chávez, Cárdenas y Benito, 2005).

En la actualidad la mayoría de los buscadores están basados en la búsqueda por palabras clave dentro de documentos, sin tomar en cuenta el contexto de dichas palabras. Para Zhu, Ma, Liu, Mao y Yang (2010) En promedio, la posibilidad de que dos personas usen los mismos términos para describir un objeto es entre el 7% y el 18%, por tal motivo se buscan alternativas para representar, almacenar y organizar los datos, es necesario proveer de mecanismos y tecnologías para definir la semántica de los hiperdocumentos, meta-información que posteriormente podrá ser utilizada por agentes y buscadores inteligentes con el fin de ofrecer resultados precisos y contextualizados (Berners-Lee, 1998). Por tanto, las ontologías juegan un rol importante en aplicaciones en el campo de la web semántica.

Lo anterior es sólo hablando de la recuperación de información en general, en el caso de la información geográfica en particular se propone el uso de técnicas para describir y representar la semántica del dominio de interés, compuesto por conceptos geográficos y relaciones espaciales (Fonseca, Egenhofer, Agouris, y Câmara, 2002).

Uno de los principales desafíos de los sistemas de información es que tengan la capacidad de intercambiar información entre ellos. La principal razón reside en la creciente demanda de nuevos servicios de información, los cuales a su vez son dependientes de sistemas con distintas características. Además, resulta

importante utilizar aquella información que ya está almacenada en repositorios heterogéneos, con el objeto de desarrollar estos servicios, es aquí donde los servicios Web juegan un papel preponderante, al permitirnos una amplia disponibilidad en la Web de los recursos que prestan, para que puedan ser accedidos desde distintas aplicaciones.

1.1 Antecedentes

Existen distintos sistemas de software dedicados a la búsqueda y recuperación de información, los cuales son utilizados para recuperar e interactuar con información que es almacenada o direccionada por dichos sistemas, los cuales usualmente se encargan de procesar la información mediante consultas a documentos como artículos, libros, reportes, mapas, entre otros (González, 2009). Por tanto, podemos destacar contenidos en grandes bases de datos o mapas digitales dentro de páginas web. Hoy en día se pretende que el resultado de estas consultas sea el resultado que mejor responda a las consultas hechas por el usuario y que dichas respuestas puedan ser utilizadas más adelante (Sarabia, 2008).

Los sistemas de búsqueda de información tradicionales presentan deficiencias al momento de hacer la recuperación de dicha información, debido a que en el momento de hacer consultas, se ve afectada la precisión de la información recuperada, además que ésta carece de semántica en los datos (Fonseca, 2002). Para evitar este problema, en el presente trabajo se hace uso de la similitud semántica, la cual nos permite brindar respuestas más aproximadas a lo que el usuario espera y evitar respuestas vacías.

Las tendencias en cuanto al uso de tecnologías para el almacenamiento y representación de la información con contenido semántico, apuntan hacia el uso de bases de datos conceptuales u ontologías, las cuales han tomado gran auge en algunas áreas de investigación, debido a que permiten la formalización de la información para convertirla en conocimiento. Así mismo, disminuimos las dificultades que existen en el manejo de la información existente, la cual podemos

encontrar en una gran variedad de formatos, lo que comúnmente se denominan fuentes de datos heterogéneas.

En la búsqueda de técnicas para la recuperación de información que más se aproxime a las consultas de los usuarios, han surgido algunas técnicas que hacen uso del contenido semántico de la información para evaluar la similitud del concepto buscado con el contenido almacenado en el sistema, lo cual permite establecer una medida de similitud y así poder aproximar al máximo la precisión de los resultados que se han de presentar al usuario (Gonzalez, 2009).

En los últimos años las tecnologías para acceder a la información geográfica se han centrado en la combinación de mapas digitales y bases de datos, los cuales caracterizan la mayoría de los Sistemas de Información Geográfica (GIS - por sus siglas en inglés).

En el presente trabajo se propone una metodología para la realización de consultas semánticas que propongan como resultado una respuesta geográfica que se adecúe a las necesidades del usuario, evaluando algunas variables del entorno definidas por el mismo usuario. Estas variables serán procesadas como el perfil del usuario, ya que cada usuario tiene un conjunto de necesidades y características diferentes, por lo cual es necesario que se definan estos elementos del mismo.

1.2 Planteamiento del problema

La recuperación de información que realizan comúnmente los sistemas de búsqueda de información se centra en la búsqueda estadística de palabras clave dentro de documentos, sin tomar en cuenta el contexto de dichas palabras, es por ello que el presente trabajo de investigación se enfoca en la recuperación semántica de información, compuesta por un elemento descriptivo o atributivo y un elemento espacial. Dicho de otra manera, lo que se pretende es recuperar la información geográfica (ubicación y/o descripción) más similar a la respuesta esperada por el usuario a una consulta dada, considerando un perfil de usuario.

1. Introducción

La recuperación semántica de información geográfica que se pretende implementar corresponde a la recuperación de entidades geográficas que ayuden en la toma de decisiones a los usuarios al momento de visitar algún sitio de interés, ya que el objetivo de este trabajo es mostrar a los usuarios entidades que sean lo más parecidas posible a las entidades que describe el usuario para su búsqueda.

En este trabajo de investigación específicamente nos enfocamos en las entidades que nos pueden ayudar a solucionar el problema en el que la mayoría de jóvenes que terminan sus estudios de nivel medio superior no cuentan con información de apoyo al momento de decidir en qué institución continuar sus estudios de nivel superior. Se trata concretamente de contar con un repositorio de información sobre las instituciones de nivel superior que se encuentren registradas en El Instituto Nacional de Estadística, Geografía e Informática (INEGI) y que se encuentren en la delegación Gustavo A. Madero del Distrito Federal, como caso de estudio.

En el INEGI existe el registro de instituciones de educación clasificadas por nivel educativo, en este caso se pretende ubicar el registro de las instituciones de educación de nivel superior, de las cuales deseamos conocer las carreras que ofrecen y su localización.

Conocer la dirección de cada institución nos permitirá dar a conocer su ubicación en un mapa, así mismo conocer las carreras en las que se especializa a los alumnos nos permitirá inferir los posibles campos de acción de los mismos, lo cual puede ser de gran utilidad al momento de sugerir alguna institución al interesado.

La información relativa a la localización de cada institución de educación superior es de vital importancia para la aplicación, los datos que componen esta información están almacenados bajo cierta estructura y son conocidos como datos geoespaciales, entendiéndose así; que un dato geoespacial es aquel que brinda información acerca de la ubicación espacial sobre el plano terrestre de algún objeto en particular.

Para presentar una respuesta al usuario, se tendrá que hacer uso de datos de otras fuentes, lo cual es conocido como *mashup*¹, en el caso particular de hacer uso de datos geoespaciales como de Google Maps se conoce como *geomashup*², ésta es parte importante en la solución del problema debido a que se hará un *ranking* de las instituciones que cumplan con los requerimientos del usuario, con respecto a su ubicación y la localización del mismo, dicho *ranking* estará evaluado en función de la carrera buscada por el usuario y las carreras ofrecidas por la universidad, así como por la distancia entre la ubicación del usuario y la institución, además del costo de la colegiatura a pagar en caso de ser una institución privada.

Uno de los principales retos de esta investigación es la integración de la información dentro de una base de conocimiento que nos permita inferir lo que buscamos. En este caso en particular, la información referente al nombre, ubicación y carreras de cada institución se encuentra en formato de texto, lo cual dificulta la integración de la información. Así mismo, es necesario convertir cada una de las direcciones de las instituciones a su coordenada espacial para poderla ubicar sobre una representación geográfica.

1.3 Objetivos

1.3.1 Objetivo General

Implementar una metodología con un enfoque semántico para recuperar información geográfica que satisfaga la consulta de un usuario, proporcionando como resultado la ubicación y/o entidad geográfica correspondiente con los criterios de búsqueda especificados por un perfil de usuario, mediante la búsqueda semántica de la información y su visualización por medio de un servicio web.

1.3.2 Objetivos específicos

Para cumplir con el objetivo general de esta investigación, es necesario comenzar con una serie de objetivos específicos, los cuales se listan a continuación.

¹ Se entiende por mashup un recurso en línea que combina otros recursos y servicios.

² Es un caso particular de mashup en el que se utilizan recursos geoespaciales.

- Establecer los criterios de búsqueda semántica que intervienen en el caso de estudio.
- Identificar y estructurar la información que corresponde al caso de estudio.
- Construir la base de datos espacial que contenga los datos geoespaciales de la información almacenada en el sistema.
- Identificar las técnicas de recuperación semántica de información.
- Aplicar las técnicas de recuperación semántica de información a la información geográfica almacenada, para que los resultados de la búsqueda sean similares a los que el usuario está buscando.
- Diseñar e implementar el servicio web para la recuperación semántica y visualización de información geográfica.
- Analizar y hacer el ranking de los resultados obtenidos.

1.4 Motivación y justificación

Actualmente existe una gran cantidad de información disponible en diversas fuentes de datos que por lo general están disponibles en textos expresados en lenguaje natural. De acuerdo con Egenhofer (2002) debido al constante crecimiento de la World Wide Web los métodos tradicionales de búsqueda y utilización de la información son frecuentemente insuficientes, los buscadores convencionales utilizan herramientas estadísticas y la potencia de las computadoras para dar acceso a la información de manera eficiente. En muchas ocasiones estas herramientas no presentan resultados debido a que la búsqueda de información está basada en el uso de palabras clave; en consecuencia, los resultados de la búsqueda dependerán de las palabras elegidas por el usuario. Estas búsquedas suelen proporcionar resultados pobres, ya que muchas veces el vocabulario del usuario que realiza la consulta no coincide con el vocabulario de las personas que redactan los documentos que contienen la información.

Con la llegada de la Web semántica, según Egenhofer (2002) se esperan mejores métodos de recuperación de información, explorando la semántica de la información en el proceso de búsqueda. Por tanto, una búsqueda semántica o

conceptual se basa en interpretar la consulta del usuario para obtener los conceptos de interés y posteriormente buscarlos en documentos indexados mediante el uso de estándares.

Otro aspecto importante de acuerdo con Gone y Shade (2008) se refiere a la heterogeneidad semántica y la falta en la automatización en el tratamiento de la información, lo cual representan los principales problemas que enfrentamos al utilizar los servicios web geoespaciales.

Las investigaciones y desarrollos en el campo de la Web semántica han dado como resultado múltiples estándares para la elaboración y almacenamiento de información, lo cual facilita la aplicación de técnicas de recuperación semántica de la misma (Peis, Herrera-Viedma, Hassan, y Herrera, 2003). Así mismo, Egenhofer (2002) ha encontrado como uno de los principales desafíos el trabajar con recuperación de información geoespacial, mediante la semántica espacial y ontologías terminológicas.

Los avances científicos y tecnológicos han propiciado un gran incremento en la obtención de datos geoespaciales, los cuales para ser utilizados por los usuarios es necesario que los servidores en que se almacenan sean dotados de técnicas de recuperación de información que faciliten el acceso y las herramientas que permitan la recuperación de información que se adecúe a su consulta, lo cual es en gran medida lo que se propone en el presente trabajo de investigación.

1.5 Hipótesis

La recuperación semántica de información garantiza resultados más apegados a las consultas realizadas por los usuarios, que las consultas convencionales realizadas con palabras clave, además de permitir una mejor exploración de la información disponible y evitando la ausencia de respuestas.

1.6 Organización de la tesis

Capítulo 1: En este capítulo se presenta una introducción general al trabajo de tesis, también los antecedentes y la justificación de por qué realizar este trabajo,

también se presenta el planteamiento del problema, los objetivos que se pretenden alcanzar, así como la hipótesis que conducirá esta investigación.

Capítulo 2: Presenta el estado del arte, el cual consiste en una revisión de los trabajos relacionados con esta tesis, con la finalidad de aterrizar la presente investigación. Se hace una breve introducción al problema de la recuperación semántica de información, así como al tratamiento de información compuesta de un elemento descriptivo y un elemento espacial, así mismo abordamos el marco teórico del trabajo de investigación propuesto, en el cual se establecen los conceptos y fundamentos acerca de la recuperación de información, así como de las ontologías y cómo éstas se pueden emplear en la búsqueda y recuperación de información. También se hace una pequeña descripción de las herramientas a utilizar en la presentación de resultados.

Capítulo 3: En el capítulo tres se establece, define y describe la metodología propuesta para la solución del problema, así como la estructura del sistema propuesto.

Capítulo 4: En este capítulo se presentan y analizan los resultados obtenidos en el servicio web de recuperación de información semántica.

Capítulo 5: En este capítulo se exponen las conclusiones obtenidas durante el desarrollo de la investigación, así mismo se proponen algunos trabajos futuros.

CAPÍTULO 2. ESTADO DEL ARTE

2. Estado del Arte

Para el estudio del estado del arte de una disciplina es importante comenzar sentando las bases de los conceptos a desarrollar, por lo cual en el presente capítulo se describe la recuperación semántica de información, la recuperación de información geográfica, además de los servicios web tanto de recuperación de información como los enfocados a la recuperación semántica de la información. A su vez, se mencionan algunas aplicaciones que trabajan con recuperación semántica de información y recuperación de información geográfica.

2.1 Información

2.1.1 Información web

Para Chávez (2005) en los últimos años se ha constatado un gran crecimiento en los volúmenes de información disponibles en la web. Actualmente existe una gran cantidad de información disponible en diversas fuentes de datos que por lo general se encuentran en textos expresados en lenguaje natural, esto trae consigo problemas al momento de hacer la búsqueda de información, intentando discriminar la información de menor relevancia y recuperar la información más relevante. Esto hace necesario el desarrollo de herramientas que permitan realizar búsquedas de manera rápida y eficiente sobre tal volumen de información.

2.1.2 Información geográfica

A diferencia de la información en general, de acuerdo con Jones and Purves (2007) en los últimos años las tecnologías para acceder a la información geográfica se han enfocando en la combinación de mapas digitales y bases de datos, los cuales caracterizan la mayoría de los GIS.

La información geoespacial la podemos encontrar almacenada en una gran variedad de formatos disponibles y en un sin número de documentos como libros,

reportes, artículos, imágenes y mapas; los cuales pueden estar disponibles en bases de datos y páginas web.

2.2 Recuperación de Información

2.2.1 Búsqueda y recuperación de información

La *recuperación de información* según Manning (2008) consiste en encontrar material (generalmente documentos) de una naturaleza no estructurada (generalmente texto) que satisface una necesidad de información desde dentro de grandes colecciones (normalmente almacenada en las computadoras). Al decir que la información es no estructurada, se refiere a que los datos no son claros o no tienen una estructura semántica abierta y fácil de entender para una máquina.

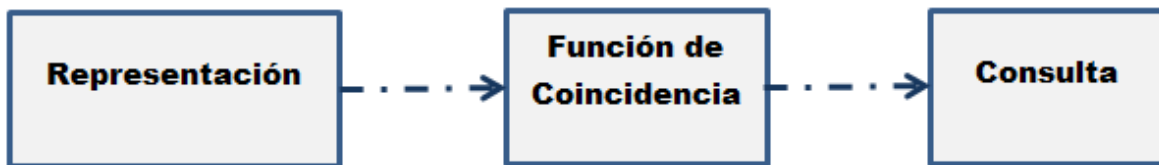


Figura 2.1 Modelo simple de recuperación de información

La recuperación de información se refiere a encontrar la información relevante basada en una consulta del usuario (Dominich, 2008). El principal desafío de la recuperación de información es que “la computadora devuelva la información que queremos, no la que pedimos, es decir, que se requiere que la computadora razone intuitivamente” Janowicz (2001, p.4). Un posible enfoque para el manejo semántico de la información es estructurar dicha información en ontologías (Wache, Voegele, Visser, Stuckenschmidt, Schuster, Neumann, y Hübner, 2001).

El campo de la recuperación de información también cubre el soporte de usuarios en la búsqueda o filtrado de colecciones de documentos o procesamiento fuerte a un conjunto de documentos.

Baeza-Yates (1999) define el problema de la recuperación de información como “dada una necesidad de información (consulta + perfil del usuario) y un conjunto de documentos, ordenar los documentos de más a menos relevantes para esa

necesidad y presentar un subconjunto de aquellos de mayor relevancia". En la solución de este problema se identifican dos grandes etapas:

1. Elección de un modelo que permita calcular la relevancia de un documento frente a una consulta.
2. Diseño de algoritmos y estructuras de datos que implementen este modelo de forma eficiente.

Baeza-Yates se preocupa especialmente de las estructuras de datos y métodos de acceso a los mismos. Curiosamente, a la hora de definir la recuperación de información, en lugar de proponer una definición propia, emplea la elaborada por Salton: "la recuperación de la información tiene que ver con la representación, almacenamiento, organización y acceso a los ítem de información".

2.2.2 Búsqueda y recuperación semántica de información

Los sistemas de recuperación de información convencionales tratan de ordenar y asimilar los documentos basados en la relevancia para el usuario. En los casos en que el número de documentos relevantes es elevado y además la mayoría de esos documentos son redundantes es necesario utilizar medios más allá de la relevancia pura para hacer el ranking de documentos (Van Kreveld, Reinbacher, Arampatzis and Van Zwol, (2005).

El uso de ontologías vino a cubrir las limitaciones de las búsquedas basadas en palabras clave, las búsquedas semánticas fueron la principal motivación para poder visualizar lo que sería la Web semántica, partiendo de esto, se propone un modelo para explotar ontologías basadas en el conocimiento para proveer las búsquedas en grandes repositorios de documentos (Vallet, Fernández, y Castells, 2005).

En el caso de algunos volúmenes de documentos del orden de Gigabytes o Terabytes se han desarrollado algunas técnicas de búsqueda semántica, las cuales consisten en miles de instancias de ontologías, clases y relaciones de complejidad arbitraria. Estas técnicas asumen que la respuesta a una consulta

2. Estado del arte

siempre tiene el 100% de precisión, pero no asumen la posibilidad de respuestas aproximadas. Según Vallet (2005) Este modelo solo tendría sentido si todo el *corpus* de documentos se pudiera representar en una ontología. En el caso de estas herramientas existen algunas limitantes como:

- No podría representar toda la información disponible en la Web en una sola ontología.
- Cada documento tiene su propio valor y no sería equivalente a sumar todos como piezas diferentes.
- No se podría establecer un ordenamiento de las respuestas dadas a una consulta, debido a que todos los resultados apuntarían al mismo documento.

En el trabajo de Vallet et al. (2005) se propone un modelo de recuperación basado en ontologías, orientado a la explotación de ontologías de dominio y bases de conocimiento, para apoyar la búsqueda semántica en repositorios de documentos, cómo se muestra en la Figura 2.2.

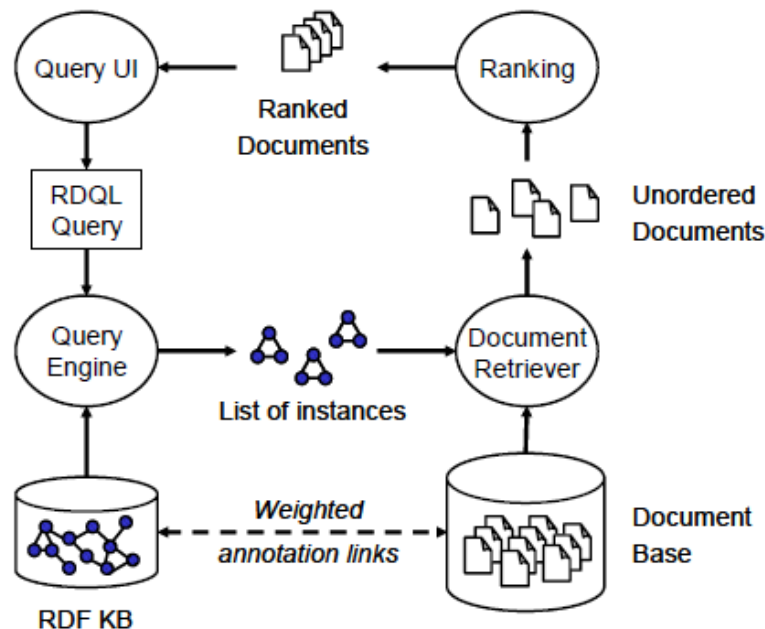


Figura 2.2 Modelo de recuperación de información basado en ontologías

2.2.3 Recuperación de Información geográfica

La Recuperación de Información Geográfica (GIR - por sus siglas en inglés) incluye todas las áreas del conocimiento, así como los elementos que componen el núcleo de la recuperación de información, como son: los sistemas de información y motores de búsqueda, pero con un énfasis especial en la información del dominio geográfico. Algunas de las técnicas que son utilizadas para la recuperación de información geográfica incluyen la extracción de entidades geográficas, análisis semántico, bases de conocimiento geográfico, adquisición de conocimiento, grafos, análisis formal de conceptos y técnicas de expansión y flexibilización de consultas (Lutz y Klien, 2006).

De acuerdo con Jones and Purves (2007) se estima que entre el 13% y 15% de las consultas realizadas en los motores de búsqueda contienen el nombre de algún lugar, esto ha motivado a hacer más accesible la información geográfica en la Web, además de incentivar el crecimiento de la recuperación de información geográfica como área académica.

Un claro ejemplo de recuperación de información geográfica es el servicio que presta el sistema GeoNames el cual es una integración de datos geográficos, tales como nombres de lugares en varios idiomas, la elevación, la población y otros de diversas fuentes. En donde los usuarios pueden editar manualmente, corregir y añadir nuevos nombres de lugares con una interfaz fácil de usar.

Según Ortega, Cumbreiras, Vega, and López (2008) existen una amplia variedad de enfoques para resolver la tarea GIR, que van desde aproximaciones simples de recuperación de información sin indexación de términos geográficos a arquitecturas que hacen uso de técnicas de procesamiento del lenguaje natural para extraer localizaciones e información topológica de los documentos y las consultas. Algunas de las técnicas usadas en la actualidad incluyen extracción de entidades geográficas, análisis semántico, bases de conocimiento geográfico (como ontologías, tesauros o gazetteers), técnicas de expansión de consultas y desambiguación geográfica.

En la Figura 2.3 se puede observar la arquitectura básica empleada en el sistema GIR GeoUJA para la recuperación de información geográfica, propuesto por el mismo grupo de desarrollo.

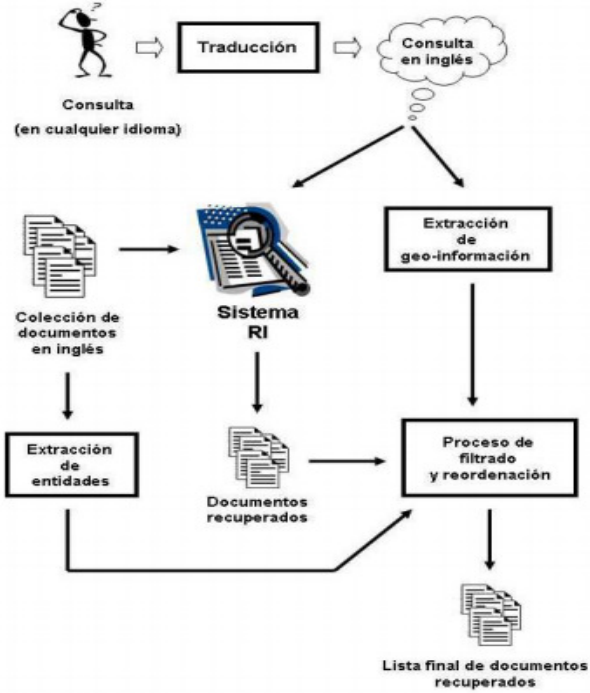


Figura 2.3 Arquitectura básica del sistema GIR GeoUJA

2.2.4 Aspectos de la recuperación semántica de información geográfica

Algunos de los aspectos a tomar en cuenta para la recuperación de la información geográfica según Jones and Purves (2007) son:

- Detectar referencias geográficas.
- Desambiguación de nombres de lugares.
- Terminología geográfica vaga.
- Ordenamiento de relevancia geográfica.
- Interfaces de usuario.
- Estudios de usuario y evaluación.

En la figura 2.7 se muestra un enfoque propuesto para la recuperación de información geográfica basada en ontologías propuesto por Jones and Purves (2007).

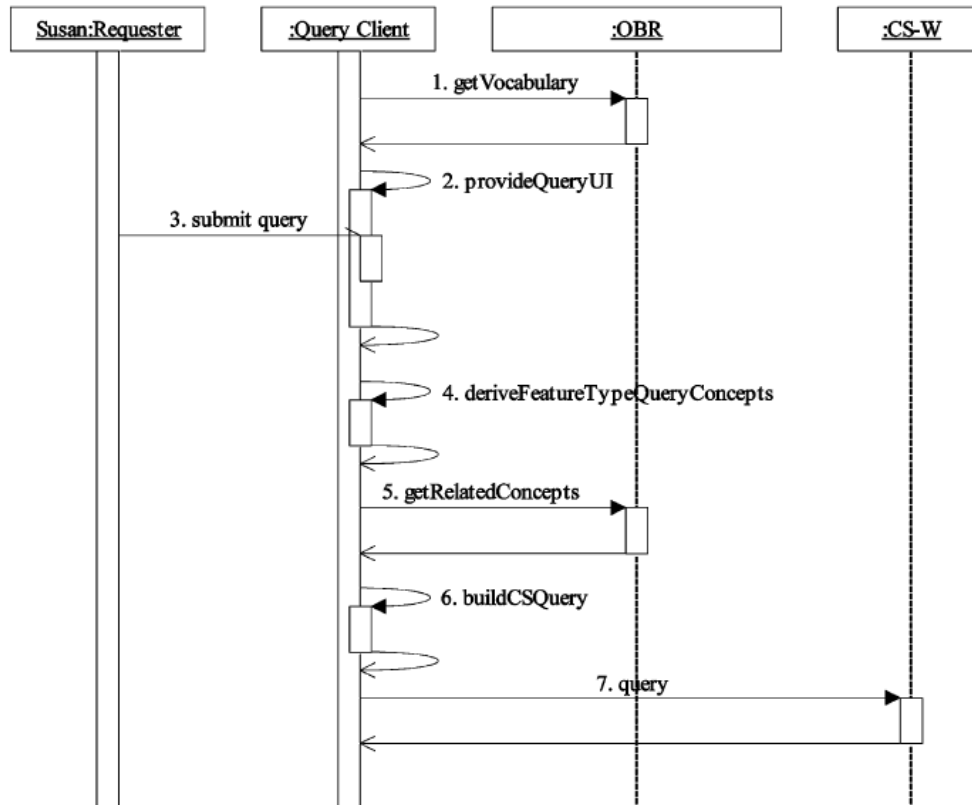


Figura 2.4 Diagrama de secuencia que ilustra los pasos del 1 al 7 en el enfoque propuesto para la recuperación de información geográfica basada en ontologías.

2.3 Web semántica

La Web Semántica “es una extensión de la Web actual en la que se proporciona la información con un significado bien definido, y se mejora la forma en la que las máquinas y las personas trabajan en cooperación” (Berners-Lee, 2001, p.25).

La World Wide Web (WWW) ha desarrollado un medio donde las personas puedan acceder a un cúmulo de documentos para ser leídos, en lugar de hacer que esa información pueda ser manipulada automáticamente. Con el aumento de páginas web dirigidas a las computadoras y añadiendo documentos únicamente para los equipos (es decir, información que pueda ser entendida por el ordenador),

podremos transformar la Web en la Web semántica, para el Consorcio Universitario para la Ciencia de la Información Geográfica (UCGIS) las ontologías juegan un rol crítico asociando el significado a los datos, de tal manera que las computadoras puedan entender el proceso de los datos automáticamente (Fonseca and Sheth, 2002).

La búsqueda semántica se podría ver como un complemento de las búsquedas basadas en palabras clave siempre y cuando las ontologías y metadatos disponibles no sean suficientes (Vallet et al. 2005). Además están de acuerdo con que la inferencia es una herramienta útil para llenar las lagunas de conocimiento y completar la información que falta.

La Web semántica comienza con la idea de tener meta-información de los documentos disponibles en la Web. La idea consiste en añadir metadatos, anotaciones semánticas y ontologías a la información existente en la Web, para lo cual se contempla que la Web semántica contenga cuatro elementos principales.

a) Expresar significado

La mayoría de los contenidos Web actuales están diseñados para ser leídos e interpretados por los usuarios, además de que la estructura de estos contenidos no permite que los programas de software los manipulen significativamente. En general, las computadoras no tienen una forma segura de procesar la semántica, por lo que esta Web propone una estructura para el contenido principal de las páginas Web, creando así un entorno en el que los agentes de software puedan interactuar de una página a otra fácilmente y puede llevar a cabo tareas más complejas para los usuarios.

b) Representación de conocimiento

Las computadoras deben tener acceso a información bien estructurada y además contar con las reglas de inferencia que permitan a los equipos razonar automáticamente para que la Web semántica funcione. Además, la Web semántica se encarga de resolver los problemas de representación de

conocimiento que enfrentan los sistemas actuales de representación de conocimiento, los cuales típicamente se encuentran centralizados y se requiere que todos compartan exactamente la misma definición de conceptos comunes.

c) Ontologías

Para que una máquina pueda interpretar la información disponible es necesario que ésta se encuentre estructurada de una forma clara y estandarizada, una manera de hacerlo es mediante ontologías, las cuales en Inteligencia Artificial y buscadores web, se refieren a un archivo o documento que define formalmente la relación entre términos. Esta estructura ayuda en la desambiguación de palabras y establecer la relación entre términos diferentes.

c) Agentes

El último elemento de la Web semántica son los agentes, los cuales son programas creados por los usuarios para coleccionar contenido de diferentes fuentes, procesar la información e intercambiarla con otros programas. La efectividad de estos agentes de software dependerá de lo entendible que sea el contenido Web y los servicios automáticos de razonamiento que estén disponibles.

De acuerdo con González (2009) este enfoque de la búsqueda semántica plantea que un buscador ya no sea una interfaz para recuperar documentos, sino un medio para realizar tareas basado en la Web.

2.3.1 Web semántica espacial

La Web semántica requiere que los contenidos de la Web tengan en su estructura descripciones enriquecidas por meta-información y así facilitar la búsqueda y recuperación de información y la interacción entre computadoras (Sarabia 2008) lo cual daría lugar a una Web más inteligente.

La iniciativa de la Web semántica geoespacial se enfoca específicamente en el tratamiento de la información geográfica que no es recuperada por los buscadores

de la Web semántica básica. Para lo cual UCGIS propone tres dimensiones básicas para la información geográfica en la Web semántica:

- *Profesional*: Información geográfica estructurada, almacenada en bases de datos espaciales, las cuales son indexadas o descritas en páginas Web.
- *Sencilla*: La recuperación de información geográfica no estructurada, subyacente e informal en páginas Web.
- *Científica*: Información geográfica en artículos científicos, modelos y teorías.

De acuerdo con Angela Schwering (2008) es difícil representar la semántica de objetos y conceptos geoespaciales, debido a algunas características especiales tales como que: los objetos geoespaciales son descritos típicamente por propiedades como la forma, tamaño y ubicación; además que las relaciones con otros objetos forman parte esencial de su representación semántica.

Para Egenhofer (2002) completar la Web semántica geoespacial requiere de una apretada agenda de trabajo de investigación en la cual hay dos tareas principalmente aparentes:

- Una forma canónica sobre cómo hacer las consultas sobre los datos geoespaciales.
- Investigar métodos para evaluar la semántica de fuentes de datos disponibles, para ver si la estructura semántica puede ser explotada por alguna tarea de consulta geoespacial particular.

2.4 Similitud semántica

Goldstone (2005) afirma que los humanos utilizan la similitud para recuperar y almacenar información para comparar nuevas situaciones con experiencias en el pasado. Además, Schewering (2008) menciona que los psicólogos consideran los juicios de similitud probablemente como los principales constructores del conocimiento humano.

2. Estado del arte

Los primeros aportes al tema de similitud se remontan a los años 60's, en donde para determinar la similitud de un objeto o entidad con otro se parte de las características comunes y diferentes entre dichos objetos.

Estos aportes a la similitud se basan en experimentos realizados manualmente por grupos de personas a los que se les encomendaba evaluar la similitud entre pares de sustantivos o nombres, en estos experimentos la evaluación constaba de un valor numérico, en un rango que iba de 0.0 para describir la ausencia de semejanza, a un valor 4.0 que era indicio de existencia. A partir de estos experimentos se derivaron una serie de propuestas para medir la similitud entre elementos de estructuras conceptuales taxonómicas.

Algunos enfoques miden la similitud de conceptos orientados principalmente en *corpus* de documentos y ontologías, como es el caso de las métricas de similitud basadas en Métodos Vectoriales y Métodos Probabilísticos.

Dentro de los métodos Vectoriales tenemos medidas de similitud tales como:

- Simple Matching Coefficient
- Coeficiente de Dice
- Coeficiente de Jaccard
- Coeficiente de Traslape
- Similitud del Coseno

Dentro de los métodos probabilísticos se encuentran:

- Divergencia de Kullback-Leibler (KL)
- Divergencia de Jensen-Shannon
- Norma Manhattan (L1)

Los métodos anteriores no se abordan específicamente en este trabajo, nos enfocamos a las medidas de similitud orientadas a la recuperación de información en ontologías, apegándonos a la clasificación de medidas de similitud propuesta por Blanchard, Harzallah, Briand, & Kuntz, (2005), quienes consideran para esta clasificación las primitivas básicas de las ontologías; conceptos y sus relaciones.

Según Blanchard, et al. (2005) para definir una métrica de similitud generalmente se especifican tres características.

Fuentes de Información. Cada medida considerada es basada en una ontología dada, algunas de ellas requieren un *corpus* de documentos de texto.

Principios. La mayoría de las métricas están basadas en principios axiomáticos, un ejemplo es la longitud del camino más corto.

Clases semánticas. En la literatura encontramos diferentes clases: distancia semántica, similitud semántica y relación semántica entre dos conceptos de la misma ontología.

2.4.1 Clasificación de medidas de similitud

Aunque principalmente la mayoría de propuestas están orientadas bajo el marco de WordNet. Generalmente podemos identificar dos enfoques de similitud en taxonomías(Schwering 2008):

- Semejanzas y diferencias
- Distancia semántica o conceptual

Estos enfoque se pueden clasificar a como se muestra en la Figura 2.5.

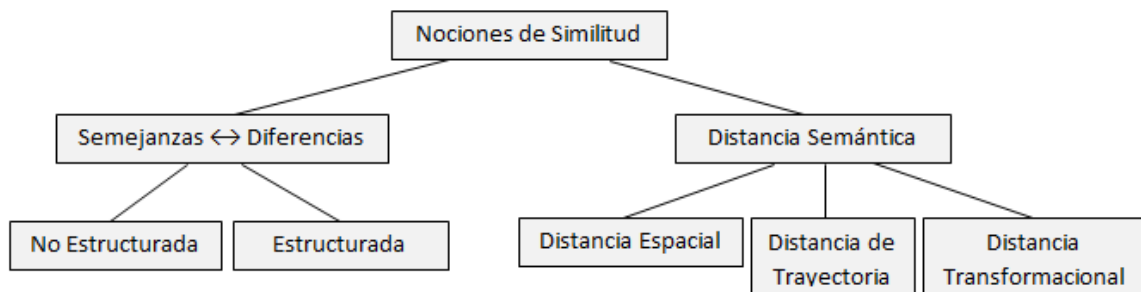


Figura 2.5 Diferentes nociones de similitud

En el caso de la noción de semejanzas y diferencias, entre dos representaciones conceptuales, toman un indicador de similitud, más coincidencias o menos

diferencias, el más alto indica la similitud, utilizando medidas basadas en comparaciones estructuradas o no-estructuradas.

La distancia semántica puede estar definida como distancia de recorrido, distancia espacial y distancia transformacional. Para aplicar la distancia semántica como una noción de la similitud semántica, todos los conceptos deben ser representados en una estructura común con alguna métrica específica, algunas medidas de similitud utilizan un espacio multidimensional como marco de referencia y una medida de distancia. La distancia semántica en un árbol o red estructurada que está definida por la longitud del camino más corto entre los nodos. La distancia transformacional es común en el uso de representaciones basadas en conjuntos de transformaciones, las cuales son compuestas y ejecutadas, una después de otra, la distancia se mide por medio del número de transformaciones o su complejidad.

(Blanchard, et al. 2005) citan algunas propuestas de métricas orientadas a medir la distancia semántica y otras orientadas a medir las semejanzas y diferencias entre conceptos, como se muestra a continuación:

Distancia de Rada. Se basa en el camino más corto entre dos conceptos c_1 y c_2 en una ontología restringida a enlaces taxonómicos.

Similitud de Resnik. Se establece en la siguiente hipótesis: entre más información tienen en común dos conceptos, más similares serán.

Similitud de Leacock y Chodorow. Corresponde a una transformación de la distancia de Rada en una similitud. El camino más corto entre dos conceptos de la ontología restringida a enlaces taxonómicos, se normaliza mediante la introducción de una división por el doble de la profundidad máxima jerarquía.

Similitud de Wu y Palmer. Es una medida entre los conceptos de una ontología restringida a enlaces taxonómicos. Se le han agregado dos parámetros que son la longitud de las dos rutas de c_i a c_j y desde c_j a c_i .

Similitud de Lin. Lin deduce de un enfoque axiomático una medida basada en una ontología restringida a enlaces taxonómicos y un *corpus*. Esta similitud tiene en cuenta la información compartida por dos conceptos como Resnik, además integra las diferencias entre ellos.

Distancia de Sussna. Se basa en todos los posibles enlaces. Para cada relación r , definimos un peso w de la relación r que va de c_i a c_j de un intervalo dado. Este peso se calcula con la densidad local que corresponde al número de relaciones del tipo R que van desde c_i .

Para Schwering (2008) las propiedades de la similitud semántica se pueden definir por tres axiomas, los cuales han sido muy criticados, debido a que presuponen las métricas en las distancias conceptuales.

Axioma 1. Minimalidad. Este axioma indica que, si la distancia entre dos objetos es cero, entonces ambos conceptos son conceptualmente iguales, es decir, se trata del mismo concepto (ver Ecuación 1).

$$d(i, j) = 0 \Rightarrow i = j \wedge d(i, j) \geq 0 \quad (1)$$

Hay críticos de las métricas de similitud quienes afirman que este axioma de minimalidad no se sostiene como axioma de similitud.

Axioma 2. Simetría. Este axioma indica que la distancia entre dos conceptos es la misma, partiendo de cualquiera de los dos conceptos y yendo al otro concepto (ver Ecuación 2).

$$d(i, j) = d(j, i) \quad (2)$$

Donde i, j son conceptos, $d(i, j)$ es la distancia conceptual entre ambos, partiendo del objeto i y $d(i, j)$ la distancia conceptual entre ambos partiendo del objeto j .

Axioma 3. Triángulo de desigualdad. Este axioma indica que la distancia entre dos conceptos es menor o igual a la distancia entre ambos conceptos, vía un tercer concepto, lo que se muestra en la Ecuación 3.

$$d(i, j) + d(j, k) \geq d(i, k) \quad (3)$$

Donde i, j, k son conceptos, $d(i, j), d(j, k), d(i, k)$ representan a la distancia entre los conceptos.

Algunas otras propiedades más de la similitud semántica son las siguientes:

La similitud semántica depende del contexto. La comparación semántica entre dos objetos, no sólo depende de sus características, o a la clase que pertenezca, depende también de la motivación y aplicación del usuario (Janowicz 2008).

La similitud semántica depende de la representación. Las teorías de similitud están relacionadas directamente con un lenguaje de representación en particular, en una representación computacional, como puede ser una descripción lógica de los objetos; sin embargo, los objetos y clases a las que pertenecen, cambian con el tiempo, así como sus representaciones computacionales (Raubal, 2008).

Similitud semántica, usabilidad y plausibilidad cognitiva. Comparando las medidas de clasificación computacionales con las medidas de similitud humanas, se define una estrategia para evaluar las medidas de similitud semántica, obteniendo clasificaciones para realizar comparaciones a nivel individual.

2.4.2 Distancia semántica

El enfoque propuesto por Tversky (1977) en el que define a una medida de similitud como un proceso de emparejamiento (acoplamiento) de entidades, esto produce un valor de similitud, no solo de semejanzas, sino también de diferencias entre entidades.

Una primera estrategia para un modelo basado en entidades es determinar una distancia semántica entre conceptos, como una distancia Euclidiana en un espacio semántico multidimensional, donde los ejes en este espacio describen las características de los conceptos.

Otro enfoque propone un modelo distancia-densidad basado en una función de distancia para la evaluación de la similitud que complementa la distancia con la densidad del espacio.

Con un enfoque muy diferente, los científicos de la computación han definido las medidas de similitud como medidas cuya estrategia básica usa las relaciones semánticas entre conceptos. Estas relaciones semánticas están típicamente organizadas en una red semántica, los enlaces (links) entre nodos denotan conceptos.

La distancia semántica resulta una intuitiva y directa forma de evaluar la similitud en una red semántica jerárquica. Para una red semántica con una relación sólo de "es-un", Rada et al. (1989), citado por Rodríguez et al. (1999) concluye que la relación semántica y la distancia semántica son equivalentes y se puede utilizar la segunda como una medida de la primera. Aquí se define la distancia conceptual como el tamaño de la ruta más corta entre dos nodos en una red semántica. Esta función de distancia conceptual satisface las propiedades de similitud propuestas anteriormente como son la minimalidad, simetría y desigualdad triangular.

Un enfoque más es el que propone Rodríguez et al. (1999) en donde nuevos conceptos pueden ser agregados o eliminados de la ontología, además puede haber diferentes maneras de clasificar un concepto basado en la aplicación específica y el grado de detalle de la definición del concepto.

Schewering (1998) y Rodríguez et al. (1999) sostienen que la evaluación de la similitud no es siempre una evaluación de simetría sino el resultado de las semejanzas y diferencias entre dos conceptos y la relevancia de las características que los diferencian (funciones, partes y atributos) entre los dos conceptos. Además, las relaciones de "es-un" complementadas con las relaciones "parte-todo" que genera una ontología reflejan mejor las interrelaciones entre conceptos.

Aunque los modelos de distancia semántica han sido empleados en un gran número de experimentos y han demostrado ser muy adecuados para un dominio

específico. La desventaja de estos modelos es que son altamente sensibles a la arquitectura de la red semántica predefinida.

En un escenario real, los nodos adyacentes no son necesariamente iguales, los conceptos de la mitad (o en medio) son más cercanos a muchos otros.

Lee (1993) propone ponderar los arcos que unen a los conceptos, para hacer más eficiente el modelo de distancia y así conseguir lo que se conoce como una red de densidad.

Otro modelo para medir la similitud es un modelo de información-contenido en el cual, las clases deben contener información de valor (como sinónimos), lo que genera que los resultados sean pesados y difíciles de comparar con otros conceptos.

Otro enfoque proveniente del dominio lingüístico el cual propone una medida de similitud se define en términos de sustitución de conceptos en sentencias o frases, el problema de este tipo de similitud es definir una forma sistemática de calcularla.

Para medir la distancia semántica entre conceptos, tomando como base el trabajo de Rodríguez et al. (1999) se menciona que el vecino semántico N de una entidad de clase a^0 es el conjunto de entidades C_i^0 cuya distancia $d()$ a la entidad a^0 es menor o igual a un entero no negativo r , llamado el radio de *vecindad semántica* (ver Ecuación 4).

$$N(a^0, r) = \{C_i^0\} \text{ donde } \forall i \ d(a^0, C_i^0) \leq r \quad (4)$$

La distancia entre dos entidades de clases en la ontología es medida por la ruta más corta, la cual se forma por el menor número de arcos no dirigidos que conectan esas entidades. Estos arcos representan subclases, superclases o relaciones del tipo *parte-todo*. La distancia es una métrica que satisface la propiedad de minimalidad, además que la distancia igual a 0 significa igualdad.

Un enfoque muy parecido es la manera en que Centeno, Kloos, Del Toro, and Gaedke, (2007), propone medir la similitud semántica, para lo cual propone dos

casos; en el primero, sólo se mide la similitud entre dos conceptos c_0 y c_1 donde c_1 es un subconcepto de c_0 y el segundo con c_1 y c_2 donde c_1 y c_2 no son subclases uno del otro.

En el primer caso utiliza los conceptos c_0 y c_1 , en donde c_1 es subclase de c_0 . Esto no implica que c_1 sea un hijo directo de c_0 en la ontología, debido a la transitividad de la propiedad *sub-ClassOf*. Con estas premisas en mente, la medida de similitud que se propone se puede expresar como se muestra en la Ecuación 5.

$$sim(c_0 ; c_1) = \frac{K * N_0}{K * N_0 + \log(\frac{E_0}{E_1})} \quad (5)$$

En donde N_0 es el número de nodos que hay entre el nodo raíz y c_0 , lo cual permite introducir la noción de que descender un nivel en la ontología supone mayor diferencia semántica en la parte alta de la ontología (la de los conceptos más genéricos) que en la parte baja (donde están los conceptos más específicos). El cociente $\frac{E_0}{E_1}$ representa la relación entre la información que aporta c_0 y la información que aporta c_1 .

Si suponemos que el nodo raíz contiene el 100% de la información presente en la ontología, cada uno de los nodos inferiores contendrá una porción de esa información. De tal manera que entre menos subclases tenga una clase más parecidas serán éstas a la clase padre.

Para el segundo caso con c_1 y c_2 que no son subclase uno del otro. Para ello, es preciso introducir el concepto de distancia. La medida propuesta en la Ecuación 6 da lugar a un valor entre 0.0 (ninguna similitud) y 1.0 (mismo concepto). De este modo, la distancia *dist* entre dos conceptos se puede definir a partir de su similitud semántica *ssim* como:

$$dist(c_1 , c_2) = \frac{1}{ssim(c_1 , c_2)} - 1 \quad (6)$$

Cuando c_1 y c_2 sean el mismo concepto, *ssim* valdrá 1 y la distancia entre ambos será 0. Análogamente, cuando *ssim* valga 0, la distancia entre ambos será infinita.

Con esto en mente podemos calcular la similitud entre c_1 y c_2 a partir de la distancia entre ambos.

De acuerdo con Tversky (1977), la similitud podría verse mejor como la comparación de las características en común y las diferencias entre conceptos que el cálculo de la distancia entre conceptos. La función de la distancia semántica entre conceptos para Tversky se define mediante tres axiomas, presentados a continuación.

- Minimalidad $\delta(a, b) \geq \delta(a, a) = 0$ (7)

- Simetría $\delta(a, b) = \delta(b, a)$ (8)

- Triángulo de desigualdad $\delta(a, b) + \delta(b, c) \geq \delta(a, c)$ (9)

También propone hacer la coincidencia (matching) de propiedades entre conceptos, por lo general las propiedades de las entidades son variables binarias o variables con un valor ordinal, pero no se restringen únicamente a esos valores, ya que pueden ser variables ordinales o cardinales (dimensiones). La función para hacer la igualdad de entidades que propone Tversky se muestra en la Ecuación 10.

$$s(a, b) = (A \cap B, A - B, B - A) \quad (10)$$

Donde:

- s es la similitud.
- a, b son los conceptos a los que se quiere aplicar el *matching*.
- A Es el conjunto de las propiedades del concepto de la clase a .
- B Es el conjunto de las propiedades del concepto de la clase b .
- Por tanto, la similitud se expresa como una función f de tres argumentos:
- $A \cap B$ Las propiedades que son comunes entre A y B .
- $A - B$ Las propiedades que tiene A pero no tiene B .
- $B - A$ Las propiedades que tiene B pero no tiene A .

El trabajo de Rodríguez et al. (2002) tiene bases lingüísticas bien definidas, debido al uso de hipónimos y sinónimos. Las conclusiones de sus estudios indican que el modelo utilizado tiene buenos resultados siempre y cuando las ontologías tengan una representación completa y detallada de las clases y entidades.

Uno de los principales problemas en el diseño de la ontología es que ésta debe ser escalable; además de permitir la integración de nuevos conceptos y entidades. Una buena estrategia para la integración de ontologías ya existentes es el mapeo de dichas ontologías a una ontología más general.

2.5 Servicios WEB

Los servicios Web (SW) son entidades computacionales que ofrecen alguna funcionalidad y que permiten el intercambio de datos entre aplicaciones. La interoperabilidad se consigue mediante la adopción de estándares abiertos (Shen, Duan, Ren, and Yang, 2010).

Las organizaciones OASIS (Organización para el Avance de Estándares de Información Estructurada) (OASIS, 2013) y W3C (Consortio de la World Wide Web) (W3C, 2013), son los comités responsables de la arquitectura y reglamentación de los servicios Web. Para mejorar la interoperabilidad entre distintas implementaciones de servicios Web se ha creado el organismo WS-I, encargado de desarrollar diversos perfiles para definir de manera más exhaustiva estos estándares.

Los servicios Web permiten a las compañías publicar componentes y servicios en un directorio en el que otras aplicaciones Web pueden buscar e implementar nuevos servicios a través de una llamada.

Las tecnologías previas al desarrollo de los SW fueron conocidas como objetos distribuidos, dentro de las que se encuentran DCOM de Microsoft, RMI de Sun (RMI, 2013), y CORBA de OMG (Object Management Group) (CORBA,2013). Sin embargo, a pesar de las alianzas de las compañías para conseguir interoperabilidad entre diferentes aplicaciones, y la no participación de Microsoft

2. Estado del arte

en el desarrollo de CORBA, ocasionó que se frenara el desarrollo de esta tecnología para dar paso a los SW, que aunque se fundamentan en nuevas características, siguen buscando la interacción entre aplicaciones.

Muchos esfuerzos se han hecho y se están realizando en la actualidad para definir las diferentes especificaciones y arquitecturas que posibiliten la expansión de este nuevo tipo de aplicaciones denominadas “Servicios Web”. Grandes empresas como Microsoft IBM, Intel, y consorcios como W3C están profundamente involucradas en esta investigación.

W3C aporta la siguiente definición en su grupo de investigación de SW: “*Un servicio de Web es un sistema de software diseñado para soportar la interacción entre dos máquinas a través de una red. Posee una interfaz descrita en un formato que puede ser procesado por una máquina (específicamente WSDL). Otros sistemas pueden interactuar con el Servicio Web en la manera prescrita por su descripción utilizando mensajes SOAP, típicamente transportados utilizando HTTP y serializados con XML, en conjunción con otros estándares relacionados con la Web*” (W3C, 2004) el esquema general de un servicio web se muestra a continuación en la Figura 2.6.

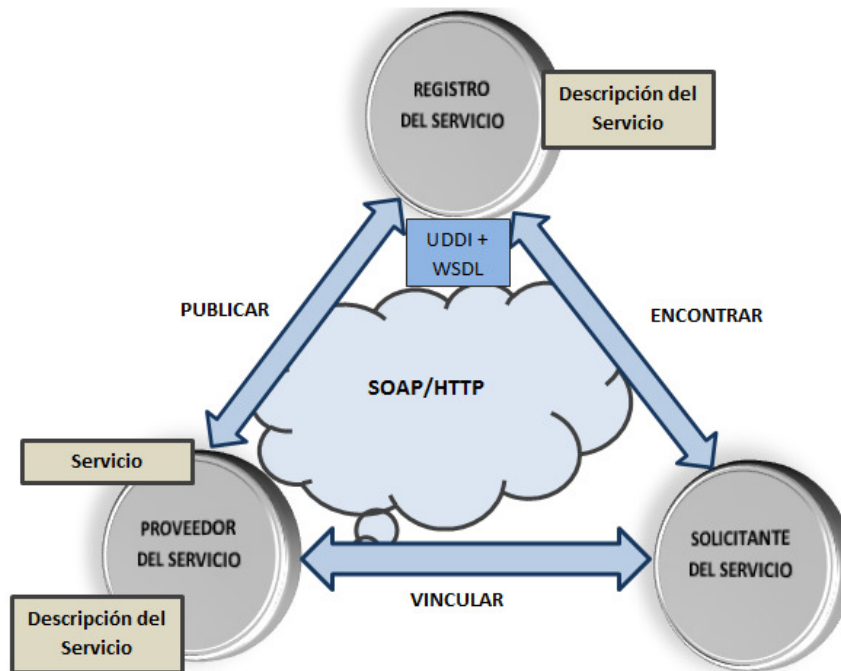


Figura 2.6 Relación de los componentes de un servicio web

Examinando los componentes de un servicio Web podemos encontrar los siguientes elementos (W3C, 2004):

- **Servicio.** La aplicación es ofrecida para ser utilizada por solicitantes que completan los requisitos especificados por el proveedor de servicios. La implementación se realiza sobre una plataforma accesible en la red. El servicio se describe a través de un lenguaje de descripción de servicio. Tanto la descripción como las políticas de uso han sido publicadas de antemano en un registro.
- **Proveedor de Servicio.** Desde el punto de vista comercial, es quien presta el servicio. Desde el punto de vista de arquitectura, es la plataforma que provee el servicio.
- **Registro de Servicios.** Es un depósito de descripciones de servicios que puede ser consultado, donde los proveedores de servicios publican sus servicios y los solicitantes encuentran los servicios y detalles para utilizar dichos servicios.
- **Solicitante de servicios.** Desde el punto de vista comercial, la empresa que requiere cierto servicio. Desde el punto de vista de la arquitectura, la aplicación o cliente que busca e invoca un servicio.

2.5.1 Servicios Web semánticos

- **Ontologías.** Proveen el conocimiento de soporte para el objetivo y la descripción del servicio web
- **Objetivos.** Describe la funcionalidad y forma de interacción con el servicio Web.
- **Descripción de servicio web.** Especifica la funcionalidad y los medios de interacción provistos por el servicio Web.
- **Mediadores.** Conectan diferentes elementos de servicios Web de modelado de ontologías y resuelven la heterogeneidad en la representación de los datos.

Los servicios web semánticos son definidos como WSMO y WSML comprenden los diferentes formalismos con el fin de investigar su aplicabilidad a la descripción de los Servicios Web Semánticos.

2.5.2 Servicios Web Geoespaciales

Los Servicios Web Geoespaciales (GWS - por sus siglas en inglés) tienen el propósito de permitir la integración de geoprocesamiento distribuido provisión de datos funcionales (OGC, 2001). De acuerdo con Gone and Schade (2008) a través de GWS, los componentes de los Sistemas de Información Geográfica (SIG) se descomponen y se convierten en accesibles a través de las Infraestructuras de Datos Espaciales (IDE). Los componentes separados se pueden acceder e integrar en una red. El Open Geospatial Consortium (OGC) normaliza ampliamente interfaces de servicios Web como un marco evolutivo de GWS. Uno de los estándares OGC más destacados es el de especificación e implementación Web Feature Service (WFS).

Los servicios web de mapas (WMS - por sus siglas en inglés) y los servicios basados en entidades (WFS) son dos de los servicios web más utilizados. Estos dos servicios tienen un papel importante en cuanto a la interoperabilidad y compartir información espacial. WMS/WFS han sido utilizados de diferentes maneras por sus proveedores debido a que OGC no ha establecido alguna especificación estándar para su administración (Shen, 2010).

Servicio web de mapas (WMS) es probablemente el estándar más utilizado para recuperar imágenes de mapas geo-referenciados a través de internet, a diferencia del WMS el servicio web basado en entidades (WFS) presenta la información de algún objeto geográfico y así mismo permite la actualización, creación y borrado de dichos objetos geográficos.

2.6 Sistemas de información geográfica

Un GIS es un sistema de información particular, el cual no se compone únicamente del software para el manejo de la información geográfica, sino que

está compuesto de software, hardware y datos; los cuales permiten la captura, análisis, manejo y visualización de cualquier tipo de información geográficamente referenciada (Tomlinson, 2007).

Los sistemas de información geográfica a través del tiempo han ayudado a resolver problemas específicos, ya que sus herramientas ayudan a ubicar las entidades u objetos afectados por ciertos fenómenos, así también los GIS combinados con algunas otras herramientas de análisis pueden ayudar a resolver distintos tipos de problemas, cómo en el apoyo a la toma de decisiones, establecimiento de rutas comerciales, predicción de propagación o afectación de algún fenómeno natural o social, entre muchos otros.

2.6.1 Componentes y características de un GIS

Un GIS es un sistema complejo de partes interconectadas, por lo cual debe estar compuesto de seis elementos que se listan a continuación.

2.6.1.1 Componentes básicos de un GIS

De acuerdo con Tomlinson (2007), los componentes básicos de un GIS son:

Hardware: Un GIS trabaja con grandes volúmenes de información en forma de mapas, imágenes, bases de datos, entre otras; lo cual hace que el procesamiento de la misma requiera explotar la potencia de algunas computadoras y en algunos casos una buena infraestructura de cómputo para manejar el sistema.

Software: Los programas de software son los encargados de proporcionar las funciones para el análisis y para crear los productos de información que se requieren al crear un GIS.

Datos: Una vez que se identifica que productos informativos se requiere que el GIS provea, es necesario planear la adquisición de los datos que se necesitan para obtener esos resultados. En muchas ocasiones, el formato en que se encuentran los datos que utilizamos, son los que influyen en la elección del software a utilizar.

Productos informativos: Se refiere a los productos que se desea obtener de un GIS ya sean mapas, imágenes, informes, gráficos, o cualquier combinación de estos elementos.

Los dos elementos restantes se refieren a la parte operativa del sistema de información geográfica; es decir, un elemento son los procedimientos que rigen la manera en que los usuarios deben operar el sistema y el otro elemento es el recurso humano, ya que para el buen funcionamiento de un GIS es necesario que las personas que se encargan de operarlo tengan la capacitación necesaria para hacerlo.

2.7 Ontologías

Según Wache et al. (2001) para compartir información, es necesario resolver algunos problemas, como la interoperabilidad entre los sistemas encargados de la recuperación de información y el sistema que hace la consulta. La interoperabilidad no requiere acceso completo a los datos, sino que requiere que los datos a los que accede puedan ser procesados e interpretados por el sistema remoto. De aquí que se originen problemas de heterogeneidad estructural en los que diferentes sistemas de información almacenan sus datos en diferentes estructuras y heterogeneidad semántica, la cual considera el contenido de un elemento de información y que diferentes sistemas los interpretan de diferente manera.

Para lograr la interoperabilidad semántica en un sistema de información heterogéneo, el significado de la información que se procesa en el sistema debe ser entendido de igual manera por todos los sistemas, ya que de hacer una interpretación diferente de la información se podrían generar conflictos como:

- **Confusión:** Cuando los elementos de información parecen tener el mismo significado pero en realidad es diferente, por ejemplo, a los diferentes contextos temporales.
- **Escala:** Se producen cuando los diferentes sistemas de referencia se utilizan para medir un valor, por ejemplo para las distancias o monedas.

- **Nombres:** Se presentan cuando los esquemas de nomenclatura de información difieren significativamente. Un fenómeno frecuente es la presencia de homónimos y sinónimos.

Para poder superar los problemas antes mencionados relacionados con la heterogeneidad semántica, Lutz y Klein (2006) y Wache et al. (2001) están de acuerdo con que el uso de ontologías es un enfoque posible. Además Uschold y Grüninger (1996) describen a la interoperabilidad como una aplicación clave de las ontologías, por lo que es necesario erradicar esos problemas de interoperabilidad.

La definición declarativa para ontologías más consolidada es la propuesta por Gruber (1995) y Studer, Benjamins, and Fensel (1998) para quienes una ontología es *“la especificación formal y explícita de una conceptualización compartida, donde la conceptualización puede ser entendida como la forma de pensar a cerca de cierto dominio”*. Para Lutz y Klein (2006) el uso de ontologías permite enriquecer la descripción de las fuentes de información, así la semántica del contenido llega a ser entendible por la computadora y los usuarios tienen la capacidad de hacer consultas concisas y completas. Además, se puede utilizar el razonamiento lógico para descubrir relaciones implícitas entre los términos buscados y descripciones de la información.

Según Wache (2001) hay distintos enfoques para implementar una ontología. Una ontología simple, múltiples ontologías o híbridas, para el enfoque de una ontología simple, se utiliza una ontología global que provee un vocabulario compartido para la especificación de la semántica y todas las fuentes de información están relacionadas a la misma ontología, esta única ontología también puede ser la combinación de múltiples ontologías especializadas.

En un enfoque de múltiples ontologías cada fuente de información se describe por su propia ontología. Por lo que no se requiere que todas las fuentes de información converjan en una sola ontología, además de que es posible que las ontologías no compartan la misma terminología.

Las ontologías pueden ser altamente informales, si son expresadas en lenguaje natural.

- Semi-informales, si son expresadas en una forma restringida y estructurada del lenguaje natural.
- Semi-formales, si son expresadas en un lenguaje artificial y definido formalmente (Ontolingua, OWL).
- Rigurosamente formales, si proporcionan términos bien definidos con una semántica formal, teoremas y completitud.

2.7.1 Construcción de una ontología

Aquí se muestran algunos criterios y un conjunto de principios que han probado ser útiles en el diseño de una ontología (Gómez-Pérez, 1997).

1. *Claridad y objetividad.* La ontología debe comunicar con eficacia el significado deseado de los términos definidos. Las definiciones deben ser objetivas y en un lenguaje natural.
2. *Coherencia.* La ontología debe permitir que las inferencias sean consistentes con las definiciones.
3. *Completa.* La ontología debe expresar definiciones completas sobre definiciones parciales.
4. *Extensibilidad.* Una ontología se debe diseñar anticipando el uso de vocabulario compartido; es decir, se debe poder definir términos nuevos para aplicaciones especiales basadas en el vocabulario existente, de manera que no se requieran revisar las definiciones existentes.
5. *Minimizar el compromiso ontológico.* Una ontología puede requerir un mínimo de compromiso ontológico, el cual sea suficiente para soportar las actividades relacionadas con compartir el conocimiento. Debido a que el compromiso ontológico está basado en el uso constante del vocabulario, se puede reducir al mínimo, especificando una teoría más débil (que se permite en la mayoría de los modelos) y definiendo solamente esos términos que sean esenciales para la comunicación de conocimiento consistente.

6. *Principio de distinción ontológica.* Significa que las clases en una ontología deben ser disjuntas, para lo cual se utiliza un criterio denominado de identidad.
7. *Diversificación de jerarquías.* Para incrementar la fortaleza de los mecanismos de herencia múltiple.
8. *Modularidad.* Para minimizar la unión entre módulos.
9. *Minimizar la distancia semántica entre conceptos hermanos.* Conceptos similares son agrupados y representados como subclases de una clase y pueden ser definidas usando las mismas primitivas, donde los conceptos son poco similares y representados en la jerarquía.
10. *Estandarizar.* Nombres hasta donde sea posible.

2.7.2 Tipos de ontologías

Los tipos más comunes de ontologías, de acuerdo con su nombre, dan una idea del conocimiento que incluyen (Corcho, Fernández-López, y Gómez-Pérez, 2003).

- Ontologías que representan el conocimiento capturan las primitivas de representación para formalizar el conocimiento en paradigmas que representen el conocimiento. El ejemplo más representativo de este tipo de ontologías es “Frame-Ontology”, la cual captura las primitivas de representación (clases, instancias, etc.), basadas en determinados lenguajes.
- Ontologías generales o comunes. Incluyen vocabularios relacionados con cosas, eventos, tiempo, espacio, funciones, causalidad, conductas, etc.
- Meta-ontologías. También llamadas como ontologías genéricas o “Core Ontologies”, las cuales se reutilizan a través de diversos dominios. El ejemplo más representativo puede ser una ontología de mereología, la cual incluye el término parte de (*part-of*).
- Ontología de Dominio. Describe el vocabulario relacionado a un dominio genérico como medicina o automóviles y sus relaciones, de las actividades que tienen lugar en este dominio, y de las teorías y principios elementales que gobiernan dicho dominio.

- **Ontología de Tarea.** Describe un vocabulario sistematizado de los términos usados para resolver problemas asociados con tareas que pueden o no pertenecer al mismo dominio. Estas ontologías proveen un conjunto de términos que describen genéricamente como resolver un tipo de problemas.
- **Ontología de Tarea de Dominio.** Son ontologías de tarea reutilizables en un dominio en específico, pero no a través de otros dominios, por ejemplo, una ontología de tarea de dominio en el dominio médico, puede incluir términos relacionados con el calendario de una operación.
- **Ontología de Aplicación,** contienen el conocimiento necesario para modelar un dominio en particular. Describen conceptos que dependen tanto de un dominio como de una tarea en particular, los cuales frecuentemente son especializaciones de ambas ontologías. A menudo, estos conceptos corresponden a los roles desempeñados por entidades del dominio mientras realizan cierta actividad.
- **Ontologías de Información** especifican la estructura de almacenamiento de bases de datos. Ofrecen un marco para el almacenamiento estandarizado de información.

2.7.3 Metodologías para construir una ontología

Uno de los principales problemas en el diseño de la ontología es que sea escalable además de permitir la integración de nuevos conceptos y entidades. Una buena estrategia para la integración de ontologías ya existentes es el mapeo de dichas ontologías a una ontología más general según Rodríguez y Egenhofer (2002).

Una manera de facilitar la escalabilidad e integración de ontologías es seguir una metodología para su desarrollo, algunos trabajos proponen una serie de pasos y metodologías para el diseño de ontologías, entre los más representativos se describen en (Corcho et al., 2003).

2.7.3.1 *Methontology*

La metodología Methontology (Fernández-López, 1999) fue diseñada en el Laboratorio de la Universidad Politécnica de Madrid. Esta metodología permite la construcción de ontologías a un nivel de conocimiento e incluye las siguientes tareas:

- La identificación del proceso de diseño de la ontología.
- Un ciclo de vida basado en prototipos.
- Técnicas particulares para la realización de cada actividad.

Para la construcción de la ontología, en esta metodología se consideran los siguientes pasos:

- Actividades para el manejo del proyecto, que incluye,
 - Planeación, se deben identificar cada una de las tareas, su duración y recursos destinados a la misma.
 - Control, se debe garantizar que las actividades se realicen.
 - Aseguramiento de la calidad, la calidad de cada producto, consecuencia de las diversas actividades debe ser satisfactoria.
- Actividades orientadas al diseño, entre las que se incluye,
 - Especificación, donde la ontología se diseña de acuerdo con los objetivos específicos y al uso que se le dará.
 - Conceptualización, se estructura el dominio del conocimiento como modelos significativos a un nivel de conocimiento.
 - Formalización, se transforma el nivel conceptual a un modelo formal.
 - Implementación, se construye el modelo de la ontología en un modelo computacional.
 - Mantenimiento, se actualiza y corrige la ontología.
- Actividades de soporte, las cuales incluyen una serie de actividades que se realizan al mismo tiempo a las actividades orientadas al diseño, mientras que la ontología no se ha construido, estas actividades son:
 - Adquisición de Conocimiento, se obtiene el conocimiento de un dominio determinado.

- Evaluación, se evalúan las ontologías existentes, su software asociado y documentación, dentro de su ambiente de desarrollo.
- Integración, se construye la ontología, en su caso utilizando ontologías ya existentes.
- Documentación, se documentan cada una de las fases de construcción y productos generados durante la construcción de la ontología.
- Manejo de la configuración, se registran todas las versiones de la documentación, software y código de la ontología.

2.7.4 Lenguajes de Ontologías

Los lenguajes para la construcción de ontologías tienen sus inicios en la década de los 90's en el campo de la inteligencia artificial, los cuales se basaban en paradigmas de representación del conocimiento y son considerados como los precursores de los actuales lenguajes para la construcción de ontologías. Algunos de estos lenguajes son descritos en Corcho et al. (2003).

- **KIF** (Formato para Intercambio de Conocimiento). Su modelo de conocimiento incluye lógica de primer orden, complementado con marcos.
- **Ontolingua** (basado en KIF). Se convirtió en un estándar por la comunidad ontológica en la década de los noventa, incluye una perspectiva orientada a objetos, una extensión basada en marcos y una librería de ontologías.
- **LOOM**. Fue creado para construir bases de conocimiento en general y no ontologías. Está basado en lógica descriptiva, reglas de producción y provee clasificación automática.
- **OKBC** (protocolo para la Conectividad de Bases de Conocimiento Abierto). Este protocolo permite acceder a bases de conocimiento almacenadas en diferentes sistemas de representación de conocimiento y las cuales pueden estar basadas en diferentes paradigmas de representación de conocimiento.
- **SHOE** (Simple HTML Ontology Extension). Es un lenguaje de representación del conocimiento diseñado para la Web, basado en la

asociación de contenido de páginas Web, vinculándolas con una o varias ontologías. Se trata de un anotador de contenido asociado a páginas Web, su modelo de conocimiento son los marcos y está basado en la sintaxis de XML, así como XOL y OML.

- **OML.** Fue parcialmente desarrollado con base en SHOE.
- **XOL.** Lenguaje para intercambio de ontologías basado en XML.
- **RDF** (Resource Description Framework). Es un marco para la descripción de recursos, su modelo de datos compuesto por: Objetos considerados como recursos definidas como expresiones RDF descritas mediante URI's y opcionalmente anclados a ID's. Propiedades que describen las características, relaciones y atributos o aspectos específicos de los recursos. De éste se deriva RDF Schema o RDFS, en el cual las primitivas son agrupadas en: clases, propiedades, contenedores de clases y propiedades, colecciones, vocabulario de refinación (transformación de algo abstracto en concreto) y propiedades de utilidad.
- **OWL** (Ontology Web Language). Es un lenguaje de marcado para publicar y compartir datos utilizando ontologías a través de la Web, añade más vocabulario para describir: clases (conceptos), propiedades (de tipos de datos y de objetos), relaciones entre clases, individuos (instancias). OWL proporciona tres sub-lenguajes, cada uno con nivel de expresividad mayor que el anterior, diseñados para ser utilizados por comunidades específicas de desarrolladores y usuarios.
 - a. **OWL Lite.**
 - b. **OWL DL.**
 - c. **OWL Full.**

2.8 Trabajos relacionados

Hiramatsu y Reitsma (2004) realizaron un trabajo en el que se presenta un enfoque de referencias geográficas que utilizan las tecnologías de la Web Semántica. Este enfoque consiste en dos partes, la creación de una ontología geográfica y sus instancias, y la aplicación de dos herramientas que utilizan la ontología. La ontología geográfica incluye clases típicas de entidades geográficas y las relaciones geográficas (por ejemplo, relaciones topológicas, de distancia y relaciones de dirección). “La ontología geográfica está definida en OWL para poder utilizarse con otras ontologías geográficas que estén basadas en RDF” según encontramos en Hiramatsu (2004). La primera herramienta es un servicio Web que nos permite calcular las relaciones geográficas entre las instancias geográficas utilizando sus coordenadas. Son accesibles mediante SOAP basado en descripciones de servicios en WSDL para que poder utilizarse en otras aplicaciones como: editores RDF, motores de inferencia, y motores de búsqueda.

La segunda herramienta es un *plug-in* para un SIG de código abierto ligero que genera RDF para describir las relaciones espaciales entre los elementos geográficos.

Mediante la primera herramienta se pueden hacer consultas de entidades de la ontología mediante sus propiedades, tales como la distancia con otra entidad, la orientación, incluso mediante las 8 relaciones entre entidades como: igualdad, disjuntos, cruza con, interseca, toca, dentro de, contiene y cubre. Este trabajo está enfocado en la semántica entre entidades geográficas.

2.9 Aplicaciones relacionadas

2.9.1 Servicios web de recuperación de información

Para Shen (2010) gracias al desarrollo de SOA (Arquitectura Orientada al Servicio - por sus siglas en inglés), el desarrollo de software ha tomado un nuevo rumbo. También afirma que al combinar SOA con un Sistema de Información Geográfica, nos acercamos a lo que se conoce como un Servicio de Información Geográfica.

Actualmente la implementación de SOA en la Web se caracteriza como Servicios Web. Existen aplicaciones que se enfocan en la recuperación de información geográfica como:

- **Google Maps** Existen en la actualidad varias aplicaciones orientadas a la búsqueda de información espacial. Google Maps es una de las más comunes en cuanto a la búsqueda y localización de empresas y lugares de interés, la cual se enfoca en la realización de *mashups* de distintas fuentes de información.
- **Nestoria** Permite la búsqueda y localización de inmuebles para compra y/o renta.

2.9.2 Recuperación semántica de información

Otras aplicaciones hacen uso de técnicas de recuperación semántica de información como:

- **Kartoo.** Clasifica los resultados en categorías que permiten al usuario seleccionar rápidamente el que le interesa, dicha clasificación se basa en la recuperación semántica de información (KARTOO, 2004).
- **Ontogator.** Emplea las búsquedas basadas en ontologías para la recuperación de imágenes mediante una consulta semántica a la descripción de las imágenes almacenada en una base de datos (Hyvonen, Saarela, y Viljanen, 2003).
- **Neptuno.** Esta aplicación hace uso de ontologías para la descripción de información, la cual es manejada mediante un módulo de búsqueda semántica y un módulo de presentación y visualización de ontologías (Castells, 2004).
- **SABIOS.** Con el objetivo de aplicar técnicas de la Web Semántica al proceso de búsqueda de información, surge el proyecto SABIOS; un sistema multi-

agente de recuperación semántica especializado en los temas de hábitat y artes (Giraldo, Luna, y Ceballos, 2006).

- **Swoogle.** Es un rastreador de indexación basado en un sistema de reproducción de la Web Semántica. Extrae los metadatos para cada documento recuperado y calcula las relaciones entre los documentos (Ding, Finin, Joshi, Pan, Cost, Peng & Sachs, 2004).
- **DIP.** (Data, Information and Process Integration with Semantic Web Services) El objetivo de DIP es desarrollar y extender la Web Semántica y las tecnologías de los Servicios Web para producir una nueva infraestructura tecnológica para los Servicios de la Web Semántica (Cabral, Domingue, Galizia, Gugliotta, Tanasescu, Pedrinaci, & Norton, 2006).
- **Esperonto Project.** Es un proyecto que fue desarrollado entre el 2002 y el 2005 dentro del 5º Programa Marco de la Comisión Europea, y cuyo objetivo era desarrollar un puente entre la Web actual y la Web Semántica (Declerck, Contreras, Corcho, & Crispi, 2004).
- **DBin.** Es una nueva clase de aplicación Web: un P2P Web Semántico con una filosofía para “compartir archivos” y/o un cliente de grupos de noticias semántico. Este sistema es similar a un cliente para compartir archivos, DBin conecta directamente a otros puntos. En lugar de archivos, se descarga “información relevante” sobre los temas que se especifiquen. De forma más técnica, permite compartir y recibir “información estructurada semánticamente”, utilizando RDF y otros componentes de la Web Semántica.

2.9.3 Sistemas orientados a la recuperación de información geográfica

Existen algunas aplicaciones que sólo están orientadas a la recuperación de información geográfica, tales como: GeoNames, Google Maps, Yahoo Maps, Bing Maps, Waze, Foursquare, entre muchos otros.

2. Estado del arte

Una característica en común de algunos de estos sistemas es que utilizan perfiles de usuario para representar las necesidades de información e intereses de los mismos.

CAPÍTULO 3. METODOLOGÍA

3. Metodología

La presente metodología propone la integración de la información para su análisis semántico, así como la elaboración de un modelo persistente de datos para un posterior análisis de los mismos. Esta metodología está compuesta por tres etapas: Conceptualización, Recuperación y Visualización.

Dentro de la etapa de *Conceptualización* se realiza la integración de la información que se desea manejar dentro del sitio Web, para su análisis y la posterior recuperación semántica de información geográfica mediante un modelo persistente de datos, así como el establecimiento de los criterios de búsqueda, acordes a los atributos de la información con la que se cuenta. En la etapa de *Recuperación* se consideran los criterios de búsqueda para hacer la recuperación semántica de la información, aplicando las medidas de similitud que garanticen que los resultados sean lo más apegados a los deseos del usuario. En la etapa de *Visualización* se propone un ordenamiento (*ranking*) de los resultados obtenidos y la representación gráfica de los mismos al usuario mediante un mapa digital.

3.1 Suposiciones Iniciales

Es difícil en gran medida definir la información a utilizar para ciertos análisis, por lo que para realizar la recuperación semántica de información geográfica en nuestro caso de estudio se establecen los siguientes supuestos.

- La información acerca de las instituciones de educación superior ya se encuentra integrada en el sistema, por lo que el mismo no se hará cargo de actualizar o recabar nuevamente dicha información.
- La estructura jerárquica definida para la información, tiene su origen en el catálogo que propone el ENOE (Encuesta Nacional de Ocupación y Empleo).

3.2 Marco de trabajo

Como se mencionó anteriormente, la metodología está compuesta de tres etapas: Conceptualización, Recuperación y Visualización. Dicho de otra manera, el esquema general del trabajo consiste en integrar la información referente a las instituciones de educación superior, posteriormente realizamos la búsqueda y recuperación de la información mediante algunos mecanismos de similitud, para más adelante integrarla en una interfaz Web, la cual permita al usuario tener una mejor alternativa en su toma de decisiones. En la Figura 3.1 se muestra el marco general de trabajo con las etapas y tareas involucradas en el proceso

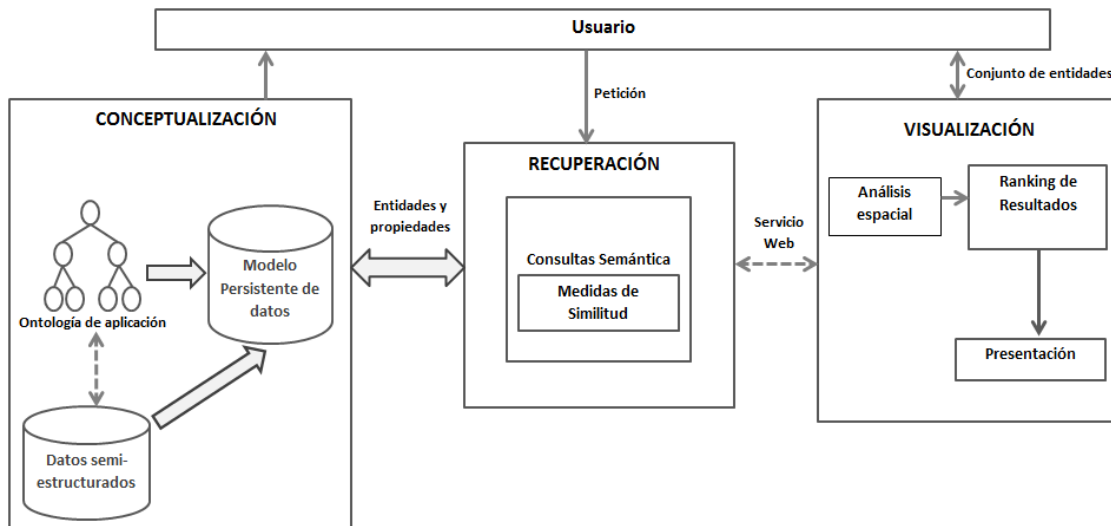


Figura 3.1 Marco de trabajo propuesto

3.3 Etapa de Conceptualización

Esta etapa consiste en formalizar la información con la que se cuenta inicialmente como una percepción informal del dominio, originando que ésta se transforme en una especificación formal por medio de la definición de una ontología de aplicación. En la Figura 3.2 se muestra un diagrama general de las tareas que conlleva esta etapa.

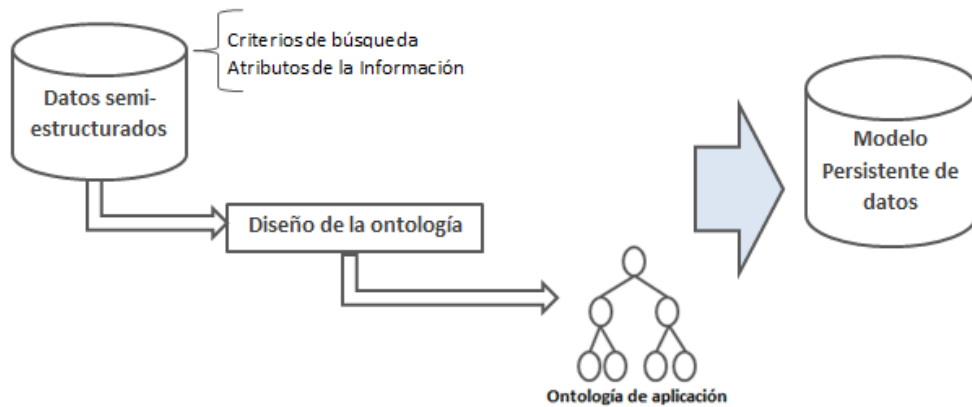


Figura 3.2 Etapa de Conceptualización

De acuerdo con la Figura 3.2, para iniciar con esta etapa se cuenta con una colección de datos semi-estructurados, los cuales están organizados como atributos de las entidades que componen la información. Además, estos atributos son los que rigen los criterios de búsqueda en el sistema de recuperación, con base en los atributos de la información y los datos semi-estructurados se realiza el diseño de la ontología de aplicación que contiene nuestra información. Posteriormente, se da inicio a la integración de todas las entidades y sus respectivas propiedades y relaciones dentro de la ontología, a lo que se conoce como el proceso de síntesis o poblado de la ontología. Una vez que se tienen los datos integrados en la ontología, es necesario crear un modelo persistente de datos, el cual permite utilizar razonamiento mediante motores de inferencia para dar pie a pasar a la etapa de recuperación semántica de la información.

3.3.1 Datos semi-estructurados

De acuerdo con Visser (2002) la información no siempre se puede encontrar en una misma fuente de datos, es por eso que es necesaria una preparación de los datos antes de hacer la integración y análisis. La información para nuestro caso de estudio se encuentra como un conjunto de datos semi-estructurados en una hoja de cálculo. Dicha información está compuesta por el nombre, tipo, dirección, carreras y costo de la colegiatura en cada universidad, en caso de las demás entidades, se pueden utilizar los mismos atributos, excepto el de carreras.

3.3.2 Criterios de Búsqueda

Así, los criterios de búsqueda introducidos por los usuarios están limitados a los atributos de la información con que se cuenta, de esta manera el usuario sólo podrá requerir las características de las entidades que se encuentren almacenadas en la base de conocimiento.

3.3.3 Atributos de la información

Para cada entidad con la que se cuenta en el repositorio de información, se tienen los atributos de nombre, ubicación, costo aproximado y tipo de entidad (si es pública o privada para el caso de estudio específico) y la información de las carreras que se imparten en cada entidad nos sirve para poder hacer la clasificación de universidades. En el caso de cualquier entidad de otro tipo, la clasificación se puede llevar a cabo con las diferentes áreas en las que pudieran encontrarse; cabe mencionar que una entidad no es específica de una sola área.

3.3.4 Ontología de aplicación

Con los atributos de la información y una metodología definida se procede al diseño de la ontología de aplicación. Para la construcción de la ontología se revisaron algunas metodologías como la que proponen Uschold and King's (1995), citado por Corcho et al. (2003) en la que describen cuatro actividades principales para la creación de ontologías como; definir el propósito de la ontología, construirla, evaluarla y documentarla. Grüninger y Fox (1995), citado por Corcho et al. (2003) proponen una metodología inspirada en el desarrollo de sistemas basados en conocimiento utilizando una lógica de primer orden, la cual consiste en seguir una estrategia *bottom-up* de desarrollo. Otras metodologías emplean estrategias *top-down* para el desarrollo de ontologías y finalmente METHONTOLOGY (Fernández-López, 1999) propone una serie de actividades no estrictamente secuenciales para el desarrollo de ontologías, dichas actividades se muestran en la Figura 3.3.

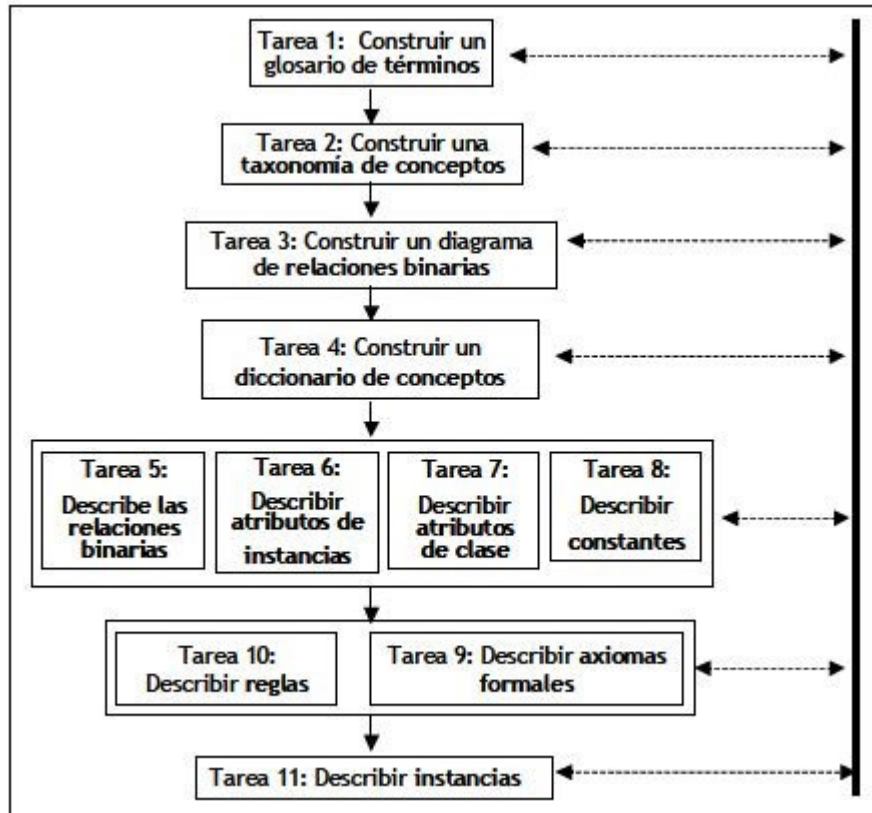


Figura 3.3 Tareas de la metodología METHONTOLOGY

Tarea 1: Construir el glosario de términos

En el glosario de términos especificado para el desarrollo de la ontología de entidades geográficas y en específico de las universidades del Distrito Federal de la delegación Gustavo A. Madero se encuentran todos los términos relevantes del dominio (conceptos, instancias, atributos, relaciones entre conceptos, etc.), así como una breve descripción en lenguaje natural de cada término.

En esta etapa es importante diferenciar si varios términos se refieren al mismo componente, en tal caso, el desarrollador debe tener la capacidad de identificar si se trata de sinónimos (ver Tabla 3.1).

3. Metodología

Tabla 3.1 Glosario de términos

| Nombre | Sinónimos | Acrónimos | Descripción | Tipo |
|---|-----------|-----------|--|----------|
| Entidad | | | Es el concepto principal, del cual se derivan todos los demás conceptos, así podrían ser: escuelas, restaurantes, hoteles, comercios, museos, para nuestro caso de estudio son Universidades | |
| Universidad | | | Es una institución de educación superior, dividida en facultades según las especialidades de estudio que la misma pueda ofrecer. | Concepto |
| Tipo de Universidad | | | Podemos encontrar universidades públicas o privadas. | Atributo |
| Carrera | | | Hace alusión a los distintos cursos académicos que se dictan en las instituciones de educación superior y que sirven para que los estudiantes se especialicen en alguna profesión. (Las carreras se encuentran clasificadas por áreas de estudio). | Concepto |
| Arquitectura, Urbanismo, Diseño Industrial, de Interiores, Textil y Grafico | | | carreras de nivel licenciatura que preparan profesionistas con conocimientos para proyectar, diseñar, dirigir y administrar la construcción de casas habitación, edificios e instalaciones, comprende, además, las carreras relacionadas con el diseño y decoración de espacios interiores y exteriores para crear ambientes agradables y funcionales en casas habitación, | Concepto |
| Biología, Biotecnología, Ecología, Ingeniería Ambiental, Ciencias atmosféricas y Ciencias del Mar | | | carreras de nivel licenciatura que preparan profesionistas con conocimientos sobre el estudio de la vida en cualquiera de sus manifestaciones y niveles de organización, mediante el trabajo de campo y laboratorio | Concepto |
| Ciencias Agropecuarias, Forestales y Pesqueras. | | | carreras de nivel licenciatura que preparan profesionistas con conocimientos prácticos, teóricos y metodológicos sobre el comportamiento genético, los sistemas de reproducción y productividad de las diferentes especies vegetales del país, sus | Concepto |

3. Metodología

| | | | | |
|--|--|--|---|----------|
| | | | necesidades de insumos, técnicas avanzadas y equipo de materiales modernos que coadyuven a incrementar la productividad agrícola. Asimismo, se incluyen las carreras que tienen como propósito intensificar la producción de hortalizas, plantas medicinales, de ornato, de jardinería e industriales | |
| Ciencias de la Salud, Nutrición y Biomédicas | | | carreras de nivel licenciatura que preparan profesionistas con conocimientos y habilidades para atender, de manera integral, los problemas de salud enfermedad de un individuo o comunidad por medio de su diagnóstico, tratamiento, prevención y rehabilitación | Concepto |
| Ciencias Humanísticas | | | carreras de nivel licenciatura que preparan profesionistas para el estudio racional del pensamiento humano y sus formas generales, el análisis reflexivo sobre el sentido y significado del hombre, y su realidad | Concepto |
| Ciencias Químicas | | | carreras de nivel licenciatura que proporcionan conocimientos sobre la composición, estructura, propiedades, reacciones y las transformaciones de la materia y la energía que se producen al combinar diferentes sustancias, con el objetivo de aplicarlos en la transformación de los recursos naturales para la producción de innumerables productos útiles para la humanidad | Concepto |
| Ciencias Sociales, Políticas, Administración Pública, Relaciones Internacionales, Comunicación Derecho y Geografía | | | Carreras que forman profesionistas que estudian la realidad social, analizan y explican el origen, el desarrollo y los cambios de los fenómenos sociales, en función de la interacción de individuos, grupos, clases, organizaciones y estructuras que dan lugar a las instituciones que integran la sociedad, determinando los efectos de tales asociaciones. | Concepto |
| Disciplinas Artísticas | | | carreras de nivel licenciatura que preparan profesionistas con los conocimientos técnicos para expresar las ideas o los | Concepto |

3. Metodología

| | | | | |
|--|--|--|--|----------|
| | | | sentimientos mediante la creación de obras artísticas como la pintura, dibujo, escultura, fotografía, cerámica y grabado a partir del análisis de la realidad social y el medio que lo circunda, comprende, además, las carreras relacionadas con la creación, adaptación e interpretación de obras musicales por medio de la técnica y características melódicas de uno o varios instrumentos musicales. | |
| Economía, Administración Contaduría y Turismo | | | carreras de nivel licenciatura cuyo objetivo es analizar los aspectos relacionados con la producción, distribución y consumo de bienes y servicios requeridos por la sociedad, además de investigar las causas de la concentración del ingreso, la dependencia económica, las variables del empleo, las políticas de crédito fiscal y de comercio interior y exterior, proponiendo soluciones a los problemas que en dichos campos se presenten. | Concepto |
| Educación y Pedagogía. | | | carreras de nivel licenciatura que preparan profesionistas para desempeñar actividades de docencia en los niveles de educación básica: preescolar, primaria y secundaria | Concepto |
| Ingenierías (Civil, Extractiva, Metalúrgica, Computación, Informática, Eléctrica, Electrónica, Mecánica, Industrial, Transportes, Aeronáutica Y Topográfica) | | | carreras de nivel licenciatura que brindan los conocimientos para planear, diseñar, ejecutar, operar y dar mantenimiento a obras de ingeniería civil, además de carreras que forman expertos en técnicas de extracción y explotación de recursos naturales así como profesionistas capaces de diseñar e implementar soluciones basadas en sistemas de información, entre muchas otras actividades. | Concepto |
| Matemáticas, Física Y Astronomía | | | Carreras de nivel licenciatura que proporcionan conocimientos de matemáticas, aplicables a la investigación, diseño, descripción y realización de modelos y contenidos matemáticos, así como carreras orientadas a las propiedades físicas de la materia y al estudio de los cuerpos | Concepto |

3. Metodología

| | | | | |
|-------------|--|--|---|--|
| | | | celestes. | |
| [Carreras] | | | Es un conjunto de las carreras que se encuentran distribuidas en cada una de las áreas que se describieron anteriormente. | |

Tarea 2: Construir la taxonomía de conceptos

Una vez que se identifican los términos en el glosario, se describe la taxonomía de los mismos, la cual define la jerarquía entre conceptos, por lo que sólo se hace entre los términos del glosario que son del tipo concepto. METHONTOLOGY propone que se utilicen las relaciones taxonómicas de *Subclase-de*, *Descomposición-Disjunta*, *Descomposición-Exhaustiva*, y *Partición*, así como relaciones *ad hoc*. En la Figura 3.4 se muestra la taxonomía de conceptos diseñada.

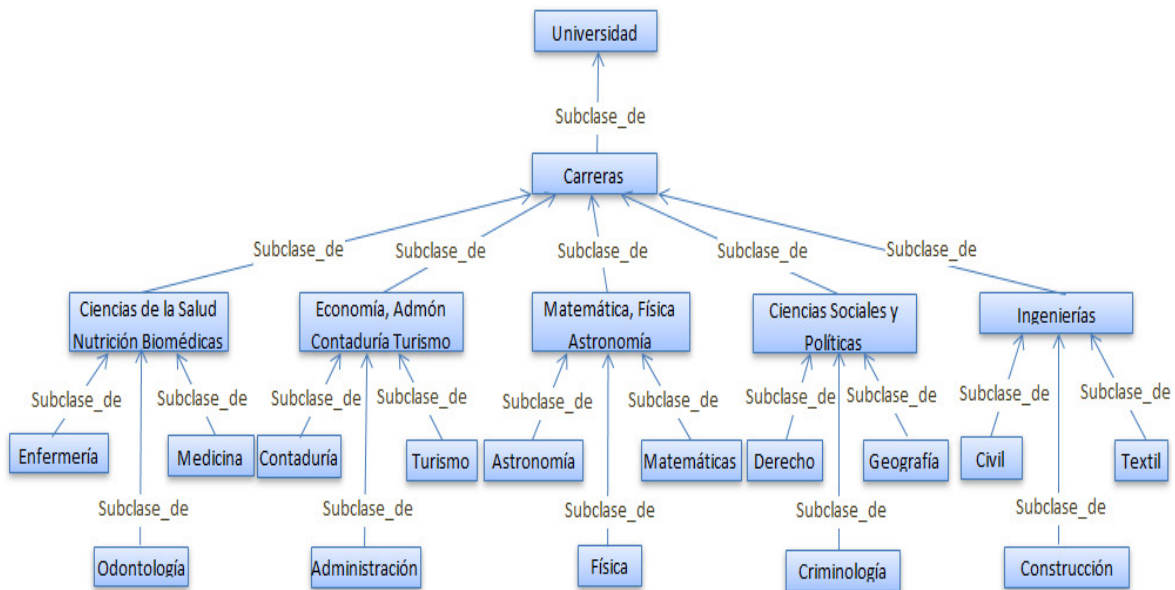


Figura 3.4 Fragmento de la taxonomía de conceptos

Tarea 3: Construir diagrama de relaciones binarias

Una vez establecida la taxonomía de conceptos, se propone establecer un diagrama de relaciones binarias, el objetivo de este diagrama es establecer las relaciones entre cada uno de los conceptos de la taxonomía. A continuación se presenta un diagrama de las principales relaciones entre los conceptos (ver Figura

3.5), en el cual se observan las relaciones “*compuesto de*” y su relación inversa “*es un*”, “*es parte de*” y su relación inversa “*tiene*”.

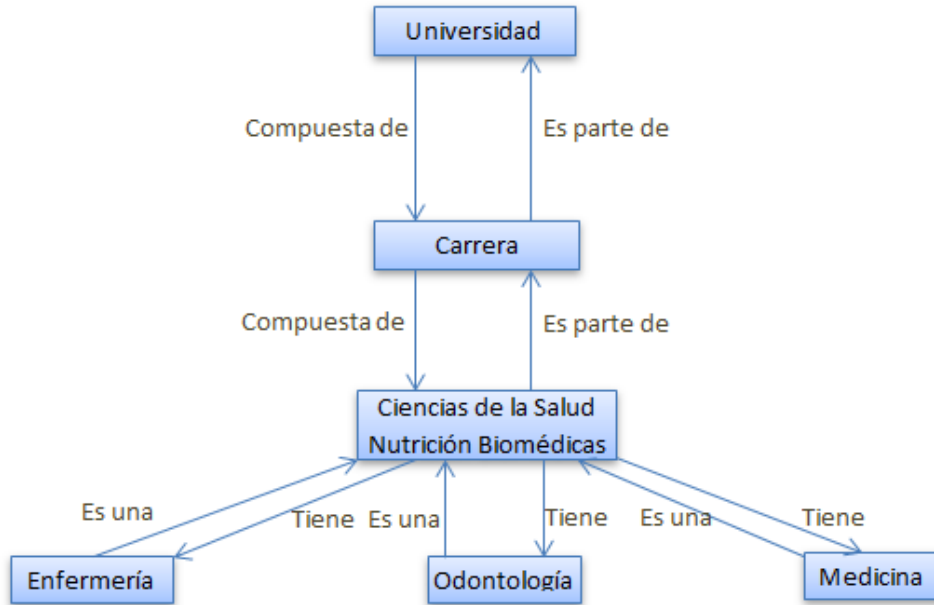


Figura 3.5 Fragmento del diagrama de relaciones binarias

Tarea 4: Construir el diccionario de conceptos

En el diccionario de conceptos se especifican las propiedades que describen a cada concepto de la taxonomía, además de describir las relaciones identificadas en el diagrama de relaciones binarias y las instancias de cada concepto. Las relaciones especificadas para cada concepto son aquellas en las que el concepto es el origen de la misma (ver Tabla 3.2).

Tabla 3.2 Diccionario de conceptos

| Nombre del concepto | Instancias | Atributos de Clase | Atributos de la Instancia | Relaciones |
|----------------------|------------|--------------------|--|------------|
| Ingeniería Eléctrica | ESIME | | Nombre de la institución Tipo de Institución Costo de la colegiatura Ubicación Dirección | Tiene |

3. Metodología

| | | | | |
|---------------|--------------------------|--|--|-------|
| Psicología | Universidad del Tepeyac | | Nombre de la institución Tipo de Institución Costo de la colegiatura Ubicación Dirección | Tiene |
| Mercadotecnia | Universidad Justo Sierra | | Nombre de la institución Tipo de Institución Costo de la colegiatura Ubicación Dirección | Tiene |
| Derecho | Colegio AMAUTA | | Nombre de la institución Tipo de Institución Costo de la colegiatura Ubicación Dirección | Tiene |

Tarea 5: Describir las relaciones binarias *ad hoc*

En este apartado se muestran las descripciones en detalle de todas las relaciones binarias identificadas en el diagrama de relaciones binarias e incluidas en el diccionario de conceptos. Además, se especifican el nombre de la relación, así como los nombres de sus conceptos origen y destino, su cardinalidad y su relación inversa, si existe (ver Tabla 3.3).

Tabla 3.3 Relaciones binarias *ad hoc*

| Nombre de la relación | Concepto origen | Cardinalidad Máxima | Concepto destino | Relación Inversa |
|-----------------------|-----------------|---------------------|-------------------|------------------|
| Compuesta de | Universidad | N | Carrera | Es parte de |
| Es una | Carrera | N | Área Conocimiento | Compuesta de |
| Tiene | [Instancia] | N | [Carrera] | Está en |

Tarea 6: Describir los atributos de instancia

En tarea se describen algunos de los atributos de instancia incluidos en el diccionario de conceptos. Cada fila de atributos de instancia contiene la

3. Metodología

descripción detallada de un atributo de instancia, se debe especificar su nombre, el concepto al que pertenece el atributo, su tipo de valor, su rango de valores y su cardinalidad mínima y máxima (ver Tabla 3.4).

Tabla 3.4 Atributos de instancia

| Nombre de Atributo | Concepto | Tipo de Valor | Rango de valores | Cardinalidad |
|--------------------|---------------|---------------|-------------------|--------------|
| Nombre | [Institución] | String | ----- | (1,1) |
| Tipo | [Institución] | String | Pública o Privada | (1,1) |
| Costo | [Institución] | Double | ----- | (1,1) |
| Ubicación | [Institución] | String | ----- | (1,1) |
| Dirección | [Institución] | String | | (1,1) |

Tarea 7: Describir los atributos de clase

En esta tabla se describen en detalle los atributos de clase incluidos en el diccionario de conceptos. Para cada atributo de clase se debe definir la siguiente información; el nombre del atributo, el nombre del concepto donde se define el atributo, su tipo de valor, cardinalidad y valor o valores (ver Tabla 3.5).

Tabla 3.5 Atributos de Clase

| Nombre de Atributo | Concepto | Tipo de Valor | Cardinalidad | Valores |
|--------------------|---------------|---------------|--------------|-------------------------|
| Nombre | [Institución] | String | 1:1 | [Nombre Institución] |
| Tipo | [Institución] | String | 1:1 | Pública o Privada |
| Costo | [Institución] | Double | 1:1 | [Costo] |
| Ubicación | [Institución] | String | 1:1 | [Coordenadas] |
| Dirección | [Institución] | String | 1:1 | [Dirección Institución] |

Tarea 8: Describir las constantes

En esta tarea se describe en detalle un fragmento de cada una de las constantes identificadas en el glosario de términos. Para cada constante, se especifica su nombre, tipo de valor y unidad de medida, en el caso de constantes numéricas. En

esta ontología no se desarrolla esta tarea, debido a que no se manejan constantes. Una vez que se han desarrollado las tareas anteriores, se procede a describir los axiomas formales y las reglas en paralelo.

Tarea 9: Definir axiomas formales

En este apartado se describen los axiomas formales que son necesarios en la ontología y describirlos de manera precisa. Para cada definición de axioma formal, se propone especificar: nombre, descripción en lenguaje natural, expresión lógica que define de manera formal el axioma utilizando lógica de primer orden, y los conceptos, atributos y relaciones utilizadas en el axioma, así como las variables utilizadas (ver Figura 3.6).

Tabla 3.6 Axiomas formales

| Nombre Axioma | Descripción | Expresión | Concepto | Relaciones | Variables |
|--------------------------|---|--|----------|----------------------|-----------|
| Tiene carreras | Una institución debe tener al menos una carrera | Existe (?X,?Y) (carrera(?X) e Institución(?Y) y tiene(?Y,?X) | 1:1 | [Nombre Institución] | |
| Carrera en una sola área | Cada carrera debe estar únicamente en una área | Existe (?X, ?Y) | 1:1 | Pública o Privada | |
| Costo | [Institución] | Double | 1:1 | [Costo] | |

Tarea 10: Definición de reglas

De la misma manera que en la definición de axiomas formales, en esta tarea se deben identificar en primer lugar que reglas se necesitan en la ontología y posteriormente describirlas en la tabla de reglas. Para cada regla se propone incluir la siguiente información: nombre, descripción en lenguaje natural, expresión que describe formalmente la regla, y conceptos, atributos y relaciones utilizados en la regla, así como las variables utilizadas.

3. Metodología

Por tanto, se propone especificar las expresiones de las reglas, utilizando el formato si <condiciones> entonces <consecuente>. La parte izquierda de la regla es una conjunción de condiciones simples, mientras que la parte derecha es una simple expresión de un valor de la ontología (ver Tabla 3.7).

Tabla 3.7 Reglas

| Nombre de Atributo | Concepto | Tipo de Valor | Cardinalidad | Valores |
|--------------------|---------------|---------------|--------------|----------------------|
| Nombre | [Institución] | String | 1:1 | [Nombre Institución] |
| Tipo | [Institución] | String | 1:1 | Pública o Privada |
| Costo | [Institución] | Double | 1:1 | [Costo] |

Tarea 11. Descripción de instancias

Una vez que el modelo conceptual de la ontología ha sido generado, se pueden definir las instancias que aparecen en el diccionario de conceptos. Para cada instancia se define: su nombre, el nombre del concepto al que pertenece y los valores de sus atributos de instancia, si se conocen, en nuestro caso, el nombre, el tipo y el costo aproximado de la colegiatura de cada institución de educación de nivel superior de la Delegación Gustavo A. Madero en el Distrito Federal.

Una vez teniendo el diseño de la ontología, se procede a la codificación con la herramienta Protégé, el cual es un editor de ontologías *open source*, además permite la codificación de la ontología en los lenguajes más utilizados para el desarrollo de las mismas, como son RDF (Resource Description Framework) y OWL (Ontology Web Language).

RDF es un modelo estándar para el intercambio de datos en la Web. Se basa en tripletas que representan expresiones del tipo *sujeto-predicado-objeto* (quién hace qué con algo). OWL es el lenguaje propuesto por W3C para representar ontologías en la Web semántica, y se ha construido como una extensión de RDF.

El lenguaje OWL es capaz de expresar *lógica de descripciones*, donde con "lógica de descripciones" (también llamada "**Lógica Descriptiva**") se define una familia

3. Metodología

de formalismos para representar y razonar sobre clases complejas de individuos (conceptos) y sus relaciones (roles). OWL es el lenguaje que se utiliza para definir la ontología en el presente trabajo. En la Figura 3.6 se muestra la implementación de la ontología mediante la herramienta Protégé³.

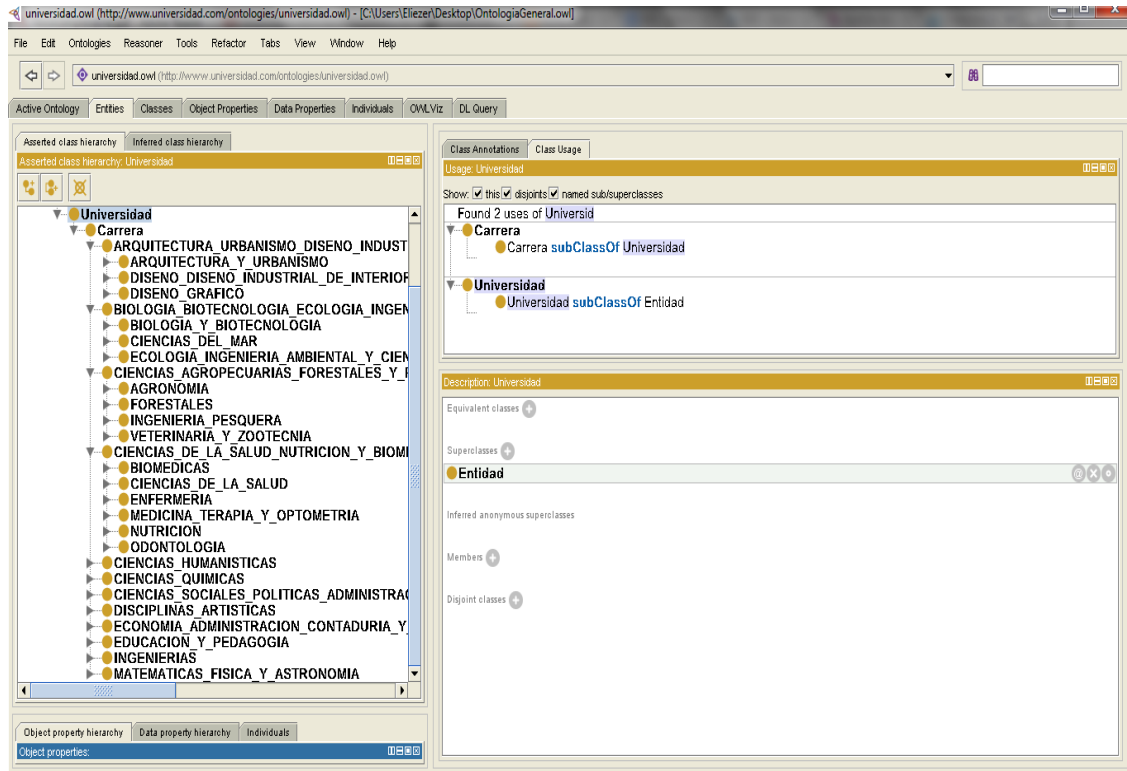


Figura 3.6 Vista de un fragmento de la ontología en Protégé

3.3.5 Poblado de la ontología

Una vez terminada la implementación del diseño de la ontología con el editor Protégé, se procede a cargar todas las instancias con el mismo editor. Cada instancia pertenece a una o más clases en la jerarquía de la ontología; además de que cada una tiene los atributos de costo, tipo, dirección y ubicación. En la Figura 3.7 se muestra un fragmento de las entidades con que se ha poblado a la ontología, mediante la herramienta Protégé.

³ Protégé es un editor de ontologías de código abierto y un marco de trabajo basado en conocimientos.

3. Metodología

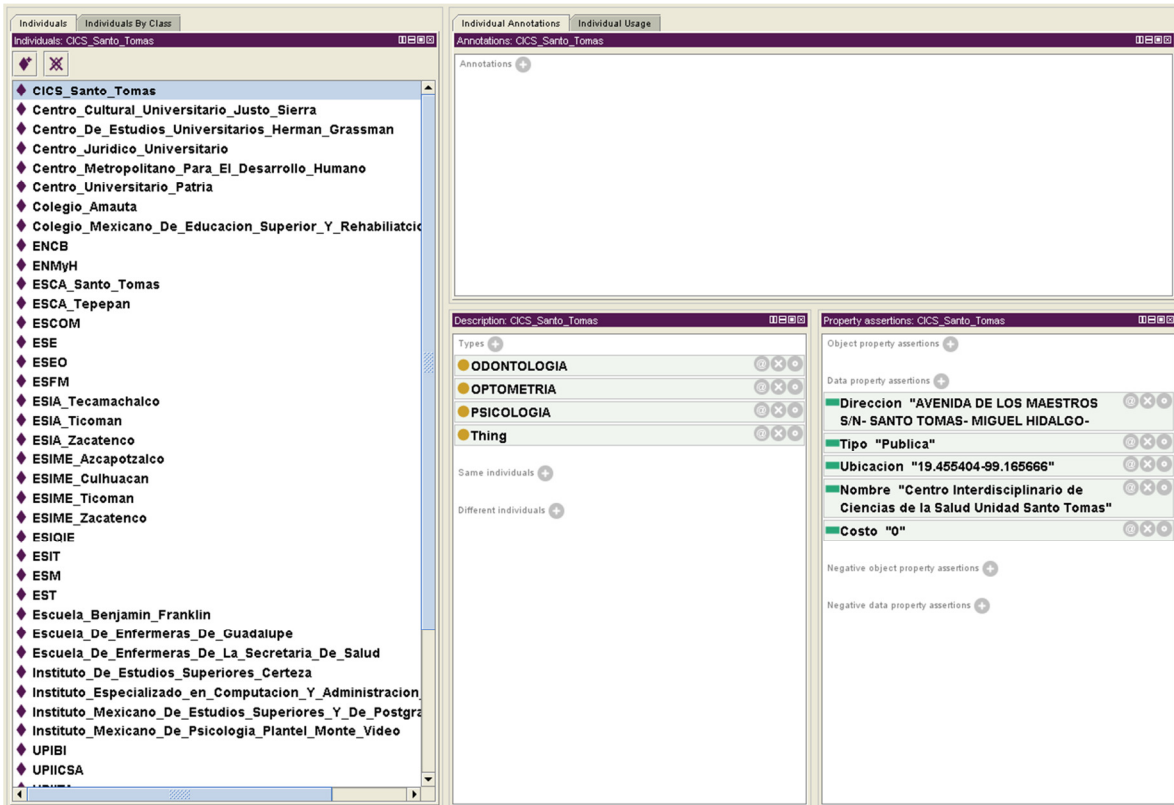


Figura 3.7 Segmento de individuos de la ontología de universidades del D.F.

En la figura 3.8 se muestra un ejemplo de una instancia de la ontología con sus propiedades correspondientes y los elementos relacionados.

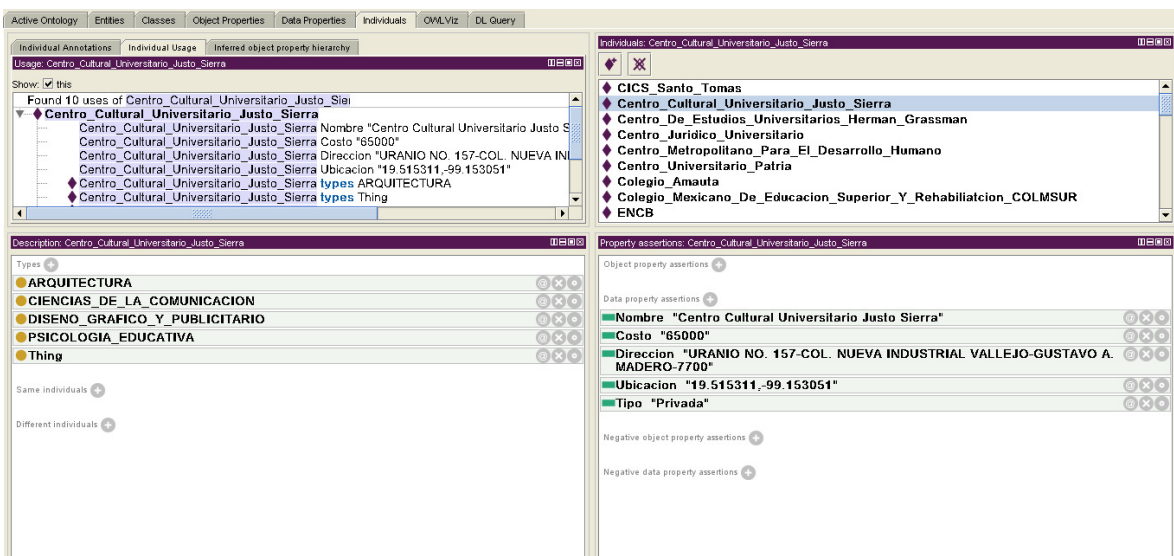


Figura 3.8 Una instancia de la ontología con sus propiedades y elementos relacionados

3.3.6 Modelo persistente de datos

Un modelo persistente de datos nos permite disponer de un espacio de almacenamiento en memoria, lo cual da lugar a tener un objeto desde el punto de creación hasta el final del programa o hasta que el programador lo destruya, la persistencia permite al programador almacenar, transferir y recuperar el estado de los objetos (Atkinson, Daynes, Jordan, Printezis, & Spence, 1996).

Una vez que hemos creado nuestra ontología es necesario almacenarla de tal manera que podamos generar automáticamente una representación conceptual explícita de las relaciones, de acuerdo con [Heymans, et al, 2007]; existen dos clases de almacenamiento persistente: nativo y en bases de datos. La primera se construye directamente en el sistema de archivos, organizándose bien como archivos que almacenan sentencias directamente e índices para su acceso normalmente mediante tripletas, o bien jerárquicamente empleando aserciones de conceptos. La segunda clase organiza la información en diferentes tablas utilizando almacenamiento genérico. En la Figura 3.9 podemos observar la clasificación descrita anteriormente del modelo persistente de datos.

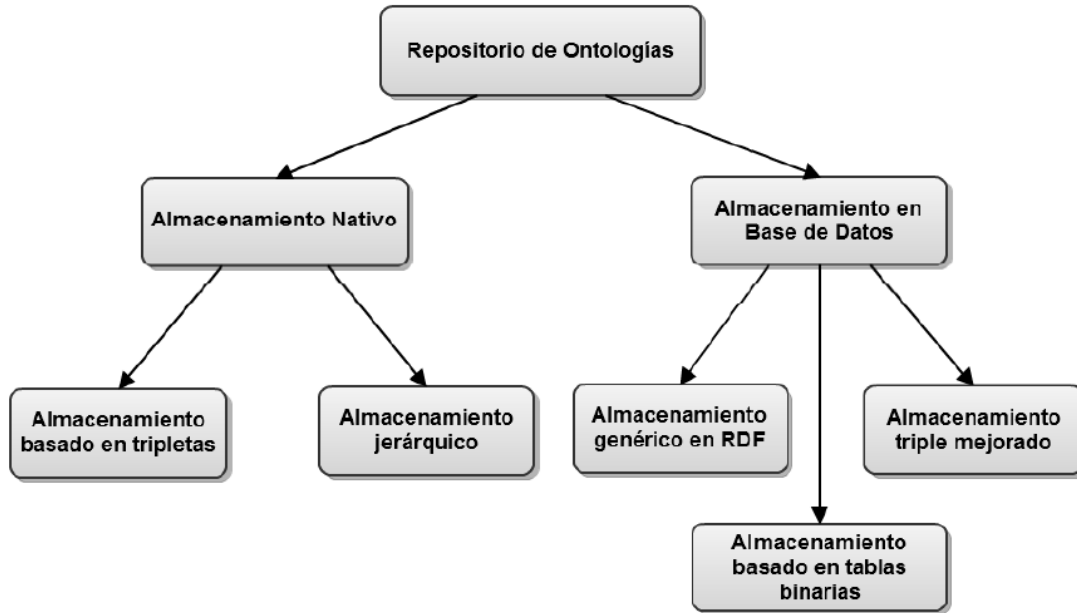


Figura 3.9 Esquema general de almacenamiento persistente

Pero la forma en la que se guardan los datos no resume toda la filosofía del almacenamiento: también el modo en el que la información es procesada. En este trabajo nos auxiliamos del API de Jena⁴ para Java que nos permite acceder al modelo persistente para obtener los recursos OWL transformados en el modelo tabular, el cual nos permitirá procesar la información contenida en él, tanto para su almacenamiento, como para su consulta y manipulación. A continuación se describen los pasos para generar el modelo persistente con Jena:

1. Se crea un modelo vacío para definir los recursos que serán almacenados, tal y como se muestra en las siguientes sintaxis:

```
OntModel model= null;
```

2. Este modelo que se creó es una extensión del OWL, el cual proporciona opciones adicionales para manejar los datos provenientes de la ontología.

```
this.model=ModelFactory.createOntologyModel (OntModelSpec.OWL_MEM_RULE_INF);
```

⁴ Jena es un marco de trabajo Java para crear aplicaciones de Web Semántica. Jena proporciona un conjunto de herramientas y bibliotecas de Java para ayudar a desarrollar la web semántica y aplicaciones enlace de datos, herramientas y servidores

3. Ahora al modelo creado se le debe de indicar la ontología con la cual será poblado, por tanto:

```
Java.io.InputStream in = FileManager.get ().open("C:/Users/Eliezer/Desktop/Universidades.owl");  
this.model.read (in, "");
```

Al realizar este último paso mediante el lenguaje de consultas SPARQL, podemos efectuar consultas de las instancias de la ontología sobre el modelo generado.

Por ejemplo, la consulta para obtener todas las instituciones de educación de nivel superior que en contengan "Inst" en el nombre del establecimiento sería de la siguiente forma:

```
PREFIX prefijo: <http://www.owl-ontologies.com/Universidades.owl#>  
  
SELECT ?Entidad ?nombre  
  
WHERE { ?unidad prefijo: nombre ?nombre.FILTER regex  
  
        (?nombre,\"inst\", \"i\") }
```

Entonces para ejecutar esta consulta empleando el modelo generado se deben seguir los siguientes pasos:

1. Para crear la consulta es necesario crear un *queryString*:

```
Query query = QueryExcecution.create (queryString);
```

2. Empleando el objeto *query* y el modelo se crea una *QueryExecution*:

```
QueryExecution qe = QueryExecutionFactory.create (query, this.model);
```

3. De tal forma de que con *QueryExecution* se puede ejecutar la consulta y obtener el resultado para almacenarlo en un *ResultSet*:

```
ResultSet results = qe.execSelect ();
```

Por tanto, el proceso de modelos persistente nos permite hacer la consulta de la ontología cargada en memoria, mediante una estructura basada en SPARQL, en

donde esta consulta tiene acceso directamente a la ontología de aplicación implementada en lenguaje OWL, utilizando los métodos y librerías de Jena para manipular el modelo persistente generado.

3.4 Etapa de Recuperación

La recuperación de información se realiza con base en una consulta dada. En el caso de una consulta general del usuario, ésta se transforma en una consulta semántica, la cual utiliza medidas de similitud semántica. En el caso concreto de esta aplicación, se utiliza la distancia semántica entre conceptos en la modalidad de distancia de trayectoria de un concepto a otro, para hacer la recuperación de información desde el modelo persistente de datos.

El resultado de estas consultas se presenta mediante un servicio Web, el cual debe estar disponible con el fin de que cualquier aplicación cliente pueda hacer la recuperación de entidades geográficas mediante él. En la Figura 3.10 se muestra el proceso de recuperación llevado a cabo.

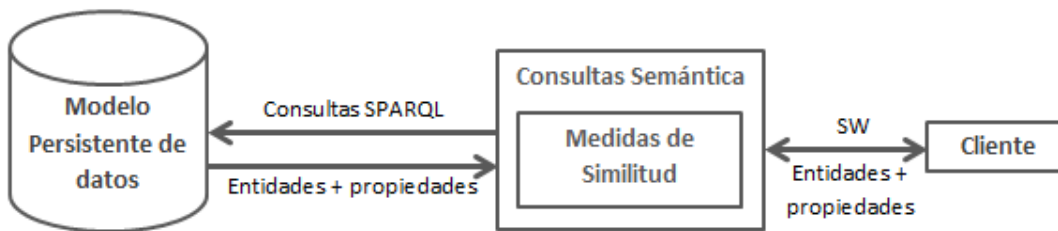


Figura 3.10 Tareas de la etapa de Recuperación

3.4.1 Cliente

Comenzamos por establecer la función del cliente, debido a que a partir de él se define la recuperación de información que se requiere, el cliente se refiere a la aplicación que se comunicará con el sistema de recuperación de información mediante un servicio web, dicha comunicación se lleva a cabo introduciendo un requerimiento de información o consulta desde el cliente, posteriormente esta consulta será interpretada por el servicio web, para posteriormente pasarla al sistema de recuperación, mediante un mensaje en formato SOAP. Una vez

establecida la consulta por el cliente e interpretada por el servicio web, pasamos a la tarea de realizar la consulta semántica (que se explica en la siguiente sección) la cual devuelve al cliente como resultado un conjunto de entidades geográficas y sus propiedades.

3.4.2 Consulta Semántica

Una vez que tenemos nuestro modelo persistente de datos, podemos hacer las consultas semánticas, para esto ya se ha recibido una consulta desde la aplicación cliente.

La consulta semántica implica identificar en nuestro repositorio de conocimiento si existen entidades con las características que se especifican en la consulta o identificar aquellas que más se acerquen a la descripción especificada por el cliente, para realizar esta tarea es necesario acceder a nuestro modelo persistente de datos mediante consultas con el lenguaje SPARQL para ir identificando los resultados que más se adecuen al resultado esperado. Esto lo podemos lograr aplicando alguna medida de similitud al elemento de consulta y cada uno de los elementos del repositorio de conocimiento para lo cual el lenguaje SPARQL es nuestro principal medio de comunicación entre estos elementos. En el presente trabajo de investigación ésta consulta semántica se hizo aplicando la medida de similitud conocida como distancia semántica entre conceptos, la cual se explica a continuación.

3.4.2.1 Medida de Similitud

En una consulta de información por palabras clave en caso de no encontrar coincidencias exactas tiende a devolver resultado vacíos, en cambio una consulta semántica busca la manera de tener los resultados que más se adecuen a lo que el usuario desea, de esta manera evitamos tener resultados vacíos y poder ofrecer a los usuarios soluciones alternativas.

La manera en que se recuperan los resultados que más se adecuen a la consulta del usuario, es identificando si existen coincidencias exactas del concepto o

3. Metodología

instancia buscados dentro de nuestra base de conocimiento, en ese caso el resultado es exacto, en caso contrario, se busca en los conceptos más cercanos al concepto buscado, así, lo que se recupera son las instancias de los conceptos con las distancias más cortas al concepto buscado.

Para hacer el cálculo de la distancia semántica entre conceptos en el desarrollo de éste trabajo se utilizó la medida de trayectoria entre conceptos, cuyo enfoque fue propuesto por Centeno et al. (2007), en donde un concepto que es subconcepto de otro pueden tener distancia semántica igual a cero, entonces se dice que tienen coincidencia exacta, ésta distancia va creciendo cuando descendemos entre los subconceptos de un concepto, o igualmente con la distancia entre conceptos que no son subconceptos entre sí.

La manera en que se implementó esta medida de similitud fue aplicando un algoritmo recursivo, en donde se hace la búsqueda de las instancias de un concepto dado, en caso de no encontrar instancias, se hace el recorrido primero en profundidad de los subconceptos del concepto buscado. En caso de no encontrar ninguna instancia del concepto ni de sus subconceptos, continuamos con la búsqueda con el concepto padre y todos sus subconceptos y así hasta recorrer toda la ontología de ser necesario.

La única manera de obtener un resultado vacío es que la ontología no tenga ninguna instancia.

La recuperación de información está implementada mediante un algoritmo recursivo, el cual hace el recorrido de la ontología hasta encontrar nodos hoja de la ontología o encontrar las instancias que están a la misma distancia de un nodo hoja.

Pseudocódigo del algoritmo utilizado:

1. Así, siguiendo la idea de Centeno et al. (2007) primero verificamos si la coincidencia es exacta y recuperamos las instancias del concepto que pasamos como parámetro de la búsqueda.

3. Metodología

2. Si no se encuentran coincidencias exactas, pasamos a verificar el siguiente nivel, los subconceptos del concepto que inició la búsqueda, para lo cual, listamos todos sus subconceptos.
 - a. Una vez listados los subconceptos, aplicamos la misma función de recuperación para cada uno de ellos.
 - b. Recuperamos las instancias de cada uno de los subconceptos listados en caso de que las tengan, en caso contrario, pasamos al siguiente nivel de subconceptos
 - c. Si llegamos a los nodos hoja, y no tenemos resultados, ahora buscamos el nodo padre del concepto que inició la búsqueda y ahora recorreremos a todos sus subconceptos en búsqueda de información.
3. El recorrido de la ontología termina cuando encontramos el primer grupo de instancias o instancia que satisfaga nuestra consulta semántica.

Los algoritmos 1 y 2 muestran cómo se desarrolló la noción del algoritmo propuesto por Centeno et al. (2007).

Algoritmo 1 Trayectoria entre conceptos

Entrada: Un concepto, para conocer sus instancias.

Salida: La lista de instancias del concepto buscado, en caso de que las tenga, o una lista de las instancias relacionadas.

Si Instancias de concepto (Concepto) != null entonces

return Instancias de concepto (Concepto)

Si no

Si Listar subClases (Concepto) != null entonces

Para cada (subClase) hacer

Trayectoria entre conceptos (subClase)

Fin

Si no

Trayectoria entre conceptos (Clase_Padre)

Fin

Fin

Algoritmo 2 Instancias de concepto

Entrada: Un concepto, para conocer sus instancias.

Salida: La lista de instancias del concepto buscado, en caso de que las tenga, o una lista vacía en caso contrario.

Si concepto tiene instancias **entonces**

return instancias

Si no

return null

Fin

El lenguaje de reglas de la web semántica SWRL nos permite escribir reglas que se pueden expresar en términos de conceptos OWL para proporcionar más capacidades de razonamiento deductivo que OWL solo. Semánticamente, SWRL está construido sobre la misma descripción lógica OWL y es de gran ayuda al momento de hacer inferencias en el modelo persistente.

El sistema de inferencia Jena está diseñado para permitir una amplia gama de motores de inferencia o razonadores que se conectan a Jena. Estos motores se utilizan para inferir afirmaciones adicionales que se implicaron de alguna base RDF junto con cualquier información ontología opcional, los axiomas y las reglas asociadas con el razonador. El principal uso de este mecanismo es apoyar el uso de los lenguajes como RDFS y OWL que permiten deducir nuevos hechos de los datos de instancia y las descripciones de clase. Jena incluye un motor de reglas genérico que se puede utilizar para varias tareas de procesamiento o transformación RDF.

En ésta aplicación se empleó el *Transitive reasoner* que viene incluido en Jena, el cual permite almacenar y recorrer las clases y sus respectivas propiedades. Es así como podemos hacer el recorrido de los conceptos dentro de la ontología viendo

las relaciones de *subClass* o *superClass* de una clase con otra, con los métodos que ofrece Jena para tal fin.

El transitive reasoner de Jena fue de utilidad al momento de implementar el algoritmo de trayectoria entre conceptos, para ubicar a los conceptos cercanos del concepto buscado y extraer sus instancias correspondientes, ya que mediante el método *listSubclases* permite el acceso a cada una de las subclases del concepto buscado, así como el método *getSuperClass* nos da accesos a la clase padre del mismo y poder recuperas sus instancias mediante el método *listInstances*.

Algoritmo 3 Trayectoria entre conceptos (implementando el swrl de Jena)

Entrada: Un concepto, para conocer sus instancias.

Salida: La lista de instancias del concepto buscado, en caso de que las tenga, o una lista de las instancias relacionadas.

Si Instancias de concepto (Concepto) != null entonces

return Instancias de concepto (Concepto.listInstances())

Si no

Si Listar subClases (Concepto.listSubClasses()) != null entonces

Para cada (subclase)**hacer**

Trayectoria entre conceptos (Concepto.listSubClasses().next)

Fin

Si no

Trayectoria entre conceptos (Concepto.getSuperClass())

Fin

Fin

3.4.3 Consultas SPARQL

SPARQL es un lenguaje diseñado para consultar y modificar ontologías, se trata de un lenguaje con bastante similitud a SQL, la diferencia principal es que SQL supone que los datos están implementados en tablas y SPARQL supone que los datos están implementados en grafos. Jena suministra una implementación de

3. Metodología

SPARQL. SPARQL es un lenguaje de consulta y un protocolo para el acceso RDF diseñado por el Grupo de trabajo W3C.

Mediante SPARQL accedemos a nuestro modelo persistente de datos y recuperamos la información específica que necesitamos en la consulta mediante los métodos que propone para facilitar el acceso a cada uno de los elementos de la ontología, para posteriormente evaluar estos resultados mediante las medidas de similitud de distancia semántica.

Como un lenguaje de consulta, SPARQL es "orientado a datos", ya que sólo consulta la información contenida en los modelos, no hay inferencia en el propio lenguaje de consulta. El modelo de Jena puede ser "inteligente", ya que proporciona la impresión de que existen ciertas tripletas pero las crea bajo demanda, a través de algunas técnicas de razonamiento OWL. SPARQL no hace otra cosa que tomar la descripción de lo que quiere la aplicación, en la forma de una consulta, y devuelve esa información, en forma de un conjunto de enlaces o un grafo RDF.

Primeramente, se crea el modelo persistente de datos, como se indica anteriormente en el apartado Modelo persistente de datos.

```
String queryString =  
    "PREFIX onto: <http://www.universidad.com/ontologies/universidad.owl#> "  
    + "PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#> "  
    + "PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> "  
    + "SELECT * "  
    + "FROM <http://www.universidad.com/ontologies/universidad.owl> " "  
    + "WHERE { $instancia rdf:type onto: PSICOLOGIA} ";
```


3. Metodología

En las primeras líneas encontramos la palabra clave *PREFIX*. Que asocia una URI a una etiqueta, que se usará en adelante para describir el espacio de nombres (*namespace*⁵). Pueden incluirse varias de estas etiquetas en una misma consulta.

El inicio de la consulta queda marcado por la palabra clave *SELECT*. Semejante a su uso en SQL, sirve para definir los datos que deben ser devueltos en la respuesta. En el ejemplo se devolverán todos los campos del elemento.

La palabra clave *FROM* identifica los datos sobre los que se ejecutará la consulta. En este ejemplo, se limitará al RDF de Universidad. Una consulta puede incluir varios *FROM*.

La palabra clave *WHERE* indica el patrón sobre el que se filtrarán las tripletas del RDF, en este ejemplo se filtran las entidades del concepto **PSICOLOGIA**.

```
Query query = QueryFactory.create(queryString);  
  
QueryExecution qe = QueryExecutionFactory.create(query, modelo);  
  
com.hp.hpl.jena.query.ResultSet results = qe.execSelect();
```

Con estas líneas ejecutamos la consulta, la cual para este ejemplo, arroja todas las entidades del concepto PSICOLOGIA, tal como se muestra en la Figura 3.11.

| Nombre | Costo_Anual | Ubicacion |
|--|-------------|-----------------------|
| "Universidad Del Tepeyac" | "10200" | "19.491055-99.124306" |
| "Universidad Insurgentes Norte" | "50000" | "19.473099-99.131931" |
| "Universidad Justo Sierra" | "65000" | "19.512457-99.124379" |
| "Centro Interdisciplinario de Ciencias de la Salud Unidad Santo Tomas" | "0" | "19.455404-99.165666" |

Figura 3.11 Resultado del ejemplo de consulta SPARQL

3.4.4 Servicio WEB

Los servicios Web son entidades computacionales que ofrecen alguna funcionalidad que tiene un valor asociado en un dominio dado. Un modelo de

⁵ Namespace es un conjunto de nombres en el cual todos los nombres son únicos.

3. Metodología

servicio Web es una descripción formal de la funcionalidad de los servicios Web, en términos de capacidad, y los métodos para interactuar con él, en términos de una interfaz.

Este servicio web permite que la recuperación de información esté disponible en la web para cualquier aplicación que desee hacer la recuperación semántica de información, basta con conocer la ubicación del servicio en la web para poder establecer la conexión y utilizar los métodos de recuperación que presenta el mismo servicio web, para comprender un poco más el funcionamiento de un servicio web podemos referirnos a la Figura 3.12.

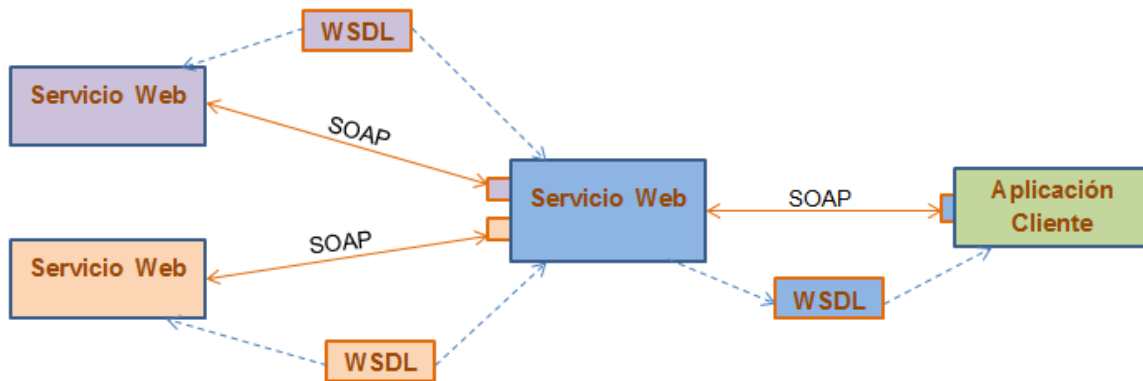


Figura 3.12 Esquema básico de un servicio Web

El esquema básico de un servicio web consiste en que un usuario (que juega el papel de cliente dentro de los Servicios Web), a través de una aplicación, solicita información a un servicio web, en nuestro caso al de recuperación semántica de información geográfica. Este servicio de recuperación ofrece a su cliente (usuario) la información que más se asemeje a la que el usuario solicita. Para proporcionar al cliente la información que necesita, el servicio puede hacer la conexión a otros servicios Web, en nuestro caso, el servicio sólo consulta a la ontología que se desarrolló al comienzo de la metodología, la cual se encuentra almacenada en un servidor que la hace disponible para las consultas. Por último, el usuario visualiza mediante su aplicación la información que el servicio le proporcione como resultado de su consulta.

3. Metodología

En todo este proceso intervienen una serie de tecnologías que hacen posible esta circulación de información. Por un lado, está SOAP (Protocolo Simple de Acceso a Objetos). Se trata de un protocolo basado en XML que permite la interacción entre varios dispositivos y que tiene la capacidad de transmitir información compleja. Los datos pueden ser transmitidos a través de HTTP, SMTP, etc. SOAP especifica el formato de los mensajes. El mensaje SOAP está compuesto por un envelope (sobre), cuya estructura está formada por los siguientes elementos: header (cabecera) y body (cuerpo), tal como se muestra en la Figura 3.13.

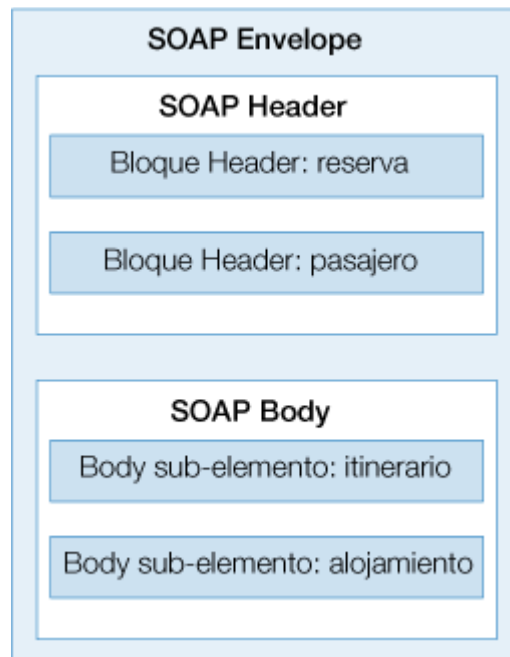


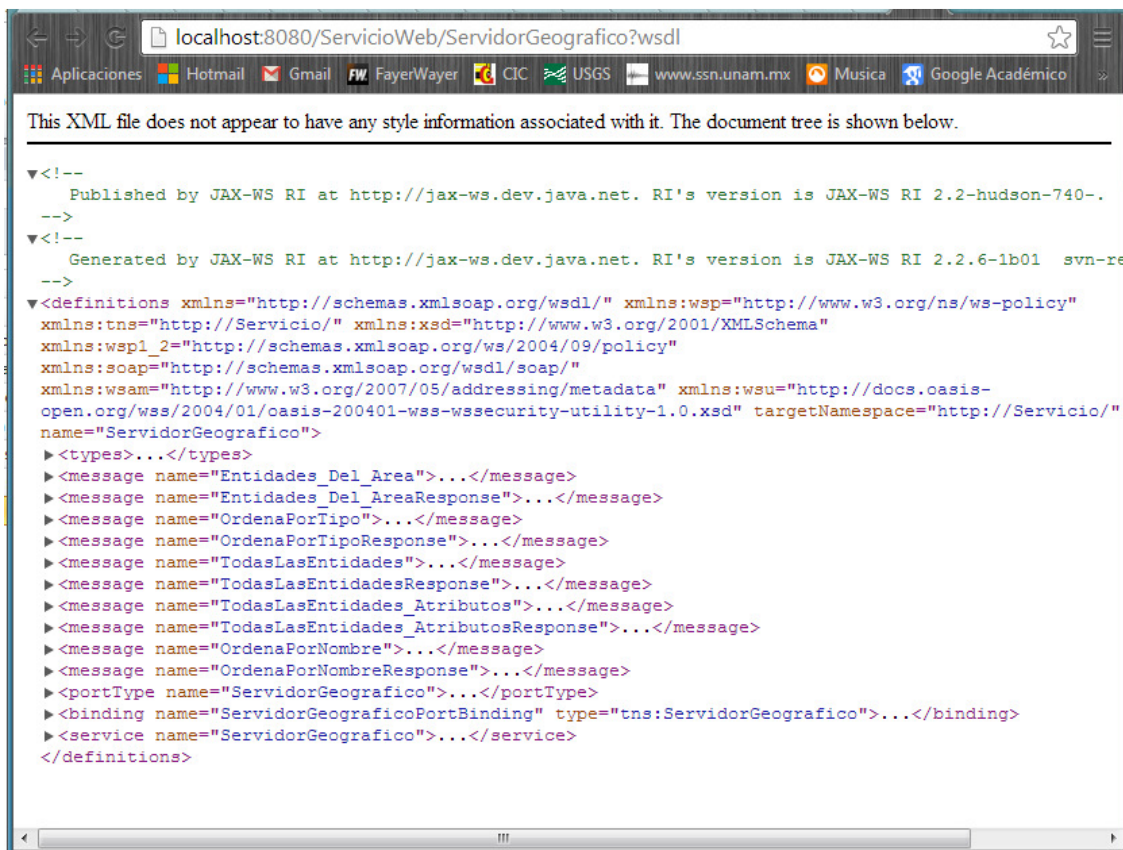
Figura 3.13 Estructura general de SOAP

Por otro lado, WSDL (Lenguaje de Descripción de Servicios Web), permite que un servicio y un cliente establezcan un acuerdo en lo que se refiere a los detalles de transporte de mensajes y su contenido, a través de un documento procesable por dispositivos. WSDL representa una especie de contrato entre el proveedor y el que solicita. WSDL especifica la sintaxis y los mecanismos de intercambio de mensaje.

3. Metodología

En cuanto a la aplicación cliente dependiendo del usuario y sus necesidades cada usuario tendrá su aplicación cliente que se adecúe a sus necesidades y además pueda interactuar con el servicio.

Esta descripción la podemos ver en la Figura 3.14 en la cual se muestra la pantalla principal de una conexión a la descripción del servicio web, dicha descripción se utiliza para facilitar el uso de cada uno de los métodos ya que permite identificar los elementos como el nombre, parámetros de entrada y tipo de retorno de cada método.



The image shows a web browser window with the address bar containing 'localhost:8080/ServicioWeb/ServidorGeografico?wsdl'. The main content area displays the following XML code:

```
This XML file does not appear to have any style information associated with it. The document tree is shown below.
<!--
  Published by JAX-WS RI at http://jax-ws.dev.java.net. RI's version is JAX-WS RI 2.2-hudson-740-.
-->
<!--
  Generated by JAX-WS RI at http://jax-ws.dev.java.net. RI's version is JAX-WS RI 2.2.6-1b01 svn-re
-->
<definitions xmlns="http://schemas.xmlsoap.org/wsdl/" xmlns:wsp="http://www.w3.org/ns/ws-policy"
  xmlns:tns="http://Servicio/" xmlns:xsd="http://www.w3.org/2001/XMLSchema"
  xmlns:wsp1_2="http://schemas.xmlsoap.org/ws/2004/09/policy"
  xmlns:soap="http://schemas.xmlsoap.org/wsdl/soap/"
  xmlns:wsam="http://www.w3.org/2007/05/addressing/metadata" xmlns:wsu="http://docs.oasis-
  open.org/wss/2004/01/oasis-200401-wss-wssecurity-utility-1.0.xsd" targetNamespace="http://Servicio/"
  name="ServidorGeografico">
  <types>...</types>
  <message name="Entidades_Del_Area">...</message>
  <message name="Entidades_Del_AreaResponse">...</message>
  <message name="OrdenaPorTipo">...</message>
  <message name="OrdenaPorTipoResponse">...</message>
  <message name="TodasLasEntidades">...</message>
  <message name="TodasLasEntidadesResponse">...</message>
  <message name="TodasLasEntidades_Atributos">...</message>
  <message name="TodasLasEntidades_AtributosResponse">...</message>
  <message name="OrdenaPorNombre">...</message>
  <message name="OrdenaPorNombreResponse">...</message>
  <portType name="ServidorGeografico">...</portType>
  <binding name="ServidorGeograficoPortBinding" type="tns:ServidorGeografico">...</binding>
  <service name="ServidorGeografico">...</service>
</definitions>
```

Figura 3.14 WSDL del servicio web de recuperación semántica

En éste servicio web encontramos el nombre de los métodos que realizan la recuperación semántica de entidades geográficas desde nuestro repositorio, y algunos otros métodos que ayudan en el ranking de los resultados, métodos como el ordenamiento por costo y ordenamiento por tipo de entidad, dichos métodos son

accedidos mediante una conexión de la aplicación cliente, como se comentó anteriormente.

3.5 Etapa de Visualización

En esta etapa se llevan a cabo las tareas de visualización, análisis, ordenación y presentación de resultados encontrados durante la recuperación semántica. En la Figura 3.15 se muestran las tareas desarrolladas para esta etapa.

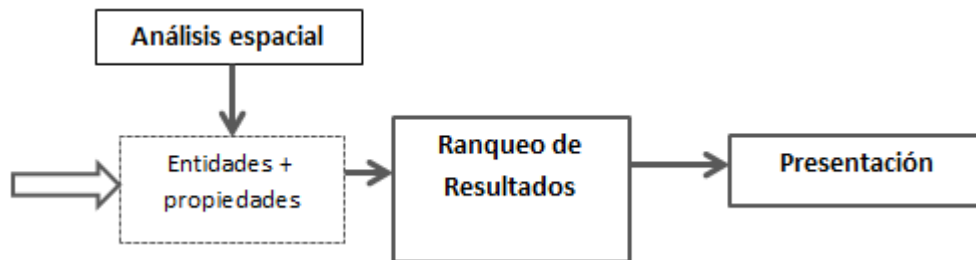


Figura 3.15 Tareas de la etapa de Visualización

3.5.1 Entidades + Propiedades

Es el conjunto de entidades y sus respectivas propiedades que se ha recuperado como resultado a una consulta del usuario, las cuales serán presentadas al usuario mediante una interfaz web, para lo cual, aplicamos los métodos de irdenamiento del servicio web y mediante el API de Google maps calculamos la distancia desde el usuario hasta cada una de dichas entidades.

3.5.2 Análisis Espacial

Más que un análisis espacial se hace el análisis de los atributos de las entidades recuperadas y las especificaciones del usuario para no dejar de lado el contexto, ya que éste también influye en la recuperación de información que el usuario desea. Autores como Rodríguez et al. (1999), Egenhofer (2002) y Schwering (2008) afirman que la distancia semántica puede ser complementada con las similitudes entre los atributos, funciones y relaciones entre las entidades.

En este caso, la recuperación de algunos elementos como la ruta que va desde el lugar en donde se encuentra el usuario que hace la solicitud de información, hasta

3. Metodología

la entidad o entidades recuperadas, mediante las herramientas que nos proporciona el API de Google Maps para tal fin, tendremos más elementos que faciliten al usuario la toma de decisiones.

El API de rutas de Google es un servicio que utiliza una solicitud HTTP para calcular rutas para llegar de una ubicación a otra. Puedes buscar rutas de varios métodos de transporte, como en transporte público, en auto particular, a pie o en bicicleta. Las rutas pueden especificar los orígenes, los destinos y los puntos intermedios como cadenas de texto o como coordenadas de latitud/longitud. El API de rutas puede devolver rutas segmentadas mediante una serie de puntos de referencia.

En la aplicación se implementa el cálculo de estas distancias mediante el uso del API de google maps con rutas en transporte particular, para lo cual se especifica la ubicación de la que desea partir el usuario, en el que mediante un mapa, el usuario puede seleccionar el punto de partida y posteriormente se calcula la ruta hacia cada una de las entidades recuperadas mediante la interfaz implementada y los resultados obtenidos mediante el servicio de recuperación de entidades geográficas. En la Figura 3.16 se observa un esquema básico de cómo se realiza el cálculo de rutas mediante el API de Google Maps.

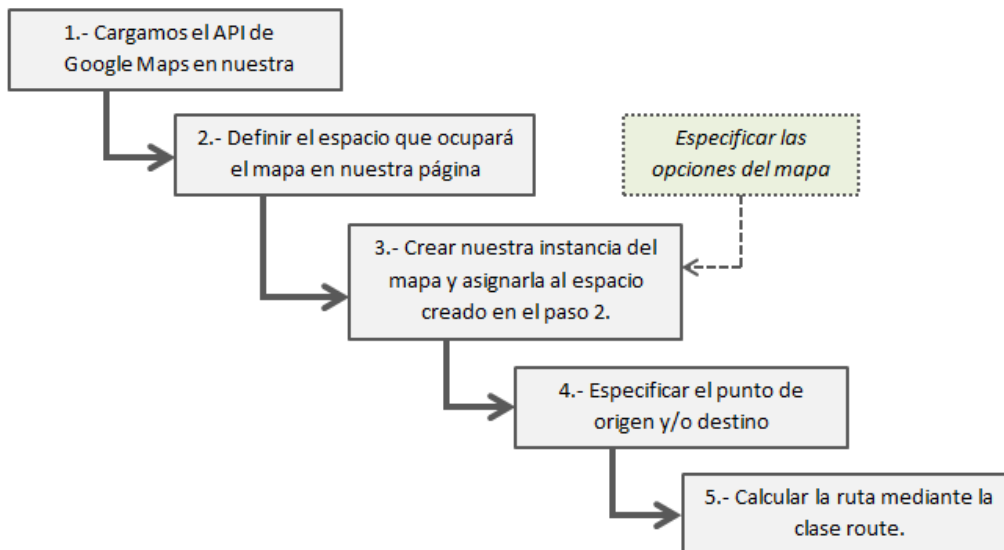


Figura 3.16 Esquema básico de cálculo de rutas mediante el API de Google Maps

1. Como primer paso, indicamos en nuestra web que haremos uso del API de Google Maps mediante la siguiente instrucción.

```
<script type="text/javascript" src="http://maps.google.com/maps/api/js?sensor=false"></script>
```

La URL incluida en la etiqueta `script` indica la ubicación de un archivo JavaScript que carga todos los símbolos y las definiciones para el uso del API de Google Maps, El parámetro `sensor` de la URL es obligatorio e indica si esta aplicación utiliza un sensor (por ejemplo, un localizador GPS) para determinar la ubicación del usuario, en éste caso no se utiliza ninguno.

2. Para que el mapa aparezca en una página web, se debe reservar un lugar para él. Normalmente, lo hacemos mediante la creación de un elemento `div` con un identificador que nos sirve como referencia a este elemento en el modelo de objetos de documento (DOM) del navegador, además es necesario darle el alto y ancho a este objeto, como se muestra en el siguiente ejemplo.

```
<div id="espacio_mapa" style="width: 100%; height: 100%"></div>
```

3. La clase de JavaScript que representa a los mapas es `Map`. Cada objeto de esta clase define un único mapa en una página. Creamos una nueva instancia de esta clase mediante el operador `new` de JavaScript.

Al crear una nueva instancia de mapa, se especifica un elemento HTML `<div>` en la página como contenedor para el mapa, en nuestro caso el “`espacio_mapa`” creado en el paso anterior. Los nodos HTML son elementos secundarios del objeto `document` de JavaScript. Se obtiene una referencia a este elemento mediante el método `document.getElementById()`.

```
Var map = new google.maps.Map(document.getElementById("espacio_mapa"), mapOptions);
```

Este código permite definir una variable (denominada `map`) y asignar dicha variable a un nuevo objeto `Map`, además de transmitir opciones definidas en el

objeto `mapOptions`. Estas opciones se utilizarán para inicializar las propiedades del mapa, como centrado, nivel de zoom, entre otras.

4. Para poder calcular la ruta mediante el API de Google Maps es necesario tener el punto de partida y el punto destino mediante latitud/longitud. Para especificar el punto de partida indicamos al mapa mediante un método *listener* que capture las coordenadas en las que nos posicionamos al hacer click para indicar nuestra ubicación. Esto se logra mediante la función que se muestra a continuación.

```
google.maps.event.addListener(map, "click", function(event) {
    var myLatLng = event.latLng;
    var lat = myLatLng.lat();
    var lng = myLatLng.lng();
    if (!puntoA) {
        puntoA = new google.maps.Marker({
            position: event.latLng,
            map: map
        });
        document.getElementById('lat').value = lat;
        document.getElementById('long').value = lng;
    } else {
        pos= new google.maps.LatLng(lat,lng);
        CambiaMarcador(puntoA,pos);
        document.getElementById('lat').value = lat;
        document.getElementById('long').value = lng;
    }
});
```

En la Figura 3.17 podemos observar un ejemplo de cómo se especifica el punto de partida que indica el usuario mediante un mapa de Google.

3. Metodología

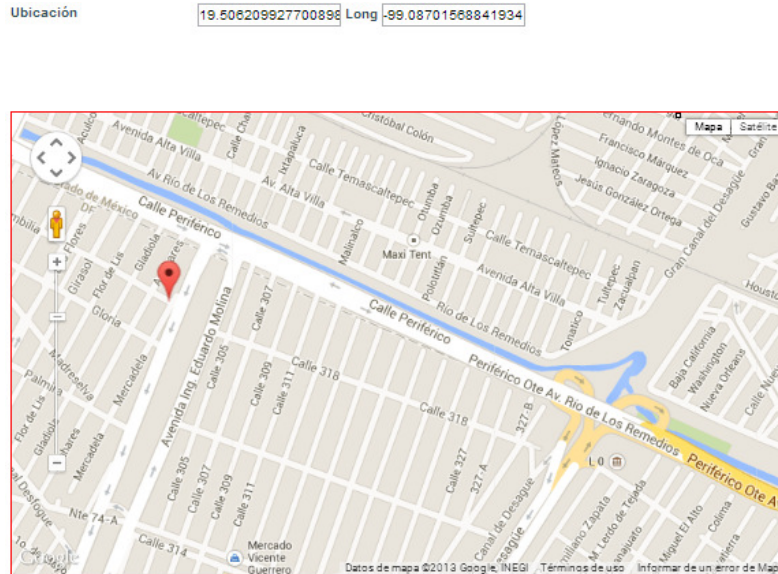


Figura 3.17 Indicar la posición a través de un mapa de google

5. Para calcular la ruta mediante las herramientas del API de Google Maps es necesario invocar la clase *route* y verificar si se puede trazar una ruta entre los puntos especificados entre el origen y destino, en caso afirmativo podemos extraer la distancia y duración de dicha ruta.

```
function Ruta(map, origen, destino) {  
  var request = {  
    origin:origen,  
    destination:destino,  
    //waypoints:waypts,  
    travelMode: google.maps.TravelMode.DRIVING  
  };  
  directionsService.route(request, function(response, status) {  
    if (status == google.maps.DirectionsStatus.OK) {  
      directionsDisplay.setDirections(response);  
      var route = response.routes[0];  
      var summaryPanel = document.getElementById('directions_panel');  
      summaryPanel.innerHTML = '';  
      for (var i = 0; i < route.legs.length; i++) {  
        summaryPanel.innerHTML += 'Duración: ' + route.legs[i].duration.text ;  
        summaryPanel.innerHTML += 'Distancia: ' + route.legs[i].distance.text;  
      }  
    }  
  });  
};
```

En la Figura 3.18 se muestra un ejemplo del cálculo de ruta obtenido mediante el API de Google Maps a como se obtienen en la etapa de visualización del servicio de recuperación semántica de información geográfica.

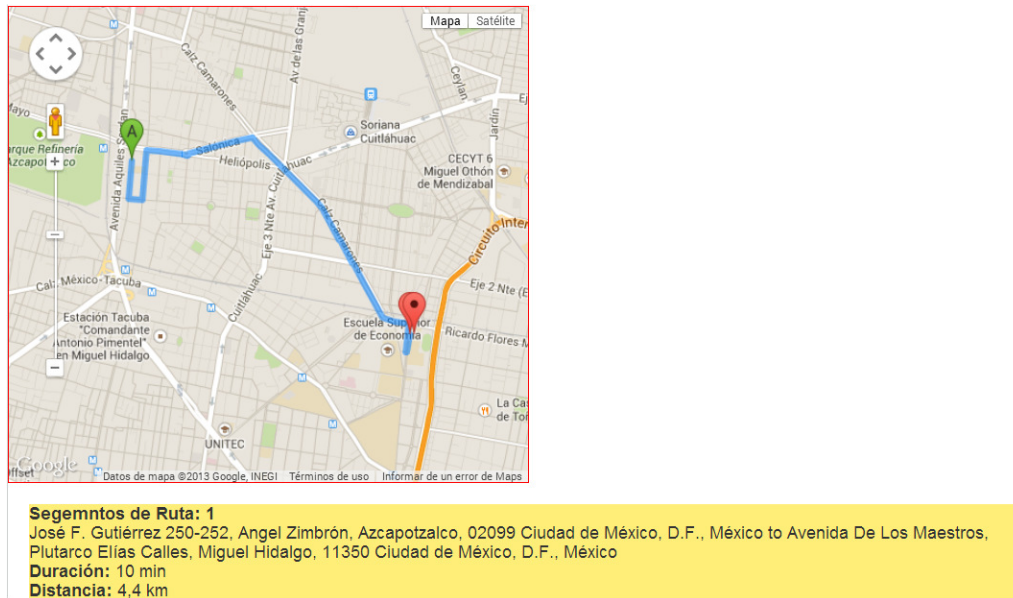


Figura 3.18 Ejemplo de cálculo de distancia entre dos puntos mediante el API de google maps

3.5.3 Ranking de resultados

De acuerdo con Zhu et al. (2010), una consulta de *ranking* relacional (consulta Top-N) q es encontrar un conjunto ordenado de N respuestas que están más cerca de q según una función de *ranking* dada. Las investigaciones sobre las consultas de **Top-N** se han intensificado desde finales de los noventa, y la mayoría implican atributos numéricos y utilizar una función de distancia numérica para reducir el conjunto de resultados de una consulta convencional para algunas de las respuestas más relevantes.

Los métodos de ordenamiento por costo y por tipo de entidad que presta el servicio de recuperación, nos facilitan la tarea de ranking u ordenamiento de los resultados ya que éstos son ordenados en grupos de acuerdo al tipo de entidad y de manera ascendente de acuerdo al costo.

3. Metodología

Un servicio Web por sí solo no puede hacer nada, es necesario que un usuario, en este caso mediante una aplicación cliente, lleve a cabo las peticiones o consultas de información. Para el servicio de recuperación semántica de información geográfica, un usuario hace la consulta mediante la aplicación, la cual es contestada por el servicio Web mediante una respuesta en formato SOAP, la cual es interpretada y representada por la aplicación cliente. En nuestra aplicación cliente, se realiza un ranking de los resultados de acuerdo con los campos del costo, el tipo y la distancia que hay del usuario a la entidad y que se ha calculado anteriormente.

En la Figura 3.18 se muestra un ejemplo del ranking de resultados, agrupando las entidades recuperadas por tipo de entidad: pública o privada y ordenadas de manera ascendente por el costo.

| NOMBRE | TIPO | COSTO | CARRERA |
|--|---------|-------|----------------------|
| Centro Interdisciplinario de Ciencias de la Salud Unidad Santo Tomas | Publica | 0 | PSICOLOGIA |
| Instituto Mexicano De Psicologia Plantel Montevideo | Publica | 0 | PSICOPEDAGOGIA |
| Universidad De Ciencias Y Administracion | Publica | 0 | PSICOLOGIA_SOCIAL |
| Universidad Del Tepeyac | Privada | 10200 | PSICOLOGIA |
| Universidad ICEL | Privada | 30000 | PSICOLOGIA_EDUCATIVA |
| Universidad ICEL | Privada | 30000 | PSICOLOGIA_SOCIAL |
| Universidad Insurgentes Norte | Privada | 50000 | PSICOLOGIA |
| Universidad Justo Sierra | Privada | 65000 | PSICOLOGIA |
| Centro Cultural Universitario Justo Sierra | Privada | 65000 | PSICOLOGIA_EDUCATIVA |

Figura 3.19 Ranking de resultados mediante el costo y tipo de entidad

3.5.4 Presentación

Una vez hecho el ordenamiento de los resultados obtenidos del servicio Web, éstos se presentan mediante una interfaz Web al usuario.

Para mostrar los resultados mediante una interfaz web seguimos una serie de pasos los cuales involucran la conexión del servicio web, la invocación de los métodos necesarios para la recuperación semántica, formato de resultados, entre otros como se muestra en la Figura 3.20.

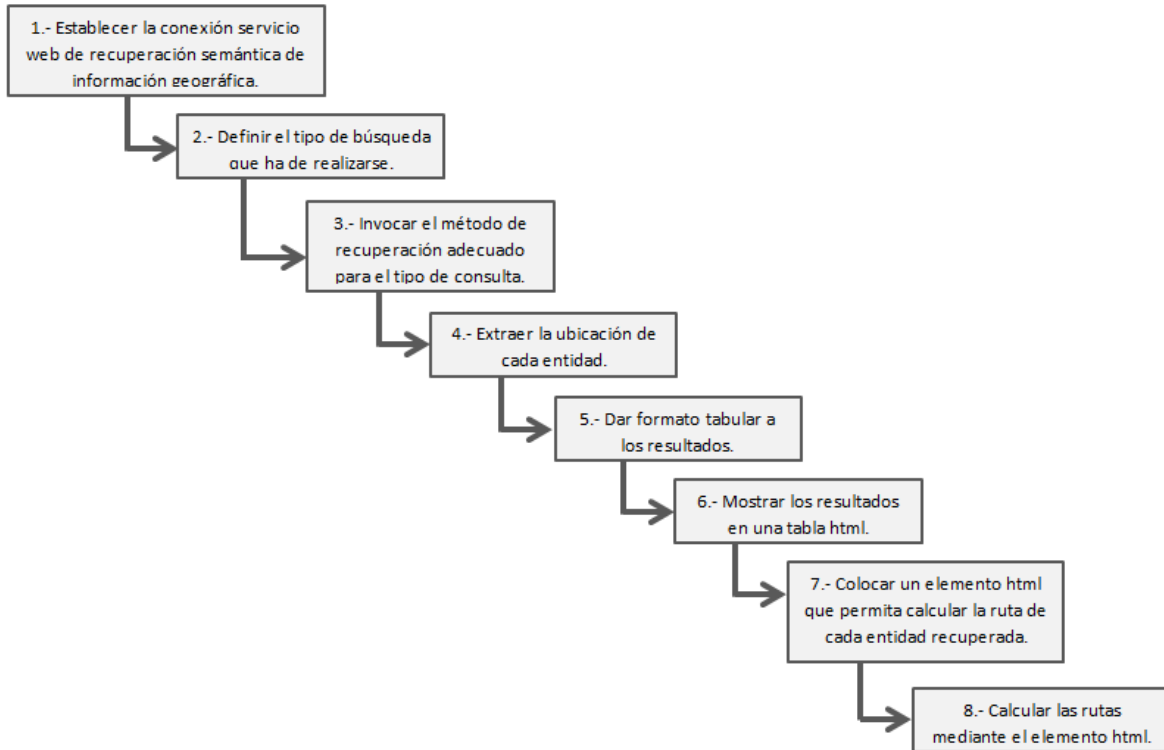


Figura 3.20 Esquema básico de la presentación de resultados al usuario

La interfaz consta de una tabla con el nombre de las entidades obtenidas y cada uno de los atributos de la misma, además de un botón que nos permite hacer el cálculo de la distancia y trazar la ruta que hay de la posición del usuario hacia cada una de dichas entidades. Cabe mencionar que el usuario introduce su ubicación mediante el uso de una de las herramientas de google maps que permite colocar un marcador en el mapa, especificando la ubicación del usuario.

Además de los resultados obtenidos y ordenados en una tabla, cada una de las entidades se presenta mediante un marcador en un mapa digital, así como la ruta desde el punto especificado por el usuario hacía cada entidad, el tiempo estimado y algunos puntos intermedios de la ruta.

CAPÍTULO 4. RESULTADOS

4. Resultados

En este capítulo se pretende dar una vista general de los resultados de algunos experimentos realizados a la implementación de la metodología propuesta. Los experimentos consisten en revisar algunos de los distintos escenarios que pueden resultar de una consulta dada. Así mismo se presentan algunas interpretaciones de los resultados a dichos experimentos.

Cabe mencionar que el sistema desarrollado funciona para instancias de la delegación Gustavo A. Madero, debido a que los datos obtenidos pertenecen a esta delegación y al ser un servicio únicamente de recuperación de información, se dotó de información a la ontología manualmente.

El servicio de recuperación está implementado con base en un conjunto de consultas que además de ser controladas, son guiadas de acuerdo con la categorización de conceptos dentro de nuestro repositorio de información, esto permite que los conceptos buscados se encuentren dentro del repositorio, pero no quiere decir que el servicio Web esté restringido a este tipo de búsqueda.

4.1 Escenarios de consulta

- Consulta una entidad
- No se selecciona ningún concepto
- Consulta un concepto
- Consulta en jerarquía de conceptos
- Consulta un concepto sin instancias

4.1.1 Consulta una entidad

En esta consulta, se recuperan las propiedades de alguna entidad en particular y se hace un listado de los conceptos que están ligados con la misma, ésta es una

4. Resultados

buena opción cuando ya tenemos una idea de la entidad que deseamos conocer. En la Figura 4.1 se muestra un ejemplo de cómo se indica este tipo de consulta.

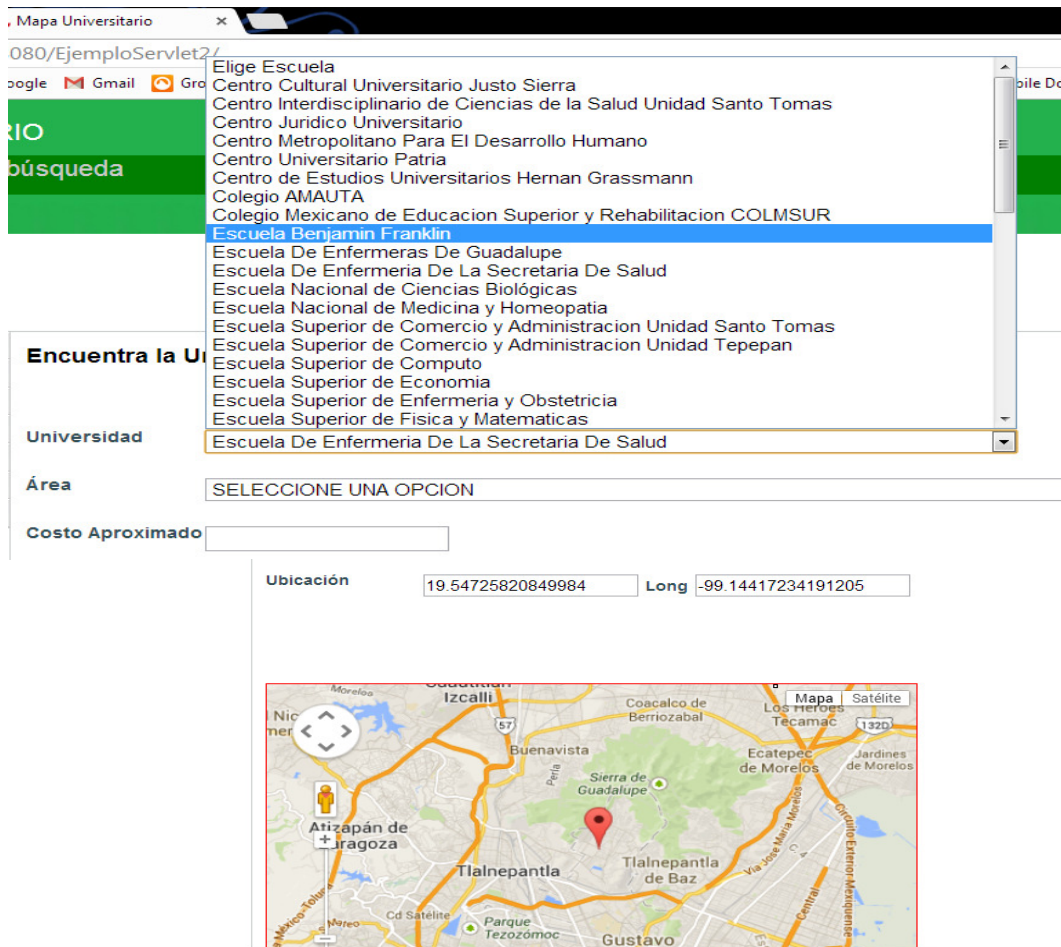


Figura 4.1 Consulta de una entidad en específico.

Es necesario dar nuestra ubicación en el mapa, si deseamos más adelante calcular la ruta que hay de nuestra posición a la ubicación de las entidades encontradas en la recuperación de información. En la Figura 4.2 se muestra la tabla con los resultados de la consulta anterior y el respectivo mapa para calcular las rutas a cada entidad. Asimismo, en la Figura 4.3 se muestra el cálculo de la ruta con los tiempos de recorrido, así como la distancia en cuestión.

4. Resultados

| ID | NOMBRE | TIPO | COSTO ANUAL APROX. | UBICACION | CARRERA |
|----------------------|---------------------------|---------|--------------------|---|----------------|
| Ruta | Escuela Benjamin Franklin | Privada | 30000 | INSURGENTES NORTE NO. 1423-COL. GUADALUPE INSURGENTES-GUSTAVO A. MADERO-07870 | CONTADURIA |
| Ruta | Escuela Benjamin Franklin | Privada | 30000 | INSURGENTES NORTE NO. 1423-COL. GUADALUPE INSURGENTES-GUSTAVO A. MADERO-07870 | ADMINISTRACION |
| Ruta | Escuela Benjamin Franklin | Privada | 30000 | INSURGENTES NORTE NO. 1423-COL. GUADALUPE INSURGENTES-GUSTAVO A. MADERO-07870 | MERCADOTECNIA |
| Ruta | Escuela Benjamin Franklin | Privada | 30000 | INSURGENTES NORTE NO. 1423-COL. GUADALUPE INSURGENTES-GUSTAVO A. MADERO-07870 | INFORMATICA |
| Ruta | Escuela Benjamin Franklin | Privada | 30000 | INSURGENTES NORTE NO. 1423-COL. GUADALUPE INSURGENTES-GUSTAVO A. MADERO-07870 | DERECHO |

Figura 4.2 Tabla con los atributos y conceptos de la entidad consultada.

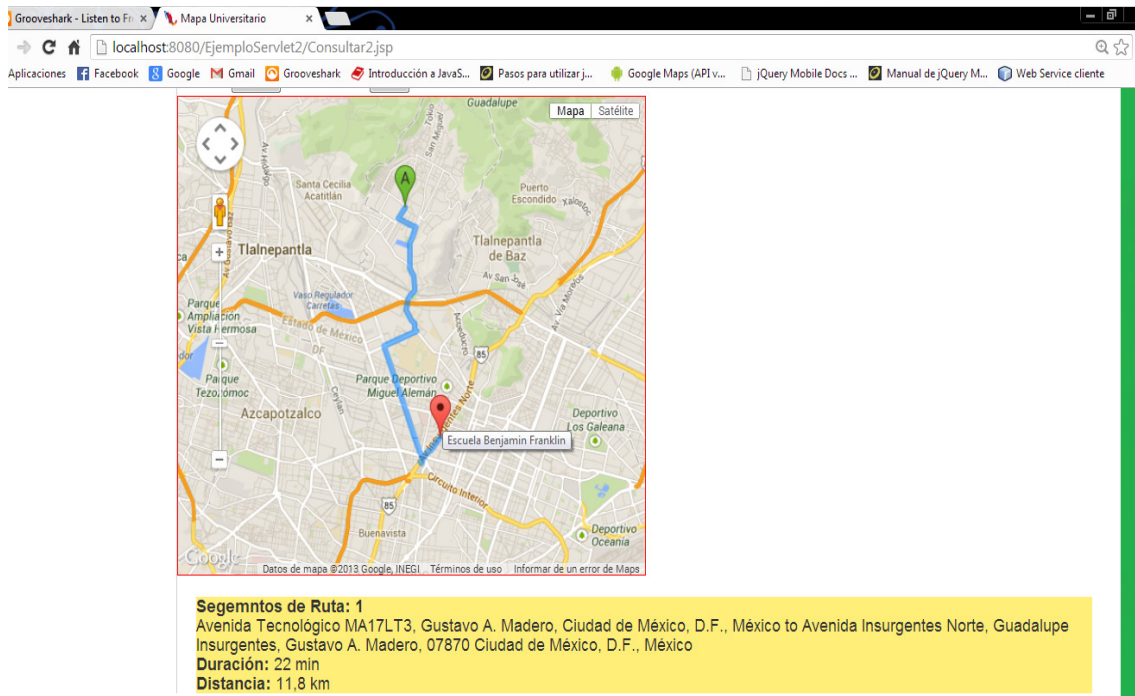


Figura 4.3 Mapa para el cálculo de rutas, distancia, duración y puntos intermedios.

Cuando se habla en la metodología del análisis espacial para el ranking de los resultados, se refiere a las técnicas implementadas del API de Google Maps para

4. Resultados

la medición de la distancia que hay de la posición proporcionada por el usuario hacia cada una de las entidades encontradas por el servicio de recuperación.

Una de las limitantes encontradas para poder hacer completamente el ranking de entidades propuestas en la metodología, fue que la recuperación de información se lleva a cabo de lado del servidor de la aplicación, mientras que el cálculo de la distancia se lleva a cabo del lado del cliente.

4.1.2 No se selecciona ningún concepto

Un segundo escenario es cuando se propone una búsqueda sin alguna área seleccionada, en la cual el servicio de recuperación devuelve las entidades que se encuentren más cerca del concepto principal. En este caso, el concepto principal es universidades y de acuerdo con el algoritmo propuesto, se listan los subconceptos de la clase *Universidades*, de los cuales recuperamos todas sus instancias en caso de que las tengan. En caso contrario, listamos los subconceptos de cada uno de los subconceptos de la clase *Universidades* y de igual manera recuperamos sus instancias. Para realizar lo anterior, se llevó a cabo un proceso de inferencia basado en el razonador utilizado en la ontología.

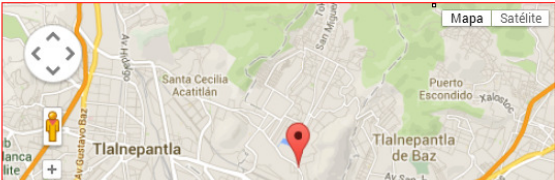
Encuentra la Universidad a tu medida

Universidad

Área

Costo Aproximado

Ubicación Long



Mapa Satélite

Figura 4.4 Esquema de una consulta sin concepto de búsqueda.

4. Resultados

En la Figura 4.4 podemos observar cómo se especifica una consulta sin indicar algún concepto de búsqueda, sólo basta con indicar la ubicación del usuario. Así mismo, en la Figura 4.5 se presenta el resultado de dicha consulta.

| ID | NOMBRE | TIPO | COSTO ANUAL APROX. | UBICACION | CARRERA |
|-------------------------------|---|---------|--------------------|---|-----------------------|
| <input type="checkbox"/> Ruta | Escuela Superior de Turismo | Publica | 0 | AV. LUIS ENRIQUE ERRO S/N-UNIDAD ZACATENCO-GUSTAVO A. MADERO-07738 | TURISMO |
| <input type="checkbox"/> Ruta | Universidad Del Tepeyac | Privada | 10200 | CALLAO NO. 802-COL. LINDAVISTA-GUSTAVO A. MADERO-7300 | TURISMO |
| <input type="checkbox"/> Ruta | Intituto Especializado en Computacion y Administracion Gauss Jordan | Privada | 25000 | DVORAK NO. 59-COL. VALLEJO-GUSTAVO A. MADERO-07870 | TURISMO |
| <input type="checkbox"/> Ruta | Universidad Insurgentes Norte | Privada | 50000 | AVENIDA ACUEDUCTO NO. 13-COL. SAN PEDRO ZACATENCO-GUSTAVO A. MADERO-07360 | MERCADOTECNIA |
| <input type="checkbox"/> Ruta | Universidad Del Tepeyac | Privada | 10200 | CALLAO NO. 802-COL. LINDAVISTA-GUSTAVO A. MADERO-7300 | MERCADOTECNIA |
| <input type="checkbox"/> Ruta | Universidad Tecnologica Roosevelt | Privada | 0 | CALZADA DE LOS MISTERIOS NO. 801-COL. TEPEYAC INSURGENTES-GUSTAVO A. MADERO-07020 | MERCADOTECNIA |
| <input type="checkbox"/> Ruta | Universidad Justo Sierra Ticoman | Privada | 65000 | AV. ACUEDUCTO 914-COL. LAGUNA TICOMAN-GUSTAVO A. MADERO-7340 | MERCADOTECNIA |
| <input type="checkbox"/> Ruta | Escuela Benjamin Franklin | Privada | 30000 | INSURGENTES NORTE NO. 1423-COL. GUADALUPE INSURGENTES-GUSTAVO A. MADERO-07870 | MERCADOTECNIA |
| <input type="checkbox"/> Ruta | Escuela Superior de Economia | Publica | 0 | PLAN DE AGUA PRIETA 66-COL. PLUTARCO ELIAS CALLES-MIGUEL HIDALGO-11350 | ECONOMIA |
| <input type="checkbox"/> Ruta | Universidad De Ciencias Y Administracion | 0 | 0 | GUILLERMO MASSIEU HELGUERA NO. 103-COL. SAN JOSE TICOMAN-GUSTAVO A. MADERO-7320 | CONTADURIA_Y_FINANZAS |
| <input type="checkbox"/> Ruta | Universidad Del Tepeyac | Privada | 10200 | CALLAO NO. 802-COL. LINDAVISTA-GUSTAVO A. MADERO-7300 | CONTADURIA_Y_FINANZAS |
| <input type="checkbox"/> Ruta | Universidad Del Tepeyac | Privada | 10200 | CALLAO NO. 802-COL. LINDAVISTA-GUSTAVO A. MADERO-7300 | DISENO_GRAFICO |

Figura 4.5 Entidades que cumplen con los criterios de búsqueda.

El resultado corresponde a las entidades con la distancia más corta que hay del concepto principal *Universidades*, a los primeros conceptos con instancias dentro de la jerarquía, lo cual se muestra en la Figura 4.6.

4. Resultados

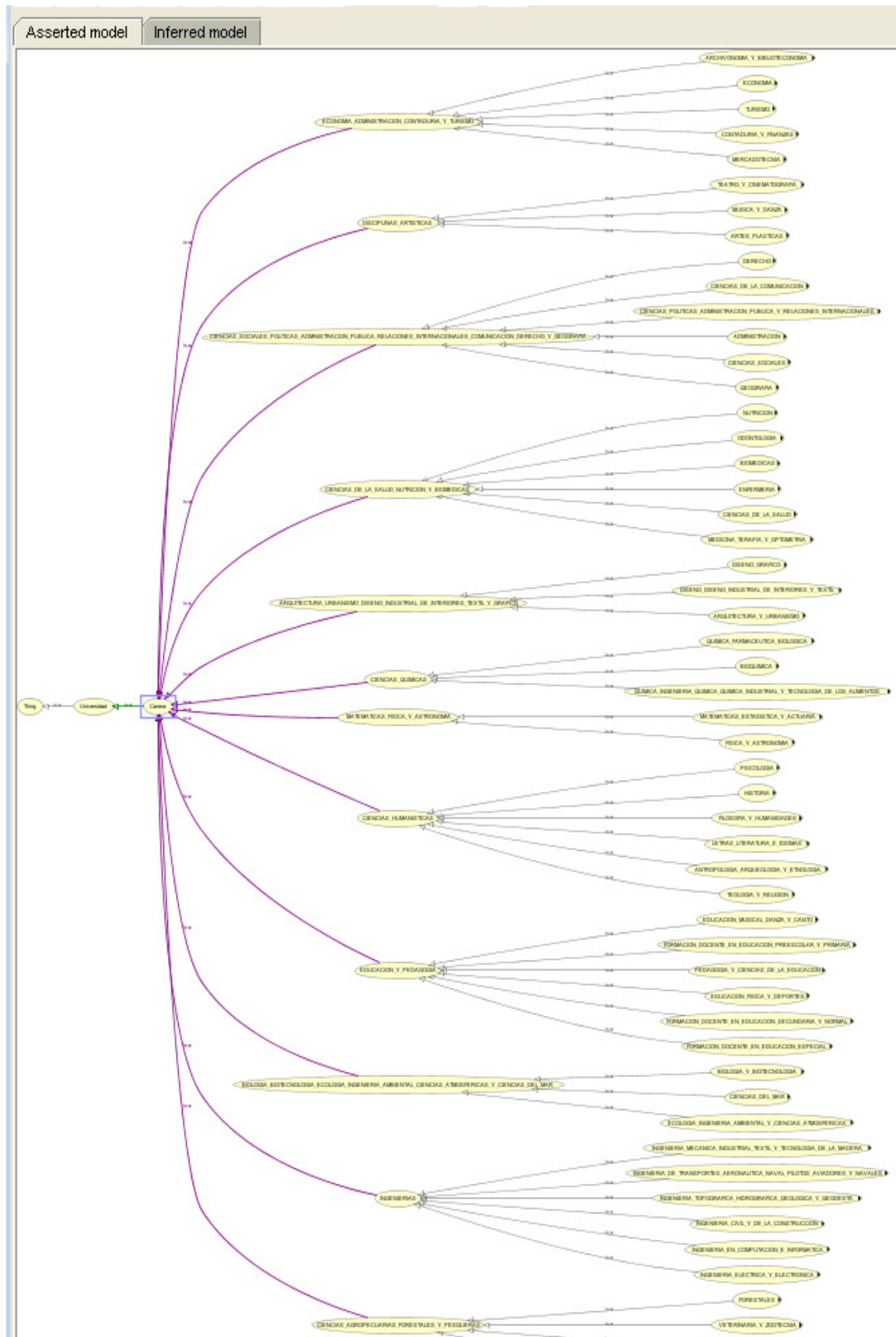


Figura 4.6 Conceptos de la ontología más cercanos al concepto raíz.

4.1.3 Consulta un concepto

En un tercer escenario se realiza una búsqueda por algún concepto en particular, se recuperan las instancias del mismo concepto en caso de que las tenga; en caso de que no se tenga ninguna, verificamos si el concepto tiene subconceptos y se procede a recuperar las instancias en caso de que las tengan. De esta manera, continuamos descendiendo en la jerarquía hasta encontrar instancias o algún nodo hoja. La Figura 4.7 muestra cómo se especifica una consulta de este tipo.

Encuentra la Universidad a tu medida

| | |
|------------------|--|
| Universidad | <input type="text" value="Elige Escuela"/> |
| Área | <input type="text" value="CIENCIAS_HUMANISTICAS"/> |
| Sub área | <input type="text" value="SELECCIONE UNA OPCION"/> |
| Costo Aproximado | <input type="text"/> |
| Ubicación | <input type="text" value="19.533345576583905"/> Long <input type="text" value="-99.13596380793024"/> |

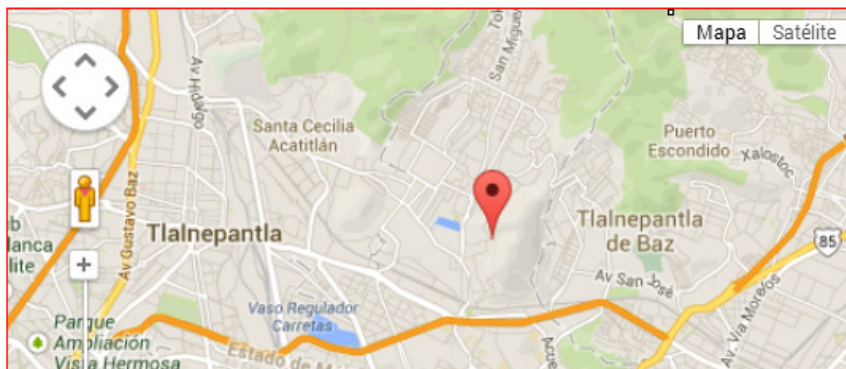


Figura 4.7 Consulta las entidades de un concepto en específico.

En la Figura 4.8 podemos apreciar las instancias recuperadas, analizando los subconceptos del concepto buscado.

4. Resultados

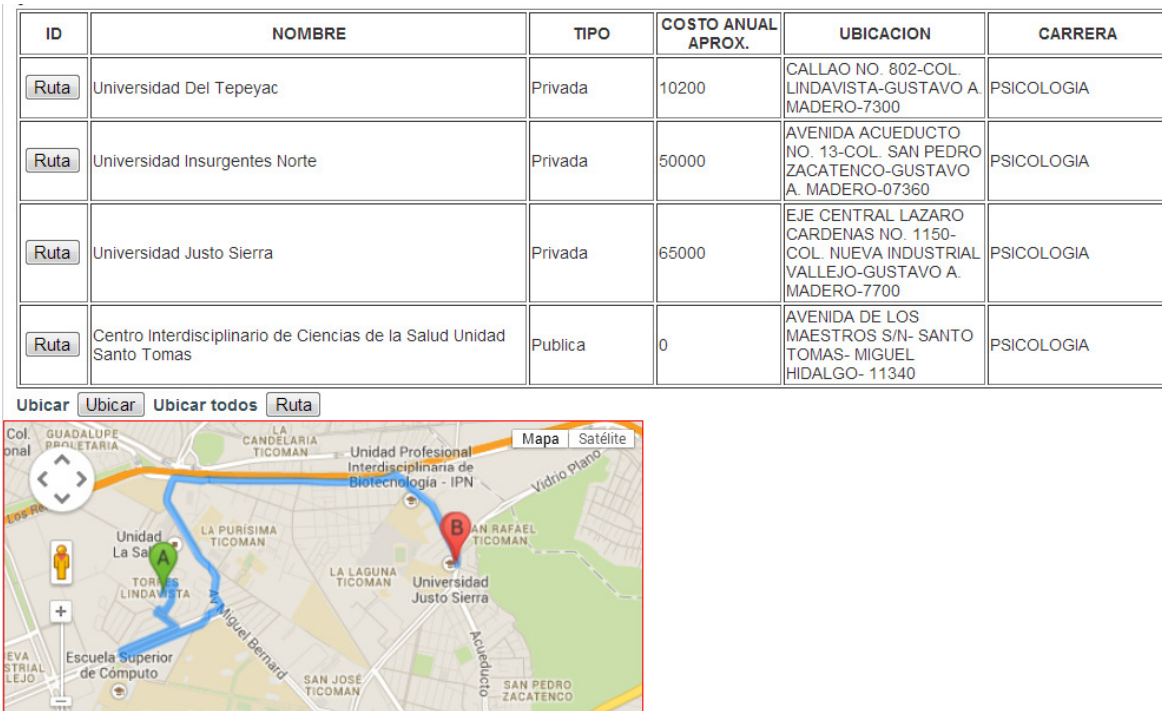


Figura 4.8 Resultado de la consulta de un concepto en específico.

En la Figura 4.9 se presenta el fragmento de la ontología que contiene el concepto buscado y sus subconceptos inmediatos, de los cuales el único que cuenta con instancias es *psicología*. Si esta instancia fue comparada con la tabla de resultados, obteniendo el mismo valor.

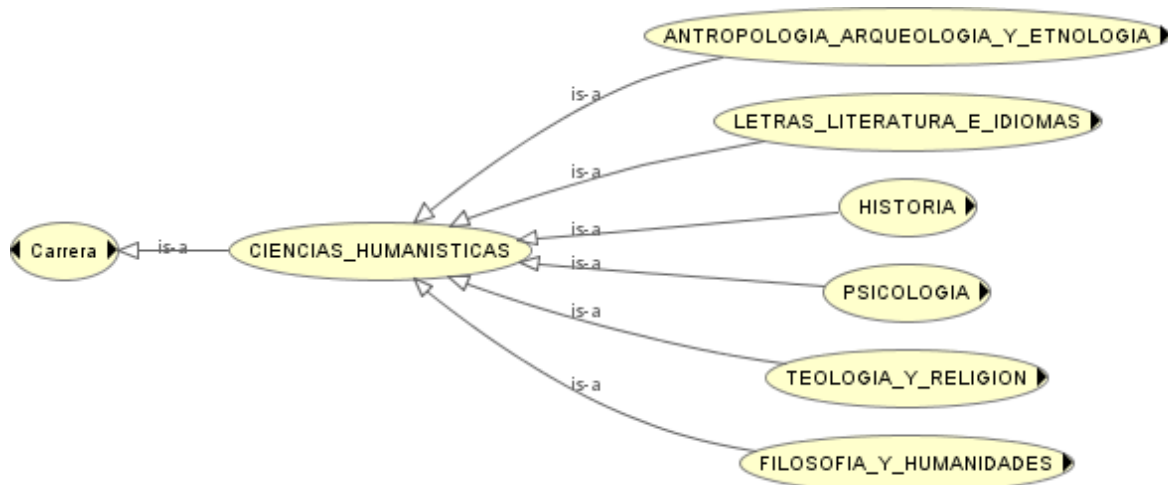


Figura 4.9 Fragmento de la ontología con el concepto buscado y sus subconceptos.

4.1.4 Consulta en jerarquía

Para la búsqueda de las entidades de un subconcepto en específico, al igual que en el caso anterior, es necesario verificar si ese subconcepto tiene instancias; en ese caso, se recuperan las propiedades de cada una. En caso contrario, se busca en sus subconceptos hasta encontrar instancias o el final de los subconceptos, en caso de llegar al final de la ontología sin tener ningún resultado. Posteriormente se verifica el super concepto (concepto padre) del concepto buscado, y entonces se procede a realizar la búsqueda de instancias en ese concepto y demás subconceptos en caso de aún no haber recuperado ninguna instancia. En la Figura 4.10 se muestra un ejemplo de la consulta de un subconcepto de la clase de *ciencias sociales*, el cual si cuenta con instancias (ver Figura 4.11).

Encuentra la Universidad a tu medida

Universidad

Área

Sub área

Carrera

Costo Aproximado

Ubicación Long

Figura 4.10 Consulta de una subclase

4. Resultados

| ID | NOMBRE | TIPO | COSTO ANUAL APROX. | UBICACION | CARRERA |
|------|---|---------|--------------------|---|---------|
| Ruta | Universidad Tecnologica Roosevelt | Privada | 0 | CALZADA DE LOS MISTERIOS NO. 801-COL. TEPEYAC INSURGENTES-GUSTAVO A. MADERO-07020 | DERECHO |
| Ruta | Universidad ICEL | Privada | 30000 | AV. CANTERA NO. 16-COL. ESTANZUELA-GUSTAVO A. MADERO-7060 | DERECHO |
| Ruta | Universidad Justo Sierra | Privada | 65000 | EJE CENTRAL LAZARO CARDENAS NO. 1150-COL. NUEVA INDUSTRIAL VALLEJO-GUSTAVO A. MADERO-7700 | DERECHO |
| Ruta | Centro Universitario Patria | Privada | 23000 | CARLOTA NO.68-GUADALUPE TEPEYAC-GUSTAVO A. MADERO-7840 | DERECHO |
| Ruta | Escuela Benjamin Franklin | Privada | 30000 | INSURGENTES NORTE NO. 1423-COL. GUADALUPE INSURGENTES-GUSTAVO A. MADERO-07870 | DERECHO |
| Ruta | Universidad Del Tepeyac | Privada | 10200 | CALLAO NO. 802-COL. LINDAVISTA-GUSTAVO A. MADERO-7300 | DERECHO |
| Ruta | Centro de Estudios Universitarios Hernan Grassmann | Privada | 83000 | CALLE MALAQUITA NO. 93-COL. ESTRELLA-GUSTAVO A. MADERO-07810 | DERECHO |
| Ruta | Universidad De Ciencias Y Administracion | 0 | 0 | GUILLERMO MASSIEU HELGUERA NO. 103-COL. SAN JOSE TICOMAN-GUSTAVO A. MADERO-7320 | DERECHO |
| Ruta | Colegio AMAUTA | Privada | 24000 | CALZADA DE GUADALUPE NO. 183-COL. 7 DE NOVIEMBRE-GUSTAVO A. MADERO-07840 | DERECHO |
| Ruta | Intituto Especializado en Computacion y Administracion Gauss Jordan | Privada | 25000 | DVORAK NO. 59-COL. VALLEJO-GUSTAVO A. MADERO-07870 | DERECHO |
| Ruta | Instituto Mexicano De Estudios Superiores Y De Postgrado | Publica | 0 | AV. MONTEVIDEO NO. 635-COL. SAN BARTOLO ATEPEHUACAN-GUSTAVO A. MADERO-7730 | DERECHO |

Figura 4.11 Instancias Recuperadas de la consulta de un subconcepto

En caso de que el subconcepto buscado no tenga instancias, es necesario buscar en los conceptos más cercanos.

En la Figura 4.12 se especifica la búsqueda de un concepto del cual no se tienen instancias, por lo tanto, hay que realizar un recorrido a través de su concepto padre, tal como se muestra en las Figuras 4.13 y 4.14.

Asimismo, en la Figura 4.14 se observan los conceptos que derivan del mismo concepto que deriva el concepto buscado.

Encuentra la Universidad a tu medida

Universidad:

Área:

Sub área:

Carrera:

Costo Aproximado:

Ubicación: Long:

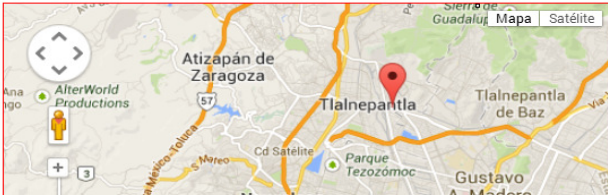


Figura 4.12 Consulta de un subconcepto que no tiene instancias



Figura 4.13 Fragmento de la ontología que indica las clases de las cuales deriva el concepto buscado.

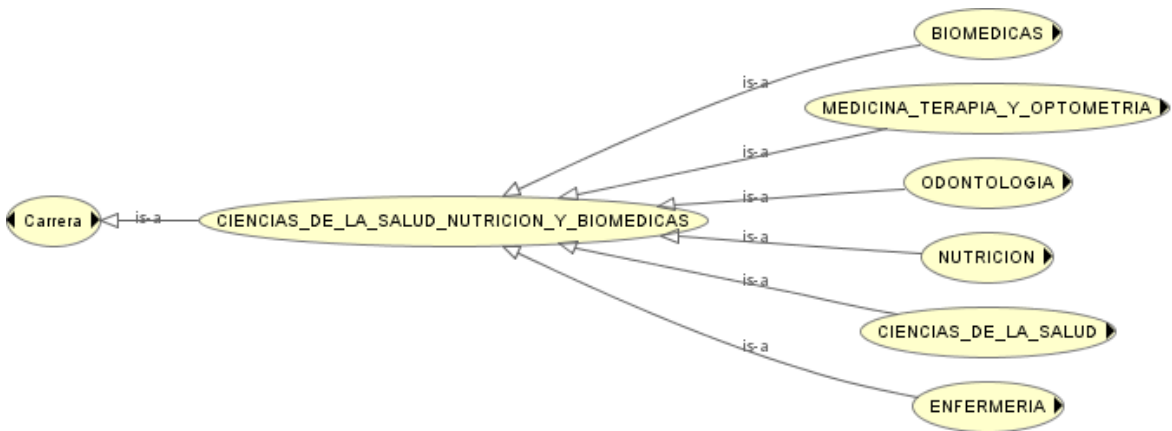


Figura 4.14 Conceptos derivados de la misma clase que el concepto buscado

En la Figura 4.15 se describen las instancias recuperadas de los conceptos relacionados con el concepto buscado específicamente.

4. Resultados

| ID | NOMBRE | TIPO | COSTO ANUAL APROX. | UBICACION | CARRERA |
|----------------------|--|---------|--------------------|---|-------------|
| Ruta | Universidad De Ciencias Y Administracion | 0 | 0 | GUILLERMO MASSIEU HELGUERA NO. 103-COL. SAN JOSE TICOMAN-GUSTAVO A. MADERO-7320 | ODONTOLOGIA |
| Ruta | Centro Interdisciplinario de Ciencias de la Salud Unidad Santo Tomas | Publica | 0 | AVENIDA DE LOS MAESTROS S/N- SANTO TOMAS- MIGUEL HIDALGO- 11340 | ODONTOLOGIA |
| Ruta | Universidad Justo Sierra Ticoman | Privada | 65000 | AV. ACUEDUCTO 914-COL. LAGUNA TICOMAN-GUSTAVO A. MADERO-7340 | NUTRICION |
| Ruta | Universidad De Ciencias Y Administracion | 0 | 0 | GUILLERMO MASSIEU HELGUERA NO. 103-COL. SAN JOSE TICOMAN-GUSTAVO A. MADERO-7320 | NUTRICION |

Figura 4.15 Entidades recuperadas de los conceptos relacionados con el concepto buscado

4.1.5 Consulta de un concepto sin instancias

Para el caso de una búsqueda, en la cual sea el último nivel de la jerarquía y no se encuentren instancias de un concepto en particular; es necesario buscar en los conceptos que derivan del mismo concepto padre del que desciende el concepto buscado, tal como se muestra en la Figura 4.16.

De igual manera, en la Figura 4.17 se presenta la tabla con los resultados de esta consulta en particular.

Encuentra la Universidad a tu medida

Universidad

Área

Sub área

Carrera

Costo Aproximado

Ubicación Long

Figura 4.16 Consulta de un concepto sin subconceptos

4. Resultados

| ID | NOMBRE | TIPO | COSTO ANUAL APROX. | UBICACION | CARRERA |
|------|--|---------|--------------------|---|----------------------|
| Ruta | Universidad Del Tepeyac | Privada | 10200 | CALLAO NO. 802-COL. LINDAVISTA-GUSTAVO A. MADERO-7300 | PSICOLOGIA |
| Ruta | Universidad Insurgentes Norte | Privada | 50000 | AVENIDA ACUEDUCTO NO. 13-COL. SAN PEDRO ZACATENCO-GUSTAVO A. MADERO-07360 | PSICOLOGIA |
| Ruta | Universidad Justo Sierra | Privada | 65000 | EJE CENTRAL LAZARO CARDENAS NO. 1150-COL. NUEVA INDUSTRIAL VALLEJO-GUSTAVO A. MADERO-7700 | PSICOLOGIA |
| Ruta | Centro Interdisciplinario de Ciencias de la Salud Unidad Santo Tomas | Publica | 0 | AVENIDA DE LOS MAESTROS S/N- SANTO TOMAS- MIGUEL HIDALGO- 11340 | PSICOLOGIA |
| Ruta | Universidad ICEL | Privada | 30000 | AV. CANTERA NO. 16-COL. ESTANZUELA-GUSTAVO A. MADERO-7060 | PSICOLOGIA_EDUCATIVA |
| Ruta | Centro Cultural Universitario Justo Sierra | Privada | 65000 | URANIO NO. 157-COL. NUEVA INDUSTRIAL VALLEJO-GUSTAVO A. MADERO-7700 | PSICOLOGIA_EDUCATIVA |
| Ruta | Instituto Mexicano De Psicologia Plantel Montevideo | | 0 | AV. MONTEVIDEO NO. 635-COL. SAN BARTOLO ATEPEHUACAN-GUSTAVO A. MADERO-7730 | PSICOPEDAGOGIA |
| Ruta | Universidad ICEL | Privada | 30000 | AV. CANTERA NO. 16-COL. ESTANZUELA-GUSTAVO A. MADERO-7060 | PSICOLOGIA_SOCIAL |
| Ruta | Universidad De Ciencias Y Administracion | 0 | 0 | GUILLERMO MASSIEU HELGUERA NO. 103-COL. SAN JOSE TICOMAN-GUSTAVO A. MADERO-7320 | PSICOLOGIA_SOCIAL |

Figura 4.17 Resultado de la consulta de un concepto sin subconceptos y sin instancias

4.1.6 Consulta de un concepto sin instancias en Jerarquía

En caso de buscar en algún subconcepto, el cual no tenga instancias, además todos sus subconceptos tampoco tengan instancias, se procede a verificar en el nivel superior y sus respectivos subconceptos, y así sucesivamente hasta encontrar instancias, como se muestra en las Figuras 4.18, 4.19 y 4.20, donde el concepto *Fruticultura* no presenta instancias.



Figura 4.18 Ejemplo de un subconcepto sin subconceptos y sin instancias.

4. Resultados

Posteriormente, verificamos en el concepto *agronomía* que en este caso es el superconcepto de *fruticultura*, el cual tampoco contiene ninguna instancia. Por tanto, retrocedemos un nivel más en la jerarquía para verificar el concepto de *Ciencias Agropecuarias forestales y pesqueras*, de las cuales, ninguno de sus subconceptos tiene instancias, así retrocedemos un nivel más, donde se encuentran resultados, ya que llegamos hasta el primer nivel en la jerarquía, ahora sólo recuperaremos las instancias de los conceptos más cercanos a la raíz de la jerarquía, tal como se hizo en el segundo escenario de consultas.

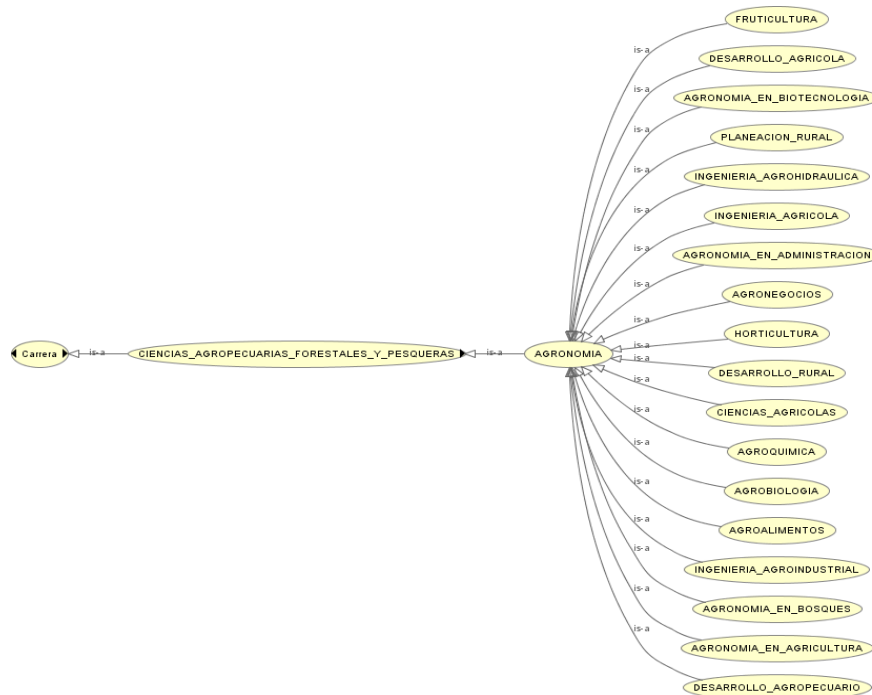


Figura 4.19 Conceptos más cercanos al concepto buscado

4. Resultados

| ID | NOMBRE | TIPO | COSTO ANUAL APROX. | UBICACION | CARRERA |
|----------------------|---|---------|--------------------|---|-----------------------|
| Ruta | Escuela Superior de Turismo | Publica | 0 | AV. LUIS ENRIQUE ERRO S/N-UNIDAD ZACATENCO-GUSTAVO A. MADERO-07738 | TURISMO |
| Ruta | Universidad Del Tepeyac | Privada | 10200 | CALLAO NO. 802-COL. LINDAVISTA-GUSTAVO A. MADERO-7300 | TURISMO |
| Ruta | Intituto Especializado en Computacion y Administracion Gauss Jordan | Privada | 25000 | DVORAK NO. 59-COL. VALLEJO-GUSTAVO A. MADERO-07870 | TURISMO |
| Ruta | Universidad Insurgentes Norte | Privada | 50000 | AVENIDA ACUEDUCTO NO. 13-COL. SAN PEDRO ZACATENCO-GUSTAVO A. MADERO-07360 | MERCADOTECNIA |
| Ruta | Universidad Del Tepeyac | Privada | 10200 | CALLAO NO. 802-COL. LINDAVISTA-GUSTAVO A. MADERO-7300 | MERCADOTECNIA |
| Ruta | Universidad Tecnologica Roosevelt | Privada | 0 | CALZADA DE LOS MISTERIOS NO. 801-COL. TEPEYAC INSURGENTES-GUSTAVO A. MADERO-07020 | MERCADOTECNIA |
| Ruta | Universidad Justo Sierra Ticoman | Privada | 65000 | AV. ACUEDUCTO 914-COL. LAGUNA TICOMAN-GUSTAVO A. MADERO-7340 | MERCADOTECNIA |
| Ruta | Escuela Benjamin Franklin | Privada | 30000 | INSURGENTES NORTE NO. 1423-COL. GUADALUPE INSURGENTES-GUSTAVO A. MADERO-07870 | MERCADOTECNIA |
| Ruta | Escuela Superior de Economia | Publica | 0 | PLAN DE AGUA PRIETA 66-COL. PLUTARCO ELIAS CALLES-MIGUEL HIDALGO-11350 | ECONOMIA |
| Ruta | Universidad De Ciencias Y Administracion | 0 | 0 | GUILLERMO MASSIEU HELGUERA NO. 103-COL. SAN JOSE TICOMAN-GUSTAVO A. MADERO-7320 | CONTADURIA_Y_FINANZAS |
| Ruta | Universidad Del Tepeyac | Privada | 10200 | CALLAO NO. 802-COL. LINDAVISTA-GUSTAVO A. MADERO-7300 | CONTADURIA_Y_FINANZAS |
| Ruta | Universidad Del Tepeyac | Privada | 10200 | CALLAO NO. 802-COL. LINDAVISTA-GUSTAVO A. MADERO-7300 | DISENO_GRAFICO |

Figura 4.20 Resultado de la consulta de un concepto sin instancias en sus subconceptos ni súper-conceptos.

Los resultados obtenidos en cada una de las consultas tratan de mantenerse lo más cerca posible del concepto buscado, inicialmente debido a que entre más nos alejamos del mismo, mayor será el número de elementos que tendríamos que analizar.

CAPÍTULO 5. CONCLUSIONES Y TRABAJOS FUTUROS

5. Conclusiones y trabajos futuros

A continuación se describen las principales conclusiones en torno a los resultados obtenidos con el presente trabajo de investigación, enmarcado en el ámbito de la recuperación de información geográfica (GIR), con especial enfoque en la recuperación semántica de información geoespacial, así como temas de investigación futuros.

1.1 Conclusiones

- La recuperación semántica de información como un área pujante en el ámbito de la investigación facilita el procesamiento de grandes volúmenes de información, de los cuales nos podemos guiar para poder implementar o mejorar las técnicas para hacer la medición de la distancia semántica entre conceptos, o bien, de la similitud semántica en general.
- La definición del posible uso que se le dará a la información tanto geográfica como descriptiva que pretendemos almacenar, es un paso esencial al momento de diseñar nuestra ontología, ya que de ello dependerá la manera en que podamos recuperar o inferir dicha información.
- La técnica de almacenamiento de la información geográfica que deseamos utilizar en cualquier SIG, tendrá que ser el adecuado dependiendo del tipo de información que queramos manejar en el sistema, en el caso particular de esta aplicación de recuperación se almacenaron directamente en la ontología, debido a que sólo se guardó latitud/longitud, en caso de que se requieran almacenar polígonos, líneas o algunas relaciones topológicas, debería analizarse si es posible hacerlo en una ontología o es necesario relacionar la información descriptiva con una base de datos geográfica.

5. Conclusiones y trabajos futuros

- Con la implementación de la metodología propuesta se incorporan varias herramientas que se encuentran en pleno desarrollo, como lo son los servicios Web, los cuales proponen un gran avance en la incorporación de aplicaciones que desean compartir sus recursos, esto implica la posibilidad de evitar la redundancia de información y el crecimiento colaborativo de los mismos. La API de Google Maps es otro ejemplo de estas herramientas, la cual cada día presenta nuevas funcionalidades, como el marcado de sitios de interés, la geocodificación, entre muchas otras.
- La funcionalidad de los servicios Web está basada en gran medida en la disponibilidad de los recursos que tenemos. En nuestro caso al manejarlos mediante un conjunto de métodos directamente desde la Web, nos abrirían las puertas para hacer la integración de grandes volúmenes de información de diferentes colaboradores.

1.1 Aportaciones

- Metodología para la recuperación semántica de información geográfica que permite inferir información adicional
- Se desarrolló una ontología de aplicación enfocada en Educación Superior
- La recuperación se realiza exclusivamente sobre la representación conceptual utilizando el algoritmo de trayectoria de conceptos
- Se implementó el servicio WSDL
- Se desarrolló un caso de estudio para instituciones educativas de nivel superior de la delegación Gustavo A. Madero

1.2 Limitaciones

- Un servicio Web por lo general debe tener las funcionalidades que permitan hacer el registro, actualización y recuperación de entidades (sistema de administración), en nuestro caso, sólo permite hacer la recuperación.
- El análisis semántico se limita al algoritmo de trayectoria de conceptos
- Es complicado actualizar la representación conceptual

1.3 Trabajos futuros

- Agregar otras medidas de similitud, las cuales permitirían hacer una recuperación más adecuada a lo que desea el usuario.
- Dotar al servicio Web de un sistema de administración que permita dar de alta y actualización entidades dentro de nuestro repositorio de información, así como el registro de nuevos conceptos dentro de la ontología, lo que nos permitiría tener un repositorio de información integral.
- Explotar las capacidades del diseño de la ontología, cargando entidades no sólo de instituciones de nivel superior, sino entidades de distintos ámbitos como son, comercios, educación en general, instituciones de gobierno, entre muchas otras.
- Agregar otros métodos de inferencia para la generación de conocimiento
- Enfocar este sistema a una aplicación *question answering* (QA).

5. Conclusiones y trabajos futuros

Referencias

Atkinson, et al. (1996)

- Atkinson, M. P., Daynes, L., Jordan, M. J., Printezis, T., & Spence, S. (1996). An orthogonally persistent Java. *ACM Sigmod Record*, 25(4), 68-75.

Baeza-Yates and Ribeiro-Neto, (1999)

- Baeza-Yates, R., & Ribeiro-Neto, B. (1999). *Modern information retrieval* (Vol. 463). New York: ACM press.

Berners-Lee, et al. (2001)

- Berners-Lee, T. Hendler, J. and Lassila, O. "The Semantic Web: A new form of Web content that is meaningful to computers will unleash a revolution of new possibilities", *Scientific American*. 2001.

Blanchard, et al. (2005)

- Blanchard, E., Harzallah, M., Briand, H., & Kuntz, P. (2005). A Typology Of Ontology-Based Semantic Measures. In EMOI-INTEROP.

Brisaboa, et al. (2010)

- Brisaboa, N. R., Luaces, M. R., Places, Á. S., & Seco, D. (2010). Exploiting geographic references of documents in a geographical information retrieval system using an ontology-based index. *Geoinformatica*, 14(3), 307-331.

Cabral, et al. (2006)

- Cabral, L., Domingue, J., Galizia, S., Gugliotta, A., Tanasescu, V., Pedrinaci, C., & Norton, B. (2006). IRS-III: a broker for semantic web services based applications. In *The Semantic Web-ISWC 2006* (pp. 201-214). Springer Berlin Heidelberg.

Castells, (2004)

- Castells, P. (2004). Neptuno: tecnologías de la web semántica para una

hemeroteca digital. Article Llarg. Interacci3n 2004, Lleida. Maig 2004.

Centeno, et al. (2007)

- Centeno, V. L., Kloos, C. D., Del Toro, J. M. B., & Gaedke, M. (2007). Web accessibility evaluation via XSLT. In *Web Information Systems Engineering–WISE 2007 Workshops* (pp. 459-469). Springer Berlin Heidelberg.

Ch3vez, et al. (2005)

- Ch3vez, M. E., C3rdenas, O., and Benito, O. (2005) La Web Sem3ntica. *Risi* 2(3), 43-54.

Corcho, et al. (2003)

- Corcho, O., Fern3ndez-L3pez, M., G3mez-P3rez, A. (2003). “Methodologies, tools and languages for building ontologies. Where is their meeting point ?”, *Data & Knowledge Engineering* 46:41-64, 2003.

De Bruijn, et al. (2005)

- De Bruijn, J., Lausen, H., Polleres, A., and Fensel, D. (2005) The Web service modeling language WSML: an overview, DERI technical report 2005-06-16. Available at: <http://www.wsmo.org/wsml/wsml-resources/deri-tr-2005-06-16.pdf>

Declerck, et al. (2004)

- Declerck, T., Contreras, J., Corcho, 3., & Crispi, C. (2004). Text-based Semantic Annotation Service for Multimedia Content in the Esperanto project.

Ding, et al. (2004)

- Ding, L., Finin, T., Joshi, A., Pan, R., Cost, R. S., Peng, Y., & Sachs, J. (2004). Swoogle: A semantic web search and metadata engine. In *Proc. 13th ACM Conf. on Information and Knowledge Management* (pp. 65-659).

Dominich, (2008)

- Dominich, S. (2008) *The Modern Algebra of Information Retrieval*. Springer, Berlin, 1 edition, 2008.

Egenhofer (2002)

- Egenhofer, M.J. (2002). Toward the Semantic Geospatial Web. In Proceedings of the Tenth ACM International Symposium on Advances in Geographic Information Systems, McLean, Virginia.

Fensel (2003)

- Fensel, D. (2003). Ontologies: Silver Bullet for Knowledge Management and Electronic Commerce, Springer-Verlag, Berlin, 2001. 2nd Edition, Springer-Verlag, Berlin, 2003

Fonseca and Davis, (1999)

- Fonseca, F. and Davis C. (1999). Using the Internet to Access Geographic Information: An Open GIS Prototype. Interoperating Geographic Information Systems. Springer US, 1999. 313-324.

Fonseca, et al. (2002)

- Fonseca, F., Egenhofer, M. J., Agouris, P. and Câmara, G. (2002), Using Ontologies for Integrated Geographic Information Systems. Transactions in GIS, 6: 231–257. doi: 10.1111/1467-9671.00109

Fonseca and Sheth, (2002)

- Fonseca, F., and Sheth, A. (2002). The Geospatial Semantic Web. *The Handbook of Geographic Information Science*, 367-376.

Giraldo, et al. (2006)

- Giraldo, J. D., Luna, J. A. G., & Ceballos, J. C. (2006). Desarrollo de un módulo de visualización y navegación de información basada en ontologías para el sistema de recuperación semántica SABIOS. *Revista Avances en Sistemas e Informática*. Jun. vol. 3, no. 1; p. 57–61.

Goldstone and Son, (2005)

- Goldstone, R. L. and Son, J. (2005). Similarity. In Holyoak K and Morrison R (ed) *Cambridge Handbook of Thinking and Reasoning*. Cambridge,

Cambridge University Press: 13–36.

Gómez-Pérez, (1997)

- Gómez-Pérez A. (1997). “Ontological Engineering: A State of The Art”, Facultad de Informática, Universidad Politécnica de Madrid, 1997.

Gone and Shade, (2008)

- Gone, M. and Shade, S. (2008). “Towards semantic composition of geospatial web services using WSMO in comparison to BPEL”. *International Journal of Spatial Data Infrastructures Research*, Vol. 3, 192-214

Gonzáles, (2009)

- Gonzáles, J. (2009), *Uso de Información semántica para la mejora de la recuperación de información en la web*. Valencia, España: Universidad de Sevilla.

Googchild, et al. (1999)

- Googchild M, Egenhofer M, Fegeas R, and Kottman C, (Eds) *Interoperating Geographic Information Systems*. pp. 313-324, Kluwer Academic Publisher, Norwel, MA

Griber, (1995)

Gruber, T.R., (1995,) *Toward principles for the design of ontologies used for knowledge sharing*. *International Journal of Human–Computer Studies*, 43, pp. 907–928.

Heymans, et al. (2008)

- Heymans, S., Ma, L., Anicic, D., Ma, Z., Steinmetz, N., Pan, Y., & Keller, U. (2008). *Ontology reasoning with large data repositories*. In *Ontology Management* (pp. 89-128). Springer US.

Hiramatsu and Reitsma, (2004)

- Hiramatsu, K., & Reitsma, F. (2004). *GeoReferencing the Semantic Web: ontology based markup of geographically referenced information*. In *Joint*

EuroSDR/EuroGeographics workshop on Ontologies and Schema Translation Services (pp. 15-16).

Hyvonen, et al. (2003)

- Hyvonen, E., Saarela, S., Viljanen, K. (2003). Ontogator: Combining view- and ontology-based search with semantic browsing, in 'Proceedings of XML Finland 2003, Open Standards, XML and the Public Sector, Kuopio.

Janowicz, et al. (2001)

- Janowicz, K., Raubal, M. and Kuhn, W. (2001). The semantics of similarity in geographic information retrieval. *Journal of Spatial Information Science*, 2: 29–57.

Jones, (2008)

- Jones, C.B. (2008). and R.S. Purves. Geographical information retrieval. *International Journal of Geographical Information Science*, 22: 219–228.

Lutz and Klien, (2006)

- Lutz M. and Klien, E. (2006). Ontology-based retrieval of geographic information, *International Journal of Geographic Information Science*, Vol. 20(3):233–260.

Maedche and Staab, (2002)

- Maedche, A. and Staab, S. (2002). Measuring similarity between ontologies. In *Knowledge Engineering and Knowledge Management: Ontologies and the Semantic Web*, number 2473 in *Lecture Notes in Computer Science*, pages 251-263. Springer.

Manning, et al. (2008)

- Manning, C. D., Raghavan, P., & Schütze, H. (2008). *Introduction to information retrieval* (Vol. 1). Cambridge: Cambridge University Press.

Ortega, et al. (2008)

- Ortega, J. M. P., Cumberras, M. A. G., Vega, M. G., & López, L. A. U. (2008).

Sistemas de Recuperación de Información Geográfica multilingües en CLEF. *Impresos de Inscripción*, 40, 129-136.

Peis, et al. (2003)

- Peis, E., Herrera-Viedma, E., Hassan, Y., and Herrera, J. C. (2003): Ontologías, metadatos y agentes: Recuperación semántica de la información. In: Proc. Tratamiento y Recuperación de la Información, España, vol. (1), pp. 157–165.

Rodríguez and Egenhofer, (2002)

- Rodríguez, M.A. and Egenhofer, M.J. (2002.) Determining semantic similarity among entity classes from different ontologies. *IEEE Transactions on Knowledge and Data Engineering*.

Rodríguez, et al. (1999)

- Rodríguez, M.A., Egenhofer, M.J., and Rugg, R. (1999). Assessing Semantic Similarities Among Geospatial Feature Class Definitions, in: A. Vckovski, K. Brassel, and H.-J. Schek, Eds., *Interoperating Geographic Information Systems—Second International Conference, INTEROP'99*, Zurich, Switzerland, vol. 1580, *Lecture Notes in Computer Science*, Springer-Verlag, pp. 189-202.

Rodríguez and Ronda León, (2005)

- Rodríguez Perojo, K. and Ronda León R. (2005) Web semántica: un nuevo enfoque para la organización, la recuperación de la información en la Web. *ACIMED*, vol. 13, no. 6.

Roman, (2005)

- Roman, D. (2005) “Web Service Modeling Ontology, Applied Ontology”, IOS Press, vol. 1, no. 1.

Sarabia López, (2008)

- Sarabia López, G. Búsqueda y Ponderación de información Contenida en Bases de Datos Espaciales, Utilizando Jerarquías. México D.F., Presentada

en el Centro de Investigación en Computación CIC-IPN para obtención del grado de Maestro en ciencias de la Computación.

Schwering, (2008)

- Schwering, A. (2008). Approaches to semantic similarity measurement for geo-spatial data - a survey. *Transactions in GIS*, 12(1):5–29.

Shen, et al. (2010)

- Shen, L., Duan, W., Ren, Y., & Yang, C. (2010). Management Service on WMS/WFS services. In *Geoinformatics, 2010 18th International Conference on* (pp. 1-5). IEEE.

Studer, et al. (1998)

- STUDER, R., BENJAMINS, V.R. and FENSEL, D., (1998), Knowledge engineering: principles and methods. *Data and Knowledge Engineering*, 25, pp. 161–197.

Tomlinson, (2007)

- Tomlinson, R. (2007). *Pensando en sig. Planificación del sistema de información geográfica dirigida a gerentes*. USA. ESRI press.

Tramullas, (1999)

- Tramullas, J. 1999, Agentes y ontologías para el tratamiento de información: clasificación y recuperación en Internet., 1999. In *Actas del IV Congreso ISKO-España, Granada, ISKO España*. pp.247-252.

Tversky, (1977)

- Tversky, A. (1977). Features of similarity. *Psychological review*, 84(4), 327.

Uschold and Gruninger, (1996)

- Uschold, M., & Gruninger, M. (1996). Ontologies: Principles, methods and applications. *Knowledge engineering review*, 11(2), 93-136.

Vallet, et al. (2005)

- Vallet, D., Fernández, M., Castells, P. (2005). *An Ontology-Based Information*

Retrieval Model. 2nd European Semantic Web Conference. LNCS Vol. 3532 455-470.

Van Kreveld, et al. (2005)

- Van Kreveld, M., Reinbacher, I., Arampatzis, A., & Van Zwol, R. (2005). Multi-Dimensional Scattered Ranking Methods for Geographic Information Retrieval*. *GeoInformatica*, 9(1), 61-84.

Visser, et al. (2002)

- Visser, U., Stuckenschmidt, H., Schuster, G., Vogele, T.(2002). Ontologies for geographic information processing. *Computers and Geoscience* 28, pp. 103–117.

Wache, et al. (2001)

- Wache, H., Vögele, T., Visser, U., Stuckenschmidt, H., Schuster, G., Neumann, H., and Hübner, S. (2001). Ontology-based integration of information—a survey of existing approaches. In *IJCAI-01 Workshop: Ontologies and Information Sharing*, Seattle, WA, pp. 108–117.

Zapater, (2005)

- Zapater, J. J. (2005). *Ontologías para servicios web semánticos de información de tráfico: descripción y herramientas de explotación*. Valencia, España: Universidad de Valencia.

Zhao, et al. (2008)

- Zhao, T., Zhang, C., Wei, M., Peng, Z.-R., 2008. Ontology-based geospatial data query and integration. *Lecture Notes in Computer Science LNCS5266: Geographic Information Science* 5266, 370–392.

Zhu, et al. (2010)

- Zhu L., Ma Q., Liu C., Mao G. and Yang W.(2010). “Semantic-distance based evaluation of ranking queries over relational databases,” *Journal of Intelligent Information Systems*, Vol. 35, No. 3, pp. 415-445.

Berners-Lee, Tim. (1998). Semantic Web Road Map. Disponible en:
<http://www.w3.org/DesignIssues/Semantic.html>.

CORBA. (s.f) <http://www.corba.org/corba-e/index.htm>

Geonames. (s.f) <http://www.geonames.org/>

Jena (s.f) <http://jena.apache.org/documentation/>

KARTOO (2004), 'Kartoo, versión 4'. *En línea: <<http://www.kartoo.com>> C02/04

OASIS. (s.f). Recuperado el 3 septiembre de 2013, <https://www.oasis-open.org/>

PROTEGÉ. (s.f) <http://protege.stanford.edu/>
<http://jena.apache.org/documentation/>

RMI. (s.f) <http://docs.oracle.com/javase/tutorial/rmi/overview.html>)

W3C. (s.f). Recuperado el 07 de Julio de 2013 <http://www.w3.org/Consortium/>

SIRVOES (s.f). <http://www.sirvoes.sep.gob.mx/sirvoes/ServletFiltro>