



INSTITUTO POLITÉCNICO NACIONAL
CENTRO DE INVESTIGACIÓN EN COMPUTACIÓN
LABORATORIO DE PROCESAMIENTO INTELIGENTE DE
INFORMACIÓN GEOESPACIAL



Recuperación semántica personalizada de información
geográfica

T E S I S

QUE PARA OBTENER EL GRADO DE

MAESTRO EN CIENCIAS DE LA COMPUTACIÓN

PRESENTA:

IVAN RUIZ SALAZAR

DIRECTORES DE TESIS:

Dr. MARCO ANTONIO MORENO IBARRA

Dr. JOSÉ GIOVANNI GUZMÁN LUGO

MÉXICO, D.F.

Julio 2012



INSTITUTO POLITÉCNICO NACIONAL
SECRETARÍA DE INVESTIGACIÓN Y POSGRADO

SIP-14 bis

ACTA DE REVISIÓN DE TESIS

En la Ciudad de México, D.F. siendo las 08:00 horas del día 08 del mes de Junio de 2012 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:

"Recuperación semántica personalizada de información geográfica"

Presentada por el alumno:

RUIZ

Apellido paterno

SALAZAR

Apellido materno

IVÁN

Nombre(s)

Con registro:

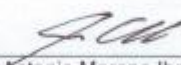
A	1	0	0	3	3	0
---	---	---	---	---	---	---

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

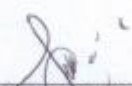
Después de intercambiar opiniones los miembros de la Comisión manifestaron **APROBAR LA TESIS**, en virtud de que satisface los requisitos señalados por las disposiciones reglamentarias vigentes.

LA COMISIÓN REVISORA

Directores de Tesis


Dr. Marco Antonio Moreno Ibarra


Dr. José Giovanni Guzmán Lugo


Dr. Cornelio Yáñez Márquez

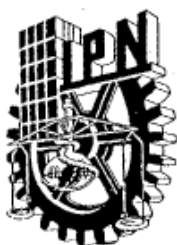

Dr. Rolando Quintero Téllez


Dr. Miguel Jesús Torres Ruiz

PRESIDENTE DEL COLEGIO DE PROFESORES


Dr. Luis Alfonso Villa Vargas

DIRECCIÓN



INSTITUTO POLITÉCNICO NACIONAL
SECRETARÍA DE INVESTIGACIÓN Y POSGRADO

CARTA CESIÓN DE DERECHOS

En la Ciudad de *México* el día *8* del mes de *Junio* del año *2012*, el que suscribe ***Ivan Ruiz Salazar*** alumno del Programa de *Maestría en Ciencias de la Computación* con número de registro *A100330*, adscrito al *Centro de Investigación en Computación*, manifiesta que es autor intelectual del presente trabajo de Tesis bajo la dirección del ***Dr. Marco Antonio Moreno Ibarra*** y del ***Dr. José Giovanni Guzmán Lugo***, y cede los derechos del trabajo intitulado *Recuperación semántica personalizada 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 *ivanruizs@gmail.com*. Si el permiso se otorga, el usuario deberá dar el agradecimiento correspondiente y citar la fuente del mismo.

Ivan Ruiz Salazar,

Nombre y firma

AGRADECIMIENTOS

A mi Mami, por su: Amor, Bondad, Tolerancia, Enseñanza, Risa y Apoyo.

A mis Hermanos Dany y Adan por ser mis amigos de toda una vida, por extenderme su mano cuando los necesito, por estar presentes en los mejores y en los peores momentos de mi vida.

A mi Padre por su esfuerzo para apoyarme.

A mis Abuelitos Esperanza, Eliseo†, Juanita† y Manuel†, por ser pilares en mi formación y por los momentos tan felices que pasamos juntos.

A mis Directores de Tesis por su confianza y guía en el transcurso de estos años, en los cuales con paciencia y dedicación compartieron gran parte de su esencia y conocimientos.

A los Bonobos por su amistad y sus atinados consejos.

A mis amigos y familiares, por aceptarme como soy y compartir un sinfín de experiencias inolvidables.

A Docentes y personal Administrativo del CIC, porque en este tiempo han pasado a ser amigos y cómplices de mi desarrollo profesional.

Al CIC por abrirme las puertas a este mundo de la investigación y proporcionarme las herramientas para seguir cumpliendo mis metas profesionales.

Al CONACYT por el apoyo sin igual para el desarrollo de este trabajo de investigación.

Al IPN por acobijarme estos ya casi 10 años.

RESUMEN

Esta tesis presenta una metodología para personalizar la recuperación de información geográfica con un enfoque semántico. Hoy en día el papel del enfoque semántico dentro de la gestión de información está en total crecimiento, donde el significado se asocia con los datos y los términos utilizados en las consultas. La investigación dentro de esta rama de la computación para la recuperación de información geográfica, a dado lugar a proyectos que realizan esfuerzos para contribuir a resolver problemas como la similitud entre objetos geográficos, la relevancia de objetos geográficos dentro de la recuperación, el enriquecimiento de la recuperación con base en un contexto, etc. Por otro lado, las preferencias del usuario en trabajos recientes, se ven utilizadas para mejorar la recuperación de la información, debido a que la personalización juega un factor fundamental. Estas se utilizan para delimitar las necesidades del sistema, para que pueda tomar decisiones y entregar resultados útiles al usuario, acotando un dominio de trabajo con base en sus preferencias y necesidades, evitando resultados sin sentido y trabajo extra que no lleva a ningún resultado para los fines del usuario.

La metodología se divide en tres etapas, la primera de ellas es la Conceptualización, la cual consiste en construir una ontología donde se identifican las propiedades, las relaciones y las entidades involucradas, con lo que se realiza la descripción del dominio, la siguiente etapa es la Caracterización del usuario, dedicada a la identificación de conceptos relevantes dentro de la ontología, para definir las preferencias del usuario, dando pie a la generación de su perfil con base en la conceptualización. Por último tenemos la etapa de Recuperación, centrada en la búsqueda de los objetos geográficos que son relevantes para el usuario con base en su perfil y en su ubicación. La investigación es aplicada en un sistema que toma datos del INEGI, y retorna resultados generados por la implantación de la metodología propuesta en este trabajo. Por último se evalúan los resultados generados y se dan conclusiones con base en todo el trabajo de investigación.

ABSTRACT

This dissertation presents a method for geographic information retrieval based on user-profiling, with a semantic approach. Up-to-date, the role of the semantic approach in information management is increasingly, in which the meaning is associated with data and the terms used in queries. The research in field of geographic information retrieval, has given rise to projects that are making efforts to solve problems such as the similarity between geographic objects, the relevance of geographic information within the recovery, enriching of the retrieval based on a context, and so on. Moreover, the user preferences in recent works are used to improve information retrieval, because the profiling plays a crucial issue in this field. These are used to define their needs to the system, so that it can make decisions and deliver useful results to the user, delimiting a domain of work based on their preferences and needs, avoiding meaningless results and extra work that does not lead to any results for the user's purpose.

The methodology is composed of three stages: the first is the conceptualization, which consists of building an ontology that identifies properties, relationships and the entities involved, thus making the description of the domain, the second stage is the characterization of the user, which is focused on the identification of relevant concepts in the ontology in order to define user preferences, giving rise to the generation of a profile in relation to the conceptualization made in the first stage. Finally, we have the retrieval stage centered in the search of geographical objects that are relevant to the user, according to their profile and location.

The approach is applied in a system that takes data from INEGI, and returns results based on the methodology proposed in this work. Finally, we evaluate the generated results and conclusions are given based on the entire research work.

ÍNDICE

RESUMEN	IX
ABSTRACT	XI
ÍNDICE.....	XIII
ÍNDICE DE FIGURAS	XVII
ÍNDICE DE TABLAS	XIX
CAPÍTULO 1 INTRODUCCIÓN.....	1
1.1 DESCRIPCIÓN DEL PROBLEMA	2
1.2 OBJETIVO.....	4
1.3 METAS	4
1.4 HIPÓTESIS	5
1.5 ALCANCES	5
1.6 ORGANIZACIÓN DEL DOCUMENTO	6
CAPÍTULO 2 ESTADO DEL ARTE	7
2.1 RECUPERACIÓN DE LA INFORMACIÓN	7
2.1.1 <i>El problema de la IR</i>	7
2.1.2 <i>Arquitectura de un sistema IR</i>	8
2.1.3 <i>Modelos de IR</i>	10
2.1.3.1 <i>Caracterización de un modelo en IR</i>	10
2.1.3.2 <i>Clasificación de los modelos de IR</i>	11
2.2 RECUPERACIÓN DE INFORMACIÓN GEOGRÁFICA	12
2.2.1 <i>Análisis de la Consulta</i>	13
2.2.2 <i>Módulo de IR</i>	14
2.2.3 <i>Ranking</i>	14
2.3 SIMILITUD SEMÁNTICA EN LOS GIR.....	15
2.3.1 <i>Análisis semántico</i>	16
2.3.1.1 <i>Semántica Geoespacial</i>	16
2.3.2 <i>Similitud Semántica</i>	16
2.3.2.1 <i>Funciones de Similitud</i>	17

2.3.2.2 La Similitud en la Recuperación de Información basada en la Semántica...	18
2.3.2.3 Recuperación basada en Ontologías	20
2.4 PERSONALIZACIÓN.....	21
2.4.1 Perfil de Usuario	22
2.4.2 Diseño centrado en el usuario.....	22
2.5 DISCUSIÓN SOBRE EL ESTADO DEL ARTE	24
CAPÍTULO 3 METODOLOGÍA	27
3.1 MARCO DE TRABAJO	27
3.2 CONCEPTUALIZACIÓN.....	30
3.2.1 Jerarquía de conceptos.....	30
3.3 CARACTERIZACIÓN	36
3.3.1 Implementación del Perfil de Usuario.....	37
3.4 RECUPERACIÓN.....	38
3.4.1 Arquitectura.....	38
3.4.1.1 Expansión de la Consulta	40
3.4.1.2 Base de Datos Espacial.....	43
3.4.1.3 Vecindad.....	44
CAPÍTULO 4 IMPLANTACIÓN, PRUEBAS Y RESULTADOS	47
4.1 CASO DE ESTUDIO.....	47
4.2 CONCEPTUALIZACIÓN: ONTOLOGÍA DE APLICACIÓN	47
4.2.1 Lenguaje de Representación	48
4.2.2 Implementación de la Ontología	49
4.3 CARACTERIZACIÓN: IMPLEMENTACIÓN DEL PERFIL DE USUARIO	53
4.4 RECUPERACIÓN:IMPLEMENTACIÓN DE UBICACIÓN, BDE, SIMILITUD Y RESULTADOS	55
4.4.1 Ubicación del Usuario	56
4.4.2 Base de Datos Espacial	57
4.4.3 Implementación de la Similitud	60
4.4.4 Expansión de la consulta.....	61
4.5 APLICACIÓN.....	63
4.5.1 Interfaz Gráfica de Usuario	64
4.5.2 Caso de Uso.....	65
4.6 ANÁLISIS DE PRUEBAS Y RESULTADOS DE LA IMPLANTACIÓN DE LA METODOLOGÍA..	68
4.6.1 Personalización y Semántica en las Pruebas y Resultados.....	69
4.6.2 Precision y Recall.....	74
4.6.3 Experimentos Generales.....	77

CAPÍTULO 5 CONCLUSIONES Y TRABAJO FUTURO	81
5.1 CONCLUSIONES.....	81
5.2 CONTRIBUCIONES	83
5.3 TRABAJOS FUTUROS.....	84
REFERENCIAS	87
ANEXO AMARCO TEÓRICO	95
A.1 PRECISION Y RECALL.....	95
A.2 CLASIFICACIÓN DE LAS ONTOLOGÍAS.....	96
A.3 OWL	97
A.4 BDE	98
A.5 DENUÉ.....	99
A.6 SCIAN	100
A.7 PROTÉGÉ	100
A.8 GOOGLE MAPS Y FUSION TABLES	101
A.9 ZK FRAMEWORK	102
A.10 HERMIT	103
A.11 RESNIK: SIMILITUD SEMÁNTICA.....	104
ANEXO B CÓDIGO FUENTE	105

ÍNDICE DE FIGURAS

CAPÍTULO 1 INTRODUCCIÓN

Figura 1.1 Problema de contexto, coherencia, relevancia y preferencias en la GIR	2
Figura 1.2 Problema de distancia en los resultados de la GIR con relación al usuario	3

CAPÍTULO 2 ESTADO DEL ARTE

Figura 2.1 Arquitectura de un sistema de IR	9
Figura 2.2 La función de ranking $R(q_i, d_j)$	11
Figura 2.3 Clasificación de los IR	12
Figura 2.4 Arquitectura general de un sistema GIR	13
Figura 2.5 Recuperación basada en subsumción y similitud	19
Figura 2.6 Metodología basada en una ontología para la GIR	21
Figura 2.7 Actividades del diseño centrado en el usuario	23

CAPÍTULO 3 METODOLOGÍA

Figura 3.1 Entrada y salida de la Metodología	28
Figura 3.2 Metodología	29
Figura 3.3 Etapa 1: Conceptualización	31
Figura 3.4 Estructura del SCIAN.	32
Figura 3.5 Estructurade Actividades	34
Figura 3.6 Ejemplo de Categoría	35
Figura 3.7 Perfil de Usuario formado por conceptos de la ontología	36
Figura 3.8 Arquitectura de la etapa de recuperación	38
Figura 3.9 Conceptos similares con base en la jerarquía de conceptos.	40
Figura 3.10 Método de expansión en la recuperación.	41
Figura 3.11 Jerarquía de conceptos ejemplo	42
Figura 3.12 Retorno de la función $Expansión_Conceptos(P_1)$	42
Figura 3.13 Recuperación de objetos geográficos con base en una vecindad.	44

CAPÍTULO 4 IMPLANTACIÓN Y RESULTADOS

Figura 4.1 Lenguajes de marcado para ontologías.	48
Figura 4.2 Desarrollo de la Ontología en Protégé	50
Figura 4.3 Relación <i>es-un</i> entre conceptos de la ontología	51
Figura 4.4 Código en OWL de la Ontología	51
Figura 4.5 Representación de conceptos usados en la Ontología.....	53
Figura 4.6 Perfil de Usuario formado por conceptos de la ontología.....	54
Figura 4.7 Perfiles de Usuario en la implementación.....	54
Figura 4.8 Visualización de conceptos en la implementación	55
Figura 4.9 Sistema de coordenadas de Google Maps.....	56
Figura 4.10 Ubicación del usuario en la Implementación	57
Figura 4.11 Datos del DENUE delegación Gustavo A. Madero en GvSIG	58
Figura 4.12 Base de Datos de atributos del DENUE.....	59
Figura 4.13 Base de Datos en Google Fusion Tables.....	59
Figura 4.14 Salida en consola de los conceptos por similitud en la ontología	61
Figura 4.15 Interfaz de la implementación de la recuperación	62
Figura 4.16 Implementación de la Metodología.....	63
Figura 4.17 Etapas de la Metodología en la Implementación	64
Figura 4.18 Selección de Ubicación en la Implementación	65
Figura 4.19 Selección de Perfil de Usuario Estudiante	66
Figura 4.20 Resultados de la implementación de la Recuperación	66
Figura 4.21 Visualización de resultados perfil Estudiante	67
Figura 4.22 Visualización de resultados perfil Profesor.....	67
Figura 4.23 Análisis resultados perfil Estudiante	69
Figura 4.24 Análisis resultados perfil Profesor.	70
Figura 4.25 Resultados de la expansión de conceptos	72
Figura 4.26 Objetos geográficos recuperados con base en la expansión de conceptos	72
Figura 4.27 Vecindad en la recuperación	73
Figura 4.28 Datos para la evaluación de la recuperación	75
Figura 4.29 Gráfica de evaluación de <i>Precision</i>	76

Figura 4.30 Gráfica de evaluación de <i>Recall</i>	77
Figura 4.31 Experimentos generales.	78
Figura 4.32 Gráfica de evaluación de <i>Precision</i> para experimentos generales	79
Figura 4.33 Gráfica de evaluación de <i>Recall</i> para experimentos generales	79

ÍNDICE DE TABLAS

CAPÍTULO 3 METODOLOGÍA

Figura 3.1 Número de Categorías del SCIAN.....	33
Figura 3.2 Actividades del SCIAN.....	33
Figura 3.3 Diccionario de datos DENUÉ (fragmento).....	44

CAPÍTULO 4 IMPLANTACIÓN Y RESULTADOS

Tabla 4.1 Expansión de Conceptos.....	71
Tabla 4.2 Objetos geográficos recuperados por concepto del perfil Estudiante.	74

CAPÍTULO 1

INTRODUCCIÓN

La Recuperación de Información Geográfica (Geographic Information Retrieval, GIR) ha sido y sigue siendo mejorada [Jones, Purves, 2008]. Hoy en día el papel del enfoque semántico dentro de la gestión de información está en total crecimiento, donde el significado se asocia con los datos y los términos utilizados en las consultas [Baeza-Yates, 2011]. La GIR, está en gran parte basada en medidas métricas (i.e. cuantitativas), pero la gente suele expresar y comprender las relaciones espaciales a través del lenguaje natural [Montello, et al., 2003]. La similitud semántica está estrechamente relacionada con una rama de la Inteligencia Artificial, las *ontologías*, las cuales aplicadas a la recuperación de la información han mostrado mejorar los resultados para el usuario [Janowicz, et. al., 2011]. La investigación dentro de esta rama para la GIR a dado lugar a proyectos [Gracia-Cumbreras, et. al., 2009] [Zhan, et. al., 2008] que realizan esfuerzos para contribuir a resolver problemas como la similitud entre objetos geográficos, la relevancia de objetos geográficos dentro de la recuperación, la eliminación de resultados erróneos, enriquecimiento de la recuperación con base en un contexto, etc.

Por otro lado, la investigación sobre Perfiles de Usuario (User Profiling, UP), se ha extendido en la Recuperación de la Información (Information Retrieval, IR), debido a que la personalización juega un factor fundamental [Challam, 2007] [Wealands, 2007]. Las preferencias del usuario se utilizan para delimitar sus necesidades al sistema, para que este pueda tomar decisiones y entregar resultados útiles al usuario, acotando un dominio de trabajo con base en sus preferencias y necesidades, evitando resultados sin sentido y trabajo extra que no lleva a ningún resultado para los fines del usuario.

En este trabajo se propone una metodología para personalizar la Recuperación de Información Geográfica con un enfoque semántico. Teniendo como base la conceptualización del dominio con una ontología y tomando las preferencias del usuario para ofrecer resultados relevantes y coherentes sobre un contexto bien definido. Ademástomando en cuenta la posición del usuario, donde las preguntas: ¿Quién? y ¿Dónde?, personalizan la recuperación.

Como caso de estudio de la presente tesis se tomaron datos del Directorio Estadístico Nacional de Unidades Económicas, previamente organizados.

1.1 Descripción del problema

En la actualidad, cuando un usuario realiza una consulta en los distintos buscadores de información, y más particularmente en sistemas de información geográfica, generalmente se introducen términos o conceptos relacionados con la información que se desea obtener tras ejecutar la consulta.

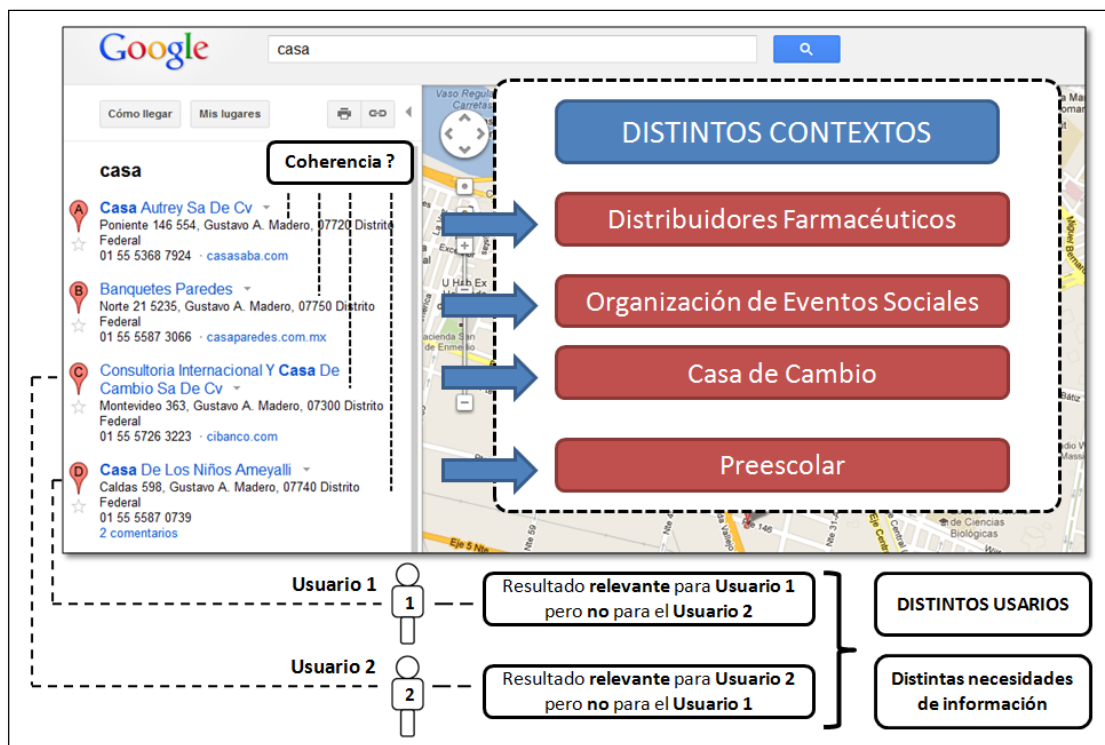


Figura 1.1 Problema de contexto, coherencia, relevancia y preferencias en la GIR.

Estos conceptos son analizados y buscados dentro de una colección de objetos geográficos y se devuelven al usuario generalmente los objetos geográficos que coinciden con base en la sintaxis de los términos de la consulta.

En los resultados, se pueden observar problemas como: objetos geográficos sin relevancia para el usuario, resultados con distintos contextos, coherencia en los resultados, pérdida de resultados probablemente relevantes, no toman en cuenta las preferencias del usuario (Fig. 1.1), la distancia de los objetos geográficos recuperados con relación al usuario (Fig. 1.2), donde posiblemente debido a la lejanía de los resultados estos pierdan relevancia para el usuario.

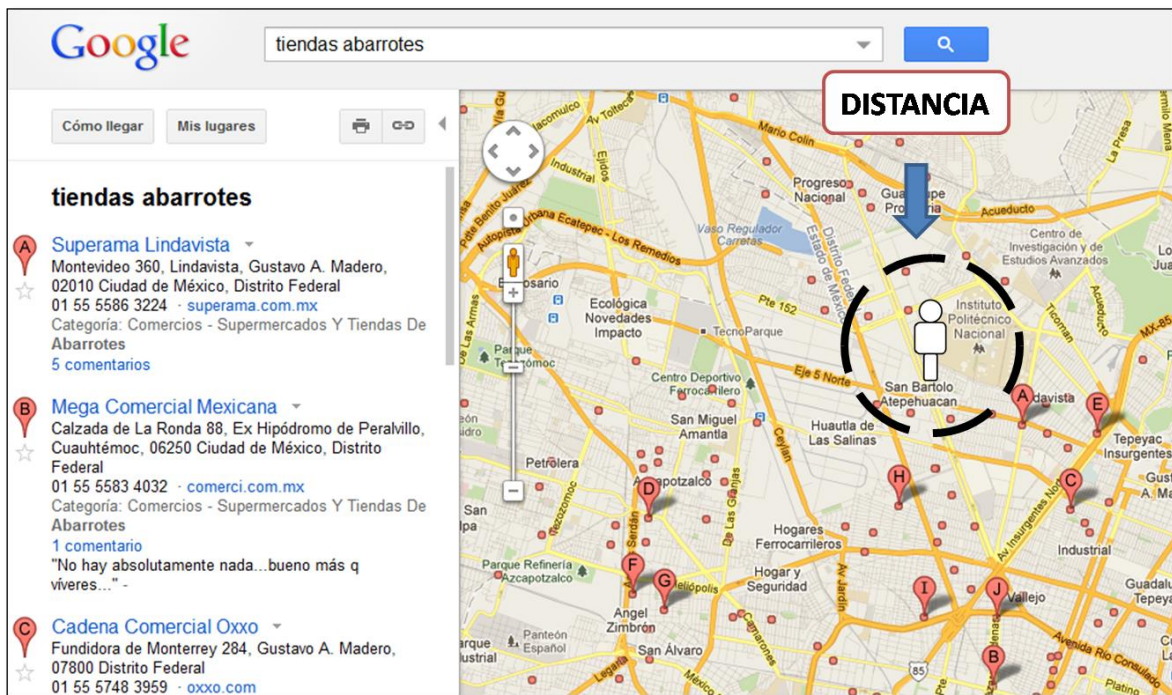


Figura 1.2 Problema de distancia en los resultados de la GIR con relación al usuario.

Debido a estos problemas proponemos la metodología de este trabajo de investigación para contribuir con un enfoque para su solución.

1.2 Objetivo

Diseñar e implementar una metodología enfocada a la Recuperación de Información Geográfica involucrando una conceptualización del dominio geográfico, la posición del usuario y una caracterización del mismo para personalizar la recuperación, generando así resultados relevantes e inteligentes para el usuario.

1.3 Metas

Las metas alcanzadas dentro de este trabajo son:

- Búsqueda y análisis de trabajos relacionados con esta tesis.
- Definir una metodología para personalizar la recuperación de información con un enfoque semántico.
- Análisis de la conceptualización de un dominio de trabajo por medio de las ontologías para el caso de estudio.
- Integración del Perfil de Usuario por medio de la selección de preferencias en la Recuperación de Información Geográfica.
- Análisis y selección de tecnologías adecuadas para la implementación de la metodología.
- Selección e implementación de las medidas de similitud entre conceptos de la ontología para el enriquecimiento de las consultas.
- Descripción e implementación del caso de estudio dentro del marco de la metodología.
- Evaluación de los resultados generados tras la implantación de la metodología.

1.4 Hipótesis

Para el desarrollo de esta tesis se plantearon las siguientes hipótesis:

- Por medio de la conceptualización del dominio geográfico, se pueden establecer relaciones entre conceptos de objetos geográficos y proveer información sobre dichos objetos, teniendo así una base de conocimiento para realizar la recuperación de información.
- La ubicación del usuario para la recuperación de la información proporciona un filtro para la recuperación de objetos geográficos de relevancia para el usuario.
- La similitud semántica proporciona un mecanismo para efectuar una recuperación de información geográfica con base en un contexto definido por el dominio de trabajo, sin dar resultados no relacionados entre sí.
- La caracterización del usuario, permite una clasificación en la cual se consideran aspectos relevantes de cada tipo de usuario, que aplicados en la recuperación de información geográfica, brindan una recuperación en base a la relevancia de los objetos geográficos según sea el perfil del usuario.

1.5 Alcances

Esta sección se incluye debido a la necesidad de describir los tópicos que se desarrollaron y aquellos que hizo falta cubrir, para una mejor percepción de los capítulos siguientes.

La metodología propuesta se basa en tres etapas: la primera de ellas es la *Conceptualización*, está orientada a construir una ontología que identifique las propiedades, relaciones y las entidades implicadas, con el fin de describir un dominio de trabajo, la segunda etapa es la *Caracterización de los usuarios*, basada en la identificación de los conceptos relevantes en la ontología para definir las preferencias del usuario, el cual define un perfil de usuario y por último la etapa de *Recuperación*, se centra en encontrar los datos geográficos que son relevantes para el usuario de acuerdo a su perfil y su ubicación, si no se encuentran resultados

que cumplan dichos criterios, interviene el proceso de la recuperación por la medida de similitud semántica dentro de la conceptualización hizo en la primera etapa.

La implementación se desarrolló con herramientas de software libre como lo son Google Fusion Tables, Google Maps, el Framework ZK de Java, OWL API y con un IDE para desarrollo web Eclipse, debido a la interoperabilidad, disponibilidad, y con fines de obtener una aplicación con la arquitectura cliente-servidor y aprovechar las nuevas herramientas que se están ofertando para los Sistemas de Información Geográfica.

Los resultados que se muestran a partir de la aplicación desarrollada están basados en datos del INEGI que fueron proporcionados para fines de esta investigación y de los cuales se hablará en capítulos posteriores.

1.6 Organización del documento

En el Capítulo 2 se presenta el estado de arte del área de investigación dentro de la cual se encuentra la presente tesis. El Capítulo 3 es la parte más importante de este documento. Es ahí donde se presenta la metodología propuesta. El Capítulo incluye la descripción de cada una de las etapas que componen el método propuesto, las cuales se retoman y se detallan en el Capítulo 4. La implementación y los resultados experimentales, así como la discusión de los mismos, se presentan en el Capítulo 4; y en el Capítulo final, el 5, se exponen las conclusiones y recomendaciones para el trabajo futuro. En el Anexo A se describen los conceptos utilizados para el desarrollo de esta tesis como marco teórico.

CAPÍTULO 2

ESTADO DEL ARTE

El capítulo que enseguida se presenta, muestra una serie de trabajos relacionados, relevantes para el desarrollo de esta tesis como lo son la Recuperación de Información (Information Retrieval, IR) y su papel en el contexto Geográfico, trabajos enfocados en Semántica dentro de la IR y en la Personalización de servicios, de los cuales se menciona de manera breve su funcionamiento para finalizar en una sección de discusión donde se analizaran sus ventajas y desventajas dando una perspectiva de la invitación que hacen estos trabajos para realizar esta investigación.

2.1 Recuperación de la Información

La Recuperación de la Información es un área muy grande dentro de las ciencias de la computación que se enfoca principalmente en proveer a los usuarios un fácil acceso a la información de su interés [Baeza-Yates, 2011]. Hoy en día la investigación en IR incluye el modelado, búsquedas WEB, clasificación de texto, interfaces de usuario, visualización de datos, filtración, etc., donde debido al gran crecimiento actual de información y tipos de datos, se están abriendo nuevos caminos para la investigación.

2.1.1 El problema de la IR

En la actualidad los usuarios de sistemas de IR, tienen necesidades de información de complejidad variable, en el caso más simple están buscando el enlace de la página principal de una compañía o una institución, y en los casos más sofisticados buscan información para realizar tareas asociadas con su trabajo o sus necesidades inmediatas, donde la descripción que el usuario proporciona para la búsqueda por lo regular no provee la mejor formulación

de una consulta para un sistema IR. Con lo que genera la necesidad de traducción de dicha descripción a un conjunto de palabras clave que resumen las necesidades de información del usuario, una vez dada la consulta del usuario el objetivo fundamental del sistema de IR es recuperar información que es útil o relevante para el usuario.

La IR para ser eficaz en su intento de satisfacer las necesidades de información del usuario, debe de alguna manera *interpretar* el contenido de los elementos de información i.e., los documentos en una colección, y ordenarlos en base al grado de relevancia para el usuario. La *interpretación* del contenido de un documento implica extraer información sintáctica y semántica del texto del documento y utilizar esta información para que coincida con la información que necesita el usuario. La dificultad no solo está en saber cómo extraer información de los documentos, sino también en saber cómo usarla para decidir la relevancia, es decir, la noción de relevancia es el *problema* central dentro de la IR.

Una cuestión principal es que la relevancia es una evaluación personal que depende de la tarea a resolver y su contexto, por ejemplo, la relevancia puede cambiar con el tiempo (e.g., se obtiene nueva información), con la localización (e.g., la respuesta más relevante es la más cercana), etc., en este sentido no existe un sistema de recuperación de información que pueda proporcionar respuestas o resultados perfectos para todos los usuarios todo el tiempo.

2.1.2 Arquitectura de un sistema IR [Baeza-Yates, 2011]

El primer paso para el desarrollo de un sistema de IR es tener la colección de documentos, la cual puede formarse de un conjunto de documentos privados u obtenidos de la Web. La colección de documentos es almacenada y usualmente se le llama *repositorio central*. Los documentos en el repositorio central necesitan ser indexados para una recuperación y clasificación rápida, la estructura de indexación más utilizada es *inverted index*, compuesta por distintas palabras de la colección y por cada palabra existe una lista de documentos que la contiene. Teniendo en cuenta que la colección de documentos está indexada, el proceso de recuperación puede iniciarse, el cual consiste en la recuperación de documentos que satisfacen la consulta del usuario. Para buscar, el usuario primero especifica una consulta que refleja la información que necesita, enseguida dicha consulta es analizada y expandida

(variantes ortográficas de la palabra de consulta). La consulta expandida, la cual mencionaremos como el sistema de consulta, se procesa contra la colección de documentos ya indexados para recuperar un subconjunto de todos los documentos. Posteriormente los documentos recuperados son clasificados (ranking).

El propósito del ranking es identificar los documentos que tienen más probabilidad de ser considerados relevantes para el usuario, y esto constituye la parte más crítica de un sistema IR, lo anterior lo podemos ver descrito en la Figura 2.1.

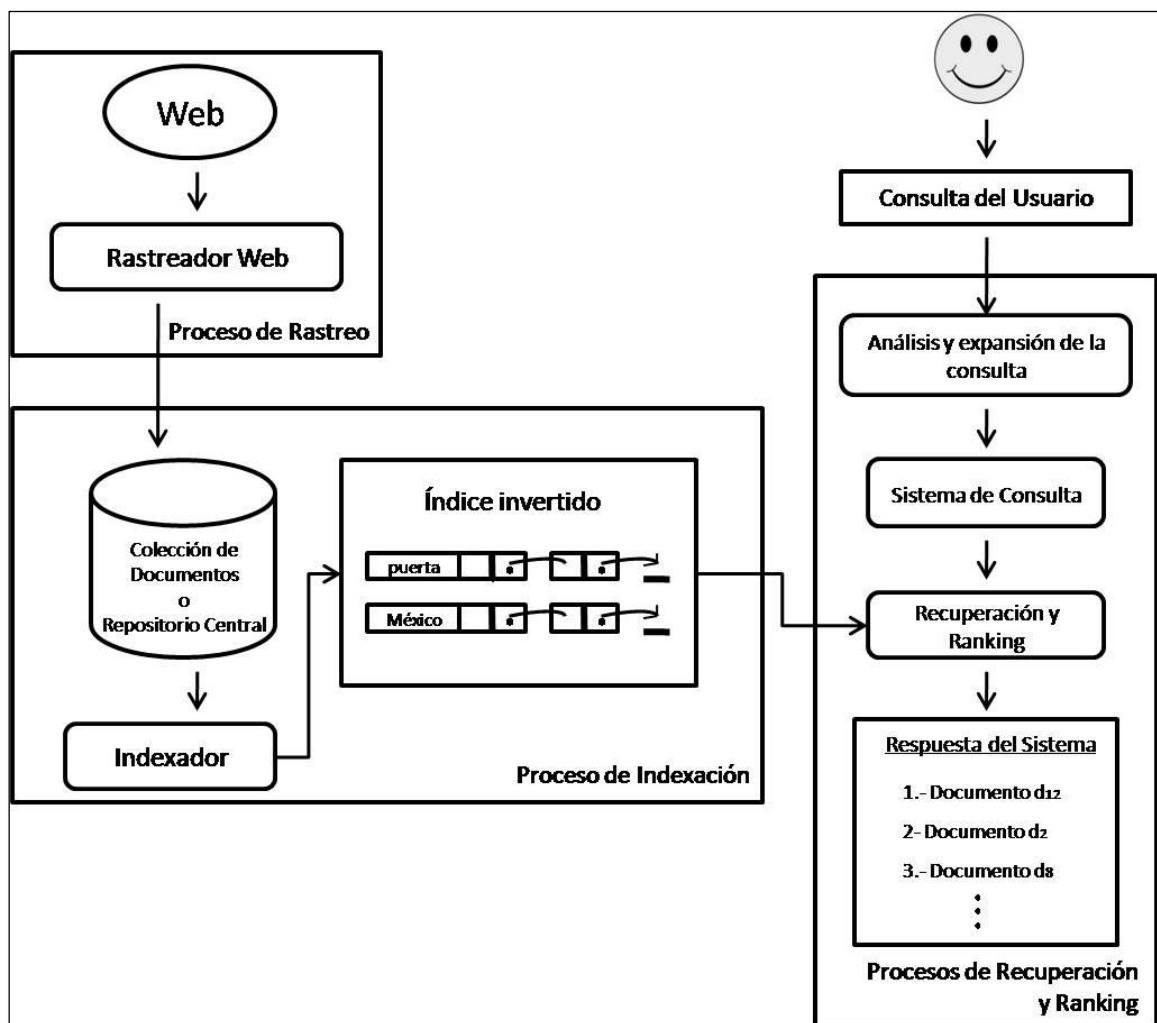


Figura 2.1 Arquitectura de un sistema de IR [Baeza-Yates, 2011].

2.1.3 Modelos de IR

El modelado en los IR es un proceso que tiene como objetivo producir una función de clasificación (ranking) para los documentos dentro de una colección. Este proceso se divide principalmente en dos tareas: (a) la concepción de un marco lógico para la representación de documentos y consultas, y (b) la definición de una función de clasificación que calcula un rank para cada documento en relación a una consulta determinada.

2.1.3.1 Caracterización de un modelo en IR

Las premisas fundamentales que forman las bases del algoritmo de ranking determinan el modelo de IR. La caracterización propuesta por [Baeza-Yates, 2011] es la siguiente:

Un modelo de Recuperación de la Información es una 4-tupla $[\mathbf{D}, \mathbf{Q}, \mathcal{F}, R(q_i, d_j)]$ donde:

1. $\mathbf{D}$ es un conjunto compuesto del punto de vista lógico (o representaciones) de los documentos en una colección.
2. $\mathbf{Q}$ es un conjunto compuesto del punto de vista lógico (o representaciones) de las necesidades de información del usuario.
3. $\mathcal{F}$ es un entorno (marco de trabajo) para modelar representaciones de documentos, consultas, y sus relaciones, tales como conjuntos y relaciones booleanas, vectores y operaciones de álgebra lineal, espacios de muestra y distribuciones de probabilidad.
4. $R(q_i, d_j)$ es la función de ranking que asociada a un número real con una representación de la consulta $q_i \in \mathbf{Q}$ y una representación del documento $d_j \in \mathbf{D}$. Tal ranking define un ordenamiento entre los documentos con respecto a la consulta q_i .

Dada una representación de la consulta y del documento, tales como q_i and d_j , la función de ranking $R(q_i, d_j)$ asigna un rank (un número real) para el documento d_j con respecto a la consulta q_i , como se muestra en la Figura 2.2.

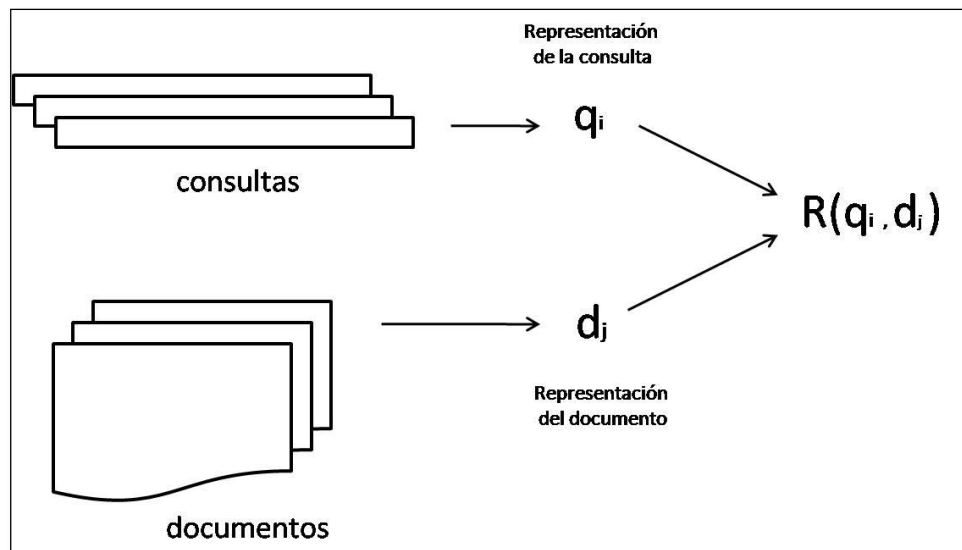


Figura 2.2 La función de ranking $R(q_i, d_j)$.

2.1.3.1 Clasificación de los modelos de IR

Los modelos de IR están basados fundamentalmente en texto, los cuales usan el texto de los documentos para el proceso de ranking. En la Web, sin embargo, también es necesario el uso de la información sobre la estructura de enlaces de la red para alcanzar una buena clasificación.

Por otro lado, los objetos multimedia, están codificados de forma diferente a los documentos de texto. Las imágenes están codificadas como mapas de bits, los objetos video están codificados como flujos temporales de imágenes, y los objetos audio están codificados como flujos discretizados de sonido. A causa de estas particularidades en su representación, los objetos multimedia se clasifican de forma diferente. Dado a estas características, podemos distinguir tres tipos principales de modelos de IR: los basados en texto, los basados en enlaces, y los basados en objetos multimedia, como se muestra en la Figura 2.3.

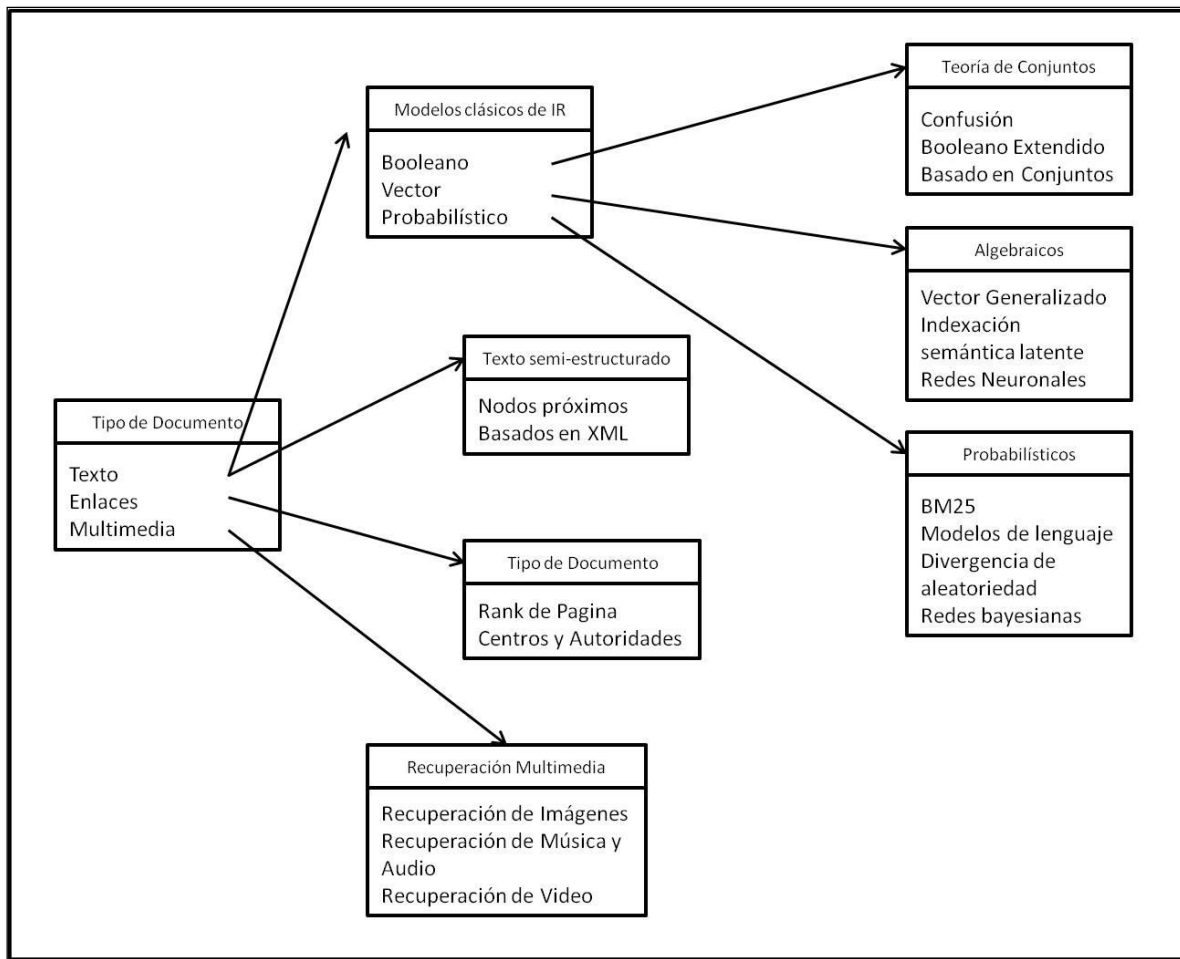


Figura 2.3 Clasificación de los IR

En la primera categoría, el texto es modelado como una secuencia de palabras (Fig. 2.3), en la segunda categoría, componentes estructurales del texto como el título, secciones, subsecciones, y los párrafos son una parte integral del modelo, que generalmente se conoce como semi-estructurada, ya que también incluye texto no estructurado.

2.2 Recuperación de Información Geográfica

La Recuperación de Información Geográfica puede ser vista como una rama especializada de la IR tradicional. Esto incluye todas las áreas de investigación que tradicionalmente han formado el núcleo de la IR, pero además tiene un énfasis en la indexación y recuperación geográfica [Larson, 1995].

La información que tiene relación con el espacio geográfico es llamada información georeferenciada. La información georeferenciada se encuentra en todos los tipos de medios, no sólo de datos estructurados como los mapas.

La mayoría de la información disponible en las bibliotecas digitales y en Internet está georeferenciada, aunque la mayoría no es denotada en términos de coordenadas geográficas. La ubicación geográfica está dada por coordenadas (latitud, longitud). La GIR requiere que los nombres de lugares y frases en donde se incluyen referencias directas o indirectas sean traducidos a huellas geográficas que puedan ser indexadas. La Figura 2.4 muestra los componentes de un Sistema GIR.

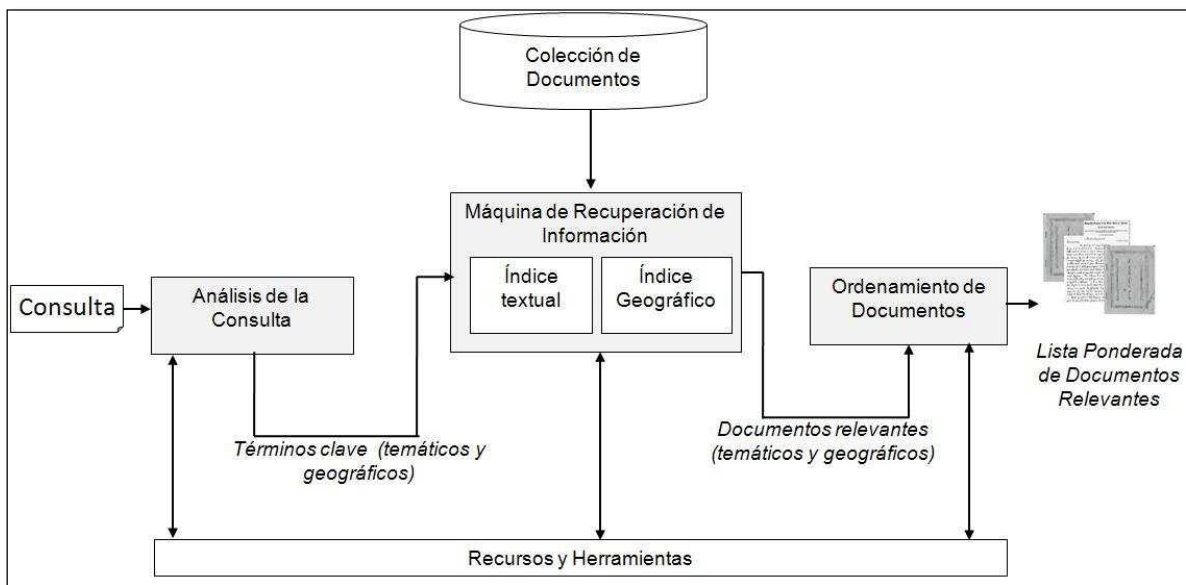


Figura 2.4 Arquitectura general de un sistema GIR.

2.2.1 Análisis de la Consulta

El primer módulo, análisis de la consulta, tiene por objetivo extraer de manera automática los términos clave de la consulta. Generalmente, estos términos son clasificados en términos temáticos y términos geográficos. Por ejemplo, en la consulta “Coche bomba en México”, la clasificación sería como, temáticos: *coche bomba*, geográficos: *México*.

Una de las formas más comunes para resolver el problema de identificar los términos geográficos contenidos en la consulta es por medio de un *identificador de entidades nombradas* (NER¹), sin embargo solo reconocerá los términos geográficos, dando lugar al problema de la ambigüedad geográfica donde una de las técnicas para lograr la desambiguación es por medio de métodos basados en reglas. El más común y sencillo de hacer es contar con un “gazetter básico” dentro del cual habrá una única entrada o localidad para cada una de las posibles entidades identificadas. Es importante tener en cuenta que el problema de ambigüedad no sólo se presenta entre lugares que tienen el mismo nombre, sino también con los nombres de organizaciones y/o nombres de personas.

Otra parte importante de la consulta es la relación espacial. Una relación espacial representa una restricción para las consultas geográficas la cual “selecciona” o “direcciona” una región geográfica. Este tipo de selección puede ser tan simple como “Lagos en el Estado de México”, o más complejas como “encuentra el camino más corto de Veracruz a Los Mochis”. Las relaciones espaciales están divididas en tres categorías [Gracia-Cumbreras, et. al., 2009]: a) topológicas (e.g., museos en el D.F.); b) métricas (e.g., no muy lejos del D.F.); c) proyectivas (e.g., a la derecha del camino).

2.2.2 Módulo de IR

De manera general en el módulo de IR hace uso de índices separados [Martins, et. al., 2005], i.e., un índice temático, el cual contiene todos aquellos términos que no correspondan a una referencia geográfica (términos temáticos), y uno o más índices geográficos, los cuales contienen solo nombres de referencias geográficas (términos geográficos). En este modulo se emplean los identificadores de entidades nombradas, técnicas de desambiguación, gazetters, y en un plano complejo el uso de expansión de la consulta para enriquecer la recuperación.

2.2.3 Ranking

Este módulo es aplicado al conjunto de documentos producto del modulo de IR y tiene como objetivo ordenar dichos documentos en función de algún criterio o métrica, que se define

¹ Por sus siglas en ingles Name Entity Recognition

para la recuperación. En el contexto de la GIR este ordenamiento se lleva a cabo no sólo tomando en cuenta la similitud temática de los documentos con las consultas, sino también tomando en cuenta una similitud o cercanía geográfica.

Algunos trabajos [Larson, 2005] [Ferrés, et. al., 2005] parten de la existencia de dos o más listas de documentos recuperados, donde simplemente se hace uso de operadores lógicos entre las distintas listas (e.g., AND), de esta forma la lista resultante entrega al usuario los documentos que aparecen en todas las listas. Por otro lado se encuentran metodologías basadas en el filtrado de documentos [Ferrés, Rodríguez, 2007] [Hauff, et. al., 2007], donde se identifican los términos geográficos de los documentos y son buscados dentro de un catalogo geográfico, para que un documento pase el filtro debe contener al menos un término geográfico que coincide con los que aparecen en la consulta o que tenga una similitud.

Por último las técnicas complejas [Larson, Frontiera, 2004] [Martins, et. al., 2005] que consisten en definir una función de ordenamiento con base en distintos criterios como: Similitudes geométricas, Distancia topológica a través de relaciones jerárquicas, Distancias Espaciales, Población compartida, Adyacencia dentro de una ontología. Una vez calculadas estas diferentes medidas, estas son empleadas ya sea de manera individual o combinadas de tal forma que al final se tenga un único valor para el término *Similitud_{geométrica}* (q, d) dando la similitud geográfica entre la consulta y el documento.

2.3 Similitud Semántica en los GIR

Dentro del proceso de IR en la recuperación de información geográfica existen diferentes metodologías para clasificar los documentos de tal forma que al usuario obtenga como resultado los documentos relevantes en base a la consulta que realice, actualmente uno de los enfoques más sobresalientes y utilizados para generar la recuperación es la *similitud semántica*.

2.3.1 Análisis semántico

En comparación con técnicas anteriores de recuperación de información basadas en palabras clave que dependen de la sintaxis, hay un papel creciente de los enfoques semánticos para la gestión de la información donde el significado se asocia con los datos y términos utilizados en las consultas [Shah, 1998].

Algunos de los aspectos más notables de los enfoques semánticos, son el desarrollo de ontologías y la anotación semántica de los datos. Una ontología es una especificación formal y explícita de una conceptualización compartida [Gruber, 1995], "conceptualización" se refiere a un modelo abstracto de algún fenómeno en el mundo, identificando los conceptos relevantes de ese fenómeno, "explícitos" significa que el tipo de conceptos utilizados, y las limitaciones de su uso son explícitamente definidos; "formal" se refiere al hecho de que la ontología debe ser legible por máquina, "compartida" refleja la idea de que una ontología es una captura consensuada de conocimiento, es decir, que el conocimiento es aceptado por un grupo [Qiuyan, Guisheng, 2008]; las ontologías son el corazón de la mayoría de los enfoques semánticos [De Bruijn, 2003].

2.3.1.1 Semántica Geoespacial

Análisis geoespaciales convencionales se basan principalmente en medidas métricas (e.g., cuantitativa). Pero la gente suele expresar y comprender las relaciones espaciales a través del lenguaje natural en lugar de medidas métricas, por lo tanto, es importante establecer una semántica geoespacial, para la realización de consultas espaciales utilizando referencias temporales (e.g., cerca, lejos, cerca del mediodía) y otras representaciones no métricas de la información [Arpinar, et. al., 2004]. Estos incluyen el uso de modificadores cualitativos (e.g., muy, casi, un poco), relaciones de proximidad (e.g., A esta cerca de B), las referencias espaciales (e.g., el lado este de la ciudad, al oeste del río).

2.3.2 Similitud Semántica

La similitud ha sido ampliamente aplicada en GIScience [Janowicz, et. Al., 2011]. Basado en el modelo de función de [Tversky, 1977], Rodríguez y Egenhofer [Rodríguez, Egenhofer, 2004] desarrollaron "*the matching distance similarity measure (MDSM)*". La similitud se ha aplicado a diversas tareas en muchos campos. En consecuencia no hay una descripción precisa independiente a la aplicación de las medidas de similitud. Incluso para la

recuperación de información basada en semántica, se han propuesto distintas medidas de similitud. Esto hace que la selección de una medida adecuada para una aplicación en particular sea una tarea difícil. También se plantea la cuestión de cómo comparar las teorías existentes. Mediante el examen de varias de estas medidas en diferentes dominios se encontraron patrones genéricos tal que en conjunto forman un marco para describir cómo se calcula la similitud [Janowicz, et. Al., 2011] [Janowicz, et. Al., 2009].

En [Schwering, 2008], se encuentran un conjunto de trabajos representativos en la rama de la similitud semántica con el enfoque a datos geoespaciales, donde hace una clasificación de los diferentes modelos de similitud semántica y los compara.

2.3.2.1 Funciones de Similitud

Dependiendo del lenguaje de representación y aplicación, diferentes funciones de similitud tienen que ser aplicadas. En la mayoría de los casos, cada función de similitud se ocupa de la normalización (con valores entre 0 y 1).

En el caso de MDSM [Rodríguez, Egenhofer, 2004], las características se distinguen en diferentes tipos durante el proceso de alineación: partes, atributos y funciones. A pesar de que un contexto de ponderación se calcula para cada uno de estos tipos, la misma función de similitud se aplica a todos ellos.

La medida de la distancia de correspondencia de similitud semántica consiste en un modelo computacional que evalúa similitud, combinando la comparación de distinción de características clasificadas por tipos. La función global de similitud $S(c_1, c_2)$ es una sumatoria del peso de los valores de similitud de partes, funciones y atributos (ver, ecuación 2.1):

$$S(c_1, c_2) = \omega_p \cdot Sp(c_1, c_2) + \omega_f \cdot Sf(c_1, c_2) + \omega_a \cdot Sa(c_1, c_2) \quad (2.1)$$

Donde ω_p , ω_f , y ω_a son la suma de los valores similares por partes, funciones, y atributos respectivamente. Estas sumatorias de pesos definen la importancia relativa de partes, funciones, y atributos que pueden variar entre los diferentes contextos. La suma de los pesos debe ser igual a 1. Para cada tipo de característica de distinción se usa la función similar

$S_t(c_1, c_2)$. Donde c_1 y c_2 son dos clases de entidades, t simboliza el tipo de característica, y C_1 y C_2 son los respectivos conjuntos de características de tipo t para c_1 y c_2 (ver, ecuación 2.2):

$$S_t(c_1, c_2) = \frac{|C_1 \cap C_2|}{|C_1 \cap C_2| + \alpha(c_1, c_2) \cdot |C_1 \setminus C_2| + (1 - \alpha(c_1, c_2)) \cdot |C_2 \setminus C_1|} \quad (2.2)$$

La ecuación 2.2 describe una función de similitud no simétrica para cada tipo de entidad. $S_t(c_1, c_2)$ define la similitud tipo t entre la entidad de clases c_1 y c_2 . C_1 y C_2 son los conjuntos de características de tipo t para c_1 y c_2 , mientras que $|C_1 \cap C_2|$ es la cardinalidad de la intersección de conjuntos y $|C_1 \setminus C_2|$ es la cardinalidad de la diferencia de conjuntos. La importancia relativa α (ecuación 2.3) de las distintas características de tipo t se define en términos de la distancia d entre c_1 y c_2 dentro de una jerarquía que tiene relaciones taxonómicas en cuenta.

La superclase inmediata común corresponde a la cota superior mínima (*lub*, least upper bound) entre dos clases de entidad en un conjunto parcialmente ordenado (Birkhoff 1967). La distancia se define como $d(c_1, c_2) = d(c_1, lub) + d(c_2, lub)$.

$$\alpha(c_1, c_2) = \begin{cases} \frac{d(c_1, lub)}{d(c_1, c_2)}, & d(c_1, lub) \leq d(c_2, lub) \\ 1 - \frac{d(c_1, lub)}{d(c_1, c_2)}, & d(c_1, lub) > d(c_2, lub) \end{cases} \quad (2.3)$$

Este trabajo no es utilizado para nuestra investigación pero toma a la distancia entre conceptos como referencia para la similitud.

2.3.2.2 La Similitud en la Recuperación de Información basada en la Semántica

En contraste con los enfoques puramente sintácticos, la recuperación de información basada en semántica toma las conceptualizaciones subyacentes en cuenta para calcular la relevancia y por lo tanto mejora la búsqueda y la navegación a través de los datos estructurados. En general, se puede distinguir entre dos enfoques para la recuperación de conceptos: los que se basan en un razonamiento de subsumción (*subsumption*) clásico y los que se basan en medidas de similitud semántica [Janowicz, et. al., 2008]. Simplificando, el razonamiento de

subsumción clásico se puede aplicar a la búsqueda vertical, mientras la similitud funciona mejor para la búsqueda horizontal (Figura 2.5).

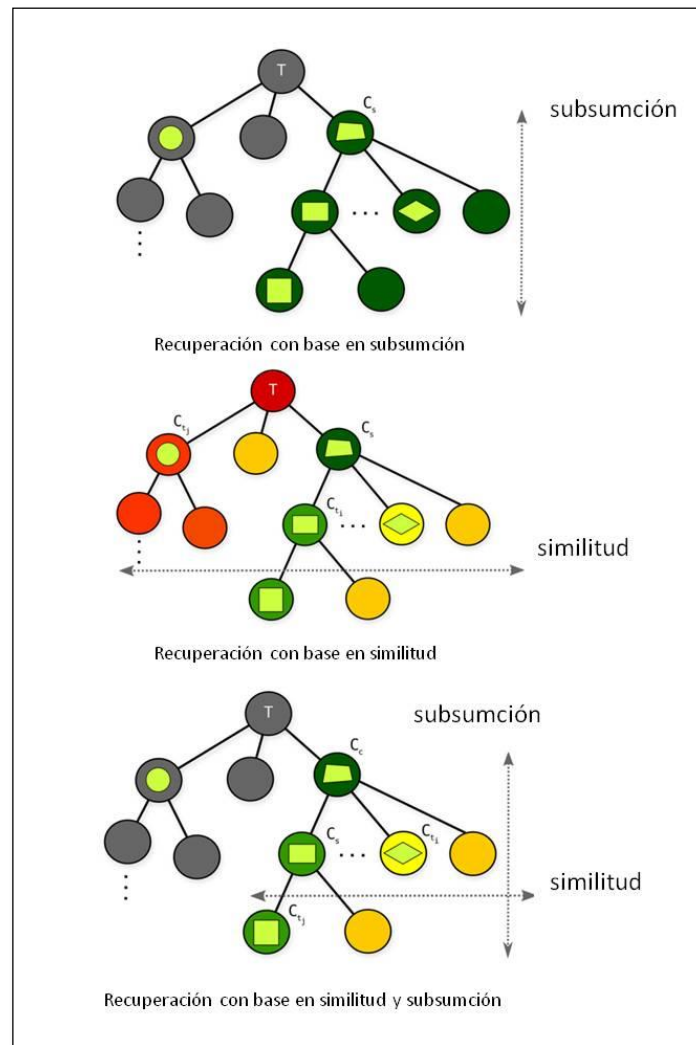


Figura 2.5 Recuperación basada en subsumción y similitud.

Este enfoque es el que se va a trabajar en la metodología de este trabajo de investigación, mas adelante hablaremos al respecto.

2.3.2.3 Recuperación basada en Ontologías

Las ontologías son la clave de la descripción semántica de la información geográfica información. La Ontología de información geográfica se ha creado para describir la semántica de las características geográficas no espaciales, lo que define los conceptos fundamentales de las características geográficas y las relaciones de ellos [Zhan, et. al., 2008]. La ontología de relaciones espaciales y la ontología de geometría se construyen para describir la semántica espacial de los objetos geográficos.

Las ontologías de relaciones espaciales definen las relaciones topológicas, relaciones de dirección y de distancia. Las ontologías de geometría definen los conceptos y las relaciones geométricas de punto, línea, polígono, superficie, etc.

En primer lugar, los conceptos fundamentales en el dominio de la ciencia de información geográfica son abstractos. Entonces, las relaciones de conceptos, las relaciones geométricas, las relaciones espaciales, las relaciones de localización son modeladas y establecidas para su formalización. Los conceptos incluyen las relaciones "Subclase de", "parte-todo", "miembro de", "instancia de" y "dependencia".

Un método desarrollado con enfoque ontológico para la GIR, se muestra en la Figura 2.6. Hay tres pasos principales en el método. El primer paso del usuario, los términos de búsqueda se asignan a los conceptos de la ontología de dominio. El segundo paso, el conceptos se amplían sobre la base de la jerarquía de los conceptos de dominio de la ontología. Si los conceptos tienen subclases, entonces las subclases son añadidas para expandir la consulta. De lo contrario, la superclase está marcada. Si hay superclase, entonces la superclase se añade a ampliar la consulta. El tercer paso, basado en la expansión de la consulta, las descripciones adecuadas de información geográfica son buscar y devolver a los usuarios. Si los resultados son adecuados, la búsqueda es terminada. De lo contrario, la repetición de la segunda etapa es necesaria.

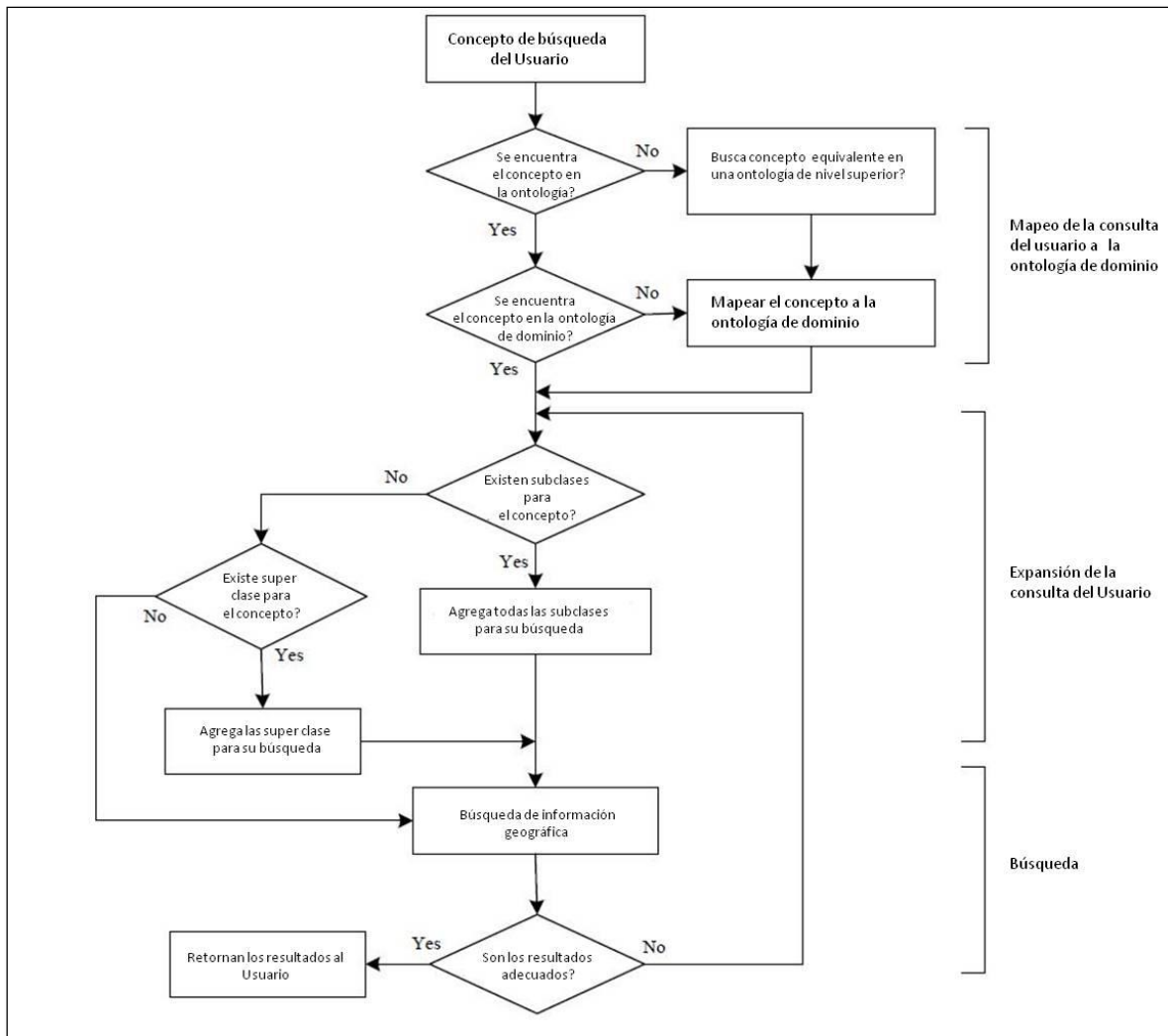


Figura 2.6 Metodología basada en una ontología para la GIR [Qin Zhan, et al. 2008]

2.4 Personalización

La personalización de la búsqueda y la navegación proporciona información que coincide con los intereses personales de los usuarios, por lo tanto provee un acceso más eficaz y eficiente a la información [Trajkova, Gauch, 2004]. Una característica clave en el desarrollo de las aplicaciones personalizadas es la construcción de perfiles de usuario que representen los intereses de un usuario con exactitud.

2.4.1 Perfil de Usuario

La noción del perfil de usuario se ha introducido con el fin de registrar el contexto de usuario y personalizar las aplicaciones para ser adaptadas a las necesidades de los usuarios [Golemati, 2007]. La necesidad de sistemas de software para adaptarse a los usuarios, ha sido reconocida en muchas áreas de aplicación y la investigación sobre perfiles de usuario y el contexto se ha extendido a muchas disciplinas que se ocupan del desarrollo de sistemas informáticos [Kobsa, 1993]. De acuerdo con [Golemati, 2007], un modelo de usuario contiene toda la información que el sistema conoce del usuario. En general, se inicia ya sea con los valores por defecto o por consulta del usuario. Una visión general de los métodos para la construcción de un perfil de usuario se presenta en [Rich, 1983]. Los tópicos en cuanto al modelado del usuario se presentan en [Kobsa, 1993], concentrándose en el modelado del conocimiento del usuario, los planes y las preferencias en un dominio.

2.4.2 Diseño centrado en el usuario

UCD (User-Centred Design) es un método empleado en el diseño de sistemas, habiendo sido desarrollado bajo la premisa de asegurar el éxito comercial y utilidad de un sistema, todo el diseño de actividades debe colocar al usuario final como su foco, de modo que el producto final es fácil utilizar y en última instancia, se ajuste a sus necesidades [Gould, Lewis, 1985] [Butler, 1996] [Mayhew, 1999]. En esencia, un enfoque de UCD se ocupa de las preguntas: “¿Cómo puedo entender al usuario?” y “¿Cómo puedo asegurar que este entendimiento se refleje en mi sistema?” [Holtzblatt, Beyer, 1993] [Wealands, Miller, 2007]. El estándar describe cuatro actividades principales de la UCD, que se llevan a cabo iterativamente hasta que los objetivos definidos se han cumplido. Estos se presentan en la Figura 2.7 y se describen a continuación:

1. Entender y especificar el contexto de uso incorporando las características del usuario, los objetivos o tareas, y el ambiente de uso, con el fin de apoyar la especificación de requisitos de usuario y proporcionar una base para las actividades de evaluación posteriores.

2. Especificar las necesidades de los usuarios que utilizan el contexto de uso previamente definido.
3. Producir soluciones de diseño basadas en los objetivos de diseño, las limitaciones y la incorporación de conocimientos HCI² (en relación con el diseño visual de la interacción, diseño, usabilidad, etc.).
4. Evaluar los diseños frente a los requisitos en todo el desarrollo, el empleo de prototipos y la aplicación de los criterios desarrollados previamente. Importante para determinar el grado a los objetivos que se han cumplido y la obtención de información relativa a mejoras de diseño.

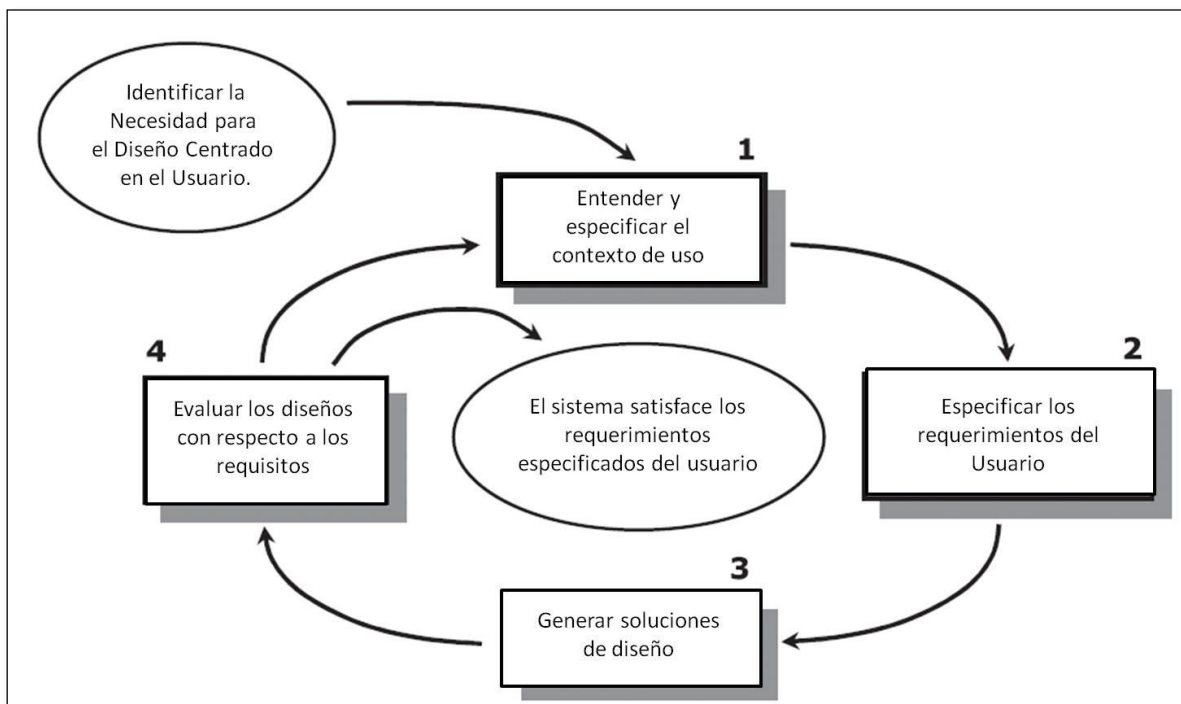


Figura 2.7 Actividades del diseño centrado en el usuario [Jokela, et. al., 2003]

²Por sus siglas en ingles Human-Computer Interaction

2.5 Discusión sobre el Estado del Arte

Los trabajos mencionados en esta sección resaltan la importancia del usuario para el diseño de los sistemas de información de forma general, y en el caso de esta investigación, con un enfoque en la personalización de la recuperación de información geográfica semántica.

En la primera sección del estado del arte se describieron las generalidades de la recuperación de información, mostrando el problema de la IR, arquitecturas generales, así como su clasificación con base en el modelado (texto no estructurado, estructurado, la web, multimedia), donde claro faltaría tratar temas como las interfaces de usuario en estos sistemas, paradigmas para la evaluación de la recuperación, e inclusive de algunos de los motores para la recuperación.

Posteriormente se presentan los fundamentos de la recuperación de información geográfica, y se deja en claro que el trabajo con información geoespacial requiere de un análisis y un tratamiento específico para este tipo de recuperación, como la división en los índices textuales y geográficos dentro del módulo de recuperación. Al tener este tipo de información se generan otro tipo de problemas con la ambigüedad de los datos, problemas en la expansión de la consulta para una mejora en la recuperación, y tener en cuenta criterios adicionales para la recuperación, como lo es la cercanía de los objetos geográficos, y las relaciones espaciales de los datos. Dando paso a soluciones propuestas como la similitud temática, los diccionarios geográficos (*gazetters*), técnicas de filtrado, y a su vez dejando la puerta abierta para nuevas soluciones a los problemas mencionados.

Por otro lado uno de los enfoques más utilizados en la actualidad dentro de los GIR es la similitud semántica, proporcionando al proceso de recuperación una mejora al tener conceptualizado un dominio de trabajo, dando pie a una base de conocimiento para la recuperación. La herramienta que hace posible este tipo de enfoque es la Ontología, describe la semántica de las características geográficas espaciales (relaciones espaciales y de

geometría) o no espaciales (temáticas), ayudando principalmente a la expansión de la consulta para mejorar la recuperación.

El tema de la personalización es parte importante de este trabajo, y como se describe en el estado del arte, de manera general para los sistemas actuales. El hecho de tener información de las preferencias y características del usuario, acota y mejora el proceso de recuperación ya que pueden ser filtrados documentos u objetos irrelevantes para el usuario, satisfaciendo las necesidades de información del usuario en una forma más específica y detallada con base en su perfil, tal que el sistema se centra completamente en el usuario.

Por otro lado los trabajos llevados a cabo en el laboratorio de Procesamiento Inteligente de Información Espacial, al cual pertenezco, se han enfocado también en la personalización y recuperación [Ibarra, 2009] [Renteria, 2009]. A diferencia de estos trabajos, en esta tesis se utiliza una medida de similitud jerárquica, y una selección de conceptos para conformar el perfil de usuario.

Por último, no cabe duda que existen diversas investigaciones en cada una de las áreas mencionadas, y que aun así siguen teniendo problemas a resolver, de forma general algunas de las investigaciones se están llevando por separado [Ferrés, et. al., 2005] [Hauff, et. al., 2007] [Larson, Frontiera, 2004] [Rich, 1983], y en distintos campos, aunque en algunos trabajos se esté empezando a dar esa integración [Janowicz, et. al., 2011] [Zhan, et. al., 2008] [Wealands, Miller, 2007], aún sigue débil el lazo entre estas disciplinas, y es aquí donde entra el desarrollo de esta investigación para proporcionar una metodología con estos enfoques.

CAPÍTULO 3

METODOLOGÍA

En este capítulo se describe la metodología desarrollada en el presente trabajo de investigación, la cual propone personalizar la GIR con base en las preferencias y ubicación del usuario.

3.1 Marco de Trabajo

La metodología propuesta en el presente trabajo de investigación está compuesta de 3 etapas: Conceptualización, Caracterización y Recuperación. De manera general la metodología trabaja con la abstracción de la realidad en una Base de Datos Espacial (BDE³) y la ubicación del usuario como datos de entrada (Figura 3.1). En la BDE se tiene una colección de objetos geográficos, la cual representa el universo de búsqueda para la etapa de recuperación. Gracias a la etapa conceptualización (que como su nombre lo indica, genera una conceptualización) y a la etapa de caracterización (genera un perfil de usuario), se puede efectuar una expansión de la consulta para enriquecer los resultados de una recuperación de forma personalizada, esto se explica a detalle en la sección correspondiente de esta etapa. La etapa de recuperación genera como salida un conjunto de objetos geográficos con base en una vecindad determinada por la ubicación del usuario. El conjunto de objetos geográficos que se obtiene como resultado final, después de finalizar la metodología, tendrá relación con las preferencias del usuario, ya que la recuperación está en función de la conceptualización; sin embargo, es importante resaltar que no todos los elementos del conjunto resultante son relevantes para el usuario. Lo anterior se podrá comprobar en los resultados descritos en el siguiente capítulo.

³ Ver Anexo A.

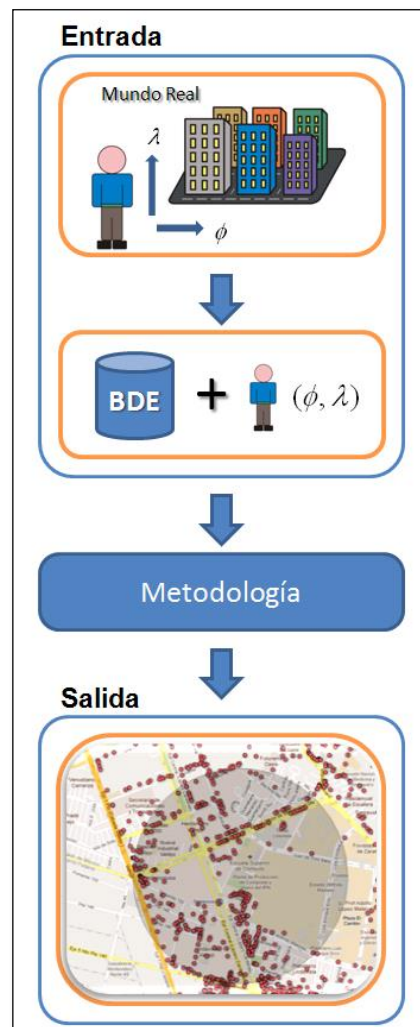


Figura 3.1 Diagrama general de la metodología propuesta.

En la Figura 3.2 se muestra de forma general las tres etapas de la metodología que permite generar resultados relevantes para un usuario. A diferencia de una recuperación de información convencional, el uso de un método de expansión, permite incrementar el número de objetos geográficos que se recuperan al momento de que el usuario desea hacer una consulta. Por medio de dos métricas, de nombre *Precision* y *Recall*, se hace una evaluación de los objetos geográficos devueltos por la metodología.

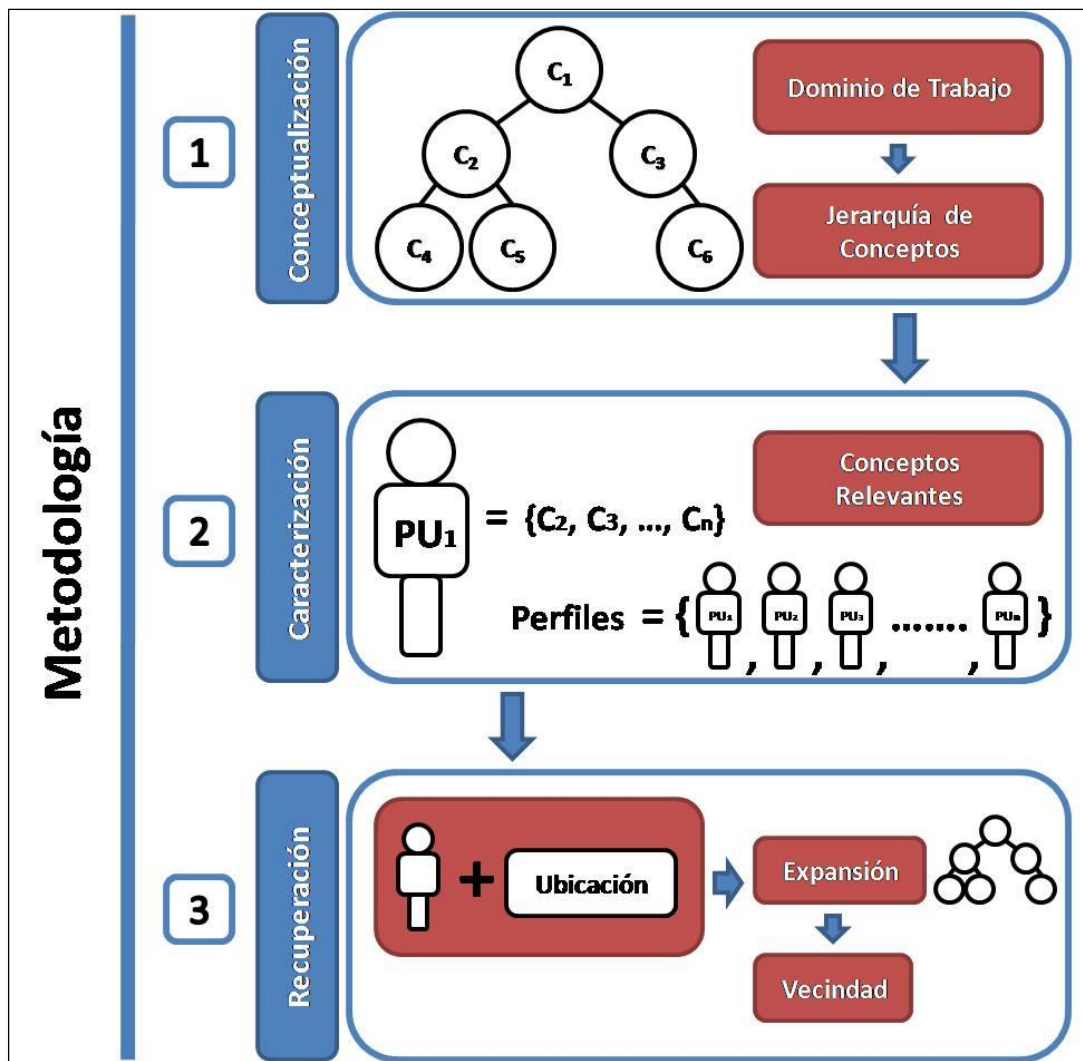


Figura 3.2 Metodología.

Conceptualización: en esta etapa se describe el dominio de trabajo por medio de una ontología de aplicación, la cual representa a una jerarquía de conceptos, usando la relación *es-un*. La ontología proporciona una base de conocimiento computable [Qiuyan, 2008], y es la clave de la descripción semántica de la información geográfica [Zhan, et. al., 2008]. Esta base de conocimiento posteriormente va a ser utilizada en la Caracterización para generar el perfil de usuario, y en la Recuperación para aplicar la función de expansión de conceptos (*Expansión_Conceptos*).

Caracterización: la personalización en esta investigación se centra en las preferencias del usuario con base en un contexto definido [Ahu Sieg, 2007]. Dichas preferencias construyen el perfil del usuario, el cual está compuesto por un conjunto de conceptos relevantes para el usuario, extraídos de la conceptualización (jerarquía de conceptos).

Recuperación: su objetivo es obtener los objetos geográficos que son relevantes para el usuario con base en su ubicación. Para que la recuperación de la información sea eficaz en su intento de satisfacer las necesidades de información del usuario, debe de alguna manera de interpretar el contenido de los elementos de la información [Baeza-Yates, 2011]. La interpretación en la metodología de este trabajo se da con base en la conceptualización, ya que esta etapa brinda el conocimiento de los objetos geográficos con base en dominio de trabajo. La caracterización por su parte, brinda una recuperación de objetos geográficos relevantes para el usuario, que junto con su ubicación la recuperación se delimita a una vecindad. Por último, la recuperación cuenta con un módulo de expansión, el cual se aplica entre el perfil del usuario, la jerarquía de conceptos y los datos disponibles en el área de interés. La etapa de recuperación genera como salida un conjunto de objetos geográficos.

3.2 Conceptualización

En esta etapa se describe el dominio de trabajo por medio de una jerarquía de clases, la cual está compuesta por conceptos y la relación *es-un*.

3.2.1 Jerarquía de conceptos

Las ontologías son un tipo especial de recurso de conocimiento, que van desde las taxonomías de conceptos simples, como las jerarquías de dominios disponibles en los motores de búsqueda, a ontologías complejas integradas en los sistemas formales con las capacidades de razonamiento [Bulskov, 2006]. En la metodología, el objetivo principal de utilizar una ontología, es pasar de una evaluación de las consultas con base en palabras a una evaluación con base en conceptos. En la Figura 3.3, se muestra un bosquejo del mundo real que por medio del proceso de conceptualización se ordena y clasifica ese conocimiento con base en

un dominio de trabajo para poder formalizarlo posteriormente. Para la metodología, la ontología de aplicación describe el mundo real por medio de conceptos (c_1, c_2, c_n) organizados, para un dominio de trabajo en particular.

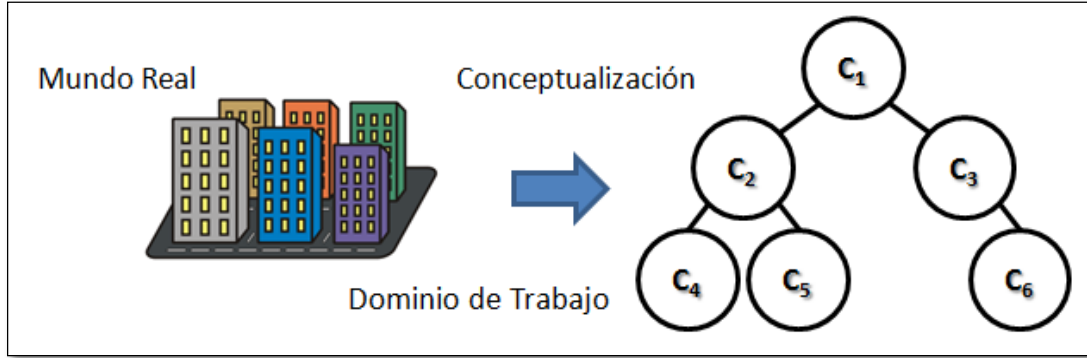


Figura 3.3 Etapa 1: Conceptualización.

Acorde con [Moreno-Ibarra, et. al., 2011] una jerarquía es una 2-tupla:

$$H (C_H, R_H) \quad (3.1)$$

Donde C_H es un conjunto de conceptos y R_H es un conjunto de relaciones de la forma apb , donde $a, b \in C_H$ y p es una relación $\rho: C_H \times C_H$, de la forma apc , $apb \in R_H$, por lo tanto $b = c \quad \forall a \in C_H$.

El conjunto de conceptos tendrá un orden jerárquico con base en el dominio de trabajo, el cual estará definido por el propósito y el contexto de la recuperación. Ejemplificando esta etapa tenemos al [SCIAN, 2007], que es una clasificación de actividades económicas, en la que los conceptos de actividad económica y de unidad económica son básicos para comprender mejor al clasificador. Una unidad económica es una entidad productora de bienes o servicios. Y una actividad económica es un conjunto de acciones realizadas por una unidad económica con el propósito de producir o proporcionar bienes y servicios. Es decir, ambos conceptos se definen uno en términos del otro. Este concepto de actividad económica —que

es congruente con el que presenta el Manual del Sistema de Cuentas Nacionales de 1993⁴ y con el de la Clasificación Industrial Internacional Uniforme— considera al sujeto (“una entidad productora”) que la realiza, es decir, no concibe la actividad en abstracto. Es importante recalcar este hecho porque las actividades que el SCIAN enuncia en su estructura fueron definidas con base en la similitud de los procesos de producción que se llevan a cabo en las unidades económicas⁵.

La estructura del SCIAN MÉXICO 2007 consta de cinco niveles de agregación: sector, subsector, rama, subrama y clase de actividad económica. El sector es el nivel más general; la clase, el más desagregado. El sector se divide en subsectores. Cada subsector está formado por ramas de actividad, las cuales se dividen en subramas. Las clases, por su parte, son desgloses de las subramas (Fig. 3.4).

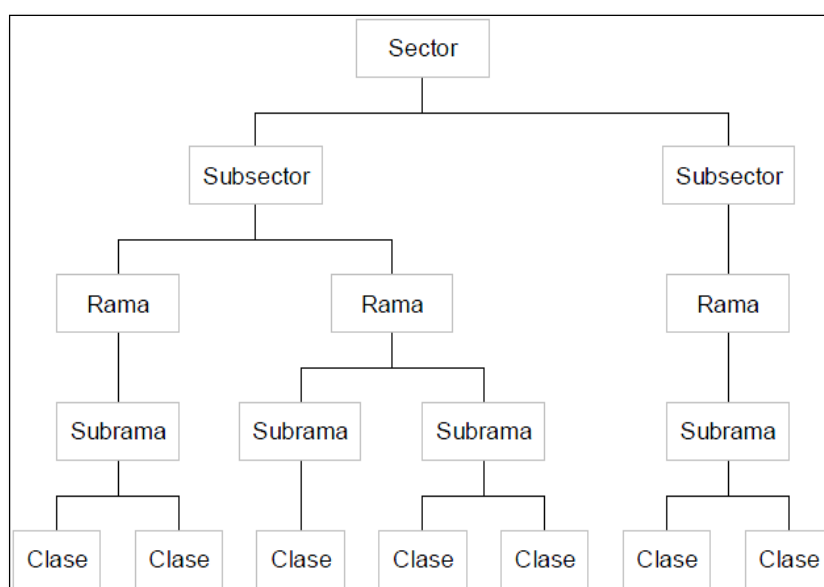


Figura 3.4 Estructura del SCIAN.

La Tabla 3.1 muestra el número de categorías en cada nivel de agregación. Para la implementación de la metodología tomaremos siete sectores con sus respectivas ramas.

⁴ Elaborado por la Comisión de las Comunidades Europeas, la Organización para la Cooperación y el Desarrollo Económico, el Fondo Monetario Internacional, el Banco Mundial y la Organización de las Naciones Unidas.

⁵ Para la metodología propuesta en este trabajo es parte fundamental la similitud definida en el SCIAN.

Tabla 3.1 Número de Categorías del SCIAN.

Nivel de Agregación	Numero de Categorías
Sector	20
Subsector	94
Rama	304
Subrama	617
Clase	1049

El orden de presentación de los sectores en el SCIAN, tiene su base en la agrupación tradicional de actividades económicas en tres grandes grupos: actividades primarias, secundarias y terciarias.

Tabla 3.2 Actividades del SCIAN.

Agrupación tradicional	Sector	
Actividades primarias	11	Agricultura, cría y explotación de animales, aprovechamiento forestal, pesca y caza
Actividades secundarias	21	Minería
	22	Generación, transmisión y distribución de energía eléctrica, suministro de agua y de gas por ductos al consumidor final
	23	Construcción
	31-33	Industrias manufactureras
Actividades terciarias	43	Comercio al por mayor
	46	Comercio al por menor
	48-49	Transportes, correos y almacenamiento
	51	Información en medios masivos
	52	Servicios financieros y de seguros
	53	Servicios inmobiliarios y de alquiler de bienes muebles e intangibles
	54	Servicios profesionales, científicos y técnicos
	55	Corporativos
	56	Servicios de apoyo a los negocios y manejo de desechos y servicios de remediación
	61	Servicios educativos
	62	Servicios de salud y de asistencia social
	71	Servicios de esparcimiento culturales y deportivos, y otros servicios recreativos
	72	Servicios de alojamiento temporal y de preparación de alimentos y bebidas
	81	Otros servicios excepto actividades gubernamentales
	93	Actividades legislativas, gubernamentales, de impartición de justicia y de organismos internacionales y extraterritoriales

Como se aprecia en la Tabla 3.2, el primer sector corresponde a las actividades primarias, es decir, las que se relacionan con el aprovechamiento directo de los recursos naturales, como el suelo, el agua, la flora y la fauna. El segundo grupo de sectores comprende las actividades secundarias, mediante las cuales se efectúa la transformación de todo tipo de bienes o productos –sea que éstos provengan del sector primario o del mismo secundario– en otros nuevos o diferentes. El último grupo de actividades corresponde a las terciarias, que se refieren al comercio y a los servicios.

Las actividades tienen la siguiente estructura (Fig. 3.5):

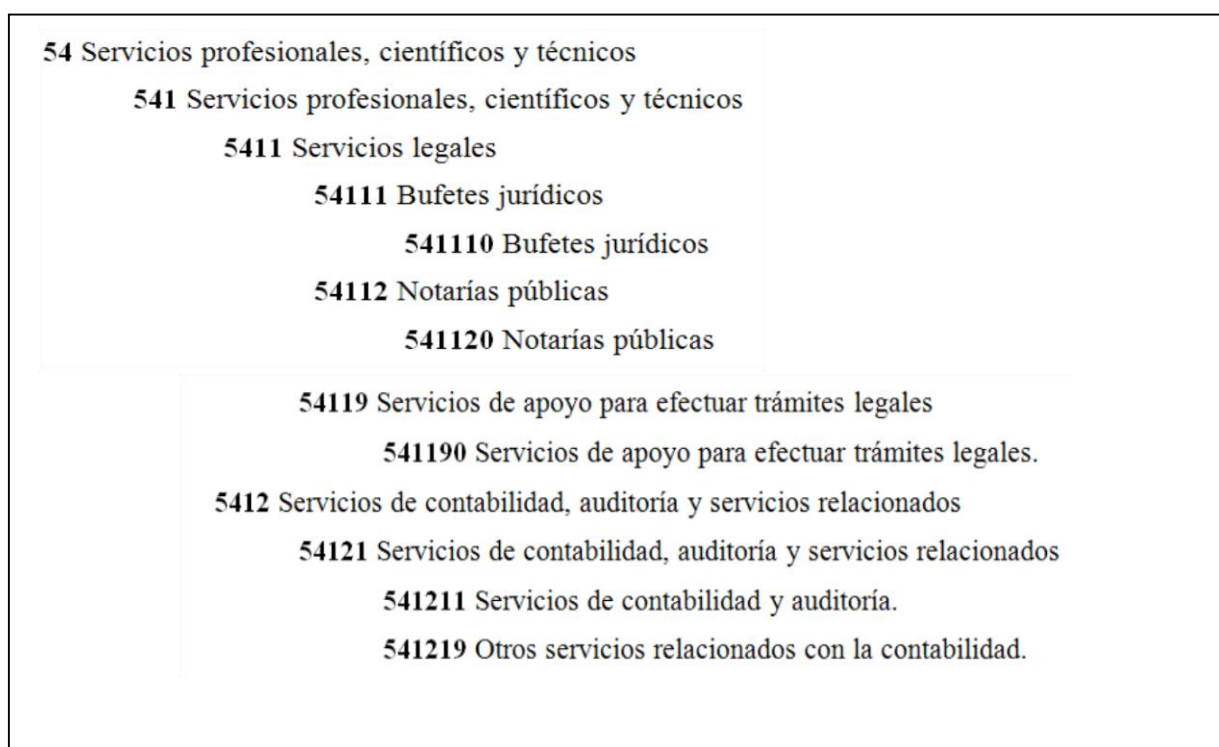


Figura 3.5 Estructura de Actividades.

Para cada categoría del SCIAN, desde el sector hasta la clase de actividad económica, se presenta: a) el código numérico; b) el título o nombre de la categoría, y c) su descripción propiamente dicha. Considérese el siguiente ejemplo: **311222** Elaboración de aceites y grasas vegetales comestibles (Fig. 3.6).

<i>Nivel</i>	<i>Número de dígitos</i>	<i>Código</i>	<i>Ejemplos de categorías</i>
Sector	2	31-33	Industrias manufactureras
Subsector	3	311	Industria alimentaria
Rama	4	3112	Molienda de granos y de semillas y obtención de aceites y grasas
Subrama	5	31122	Elaboración de almidones, aceites y grasas vegetales comestibles
Clase	6	311222	Elaboración de aceites y grasas vegetales comestibles ^{MEX.}

Figura 3.6 Ejemplo de Categoría.

Los dígitos del código de cada actividad económica permiten ver los niveles más agregados en los que está incluida. Así, por ejemplo, la clase de actividad “Elaboración de aceites y grasas vegetales comestibles” está formando parte de las siguientes categorías: del sector **31-33**, del subsector **311**, de la rama **3112** y de la subrama **31122**. Los códigos tienen otras particularidades: en el nivel de clase de actividad, un cero como último dígito indica que se trata de la única clase de la subrama. Y en los niveles que van desde el subsector hasta la clase de actividad, un nueve como último dígito señala, por lo regular, que el contenido de la categoría en cuestión es residual con respecto a las otras dentro de su mismo nivel. Por lo tanto para la conceptualización tendríamos de manera general:

$$C_H = \{\text{Sector, Subsector, Rama, Subrama, Clase}\}$$

$$R_H = \{\text{es-un}\}$$

De esta forma la jerarquía de conceptos queda construida y donde posteriormente C_H estará compuesto por los conceptos del SCIAN:

$$C_H = \{\text{Servicios profesionales científicos y técnicos, Servicios legales, Bufetes jurídicos, Notarías Públicas, Servicios de contabilidad y auditoría, Otros servicios relacionados con la contabilidad,}\}$$

En la salida de esta etapa tenemos a C_H y R_H , la cual sirve de entrada para la etapa de caracterización, ya que para la generación del perfil de usuario se toman los conceptos de C_H

y para la etapa de recuperación, donde se utiliza la jerarquía de conceptos en el proceso de expansión.

3.3 Caracterización

En el presente trabajo de investigación se persigue realizar una personalización de las preferencias del usuario con base en un contexto definido [Sieg, 2007]. Dichas preferencias construyen el perfil del usuario, el cual está compuesto por un conjunto de conceptos relevantes para el usuario, existentes en la conceptualización (jerarquía de conceptos).

La personalización de la búsqueda y la navegación proporciona información que coincide con los intereses personales de los usuarios, por lo tanto provee un acceso más eficaz y eficiente a la información. Para la personalización se está utilizando el Diseño Centrado en el Usuario [Wealands, 2007], para esto empleamos un enfoque de perfiles de usuario. Esta etapa se enfoca a tomar las preferencias del usuario para personalizar la recuperación, para esto se genera un perfil de usuario, el cual está compuesto por conceptos obtenidos de C_H (Fig. 3.7) conjunto generado por la etapa de conceptualización.

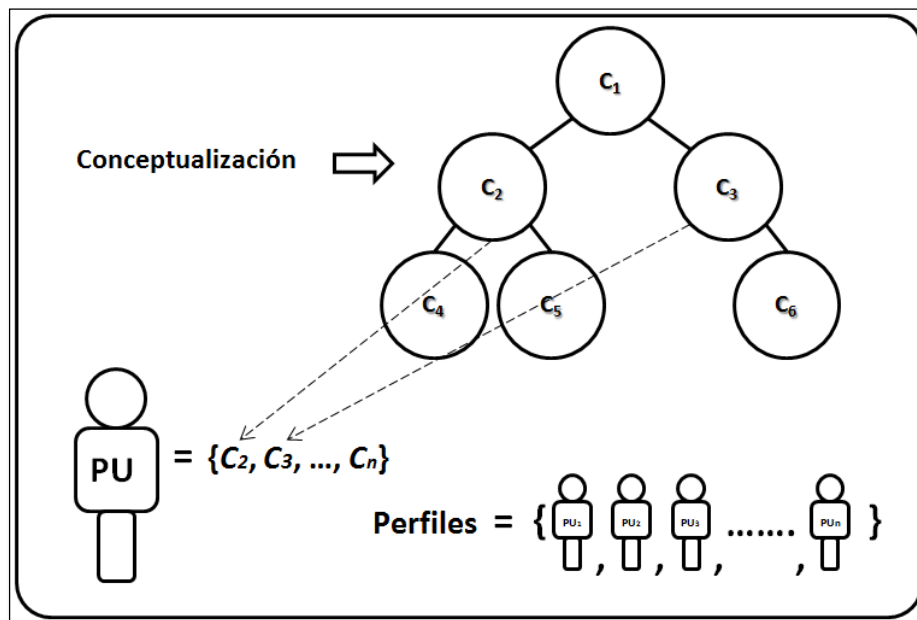


Figura 3.7 Perfil de Usuario formado por conceptos de la ontología.

3.3.1 Implementación del Perfil de Usuario

El perfil de usuario se ha introducido con el fin de registrar las preferencias del usuario y personalizar las aplicaciones para ser adaptadas a las necesidades de los usuarios. La construcción de los perfiles de usuario se define con base en la expresión matemática 3.1 como se muestra a continuación:

$$P = \{c_1, c_2, c_3, \dots, c_k\} \mid c \in C_H ; k > 0 \quad (3.2)$$

Donde; P es el perfil de usuario compuesto por un conjunto de conceptos de jerarquía de conceptos, c es un concepto que relevante para el usuario que pertenece al conjunto C_H , y forma parte del conocimiento de las preferencias del usuario. Por lo tanto.

$$P \subseteq C_H \quad (3.3)$$

Por otro lado se tiene un conjunto de perfiles de usuario para ejecutar la recuperación:

$$Perfiles = \{P_1, P_2, P_3, \dots, P_l\} \mid l > 0 \quad (3.4)$$

Donde; $Perfiles$ es el conjunto de perfiles de usuario para ejecutar la recuperación. Esta etapa tiene como salida las preferencias del usuario como un conjunto de conceptos, los cuales conforman las preferencias de información del usuario.

Ejemplificando esta etapa los perfiles de usuario estarán compuestos por conceptos de C_H , para formar dos perfiles tenemos que:

$$P_1 = \{\text{Bufetes jurídicos, Notarías publicas}\}$$

$$P_2 = \{\text{Servicios de contabilidad y auditoría, Otros servicios relacionados con la contabilidad}\}$$

Por lo tanto:

$$Perfiles = \{P_1, P_2\}$$

La salida de esta etapa forma parte de la entrada para la etapa de Recuperación. En la siguiente y última etapa se mostrarán detalles de implementación.

3.4 Recuperación

El objetivo de esta etapa es obtener los objetos geográficos que son relevantes para el usuario con base en su ubicación. La conceptualización, caracterización y la ubicación del usuario, forman parte de la entrada para esta etapa, y responden a las preguntas ¿Quién? y ¿Dónde está?, las cuales personalizan la recuperación.

3.4.1 Arquitectura

La arquitectura de la etapa de recuperación (Fig. 3.8), se encuentra basada en el sistema de recuperación de información general mostrado en [Baeza-Yates, 2011], donde el primer paso para el desarrollo de la recuperación es tener la colección de documentos, esta colección la tenemos dada por la BDE en conformada por un conjunto de objetos geográficos (colección de documentos en IR).

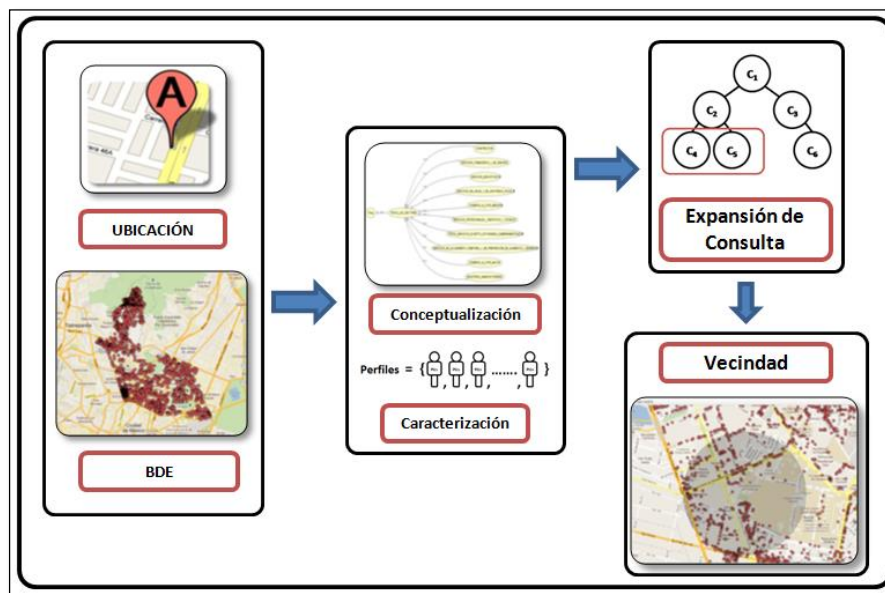


Figura 3.8 Arquitectura de la etapa de recuperación.

Los objetos geográficos en el repositorio central necesitan ser indexados para una recuperación y clasificación rápida, nuestra indexación está dada por el SCIAN⁶ del cual se habla a detalle en el siguiente capítulo. Con lo anterior el proceso de recuperación puede iniciarse, y consiste en la recuperación de documentos que satisfacen la consulta del usuario [Baeza-Yates, 2011], que para el caso de esta metodología a los documentos los llamaremos objetos geográficos. Para la búsqueda, el usuario primero especifica una consulta que refleja la información que necesita, en la metodología desarrollada la consulta se efectúa por medio del perfil de usuario y su ubicación. Enseguida dicha consulta es analizada y expandida (variantes ortográficas de la palabra de consulta), en este bloque (expansión de la consulta) utilizamos la conceptualización para poder efectuar un proceso de similitud (*Expansión _Conceptos*) y expandir la consulta. Posteriormente la consulta expandida se procesa contra la colección de objetos geográficos ya indexados para recuperar un subconjunto de dicha colección, esto con base en una vecindad determinada por la ubicación del usuario.

Siguiendo el modelo de recuperación de [Baeza-Yates, 2011] esta etapa propone lo siguiente:

Un modelo de Recuperación de la Información es una 4-tupla $[D, Q, F, R(q_i, d_j)]$ donde:

1. **D** es un conjunto compuesto del punto de vista lógico (o representaciones) de los documentos en una colección. Nuestra metodología toma al conjunto colección de objetos geográficos.
2. **Q** es un conjunto compuesto del punto de vista lógico (o representaciones) de las necesidades de información del usuario. Nuestra metodología toma al conjunto *Perfiles*.
3. **F** es un entorno (marco de trabajo) para modelar representaciones de documentos, consultas, y sus relaciones, tales como conjuntos y relaciones booleanas, vectores y

⁶Sistema de Clasificación Industrial de América del Norte, clasifica las actividades económicas.

operaciones de álgebra lineal, espacios de muestra y distribuciones de probabilidad. Nuestra metodología toma a la 2-tupla $H (C_H, R_H)$.

4. $R (q_i, d_j)$ es la función de ranking que asociada a un número real con una representación de la consulta $q_i \in \mathbf{Q}$ y una representación del documento $d_j \in \mathbf{D}$. Tal ranking define un ordenamiento entre los documentos con respecto a la consulta q_i . Para el caso de nuestra metodología esta función la sustituimos por la de *Expansión _Conceptos*, la cual retorna un conjunto de conceptos relacionados semánticamente para ejecutar la recuperación. Este método se describe enseguida.

3.4.1.1 Expansión de la Consulta

Es aquí donde se utiliza el conjunto de conceptos de la etapa de conceptualización (C_H), el conjunto de conceptos del perfil de usuario (P). Analizando la literatura, se encuentran trabajos como [Rodríguez, Egenhofer, 2004] y el de [Janowicz, 2008], donde la recuperación de información basada en semántica toma las conceptualizaciones en cuenta para calcular la relevancia de los conceptos. Tomando en cuenta un enfoque jerárquico, se puede distinguir entre dos enfoques para la recuperación de conceptos: los que se basan en un razonamiento de subsumción (*subsumption*) clásico y los que se basan en medidas de similitud. Simplificando, el razonamiento *subsumption* se puede aplicar a la búsqueda vertical, mientras similitud funciona mejor para la búsqueda horizontal (Fig. 3.9).

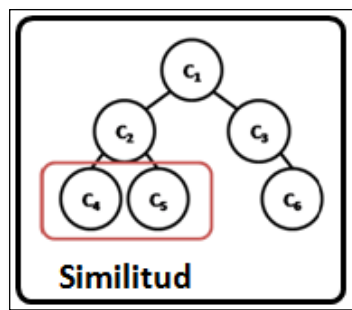


Figura 3.9 Conceptos similares con base en la jerarquía de conceptos.

Aplicando el enfoque de similitud, en la metodología vamos a trabajar con el concepto de similitud semántica ya que estamos tomando la jerarquía de conceptos generada por la

conceptualización, en la cual los conceptos ya tienen relacionado un significado. La función de expansión (*Expansión_Conceptos*) se describe en la Figura 3.8, siendo el retorno de este, un conjunto de conceptos similares semánticamente que enriquecerán la consulta y los resultados en el proceso de recuperación.

```

1  Expansión_Conceptos ( P )
2  {
3    Entrada: Conjunto de conceptos del perfil de usuario P.
4    Salida: Conjunto de conceptos similares con base en la conceptualización R.
5     $R \leftarrow \emptyset$ 
6    Hacer mientras  $P \neq \emptyset$ 
7       $x \leftarrow \text{primero}(P)$ 
8       $P \leftarrow P - \{x\}$ 
9       $X \leftarrow \text{hermanos}(x)$ 
10      $H \leftarrow \text{hijos}(x)$ 
11      $R \leftarrow R \cup \{x\}$ 
12     Para cada concepto  $s \in X$ 
13       Si  $R \cap \{s\} = \emptyset$ 
14          $P \leftarrow P \cup \{s\}$ 
15       Fin-si
16     Fin-para
17     Para cada concepto  $h \in H$ 
18       Si  $R \cap \{h\} = \emptyset$ 
19          $P \leftarrow P \cup \{h\}$ 
20       Fin-si
21     Fin-para
22   Fin-mientras
23   Retorna R;
24 }
```

Figura 3.10 Método de expansión en la recuperación.

Para ejemplificar este el proceso de expansión tomamos a P_1 , ejecutamos la función propuesta con base en la jerarquía de las actividades (Fig. 3.11):

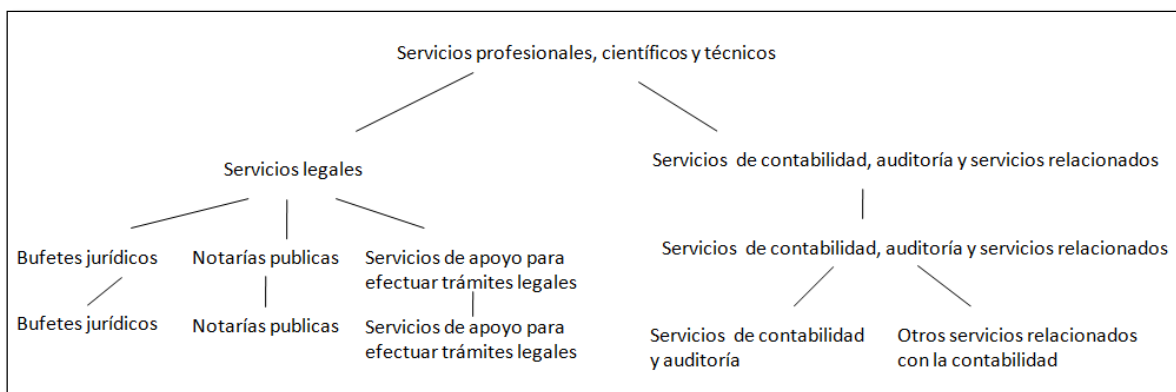


Figura 3.11 Jerarquía de conceptos ejemplo.

Como resultado de ejecutar la función (Fig. 3.12) tenemos:

$Expansión_Conceptos(P_1) = \{\text{Bufetes jurídicos, Notarías Públicas, Servicios de apoyo para efectuar trámites legales}\}$

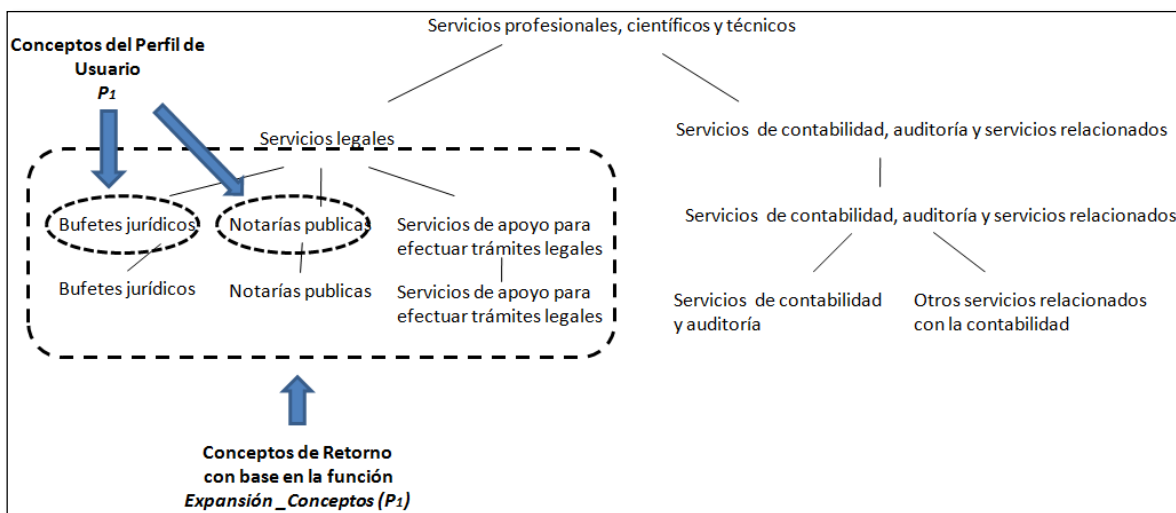


Figura 3.12 Retorno de la función $Expansión_Conceptos(P_1)$.

Con el conjunto de conceptos resultado de la expansión y la vecindad (determinada con la ubicación) se ejecuta la recuperación de los objetos geográficos, y el resultado es el que se muestra y devuelve al usuario.

3.4.1.2 Base de Datos Espacial (BDE).

La colección de objetos geográficos mencionada en la arquitectura, está dada por una Base de Datos Espacial (BDE). Una base de datos espacial es un sistema administrador de bases de datos que maneja datos existentes en un espacio o datos espaciales, y cuenta con las herramientas para trabajar este tipo de datos (Ver Anexo A). La construcción de una base de datos geográfica implica un proceso de abstracción para pasar de la complejidad del mundo real a una representación simplificada que pueda ser procesada por el lenguaje de las computadoras.

En el caso de esta metodología la BDE tiene que ser diseñada de tal forma que se pueda tener una indexación de los objetos geográficos, los atributos espaciales y temáticos deben ser diseñados con base en la conceptualización para su posterior recuperación, esto se explica y ejemplifica a detalle en el siguiente capítulo.

El Directorio Estadístico Nacional de Unidades Económicas [DENUE, 2011] ofrece información sobre la identificación y ubicación de todos los establecimientos activos en el territorio nacional con base en el Sistema de Clasificación Industrial de América del Norte 2007 [SCIAN, 2007].

Los datos del DENUE contienen a toda la República Mexicana, de los cuales solo vamos a trabajar con la Delegación Gustavo A. Madero para efectos de la implementación (Fig. 4.14). El diccionario de datos se encuentra en [SCIAN, 2007], el fragmento de relevancia para la implementación se muestra en la Tabla 3.3.

La BDE es utilizada en la recuperación como colección de objetos geográficos, los cuales son instancias de la ontología de aplicación, por lo tanto lo que se hace es un mapeo de los conceptos de la ontología (expandidos) contra las instancias de la BDE, finalmente estas son las que se muestran al usuario como resultado.

Tabla 3.3 Diccionario de datos DENUE (fragmento).

Nombre de la columna en el archivo CSV	Nombre de la columna en el archivo DBF	Tipo de dato	Longitud	Descripción
Nombre de la Unidad Económica	nom_estab	alfanumérico	75	Es el nombre comercial o nombre exterior con el que se identifica o anuncia la unidad económica. Generalmente es fácil reconocerlo porque está visible y escrito en rótulos, fachadas o anuncios luminosos. Ejemplo: ABARROTES LUPITA. Cuando la unidad económica no tiene ningún apelativo comercial, entonces se registra con el nombre genérico de la actividad o la descripción de la actividad económica, ejemplo: ABARROTES o Comercio al por menor en tiendas de abarrotes, ultramarinos y misceláneas.
Código de la clase de actividad	clase_act	alfanumérico	6	La clasificación de las actividades desarrolladas por las unidades económicas debe hacerse utilizando el SCIAN y sus actualizaciones, con base en lo dispuesto en el acuerdo para el uso del SCIAN y en la recopilación, análisis y presentación de estadísticas económicas, publicado en el Diario Oficial de la Federación del 10 de julio de 2009.
Nombre de clase de la actividad	desc_act	alfanumérico	160	Descripción del código de actividad conforme al SCIAN 2007.
Latitud	latitud	alfanumérico	12	Es la distancia que existe entre la Unidad económica y el ecuador, medida sobre el meridiano que pasa por dicho punto.
Longitud	longitud	alfanumérico	12	Es la distancia que existe entre la Unidad económica y el meridiano de Greenwich, medida sobre el paralelo que pasa por dicho punto.

3.4.1.3 Vecindad

La ubicación del usuario es parte de los datos de entrada para la metodología, esta funciona para delimitar la recuperación de objetos geográficos. En la Figura 3.13 se muestra gráficamente la colección de objetos geográficos en la parte izquierda y el conjunto de objetos geográficos producto de delimitar la recuperación con base en la ubicación del usuario, estableciendo con esto una vecindad.

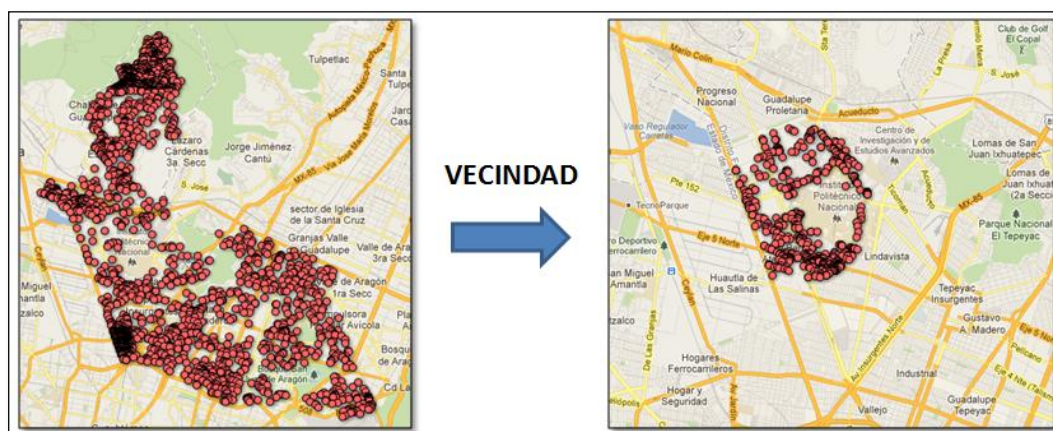


Figura 3.13 Recuperación de objetos geográficos con base en una vecindad.

La vecindad, parte teniendo en cuenta que los objetos geográficos recuperados son relevantes con base a la distancia del usuario. En la implantación de la metodología se muestra de manera explícita la relevancia de la ubicación.

En el siguiente capítulo abordaremos la aplicación desarrollada para probar la metodología y analizar los resultados, se verá a detalle cada una de las etapas de la metodología.

CAPÍTULO 4

IMPLANTACIÓN, PRUEBAS Y RESULTADOS

En este capítulo se presentan los resultados obtenidos con la implantación de la metodología propuesta. Se describe la implementación y su relación con la metodología, así como la interfaz de usuario desarrollada. Para la implementación, se debe tener en cuenta que no se está generando la ontología de aplicación de manera automática, sino que esta va a ser desarrollada con base en un dominio determinado.

4.1 Caso de Estudio

El Directorio Estadístico Nacional de Unidades Económicas [DENUE, 2011] ofrece información sobre la identificación y ubicación de todos los establecimientos activos en el territorio nacional con base en el Sistema de Clasificación Industrial de América del Norte 2007 [SCIAN, 2007], constituye una herramienta fundamental para la toma de decisiones en los ámbitos público y privado (Ver Anexo A). El caso de estudio que implementará la metodología es con base en el DENUE, donde el usuario podrá hacer una consulta de los establecimientos, obteniendo como resultado un conjunto de locales o negocios generados por la metodología.

4.2 Conceptualización: Ontología de Aplicación

El DENUE contiene información de identificación y ubicación de cada unidad económica. Entre los datos de identificación están el nombre de la unidad económica, denominación o

razón social y tipo de sociedad, estrato de personal ocupado, la actividad principal de la unidad económica (el código y descripción de clase de actividad económica) y tipo de unidad económica. El código de la clase de actividad fue asignado con base en el SCIAN 2007, considerando tanto los productos, (bienes y servicios) desarrollados por el establecimiento como sus procesos de producción; mientras que, con base en el número de trabajadores reportado por las unidades económicas, se clasificaron las unidades en los estratos de personal ocupado definidos para el DENE. En pocas palabras el DENE ya proporciona una clasificación a su información, elaborada por expertos y siguiendo normas para su desarrollo. Lo anterior permite llevar a un lenguaje formal y desarrollar la ontología con la información que proporciona el DENE. En la metodología, el objetivo principal de utilizar una ontología, es pasar de una evaluación de las consultas con base en palabras a una evaluación con base en conceptos.

4.2.1 Lenguaje de Representación

A principios de 1990, un conjunto de lenguajes de representación de ontologías fueron desarrollados. Algunos se basan en la lógica de primer orden, principalmente basados en el formato de intercambio de conocimiento (KIF, Knowledge Interchange Format) [Genesereth, 1991], mientras que otros se basan en marcos combinados con la lógica de primer orden, por ejemplo, el lenguaje Ontolingua [Gruber, 1992]. El Internet ha dado lugar a una serie de lenguajes de representación basados en la web (Fig. 4.1), que para efectos de la implementación nos enfocaremos en OWL (OWL, Web Ontology Language). El lenguaje OWL [Dean, Schreiber, 2003] ha sido creado por el grupo de trabajo WebOnt de la W3C (W3C, World Wide Web Consortium).

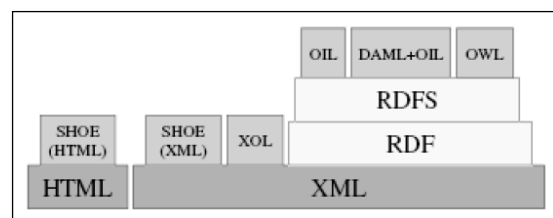


Figura 4.1 Lenguajes de marcado para ontologías.

El lenguaje OWL pertenece a una familia de lenguajes de representación del conocimiento para la creación de ontologías (Ver Anexo A). OWL puede describir también propiedades, instancias, relaciones y todo lo definido en las especificaciones de la W3C, además de contar con herramientas para su análisis y representaciones como razonadores y herramientas de representación gráfica.

4.2.2 Implementación de la Ontología

El desarrollo de la ontología se centra en la estructura del [SCIAN, 2007], de la cual se toman las actividades económicas, para la formalización utilizaremos OWL. Para este trabajo, la ontología es una base de conocimientos de un dominio de trabajo. En este sentido, una ontología captura una visión del mundo, admite consultas sobre el contenido en la base de datos, define la semántica de forma independiente de la representación de datos, y refleja la relevancia de los datos sin necesidad de acceder a ellos [Goñi, et. al., 1998]. Las relaciones semánticas son una forma típica de describir el conocimiento acerca de los conceptos. En el lenguaje natural de comunicación, por ejemplo, sinonimia⁷, antonimia⁸, hiponimia⁹, meronimia¹⁰ y la implicación son ejemplos de las relaciones semánticas utilizadas para definir los términos [Miller, 1995].

Nos referimos a las clases de entidad por medio de palabras o conjuntos de sinónimos, los cuales están relacionados entre sí por relaciones de hiponimia y de meronimia. La relación de hiponimia, generalmente llamada relación *es-un* [Smith, 1977], es la relación más común

⁷ Es una relación de semejanza de significados entre determinadas palabras.

⁸ Es una relación donde las palabras tienen un significado opuesto o contrario.

⁹ En semántica lingüística, se denomina hipónimo (del griego, que literalmente significa 'pocos nombres') a aquella palabra que posee todos los rasgos semánticos, de otra más general (su hiperónimo) pero que añade en su definición otros rasgos semánticos que la diferencian de la segunda. En el mundo de las Tecnologías de Información se podría encontrar el símil en la Orientación a Objetos, donde el hipónimo podría ser una clase (u objeto dependiendo del caso) y el hiperónimo la superclase (o clase).

¹⁰ La meronimia es una relación semántica no-simétrica entre los significados de dos palabras dentro del mismo campo semántico. Se denomina merónimo a la palabra cuyo significado constituye una parte del significado total de otra palabra, denominada ésta holónimo. Por lo tanto: X es merónimo de Y si X forma parte de Y.

que se utiliza en una ontología. La relación *es-un* es transitiva¹¹ y asimétrica¹², define una estructura jerárquica donde los términos heredan todas las características de sus términos de orden superior [Rodríguez, Egenhofer, 2004]. La relación *es-un* es utilizada para definir las relaciones entre conceptos dentro de la ontología de aplicación para nuestra metodología. Para el desarrollo de la ontología se utilizó Protégé (Ver Anexo A). Protégé es una plataforma de código abierto que provee una creciente comunidad de usuarios con un conjunto de herramientas para la construcción de modelos de dominio y las aplicaciones basadas en el conocimiento con ontologías. En su esencia, Protégé implementa un amplio conjunto de estructuras de modelado del conocimiento y las acciones que apoyan la creación, visualización y manipulación de ontologías en diversos formatos de representación. Además, Protégé se puede ampliar por medio de una arquitectura plug-in y una API (Application Programming Interface) basada en Java [Stanford, 2012].

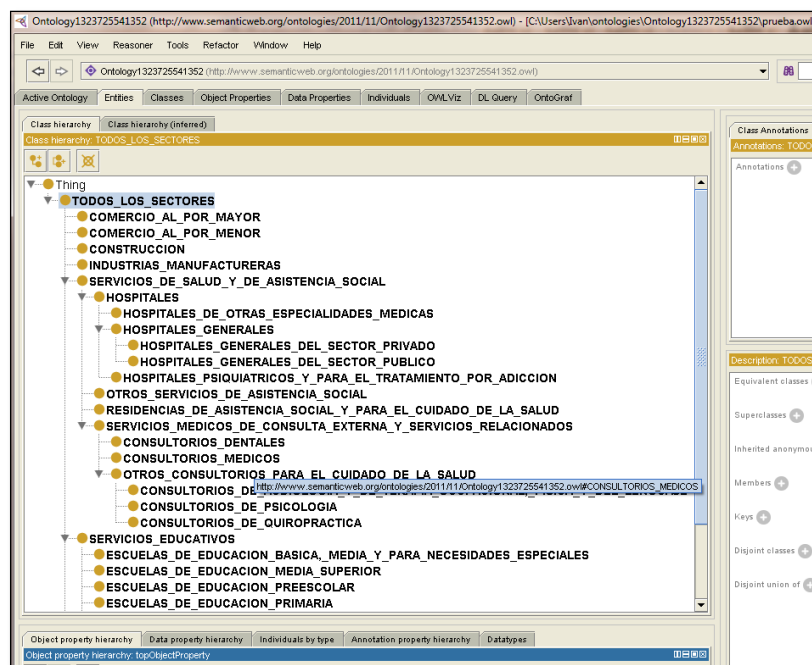


Figura 4.2 Desarrollo de la Ontología en Protégé.

Únicamente trabajamos una parte de las actividades del SCIAN para desarrollar la ontología en Protégé (Fig. 4.2) debido a que en la implantación de los perfiles de usuario no se

¹¹ Una relación es transitiva cuando se cumple: siempre que un elemento se relaciona con otro y éste último con un tercero, entonces el primero se relaciona con el tercero.

¹² Es una relación que se establece entre conceptos que están en desigualdad de nivel o jerarquía.

seleccionan todas las actividades. La relación *es-un* la podemos ver gráficamente (Fig. 4.3), y representa la relación que tienen los conceptos dentro de la ontología utilizando la estructura del SCIAN, lo que permite establecer la semántica de los conceptos y la base de conocimientos para la recuperación inteligente de información.

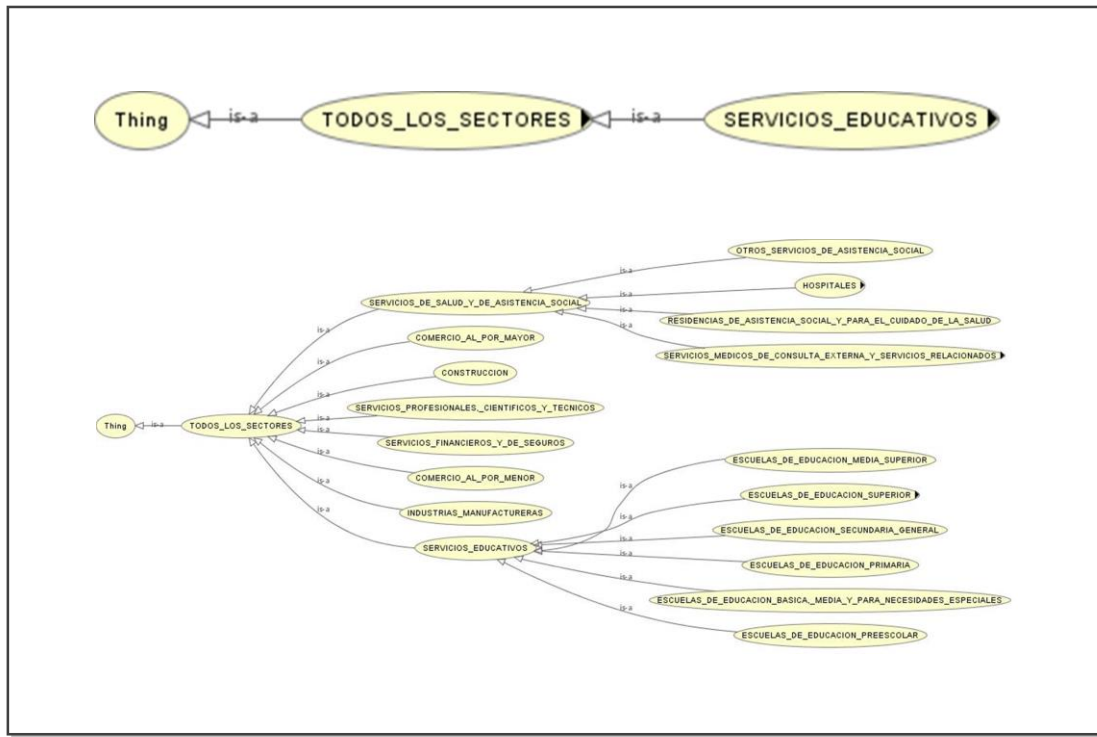


Figura 4.3 Relación *es-un* entre conceptos de la Ontología.

La formalización por medio del OWL se llevó a cabo empleando Protégé con base en los datos del DENUE, generando un archivo owl, a continuación se muestra un fragmento del mismo:

```

1  <?xml version="1.0"?>
2
3
4  <!DOCTYPE Ontology [
5    <!ENTITY xsd "http://www.w3.org/2001/XMLSchema#" >
6    <!ENTITY xml "http://www.w3.org/XML/1998/namespace" >
7    <!ENTITY rdfs "http://www.w3.org/2000/01/rdf-schema#" >
8    <!ENTITY rdf "http://www.w3.org/1999/02/22-rdf-syntax-ns#" >
9  ]>
10
11
12  <Ontology xmlns="http://www.w3.org/2002/07/owl#"
13    xml:base="http://www.semanticweb.org/ontologies/2011/11/Ontology1323725541352.owl"
14    xmlns:xsd="http://www.w3.org/2001/XMLSchema#"
15    xmlns:xml="http://www.w3.org/XML/1998/namespace"
16    xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
17    xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
18    ontologyIRI="http://www.semanticweb.org/ontologies/2011/11/Ontology1323725541352.owl">
19    <Prefix name="rdf" IRI="http://www.w3.org/1999/02/22-rdf-syntax-ns#" />
20    <Prefix name="" IRI="http://www.w3.org/2002/07/owl#" />
21    <Prefix name="xsd" IRI="http://www.w3.org/2001/XMLSchema#" />
22    <Prefix name="rdfs" IRI="http://www.w3.org/2000/01/rdf-schema#" />
23    <Prefix name="owl" IRI="http://www.w3.org/2002/07/owl#" />
24    <Declaration>

```

Figura 4.4 Código en OWL de la Ontología.

```

25   <Class IRI="#COMERCIO_AL_POR_MAYOR"/>
26 </Declaration>
27 <Declaration>
28   <Class IRI="#COMERCIO_AL_POR_MENOR"/>
29 </Declaration>
30 <Declaration>
31   <Class IRI="#CONSTRUCCION"/>
32 </Declaration>
33 <Declaration>
34   <Class IRI="#CONSULTORIOS_DENTALES"/>
35 </Declaration>
36 <Declaration>
37   <Class IRI="#CONSULTORIOS_DE_AUDIOLOGIA_Y_DE_TERAPIA_OCUPACIONAL,_FISICA_Y_DEL LENGUAJE"/>
38 </Declaration>
39 <Declaration>
40   <Class IRI="#CONSULTORIOS_DE_PSICOLOGIA"/>
41 </Declaration>
42 <Declaration>
43   <Class IRI="#CONSULTORIOS_DE_QUIROPRACTICA"/>
44 </Declaration>
45 <Declaration>
46   <Class IRI="#CONSULTORIOS_MEDICOS"/>
47 </Declaration>

```

↓

```

200 <SubClassOf>
201   <Class IRI="#HOSPITALES_GENERALES_DEL_SECTOR_PUBLICO"/>
202   <Class IRI="#HOSPITALES_GENERALES"/>
203 </SubClassOf>
204 <SubClassOf>
205   <Class IRI="#HOSPITALES_PSIQUIATRICOS_Y_PARA_EL_TRATAMIENTO_POR_ADICCION"/>
206   <Class IRI="#HOSPITALES"/>
207 </SubClassOf>
208 <SubClassOf>
209   <Class IRI="#INDUSTRIAS_MANUFACTURERAS"/>
210   <Class IRI="#TODOS_LOS_SECTORES"/>
211 </SubClassOf>
212 <SubClassOf>
213   <Class IRI="#OTROS_CONSULTORIOS_PARA_EL_CUIDADO_DE_LA_SALUD"/>
214   <Class IRI="#SERVICIOS_MEDICOS_DE_CONSULTA_EXTERNA_Y_SERVICIOS_RELACIONADOS"/>
215 </SubClassOf>
216 <SubClassOf>
217   <Class IRI="#OTROS_SERVICIOS_DE_ASISTENCIA_SOCIAL"/>
218   <Class IRI="#SERVICIOS_DE_SALUD_Y_DE_ASISTENCIA_SOCIAL"/>
219 </SubClassOf>
220 <SubClassOf>
221   <Class IRI="#RESIDENCIAS_DE_ASISTENCIA_SOCIAL_Y_PARA_EL_CUIDADO_DE_LA_SALUD"/>
222   <Class IRI="#SERVICIOS_DE_SALUD_Y_DE_ASISTENCIA_SOCIAL"/>
223 </SubClassOf>
224 <SubClassOf>
225   <Class IRI="#SERVICIOS_DE_SALUD_Y_DE_ASISTENCIA_SOCIAL"/>
226   <Class IRI="#TODOS_LOS_SECTORES"/>
227 </SubClassOf>
228 <SubClassOf>
229   <Class IRI="#SERVICIOS_EDUCATIVOS"/>
230   <Class IRI="#TODOS_LOS_SECTORES"/>
231 </SubClassOf>
232 <SubClassOf>
233   <Class IRI="#SERVICIOS_FINANCIEROS_Y_DE_SEGUROS"/>
234   <Class IRI="#TODOS_LOS_SECTORES"/>
235 </SubClassOf>
236 <SubClassOf>
237   <Class IRI="#SERVICIOS_MEDICOS_DE_CONSULTA_EXTERNA_Y_SERVICIOS_RELACIONADOS"/>
238   <Class IRI="#SERVICIOS_DE_SALUD_Y_DE_ASISTENCIA_SOCIAL"/>
239 </SubClassOf>
240 <SubClassOf>
241   <Class IRI="#SERVICIOS_PROFESIONALES,_CIENTIFICOS_Y_TECNICOS"/>
242   <Class IRI="#TODOS_LOS_SECTORES"/>
243 </SubClassOf>
244 </Ontology>

```

Figura 4.4 (Cont...) Código en OWL de la Ontología.

Cabe resaltar que para fines de la implementación de la metodología, en la formalización, no estamos utilizando propiedades, atributos o instancias ya que trabajaremos con una BDE en la que vamos a llevar a cabo la recuperación (Fig. 4.5).

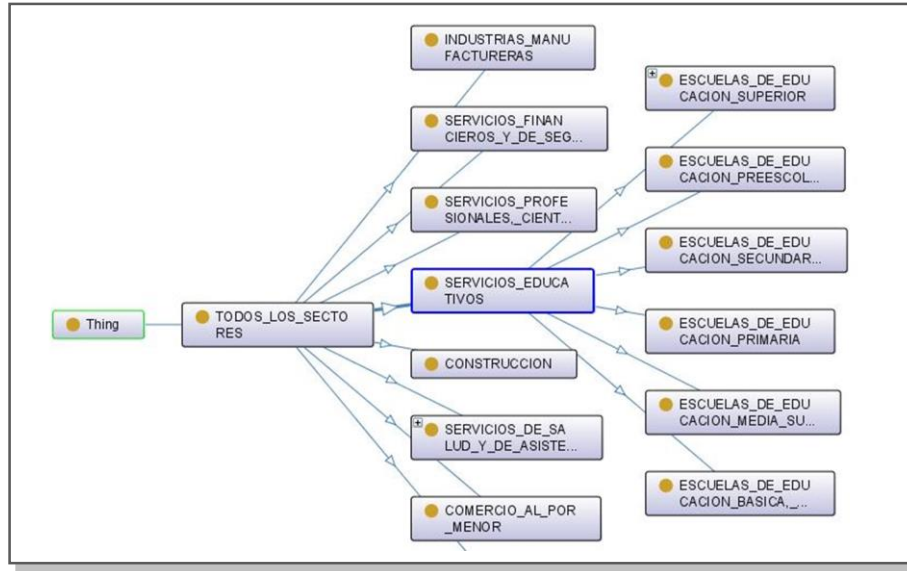


Figura 4.5 Representación de conceptos usados en la Ontología.

La etapa de conceptualización genera como salida una descripción del dominio de trabajo de manera formal para que pueda ser procesada por la computadora, la cual también es la entrada para la etapa de caracterización y de la etapa de recuperación.

4.3 Caracterización: Implementación del Perfil de Usuario

Para implementar esta etapa, se generaron perfiles de usuario, los cuales contienen conceptos obtenidos de la ontología (Fig. 4.6). Para efectos de la implementación, definiremos dos perfiles de usuario (Fig. 4.7), tomando conceptos generados por la Conceptualización:

- a) **Estudiante:** {ESCUELAS DE EDUCACION SUPERIOR, COMERCIO AL POR MENOR DE ABARROTES Y ALIMENTOS, RESTAURANTES CON SERVICIO COMPLETO}

- b) **Profesor:** {HOSPITALES GENERALES, REPARACION MECANICA EN GENERAL DE AUTOMOVILES Y CAMIONES, COMERCIO AL POR MENOR DE ROPA EXCEPTO DE BEBE Y LENCERIA}

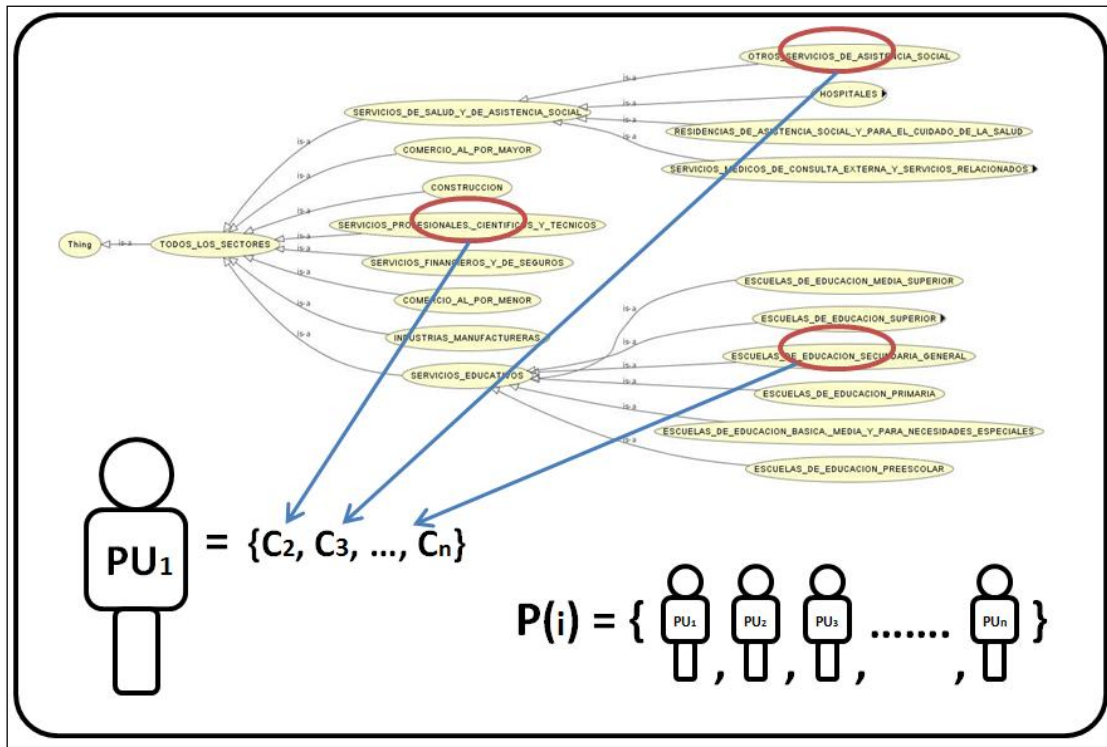


Figura 4.6 Perfil de Usuario formado por conceptos de la ontología.

La visualización de conceptos se lleva a cabo por medio de una interfaz de tipo arborescente, para mostrar los conceptos y su nivel jerárquico dentro de la ontología (Fig. 4.8).

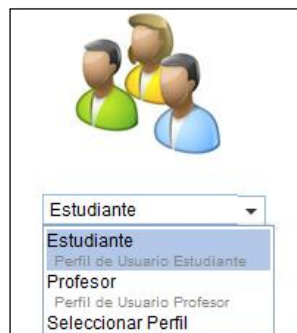


Figura 4.7 Perfiles de Usuario en la implementación.

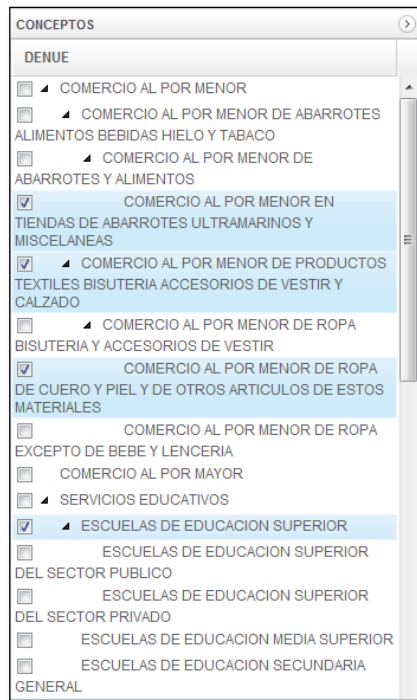


Figura 4.8 Visualización de conceptos en la implementación.

Esta interfaz tiene como salida las preferencias del usuario como un conjunto de conceptos, los cuales conforman las preferencias de información del usuario formando parte de la entrada para la implementación de la recuperación.

4.4 Recuperación: Implementación de Ubicación, BDE, Expansión y Resultados

En esta etapa, la implementación fue desarrollada para devolver al usuario los objetos geográficos relevantes con base en su perfil y a su ubicación. La conceptualización, caracterización y la ubicación del usuario, forman parte de la entrada de datos, respondiendo a las preguntas ¿Quién? y ¿Dónde?, las cuales personalizan la recuperación.

La implementación se basa en la arquitectura de la etapa de recuperación (Fig. 3.8). Los objetos geográficos en el repositorio central necesitan ser indexados para una recuperación y clasificación rápida, nuestra indexación está dada por el SCIAN. Con lo anterior el proceso de recuperación puede iniciarse, y consiste en la recuperación de objetos geográficos que satisfacen la consulta del usuario [Baeza-Yates, 2011]. Para la búsqueda, el usuario primero

especifica una consulta que refleja la información que necesita, en la metodología desarrollada la consulta se efectúa por medio del perfil de usuario y su ubicación. Enseguida dicha consulta es analizada y expandida (variantes ortográficas de la palabra de consulta), en este bloque (expansión de la consulta) utilizamos la conceptualización para poder efectuar un proceso de similitud y expandir la consulta. Posteriormente la consulta expandida se procesa contra la colección de objetos geográficos ya indexados para recuperar un subconjunto de estos, con base en la ubicación del usuario (vecindad de objetos geográficos).

4.4.1 Ubicación del Usuario

La ubicación del usuario es parte de los datos de entrada para la metodología, para la implementación se hace uso del API de Google Maps y el Framework Java ZK para desarrollo Web (Ver Anexo A). Por medio de una interfaz grafica, el usuario puede ingresar su ubicación como ya sea por medio de un clic de su ubicación en el mapa, o ingresando sus coordenadas. Las coordenadas están expresadas usando números decimales separados por coma. La latitud siempre precede la longitud. La latitud es positiva si va después del punto mostrado en el mapa y negativa si va antes. La longitud es positiva si va arriba del punto y negativa si va debajo (Fig. 4.9).

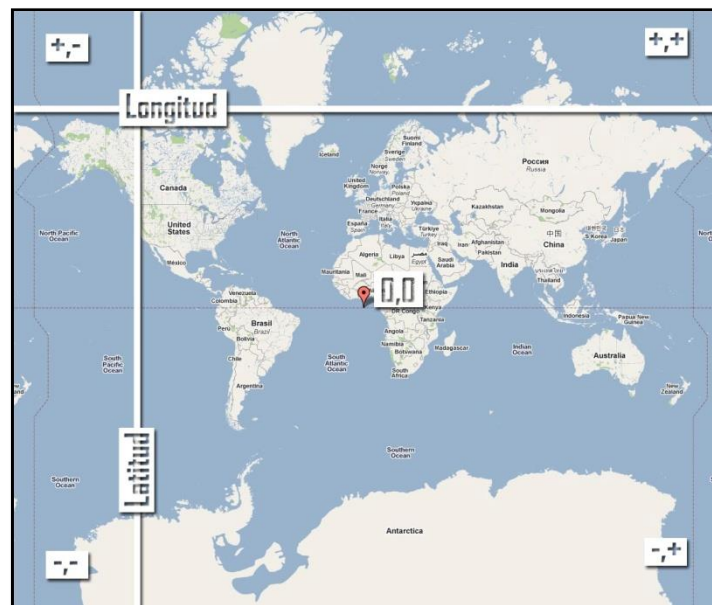


Figura 4.9 Sistema de coordenadas de Google Maps.

Por ejemplo de manera general en los mapas, las coordenadas están expresadas en grados, así que la posición de Puerto Rico sería: 18°14'70" N 66°29'68" W. La forma de convertir estos datos a decimales sería:

$$(18^{\circ}14'70'' \text{ N}) = (18 + (14 / 60) + (70 / 3600)) = 18.252$$

$$(66^{\circ}29'68'' \text{ W}) = -(66 + (29 / 60) + (68 / 3600)) = -66.8627$$

La longitud se multiplica por negativo, porque está a la izquierda (oeste) del punto 0,0. Dentro de la implementación se cuenta con una interfaz grafica para seleccionar la ubicación sin tener que efectuar operaciones, pero en caso de que se tengan los datos de la latitud y la longitud, estos se pueden ingresar manualmente también (Fig. 4.10).

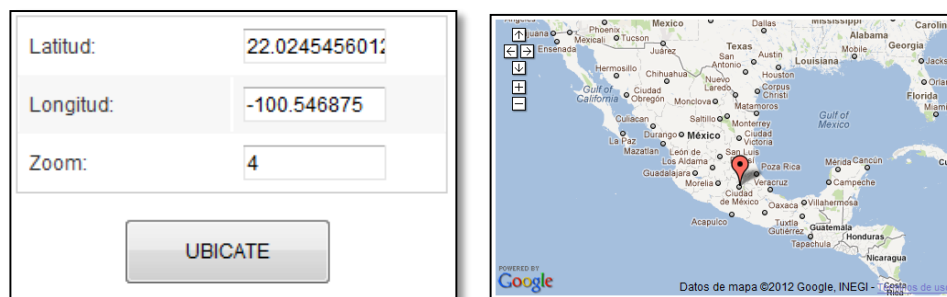


Figura 4.10 Ubicación del usuario en la Implementación.

4.4.2 Base de Datos Espacial

La colección de documentos mencionada en la arquitectura, está dada por una Base de Datos Espacial. Una base de datos espacial (Spatial DataBase, SDB) es un sistema administrador de bases de datos que maneja datos existentes en un espacio o datos espaciales, y cuenta con las herramientas para trabajar este tipo de datos (Ver Anexo A).

El DENUe proporciona la colección de datos, por medio de un *shapefile*. Un Shapefile es un formato vectorial de almacenamiento digital donde se guarda la localización de los elementos geográficos y los atributos asociados a ellos.

El DENUe proporciona estos archivos, los cuales trabajamos con GvSIG (Fig. 4.11), una herramienta que permite gestionar datos espaciales y realizar análisis a partir de estos (Ver Anexo A).

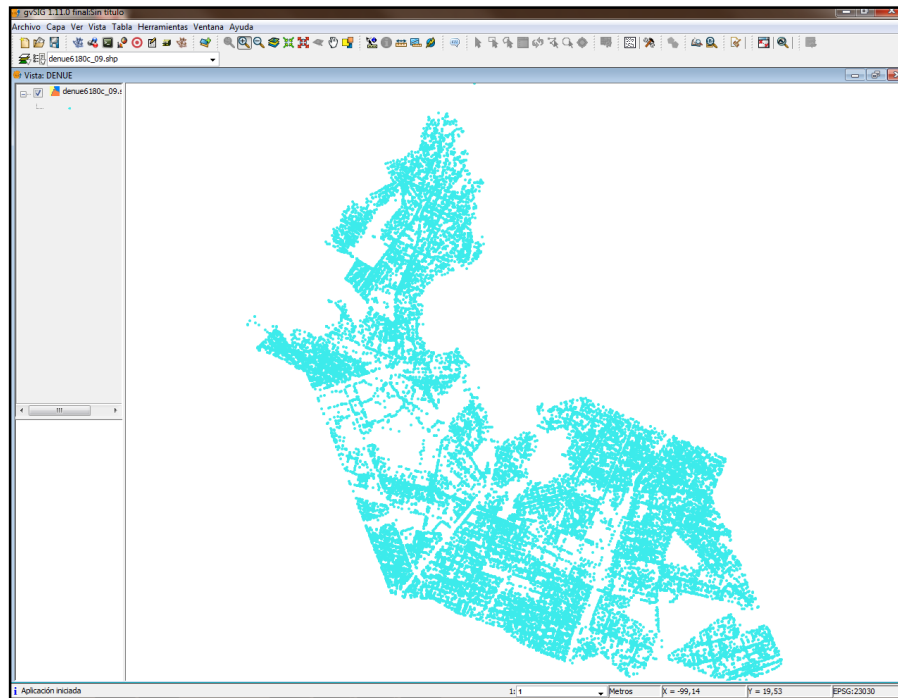


Figura 4.11 Datos del DENUe delegación Gustavo A. Madero en GvSIG.

La BDE se puede visualizar, editar y consultar por medio del GvSIG (Fig. 4.12). En la implementación la BDE que se ocupa, es una tecnología de Google llamada Google Fusion Tables [GFT, 2009]. GFT es un servicio gratuito de Google que permite visualizar, analizar, editar, compartir y combinar datos geoespaciales con la característica que están en la nube¹³. Los formatos de datos estándar para la importación son los datos de hojas de cálculo (incluida la exportación directa desde Google Docs), CSV (comma-separated values) y archivos KML (Keyhole Markup Language).

Para agregar los datos a GFT, primeramente se deben convertir los datos de Shapefile a KML, para lo cual se utilizó GvSIG. KML es un lenguaje de marcado basado en XML para representar datos geográficos en tres dimensiones. Un archivo KML especifica una

¹³Referencia al cómputo en nube.

característica (un lugar, una imagen o un polígono). Contiene título, una descripción básica del lugar, sus coordenadas (latitud y longitud) y alguna otra información. Una vez obtenido el archivo KML lo podemos importar a GFT (Fig. 4.13).

Las consultas se harán directamente en GFT utilizando su API junto con el API de Google Maps (Ver Anexo A), para poder mostrar los resultados.

NOM_ESTAB	NOM_PROPIE	CLASE_ACT	DESC_ACT	PERS_OCUP	TIPO_CALLE	CALLE	NUMERO_EXT	ED
FERRO COL...	FERRO MEXICANA SA ...	434222	COMERCIO AL POR MAYOR DE PRODUCTOS QUI...	101 A 250 P...	CALLE	ORIENTE 171	450	
2 HERMANOS		461110	COMERCIO AL POR MENOR EN TIENDAS DE ABA...	0 A 5 PERS...	CALLE	EVA SÁMAN...	M11L25	
RECONTRU...		811113	RECTIFICACION DE PARTES DE MOTOR DE AUT...	0 A 5 PERS...	AVENIDA	CENTENARIO	1801	
PESCADERI...		722219	OTROS RESTAURANTES CON SERVICIO LIMITADO	0 A 5 PERS...	CALLE	GENERAL VI...	124A	
MERCADO R...		811211	REPARACION Y MANTENIMIENTO DE EQUIPO EL...	0 A 5 PERS...	EJE VIAL	2 ORIENTE (...)	SN	
SOPE TACO		722212	RESTAURANTES DE COMIDA PARA LLEVAR	0 A 5 PERS...	CALLE	XOCHIQUET...	L5	
ASEO DE CA...		812130	SANITARIOS PUBLICOS Y BOLERIAS	6 A 10 PERS...	CALLE	NINGUNO	4709	
TEQUILA O...	TEQUILA ORENDAIN D...	431212	COMERCIO AL POR MAYOR DE VINOS Y LICORES	11 A 30 PER...	CALLE	PERNAMBUCO	775	
CARNICERI...		461121	COMERCIO AL POR MENOR DE CARNES ROJAS	0 A 5 PERS...	CALLE	XOCHIQUET...	164 B	
OFICINA CE...	INSTITUTO NACIONAL...	931210	ADMINISTRACION PUBLICA EN GENERAL	6 A 10 PERS...	CALZADA	SAN JUAN D...	311	
HAMBURGU...		722211	RESTAURANTES DE AUTOS					
SALON DE F...		531113	ALQUILER SIN INTERMEDIA					
RIMINI CAFE		311520	ELABORACION DE HELADOS					
BODEGA TA...		468211	COMERCIO AL POR MENOR					
ESTETICA A...		812110	SALONES Y CLINICAS DE BE					
RESTAURAN...		722212	RESTAURANTES DE COMIDA					
PROLIMEX P...		467115	COMERCIO AL POR MENOR					
SERVICIO P...		811112	REPARACION DEL SISTEMA					
ALCOHOLIC...		624191	AGRUPACIONES DE AUTOA					
VENTA DE B...		463215	COMERCIO AL POR MENOR					
TALLER DE ...		316999	FABRICACION DE OTROS P					
TORTILLERI...		311830	ELABORACION DE TORTILL					
MISCELANE...		461110	COMERCIO AL POR MENOR					
MATERIAS P...		465914	COMERCIO AL POR MENOR					
TINTORERI...		812210	LAVANDERIAS Y TINTORER					
INTERNET	PRICILA SANCHEZ	561432	SERVICIOS DE ACCESO A C					
MICHELADAS		722219	OTROS RESTAURANTES CO					
VERDURAS ...		461130	COMERCIO AL POR MENOR					

Figura 4.12 Base de Datos de atributos del DENUE.

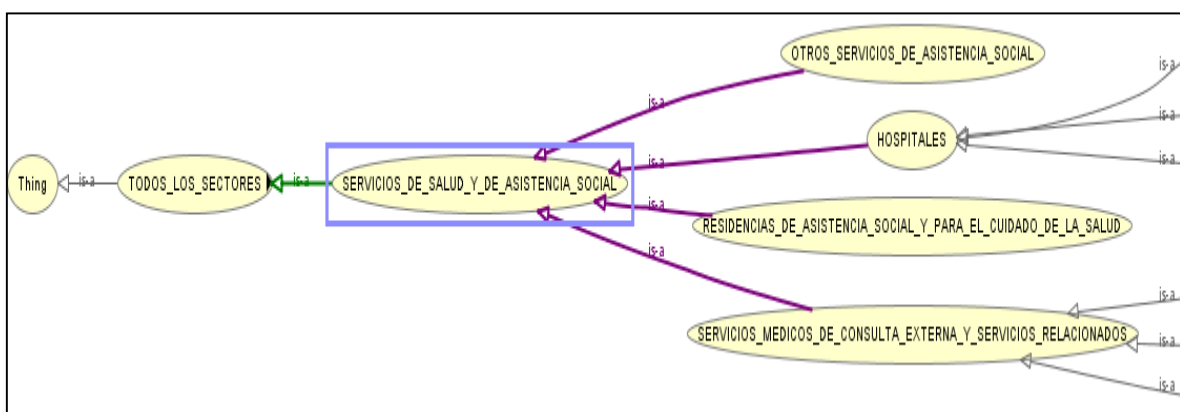


Figura 4.13 Base de Datos en Google Fusion Tables.

4.4.3 Implementación de la Similitud

Para poder llevar a la implementación el enfoque jerárquico de similitud utilizado en la metodología, utilizaremos la herramienta OWL API (Ver Anexo A). Esta API es para Java¹⁴, con ella se puede crear, editar, manipular ontologías de tipo OWL y es de código abierto. Además de esta API trabajaremos con un *reasoner*¹⁵ llamado Hermit [Hermit, 2012], razonador para ontologías en OWL.

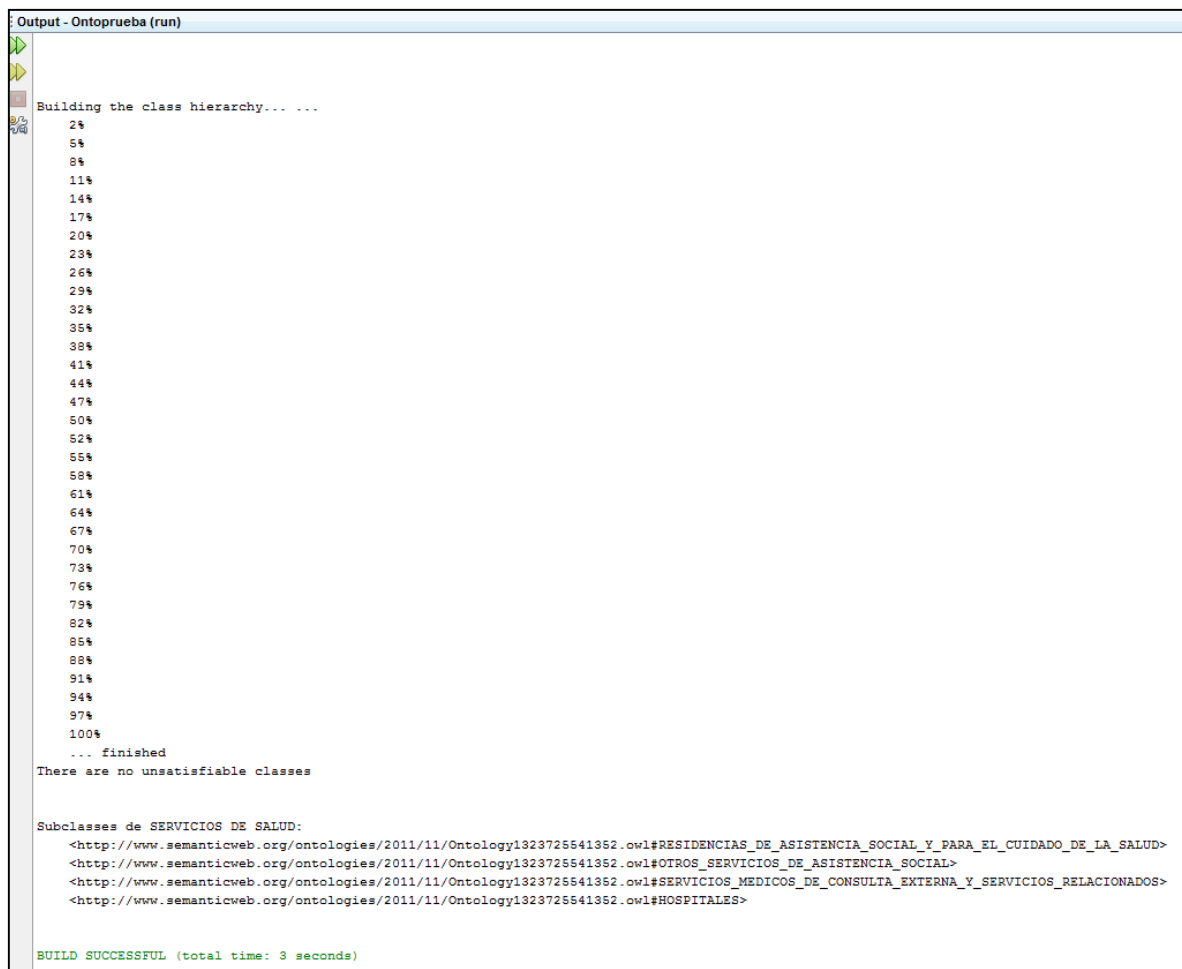
Dado un archivo OWL, Hermit puede determinar si la ontología es consistente, identificar las relaciones entre las clases, entre otras funcionalidades. Gracias a que Hermit tiene un API en Java, podemos utilizar sus Clases para poder trabajar la implementación en un entorno compatible con Java. En la implementación para la navegación de la ontología, se utiliza el API del *reasoner*, en específico la clase OWLReasoner, esta clase proporciona dos métodos que utilizamos para implementar la similitud, *getSubClasses()*, para obtener los conceptos hijo de un concepto padre, y el método *getSuperClasses()*, para obtener el concepto padre de un concepto hijo (Fig. 4.14). Con estos métodos podemos establecer la similitud semántica, donde se busca el concepto padre para posteriormente saber cuáles son sus conceptos hijo, y que a su vez son hermanos del concepto que se proporcionó al método, obteniendo así sus conceptos similares. Después de la ejecución se obtiene un resultado como se muestra en la figura siguiente:



(a) Archivo OWL (representación gráfica).

¹⁴Lenguaje de programación orientado a objetos.

¹⁵ Motor de inferencia lógica.



```

Output - Ontopueba (run)
Building the class hierarchy... ..
2%
5%
8%
11%
14%
17%
20%
23%
26%
29%
32%
35%
38%
41%
44%
47%
50%
52%
55%
58%
61%
64%
67%
70%
73%
76%
79%
82%
85%
88%
91%
94%
97%
100%
... finished
There are no unsatisfiable classes

Subclasses de SERVICIOS DE SALUD:
<http://www.semanticweb.org/ontologies/2011/11/Ontology1323725541352.owl#RESIDENCIAS_DE_ASISTENCIA_SOCIAL_Y_PARA_EL_CUIDADO_DE_LA_SALUD>
<http://www.semanticweb.org/ontologies/2011/11/Ontology1323725541352.owl#OTROS_SERVICIOS_DE_ASISTENCIA_SOCIAL>
<http://www.semanticweb.org/ontologies/2011/11/Ontology1323725541352.owl#SERVICIOS_MEDICOS_DE_CONSULTA_EXTERNA_Y_SERVICIOS_RELACIONADOS>
<http://www.semanticweb.org/ontologies/2011/11/Ontology1323725541352.owl#HOSPITALES>

BUILD SUCCESSFUL (total time: 3 seconds)
  
```

(b) Salida en consola de las clases similares utilizando el *reasoner*.

Figura 4.14 Salida en consola de los conceptos por similitud en la ontología.

Este proceso se lleva a cabo para el conjunto de conceptos que conforman el perfil de usuario, devolviendo un conjunto nuevo de conceptos, donde se encuentran los conceptos del perfil de usuario más los conceptos similares semánticamente hablando, dicho conjunto pasa a la expansión de la consulta para continuar con la etapa de recuperación.

4.4.4 Expansión de la consulta

En este proceso se ejecuta la consulta con el conjunto de conceptos resultado del proceso de similitud para generar la salida de la metodología (Fig. 4.15). Por último este proceso se encarga de ejecutar la consulta.

Para consultar los datos, se utiliza la siguiente sintaxis en una petición HTTP GET:

```
https://www.google.com/fusiontables/api/query?sql=<statement>
```

```
SELECT <column_spec> {, <column_spec>}* FROM <table_id>

{ WHERE <filter_condition> | <spatial_condition> { AND <filter_condition> }* }

{ GROUP BY <column_name> {, <column_name>}* }

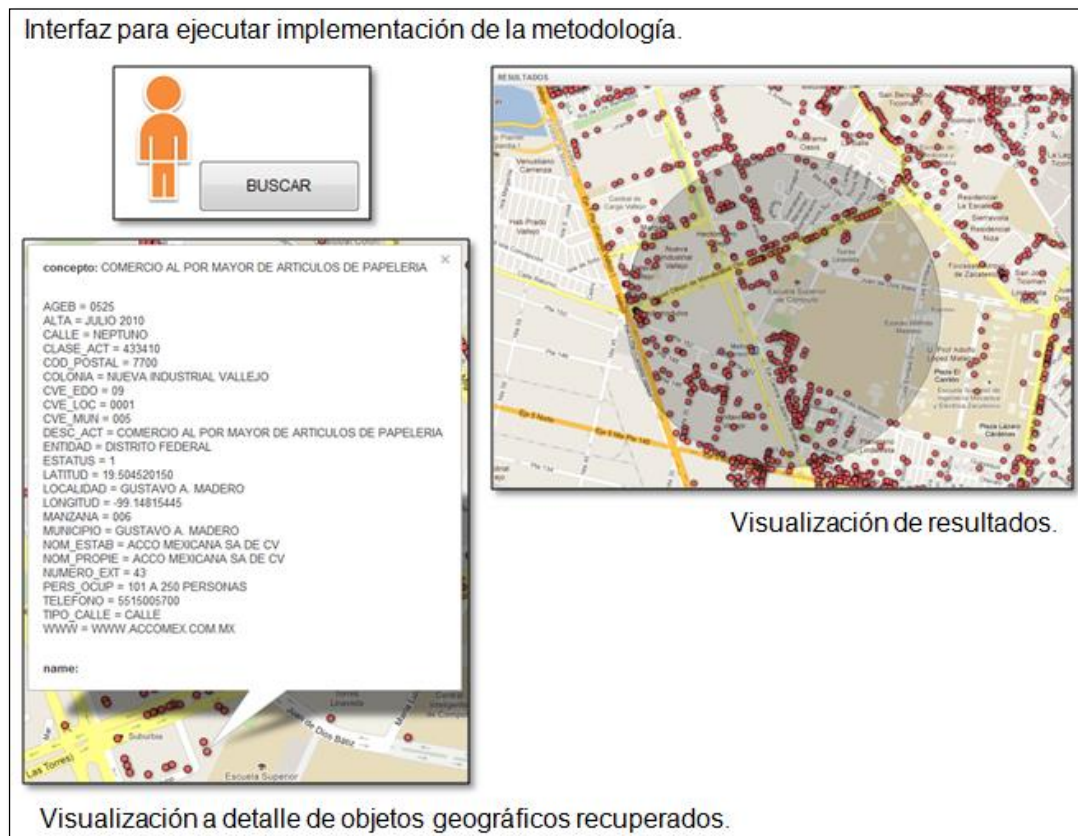
{ ORDER BY <column_spec> { ASC | DESC } | <spatial_relationship> }

{ OFFSET <number> }

{ LIMIT <number> }
```

De esta manera los resultados obtenidos se agregan como capas para su visualización con Google Maps.

Interfaz para ejecutar implementación de la metodología.



concepto: COMERCIO AL POR MAYOR DE ARTICULOS DE PAPELERIA

AGEB = 0525
 ALTA = JULIO 2010
 CALLE = NEPTUNO
 CLASE_ACT = 433410
 COD_POSTAL = 7700
 COLONIA = NUEVA INDUSTRIAL VALLEJO
 CVE_EDO = 09
 CVE_LOC = 0001
 CVE_MUN = 005
 DESC_ACT = COMERCIO AL POR MAYOR DE ARTICULOS DE PAPELERIA
 ENTIDAD = DISTRITO FEDERAL
 ESTATUS = 1
 LATITUD = 19.504520150
 LOCALIDAD = GUSTAVO A. MADERO
 LONGITUD = -99.14815445
 MANZANA = 006
 MUNICIPIO = GUSTAVO A. MADERO
 NOM_ESTAB = ACCO MEXICANA SA DE CV
 NOM_PROPIE = ACCO MEXICANA SA DE CV
 NUMERO_EXT = 43
 PERS_OCUP = 101 A 250 PERSONAS
 TELEFONO = 5515005700
 TIPO_CALLE = CALLE
 WWW = WWW.ACCOMEX.COM.MX

name:

Visualización de resultados.

Visualización a detalle de objetos geográficos recuperados.

Figura 4.15 Interfaz de la implementación de la recuperación.

El proceso de ranking se ejecuta al mismo tiempo que la consulta, ya que esta devuelve un conjunto de objetos con base en la ubicación del usuario, los cuales ya pasaron por el proceso de similitud, es decir, que ya solo se encuentran objetos geográficos (conceptos) relevantes para el usuario, la implementación no muestra una lista de los objetos sino que siempre está retornando los objetos relevantes para el usuario en un radio cercano (1km. en la implementación).

4.5 Aplicación

Para la implantación de la metodología se desarrolló una aplicación para la Web, con una arquitectura Web service con el framework para Java ZK (Fig. 4.16), en la cual se integran los procesos que involucran la implementación de cada etapa de la metodología (Ver Anexo B).

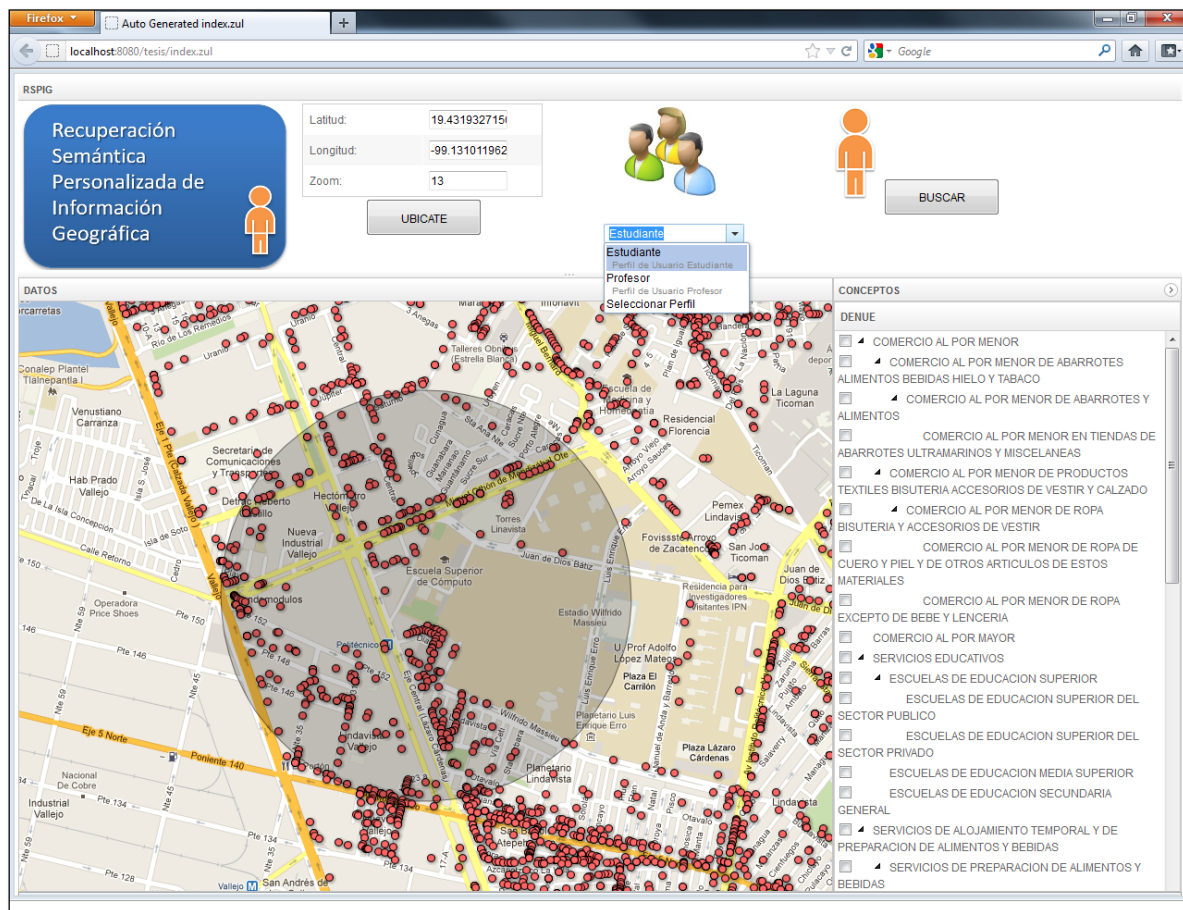


Figura 4.16 Implementación de la Metodología.

4.5.1 Interfaz Gráfica de Usuario

La interfaz gráfica de usuario está diseñada con base en las etapas de la metodología (Fig. 4.17). En la sección *a* de la figura 4.17, se desarrolló una interfaz que representa los conceptos dentro de la ontología, por medio de un componente de estilo arborescente con la característica de selección por concepto, que mostrará al usuario los conceptos relevantes seleccionados para la recuperación. En la sección *b* de la misma, se está implementando la etapa de caracterización por medio de una interfaz que permite seleccionar un perfil de usuario, mismo que está relacionado con los conceptos de la ontología. Por último, en la sección *c* se muestra la interfaz para el ingreso de la ubicación del usuario de manera que puede ingresar la latitud y la longitud si es que se conocieran, o se puede ubicar de manera interactiva por medio de un mapa, para posteriormente ejecutar la etapa de recuperación por medio de un botón (**BUSCAR**) y mostrar los resultados en un mapa para una visualización práctica y sencilla de manipular.

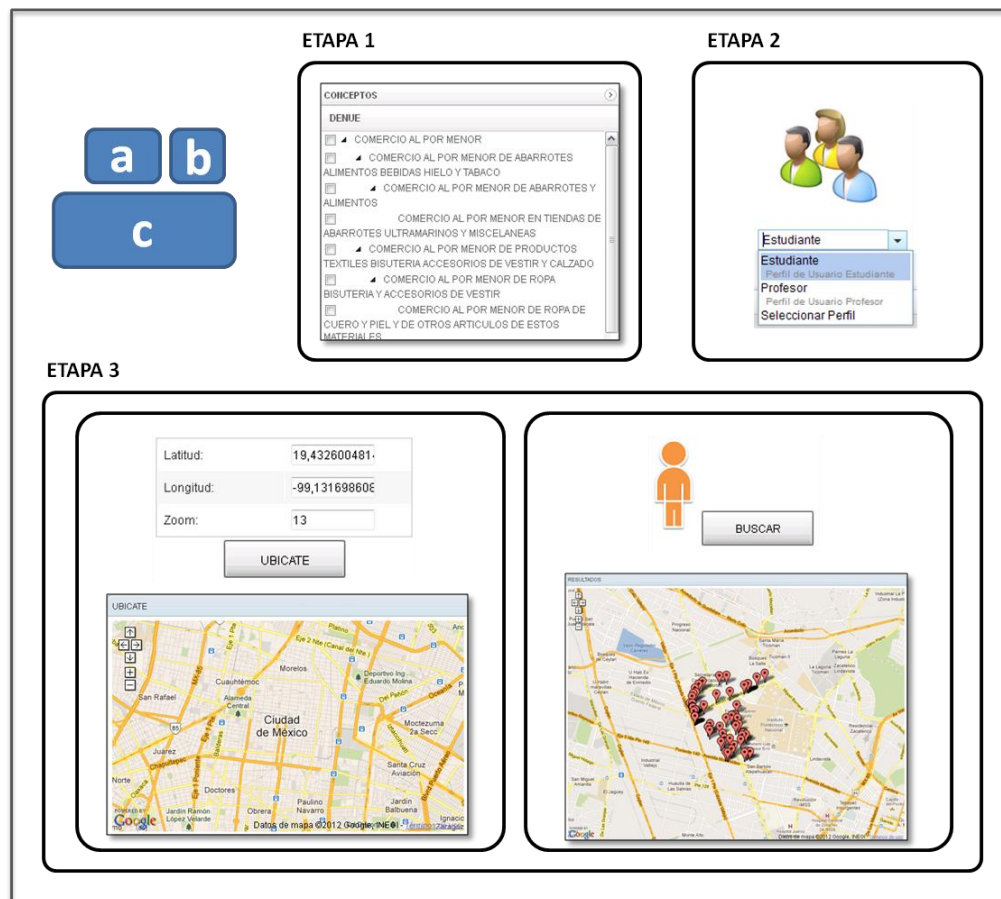


Figura 4.17 Etapas de la Metodología en la Implementación.

4.5.2 Caso de Uso

Para ejemplificar el funcionamiento de la implementación, desarrollaremos un caso de uso, donde como datos de entrada tenemos:

Perfiles de Usuario: Estudiante, Profesor.

Ubicación: CIC- IPN, Latitud-Longitud: (19.5030, -99.1475)

En primer lugar ingresamos la ubicación del usuario, y la tomaremos como referencia para la ejecución de la metodología con los perfiles de entrada como se muestra en la siguiente figura:

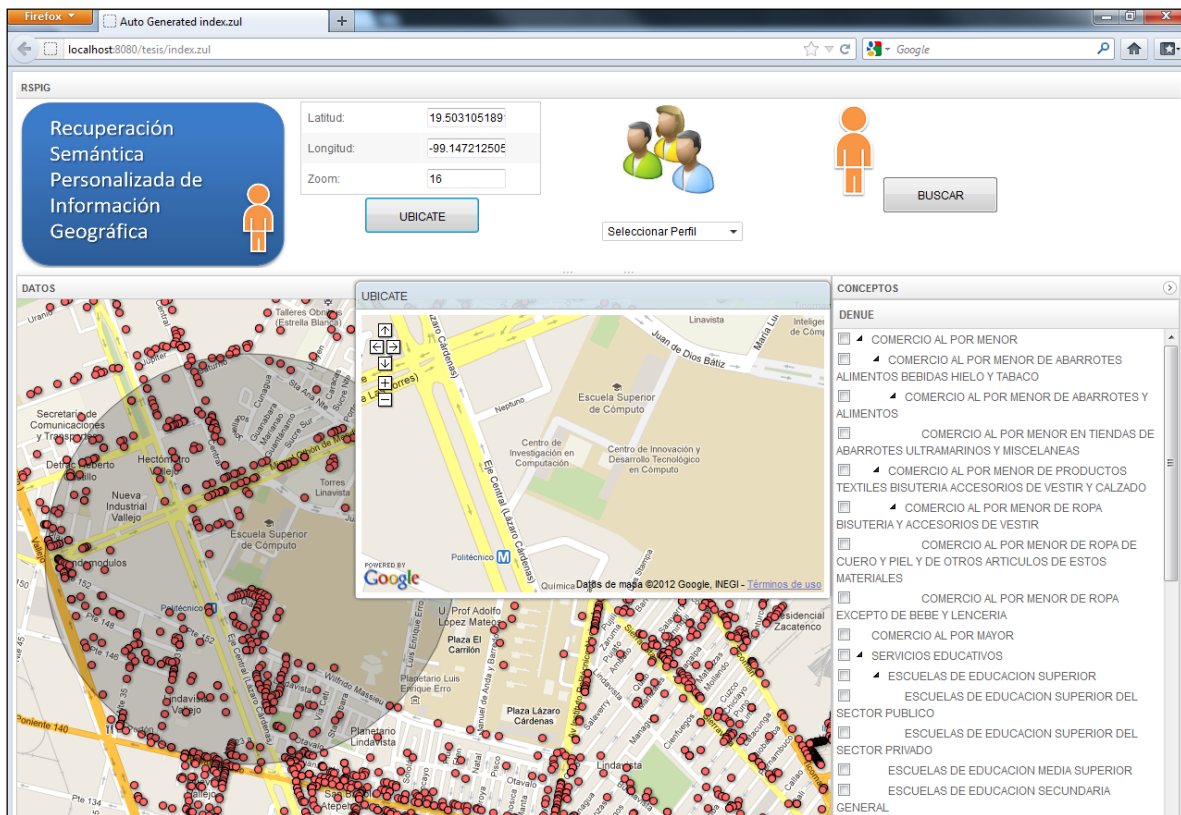


Figura 4.18 Selección de Ubicación en la Implementación.

En seguida, seleccionamos el perfil, en este caso seleccionamos primero el de Estudiante, como se muestra a continuación:

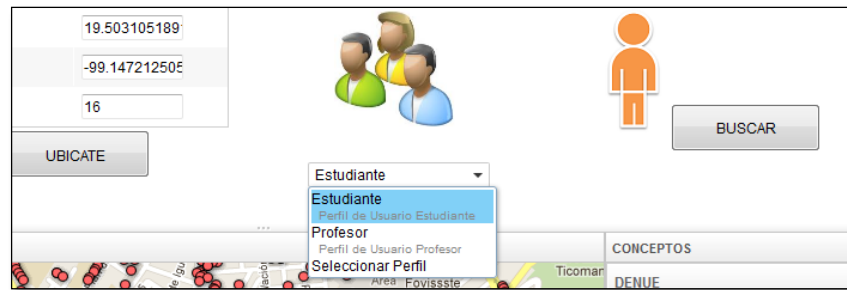


Figura 4.19 Selección de Perfil de Usuario Estudiante.

Debido a que ya se cuenta con los datos de entrada para poder ejecutar la recuperación, damos clic en el botón de *Buscar* en la implementación, y observamos los resultados en la siguiente figura:

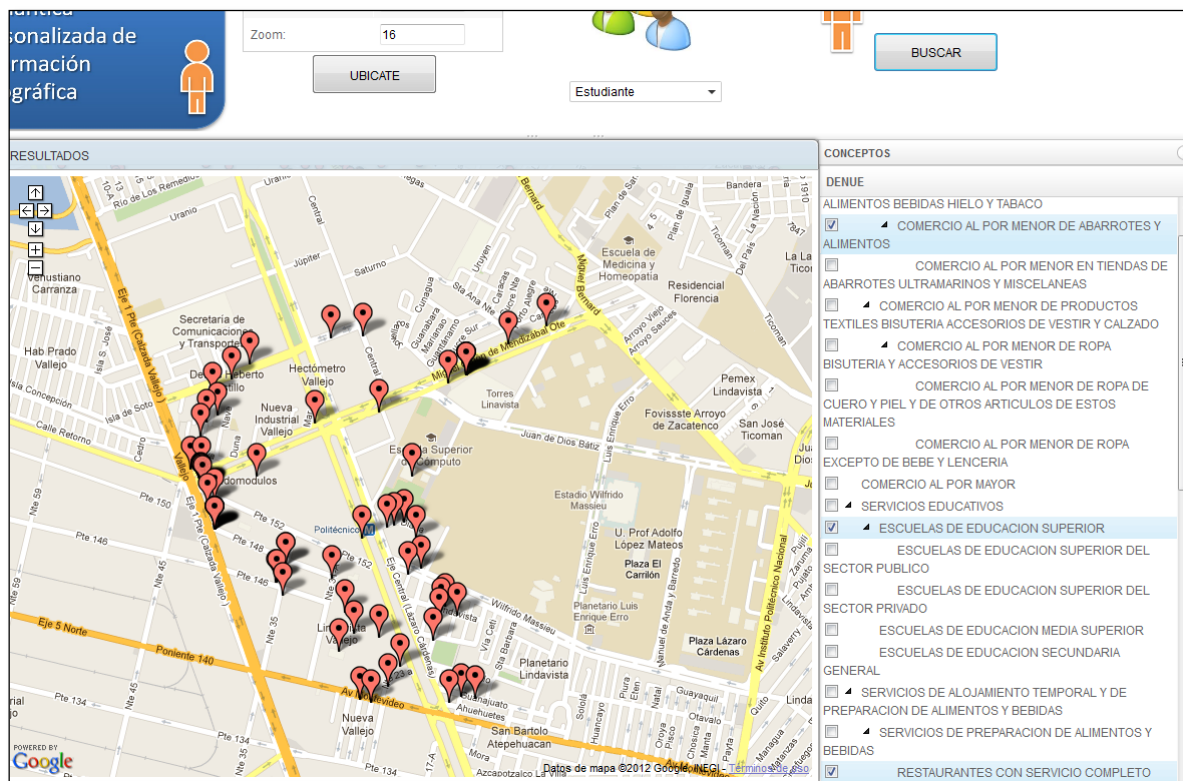


Figura 4.20 Resultados de la implementación de la Recuperación.

En la Figura 4.18, se pudo observar que en la sección de conceptos, no se encuentra seleccionado ningún concepto, al dar clic en el botón de *BUSCAR* se ejecuta la recuperación, y como se puede ver en la Figura 4.23 se muestran los resultados de la recuperación, en la sección de *CONCEPTOS* se seleccionan solo los conceptos que definen al perfil Estudiante, y por otro lado, en la sección de *RESULTADOS* se puede interactuar con los objetos geográficos recuperados y con el mapa.

Cada punto en rojo de los resultados, representa un objeto geográfico resultado de la implementación de la metodología, que al dar clic sobre él, despliega la información de los atributos del objeto geográfico.

Con respecto al mapa, podemos movernos en cualquier sentido, y hacer zoom, todo esto para poder visualizar mejor los resultados (Fig. 4.21).

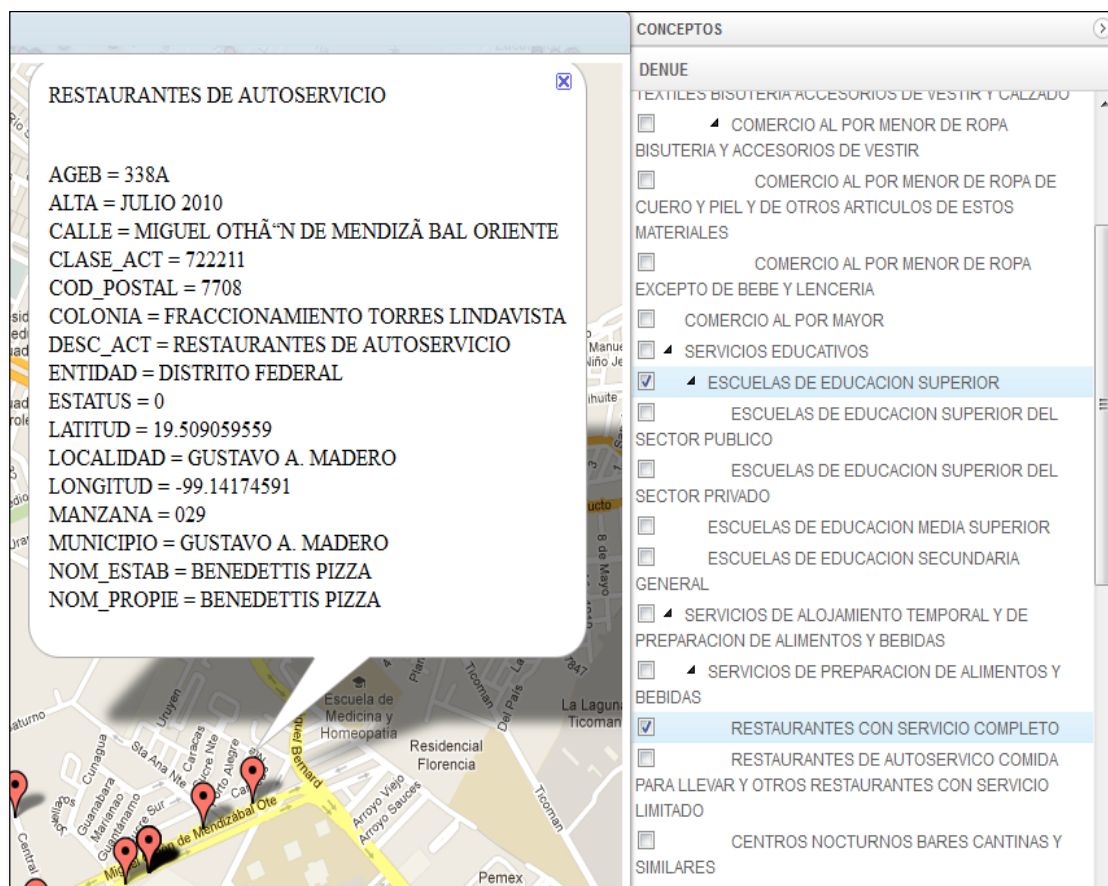


Figura 4.21 Visualización de resultados perfil Estudiante.

Este proceso también es ejecutado para el perfil de Profesor, obteniendo los resultados que se muestran en la Figura 4.22. De esta manera es que la implementación funciona, y retorna los resultados generados por la metodología propuesta en este trabajo de investigación.

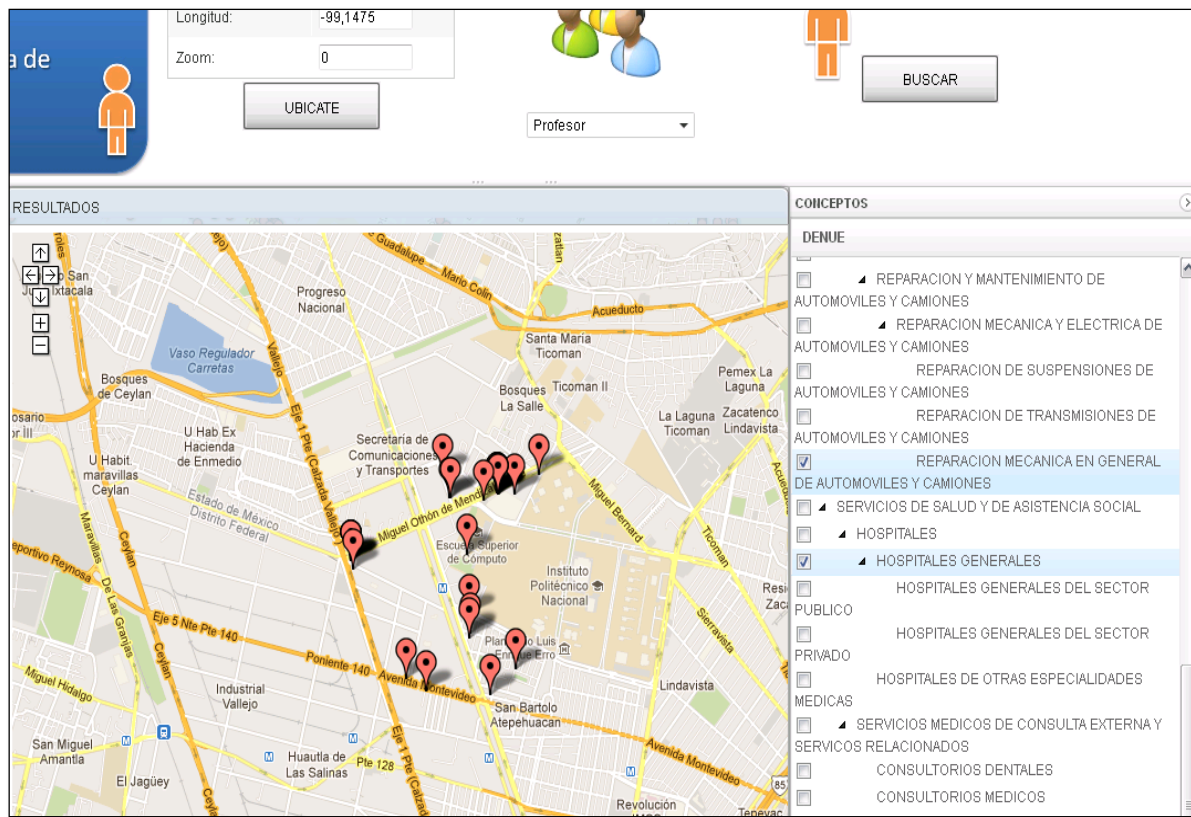


Figura 4.22 Visualización de resultados perfil Profesor.

A continuación pasaremos a analizar los resultados, veremos objetos geográficos generados gracias al uso de la semántica dentro de la metodología y evaluaremos los resultados por medio de los métodos de *Precision* y *Recall*.

4.6 Análisis de Pruebas y Resultados de la Implantación de la Metodología

Para el caso de uso desarrollado en la sección 4.5.2 se realizaron pruebas de la implementación tomando en cuenta dos perfiles y una ubicación en común. Primeramente, nos centraremos en analizar el impacto de la personalización y la semántica en los resultados obtenidos, para posteriormente evaluar el conjunto de resultados.

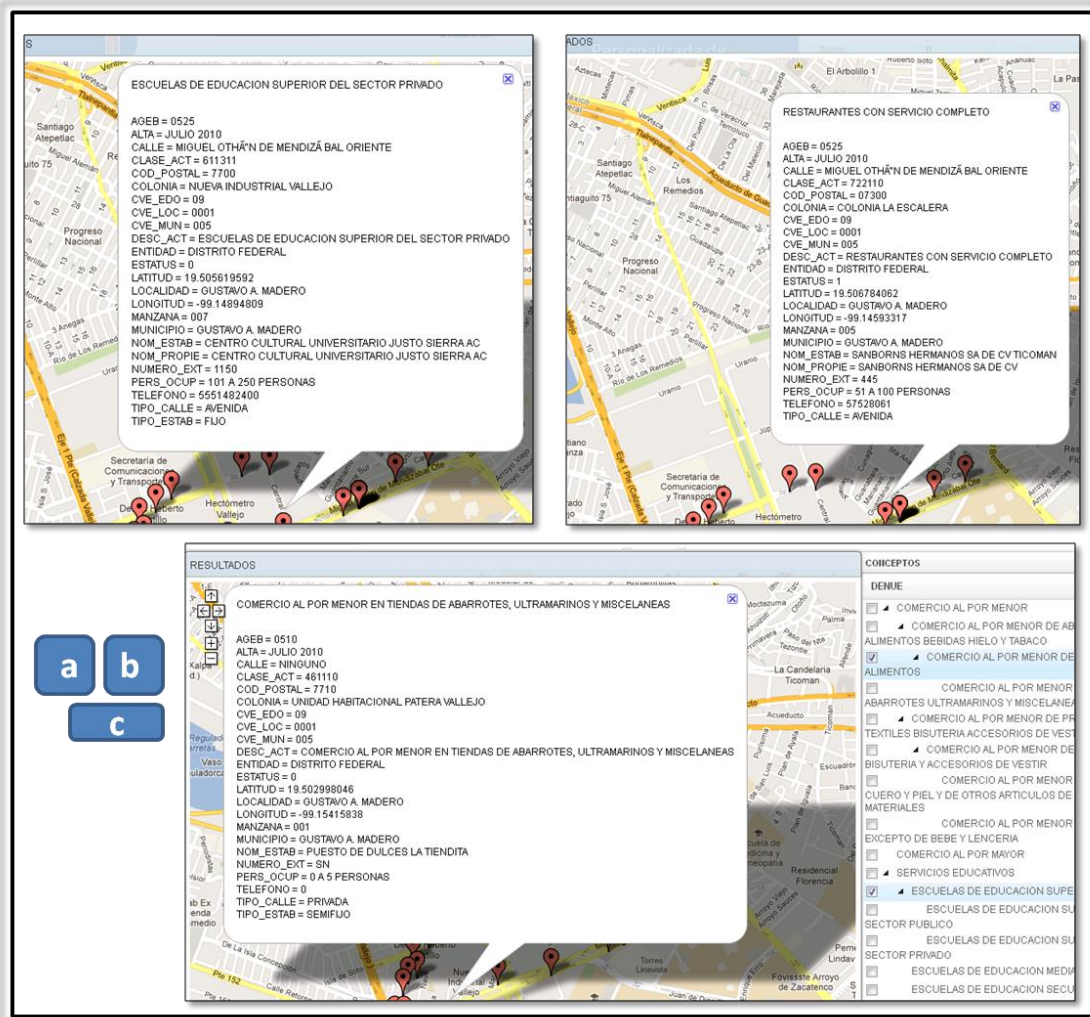


Figura 4.23 Análisis resultados perfil Estudiante.

4.6.1 Personalización y Semántica en las Pruebas y Resultados

En la implementación, por medio del *perfil de usuario* se define un conjunto de conceptos relevantes para efectuar la recuperación, es aquí donde la personalización está influyendo en los resultados, ya que como se muestra en la Figura 4.23 y 4.24, los objetos geográficos resultantes, tienen una relación directa con las preferencias del usuario.

Por otro lado, también podemos observar que algunos de los conceptos recuperados no son idénticos a los conceptos que conforman el perfil de usuario, aquí es donde se refleja el efecto de integrar la similitud semántica en la recuperación (Fig. 4.23 sección a, Fig. 4.24 sección c).

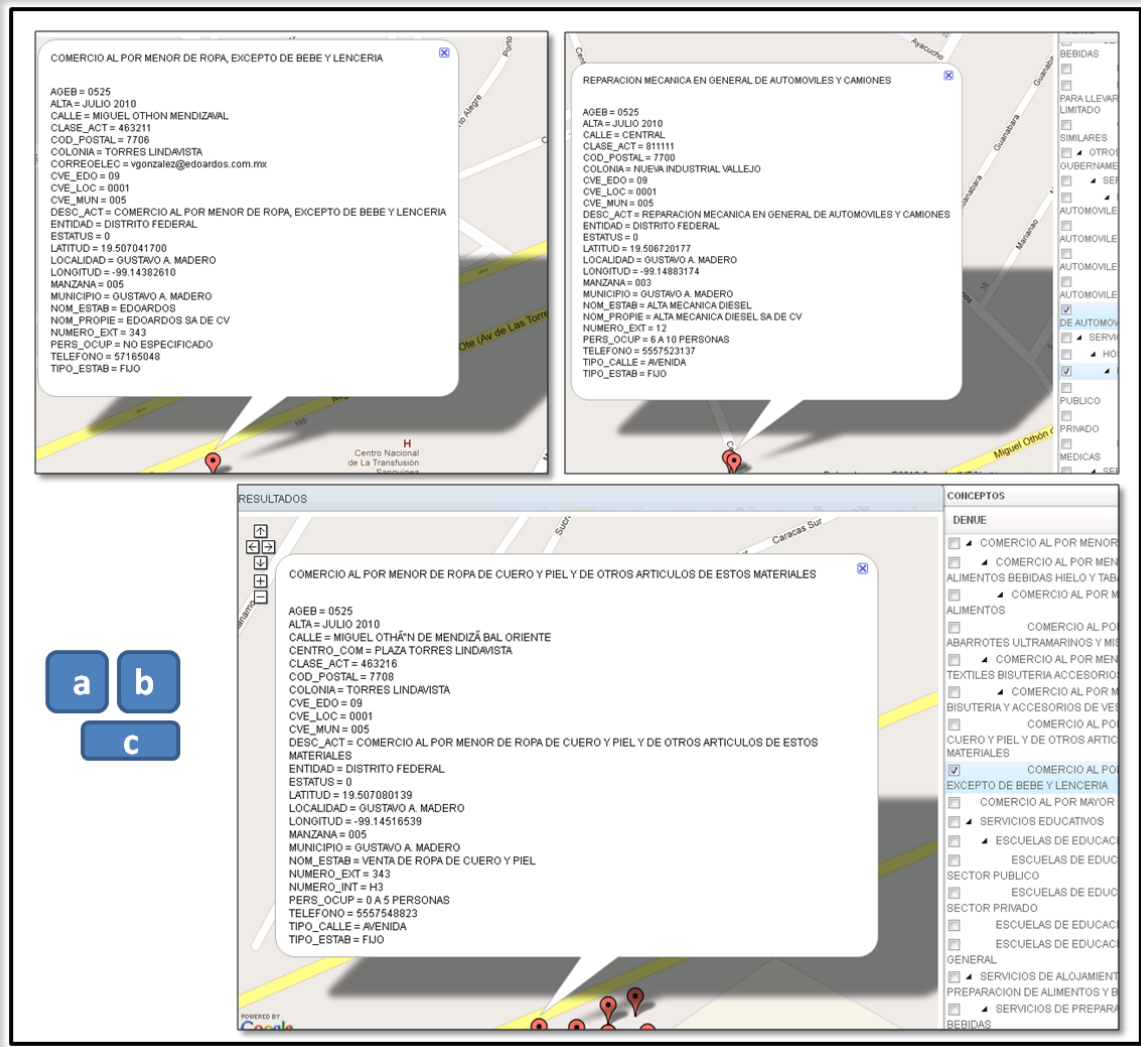


Figura 4.24 Análisis resultados perfil Profesor.

En la etapa de recuperación realizamos un proceso de expansión de la consulta, la cual se lleva a cabo al tomar los conceptos relevantes tomados del perfil de usuario y aplicarles el proceso de similitud semántica que proponemos en la metodología, dicho método devuelve como resultado un conjunto de conceptos similares con base en la conceptualización, enriqueciendo la consulta, para que ahora la consulta ya expandida, sea la entrada para ejecutar la recuperación y mejorar los resultados. La siguiente tabla muestra los conceptos producto de la expansión de la consulta por medio de la similitud semántica desarrollada en la metodología.

Tabla 4.1 Expansión de Conceptos.

PERFIL	CONCEPTOS DEL PERFIL	CONCEPTOS EXPANDIDOS
Estudiante	• ESCUELAS DE EDUCACION SUPERIOR • COMERCIO AL POR MENOR DE ABARROTES Y ALIMENTOS • RESTAURANTES CON SERVICIO COMPLETO	• RESTAURANTES DE AUTOSERVICIO • RESTAURANTES DE COMIDA PARA LLEVAR • ESCUELAS DE EDUCACION SUPERIOR DEL SECTOR PRIVADO • COMERCIO AL POR MENOR EN TIENDAS DE ABARROTES ULTRAMARINOS Y MISCELANEAS • OTROS RESTAURANTES CON SERVICIO LIMITADO
Profesor	• HOSPITALES GENERALES • REPARACION MECANICA EN GENERAL DE AUTOMOVILES Y CAMIONES • COMERCIO AL POR MENOR DE ROPA EXCEPTO DE BEBE Y LENCERIA	• COMERCIO AL POR MENOR DE ROPA DE CUERO Y PIEL • COMERCIO AL POR MENOR DE ROPA DE BEBE

Al analizar la Tabla 4.1, podemos observar que los conceptos expandidos tienen similitud con los conceptos del perfil, para que al ejecutar la recuperación, en los resultados no se tengan objetos geográficos sin relación alguna para el usuario.

Por otro lado también podemos ver que para el caso del perfil de Estudiante los conceptos expandidos representan poco más del doble de los conceptos del perfil (62.5%) y para el caso del perfil Profesor son poco menos de la mitad (40%), estos resultados influyen directamente en la recuperación (Fig. 4.25).

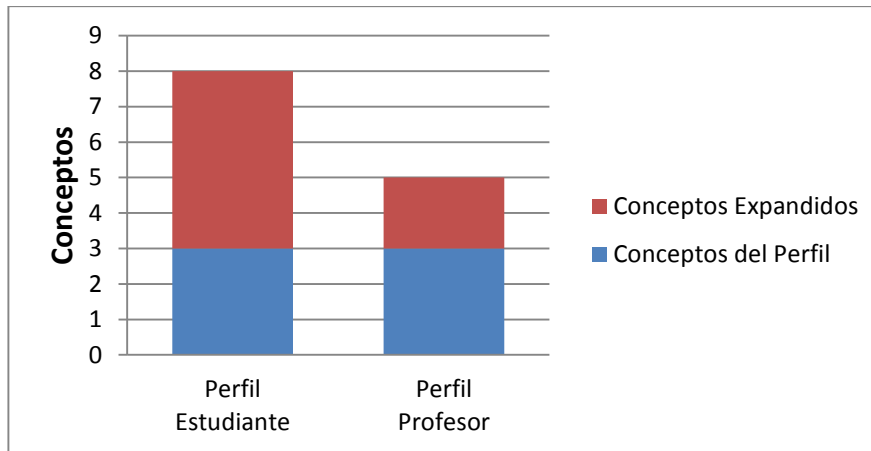


Figura 4.25 Resultados de la expansión de conceptos.

Debido a que en el proceso de recuperación se toma al conjunto de conceptos ya expandido, el conjunto de objetos geográficos resultado en el peor de los casos se queda con los mismos objetos sin aplicar la expansión. En la mayoría de los casos recupera más, ampliando los resultados de la recuperación y gracias a la similitud semántica, dichos resultados tienen una alta probabilidad de ser relevantes para el usuario.

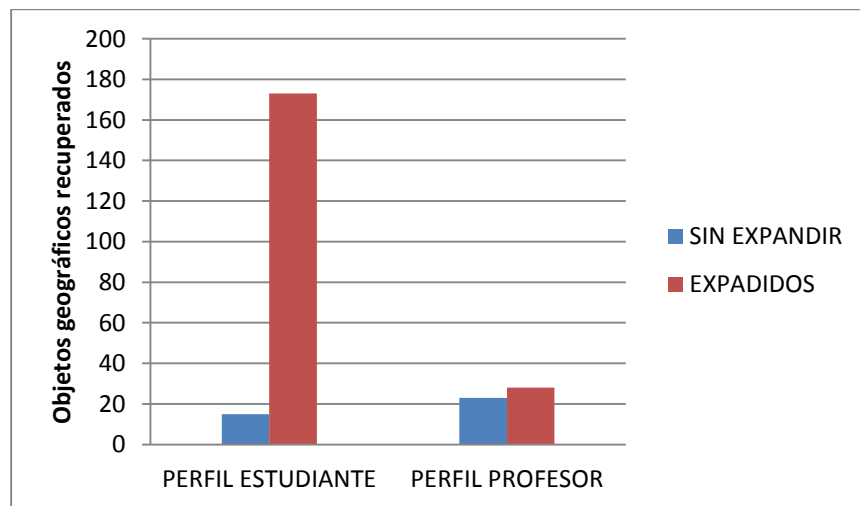


Figura 4.26 Objetos geográficos recuperados con base en la expansión de conceptos.

En la Figura 4.26 se muestra gráficamente el número de objetos geográficos recuperados aplicando la expansión de la consulta y sin la expansión, se puede ver que para el caso del perfil de Estudiante los objetos recuperados son notablemente más con el proceso de expansión que sin él, cuantitativamente son el 91.32% del total de objetos recuperados

ejecutando la expansión, y para el caso del perfil de profesor representan el 17.85% del total, cantidad menor pero aún así está contribuyendo a los resultados de la recuperación, y como ya se mencionó en el peor de los casos podría ser un 0%, dato que posteriormente vamos a ver es poco probable, pero que si se diera el caso aun así se recuperarían los objetos con base en las preferencias del perfil de usuario.

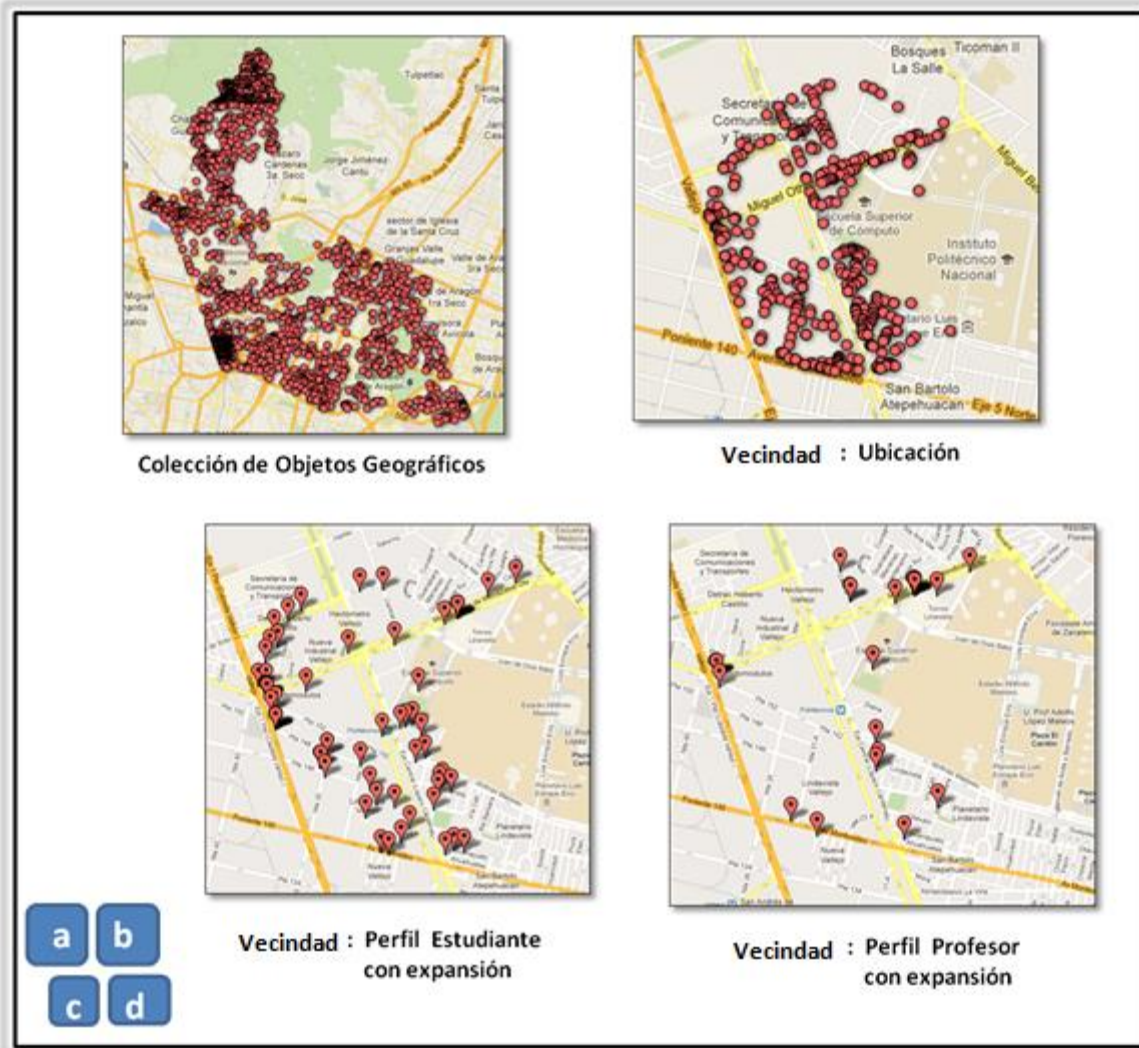


Figura 4.27 Vecindad en la recuperación.

En la metodología se delimita la recuperación con base en la vecindad, por la Ubicación y por el Perfil de usuario (Fig. 4.27). Por medio de la ubicación se recuperan los objetos geográficos a un kilómetro de distancia del usuario (Fig. 4.27sección a), para posteriormente aplicar la recuperación con base en el Perfil (Fig. 4.27 sección c y d), donde ya solo se

recuperan objetos geográficos a una distancia de un kilómetro, de los cuales solo son recuperados los objetos geográficos relevantes con base en el perfil de usuario y aplicando la similitud semántica, que para el ejemplo que se ha estado desarrollando y para más casos, gracias a la similitud semántica no se tiene un resultado nulo o vacío (en la mayoría de los casos). En la Tabla 4.2 se muestra el número de objetos geográficos recuperados por concepto, donde dos de los conceptos sin expandir (conceptos del perfil) tienen cero objetos recuperados lo cual indica un resultado nulo o vacío para el usuario, pero gracias al proceso de similitud podemos ver en la tabla que se recuperan objetos geográficos de los conceptos expandidos, es aquí también donde se encuentra un aporte más a la recuperación con base en la metodología.

Tabla 4.2 Objetos geográficos recuperados por concepto del perfil Estudiante.

SIN EXPANDIR	Objetos Recuperados	EXPANDIDOS	Objetos Recuperados
• ESCUELAS DE EDUCACION SUPERIOR	0	• RESTAURANTES DE AUTOSERVICIO	4
• COMERCIO AL POR MENOR DE ABARROTES Y ALIMENTOS	0	• RESTAURANTES DE COMIDA PARA LLEVAR	58
• RESTAURANTES CON SERVICIO COMPLETO	15	• ESCUELAS DE EDUCACION SUPERIOR DEL SECTOR PRIVADO	1
		• COMERCIO AL POR MENOR EN TIENDAS DE ABARROTES ULTRAMARINOS Y MISCELANEAS	63
		• OTROS RESTAURANTES CON SERVICIO LIMITADO	47

4.6.2 Precision y Recall

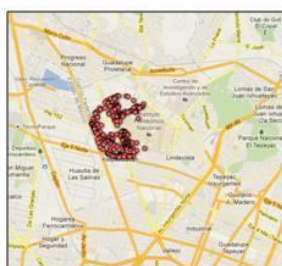
Para poder evaluar la recuperación con base en la metodología desarrollada, ocuparemos las métricas de evaluación para la recuperación de *Precision* y *Recall* (Ver Anexo A). En la Figura 4.28 se muestran los datos obtenidos para proceder a la evaluación.

Colección Objetos Geográficos = COG

Objetos RElevantes = OREL

Objetos Relevantes Recuperados = ORR

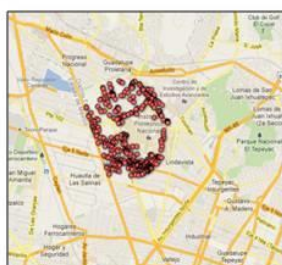
Objetos Recuperados = OR



Datos	Estudiante	Profesor
COG	749	
ORR	169	28
OREL	188	32
OR	208	37
Vecindad = 1 Km.		

Precision
P = 0.812 Estudiante
P = 0.756 Profesor

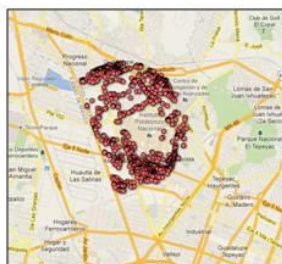
Recall
R = 0.898 Estudiante
R = 0.875 Profesor



Datos	Estudiante	Profesor
COG	1633	
ORR	408	107
OREL	495	126
OR	547	146
Vecindad = 1.5 Km.		

Precision
P = 0.745 Estudiante
P = 0.732 Profesor

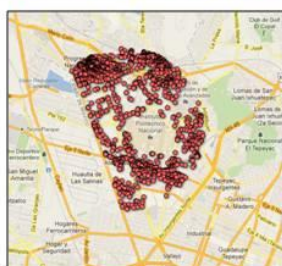
Recall
R = 0.824 Estudiante
R = 0.849 Profesor



Datos	Estudiante	Profesor
COG	3390	
ORR	481	134
OREL	586	158
OR	633	210
Vecindad = 2 Km.		

Precision
P = 0.759 Estudiante
P = 0.638 Profesor

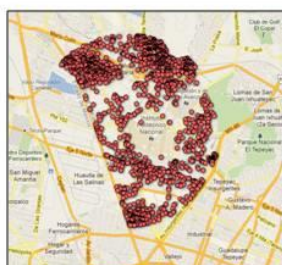
Recall
R = 0.820 Estudiante
R = 0.848 Profesor



Datos	Estudiante	Profesor
COG	6274	
ORR	561	298
OREL	691	358
OR	1150	634
Vecindad = 2.5 Km.		

Precision
P = 0.487 Estudiante
P = 0.470 Profesor

Recall
R = 0.811 Estudiante
R = 0.832 Profesor



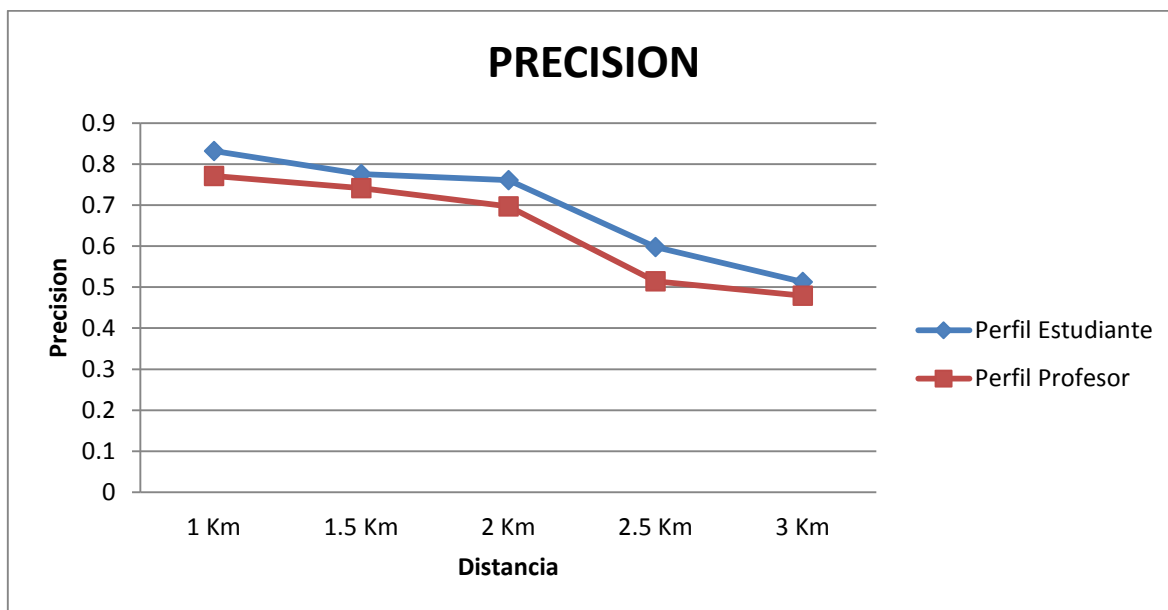
Datos	Estudiante	Profesor
COG	9764	
ORR	825	374
OREL	964	423
OR	1933	805
Vecindad = 3 Km.		

Precision
P = 0.426 Estudiante
P = 0.464 Profesor

Recall
R = 0.855 Estudiante
R = 0.884 Profesor

Figura 4.28 Datos para la evaluación de la recuperación.

Para las métricas de *Precision* y *Recall* tomamos como el número de documentos recuperados (en nuestro caso en lugar de documentos son objetos geográficos) a la distancia ya que dicho número se encuentra en función de esta. Para determinar la relevancia de los objetos geográficos, se utilizó la medida de similitud semántica propuesta en [Resnik, 1999], basada en el contenido de la información usando una jerarquía y la teoría de la información para determinar la similitud semántica, la idea principal es que entre más información compartan dos conceptos, mayor similitud poseen (Ver Anexo A). De esta manera la relevancia queda en función de la medida de similitud semántica de Resnik, y no en un criterio subjetivo, la cual fue seleccionada para seleccionar objetivamente cuales son los conceptos relevantes, la relevancia está en función de la similitud semántica de los conceptos que son parte de la recuperación, el umbral para la similitud semántica que se tomó con base en la medida propuesta por Resnik, fue de 0.7 a 1. Los resultados para la métrica de *Precision* se muestran en la Figura 4.29.

**Figura 4.29** Gráfica de evaluación de *Precision*.

La gráfica de la métrica de *Precision*, muestra un patrón lógico en la recuperación de información [Baeza-Yates, 2011] ya que entre más objetos geográficos se encuentren en la colección, la precisión disminuye.

Para el caso de la gráfica de la métrica de *Recall* (Fig. 4.30), se obtienen resultados que permanecen en un rango del 82% al 94% de objetos recuperados relevantes para el usuario, sin importar el tamaño de la colección de objetos geográficos. Lo cual indica que los resultados de la recuperación muestran que se están recuperando un alto porcentaje de los objetos geográficos relevantes totales en la colección. Este es uno de los objetivos de la metodología y se está cumpliendo como indican los resultados.

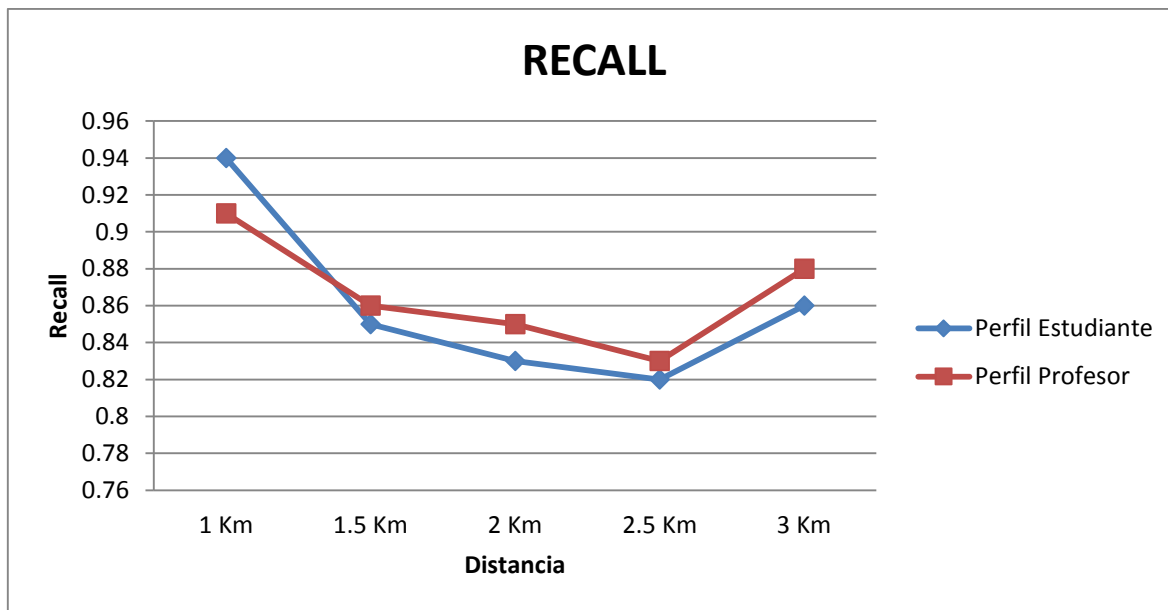


Figura 4.30 Gráfica de evaluación de *Recall*.

4.6.3 Experimentos Generales

Para seguir evaluando la metodología, se llevaron a cabo una serie de experimentos modificando la entrada de datos a la implementación. En la Figura 4.31 se muestran algunas zonas de prueba, los datos que se estuvieron modificando fueron: la ubicación y los conceptos del perfil de manera aleatoria para llevar a cabo cincuenta consultas más con la estructura de ejecución como se desarrolló en el caso de uso descrito anteriormente.



Figura 4.31 Experimentos generales.

De los experimentos se promediaron las métricas de *Precision* y de *Recall* y se obtuvieron los siguientes resultados. De la misma forma que en ejemplo trabajado, la relevancia se determinó por medio de la similitud semántica de Resnik.

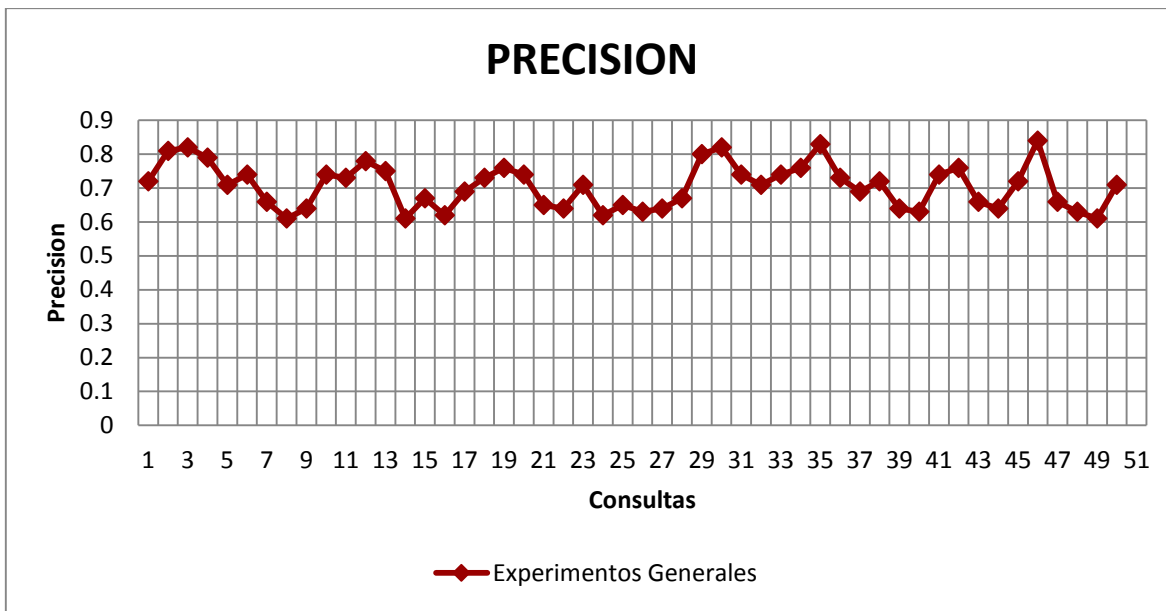


Figura 4.32 Gráfica de evaluación de *Precision* para experimentos generales.

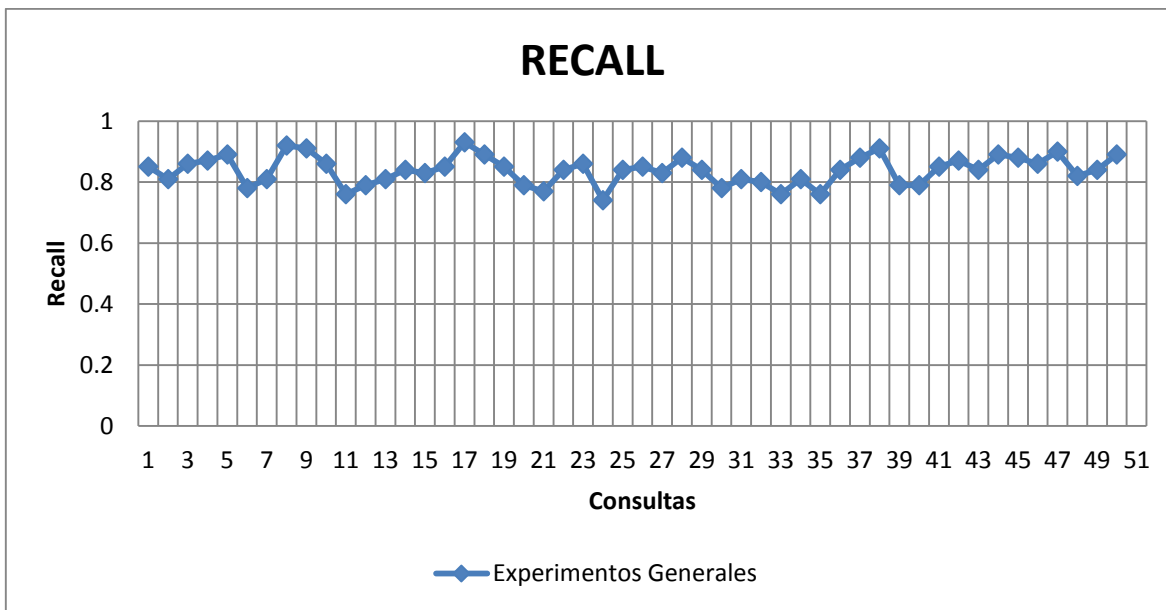


Figura 4.33 Gráfica de evaluación de *Recall* para experimentos generales.

Por último, cabe resaltar que se podrían hacer más experimentos, pero estos se consideraron suficientes debido a los mismos resultados que mostraron un comportamiento parecido, para poder sacar una muestra y poder realizar conclusiones de este trabajo de investigación, las cuales se describen en el siguiente capítulo.

CAPÍTULO 5

CONCLUSIONES Y TRABAJO

FUTURO

Esta tesis propone una metodología para personalizar la Recuperación de Información Geográfica con un enfoque semántico, con base en la conceptualización de un dominio de trabajo, tomando las preferencias del usuario y su ubicación para ofrecer resultados relevantes y coherentes. Adicionalmente, la metodología se puede aplicar en diferentes casos de uso. Cabe mencionar que en un principio se empezó a trabajar en un caso para la mitigación de los desastres naturales, que debido a los datos que se tenían del DENUe y a su estructura, se optó por cambiar el caso por el que se presenta en este trabajo.

5.1 Conclusiones

Con base en los resultados obtenidos con la implantación de la metodología, se concluye de maneja general, que la personalización y el enfoque semántico que se dio en este trabajo, están reflejados en los resultados de la recuperación en las consultas que se ejecutaron. El primer impacto se da en el número de objetos geográficos recuperados, debido al proceso de expansión de la consulta, la cual implica ejecutar el método de similitud semántica propuesto. Gracias a este proceso en la mayoría de las consultas, el número de objetos recuperados con base en los conceptos expandidos representaban en promedio un 35%. Dato que es notablemente mayor en ciertas consultas donde incluso representaba el 100% de los objetos recuperados. En este caso en particular, podemos concluir que la metodología da soporte para casos donde el resultado fuera nulo o vacío, si solo se basara en una recuperación sin el

proceso de expansión. Recordando que en el peor de los casos este pudiera ser el resultado, pero que como se vio en las pruebas y resultados, es un caso no muy cotidiano.

Por otro lado, el conjunto de objetos geográficos recuperado, tiene una característica importante y representativa en el desarrollo de esta investigación, estamos hablando de la relación entre objetos geográficos recuperados.

Debido a la descripción de un dominio de trabajo, se acota la recuperación de información geográfica, que en conjunto de las preferencias del usuario, originan resultados coherentes para el usuario siendo la mayoría de estos relevantes. La relevancia de estos resultados la pudimos evaluar por medio de las métricas de *Precision* y de *Recall*.

En la evaluación de los resultados, nos encontramos con un comportamiento lógico con la métrica de Precision, donde entre más grande es la colección de documentos, la precisión iba disminuyendo. Pero en el caso del Recall donde se involucra a los objetos geográficos relevantes, se obtienen resultados que permanecen en un rango del 80% al 90% de objetos recuperados relevantes para el usuario, sin importar el tamaño de la colección de objetos geográficos. Es en este punto donde, contribuyen las etapas de conceptualización y caracterización, ya que por medio de esta última, estamos personalizando la recuperación, y debido a la primer etapa, trabajamos sobre un mismo contexto, como resultado se tienen objetos geográficos relevantes para el usuario, atacando problemas como los resultados sin importancia para el usuario y de búsquedas sin sentido. Por otro lado cabe resaltar que en la literatura no se utiliza una medida de similitud semántica para calcular la relevancia de los objetos geográficos en la colección así como también en los resultados, propuesta que se incluye en este trabajo. Definitivamente el umbral que se tome para determinar si contaba un concepto como relevante o no con base a la similitud semántica propuesta por Resnik, esta directamente ligado a los resultados, e incluso se podría graduar en posibles trabajos futuros.

De manera general, con las pruebas y resultados que se llevaron a cabo, podemos establecer que existe una mejora en la recuperación de información geográfica, basándonos en las

métricas de *Precision* y de *Recall*. Y por último, con este trabajo se muestra una perspectiva más para efectuar recuperación de información geográfica.

5.2 Contribuciones

Las contribuciones de la presente tesis son:

- Una metodología para realizar recuperación de información geográfica, de manera personalizada con un enfoque semántico.
- Se definió un método para aplicar la similitud semántica.
- Se implementó un caso de estudio con datos reales del DENE.
- Se diseñó y desarrolló un sistema para implantar la metodología y realizar pruebas.
- Se utilizó un método de similitud semántica para calcular la relevancia de los objetos geográficos para efectuar las métricas de evaluación de *Precision* y *Recall*.

Se ha publicado el siguiente trabajo derivado de la presente investigación:

- XIII Conferencia Iberoamericana en Sistemas de Información Geográfica, “Recuperación Semántica de Información Basada en Perfiles de Usuario aplicado a Servicios Basados en Localización”, 2011. Mismo que fue seleccionado como capítulo del libro: “La innovación geotecnológica como soporte para la toma de decisiones en el desarrollo territorial”, pág. 85 -98, ISBN: 978-607-00-4404-5, 2011.

5.3 Trabajos futuros

Los trabajos que vemos con factibilidad de realizar a mediano plazo son los siguientes:

- Uno de los principales obstáculos que presenta el desarrollar ontologías es que no existen conceptualizaciones muy detalladas sobre datos geográficos, lo que dificulta su desarrollo. Se tiene que trabajar en la infraestructura.
- Uno de los procesos más pesados es llegar a la conceptualización. En este trabajo se realizó manualmente y con datos previamente estructurados. Un trabajo a futuro, consistiría en automatizar este proceso.
- En cuanto a la construcción del perfil de usuario, se pueden abordar problemas como la generación automática del perfil, o un perfil dinámico que esté en constante cambio con base en consultas previas, ubicación del usuario, variables espacio temporales, etc.
- Diseño y desarrollo de métricas para establecer la Similitud Semántica, este punto definitivamente es crucial en la recuperación.
- Diseño y desarrollo de metodologías para graduar la Similitud Semántica de manera dinámica.
- Diseño y desarrollo de metodologías para hacer dinámica la vecindad para la recuperación.

REFERENCIAS

[Arpinar, 2004] Geospatial Ontology Development and Semantic Analytics , I. Budak Arpinar, Amit Sheth & Cartic Ramakrishnan, Handbook of Geographic Information Science, Eds: J. P. Wilson and A. S. Fotheringham, Blackwell Publishing , 2004.

[Baeza-Yates, 2011] Modern Information Retrieval, the concepts and technology behind search, second edition, Ricardo Baeza-Yates, Berhier Ribeiro-Neto, 2011.

[Bulskov, 2006] Ontology-based Information Retrieval, Henrik Bulskov Styltsvig, Computer Science Section Roskilde University, Denmark, 2006.

[Challam, 2007]Challam, V., S. Gauch, and A. Chandramouli (2007). Contextual information retrieval using ontology based user profiles. In RIAO 2007.

[Dean y Schreiber, 2003] Dean M, Schreiber G (2003) OWL Web Ontology Language Reference. W3C Working Draft.<http://www.w3.org/TR/owl-ref/>.

[DENUE, 2011] El Directorio Estadístico Nacional de Unidades Económicas (DENUE, 03/2011), <http://www.inegi.org.mx/sistemas/denue/Default.aspx>, INEGI, 2011.

[Ferrés, 2005] Daniel Ferrés, Alicia Ageno, and Horacio Rodríguez. The geotalp-ir system at geoclef 2005: Experiments using a qa-based ir system, linguistic analysis, and a geographical thesaurus. In Accessing Multilingual Information Repositories: 6th Workshop of the Cross-Language Evaluation Forum, CLEF 2005, volume 4022 of Lecture Notes in Computer Science, pages 947–955. Springer, 2006.

[Ferrés, 2007] Daniel Ferrés and Horacio Rodríguez. Talp at geoclef 2007: Results of a geographical knowledge filtering approach with terrier. In *Advances in Multilingual and Multimodal Information Retrieval: 8th Workshop of the Cross-Language Evaluation Forum, CLEF 2007*, volume 5152 of *Lecture Notes in Computer Science*, pages 830–833. Springer, 2008.

[Genesereth, 1991] M. R. Genesereth. Knowledge interchange format. In J. Allen, R. Fikes, and E. Sandewall, editors, *Principles of Knowledge Representation and Reasoning: Proc. of the Second International Conference (KR'91)*, pages 599-600. Kaufmann, San Mateo, CA, 1991.

[GFT, 2009] Google Fusion Tables, www.google.com/fusiontables.

[Golemati, 2007] Creating an Ontology for the User Profile: Method and Applications Maria Golemati, Akrivi Katifori, Costas Vassilakis, George Lepouras, Constantin Halatsis.

[Golemati, 2007] M. Golemati, A. Katifori, C. Vassilakis, G. Lepouras, C. Halatsis, User Profile Ontology version 1, <http://oceanis.mm.di.uoa.gr/pened/?category=publications>.

[Goñi, et. al., 1998] Goñi, A., Mena, E. and Illarramendi, A. (1998). Querying Heterogeneous and Distributed Data Repositories Using Ontologies. In Charrel, P.-J. and Jaakkola, H. (eds.), *Information Modelling and Knowledge Base IX*, IOS Press: 19-34.

[Gracia-Cumbreras, 2009] Miguel Á. Gracia-Cumbreras, José M. Perea-Ortega, Manuel García-Vega, and L. Alfonso Ureña-López. Information retrieval with geographical references. Relevant documents filtering vs. query expansion. *Information Processing and Management*, 45(5):605–614, 2009.

[Guarino, 1998] Formal Ontology and Information Systems, Nicola Guarino, *Proceedings of FOIS'98*, Trento, Italy, 6-8 June 1998. Amsterdam, IOS Press, pp. 3-15.

[Gruber, 1992] Thomas Robert Gruber. *Ontolingua: A mechanism to support portable ontologies*, 1992.

[Guarino, 1998] Nicola Guarino, *Formal Ontology and Information Systems*, Proceedings of FOIS'98, Trento, Italy, 6-8 June 1998. Amsterdam, IOS Press, pp. 3-15.

[Hauff, 2006] Claudia Hauff, Dolf Trieschnigg, and Henning Rode. University of twente at geoclef 2006: geofiltered document retrieval. In *Evaluation of Multilingual and Multi-modal Information Retrieval: 7th Workshop of the Cross-Language Evaluation Forum, CLEF 2006*, volume 4730 of *Lecture Notes in Computer Science*, pages 958–961. Springer, 2007.

[Hermit, 2012] Hermit OWL Reasoner, <http://www.hermit-reasoner.com/>.

[Hill, 2000] Hill, L., "Core elements of digital gazetteers: Placenames, categories and footprints" Borbinha, J. and Baker, T. (Eds.) *Research and Advanced Technology for Digital Libraries*, proceedings 2000.

[Ibarra, 2009] José Jonathan Ibarra Vargas, *Análisis espacial en espacios conceptuales*, tesis de maestría, 2009.

[Jokela, et. al., 2003] Jokela T, Iivari N, Matero J, and Karukka M 2003, The standard of user-centred design and the standard definition of usability: Analyzing ISO 13407 against ISO 9241–11. In *Proceedings of the Latin American Conference on Human-Computer Interaction*, Rio de Janeiro, Brazil:53–60.

[Janowicz, 2008] Janowicz, K., Wilkes, M., and Lutz, M. Similarity-based information retrieval and its role within spatial data infrastructures. In *Proc. 5th International Conference on Geographic Information Science (GIScience) (2008)*, Springer, pp. 151–167. doi:10.1007/978-3-540-87473-7 10.

[Janowicz, 2009] Janowicz, K., and Wilkes, M. Sim-Dla: A Novel Semantic Similarity Measure for Description Logics Reducing Inter-concept to Inter-instance Similarity. In Proc. 6th Annual European Semantic Web Conference (ESWC) (2009), L. Aroyo, P. Traverso, F. Ciravegna, P. Cimiano, T. Heath, E. Hyvoenen, R. Mizoguchi, E. Oren, M. Sabou, and E. P. B. Simperl, Eds., vol. 5554 of Lecture Notes in Computer Science, Springer, pp. 353–367. doi:10.1007/978-3-642-02121-3 28.

[Janowicz, 2011] The semantics of similarity in geographic information retrieval, Krzysztof Janowicz, Martin Raubal, and Werner Kuhn, JOURNAL OF SPATIAL INFORMATION SCIENCE Number 2 (2011), pp. 29–57.

[Jones,Purves, 2008] Jones, C. B., and Purves, R. S. Geographical information retrieval. International Journal of Geographical Information Science 22, 3 (2008), 219–228. doi:10.1080/13658810701626343.

[Kobsa, 1993] A. Kobsa, User Modelling: Recent work, prospects and hazards, Adaptive User Interfaces: Principles and Practices (Schneider-Hufschmidt, T. Khme, U. Malinowski, eds. 1993).

[Larson, 1995] Larson, Ray R., "Geographic Information Retrieval and Spatial Browsing" Linda C. Smith, Mike Gluck (Eds) Geographic Information Systems Patrons Maps and Spatial Information, 1995 pp 81-123

[Larson, 2004] Ray R. Larson and Patricia Frontiera. Ranking and representation for geographical information retrieval. In Proceedings of Workshop on Geographic Information Retrieval SIGIR 04, Sheffield, UK, 2004. Extended abstract.

[Larson, 2005] Ray R. Larson. Cheshire ii at geoclef 2005: Fusion and query expansion for gir. In Working notes for the CLEF 2005 Workshop, Wien, Austria, September 2005.

[Martins, 2005] Bruno Martins, Mário J. Silva, and Leonardo Andrade. Indexing and ranking in geo-ir systems. In Proceedings of Workshop on Geographic Information Retrieval, GIR'05, pages 31–34, Bremen, Germany, 2005. ACM Press.

[Miller, 1995] Miller, G. (1995). WordNet: A Lexical Database for English. Communications of the ACM, 38(11), 39-41.

[Montello, et al., 2003] Montello, D., Goodchild, M., Gottsegen, J. and Fohl, P., 2003, Where's downtown?: Behavioral methods for determining referents of vague spatial queries. SpatialCognition and Computation, 3, pp. 185–204.

[Moreno-Ibarra, et. al., 2011] Moreno-Ibarra, Marco; Torres, Miguel; Quintero, Rolando; Guzman, Giovanni; Menchaca-Mendez, Rolando: Semantic assessment of similarity between raster elevation datasets, Revista Facultad de Ingeniería Universidad de Antioquia, núm. 59, junio, 2011, pp. 37- 46.

[Noy, 2001] Ontology Development 101: A Guide to Creating Your First Ontology, Natalya F. Noy and Deborah L. McGuinness Stanford University, Stanford, CA, 2001.

[Qiuyan, 2008] Measuring Semantic Similarity in Ontology and Its Application in Information Retrieval, Sheng Qiuyan, Ying Guisheng, Congress on Image and Signal Processing, 2008.

[Renteria, 2009] Walter Renteria Agualimpia, Recuperación controlada de información cualitativa desde repositorios de datos, tesis de maestría, 2009.

[Resnik, 1999] P. Resnik : "Semantic Similarity in a Taxonomy: An Information-Based Measure and its Application to Problems of Ambiguity in Natural Language", (1999), Volume 11, pages 95-130.

[Rich, 2004] E. Rich, Users are individuals: individualizing user models, *International Journal of Man-machine Studies* 18(3), 199—214.

[Rodríguez, Egenhofer, 2004] Rodríguez, A., and Egenhofer, M. Comparing geospatial entity classes: an asymmetric and context-dependent similarity measure. *International Journal of Geographical Information Science* 18, 3 (2004), 229–256. doi:10.1080/13658810310001629592.

[Schwering, 2008] Angela Schwering : Approaches to Semantic Similarity Measurement for Geo-Spatial Data: A Survey, *Transactions in GIS*, 2008, 12(1): 5–29.

[SCIAN, 2007] Sistema de Clasificación Industrial de América del Norte 2007, <http://www.inegi.org.mx/sistemas/scian/default.aspx>, INEGI, 2007.

[Sieg, 2007] Representing Context in Web Search with Ontological User Profiles, Ahu Sieg, Bamshad Mobasher, and Robin Burke, *Modeling and Using Context In Modeling and Using Context - 6th International and Interdisciplinary Conference, CONTEXT'07*, Roskilde, Denmark, Vol. 4635 (2007), pp. 439-452. doi:10.1007/978-3-540-74255-5_33.

[Stanford, 2012] <http://protege.stanford.edu/>.

[Torres, 2008] M. Torres, R. Quintero, S. Levachkine, Giovanni Guzmán & M. Moreno, “Towards a Methodology to Conceptualize the Geographic Domain”, *Advances in Artificial Intelligence in Alexander Gelbukh, Eduardo F. Morales (Eds.), Lecture Notes in Artificial Intelligence*, Vol. 5317, Springer-Verlag, Berlin Heidelberg, 2008, ISSN: 0302-9743, pp. 111-122.

[Torres, 2009] Ontological representation based on semantic descriptors applied to geographic objects. 2009.

[Trajkova, 2006] Improving Ontology-Based User Profiles, Joana Trajkova, Susan Gauch, *Electrical Engineering and Computer Science University of Kansas*.

[Tversky, 1977] Tversky, A. Features of similarity. *Psychological Review* 84, 4 (1977), 327–352.doi:10.1037/0033-295X.84.4.327.

[Wealands, 2007] User Assessment as Input for Useful Geospatial Representations within Mobile Location-Based Services, Karen Wealands, Suzette Miller, *Transactions in GIS* , 2007, 11(2): 283–309.

[Wikipedia, 2012] http://en.wikipedia.org/wiki/Web_Ontology_Language

[Zhan, 2008] Ontology-based semantic description model for discovery and retrieval of geospatial information, Qin Zhan, Xia Zhang, Deren Li, *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*.Vol. XXXVII. Part B4. Beijing 2008.

ANEXO A – MARCO TEÓRICO

A.1 Precision y Recall

La métrica de Precision es la fracción de casos recuperados que son relevantes, mientras que la de Recall es la fracción de casos relevantes que se recuperan. Ambas métricas están diseñadas para medir la relevancia, que en el contexto de la recuperación de información, se definen en términos de un conjunto de documentos recuperados (por ejemplo, la lista de los documentos producidos por un motor de búsqueda en la web para una consulta) y una serie de documentos relevantes (por ejemplo, la lista de todos los documentos en Internet que son relevantes para un tema determinado).

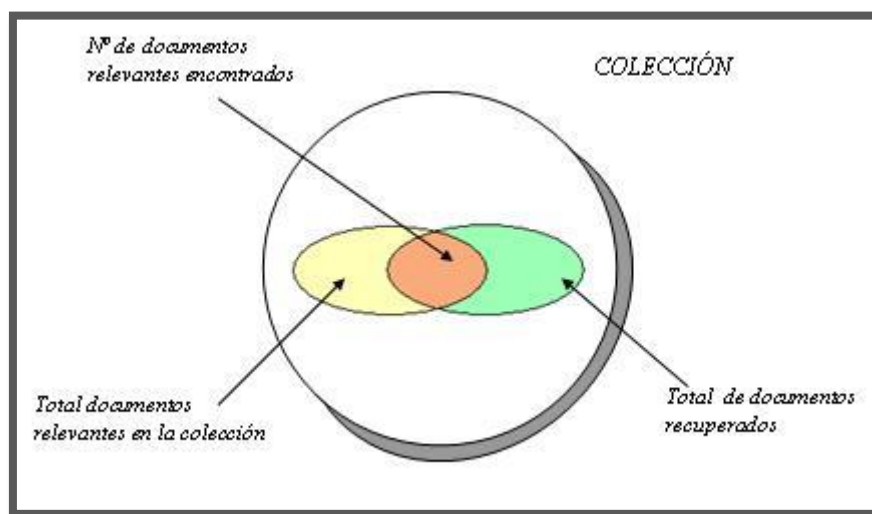


Figura A.1 Conjuntos de documentos para efectuar Precision y Recall.

Con base en la Fig. A.1 se definen las siguientes ecuaciones para el cálculo de estas métricas:

$$Precision = \frac{|{\text{documentos relevantes encontrados}}|}{|{\text{total de documentos recuperados}}|}$$

$$Recall = \frac{|{\text{documentos relevantes encontrados}}|}{|{\text{documentos relevantes en la colección}}|}$$

A.2 Clasificación de las Ontologías

Las ontologías pueden ser clasificadas de acuerdo a su dependencia de una tarea específica o de un punto de vista (Figura A.2), se describen brevemente a continuación [Guarino, 1998]:

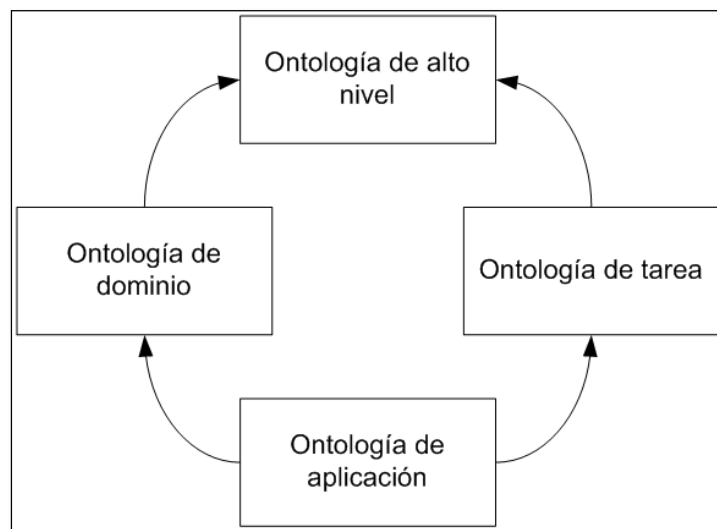


Figura A.2 Tipos de ontologías de acuerdo a su nivel de dependencia para una tarea en particular (Las líneas representan relaciones de especialización). [Guarino, 1998]

- **Ontología de alto nivel (*Top-level Ontology*)**. Describe conceptos muy generales, tales como espacio, tiempo, materia, objeto, eventos, acciones, etc., que son independientes de un dominio o problema en particular; por lo tanto, parece razonable, por lo menos en teoría, tener ontologías a nivel superior unificadas para grandes comunidades de usuarios.
- **Ontología de Dominio (*Domain Ontology*)**. Describe el vocabulario relacionado con un dominio genérico, tales como medicina o automóviles, especializando los términos introducidos en la ontología de nivel superior.
- **Ontología de Tarea (*Task Ontology*)**. Describe el vocabulario relacionado con una tarea en específico, tales como un diagnóstico o ventas, especializando los términos introducidos en la ontología de nivel superior.
- **Ontología de Aplicación (*Application Ontology*)**. Describe conceptos tanto de un dominio como de una tarea en particular, que frecuentemente son especializaciones de ambos. Estos conceptos corresponden a los roles desempeñados por las entidades del dominio mientras realizan cierta actividad, como unidad reemplazable o componente de repuesto.

A.3 OWL

OWL (Web Ontology Language), es junto con RDF las recomendaciones oficiales del W3C para la Web Semántica. Es un conjunto de elementos y atributos XML con significado definido, usados para describir términos y relaciones entre ellos. El lenguaje OWL es una familia de lenguajes de representación del conocimiento para la creación de ontologías. Se caracterizan por el uso de la especificación RDF (Resource Description Framework) con base en XML para la Web Semántica. OWL está avalado por la W3C, [Wikipedia, 2012]. Los archivos OWL son documentos RDF. Por lo tanto el elemento raíz de una ontología OWL es un `rdf:RDF`. Por ejemplo:

```
<rdf:RDF
  xmlns:owl ="http://www.w3.org/2002/07/owl#"
  xmlns:rdf ="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
  xmlns:xsd ="http://www.w3.org/2001/XMLSchema#">
```

Un archivo OWL puede comenzar con una colección de afirmaciones para fines de mantenimiento. Estas afirmaciones se agrupan bajo un elemento `owl:Ontology`, que contiene comentarios, control de versiones y la inclusión de otras ontologías. Por ejemplo:

```
<owl:Ontology rdf:about="">
  <rdfs:comment>An example OWL ontology</rdfs:comment>
  <owl:priorVersion
    rdf:resource="http://www.mydomain.org/uni-ns-old"/>
  <owl:imports
    rdf:resource="http://www.mydomain.org/persons"/>
  <rdfs:label>University Ontology</rdfs:label>
</owl:Ontology>
```

Las clases son definidas usando un elemento owl: Class. Por ejemplo:

```
<owl:Class rdf:ID="associateProfessor">
    <rdfs:subClassOf rdf:resource="#academicStaffMember"/>
</owl:Class>
```

OWL puede describir también propiedades, instancias, relaciones y todo lo definido en las especificaciones de la W3C, además de contar con herramientas para su análisis y representaciones como razonadores y herramientas de representación gráfica.

WEB: <http://www.w3.org/TR/owl-features/>

OWL API

OWL API, es una interfaz de Java y de ejecución para el Lenguaje de Ontologías Web del W3C (OWL), que se utiliza para representar ontologías de Web Semántica. La API está enfocada hacia OWL DL y OWL 2 la próxima. La API de OWL también ofrece enlaces específicos por el hecho de razonadores OWL DL ++ y pellets.

WEB: <http://owlapi.sourceforge.net/>

A.4 BDE

Una Base de Datos Espacial (BDE) permite describir los objetos espaciales que la forman a través de tres características básicas: atributos, localización y topología. Los atributos representan características de los objetos que nos permiten saber qué es lo que son. La localización, representada por la geometría del objeto y su ubicación espacial de acuerdo a un sistema de referencia, permite saber dónde está el objeto y qué espacio ocupa. Por último, la topología definida por medio de las relaciones conceptuales y espaciales entre los objetos, permite mejorar la interpretación semántica del contexto y establecer ciertas jerarquías de elementos a través de sus relaciones. Por otro lado, desde el punto de vista tecnológico, una

Infraestructura de Datos Espaciales debe incluir datos y atributos geográficos, metadatos, métodos de búsqueda, de visualización y mecanismos para proporcionar acceso a los datos espaciales. geográficos, metadatos, métodos de búsqueda, de visualización y mecanismos para proporcionar acceso a los datos espaciales.

A.5 DENUE

El Directorio Estadístico Nacional de Unidades Económicas (DENUE, 03/2011) ofrece información sobre la identificación y ubicación de todos los establecimientos activos en el territorio nacional, de acuerdo a los datos recabados durante la Actualización de Unidades Económicas, 2010; por ello, constituye una herramienta fundamental para la toma de decisiones en los ámbitos público y privado. Entre la información contenida en el Directorio Estadístico Nacional de Unidades Económicas (DENUE) 03/2011, se hallan los datos actualizados de identificación y ubicación de las unidades económicas más importantes del país, las cuales cumplen por lo menos uno de los siguientes criterios:

- Ingresos anuales iguales o mayores a 50 millones de pesos
- Personal ocupado igual o mayor a 100 personas
- Establecimientos que forman parte de una empresa nacional (comparten la misma razón social y están ubicados en más de una entidad)

Estos datos actualizados se obtuvieron en una visita directa a esas unidades económicas, es importante señalar que dicha visita se realizó entre octubre y noviembre del 2010; y en la ficha técnica de cada unidad económica aparece la leyenda "marzo, 2011".

WEB: <http://www.inegi.org.mx/Sistemas/denue/Default.aspx>

A.6 SCIAN

En México, el clasificador oficial de actividades económicas es el Sistema de Clasificación Industrial de América del Norte (SCIAN), construido con Estados Unidos y Canadá, que tienen sus propias versiones nacionales de este clasificador.

El SCIAN 2007 se divide en 20 sectores de actividad en el nivel más general, 94 subsectores, 304 ramas, 617 subramas y, en su nivel más detallado, en 1049 clases de actividad. El SCIAN, permite localizar la actividad de su interés, con su código y descripción, además de conocer los productos, actividades y nombres genéricos de establecimientos que corresponden a cada actividad en su nivel más detallado. Es importante tener presente que cuando los negocios realizan más de una actividad, se clasifican por la más relevante según el ingreso que ésta genera.

El objetivo del SCIAN MÉXICO es proporcionar un marco único, consistente y actualizado para la recopilación, análisis y presentación de estadísticas de tipo económico, que refleje la estructura de la economía mexicana. El SCIAN MÉXICO es la base para la generación, presentación y difusión de todas las estadísticas económicas del INEGI. Su adopción por parte de otras instituciones que también elaboran estadísticas permitirá homologar la información económica que se produce en el país, y con ello contribuir a la de la región de América del Norte.

WEB: <http://www.inegi.org.mx/sistemas/scian/default.aspx>

A.7 Protégé

Protégé es una plataforma de código libre, basada en Java, que provee herramientas para construir modelos de dominio y aplicaciones basadas en conocimiento, usando ontologías. Implementa un conjunto de estructuras para modelado de conocimiento y acciones que soportan la creación, visualización y manipulación de ontologías en varios formatos de representación. Se puede personalizar para proveer un soporte amigable del dominio para crear modelos de conocimiento e introducir datos. También se puede expandir su funcionalidad por medio de otros componentes basados en Java.

WEB: <http://protege.stanford.edu/>

A.8 Google Maps y Fusion Tables

Google Maps

Google Maps es un servicio de Google que ofrece imágenes vía satélite de todo el planeta, combinadas, en el caso de algunos países, con mapas de sus ciudades, lo que unido a sus posibilidades de programación abierta ha dado lugar a diversas utilidades ofrecidas desde numerosas páginas web.

Desde su lanzamiento en febrero de 2005, la aplicación cartográfica de Google ha conmocionado a la comunidad de desarrolladores. Si bien sus principios técnicos de base eran ya conocidos, incluso utilizados desde hacía tiempo, la aplicación de Google los combina de manera inteligente, y sobre todo ofrece una accesibilidad sin igual.

Varios aspectos de Google Maps son los responsables de su facilidad de uso por cualquier usuario: el sistema de deslizamiento de imagen, acoplado a la carga dinámica de nuevas imágenes; la adaptación del mapa al tamaño de ventana del navegador; la interfaz minimalista; la posibilidad de cambiar de tipo de mapa en un clic. Como todas las demás aplicaciones Google, Maps descansa poderosamente sobre la utilización de JavaScript. La carga y el deslizamiento de imagen no podrían efectuarse sin este código. Según el nivel de zoom, un mapa podría descomponerse en varias decenas de miles de cuadrados. Esta descomposición se realiza automáticamente por un script del lado servidor. Cuando el utilizador hace deslizar la imagen en un sentido, JavaScript calcula cuales son los cuadrados involucrados, envía al servidor una pregunta sobre las imágenes en cuestión, y las coloca alineadas junto a las otras. Cada cuadrado es almacenado en un fichero cuyo nombre indica su longitud, su latitud, y el valor del zoom. Recuperar estas informaciones para todos los cuadrados a colocar, no es sino una cuestión de derivación de los datos conocidos para un solo cuadrado. Todo ello necesita, por tanto, una sincronización perfecta entre lo que puede preguntar el código Java Script, y lo que está efectivamente disponible en el servidor. El

programa de recorte inicial de la imagen completa es por tanto extremadamente importante, porque es el responsable, para la precisión del corte y el rigor de su etiquetamiento, del buen desenvolvimiento de los acontecimientos desencadenados por las acciones del usuario

WEB: <https://developers.google.com/maps/>

Google Fusion Tables

Google fusion tables es una potente herramienta gratuita que Google pone a disposición pública para generar, compartir y publicar información en forma de gráficos o mapas. Para utilizar este servicio únicamente necesitamos tener una cuenta de correo con Google, de esta manera podremos crear nuestras propias tablas y representar nuestros datos y acceder a las diferentes tablas que tengan carácter público.

Google Fusion Tables nos permite generar nuestras bases de datos de diferentes formas: podemos crear tablas desde cero o importar archivos ya existentes, como .kml, .csv y hojas de cálculo generadas en Google Docs.

WEB: <http://www.google.com/fusiontables/Home/>

A.9 ZK Framework

ZK es un framework de aplicaciones web en AJAX, completamente en Java de software de código abierto que permite una completa interfaz de usuario para aplicaciones web sin usar JavaScript y con poca programación.

El núcleo de ZK es un mecanismo conducido por eventos basado en AJAX, sustentado sobre 70 componentes XUL y 80 componentes XHTML, y un lenguaje de marcación para diseñar interfaces de usuario. Los programadores diseñan las páginas de su aplicación en componentes XUL/XHTML ricos en características, y los manipulan con eventos disparados

por la actividad del usuario final. Es similar al modelo de programación encontrado en las aplicaciones basadas en GUI de escritorio.

ZK utiliza el acercamiento llamado centrado-en-el-servidor para la sincronización de componentes y el pipelining entre clientes y servidores se haga automáticamente por el motor, y los códigos de Ajax sean completamente transparentes para los desarrolladores de aplicaciones web. Por lo tanto, los usuarios finales obtienen una interacción y respuesta similar a las de una aplicación de escritorio, mientras que la complejidad del desarrollo es similar a la que tendría la codificación de aplicaciones de escritorio.

WEB: <http://www.zkoss.org/>

A.10 Hermit

Hermit es un razonador de ontologías por escrito utilizando el Lenguaje de Ontologías Web (OWL). Dado un fichero OWL, Hermit puede determinar si la ontología es consistente, identificar las relaciones entre las clases de subsunción, y mucho más.

Hermit es el primero a disposición del público razonador OWL basado en original "hypertableau" cálculo, que proporciona un razonamiento mucho más eficiente que cualquier algoritmo previamente conocido. Hermit utiliza la semántica directa y pasa todas las pruebas de conformidad de OWL 2 para razonadores semántica directa.

WEB: <http://www.hermit-reasoner.com/>

A.11 Resnik: Similitud Semántica

Uno de los algoritmos pioneros y más destacados es el propuesto en [Resnik, 1999], el cual propone que la similitud entre dos conceptos c_1 y c_2 de una estructura taxonómica, puede ser obtenida mediante la expresión:

$$sim(c_1, c_2) = \max_{c \in S(c_1, c_2)} (-\log p(c))$$

En donde $S(c_1, c_2)$ representa el conjunto de conceptos de los cuales tanto c_1 como c_2 descienden. Mientras que $p(c)$ es la probabilidad del concepto c . El término concepto hace referencia al conjunto de términos que apuntan a una misma idea. Ahora bien, para estimar la probabilidad de un concepto c , Resnik utiliza la frecuencia de aparición de los términos de ese concepto en el Brown Corpus of American English (compilación de textos en inglés de aproximadamente un millón de palabras seleccionadas de una amplia variedad de fuentes).

ANEXO B – CÓDIGO FUENTE

B.1 Archivo OWL

```
<?xml version="1.0"?>

<!DOCTYPE Ontology [
<!ENTITY xsd "http://www.w3.org/2001/XMLSchema#" >
<!ENTITY xml "http://www.w3.org/XML/1998/namespace" >
<!ENTITY rdfs "http://www.w3.org/2000/01/rdf-schema#" >
<!ENTITY rdf "http://www.w3.org/1999/02/22-rdf-syntax-ns#" >
]>

<Ontology xmlns="http://www.w3.org/2002/07/owl#"

xml:base="http://www.semanticweb.org/ontologies/2011/11/Ontology132372554
1352.owl"
    xmlns:xsd="http://www.w3.org/2001/XMLSchema#"
    xmlns:xml="http://www.w3.org/XML/1998/namespace"
    xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
    xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"

ontologyIRI="http://www.semanticweb.org/ontologies/2011/11/Ontology132372
5541352.owl">
<Prefix name="rdf" IRI="http://www.w3.org/1999/02/22-rdf-syntax-ns#"/>
<Prefix name="" IRI="http://www.w3.org/2002/07/owl#"/>
<Prefix name="xsd" IRI="http://www.w3.org/2001/XMLSchema#"/>
<Prefix name="rdfs" IRI="http://www.w3.org/2000/01/rdf-schema#"/>
<Prefix name="owl" IRI="http://www.w3.org/2002/07/owl#"/>
<Declaration>
<Class IRI="#COMERCIO_AL_POR_MAYOR"/>
</Declaration>
<Declaration>
<Class IRI="#COMERCIO_AL_POR_MENOR"/>
</Declaration>
<Declaration>
<Class IRI="#CONSTRUCCION"/>
</Declaration>
<Declaration>
<Class IRI="#CONSULTORIOS_DENTALES"/>
</Declaration>
<Declaration>
<Class
IRI="#CONSULTORIOS_DE_AUDIOLOGIA_Y_DE_TERAPIA_OCUPACIONAL,_FISICA_Y_DEL_L
ENGUAJE"/>
</Declaration>
<Declaration>
<Class IRI="#CONSULTORIOS_DE_PSICOLOGIA"/>
</Declaration>
```

```

<Declaration>
<Class IRI="#CONSULTORIOS_DE_QUIROPRACTICA"/>
</Declaration>
<Declaration>
<Class IRI="#CONSULTORIOS_MEDICOS"/>
</Declaration>
<Declaration>
<Class
IRI="#ESCUELAS_DE_EDUCACION_BASICA,_MEDIA_Y_PARA_NECESIDADES_ESPECIALES"/
>
</Declaration>
<Declaration>
<Class IRI="#ESCUELAS_DE_EDUCACION_MEDIA_SUPERIOR"/>
</Declaration>
<Declaration>
<Class IRI="#ESCUELAS_DE_EDUCACION_PREESCOLAR"/>
</Declaration>
<Declaration>
<Class IRI="#ESCUELAS_DE_EDUCACION_PRIMARIA"/>
</Declaration>
<Declaration>
<Class IRI="#ESCUELAS_DE_EDUCACION_SECUNDARIA_GENERAL"/>
</Declaration>
<Declaration>
<Class IRI="#ESCUELAS_DE_EDUCACION_SUPERIOR"/>
</Declaration>
<Declaration>
<Class IRI="#ESCUELAS_DE_EDUCACION_SUPERIOR_DEL_SECTOR_PRIVADO"/>
</Declaration>
<Declaration>
<Class IRI="#ESCUELAS_DE_EDUCACION_SUPERIOR_DEL_SECTOR_PUBLICO"/>
</Declaration>
<Declaration>
<Class IRI="#HOSPITALES"/>
</Declaration>
<Declaration>
<Class IRI="#HOSPITALES_DE_OTRAS_ESPECIALIDADES_MEDICAS"/>
</Declaration>
<Declaration>
<Class IRI="#HOSPITALES_GENERALES"/>
</Declaration>
<Declaration>
<Class IRI="#HOSPITALES_GENERALES_DEL_SECTOR_PRIVADO"/>
</Declaration>
<Declaration>
<Class IRI="#HOSPITALES_GENERALES_DEL_SECTOR_PUBLICO"/>
</Declaration>
<Declaration>
<Class
IRI="#HOSPITALES_PSIQUIATRICOS_Y_PARA_EL_TRATAMIENTO_POR_ADICCION"/>
</Declaration>
<Declaration>
<Class IRI="#INDUSTRIAS_MANUFACTURERAS"/>
</Declaration>
<Declaration>
<Class IRI="#OTROS_CONSULTORIOS_PARA_EL_CUIDADO_DE_LA_SALUD"/>
</Declaration>

```

```

<Declaration>
<Class IRI="#OTROS_SERVICIOS_DE_ASISTENCIA_SOCIAL"/>
</Declaration>
<Declaration>
<Class
IRI="#RESIDENCIAS_DE_ASISTENCIA_SOCIAL_Y_PARA_EL_CUIDADO_DE_LA_SALUD"/>
</Declaration>
<Declaration>
<Class IRI="#SERVICIOS_DE_SALUD_Y_DE_ASISTENCIA_SOCIAL"/>
</Declaration>
<Declaration>
<Class IRI="#SERVICIOS_EDUCATIVOS"/>
</Declaration>
<Declaration>
<Class IRI="#SERVICIOS_FINANCIEROS_Y_DE_SEGUROS"/>
</Declaration>
<Declaration>
<Class
IRI="#SERVICIOS_MEDICOS_DE_CONSULTA_EXTERNA_Y_SERVICIOS_RELACIONADOS"/>
</Declaration>
<Declaration>
<Class IRI="#SERVICIOS_PROFESIONALES,_CIENTIFICOS_Y_TECNICOS"/>
</Declaration>
<Declaration>
<Class IRI="#TODOS_LOS_SECTORES"/>
</Declaration>
<SubClassOf>
<Class IRI="#COMERCIO_AL_POR_MAYOR"/>
<Class IRI="#TODOS_LOS_SECTORES"/>
</SubClassOf>
<SubClassOf>
<Class IRI="#COMERCIO_AL_POR_MENOR"/>
<Class IRI="#TODOS_LOS_SECTORES"/>
</SubClassOf>
<SubClassOf>
<Class IRI="#CONSTRUCCION"/>
<Class IRI="#TODOS_LOS_SECTORES"/>
</SubClassOf>
<SubClassOf>
<Class IRI="#CONSULTORIOS_DENTALES"/>
<Class
IRI="#SERVICIOS_MEDICOS_DE_CONSULTA_EXTERNA_Y_SERVICIOS_RELACIONADOS"/>
</SubClassOf>
<SubClassOf>
<Class
IRI="#CONSULTORIOS_DE_AUDIOLOGIA_Y_DE_TERAPIA_OCUPACIONAL,_FISICA_Y_DEL_L
ENGUAJE"/>
<Class IRI="#OTROS_CONSULTORIOS_PARA_EL_CUIDADO_DE_LA_SALUD"/>
</SubClassOf>
<SubClassOf>
<Class IRI="#CONSULTORIOS_DE_PSICOLOGIA"/>
<Class IRI="#OTROS_CONSULTORIOS_PARA_EL_CUIDADO_DE_LA_SALUD"/>
</SubClassOf>
<SubClassOf>
<Class IRI="#CONSULTORIOS_DE_QUIROPRACTICA"/>
<Class IRI="#OTROS_CONSULTORIOS_PARA_EL_CUIDADO_DE_LA_SALUD"/>
</SubClassOf>

```

```

<SubClassOf>
<Class IRI="#CONSULTORIOS_MEDICOS"/>
<Class
IRI="#SERVICIOS_MEDICOS_DE_CONSULTA_EXTERNA_Y_SERVICIOS_RELACIONADOS"/>
</SubClassOf>
<SubClassOf>
<Class
IRI="#ESCUELAS_DE_EDUCACION_BASICA,_MEDIA_Y_PARA_NECESIDADES_ESPECIALES"/
>
<Class IRI="#SERVICIOS_EDUCATIVOS"/>
</SubClassOf>
<SubClassOf>
<Class IRI="#ESCUELAS_DE_EDUCACION_MEDIA_SUPERIOR"/>
<Class IRI="#SERVICIOS_EDUCATIVOS"/>
</SubClassOf>
<SubClassOf>
<Class IRI="#ESCUELAS_DE_EDUCACION_PREESCOLAR"/>
<Class IRI="#SERVICIOS_EDUCATIVOS"/>
</SubClassOf>
<SubClassOf>
<Class IRI="#ESCUELAS_DE_EDUCACION_PRIMARIA"/>
<Class IRI="#SERVICIOS_EDUCATIVOS"/>
</SubClassOf>
<SubClassOf>
<Class IRI="#ESCUELAS_DE_EDUCACION_SECUNDARIA_GENERAL"/>
<Class IRI="#SERVICIOS_EDUCATIVOS"/>
</SubClassOf>
<SubClassOf>
<Class IRI="#ESCUELAS_DE_EDUCACION_SUPERIOR"/>
<Class IRI="#SERVICIOS_EDUCATIVOS"/>
</SubClassOf>
<SubClassOf>
<Class IRI="#ESCUELAS_DE_EDUCACION_SUPERIOR_DEL_SECTOR_PRIVADO"/>
<Class IRI="#ESCUELAS_DE_EDUCACION_SUPERIOR"/>
</SubClassOf>
<SubClassOf>
<Class IRI="#ESCUELAS_DE_EDUCACION_SUPERIOR_DEL_SECTOR_PUBLICO"/>
<Class IRI="#ESCUELAS_DE_EDUCACION_SUPERIOR"/>
</SubClassOf>
<SubClassOf>
<Class IRI="#HOSPITALES"/>
<Class IRI="#SERVICIOS_DE_SALUD_Y_DE_ASISTENCIA_SOCIAL"/>
</SubClassOf>
<SubClassOf>
<Class IRI="#HOSPITALES_DE_OTRAS_ESPECIALIDADES_MEDICAS"/>
<Class IRI="#HOSPITALES"/>
</SubClassOf>
<SubClassOf>
<Class IRI="#HOSPITALES_GENERALES"/>
<Class IRI="#HOSPITALES"/>
</SubClassOf>
<SubClassOf>
<Class IRI="#HOSPITALES_GENERALES_DEL_SECTOR_PRIVADO"/>
<Class IRI="#HOSPITALES_GENERALES"/>
</SubClassOf>
<SubClassOf>
<Class IRI="#HOSPITALES_GENERALES_DEL_SECTOR_PUBLICO"/>

```

```

<Class IRI="#HOSPITALES_GENERALES"/>
</SubClassOf>
<SubClassOf>
<Class
IRI="#HOSPITALES_PSIQUIATRICOS_Y_PARA_EL_TRATAMIENTO_POR_ADICCION"/>
<Class IRI="#HOSPITALES"/>
</SubClassOf>
<SubClassOf>
<Class IRI="#INDUSTRIAS_MANUFACTURERAS"/>
<Class IRI="#TODOS_LOS_SECTORES"/>
</SubClassOf>
<SubClassOf>
<Class IRI="#OTROS_CONSULTORIOS_PARA_EL_CUIDADO_DE_LA_SALUD"/>
<Class
IRI="#SERVICIOS_MEDICOS_DE_CONSULTA_EXTERNA_Y_SERVICIOS_RELACIONADOS"/>
</SubClassOf>
<SubClassOf>
<Class IRI="#OTROS_SERVICIOS_DE_ASISTENCIA_SOCIAL"/>
<Class IRI="#SERVICIOS_DE_SALUD_Y_DE_ASISTENCIA_SOCIAL"/>
</SubClassOf>
<SubClassOf>
<Class
IRI="#RESIDENCIAS_DE_ASISTENCIA_SOCIAL_Y_PARA_EL_CUIDADO_DE_LA_SALUD"/>
<Class IRI="#SERVICIOS_DE_SALUD_Y_DE_ASISTENCIA_SOCIAL"/>
</SubClassOf>
<SubClassOf>
<Class IRI="#SERVICIOS_DE_SALUD_Y_DE_ASISTENCIA_SOCIAL"/>
<Class IRI="#TODOS_LOS_SECTORES"/>
</SubClassOf>
<SubClassOf>
<Class IRI="#SERVICIOS_EDUCATIVOS"/>
<Class IRI="#TODOS_LOS_SECTORES"/>
</SubClassOf>
<SubClassOf>
<Class IRI="#SERVICIOS_FINANCIEROS_Y_DE_SEGUROS"/>
<Class IRI="#TODOS_LOS_SECTORES"/>
</SubClassOf>
<SubClassOf>
<Class
IRI="#SERVICIOS_MEDICOS_DE_CONSULTA_EXTERNA_Y_SERVICIOS_RELACIONADOS"/>
<Class IRI="#SERVICIOS_DE_SALUD_Y_DE_ASISTENCIA_SOCIAL"/>
</SubClassOf>
<SubClassOf>
<Class IRI="#SERVICIOS_PROFESIONALES,_CIENTIFICOS_Y_TECNICOS"/>
<Class IRI="#TODOS_LOS_SECTORES"/>
</SubClassOf>
</Ontology>

```

```

<!-- Generated by the OWL API (version 3.2.3.1824)
http://owlapi.sourceforge.net -->

```

B.2 Similitud.java

```
import org.semanticweb.owlapi.apibinding.OWLManager;
import org.semanticweb.owlapi.io.*;
import org.semanticweb.owlapi.model.*;
import org.semanticweb.owlapi.util.SimpleIRIMapper;
import org.semanticweb.owlapi.apibinding.OWLManager;

import org.semanticweb.owlapi.reasoner.OWLReasonerFactory;
import org.semanticweb.owlapi.reasoner.OWLReasoner;

import org.semanticweb.Hermit.Reasoner;

import java.io.File;
import java.io.FileNotFoundException;
import java.io.IOException;
import java.net.URI;
import java.net.UnknownHostException;
import java.util.Collections;
import java.util.HashSet;
import java.util.Map;
import java.util.Set;
import org.semanticweb.owlapi.reasoner.ConsoleProgressMonitor;
import org.semanticweb.owlapi.reasoner.InferenceType;
import org.semanticweb.owlapi.reasoner.Node;
import org.semanticweb.owlapi.reasoner.NodeSet;
//import org.semanticweb.owlapi.reasoner.OWLReasoner;
import org.semanticweb.owlapi.reasoner.OWLReasonerConfiguration;
import org.semanticweb.owlapi.reasoner.SimpleConfiguration;
import org.semanticweb.owlapi.util.OWLOntologyWalker;
import org.semanticweb.owlapi.util.OWLOntologyWalkerVisitor;

/**
 *
 * @author Ivan
 */
public class Main {

    public static void main(String[] args) {

        try {

            OWLOntologyManager manager =
                OWLManager.createOWLOntologyManager();

            File file = new
                File("C:/Users/Ivan/Desktop/ponto/prueba.owl");

            OWLOntology denue =
                manager.loadOntologyFromOntologyDocument(file);

            System.out.println("Loaded ontology: " + denue);

            IRI documentIRI = manager.getOntologyDocumentIRI(denue);
```

```

System.out.println("    from: " + documentIRI);
    //Imprime las clases
for (OWLClass cls : denue.getClassesInSignature()) {
    System.out.println(cls);
}

    OWLReasonerFactory reasonerFactory = new
Reasoner.ReasonerFactory();
    ///////////////
    OWLReasoner r = reasonerFactory.createReasoner(denue);
r.precomputeInferences(InferenceType.CLASS_HIERARCHY);

boolean consistent = r.isConsistent();
System.out.println("Verifica Consistencia: " + consistent);
System.out.println("\n");

    ConsoleProgressMonitor progressMonitor = new
ConsoleProgressMonitor();
    OWLReasonerConfiguration config = new
SimpleConfiguration(progressMonitor);
    OWLReasoner reasoner = reasonerFactory.createReasoner(denue,
config);

reasoner.precomputeInferences();

boolean consistent2 = reasoner.isConsistent();
    // System.out.println("Consistent: " + consistent2);
System.out.println("\n");

    ///////
    Node<OWLClass> bottomNode =
reasoner.getUnsatisfiableClasses();

    Set<OWLClass> unsatisfiable =
bottomNode.getEntitiesMinusBottom();
if (!unsatisfiable.isEmpty()) {
    System.out.println("The following classes are unsatisfiable: ");
    for (OWLClass cls : unsatisfiable) {
        System.out.println("    " + cls);
    }
} else {
    System.out.println("There are no unsatisfiable classes");
}
System.out.println("\n");

    ///////////////////////////////////////////////////

    OWLDataFactory fac = manager.getOWLDataFactory();
    OWLClass servsalud =
fac.getOWLClass(IRI.create("http://www.semanticweb.org/ontologies/2011/11
/Ontology1323725541352.owl#SERVICIOS_DE_SALUD_Y_DE_ASISTENCIA_SOCIAL"));

    NodeSet<OWLClass> subClses =
reasoner.getSubClasses(servsalud, true);

    Set<OWLClass> clses = subClses.getFlattened();

```

```

System.out.println("Subclasses de SERVICIOS DE SALUD: ");
for (OWLClass cls : clses) {
    System.out.println("    " + cls);
}
System.out.println("\n");

// Look up and print all direct subclasses for all classes
// a NodeSet represents a set of Nodes.
// a Node represents a set of equivalent classes
/*for (OWLClass c : denue.getClassesInSignature()) {
    // the boolean argument specifies direct subclasses
    NodeSet<OWLClass> subClasses = r.getSubClasses(c, true);
    for (OWLClass subClass : subClasses.getFlattened())
        System.out.println(labelFor(subClass,denue)+ " subclass of " +
            labelFor(c,denue));
}*/
//////////

        } catch (OWLOntologyCreationIOException e) {
            // IOExceptions during loading get wrapped in an
            OWLOntologyCreationIOException
                IOException ioException = e.getCause();
            if (ioException instanceof FileNotFoundException) {
                System.out.println("Could not load ontology. File not found: " +
                    ioException.getMessage());
            } else if (ioException instanceof UnknownHostException) {
                System.out.println("Could not load ontology. Unknown host: " +
                    ioException.getMessage());
            } else {
                System.out.println("Could not load ontology: " +
                    ioException.getClass().getSimpleName() + " " + ioException.getMessage());
            }
        } catch (UnparsableOntologyException e) {
            // If there was a problem loading an ontology because there
            // are syntax errors in the document (file) that
            // represents the ontology then an
            UnparsableOntologyException is thrown
            System.out.println("Could not parse the ontology: " + e.getMessage());
            // A map of errors can be obtained from the exception
            Map<OWLParse, OWLParseException> exceptions =
            e.getExceptions();
            // The map describes which parsers were tried and what the
            errors were
            for (OWLParse parser : exceptions.keySet()) {
                System.out.println("Tried to parse the ontology with the " +
                    parser.getClass().getSimpleName() + " parser");
                System.out.println("Failed because: " +
                    exceptions.get(parser).getMessage());
            }
        } catch (UnloadableImportException e) {
            // If our ontology contains imports and one or more of the
            imports could not be loaded then an
            // UnloadableImportException will be thrown (depending on the
            missing imports handling policy)

```

```

System.out.println("Could not load import: " +
e.getImportsDeclaration());
    // The reason for this is specified and an
OWLOntologyCreationException
    OWLOntologyCreationException cause =
e.getOntologyCreationException();
System.out.println("Reason: " + cause.getMessage());
    } catch (OWLOntologyCreationException e) {
System.out.println("Could not load ontology: " + e.getMessage());
    }

    }
}

```

B.3 Aplicación ZK (interfaz de usuario)

```

<?page title="Auto Generated index.zul"?>
<window id="principal" border="normal" width="100%" height="900px">
<borderlayout>
<center title="RESULTADOS">
<div id="resultado" style="background:#ffffff; height:100%" >

<html><![CDATA[

<head>
<script type="text/javascript"
src="http://maps.google.com/maps/api/js?sensor=false"></script>
<script type="text/javascript">
    function initialize() {
        var tableid = 2134897;
        var layer;
        var latlng = new google.maps.LatLng(19.5030, -99.1475);
        var myOptions = {
            zoom: 15,
            center: latlng,
            disableDefaultUI: true,
            mapTypeId: google.maps.MapTypeId.ROADMAP
        };
        var map = new google.maps.Map(document.getElementById("map_canvas"),
myOptions);

        layer = new google.maps.FusionTablesLayer(1953018, {
            suppressInfoWindows: false
        });
        layer.setMap(map);

        var circle = new google.maps.Circle({
            center: new google.maps.LatLng(19.5030, -99.1475),
            radius: 1000,
            map: map,
            fillOpacity: 0.2,

```

```

        strokeOpacity: 0.5,
        strokeWeight: 1
    });

}

function gen() {
    layer.setQuery("select conceptos from 1953018");
}

</script>
</head>
<body onload="initialize()">

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

</body>
]]></html>
</div>
</center>

<east title="CONCEPTOS" size="30%" flex="true" splittable="false"
collapsible="true">
<div id="conceptosgir" style="background:#FFFFFF; height:100%" >

<tree id="tree" rows="18" multiple="true" checkmark="true">
<treecols>
<treecol label="DENUE"/>
</treecols>
<treechildren>
<treeitem>
<treerow>
<treecell label="COMERCIO AL POR MENOR"/>
</treerow>
<treechildren>
<treeitem>
<treerow>
<treecell label="COMERCIO AL POR MENOR DE ABARROTES ALIMENTOS BEBIDAS
HIELO Y TABACO"/>
</treerow>
<treechildren>
<treeitem>
<treerow>
<treecell label="COMERCIO AL POR MENOR DE ABARROTES Y ALIMENTOS"/>
</treerow>
<treechildren>
<treeitem>
<treerow>
<treecell label="COMERCIO AL POR MENOR EN TIENDAS DE ABARROTES
ULTRAMARINOS Y MISCELANEAS"/>
</treerow>
</treeitem>
</treechildren>
</treeitem>
</treechildren>
</treeitem>

```

```
<treeitem>
<treerow>
<treecell label="COMERCIO AL POR MENOR DE PRODUCTOS TEXTILES BISUTERIA
ACCESORIOS DE VESTIR Y CALZADO"/>
</treerow>
<treechildren>
<treeitem>
<treerow>
<treecell label="COMERCIO AL POR MENOR DE ROPA BISUTERIA Y ACCESORIOS DE
VESTIR"/>
</treerow>
<treechildren>
<treeitem>
<treerow>
<treecell label="COMERCIO AL POR MENOR DE ROPA DE CUERO Y PIEL Y DE OTROS
ARTICULOS DE ESTOS MATERIALES"/>
</treerow>
</treeitem>
<treeitem>
<treerow>
<treecell label="COMERCIO AL POR MENOR DE ROPA EXCEPTO DE BEBE Y
LENCERIA"/>
</treerow>
</treeitem>
</treechildren>
</treeitem>
</treechildren>
</treeitem>
</treechildren>
</treeitem>
<treeitem>
<treerow>
<treecell label="COMERCIO AL POR MAYOR"/>
</treerow>
</treeitem>
<treeitem>
<treerow>
<treecell label="SERVICIOS EDUCATIVOS"/>
</treerow>
<treechildren>
<treeitem>
<treerow>
<treecell label="ESCUELAS DE EDUCACION SUPERIOR"/>
</treerow>
<treechildren>
<treeitem>
<treerow>
<treecell label="ESCUELAS DE EDUCACION SUPERIOR DEL SECTOR PUBLICO"/>
</treerow>
</treeitem>
<treeitem>
<treerow>
<treecell label="ESCUELAS DE EDUCACION SUPERIOR DEL SECTOR PRIVADO"/>
</treerow>
</treeitem>
</treechildren>
</treeitem>
```

```
<treeitem>
<treerow>
<treecell label="ESCUELAS DE EDUCACION MEDIA SUPERIOR"/>
</treerow>
</treeitem>
<treeitem>
<treerow>
<treecell label="ESCUELAS DE EDUCACION SECUNDARIA GENERAL"/>
</treerow>
</treeitem>
</treechildren>
</treeitem>
<treeitem>
<treerow>
<treecell label="SERVICIOS DE ALOJAMIENTO TEMPORAL Y DE PREPARACION DE
ALIMENTOS Y BEBIDAS"/>
</treerow>
<treechildren>
<treeitem>
<treerow>
<treecell label="SERVICIOS DE PREPARACION DE ALIMENTOS Y BEBIDAS"/>
</treerow>
<treechildren>
<treeitem>
<treerow>
<treecell label="RESTAURANTES CON SERVICIO COMPLETO"/>
</treerow>
</treeitem>
<treeitem>
<treerow>
<treecell label="RESTAURANTES DE AUTOSERVICIO COMIDA PARA LLEVAR Y OTROS
RESTAURANTES CON SERVICIO LIMITADO"/>
</treerow>
</treeitem>
<treeitem>
<treerow>
<treecell label="CENTROS NOCTURNOS BARES CANTINAS Y SIMILARES"/>
</treerow>
</treeitem>
</treechildren>
</treeitem>
</treechildren>
</treeitem>
<treeitem>
<treerow>
<treecell label="OTROS SERVICIOS EXCEPTO ACTIVIDADES GUBERNAMENTALES"/>
</treerow>
<treechildren>
<treeitem>
<treerow>
<treecell label="SERVICIOS DE REPARACION Y MANTENIMIENTO"/>
</treerow>
<treechildren>
<treeitem>
<treerow>
<treecell label="REPARACION Y MANTENIMIENTO DE AUTOMOVILES Y CAMIONES"/>
</treerow>
```

```

<treechildren>
<treeitem>
<treerow>
<treecell label="REPARACION MECANICA Y ELECTRICA DE AUTOMOVILES Y
CAMIONES"/>
</treerow>
<treechildren>
<treeitem>
<treerow>
<treecell label="REPARACION DE SUSPENSIONES DE AUTOMOVILES Y CAMIONES"/>
</treerow>
</treeitem>
<treeitem>
<treerow>
<treecell label="REPARACION DE TRANSMISIONES DE AUTOMOVILES Y CAMIONES"/>
</treerow>
</treeitem>
<treeitem>
<treerow>
<treecell label="REPARACION MECANICA EN GENERAL DE AUTOMOVILES Y
CAMIONES"/>
</treerow>
</treeitem>
</treechildren>
</treeitem>
</treechildren>
</treeitem>
</treechildren>
</treeitem>
</treechildren>
</treeitem>
<treerow>
<treecell label="SERVICIOS DE SALUD Y DE ASISTENCIA SOCIAL"/>
</treerow>
<treechildren>
<treeitem>
<treerow>
<treecell label="HOSPITALES"/>
</treerow>
<treechildren>
<treeitem>
<treerow>
<treecell label="HOSPITALES GENERALES"/>
</treerow>
<treechildren>
<treeitem>
<treerow>
<treecell label="HOSPITALES GENERALES DEL SECTOR PUBLICO"/>
</treerow>
</treeitem>
<treeitem>
<treerow>
<treecell label="HOSPITALES GENERALES DEL SECTOR PRIVADO"/>
</treerow>
</treeitem>
</treechildren>

```

```

</treeitem>
<treeitem>
<treerow>
<treecell label="HOSPITALES DE OTRAS ESPECIALIDADES MEDICAS"/>
</treerow>
</treeitem>
</treechildren>
</treeitem>
<treeitem>
<treerow>
<treecell label="SERVICIOS MEDICOS DE CONSULTA EXTERNA Y SERVICIOS
RELACIONADOS"/>
</treerow>
<treechildren>
<treeitem>
<treerow>
<treecell label="CONSULTORIOS DENTALES"/>
</treerow>
</treeitem>
<treeitem>
<treerow>
<treecell label="CONSULTORIOS MEDICOS"/>
</treerow>
</treeitem>
</treechildren>
</treeitem>
<treeitem>
<treerow>
<treecell label="OTROS SERVICIOS DE ASISTENCIA SOCIAL"/>
</treerow>
</treeitem>
</treechildren>
</treeitem>
<treeitem label="CONSTRUCCION"/>
<treeitem label="INDUSTRIAS MANUFACTURERAS"/>
<treeitem label="SERVICIOS PROFESIONALES CIENTIFICOS Y TECNICOS"/>
<treeitem label="SERVICIOS FINANCIEROS Y DE SEGUROS"/>
</treechildren>
</tree>
</div>
</east>
<north title="RSPIG" size="210px" flex="true" border="0"
splittable="true" collapsible="false">
<div id="posper" style="background:#FFFFFF; height:100%" >

<tablelayout columns="4">
<tablechildren rowspan="1">
<panel border="none" width="300px" >
<panelchildren>
<image src="/images/imgtesis.jpg" width="290px" height="180px"/>
</panelchildren>
</panel>
</tablechildren>
<tablechildren>
<panel height="180px" width="260px" border="none">
<panelchildren>
<grid><rows>

```

```

<row>Latitud: <doublebox id="mylat" value="0" /></row>
<row>Longitud: <doublebox id="mylng" value="0" /></row>
<row>Zoom: <intbox id="myzoom" value="0" /></row>
</rows></grid>
<image src="/images/esp2.jpg" width="260px" height="2px"/>
<image src="/images/esp2.jpg" width="65px" height="2px"/>
<button width="125px" height="40px" label="UBICATE"
onClick="win.setSizable(true);win.doPopup();win.position =
'center';"/>
</panelchildren>
</panel>
</tablechildren>
<tablechildren>
<panel height="180px" width="240px" border="none">
<panelchildren>
<image src="/images/esp.jpg" width="80px" height="20px"/>
<image src="/images/up.jpg" width="100px" height="100px"/>
<image src="/images/esp.jpg" width="180px" height="20px"/>
<image src="/images/esp.jpg" width="58px" height="20px"/>
<combobox>
<comboitem label="Estudiante"
description="Perfil de Usuario Estudiante" />
<comboitem label="Profesor"
description="Perfil de Usuario Profesor" />
<comboitem label="Seleccionar Perfil"
description="" />
</combobox>
</panelchildren>
</panel>
</tablechildren>
<tablechildren rowspan="1">
<panel border="none" width="300px" >
<panelchildren>
<!-- <image src="/images/girb.png" width="53px" height="106px"/> -->
<image src="/images/esp.jpg" width="60px" height="60px"/>
<image src="/images/girb.png" width="53px" height="106px"/>
<button width="125px" height="40px" label="BUSCAR"/>
</panelchildren>
</panel>
</tablechildren>
</tablelayout>
</div>
</north>
</borderlayout>
<window id="win" title="UBICATE" border="normal">
<script type="text/javascript"
content="zk.googleAPIkey='ABQIAAAAmGxmYR57XDAbAumS9tV5fxQXyy1OlR69a1vFTcU
cpV6DXdesOBSEHfkcwSzwEwBT7UzVx8ep8vjA'" />
<vbox>
<gmaps id="mymap" width="500px" height="300px" showSmallCtrl="true">
<attribute name="onMapMove">
mylat.setValue(self.getLat());
mylng.setValue(self.getLng());
</attribute>
<attribute name="onMapZoom">
myzoom.setValue(self.getZoom());
</attribute>

```

```
<attribute name="onMapClick">
Gmarker gmarker = event.getGmarker();
if (gmarker != null) {
gmarker.setOpen(true);
}
</attribute>

<gmarker id="mymark" lat="19.5030" lng="-99.1475">
<attribute name="content">
Centro de Investigacion en Computacion, &lt;a
href="http://www.cic.ipn.mx"&gt;CIC&lt;/a&gt;
</attribute>
</gmarker>
</gmaps>
</vbox>
</window>
</window>
```