



**INSTITUTO POLITÉCNICO NACIONAL**  
**CENTRO DE INVESTIGACIÓN EN COMPUTACIÓN**  
**LABORATORIO DE PROCESAMIENTO INTELIGENTE DE**  
**INFORMACIÓN GEO-ESPACIAL**



**RECUPERACIÓN Y PONDERACIÓN DE**  
**INFORMACIÓN GEOGRÁFICA DESDE**  
**REPOSITORIOS NO ESTRUCTURADOS**  
**CONDUCCIDAS POR ONTOLOGÍAS**

**T E S I S**

QUE PARA OBTENER EL GRADO DE  
**DOCTOR EN CIENCIAS DE LA COMPUTACIÓN**

PRESENTA:

**MIGUEL FÉLIX MATA RIVERA**

DIRECTOR DE TESIS: **Dr. SERGUEI LEVACHKINE**

MÉXICO, DF.

Abril 2009

## *A G R A D E C I M I E N T O S*

*A DIOS, porque este triunfo académico se forjó desde que era dependiente de mostrador. Me has recompensado de tantas dificultades. Sea éste un testimonio de tu infinita sabiduría.*

*A mis dos ángeles que tengo en el cielo, mi papá y mi abuelita ¡Miren hasta donde me ha llevado esta necesidad por estudiar! En mi corazón llevo su capacidad de amar y su espíritu de sacrificio.*

*A mi mamá, la adversidad nos golpeó, pero tú hiciste posible este gran logro ¡Te quiero Muchísimo!*

*A mis hermanas por su amor y cariño en las épocas de tribulación y soledad, estamos de pie, les amo.*

*A la mujer que ha transformado y alegrado ampliamente mi vida ¡Te Amo Mi PAo!*

*A mis sobrinos, sus ocurrencias y diabluras me recuerdan que cualquier sueño es realizable.*

*A mi Tía Rosa y mis primas por estar en los momentos difíciles. Tía, su apoyo permitió que este sueño se hiciera realidad. Mi agradecimiento infinito.*

*Al Dr. Serguei, por las revisiones a artículos y comentarios para culminar esta tesis.*

*A Walter quien me ha mostrado la esencia de la amistad, tanto en las buenas como en las malas.*

*A Hugo, su amistad "a distancia" fue el pan cuando había que probar la sal académica.*

*A Saúl Puga quien me ha apoyado para cumplir este sueño académico, mi agradecimiento total.*

*A Roberto, por escucharme, tu entusiasmo por la computación, me recordó porque estoy aquí.*

*A mis alumnos, ustedes son mi principal motivación académica y mis mejores maestros.*

*A la familia Cortez Herrera por sus atenciones y hospitalidad en esta etapa final de tesis.*

*A los Geos y Geocitos, Child, Gera y Jhon, su amistad no llegó tarde, sólo que ya no la esperaba, las vacaciones y pláticas en la sobremesa, son invaluableles.*

*A mi alma Mater, el IPN, quien me lo ha dado todo, desde niño soñé ser parte de esta gran casa de estudios, me ha dado una identidad, pero jamás imaginé que la recompensa sería tan grande.*

*A los profesores que con su fervor y pasión académica me contagiaron y marcaron mi vida académica.*

*A todos los organismos e instituciones que nos permiten alcanzar un sueño académico, apoyándonos con becas y financiamiento para proyectos de investigación.*

*A la comunidad científica GIS, sus revisiones y comentarios le dieron vida a esta tesis.*

*A la Adversidad, gracias a ti he vivido intensamente ¡Vamos por más!*

*Mis Colores son el verde y el blanco, porque  
el rojo lo llevo en la sangre.*

*¡Viva el Politécnico!*

# RESUMEN

En esta investigación se muestra una metodología integral para la recuperación de información geográfica (*iGIR*) y un método para ponderar los resultados de la recuperación (*iRank*). La metodología considera consultas espaciales que contengan tres elementos, los cuales son objetos o lugares geográficos y una relación espacial entre estos objetos. El resultado de la búsqueda es el conjunto de documentos con datos geográficos, que tiene la mayor similitud, de acuerdo a las características evaluadas.

El éxito de la recuperación se basa en la integración de tres criterios de búsqueda para tres fuentes de datos geográficas heterogéneas: Ontologías Geográficas (*GeoOntologías*), diccionarios de datos y archivos vectoriales. Las cuales describen a los objetos geográficos utilizando la topología, los atributos geográficos y la semántica espacial.

Los resultados obtenidos demuestran que al combinar las diferentes fuentes de información y los criterios de búsqueda, se puede obtener información relevante que podría ser omitida al recuperarse de forma independiente en cada fuente. Sin embargo, el nivel de precisión en la recuperación se determina con el perfil del usuario, debido a que la experiencia de éste, determina el grado de formalidad en efectuar una consulta. El apropiado procesamiento de la intención de búsqueda y el perfil resulta en una mejor recuperación y ponderación.

Los documentos recuperados se ponderan con el método *iRank*, el cual integra tres métodos de ponderación geográfica. El proceso de integración se basa en el método de confusión. Nuevamente, la ponderación integral resultó mejor que la ponderación aislada y ofreció mayor certidumbre de la relevancia. Finalmente, lo novedoso de esta tesis, es el enfoque integrador que procesa la semántica espacial y la intención del usuario basada en su perfil.

# TABLA DE CONTENIDO

---

CAPÍTULO 1	INTRODUCCIÓN.....	1
1.1	Objetivo general.....	7
CAPÍTULO 2	ESTADO DEL ARTE.....	13
2.1	Enfoque General	14
2.2	Recuperación de información clásica	19
2.3	Recuperación de información clásica y geográfica usando similitud.....	22
2.4	Recuperación de información clásica usando semántica y contexto.....	26
2.5	Recuperación de información geográfica y clásica en la web....	31
2.6	Recuperación de información geográfica usando relaciones espaciales.....	38
2.7	Recuperación y ponderación de información con enfoques híbridos.....	40
2.8	Recuperación de información apoyada en la construcción de ontologías.....	44
2.9	Comentarios Finales.....	45
CAPÍTULO 3	MARCO TEÓRICO.....	48
3.1	Recuperación de Información.....	49
3.2	Información Geográfica y Espacial.....	57
3.3	Fuentes de Datos Geográficos.....	61
3.4	Fuentes de Datos Vectoriales.....	64
3.5	Ontologías como Fuente de datos.....	66
3.6	Consultas Geográficas y Espaciales.....	70
3.7	Indexado geográfico y espacial.....	80
3.8	Análisis Espacial.....	83
3.9	Redes semánticas.....	90

3.10	Jerarquías y Teoría de Confusión.....	96
3.11	Teoría de Confusión.....	99
3.12	Comentarios Finales.....	103
<b>CAPÍTULO 4</b>	<b>METODOLOGÍA DE RECUPERACIÓN.....</b>	<b>106</b>
4.1	Primera fuente de datos geográfica: GeoOntologías.....	108
4.2	Segunda fuente de datos geográfica: Diccionarios geográficos...	117
4.3	Tercera fuente de datos geográfica: Archivos topológicos.....	120
4.4	Estrategia de Recuperación de Información Geográfica ( <i>iGIR</i> )...	123
4.5	Algoritmo de exploración de GeoOntología (OntoExplora).....	128
4.6	Recuperación de información geográfica en las GeoOntologías.	132
4.7	Recuperación de información geográfica en diccionarios geográficos.....	135
4.8	Recuperación de información geográfica en archivos topológicos.....	138
4.9	Integración y despliegue de resultados.....	140
<b>CAPÍTULO 5</b>	<b>METODOLOGÍA DE PONDERACIÓN.....</b>	<b>152</b>
5.1	Ponderación Clásica y Ponderación Geográfica.....	153
5.2	<i>iRank</i> : Ponderación integrando topología, geografía y semántica espacial.....	156
5.3	Ponderación Conceptual.....	159
5.4	Ponderación Geográfica.....	164
5.5	Ponderación Topológica.....	167
5.6	Integración de la ponderación.....	173
<b>CAPÍTULO 6</b>	<b>PRUEBAS Y RESULTADOS.....</b>	<b>177</b>
6.1	Pruebas para la recuperación y ponderación conceptual.....	179
6.2	Pruebas con Consultas en los diccionarios de datos.....	182
6.3	Recuperación en archivos topológicos.....	184
6.4	Resultado de la recuperación en formato KML.....	187

6.5	Resultados finales ponderados ( <i>iGIR</i> + <i>iRank</i> ).....	188
<b>CAPÍTULO 7</b>	<b>CONCLUSIONES Y TRABAJO FUTURO.....</b>	<b>198</b>
7.1	Conclusiones generales.....	199
7.2	Conclusiones particulares.....	200
7.3	Productos derivados de la tesis.....	203
7.4	Aportaciones.....	204
7.5	Trabajos a futuro.....	204
<b>ANEXO A</b>	<b>GeoOntologías.....</b>	<b>217</b>
<b>ANEXO B</b>	<b>Código Fuente.....</b>	<b>303</b>

# ÍNDICE

---

CAPÍTULO 1	INTRODUCCIÓN.....	1
1.1	Objetivo general .....	7
1.2	Objetivos particulares .....	8
1.3	Metas .....	9
1.4	Hipótesis .....	10
1.5	Notaciones del documento .....	11
CAPÍTULO 2	ESTADO DEL ARTE .....	13
2.1	Enfoque General .....	14
2.2	Recuperación de información clásica.....	19
2.2.1	Baeza Yates y Ribeiro-Neto (1999) .....	19
2.2.2	Kobayashi y Takeda (2000).....	20
2.2.3	Braga et al (2001) .....	21
2.3	Recuperación de información clásica y geográfica usando similitud.....	22
2.3.1	Varelas et al (2005).....	22
2.3.2	L. Andrade y M. Silva (2006) .....	23
2.3.3	Levachkine et al (2004) .....	24
2.4	Recuperación de información clásica usando semántica y contexto.....	26
2.4.1	Rocha et al (2004) .....	27
2.4.2	Reiner Kraft et al (2005).....	28
2.4.3	Guha et al (2003) .....	28
2.4.4	Peter Anick (1997) .....	29
2.4.5	Max Egenhofer (2002) .....	30
2.5	Recuperación de información geográfica y clásica en la web.....	31
2.5.1	Hawking et al (2001) .....	31
2.5.2	Sergey Brin et al (1998) .....	32
2.5.3	Zhou et al (2005) .....	33
2.5.4	D. Santos y M. Chaves (2006).....	35
2.5.5	Jones et al (2002).....	35
2.5.6	T.Delboni et al (2005) .....	36
2.5.7	Vaid et al (2005) .....	37
2.5.8	Jens Graupmann and Ralf Schenkel (2006) .....	38
2.6	Recuperación de información geográfica usando relaciones espaciales .....	38
2.6.1	Walker et al (2005) .....	39
2.7	Recuperación y ponderación de información con enfoques híbridos.....	40
2.7.1	G. Cai (2002) .....	41
2.7.2	Wang et al (2005) .....	41
2.7.3	Paul Clough (2005).....	42
2.7.4	Zhisheng et al (2006).....	43

2.8	Recuperación de información apoyada en la construcción de ontologías .....	44
2.8.1	Sang Ok et al (2003).....	44
2.9	Comentarios Finales .....	45
CAPÍTULO 3 MARCO TEÓRICO .....		48
3.1	Recuperación de Información.....	49
3.1.1	Recuperación de Información Geográfica (GIR) .....	50
3.1.2	Recuperación de información geográfica por interpretación de datos .....	56
3.2	Información Geográfica y Espacial .....	57
3.2.1	Datos Geográficos .....	58
3.2.2	Datos Espaciales .....	61
3.3	Fuentes de Datos Geográficos .....	61
3.4	Fuentes de Datos Vectoriales .....	64
3.5	Ontologías como Fuente de datos.....	66
3.5.1	Ontologías en función de la conceptualización .....	67
3.5.2	Ontologías en función de axiomas y lógica de primer orden .....	68
3.5.3	Clasificación de ontologías.....	69
3.6	Consultas Geográficas y Espaciales .....	70
3.6.1	Tipos de Consultas Espaciales.....	72
3.6.2	Exploración o navegación espacial.....	78
3.7	Indexado geográfico y espacial .....	80
3.8	Análisis Espacial .....	83
3.8.1	Relaciones Espaciales.....	83
3.8.2	Semántica Espacial .....	87
3.8.3	La Web Semántica Geoespacial .....	89
3.9	Redes semánticas .....	90
3.10	Jerarquías y Teoría de Confusión .....	96
3.10.1	Jerarquía simple, ordenada, porcentual y mixta .....	98
3.11	Teoría de Confusión .....	99
3.11.1	Confusión de usar r en vez de s para jerarquías simples .....	99
3.11.2	Confusión de usar r en vez de s, para jerarquías ordenadas .....	100
3.11.3	Confusión de usar r en vez de s, para jerarquías porcentuales .....	100
3.11.4	Confusión de usar r en vez de s, para jerarquías mixtas.....	101
3.11.5	El conjunto de valores que son iguales a otros, dada una confusión.....	101
3.11.6	Objetos idénticos, muy similares, algo similares .....	102
3.12	Comentarios Finales .....	103
CAPÍTULO 4 METODOLOGÍA DE RECUPERACIÓN.....		106
4.1	Primera fuente de datos geográfica: GeoOntologías .....	108
4.2	Segunda fuente de datos geográfica: Diccionarios geográficos .....	117
4.3	Tercera fuente de datos geográfica: Archivos topológicos .....	120
4.4	Estrategia de Recuperación de Información Geográfica ( <i>iGIR</i> ).....	123
4.4.1	Estructura de Consultas. ....	125
4.5	Algoritmo de exploración de GeoOntología (OntoExplora) .....	128
4.5.1	Coincidencia completa .....	129
4.5.2	Coincidencia parcial .....	129
4.5.3	Coincidencia atómica .....	130
4.5.4	Coincidencia nula .....	130
4.6	Recuperación de información geográfica en las GeoOntologías.....	132

4.7	Recuperación de información geográfica en diccionarios geográficos .....	135
4.8	Recuperación de información geográfica en archivos topológicos .....	138
4.9	Integración y despliegue de resultados .....	140
4.10	Interfaz para recuperar información geográfica .....	142
CAPÍTULO 5 METODOLOGÍA DE PONDERACIÓN.....		152
5.1	Ponderación Clásica y Ponderación Geográfica.....	153
5.2	<i>iRank</i> : Ponderación integrando topología, geografía y semántica espacial..	156
5.3	Ponderación Conceptual .....	159
5.4	Ponderación Geográfica .....	164
5.5	Ponderación Topológica.....	167
5.5.1	Inclusión (Inclusion).....	167
5.5.2	Hermanos (Siblings).....	168
5.5.3	Proximidad (Proximity).....	168
5.6	Integración de la ponderación.....	173
CAPÍTULO 6 PRUEBAS Y RESULTADOS .....		177
6.1	Pruebas para la recuperación y ponderación conceptual .....	179
6.2	Pruebas con Consultas en los diccionarios de datos.....	182
6.3	Recuperación en archivos topológicos .....	184
6.4	Resultado de la recuperación en formato KML.....	187
6.5	Resultados finales ponderados ( <i>iGIR</i> + <i>iRank</i> ) .....	188
CAPÍTULO 7 CONCLUSIONES Y TRABAJO FUTURO.....		198
7.1	Conclusiones generales .....	199
7.2	Conclusiones particulares .....	200
7.3	Productos derivados de la tesis.....	203
7.4	Aportaciones.....	204
7.5	Trabajos a futuro.....	204
ANEXO A. GeoOntologías.....		217
ANEXO B. Código Fuente .....		303
B.1.	Generador de KML.....	304
B.2.	Procesador de nombres usando Geonames.....	305
B.3.	Página Índice .....	306
B.4.	Recuperación Ontológica .....	307
B.5.	Recuperación Geográfica .....	308
B.6.	Recuperación Topológica.....	309
B.7.	Página Principal.....	310
B.8.	Código Ruby para extraer contenido de Wikipedia (1).....	311
B.9.	Código Ruby para extraer contenido de Wikipedia (2).....	311
B.10.	Código Ruby para extraer contenido de Wikipedia (3).....	312
B.11.	Código Ruby para extraer contenido de Wikipedia (4).....	313

# RESUMEN

En esta investigación se muestra una metodología integral para la recuperación de información geográfica (*iGIR*) y un método para ponderar los resultados de la recuperación (*iRank*). La metodología considera consultas espaciales que contengan tres elementos, los cuales son objetos o lugares geográficos y una relación espacial entre estos objetos. El resultado de la búsqueda es el conjunto de documentos con datos geográficos, que tiene la mayor similitud, de acuerdo a las características evaluadas.

El éxito de la recuperación se basa en la integración de tres criterios de búsqueda para tres fuentes de datos geográficas heterogéneas: Ontologías Geográficas (*GeoOntologías*), diccionarios de datos y archivos vectoriales. Las cuales describen a los objetos geográficos utilizando la topología, los atributos geográficos y la semántica espacial.

Los resultados obtenidos demuestran que al combinar las diferentes fuentes de información y los criterios de búsqueda, se puede obtener información relevante que podría ser omitida al recuperarse de forma independiente en cada fuente. Sin embargo, el nivel de precisión en la recuperación se determina con el perfil del usuario, debido a que la experiencia de éste, determina el grado de formalidad en efectuar una consulta. El apropiado procesamiento de la intención de búsqueda y el perfil resulta en una mejor recuperación y ponderación.

Los documentos recuperados se ponderan con el método *iRank*, el cual integra tres métodos de ponderación geográfica. El proceso de integración se basa en el método de confusión. Nuevamente, la ponderación integral resultó mejor que la ponderación aislada y ofreció mayor certidumbre de la relevancia. Finalmente, lo novedoso de esta tesis, es el enfoque integrador que procesa la semántica espacial y la intención del usuario basada en su perfil.

# ABSTRACT

This research presents an integral methodology for geographic information retrieval (*iGIR*) and a method to rank the retrieved results (*iRank*). The approach processes spatial queries that contain geographical objects and spatial relationships between these objects. The search results are a set of the highest similarity documents, which contain geographic data, according to the query characteristics.

The retrieval is based on the integration of three search criteria, according to three different heterogeneous sources of geographic data: Geographical Ontologies (*GeoOntologies*), geographic dictionaries and vector files. They represent the geographical objects, taking into account the essence and nature of spatial data. In particular, we use the topology, the geographic properties, and the spatial semantics.

The results show that integrating different information sources and search criteria, it is possible to obtain relevant geographic information, which is usually omitted when this is retrieved independently on each source. The level of accuracy in the recovery is determined by the user's profile, because the user's experience determines the degree of formality in a query expression. The adequate processing of query according to the user's intention and his/her profile results in a better retrieval and ranking.

The retrieved documents are ranked using *iRank*, which integrates three geographical ranking methods. The integration process is based on the previously developed theory of confusion. The results of integral ranking are better than those obtained separately. In addition, *iRank* offers higher precision of relevance. Finally, the novelty of this thesis is the approach that processes the spatial semantics and the intention of search based on the user's profile.

# CAPÍTULO 1

## INTRODUCCIÓN

*Si una imagen dice más que mil palabras, entonces un MAPA dice más que mil imágenes*

Actualmente, para recuperar información geográfica se requieren considerar diferentes aspectos, tales como el procesamiento de la consulta (semántico ó sintáctico), la organización (estructura, indexación) de las fuentes de datos, y los formatos de representación (vectorial, raster). Además, de los procesos de clasificación y ponderación de resultados. Sin soslayar, los problemas relacionados con el procesamiento de la información imprecisa<sup>1</sup>, las fuentes de datos no estructuradas, y la intención de búsqueda de cualquier usuario (experto o neófito). Este escenario es típico en México, donde aún no se cuenta con un marco de trabajo que permita integrar de forma transparente estos datos, básicamente porque las instituciones, dependencias gubernamentales, y el mundo empresarial aún carece de la tecnología necesaria para lograrlo (e.g. *Semantic Web Services* o *Geo-Web Services*).

---

<sup>1</sup> El término información imprecisa, se refiere a dos aspectos: el primero, cuando una consulta no refleja de forma clara la necesidad del usuario (ausencia de parámetros suficientes). El segundo ocurre cuando en una fuente de datos, se describe de manera muy pobre a un objeto geográfico.

Por otra parte, las búsquedas actuales requieren una selección o coincidencia exacta. Es decir, si busco “carretera” debo encontrar en las fuentes de datos la palabra o concepto “carretera”. Sin embargo cuando no existe una coincidencia exacta, el resultado son respuestas vacías. Mientras que al recuperar se ofrecen aproximaciones a las respuestas,

Por ejemplo, si se busca carretera y no existen instancias o información que coincida con el criterio de búsqueda, entonces se reciben respuestas vacías. En este caso, es más útil, recibir respuestas aproximadas. Estas aproximaciones deben ser de acuerdo a aspectos geográficos, así como relaciones espaciales y semánticas.

Por ejemplo, para el objeto “carretera federal” se tienen las siguientes aproximaciones: “vías de comunicación”, “calle 67 *conecta* calle norte”, “avenida López Mateos *en* Guadalajara”, “camino”, etc. Las cuales se obtienen por medio de relaciones semánticas o topológicas que están asociadas al objeto geográfico.

Además, al realizar una búsqueda generalmente se efectúa sobre una fuente de datos ya estructurada (indexada) la cual fue diseñada para agilizar el proceso de búsqueda. Mientras que un sistema de recuperación generalmente no tiene organizada (indexada) la información para agilizar dicho proceso. Esto debido a que recuperar implica utilizar algoritmos que midan relevancia y establezcan grados de aproximación. El mejor ejemplo, lo tenemos con *Google*<sup>2</sup> que es conocido como un buscador y no estrictamente como un recuperador de información (la búsqueda es parte del proceso de recuperación).

Actualmente, la línea de investigación de recuperación de información geográfica enfrenta diversos retos dentro de los cuales destacamos los siguientes:

- Extensión del enfoque sintáctico al semántico y espacial.
- Heterogeneidad en las fuentes de información.

---

<sup>2</sup> El buscador *Google* [31], [www.google.com](http://www.google.com) y [www.google.com.mx](http://www.google.com.mx)

- Intención del usuario expresada en las consultas.
- La extracción de términos geográficos desde fuentes estructuradas, y de reto aún mayor, desde fuentes no estructuradas de acuerdo a criterios espaciales.
- La identificación y eliminación de ambigüedades en el ámbito geográfico para los procedimientos de extracción.
- Tipos de consultas, de acuerdo a los siguientes criterios:
  - Tipo de parámetros de entrada: espaciales, geográficas, topológicas, etc.
  - Cardinalidad: número de elementos u objetos que conforman la consulta.
  - Estructura y formato: texto libre, operadores espaciales o geográficos, etc.
- Metodologías para el almacenamiento eficiente de ubicaciones y sus relaciones.
- Técnicas de ponderación.

Adicionalmente, mencionamos algunos aspectos que son abordados en esta tesis y que se resumen en:

- La integración de fuentes heterogéneas de información geográfica (con semántica espacial implícita y explícita) a partir de una consulta geográfica. Este enfoque permite encontrar de forma más precisa, las respuestas más relevantes para una consulta. Además, la ponderación de resultados basándose en la integración de criterios espaciales. Finalmente, la consideración de algunas técnicas que permiten al usuario interactuar y explorar los resultados de consultas para sistemas GIR.

Por ejemplo, hoy en día, se realizan consultas con criterios geográficos, al buscar lugares turísticos, localizar servicios, ubicar puntos de interés, establecer cercanía y afectación a causa de ciertos fenómenos, entre muchos otros. Sin embargo, estas tareas se realizan de forma aislada, a través de diferentes sistemas y con un considerable esfuerzo por parte del usuario.

Por otra parte, hasta la fecha se cuenta con una inmensa cantidad de información relativa a un término geográfico o espacial<sup>3</sup>. Estos datos están dispersos y disponibles en numerosos formatos y representaciones. Por lo tanto, para consultarlos, obtenerlos, o visualizarlos se requieren de fuentes de información tales como: bases de datos, archivos planos, servicios Web, entre otros, así como de algoritmos de inteligencia y de procesamiento semántico. Entonces, para expresar una consulta, se requiere un conocimiento a priori de lo que se busca (intención del usuario). Lo cual permite establecer los parámetros de búsqueda adecuados y recibir, en consecuencia, respuestas satisfactorias.

Entonces, contar con un sistema que permita realizar alguna de estas tareas de forma integral, resulta útil para cualquier usuario. Por lo tanto, procesar las búsquedas orientadas a localización, ubicación o con algún criterio geográfico-espacial (típicamente expresados en lenguaje natural) son, hoy en día, una tarea que demanda atención en el ámbito de GIR. Además, en este sentido, resolver satisfactoriamente una consulta geográfica es aún costoso en términos de tiempo y procesamiento.

Por ejemplo, para satisfacer la consulta geográfica:  $Q_{GI} = \{\text{Ríos en Jalisco}\}$  se requieren procesar numerosas capas de datos espaciales, así como documentos relativos a los objetos geográficos Ríos, Jalisco y la relación espacial “en”. Donde el procesamiento tome en cuenta relaciones, propiedades, conceptos y representación de un objeto geográfico (extender el enfoque sintáctico, hacia el semántico y espacial). Esto es requerido porque a la fecha, la mayoría de los sistemas de búsqueda y de recuperación de información, resuelven consultas basándose en la técnica de coincidencia de palabra (*keyword-matching*).

Es decir, se basa en encontrar palabras idénticas a través del procesamiento de texto. En consecuencia, se obtienen resultados irrelevantes porque la palabra es ambigua. Un ejemplo

---

<sup>3</sup> Un término geográfico es aquel que tiene asociado un sistema de coordenadas (e.g. geográficas) y que permiten medir una distancia; mientras que un término espacial se refiere a relaciones entre objetos geográficos (e.g. adyacencia, contenido) independientemente del sistema de coordenadas que maneje.

de ello lo tenemos con palabras que representan una relación de distancia, tales como “cerca”. Donde, el enfoque sintáctico no puede establecer si la palabra “cerca” se refiere a un tiempo o una distancia, o si su sentido cambia con relación al objeto espacial que lo involucra.

Para describir esto, consideremos los dos sentencias siguientes: {“*El profesor está cerca de la solución*”} comparado contra: {“*La carretera 45 está cerca del Aeropuerto*”} en ambos casos la semántica de la relación “cerca” es diferente, una habla de distancia, mientras que la otra se refiere al tiempo o habla en sentido figurado. Además de que no se abunda en detalles para poder medir “la cercanía”. Pero, en ambos casos, la semántica está definida por el contexto (palabras y relaciones vecinas a “cerca”) así como a sentidos de las palabras. Este es un claro ejemplo, donde el enfoque sintáctico resulta insuficiente para procesar consultas geográficas, y por ello requerimos integrar relaciones espaciales y semánticas.

Otro ejemplo, lo tenemos al utilizar el buscador *Google* con la consulta  $Q_{G1} = \{\text{Ríos en Jalisco}\}$  obteniéndose respuestas que pertenecen a diferentes áreas y contextos tales como: “*Ríos Martínez fue apresado*” (noticia), “*Ríos de color púrpura*” (título de una película), “*Los ríos son corrientes de*” (definición), entre muchos otros más. La Figura 1-1 muestra estos resultados usando el motor de búsqueda *Google*.

Descripción	<a href="#">México: Explora</a> <b>JALISCO RIOS Y MONTAÑAS.</b> ¿Qué sabes de los ríos y montañas de Jalisco? Er el relieve de Jalisco hay dos grandes sistemas montañosos o conjunto de sierras: .. <a href="http://www.elbalero.gob.mx/explora/html/jalisco/orografia.html">www.elbalero.gob.mx/explora/html/jalisco/orografia.html</a> - 8k - <a href="#">En caché</a> - <a href="#">Páginas similares</a>
Deportes	<a href="#">Fotografías</a> ? <b>¿Ríos?</b> La final de la categoría sub18, la disputaron Raúl <b>Ríos de Jalisco</b> y Horacio Najara de Sinaloa. Carlos Chiu, del Distrito Federal y Gabriel Calderón, ... <a href="http://www.deporte.org.mx/eventos/on2004/paginas/fotografias.asp?clv=862&amp;l=g&amp;nm=2004+Olimpiada+Nacional+...">www.deporte.org.mx/eventos/on2004/paginas/fotografias.asp?clv=862&amp;l=g&amp;nm=2004+Olimpiada+Nacional+...</a> - 14k - <a href="#">En caché</a> - <a href="#">Páginas similares</a>
Definición	<a href="#">Jalisco - Wikipedia, la enciclopedia libre</a> El relieve de Jalisco favorece la fromación de ríos y lagos. Sus principales ríos son el río Lerma, el río de San Juan de los Lagos, el río San Miguel, ... <a href="http://es.wikipedia.org/wiki/Jalisco">es.wikipedia.org/wiki/Jalisco</a> - 52k - <a href="#">En caché</a> - <a href="#">Páginas similares</a>
¿Ríos o lagos?	<a href="#">Lago de Chapala</a> A partir de la fisiografía del estado de Jalisco, se supone la posibilidad de ... Todos estos ríos fluyen a través de capas gruesas de material volcánico de ... <a href="http://www.semarnat.gob.mx/regiones/chapala/fisiografia.shtml">www.semarnat.gob.mx/regiones/chapala/fisiografia.shtml</a> - 20k - <a href="#">En caché</a> - <a href="#">Páginas similares</a>
Apellido (noticia)	<a href="#">Alfredo Ríos, "El Conde"</a> ALFREDO RÍOS, "El Conde". Guadalajara (México), el 23 de diciembre de 1974. Debut como novillero: Valle de Guadalupe, Jalisco, el 12 de enero de 1991. ... <a href="http://www.portaltaurino.com/matadores/el_conde.htm">www.portaltaurino.com/matadores/el_conde.htm</a> - 16k - <a href="#">En caché</a> - <a href="#">Páginas similares</a>
¿Coincidencia exacta?	

Figura 1-1 Resultados para  $Q_{G1} = \{Ríos \text{ en Jalisco}\}$  usando Google.

Como se muestra en la Figura 1-1. Se resalta el problema de la ambigüedad de palabra, donde Río es tratado como palabra y no como un objeto geográfico cuya representación geográfica es lineal y es descrito por ciertas propiedades y relaciones.

Otro aspecto a resaltar radica en el usuario, quien debe invertir mucho tiempo en la exploración y navegación de resultados. En este caso, el éxito de la tarea ya no depende del sistema, sino de la habilidad del usuario para encontrar el resultado esperado. ¿Cómo saber si es mejor respuesta el “Río Papaloapan” al “Río Pánuco” ó “Laguna de Catemaco”?

Basándonos en los inconvenientes anteriores, una alternativa de solución la encontramos al hacer procesamiento semántico (conceptos y relaciones) lo cual permite una recuperación más precisa, ya que se evita la ambigüedad (una palabra puede ser ambigua mientras que un concepto no lo es). Por lo tanto, se requiere utilizar una estructura de datos que maneje

conceptos. Para que a través de ellos se guíe hacia una mejor recuperación. Ese tipo de estructuras son las ontologías.

Estas estructuras, permiten definir cuales conceptos describen a un término espacial. Facilitando la tarea de recuperación, sin soslayar el objetivo primordial de IR y GIR que es: Encontrar, *una aproximación al mejor resultado*, o *el mejor resultado para una consulta*. En este sentido, se requiere una aproximación a un resultado cuando no existen instancias que correspondan con el criterio de consulta. Para evitar recibir respuestas vacías (del tipo “*no match found*”).

Bajo este enfoque, las respuestas aproximadas tienen un grado de relevancia. Por ejemplo, considerando la consulta  $Q_{G2} = \{\text{Sitios arqueológicos en Área Metropolitana}\}$  un resultado relevante es “*Teotihuacan en Estado de México*” mientras que un resultado con menor relevancia es “*Museos en el DF*”. Es decir, que mediante las relaciones espaciales y semánticas podemos establecer un grado de relevancia. En el caso anterior, los museos aparecen en las respuestas, porque están relacionados semánticamente con sitios arqueológicos (relación semántica de *hiperonimia*<sup>4</sup>) y además están relacionados espacialmente (relación topológica “*contiene*”). Por lo tanto, se requiere un enfoque de GIR el cual integre relaciones semánticas y espaciales para resolver problemas donde el enfoque sintáctico ha sido insuficiente, así como un algoritmo de ponderación que se base en los aspectos anteriores, y donde el usuario pueda explorar los resultados de acuerdo a lo que espera.

## 1.1 Objetivo general

En la presente sección se presentan los objetivos y metas que conforman esta tesis doctoral.

---

<sup>4</sup> Una relación de hiperonimia es aquella que representa un tipo\_de. Por ejemplo Árbol y Pino están unidos bajo la relación de hiperonimia, ya que un Pino es un tipo de Árbol.

Recuperar información geográfica desde repositorios<sup>5</sup> geográficos heterogéneos conducida por ontologías, integrando las relaciones semánticas y espaciales asociadas a objetos geográficos.

Teniendo como base un conjunto específico de conceptos y términos geográficos utilizados en consultas espaciales y geográficas. En particular aquellas basadas en relaciones geográficas y espaciales.

- Las relaciones geográficas son datos asociados a un sistema de coordenadas, los cuales permiten medir una distancia y establecer una dirección
- Las relaciones espaciales incluyen datos obtenidos a través de relaciones topológicas (e.g. “contenido en”, “adyacencia”).

Adicionalmente, se ponderarán los resultados con base en la teoría de confusión, y de acuerdo a criterios geográficos.

## **1.2 Objetivos particulares**

- Identificar y analizar las relaciones semánticas y espaciales que existen entre objetos geográficos.
- Identificar y analizar las propiedades que definen a un objeto geográfico.
- Analizar las principales, técnicas y mecanismos de búsqueda-recuperación para datos en general.
- Explorar las propiedades y relaciones presentes en ontologías, con el fin de procesar y recuperar el conocimiento representado en la misma.

---

<sup>5</sup> Los repositorios son colecciones de datos. Cada dato o elemento en un repositorio es generalmente asociado con un nombre. Entonces, se puede encontrar un elemento por medio de su nombre. De forma adicional, pueden mantener índices para agilizar el proceso de búsqueda [73].

- Establecer el mapeo (representación semántica) de los términos espaciales y geográficos para una tarea u operación de análisis espacial (e.g. el mapeo para el término *cerca* correspondería a una tarea de análisis de *buffer* u operación *surround*).
- Construcción de una *GeoOntología* que describa las principales relaciones y propiedades de un objeto *geo-espacial*.
- Definir un algoritmo que permita explorar la Ontología generada, para extraer la representación conceptual de un objeto geográfico.
- Definir un *mecanismo basado en la teoría de confusión* que permita ponderar los resultados con base en las propiedades, relaciones semánticas y espaciales de objetos geográficos.

### 1.3 Metas

- Establecer una metodología para realizar recuperación de información a través de la integración de fuentes de datos heterogéneas (archivos vectoriales y diccionarios de datos) y conducida por ontologías.
- Analizar consultas espaciales que consideren la intención del usuario.
- Establecer una técnica para recuperar información geográfica a través de la integración de relaciones semánticas y espaciales de objetos geográficos.
- Definir la representación en alguna estructura o formato de datos para las propiedades y relaciones de los objetos geográficos.
- Establecer un mecanismo para la construcción y modificación de ontologías, utilizando relaciones que describan objetos espaciales.
- Establecer un algoritmo de exploración de ontologías el cual permita recuperar propiedades y relaciones de objeto geográficos.
- Establecer e implantar el mecanismo de ponderación de resultados de acuerdo a la teoría de confusión.
- Evaluar los resultados obtenidos con la estrategia de recuperación propuesta, bajo un parámetro que permita establecer la relevancia del resultado.

- Mostrar de acuerdo al paso anterior, las mejoras y conclusiones relacionadas con el proceso de recuperación propuesto.
- Realizar pruebas de desempeño y funcionalidad de la estrategia de recuperación propuesta.

#### **1.4 Hipótesis**

La presente tesis tiene como fundamento las siguientes hipótesis:

- La recuperación de información geográfica puede mejorarse al integrar las relaciones geográficas y topológicas asociadas a los objetos espaciales.
- Estas relaciones se encuentran implícitas en diversas fuentes de datos heterogéneas por lo que su extracción e integración generará respuestas de mayor relevancia.
- Las fuentes de información vectorial contienen relaciones topológicas que al ser consideradas en el proceso de recuperación, permitirán incluir objetos que con otros enfoques son ignorados u omitidos.
- Otras fuentes de información, como los diccionarios de datos vectoriales definen el dominio, propiedades y relaciones de un objeto geográfico. Estas relaciones permiten establecer criterios de búsqueda más precisos (saber que un río puede ser perenne, intermitente o seco; o que un “Aeropuerto” se puede representar con puntos o polígonos).
- Las ontologías representan conocimiento, por lo tanto la exploración de las mismas permiten conocer el comportamiento o estados que presenta un objeto geográfico (e.g. cuando un aeropuerto tiene representación puntual, se refiere a la edificación encargada de sus funciones operativas, mientras que si tiene representación poligonal se refiere al espacio aéreo de navegación).
- La integración de fuentes de datos heterogéneas enriquece el grado de conocimiento que se tiene para un objeto geográfico (e.g. Topológicamente se puede saber si un río atraviesa un estado, geoméricamente se determina su longitud, y conceptualmente se puede definir una clasificación para dicho río o su comportamiento en cierta época del año (seco, intermitente o perenne).

- La ponderación de resultados requiere un criterio geográfico y no únicamente de ocurrencia de palabra (e.g. la *carretera Federal 117* es más relevante que la *carretera Federal 54* si se considera su ubicación, número de sitios que atraviesa, o sitios que conectan. Donde dichos sitios también tienen una relevancia geográfica).

A continuación se indican las notaciones y herramientas que se emplean en el documento de esta tesis.

## 1.5 Notaciones del documento

Antes de explicar el desarrollo de esta tesis, conviene indicar las notaciones que describen a cada uno de los elementos involucrados en el desarrollo del documento. La Tabla 1-1 muestra la descripción de las notaciones de este documento de tesis.

**Tabla 1-1 Notaciones del documento**

Elemento	Notación	Ejemplo
Consulta espacial	$Q_G$	$Q_{G1} = \{\text{Ríos en Jalisco}\}$
Documento Geográfico Recuperado	$D_G$	$D_{G1} = \langle \text{Circuito interior conectan Aeropuerto Benito Juárez} \rangle$
Elemento de una consulta espacial	$\{\}$	$\{\text{Ríos}\}$
Elemento de un documento recuperado	$\langle \rangle$	$\langle \text{Circuito Interior} \rangle$
Nombre de programa, formato o de un objeto geográfico	<i>Letra cursiva</i>	<i>Carreteras</i>

En la Tabla 1-1 se utilizando diferentes de texto. Por ejemplo se utiliza la fuente itálica para denotar el nombre de un programa, de un formato de datos, o de un objeto geográfico. Por otra parte, las herramientas y lenguajes de programación que se utilizaron para desarrollar esta tesis, se muestran en la Tabla 1-2.

**Tabla 1-2 Herramientas y lenguajes de programación**

Herramientas y Lenguajes de programación	Descripción
<b>Ruby</b>	Utilizado para explorar y extraer datos de las páginas de <i>Wikipedia</i> .
<b>Java</b>	Lenguaje de programación utilizado para implementar la metodología integral de esta tesis, denominada <i>iGIR</i> , explorar documentos XML, y tablas de archivos topológicos.
<b>Aplicación WordNet</b>	Herramienta utilizada para explorar las relaciones semánticas asociadas a conceptos de objetos geográficos
<b>XML</b>	Formato que facilita la exploración, búsqueda y recuperación desde diccionarios de de datos y <i>TopologyFiles</i> . Formato utilizado para la generación de las Ontologías geográficas ( <i>GeoOntologías</i> .)
<b>API Google Maps</b>	Herramienta que permite desplegar un objeto geográfico a través de sus coordenadas geográficas. El despliegue es vía la Web y es posible visualizar información descriptiva y otros objetos geográficos (Por ejemplo, calles) en una zona geográfica.
<b>Google Earth</b>	Herramienta de escritorio (visor geográfico) que permite desplegar un objeto geográfico a través de sus coordenadas geográficas. El despliegue es en un visor que requiere una conexión a Internet. Adicionalmente se pueden realizar representación en 3D para una mejor presentación de la información geográfica.
<b>ArcView 3.2</b>	GIS para realizar tareas de análisis espacial en diferentes formatos de datos geográficos. En particular, en esta tesis se trabaja con el formato de de datos geográficos denominado <i>shapefiles</i> .

Cada una de estas herramientas como se observa, fue empleada con diversos propósitos, por ello se hace una breve descripción de las mismas para ubicarlas en el presente documento de tesis.



## CAPÍTULO 2

# ESTADO DEL ARTE

*Una palabra mal colocada estropea el más bello pensamiento*  
*Voltaire*

En esta sección se presenta un panorama global del estado del arte en el área de la recuperación de información (IR). De manera particular, se proporciona una visión general del área de recuperación de información geográfica (GIR) la cual por ser multidisciplinaria y de reciente nacimiento involucra diferentes líneas de investigación y la convierte en tema de frontera en el ámbito de la investigación en GIR.

Adicionalmente, este apartado describe los trabajos más relevantes que se han desarrollado de forma interdisciplinaria para recuperar información geográfica. Se comentan y discuten los enfoques, técnicas, procedimientos y sistemas que se han propuesto e implantado. Así como también se señalan las debilidades, aportaciones y áreas de oportunidad actuales en GIR.

Esta sección se divide como sigue, en la sección 2.1 se describe el enfoque general de este trabajo y su relación con IR y en específico para GIR. Mientras que, desde la sección 2.2.1 y hasta la sección 2.3.3, se comentan los diversos trabajos relacionados con la recuperación de información y con GIR. Finalmente la sección 2.9 ofrece las conclusiones obtenidas, las cuales sirvieron de referencia para definir y delimitar los alcances de esta tesis.

## **2.1 Enfoque General**

Teniendo como base el análisis realizado sobre los artículos publicados en las áreas de recuperación de información clásica<sup>6</sup> (IR) y de recuperación de información geográfica (GIR). Se realizarán comentarios acerca de los artículos y trabajos; destacando sus alcances, avances, enfoques, ventajas y desventajas, así como sus aportaciones científicas. En la mayoría de los artículos revisados predominan los enfoques que trabajan con sentidos de las palabras, así como con relaciones semánticas de palabras para datos geográficos en IR y GIR. También, están aquellos cuya meta consiste en resolver la interoperabilidad y la integración de datos desde fuentes heterogéneas.

Otras investigaciones trabajan para mejorar las capacidades (desempeño, relevancia, ponderación) de los sistemas IR. Muchos de estos trabajos se apoyan en estructuras tales como corpus, jerarquías, taxonomías, ontologías y combinaciones entre éstos para realizar IR y GIR. En esta misma línea, destacan los enfoques basados en aspectos cuantitativos y heurísticas.

La mayoría de estas investigaciones se basan en el modelo vectorial, procesos heurísticos y de inteligencia artificial. Aunque, también existen trabajos que usan ontologías como herramientas auxiliares en alguna fase de la recuperación de la información (por ejemplo: en el indexado).

---

<sup>6</sup> Utilizaremos el término Información clásica para referirnos a la recuperación de información tradicional, es decir aquella que realiza la recuperación basándose en la naturaleza del texto y la lingüística.

Por otra parte, cabe señalar que los trabajos que manejan ontologías, procesan el conocimiento de éstas, a partir de la implementación y diseño de las mismas. Por ejemplo *WordNet*<sup>7</sup> es una Ontología cuya implementación está basada en relaciones semánticas de acuerdo a sentidos de palabras. Entonces, el conocimiento se procesa a través de dichas relaciones. También, existen ontologías particulares y generales que son implementadas usando lenguajes como *OWL*<sup>8</sup>. No obstante, las ontologías se pueden almacenar utilizando estructuras tales como: árboles, redes semánticas, grafos, entre muchas otras.

Además, existen enfoques que integran metodologías con soporte de redes semánticas, algoritmos bayesianos, redes neuronales y ontologías. Sin embargo, estos trabajos utilizan en su mayoría datos geocodificados<sup>9</sup> o con un formato definido. Además, en estos trabajos se requiere tener un estándar en el formato de datos para que sean fácilmente integrados a otros sistemas, además de disponer de una infraestructura de servicios Web, lo cual es parcialmente alcanzado en Europa o Estados Unidos de América.

Por lo tanto, un enfoque como el anterior, no es factible en países como México, donde los ***principales proveedores de información geográfica carecen de desarrollos de servicios Web***. En general, los sistemas que utilizan servicios Web para tareas de localización se han denominado aplicaciones de búsquedas locales. El mejor ejemplo de este tipo de aplicaciones son las que son ofrecidas por los tres gigantes informáticos: (*Google, Yahoo* y *Microsoft*)<sup>10</sup>.

Sin embargo, a pesar de que estos servicios han tenido gran aceptación y ofrecen desarrollo continuo, están basados en geocodificación y vinculados a los resultados previamente indexados. Es decir, se basan en un algoritmo basado en geocodificación y no mediante relaciones semánticas, por lo tanto no existe forma de explotar relaciones semánticas en estas aplicaciones.

---

<sup>7</sup> <http://wordnet.princeton.edu/>

<sup>8</sup> Ontology Web Language. <http://www.w3.org/TR/owl-features/>

<sup>9</sup> Una geocodificación consiste en asociar un código a un área geográfica o sitio en la tierra (e.g. el código postal).

<sup>10</sup> [\(http://maps.live.com/\)](http://maps.live.com/) ( <http://local.yahoo.com/> ) , (<http://local.google.com/> )

No obstante, algunos de estos servicios de localización son soportados por ontologías de dominio espacial y ontologías espaciales de otras fuentes, donde cierta relevancia es alcanzada, pero con la desventaja de que no consideran el análisis espacial que se le aplica a los datos. Además, de que no consideran la topología de los datos.

Por otro lado, están los denominados servicios RSS<sup>11</sup>, estos servicios aprovechan que esta implementación de XML<sup>12</sup> permite difundir información desde diferentes fuentes, pero su funcionamiento, también, está basado en la geocodificación o georeferenciación<sup>13</sup> de los datos. Es decir, si no se tiene el dato geocodificado, entonces este dato no se puede utilizar para tareas de localización usando servicios RSS. En este sentido, la proliferación de estos servicios son el principio del desarrollo de lo que se denominó como la Web Semántica (*Semantic Web*) [23] la cual ha mostrado avances considerables, sin embargo aunque se va consolidando, aún es un reto y no una realidad.

La Web semántica, también ha sido considerada como una de las soluciones para ciertos problemas en IR, esto debido a que la Web Semántica está orientada hacia la explotación semántica de los datos: en donde se realizan consultas y búsquedas a través del significado de los datos, los cuales aparecen en las consultas (*queries*) y los documentos Web. Cabe destacar, en este punto que bajo este esquema se han realizado muy pocos trabajos enfocados a la **semántica de las tareas y procesos en los cuales los datos participan**. Es decir, **extendiendo el enfoque sintáctico al enfoque semántico**, lo cual permite emitir resultados de acuerdo al significado y al contexto de la consulta.

Hoy en día, algunos de los sistemas Web tradicionales ya ofrecen algunas características o servicios basados en la semántica, pero únicamente para datos convencionales (no espaciales). Es decir, que dichos servicios no son factibles de implantarse en un dominio geo-espacial, ya que la naturaleza de los datos espaciales, implica considerar aspectos que en los datos convencionales no son requeridos (e.g. representación, sistema de coordenadas, etc.)

---

<sup>11</sup> (Really Simple Syndication) Es un formato de datos que es utilizado para redifundir contenidos a suscriptores de un sitio Web.

<sup>12</sup> *eXtensible Markup Language*, <http://www.w3.org/>

<sup>13</sup> Un dato georeferenciado es aquél que está asociado con una sistema de coordenadas o proyección en un mapa.

En el caso de GIR y GIS en la Web, también la *GeoSpatial Semantic Web* acuñada así por M. Egenhofer, ha mostrado algunas aportaciones en investigación, pero aún está en proceso de construcción y hoy en día, es considerado uno de los retos más grandes en GIS y GIR, y donde por supuesto se encuentran las áreas de oportunidad para realizar investigación que puede tener alto impacto. Esta tesis trabaja en esta dirección, donde los resultados obtenidos permitirán contribuir en la consolidación de esta nueva línea de investigación.

En otro orden de ideas, en IR y GIR también se trabaja en las fuentes de información (bases de datos, documentos Web y de texto plano, archivos vectoriales). En aspectos tales como la explotación de sus estructuras, propiedades y formatos. Por ejemplo, en el caso de la Web, recuperar información implica considerar la estructura del documento, de acuerdo a formatos tales como: HTML<sup>14</sup>, XML, y GML<sup>15</sup> mientras que en el caso de una base de datos se consideran los modelos de datos {*Relacional, Orientado a Objetos, Abstract Data Types (ADT's)*}.

El escenario se complica si se consideran repositorios no estructurados, donde recuperar información, requiere que previamente se hayan organizado, ordenado o indexado los datos. Además de que, los resultados que ofrecen los sistemas actuales (por ejemplo: los buscadores) se presentan con base en la naturaleza textual de la consulta (palabras clave). Lo cual es insuficiente para el dominio geográfico. Lo anterior lo ilustraremos con la siguiente consulta  $Q_{G3} = \{\text{Hoteles cerca Aeropuerto Benito Juárez}\}$  donde la relación “*cerca*” implica una distancia o tiempo. Por lo tanto “*cerca*” debe ser procesada semánticamente por relaciones espaciales, ya que los métodos tradicionales (e.g. sinonimias, clasificaciones, tesauros, etc.) no permiten ofrecer un resultado de acuerdo al espacio geográfico.

---

<sup>14</sup> Hypertext Markup Language, <http://www.w3.org/MarkUp/>

<sup>15</sup> Geography Markup Language, <http://www.opengis.net/gml/>

Por ejemplo, “cerca” puede aparecer en un fragmento como éste: {“él está cerca de la solución del problema matemático”}, el cual resulta irrelevante en el dominio geográfico. Sin embargo, también “cerca” puede describirse en fragmentos como: “Río X está cerca de una población rural”, “El huracán se aproxima a la población rural a una gran velocidad”, “El fenómeno meteorológico rodeará la península de Yucatán” donde cada uno de estos documentos hablan de la relación de “cerca”, pero la palabra no aparece de forma explícita. En este caso, los métodos (sintácticos) no podrían discernir entre la relevancia de un documento que utiliza el término “cerca” en sentido figurado y otro que lo utiliza de acuerdo al dominio geográfico (Esta tarea requeriría procesamiento de lenguaje natural).

Es por ello que para ofrecer resultados de acuerdo al contexto y dominio geográfico (e.g. Hidrología, Topografía) se requiere considerar otros aspectos como son: las primitivas de representación: puntos (*P*), líneas (*L*) y polígonos (*Pol*) ya que la forma de procesarlas puede filtrar los documentos irrelevantes para una consulta geográfica. Puesto que es diferente medir la “cercanía” entre objetos puntos, que entre objetos líneas o polígonos (se requieren tareas adicionales de análisis espacial).

Por ejemplo, para conocer la “cercanía” que existe entre un Hotel (*P*) y un Aeropuerto (*P*) se requieren conocer las vías de comunicación (*L*) entre estos. Por lo tanto, se requiere realizar una operación espacial de sobreposición (*Overlay*) entre Hoteles, Ríos y Aeropuertos. Mientras que si lo que se desea saber es el grado de afectación ante el desborde de un Río (*L*) sólo se requiere una operación espacial de buffer<sup>16</sup>. Los procesos anteriores se pueden complicar si, además, se consideran las propiedades geográficas (por ejemplo: los sistemas de coordenadas) donde se requerirían conversiones adicionales, para trabajar con los datos, y posteriormente poder procesarlos bajo un enfoque semántico. Sin embargo, en la actualidad los sistemas de recuperación de información, buscadores y otros sistemas Web no ofrecen soporte para procesar e interpretar búsquedas geográficas basadas en aspectos semánticos.

---

<sup>16</sup> Análisis de proximidad o *buffer*: permite medir cercanía, influencia o afectación por un fenómeno climatológico

Esto se debe en gran medida a que los formatos o estructura de los documentos no consideran el comportamiento o el significado de los datos. Además, las representaciones de datos geográficos son en su mayoría de carácter cuantitativo. Lo cual representa un reto importante ya que en las consultas *geo-espaciales* son utilizados términos cualitativos y no cuantitativos (e.g. preposiciones).

A continuación se presenta una breve descripción de los trabajos más sobresalientes en cada una de las categorías mencionadas.

## **2.2 Recuperación de información clásica**

En esta sección se comentan cuatro trabajos referidos a la recuperación de información clásica o tradicional. Es decir, La recuperación basada en procesamiento de palabra y aspectos lingüísticos. Estos artículos explican los procesos involucrados para realizar recuperación de información clásica (IR) y recuperación de información geográfica (GIR). Lo cual permite establecer las diferencias existentes en IR y en GIR.

### **2.2.1 Baeza Yates y Ribeiro-Neto (1999)**

En este trabajo, Baeza y Ribeiro-Neto [1] hacen un resumen de lo más relevante que se ha desarrollado en el campo de IR. Se describen técnicas, mecanismos, metodologías y herramientas para realizar una mejor recuperación de la información. También, mencionan ventajas y desventajas de los principales enfoques empleados en IR: el modelo vectorial y el probabilista. Se puntualizan los problemas y tendencias que existen en cada uno de los enfoques. Se destacan las aplicaciones basadas en los enfoques descritos.

En relación a este trabajo se pueden tomar las bases y las diferentes aportaciones de IR, para adaptarlas o extenderlas para definir nuevas técnicas de acuerdo a la naturaleza de los datos espaciales y geográficos. Considerando los componentes que conforman un sistema IR.

### 2.2.2 Kobayashi y Takeda (2000)

En este artículo Kobayashi y Takeda [4] realizan un estudio de las tecnologías y sistemas que actualmente son útiles para la búsqueda y recuperación de la información en la Web. También analizan el desempeño de los mismos de acuerdo a la perspectiva del usuario y al uso de recursos.

Se reporta que el motor de búsqueda es uno de los componentes de mayor investigación para IR en la Web. En donde la mayoría de los algoritmos de recuperación son sintácticos. La investigación está dirigida principalmente a: mejorar la calidad en los resultados (reducir ruido y vínculos rotos) y aumentar la velocidad de recuperación. Considerando, consultas simples, compuestas, híbridas y personalizadas. También detallan como medir el desempeño de un sistema IR utilizando estadísticas y los tres parámetros tradicionales: velocidad, *precision*, y *recall*. Estos últimos se calculan con las siguientes formulas:

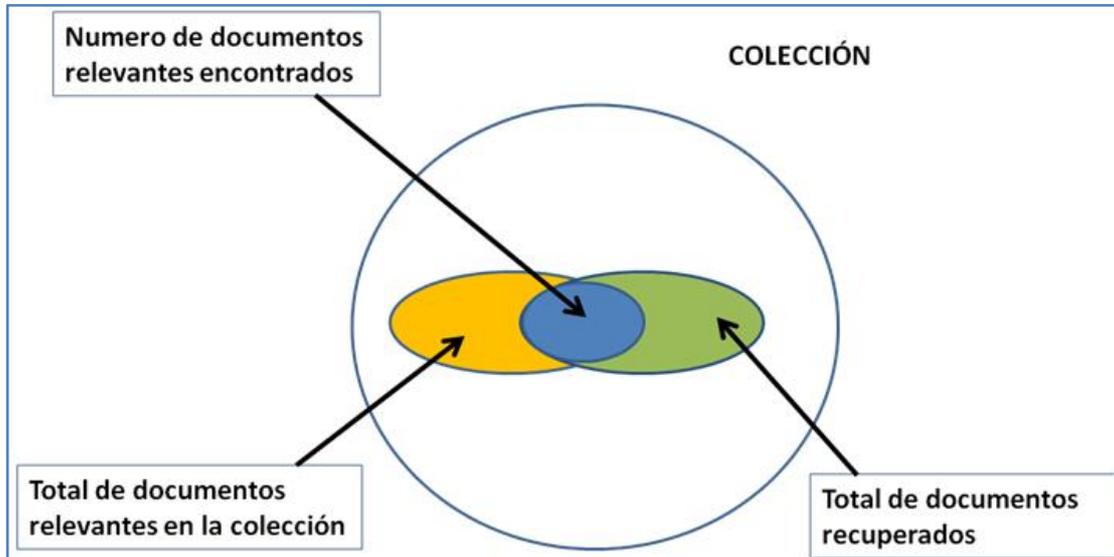
**Precision:** La proporción de documentos relevantes que fueron recuperados de todos los documentos recuperados. La cual se calcula con la siguiente fórmula:

$$Precision = \frac{|{\{documentos\ relevantes \cap documentos\ recuperados\}}|}{|{\{documentos\ recuperados\}}|}$$

**Recall:** La proporción de documentos relevantes que son recuperados, sin considerar todos los documentos relevantes disponibles. La cual se calcula con la siguiente fórmula:

$$Recall = \frac{|{\{documentos\ relevantes \cap documentos\ recuperados\}}|}{|{\{documentos\ relevantes\}}|}$$

La Figura 2-1 muestra de forma gráfica los aspectos involucrados en estas fórmulas.



**Figura 2-1** Representación gráfica de Recall y Precision.

Se hace hincapié en que algunos enfoques están dirigidos a obtener *precision* en los resultados, mientras que otros se centran en conseguir páginas *Hub* (aquellas que contienen vínculos a páginas relevantes respecto a la información buscada).

Se indican dos de las mayores diferencias entre la IR e IR en la Web: 1) concurrencia en la búsqueda. 2) El número de documentos que pueden ser accedidos y ponderados. Por otra parte, la recuperación en la Web, no está organizada, indexada ni estructurada. Además, una tarea que determine cuales páginas son relevantes para ser indexadas, o cuales son candidatas para asignarles un peso (agruparlas) son procesos de alta exigencia y costo computacional, por lo cual encontrar la solución que represente el menor costo aun no es alcanzada.

### 2.2.3 Braga et al (2001)

En este trabajo Braga *et al* [6] se enfocan en mejorar el proceso de recuperar información específica. En este caso de información relacionada con componentes de software. Para ello realizan un análisis de la búsqueda actual para encontrar componentes de software.

Encontrando que las interrelaciones existentes entre componentes de software pueden ser explotadas para mejorar el proceso de recuperación. También, consideran que los componentes están atados a un dominio de aplicación específico para que puedan ser reutilizados. Por ello, se utiliza una Ontología de dominio para que actúe como una técnica que especifique acuerdos entre componentes de usuarios y proveedores. Utilizando las relaciones semánticas de Hiperonimia, Hipónimia, y Sinónimia.

Resumiendo, el trabajo se enfoca en el problema de interoperabilidad entre repositorios de componentes de software. Usa una arquitectura de capas, donde una de ellas está basada en intermediarios y ontologías que proporcionan el vínculo entre diferentes componentes a los conceptos del dominio. La identificación de componentes relacionados y la organización del dominio, la efectúa cada intermediario. Donde cada uno de ellos engloba una ontología de dominio y proporciona el mapeo a sus respectivos repositorios de componentes.

## **2.3 Recuperación de información clásica y geográfica usando similitud**

En esta sección se comentan los trabajos que utilizan algún mecanismo de similitud para recuperar y/o ponderar información clásica o geográfica. Estos mecanismos también pueden ser utilizados para ponderar información. En especial existen trabajos enfocados a medir similitud por nombre de palabra, por su semántica lingüística o por sus propiedades geográficas y semántica espacial.

### **2.3.1 Varelas et al (2005)**

En este trabajo Varelas *et al* [5] proponen e implementan un modelo para recuperar información de imágenes y documentos en la Web. Utilizando similitud semántica entre conceptos que no son léxico-gráficamente similares. La determinación de la similitud semántica se realiza con un mapeo entre conceptos y relaciones presentes en la ontología de *WordNet*.

Se mide similitud con un proceso que detectan similitudes semánticas entre documentos, aunque no contengan términos lexicográficos similares o idénticos. Los resultados ofrecen una mejora mostrada con gráficas de *recall* y *precision* en relación a otros sistemas de recuperación de imágenes. El modelo se conforma por un módulo de *crawler* (que almacenó 1,5 millones de páginas con imágenes), un módulo de análisis (extrayendo texto y enlaces presentes en cada documento), un módulo de almacenamiento y un módulo de consultas (por palabra clave o texto libre). Dentro de las principales contribuciones que ofrece este trabajo se encuentran: 1) Un *framework* y un sistema implementado para evaluar el desempeño de diversos métodos de similitud semántica usando *WordNet*. 2) Un modelo de recuperación de información basado en la integración de métodos de similitud semántica.

El modelo propuesto se integro con el modelo vectorial<sup>17</sup> en un sistema para recuperar páginas Web e imágenes en la Web. La relevancia de los resultados se midió comparando el modelo propuesto contra el modelo vectorial, a juicio de cinco personas (árbitros). Utilizando veinte consultas que contienen desde uno hasta cuatro términos (no es posible establecer si con un mayor número de términos los resultados serán satisfactorios). Se trabaja con la relación *Is-a* disponible en *WordNet*, y no se pueden emplear frases o términos compuestos (solo palabras-sustantivos).

### **2.3.2 L. Andrade y M. Silva (2006)**

En [76] se describe un operador de similitud geográfica, este operador calcula la relación que existe entre dos lugares geográficos, así como el método para combinar un enfoque geográfico con un enfoque basado en texto para realizar ponderación. La evaluación utiliza la colección de datos del GeoCLEF 2005. Además, definen una estrategia para combinar la ponderación tanto en texto como en geografía. Se utiliza una ontología construida

---

<sup>17</sup> El modelo vectorial se basa en definir un conjunto de palabras útiles (keywords (palabras clave), términos) para con base en este realizar recuperación de documentos. Los documentos se modelan como un vector de términos.

previamente, la cual se explora mediante reglas definidas para asignar la relevancia a los documentos utilizando pesos.

Las pruebas se realizan usando datos del GEOCLEF 2005, mostrando diversas tablas para enfoques de relevancia basados en texto contra el propuesto basado en geografía. Las conclusiones obtenidas son que la ponderación geográfica por similitud y alcance es efectiva únicamente para ciertas consultas geográficas, principalmente las de proximidad. Además el balance óptimo entre ponderación geográfica y textual depende de la consulta. La ponderación textual ofrece buenos resultados sobre la ponderación geográfica, siempre y cuando se procese sobre un número grande de términos geográficos.

### **2.3.3 Levachkine et al (2004)**

En el trabajo publicado en [19] Levachkine et al., se describen una forma de representar, analizar, procesar y medir variables cualitativas a través de jerarquías. El artículo en extenso propone este tipo de representación como una nueva estructura de datos, la cual es más simple que las ontologías. De igual forma, las jerarquías son más fáciles de entender y las extensiones a búsquedas y respuestas imperfectas son más sencillas.

Adicionalmente, se hace un énfasis primordial a variables cualitativas, las cuales toman valores simbólicos. Estos valores algunas veces pueden acomodarse en capas o niveles de detalle. Por ejemplo, tenemos para la variable *lugar\_origen*

En el nivel 1 toma los siguientes valores: *Europa, África,...*,

En nivel 2 tomas los valores: *Francés, Alemán,....*,

En el nivel 3 toma los valores: *Californiano, Texano,....*,

En esencia, aquí se considera un dato como una entidad relacional, la cual depende de un contexto<sup>18</sup> particular. Cabe señalar que muchos trabajos sobre procesamiento de datos

---

<sup>18</sup> El *contexto* depende de un ambiente particular (representación, espacio, dominio del sujeto, etc.), el cual se encuentra inmerso en el dominio de los datos. La relación entre los elementos de datos depende de éste. Por ejemplo, los colores *pálido* y *beige* podrían ser mucho más cercanos (indistinguible) en un contexto, mientras en otro serían diferentes (lejanos).

cualitativos omiten usualmente el contexto del problema; es decir, intentan buscar soluciones generales para todos los contextos de aplicación. Por otro lado, se hace uso de las jerarquías para medir la similitud y disimilitud entre valores cualitativos, intentando conservar el contexto. Para extender la noción de jerarquía, es necesario proporcionar una herramienta adecuada para análisis de datos cualitativos, procesamiento y clasificación; ya que las jerarquías encapsulan las relaciones (algunas veces ordenadas) entre particiones de un conjunto de datos y por lo tanto mantienen fácilmente el contexto del problema.

Dentro de las principales contribuciones de este trabajo, se tiene que contiene un cálculo de predicados sobre jerarquías, el cual puede ser utilizado para formalizar las consultas soportadas por ontologías. Este último punto es una aportación de esta tesis doctoral.

Se propone un método para graduar los errores en consultas, utilizando jerarquías y conjuntos ordenados. Basándose en el hecho de que frecuentemente los valores cualitativos tienen un orden o un nivel jerárquico (muy corto, corto, medio alto, alto). Por tal motivo, cuando se realizan algunas consultas sobre estas jerarquías pueden presentarse pequeños errores; por lo tanto se proponen una serie de métodos para medir el grado de confusión que puede existir cuando se recuperan datos por medio de consultas. Estos errores pueden suscitarse, cuando se realiza una consulta, ya que ésta puede presentar diversos datos en diferentes niveles de una jerarquía.

En esencia para solucionar este tipo de confusiones es necesario contar con un conjunto ordenado de los elementos, esto con el objetivo de evitar en la medida de lo posible el grado de error (*confusión*) en los datos recuperados de una consulta.

En otro enfoque, relacionado con jerarquías en [19], se muestra una forma de cómo introducir conjuntos ordenados o arreglados en jerarquías, las cuales pueden ser utilizadas para diversas tareas:

- Para comparar dos valores, tales como Madrid y Ciudad de México y para medir su confusión. Por ejemplo, para contestar a la consulta ¿Cuál es la capital de España?

- Para comparar la similitud entre dos objetos, utilizando los conceptos de identidad, muy similar, similar, etc., entre diversos objetos.
- Para encontrar la cercanía de un objeto  $o$  que se pueda adecuar a un predicado.
- Para recuperar objetos que no se adaptan a un predicado dado por un umbral o confusiones acumuladas.
- Para manejar en forma parcial el conocimiento.

En los puntos a destacar de este trabajo encontramos que expresan que las jerarquías realizan buenas aproximaciones para utilizar granularidad de valores cualitativos (en conjuntos ordenados) los cuales proporcionan resultados adecuados en la recuperación de objetos. Adicionalmente, los conjuntos ordenados añaden un refinamiento adicional a la *precision*, con lo cual la confusión puede ser medida y utilizada.

Además de que esta estructura puede ser empleada como un clasificador de patrones supervisado, utilizando las definiciones de similitud, identidad, etc. Con el objetivo de medir la cercanía entre dos objetos. Pero, como el trabajo tiene como base el uso de jerarquías, no se describen procesos o metodologías para que la propuesta se pueda aplicar a ontologías. Es por ello que en esta tesis se desarrolla en la metodología un mecanismo para poder utilizar ontologías en lugar de jerarquías.

## **2.4 Recuperación de información clásica usando semántica y contexto**

En esta sección se discuten los trabajos que emplean semántica lingüística, espacial, y el contexto para recuperar información textual o geográfica. La revisión de estos trabajos permite tener un panorama claro del papel que actualmente desempeña el procesamiento de la semántica y del contexto. Como resultado de estos trabajos se derivan nuevas líneas de investigación como la del trabajo de la *Web semántica geoespacial*. Finalmente los trabajos comentados en esta sección permiten visualizar como recuperar información relevante que con otros mecanismos es omitida.

### 2.4.1 Rocha et al (2004)

En este artículo [8] se presenta una arquitectura de búsqueda que combina técnicas de búsqueda clásicas con técnicas de “*spread activation*” aplicadas a un modelo semántico para un dominio específico.

La técnica de “*spread activation*” permite, principalmente, encontrar conceptos relacionados en la ontología, logrando esto a través de un conjunto inicial de conceptos (proporcionado) y sus correspondientes valores de activación iniciales. Estos valores iniciales son obtenidos de los resultados de búsqueda clásica aplicados a los datos, los cuales son asociados con los conceptos presentes en la ontología. Para la realización de pruebas se implementaron dos escenarios de propagación basado en enfoques simbólicos y sub-simbólicos, denominada activación híbrida. El segundo escenario se realizó con un enfoque único, que puede ser simbólico o sub-simbólico. En este último los resultados se indican como positivos. Con base en los resultados se concluyó que la activación híbrida propuesta, alcanzó mejores resultados que el resto de los enfoques.

El sistema usa en enfoque de expansión de consulta donde el primer conjunto de expansiones busca mejorar la funcionalidad existente a través del uso de diferentes pesos. Mientras que el segundo conjunto de expansiones debe agregar nuevas características tales como: incorporación en una categoría basada en un peso, mapeo de pesos de acuerdo al contexto, retroalimentación por relevancia, etc.

Finalmente, una característica interesante de la técnica de “*spread activation*” es que es posible proponer un conjunto de conceptos que se estimen estén fuertemente conectados para un concepto dado, aunque no exista una relación explícita entre los conceptos almacenados en la base de conocimiento. Pero, no se ofrecen mayores detalles acerca de cuál sería el criterio que permitiría “estimar” un fuerte acoplamiento o conexión entre ciertos conceptos.

#### **2.4.2 Reiner Kraft et al (2005)**

En este artículo, Reiner et al [60] proporcionan un panorama general del sistema *Y!Q Contextual Search*, , subrayando las técnicas utilizadas para capturar el contexto de una búsqueda (el contexto es un conjunto de palabras relacionadas con los términos que conforman la consulta) y con ello mejorar la recuperación de información .

La búsqueda contextual intenta capturar de mejor manera las necesidades de información del usuario. A través de expandir la consulta con información contextual. El artículo se enfoca en como capturar el contexto de búsqueda con alta calidad, y cómo utilizar este contexto para mejorar la relevancia de los resultados. Para el primer problema, Y!Q presenta un elemento gráfico que captura el contexto de búsqueda y proporciona acceso a su funcionalidad en el punto de interés de la consulta. Y!Q utiliza una red semántica para analizar el contexto de búsqueda, resolviendo posibles ambigüedades en los términos, y generando un compendio contextual que consiste de sus conceptos clave.

Este compendio, es enviado a un planificador de consultas y marco de reescritura para expandir la consulta del usuario con términos de contexto relevantes, y así mejorar la relevancia global de la búsqueda. Los resultados de Y!Q se comparan con los resultados del buscador Yahoo! En donde Y!Q obtiene una mayor relevancia. En particular, indican que para consultas ambiguas el contexto ayudo a proporcionar mayor relevancia y dirigir resultados de acuerdo al contexto.

#### **2.4.3 Guha et al (2003)**

En este artículo Guha et al [10] presentan una aplicación de búsqueda semántica diseñada para mejorar los resultados obtenidos por la búsqueda tradicional en la Web. El sistema considera la notación del *query* de búsqueda e incrementan los resultados de la búsqueda usando datos relevantes obtenidos de diferentes fuentes. Subrayando que la explotación de la semántica de cada término de una consulta mejorara la recuperación de información.

Se enfatiza el hecho de que los resultados relevantes aumentan para una búsqueda que considera la semántica. Cabe destacar que la dirección de este trabajo está enfocada en el componente de búsqueda basada en texto y se propone un sistema híbrido que procese palabras clave y la semántica de una notación particular para una consulta específica. Finalmente, este trabajo requiere ser adaptado para ser empleado en el dominio geográfico.

#### **2.4.4 Peter Anick (1997)**

En este trabajo Peter Anick [2] expone la sinergia entre la agrupación de documentos y el análisis de frases para construir un sistema automático de recuperación de información basado en contexto. Donde un contexto consiste en dos componentes: un conjunto de artículos relacionados lógicamente (su extensión), y un pequeño conjunto de conceptos resultante, el cual es representado por palabras y frases que son organizados por grupos de palabras clave (su intención). El sistema propuesto presenta los resultados con base en el contexto que mejor concuerde con una consulta de usuario expresada en lenguaje natural.

Adicionalmente, el usuario puede elegir un contexto y manipular el componente intencional para explorar la extensión del contexto y realizar nuevas búsquedas en la base de datos. Es decir, usa un mecanismo de retroalimentación que permite a los usuarios una forma natural de refinar consultas que son expresadas de forma imprecisa, además de ser una herramienta auxiliar para distinguir entre las tareas de buscar y explorar.

En este trabajo la recuperación de la información se ha enfocado en dos tareas: 1) Las formas en que automática o semi-automáticamente se expande la consulta para hacerla más efectiva. 2) algoritmos de ponderación, que permitan incrementar la probabilidad de que los más relevantes sean los primeros en mostrarse.

#### 2.4.5 Max Egenhofer (2002)

En este artículo [17] Max Egenhofer describe los elementos necesarios para construir lo que él denomina como la *Web semántica Geoespacial*, haciendo énfasis en aspectos para la construcción de la semántica formal, así como el desarrollo de múltiples ontologías espaciales y de términos<sup>19</sup>.

Indicando, además que se requieren mecanismos para que la semántica se represente de la tal forma, que sea procesable tanto por personas como por computadoras, así como el procesamiento de consultas espaciales apoyadas por ontologías. Además, se mencionan aspectos tales como la medición de los resultados recuperados basados en la concordancia entre la necesidad de información expresada y la semántica disponible en las fuentes de información y sistemas de búsquedas.

Básicamente, se detalla y describe un nuevo marco de trabajo para la recuperación de información geográfica basada en la semántica de ontologías espaciales y de términos. Adicionalmente, el artículo visualiza y describe la representación de la semántica en diferentes componentes del proceso de recuperación (personas, interfaces, sistemas de búsquedas, y fuentes de información). Se enfatiza en la necesidad de la creación de servicios, en la carencia del procesamiento semántico de datos geográficos, así como también en como el papel de las ontologías puede ayudar a construir el equivalente a la Web Semántica. Finalmente, se menciona que la *Web semántica geoespacial* permitirá a los usuarios recuperar de forma más precisa los datos que ellos necesitan, basándose en la semántica asociada a dichos datos. Se subrayan diversos retos para lograr esto, en esencia son áreas de oportunidad en GIR.

---

<sup>19</sup> Una ontología espacial es aquella que contiene conceptos y relaciones de un dominio geográfico (e.g. Hidrología) mientras que las ontologías basadas en términos se enfocan en la naturaleza del texto (e.g. Wordnet )

## **2.5 Recuperación de información geográfica y clásica en la web**

En esta sección se comentan los trabajos que recuperan información geográfica y clásica desde la Web. En donde se proponen diversos métodos para resolver problemas de ubicación, de indexación y de recuperación basada en aspectos geográficos.

### **2.5.1 Hawking et al (2001)**

En este trabajo Hawking et al [7] realizan una evaluación de veinte máquinas de búsqueda con el objetivo de evaluar la calidad de cada una de éstas. La evaluación se realizó con base en un conjunto de cincuenta y cuatro *queries* tomados de los archivos *log* de ciertos buscadores Web.

Además, para la evaluación se tomaron en cuenta sólo los primeros siete resultados de cada máquina de búsqueda. Particularmente, el estudio presentado y sus predecesores manejan consultas que se asume son obtenidas de la necesidad de encontrar una selección de documentos relevantes para un tema específico. Enfatizan en que la meta es encontrar nuevas técnicas para implementarse en los buscadores actuales, así como mejorar sistemas que usan metodologías bien conocidas de IR. Las medidas usadas para la evaluación de los motores de búsqueda se relacionan principalmente con el desempeño del *crawler*, y se concluye que ninguna de las máquinas de búsqueda más populares explora más del 16% de su estimado total (aproximadamente 800 millones de páginas Web indexadas) para satisfacer una consulta.

Los metabuscadores no indexan documentos por sí mismos sino que reenvían las consultas a las máquinas de búsqueda más populares y conforman una lista de resultados. Donde el criterio de comparación para las máquinas de búsqueda consideradas usa un rango de medidas obtenidas a través de juicios de relevancia binaria. Lo cual en IR es una desventaja ya que no hay forma de aproximar, es decir en un enfoque booleano, solo se puede categorizar en dos valores: bueno o malo. Básicamente lo que se hizo fue medir precisión

de una lista de resultados obtenida por once máquinas de búsqueda, donde destacan: Yahoo!, Google y Microsoft.

### **2.5.2 Sergey Brin et al (1998)**

En este trabajo Sergey Brin y Lawrence Page [9] describen a uno de los buscadores de mayor auge y popularidad en la actualidad: *Google*, básicamente es la descripción del prototipo de lo que denominaron una máquina de búsqueda de gran escala, cuya principal característica es explotar ampliamente la estructura presente en el hipertexto. Se destaca que *Google* se diseñó para explorar e indexar la Web de forma eficiente en comparación a como lo hacían los sistemas existentes de aquella época. El prototipo manejaba 24 millones de páginas Web residentes en una base de datos de hiperenlaces.

La justificación para desarrollar una máquina de búsqueda de gran escala se basó en el hecho de que las máquinas de búsqueda indexan cientos de millones de documentos Web, además de responder a decenas de millones de consultas al día. Por lo cual los requerimientos y características de un buscador debían considerar y resolver estos aspectos, destacando que la investigación relativa a resolver este tipo de problemáticas era escasa (en el año en que este trabajo se publicó). Además que se indica que este trabajo fue la primera descripción en profundidad de una máquina de búsqueda de gran escala. *Google* es una arquitectura completa para coleccionar páginas Web con la capacidad de realizar consultas sobre dichos documentos, proporcionando resultados de alta calidad para una búsqueda específica en la World Wide Web.

Otra aportación consiste en la explotación y utilización de la información adicional presente en el hipertexto para producir mejores resultados a la búsqueda. Con esta aportación se generaron, además, nuevos retos relacionados con la explotación de dicha información adicional en el hipertexto. Otro detalle, es el enfoque mostrado para tratar de forma efectiva con colecciones de hipertexto libres, como es el caso de la Web donde cualquiera puede, prácticamente, publicar lo que desea. Por otra parte, el prototipo emplea diversas técnicas para mejorar la calidad de los resultados, como son: *PageRank*, el punto de referencia en el

texto (*anchor text*), e información de proximidad. Otro aspecto de interés es la implementación de algoritmos inteligentes, cuya meta es decidir cuales páginas antiguas deben ser nuevamente recuperadas desde la Web y cuales páginas nuevas deben ser incluidas dentro del conjunto ya existente de documentos Web.

### 2.5.3 Zhou et al (2005)

En este trabajo Zhou et al [13] se enfocan en las búsquedas en la Web considerando su ubicación y no la descripción de su ubicación, es decir, el lugar donde residen o están alojadas las páginas Web. Por ejemplo: la página de la embajada mexicana, en un servidor de Canadá. El objetivo consiste en encontrar información contenida en páginas Web cuyos temas están relacionados con un lugar o región.

Justifican su investigación en el hecho de que los motores de búsqueda convencionales trabajan con indexado orientado a conjuntos, mientras que la información relacionada con una ubicación es de dos dimensiones en el espacio Euclidiano. Por lo tanto, se enfocan en la representación eficiente de los atributos de ubicación presentes en documentos Web. Además de establecer una forma de combinar ambos tipos de indexado (textual y de localización). La propuesta consiste en una estructura de indexado híbrida la cual integra archivos invertidos y árboles R<sup>\*20</sup>.

Donde un archivo invertido (índice invertido) es una estructura de índice que almacena el mapeo de palabras a sus ubicaciones en un documento o conjunto de documentos lo cual permite búsqueda de texto libre. Por ejemplo, dados los textos  $T_0 = \text{"esto es que esto es"}$ ,  $T_1 = \text{"que es esto"}$  y  $T_2 = \text{"esto es una manzana"}$ , el archivo invertido es:

---

<sup>20</sup> Los árboles-R\* o R\*-árboles son estructuras de datos de tipo árbol similares a los árboles-B, con la diferencia de que se utilizan para métodos de acceso espacial, es decir, para indexar información multidimensional. Por ejemplo, las coordenadas (x, y) de un lugar geográfico.

"Una":	{2}
"manzana":	{2}
"es":	{0, 1, 2}
"esto":	{0, 1, 2}
"que":	{0, 1}

**Figura 2-2** Estructura de un archivo invertido

Tal y como se aprecia en la Figura 2-2 se almacena el número de texto (documento) en el cual una palabra o término aparece. Entonces en este trabajo se integra archivos invertidos y árboles  $R^*$  para gestionar las consultas relacionadas con ubicación y consultas textuales. Además, se indican que se estudiaron las posibles combinaciones resultando en tres esquemas: 1) archivos invertidos e índices de árbol  $R^*$ , 2) primero archivo invertido y después árboles  $R^*$ , 3) primero árboles  $R^*$  y después archivos invertidos.

De manera adicional, se propone un esquema para validar el desempeño de la estructura de indexación propuesta, la cual consiste en un motor de búsqueda Web conformado a su vez en cuatro módulos: 1) Un extractor el cual detecta alcances geográficos de las páginas Web y representa el ámbito geográfico con base en coordenadas geográficas. 2) Un indexador el cual construye las estructuras de indexado híbridas para integrar información de ubicación y textual. 3) Un ponderador que trabaja con base en relevancia geográfica y no geográfica. 4) Una interfaz que permite visualizar la relevancia de los resultados geográficos y no geográficos.

En esencia, el trabajo se centra en la hipótesis que si se busca cierta información desde la ciudad X, es más probable que el usuario se interese solo en los sitios de la ciudad X, por lo cual se descartan las ciudades Y ó Z. La desventaja aquí, es que se requiere tener la información de la red y además procesar la información de archivos log, para determinar desde donde está accediendo a un cierto sitio un usuario.

#### **2.5.4 D. Santos y M. Chaves (2006)**

En este trabajo [2] se discuten los métodos usados en GIR para realizar geo-indexación con páginas Web, utilizando el método de *Geo-scoping (Grounding)*. En donde demuestran que el método no ofrece los mejores resultados cuando la página Web de una ciudad X describe una ciudad Z, pero la página Web no está alojada en la ciudad Z.

Por ejemplo, un sitio mexicano cuya página habla de Suiza, y por otro lado, la página de una tienda situada en Suiza y que vende zapatos en ese país. Se discute el problema de lo que significa “lugar” en lenguaje natural. Los autores investigan el rol de “lugar” en lenguaje natural integrando ocho aspectos, que han sido tratados de forma independiente en otros trabajos. Se apoyan en una máquina de búsqueda existente, una ontología de lugares, un sistema de reconocimiento de nombres de entidades (NER) y una colección de datos en portugués. Las pruebas que realizan son reportadas de forma empírica, considerando: distribución de distritos y división política de Portugal, una geo-Ontología basada en mapas, y conceptos geográficos de textos portugueses. Estos últimos son dependientes de la cultura (serían diferentes para otros países).

Los resultados indican cuantos documentos mencionan sitios geográficos por nombre, cuantos se repiten, tipo de sitios se refiere (ciudades, villas, etc.) y el porcentaje de ambigüedad para los nombres de lugares y de organizaciones. Finalmente, se concluye que los diferentes roles para nombres de lugares en lenguaje natural, son dependientes del lenguaje y de la cultura. Además de que son expresados de forma imprecisa, y son dependientes del contexto.

#### **2.5.5 Jones et al (2002)**

En este artículo Jones et al [3] reportan algunos métodos para extraer información geográfica desde páginas Web como parte del proyecto SPIRIT (*Spatially-Aware Information Retrieval on the Internet*). Este proyecto consiste en la recuperación de la

información espacial a través de ontologías geográficas, donde incluso ontologías como *WordNet* (basada en sentidos de palabras en inglés) han sido ampliamente explotadas para integrarse a este proyecto.

El objetivo del proyecto es enriquecer los sitios Web usando conceptos espaciales y construyendo conjuntos de datos espaciales que estén visibles y disponibles en Internet. Así como realizar recuperación de información considerando criterios espaciales y geográficos. El proyecto resultó innovador, ya que a través de conceptos se recupera la información. Sin embargo, en el marco general de estos conceptos sólo se consideran atributos descriptivos definidos para cada objeto geográfico tales como: nombres, direcciones, código postal, números telefónicos, etc. Y hasta el momento, no reportan atributos espaciales o la consideración de relaciones espaciales (enfocadas a topología) entre los objetos geográficos. Los esfuerzos hasta el día de hoy son enfocados a criterios geográficos y de localización.

En esencia, este trabajo describe una herramienta espacial para extraer *metadatos*, y un programa de geocodificación que asigna coordenadas espaciales a cada una de las localidades extraídas por algún método específico. Este último aspecto permite ubicarlo en la categoría de sistemas de georeferenciación. El proyecto descrito es un sistema GIR que se apoya en ontologías léxicas y jerárquicas, tesauros, ontologías espaciales y dominios geográficos. Muy al estilo de *WordNet*. En fechas recientes han publicado trabajos enfocados con el manejo de relaciones espaciales y algunos enfoques semánticos, pero el procesamiento de relaciones se lleva a cabo de forma aislada y sin considerar tareas de análisis espacial.

#### **2.5.6 T.Delboni et al (2005)**

En este documento [14], se explora el uso de las expresiones en lenguaje natural para realizar búsquedas geográficas en la Web, en particular, enfatizan en el hecho de que no se utilizan datos geocodificados. Las consultas están referidas a la localización de sitios usando términos utilizados para referirse a lugares y su ubicación sin mencionar

coordenadas, por ejemplo: la consulta “Aeropuertos cercanos a Brasil”. Estas expresiones denotan la posición de un sujeto con respecto a un punto de interés (*landmark*), por ejemplo: el edificio de gobierno es un punto de interés o referencia para ubicar otro sitio cercano a éste.

La justificación para procesar este punto de interés como el componente que permita procesar la semántica de los términos relacionados con cercanía, reside en la aseveración de que es una fuente valiosa del contexto geográfico que está incrustado en muchos documentos Web. El enfoque propuesto, tiene la finalidad de guiar hacia una técnica basada en la expansión de consultas, la cual utiliza una maquina de búsqueda que trabaja bajo un mecanismo de *keyword-matching*. Se presenta una interfaz del sistema prototipo para sitios en Brasil. En resumen, en este trabajo se explora el uso de las expresiones que de forma común una persona utiliza para ubicar un lugar, en particular usando puntos de referencia, enfatizan en el hecho de que no se utilizan datos geocodificados. Básicamente lo que se hace es asociar a términos de localización un valor numérico. Por ejemplo: si se habla de cercanía se asigna mediante un proceso, un valor a este término para poder satisfacer el significado de cercanía.

El trabajo pertenece al enfoque de expansión de consulta. Una de las desventajas reside en saber cuántos términos son necesarios en la expansión de consulta, y por otro lado, que existe dependencia de un motor de búsqueda basado en palabra clave. Así que el trabajo ofrece mejoras en la relevancia de resultados, pero estas mejoras son basadas en procesos estadísticos y aspectos cuantitativos.

### **2.5.7 Vaid et al (2005)**

El presente artículo [16] presenta tres métodos para realizar indexado tanto espacial como textual. Está orientado hacia máquinas Web de búsqueda, considerando que éstas tratan los términos geográficos en la misma forma que otros términos, lo cual genera que los resultados no sean los requeridos por el usuario. Entonces bajo esta premisa se propone asociar métodos de indexado espacial con los de indexado textual, explotando los

procedimientos de geo-tagging para categorizar documentos con respecto al espacio geográfico. Sin embargo, los esquemas son comparados experimentalmente con máquinas de búsqueda convencionales (que usan indexado textual) para mostrar su desempeño y velocidad con respecto a aquellas que utilizan únicamente indexado textual. Además, las pruebas no fueron evaluadas mediante criterios espaciales, sino que se usaron algunos buscadores que explotan el texto, por lo tanto es claro que la relevancia de los resultados mejora, pero es necesario medir esta relevancia con otros sistemas que si consideren aspectos geográficos y no solo mediante los aspectos del texto.

### **2.5.8 Jens Graupmann and Ralf Schenkel (2006)**

En este trabajo [75] describen un motor de búsqueda para consultas geográficas en la Web. El cual incluye una interfaz visual de los resultados discretos o expandidos para una consulta específica. Contrario a otros enfoques, el artículo no asigna *footprints* a los documentos, pero si considera el contexto de la información geográfica de forma general, con lo que se permite evaluar consultas de granularidad fina, esto al nivel de fragmentos de documentos. No permite hacerlo con documentos completos.

El trabajo se enfoca en consultas geográficas la cuales combinen restricciones sobre el contenido de una página Web con al menos una restricción geográfica, por ejemplo usando rangos (“*between Paris and Nancy*”) e imprecisión (“*near London*”). El motor de búsqueda presentado es una extensión a un motor de búsqueda ya existente (SphereSearch Engine, SSE) donde SSE integra técnicas de IR y de extracción de información (IE). La propuesta de solución consiste en realizar anotaciones de información geográfica dentro de una página y agregándola en el contexto de un contenido que coincidió (match) al explorarlo, donde posiblemente se incluye conocimiento externo como una jerarquía de ubicaciones.

### **2.6 Recuperación de información geográfica usando relaciones espaciales**

En esta sección se describen los trabajos relacionados con recuperar información geográfica desde una fuente de datos vectorial. Asimismo se describen enfoques que explotan las relaciones espaciales entre objetos geográficos. Cabe señalar que son muy pocos los

trabajos enfocados a procesar las relaciones espaciales. Siendo entonces esta una área de oportunidad importante en GIR.

### **2.6.1 Walker et al (2005)**

En este trabajo Walker et al [11], proponen el uso de algoritmos de aprendizaje bayesianos<sup>21</sup> como una alternativa para mejorar la recuperación de información geográfica. Cabe destacar que este artículo utiliza las relaciones espaciales como un factor de importancia para recuperar información geográfica.

La propuesta considera la tarea de análisis espacial más común en GIS (la sobreposición). La cual se realiza sobre un conjunto de temas (capas de datos) también conocidos como “workspaces” o “projects” en software comercial. La propuesta está dirigida a automatizar tareas de análisis espacial vertical, es decir aquellas que se aplican sobre un conjunto de capas de datos. En particular, es una alternativa de solución al problema de decidir cuales temas (capas) deben ser incluidos en un mapa (workspace) para realizar una tarea de análisis espacial específica. Por ello, enfatizan que las técnicas actuales de GIR, recuperan las capas de datos aisladamente y se agregan manualmente a un mapa. Pero, resulta más significativo analizarlas en conjunto.

Sin embargo, para definir las capas de datos que conformarán un mapa es necesario recuperar la información más relevante de entre todas las capas disponibles, y es en este punto donde tiene origen este trabajo.

No obstante, en la práctica los temas de un mapa no son analizados de forma individual, entonces su propuesta consiste en analizar el conjunto de temas que conforman un mapa previamente creado. El propósito de dicho análisis es extraer las relaciones espaciales presentes en estas capas de datos, para después explotarlas y a partir de ellas inferir cuales capas de datos son las más ideales para la construcción de un mapa. La justificación para argumentar que la extracción de las relaciones espaciales de los temas que conforman un

---

<sup>21</sup> Trabajan con base en probabilidades que son establecidas en cada nodo que conforma a una red.

mapa permitirá automatizar dicho proceso, se basa en que el analista GIS construye mapas de acuerdo al tipo de análisis que realizará. Entonces para tareas de análisis diferentes, los mapas contendrán algunos de los temas usados en otra tarea, es decir coincidirán en las capas usadas. Además, hacen referencia a que las redes bayesianas han sido adoptadas para asignar relevancia a un tema, de forma tal que el enfoque se puede usar para automatizar la creación de mapas.

En esencia, el trabajo propone tres algoritmos espaciales de aprendizaje de redes bayesianas, los cuales incorporan las relaciones espaciales presentes en los temas, en el proceso de aprendizaje. Las redes bayesianas resultantes fueron cargadas en una maquina de inferencia que se utilizó para recuperar todos los temas relevantes, de un conjunto de prueba, para una consulta de usuario.

El rendimiento de los algoritmos de aprendizaje espacial bayesiano se evaluó y comparó contra los algoritmos convencionales (aquellos que no manejan aspectos espaciales). Por otra parte, dentro de las desventajas, es que los algoritmos de aprendizaje de redes bayesianas ofrecen mejores resultados (en el aprendizaje) si el número de datos es grande. Mientras que su desempeño es más pobre cuando los datos son escasos. Además, los modelos probabilistas en el área de recuperación de la información ofrecen menores ventajas comparadas con el modelo vectorial.

Sin embargo, una aportación interesante en el trabajo es que la recuperación de la información, se realiza explotando y trabajando con datos vectoriales y sobre la tarea de análisis espacial de sobreposición, y es el único que ha trabajado con este enfoque.

## **2.7 Recuperación y ponderación de información con enfoques híbridos**

En esta sección se comentan los artículos que abordan la recuperación de información geográfica a partir de los métodos definidos en IR. Algunos trabajos extienden la funcionalidad para trabajar con datos geográficos. Mientras que otros proponen esquemas que mezclan o combinan medidas de IR con métodos de GIR.

### **2.7.1 G. Cai (2002)**

En este trabajo G Cai et al [12] discuten acerca de las ventajas y limitaciones de los sistemas actuales de recuperación de información geográfica (GIR) y los tradicionales (IR). En especial, para el tratamiento de información geográfica multimedia, por lo que propone un modelo de recuperación, el cual integra el indexado geográfico con el modelo vector-espacio. El primero usa coordenadas geográficas, mientras el segundo utiliza palabras clave. El propósito consiste en tener una mejor representación de información espacial.

La ponderación de documentos se realiza a través de la medición de la similitud consulta-documento, tomando en cuenta el grado de relevancia en el dominio temático y espacial. Se ofrece el soporte visual para la consulta y navegación considerando que el modelo reconoce las diferencias fundamentales en la configuración y dimensión del espacio temático y geográfico. El sistema utiliza bibliotecas digitales espaciales. Sin embargo, se requiere de mayor trabajo para que el modelo descrito sea establecido como un modelo de recuperación de información, básicamente porque integrar multimedios es una tarea que representa mayores retos, ya que la semántica de los diferentes medios no está disponible y hay que construirla, lo que impide que estos se integren de forma transparente.

### **2.7.2 Wang et al (2005)**

En el trabajo de Wang et al [74] se describen tres diferentes formas de ubicación geográfica, que pueden ser definidas a través del uso de páginas Web y datos relacionados. Estas son, la ubicación física del proveedor propietario del contenido Web, la ubicación del contenido ( el lugar descrito por el contenido), y finalmente la ubicación desde la cual el contenido es entregado, esto es definido como la región dentro de la cual el contenido es de interés. Tanto el proveedor como la ubicación del contenido son calculados a través del geo-parsing de documentos para referencias de lugares geográficos que están en diccionarios geográficos.

Un conjunto de reglas son utilizadas para distinguir entre sitios, donde dichos sitios son comúnmente usados para referirse a un proveedor o ubicación de contenido. Entonces, por ejemplo, si se encuentra que la ubicación de un proveedor aparece comúnmente en múltiples páginas y aparece de forma más frecuente en el encabezado o al pie de una página. Entonces se le considera a esta característica un candidato. Por lo tanto, teniendo identificados los posibles candidatos para ubicación del proveedor, entonces se manejan a través de un vector (similar al método vectorial de IR) para extraer la probable ubicación del proveedor.

También trabajan sobre el problema de que muchas páginas Web se refieren a múltiples ubicaciones, entonces describen un enfoque para identificar sitios dominantes a través del uso de un árbol geográfico de pesos basado en las asociaciones existentes entre las ubicaciones físicas (e.g. uso de suelo). Todas estas ubicaciones aparecen dentro de la página Web (se explora la página para identificar estas ubicaciones o sitios).

Finalmente, la ubicación del contenido es calculado usando un algoritmo similar, pero usando ubicaciones del usuario (obtenidos de los archivos *log* de servidores) y ubicaciones de hiperenlaces como fuentes de datos adicionales. Se hace uso de una aplicación que utiliza los algoritmos previamente descritos, pero el conjunto de datos utilizados es relativamente pequeño. Entonces los resultados iniciales muestran que, primeramente, distinguir entre estos diferentes tipos de ubicaciones es más provechoso ya que el 76% de las ubicaciones de contenidos son diferentes a la real ubicación del proveedor. Además, en las pruebas de datos que usaron los algoritmos aparecen buenos resultados en términos de *precision* y *recall*.

### **2.7.3 Paul Clough (2005)**

En este trabajo Clough [98] se enfoca en la extracción de la ubicación geográfica de páginas Web, utilizando conjuntos de datos provenientes fuentes de datos homogéneas. Sin

embargo, el enfoque se centra en identificar correctamente el contenido geográfico de cualquier tipo, almacenado en las páginas Web. Empleando técnicas tanto de reconocimiento de objetos geográficos por su nombre (*Named Entity Recognition*) hasta el análisis sintáctico- geográfico (*geo-parsing*) de documentos.

Teniendo nombres de lugares relacionados con el uso de suelo, se trabaja con el problema de ambigüedad. Entonces se asume que la ubicación más comúnmente referida a un nombre de lugar es aquél con la jerarquía administrativas más corta (e.g. *County-> Region->Town*). El enfoque es suficiente para ser aplicado a un rango de de diferentes países europeos, dado que las reglas aplicadas son independientes del lenguaje. Las pruebas se realizaron para cuatro países europeos, se presenta un análisis de lugares identificados como únicos en una colección de 800,000 páginas Web. Y se comenta que se tiene un promedio de alrededor de tres referencias geográficas por página.

#### **2.7.4 Zhisheng et al (2006)**

El artículo de [100] está enfocado a la búsqueda local, a lo que se ha denominado servicios de búsqueda basados en localización. Los autores presentan un sistema GIR que utiliza ubicaciones implícitas, el propósito del sistema consiste en mejorar el desempeño en la recuperación utilizando nombres de ubicaciones. En particular presentan un método de geo-indexado usando ubicaciones implícitas. Las pruebas las realizaron basándose en el *Geo-CLEF 2006*. Un ejemplo de una ubicación es: “Norte América” que puede ser vista como la ubicación implícita para “Canadá”

Dentro de los puntos a mencionar, es que evitan usar la técnica de expansión de consulta, debido a las desventajas que ésta representa. Su propuesta consiste en indexar ubicaciones explícitas e implícitas en conjunto, donde a estas ubicaciones se les asignan diferentes marcadores geográficos. Utilizan además geo-índices, utilizando archivos invertidos y el índice por rejilla, el cual divide la superficie de la tierra en una rejilla de 1000 X 2000. Dentro de los resultados obtenidos se muestra un mejor desempeño con el método propuesto en comparación con métodos anteriores utilizados en *Geo-CLEF 2006* y

finalmente se concluye que el proceso propuesto es más rápido en relación al enfoque de expansión de consulta.

## **2.8 Recuperación de información apoyada en la construcción de ontologías**

Esta sección comenta los trabajos que han utilizado alguna metodología o mecanismo para construir o procesar ontologías. Las fuentes de información utilizadas describen los datos principalmente mediante palabras.

### **2.8.1 Sang Ok et al (2003)**

En este artículo Sang Ok et al [15], proponen un método semi-automático para la construcción de una ontología usando agrupaciones de palabras (*hub words*).

El enfoque tiene como hipótesis la definición de ciertas palabras (*hub words*) las cuales están relacionadas con un conjunto de otras palabras. Estas *hub words* son determinadas por la frecuencia de aparición de los términos en un documento. Un detalle interesante es que se propone un proceso de construcción para una ontología usando *hub words* y un método automático para la extensión de la ontología (agregando relaciones). Además de que resaltan que la ontología propuesta se puede usar como un archivo de indexado en IR, justificando esta afirmación en el hecho de que una ontología puede ofrecer mayor información semántica que los archivos índices.

El artículo describe la construcción de la ontología y su extensión, así como el proceso de recuperación de información. La construcción de la ontología se realiza en tres pasos: 1) encontrar las palabras con mayor índice de frecuencia en una colección de texto (sustantivos), 2) Construir manualmente la ontología donde los nodos principales son las *hub words*, 3) Extender automáticamente la ontología (agregar las relaciones), este proceso se realiza agregando las palabras que tengan relaciones (acorde al dominio) con las *hub words* seleccionadas.

Por otra parte, el proceso de extracción de *hub words* se basa en técnicas conocidas de IR, sin embargo la aportación consiste en considerar el dominio del problema, ya que las *hub words* seleccionadas para un dominio económico serán diferentes para un dominio como la medicina. Por ejemplo, en un dominio económico la palabra “*company*” tiene muchas relaciones con palabras tales como: “*share*”, “*interests*”, “*stock*”, “*trade*”, “*lawyer*”,...entre otras, pero en el dominio de la medicina tiene pocas relaciones.

El proceso de extensión de la ontología consiste en agregar relaciones, siguiendo los mismos pasos que en la construcción de la ontología, pero considerando los verbos que aparecen entre *hub words* y las no *hub words*. El primer paso extrae los sustantivos vecinos de las *hub words* (se aplican reglas de análisis de enunciados) y el segundo paso establece dicha relación. Por ejemplo, si el verbo “*belong*” o “*include*” aparece entre dos sustantivos, entonces la relación “*belong to*” se considera como candidata a relación entre las *hub words* que contienen dichos sustantivos.

Se utiliza RDF para realizar las consultas en la ontología y se presentan gráficas que muestran los resultados obtenidos. Un detalle de este trabajo es que no se explica cómo se establecieron las reglas de extracción para las ontologías (las cuales serían muy valiosas). Además de que las técnicas usadas son estrictamente de análisis de texto.

## **2.9 Comentarios Finales**

En síntesis, el estado del arte muestra de manera clara que los desarrollos enfocados a GIR, consideran la explotación de las relaciones geométricas y topológicas de forma parcial o en su defecto sin profundizar en su contenido semántico. Es decir, no consideran las relaciones entre objetos espaciales, cuando éstos participan en una tarea de *análisis espacial* (e.g. *sobreposición*). Esto permite explorar este terreno para obtener resultados que puedan mejorar la relevancia de resultados para una consulta espacial.

Otro punto a favor de esta investigación, es que los trabajos relacionados con la ponderación de resultados, ofrecen mecanismos basados en técnicas de inteligencia artificial, pero no se enfocan en las relaciones semánticas para llevarla a cabo. En esta tesis la ponderación se basa en un mecanismo de cercanía entre conceptos, basado en un trabajo previo [19]. Este considera la semántica de las relaciones y sus conceptos. En igual forma, otros trabajos señalan y enfatizan la importancia de las ontologías en este proceso, y el rol que desempeñan en este ámbito. Sin embargo, no hay trabajos que la utilicen como el mecanismo que conduzca o guíe un proceso de recuperación.

Otro punto importante a favor de esta tesis, lo encontramos al considerar la semántica del análisis espacial que se aplica a datos geográficos, ya que a través de un análisis espacial se puede obtener nuevos datos, y en consecuencia derivar un “nuevo” conocimiento. Además de permitir procesar propiedades y relaciones de diversos fenómenos relacionados a un término geográfico para un área común.

Por esta razón, esta tesis considera criterios topológicos y geométricos para interpretar el significado de una consulta *geo-espacial*. Lo cual incluye aspectos tales como relaciones espaciales semánticas, descripciones, restricciones en representación y topología, entre otros atributos existentes en fuentes geográficas.

Uno de los beneficios que aporta el desarrollo de mecanismos de recuperación como el de este trabajo, es que el conjunto de resultados podrá ser utilizado en la asistencia de la toma de decisiones.

Además de que en GIR existen otros retos como garantizar interoperabilidad (no importa el formato y fuente de los datos). Como también, ofrecer resultados relevantes (datos en diferentes representaciones o aproximados).

Otra aportación es que debido a que las consultas ( $Q_G$ ) contienen expresiones con alto contenido de semántica espacial (*cerca, alrededor, en, etc.*) Entonces, realizar una interpretación automatizada de esta semántica espacial, es muy útil ya que los resultados que se obtengan serán de acuerdo al contexto geográfico. Lo cual permitirá recuperar documentos de acuerdo a su relevancia geográfica.

También, otro punto a favor es que no existen trabajos que realicen la recuperación desde repositorios no estructurados. Este escenario es típico en México, ya que instituciones como el INEGI<sup>22</sup> o el IFE<sup>23</sup> no ofrecen *servicios Web* en la actualidad, o los datos publicados no se encuentran en un formato estándar o estructurado como es el caso de XML, lo cual facilitaría su explotación y procesamiento.

También, se indica que el desarrollo de la presente tesis se enfoca en el procesamiento de documentos no estructurados (es decir no organizados con criterios espaciales) y con apoyo de fuentes de datos estructuradas. Mientras que la línea de investigación a la que está dirigido el trabajo actual es en la de recuperación de información geográfica. En particular está enfocada a: *“recuperación de información geográfica basada en la integración de fuentes de datos geográficas heterogéneas”*.

El área de aplicación se enfoca a: **la recuperación de información geográfica conducida por ontologías y ponderación integral basada en el método de confusión.**

Finalmente se subraya que el sistema **no compite contra Google o Yahoo!** es decir, la meta consiste en recuperar documentos asociados a datos geográficos, o que contienen alguna referencia geográfica.

---

<sup>22</sup> El Instituto Nacional de Estadística, Geografía e Informática (INEGI) es un órgano desconcentrado de la Secretaría de Hacienda y Crédito Público, con autonomía técnica y administrativa, en la República Mexicana.

<sup>23</sup> Instituto Federal Electoral, <http://www.ife.org.mx/portal/site/ife>

# CAPÍTULO 3

## MARCO TEÓRICO

*La educación debe ser impartida, desde la cuna hasta la tumba*

*José Martí*

En este capítulo se describen conceptos y términos que se utilizarán en la metodología desarrollada para esta tesis. En especial, se describirán los aspectos relacionados con la recuperación de información geográfica desde fuentes de datos geográficas tales como son ontologías geográficas, archivos vectoriales y diccionarios de datos. Se explican los conceptos afines a las relaciones y propiedades de objetos geográficos almacenadas en las tres fuentes anteriores. En donde se almacena la semántica de objetos geográficos de forma implícita y con diferente nivel de detalle.

También, se explican conceptos involucrados en los procesos de extracción, búsqueda y ponderación de datos. Para finalizar, se ofrecen las definiciones de un mecanismo de ponderación que se basa en un trabajo previo denominado: teoría de confusión.

Este capítulo se organiza de la siguiente forma: la sección 3.1 explica lo que significa IR y GIR. La sección 3.2 trata aspectos de la información geográfica y espacial. La sección 3.3 explica las fuentes de datos geográficos. La sección 3.4 describe las fuentes de datos vectoriales. La sección 3.5 aborda las ontologías y su relación con GIR. La sección 3.6 habla de las consultas espaciales y geográficas. Mientras que en 3.7 y 3.8 se encargan del análisis e indexado espacial. La sección 3.9 explica las redes semánticas. Además, las secciones 3.10 y 3.11 explican las jerarquías y la teoría de confusión. Finalmente la sección 3.12 esboza las conclusiones obtenidas.

### **3.1 Recuperación de Información**

Existen numerosas definiciones con respecto a la recuperación de información, sin embargo dentro de los trabajos revisados, no se encuentra una definición que sea aceptada de forma general. Por lo tanto, mostraremos algunos extractos de dichas definiciones para lo que se conoce como recuperación de información clásica y nuestra propuesta de lo que representa recuperar información geográfica.

La recuperación de información, por sus siglas en inglés (IR), es la ciencia de la búsqueda de información en documentos, textos, imágenes, sonido o datos de otras características, de manera pertinente y relevante. En fuentes tales como bases de datos, internet, intranet, entre otros. La recuperación de información es un estudio interdisciplinario. Abarca tantas disciplinas que genera normalmente un conocimiento parcial desde una u otra perspectiva. Algunas de las disciplinas que participan en estos estudios son la psicología cognitiva, la arquitectura de la información, diseño de la información, el comportamiento humano hacia la información, la lingüística, la semiótica, informática, biblioteconomía y documentación.

Los buscadores, tales como *Google*, *Yahoo* y *Live.com*, son algunas de las aplicaciones que surgieron como resultado de investigaciones en la recuperación de información.

Básicamente, en IR tradicional se requieren considerar los siguientes elementos: un vocabulario (lista de términos en lenguaje natural), un algoritmo que incluya las reglas lógicas de la búsqueda {Tabla de verdad} y una valoración de los resultados o cantidad de información lograda o posible. Además, se requiere de un motor de búsqueda, el cual permita plantear una pregunta con no menos de dos términos y mostrar los resultados mínimos para dicha pregunta (consulta) [29]. Adicionalmente, se requieren de mecanismos de ponderación para el despliegue de resultados al usuario.

### **3.1.1 Recuperación de Información Geográfica (GIR)**

La Recuperación de información geográfica GIR (por sus siglas en inglés) se encarga de proporcionar accesos a fuentes de información geográfica y espacial. Esta incluye todas las áreas que tradicionalmente han sido el núcleo de la investigación en IR, con un énfasis o extensión hacia aspectos geográficos y espaciales, tanto en procesos de ponderación, almacenamiento y de indexado [30]. GIR es un tema interdisciplinario de reciente nacimiento, pero de rápido desarrollo académico y comercial. En donde se procesan datos por medio de alguna noción de la relevancia geográfica en la información. La investigación actual enfrenta varios retos, de entre los cuales se destacan:

- La extracción de términos geográficos de datos estructurados, y de reto aun mayor, de datos no estructurados.
- La identificación y eliminación de ambigüedades en los procedimientos de extracción.
- Metodologías para almacenamiento eficiente de ubicaciones y sus relaciones.
- Desarrollo de máquinas de búsqueda y algoritmos para aprovechar las características de la información geográfica.
- La combinación de relevancia geográfica y contextual para aportar una relevancia significativa a los documentos; así como, técnicas para permitir al usuario interactuar y explorar los resultados de consultas para sistemas GIR.

En este punto, conviene distinguir lo que es IR y recuperación de datos (ver sección 3.1.3)

En IR el modelo subyacente que proporciona acceso a los documentos<sup>24</sup> es probabilista. Se interesa con cuestiones subjetivas e indeterminadas. Es decir, si un documento es relevante (a un cierto grado) para el usuario y su consulta. Mientras que la recuperación de datos es determinista con respecto a operaciones de recuperación. Si un documento cumple las condiciones especificadas en una consulta, entonces éste es por definición “relevante”.

En GIR se enfoca en ambos aspectos: recuperación determinista (e.g. encontrar todos los conjuntos de datos que contienen información sobre una coordenada particular) y recuperación probabilística (tales como encontrar todos los municipios cerca de un río) [31, 32]. GIR se enfoca en los problemas de encontrar fuentes de información que están relacionadas con ubicaciones geográficas. Debido a que la mayoría de máquinas de búsqueda tratan la terminología geográfica al igual que otra terminología, lo cual trae como consecuencia que se recuperen documentos irrelevantes. Actualmente, GIR trabaja en los siguientes retos:

- Existen muchos lugares con el mismo nombre (documentos que se refieren al lugar equivocado son recuperados).
- Existen muchos usos para nombres de lugares. Por ejemplo, se usan nombres de personas y de organizaciones para referirse a un lugar.
- Existen consultas que incluyen preposiciones espaciales, tales como cerca, afuera, que requieren consideraciones especiales en el ámbito geográfico. Por ejemplo técnicas de geo-parsing<sup>25</sup> y , geo-coding<sup>26</sup>

---

<sup>24</sup> Se utiliza el término documento para representar cualquier elemento de interés potencial en una colección o base de datos , sin considerar el contenido -- texto, imágenes, mapas, video, etc. -- o la forma – papel o digital

<sup>25</sup> El propósito consiste en detectar terminología geográfica.

<sup>26</sup> Consiste en adjuntar una referencia única de ubicación para un documento.

Una de las metas en GIR es la de proporcionar acceso a fuentes de información georeferenciada. Por lo tanto, podemos considerar a la recuperación de información geográfica como una especialización de la recuperación de información clásica. Esto incluye todas las áreas que tradicionalmente conforman el núcleo de la investigación en IR, con especial énfasis en el indexado y recuperación de información orientada espacial y geográficamente. Estos procesos se realizan de manera común a través de consultas espaciales (*Spatial Query*) o de consultas geográficas (ambos tipos de consultas son tratados en la sección 3.4).

En la literatura computacional han existido distinciones entre IR y recuperación de datos (data retrieval) este último término está asociado con sistemas manejadores de bases de datos (SMDB). En la práctica esta distinción es más de grado que de tipo. En la Figura 3-1 se muestra un espectro de varios atributos relacionados con la recuperación de información y la recuperación de datos. El estudio de estos atributos permitirá establecer la diferencia entre estos.



**Figura 3-1** Espectro: Recuperación de información vs. Recuperación de datos.

Como se observa en la Figura 3-1 el modelo utilizado en IR para el acceso a los documentos es *probabilista*. Está referido con cuestiones subjetivas e indeterminadas como pueden ser: que un documento satisfaga, *en cierto grado*, las necesidades de información de un usuario (que sea relevante para el usuario de acuerdo a su consulta-requerimiento). Por otro lado, la recuperación de datos, es *determinista* considerando operaciones de recuperación. Es decir, si un documento cumple las condiciones especificadas en la *consulta* del usuario, entonces es por definición “relevante”.

Entonces, cuando nos referimos a recuperación de información geográfica hablamos de ambos tipos de modelos de recuperación tanto determinista (e.g. encontrar todos los conjuntos de datos que contienen “cualquier” información para una coordenada en particular) como probabilista (e.g. encontrar todas las ciudades cerca de un gran río).

De esta manera, para recuperar información se requiere de un indexado tanto para garantizar un acceso eficiente a grandes bases de datos, como para organizar y limitar el conjunto de elementos de la base de datos que sean accesibles. La mayoría de los sistemas IR obtienen sus elementos índices a partir del contenido de los elementos a ser indexados. Esta obtención puede ser una simple extracción (tal como extraer palabras clave, “keywords”, de un texto) o extracción por inferencia (por ejemplo: el mapeo de una palabra a los términos en un tesaurus) o podría ser un análisis inteligente de asignación de elementos índice (e.g. asignar encabezados, “títulos”, de los temas en un documento).

Mientras que en la recuperación de datos, el elemento por sí mismo, en su totalidad, es la unidad de indexación. Claro, que en la escala de la Figura 3-1 esto no es lineal o continuo dado que ambos tipos de indexado podrían estar presentes dentro del mismo sistema. En GIR ambos extremos de la escala están mezclados. Por ejemplo, usando indexado inteligente (e.g. asignar las coordenadas de un cuadro de selección a una fotografía aérea), e indexado por inferencia (asignar coordenadas a los lugares que son mencionados en un texto).

En la recuperación actual de elementos de una base de datos, los algoritmos utilizados para *coincidir*<sup>27</sup> una consulta con los elementos del índice (o el contenido de una base de datos) están basados en el modelo de recuperación tradicional. Los modelos de recuperación de información guían a una clase de algoritmos de recuperación que son probabilistas por naturaleza, y pueden involucrar el cálculo de probabilidades y el uso de métodos de inferencia estadísticos, aunque también pueden tener un enfoque basado en otro modelo del espacio de documentos (e.g en [49]).

Estos modelos están enfocados a encontrar todas las coincidencias (parciales) potenciales entre una *consulta*<sup>28</sup> y un documento, mientras que la ponderación de éstas trabaja con base en parámetros que miden el “grado de coincidencia”, de forma tal que las “mejores” coincidencias reciben las más altas ponderaciones.

Los algoritmos de recuperación de datos son deterministas, y entonces demandan una *coincidencia* exacta entre la especificación de la *consulta* y el contenido de la base de datos. La lógica de Boole es utilizada en el procesamiento de los lenguajes basados en consultas (prácticamente en todos los sistemas comerciales manejadores de bases de datos) como también en los catálogos en línea y sistemas de recuperación de información, este también es un algoritmo determinista. En GIR la coincidencia determinista aproximada, parcial y precisa son de utilidad en el procesamiento de *consultas* espaciales y geográficas [50].

Las consultas en sistemas de IR son generalmente expresadas como enunciados en lenguaje natural de acuerdo a las necesidades de los usuarios que buscan información. Estas *consultas* son intrínsecamente imprecisas y pueden ser ambiguas. Mientras que en la recuperación de datos la *consulta* es típicamente expresada en algún tipo de lenguaje estructurado cuya sintaxis es precisa y de características semánticas.

---

<sup>27</sup> El término coincidencia será utilizado como el equivalente al término *matching* el cual es ampliamente usado en las áreas de lenguajes de programación, bases de datos, ingeniería de software, entre otras.

<sup>28</sup> El término *consulta* es usado como el equivalente al término *query* utilizado en base de datos, aunque en este caso existen diferentes estructuras para definirlo.

Por lo tanto, cuando el objetivo consiste en recuperar todas las unidades de la base de datos que coinciden exactamente con las especificaciones de la *consulta*, el problema de la ambigüedad en el enunciado de la *consulta* no se presenta (respecto a lo que se busca).

En consecuencia, los tipos de consulta pueden reflejar los modelos subyacentes de los sistemas de recuperación. En la recuperación de la información las consultas son consideradas como una “pista o clave” acerca de lo que el usuario podría considerar un elemento relevante de la base de datos, y adicionalmente esta recuperación está basada en que también un elemento coincide con la “clave”.

Típicamente, los resultados de una búsqueda son presentados en un orden ponderado, donde el criterio de ponderación se basa en el grado de “coincidencia” entre la *consulta* y el elemento de la base de datos.

En la *recuperación de datos* la consulta es tratada como una precisa especificación de los elementos deseados de la base de datos y la recuperación está basada en una correspondencia exacta entre el elemento y la *consulta*, a menos de que sea explícitamente indicado por el sistema o por el usuario como parte de la *consulta*, no hay ponderación u orden impuesto sobre los resultados para una *consulta* de recuperación de datos. Adicionalmente, la recuperación de información geográfica, es conocida como un área de investigación aplicada, la cual combina aspectos de investigación de otras áreas tales como DBMS, interfaces de usuario, GIS, y por supuesto IR. También trabaja con aspectos tales como el indexado, búsqueda, recuperación y navegación, exploración en diversas y múltiples fuentes de información *geo-referenciada*, y al diseño de sistemas que cumplan estas tareas de forma efectiva y eficiente [50].

### 3.1.2 Recuperación de información geográfica por interpretación de datos

Actualmente, en GIR se ha enfocado únicamente en la recuperación de datos y no en los procesos aplicados a los mismos (interpretación de datos). Por otra parte, no es lo mismo recuperar numerosas capas de datos, que recuperar una capa de datos de forma aislada. Es decir, si se consideran los procesos que afectan a los datos en conjunto (como en el análisis espacial) se pueden obtener resultados de mayor relevancia ya que se están considerando más aspectos en la recuperación.

Entonces, lo que se desea resaltar es que los enfoques se han dirigido a recuperar datos para posteriormente aplicarles análisis y procesos. Pero, resultaría más útil recuperar datos basándose en dichos análisis y procesos. Por lo tanto, lograr una recuperación basado en la interpretación de los datos representa una mayor utilidad y extendería el enfoque a usuarios comunes.

Por ejemplo, si se desea conocer donde existen terrenos fértiles para el cultivo de maíz, una recuperación de datos, arrojaría capas de datos de terrenos, clima, áreas de cultivo, entre otras. En este caso el usuario tiene que tener un conocimiento en GIS para integrar e interpretar los datos.

Pero, si la recuperación se basa en interpretación, entonces se arrojan datos que ya incluyen las capas resultantes de un análisis para encontrar terrenos fértiles. En otras palabras, *el usuario recibe la información integrada y procesada*, facilitando muchas de sus tareas. Además, si consideramos los enfoques conducidos por ontologías, en donde los datos se procesan de acuerdo a sus relaciones semánticas y a sus propiedades. Entonces, se pueden obtener resultados que con los enfoques sintácticos son omitidos. Por lo tanto, los trabajos que se dirigen en esta dirección permitirán mejorar los procesos de recuperación y consulta de datos.

### 3.2 Información Geográfica y Espacial

Debido a que en la literatura de GIS, se manejan con intensa frecuencia los términos: geográfico, espacial y geoespacial, entonces se requiere establecer una definición para los mismos. Sin embargo, en el caso del término *geoespacial*, aún no existe un consenso para aceptar una definición formal. Por ello, se citan varias definiciones relativas a dicho término, destacando las de OGC y Wikipedia.

Geoespacial es un término ampliamente utilizado para describir la combinación de software espacial y métodos analíticos con conjuntos de datos geográficos o de terreno. El término también es muy usado en conjunción con GIS y Geomática<sup>29</sup>. Muchos productos GIS aplican el término análisis geoespacial en un contexto limitado. En el caso de los GIS basados en datos vectoriales. Esto típicamente significa operaciones tales como sobreposición de mapas, buffer, y operaciones similares básicas. Esto refleja el uso del término análisis espacial dentro en la OGC “simple feature specifications”.

Para los GIS basados en *Raster*, ampliamente usados en las ciencias del medio ambiente y la percepción remota. El término típicamente significa un rango de acciones aplicadas a celdas de rejilla en uno o más mapas (o imágenes) frecuentemente involucrando operaciones de filtración o algebraicas (álgebra de mapas) [29].

Es por ello, que en esta tesis nos basaremos tanto en la definición anterior como en la que ofrece la OGC<sup>30</sup> para lo que denominó una ubicación geoespacial. La razón de esta decisión se debe a que uno de los tipos de consulta que se consideran en esta tesis incluyen criterios de localización. Además, la ubicación geoespacial es una propiedad fundamental para modelar el mundo en una forma coherente e intuitiva. La ubicación y el tiempo pueden ser explotados como un tema unificado para el mejor entendimiento del contexto de la

---

<sup>29</sup> Geomática es un término utilizado en algunas comunidades de GIS, el término proviene de la contracción de las palabras GeoInformación Automática. Muy similar al término informática (información automática)

<sup>30</sup> Open Geospatial Consortium, [www.opengis.org](http://www.opengis.org)

mayoría de los fenómenos reales y abstractos. La localización es contextualmente simple e intuitiva para la mayoría de las personas. Por ejemplo: la gente puede relacionar en donde se encuentran sobre un mapa, seguir direcciones a un lugar, facilitar la comprensión de contexto espacial en su ámbito local, entre muchas otras.

Por otra parte, la localización geoespacial es considerada ubicua, debido a que puede representarse en un mapa o en una imagen; también codificada como una dirección, código postal, o número telefónico; así como también en un lugar de referencia o evento en un texto; o en cualquiera de las diversas formas de representación para objetos en el planeta.

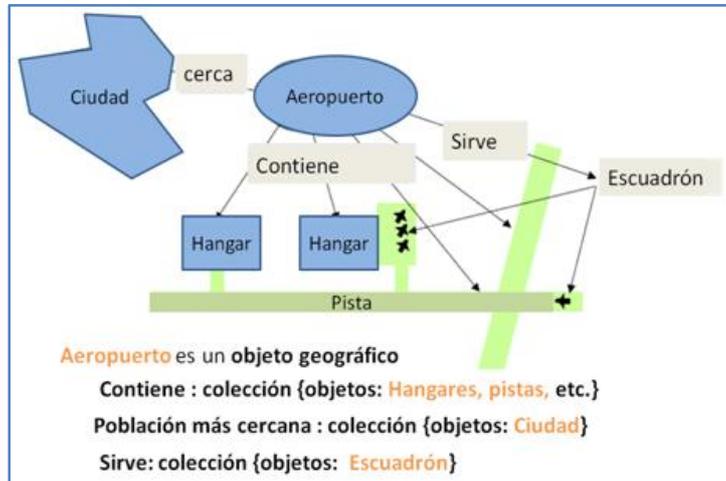
Entonces, está claro que la información geoespacial es muy amplia, y que su procesamiento y extracción se dificulta cuando no están estructuradas para procesar la semántica. Es decir, que debido a que las grandes fuentes de información no están relacionadas, no se puede explotar el conocimiento almacenado en ellas de mejor forma. Entonces, si estas fuentes de información fueran relacionadas y organizadas mediante sus propiedades geoespaciales y temporales, se podrían procesar y descubrir nuevas relaciones semánticas, lo cual permitirá mejorar muchos procesos y servicios a diversos niveles. Este es parte del fundamento teórico que motivo al desarrollo de esta tesis.

### **3.2.1 Datos Geográficos**

En este documento se utiliza el término objeto o dato geográfico para referirnos a *Feature* en el ámbito de GIS. Las características que definen a un objeto geográfico, de acuerdo a la OGC son que debe existir un único objeto geográfico relacionado con cada entidad del mundo real. Donde la granularidad es determinada por el usuario. Por “objeto geográfico” nos referimos a la unidad esencial de datos con la cual es conformada una base de datos geográfica. Este objeto geográfico puede manejar toda la responsabilidad de información en cascada acerca de una entidad del mundo real, esto incluye actualizaciones, creación o eliminación de atributos, metadatos e historial.

- Por “entidad del mundo real” nos referimos a cualquier cosa o fenómeno del cual se tenga datos almacenados.
- Los objetos tiene un único y persistente ID, el cual podría no ser configurable por el ser humano. Este ID puede ser usado para localizar el objeto geográfico.
- Los objetos geográficos tienen descripciones, las cuales a su vez no tienen restricciones únicas.
- Los objetos geográficos se componen de uno o más atributos (propiedades con nombre, tipo, valor, y [alguna descripción opcional]). Los atributos pueden ser nulos. La semántica de un atributo requiere que esta debe ser identificada unívocamente por su nombre. Debe ser posible preguntar por un atributo por nombre y descubrir lo siguiente:
  - Tipo de dato (e.g., entero, real, geometría, raster, referencias, etc.). El valor de regreso debe ser el valor raíz del atributo, o un valor derivado, dependiendo de los privilegios de acceso del usuario que expreso la consulta.
  - Tipo de colección (aplicable a objetos geográficos que son colecciones). Los objetos que sin colecciones siempre tiene uno o más objetos geográficos hijos. Los Padres de objetos geográficos contienen referencia a sus hijos.

Por ejemplo, la Figura 3-2, el objeto padre de aeropuerto tiene objetos hijos Hangares y Pistas de aterrizaje, presta servicio a objetos escuadron, y está cerca de otro objeto que es una población denominada ciudad.



**Figura 3-2** Objeto geográfico Aeropuerto, imagen de [39].

Los términos "comprende, presta servicio y cerca", son ejemplos de tipos de colección (i.e., tipos de relaciones de objetos que pertenecen a la colección). Un objeto geométrico es una combinación de una coordenada geométrica y de un sistema de referencia de coordenadas. En general, un objeto geométrico es un conjunto de puntos geométricos, representado por posiciones directas. Una posición directa guarda las coordenadas para una ubicación dentro de algún sistema de referencia de coordenadas [33].

Por otra parte, los datos geográficos son mucho más que imágenes electrónicas de mapas. Los datos geográficos que describen nuestro mundo permiten la planificación de ciudades, predicción de inundaciones y descongestionamiento, servicios de ruteo para emergencia, evaluaciones ambientales, monitoreo de vientos e identificación de patrones entre muchas otras aplicaciones. Los datos geográficos se procesan con software GIS el cual, como uno de sus aspectos de su funcionamiento produce mapas.

En el caso de México, existen diversas dependencias gubernamentales, así como empresas del ámbito privado que producen datos geográficos en forma de cartografía digital. Toda esta información a través de dos clasificaciones de formatos, conocidas como formato *raster* y formato *vectorial*.

### 3.2.2 Datos Espaciales

Los datos espaciales refieren a entidades o fenómenos que cumplen los siguientes principios básicos:

- Tienen posición absoluta: sobre un sistema de coordenadas (x, y, z)
- Tienen una posición relativa: frente a otros elementos del paisaje (topología: incluido, adyacente, cruzado, etc)
- Tienen una figura geométrica que las representan (punto, línea, polígono)
- Tienen atributos que lo describen (características del elemento o fenómeno)

### 3.3 Fuentes de Datos Geográficos

En la presente sección, se definen y explican las fuentes de datos utilizadas en esta tesis. Básicamente, son tres fuentes de datos geográficas: diccionarios, ontologías y archivos topológicos.

Comenzando describiendo los diccionarios de datos. En particular se consideraron aquellos que son proporcionados por el INEGI (Instituto Nacional de Geografía e Informática). Estos diccionarios de datos son clasificados por el INEGI de acuerdo a diversos temas y aplicaciones, para los propósitos de esta tesis se consideraron dos tipos de diccionarios de datos: 1) Diccionarios de Datos Toponímicos y 2) Diccionarios de Datos Vectoriales. Donde los diccionarios toponímicos son elaborados con base en la toponimia, es por ello que definiremos que es la toponimia, y posteriormente definiremos al diccionario toponímico.

La **toponimia** u **onomástica geográfica** es una disciplina que consiste en el estudio y origen de los nombres propios de un lugar. Los topónimos, nombres de lugar, en ocasiones tienen su origen en apellidos o nombres propios de personas, pero habitualmente su origen

está en algún aspecto físico del lugar que designan. La palabra toponimia proviene etimológicamente del griego τόπος *topos*, lugar, y ονομα *ōnoma*, nombre. La toponimia, además de estudiar el origen de los nombres, analiza los símbolos de los nombres de un lugar. Ejemplo: Ocotlán significa “junto a los pinos”; también se ha interpretado como “lugar de pinos u ocotes”.

Los diccionarios de datos toponímicos forman parte de la componente alfanumérica del de la base de datos geográfica (BDG) del INEGI. El diccionario consigna información alfanumérica relacionada directamente con los nombres geográficos que aparecen en la carta Topográfica, en sus diversas escalas, y en la carta *barimétrica*, escala 1: 1000 000; además, ofrece información básica referente a las localidades representadas en el componente vectorial de la base de datos.

Adicionalmente, estos diccionarios de Datos son generados a escalas 1:50,000 [40],[41] y 1:250,000 [43] y la referencia del modelo vectorial [42]. La utilización de estos documentos proporciona la base para identificar que objetos existen en las bases de datos espaciales generadas en México.

Cabe señalar, en este punto la importancia que tienen los nombres geográficos para referirse o designar un rasgo dentro del ámbito geográfico es indiscutible, pues nos permite, entre otras cosas, identificarlo en forma rápida y objetiva, en ocasiones, incluso, lo describen. En forma general los nombres geográficos, como elementos referenciales del paisaje geográfico, son creados y utilizados por el común de la gente, en ocasiones también, por usuarios especializados en investigaciones geográficas, y forma parte del acervo cultural de México. Entonces un diccionario toponímico es elaborado de acuerdo a la descripción anterior, incluyendo la definición de entidades, dentro de las cuales destacamos las siguientes:

**Nombre geográfico:** entidad que agrupa dentro de siete clases, definidas con base en términos genéricos asociados a todos los nombres geográficos, y establece una serie de características comunes a todos ellos.

**Localidad:** esta entidad ha sido considerada dentro del diccionario de manera integradora, para consignar dentro de un grupo de atributos características básicas de las localidades que aparecen en la carta topográfica, en sus diversas escalas. La Figura 3-3 muestra de forma gráfica un fragmento de la estructura de un diccionario de datos del INEGI.

<b>ATRIBUTOS</b>			
<b>DOMINIO FIJO</b>			
<b>TIPO DE AEROPUERTO</b>			
DOMINIO DE VALORES :			
Internacional : Permite vuelos con cubrimiento internacional.			
Nacional : Permite vuelos con cubrimiento nacional.			
Local : Sólo permite vuelos con cubrimiento regional.			
<b>DOMINIO VARIABLE</b>			
<b>IDENTIFICADOR DE AEROPUERTO</b> : Un número secuencial que se incrementa con cada ocurrencia			
DOMINIO DE VALORES :			
En el intervalo de 1 a N.			
<b>RESTRICCIONES DE INTEGRIDAD</b>			
<b>COMBINACIÓN(ES) AUTORIZADA(S) DE VALORES DE ATRIBUTOS</b>			
Internacional.			
Nacional.			
Local.			
<b>CALIFICADOR AUTORIZADO DE REPRESENTACIÓN(ES) GEOMÉTRICA(S)</b>			
Definida.			
Virtual.			
<b>RELACIONES</b>			
Aeropuerto (P)	Conecta	Camino (L)	
Aeropuerto (P)	Conecta	Carretera (L)	
Aeropuerto (A)	Conecta	Camino (L)	
Aeropuerto (A)	Conecta	Carretera (L)	
Aeropuerto (A)	Comparte	Calle (L)	
Aeropuerto (A)	Comparte	Camino (L)	
Aeropuerto (A)	Comparte	Carretera (L)	
Aeropuerto (A)	Comparte	Área de cultivo (A)	
Aeropuerto (A)	Comparte	Área urbana (A)	
Aeropuerto (A)	Comparte	Área verde urbana (A)	

**Figura 3-3** Fragmento de la estructura de un diccionario de datos del INEGI.

A continuación, se citan las definiciones que la OGC proporciona para un diccionario geográfico, la de un diccionario de datos y para un geocoder, que son términos involucrados en este apartado.

Un diccionario geográfico recupera las geometrías para uno o más objetos geográficos, dado un identificador de objeto (cadenas de texto). Los identificadores son cualquier

palabra o términos que describen objetos geográficos, tales como un conjunto de nombres de lugares o referencias de lugar. Cada instancia o diccionario geográfico tiene un vocabulario asociado o identificador. Un diccionario geográfico podría aplicar a una región dada, tal como un país, o alguna otra agrupación de objetos geográficos especializados. Un diccionario geográfico también puede ser visto como un caso especial de un geocoder.

Un geocoder transforma una descripción de localización de un objeto geográfico, tal como un nombre de lugar, dirección de una calle o código postal, en una descripción normalizada de la ubicación, la cual incluye una coordenada geométrica. Un servicio de geocoder recibe una descripción de ubicación de un objeto geográfico como entrada y proporciona una dirección normalizada con una geometría como salida. Adicionalmente los diccionarios geográficos y la geocodificación han sido utilizados en diversos trabajos para resolver la ambigüedad [33].

### **3.4 Fuentes de Datos Vectoriales**

Los datos vectoriales tratan con fenómenos discretos, cada uno de los cuales es concebido como un objeto geográfico. Las características espaciales de un fenómeno discreto del mundo real son representados por un conjunto de uno o más primitivas geométricas (puntos, curvas, superficies, o sólidos). Otras características del fenómeno son almacenadas como atributos del objeto. Generalmente, un objeto geográfico es asociado con un conjunto de valores de atributo. [33].

Mientras que la geometría proporciona el medio para la descripción cuantitativa, a través de coordenadas y funciones matemáticas, de las características espaciales de objetos, incluyendo, dimensión, posición, tamaño, forma, y orientación. Las funciones matemáticas utilizadas para describir la geometría de un objeto dependen del tipo de sistema de coordenadas utilizado para definir la posición espacial. La geometría es el único aspecto de la información geográfica que cambia cuando la información es transformada de un sistema de referencia geodésico a otro.

La topología trata con las características de las figuras geométricas que permanecen invariantes si el espacio es deformado elásticamente y continuamente. Por ejemplo, cuando los datos geográficos son transformados de un sistema a otro. Dentro del contexto de la información geográfica, la topología es comúnmente utilizada para describir la conectividad de un grafo  $n$ -dimensional, una propiedad que es invariante bajo una transformación constante del grafo.

También el uso más productivo de la topología se debe a que el procesamiento de la geometría computacional es más rápido. Los cálculos geométricos tales como contenido (punto-en-polígono), adyacencia, límites, y trazado en redes son de alto costo computacional. Por esa razón, las estructuras combinatorias conocidas como de topología complejas son construidas para convertir los algoritmos de geometría computacional en algoritmos combinatorios. Otro propósito es, dentro de la información geográfica, relacionar instancias de objetos independientemente de su geometría [33].

Por otra parte, muchos términos se refieren a las ubicaciones cerca de la superficie terrestre, e.g., identificadores y nombres de lugares. Estos últimos son ambiguos. Por ejemplo, “Springfield”, requiere información adicional para ser resuelta dentro de una ubicación específica (muchos lugares se llaman igual en Estados Unidos). El mismo caso aplica a “San José” en el continente americano, y “Lázaro Cárdenas” e “Insurgentes” en la República Mexicana. Este tipo de terminología requiere ser explotada con un enfoque semántico y espacial, de tal forma que los resultados del procesamiento de éstas, sean de mayor relevancia. Este tipo de situaciones han sido exploradas en trabajos previos como en [2] y [79] así como, en los congresos y workshops de GIR en sus recientes ediciones [80], [81]. Sin embargo, en el caso de México, no existen trabajos al respecto, y por ello estas requieren ser estudiadas y analizadas (ver capítulos 2 y 3).

### 3.5 Ontologías como Fuente de datos

Las ontologías en el área computacional, tienen un auge importante al día de hoy. Para comenzar ofreceremos definiciones generales, así como las más citadas de acuerdo a los precursores de este concepto.

- Una ontología hace referencia al intento de formular un exhaustivo y riguroso esquema conceptual dentro de un dominio dado, con la finalidad de facilitar la comunicación y compartir la información entre diferentes sistemas [29].
- “Una Ontología es el estudio metafísico<sup>31</sup> de la naturaleza del ser y su existencia” [26].
- Una Ontología es una especificación explícita de una conceptualización<sup>32</sup> [24].
- “Campo de la metafísica que investiga y explica la naturaleza, las características y las relaciones esenciales de todos los seres, como tales, o los principios y las causas de ser” [46].

Podemos entonces decir, de forma general, que una Ontología es una descripción formal de conceptos en el dominio de un discurso. También son teorías que especifican un vocabulario relativo a un cierto dominio. Este vocabulario define entidades, clases, propiedades, predicados y funciones, además de las relaciones entre estos componentes. Las ontologías toman un papel clave en la resolución de interoperabilidad semántica entre sistemas de información y su uso dentro del contexto.

---

<sup>31</sup> Metafísica. El estudio filosófico del ser y del saber.

<sup>32</sup> El cuerpo del conocimiento representado formalmente esta basado en una conceptualización: los objetos, conceptos y otras entidades que se asume existen en un área de interés así como las relaciones que existen entre ellas [47]. También, una conceptualización es una vista abstracta simplificada del mundo que se quiere representar para algún propósito.

### 3.5.1 Ontologías en función de la conceptualización

En 1995, Thomas Gruber presenta a las ontologías como un medio para reutilizar y compartir el conocimiento entre aplicaciones [24]. Cuya definición dice: “*Una Ontología es una especificación explícita, formal de una conceptualización compartida*”. De donde se realizan las siguientes observaciones:

Una *conceptualización* se refiere a un modelo abstracto de algún fenómeno en el mundo, el cual identifica los conceptos relevantes de dicho fenómeno. *Explícito* significa que los tipos de conceptos utilizados, y las restricciones en su utilización, están claramente definidos. *Formal* se refiere al hecho de que la Ontología debe ser legible y procesable por los programas (machine- readable). Mientras que *Compartido* significa que la Ontología surge a partir de un consenso entre muchas partes.

Considerando esta definición una base de datos puede ser una Ontología. Es una representación abstracta de un fenómeno del mundo real. Y también es explícita y procesable por programas. Sin embargo, se puede argumentar si representa conocimiento compartido. Debido a que los esquemas de base de datos son típicamente desarrollados para uno o un conjunto limitado de aplicaciones. Mientras que una Ontología tiene que ser un acuerdo entre muchas partes (y no solo una).

En el sentido filosófico, podemos referir a una ontología como un sistema particular de categorías para una cierta visión del mundo. Formalmente, una ontología es la declaración de una teoría lógica. Las ontologías se comparan frecuentemente con jerarquías taxonómicas de clases por las definiciones de clase y las relaciones *subsumption* (de una categoría general a una particular), pero las ontologías no necesitan ser limitadas a esas formas.

### 3.5.2 Ontologías en función de axiomas y lógica de primer orden

Por otra parte, otro artículo de gran relevancia para ontologías, es el de Nicola Guarino [45]. En donde, se describe la introducción formal de los principios básicos para ingeniería de conocimiento con el fin explorar las relaciones entre Ontología y representación de conocimiento. También, se definen algunos conceptos como Ontología, compromiso ontológico y conceptualización. Además de basarse en axiomas y lógica de primer orden.

Debido a que una Ontología es una especificación de una conceptualización compartida, los expertos en el dominio, los usuarios y diseñadores necesitan ponerse de acuerdo sobre el conocimiento especificado en la Ontología, para que la Ontología resulte útil. Es difícil lograr tal acuerdo. Por lo tanto, resulta ventajoso dividir el conocimiento en capas para diferentes ontologías basándose en su generalidad, de tal forma que no todo mundo tenga que estar de acuerdo con todas las ontologías. Únicamente requieren estar de acuerdo en las ontologías de nivel superior, las cuales son utilizadas en ontologías de dominio específico y de aplicación [84]. En donde se sugiere que el conocimiento se establezca en al menos tres capas: “*Language Knowledge*”, “*Domain Knowledge*” y “*Application Knowledge*”

Por su parte, Guarino [45] también identifica tres capas de conocimiento correspondientes a tres diferentes tipos de Ontología, basándose en sus niveles de generalidad, nombrándolas:

- ***Top-level ontologies***: Describen conceptos generales, independientes de cualquier tarea o dominio particular. Ejemplos de estas son *WordNet* [26] y *Cyc* [85]. Sin embargo, *WordNet* si contiene algunas relaciones dependientes del dominio [86], por su parte *Cyc* contiene algunas micro-teorías, las cuales son dependientes del dominio. Entonces, al parecer, una separación estricta no siempre es posible.
- ***Domain ontologies y task ontologies***: Describen, respectivamente, conceptos generales para un dominio particular y conceptos genéricos para una tarea general. Un ejemplo es la

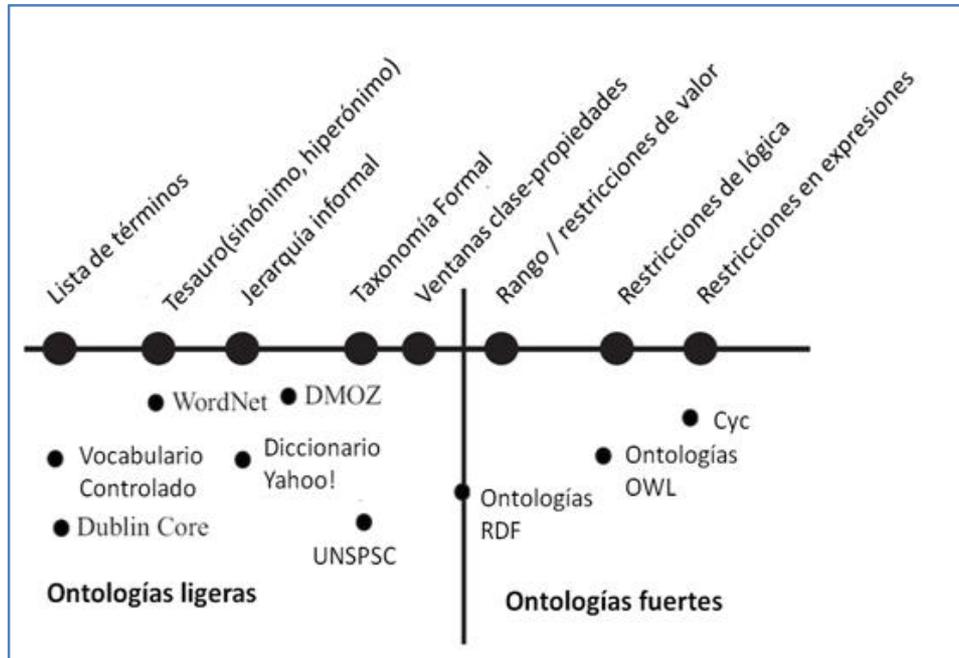
ontología empresarial [87], aunque se podría discutir que esta es una Ontología “Top-level”, porque puede ser aplicada a diferentes empresas, operando en diferentes dominios.

- **Application ontologies**: Describen conceptos que dependen del dominio y la tarea. En [88] y [89] se describe un enfoque ligeramente diferente al distinguir los tipos de ontologías. Haciendo una distinción entre conocimiento fijo y conocimiento en la resolución de problemas. Los niveles de generalidad distinguidos para ontologías de conocimiento fijo corresponden en grandes rasgos a los niveles distinguidos por Guarino.

Por otra parte, existen muchos tipos de ontologías, construidas para muchas áreas de aplicación, por ejemplo *WordNet* y *Cyc*, tiene variaciones en el grado de detalle que expresan. La Ontología de *WordNet* [94] es muy amplia (considera todo el idioma inglés) pero el nivel de detalle es bajo. Solo tiene descripciones en lenguaje natural de los términos. *Cyc* [95] es una Ontología de alcance amplio, que intenta capturar todo el conocimiento en sentido común (espacio, tiempo) pero con alto nivel de detalle. Con muchas relaciones formales entre términos, y procesables por programas.

### **3.5.3 Clasificación de ontologías**

Esta clasificación abarca varios niveles de expresividad, basándose en el espectro de Ontologías que presento McGuinness [83], la cual aparece en la Figura 3-4.



**Figura 3-4 Espectro de Ontologías.**

Finalmente, en términos prácticos, el desarrollo de una Ontología incluye: definir clases en la Ontología, colocar las clases en un jerarquía de taxonomías (subclase-superclase), definir [slots (o atributos)], describir los valores permitidos para esos [slots], y rellenar los valores de los slots con ejemplos.

### **3.6 Consultas Geográficas y Espaciales**

Referirse a los términos consultas geográficas y consultas espaciales implica consultar a una base de datos indexada espacialmente. Esta indexación se efectúa con base en las relaciones entre elementos particulares de la base de datos y dentro de un sistema de coordenadas particular (o sistemas de coordenadas compatibles). Las consultas espaciales o búsquedas espaciales (*spatial querying*) son un término más general en el campo de la geoinformática.

Una *consulta* espacial puede ser definida como aquella que involucre relaciones espaciales (**intersección, contenido, con límites, en los linderos, adyacencia, proximidad**) de entidades geoméricamente definidas y localizadas en el espacio [48] sin considerar la naturaleza del sistema de coordenadas. Esto puede ser sustentado considerando que el modelo de espacio vectorial de IR es un sistema de búsquedas espaciales, donde el espacio y las coordenadas son definidos por la ocurrencia y frecuencia de uso para un término dentro de una colección de documentos. Mientras que una *consulta geográfica*, requiere involucrar el espacio, pero a través de sistemas de coordenadas. En otras palabras, efectuar *consultas geográficas* asume que el espacio es delineado por los sistemas de coordenadas bien definidos del “mundo real”. Como se ha señalado en [51], existen muchas características de los datos geográficos que requieren estructuras de datos y métodos de acceso espacial.

En general, **las relaciones geográficas en los sistemas de coordenadas impuestos en el mundo real son relaciones geométricas** [50]. Dentro de un marco de trabajo geométrico, donde la dirección y distancia pueden ser medidas en una escala continua, muchos tipos de relaciones entre objetos definidos dentro de este espacio pueden ser definidos usando la geometría.

Se puede citar un ejemplo para ilustrar esto: si tenemos las coordenadas en latitud-longitud de Chicago (4152'N 8737'W) y Nueva York (4040'N 7358'W), un cálculo simple y sencillo puede darnos la distancia entre las dos ciudades. Utilizando el método del gran círculo, la distancia es  $(1/2 \cdot 7915.6) \cdot (0.86838) \cdot (D / 180)$ , donde 7915.6 es el diámetro de la tierra en millas, 0.86838 es el radio transformado de millas a millas náuticas, y  $\cos D = \sin \text{Latitude}_1 \sin \text{Latitude}_2 + \cos \text{Latitude}_1 \cos \text{Latitude}_2 \cos (\text{Longitude}_1 - \text{Longitude}_2)$ , o casi 651 millas náuticas. Empleando únicamente las coordenadas es sencillo determinar otras relaciones entre las ciudades: e.g., Chicago está al Oeste y Norte de Nueva York.

Las relaciones espaciales pueden ser tanto geométricas como topológicas (relacionadas espacialmente pero, sin dirección absoluta o distancia que sea pueda medir). Algunos ejemplos de relaciones topológicas incluyen propiedades tales como la *adyacencia*, *conectividad*, y *contenido en*. Por ejemplo, para conocer si un edificio está dentro o fuera de los límites de la ciudad de Chicago, se debe construir con la relación del edificio una delimitación arbitraria (boundary), sin que la distancia o dirección entre los dos sea considerada. Además, las direcciones podrían no tener relación particular con algún sistema de coordenadas dentro del cual estén involucradas. Por ejemplo, "Izquierda" y "Derecha" son direcciones válidas solamente en relación al punto de referencia del observador y no tendrán ninguna relación con "Norte" u "Oeste".

Las *consultas espaciales* y *geográficas* combinan elementos geográficos y topológicos. Los autores de [52] sugieren que hay dos clases primarias de solicitudes realizadas por los usuarios: la consulta “¿*Que hay (está) aquí?*” y la consulta “¿*Dónde está esto?*”. El primer tipo de consulta tiene como meta descubrir información disponible acerca de una ubicación particular, mientras que el segundo tiene como meta descubrir donde ocurre cierto fenómeno. Dentro de esta simple clasificación de consultas geográficas y espaciales, existe una gran cantidad de diferentes tipos de consultas, que se distinguen por la forma en que las ubicaciones son definidas. Estos tipos de consultas son discutidos en [48] y se describen brevemente en la siguiente sección.

### **3.6.1 Tipos de Consultas Espaciales**

Los diferentes tipos de consultas espaciales utilizados en sistemas de información pueden ser arbitrariamente complejos en cuanto a los tipos de información requerida, las delimitaciones en las áreas, periodos de tiempo, etc. Y muchas otras condiciones (espaciales o no) que podrían ser específicas en tales consultas.

Si nos concentramos únicamente en los aspectos geográficos o espaciales de la consulta, las podemos distinguir basándonos en el tipo de información proporcionada por el usuario en dicha consulta. De tal forma que podemos considerar cinco tipos de consultas espaciales: 1) Consultas de punto en polígono, 2) Consultas de región, 3) Consultas de distancia y zona de proximidad (Buffer), 4) Consultas de ruta, 5) Consultas multimedia.

El primer tipo de “consulta” es probablemente el más simple para procesar y describir. Esta es la consulta *Punto-en-polígono*, la cual es mostrada en la Figura 3-5.



**Figura 3-5 Consulta de tipo “Punto en polígono”**

La Figura 3-5 esencialmente formula la pregunta:

"¿Que hay/tenemos en este punto X, Y en el sistema de coordenadas actual?"

Otro ejemplo sería (en el contexto de una biblioteca digital) ¿Cuáles imágenes satelitales están disponibles para mostrar un punto en particular? o ¿Cuáles documentos describen el lugar indicado por el punto?

Este tipo de consultas requieren cualquier objeto geo-referenciado o conjunto de datos geográficos que contenga, se refiera o este “cerca” de un punto en particular sobre la

superficie de la tierra. Este tipo de consultas espaciales son los más precisos. La Figura 3-5 y la Figura 3-6 muestran las consultas “punto en polígono” y de “tipo Región”. En la Figura 3-6 se muestra el tipo de consulta *Región*.



**Figura 3-6** Consulta de Tipo Región.

Un ejemplo de la consulta de tipo *Región*, presentado en la Figura 3-6 es "¿Que tenemos/hay en esta región?" en lugar de referirnos un punto particular del espacio de coordenadas, una consulta de tipo *Región* define un polígono en dicho espacio y solicita información concerniente a cualquier cosa que este contenida en la región, adyacente a ella, o que sobrepone el área poligonal definida.

En este caso, existen numerosas variantes o restricciones potenciales que pueden ser aplicadas por ejemplo, un usuario puede preguntar:

- ¿Cuales puntos se encuentran dentro de la región?
- ¿Cuáles líneas (límites, ríos, etc.) se encuentran dentro o cruzan la región?
- ¿Cuales áreas (o conjuntos de datos regionales) sobreponen esta región?
- ¿Cuales áreas se encuentran completamente dentro de esta región?
- ¿Cuáles áreas comparten un límite fronterizo con esta región?

Cualquier combinación de elementos o criterio acerca del contenido puede ser indicado, de acuerdo a las necesidades de un usuario particular. Por lo tanto, una consulta de región

puede referirse a cualquier polígono, desde formas regulares tales como rectángulos o incluso círculos (lo cual sería equivalente a una consulta de zona Buffer sobre un punto como se describe más adelante) hasta formas irregulares como los límites de una ciudad, o cualquier conjunto de puntos arbitrarios definiendo un polígono cerrado. El criterio acerca del contenido no requiere ser precisado, pero podría usar interpretaciones probabilistas o difusas de cosas tales como el área máxima o mínima de superposición para un objeto que se considera incluir en el área indicada, o las áreas de cobertura para datos particulares que son candidatos para la recuperación [53].

El siguiente tipo de consulta es el de *distancia* o *zona Buffer*, mostrado en la Figura 3-7.



**Figura 3-7** Consulta de tipo Distancia y Zona Buffer.

Un ejemplo de consulta zona Buffer, mostrado en la Figura 3-7 es, "¿Qué hay dentro de una distancia fijada para un objeto (punto, línea o polígono)." Obviamente existen diferentes y numerosos pasos de procesamiento involucrados si el objeto usado como base para una consulta de zona buffer es un punto, una línea, o un polígono.

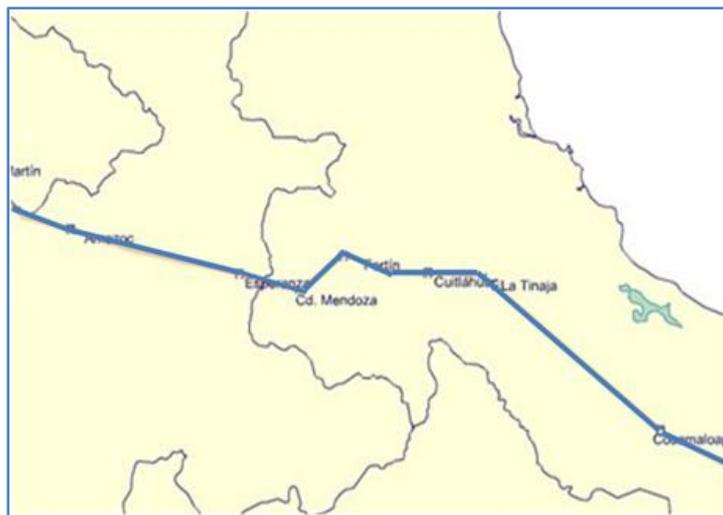
Algunas consultas de este tipo son:

- ¿Cuales ciudades se encuentran a 40 millas al norte y sur del límite de este estado?" (Figura 3-7).
- ¿Cuales plantas industriales se encuentran a 2 millas de este Río?"
- ¿Cuáles flujos-corrientes se encuentran a 10 Km. de esta carretera?"

- ¿Cuáles minas están dentro de un radio de 5 km de esta ciudad?"

La zona buffer no necesita ser exacta, e.g., "¿Cuáles datos describen el área cerca de este punto?" para ofrecer una respuesta ciertos trabajos proponen una función probabilista o "difusa" basada en la ubicación de los objetos de base de datos. Para tales consultas, una lista ponderada de objetos de la base de datos ordenados por "cercanía" al punto, podría ser una mejor respuesta que una definición arbitraria de una distancia.

Las consultas de ruta son una forma especializada de una consulta espacial, la cual requiere la presencia de una estructura de red en los datos espaciales o geográficos. Las redes son conjuntos de segmentos de líneas interconectados, las cuales representan cosas como caminos, petróleo o tuberías de agua, etc. La Figura 3-8 ilustra este tipo de consulta.

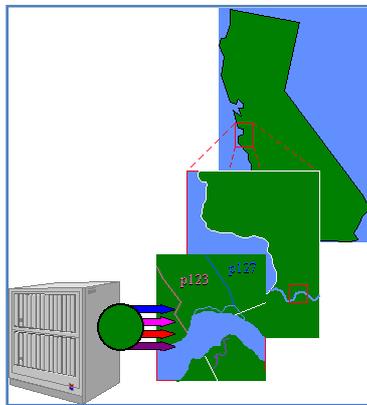


**Figura 3-8** Consulta de tipo ruta.

Una consulta típica de ruta involucra encontrar la ruta más corta desde un punto en la red hacia otro. Por ejemplo, "¿Cuál es la ruta más corta de Cancún a Cabo San Lucas?" (Ver Figura 3-8). Además, se debe considerar que las consultas de ruta pueden llegar a ser más complejas (y sin certeza) que las consultas multimedia, esto ocurre cuando otros criterios aparte de la dirección y la distancia están involucrados. Por ejemplo, la consulta "¿Cuál es la ruta más rápida de Cancún a Cabo San Lucas?" para ofrecer una respuesta se requiere

mayor información, como los límites de velocidad y las condiciones de tráfico sobre diferentes rutas de la red para proporcionar incluso una respuesta aproximada a la pregunta.

Mientras que, las consultas multimedia combinan múltiples fuentes de información georeferenciada para poder ofrecer una respuesta a una consulta. Esto puede incluir mapas múltiples (o varias capas de mapas dependiendo del tipo de sistema usado para resolver la consulta), además de información no referenciada ni presente en un mapa, tales como registros de propiedad para predios o lotes particulares de la tierra. La Figura 3-9 ilustra una consulta de este tipo.



**Figura 3-9 Consulta de tipo Multimedia.**

Por ejemplo, considerando la Figura 3-9 una posible consulta sería "¿Cuales son los nombres de los granjeros afectados por inundaciones en las ciudades de Monterrey y Santa Cruz?" Responder a esta consulta no solamente involucra información de un mapa, tales como los linderos de una ciudad y ubicaciones de ríos, sino también información del catastro para mostrar quien es el propietario de los predios a lo largo de ríos específicos en las áreas afectadas por inundaciones.

En esta consulta particular, las operaciones complejas son comúnmente requeridas, tales como la combinación de fotografías aéreas o satelitales así como los datos de percepción remota (mostrando la extensión de la inundación) e información de mapa y catastro, con frecuencia de diferentes bases de datos, con diferentes medidas, escalas y niveles de detalle.

Los tipos de consultas descritos con anterioridad pueden ser combinados en una búsqueda compleja. Por ejemplo, la consulta: "*¿Cuales corrientes o ríos fluyen a través del estado en el cual existe la ciudad de Richmond (California)?*" requeriría una búsqueda *punto-en-poligono* de información de estados para localizar el estado que contiene la ciudad, y una búsqueda de región para identificar las corrientes y ríos que intersectan el área de la provincia. Obviamente, cualquier sistema GIR debe combinar el texto o recuperación basada en conceptos asociada con información convencional y sistemas de bases de datos con los tipos de consultas espaciales descritas anteriormente.

Cualquier sistema de información multimedia podría incluir una amplia variedad de información espacial y no espacial que podría tener una asociación geográfica, si no una localización precisa [55]. Por ejemplo en [56], proporcionan un diseño para un sistema que combina características espaciales, textuales, y recuperación basada en conceptos.

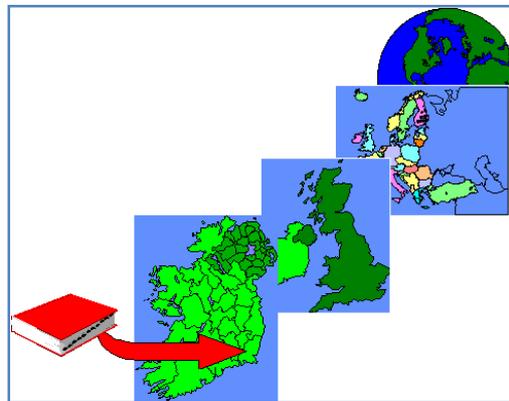
### **3.6.2 Exploración o navegación espacial**

Buscar en una base de datos geográficamente indexada o biblioteca digital es una actividad que asume que el usuario tiene una buena noción de lo que busca, y es capaz de especificar lo que necesita en alguna forma. La mayoría de las consultas de ejemplos anteriores reflejan esta situación. Otro tipo de búsqueda es mucho menos dirigida, y asume que los usuarios tienen alguna noción del tipo de información deseada, pero podría suceder que ellos sean capaces de especificar esta información en un lenguaje de consultas. Lo que es necesario en estos casos es tener la capacidad para explorar la base de datos geográficamente, sin requerir una formulación explícita de la consulta.

Esta exploración espacial combina consultas espaciales "ad hoc" con el despliegue de mapas digitales interactivos que permiten al usuario explorar la dimensión geográfica de la información en una base de datos o biblioteca digital. Por ejemplo, en [56] describen la exploración espacial usando el concepto de "hypermap". En las bases de datos de

hipertexto<sup>33</sup> cada documento (o nodo) puede contener muchos enlaces a otros documentos en una amplia gama de medios<sup>34</sup> y el usuario podría ver cualquier documento referenciado simplemente al seleccionar el enlace en el documento actual.

En un *hypermap*, los enlaces son representados por un icono o footprint (un polígono que resalta el contorno del área descrita por el objeto enlazado al icono), y la seleccionarlo se presenta en pantalla el documento referenciado por dicho enlace. La Figura 3-10 muestra un ejemplo de hypermap.



**Figura 3-10 Exploración Espacial.**

En la Figura 3-10 se muestra una secuencia de mapas que podrían ser mostrados a alguien que explora un sistema GIS, yendo desde una vista global de Europa, y luego hacia el Reino Unido, después hacia Irlanda, y finalmente hacia un icono particular sobre el mapa de Irlanda que representa un libro relacionado con la región.

Existen numerosas ventajas en los sistemas de exploración espaciales y el concepto de *hypermap* como una interfaz de usuario "metáfora". Estos sistemas son con frecuencia muy intuitivos y comprensibles<sup>35</sup> y puede proporcionar interacción directa en la búsqueda y la exploración, contrario a la especificación de nombres y coordenadas. En la mayoría de los casos, para los propósitos de especificación en exploración y búsqueda, el mapa digital

---

<sup>33</sup> El mejor ejemplo del hipertexto, actualmente, es la www.

<sup>34</sup> Texto, imágenes, video, clips, de sonido etc.

<sup>35</sup> Asumiendo que el usuario tiene alguna noción de geografía.

desplegado al usuario no necesita ser altamente detallado, ni requiere la precisión que ofrece un GIS.

Aunque, también hay numerosos problemas potenciales, o requerimientos para tales sistemas. Un problema es el agrupamiento desordenado en el despliegue. Si todos los iconos o footprints (representando a todos los documentos de una gran BD asociada con cualquier área geográfica visible en el mapa digital) son mostrados simultáneamente, el mapa podría desaparecer en su totalidad por debajo de una pila de iconos.

Otro requerimiento obvio para la exploración espacial y que debe realizarse es el indexado basado en coordenadas de la base de datos. En la siguiente sección, los métodos de indexado automático y geo-referencia automática de documentos de texto son analizados.

### **3.7 Indexado geográfico y espacial**

Una de las mayores fuentes de información es el texto en una amplia variedad de formas y de diferentes fuentes. Estos elementos de texto podrían incluir documentos que contienen solo texto tales como revistas científicas o artículos de enciclopedias, libros, reportes técnicos, y documentos más especializados tales como reportes de impacto ambiental (EIRs), leyes y legislación. Muchos de estos documentos describen, discuten o se refieren a lugares particulares o regiones. La ubicación geográfica es un criterio usado con frecuencia, o incluso es criterio primario cuando se busca información en un sistema o biblioteca digital.

En la práctica tradicional, en los catálogos de bibliotecas, las referencias geográficas han sido una forma común para asignar un punto de acceso a los documentos (primeramente libros y mapas), pero la asignación se efectúa a criterio del catalogador, es decir si la identificación geográfica se considera un mecanismo importante para acceder al documento. Aunque esto podría ser posible en un principio para que los catalogadores evalúen cada elemento que entra a la biblioteca digital con referencias geográficas en su contenido, sin embargo un mecanismo tan detallado sería prohibitivamente costoso.

La automatización de este indexado y catalogado es uno de las metas en sistemas IR y bibliotecas digitales. Un componente importante de este indexado automático es el desarrollo de métodos que puedan automáticamente *geo-referenciar* documentos de texto. La *geo-referenciación* automática, significa automáticamente indexar y recuperar un documento de acuerdo a las ubicaciones geográficas discutidas, desplegadas, o de otra forma asociadas con su contenido.

En la mayoría de los sistemas de IR para texto plano y bibliográficos, las búsquedas con un componente geográfico, como punto-en-polígono, consultas de región o multimedia “*localiza cualquiera de los documentos cuyo contenido este referido a la ubicación XY*”, no están soportados directamente por el indexado, consulta, o funciones de interfaces de usuarios.

En lugar de ello, estas búsquedas se basan en el indexado y especificación de la consulta de nombres de lugares, ya sea proporcionado por catálogos o extraídos del texto por si mismos, esencialmente como efecto lateral del indexado de palabras clave. Incluso en casos donde un documento es manual y meticulosamente indexado, los términos geográficos del indexado consisten de cadenas de texto (como nombres de autoridades) los cuales tienen numerosos problemas bien documentados con ambigüedad, sinonimia, y con cambios de los nombres a través del tiempo [54,57]. Específicamente, los principales problemas son:

- Los nombres no son únicos: San José es un nombre de ciudad muy común en América Central y Sudamérica, como también en California. En México muchos estados tienen avenidas llamadas Lázaro Cárdenas y López Mateos. Sin modificaciones adicionales, muchos nombres de lugares son ambiguos.
- Los lugares cambian en el tiempo en aspectos como: tamaño, formas y nombres.
- Los cambios políticos en el mundo se mueven mucho más rápido que los cambios geológicos, y los límites, ciudades y nombres de regiones, incluso la existencia de entidades políticas podrían cambiar en cualquier momento.

- Variaciones en la escritura: nombres locales para una región podrían diferir desde las formas comunes en inglés, y también podrían existir variaciones en la escritura o referencia del nombre sobre el tiempo (e.g. Peking, Beijing).
- Algunos nombres de lugares en los textos son simplemente convenciones temporales para estudios específicos, contextos históricos. Por ejemplo, ciertos nombres pueden ser creados por estudiantes para describir un área o región (áreas de estudio, campos de batalla, etc.) que no son parte de los nombres de la división política de una región, pero son definidos en forma precisa para los propósitos del estudio o investigación (e.g. GeoS en el CIC, MAGNO en el CIC).
- Los nombres de lugares que son utilizados para describir ubicaciones referenciadas o que son aludidos en alguna forma en los documentos. Las coordenadas geográficas de los lugares pueden ser usadas para proporcionar un mejor acceso a aquellos documentos que traten acerca de estas ubicaciones. Las coordenadas geográficas tienen numerosas ventajas sobre los nombres.
- Son persistentes sin importar el nombre, división política u otras modificaciones. Una ubicación geográfica especificada por coordenadas no es dependiente de las peculiaridades políticas, guerras, o sinonimia.
- Estas pueden estar simplemente conectadas a interfaces de exploración espacial y datos de GIS. Entonces, las ubicaciones basadas en coordenadas, o la representación de los documentos asociados con estas ubicaciones, puede ser desplegada y superpuesta sobre mapas digitales.
- Proporcionan un marco de trabajo consistente para aplicaciones GIS y consultas espaciales. Considerando las coordenadas geográficas para un objeto (si es especificado como un punto o como una región poligonal) como entradas indexadas permitirán realizar consultas espaciales precisas o aproximadas en la base de datos, utilizando todos los tipos de búsqueda espacial anteriormente descritas.

- Uno de los retos actuales consiste en proporcionar indexado automático confiable para ubicaciones geográficas, basándose en los nombres que ocurren en un texto.

### **3.8 Análisis Espacial**

El análisis espacial (AE) incluye cualquiera de las técnicas formales las cuales estudian entidades usando sus propiedades topológicas, geométricas y geográficas [29]. El AE, es con frecuencia referido como modelado. Se refiere al análisis de fenómenos distribuidos en el espacio y que tengan dimensiones físicas (la ubicación de, cercano a, o la orientación de un objeto respecto a otro; relación con un área de un mapa referenciado o relacionado a una ubicación específica en la superficie del planeta).

Es también, el proceso de extraer o crear nueva información acerca de un conjunto de objetos geográficos para realizar una inspección de rutina, evaluación, valoración, análisis o modelado de datos en un área geográfica basado en criterios preestablecidos y criterios computacionales, así como estándares. El análisis espacial es además, un modelo de interpretación de resultados útiles para evaluación de disponibilidad y capacidad, para estimación y predicción, y para entendimiento.

En GIS, existen cuatro tipos de análisis espacial: análisis por sobreposición y contigüidad, análisis de superficies, análisis lineal, y análisis raster. Este incluye funciones GIS como sobreposición topológica, generación de buffers, y modelado de redes [34]. Un ejemplo de análisis espacial se presenta en la siguiente situación: un estudio de salud podría describir la ubicación espacial de humanos con un punto colocado en donde ellos viven, o con un punto situado donde ellos trabajan, o usando una línea para describir sus recorridos semanales.

#### **3.8.1 Relaciones Espaciales**

Las relaciones espaciales se definen a partir de las primitivas de representación (puntos, líneas y polígonos) es decir que existen relaciones espaciales que son válidas entre las diferentes combinaciones de estas primitivas. Por ese motivo, en investigaciones previas se

definen las relaciones espaciales para primitivas de tipo línea y regiones, puntos y líneas; en donde destacan las siguientes: “dentro de”, “fuera de”, “intersecta”, “dentro”, “sobreposición” e “inclusión”. Todas estas aparecen en [63], [64], [65].

En otro orden de ideas, las teorías generales para relaciones espaciales se trataron con mucha frecuencia en la literatura de GIS desde [66], [67], [68], [69], [70]. Por ejemplo Boyle et al [66], señalaron la carencia de una teoría general como uno de los mayores problemas para el desarrollo de GIS. Mientras que Abler [67] señaló que había conceptos espaciales fundamentales que no habían sido formalizados matemáticamente.

Abundando que las relaciones cardinales ( norte, sur, este, oeste) son conceptos relativos, como en la geografía son las ideas básicas para “cerca”, “lejos”, “adyacente”, “a la izquierda”, “a la derecha”, “dentro”, “afuera”, “abajo”, “arriba”, “encima”, y “debajo”. También cabe señalar que la mayoría de las categorizaciones de relaciones espaciales distinguen entre relaciones topológicas, tales como sobreposición y relaciones métricas (distancia y direcciones) [71], [72]. Las relaciones espaciales frecuentemente son utilizadas como, relaciones de dirección y relaciones de distancia.

En esta tesis tomaremos como base el modelo: *Dimensionally-Extended 9 Intersection Matrix (DE-9IM)*. El cual es una especificación de la OGC. La Tabla 3-1 describe estas relaciones y el significado asociado a ellas.

**Tabla 3-1.** Relaciones espaciales entre objetos espaciales representados por una geometría.

Método	Descripción
(Igual) Equals	Las Geometrías de los objetos espaciales son iguales topológicamente.
(Disjunto) Disjoint	Las Geometrías de los objetos espaciales no tienen un punto en común.
(Intersecta) Intersects	Las Geometrías de los objetos espaciales tienen al menos un punto en común (la inversa de disjunto).
(Contacto) Touches	Las Geometrías de los objetos espaciales tiene al menos un borde en común pero no puntos interiores.
(Cruza) Crosses	Las Geometrías de los objetos espaciales comparten algunos pero no todos los puntos interiores, y la dimensión de la intersección es menor que la de alguna de las geometrías de los objetos espaciales.
(Dentro) Within	Geometría del objeto A está en el interior de la Geometría del objeto B.
(Contiene) Contains	Geometría del objeto B esta en el interior de la Geometría del objeto A (la inversa de Within)
(Sobrepone)Overlaps	Las Geometrías de los objetos espaciales comparten algunos, pero no todos los puntos en común, y la intersección tiene la misma dimensión que las Geometrías de los objetos espaciales.

Otro aspecto importante es que los datos geográficos no existen aislados en el espacio, es decir, tan importante es localizarlos como descubrir y representar las relaciones entre los diferentes datos. Esto se muestra en la

**Tabla 3-2** Ejemplos de análisis espacial

Análisis	Pregunta General	Ejemplo
Condición	¿Qué es lo que está ...?"	¿Cuál es la población de esta ciudad?
Localización	¿Dónde está...?	¿Cuáles áreas tienen una declividad superior a 20%?
Tendencia	¿Qué cambió...?	¿Esta tierra era productiva hace 5 años?
Derrotero	¿Por dónde ir.. ?	¿Cuál es la mejor ruta para el metro?
Patrones	¿Cuál es el patrón....?	¿Cómo es la distribución del dengue en Europa?
Modelos	¿Qué ocurre si...?	¿Cuál es el impacto en el clima si se desforesta el Amazonas?

Consideremos un ejemplo concreto para explicar los conceptos de análisis espacial (en este caso realizado manualmente). En 1854, Londres estaba sufriendo una grave epidemia de cólera, enfermedad sobre la cual en aquella época no se conocía la vía de contaminación. Entonces, cuando ya habían ocurrido más de 500 muertes, el doctor John Snow tuvo repentinamente una idea: colocar en el mapa de la ciudad la localización de los enfermos de cólera y de los pozos de agua (en aquella época la fuente principal de agua de los habitantes de la ciudad). El mapa obtenido es mostrado en la Figura 3-11(E. Tufte, 1983).



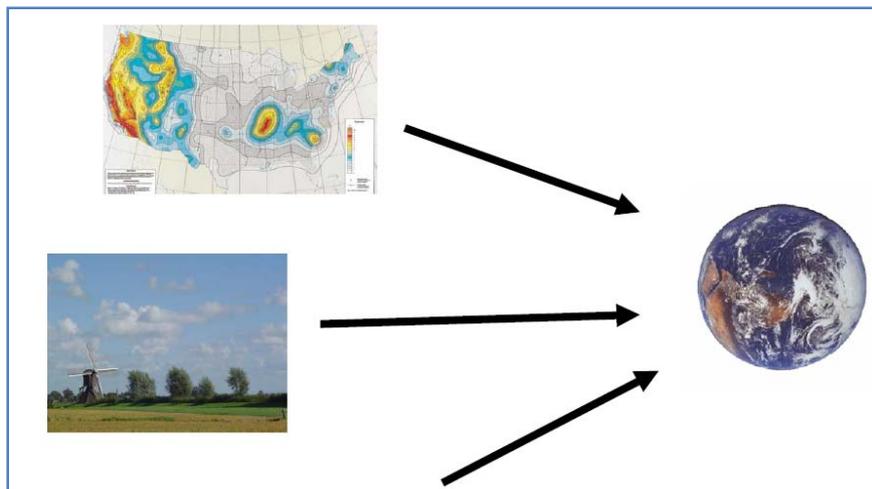
**Figura 3-11** Mapa de Londres con casos de cólera (puntos) y pozos de agua (cruces).

Con la distribución espacial de los datos, se percibió que la mayoría de los casos estaban concentrados en torno del pozo de la “Broad Street” y se ordenó que fuera lacrado, lo que contribuyó de forma notable para revelar las causas de la epidemia. Este caso suministró evidencia empírica para la hipótesis (después comprobada) de que el cólera es transmitido por la ingestión de agua contaminada. Esta es una situación típica, donde la relación espacial entre los datos difícilmente sería inferida por un simple listado de los casos de cólera y de los pozos. El mapa del doctor Snow pasó para la Historia como uno de los primeros ejemplos que ilustra el poder explicativo del análisis espacial.

### 3.8.2 Semántica Espacial

Aunque existen numerosos trabajos que hacen uso del término semántica espacial, aun no existe un consenso para establecer una definición formal y que en consecuencia sea aceptada por la comunidad científica en general. Es por ello que describiremos algunos ejemplos y definiciones de los trabajos relacionados con semántica espacial para definirla.

En sentido amplio, la semántica espacial se encarga del significado espacial y/o geográfico de los datos. Por ejemplo, el nombre de un objeto geográfico y sus posibles representaciones (fotos, datos vectoriales, imágenes raster, etc). Así como las relaciones que tienen con otros objetos (si están cerca en términos de distancia o tiempo, si comparten un punto, si una carretera los conecta, etc.) La Figura 3-12 es un ejemplo de lo anteriormente descrito y gráficamente representa lo que se entiende por semántica espacial. En donde la geografía del planeta Tierra es representada usando un mapa utilizando una fotografía, ambas representan el mismo objeto geográfico (el planeta) pero la semántica espacial que se despliega es diferente (en el mapa se enfatizan las temperaturas de ciertas regiones, mientras en la fotografía se observan áreas verdes y árboles de uno de estas regiones).



**Figura 3-12** Semántica de la información espacial.

Por otra parte, están las denotaciones que tienen esos datos, es decir la semántica se describe a partir de metadatos con respecto a descripciones, atributos y aspectos espaciales de los objetos geográficos. También, es comúnmente citada la estadística que hasta el 80% de la información está referenciada espacialmente en alguna forma. Sin embargo, los aspectos espaciales de mucha de esa información no son utilizables para la computación debido a su inescrutable semántica (no está expresada o es expresada informalmente).

Además de las limitaciones de los razonadores y lenguajes formales para representar la semántica espacial. Ya que actualmente, descubrir, procesar, analizar y visualizar datos geográficos, aun requiere expertos que puedan entender el significado de los datos espaciales y que puedan cumplir con dichas tareas de forma inteligente.

### **3.8.3 La Web Semántica Geoespacial**

El concepto de la Web Semántica Geoespacial [17] es un término que surgió a raíz del interés de la comunidad científica de GIS, por la Web y la Web Semántica.

La Web semántica espacial promete mejores métodos de recuperación al incorporar semántica a los datos y explotándola durante el proceso de búsqueda. Este desarrollo requiere especial atención desde la perspectiva geoespacial, para garantizar que las particularidades del significado geoespacial sean capturados apropiadamente. La creación de la “Semantic Geospatial Web” necesita el desarrollo de múltiples ontologías espaciales y de términos, cada una de ellas con semántica formal. La representación de esta semántica debe ser tal que esté disponible para procesamiento en las computadoras, y para que las personas la entiendan.

Y el procesamiento de las consultas espaciales, se debe hacer contra estas ontologías, mientras que la evaluación de los resultados recuperados basados en la coincidencia (*matching*) entre la semántica de la necesidad de información expresada (consulta) y la semántica disponible de los recursos de información (estructura y organización) y sistemas de búsqueda (e.g. Google).

Esto conducirá a un nuevo marco de trabajo para la recuperación de información espacial, el cual está basado en la semántica de ontologías espaciales y de términos. A través de representar explícitamente el rol de la semántica en los diferentes componentes del proceso de recuperación (personas, interfaces, sistemas de búsqueda, y fuentes de información). Finalmente, la perspectiva es que la Web Semántica Geoespacial permitirá a los usuarios

recuperar de forma más precisa los datos que requieren, basándose en la semántica asociada con estos datos.

### **3.9 Redes semánticas**

En esta sección comentaremos aspectos teóricos y trabajos de investigación relacionados con redes semánticas. Estos trabajos e investigaciones representan el primer antecedente del área de recuperación de información por medio del procesamiento semántico de la información. Además, estas investigaciones dieron origen al procesamiento semántico de la información en estructuras de datos que contienen relaciones semánticas. Estos trabajos están relacionados con la presente tesis porque muestran diversas formas de procesar relaciones semánticas. La diferencia principal es que, en esta tesis, las relaciones semánticas se almacenan en ontologías.

Comenzaremos con el trabajo descrito en [102], en el cual la investigación se enfoca en que los contenidos de la memoria humana sean organizados y accedidos en términos de relaciones de significado. En este sentido, cualquier tipo de relación de significado puede enlazar dos o más entradas en el sistema y los enlaces tienen etiquetas que pueden ser cambiadas. Esto se considera legal para una secuencia de enlaces recursivos. Donde las “esferas de activación” conducen desde diferentes entradas que pueden intersectar y unificar actividades previas para crear patrones complejos de activación. Este modelo puede considerarse como conexionista con localización automática de entradas nuevas.

Por otra parte, el término “*Semantic Memory*”<sup>36</sup> es utilizado para describir un tipo de memoria con un cierto grado de conciencia. Algunas preguntas que originaron esta investigación son: ¿Cómo es organizada la información semántica en la memoria de las personas? En otras palabras, ¿Qué orden de formato de representación puede permitir el significado de las palabras? Para responder estas preguntas se propone un modelo de la estructura de la memoria, denominado *Modelo de Memoria*, el cual consiste de una masa de nodos interconectados por diferentes tipos de enlaces asociativos. Cada nodo puede ser

---

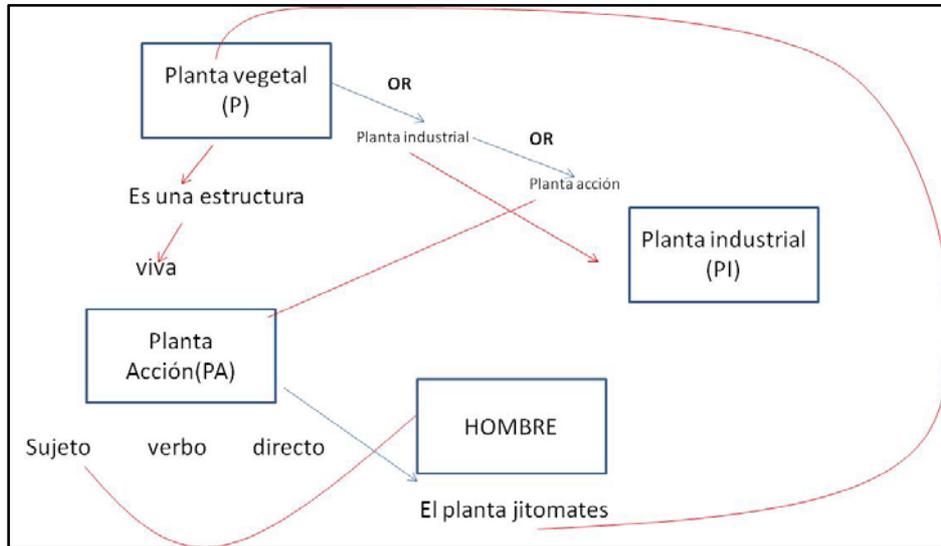
<sup>36</sup> “*Semantic Memory*” traducción al español “*Memoria Semántica*”.

pensado como el nombre de una palabra en algún idioma. La característica más importante es que un nodo puede estar relacionado al significado (concepto) del nombre de la palabra en alguna de las dos formas siguientes.

El nodo se relaciona directamente. Por ejemplo, sus enlaces asociativos pueden conducir directamente en una configuración de otros nodos que representan el significado del nombre de la palabra (*nodo tipo*). El nodo en la memoria se refiere indirectamente al concepto de la palabra, teniendo un tipo especial de enlace asociativo que apunta al nodo del tipo de concepto (*nodo simbólico*). Para cualquier significado de una palabra debe haber exactamente sí y solo sí un tipo en la memoria. Sin embargo, existirán muchos nodos simbólicos esparcidos a través de la memoria, cada uno con un apuntador al mismo *nodo tipo* único para el concepto.

En el *Modelo de Memoria*, los componentes utilizados para construir un concepto son representados por los nodos simbólicos que llaman a otros conceptos; mientras que el significado de la configuración del concepto es representado por la estructura particular de inter-enlaces de conexiones de *nodos simbólicos* a cada uno. Esto se realiza para pensar en la configuración de los nodos simbólicos inter-enlazados que representan un concepto único como incluir un *plano* en la memoria. Cada nodo simbólico en la memoria entera se ubica en algún plano, tal que tiene un apuntador a su enlace asociativo especial “fuera del plano” a su nodo tipo y otros apuntadores de enlaces asociativos dentro del plano a otros nodos simbólicos que incluyen la configuración.

En la Figura 3-13 se muestran los planos de tres conceptos de palabras correspondientes a tres significados de “*planta*”. Las palabras “*planta vegetal*”, “*planta industrial*” y “*planta (verbo)*” posicionadas en la parte superior de los tres planos representan los *nodos tipo*. Cada otra palabra localizada en algunos de los planos de la Figura 3-13 representa un nodo simbólico. Las flechas que no terminan desde los *tokens* indican que cada uno tiene apuntador especial que lo conduce a su plano y a su tipo de definición. Por ejemplo, para un *nodo tipo* situado en la parte superior de su propio plano, existe un lugar en alguna parte de la memoria. Cada uno de estos planos, es por sí mismo completo.



**Figura 3-13** Representación de tres planos para el significado de *Planta*

La estructura general del modelo de memoria consiste en planos, cada uno elaborado a partir de un *nodo tipo* y un número de nodos simbólicos. Es necesario determinar el formato de los nodos por sí mismos y las variabilidades específicas de enlaces asociativos entre los nodos a ser utilizados dentro del plano.

La restricción más importante determina el crecimiento de aseveraciones necesarias para continuar considerando las propiedades de la memoria semántica de los seres humanos. En este caso, el modelo debe estar disponible para enlazar nodos junto con las configuraciones que han variado y enriquecido las ideas expresadas en lenguaje natural. Este modelo debe representar toda la información en una forma lo suficientemente estandarizada para permitir procesar reglas que puedan ser especificadas explícitamente, sino éste no será manejable como una teoría de memoria para un lenguaje por sí misma.

Para concluir, la organización de una memoria será útil para auxiliar en las tareas de rendimiento semántico y constituir además, una descripción razonable de la organización en general de la memoria humana para un material semántico. Por lo tanto, con este tipo de

modelo, se propone una forma para estructurar el conocimiento y almacenarlo a largo plazo, en donde se propone que los conceptos sean vistos como planos.

En otro trabajo [103], se describe que es posible realizar una inferencia basada en una estructura mental a través de reacciones en el tiempo. En este caso, la técnica de procesamiento de redes semánticas puntualiza los significados de las palabras que están embebidas en las redes de otros significados. El conocimiento es validado y adquiere significado a través de la correlación con otro conocimiento; en donde el tiempo puede jugar un rol importante. Los enlaces entre la información en una red semántica son propósitos cualitativos, por lo tanto estos enlaces adquieren un valor semántico.

En este modelo los conceptos son representados por nodos que están interconectados dentro de la red. Los nodos son accedidos cuando se escuchan y entonces se activan en la memoria, causando información que está correlacionada al concepto a ser seleccionado. El enlace “*is-a*” es el más común en el modelo de red semántica. Los nodos dentro de la red que están conectados por este enlace tienen un tipo específico de relación; es decir, una jerarquía. Por lo tanto, el concepto en el nodo de nivel más bajo es una forma/tipo del concepto en el nodo de nivel superior.

Mientras que en [104] una red semántica es una notación gráfica que se utiliza para representar conocimiento mediante patrones de nodos interconectados por arcos. Éstas son consideradas representaciones gráficas declarativas que pueden ser orientadas para representar conocimiento o para sistemas de soporte automático enfocados al razonamiento acerca del conocimiento. Algunas versiones de estas representaciones son algo informales; sin embargo, otras están definidas en sistemas de lógica. De acuerdo a [104] existen seis tipos de redes semánticas: redes de definición, redes de aseveración, redes de implicación, redes de ejecución, redes de aprendizaje, y redes híbridas.

Por otra parte, una *red semántica* o *esquema de representación en red* es una forma de representación de conocimiento en que las interrelaciones entre diversos conceptos o elementos semánticos le dan la forma de un grafo. Estas redes pueden ser visualizadas

como grafos, aunque algunas veces pueden ser también árboles. Las redes semánticas pueden ser mapas *conceptuales*<sup>37</sup> y *mentales*<sup>38</sup>.

En un grafo o red semántica, los elementos semánticos se representan por **nodos**. Dos elementos semánticos, entre los que se admite se da una relación semántica que representa la red, estarán unidos mediante una línea, flecha o enlace o **arista**. Cierta tipo de relaciones no simétricas requieren *grafos dirigidos* que usan flechas en lugar de líneas.

En este sentido, dado un conjunto de conceptos, elementos semánticos o términos relacionados mediante alguna relación semántica. Estas relaciones se representan en forma de grafo. Explícitamente, dado un conjunto de términos  $\{t_1, t_2, \dots, t_n\}$  y cierta relación semántica simétrica entre ellos se construye un grafo  $G(V, A)$ , cumpliendo las siguientes condiciones:

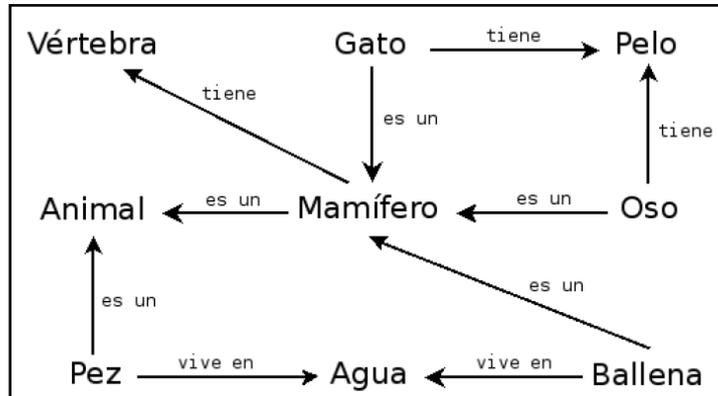
El conjunto  $V$  es el conjunto de vértices o nodos del grafo. Este conjunto estará formado por  $n$  elementos (tantos vértices como términos relacionables). Cada uno de los vértices del grafo representará uno de los términos, por lo tanto los vértices del grafo se llamarán  $t_1, t_2, \dots, t_n$ .

El conjunto  $A$  es el conjunto de aristas o líneas del grafo. Dados dos vértices (términos) del grafo  $t_i$  y  $t_j$  existirá una línea  $ij$   $a$  que une los vértices  $t_i$  y  $t_j$  sí y solo sí los términos  $t_i$  y  $t_j$  están relacionados. Si la relación no es simétrica, entonces se usan grafos dirigidos para representar la relación. Por otro lado, las redes semánticas se consideran un importante tipo de diccionario legible para las computadoras. La Figura 3-14 muestra un ejemplo de una red semántica.

---

<sup>37</sup> Un *mapa conceptual* es una técnica utilizada para la representación gráfica del conocimiento. Se considera una red de conceptos; en donde los nodos representan conceptos y los enlaces las relaciones entre éstos en forma de flechas etiquetadas.

<sup>38</sup> Un *mapa mental* es una representación abstracta de algo, en donde el sistema cognitivo del ser humano realiza una generalización y esquematiza los elementos o conceptos esenciales de ese entorno.



**Figura 3-14** Un ejemplo de red semántica

Las redes semánticas han sido muy utilizadas en inteligencia artificial para representar el conocimiento, con una gran diversificación de técnicas. Los elementos básicos de los esquemas de redes son: Estructuras de datos en nodos que representan conceptos, unidos por arcos que representan las relaciones entre los conceptos. Un conjunto de procedimientos de inferencia que operan sobre las estructuras de datos. También se pueden distinguir tres categorías de redes semánticas: redes is-a, grafos conceptuales, y redes de marcos. Las relaciones semánticas más importantes que se pueden encontrar en este tipo de estructuras son las siguientes:

- **Meronomía** (A es parte de B, i.e. B tiene A como parte de él mismo)
- **Holonomía** (B es parte de A, i.e. A tiene a B como parte de ella misma)
- **Hiponimia** (o troponimia) (A es un subordinado de B; A es del tipo de B)
- **Hiperonimia** (A es superordinado de B)
- **Sinonimia** (A denota lo mismo que B)
- **Antonimia** (A denota lo opuesto de B)

Por otra parte, diversos tipos de relaciones entre arcos pueden combinarse dentro de una misma red semántica. Por lo tanto, se puede concluir que existen ciertos tipos de arcos, entre los cuales están:

**Es-un.** Este tipo de arco se utiliza para identificar que un cierto nodo pertenece a una clase mayor de objeto.

**Tiene-un.** Este tipo de arco se utiliza para identificar que un cierto nodo tiene o pasa una cierta característica, atributo o propiedad.

Se puede concluir que a través de la propiedad de herencia, las redes semánticas tienen la capacidad de inferir conocimiento. De acuerdo con la descripción realizada anteriormente, es posible representar descripciones lógicas usando redes semánticas tal como los grafos conceptuales propuestos por [105]. Este tipo de estructuras tienen un alto poder de expresividad igual o mayor que la lógica de predicados de primer orden, de acuerdo con el análisis realizado en dicho trabajo.

### 3.10 Jerarquías y Teoría de Confusión

En este punto, se describen los conceptos de jerarquías que son utilizados como base para definir la Teoría de Confusión. En [19, 20] se establecen para el concepto de jerarquías las siguientes definiciones:

**Definición (Conjunto de elementos).** Un conjunto  $E$  cuyos elementos se definen explícitamente.

*Ejemplo:* {rojo, azul, verde, naranja, amarillo}.

**Definición (Conjunto ordenado).** Un conjunto de elementos cuyos valores están ordenados por la relación  $<$  (“menor que”)

*Ejemplo:* {muy frío, frío, templado, tibio, caliente, muy caliente}.

**Definición (Cubrimiento).**  $K$  es un cubrimiento del conjunto  $E$  si  $K$  es un conjunto de subconjuntos  $e_i \subset E$ , tal que  $\cup_i e_i = E$ .

Cada elemento de  $E$  está en algún subconjunto  $e_i \in K$ . Si  $K$  no es un cubrimiento de  $E$ , podemos hacerlo añadiendo un nuevo  $e_j$ , denominado “otros”, que contiene todos los elementos restantes de  $E$  que no pertenecen a ningún  $e_i$  previo.

**Definición (Conjunto exclusivo).**  $K$  es un conjunto exclusivo si  $e_i \cap e_j = \emptyset$ , para cada  $e_i, e_j \in K$

Sus elementos son mutuamente exclusivos. Si  $K$  no es un conjunto exclusivo, podemos hacerlo reemplazando cada par que se traslape  $e_i, e_j \in K$  con tres:  $e_i - e_j$ ,  $e_j - e_i$ , y  $e_i \cap e_j$ .

**Definición (Partición).**  $P$  es una partición del conjunto  $E$ , si  $P$  es cubrimiento y conjunto exclusivo de  $E$ .

**Definición (Variable cualitativa).** Una variable uni-valuada que toma valores simbólicos.

Su valor no puede ser un conjunto. Por simbólico entendemos que es cualitativo, opuesto a numérico, o variables cuantitativas.

**Definición (Valor Simbólico).** Un valor simbólico  $v$  representa al conjunto  $E$ , descrito por  $v \in E$ , si  $v$  puede considerado un nombre o una representación de  $E$ .

**Definición (Jerarquía).** Para un elemento del conjunto  $E$ , una jerarquía  $H$  de  $E$  es otro conjunto de elementos, donde cada  $e_i$  es un valor simbólico que representa a cualquier elemento de  $E$  o una partición; y  $\cup_i \{ \rho_i \mid \varepsilon_i \in \rho_i \} = E$  (La unión de todos los conjuntos representados por los  $e_i$  de  $E$ ).

Suponga una **Jerarquía H1**, Para un conjunto  $E = \{\text{Canada, USA, México, Cuba, Puerto\_Rico, Jamaica, Guatemala, Honduras, Costa\_Rica}\}$ , la jerarquía  $H_1$  es  $\{\text{Norte\_America, Islas\_Caribe, America\_Central}\}$ , donde  $\text{Norte\_America} \in \{\text{Canada, USA, Mexico}\}$ ;  $\text{Islas\_Caribe} \in \{\text{Islas\_Hispano-parlantes, English\_Speaking\_Island}\}$ ;  $\text{Islas\_Hispano-parlantes} \in \{\text{Cuba, Puerto\_Rico}\}$ ;  $\text{English\_Speaking\_Island} \in \{\text{Jamaica}\}$ ;  $\text{America\_Central} \in \{\text{Guatemala, Honduras, Costa\_Rica}\}$ .

Las jerarquías facilitan la comparación de los valores cualitativos que pertenecen a ella.

**Definición (Variable Jerárquica)** Una variable jerárquica es una variable cualitativa cuyos valores pertenecen a la jerarquía (El tipo de dato de una variable jerárquica es jerarquía).

*Ejemplo:* lugar\_de\_origen toma valores de  $H_1$ .

**Definición** Se utilizará la siguiente notación: a) **padre\_de** ( $v$ ). En un árbol que representa una jerarquía, el nodo **padre\_de** un nodo es aquel del cual desciende; b) los **hijos\_de** ( $v$ ) son los valores que cuelgan de  $v$ . Los nodos con el mismo padre son **hermanos**; c) **abuelo de, hermano\_de, tío\_de, ascendentes, descendientes...** son definidos, cuando ellos existen; d) La **raíz** es el nodo que no tiene padre.

### 3.10.1 Jerarquía simple, ordenada, porcentual y mixta

Una jerarquía describe la estructura de valores cualitativos en un conjunto  $E$ . Son definidas las siguientes jerarquías:

**Definición (Jerarquía Simple):** Una jerarquía simple (normal) es un árbol con raíz  $E$  y si un nodo tiene un hijo, éste forma una partición del padre.

Una jerarquía simple describe una jerarquía donde  $E$  es un conjunto (así sus elementos no se repiten ni son ordenados).

*Ejemplo:* ser vivo {animal {mamífero, pez, reptil, otro animal}, planta {árbol, otra planta}}.

**Definición (Jerarquía Ordenada):** En una jerarquía ordenada, los nodos de algunas particiones obedecen una relación de orden.

*Ejemplo* objeto {diminuto, pequeño, mediano, grande}

**Definición (Jerarquía Porcentual):** En una jerarquía porcentual, el tamaño de cada conjunto es conocido.

*Ejemplo:* Continente Americano (740M) {Norteamérica (430M) {USA(300M), Canadá(30M), México(100M)} América Central (10M), Sudamérica(300M)}.

**Definición (*Jerarquía Mixta*):** Una jerarquía mixta combina los tres casos anteriores

Para cada uno de los tipos de jerarquías se define  $conf(r,s)$  como el error de usar un valor  $r$  en vez de  $s$ .

### 3.11 Teoría de Confusión

La teoría de confusión consiste en modelar la similitud entre objetos de una o de diversas jerarquías. Se basa en una medida asimétrica y dependiente del contexto denominada confusión (de usar un valor cualitativo en vez del valor esperado o correcto) [19]. El término confusión se definió para diferenciar este enfoque de otros que se orientan a medir distancias entre conceptos (i.e., simetría, medidas independientes del contexto, cercanía, similitud).

En este caso la asimetría de la confusión se da por su definición y su dependencia al contexto por la estructura jerárquica. El concepto de confusión permite definir la cercanía a la que un objeto cumple un predicado como también derivar otras operaciones y propiedades entre valores jerárquicos.

#### 3.11.1 Confusión de usar $r$ en vez de $s$ para jerarquías simples

**Definición.** Si  $r, s \in H$ , entonces la confusión de usar  $r$  en vez de  $s$ , que se escribe como  $conf(r, s)$ , es:

- $conf(r, r) = conf(r, s) = 0$ , donde  $s$  es cualquier ascendente de  $r$ ; (regla 1)
- $conf(r, s) = 1 + conf(r, padre\_de(s))$ . (regla 2)

Para medir  $conf$ , se cuentan los enlaces *descendentes* de  $r$  a  $s$ , al valor reemplazado.

### 3.11.2 Confusión de usar r en vez de s, para jerarquías ordenadas

**Definición.** Para jerarquías simples compuestas por conjuntos ordenados, la confusión de usar  $r$  en vez de  $s$ ,  $\text{conf}^\circ(r, s)$ , se define como:

- $\text{conf}^\circ(r, r) = \text{conf}^\circ(r, \text{cualquier ascendente de } r) = 0$ ;
- Si  $r$  y  $s$  son hermanos distintos,  $\text{conf}^\circ(r, s) = 1$  Si el padre no está en un conjunto ordenado; entonces,  $\text{conf}^\circ(r, s)$  es igual a la distancia relativa de  $r$  a  $s$  es y a la vez es igual el número de pasos requeridos para llegar de  $r$  a  $s$  en el orden definido, dividido entre la cardinalidad-1 del padre; (regla 3)
- $\text{conf}^\circ(r, s) = 1 + \text{conf}^\circ(r, \text{padre\_de}(s))$ .

Esto es como  $\text{conf}$  para las *jerarquías constituidas por conjuntos*, excepto que allí el error entre dos hermanos es 1, y aquí es un número  $\leq 1$ .

Por ejemplo: Temperatura = {congelado, frío, normal, tibio, caliente, ardiendo}; en esta lista  $\text{conf}^\circ(\text{congelado}, \text{frío}) = 1/5$ , mientras que la  $\text{conf}^\circ(\text{congelado}, \text{ardiendo}) = 5/5 = 1$ .

### 3.11.3 Confusión de usar r en vez de s, para jerarquías porcentuales

Considerando la jerarquía  $H$  (de un elemento del conjunto  $E$ ) pero compuesta por un conjunto desordenado en vez de un conjunto ordenado.

**Definición.** Para conjuntos desordenados, la confusión de usar  $r$  en lugar de  $s$ ,  $\text{conf}^{\circ\circ}(r, s)$ , es:

- $\text{conf}^{\circ\circ}(r, r) = \text{conf}^{\circ\circ}(r, s) = 0$ , cuando  $s$  es cualquier ascendente de  $r$ ;
- $\text{conf}^{\circ\circ}(r, s) = 1 - \text{proporción relativa de } s \text{ en } r$ . (regla 4)

### 3.11.4 Confusión de usar r en vez de s, para jerarquías mixtas

**Definición.** Para calcular  $\text{conf}''(r, s)$  en una jerarquía mixta:

- Aplicar (regla 1) para la ruta *ascendente* de  $r$  a  $s$ ;
- En la ruta descendente, usar (regla 3) en vez de (regla 2), si  $p$  es un conjunto ordenado; o usar (regla 4) en vez de (regla 2), cuando los tamaños de  $p$  y  $q$  son conocidos.

Es decir, usar (regla 4) para las jerarquías porcentuales en lugar de (regla 2). Esta definición es consistente y reduce las definiciones previas de jerarquías porcentuales, simples, ordenadas y mixtas.

### 3.11.5 El conjunto de valores que son iguales a otros, dada una confusión

**Definición.** Un valor  $u$  es igual al valor  $v$ , dentro de una confusión dada  $\varepsilon$ , escrita  $u =_{\varepsilon} v$ , ssi  $\text{conf}(u, v) \leq \varepsilon$  (Esto significa que el valor  $u$  puede ser usado en vez de  $v$ , con un error  $\varepsilon$ ).

Ejemplo: Si  $v = \text{limón}$  (Figura 3-15), entonces (a) el conjunto de valores es igual a  $v$  con confusión 0 es  $\{\text{limón}\}$ ; (b) el conjunto de valores iguales a  $v$  con confusión 1 es  $\{\text{cítrico}, \text{limón}\}$ ; (c) el conjunto de valores iguales a  $v$  con confusión 2 es  $\{\text{planta}, \text{cítrico}, \text{pino}, \text{limón}\}$ .

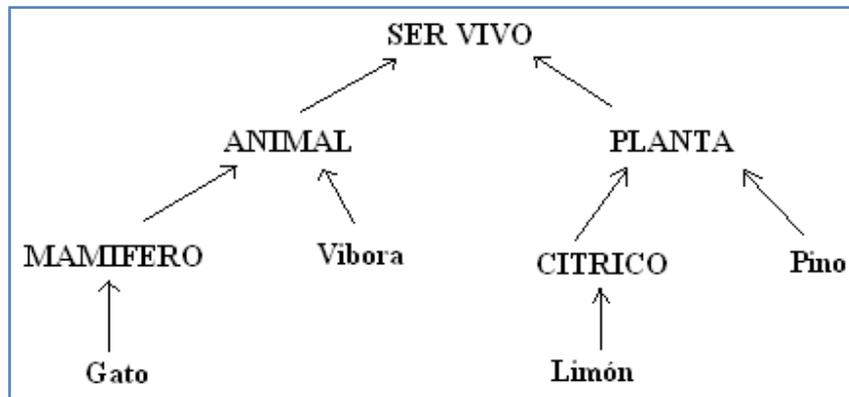


Figura 3-15 Jerarquía para un ser vivo

### 3.11.6 Objetos idénticos, muy similares, algo similares

Los objetos son entidades descritas por pares  $k$  (propiedades, valores), que en la notación se hará referencia como pares (variable, valor) (Levachkine, et al., 2007) . En bases de datos son denominados como pares (relación, atributo). Un objeto  $o$  con  $k$  pares (variable, valor) es escrita como  $(o (v_1 a_1) (v_2 a_2)... (v_k a_k))$ .

Se desea estimar el error de utilizar el objeto  $o'$  en vez del objeto  $o$ . Para un objeto  $o$  con  $k$  (tal vez jerárquica) variables  $v_1, v_2, \dots, v_k$  y valores  $a_1, a_2, \dots, a_k$ , decimos que otro objeto  $o'$  con las mismas variables  $v_1, v_2, \dots, v_k$  pero con valores  $a'_1, a'_2, \dots, a'_k$ , se presentan las siguientes definiciones:

**Definición**  $o'$  es idéntico a  $o$ , if  $a'_i = a_i$  para todas  $1 \leq i \leq k$ . Los valores correspondientes son idénticos.

Si todo lo que sabemos sobre  $o$  y  $o'$  son los valores de las variables  $v_1, v_2, \dots, v_k$ , y ambos objetos tienen valores idénticos, podemos decir que “con todo lo que sabemos,”  $o$  y  $o'$  son el mismo.

**Definición**  $o'$  es *substituto* para  $o$ , si  $\text{conf}(a'_i, a_i) = 0$  para todo  $1 \leq i \leq k$ .

No se trata de una confusión entre el valor del atributo de  $o'$  y el valor correspondiente de  $o$ . Se puede usar  $o'$  en vez del  $o$  solicitado con confusión 0.

**Definición**  $o'$  es *muy similar* a  $o$ , si  $\sum_i \text{conf}(a'_i, a_i) = 1$ .

**Definición**  $o'$  es *similar* a  $o$ , si  $\sum_i \text{conf}(a'_i, a_i) = 2$ .

**Definición**  $o'$  es *algo similar* a  $o$ , si  $\sum_i \text{conf}(a'_i, a_i) = 3$ .

**Definición** En general,  $o'$  es *similar<sub>n</sub>* a  $o$ , si  $\sum_i \text{conf}(a'_i, a_i) = n$ .

### 3.12 Comentarios Finales

En este capítulo se ofrecieron las definiciones de los conceptos involucrados en la estrategia de recuperación de esta tesis. Se mencionan conceptos relacionados con las fuentes de información, así como la terminología involucrada con las mismas. Se definen las relaciones espaciales, así como su importancia en las tareas de análisis de GIS.

Cabe destacar que el objetivo de ofrecer estas definiciones y descripciones, es reflejar la gran cantidad de riqueza semántica que implícitamente almacenan estas fuentes de datos (diccionarios geográficos, archivos vectoriales, ontologías). Donde además, para poder procesar esta semántica se debe tener un formato común que permita la extracción, consulta y búsqueda en dichas fuentes. En otras palabras, los formatos son heterogéneos lo cual complica las tareas de búsqueda y recuperación al tener que buscar transformaciones entre dichos formatos. En este sentido se propone el uso del formato XML para facilitar las tareas de búsqueda, consulta y recuperación.

Para el caso de las Ontologías Geográficas (*GeoOntologías*), se subraya que para su generación y construcción se utilizaron otras fuentes de datos. Por ejemplo, *WordNet* y de

*Wikipedia*. La elección de estas dos fuentes de datos se justifica considerando que ambas contienen un alto grado de riqueza en relaciones semánticas. Esta riqueza semántica puede emplearse en los procesos de recuperación geográfica para encontrar datos que tengan alguna relevancia geográfica.

Por lo tanto, considerando los aspectos señalados, es posible establecer que el desarrollo de una metodología que procese e integre la riqueza espacial semántica de estas fuentes de datos geográficas. Permitirá conducir a resultados de mayor relevancia en el área de la recuperación de información geográfica. Esto, motivado por la necesidad de recuperar datos geográficos que con otros enfoques son omitidos o ni siquiera son considerados. Por lo cual, se requiere interoperar u obtener datos de diferentes sistemas.

Para complicar aún más el escenario, se presenta el problema de la heterogeneidad de datos (se manejan diferentes estructuras y formatos de datos). Entonces, un mecanismo integrador de estos elementos resulta muy útil, sobre todo en países como México. Sin soslayar que existen otras alternativas de solución como las tecnologías de servicios Web semánticos y los Servicios Web Geográficos, cuyo desarrollo en México aún se encuentra en etapas iniciales, lo que limita la explotación de la información a través de estos servicios.

Es por ello, que se requieren mecanismos alternos como el que ha surgido a raíz de esta investigación y que como resultado final tiene un sistema denominado *iGIR* (*intelligent, integrated GIR*). El cual tiene como entidad principal a una Ontología, la cual conduce el proceso de recuperación con base en criterios conceptuales, topológicos y de atributos geográficos.

Adicionalmente, se describieron los fundamentos de la teoría de confusión y los elementos que la conforman. Esta sección representa el respaldo teórico para realizar la adaptación de la teoría de confusión para aplicarla en un mecanismo que permita ponderar los resultados obtenidos por el sistema *iGIR*, en otras palabras permite desplegar los resultados de acuerdo a su relevancia geográfica, conceptual y topológica.

Por último, se tienen las bases teóricas, las razones que justifican y los argumentos que sustentan el uso de un mecanismo integrador que utilice fuentes de datos geográficas y que sea conducido por una ontología geográfica (*GeoOntología*) lo cual permitirá recuperar datos y desplegarlos de acuerdo a su relevancia geográfica, lo cual es la principal meta de esta tesis.



# CAPÍTULO 4

## METODOLOGÍA DE RECUPERACIÓN

*La geografía es el integrador de la información*

*Roger Tomlinson*

Este capítulo describe la metodología seguida para desarrollar esta tesis. En particular, se explica una estrategia para recuperar información geográfica desde tres fuentes de datos heterogéneas: *GeoOntologías*, diccionarios de datos y archivos topológicos. La recuperación de información geográfica se realiza integrando los datos de estas fuentes mediante una metodología denominada *iGIR* (integración de GIR). En donde, las *GeoOntologías* utilizan un mecanismo que conduce la búsqueda en los diccionarios geográficos y archivos topológicos. Las consultas espaciales utilizadas tienen una estructura de tipo tripleta {que, relación espacial, donde}.

También, se define un algoritmo de exploración de *GeoOntologías* (*OntoExplora*) el cual actúa como un analista GIS que indica si la búsqueda de un objeto geográfico se realiza utilizando sus propiedades, relaciones topológicas o semánticas. Por ejemplo, si se buscan carreteras, la recuperación se puede hacer por el número de carriles, por las ciudades que conecta, o buscando vías de transporte. La presente tesis se enfoca en extender el enfoque sintáctico al semántico espacial para recuperar información geográfica integrando los datos de tres fuentes de información geográfica heterogénea.

Los resultados se ponderan integralmente utilizando un método denominado *iRank* el cual se explica en la sección 5.2. Dichos resultados se despliegan en formato de lista (estilo *Google*) y pueden ser descargados en cuatro formatos diferentes: KML<sup>39</sup>, *shapefiles*<sup>40</sup>, documentos Web y en *Google Maps*.

Este capítulo se estructura en la siguiente forma. Primero se explican las tres fuentes de datos geográficas. En la sección 4.1 se explica cómo se construyeron las GeoOntologías, en la sección 4.2 se describen los diccionarios geográficos y en la sección 4.3 se explican los archivos topológicos. Posteriormente, en la sección 4.4 se describe la estrategia de recuperación (*iGIR*). En la sección 4.5 se explica el algoritmo de exploración de *GeoOntologías*. Mientras que la sección 4.6 describe como se recuperan los datos en las *GeoOntologías*. En la sección 4.7 se explica la recuperación en diccionarios geográficos. La sección 4.8 detalla cómo recuperar información en los archivos topológicos. Finalmente, la sección 4.9 explica el proceso de integración de los resultados obtenidos al seguir esta metodología.

---

<sup>39</sup> (Keyhole Markup Language) es un lenguaje basado en XML. Utilizado para el despliegue de datos espaciales en 3 dimensiones en el programa Google Earth.

<sup>40</sup> shapefile formato vectorial de datos geográficos, utilizado en el software ArcGIS.

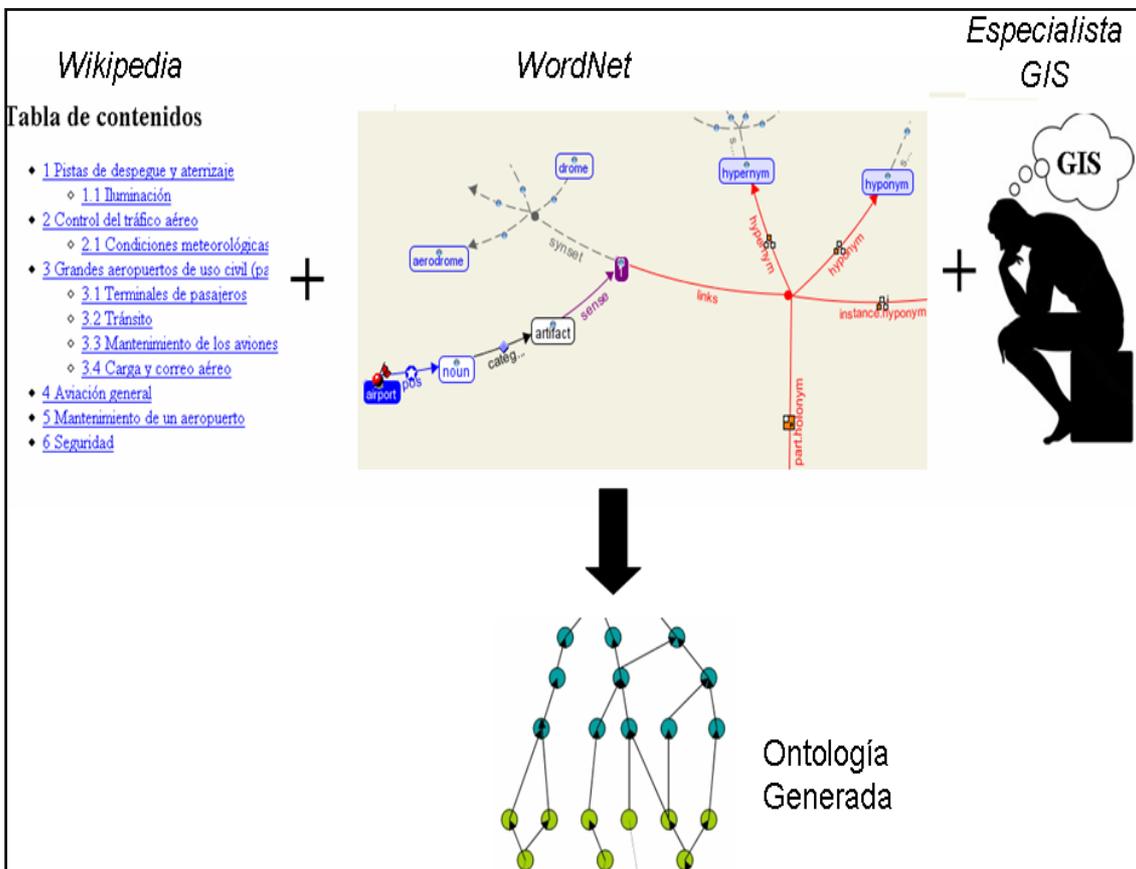
#### 4.1 Primera fuente de datos geográfica: GeoOntologías

Las ontologías geográficas (*GeoOntologías*) son la primera de las tres fuentes de datos geográficas utilizadas en esta tesis. Estas *GeoOntologías* se construyeron de forma semiautomática. Es decir, manualmente y utilizando programas de cómputo. Esta decisión, se basa en el hecho de que no hay disponibilidad de *GeoOntologías* en español para los datos geográficos utilizados en esta investigación.

Las ontologías son estructuras que representan y permiten almacenar conocimiento de acuerdo a un dominio específico [24], [45]. Entonces, con base en el análisis e investigación previa, podemos clasificar este conocimiento en tres categorías, dominio común, especializado o de propósito específico. Es decir, que una ontología puede ser definida por una comunidad científica, por una institución o gobierno, o por un grupo de personas comunes. Por ello, las *GeoOntologías* construidas para este trabajo, se apoyan en tres componentes que se mencionan a continuación:

- Documentos de *Wikipedia* (almacenan la definición de uso popular para un término que representa a un objeto geográfico) [29].
- Estructura semántica de conceptos de *WordNet* (almacena relaciones semánticas del sentido de una palabra [36]. En este caso palabras que representan un objeto espacial).
- Criterio del especialista geográfico (manejo de terminología especializada de un experto para un dominio espacial específico).

La Figura 4-1 muestra de forma gráfica, la construcción de una *GeoOntología* mediante los tres aspectos anteriores.



**Figura 4-1** Construcción de Ontología geográfica (*GeoOntología*).

Como se aprecia en la Figura 4-1, los objetos geográficos se definen con base en la terminología de uso popular, incluida en un documento de *Wikipedia*. Esta terminología, se estructura y organiza utilizando las relaciones semánticas que existen en *WordNet*. En este procedimiento, el analista GIS resuelve las discrepancias que pueden ocurrir entre *WordNet* y *Wikipedia* al definir objetos geográficos. Es decir, él es quien decide cuales conceptos y relaciones se incluyen y cuáles no, de acuerdo a su experiencia. A continuación justificaremos, de forma breve, las razones para elegir estos tres componentes como base para la construcción de las ontologías.

Los documentos de *Wikipedia* son parte de una enciclopedia de dominio popular, incluso representan un consenso entre usuarios especialistas y usuarios comunes (ver Sección 3 Marco Teórico). Mientras que *WordNet* ha organizado su información de acuerdo a relaciones semánticas para sentidos de las palabras, principalmente del idioma

inglés. Entonces, para su integración con fuentes en idioma español, se requiere realizar una tarea adicional de traducción (del inglés al español). Además de adaptaciones o adecuaciones para considerar aspectos geográficos (también expresados en el idioma español). Estas tareas, también, formaron parte de este trabajo de tesis.

La integración de estos componentes facilitará la búsqueda de información de objetos geográficos, mejorando el proceso de recuperación. Considerando criterios de especialistas (representados por la ontología); criterios semánticos (representados por relaciones de *WordNet*) y criterios de usuarios comunes (representados por *Wikipedia*). Adicionalmente, el analista GIS tiene la facultad de decidir si se incluye un concepto o no (decide la consistencia y riqueza de la información). El resultado son las *GeoOntologías* para el caso de estudio.

Las *GeoOntologías* se construyen usando archivos en formato XML. Este formato facilita la integración de información y la adición de nuevos datos [21]. Para definir la estructura del documento XML se consideraron una serie de criterios, los cuales se dividen en dos grupos: los primeros aplican a conceptos estructurados conforme a *WordNet*, mientras que el resto se aplica a los conceptos de documentos *Wikipedia*. A continuación se detallan estos aspectos:

Considerando la estructura de *WordNet*:

- Se utilizan como base las relaciones de hiperonimia, hiponimia y meronimia que ofrece *WordNet*.
- No se consideran las descripciones asociadas a cada nodo para el sentido de una palabra (son de índole sintáctica y no geográfica).
- Únicamente se consideran nodos padres de acuerdo a la relación de hiperonimia, los cuales llamaremos hiperonimos directos (es decir, el padre del concepto).

Considerando Documentos *Wikipedia*:

- Se toma como base la estructura de la tabla de contenidos para el concepto geográfico definido en el artículo de *Wikipedia*.
- De acuerdo a la relación de hiperonimia, hiponimia y meronimia, se integran conceptos y clases de los artículos de *Wikipedia*.
- Se consideran únicamente documentos que describen un objeto geográfico.
- Los vínculos cuyas etiquetas representen un aspecto geográfico o espacial, se consideran candidatos a conceptos o clases.
- Los verbos involucrados en aspectos geográficos se consideraron como candidatos a relaciones.
- Los sustantivos son considerados candidatos a conceptos.
- Los títulos y subtítulos se consideraron criterios para clasificar los conceptos.
- Si existe algún concepto o clase que se repita con uno definido en *WordNet*, entonces se elimina aquel que afecte la descripción geográfica. Esta decisión la toma el analista GIS que está construyendo la *GeoOntología*.
- Se definió una lista de sinónimos para el concepto (que representa un objeto geográfico) de acuerdo a los sinónimos que aparecen en el documento *Wikipedia*.
- La clasificación de conceptos se realiza aprovechando la estructura jerárquica que tiene un documento XML.

Para explicar cómo funciona el proceso de la generación de la ontología consideremos los objetos geográficos de la consulta  $Q_{G4} = \{\text{Vialidades conectan Aeropuertos}\}$ . La tripleta está conformada por la relación espacial {conectan}, y los objetos geográficos {Avenidas} y {Aeropuertos}. Estos objetos son buscados en *WordNet* (en forma plural o singular) para extraer las relaciones semánticas de hiperonimia, meronimia e hiponimia de {Aeropuertos} y {Vialidades}. Estas relaciones se muestran en la Tabla 4-1 y Tabla 4-2 respectivamente.

**Tabla 4-1** Relaciones semánticas extraídas de *WordNet* para el concepto Aeropuertos

HIPERONIMO	MERONIMIA	HIPONIMIA
<b>Aeropuerto,</b> <b>Aeródromo</b> => pista aterrizaje, pista de vuelo => Complejo, instalación => Artefacto => objeto => <b>Entidad</b>	<b>Aeropuerto, Aeródromo</b> Tiene parte: hangar, centro de reparaciones Tiene parte: Terminal aérea, Terminal aeropuerto Tiene parte: sala de acceso, pasillo de acceso Tiene parte: torre de control Tiene parte: plataforma de estacionamiento Tiene parte: pista de despegue Tiene parte: pista de rodaje Tiene parte: línea de vuelo	<b>Aeropuerto, Aeródromo</b> => <b>Helipuerto</b> => <b>Benito Juárez</b> <b>Aeropuerto internacional,</b> <b>aeropuerto internacional de</b> <b>la Ciudad de México</b> => objeto físico

**Tabla 4-2** Relaciones semánticas extraídas de *WordNet* para el concepto Vialidades.

HIPERONIMO	MERONIMIA	HIPONIMIA
<b>Vialidad, Boulevard</b> => Estructura => Artefacto => objeto => <b>Entidad</b>	<b>Avenida--</b> Tiene : pavimento, Tiene: sentido,	<b>Avenida</b> => <b>Boulevard</b> => <b>Avenida Insurgentes</b>

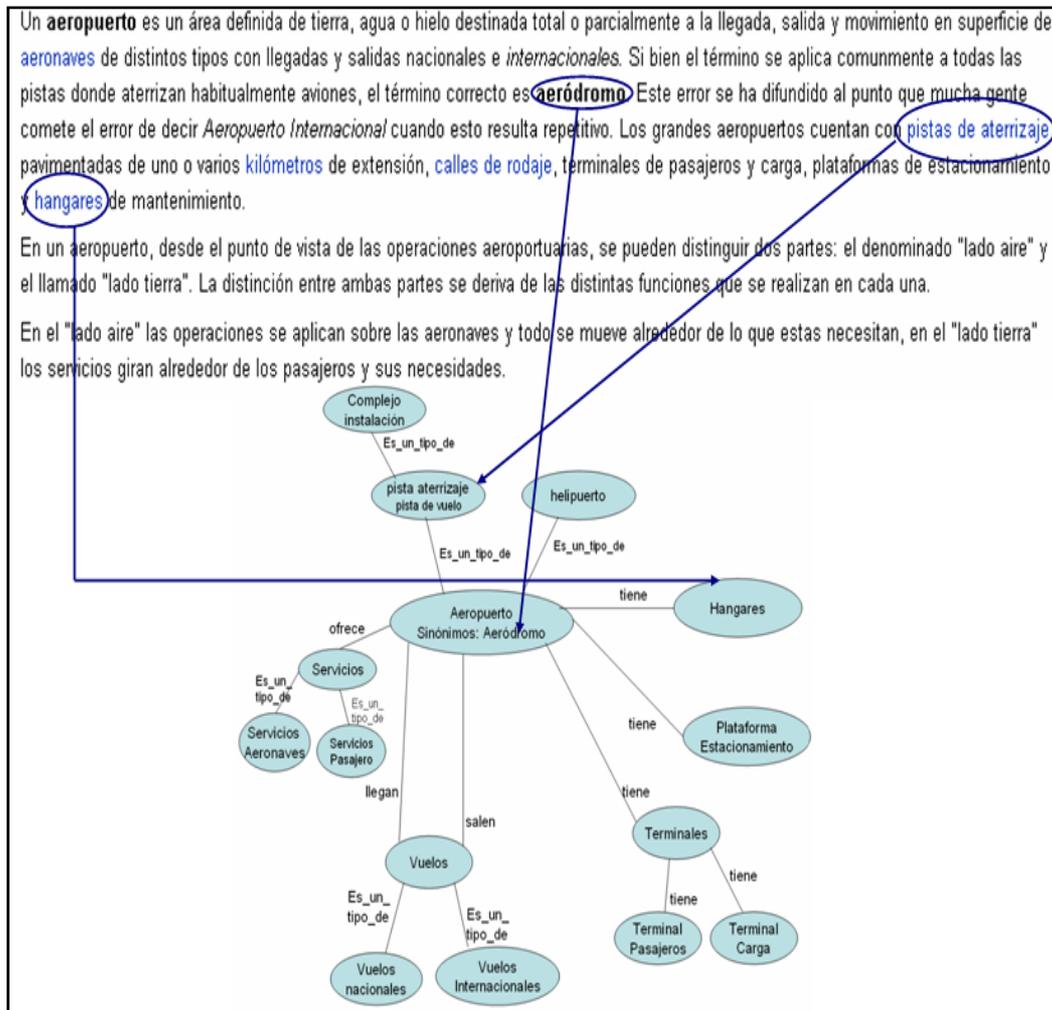
La Tabla 4-3 despliega la estructura de contenidos del documento *Wikipedia* para los conceptos {Aeropuerto} y {Vialidades}.

**Tabla 4-3** Estructura de contenidos de *Wikipedia* para el concepto Aeropuerto y Vialidades.

Tabla de contenidos para AEROPUERTO (de Wikipedia)	Tabla de contenidos para Vialidades (de Wikipedia)
<ul style="list-style-type: none"> <li>* <b>1 Pistas de despegue y aterrizaje</b> <ul style="list-style-type: none"> <li>o <b>1.1 Iluminación</b></li> </ul> </li> <li>* <b>2 Control del tráfico aéreo</b> <ul style="list-style-type: none"> <li>o <b>2.1 Condiciones meteorológicas de vuelo visual (VMC) y Condiciones meteorológicas de vuelo instrumental (IMC)</b></li> </ul> </li> <li>* <b>3 Grandes aeropuertos de uso civil (pasajeros o carga)</b> <ul style="list-style-type: none"> <li>o <b>3.1 Terminales de pasajeros</b></li> <li>o <b>3.2 Tránsito</b></li> <li>o <b>3.3 Mantenimiento de los aviones</b></li> <li>o <b>3.4 Carga y correo aéreo</b></li> </ul> </li> <li>* <b>4 Aviación general</b></li> <li>* <b>5 Mantenimiento de un aeropuerto</b></li> <li>* <b>6 Seguridad</b></li> <li>* <b>7 Sistemas de navegación/radioayudas integrados en aeropuerto</b> <ul style="list-style-type: none"> <li>o <b>7.1 VOR y L/MF</b></li> <li>o <b>7.2 ILS</b></li> </ul> </li> <li>* <b>8 Contaminación</b></li> <li>* <b>9 Aeropuertos militares (bases aéreas)</b></li> </ul>	<ul style="list-style-type: none"> <li>• <b>Historia</b></li> <li>• <b>Una vía principal de una ciudad.</b></li> <li>• <b>Una crecida excepcional de un curso de agua.</b></li> </ul>

De acuerdo a esta estructura, se enriquece la *GeoOntología* usando la jerarquía de *WordNet* y el documento de *Wikipedia* para el concepto {Aeropuertos} y para el concepto {Vialidades}.

La Figura 4-2 muestra gráficamente la *GeoOntología* generada a partir de la integración del documento *Wikipedia* de Aeropuerto y de las relaciones semánticas en *WordNet*

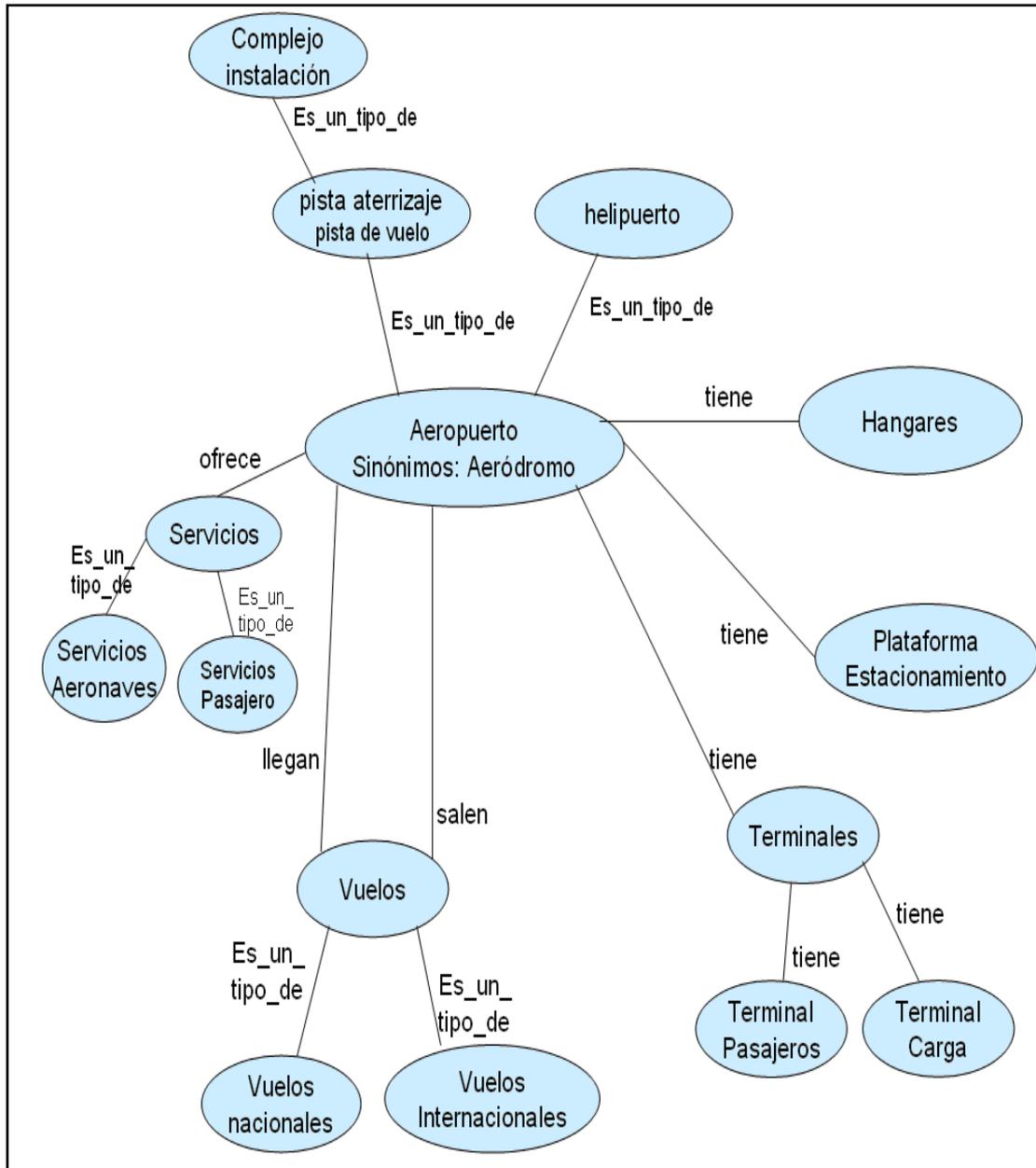


**Figura 4-2** Generación de *GeoOntología* usando *Wikipedia* y *WordNet*.

Como se puede apreciar en la Figura 4-2 se muestran las propiedades y relaciones para la clase Aeropuerto. La definición está basada en documentos *Wikipedia* (en español) los cuales contienen las descripciones populares (en lenguaje natural) que comúnmente se utilizan para objetos geográficos (e.g. Aeropuertos) por usuarios tales como estudiantes, empresarios, etc. Este proceso se realiza de forma semiautomática. Es decir, la extracción y búsqueda de documentos *Wikipedia* se realiza automáticamente mediante un programa hecho en el lenguaje Ruby (<http://www.ruby-lang.org/es/>) la extracción de información de documentos *Wikipedia*, se realiza basándose en las

relaciones semánticas de *WordNet* (donde cada término se traduce a un término equivalente en español en forma manual).

En la Figura 4-3 se muestra de forma gráfica (esta representación se hizo de forma manual) un fragmento de la *GeoOntología* que contiene el concepto aeropuerto.



**Figura 4-3** *GeoOntología* generada para aeropuerto usando *WordNet* y Wikipedia.

La *GeoOntología* mostrada en la Figura 4-3 se implementó utilizando el formato XML para permitir agregar nodos de otras fuentes y así enriquecer el conocimiento que esta representa. La Figura 4-4 muestra el fragmento de la *GeoOntología* de la Figura 4-3 en formato XML.

```

- <clase id="Aeropuerto">
  - <propiedad id="tiene">
    <clase_indiv> hangares</clase_indiv>
    <clase_indiv> Plataforma Estacionamiento</clase_indiv>
    <clase_indiv> Terminales</clase_indiv>
  </propiedad>
  - <relaciones>
    - <relacion nombre="es_un_tipo_de">
      <clase_indiv>pista aterrizaje</clase_indiv>
    </relacion>
    - <relacion nombre="es_un_tipo_de">
      <clase_indiv>helipuerto</clase_indiv>
    </relacion>
    <relacion nombre="llegan"> Vuelos</relacion>
  - <relacion nombre="ofrece">
    - <clase id="servicios">
      <propiedad/>
      - <relaciones>
        - <relacion nombre="es_un_tipo_de">
          <clase_indiv>Servicios aeronave</clase_indiv>
        </relacion>
        - <relacion nombre="es_un_tipo_de">
          <clase_indiv>Servicios pasajero</clase_indiv>
        </relacion>
      </relaciones>
    </clase>
  </relacion>
</relaciones>
</clase>

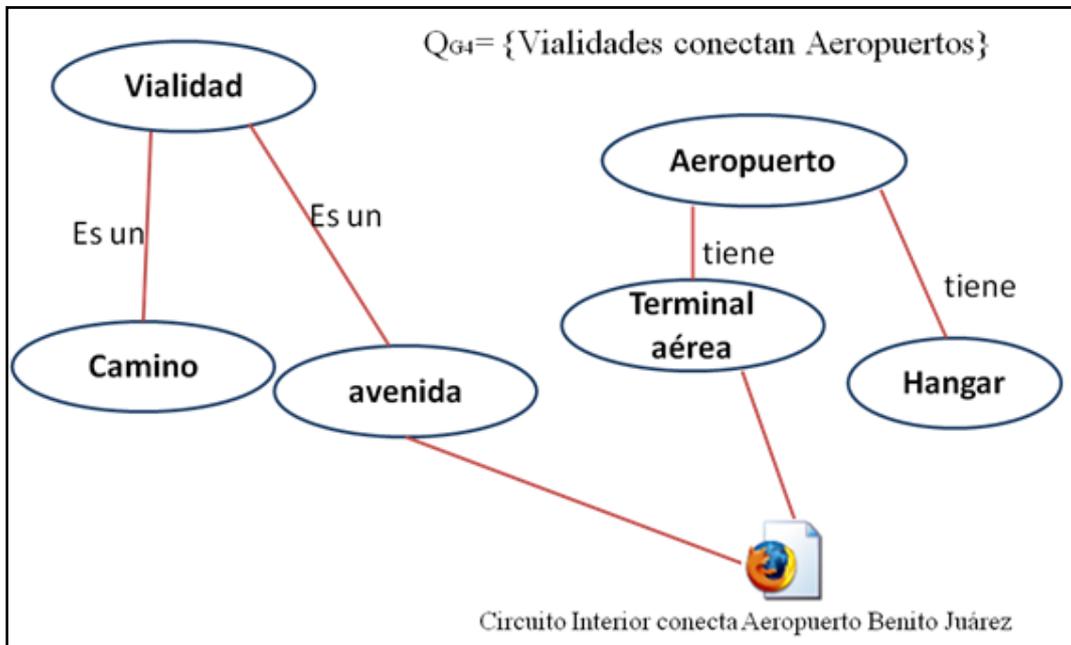
```

Figura 4-4 Fragmento de GeoOntología implementado en XML.

De esta forma es como se genera la *GeoOntología*. Para enriquecer este proceso se asociaron documentos Web a la GeoOntología, para ello se utilizó un programa (*ExploraWeb*) desarrollado en el lenguaje de programación *Ruby* [97]. Este programa extrae palabras de documentos Web que coinciden con las etiquetas de los conceptos en la *GeoOntología*. Enlazando en esta forma documentos Web con la estructura definida en la *GeoOntología*. El proceso se describe a continuación:

Consideremos la consulta  $Q_{G4} = \{\text{Vialidades conectan Aeropuertos}\}$  el programa *ExploraWeb* busca páginas Web que contengan los conceptos relacionados con estos objetos geográficos. Los cuales son obtenidos por el algoritmo *OntoExplora* (ver sección 4.5). Una vez hecho esto, todos los documentos encontrados serán asociados a los conceptos correspondientes. En este caso para  $Q_{G4}$  los documentos encontrados

(páginas Web) se asocian con los conceptos “Vialidad”, “Camino”, “Avenida” y “terminal aérea”. La Figura 4-5 muestra este ejemplo.



**Figura 4-5** Asociando documento Instancias a la GeoOntología.

De esta forma se asocian documentos Web a los conceptos que representan objetos geográficos. Estos documentos serán los resultados de la *GeoOntología* para las consultas espaciales. Esto mediante la exploración de la *GeoOntología* por el algoritmo de *OntoExplora* (el cual se explica en la sección 4.5). Así, se construyeron las GeoOntologías y los documentos asociados a éstas. Finalizando, en esta forma la descripción de la primer fuente de datos geográfica. En la sección 4.7 se explica la segunda fuente de datos, los diccionarios geográficos.

#### 4.2 Segunda fuente de datos geográfica: Diccionarios geográficos

Los diccionarios de datos están contruidos con base en las cartas cartográficas y de la bases de datos espaciales de México. Además que, los especialistas en geografía y GIS los utilizan como fuente de consulta para realizar proyectos de tipo gubernamental y privados. Adicionalmente, estos diccionarios representan un consenso por parte de la

principal institución encargada de construir y gestionar la información geográfica en México, el Instituto Nacional de Estadística e Información Geográfica, INEGI<sup>41</sup>.

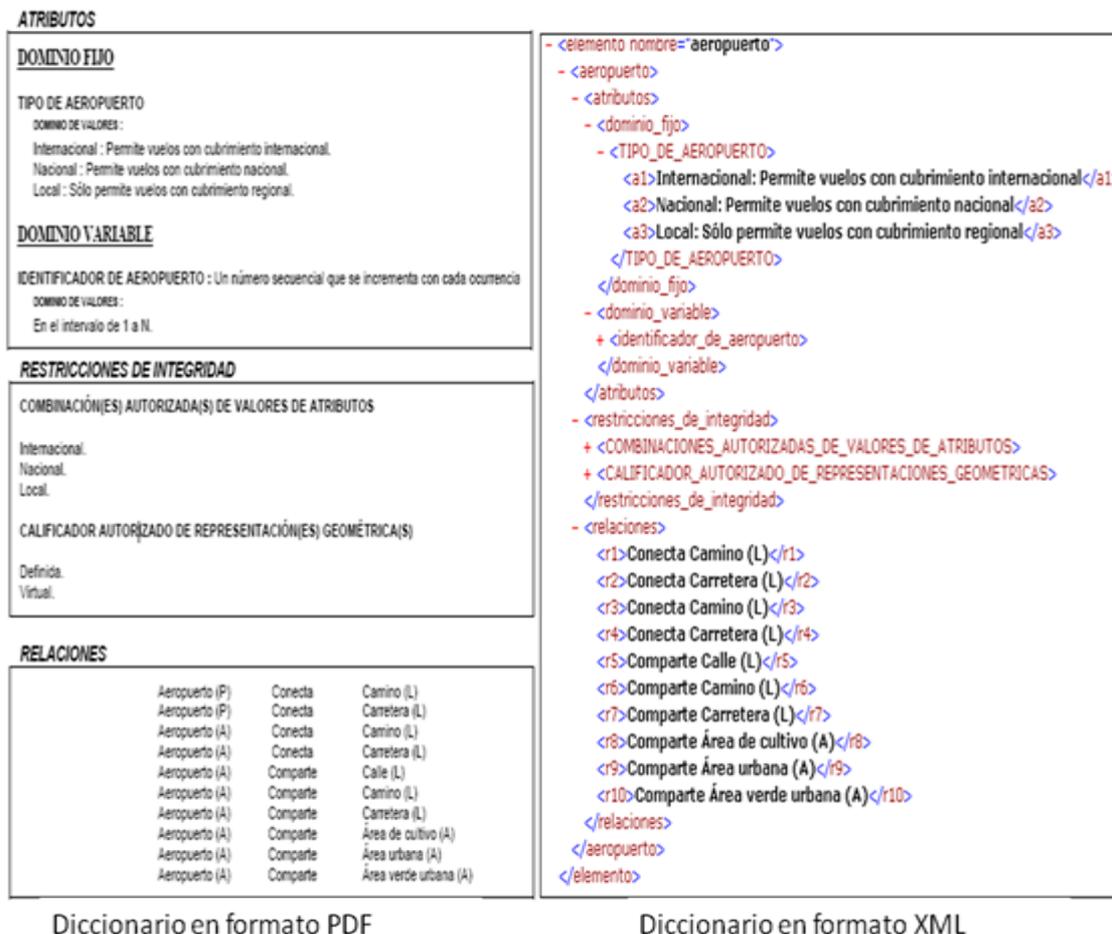
Finalmente, apuntaremos que la información almacenada en los diccionarios de datos está organizada de acuerdo a los criterios establecidos por especialistas en el área de la geografía, la cartografía y la computación. Por lo tanto, la integración de esta fuente de información permite de forma significativa establecer de forma certera los mejores parámetros de búsqueda para datos espaciales en formato vectorial o topológico.

Estos diccionarios se proporcionan en el sitio Web del INEGI ([www.inegi.gob.mx](http://www.inegi.gob.mx)), en formato PDF. Sin embargo, para propósitos de la presente tesis y para facilitar el proceso de extracción y procesamiento de estos archivos, se convirtieron a formato XML. El proceso se describe a continuación.

Cada objeto es definido, en su respectivo diccionario, de acuerdo a las siguientes categorías: dominio fijo, dominio variable, restricciones de integridad, así como las principales relaciones y sus objetos asociados. Adicionalmente, se indica el tipo de primitiva de dichos objetos (puntos, líneas, polígonos). La Figura 4-6 muestra un fragmento de la descripción para el objeto geográfico Aeropuerto en formato PDF y su correspondiente conversión a formato XML.

---

<sup>41</sup> Así como también brinda el servicio público de información estadística y geográfica sobre el territorio, la población y la economía de México



**Figura 4-6** Diccionario de datos en formato PDF y su conversión a formato XML.

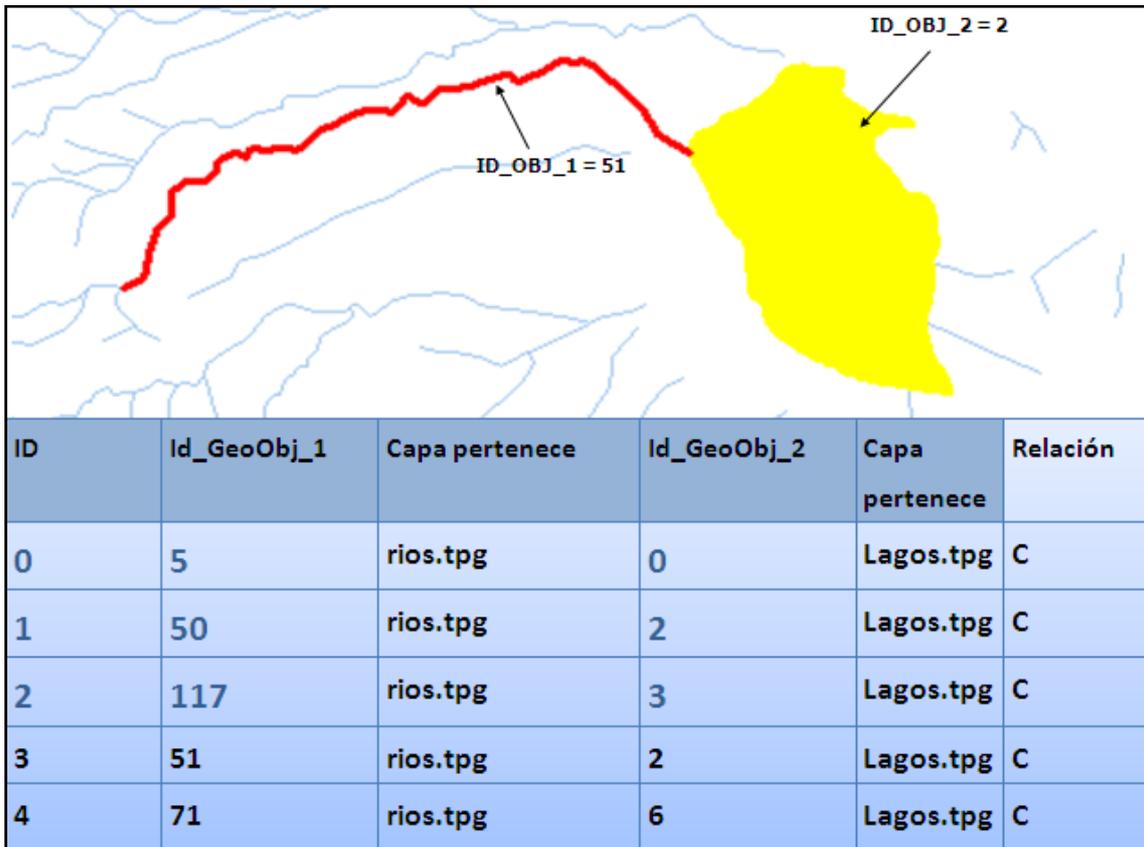
Como se observa tenemos del lado izquierdo el diccionario en formato PDF para el concepto Aeropuerto, mientras que del lado derecho se tiene el equivalente en formato XML. El proceso de conversión se realiza de forma semiautomática utilizando la API *PJX* (<http://javasource.net/open-source/pdf-libraries/pjx/>) esta API de programación permite la manipulación y exploración de un archivo PDF y utilizando la API SAX de java (<http://www.saxproject.org/>) para convertir los datos hacia XML.

Sin embargo, debido a que los archivos proporcionados en el sitio Web del INEGI, están protegidos contra escritura, se realizó un proceso manual para generar un nuevo archivo PDF, el cual contiene las secciones de interés para esta tesis, y posteriormente usando los métodos de *PJX* se construyó el respectivo XML, el cual es mostrado en la Figura 4-6.

Donde uno de los puntos a destacar, es que estos archivos almacenan relaciones de tipo topológico, utilizando sentidos de palabra del usuario neófito GIS. Por ejemplo, en la Figura 4-6 aparecen las relaciones “*Conecta*” y “*Comparte*” asociadas a objetos: camino, carretera y calle, así como área de cultivo, área urbana, y área verde urbana. Adicionalmente, se indica el tipo de primitiva (punto, línea, polígono) utilizada para representar estos objetos. Cada uno de estos apartados, se consideraron como nodos y atributos del correspondiente archivo XML. El cual, también es mostrado en la Figura 4-6. De esta forma finalizamos la descripción de cómo se utiliza y procesa la información que contienen los diccionarios geográficos. En la sección 4.3 se explica la tercera fuente de datos geográfica, los archivos topológicos. Para posteriormente, explicar la estrategia de recuperación que integra las tres fuentes de datos.

### **4.3 Tercera fuente de datos geográfica: Archivos topológicos**

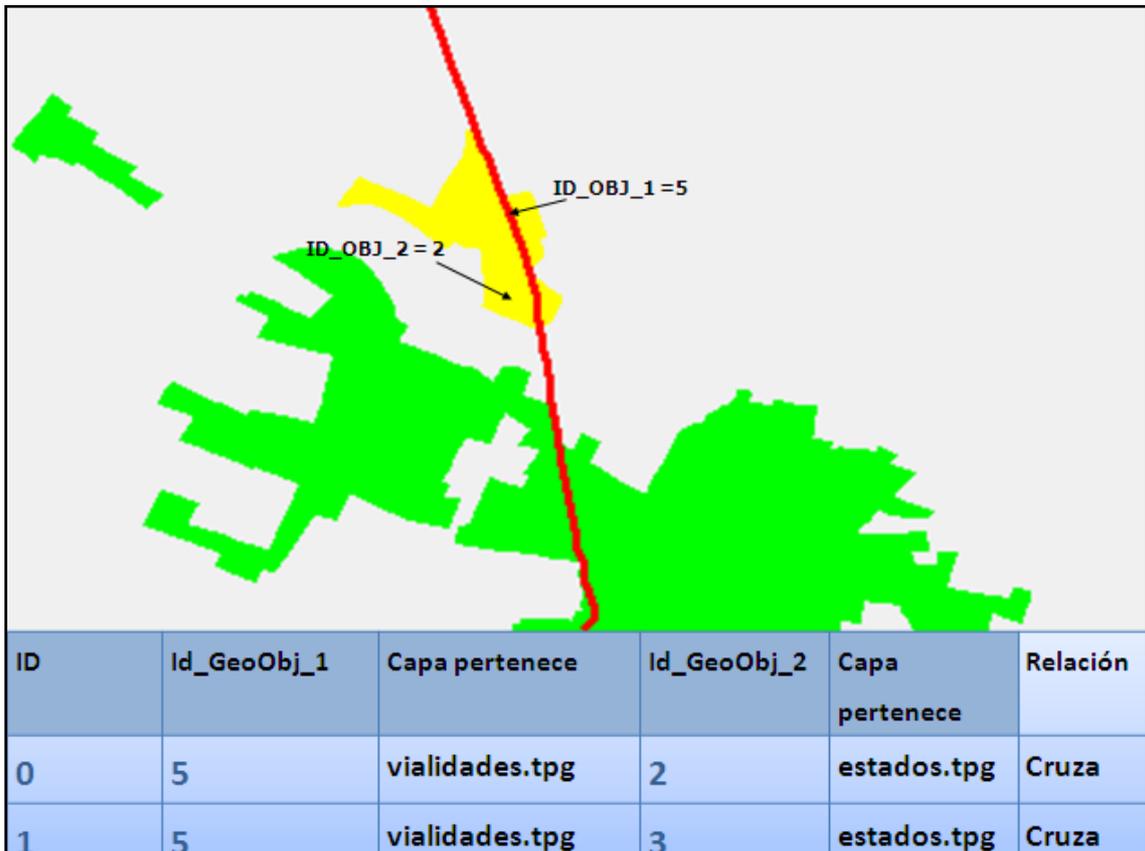
Los archivos topológicos (*TopologyFiles*) almacenan atributos descriptivos y las relaciones topológicas entre objetos geográficos. Estos archivos se generaron a partir de archivos vectoriales (*shapefiles*), el proceso de generación está basado en un trabajo previo [22]. La Figura 4-7 muestra de forma grafica la relación “conecta” entre una línea y un polígono (por ejemplo una carretera y un estado). También se observa que la relación topológica que existe ente el objeto *id\_obj\_1* y el objeto *id\_obj\_2*, se almacena en un formato de archivo denominado *TopologyFile*.



**Figura 4-7** Relación topológica “conecta” almacenada en un *TopologyFile*.

Como se aprecia en la Figura 4-7 la relación topológica se almacena en una tabla, cuyos atributos son: el número de relación (ID), los objetos geográficos (ID\_GeoObj\_1 e ID\_GeoObj\_2), la relación que existe entre estos objetos (denotada con la letra “c”) y la capa de datos a la que pertenecen.

Otro ejemplo se muestra en la Figura 4-8, para la relación topológica “cruza”.



**Figura 4-8** Relación topológica “cruza” entre dos objetos geográficos.

Como se observa en la figura un *TopologyFile* guarda en formato tabular la relación topológica entre un par de objetos geográficos. En este caso los objetos son vialidades y estados. De esta forma se muestra en la Tabla 4-4 el formato de un *TopologyFile* los cuales se utilizan para poder realizar recuperación a partir del procesamiento de estas relaciones.

**Tabla 4-4** Estructura de un *TopologyFile*

ID	Id_GeoObj_1	Capa pertenece	Id_GeoObj_2	Capa pertenece	Relación
1	5	Aeropuertos.tpg	2	Caminos.tpg	C
2	10	Calle.tpg	6	Area_Urbana.tpg	D

C=conecta, D="Dentro"

Como se muestra en la Tabla 4-4 cada registro contiene dos objetos geográficos, la capa a la que pertenece cada uno de ellos, y la relación topológica (*conecta, dentro, etc.*) que los vincula. Adicionalmente se tienen atributos descriptivos para cada objeto geográfico.

De esta forma, se finaliza con la descripción de la tercera y última fuente de datos geográfica considerada para el desarrollo de esta tesis. En la siguiente sección se define y explica la estrategia de recuperación integral de información geográfica (*iGIR*) desarrollada en esta tesis.

#### **4.4 Estrategia de Recuperación de Información Geográfica (*iGIR*)**

Actualmente, para recuperar información geográfica se requiere técnicas y mecanismos que además de procesar los términos o palabras, también procesen la esencia y naturaleza de los datos espaciales. Por ejemplo, establecer si la palabra *cerca* se refiere a un tiempo o una distancia, o si su sentido cambia con relación al objeto espacial que lo involucra. Una forma de lograrlo es procesar las relaciones semánticas y espaciales asociadas a un objeto geográfico. Esta tesis define una estrategia de GIR procesando integralmente las relaciones semánticas, las relaciones topológicas y las propiedades de los objetos geográficos (*iGIR*).

Considerando que un mismo objeto espacial es descrito con diferentes niveles de detalle y diferentes términos en un archivo vectorial, diccionario de datos, ó *GeoOntología*. Además, estos objetos son almacenados en cada fuente de datos de acuerdo a su semántica espacial, su topología y sus propiedades geográficas.

Un ejemplo real lo tenemos con el objeto *Carretera* el cual es definido en una *GeoOntología* como una vía de comunicación, mientras es descrita por su número de carriles en diccionarios geográficos. Similarmente, en un archivo topológico se indica que una carretera atraviesa algunas ciudades. Por lo tanto, una recuperación que integre estos tres aspectos permitirá incluir datos que serian omitidos con otros enfoques. Uno de los primeros productos de esta tesis considerando este hecho es un artículo publicado en [78].

*iGIR* es el nombre con el que se identifica el sistema desarrollado en esta tesis. En donde la entrada del sistema es una consulta espacial y la salida es un conjunto de documentos geográficos obtenido de las tres fuentes de información geográfica. Para ello, se definieron tres mecanismos de búsqueda para las tres fuentes de datos geográficas. Cada mecanismo funciona utilizando un criterio de coincidencia o *matching* para realizar la recuperación. El primero es por coincidencia conceptual o *matching conceptual* el cual procesa conceptos de *GeoOntologías*, el segundo es por coincidencia geográfica o *matching geográfico* el cual procesa atributos y relaciones de diccionarios geográficos. El tercero es por coincidencia topológica o *matching topológico* el cual trabaja con las relaciones topológicas de archivos topológicos.

La decisión de utilizar *GeoOntologías*, diccionarios y archivos vectoriales se debe a que los objetos geográficos son descritos en forma diferente en cada fuente. Además son utilizadas por instituciones como el INEGI, empresas comerciales de GIS y por centros de investigación. La Figura 4-9 muestra la estrategia de recuperación propuesta.

**Figura 4-9** Marco de trabajo para recuperar y ponderar información geográfica.

Tal y como se ilustra en la Figura 4-9 *iGIR* tiene como entrada, la consulta y como salida un conjunto de documentos que responden a la consulta. También se muestran tres fuentes de datos geográficas con su respectivo mecanismo de *matching*, un método de ponderación (*iRank*) y el despliegue de resultados ponderados en una interfaz Web.

El módulo de consulta representa la interfaz en donde se expresarán las consultas de tipo tripleta. Esta consulta es procesada y enviada a *OntoExplora*, un algoritmo que procesa la consulta para indicar que buscar en las *GeoOntologías*, diccionarios y archivos topológicos. Finalmente, se integran estos resultados y se ponderan usando el método *iRank* (ver sección 5.2) para desplegarlos al usuario. A continuación se describe la estructura requerida para las consultas y posteriormente se explica el algoritmo *OntoExplora*, el cual conduce la recuperación de información geográfica en *GeoOntologías*.

#### 4.4.1 Estructura de Consultas.

Para propósitos de esta tesis y delimitar el alcance de la misma, el mecanismo de recuperación se realiza a partir de consultas geográficas de tipo tripleta, cuya estructura es la siguiente:

$$Q_{Gk} = \{que; relación\ espacial; donde\}^{42}$$

Esta estructura está formada por tres elementos: 1) *que*, el cual es un objeto geográfico; 2) *relación espacial*, indica cómo se asocian dos objetos geográficos. Por ejemplo, con las relaciones *conecta* y *dentro*. 3) *donde* es una referencia geográfica u objeto geográfico.

---

<sup>42</sup> Donde k es un identificador que representa el número de consulta, G denota que la consulta es geográfica, mientras {*que*} es un objeto geográfico; {*donde*} es un objeto geográfico ó un objeto de referencia geográfica y {*relación*} es una relación espacial entre {*que*} y {*donde*}.

Por otra parte, debido a que las *relaciones espaciales* se expresan o denotan utilizando diferentes palabras, es necesario realizar equivalencias o correspondencias entre éstas. Es decir, asociar a la *relación espacial* su correspondiente operación espacial u expresión común (e.g. *cerca* se puede asociar con la operación espacial de buffer).

A continuación, se mencionan las relaciones espaciales y la terminología asociada a dichas relaciones. Estos términos o palabras representan las relaciones espaciales, de acuerdo a alguna especificación establecida, así como también aquellas que simulan y modelan el comportamiento de fenómenos geográficos con base en la experiencia humana. La determinación de estas asociaciones, fueron publicadas en [77] como producto de esta tesis. Se listan a continuación en la Tabla 4-5.

**Tabla 4-5** Términos o palabras asociadas a relaciones espaciales

Ejemplo de uso de la relación espacial	Relación espacial	Operación o relación espacial asociada
México está “junto” a Estados Unidos	Junto	Comparte limite
El Lago Chapala está “en” Jalisco	En	Dentro
La avenida insurgentes interseca avenida Tacubaya	Intersecta	Intersección
El puente San Martín “cruza” el río Grijalva	Cruza	Intersección
La carretera 45 “conecta” con la carretera 65	Conecta	Conecta

Con base en las correspondencias indicadas en la Tabla 4-5, se facilitará la tarea de búsqueda ya que se utilizará la relación u operación espacial equivalente al sentido de la palabra empleado en la consulta geográfica (noción de la geografía). Por ejemplo, para *junto* se utiliza la relación espacial *comparte límite*. De esta forma, los resultados serán documentos relevantes. A continuación, en la Tabla 4-6, se listan algunos ejemplos de consultas de tipo triplete, las cuales utilizan estas relaciones espaciales.

**Tabla 4-6** Ejemplos de Consulta tipo tripleta

Ejemplos de consultas utilizadas en esta tesis	Utilizada en sección
Q <sub>G1</sub> = {Ríos en Jalisco}	Capitulo 1 : Introducción
Q <sub>G2</sub> = {Sitios arqueológicos <b>en</b> Área Metropolitana}	Capitulo 1 : Introducción
Q <sub>G3</sub> ={Hoteles <b>cerca</b> Aeropuerto Benito Juárez}	Capitulo 1 : Introducción
Q <sub>G4</sub> = {Vialidades <b>conectan</b> aeropuertos}	Capitulo 4 : Metodología de recuperación Capitulo 5: Metodología de ponderación Capitulo 6: Pruebas y Resultados
Q <sub>G5</sub> = {Vialidades <b>cruzan</b> línea metro}	Capitulo 5 : Metodología de ponderación Capitulo 6: Pruebas y Resultados
Q <sub>G6</sub> = {Centro comercial <b>dentro</b> colonia}	Ejemplo usado en Capitulo 6: Pruebas y Resultados
Q <sub>G7</sub> = {Línea metro <b>intersecta</b> vialidades}	Ejemplo usado en Capitulo 6: Pruebas y Resultados
Q <sub>G8</sub> = {Colonias <b>comparte límite</b> colonias}	Ejemplo usado en Capitulo 6: Pruebas y Resultados
Q <sub>G9</sub> = {Línea metro <b>conecta</b> colonias}	Ejemplo usado en Capitulo 6: Pruebas y Resultados

Como se puede observar en la Tabla 4-6 algunas consultas se utilizan como ejemplos para explicar la metodología y el planteamiento del problema, en la presente tesis. En lo sucesivo, estas consultas serán utilizadas como ejemplos para explicar el proceso de recuperación y ponderación en los capítulos sucesivos de la tesis. En particular la consulta Q<sub>G4</sub> se utiliza desde el capítulo 4 hasta el capítulo 6 para explicar todo el proceso. En la siguiente sección se explica el algoritmo *OntoExplora* el cual conduce el proceso de recuperación en las GeoOntologías, y proporciona los parámetros de búsqueda para recuperar información geográfica en los diccionarios geográficos y archivos topológicos.

#### 4.5 Algoritmo de exploración de GeoOntología (OntoExplora)

La extracción de conocimiento a partir de la *GeoOntología* se realizará mediante el algoritmo de exploración (*OntoExplora*) cuya misión principal es encontrar los conceptos asociados a los objetos geográficos de una consulta.

*OntoExplora* extrae las relaciones vecinas de un concepto geográfico. En particular se extraen las relaciones directas de cada concepto. En este caso, son los que están a un nodo de distancia en el documento XML de la *GeoOntología*. Esta decisión se tomó teniendo como referencia la primera ley de la geografía, la cual dice: "*Everything is related to everything else, but near things are more related than distant things*" [58]. Por tal motivo, sólo se consideran las relaciones semánticas que están a un nodo de distancia entre los conceptos la *GeoOntología*, por ser las más cercanas y que en consecuencia guardan más relación entre conceptos.

Para reforzar esta decisión, también se considero el concepto de herencia de programación orientado a objetos, cuya definición es: La herencia es uno de los mecanismos por medio del cual una clase se deriva de otra, llamada entonces *superclase*, de manera que extiende su funcionalidad. La idea es la de partir de las situaciones más generales e ir derivando hacia las más particulares, creando categorías, de la misma forma en que se considera piensa el ser humano.

Adicionalmente, mencionaremos que el término contexto es utilizado de forma similar al término vecindad (usado en esta tesis) por el sistema Yahoo Contextual [60]. En este sistema el contexto es conformado por las relaciones vecinas de los conceptos que aparecen en un documento. La diferencia radica en que en dicho trabajo se apoyan en una red semántica, mientras que esta tesis emplea ontologías geográficas. El algoritmo de exploración (*OntoExplora*) extrae los conceptos y relaciones vecinas de cada concepto geográfico, los cuales son los parámetros utilizados para la búsqueda en diccionarios de datos y archivos vectoriales.

*OntoExplora* inicia buscando el concepto y relaciones que representan un objeto, cuando encuentra una coincidencia se extraen las relaciones asociadas y sus clases o instancias respectivas. Durante esta fase, pueden ocurrir cuatro tipos de coincidencia, los cuales se mencionan a continuación, 1) Coincidencia completa, 2) Coincidencia parcial, 3) Coincidencia atómica, y 4) Coincidencia nula. En las secciones siguientes se explica cada una de ellas.

#### **4.5.1 Coincidencia completa**

Una coincidencia completa ocurre cuando todos los elementos que conforman a una consulta de tipo tripleta son encontrados en la *GeoOntología*. Es decir, se encuentra los conceptos o instancias que representan al elemento {que} y {donde}. Además, estos conceptos se encuentran vinculados por otros conceptos o instancias a través del elemento <relación espacial>. Por ejemplo para  $Q_{G4} = \{Vialidades\ conectan\ Aeropuertos\}$  se tiene la tripleta: {que}= “Avenida”; {donde}= “Aeropuerto”; {relación espacial}= “conecta”. Entonces si cada uno de estos elementos aparece en la *GeoOntología*, tenemos una coincidencia completa y se extraen los documentos asociados dichos conceptos.

#### **4.5.2 Coincidencia parcial**

Una coincidencia parcial ocurre cuando, en la *GeoOntología*, se encuentran parejas de los elementos que conforman a la tripleta. Es decir, se encuentran clases o instancias para el elemento {que} y para el elemento {donde}. También puede ocurrir que la {relación espacial} está asociada al elemento {que} ó al elemento {donde}.

Por ejemplo, considerando la consulta  $Q_{G4} = \{Vialidades\ conectan\ Aeropuertos\}$  se tiene la tripleta: {que}= “Vialidades”; {donde}= “Aeropuertos”; {relación espacial}= “conecta”. Entonces, si se encuentra alguna de las siguientes combinaciones: “Vialidades” y “Aeropuertos”; ó se encuentra “Vialidades” y “conecta”; ó “Aeropuerto” y “conecta”. Tenemos entonces una coincidencia parcial.

### 4.5.3 Coincidencia atómica

En este caso, se tiene una coincidencia atómica cuando únicamente se encontró uno de los elementos de la tripleta. Es decir, se tiene el elemento {que} o el elemento {donde}. El elemento {relación espacial} puede ser encontrado, pero como es insuficiente para contextualizar, no se considera (e.g. la relación conecta, se puede hallar asociada a varias o todas las clases e instancias de la *GeoOntología*)

Por ejemplo, considerando la consulta  $Q_{G4} = \{\text{Vialidades conectan Aeropuertos}\}$  se tiene la tripleta: {que}= “Vialidades”; {donde}= “Aeropuertos”; {relación espacial}= “conectan”. Entonces, si se encuentra “Vialidades” o “Aeropuertos” tenemos una coincidencia atómica.

### 4.5.4 Coincidencia nula

Este último caso, se presenta cuando ningún objeto de la tripleta no se encuentra en la *GeoOntología*, lo que resultará en una respuesta vacía (“No Match Found”). En este caso, una alternativa de solución sería utilizar algoritmos de alineación o de integración con otras fuentes de información para poder responder a las consultas. Así como también utilizar la teoría de confusión para poder encontrar (usando otra fuente de datos) un concepto equivalente al buscado.

Una vez que se han definido los casos posibles para el *matching*. Se procede a continuar con la explicación de la estrategia de recuperación, la cual en esta fase se realiza explorando la *GeoOntología* (implementada en un documento XML) a través del uso de un analizador sintáctico [36]. Este analizador proporciona un conjunto de métodos los cuales permiten explorar un documento XML a través de los elementos (etiquetas o nodos) que lo conforman, facilitando, de esta forma, su exploración. El proceso de concordancia o coincidencia se realiza a nivel de los nodos que conforman el documento XML. Para tal propósito, se definió el algoritmo *OntoExplora* que se muestra en la Figura 4-10.

**Inicio**

Identifica y procesa cada elemento de la consulta geo-espacial ( $Q_{GN}$ ) {que, relación, donde}

Asigna la relación topológica, geográfica, o semántica correspondiente a relación ( Tr)

**Recuperación guiada por GeoOntología:**

Busca el concepto correspondiente para cada objeto de la consulta en la GeoOntología

If (hay\_coincidencias) Then

Extrae el contexto usando relaciones semánticas de meronimia e hiperonimia

Extrae las propiedades del concepto ( $P_{concept}$ )

Se llena el vector de vecindad (VectorVec) con las relaciones y propiedades

Se generan las consultas (QueryNuevo) usando VectorVec

Else

Envía  $Q_{GN}$  a las fuentes vectoriales y de diccionarios

**Recuperación de Diccionarios de Datos:**

Busca datos en diccionarios de acuerdo a la consulta y VectorVec

Utilizando  $Resp_{Ont}$ , recupera datos geográficos por sus propiedades y relaciones

If ( hay\_ coincidencia) Then

Recupera objetos de acuerdo a las relaciones propiedades geográficas

End If

Regresa respuestas ( $Resp_{Geo}$ )

**Recuperación de datos topológicos:**

Busca datos en topologyfiles de acuerdo a la consulta y VectorVec

Utilizando  $Resp_{Ont}$ , solicita los datos vectoriales (capas)

If ( hay\_ coincidencia ( datos vectoriales) ) Then

Recupera objetos de acuerdo a las relaciones topologicas

End If

Regresa respuestas  $Resp_{Topological}$

**Integración y ponderación:**

Integra  $Resp_{Topological}$ ,  $Resp_{Ont}$  y  $Resp_{Geo}$  en un conjunto final (FVector)

Pondera  $FS_{geobj}$  usando iRank y lo muestra al usuario

**Recuperación aislada o en dos fuentes**

Else

Recupera los datos considerando los objetos de  $Q_{GN}$

If (hay\_coincidencias) Then

Recupera los datos de acuerdo a propiedades y relaciones espaciales

Regresa las coincidencias encontradas

Else

Recupera los datos vectoriales de acuerdo a relaciones topológicas

Regresa las coincidencias

Figura 4-10 Algoritmo *OntoExplora*.

El algoritmo muestra de forma global los pasos utilizados en la estrategia de recuperación. Los detalles de cada uno de estos pasos, corresponden a los tres mecanismos de *matching* definidos para recuperar información cada una de las tres fuentes de datos geográficas. A continuación se explica el primero de estos mecanismos, denominado *matching* conceptual y que utiliza como fuente de información a las *GeoOntologías*. Posteriormente, en las secciones 4.7 y 4.8 se explican la forma en que se recuperan los datos de las otras dos fuentes de información, utilizando los mecanismos de búsqueda basados en *matching* geográfico y *matching* topológico.

#### **4.6 Recuperación de información geográfica en las GeoOntologías**

La recuperación de información en las *GeoOntologías* utiliza el mecanismo de búsqueda denominado coincidencia conceptual o *matching* conceptual (*MatchCon*) empleando las *GeoOntologías* descritas en la sección 4.1, las cuales representan la primer fuente de información. En particular, el proceso de *MatchCon* consiste en encontrar los conceptos que concuerden con los objetos geográficos de una consulta. Esto se logra explorando los nombres de los conceptos. Por ejemplo, para la consulta  $Q_{G4} = \{Vialidades\ conectan\ Aeropuertos\}$ , se buscan los nombres de conceptos, *Vialidades* y *Aeropuertos*.

Una vez que se han encontrado los conceptos coincidentes, se proceden a extraer sus conceptos y relaciones vecinas, denominadas como vecindad. La vecindad son las relaciones semánticas que unen a un concepto con otro. Por ejemplo, en la *GeoOntología* de la sección 4.1 el concepto “*Carretera*” está relacionado con los conceptos “*Camino*” y “*Vialidad*”. Un enfoque similar a esta proceso, se utilizan en el sistema *Yahoo Contextual* [60].

El proceso de *MatchCon* utiliza el algoritmo *OntoExplora* (descrito en la sección 4.5) para explorar las *GeoOntologías* en búsqueda de los conceptos que coincidan (*matching*) con una consulta y posteriormente extraer su vecindad. A continuación, se explica cómo se realiza la recuperación de información geográfica en las *GeoOntologías* usando el algoritmo *OntoExplora*.

La recuperación de información geografía comienza cuando el algoritmo *OntoExplora* recibe la consulta espacial. Enseguida, se efectúa la búsqueda de los elementos de la consulta y cuando se encuentran coincidencias, se extraen las relaciones vecinas y conceptos asociados. Los cuales se almacenan en un vector de vecindad (*VectorVec*) con base en la siguiente estructura:

*VectorVec* = (Concepto por hiperonimia, Concepto por meronimia, relaciones espaciales, relaciones semánticas, <<Atributos asociados>>}

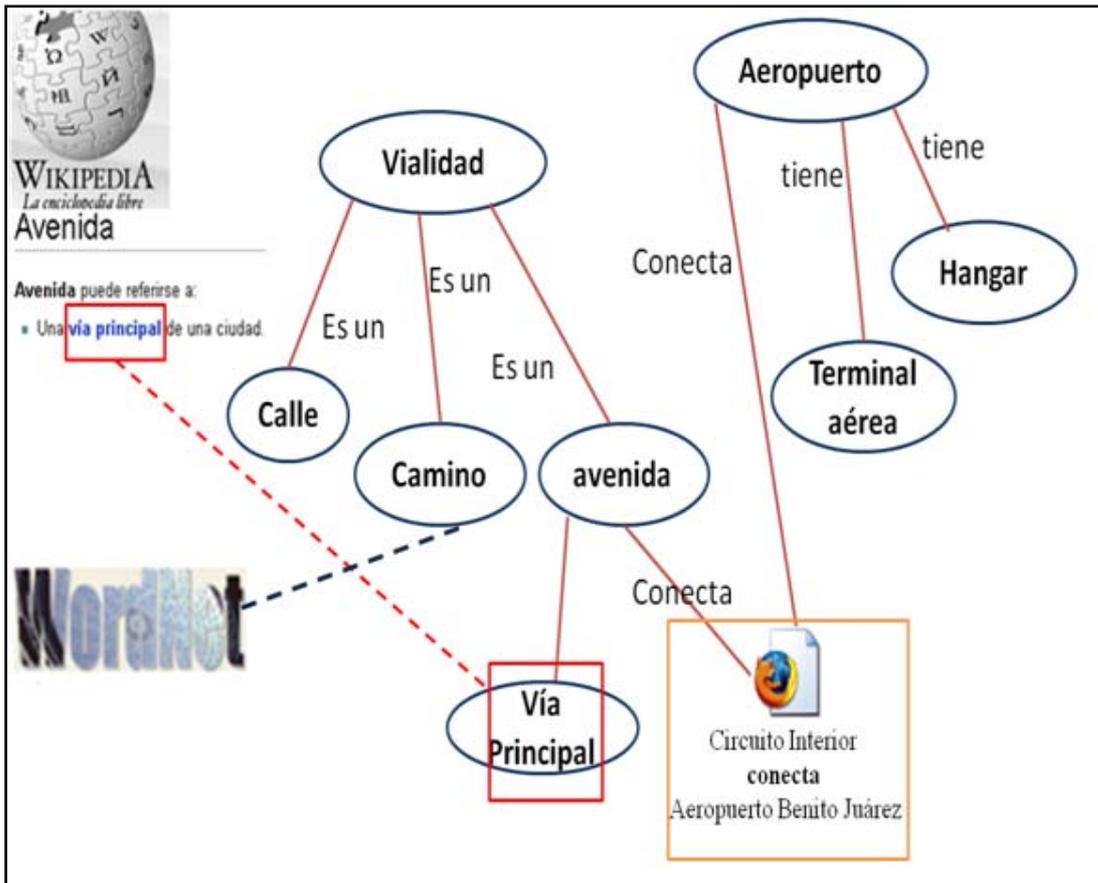
Por ejemplo, para el objeto geográfico {Avenida} se obtiene:

*VectorVec* = {(Vialidad, Camino, Avenida insurgentes, <<número de carriles, sentido >>}

La obtención de estos elementos se basa en los siguientes criterios para la extracción de relaciones vecinas:

- Si existe la relación de hiperonimia y/o meronimia, entonces
  - se extrae el concepto o jerarquía de concepto correspondiente
- De lo contrario
  - Se extraen el resto de las relaciones y conceptos
- Se integran todos los resultados en una lista

Para ilustrar el funcionamiento de estos criterios, consideremos la consulta  $Q_{G4}$  = {Vialidades conectan Aeropuertos}, donde los elementos {que} = *vialidades*, {donde} = *Aeropuertos* y {relación espacial} = *conectan*, son buscados de acuerdo al algoritmo *OntoExplora*, un fragmento de la *GeoOntología* que describe estos términos se ilustra en la Figura 4-11.



**Figura 4-11** Fragmento de GeoOntología generada usando *Wikipedia* y *WordNet*

En este escenario, se encuentra que {Vialidad} está vinculado con el concepto *avenida* el cual está asociado a la relación *conecta*. La cual a su vez está vinculada con el elemento {Aeropuerto}. Donde {conecta} y {Aeropuerto} son también parte de la tripleta, por lo cual se tiene un *matching completo* (ver sección 4.5.1) y se procede a extraer los documentos asociados a estos conceptos y que son los resultados obtenidos por *MatchCon*. En este caso, el documento asociado es <Circuito Interior conecta Aeropuerto Benito Juárez>.

De esta forma, tenemos que el vector de vecindad para la consulta  $Q_{G4}$  se conforma por:

*VectorVec* = {(*Vialidad, es un, Camino, Avenida, conecta, Aeropuerto; Avenida: Circuito Interior; Avenida: Circuito Interior conecta Aeropuerto Benito Juárez, << número de carriles, sentido >>, Aeropuerto tiene terminal aérea, Hangar, conecta Avenida; Aeropuerto: Benito Juárez, Avenida: Circuito Interior conecta Aeropuerto Benito Juárez, << Internacional, Nacional, Local >>*)}

Como se observa el vector de vecindad contiene diversos elementos que describen a cada concepto y las relaciones espaciales asociadas a ellos, mediante el procesamiento de este vector, se indica cuales conceptos y relaciones se buscarán en las otras dos fuentes de información.

El siguiente paso consiste en la integración de estas relaciones vecinas para procesar el vector de vecindad (*VectorVec*), el cual es la estructura que será enviada a las otras fuentes de información (archivos vectoriales y diccionarios de datos). El propósito de esto, es que los elementos del vector sean extraídos y buscados en cada fuente de información de acuerdo a los respectivos criterios de *matching* establecidos.

Considerando la estructura del vector, el cual contiene los conceptos vecinos, relaciones espaciales y semánticas, así como los atributos de cada objeto geográfico. En donde los atributos contenidos en el vector serán procesados en los diccionarios geográficos. Mientras que las relaciones espaciales serán procesadas en los *TopologyFiles*.

De esta forma el módulo de *matching conceptual* finaliza al haber generado este vector de vecindad y extraído los documentos asociados a cada concepto. Este vector de vecindad se utilizará para recuperar información en los diccionarios de datos y *TopologyFiles* en las dos secciones siguientes.

#### **4.7 Recuperación de información geográfica en diccionarios geográficos**

Este módulo de recuperación de información utiliza como mecanismo de búsqueda la coincidencia geográfica o *matching* geográfico (*MatchGeo*). Se utiliza como fuente de información a los diccionarios de datos del INEGI [59]. Estos diccionarios contienen información relativa a objetos geográficos. También, tienen definida una estructura, dentro de la cual destacan: una descripción de los objetos geográficos, sus principales relaciones y restricciones. También, contienen ciertas consideraciones para la representación y procesamiento de los datos. Además de que tienen un dominio definido (ver sección 3 Marco Teórico). Se describen enseguida, brevemente, las razones por las cuales, se eligió como fuente de información geográfica a los diccionarios de datos.

La recuperación en los diccionarios de datos se efectúa usando el vector de vecindad recibido desde *MatchCon*. Cada elemento del vector se extrae y se busca en los documentos XML. Por ejemplo, para *Vialidad* se buscan nombres de archivo y nodos que contengan ese nombre.

En este proceso pueden ocurrir dos escenarios:

- *MatchCon* envió el vector de vecindad con al menos un elemento (cada elemento del vector es un parámetro de búsqueda).
- *MatchCon* envía un vector de vecindad vacío, entonces se utiliza en sustitución, la consulta original.

Para el primer caso, de acuerdo al vector de vecindad recibido, se buscan los conceptos que aparecen en el vector de vecindad. Cada elemento es buscado en el documento XML, para lo cual se utiliza un analizador sintáctico de XML (SAX: Simple Api for XML) [36]. En caso de encontrarse una coincidencia, se extrae el nodo, y sus atributos relacionados (descripciones y relaciones con otros conceptos). En caso de no encontrar coincidencia, se pasa al segundo caso.

Para el caso número dos, se recibe la consulta original, y cada elemento de la tripleta es buscado en el árbol XML y se regresan las coincidencias encontradas (nodos y atributos relacionados). El resultado final también incluye el grupo de relaciones espaciales que establece el diccionario para el concepto encontrado.

Una vez definidos los criterios para la búsqueda en este tipo de fuentes de información, se ilustra el procedimiento considerando la consulta  $Q_{G4}$ . Entonces, se reciben las vecindades obtenidas por *MatchCon* y para la consulta  $Q_{G4}$ , los elementos a buscar son:

*VectorVec* = {(*Vialidad*, es un, *Camino*, *Avenida*, conecta, *Aeropuerto*; *Avenida*: *Circuito Interior*; *Avenida*: *Circuito Interior* conecta *Aeropuerto Benito Juárez*, <<número de carriles, sentido >>, *Aeropuerto* tiene terminal aérea, *Hangar*, conecta *Avenida*; *Aeropuerto*: *Benito Juárez*, *Avenida*: *Circuito Interior* conecta *Aeropuerto Benito Juárez*, <<*Internacional*, *Nacional*, *Local*>> }

Entonces, cada elemento de los vectores de vecindad es buscado ya sea como nombre del documento XML, nodo, o atributo. Conformándose una lista de coincidencias por cada elemento encontrado. En el caso de la consulta  $Q_{G4}$ , se muestran las coincidencias encontradas en este caso “Aeropuerto” y “Camino” en la Figura 4-12.

- <elemento nombre="aeropuerto">	- <elemento nombre="Camino">
- <aeropuerto>	- <Camino>
- <atributos>	- <atributos>
- <dominio_fijo>	- <dominio_fijo>
- <TIPO_DE_AEROPUERTO>	- <Tipo_de_camino>
<a1>Internacional: Permite vuelos con cubrimiento internacional	<a1>Brecha : Generalmente es posible la circulación de un vehículo.</a1>
<a2>Nacional: Permite vuelos con cubrimiento nacional</a2>	<a2>Vereda : Sólo circulan personas y animales.</a2>
<a3>Local: Sólo permite vuelos con cubrimiento regional</a3>	
</TIPO_DE_AEROPUERTO>	</Tipo_de_camino>
</dominio_fijo>	</dominio_fijo>
+ <dominio_variable>	<dominio_variable>
</atributos>	</atributos>
- <restricciones_de_integridad>	- <restricciones_de_integridad>
- <COMBINACIONES_AUTORIZADAS_DE_VALORES_DE_ATRIBUTOS>	- <COMBINACIONES_AUTORIZADAS_DE_VALORES_DE_ATRIBUTOS>
<c1>Internacional</c1>	<c1>Brecha</c1>
<c2>Nacional</c2>	<c2>Vereda</c2>
<c3>Local</c3>	
</COMBINACIONES_AUTORIZADAS_DE_VALORES_DE_ATRIBUTOS>	</COMBINACIONES_AUTORIZADAS_DE_VALORES_DE_ATRIBUTOS>
<CALIFICADOR_AUTORIZADO_DE_REPRESENTACIONES_GEOMETRICAS>	+ <CALIFICADOR_AUTORIZADO_DE_REPRESENTACIONES_GEOMETRICAS>
</restricciones_de_integridad>	</restricciones_de_integridad>
- <relaciones>	- <relaciones>
<r1>Conecta Camino (L)</r1>	
<r2>Conecta Carretera (L)</r2>	
<r3>Conecta Camino (L)</r3>	
<r4>Conecta Carretera (L)</r4>	
<r5>Comparte Calle (L)</r5>	
<r6>Comparte Camino (L)</r6>	
<r7>Comparte Carretera (L)</r7>	
	<r13>Conecta Aeropuerto (A)</r13>

Figura 4-12 Resultados de diccionarios para  $Q_{G4} = \{ \text{Vialidades conectan Aeropuertos} \}$

Como se puede observar para la consulta  $Q_{G4}$  se buscan Vialidades, pero no se encuentran coincidencias. Entonces se busca sus conceptos vecinos (Avenidas y caminos). De esta forma se encuentra *<Camino Conecta Aeropuerto>*.

Entonces se tiene que un objeto aeropuerto está asociado a un objeto Camino a través de la relación *conecta*. Este resultado se envía al módulo de recuperación topológica para buscar *TopologyFiles* que cumplan con estos criterios. Para continuar con el proceso de

recuperación, cada una de las relaciones encontradas se envía al módulo de recuperación topológica (*MatchTopology*) para buscar los objetos geográficos que coincidan con dichos criterios. En este punto, termina la fase de recuperación de información geográfica en diccionarios geográficos.

#### **4.8 Recuperación de información geográfica en archivos topológicos**

La recuperación de información en archivos topológicos (*TopologyFiles*) utiliza como mecanismo el *matching* topológico (*MatchTopology*). Los *TopologyFiles* son descritos en la sección 4.3. Estos archivos serán procesados para extraer las relaciones topológicas de los objetos geográficos. Las relaciones topológicas que almacenan los *TopologyFiles* se generan al aplicar una operación espacial de sobreposición entre dos archivos vectoriales (capas de datos). Por ejemplo, al sobreponer una capa de carreteras, con una capa de estados se puede obtener la relación cruza: “*La carretera 75 cruza por el estado de Tamaulipas*”. De esta forma, se pueden conocer: intersecciones, conexiones, objetos vecinos, etc. Básicamente, el proceso en este módulo se realiza buscando los elementos del vector de vecindad (enviado por *MatchCon*) y los elementos de la lista de relaciones espaciales (enviado por *MatchTopology*). Para el vector de vecindad, se tienen en consideración los siguientes puntos:

- Se buscan los objetos geográficos definidos en la consulta original de acuerdo al tipo especificado en el dominio fijo (e.g. para Aeropuerto se buscan: Aeropuertos Locales, Aeropuertos Nacionales, y Aeropuertos Internacionales).
- Se agrupa cada elemento del grupo dos, con todos los nombres de atributos del grupo tres ( de acuerdo a los grupos definidos en la sección 4.3.3.1)
- Si existen coincidencias, entonces los resultados se filtran utilizando los valores de atributos correspondientes.
- Para cada registro coincidente, se extrae la relación correspondiente y los registros con los cuales está vinculada esta relación.( e.g. si Camino está asociado con Aeropuerto, se extrae el registro de Aeropuerto)
- De acuerdo al grupo de relaciones, se buscan los objetos geográficos involucrados con dicha relación.

Entonces considerando el vector de vecindad de la consulta Q<sub>G4</sub>:

*VectorVec= {(Vialidad,es un, Camino,Avenida, conecta, Aeropuerto; Avenida:Circuito Interior; Avenida:Circuito Interior conecta Aeropuerto Benito Juarez, <<número de carriles, sentido >>, Aeropuerto tiene terminal aérea, Hangar, conecta Avenida; Aeropuerto: Benito Juárez, Avenida:Circuito Interior conecta Aeropuerto Benito Juarez,<<Internacional, Nacional, Local >>}*

Aplicando los criterios anteriores, tenemos que se buscan los objetos que contengan los atributos correspondientes. Por ejemplo, para el caso del objeto *Aeropuerto* se buscan los atributos: {Internacional, Nacional, Local} mientras que para *Caminos* se buscan (número de carriles y sentido).

Las coincidencias encontradas se despliegan, en este caso son *vialidades, caminos y calles* que se conectan con el Aeropuerto o conducen hacia éste. La Tabla 4-7 muestra un fragmento de resultados para la consulta Q<sub>G4</sub>.

**Tabla 4-7** Fragmento de resultados de MatchTopology

ID	Id_GeoObj_1	Capa_pertenece	Id_GeoObj_2	Capa_pertenece	Relación
1	5	Aeropuertos.tpg	2	Caminos.tpg	C
2	10	Calle.tpg	6	Area_Urbana.tpg	D
3	7	Carreteras.tpg	8	Aeropuertos.tpg	C
4	13	Calle.tpg	10	Area_Urbana.tpg	D
5	15	Aeropuertos.tpg	9	Caminos.tpg	C

Los resultados de la Tabla 4-7 muestran algunos de los objetos geográficos recuperados para la consulta Q<sub>G4</sub>. La Figura 4-13 muestra la tabla de resultados para aeropuertos involucrados y su respectivo archivo vectorial (capa de datos).

Aeropuerto				
OBJECTID	TIPO	ENTIDAD	FC	CAPA DATOS
1	Internacional	AEROPUERTO	643	airport1.shp
2	Internacional	AEROPUERTO	643	airport2.geo
3	Internacional	AEROPUERTO	643	airport1.shp
4	Internacional	AEROPUERTO	643	airport1.shp
5	Internacional	AEROPUERTO	643	airport2.geo
6	Internacional	AEROPUERTO	643	airport2.geo
7	Internacional	AEROPUERTO	643	airport2.geo
8	Internacional	AEROPUERTO	643	airport2.geo

**Figura 4-13** Resultados para el objeto aeropuerto de la consulta Q<sub>G4</sub>.

De esta forma, el proceso de *MatchTopology* termina, y se procede a ponderar y desplegar los resultados de forma integral.

#### 4.9 Integración y despliegue de resultados

La fase de integración de resultados se lleva a cabo, aprovechando las características del formato XML, agregando un nodo por cada resultado y agrupando los resultados de acuerdo al criterio de recuperación, es decir agregando los resultados obtenidos en la *GeoOntología*, enseguida los obtenidos de los diccionarios geográficos, y finalmente se incluyen los resultados obtenidos de los *TopologyFiles*. Por ejemplo, para el de la consulta Q<sub>G4</sub>, se obtiene el vector de vecindad, y se encuentra un documento que describe que la avenida <Río Churubusco> conecta con el aeropuerto internacional Benito Juárez de la ciudad de México. Este es el primer documento encontrado usando *MatchCon*, para el caso de *MatchTopology* se encuentran las descripciones y relaciones para objetos *camino*, *carreteras* y *calles*. Por último se extraen de los *TopologyFiles* a todos los objetos *camino*, *carreteras* y *calles* que conectan con un aeropuerto, integrándolos para su despliegue. La Figura 4-14 muestra la estructura general que tiene un archivo XML para procesar los resultados obtenidos de forma integrada.

### ¿Vialidades conectan Aeropuertos?

```

-<iGIR>
- <Resultados_MatchCon>
  <doc1>001</doc1>
  <doc2>002</doc2>
  </Resultados_MatchCon>
- <Resultados_MatchGeo>
  <doc1>003</doc1>
  <doc2>004</doc2>
  </Resultados_MatchGeo>
- <Resultados_MatchTopology>
  <doc1>005</doc1>
  <doc2>006</doc2>
  <doc3>007</doc2>
  <doc4>008</doc2>
  <doc5>009</doc2>
  </Resultados_MatchTopology>
</iGIR>

```

Id Documento	Recuperado desde	Documento recuperado
001	<i>GeoOntología</i>	Conecta al Viaducto Piedad y Río Churubusco con la Terminal 2
002	<i>GeoOntología</i>	El Distribuidor Vial 2, que conecta a Viaducto Piedad y Río Churubusco con la Terminal 2
003	<i>Diccionario geográfico</i>	Carretera conecta aeropuerto
004	<i>Diccionario geográfico</i>	Calle conecta calle aeropuerto
005	<i>TopologyFiles</i>	Boulevard puerto aéreo conecta Aeropuerto Benito Juárez
006	<i>TopologyFiles</i>	calle de Carlos León conecta Aeropuerto Benito Juárez
007	<i>TopologyFiles</i>	Eje uno norte conecta Terminal 1
008	<i>TopologyFiles</i>	Calle sonora conecta Terminal 2
009	<i>TopologyFiles</i>	Eje uno norte conecta Hangares

Figura 4-14 Integración de resultados.

Como se observa en la Figura 4-14 se tienen tres secciones correspondientes a los resultados encontrados en cada fuente de datos (MatchCon, MatchGeo y MatchTopology). El documento es representado con la etiqueta <doc1> donde el número 1 indica el número de documento encontrado, mientras que el valor de la etiqueta <doc1>001</doc1> en este caso es 001 significa que es el identificador asignado al documento y mediante este número se vincula con el documento especificado.

A continuación se muestran diez de los resultados obtenidos para la consulta  $Q_{G4} = \{ \text{Vialidades conectan aeropuertos} \}$ .

**Tabla 4-8** Diez documentos recuperados utilizando iGIR para la consulta Q<sub>G4</sub>.

Recuperado desde	Documento recuperado
<i>GeoOntología</i>	Conecta al Viaducto Piedad y Río Churubusco con la Terminal 2”
<i>GeoOntología</i>	El Distribuidor Vial 2, que conecta a Viaducto Piedad y Río Churubusco con la Terminal 2
Diccionario geográfico	Carretera conecta aeropuerto
Diccionario geográfico	Calle conecta calle aeropuerto
<i>TopologyFiles</i>	Boulevard puerto aéreo conecta Aeropuerto Benito Juárez
<i>TopologyFiles</i>	calle de Carlos León conecta Aeropuerto Benito Juárez
<i>TopologyFiles</i>	Eje uno norte conecta Terminal 1
<i>TopologyFiles</i>	Calle sonora conecta Terminal 2
<i>TopologyFiles</i>	Eje uno norte conecta Hangares
<i>TopologyFiles</i>	Distribuidor vial 1 conecta Terminal1

Entonces como se aprecia en la

Tabla 4-8 se tienen los documentos recuperados desde las tres fuentes de datos utilizadas. Estos resultados deben ser ponderados, el mecanismo para lograrlo se explica en el capítulo 5. A continuación se explica la interfaz del sistema y el despliegue de estos resultados.

#### 4.10 Interfaz para recuperar información geográfica

Para este trabajo se desarrolló una interfaz para mostrar los resultados de acuerdo a la estrategia de recuperación descrita. Dicha interfaz permitirá formular consultas de tipo triplete, y desplegará los resultados de la recuperación tal y como se describió en las secciones previas. Esta interfaz es una página Web, que despliega un área para formular una consulta. A partir de esta consulta se tienen tres secciones, la primera sección muestra los resultados recuperados por *MatchCon*, cada resultado (documento Web) se

despliega en formato de vínculo, en donde además se ofrece una breve descripción para cada resultado.

Mientras que, la segunda sección se despliega los resultados obtenidos a partir de diccionarios geográficos. Estos resultados se muestran en forma de lista y al estilo de vínculos Web. Cada vínculo muestra las relaciones encontradas, así como las propiedades asociadas a los objetos geográficos indicados en la consulta.

Finalmente, la tercera y última sección muestra los resultados de archivos vectoriales, desplegando sus atributos en formato de tabla, donde algunos campos tiene como valor un vínculo Web, estos vínculos representan la localización (archivo KML) o un *shapefile*®. Ambos están disponibles, de acuerdo al caso, para su descarga.

A continuación se muestra imágenes que despliegan cada una de las secciones descritas. Por ejemplo, la Figura 4-15 muestra el área de consulta inicial, a partir de la cual se mostrarán los resultados.



**Figura 4-15** Interfaz *iGIR* de Búsqueda inicial.

Como se puede apreciar en la Figura 4-15 se tiene un campo en el cual se expresa la consulta, y a partir de esta la interfaz desplegará las opciones y resultados de acuerdo a la estrategia de recuperación.

Por ejemplo, la Figura 4-16 muestra la pantalla que se desplegó para la consulta  $Q_{G4} = \{\text{avenidas conectan aeropuertos}\}$ . En este caso el sistema muestra los resultados encontrados por *MatchCon*, y ofrece al usuario la posibilidad de hacer visualizar con cada vínculo Web los resultados obtenidos por *TopologyFiles*.



The screenshot displays the iGIR web interface. At the top, there is a dark red banner with the text "Laboratorio de Procesamiento Inteligente de Información GeoEspacial" and a small logo on the left. Below the banner, the "iGIR" logo is prominently displayed in large, red, stylized letters. To the right of the logo, the word "Resultados" is written in a smaller font. Below this, a table lists search results. The table has three columns: the first column contains the search terms, the second column contains the word "Conecta", and the third column contains the name of the airport. The search terms are "Circuito Interior" and "Eje 1 norte". The airport name is "Aeropuerto Benito Juárez". Below the table, there is a search bar with the text "Vialidades conectan aeropuertos" and a "buscar" button.

Resultados		
Circuito Interior	Conecta	Aeropuerto Benito Juárez
Eje 1 norte	Conecta	Aeropuerto Benito Juárez

**Figura 4-16** Interfaz *iGIR* después de la exploración de Ontología.

Como se observa en la Figura 4-16 se muestran los resultados obtenidos desde la *GeoOntología*, cada resultado aparece como vínculo Web permitiendo mostrar los documentos de diccionarios geográficos de acuerdo a cada resultado de *MatchCon*.

Entonces, una vez seleccionado sobre cual objeto centrar la recuperación, se despliegan los resultados usando diccionarios geográficos. La Figura 4-17 despliega estos resultados, de acuerdo a la consulta  $Q_{G4}$ .

The screenshot shows the iGIR web application interface. At the top, there is a dark red header with the text "Laboratorio de Procesamiento Inteligente de Información GeoEspacial". Below the header, the iGIR logo is displayed in large red letters. To the right of the logo, there is a search bar containing the text "Vialidades conectan aeropuertos" and a "buscar" button. Below the search bar, there are two sections of results:

**Resultados**

Circuito Interior	Conecta	Aeropuerto Benito Juárez
Eje 1 norte	Conecta	Aeropuerto Benito Juárez

**Resultados de Diccionarios**

- Conecta Camino (L)
- Conecta Carretera (L)
- Conecta Camino (L)
- Conecta Carretera (L)

**Figura 4-17** Resultados en *iGIR* de acuerdo a diccionarios.

Como se aprecia en la Figura 4-17, se muestran los resultados de *Aeropuertos* que conectan con *caminos* y *carreteras*. Estos resultados aparecen debido a que en la *GeoOntología* *caminos* y *carreteras* están vinculados por el concepto padre *Vialidades*. Estas relaciones conducen a los archivos vectoriales que contienen esas relaciones topológicas.

Entonces, al seleccionar alguna de estas relaciones topológicas, se recuperan los archivos vectoriales que cumplen con el requerimiento (relación topológica entre dos objetos). La Figura 4-18 ilustra este proceso.



**Figura 4-18** Resultados *iGIR* usando archivos vectoriales.

Como se aprecia en la Figura 4-18 se tienen algunos vínculos Web, cada uno de ellos puede descargarse en formato KML o de *shapefile* .

En el caso del archivo KML se genera de forma dinámica con el propósito de visualizar la localización del objeto en el visor del programa cliente de *Google Earth*. La Figura 4-19 muestra la interfaz al seleccionar uno de los resultados y descargarlo en formato de archivo *KML*.



**Figura 4-19** Descarga de archivo KML en iGIR.

Como se ilustra en la Figura 4-19 la descarga del archivo KML se realiza al seleccionar uno de los vínculos, en este punto cabe destacar, que los archivos KML se generan únicamente cuando las fuentes de datos vectoriales contengan los atributos, *latitud* y *longitud* para los objetos geográficos en cuestión. De otra forma la opción de descarga en formato KML no estará disponible. La Figura 4-20 muestra un ejemplo de la estructura del archivo KML generado a partir de una consulta.

```
<Placemark>
<name>Aeropuerto Internacional Benito Juárez</name>
<LookAt>
<longitude>-99.07504997106334</longitude>
<latitude>19.43445392429583</latitude>
<altitude>0</altitude>
      <range>3999.999999996305</range>
<tilt>2.438163197949261e-010</tilt>
<heading>0.04768059835391107</heading>
</LookAt>
<styleurl>#default4</styleurl>
  <style>
<IconStyle>
<Icon>
<href>http://maps.google.com/mapfiles/kml/pal2/icon56.png</href>
</Icon>
</IconStyle>
</style>
      <Point>
<coordinates>-99.06789911454179,19.43715441347298,0</coordinates>
</Point>
</Placemark>
```

**Figura 4-20** Ejemplo de archivo KML que despliega el aeropuerto de México, DF.

Mientras que la Figura 4-21 muestra el despliegue del archivo KML con la ubicación geográfica del aeropuerto de la ciudad de México, como uno de los resultados de la recuperación, el despliegue se realiza usando el programa cliente de *Google Earth*.

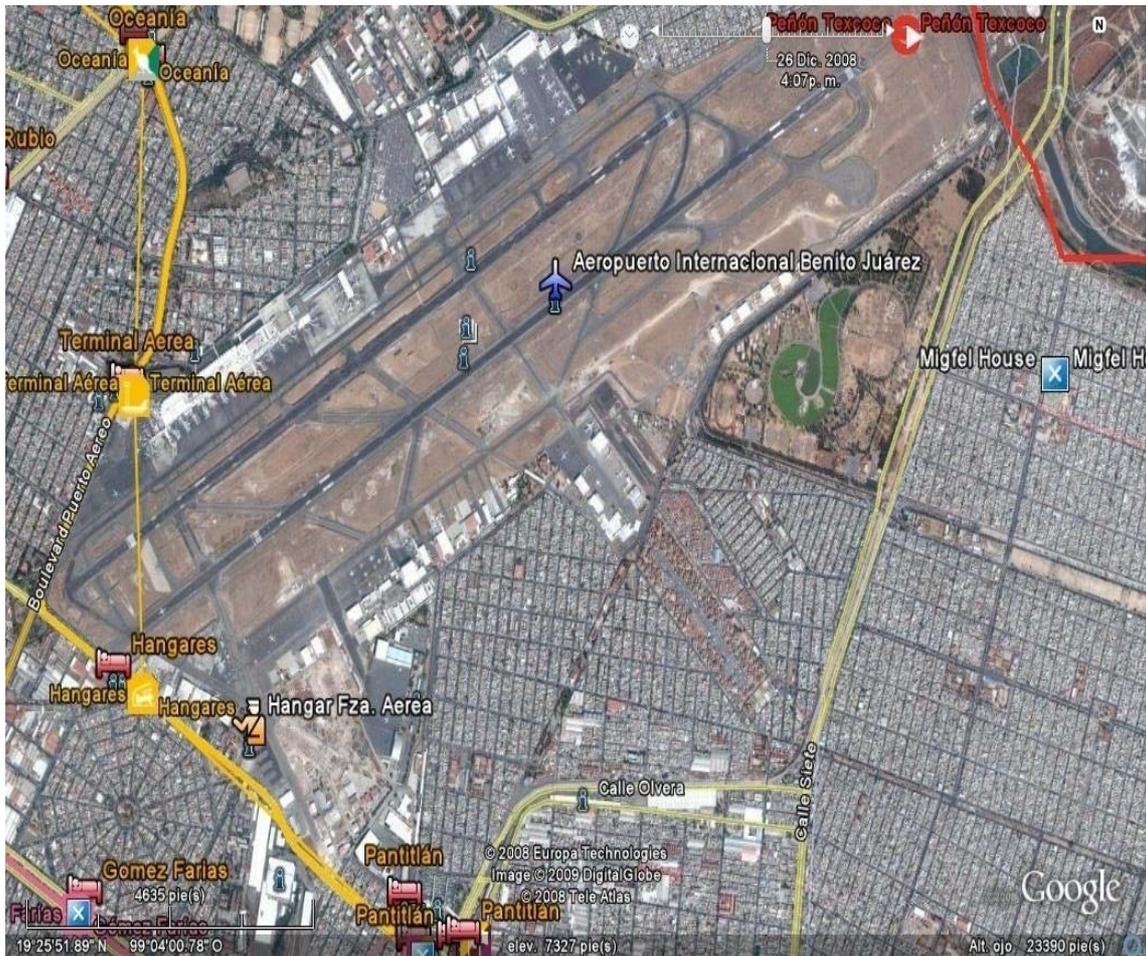


Figura 4-21 Despliegue del archivo KML de la Figura 4-20 en el programa Google Earth.

Como se aprecia en la Figura 4-21 se muestra la ubicación del aeropuerto internacional de la ciudad de México, este es uno de los resultados encontrados para la consulta Q<sub>G4</sub> y con propósitos de ilustración se utiliza el programa *Google Earth* para su visualización.

En la Figura 4-22 se muestra este mismo resultado pero ahora se despliega utilizando el programa *Google Maps* donde incluso se muestran nombres de las colonias que están alrededor del mismo.



**Figura 4-22** Despliegue del archivo KML de la Figura 4-20 en el programa Google Maps.

Como se puede apreciar en la Figura 4-21 y Figura 4-22 la misma información es desplegada con algunos datos adicionales o con diferente perspectiva, las calles y caminos obtenidos por los *shapefiles* no se muestran en estas imágenes debido a que estas capas de datos no cuentan con su posición geográfica, mientras que el aeropuerto si tiene sus atributos de ubicación (latitud y longitud). De esta forma es cómo funciona el sistema *iGIR*.

Como conclusión podemos mencionar que la recuperación de información geográfica se mejora cuando se hace de forma integral que cuando se hace de forma separada. *iGIR* garantiza que se recupera información geográfica que con otros enfoques es omitida.

Por ejemplo, cuando la recuperación se efectúa procesando la topología, únicamente se recupera información topológica y no es posible recuperar atributos geográficos o documentos con datos geográficos. Por ejemplo, la consulta  $Q_{G4} = \{ \text{Vialidades conectan Aeropuertos} \}$  intuitivamente debe responderse con datos geográficos de vialidades y de aeropuertos. Sin embargo, en las *GeoOntologías* y archivos vectoriales las vialidades se describen como calles y caminos, mientras que la relación *conecta* no es procesada espacialmente en documentos Web. Además, los aeropuertos se representan con dos diferentes primitivas de representación (punto o polígono). Estos datos son relevantes para la consulta y solo pueden ser recuperados si se integran de acuerdo a la metodología *iGIR*. En el capítulo 5 se explica el mecanismo de ponderación utilizado.

# CAPÍTULO 5

## METODOLOGÍA DE PONDERACIÓN

*La formulación de un problema, es más importante que su solución.*

*Albert Einstein*

Este capítulo describe a *iRank* un método integral para ponderar los documentos obtenidos por el sistema *iGIR* descrito en la sección 4.4. El método *iRank* integra tres criterios de ponderación para medir la relevancia de un documento geográfico. Tiene como base la métrica de confusión, la cual se adaptó para comparar pares de conceptos geográficos asociados por medio de una consulta y un documento recuperado. En particular, *iRank* utiliza un vector de vecindad obtenido por el algoritmo *OntoExplora* de la sección 4.5. Mediante este vector se integran tres criterios de ponderación basados en la topología, geografía y semántica espacial. El vector de vecindad se procesa y se obtiene un valor de ponderación integral para un documento geográfico.

Este capítulo se estructura como sigue, en la sección 5.1 se explican las diferencias entre la ponderación clásica y la geográfica. Así, como sus alcances y limitaciones actuales. La sección 5.2 describe el método de ponderación de información geográfica, *iRank*. Mientras que la sección 5.3 describe la ponderación geográfica basada en conceptos (*ConceptRank*). Por su parte las secciones 5.4 y 5.5 se encargan de la ponderación basada en la geografía (*GeoRank*) y la topología (*TopologyRank*). La sección 5.6 explica el proceso de integración de estas ponderaciones utilizando como caso de estudio varias consultas geográficas. La principal aportación de *iRank* es la ponderación integral basada en criterios semánticos, geográficos, y topológicos. Los resultados integrales muestran una mejor ponderación comparados con los obtenidos en forma aislada.

## **5.1 Ponderación Clásica y Ponderación Geográfica.**

La relevancia o importancia de un documento, consiste en medir el grado en que un documento responde a una consulta, Para lograrlo, se procesa algún parámetro en común entre la consulta y el documento. Al resultado obtenido de este procesamiento se le denomina valor de relevancia, mientras que a todo el proceso se le conoce como ponderación. Por ejemplo, para un documento de texto, el parámetro más común para medir su relevancia es procesar el número de ocurrencias de una palabra. Mientras que para un documento geográfico deben procesarse parámetros tales como la geografía y la topología.

Por otra parte, cuando muchos documentos responden a una consulta, se requiere clasificarlos de acuerdo a su valor de ponderación. Es decir, por su orden de importancia o relevancia. Este proceso de clasificación es necesario para desplegar los resultados al usuario de acuerdo a la relevancia del documento.

En el ámbito geográfico, como ya se mencionó, la relevancia de un documento se juzga utilizando diferentes criterios de ponderación geográfica. En donde destacan los que procesan las propiedades y relaciones espaciales. Sin embargo, este tipo de procesamiento se efectúa en forma aislada. Es decir, sólo se procesa por un parámetro a la vez (geografía o topología) lo cual limita considerablemente la evaluación de la relevancia del documento. Esto puede resolverse procesando de forma integral las propiedades, relaciones topológicas, y la semántica espacial que caracterizan a un objeto geográfico. Por ello se ha definido el método *iRank*, el cual pondera integralmente documentos de *GeoOntologías*, diccionarios, y archivos topológicos. *iRank* consta de tres etapas, en cada una de ellas se establece la relevancia geográfica que existe entre una consulta y un documento. En la primera etapa, la relevancia se calcula usando conceptos que están almacenados en *GeoOntologías* para documentos Web. En la segunda etapa se utilizan atributos geográficos para documentos obtenidos de diccionarios geográficos. En la última etapa se procesan relaciones topológicas para documentos que provienen de archivos vectoriales.

Uno de los primeros antecedentes para hablar de ponderación de documentos, lo tenemos en la recuperación de información (IR), donde se utilizan métodos de ponderación basados en la palabra y la lingüística. El más destacado es el del modelo vectorial [1] el cual se basa en contabilizar el número de veces que se repite una palabra en un documento, este valor es el parámetro que mide la relevancia del documento. Mientras que en el caso de los documentos Web la relevancia se mide utilizando el número de enlaces que contienen o apuntan a una página Web, la frecuencia de aparición de estos enlaces es uno de los parámetros que determinan la relevancia del documento.

En el caso de GIR, estos criterios son insuficientes, debido a que los objetos geográficos no pueden ser evaluados únicamente por palabras o enlaces Web, ya que tienen otros elementos que los denotan, como la topología y la semántica espacial. Entonces, se requiere que las métricas evalúen los documentos considerando el espacio geográfico. Esto se puede lograr procesando integralmente los rasgos, descripciones, ámbito geográfico y relaciones espaciales de un objeto. Los cuales describen y capturan

ampliamente la esencia geográfica de un objeto y en algunos casos, la forma en que las personas perciben su entorno espacial (e.g. a 5 minutos del hotel). Por lo tanto, mediante esos parámetros se puede calificar la relevancia de un documento en el ámbito de la información geográfica.

Sin embargo, procesar estos elementos requiere enfrentarse a dos aspectos: el primer aspecto se refiere a la heterogeneidad de las fuentes de información (diferentes codificaciones, formatos y representaciones). Mientras que el segundo aspecto se refiere a la semántica espacial (procesamiento y almacenamiento). El primer caso puede ser resuelto empleando enfoques de IR, sistemas de información geográfica (GIS) y XML como formato de interoperabilidad. El segundo caso es más complicado, ya que el significado geográfico de una relación u objeto debe capturarse para poder automatizar su procesamiento. Por ejemplo, el significado de cerca, puede capturarse con base en la percepción del usuario, y teniendo referencias de otros objetos. Así como también, utilizando métricas que evalúan la cercanía con base en parámetros de tiempo, distancia, o perspectiva. Esta es una muestra de lo que representa capturar y procesar la semántica espacial.

Además, la semántica espacial está presente en diferentes fuentes de información, como las GeoOntologías (ontologías geográficas) las cuales almacenan la semántica espacial de un objeto, de acuerdo a su jerarquía geográfica (Por ejemplo, Continente-País-Estado) mientras que otras guardan la semántica de acuerdo a la forma en que un usuario percibe su ámbito geográfico (Por ejemplo, la expresión “cerca de la escuela” puede significar lo mismo que la expresión “a unos pasos de la escuela”).

El panorama se complica todavía más cuando consideramos que los objetos geográficos se almacenan basándose en alguno de sus rasgos característicos. Por ejemplo, los atributos geográficos de un objeto se guardan en diccionarios, mientras que las relaciones espaciales se guardan en archivos vectoriales (dentro, conecta).

Por lo tanto, para calificar la relevancia de un documento en el ámbito geográfico, se requiere procesar los elementos que conforman al documento. Estos elementos son atributos geográficos, formato, codificación y semántica espacial. Mediante la integración de estos aspectos se puede ponderar geográficamente un documento ya que

se tiene como base el entorno geográfico y la forma en que las personas lo perciben, resultando en un proceso de ponderación enriquecido.

Como ejemplo, consideremos que en el sistema *iGIR* las consultas tienen un formato de tipo tripleta y los documentos se describen también en formato de tripleta. Entonces para la consulta: “Avenidas conectan Aeropuertos” se recupera el documento  $D_{G1}$ =“Río churubusco conecta terminal aérea”. Para evaluar la relevancia se debe comparar Río churubusco y avenidas, las relaciones “conecta” y “conectan”. Así como también, aeropuertos y terminal aérea. Esto se puede lograr preguntando a la *GeoOntología* por los conceptos asociados a estos objetos. Sin embargo, también se pueden comparar las propiedades geográficas de cada objeto y de acuerdo a su semejanza determinar su relevancia. Adicionalmente se pueden considerar la topología existente entre estos, como los caminos con los que se conectan. Esta es una muestra de cómo la integración de estas tres relevancias, permite ponderar documentos de acuerdo a la semántica y características geográficas de los objetos involucrados. Además, un documento puede describirse mediante sus relaciones topológicas, y si se pondera únicamente por los conceptos que lo describen, no se podrá obtener un valor que refleje de forma precisa su relevancia. En este caso es cuando la ponderación integral es de utilidad.

Es por ello que proponemos un método de ponderación que considera e integra los parámetros anteriores y que permite evaluar la semejanza entre un documento y una consulta. Esta evaluación, utiliza la teoría de *confusión* [19], la cual es una métrica que evalúa el “parecido conceptual” entre dos conceptos almacenados en una jerarquía. Esta métrica es extendida en este trabajo, para aplicarse usando una *GeoOntología* y evaluar la semejanza conceptual que existe entre dos objetos en el ámbito geográfico. El método de *confusion* evalúa parejas de objetos geográficos, donde el primer objeto pertenece a un documento y el segundo a una consulta. El nivel de semejanza que exista entre estos objetos es el indicador que establece la relevancia del documento.

## **5.2 *iRank*: Ponderación integrando topología, geografía y semántica espacial**

*iRank* es un método integral que pondera documentos geográficos( $D_G$ ) recuperados a partir de una consulta geográfica ( $Q_G$ ). En donde la relevancia se mide utilizando tres

parámetros que describen el espacio geográfico, estos son: la topología, la semántica, y los atributos geográficos.

La técnica de *iRank* consta de tres etapas, en cada una de ellas se compara una consulta y un documento para establecer su relevancia. En la primera etapa la comparación se lleva a cabo procesando conceptos geográficos. En la segunda etapa, se procesan atributos geográficos para efectuar la comparación. Mientras que para la última etapa, se procesan relaciones topológicas para poder realizar la comparación. Estas fases se integran a través de un vector de vecindad (explicado en la sección 4.5). La meta final consiste en obtener un valor de relevancia integral mediante los parámetros descritos.

Los documentos geográficos a ponderar pertenecen a tres fuentes heterogéneas de información geográfica, las cuales son *GeoOntologías*, archivos topológicos y diccionarios de datos toponímicos. Cada una de estas fuentes almacena los objetos espaciales con un formato y codificación diferente (sus características espaciales). Por ello para medir la relevancia entre estos objetos, se procesan usando conceptos que los describen (su semántica espacial). Este mecanismo es muy importante, ya que de esta forma no nos enfrentamos a la heterogeneidad de los datos y a la posible ambigüedad en los nombres de los mismos. De esta manera, se facilita la tarea de comparación entre consultas y documentos, y en consecuencia de la asignación de la relevancia de un documento.

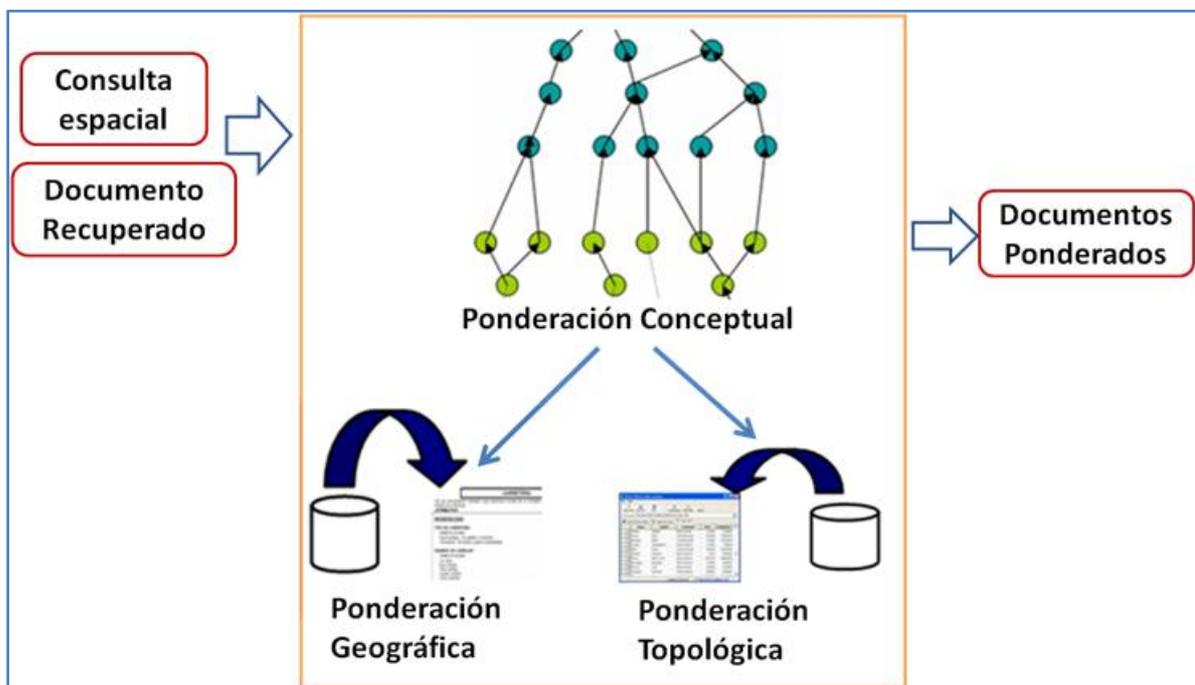
Para reforzar esta idea, consideremos como se almacena una carretera en estas fuentes de datos. Por ejemplo, en los diccionarios se almacenan sus propiedades geográficas, mientras que en un *TopologyFile* se almacenan sus relaciones topológicas. Realizar una comparación entre ellos es complicado, ya que no tienen parámetros que coincidan entre sí.

Por lo tanto utilizar su representación semántica permite medirlos aún cuando su codificación y formato sean diferentes. Debido a que el concepto carretera es el mismo tanto en diccionarios como en *TopologyFiles*.

Entonces, para comparar objetos mediante sus conceptos correspondientes, se utilizó la métrica de *confusion*, la cual es una métrica que arroja como resultado un valor que

indica que tanto se diferencia o se parece un concepto de otro. Por ejemplo, si se pregunta por *carreteras* y se obtuvieron *caminos*, se evalúa si un *camino* significa lo mismo que una *carretera*. Para llevar a cabo esta comparación se utiliza una *GeoOntología* que almacena estos conceptos geográficos (camino y carretera). La *GeoOntología* se explora y se determina la distancia conceptual que existe entre estos conceptos (el número de nodos que deben recorrerse para llegar desde el concepto camino hacia el concepto carretera). Este valor es el primer indicador de la relevancia.

En la Figura 5-1 se muestra el marco de trabajo de *iRank* y los elementos que lo conforman.



**Figura 5-1** Marco de trabajo de *iRank*.

Como se puede apreciar en la Figura 5-1, el proceso de ponderación tiene como entrada la consulta y el documento recuperado, estos se comparan entre sí y se obtiene como resultado el documento ponderado. La comparación inicia identificando cada elemento (*que, relación, donde*) tanto en la consulta y el documento. Después, se explora la *GeoOntología* para encontrar el concepto asociado a cada elemento de la tripleta. En este punto, el elemento puede ser identificado si está escrito tanto en plural como en singular. Por ejemplo, el objeto *Lago* o *Lagos* tiene asociado el concepto *Cuerpo de Agua*.

De esta forma, una vez identificado el concepto se extrae su vecindad (los conceptos enlazados por las relaciones de hiperonimia y/o meronimia). La cual se almacena en un vector de vecindad ( $V_c$ ). De tal manera que, si se recuperó el documento <Río Churubusco conecta terminal aérea>, para la consulta “Avenidas conectan Aeropuertos”, se procede a localizar los conceptos asociados al elemento <que> tanto de la consulta como del documento, que en este caso son <Río Churubusco> y <avenidas>. Una vez encontrados, se mide la distancia conceptual ente ellos (a nivel de nodos en la *GeoOntología*) al resultado obtenido se le llama relevancia conceptual (*RelCon*).

Una vez obtenida la relevancia conceptual (*RelCon*), se procede a medir la relevancia topológica (*RelTopology*) y la relevancia geográfica (*RelGeo*) que existe entre la consulta y el documento. Estas relevancias se integran y se obtiene un valor de relevancia final, este proceso de integración se explica en la sección 5.6.

### **5.3 Ponderación Conceptual**

El módulo de ponderación conceptual (conceptual Ranking), es la primera fase de *iRank*, consiste en evaluar la relevancia de un documento a través de los conceptos que lo representan. Para ello, se utilizan *GeoOntologías* que almacenan conceptos geográficos, relaciones espaciales y atributos geográficos. La descripción completa de la construcción de las *GeoOntologías* está en la sección 4.1. La Figura 5-2 muestra un fragmento de una *GeoOntología*.

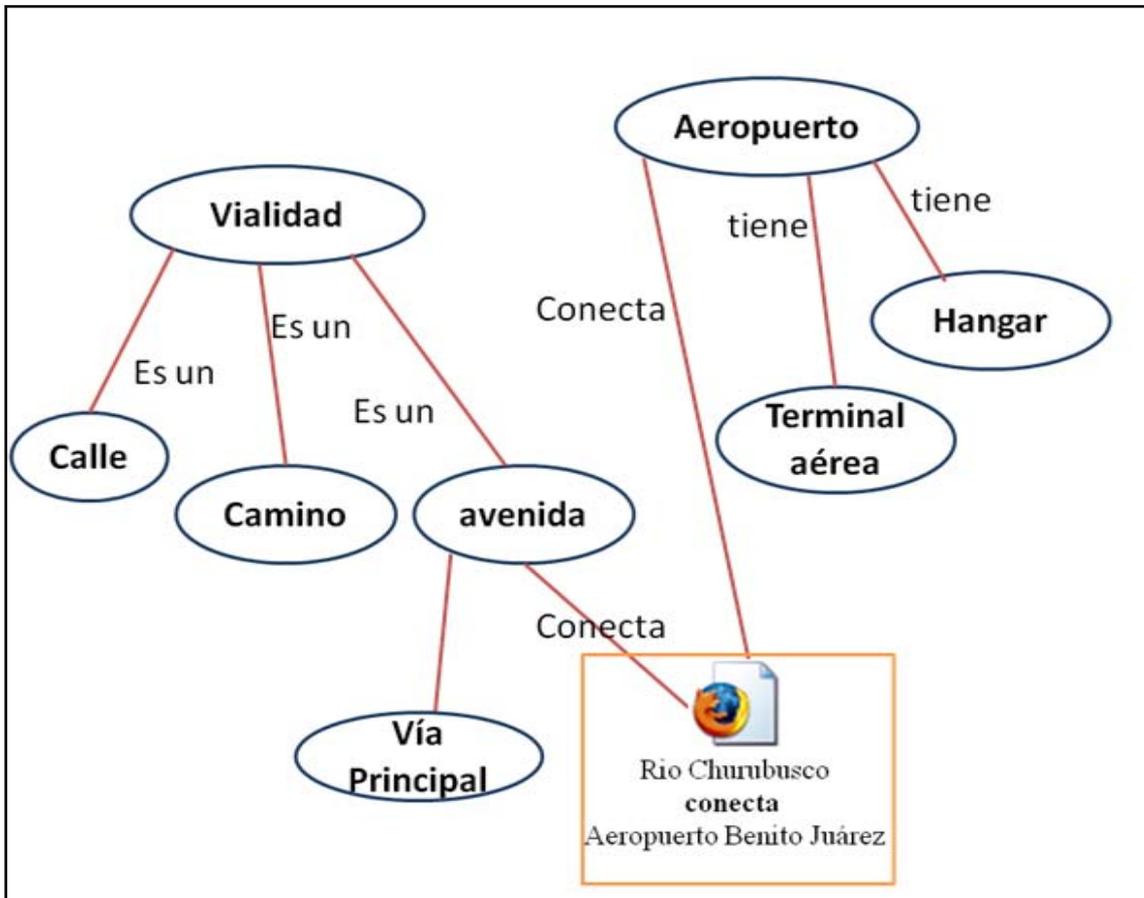


Figura 5-2 Fragmento de *GeoOntología*.

Tal y como se muestra en la Figura 5-2 cada concepto de la *GeoOntología* tiene diversas relaciones y propiedades, así como documentos asociados, donde cada uno de ellos tiene un peso inicial ( $W_i$ ). Este valor se obtuvo al procesar un conjunto de documentos de *Google*, *Yahoo!* y *Yahoo! Answers*<sup>43</sup>, utilizando un enfoque similar al del modelo vectorial [9]. Es decir, para cada consulta utilizada en este trabajo, se le preguntó a *Google* y *Yahoo! Answers*, y de los documentos obtenidos se seleccionaron, semiautomáticamente (usando un programa y de forma manual) aquellos que contenían un nombre de lugar, o el nombre de una etiqueta asociada a un concepto de la *GeoOntología*. Por ejemplo, cuando se pregunta por *Avenidas conectan Aeropuertos* para sitios de México, la mayoría de las respuestas de *Google* se referían o incluyen a la avenida <Río Churubusco> o <circuito interior>. Por lo tanto, los documentos referidos a <Río Churubusco> o <circuito interior>, se consideran de mayor relevancia (peso)

<sup>43</sup> <http://mx.answers.yahoo.com/>

para las consultas que involucren a los objetos *avenidas* y *Aeropuertos* por la relación conecta. La fórmula 5.1 indica cómo se obtiene este peso inicial.

$$W_i = \frac{F_t}{N_d} \quad (5.1)$$

Donde  $W_i$  es el peso del concepto,  $F_t$  es frecuencia de aparición del término asociado al concepto, en un documento. Por su parte,  $N_d$  es el número de documentos utilizados.

La fórmula es normalizada al intervalo de [0,1] donde 1 representa el valor de relevancia máxima, mientras que cero es el valor de relevancia mínima.

Por otra parte, los documentos obtenidos por *Yahoo! Answers* y *Google* son asociados a los conceptos de la *GeoOntología* por medio de relaciones semánticas. Por ejemplo, el documento <Río Churubusco conecta aeropuerto Benito Juárez> está asociado al concepto *avenida* y al concepto *aeropuerto* a través de la relación semántica asociada a *conecta*. Los criterios para asociar los documentos con los conceptos de la *GeoOntología*, se basan en las etiquetas de las propiedades y relaciones que describen a los conceptos. Por ejemplo, un documento que se refiere al aeropuerto está asociado al concepto *aeropuerto* a través de la etiqueta *aeropuerto*. Este proceso se realiza de forma semiautomática, utilizando un programa desarrollado en Ruby [29] el cual recupera documentos desde *Yahoo! Answers* y *Google*. En este punto, de forma manual se asocian con el concepto correspondiente de la *GeoOntología* de acuerdo al criterio del analista GIS.

Explicaremos a continuación cómo se calcula la relevancia conceptual, considerando el siguiente escenario: un analista GIS requiere hacer un análisis de afectación por las obras que se realizarán en una de las avenidas que conectan con el aeropuerto, por lo cual necesita datos geográficos adicionales de calles y otros caminos que conduzcan o se conecten con el aeropuerto. Para lo cual necesita datos geográficos de estos objetos.

Entonces expresa esta necesidad mediante la siguiente consulta:  $Q_{G4} = \{\text{vialidades conectan aeropuertos}\}$ .

El proceso inicia expresando la consulta, y consta de cuatro pasos:

- 1.- Analizar la consulta para identificar cada elemento de la tripleta.
- 2.- Identificar los conceptos asociados para el documento y la consulta.
- 3.- Extraer el vector de vecindad para el documento y para la consulta.
- 4.- Procesar los pesos y calcular la confusión conceptual.

El primero paso, permite saber cuál es el elemento que, relación y donde de la consulta y del documento.

El segundo paso utiliza el algoritmo *OntoExplora* (ver sección 4.5) para encontrar en los conceptos correspondientes para cada elemento de la tripleta. Por ejemplo, *OntoExplora* encuentra asociado al documento “circuito interior”, al concepto “avenida” a través de la relación de hiperonimia.

Entonces, en el caso de  $Q_{G4} = \{\text{vialidades conectan aeropuertos}\}$  se busca el concepto asociado a “vialidades”, “conectan” y para “aeropuertos”. Para lo cual se utilizan las etiquetas de los conceptos, tanto en forma singular como en plural (e.g. aeropuerto ó aeropuertos). Por ejemplo, la relación “junto” está asociada al concepto “adyacencia”.

El tercer paso, extrae la vecindad de la consulta y del documento usando la *GeoOntología*. Para este caso, tenemos que el concepto *vialidades* está asociado con el elemento *que*= “avenidas” de  $Q_{G4}$ . Mientras que para el documento  $D_{G1} = \langle \text{Río Churubusco conecta aeropuerto} \rangle$ , el elemento  $\langle \text{Río Churubusco} \rangle$  está asociado al concepto  $\langle \text{avenida} \rangle$ , mientras que  $\langle \text{aeropuerto} \rangle$  está vinculado a “aeropuerto Benito Juárez”, y la relación  $\langle \text{conecta} \rangle$  está asociada con *conectan*. De esta forma, el vector de vecindad ( $V_c$ ) es llenado con estos elementos. En la Figura 5-3 se muestra la *GeoOntología*, la consulta, el documento geográfico y el vector de vecindad descrito en este párrafo.

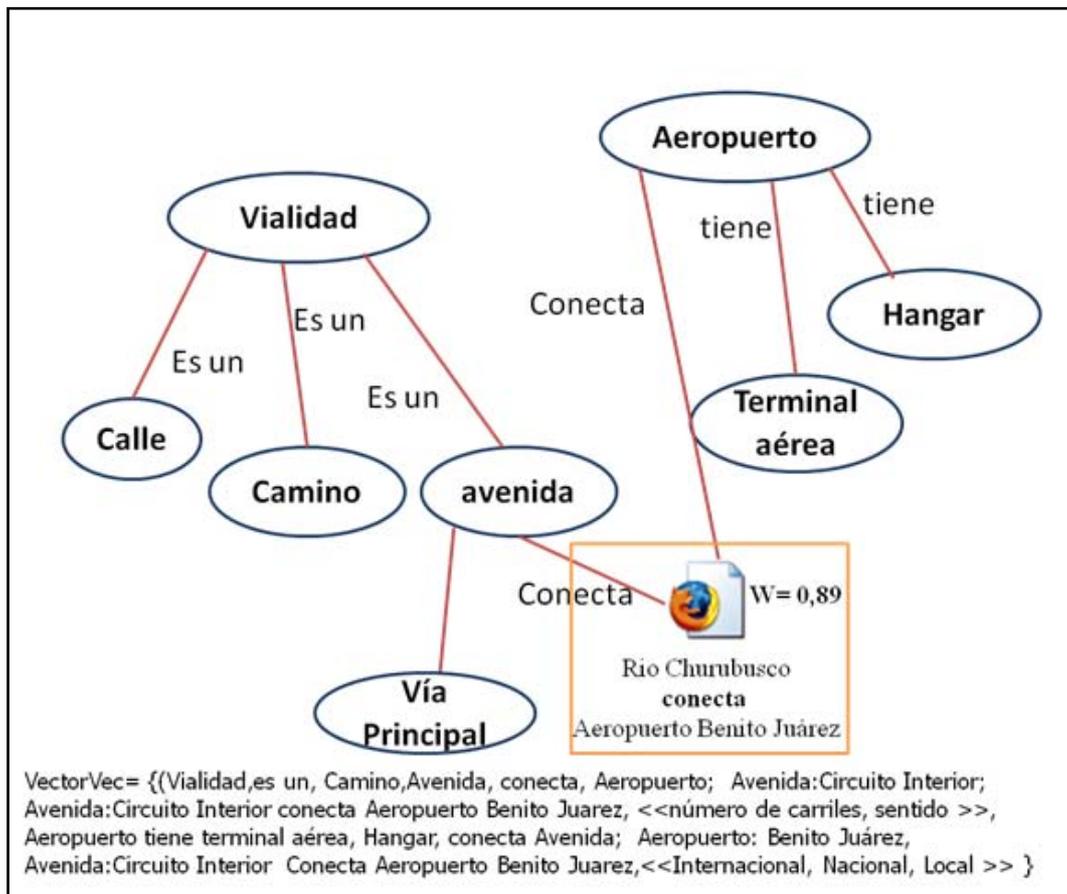


Figura 5-3 Generación del Vector de Vecindad.

Tal y como se aprecia en la Figura 5-3 el vector de vecindad, contiene los conceptos vecinos de la consulta. Se destaca el peso asociado del documento, denotado por la letra  $W$  para indicar su relevancia inicial. El cuarto y último paso, consiste en determinar la relevancia conceptual entre <avenidas conectan aeropuerto> y <Río Churubusco conecta aeropuerto Benito Juárez> para ello se aplica la Fórmula 5.2. Al valor obtenido se denomina *confusion conceptual* ( $C_c$ ).

$$Cc(ci, cj) = \frac{Wi}{D} \quad (5.2)$$

Donde,  $ci$  es el concepto de un elemento de la consulta, y  $cj$  es el concepto asociado a un elemento del documento,  $Wi$  es el peso del documento. Mientras que  $D$  es la distancia entre nodos que existe entre  $ci$  y  $cj$ .

De esta forma, al aplicar la fórmula 5.2 para el ejemplo de  $Q_{G4}$ , tenemos:

$$Cc(\textit{avenidas,rio churubusco})=\frac{W_{\textit{avenidas}}}{D}$$

$$Cc(\textit{conectan,conecta})=\frac{W_{\textit{conecta}}}{D}$$

$$Cc(\textit{aeropuertos,aeropuerto Benito Juarez})=\frac{W_{\textit{aeropuerto}}}{D}$$

Es así como se obtiene la *confusion conceptual* que existe entre el documento  $D_{G1} = \langle \text{Río Churubusco conecta aeropuerto Benito Juárez} \rangle$  y la consulta {“avenidas conectan aeropuertos”}. Finalizando de esta forma la primera fase de la ponderación.

La siguiente fase, consiste en utilizar el vector de vecindad obtenido en esta etapa, para ponderar documentos recuperados de las otras dos fuentes de información (*TopologyFiles* y *Diccionarios*). Este proceso se explica en la sección 5.4 (Ponderación Geográfica) y la sección 5.5 (Ponderación Topológica).

#### 5.4 Ponderación Geográfica

El módulo de ponderación geográfica, es la segunda fase de *iRank*. Su objetivo es establecer un valor de relevancia geográfica para un documento. Donde los documentos fueron recuperados de diccionarios geográficos. El proceso de ponderación se realiza comparando un documento y una consulta. El proceso de comparación se denomina *confusión geográfica* ( $C_G$ ) ya que compara los objetos geográficos (de la consulta y documento) de acuerdo a sus atributos geográficos.

El proceso calcula la  $C_G$  utilizando el vector de vecindad obtenido en la sección 4.6 y las propiedades geográficas de los objetos involucrados. El resultado final es la relevancia geográfica (*RelGeo*).

Por ejemplo, la consulta  $Q_{G4} = \{ \text{Vialidades conectan Aeropuertos} \}$  recuperó documentos que cumplen con los criterios geográficos. Por ejemplo: *<caminos conectan aeropuertos>*. Entonces, para evaluar la relevancia de los documentos que cumplen dichos criterios se aplica un proceso que consta de dos pasos:

- 1.- Se forman parejas entre los objetos geográficos, el primer objeto corresponde a la consulta y el segundo objeto corresponde al documento.
- 2.- Utilizando el vector de vecindad, se establece la confusión geográfica entre cada pareja de objetos, promediando su valor conceptual y el valor de comparación de sus propiedades geográficas.

Para explicar el proceso, consideraremos la consulta  $Q_{G4} = \{ \text{Vialidades conectan Aeropuertos} \}$  y la pareja de documentos recuperados de acuerdo a los siguientes criterios:  $D_{G2} = \langle \text{caminos conectan aeropuertos} \rangle$  y  $D_{G3} = \langle \text{carreteras conectan aeropuertos} \rangle$ . Al aplicarse el punto 1 para el primer documento, tenemos la pareja de objetos *<Vialidades>* vs “caminos”, la pareja de relaciones *<conectan>* vs “conectan”, y finalmente el par de objetos *<aeropuertos>* vs “aeropuertos”.

De esta manera, al aplicar el segundo paso *<caminos>* tiene asignado una relevancia conceptual ( $W_c$ ) de 0.87 mientras que “avenidas” tiene un  $W_c$  de 0.76. Se hace el promedio entre estos valores, y entre las propiedades definidas en la *GeoOntología* y las descritas en el diccionario geográfico, obteniéndose 0.81 como resultado y este es el valor de relevancia para el elemento *que*. Mientras que *<aeropuertos>* y la relación *conecta* tienen un  $W_c$  de 1 ya que se trata del mismo concepto. Se promedian estos valores aplicando la fórmula 5.3, el resultado es la relevancia geográfica (*RelGeo*).

$$Cc(Q,D) = \frac{Wcg1+Wcg2+Wcg3}{3} \quad (5.3)$$

Donde  $Wcg$  es el peso obtenido al procesar el peso conceptual y la comparación entre las propiedades geográficas de cada objeto geográfico.

Por lo tanto al aplicar la formula 5.3 al primer resultado se obtiene:

$$Cc(Q,D) = \frac{0.81+1+1}{3} = 0.93$$

Mientras que para el segundo resultado se obtiene:

$$Cc(Q,D) = \frac{0.71+1+1}{3} = 0.90$$

La Figura 5-4 muestra el proceso descrito, para la consulta  $Q_{G4}$  y la pareja de documentos comentada.

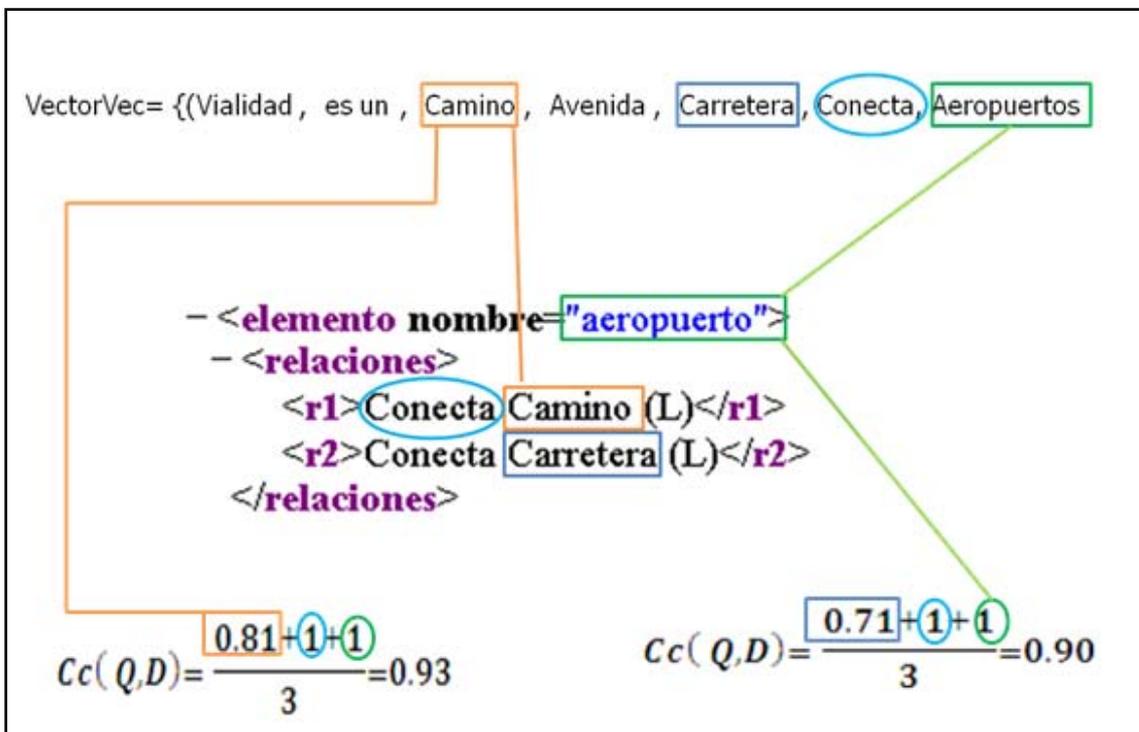


Figura 5-4 GeoRanking aplicado en Dictionarios.

Como se aprecia en la Figura 5-4 se tiene el vector de vecindad y sus valores de relevancia. Así es como se obtiene el valor de ponderación geográfica para cada documento. Finalizando de esta forma la segunda fase de la ponderación. La siguiente fase, consiste en ponderar documentos recuperados de la última fuente de información (*TopologyFiles*). Este proceso se explica en la sección 4.8 (Topological Ranking).

## 5.5 Ponderación Topológica

El módulo de ponderación topológica, es la tercera fase de *iRank*, su objetivo consiste en establecer la relevancia topológica de un documento. Para ello se combinan criterios que evalúan y procesan relaciones topológicas. En este módulo se evalúan los documentos recuperados utilizando *TopologyFiles*.

Ahora, para evaluar la relevancia, se clasifican las relaciones espaciales en tres grupos de acuerdo a lo que se define en [1] y en [26]. El primer grupo se refiere a relaciones espaciales de Inclusión (si un objeto A está contenido en un objeto B), el segundo a Proximidad (que tan cerca está el objeto A del objeto B), y el tercero a Hermandad (dos conceptos son hermanos si tienen el mismo padre). A continuación se definen las reglas que permiten evaluar estos aspectos y posteriormente calificarlos con un valor de relevancia.

### 5.5.1 Inclusión (Inclusion)

Verifica si  $Sd$  está dentro de  $Sq$ , donde  $Sd$  es el área ó extensión geográfica del documento, mientras que  $Sq$  es el área geográfica del query). Por ejemplo, el  $Sq$  de la consulta  $Q_{G4}$  es la ciudad de México y el  $Sd$  del documento  $D_{G1}$  es el aeropuerto Benito Juárez, entonces se verifica que el aeropuerto Benito Juárez este dentro de la ciudad de México. Además, se evalúa el grado de *confusión topológica* que existe entre ambos elementos. Para ello, se emplea el número de descendientes que cada concepto tiene en la *GeoOntología*. Esto se procesa aplicando la fórmula 5.4 que se define a continuación.

$$Inclusion(Sq, Sd) \begin{cases} \frac{NumDescendientes(Sd)+1}{NumDescendientes(Sq)+1} & \text{if } Sd \sqsubset Sq \\ 0 & \text{En otro caso} \end{cases} \quad (5.4)$$

La fórmula 5 regresa valores en el intervalo [0,1] el valor máximo resulta cuando ambos tienen el mismo número de descendientes ( $Sd$  está dentro de  $Sq$ ) y el mínimo cuando  $Sd$  no tiene descendientes.  $NumDescendientes(S)+1$  es el número de alcances dentro de  $S$  más el alcance en sí mismo (es decir, las relaciones “sub-region-of” en la *GeoOntología*).

### 5.5.2 Hermanos (Siblings)

Es una función binaria que verifica si  $Sq$  y  $Sd$  son hermanos en la *GeoOntología*. Por ejemplo, avenida y camino tienen el mismo padre y por lo tanto son hermanos. El valor máximo resulta cuando los elementos son hermanos, y el mínimo cuando no son hermanos. La fórmula 5.5 se aplica para este proceso.

$$Hermanos(Sq, Sd) = \begin{cases} 1, & \text{if } \exists Sx: padre(Sq) = Sx \wedge padre(Sd) = Sx; \\ 0, & \text{de otra forma} \end{cases} \quad (5.5)$$

### 5.5.3 Proximidad (Proximity)

Es la inversa de la distancia euclidiana que existe entre dos objetos, donde el primer objeto pertenece a la consulta, y la segunda al documento. Esto se define en la fórmula 5.6.

$$Proximidad(Sq, Sd) = \frac{1}{1 + \frac{Distancia(Sq, Sd)}{Diagonal(Sq)}} \quad (5.6)$$

Donde  $Sq$  es el alcance geográfico de la referencia geográfica de la consulta y  $Sd$  es el alcance geográfico del objeto descrito por el documento. Por ejemplo, para  $Q_{G4}$  el alcance es la ciudad de México, debido a que es en esta ciudad donde se encuentran líneas del metro. En adición, la distancia euclidiana es normalizada por la diagonal del MBR<sup>44</sup> (Minimum Bounding Rectangle) definido para el área geográfica de la consulta.

---

<sup>44</sup> Los MBR's facilitan el procesamiento de datos vectoriales y de las relaciones topológicas entre ellos.

Un MBR es el rectángulo mínimo que cubre el área geográfica de un objeto, donde este objeto puede ser de forma irregular o regular. En la Figura 5-5 se muestra un ejemplo del MBR generado para el aeropuerto Benito Juárez, en México, DF.



**Figura 5-5** MBR del aeropuerto Benito Juárez.

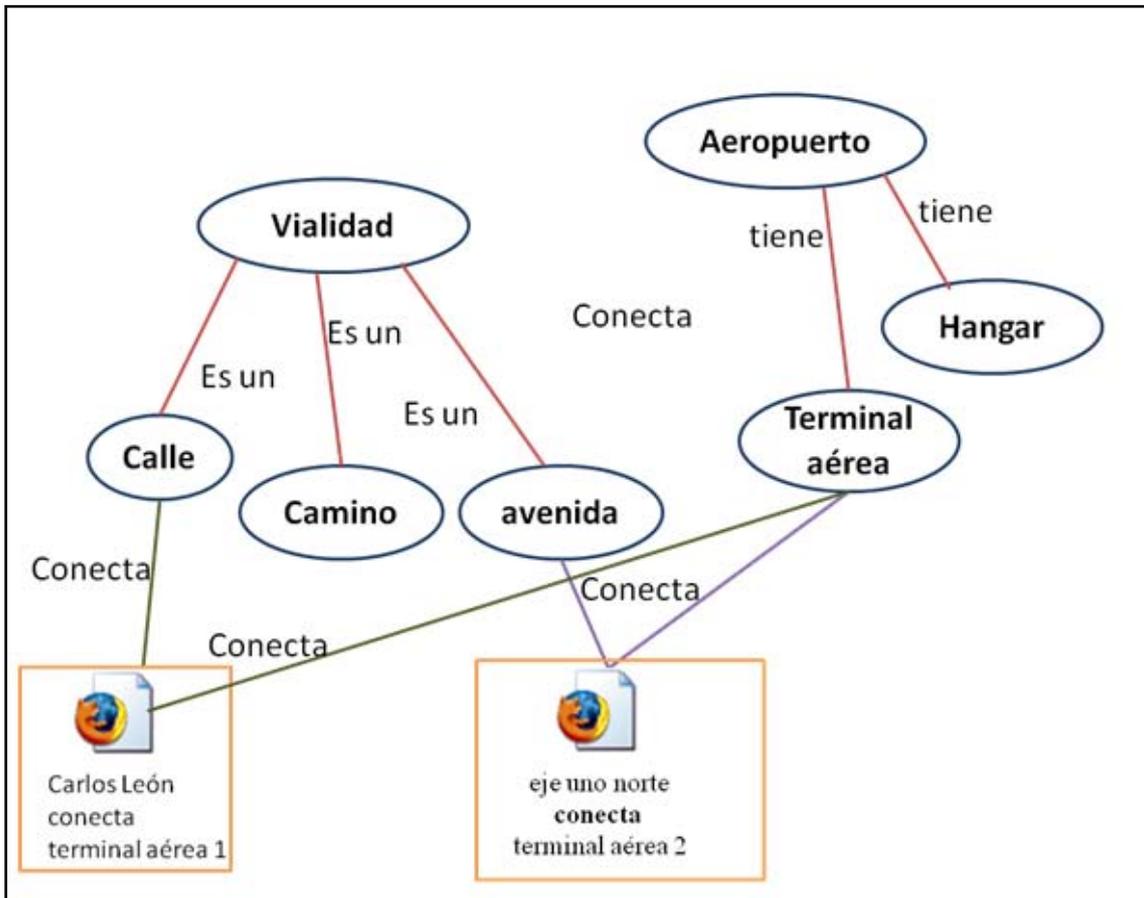
En la Figura 5-5 se muestra el área geográfica del aeropuerto Benito Juárez en México DF, el MBR es el rectángulo de color azul marino, el cual encierra totalmente la forma irregular del aeropuerto. También, se muestra la diagonal (en color rojo) de dicho MBR.

De esta forma, una vez que se han definido y descrito las fórmulas que se utilizan para ponderar los *TopologyFiles*. Se procede a explicar el procedimiento con el siguiente ejemplo:

Considerando la consulta  $Q_{G4} = \{\text{Vialidades conectan Aeropuertos}\}$  y el par de documentos recuperados,  $D_{G4} = \langle \text{eje uno norte conecta terminal aérea 2} \rangle$  y  $D_{G5} = \langle \text{Carlos León conecta terminal aérea 1} \rangle$ . Donde eje uno norte y Carlos Leon son nombres de avenidas y calles respectivamente. Ahora, procedemos a comparar estos documentos y la consulta para obtener su relevancia topológica que existe entre la consulta y el documento. Este proceso consta de cuatro pasos:

- 1.- Determinar si los objetos tienen el mismo concepto padre (usando *OntoExplore*).
- 2.- Extraer el alcance geográfico del documento y la consulta, para evaluar la proximidad, inclusión y hermandad entre ambos.
- 3.- Realizar una operación de sobreposición, entre sus alcances geográficos.
- 4.- Utilizando el resultado de la sobreposición, en conjunción con el de inclusión, hermandad, y proximidad se determina el valor de relevancia topológica.

El primer paso utiliza la *GeoOntología* para saber cuál es el concepto padre que representa al objeto. Por ejemplo, para  $D_{G4}$  el concepto *avenidas* es el que representa a eje uno norte. Y como en la consulta, se pidió *avenidas*, entonces se trata del mismo concepto. Por lo tanto los objetos geográficos tienen una *confusion* = 0, es decir, son idénticos conceptualmente porque tienen al mismo padre. La Figura 5-6 muestra gráficamente los documentos y la consulta, así como su asociación al explorar la *GeoOntología*.



**Figura 5-6** Identificando conceptos y documentos en *la GeoOntología*.

El segundo paso, consiste en extraer los alcances geográficos de la consulta y de los documentos. Para el caso de la consulta  $Q_{G4}$  se extrae el área de la ciudad de México. Mientras que para el documento  $D_{G4} = \langle \text{eje uno norte conecta terminal aérea 2} \rangle$  se extrae el área del aeropuerto. En seguida se verifica la inclusión, grado de proximidad y hermandad para ambos objetos. Para el caso de la inclusión, se obtiene que si existe inclusión (la terminal aérea 2 está dentro del aeropuerto) el grado de proximidad es igual a la unidad, mientras que la hermandad no se cumple ya que terminal aerea no es concepto hermano de aeropuerto. Entonces, teniendo los resultados de estas operaciones, se determina su relevancia que hasta este paso es de 0.90, es decir la relevancia es casi completa.

En el tercer paso, se verifica si existe sobreposición entre dos objetos, de forma que si no existe sobreposición, se determina que no existe relevancia topológica. Es decir, el valor de relevancia topológica es cero. En el caso de que si exista sobreposición, se considera el tamaño del área geográfica que se sobrepone, y el valor de dicha área determina su relevancia. Esta operación se realiza a nivel de registros en una tabla donde están almacenados los MBR de cada objeto. Para nuestro ejemplo, en la Figura 5-7 se muestran las áreas sobrepuestas de la terminal aérea 1 y la terminal aérea 2 con la del aeropuerto Benito Juárez de la ciudad de México. Los cuales son los objetos geográficos que representan a los documento  $D_{G4}$ ,  $D_{G5}$ . Como se observa se despliegan en color naranja y azul estas áreas y como son cubiertas totalmente por el MBR del aeropuerto, entonces todo el tamaño de sus áreas determina la relevancia de los respectivos documentos.



**Figura 5-7** Sobreposición para Ponderación topológica

Finalmente, en el cuarto y último paso, se organizan los resultados de acuerdo al área sobrepuesta, en forma ascendente o descendente. El proceso descrito se aplica para las relaciones asociadas a proximidad, en donde de acuerdo a la relación involucrada, se aplican las funciones definidas previamente, para obtener la relevancia topológica. En esta investigación, se trabajaron las relaciones: “dentro” y “conecta” aplicadas a escenarios de prueba con los datos disponibles en diccionarios del INEGI y en datos generados en nuestro laboratorio. Como trabajo a futuro se considera utilizar siete relaciones topológicas en diferentes casos de estudio teniendo como base el modelo “9-intersección”.

## 5.6 Integración de la ponderación

Para la integración de los resultados, consideraremos que todo el proceso fue conducido por las *GeoOntologías*. Donde el vector de vecindad se utilizó para ponderar los documentos de acuerdo a cada una de las tres fuentes de información. Por lo tanto, el promedio de estas ponderaciones es el valor de ponderación final para cada documento.

La función que calcula el valor de ponderación final dada una consulta  $Q_G$  para cada documento  $D_G$  se muestra en la Formula 5.8.

$$RelInt = \frac{RelCon(Cq,Cd) + RelGeo(Gq,Gd) + RelTopoly(Tq,Td)}{\text{Número de fuentes geograficas}} \quad (5.8)$$

*RelInt* es la relevancia integral, la cual se mide en el intervalo de valores  $0 \Rightarrow RelInt \leq 1$ , donde el valor 1 representa la relevancia completa y el valor cero significa relevancia nula. Por ejemplo, se pregunto por “avenidas”, y se respondió con “avenida eje uno norte”, entonces su relevancia es completa ya se obtuvo una avenida como respuesta. Mientras, que si se pregunta por Lago texcoco y se respondió con carretera texcoco, su relevancia es nula, ya que el objeto carretera no es un lago, ni tampoco son similares entre sí. De esta manera se ponderan los resultados, se clasifican por su categoría (semántica, topológica o geográfica) y se despliegan al usuario.

Mostramos ahora los resultados de aplicar la metodología iRank para la consulta  $Q_{G4} = \{\text{Vialidades conectan Aeropuertos}\}$ .

**Tabla 5-1** Resultados ponderados para  $Q_{G4} = \{\text{Vialidades conectan aeropuertos}\}$ .

Posición en iRank	Documento pertenece a fuente de datos	Documento recuperado usando iGIR	Ponderación en GEO RANK	Ponderación en TOPOLOGY RANK	Ponderación en CONCEPT RANK	Ponderación en iRank	
1	Topology File	Boulevard puerto aéreo conecta Aeropuerto Benito Juárez	0.75	0.84	0.96	0.85	Totalmente Relevante
2	Topology File	Calle de Carlos León conecta Aeropuerto Benito Juárez	0.68	0.95	0.87	0.83	Totalmente relevante
3	Topology File	Eje uno norte conecta Terminal 1	0.59	0.76	0.81	0.72	Casi relevante
4	Topology File	Calle sonora conecta Terminal 2	0.51	0.82	0.81	0.71	Casi relevante
5	GeoOntología	“Conecta al Viaducto Piedad y Río Churubusco con la Terminal 2”	0.61	0.38	0.90	0.63	Casi relevante
6	GeoOntología	“El Distribuidor Vial 2, que conecta a Viaducto Piedad y Río Churubusco con la Terminal 2...”	0.60	0.36	0.90	0.62	Casi relevante
7	Diccionario	Carretera conecta aeropuerto	0.65	0.46	0.36	0.49	Medio relevante
8	Diccionario	Calle conecta aeropuerto	0.63	0.43	0.32	0.46	Medio Relevante
9	Topology File	Distribuidor vial 1 conecta Terminal1	0.21	0.70	0.46	0.45	Medio relevante
10	GeoOntología	Vialidades congestionadas conectan Aeropuerto	0.14	0.12	0.46	0.24	Poco Relevante

Como se puede observar se tienen resultados de las tres fuentes de datos. En este caso para la consulta  $Q_{G4}$  predominaron los resultados provenientes de la fuente de datos denominada *TopologyFiles* esto se debe a que la relación conecta es dominante en esta fuente de datos.

La Tabla 5-1 muestra las ponderaciones otorgadas por *GeoRank*, *TopologyRank*, y *ConceptRank*. La última columna es el valor proporcionado por *iRank* y de acuerdo a este valor son desplegados los resultados. Como se observa varios documentos reciben valores de relevancia opuestos. Es decir, con un criterio la relevancia es alta mientras que con otro criterio la relevancia es media o baja (ver resultado número seis de la Tabla 5-1). *iRank* integra estas relevancias para que los documentos sean ponderados de forma global y no por un único criterio, reduciendo la posibilidad de que un documento relevante por su topología sea descartado porque fue evaluado por sus propiedades geográficas. Con ello se logra mayor precisión en la ponderación y un mayor grado de certidumbre. Por otro lado, cabe destacar, que de los documentos recuperados, se conoce a priori cuales documentos son relevantes. iGIR recupero un 75% de ellos para la consulta  $Q_{G4}$ . El restante 25 % representa resultados con relevancia mínima o nula para la consulta. Esto ocurre porque cuando no se encuentra al objeto geográfico buscado de forma directa, entonces se recuperan objetos relacionados por otras relaciones o atributos del vector de contexto.

Otra situación a discutir es la ponderación de los documentos ubicados en el lugar tres y cuatro. *GeoRank* los pondera con una relevancia media, mientras que *ConceptRank* y *TopologyRank* los califican con una relevancia alta (casi relevante). Esto implica que un resultado relevante sea omitido porque carece de propiedades geográficas. Sin embargo, posee relaciones topológicas y semánticas que lo ubican como relevante. La integración de estas medidas permite evaluarlo con mayor precisión, lo cual resulta en una mejor ponderación de acuerdo a los perfiles de diferentes usuarios y niveles de especialización. Otro aspecto a mencionar, es que únicamente dos documentos de GeoOntologías aparecen en la Tabla 5-1; **Error! No se encuentra el origen de la referencia.**, esto se debe a que la relación conecta aparece en documentos Web pero no es procesada de acuerdo a la semántica de un TopologyFile.

Como conclusión de este capítulo, mencionaremos que la ponderación integral es más útil que la ponderación aislada. Esto se demuestra cuando los objetos requeridos en la consulta se recuperaron mediante otros parámetros que describen a dichos objetos. Por ejemplo, se solicitaron {"*Vialidades conectan aeropuertos*"} pero, se recuperan <*caminos que conectan terminales aéreas*>. Medir la relevancia entre lo solicitado y lo recuperado requiere considerar las características que describen a los objetos geográficos.

Esto permite emitir un juicio de relevancia con mayor certidumbre. Es decir, si la consulta incluyo relaciones topológicas, pero el objeto recuperado es descrito por sus atributos geográficos y no por relaciones topológicas. Entonces, debe medirse su relevancia procesando tanto sus características más representativas como también aquellas que no contiene. Este procedimiento o política permite que documentos relevantes sean calificados con base en diversos criterios y no aplicando un único criterio de forma aislada o separada.

Finalmente, se comenta que el método de ponderación conceptual se baso en el método de *confusion*. En particular el método de *confusion* se adaptó para que funcionara en *GeoOntologías*. Mientras que en el caso de la ponderación también se adapto para trabaja en ontologías geográficas. Este proceso se utilizó para obtener un valor de ponderación para un documento asociado a un concepto y relaciones semánticas.



# CAPÍTULO 6

## PRUEBAS Y RESULTADOS

*Para ser un buen científico hay que saber decir "no sé" a tiempo.*

*Lee Smolin.*

En este capítulo, se presentan los resultados que se han obtenido con la estrategia de recuperación (*iGIR*) y el método de ponderación *iRank*. En particular, se muestran ejemplos de consultas que incluyen las siguientes relaciones dentro, conecta, junto, cruza e intersecta. Se muestra el proceso de recuperación a partir de una consulta y la secuencia de pasos aplicados de acuerdo a la metodología de *iGIR* y del algoritmo *OntoExplora*.

Se destaca que para cada consulta se recupera un conjunto de documentos. Los documentos se recuperan si contienen el o los objetos geográficos requeridos. También, si describen alguna propiedad de los objetos geográficos solicitados en la consulta. Así, como cuando los documentos contienen la relación espacial o topológica indicada en la consulta. Por ejemplo, para la consulta  $Q_{G5} = \{\text{Vialidades cruzan línea metro}\}$ . Los documentos recuperados contienen *objetos geográficos* de vialidades o de líneas del metro, o *propiedades geográficas* de vialidades. También, pueden contener la relación topológica *cruza* entre los objetos espaciales requeridos.

Además, se muestran los resultados ponderados usando *iRank* y clasificados por categoría. Finalmente, se ilustra el despliegue de resultados en formato KML, *Shapefiles* y documentos Web, utilizando los programas cliente *Google Earth* y *Google Maps*, el software GIS *Arcview* y un navegador Web. Los formatos utilizados son descritos en la sección 4.4 . Se manejan un total de 300 nodos relacionados con objetos geográficos en las *GeoOntologías* y en los diccionarios. Adicionalmente, se manejaron 300 objetos de archivos vectoriales (*shapefiles* y *TopologyFiles*). A continuación se seleccionaron ejemplos de la consultas que aparecen en la Tabla 4-6 del capítulo 4. Estas consultas contienen las relaciones utilizadas en esta tesis. Para cada consulta se indica el capítulo en el cual aparecen y en particular aquellas que son descritas en esta sección de resultados. Esto se despliega en la Tabla 6-1.

**Tabla 6-1** Ejemplos de consultas para realizar recuperación de información geográfica.

Ejemplos de consultas utilizadas en esta tesis	Aparece en el capítulo
Q <sub>G4</sub> = {Vialidades <b>conectan</b> aeropuertos}	Capítulo 4 : Metodología de recuperación Capítulo 5: Metodología de ponderación Capítulo 6: Pruebas y Resultados
Q <sub>G5</sub> = {Vialidades <b>cruzan</b> línea metro}	Capítulo 5 : Metodología de ponderación Capítulo 6: Pruebas y Resultados
Q <sub>G6</sub> = {Centro comercial <b>dentro</b> colonia}	Capítulo 6: Pruebas y Resultados
Q <sub>G7</sub> = {Línea metro <b>intersecta</b> vialidades}	Capítulo 6: Pruebas y Resultados
Q <sub>G8</sub> = {Colonias <b>comparte límite</b> colonias}	Capítulo 6: Pruebas y Resultados

Esta sección está estructurada como sigue, la sección 6.1 muestra los resultados obtenidos para una consulta en el módulo de *GeoOntología*, la sección 6.2 utiliza esta misma consulta con los resultados del módulo de diccionarios de datos (estos resultados son en cascada, es decir, la *GeoOntología* determina que buscar en los diccionarios). Mientras que la sección 6.3 despliega los resultados obtenidos para la consulta en los *TopologyFiles*. Posteriormente, la sección 6.4 muestra el despliegue de resultados en *Google Earth*, *Google Maps*, *Arcview* y con un navegador Web, para las consultas seleccionadas y finalmente la sección 6.5 ilustra las estadísticas finales.

## 6.1 Pruebas para la recuperación y ponderación conceptual

En esta sección explicamos los resultados obtenidos por *iGIR* e *iRank*. Donde para cada consulta el resultado es un conjunto de documentos que responden a una consulta. Cabe destacar que los documentos recuperados contienen los objetos geográficos solicitados en la consulta. También, pueden contener alguna propiedad de los objetos geográficos requeridos en la consulta. Así como también, pueden estar vinculados por la relación espacial o topológica indicada en la consulta. Por ejemplo, para la consulta  $Q_{G5} = \{\text{Vialidades cruzan línea metro}\}$  Los documentos recuperados contienen objeto geográficos de vialidades o de líneas del metro, o propiedades geográficas de vialidades. También, pueden contener la relación topológica cruza entre los objetos espaciales requeridos.

Comenzamos presentando los resultados del proceso de recuperación para las consultas  $Q_{G5} = \{\text{Vialidades cruzan línea metro}\}$  y  $Q_{G6} = \{\text{Centro comercial dentro colonias}\}$ . También, se muestran los resultados con la ponderación final de la consulta  $Q_{G4} = \{\text{Vialidades conectan aeropuertos}\}$ . El proceso para cada consulta, consiste en analizarlas para verificar que su estructura corresponde a una tripleta (ver sección 4.4.1). Después se utiliza el algoritmo *OntoExplora* (sección 4.5) para extraer el vector de vecindad donde los elementos del vector se utilizan como parámetros de búsqueda para recuperar datos de los diccionarios y *TopologyFiles*.

La Figura 6-1 muestra de forma gráfica un fragmento de la GeoOntología con un par de documentos recuperados para la consulta  $Q_{G5}$ . La representación gráfica de esta *GeoOntología* se hizo de forma manual, se utiliza XML como formato para la construcción de las *GeoOntologías*. Las cuales pueden ser visualizadas usando un navegador Web con soporte XML. Una descripción completa de las *GeoOntologías* y su integración con diccionarios se pueden consultar en el ANEXO A.

A continuación en la Figura 6-1 se muestra un fragmento de la *GeoOntología* que contiene un par de documentos recuperados para la consulta  $Q_{G5}$ .

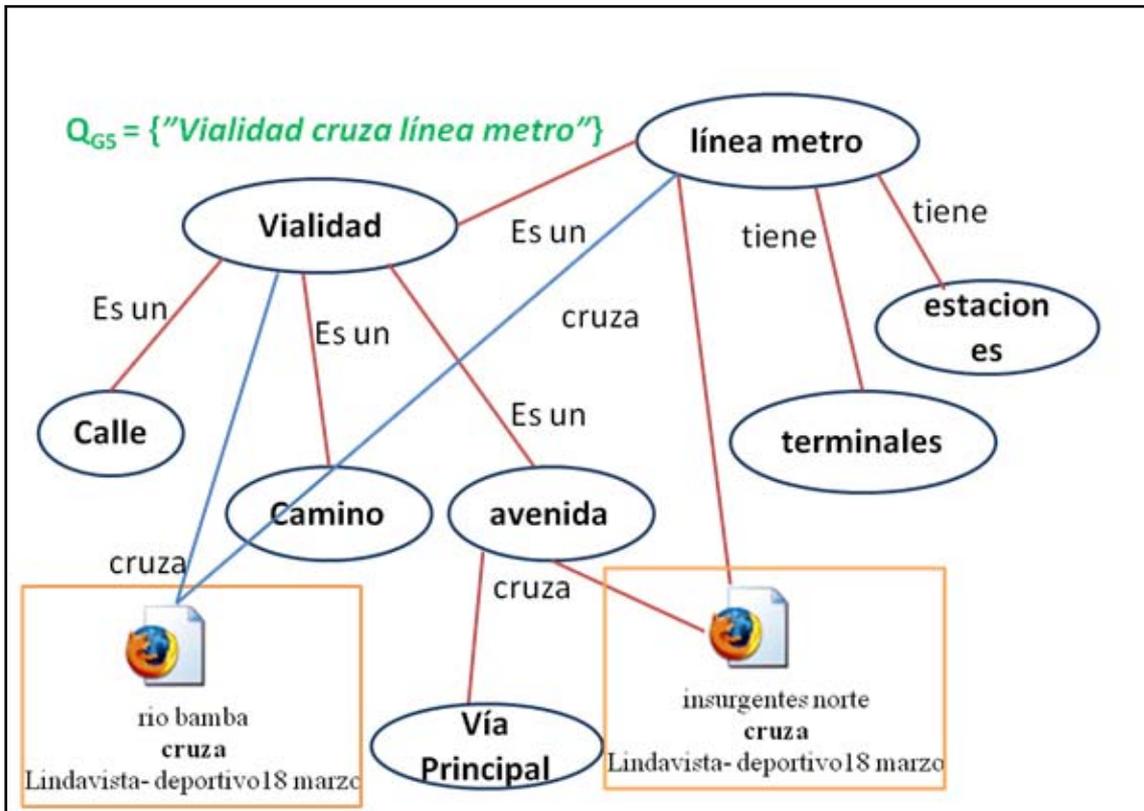
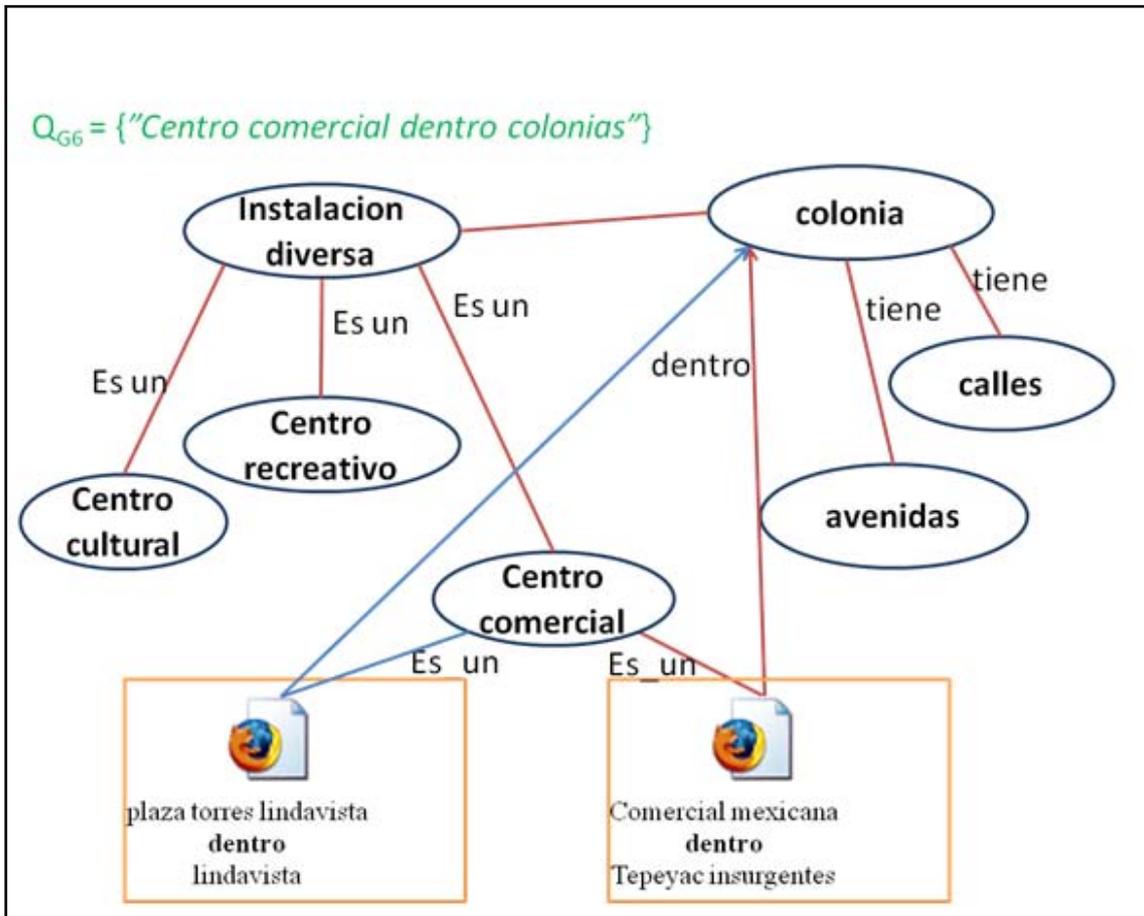


Figura 6-1 Fragmento de GeoOntología con documentos recuperados para  $Q_{G5}$ .

Como se aprecia en la Figura 6-1 se muestran los documentos *rio bamba cruza lindavista-deportivo 18 marzo* y el documento *insurgentes norte cruza lindavista-deportivo 18 marzo*. Mientras que en la Figura 6-2 se muestra la *GeoOntología* con un par de documentos recuperados para la consulta  $Q_{G6}$ .



**Figura 6-2** Fragmento de GeoOntología con documentos recuperados para  $Q_{G6}$ .

El resto de las *GeoOntologías* sigue la misma estructura y se pueden consultar en el anexo A. El algoritmo *OntoExplora* arroja como resultado un vector de vecindad para la consulta  $Q_{G5} = \{\text{Vialidades cruzan línea metro}\}$  del cual se muestran las propiedades y relaciones a buscar en los diccionarios geográficos y en los *TopologyFiles*.

*VectorVec* = [Calle, Boulevares, periféricos, ejes viales, viaductos, avenidas; tipo de calle, cruza, intersecta, estaciones, terminales, metro]

Mientras que para la consulta  $Q_{G6} = \{\text{"Centro comercial dentro colonias"}\}$  se obtuvo el siguiente vector de vecindad del cual se muestran las propiedades y relaciones a buscar en los diccionarios geográficos y en los *TopologyFiles*.

*VectorVec= [instalación diversa, tienda, supermercado, tienda autoservicio, dentro, contenido, colonia, suburbio, barrio, distritos, departamentos, colonia administrativa]*

Estos vectores de vecindad serán utilizados en los diccionarios de datos y los *TopologyFiles* para recuperar información de ellos, cabe señalar que el vector de vecindad es el mecanismo que funciona como integrador de la información contenida en estas tres fuentes de datos.

## **6.2 Pruebas con Consultas en los diccionarios de datos**

En este módulo continuaremos con el proceso de recuperación teniendo como elementos de entrada al vector de vecindad generado para la consulta  $Q_{G4}$ , así como los elementos de dicha consulta. Cada elemento del vector de vecindad es buscado dentro de los diccionarios geográficos, tanto por el nombre que los identifica así como por el conjunto de propiedades que los define.

El resultado arroja descripciones, otras propiedades y las posibles relaciones con las cuales este objeto puede existir en bases de datos geográficas y archivos *shapefile* s de acuerdo a lo definido por el INEGI.

En la Figura 6-3 se muestra las relaciones encontradas para la consulta  $Q_{G5} = \{\text{Vialidades cruzan línea metro}\}$  y el vector de vecindad:

*VectorVec= [Calle, Boulevares, periféricos, ejes viales, viaductos, avenidas; tipo de calle, cruza, intersecta, estaciones, terminales, metro]*

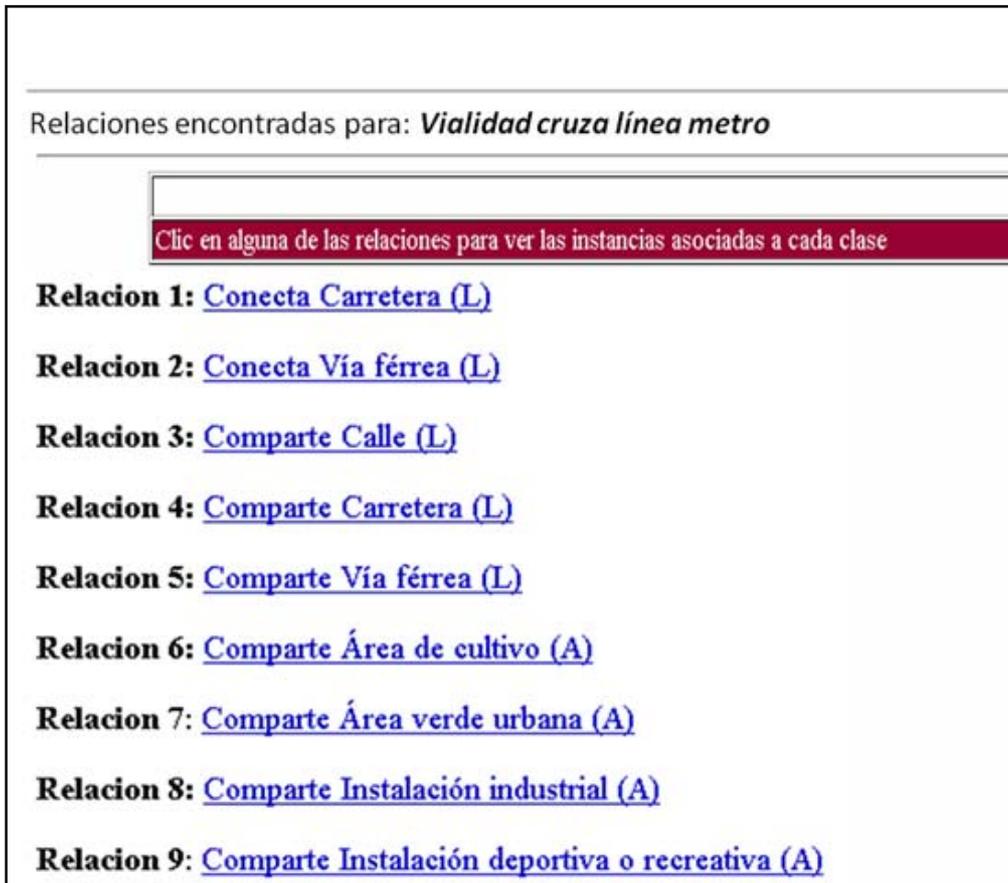


Figura 6-3 Recuperación de información geográfica en diccionarios para Q<sub>G5</sub>.

La Figura 6-3 muestra las relaciones asociadas a vialidades y líneas del metro a partir de la relación cruza e intersecta. Donde cada una de las relaciones encontradas se presenta como vínculos ya que están relacionadas con objetos geográficos almacenados en *shapefiles* y conducen hacia datos vectoriales (es la tercera y última etapa de la metodología *iGIR*). Por ejemplo, al seleccionar la relación “cruza carretera” se recuperarán los archivos vectoriales que contienen vialidades (calles, avenidas, líneas del metro) que crucen con una carretera. De forma adicional se presentan las descripciones de calles y la jerarquía en la cual están organizados en los diccionarios de datos.

En la Figura 6-4 se muestran el objeto geográfico carretera y sus atributos, de acuerdo a como están definidos en los diccionarios geográficos del INEGI.

```

- <elemento nombre="Carretera">
- <Carretera>
- <atributos>
- <dominio_fijo>
- <Tipo_de_carretera>
  <a1>Pavimentada : De asfalto o concreto.</a1>
  <a2>Terracería : De arena y grava compactada.</a2>
</Tipo_de_carretera>
+ <Numero_de_carriles></Numero_de_carriles>
- <Derecho_de_transito_de_la_carretera>
  <a1>Cuota : Con pago de peaje.</a1>
  <a2>Libre : Sin pago de peaje.</a2>
  <a3>Restringido : de acceso exclusivo</a3>
  + <Restricciones_del_atributo></Restricciones_del_atributo>
</Derecho_de_transito_de_la_carretera>
+ <Administracion_de_la_carretera></Administracion_de_la_carretera>
+ <Condicion_de_la_carretera></Condicion_de_la_carretera>
</dominio_fijo>
- <dominio_variable>
  + <identificador_de_carretera></identificador_de_carretera>
  + <Numero_de_carretera></Numero_de_carretera>
  + <Jurisdiccion></Jurisdiccion>
</dominio_variable>
</atributos>

```

Figura 6-4 Objeto geográfico carretera y sus atributos en un diccionario geográfico del INEGI.

En esta forma es como se realiza la recuperación en el módulo de diccionarios geográficos por medio del método de búsqueda denominado *MatchTopology*. La metodología *iGIR* continúa realizando recuperación desde archivos topológicos, utilizando el método de búsqueda *MatchTopology* el cual se explica en la sección 6.3.

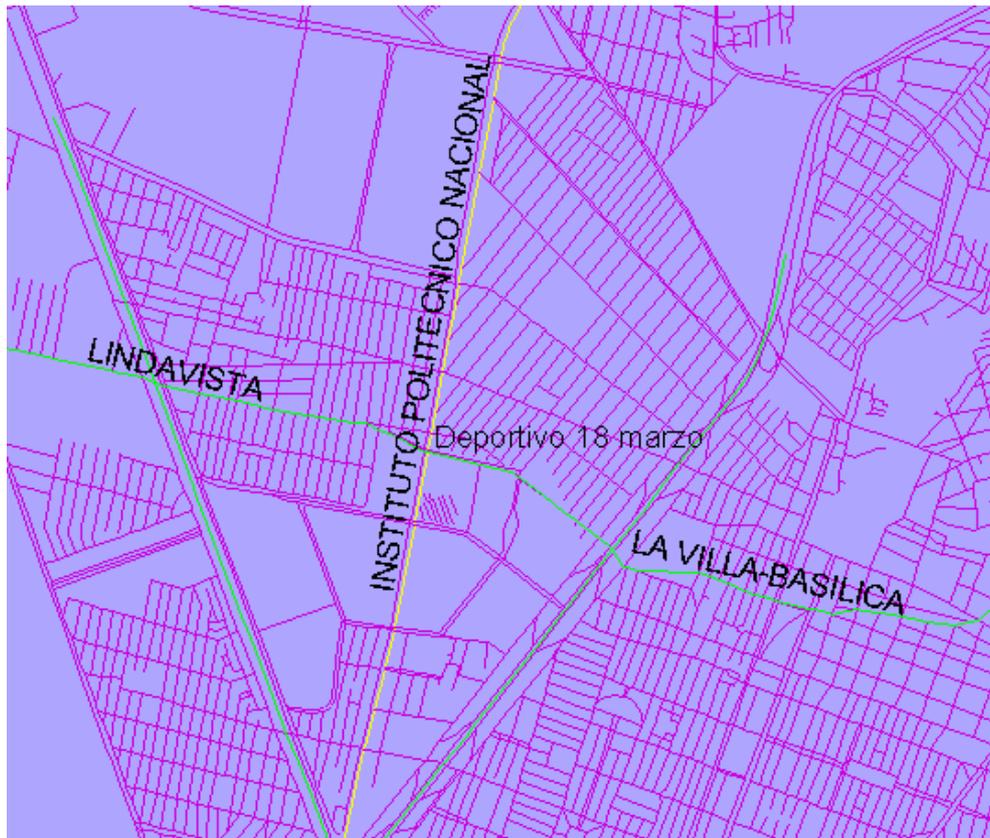
### 6.3 Recuperación en archivos topológicos

En este módulo se realiza el proceso de recuperación teniendo como elementos de entrada las relaciones contenidas en el vector de vecindad generado para la consulta  $Q_{G4}$ , así como los elementos de dicha consulta. Las relaciones asociadas a los objetos geográficos que están contenidas en el vector de vecindad son buscadas en los *TopologyFiles*, utilizando los nombres de los objetos y sus propiedades, así como también las relaciones espaciales que los asocian. En la Tabla 6-2 se muestran el conjunto de resultados finales para la consulta  $Q_{G5}$ .

**Tabla 6-2** Recuperación para  $Q_{GS} = \{\text{Vialidades cruzan línea metro}\}$  en los *TopologyFiles*

Nombre_objeto1	Capa Pertenece	Nombre_objeto2	Capa Pertenece	Relación	shapefile
INSTITUTO POLITECNICO NACIONAL	VIALIDAD	LINDAVISTA - DEPORTIVO 18 DE MARZO	LINEA_METRO	Cruza	Vialidad_ metro.shp
INSTITUTO POLITECNICO NACIONAL	VIALIDAD	LINDAVISTA - DEPORTIVO 18 DE MARZO	LINEA_METRO	Cruza	Vialidad_ metro.shp
RIOBAMBA	VIALIDAD	LINDAVISTA - DEPORTIVO 18 DE MARZO	LINEA_METRO	Cruza	Vialidad_ metro.shp
RIOBAMBA	VIALIDAD	LINDAVISTA - DEPORTIVO 18 DE MARZO	LINEA_METRO	Cruza	Vialidad_ metro.shp
BOGOTA	VIALIDAD	LINDAVISTA - DEPORTIVO 18 DE MARZO	LINEA_METRO	Cruza	Vialidad_ metro.shp
INSURGENTES NORTE	VIALIDAD	LINDAVISTA - DEPORTIVO 18 DE MARZO	LINEA_METRO	Cruza	Vialidad_ metro.shp
INSURGENTES NORTE	VIALIDAD	LINDAVISTA - DEPORTIVO 18 DE MARZO	LINEA_METRO	Cruza	Vialidad_ metro.shp
INSURGENTES NORTE	VIALIDAD	LINDAVISTA - DEPORTIVO 18 DE MARZO	LINEA_METRO	Cruza	Vialidad_ metro.shp
INSURGENTES NORTE	VIALIDAD	LINDAVISTA - DEPORTIVO 18 DE MARZO	LINEA_METRO	Cruza	Vialidad_ metro.shp

La última columna en la Tabla 6-2 muestra un vínculo a un archivo *shapefile* que contiene los datos geográficos recuperados para la consulta Q<sub>G5</sub>. En la Figura 6-5 se muestra el despliegue del archivo *shapefile* para el primer elemento de la Tabla 6-2, el archivo se visualiza usando el software GIS ArcView<sup>®</sup> de esri ([www.esri.com](http://www.esri.com)).



**Figura 6-5** Despliegue del archivo *shapefile* para el primer resultado de la Figura 6-4.

Como se puede apreciar en la Figura 6-5 se muestra en color amarillo la avenida instituto politécnico nacional (IPN) y la línea del metro que abarca las estaciones deportivo 18 marzo y lindavista, y que efectivamente la avenida IPN cruza este tramo de la línea del metro.

Adicionalmente, se puede llevar a cabo la generación de un archivo KML, para el objeto geográfico en cuestión (en este caso, líneas del metro y de vialidades). El proceso de generación de archivos KML se lleva a cabo siempre y cuando existan los atributos de latitud y longitud del objeto geográfico, que para los datos disponibles solo se cuentan

para datos cuya representación es de tipo punto o donde los datos lineales estén georeferenciados, este requisito es necesario para la generación de archivos KML.

#### 6.4 Resultado de la recuperación en formato KML

La Figura 6-6 muestra el resultado para el archivo KML generado a partir de la consulta  $Q_{G5} = \{\text{"Vialidades cruzan línea metro"}\}$ .



Figura 6-6 Vista del archivo KML para el primer resultado de la Figura 6-5.

La Figura 6-6 muestra parte del área donde se ubica la avenida instituto politécnico nacional y las estaciones del metro Deportivo 18 marzo y Lindavista, las cuales son vialidades que cruzan la línea del metro, en particular es una avenida.

Cabe señalar que si se cuenta con el nombre del objeto (por ejemplo, “*Avenida Instituto politécnico nacional*”) se puede utilizar el servicio Web de geonames<sup>45</sup> para conocer la latitud-longitud e integrarlo en los *TopologyFiles*. Permitiendo en esta forma extender la funcionalidad del sistema *iGIR*.

<sup>45</sup> <http://www.geonames.org/>

Una vez que se han mostrado los resultados de la recuperación, se procederá a explicar y mostrar cómo se ponderan estos resultados para las consultas establecidas. Los detalles completos de la ponderación se describen en la sección 5.2 y los resultados se despliegan en la sección 6.5.

### 6.5 Resultados finales ponderados (*iGIR* + *iRank*)

Los resultados de nuestro proceso de recuperación se han clasificado utilizando un rango de valores en el intervalo [0,1] de acuerdo a la Tabla 6-3.

**Tabla 6-3** Clasificación de valores de ponderación.

Rango de valores de ponderación	Descripción
$0 \leq \text{Valor de ponderación} < 0.1$	Nada relevante
$0.1 \leq \text{Valor de ponderación} < 0.25$	Poco relevante
$0.25 \leq \text{Valor de ponderación} < 0.5$	Medio relevante
$0.5 \leq \text{Valor de ponderación} < 0.75$	Casi relevante
$0.75 \leq \text{Valor de ponderación} \leq 1$	Totalmente relevante

Como se observa en la Tabla 6-3 de acuerdo al rango de valores se considera la relevancia del documento. Estos rangos de valores se propusieron para facilitar la lectura de resultados, ya que se despliegan en formato numérico y de esta forma el usuario tenga una noción de la relevancia del resultado obtenido.

Entonces, de acuerdo a esta clasificación, tenemos que un ejemplo de relevancia total ocurre cuando se busca *avenidas* y se obtienen *avenidas*. Mientras que un documento nada relevante ocurre cuando se busca *carretera* y se recuperó *población*. El resto de los rangos de valores se otorgan de acuerdo a la Tabla 6-3.

A continuación se muestran los resultados finales ponderados por *iRank*. Se muestran las consultas y el conjunto de documentos recuperados, esto se resume en dos tablas para cada consulta. La Tabla 6-4 muestra la consulta, el documento y la fuente de la cual fue recuperado. Así, como la ponderación de cada documento otorgado por *iRank*. En la se muestran los resultados ponderados por las tres medidas y por *iRank* para la consulta  $Q_{G5} = \{\text{Vialidades cruzan línea metro}\}$ .

**Tabla 6-4** Resultados para la consulta  $Q_{G5} = \{Vialidades\ cruzan\ línea\ metro\}$ .

Posición en <i>iRank</i>	Documento pertenece a fuente de datos	Documento recuperado usando iGIR	Ponderación en GEO RANK	Ponderación en TOPOLOGY RANK	Ponderación en CONCEPT RANK	Ponderación en <i>iRank</i>	
1	<i>TopologyFiles</i>	NEPTUNO intersecta LA RAZA – TLATELOLCO	0.76	0.96	0.86	0.86	Totalmente Relevante
2	<i>TopologyFiles</i>	HIDALGO intersecta JUAREZ HIDALGO	0.71	0.92	0.88	0.83	Totalmente relevante
3	GeoOntología	Avenida de los insurgentes intersecta la línea 5 del metro	0.52	0.75	0.92	0.73	Casi relevante
4	<i>TopologyFiles</i>	JUPITER intersecta LA RAZA – TLATELOLCO	0.61	0.86	0.72	0.73	Casi relevante
5	GeoOntología	Misterios cruza la línea 5 del metro	0.30	0.36	0.72	0.66	Casi relevante
6	Diccionario geográfico	Calle conecta línea metro	0.68	0.38	0.64	0.56	Casi relevante
7	<i>TopologyFiles</i>	MOSQUETA intersecta GUERRERO TLATELOLCO	0.58	0.42	0.42	0.47	Medio relevante
8	<i>TopologyFiles</i>	MOCTEZUMA intersecta HIDALGO GUERRERO	0.45	0.63	0.31	0.46	Medio Relevante
9	GeoOntología	Afecta vialidades que cruzan con línea 5 del metro	0.25	0.22	0.76	0.41	Medio relevante
10	Diccionario geográfico	Avenida comparte línea metro	0.59	0.25	0.36	0.4	Medio relevante

Los resultados que intuitivamente se esperan para la consulta  $Q_{G5}$  son todas las vialidades que cruzan una línea del metro. Si no se utilizara el método integral solo se recuperarían los TopologyFiles que contienen vialidades y líneas del metro (ver resultados 1, 2,4, 7 y 8). Sin embargo, los diccionarios también contienen documentos relevantes para la consulta  $Q_{G5}$ . Por ejemplo, atributos de vialidades como avenidas y calles (ver resultados 6 y 10). En forma similar, las *GeoOntologías* tienen información relacionada para  $Q_{G5}$ . Por ejemplo, estaciones del metro y la relación interseca (ver resultados 3,5 y 9).

En esta forma, se aprecia que la recuperación integral permite recuperar datos que de otra forma serían omitidos. El uso del vector de vecindad permite controlar cuáles documentos se recuperan. En esta forma se reduce la posibilidad de recuperar información no relevante.

Mientras que el método de ponderación, permite que la relevancia de un documento no se limite únicamente a evaluar las relaciones y objetos geográficos que se solicitan en la consulta. Por ejemplo, en el caso de la consulta  $Q_{G5} = \{\text{Vialidades cruzan línea metro}\}$  se recupero un documento que contiene una estación del metro. Si este documento se evaluara únicamente por sus atributos geográficos su relevancia sería prácticamente nula.

Para continuar el análisis, se procede a explicar los resultados obtenidos para la consulta  $Q_{G6} = \{\text{centro comercial dentro colonias}\}$ . Los cuales se muestran en la Tabla 6-5.

**Tabla 6-5** Documentos recuperados y ponderados para  $Q_{G6} = \{\text{Centro comercial dentro colonias}\}$ 

Posición en <i>iRank</i>	Documento pertenece a fuente de datos	Documento recuperado usando iGIR	Ponderación en GEO RANK	Ponderación en TOPOLOGY RANK	Ponderación en CONCEPT RANK	Ponderación en <i>iRank</i>	
1	<i>TopologyFiles</i>	Bodega aurrera dentro CTM VII culhuacan	0.76	0.98	0.94	0.89	Totalmente Relevante
2	<i>TopologyFiles</i>	Superama dentro san pablo tepetlapa	0.73	0.93	0.86	0.84	Totalmente relevante
3	GeoOntología	Tiendas autoservicio en colonias	0.60	0.55	0.98	0.71	Casi relevante
4	<i>TopologyFiles</i>	Chedrahui dentro los girasoles I	0.45	0.86	0.72	0.67	Casi relevante
5	Diccionario geográfico	Instalación diversa dentro area urbana	0.74	0.51	0.68	0.64	Casi relevante
6	Diccionario geográfico	Instalación diversa dentro colonia	0.70	0.36	0.61	0.55	Casi relevante
7	<i>TopologyFile</i>	Oxxo dentro santa ursula	0.37	0.75	0.32	0.48	Medio relevante
8	<i>TopologyFile</i>	Megacomercial dentro el caracol	0.39	0.75	0.33	0.49	Medio relevante
9	GeoOntología	Walmart en tepeyac	0.38	0.33	0.51	0.40	Medio relevante
10	GeoOntología	Superama en lindavista	0.36	0.34	0.51	0.40	Medio Relevante

Como se observa en la Tabla 6-5, la consulta  $Q_{G6} = \{\text{Centro comercial dentro colonias}\}$  recupera en su mayoría documento de archivos topológicos, debido a que el procesamiento de la relación *dentro* es muy común en estos archivos( ver resultados 1,2, 4, 7, y 8). Sin embargo la relación *dentro* no se procesa topológicamente en los documentos Web. Tampoco es frecuente (a diferencia de la relación *cerca*). Pero, existen documentos que contienen datos geográficos de centros comerciales y de colonias que son relevantes y que deben ser recuperados para esta consulta (ver resultados 9 y 10). Por ejemplo, direcciones del centro comercial y coordenadas geográficas. Los diccionarios geográficos por su parte describen estos objetos con otros nombres y relaciones (ver resultados 5 y 6).

En el caso de la ponderación, la relación *dentro* tiene alta correspondencia con la relación “*en*”. Sin embargo, la ponderación de la relación *dentro* es obtenida usando operaciones de sobreposición o de proximidad, mientras que para la relación *en*, es obtenida mediante su ubicación. La ponderación sigue caracterizándose por tener resultados opuestos (ver resultado 6 y 3) pero nuevamente la integración permite que la relevancia de un documento sea procesada con mayor certidumbre, al medir por sus características más representativas y no únicamente por las características solicitada en la consulta. A continuación se revisan los resultados para la consulta  $Q_{G7} = \{\text{Línea metro intersecta vialidades}\}$  los cuales se muestran en la Tabla 6-6.

**Tabla 6-6** Documentos recuperados y ponderados para  $Q_{G7} = \{\text{Línea metro interseca vialidades}\}$

Posición en <i>iRank</i>	Documento pertenece a fuente de datos	Documento recuperado usando iGIR	Ponderación en GEO RANK	Ponderación en TOPOLOGY RANK	Ponderación en CONCEPT RANK	Ponderación en <i>iRank</i>	
1	<i>TopologyFiles</i>	INDIOS VERDES - DEPORTIVO 18 DE MARZO INTERSECTA RICARTE	0.75	0.84	0.96	0.85	Totalmente Relevante
2	<i>TopologyFiles</i>	INDIOS VERDES - DEPORTIVO 18 DE MARZO INTERSECTA MONTIEL	0.68	0.95	0.87	0.83	Totalmente relevante
3	<i>TopologyFiles</i>	LA RAZA – TLATELOLCO INTERSECTA PAGANINI	0.59	0.76	0.81	0.72	Casi relevante
4	Diccionario geográfico	Calle comparte instalación diversa	0.51	0.82	0.81	0.71	Casi relevante
5	GeoOntología	Metro atraviesa vialidades	0.61	0.38	0.90	0.63	Casi relevante
6	Diccionario geográfico	Calle conecta carretera	0.60	0.36	0.90	0.62	Casi relevante
7	GeoOntología	Línea 6 del metro cruza avenidas	0.65	0.46	0.36	0.49	Medio relevante
8	<i>TopologyFile</i>	GUERRERO TLATELOLCO INTERSECTA MARTE	0.63	0.43	0.32	0.46	Medio Relevante
9	<i>TopologyFile</i>	GUERRERO TLATELOLCO INTERSECTA R.F.MAGON	0.21	0.70	0.46	0.45	Medio relevante
10	GeoOntología	Metro la raza cruza insurgentes	0.19	0.65	0.46	0.43	Medio relevante

En la Tabla 6-6 la recuperación de documentos para la consulta  $Q_{G7} = \{\text{Línea metro interseca vialidades}\}$  los resultados obtenidos nuevamente se ven influenciados por los archivos topológicos. Esto, se debe a que la relación cruza se procesa y aparece con poca frecuencia en los diccionarios geográficos. En el mismo caso para los documentos Web. De nueva cuenta tenemos que la integración permite recuperar documentos relevantes aunque estos no contengan explícitamente la relación solicitada o el nombre de los objetos geográficos requeridos (ver resultados 4 y 6 donde *instalación diversa* puede representar una estación del metro)

También, se observa que la relación cruza puede ponderarse de forma similar a la relación interseca. Donde la relación cruza aparece en las GeoOntologías mientras que en los archivos topológicos corresponde con la relación interseca. Podemos concluir que la ponderación integral es útil para poder medir la relevancia aunque los nombres de objetos sean diferentes y sus relaciones no coincidan completamente.

Ahora, se analizará la Tabla 6-7 en donde se muestran resultados para la consulta  $Q_{G8} = \{\text{Colonias comparte límite colonias}\}$

Tabla 6-7 Documentos recuperados y ponderados para  $Q_{G8} = \{\text{Colonias comparte límite colonias}\}$

Posición en <i>iRank</i>	Documento pertenece a fuente de datos	Documento recuperado usando iGIR	Ponderación en GEO RANK	Ponderación en TOPOLOGY RANK	Ponderación en CONCEPT RANK	Ponderación en <i>iRank</i>	
1	<i>TopologyFiles</i>	malacates comparte límite arboledas	0.71	0.96	0.94	0.87	Totalmente Relevante
2	<i>TopologyFiles</i>	lomas de cuatepec comparte límite malacates	0.73	0.96	0.89	0.86	Totalmente relevante
3	<i>TopologyFiles</i>	Parque metropolitano comparte límite lomas de cuatepec	0.71	0.96	0.89	0.85	Casi relevante
4	<i>TopologyFiles</i>	quetzalcoatl 3 comparte límite tlacaelel 2	0.71	0.96	0.75	0,68	Casi relevante
5	Diccionario geográfico	Area urbana comparte instalación diversa	0.46	0.72	0.81	0.66	Casi relevante
6	Diccionario geográfico	Area urbana comparte instalación industrial	0.46	0.72	0.81	0.66	Casi relevante
7	GeoOntología	colonia narvarte colinda colonia buenos aires	0.39	0.46	0.65	0.63	Casi relevante
8	<i>TopologyFiles</i>	juventino rosas comparte límite tepetatal	0.43	0.43	0.76	0.55	Casi relevante
9	GeoOntología	Colonia la olimpica colinda con colonia escudero	0.45	0.46	0.76	0.55	Casi relevante
10	GeoOntología	Colonia del valle colinda viaducto	0.31	0.50	0.55	0.45	Medio relevante

De esta forma, nuevamente se visualiza que cuando los documentos son calificados de forma aislada, las calificaciones de relevancia que obtienen son muy opuestas y sólo en casos como el del documento nueve la evaluación otorgada es casi la misma para dos de las medidas. Nuevamente, se tiene caso donde los valores de ponderación son opuestos. Por lo tanto, comparando *iRank* contra *GeoRank*, *TopologyRank* y *ConceptRank* se tiene que la integración de criterios permite obtener mayor certidumbre en la ponderación más adecuado, lo cual denota que *iRank* obtiene mejores resultados de forma integral que usando criterios aislados.

Como se muestran en las tablas, los valores otorgados por *iRank* resultan más útiles de acuerdo a lo que se conoce a priori, y en general *iRank* ofreció mejores resultados de ponderación al integrar los criterios que utilizando un único criterio de forma aislada. Los valores de relevancia dados por *iRank* fueron evaluados por un grupo de personas de diversas disciplinas y estudiantes del laboratorio de procesamiento inteligente de información espacial. De acuerdo al juicio de ellos, el valor asignado por *iRank* es más útil que el proporcionado por alguno de los criterios en forma aislada.

Por lo tanto, tenemos que los resultados son satisfactorios de acuerdo a las expectativas de los diferentes perfiles del usuario. Sin embargo, se requieren pruebas adicionales para evaluar desempeño en condiciones de concurrencia. Así como el uso de otros datos, provenientes de otras instituciones, en cualquiera de las tres fuentes utilizadas.



# CAPÍTULO 7

## CONCLUSIONES Y TRABAJO FUTURO

*No hay inversión más rentable que la del conocimiento*  
*Benjamín Franklin*

La presente tesis pertenece a la línea de investigación de recuperación de información geográfica (GIR). En particular, en este trabajo se recupera y pondera información de tres fuentes de datos geográficas heterogéneas, mediante un mecanismo de integración conducido por *GeoOntologías*, el cual procesa la semántica espacial, las relaciones topológicas, y atributos de objetos geográficos. El enfoque de recuperación de información extiende el mecanismo de *matching* por *palabras clave* hacia el de *matching* por atributos *geográficos* y *relaciones topológicas*. Los resultados de la recuperación y la ponderación de forma integral, fueron mejores que cuando se procesan por separado. A continuación se muestran las conclusiones, aportaciones, productos derivados y los trabajos a futuro que pueden realizarse a partir de la presente tesis y de la línea de investigación a la que pertenece.

El capítulo está estructurado como sigue: la sección 7.1 describe las conclusiones generales obtenidas durante la investigación y desarrollo de esta tesis, mientras que en la sección 7.2 se despliega en una lista las principales conclusiones obtenidas durante las pruebas y desarrollo de la tesis. La sección 0 indica los productos obtenidos, mientras que en la sección 0 se comentan las aportaciones. Por último en la sección 7.5 se esboza en forma de lista los propuestas y posibles trabajos a futuro a partir de esta tesis.

## **7.1 Conclusiones generales**

Actualmente, en el área de recuperación de información geográfica aún existen muchos retos y tareas relativas a la geografía, en las cuales es necesario investigar para encontrar mejores métodos para procesar, buscar y manejar la información geográfica. Básicamente, porque la geografía de cada país es diferente tanto a nivel de accesibilidad, creación y generación de cartografía, así como en la regulación de la misma. Esto se debe a que se ven inmiscuidos aspectos políticos, culturales, y del lenguaje, entre otros. Una solución a esta problemática se puede obtener con un enfoque integrador como el propuesto en esta tesis.

Por otra parte, en la generación y producción de información geográfica participan diversas empresas, instituciones y el propio gobierno, por lo que se requieren estandarizaciones, consensos y acuerdos. Sin embargo en México aún no existe un consenso para tal efecto, recientemente el INEGI ha comenzado a colaborar con otras instituciones, empresas y organismos para lograr una estandarización en cuanto a formatos y entonces publicar mediante servicios Web los numerosos datos que cada organismo genera o ha creado. Sin embargo, en México aún no existe algún servicio Web disponible para explotar, procesar o consultar esta información. En este aspecto, integrar los datos de diversas fuentes es una alternativa de solución para procesar la información de estas fuentes.

Por otra parte, en GIR existen varias áreas de oportunidad. Por ejemplo, en la Web se ataca el problema de consultas expresadas en lenguaje natural, mientras que en sistemas de escritorio se enfocan en tareas de análisis espacial, así como estudio de fenómenos geográficos. Utilizar el enfoque integrador de esta tesis enfocado a sistemas de escritorio y de análisis espacial permitirá mejorar la administración, procesamiento y la búsqueda de información geográfica en estos escenarios.

Otras áreas de oportunidad en GIR incluyen temas tales como el *Geo-parsing* para detectar terminología geográfica en documentos, *Geo-coding* para adjuntar una referencia a una ubicación única asociada a un documento. Así como también, buscadores que asocian al texto mapas o fuentes de información geográfica. Otros retos se enfocan en los temas de indexado espacial de documentos, la ponderación geográfica, las interfaces de usuario para GIR, y la recuperación de datos geográficos para usarse en GIS y aspectos relativos con el diseño y evaluación de sistemas GIR. De nueva cuenta, integrar estos aspectos en un sistema permitirá mejorar varios de los procesos comentados.

En esta tesis, se ofrece una solución para recuperar información geográfica desde tres fuentes de datos heterogéneas. La recuperación y ponderación se basa en una metodología que integra los datos de estas fuentes. Esta es la principal aportación de esta tesis.

## **7.2 Conclusiones particulares**

- Procesar la semántica espacial permite recuperar información geográfica de tres fuentes geográficas, indicando los parámetros de búsqueda de acuerdo a cada fuente de datos.
- Integrar fuentes de información heterogéneas por medio de los conceptos, atributos y relaciones que representan a un objeto geográfico, permiten obtener información útil, que no puede obtenerse por separado.

- Ponderar en GIR requiere métodos para medir similitud entre objetos geográficos. La medida de *confusion* es una alternativa para dichos métodos.
- *iRank* pondera integrando tres medidas, ofreciendo mejores resultados que ponderando de forma separada.
- Recuperar información geográfica implica extender los criterios de recuperación basada en palabra, hacia criterios basados en la geografía y la semántica espacial.
- Una *GeoOntología* puede emular el rol de un experto en GIS, a quien podemos preguntar acerca de un tópico y que éste nos ofrezca respuestas acertadas o relacionadas.
- Procesar relaciones topológicas permite recuperar objetos de acuerdo a las relaciones espaciales que tienen con otros objetos.
- El procesamiento de objetos geográficos, considerando su dominio de valores y atributos, permite recuperar datos de acuerdo las características que los describen.
- Recuperar documentos relacionados con los objetos geográficos de una consulta, requirió establecer una escala de acuerdo a los diferentes grados de concordancia o coincidencia para cada objeto geográfico.
- Las *GeoOntologías* contienen documentos Web porque son utilizados por el usuario neófito en GIS, y representan la noción geográfica de las personas para su entorno geográfico.

- El procesamiento de archivos vectoriales permite utilizar datos de otros organismos e instituciones para ofrecer resultados más relevantes de acuerdo a la riqueza de relaciones que presentan estos archivos.
- El enfoque semántico ofrece una extensión en el área de recuperación de información geográfica ya que permite resolver de mejor manera las consultas que con el enfoque lingüístico o basado en palabra no se han podido resolver.
- Satisfacer un criterio de búsqueda, requiere saber expresar una consulta. Sin embargo, esto se complica si el usuario es neófito en el tema. Es decir, carece del conocimiento para expresarle al buscador que tipo de información requiere, de tal forma que no dé lugar a ambigüedades o confusiones. Es por ello, que se requieren mecanismos que permitan a los no expertos, obtener resultados relevantes para las búsquedas que realizan.
- Considerar el perfil del usuario, especialista o neófito en el tema a consultar, permite ofrecer resultados de acuerdo a su grado de conocimiento cuando expresa una consulta (la intención del usuario).
- Considerar las relaciones topológicas en el proceso de ponderación, permite que los resultados se organicen de acuerdo a dichas relaciones (ocurrencia de palabra vs. ocurrencia topológica).
- *iGIR* trabaja con base en propiedades y relaciones espaciales, a diferencia de IR tradicional (e.g. coordenada geográfica vs. palabra clave)
- El método *iRank* pondera integrando tres medidas, de tal forma que se aplican criterios de ponderación que ofrecen mejores resultados que al utilizarlos de forma separada.

### 7.3 Productos derivados de la tesis

- Cuatro artículos
  - Matching concepts: an ontological approach to retrieve geographic information (Geos 2005, Ciudad de México)
  - Semantics of proximity in locative expressions (Congeo 2006, Ciudad de México)
  - Geographic information retrieval by topological, geographical and conceptual ( Geos 2007, Ciudad de México)
  - *iRank*: Integral Ranking of Geographical Information by semantic, geographic, and topological matching (IF &GIS 2009, San Petersburgo, Rusia)
  
- Dos artículos para revista
  - *iGIR*: A methodology to geographic information retrieval from heterogeneous geographic data sources.
  - *iRank*: A methodology to rank geographic information from heterogeneous geographic data sources.
  
- Dos tesis de licenciatura
  - Sistema de geolocalización para dispositivos móviles (2006)
  - Simplificación de cartografía digital en la web (2006)

#### 7.4 Aportaciones

- Una metodología de recuperación de información geográfica que integra tres fuentes geográficas heterogéneas.
- Un algoritmo que procesa semántica espacial de GeoOntologías (algoritmo *OntoExplora*).
- Integración de tres fuentes de datos geográficas heterogéneas para recuperar documentos relevantes que de otra forma sería omitidos.
- Se integraron tres fuentes de información geográfica heterogéneas, las cuales son utilizadas por tres tipos de usuarios: analista GIS, usuarios GIS, y neófitos.
- Tres tipos de *matching* para recuperar información geográfica: *matching* conceptual, geográfico y topológico.
- Una metodología de ponderación de información geográfica integrando tres fuentes de datos heterogéneas
- La adaptación del método de confusión para usarse con GeoOntologías.

#### 7.5 Trabajos a futuro

A continuación se puntualiza en los trabajos que pueden realizarse a futuro teniendo como base la presente investigación para dirigirla hacia nuevas líneas de investigación. En particular aquellas que en donde convergen la recuperación de información geográfica, los sistemas GIS, la tecnología Web y de cómputo móvil. Las cuales son propuestas como temas de tesis de licenciatura, maestría y doctorado.

○

- Establecimiento de una estrategia de recuperación de información geográfica la cual procese las nueve relaciones topológicas del modelo 9- interseccion.
- Incluir el enfoque de recuperación geográfica dentro de la construcción de la Web Semántica *Geo-Espacial*. Es decir, efectuar la búsqueda de información geográfica de acuerdo a la semántica espacial y a través de la Web. Por ejemplo: “Epidemias dentro Alemania” donde la relación *dentro* es procesada como la relación topológica *contenido*.
- Integración por medio de mecanismos y técnicas de alineación de otras ontologías, las cuales pertenezcan a otras instituciones, organismos y empresas para poder consultar la información de una forma transparente y sin explorar varios sitios a la vez.
- Incluir servicios Web geográficos para enriquecer el conjunto de resultados u ofrecer otras alternativas para cada consulta geográfica, es decir programas que procesen la semántica espacial utilizando la arquitectura orientada a servicios (SOA). Por ejemplo un servicio Web del INEGI que consulte automáticamente al servicio WEB de la SCT.
- Generar programas para automatizar la conversión de diccionarios de datos geográficos y de otras fuentes de información geográfica hacia XML, para facilitar su procesamiento integrado e interoperabilidad con otros sistemas GIR.
- Extender el mecanismo de confusión para aplicarlo en conjunto con servicios Web geográficos, permitiendo así tener una herramienta web que pueda evaluar si dos objetos geográficos significan lo mismo, de acuerdo a la conceptualización de una GeoOntología.

- El enfoque presentado puede ser implantado como una extensión en los actuales servicios basados en localización (Por ejemplo: Yahoo Local, Local Google, Windows Live Local) donde además de desplegar la ubicación geográfica, se incluya la descripción conceptual o jerarquía en el espacio geográfico (Continente- País- Estado) de un objeto geográfico.
- Construir un buscador Web o de escritorio que trabaje con asistencia de ontologías y geo-ontologías y así ofrecer la base para construir servicios web de semántica espacial en México.
- Utilizar los resultados obtenidos como tema de estudio para poder definir métodos o técnicas de indexado semántico espacial en crawlers.
- Utilizar el enfoque de mashups (mezcla de sitios Web) para extender y enriquecer el sistema *iGIR* y aplicarlo a otros escenarios. Por ejemplo, búsqueda de casas, o artículos con base en alguna propiedad geográfica o espacial (“*Casa al lado de un lago*”, “*Memorias USB con precio de 100 pesos en Ciudad de México*”).
- Integrar el sistema con el nuevo campo denominado la neo geografía (e.g. *Google Maps+eBay*) e integrando el tiempo. Por ejemplo, “laptops vendidas alrededor Lindavista en Diciembre 2008”
- *iGIR* e *iRank* pueden ser adaptados para trabajarse en conjunto con servicios basados en localización y para usuarios móviles. “Ruta a pie del hotel Royal al Centro Histórico”.
- *iGIR* puede utilizarse para asistir a sistemas que administren y procesen situaciones de emergencia, utilizando dispositivos con asistencia GPS. Por ejemplo, con consultas tales como: “Hospitales alrededor Aeropuerto”.



## REFERENCIAS

- [1] Baeza-Yates, R. and Ribeiro-Neto, B. Modern Information Retrieval. New York: ACM Press; Harlow, Addison-Wesley, 1999 XX, 513 p. ISBN 0-201-39829-X.
- [2] Diana Santos and Marcirio Silveira Chaves, The place of place in geographical IR, Proceedings of SIGIR 2006, August 10 2006, Seattle, USA.
- [3] Jones C., Purves R., Ruas A., Sanderson M., Sester M., Van Kreveld M. & Weibel R., “Spatial Information Retrieval and Geographical Ontologies An Overview of the SPIRIT Project”, in Proceedings of the 25th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, ACM Press, Tampere, Finland, 2002, 387-394
- [4] Kobayashi, M. and Takeda, K., Information Retrieval on the Web, ACM Computing Surveys (CSUR) , Volume 32 , Issue 2 (June 2000), Pages: 144 - 173, Year of Publication: 2000, ISSN:0360-0300
- [5] Varelas G, Voutsakis E, Raftopolou P, Petrakis EGM, Milios E, Semantic similarity methods in *WordNet* and their application to information retrieval on the web, Proceedings of the 7th annual ACM international workshop on Web information and data management , Bremen, Germany, Pages: 10 – 16, 2005, ISBN:1-59593-194-5.
- [6] Regina M. M. Braga, Marta Mattoso, Cláudia M. L. Werner, The Use of Mediation and Ontology Technologies for Software Component Information Retrieval, Proceedings of the 2001 symposium on Software reusability: putting software reuse in context, Toronto, Ontario, Canada, Pages: 19 – 28, 2001,ISSN:0163-5948.
- [7] Hawking, D., Craswel, N., Bailey, P., Griths, K.: Measuring Search Engine Quality, Information Retrieval. Vol. 4 (1), 33—59 (2001)
- [8] Cristiano Rocha, Daniel Schwabe, Marcus Poggi Aragao, A Hybrid Approach for Searching in the Semantic Web, Proceedings of the 13th international conference on World Wide Web, Pages: 374 - 383,2004, ISBN:1-58113-844-X, New York, NY, USA.
- [9] Sergey Brin, Lawrence Page, “The anatomy of a large scale hyper textual web search engine”, Proceedings of 7th international Web conference, Pages: 107 – 117, 1998, ISSN:0169-7552, Brisbane, Australia.

- [10] R. Guha and R. McCool and E. Miller, Semantic search, WWW2003 Proceedings of the 12th international conference on World Wide Web, pages 700--709. ACM Press, 2003.
- [11] Arron Walker, Binh Pham, Miles Moody, Spatial Bayesian Learning Algorithms for geographic information retrieval, GIS'05, November 4-5, Bremen, Germany.
- [12] G.Cai, GeoVSM: An integrated Retrieval Model for Geographic Information, in M. Egenhofer and D. Mark, Eds., *Geographic Information Science-Second International Conference*, GIScience 2002, Boulder, CO, 2002, vol. 2489, Lecture Notes in Computer Science, Springer, pp. 70-85.
- [13] Zhou Yinhua, Xie Xing, Wang Chuang, Gong Yuchang, Ma Wei-Ying, "Hybrid Index Structures for Location-based Web Search", CIKM'05, October 31-November 5, 2005, Bremen, Germany.
- [14] Delboni, T., Borges, K., Laender, A., in Geographical information retrieval, ACM-GIR'05, Bremen, Germany.
- [15] Sang Ok Koo, Soo Yeon Lim, Sang Jo Lee, "Building an Ontology Based on Hub Words for Information Retrieval," pag. 466, IEEE/WIC International Conference on Web Intelligence (WI'03), 2003.
- [16] S. Vaid, C. B. Jones, H. Joho, and M. Sanderson. Spatio-textual Indexing for Geographical Search on the Web. In Proceedings of SSTD-05, the 9th Symposium on Spatial and Temporal Databases, 2005.
- [17] Max Egenhofer, Toward the Semantic Geospatial Web, *ACM-GIS 2002*, A. Voisard and S.-C. Chen (eds.), November 2002.
- [18] Adolfo Guzman-Arenas and Victor-Polo de Gyves. A distributed digital text accessing and acquisition system, BiblioDigital. SoftwarePro International. Lecture Notes in Computer Science 3061, 274-283. (Springer Verlag 2004), ISSN 0302-9743.
- [19] Levachkine S. & Guzmán-Arenas A., "Hierarchies Measuring Qualitative Variables", in Gelbukh (Ed.), *Computational Linguistics and Intelligent Text Processing (CICLing' 2004)*, Lecture Notes in Computer Science, Vol. 2945, Springer-Verlag, 2004, 262-274.

- [20] Guzmán-Arenas A. & Levachkine S., “Graduated Errors in Approximate Queries using Hierarchies and Ordered Sets”, in Monroy R., Arroyo-Figueroa G., Sucar L.E. & Sossa-Azuela J.H. (Eds.), *Advances in Artificial Intelligence, Third Mexican International Conference on Artificial Intelligence (MICAI' 2004)*, Mexico City, Mexico, 2004, *Lecture Notes in Computer Science 2972*, Springer-Verlag, 2004, 119-128.
- [21] W3C, World Wide Web Consortium, [www.w3c.org](http://www.w3c.org).
- [22] Martínez M, *Topologic Descriptor for Topographic Maps*, Master of Science Thesis, México, 2006 (in Spanish).
- [23] Tim Berners-Lee, James Hendler and Ora Lassila, *The Semantic Web*, *Scientific American*, May 2001.
- [24] T. Gruber, *Toward principles for the design of ontologies used for knowledge sharing*, *International Journal of Human-Computer Studies archive*, Volume 43 , Issue 5-6 Nov./Dec. 1995 table of contents, Special issue: the role of formal ontology in the information technology, Pages: 907- 928, Year of Publication 1995, ISSN:1071-5819.
- [25] Cutting D.R, D.R. Karger, J.O. Pedersen and J.W. Tukey, “Scatter/Gather: A cluster-based approach to browsing large document collection, in proceedings of ACM/SIGIR '92 pp 318-329.
- [26] *WordNet*, An electronic lexical database, <http://WordNet.princeton.edu>, último acceso: 27 de junio de 2006.
- [27] Junghoo Cho, Hector Garcia –Molina, Lawrence Page, *Efficient Crawling through URL ordering*. Seventh international Web Conference (www98). Brisbane, Australia, April 14-18, 1998.
- [28] S. Lawrence and C.L. Giles. *Searching the World Wide Web*. *Journal: Science*, volume 280, pages, 98—100, 1998.
- [29] Wikipedia, la enciclopedia libre. <http://es.wikipedia.org/>
- [30] Mark Sanderson, Kalervo Järvelin, James Allan, Peter Bruza (Eds.): *SIGIR 2004: Proceedings of the 27th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, Sheffield, UK, July 25-29, 2004. ACM 2004, ISBN 1-58113-881-4.

- [31] Ray R. Larson, Geographic Information Retrieval and Spatial Browsing, University of California, Berkeley. [https://sherlock.sims.berkeley.edu/geo\\_ir/PART1.html](https://sherlock.sims.berkeley.edu/geo_ir/PART1.html), último acceso: mayo 2007.
- [32] Ray R. Larson, Patricia Frontiera, Geographic Information Retrieval (GIR) Ranking Methods for Digital Libraries, Pages: 415 – 415, Proceedings of the 2004 Joint ACM/IEEE Conference on Digital Libraries, Tuscon, AZ, USA June 07 - 11, 2004.
- [33] Open Geospatial Consortium, Inc., <http://www.opengeospatial.org/>
- [34] <http://www.answers.com/topic/spatial-analysis?cat=technology>, último acceso: mayo 2007.
- [35] Andy Carvin: Tim Berners-Lee: Weaving a Semantic Web. Digital divide network articles, 2005.
- [36] Simple API for XML, <http://sax.sourceforge.net/>
- [37] S. Lawrence. Context in web search, IEEE Data Engineering Bulletin, 23(3):25–32, 2000.
- [38] Miller, G. A. (1993) "Introduction to *WordNet*: An On-line Lexical Database", en G. Miller et al. (1993), <http://WordNet.princeton.edu/5papers.pdf>, último acceso: junio de 2007.
- [39] OGC Reference Model , Open Geospatial Consortium Inc. , Reference number: OGC 03-040, Editor: George Percivall, 2003-09-16.
- [40] Base de datos geográficos. Modelo de datos Vectorial, Instituto Nacional de Estadística Geografía e Informática, 1993.
- [41] Base de datos geográficos. Diccionario de datos topográficos, Instituto Nacional de Estadística Geografía e Informática, 1995a.
- [42] Base de datos geográficos. Diccionario de datos topográficos Escalas 1:50,000, Instituto Nacional de Estadística Geografía e Informática, 1996.
- [43] Base de datos geográficos. Diccionario de datos topográficos escala 1:250,000, Instituto Nacional de Estadística Geografía e Informática, 1995b.
- [44] Base de datos geográficos, Diccionario de datos topográficos escala 1:1, 000,000, Instituto Nacional de Estadística Geografía e Informática, 1995.

- [45] Guarino, N.: Formal ontology, conceptual analysis and knowledge representation. *International Journal of Human-Computer Studies*. Vol. 43 Num:5-6, pp. 625-640 (1995)
- [46] Webster's 1913 Dictionary,  
[http://humanities.uchicago.edu/orgs/ARTFL/forms\\_unrest/webster.form.html](http://humanities.uchicago.edu/orgs/ARTFL/forms_unrest/webster.form.html), último acceso: Febrero de 2007.
- [47] Genesereth, M. R. and Nilsson, N. J. 1987. *Logical Foundation of Artificial Intelligence*. Morgan Kaufmann, Los Altos, California.
- [48] "L. De Floriani and P. Marzano and E. \ Puppo", Spatial queries and data models, *Information Theory: a Theoretical Basis for {GIS}*, Springer-Verlag, Lecture Notes in Computer Science, N.716;ed, I. Campari A. U. Frank and U. Formentini", pages:113--138,1992.
- [49] Salton, G. (1989). *Automatic Text Processing: The Transformation, Analysis, and Retrieval of Information by Computer*. Reading, Mass. : Addison-Wesley.
- [50] Ray R. Larson , *Geographic Information Retrieval and Spatial Browsing* , University of California, Berkeley, [https://sherlock.sims.berkeley.edu/geo\\_ir/PART1.html](https://sherlock.sims.berkeley.edu/geo_ir/PART1.html), último acceso: marzo 2007.
- [51] Frank, A. U. (1991). Properties of Geographic Data: Requirements for Spatial Access Methods. In Guenther, O & Schek, H.-J. (eds) *Advances in Spatial Databases*, (pp. 225-234). Berlin : Springer-Verlag.
- [52] Frew, J., Carver, L., Fischer, C., Goodchild, M., Larsgaard, M. Smith, T. & Zheng, Q. (1995). The Alexandria Rapid Prototype: building a digital library for spatial information. In *1995 ESRI User Conference Proceedings*. Redlands, CA: Environmental Systems Research Institute, 1995. <http://www.esri.com/resources/userconf/proc95/to300/p255.html>, último acceso: noviembre 2006.
- [53] Brimicombe, A. J. (1993). Combining Positional and Attribute Uncertainty Using Fuzzy Expectation in a GIS. In *GIS/LIS Proceedings*, (pp.72-81). Bethesda, MD : American Society for Photogrammetry and Remote Sensing, American Congress on Surveying and Mapping.
- [54] Griffiths, A. (1989). SAGIS: A Proposal for a Sardinian Geographic Information System and an assessment of alternative implementation strategies. *Journal of Information Science*, 15, 261-267.

- [55] Walker, D. R. F, Newman, I. A., Medyckj-Scott, D. J. & Ruggles, C. L. N. (1992). A System for Identifying Datasets for GIS Users. *International Journal of Geographical Information System*, 6, 511-527.
- [56] Laurini, R & Thompson, D. (1992). *Fundamentals of spatial information systems*. San Diego, Calif. : Academic Press.
- [57] Holmes, D. O. (1990). Computers and geographic information access. *Meridian*, 4, 37-49.
- [58] Tobler, W. R. (1970). "A computer model simulation of urban growth in the Detroit region". *Economic Geography*, 46(2): 234-240.
- [59] Acerca del INEGI, Objetivo principal,  
<http://www.inegi.gob.mx/inegi/contenidos/espanol/instituto/default.asp?c=1503>, último acceso junio 2007.
- [60] Reiner Kraft, Farzin Maghoul, Chi Chao Chang, Y!Q: Contextual Search at the Point of Inspiration, CIKM'05, October 31–November 5, 2005, Bremen, Germany.
- [61] Natalya F. Noy and Deborah L. McGuinness, Desarrollo de Ontologías-101: Guía Para Crear Tu Primera Ontología, Stanford University, Stanford, CA, 94305, Traducido del ingles por: Erick Antezana!, September 19, 2005.
- [62] Francis, W.N, and Lucera, H. "Frequency Analysis of Englis Usage", Houghton Mifflin, Boston 1982.
- [63] Gütting R.H., Geo-relqational algebra: a model and query language for geometric database systems. In Noltemeier, H., editor, *Computational Geometry ans its applications*, New York: Springer-Verlag.
- [64] Menon, S., and T.R. Smith,. A declarative spatial query processor for geographic information systems. *Photogrammetric Engineering and Remote Sensing*, vol. 55, no11, pp 1593-1600, year: 1989.
- [65] Bennis, K., B. David, I. Morize-Quilio, J. Thévenin, and Y. Viémon, 1991. GeoGraph: a topological Storage model for extensible GIS. *Proceedings, Auto Carto 10*, pp. 349-367.

- [66] Boyle, A.R., Dangermond, J., Marble, D.F. Simonett, D.S., Smith, L.K., and Tomlinson, R.F., 1983. Final Report of a conference on the Review and Synthesis of Problems and Directions for Large Scale Geographic Information System Development, National Aeronautics and Space Administration, Contract NAS2-11246.
- [67] Abler, Ronald F., The national Science Foundations National Center for Geographic Informations and Analysis. International Journal of Geographical Information Systemas, vol. 1, no. 4, pp 303-326.
- [68] Frank, A. U., Towards a spatial theory. Proceedings, International Symposium on Geographic Information Systems: The Research Agenda, November 1987, Cristal City, Virginia, 2: 215-227.
- [69] Peuquet, D.J. Representations of Geographic space: toward a conceptual synthesis. Annals of the Association of American Geographers, vol. 78, no. 3, pp. 375-394.
- [70] NCGIA (National Center for Geographic Information and Analysis), The research plan of the National Center for Geographic Information and Analysis. International Journal of Geographical Information Systems, vol. 3, no.2, pp. 117-136. Year: 1989.
- [71] Pullar, D., and M. Egenhofer. Towards Formal Definitions of Topological Relations Among Spatial Objects, Third International Symposium on Spatial Data Handling, Sidney, Australia, pp. 225-243. Year: 1998.
- [72] Worboys, M., and S. Deen, Semantic Heterogeneity in Geographic Data bases. SIGMOD RECORD, vol. 20, no. 4, pp. 30-34. Year: 1991.
- [73] Internet GIS, Zhong-Reng Peng, Ming-Hsiang Tsou, Publisher: Jhon wiley & Sons, Inc., Hoboken, New Jersey.
- [74] Chuang Wang, Xing Xie, Lee Wang, Yansheng Lu and Wei-Ying Ma, Detecting Geographic Locations from Web Resources, Proceedings of the 2005 workshop on Geographic information retrieval, November 04-04, 2005, Bremen, Germany.
- [75] Jens Graupmann and Ralf Schenkel , GeoSphereSearch: Context-Aware Geographic Web Search, Proceedings of SIGIR 2006, August 10 2006, Seattle, USA.
- [76] Leonardo Andrade and Mário J. Silva , Relevance Ranking for Geographic IR, Proceedings of SIGIR 2006, August 10 2006, Seattle, USA.

- [77] Mata F., Levachkine S.,: Semantics of Proximity in Locative Expressions, In e-Proceedings of First International Conference, GeoS 2005, Poster Session, Mexico city, Mexico (2005)
- [78] F. Mata, Geographic Information Retrieval by Topological, Geographical, and Conceptual *Matching*, In: Proceedings of Second International Conference, GeoS 2007, Mexico city, Mexico (2007)
- [79] W. Zong, D. Wu, A. Sun, E. Lim, and D. Goh, On assigning place names to geography related web pages. In Proceedings of JCDL, 2005.
- [80] N. Cardoso, B. Martins, M. Chaves, L. Andrade, and M. Silva. The XLDB group at GeoCLEF 2005, In GeoCLEF 2005 Workshop, (2005).
- [81] ACM SIGIR Forum archive, Volume 40, Issue 2, ACM Press, New York, NY, USA, (2006)
- [82] Corcho, O., Fernandez-Lopez, M., and Gómez Perez, A.: Metodologies, tools and languages for building ontologies: where is the meeting point?, *Data and Knowledge Engineering* 46(1), 41,64 (2003)
- [83] McGuinness, D.L.: Ontologies come age, in D. Fensel, J. Hendler, H. Lieberman and W. Wahlster (eds), *Spinning the Semantic Web: Bringing the World Wide Web to Its Full Potential*, MIT Press, Cambridge, MA.
- [84] R. Meersman: Semantic Ontology Tools in IS Design. ISMIS'99 Conference in Warsaw, June, 1999.
- [85] D.B. Lenat: CYC: a large-scale investment in knowledge infrastructure, *Communications of the ACM* 38 (11) (1995)33.
- [86] R. Meersman: Semantic Ontology Tools in IS Design. ISMIS'99 Conference in Warsaw, June, 1999.
- [87] M. Uschold, M. King, S. Morale, and Y. Zorgios: The Enterprise Ontology, *The Knowledge Engineering Review* 13 1 (1998), pp. 31-89.
- [88] R. Studer, V. R. Benjamins and D. Fensel: Knowledge Engineering: Principles and methods, *Data and Knowledge engineering (DKE)*, 25(1-2): 161-197, 1998.
- [89] D. Fensel: *Ontologies: a silver bullet for Knowledge Management and Electronic Commerce*, Springer-Verlag, Berlin, 2003. To appear.

- [90] S. Staab, H.P. Schnurr, R. Studer, Y. Sure: Knowledge processes and ontologies, *IEEE Intelligent Systems* 16 (1) (2001), pp. 26-34.
- [91] D. Fensel: Problem-solving methods: Understanding, development, description, and reuse, Lecture Notes on Artificial Intelligence 1791, Springer-Verlag, Berlin, 2000.
- [92] T. R. Gruber: A translation approach to portable ontology specifications, *Knowledge Acquisition*, 5:199-220, 1993.
- [93] Alexiev V, Breu M et al. (2005): Information Integration with Ontologies. Experiences from an Industrial Showcase. Wiley & Sons: Chichester.
- [94] Christiane Fellbaum (editor). *WordNet: An Electronic Lexical Database*. MIT Press, 1998.
- [95] Lenat, D. B. and Guha, R. V. (1990). Building large knowledge-based systems: representation and inference in the Cyc project. Addison-Wesley.
- [96] R. Studer, V. R. Benjamins and D. Fensel: Knowledge Engineering: Principles and methods, *Data and Knowledge engineering (DKE)*, 25(1-2): 161-197, 1998.
- [97] El lenguaje de programación Ruby, <http://www.ruby-lang.org/es/>
- [98] Paul Clough, Extracting Metadata for Spatially-Aware Information Retrieval on the Internet, Proceedings of the 2005 workshop on Geographic information retrieval, November 04-04, 2005, Bremen, Germany.
- [99] Diana Santos and Marcirio Silveira Chaves, The place of place in geographical IR. Proceedings of SIGIR 2006, August 10 2006, Seattle, USA.
- [100] Zhisheng Li, Chong Wang, Xing Xie, Xufa Wang and Wei-Ying Ma, Indexing implicit locations for geographical information retrieval, Proceedings of SIGIR 2006, August 10 2006, Seattle, USA.
- [101] Leonardo Andrade and Mário J. Silva , Relevance Ranking for Geographic IR, Proceedings of SIGIR 2006, August 10 2006, Seattle, USA.
- [102] Quillian, M.R., "Semantic Memory", in *Semantic Information Processing*, Minsky, M. (Ed.), MIT Press, 1968, pp. 216-270.

[103] Collins, A.M. & Quillian, M.R., "Retrieval time from semantic memory", *Journal of Verbal Learning and Verbal Behavior*, Vol. 8, 1969, pp. 240-247.

[104] Shapiro, S.C., and Rapaport, W.J., "The SNePS Family" *Computers and Mathematics with Applications*, 1992, 23: 243-- 275.

[105] Sowa, J. "Knowledge Representation: Logical, Philosophical, and Computational Foundations", Pacific Grove, CA., Brook/Cole, a division of Thomson Learning, 2000.



# ANEXO A. GeoOntologías

En este anexo se muestran las *GeoOntologías* generadas y construidas en forma semiautomática de acuerdo a lo que se explica en la sección 4.1.

El formato de estas *GeoOntologías* es XML, y fueron implementadas de esa forma para facilitar interoperabilidad con otros sistemas y trabajos a futuro.

```
<?xml version="1.0" encoding="UTF-8" ?>
```

```
= <!--
```

```
Document : diccionario1-50000.xml
```

```
Created on : 15 de abril de 2007, 11:13 AM
```

```
Author : migfel
```

```
Description:
```

```
Purpose of the document follows.
```

```
-->
```

```
= <Diccionario1-50000>
```

```
= <elemento nombre="acueducto">
```

```
= <acueducto>
```

```
= <atributos>
```

```
= <dominio_fijo>
```

```
= <relacion_del_acueducto_con_el_suelo>
```

```
<w1>Subterráneo: Bajo la superficie del terreno</w1>
```

```
<w2>Superficial: Sobre la superficie del terreno</w2>
```

```
</relacion_del_acueducto_con_el_suelo>
```

```
= <condicion_del_acueducto>
```

```
<w1>En construcción: Que está en proceso de construcción</w1>
```

```
<w2>En operación: Que está en servicio o puede usarse</w2>
```

```
<w3>Fuera de uso: Que no está en uso</w3>
```

```
</condicion_del_acueducto>
```

```
</dominio_fijo>
```

```
= <dominio_variable>
```

```
= <identificador_de_acueducto>
```

```
<w1>Un número secuencial que se incrementa con cada ocurrencia</w1>
```

```
= <dominio_de_valores>
```

```
<w1>En el intervalo de 1 a N</w1>
```

```
</dominio_de_valores>
```

```
</identificador_de_acueducto>
```

```
</dominio_variable>
```

```
</atributos>
```

```
= <restricciones_de_integridad>
```

```
= <COMBINACIONES_AUTORIZADAS_DE_VALORES_DE_ATRIBUTOS>
```

```
<c1>Subterráneo, en construcción</c1>
```

```
<c2>Subterráneo, en operación</c2>
```

```
<c3>Subterráneo, fuera de uso</c3>
<c4>Superficial, en construcción</c4>
<c5>Superficial, en operación</c5>
<c6>Superficial, fuera de uso</c6>
  </COMBINACIONES_AUTORIZADAS_DE_VALORES_DE_ATRIBUTOS>
= <CALIFICADOR_AUTORIZADO_DE_REPRESENTACIONES_GEOMETRICAS>
  <a1>aproximada</a1>
  <a2>definida</a2>
  </CALIFICADOR_AUTORIZADO_DE_REPRESENTACIONES_GEOMETRICAS>
  </restricciones_de_integridad>
= <relaciones>
  <r1>Conecta Instalación de bombeo (P)</r1>
  <r2>Conecta Manantial (P)</r2>
  <r3>Conecta Tanque de agua (P)</r3>
  <r4>Conecta Acueducto (L)</r4>
  <r5>Conecta Canal (L)</r5>
  <r6>Conecta Canal (A)</r6>
  <r7>Conecta Cuerpo de agua (A)</r7>
  <r8>Conecta Estanque (A)</r8>
  <r9>Conecta Instalación de bombeo (A)</r9>
  <r10>Conecta Instalación diversa (A)</r10>
  </relaciones>
  </acueducto>
  </elemento>
= <elemento nombre="aeropuerto">
= <aeropuerto>
= <atributos>
= <dominio_fijo>
= <TIPO_DE_AEROPUERTO>
  <a1>Internacional: Permite vuelos con cubrimiento internacional</a1>
  <a2>Nacional: Permite vuelos con cubrimiento nacional</a2>
  <a3>Local: Sólo permite vuelos con cubrimiento regional</a3>
  </TIPO_DE_AEROPUERTO>
  </dominio_fijo>
= <dominio_variable>
= <identificador_de_aeropuerto>
```

<w1>Un número secuencial que se incrementa con cada ocurrencia</w1>

= <dominio\_de\_valores>

<w1>En el intervalo de 1 a N</w1>

</dominio\_de\_valores>

</identificador\_de\_aeropuerto>

</dominio\_variable>

</atributos>

= <restricciones\_de\_integridad>

= <COMBINACIONES\_AUTORIZADAS\_DE\_VALORES\_DE\_ATRIBUTOS>

<c1>Internacional</c1>

<c2>Nacional</c2>

<c3>Local</c3>

</COMBINACIONES\_AUTORIZADAS\_DE\_VALORES\_DE\_ATRIBUTOS>

= <CALIFICADOR\_AUTORIZADO\_DE\_REPRESENTACIONES\_GEOMETRICAS>

<a1>Definida</a1>

<a2>Virtual</a2>

</CALIFICADOR\_AUTORIZADO\_DE\_REPRESENTACIONES\_GEOMETRICAS>

</restricciones\_de\_integridad>

= <relaciones>

<r1>Aeropuerto Conecta Camino (L)</r1>

<r2>Aeropuerto Conecta Carretera (L)</r2>

<r3>Aeropuerto Conecta Camino (L)</r3>

<r4>Aeropuerto Conecta Carretera (L)</r4>

<r5>Aeropuerto Comparte Calle (L)</r5>

<r6>Aeropuerto Comparte Camino (L)</r6>

<r7>Aeropuerto Comparte Carretera (L)</r7>

<r8>Aeropuerto Comparte Área de cultivo (A)</r8>

<r9>Aeropuerto Comparte Área urbana (A)</r9>

<r10>Aeropuerto Comparte Área verde urbana (A)</r10>

</relaciones>

</aeropuerto>

</elemento>

= <elemento nombre="area\_de\_cultivo">

= <area\_de\_cultivo>

= <atributos>

= <dominio\_fijo>

<a1>Ninguno</a1>  
</dominio\_fijo>  
= <dominio\_variable>  
= <identificador\_de\_area\_de\_cultivo>  
<w1>Un número secuencial que se incrementa con cada ocurrencia</w1>  
= <dominio\_de\_valores>  
<w1>En el intervalo de 1 a N</w1>  
</dominio\_de\_valores>  
</identificador\_de\_area\_de\_cultivo>  
</dominio\_variable>  
</atributos>  
= <restricciones\_de\_integridad>  
= <COMBINACIONES\_AUTORIZADAS\_DE\_VALORES\_DE\_ATRIBUTOS>  
<c1>Ninguna</c1>  
</COMBINACIONES\_AUTORIZADAS\_DE\_VALORES\_DE\_ATRIBUTOS>  
= <CALIFICADOR\_AUTORIZADO\_DE\_REPRESENTACIONES\_GEOMETRICAS>  
<a1>Definida</a1>  
<a2>Virual</a2>  
</CALIFICADOR\_AUTORIZADO\_DE\_REPRESENTACIONES\_GEOMETRICAS>  
</restricciones\_de\_integridad>  
= <relaciones>  
<r1>Comparte Canal (L)</r1>  
<r2>Comparte Camino (L)</r2>  
<r3>Comparte Carretera (L)</r3>  
<r4>Comparte Corriente de agua (L)</r4>  
<r5>Comparte Límite (L)</r5>  
<r6>Comparte Vía férrea (L)</r6>  
<r7>Comparte Aeropuerto (A)</r7>  
<r8>Comparte Área urbana (A)</r8>  
<r9>Comparte Canal (A)</r9>  
<r10>Comparte Cementerio (A)</r10>  
<r11>Comparte Cuerpo de agua (A)</r11>  
<r12>Comparte Instalación diversa (A)</r12>  
<r13>Comparte Instalación industrial (A)</r13>  
<r14>Comparte Terreno sujeto a inundación (A)</r14>  
<r15>Comparte Vegetación densa (A)</r15>

</relaciones>

</area\_de\_cultivo>

</elemento>

= <elemento nombre="area\_natural\_protegida">

= <area\_natural\_protegida>

= <atributos>

= <dominio\_fijo>

= <tipo\_de\_area\_natural\_protegida>

<a1>Área de protección de flora y fauna : Contiene los hábitat de cuyo equilibrio y preservación dependen especies de flora y fauna</a1>

<a2>Área de protección de recursos naturales : Destinada a la preservación y restauración de zonas forestales y a la conservación de suelos y aguas</a2>

<a3>Monumento natural : Contiene uno o varios elementos naturales de importancia nacional; no tiene la variedad de ecosistemas ni la superficie necesaria para ser considerado con otro valor</a3>

<a4>Parque marino nacional : Zona marina y playas o áreas contiguas destinadas a la preservación de ecosistemas acuáticos</a4>

<a5>Parque nacional : Terreno forestal o representación biogeográfica de uno o más ecosistemas</a5>

<a6>Reserva de la biosfera : Área biogeográfica representativa de uno o más ecosistemas, con presencia de especies endémicas, amenazadas o en peligro de extinción. El acceso es restringido</a6>

<a7>Zona sujeta a conservación ecológica : Área con protección estatal o municipal destinada a preservar el equilibrio ecológico</a7>

</tipo\_de\_area\_natural\_protegida>

</dominio\_fijo>

= <dominio\_variable>

= <identificador\_de\_area\_natural\_protegida>

<w1>Un número secuencial que se incrementa con cada ocurrencia</w1>

= <dominio\_de\_valores>

<w1>En el intervalo de 1 a N</w1>

</dominio\_de\_valores>

</identificador\_de\_area\_natural\_protegida>

>

</dominio\_variable>

</atributos>

```
= <restricciones_de_integridad>
= <COMBINACIONES_AUTORIZADAS_DE_VALORES_DE_ATRIBUTOS>
  <c1>Área de protección de flora y fauna</c1>
  <c2>Área de protección de recursos naturales</c2>
  <c3>Monumento natural</c3>
  <c4>Parque marino nacional</c4>
  <c5>Parque nacional</c5>
  <c6>Reserva de la biosfera</c6>
  <c7>Zona sujeta a conservación ecológica</c7>
  </COMBINACIONES_AUTORIZADAS_DE_VALORES_DE_ATRIBUTOS>
= <CALIFICADOR_AUTORIZADO_DE_REPRESENTACIONES_GEOMETRICAS>
  <a1>Aproximada</a1>
  <a2>Virtual</a2>
  </CALIFICADOR_AUTORIZADO_DE_REPRESENTACIONES_GEOMETRICAS>
  </restricciones_de_integridad>
= <relaciones>
  <r1>Comparte Carretera (L)</r1>
  <r2>Comparte Límite (L)</r2>
  <r3>Comparte Vía férrea (L)</r3>
  <r4>Comparte Área natural protegida (A)</r4>
  </relaciones>
  </area_natural_protegida>
  </elemento>
= <elemento nombre="area_urbana">
= <area_urbana>
= <atributos>
= <dominio_fijo>
  <a1>Ninguno</a1>
  </dominio_fijo>
= <dominio_variable>
= <identificador_de_area_urbana>
  <w1>Un número secuencial que se incrementa con cada ocurrencia</w1>
= <dominio_de_valores>
  <w1>En el intervalo de 1 a N</w1>
  </dominio_de_valores>
  </identificador_de_area_urbana>
```

</dominio\_variable>  
</atributos>  
= <restricciones\_de\_integridad>  
= <COMBINACIONES\_AUTORIZADAS\_DE\_VALORES\_DE\_ATRIBUTOS>  
<c1>Ninguna</c1>  
</COMBINACIONES\_AUTORIZADAS\_DE\_VALORES\_DE\_ATRIBUTOS>  
= <CALIFICADOR\_AUTORIZADO\_DE\_REPRESENTACIONES\_GEOMETRICAS>  
<a1>Definida</a1>  
<a2>Virtual</a2>  
</CALIFICADOR\_AUTORIZADO\_DE\_REPRESENTACIONES\_GEOMETRICAS>  
</restricciones\_de\_integridad>  
= <relaciones>  
<r1>Conecta Calle (L)</r1>  
<r2>Conecta Camino (L)</r2>  
<r3>Conecta Carretera (L)</r3>  
<r4>Conecta Vía férrea (L)</r4>  
<r5>Comparte Bordo (L)</r5>  
<r6>Comparte Calle (L)</r6>  
<r7>Comparte Camino (L)</r7>  
<r8>Comparte Canal (L)</r8>  
<r9>Comparte Carretera (L)</r9>  
<r10>Comparte Corriente de agua (L)</r10>  
<r11>Comparte Instalación portuaria (L)</r11>  
<r12>Comparte Límite (L)</r12>  
<r13>Comparte Vía férrea (L)</r13>  
<r14>Comparte Área de cultivo (A)</r14>  
<r15>Comparte Canal (A)</r15>  
<r16>Comparte Cementerio (A)</r16>  
<r17>Comparte Cuerpo de agua (A)</r17>  
<r18>Comparte Instalación diversa (A)</r18>  
<r19>Comparte Instalación industrial (A)</r19>  
<r20>Comparte Instalación portuaria (A)</r20>  
<r21>Comparte Rasgo arqueológico (A)</r21>  
<r22>Comparte Terreno sujeto a inundación (A)</r22>  
</relaciones>  
</area\_urbana>

```
    </elemento>
= <elemento nombre="area_verde_urbana">
= <area_verde_urbana>
= <atributos>
= <dominio_fijo>
  <a1>Ninguno</a1>
  </dominio_fijo>
= <dominio_variable>
= <identificador_de_area_verde_urbana>
  <w1>Un número secuencial que se incrementa con cada ocurrencia</w1>
= <dominio_de_valores>
  <w1>En el intervalo de 1 a N</w1>
  </dominio_de_valores>
  </identificador_de_area_verde_urbana>
  </dominio_variable>
  </atributos>
= <restricciones_de_integridad>
= <COMBINACIONES_AUTORIZADAS_DE_VALORES_DE_ATRIBUTOS>
  <c1>Ninguno</c1>
  </COMBINACIONES_AUTORIZADAS_DE_VALORES_DE_ATRIBUTOS>
= <CALIFICADOR_AUTORIZADO_DE_REPRESENTACIONES_GEOMETRICAS>
  <a1>Definida</a1>
  <a2>Virtual</a2>
  </CALIFICADOR_AUTORIZADO_DE_REPRESENTACIONES_GEOMETRICAS>
  </restricciones_de_integridad>
= <relaciones>
  <r1>Comparte Calle (L)</r1>
  <r2>Comparte Aeropuerto (A)</r2>
  <r3>Comparte Área urbana (A)</r3>
  <r4>Comparte Cementerio (A)</r4>
  <r5>Comparte Cuerpo de agua (A)</r5>
  <r6>Comparte Instalación diversa (A)</r6>
  <r7>Comparte Instalación industrial (A)</r7>
  </relaciones>
  </area_verde_urbana>
  </elemento>
```

```
= <elemento nombre="arrecife_bajo">
= <arrecife_bajo>
= <atributos>
= <dominio_fijo>
= <Tipo_de_arrecife_bajo>
  <a1>Arrecife : Estructura rocosa, frecuentemente coralina</a1>
  <a2>Bajo : Fondo subacuático somero y accidentado conformado por
    materiales no consolidados</a2>
  </Tipo_de_arrecife_bajo>
  </dominio_fijo>
= <dominio_variable>
= <identificador_de_arrecife_bajo>
  <w1>Un número secuencial que se incrementa con cada ocurrencia</w1>
= <dominio_de_valores>
  <w1>En el intervalo de 1 a N</w1>
  </dominio_de_valores>
  </identificador_de_arrecife_bajo>
  </dominio_variable>
  </atributos>
= <restricciones_de_integridad>
= <COMBINACIONES_AUTORIZADAS_DE_VALORES_DE_ATRIBUTOS>
  <c1>Arrecife</c1>
  <c2>Bajo</c2>
  </COMBINACIONES_AUTORIZADAS_DE_VALORES_DE_ATRIBUTOS>
= <CALIFICADOR_AUTORIZADO_DE_REPRESENTACIONES_GEOMETRICAS>
  <a1>Aproximada</a1>
  <a2>Definida</a2>
  <a3>Virtual</a3>
  </CALIFICADOR_AUTORIZADO_DE_REPRESENTACIONES_GEOMETRICAS>
  </restricciones_de_integridad>
= <relaciones>
  <r1>Ninguna</r1>
  </relaciones>
  </arrecife_bajo>
  </elemento>
= <elemento nombre="Banco_de_material">
```

= <Banco\_de\_material>  
= <atributos>  
= <dominio\_fijo>  
= <Condición\_de\_banco\_de\_material>  
  <a1>**En operación : Que está en servicio o puede usarse**</a1>  
  <a2>**Fuera de uso : Que no está en uso**</a2>  
  </Condición\_de\_banco\_de\_material>  
  </dominio\_fijo>  
= <dominio\_variable>  
= <identificador\_de\_banco\_de\_material>  
  <w1>**Un número secuencial que se incrementa con cada ocurrencia**</w1>  
= <dominio\_de\_valores>  
  <w1>**En el intervalo de 1 a N**</w1>  
  </dominio\_de\_valores>  
  </identificador\_de\_banco\_de\_material>  
  </dominio\_variable>  
  </atributos>  
= <restricciones\_de\_integridad>  
= <COMBINACIONES\_AUTORIZADAS\_DE\_VALORES\_DE\_ATRIBUTOS>  
  <c1>**En operación**</c1>  
  <c2>**Fuera de uso**</c2>  
  </COMBINACIONES\_AUTORIZADAS\_DE\_VALORES\_DE\_ATRIBUTOS>  
= <CALIFICADOR\_AUTORIZADO\_DE\_REPRESENTACIONES\_GEOMETRICAS>  
  <a1>**Aproximada**</a1>  
  </CALIFICADOR\_AUTORIZADO\_DE\_REPRESENTACIONES\_GEOMETRICAS>  
  </restricciones\_de\_integridad>  
= <relaciones>  
  <r1>**Conecta Camino (L)**</r1>  
  <r2>**Conecta Carretera (L)**</r2>  
  </relaciones>  
  </Banco\_de\_material>  
  </elemento>  
= <elemento nombre="Bordo">  
= <Bordo>  
= <atributos>  
= <dominio\_fijo>

```

<a1>Ninguno</a1>
  </dominio_fijo>
= <dominio_variable>
= <identificador_de_bordo>
  <w1>Un número secuencial que se incrementa con cada ocurrencia</w1>
= <dominio_de_valores>
  <w1>En el intervalo de 1 a N</w1>
  </dominio_de_valores>
  </identificador_de_bordo>
  </dominio_variable>
  </atributos>
= <restricciones_de_integridad>
= <COMBINACIONES_AUTORIZADAS_DE_VALORES_DE_ATRIBUTOS>
  <c1>Ninguna</c1>
  </COMBINACIONES_AUTORIZADAS_DE_VALORES_DE_ATRIBUTOS>
= <CALIFICADOR_AUTORIZADO_DE_REPRESENTACIONES_GEOMETRICAS>
  <a1>Definida</a1>
  </CALIFICADOR_AUTORIZADO_DE_REPRESENTACIONES_GEOMETRICAS>
  </restricciones_de_integridad>
= <relaciones>
  <r1>Conecta Bordo (L)</r1>
  <r2>Conecta Corriente de agua (L)</r2>
  <r3>Comparte Calle (L)</r3>
  <r4>Comparte Camino (L)</r4>
  <r5>Comparte Carretera (L)</r5>
  <r6>Comparte Cuerpo de agua (A)</r6>
  <r7>Comparte Área urbana (A)</r7>
  </relaciones>
  </Bordo>
  </elemento>
= <elemento nombre="Calle">
= <Calle>
= <atributos>
= <dominio_fijo>
= <Tipo_de_calle>

```

**<a1>Calle de primer orden : Boulevares, periféricos, ejes viales, viaductos y avenidas principales</a1>**

**<a2>Calle de segundo orden : Avenidas y calles de menor importancia que las definidas para calles de primer orden. Vialidad principal dentro de localidades pequeñas que sirven de enlace entre carreteras que las cruzan.</a2>**

**<a3>Calle de tercer orden : Calles de tránsito local dentro de una área urbana.</a3>**

**<a4>Calle de cuarto orden : Calles de tránsito local dentro de áreas diversas.</a4>**

**</Tipo\_de\_calle>**

**</dominio\_fijo>**

**= <dominio\_variable>**

**= <identificador\_de\_calle>**

**<w1>Un número secuencial que se incrementa con cada ocurrencia</w1>**

**= <dominio\_de\_valores>**

**<w1>En el intervalo de 1 a N</w1>**

**</dominio\_de\_valores>**

**</identificador\_de\_calle>**

**</dominio\_variable>**

**</atributos>**

**= <restricciones\_de\_integridad>**

**= <COMBINACIONES\_AUTORIZADAS\_DE\_VALORES\_DE\_ATRIBUTOS>**

**<c1>Calle de primer orden.</c1>**

**<c2>Calle de segundo orden.</c2>**

**<c3>Calle de tercer orden.</c3>**

**<c4>Calle de cuarto orden.</c4>**

**</COMBINACIONES\_AUTORIZADAS\_DE\_VALORES\_DE\_ATRIBUTOS>**

**= <CALIFICADOR\_AUTORIZADO\_DE\_REPRESENTACIONES\_GEOMETRICAS>**

**<a1>Definida</a1>**

**<a2>Aproximada</a2>**

**</CALIFICADOR\_AUTORIZADO\_DE\_REPRESENTACIONES\_GEOMETRICAS>**

**</restricciones\_de\_integridad>**

**= <relaciones>**

**<r1>Conecta Calle (L)</r1>**

**<r2>Conecta Camino (L)</r2>**

**<r3>Conecta Carretera (L)</r3>**

<r4>Conecta Área urbana (A)</r4>  
<r5>Conecta Vía férrea (L)</r5>  
<r6>Comparte Bordo (L)</r6>  
<r7>Comparte Puento (L)</r7>  
<r8>Comparte Túnel (L)</r8>  
<r9>Comparte Aeropuerto (A)</r9>  
<r10>Comparte Área urbana (A)</r10>  
<r11>Comparte Área verde urbana (A)</r11>  
<r12>Comparte Cementerio (A)</r12>  
<r13>Comparte Instalación deportiva y recreativa (A)</r13>  
<r14>Calle (L) Comparte Instalación diversa (A)</r14>  
<r15>Calle (L) Comparte Instalación industrial (A)</r15>  
<r16>Calle (L) Comparte Instalación portuaria (A)</r16>  
</relaciones>  
</Calle>  
</elemento>

= <elemento nombre="Camino">

= <Camino>

= <atributos>

= <dominio\_fijo>

= <Tipo\_de\_camino>

<a1>Brecha : Generalmente es posible la circulación de un vehículo.</a1>

<a2>Vereda : Sólo circulan personas y animales.</a2>

</Tipo\_de\_camino>

</dominio\_fijo>

= <dominio\_variable>

= <identificador\_de\_>

<w1>Un número secuencial que se incrementa con cada ocurrencia</w1>

= <dominio\_de\_valores>

<w1>En el intervalo de 1 a N</w1>

</dominio\_de\_valores>

</identificador\_de\_>

</dominio\_variable>

</atributos>

= <restricciones\_de\_integridad>

= <COMBINACIONES\_AUTORIZADAS\_DE\_VALORES\_DE\_ATRIBUTOS>

<c1>Brecha</c1>

<c2>Vereda</c2>

</COMBINACIONES\_AUTORIZADAS\_DE\_VALORES\_DE\_ATRIBUTOS>

= <CALIFICADOR\_AUTORIZADO\_DE\_REPRESENTACIONES\_GEOMETRICAS>

<a1>Definida</a1>

<a2>Aproximada</a2>

</CALIFICADOR\_AUTORIZADO\_DE\_REPRESENTACIONES\_GEOMETRICAS>

</restricciones\_de\_integridad>

= <relaciones>

<r1>Conecta Aeropuerto (P)</r1>

<r2>Conecta Banco de material (P)</r2>

<r3>Conecta Cementerio (P)</r3>

<r4>Conecta Mina (P)</r4>

<r5>Conecta Pozo de explotación (P)</r5>

<r6>Conecta Ruta de embarcación (P)</r6>

<r7>Conecta Calle (L)</r7>

<r8>Conecta Camino (L)</r8>

<r9>Conecta Carretera (L)</r9>

<r10>Conecta Presa (L)</r10>

<r11>Conecta Ruta de embarcación (L)</r11>

<r12>Conecta Cementerio (A)</r12>

<r13>Conecta Aeropuerto (A)</r13>

<r14>Conecta Área urbana (A)</r14>

<r15>Comparte Bordo (L)</r15>

<r16>Comparte Presa (L)</r16>

<r17>Comparte Puente (L)</r17>

<r18>Comparte Área de cultivo (A)</r18>

<r19>Comparte Área urbana (A)</r19>

<r20>Comparte Cementerio (A)</r20>

</relaciones>

</Camino>

</elemento>

= <elemento nombre="canal">

= <canal>

= <atributos>

= <dominio\_fijo>

= <Condicion\_del\_canal>  
 <a1>En construcción : Que está en proceso de construcción.</a1>  
 <a2>En operación : Que está en servicio.</a2>  
 <a3>Fuera de Uso : Que no está en uso.</a3>  
 </Condicion\_del\_canal>  
 </dominio\_fijo>

= <dominio\_variable>

= <identificador\_de\_canal>  
 <w1>Un número secuencial que se incrementa con cada ocurrencia</w1>

= <dominio\_de\_valores>  
 <w1>En el intervalo de 1 a N</w1>  
 </dominio\_de\_valores>  
 </identificador\_de\_canal>  
 </dominio\_variable>  
 </atributos>

= <restricciones\_de\_integridad>

= <COMBINACIONES\_AUTORIZADAS\_DE\_VALORES\_DE\_ATRIBUTOS>  
 <c1>En construcción.</c1>  
 <c2>En operación.</c2>  
 <c3>Fuera de uso.</c3>  
 </COMBINACIONES\_AUTORIZADAS\_DE\_VALORES\_DE\_ATRIBUTOS>

= <CALIFICADOR\_AUTORIZADO\_DE\_REPRESENTACIONES\_GEOMETRICAS>  
 <a1>Definida.</a1>  
 <a2>Virtual.</a2>  
 </CALIFICADOR\_AUTORIZADO\_DE\_REPRESENTACIONES\_GEOMETRICAS>  
 </restricciones\_de\_integridad>

= <relaciones>  
 <r1>Conecta Instalación de bombeo (P)</r1>  
 <r2>Conecta Manantial (P)</r2>  
 <r3>Conecta Tanque de agua (P)</r3>  
 <r4>Conecta Acueducto (L)</r4>  
 <r5>Conecta CEn construcción: Que está en proceso de construcciónanal  
 (L)</r5>  
 <r6>Conecta Corriente de agua (L)</r6>  
 <r7>Conecta Canal (A)</r7>  
 <r8>Conecta Cuerpo de agua (A)</r8>

<r9>Conecta Estanque (A)</r9>  
<r10>Conecta Tanque de agua (A)</r10>  
<r11>Comparte Puentes (L)</r11>  
<r12>Comparte Túnel (L)</r12>  
<r13>Comparte Área de cultivo (A)</r13>  
<r14>Comparte Instalación de bombeo (A)</r14>  
<r15>Comparte Instalación diversa (A)</r15>  
<r16>Comparte Área urbana (A)</r16>  
<r17>Conecta Acueducto (L)</r17>  
<r18>Conecta Canal (L)</r18>  
<r19>Conecta Corriente de agua (L)</r19>  
<r20>Comparte Área de cultivo (A)</r20>  
<r21>Comparte Área urbana (A)</r21>  
<r22>Comparte Cuerpo de agua (A)</r22>  
<r23>Comparte Estanque (A)</r23>  
<r24>Comparte Instalación de bombeo (A)</r24>  
<r25>Comparte Instalación diversa (A)</r25>  
<r26>Comparte Salina (A)</r26>  
<r27>Comparte Tanque de agua (A)</r27>  
<r28>Comparte Vegetación densa (A)</r28>  
</relaciones>  
</canal>  
</elemento>

= <elemento nombre="Carretera">  
= <Carretera>  
= <atributos>  
= <dominio\_fijo>  
= <Tipo\_de\_carretera>  
<a1>Pavimentada : De asfalto o concreto.</a1>  
<a2>Terracería : De arena y grava compactada.</a2>  
</Tipo\_de\_carretera>  
= <Número\_de\_carriles>  
<a1>Un carril.</a1>  
<a2>Dos carriles.</a2>  
<a3>Tres carriles.</a3>  
<a4>Cuatro carriles.</a4>

<a5>Cinco carriles.</a5>

<a6>Seis carriles.</a6>

<a7>Más de seis carriles.</a7>

= <Restricciones\_del\_atributo>

<a1>No aplicable (N/A) : Indica que este atributo no es válido, cuando el valor del atributo Condición de la Carretera es Fuera de Uso o En construcción o cuando el valor del atributo Derecho de tránsito es Restringido.</a1>

</Restricciones\_del\_atributo>

</Número\_de\_carriles>

= <Derecho\_de\_transito\_de\_la\_carretera>

<a1>Cuota : Con pago de peaje.</a1>

<a2>Libre : Sin pago de peaje.</a2>

<a3>Restringido : de acceso exclusivo</a3>

= <Restricciones\_del\_atributo>

<a1>No aplicable (N/A) : Indica que este atributo no es válido, cuando el valor del atributo Condición de la Carretera es Fuera de Uso o En construcción.</a1>

</Restricciones\_del\_atributo>

</Derecho\_de\_transito\_de\_la\_carretera>

= <Administracion\_de\_la\_carretera>

<a1>Concesionada : Bajo el cargo de particulares, con pago de peaje.</a1>

<a2>Estatal : Bajo el cargo de la entidad federativa donde se localiza.</a2>

<a3>Federal : Bajo el cargo del gobierno federal.</a3>

<a4>Otro : El valor del atributo es diferente de los valores citados anteriormente.</a4>

= <Restricciones\_del\_atributo>

<a1>Desconocido: El valor del atributo es alguno de los citados en el dominio, pero no se sabe actualmente cual de ellos es para alguna(s) carretera(s).</a1>

<a2>No aplicable (N/A) : Indica que este atributo no es válido, cuando el valor del atributo Condición de la Carretera es Fuera de Uso o En construcción, o cuando el valor del atributo Derecho de tránsito es Restringido.</a2>

</Restricciones\_del\_atributo>

</Administracion\_de\_la\_carretera>

= <Condicion\_de\_la\_carretera>

<a1>En construcción: Que está en proceso de construcción</a1>

<a2>En operación : Que está en servicio o puede usarse</a2>

<a3>Fuera de uso : Que no está en uso</a3>

= <Restricciones\_del\_atributo>

<a1>No aplicable (N/A) : Indica que este atributo no es válido, cuando el valor del atributo Derecho de tránsito es Restringido.</a1>

</Restricciones\_del\_atributo>

</Condicion\_de\_la\_carretera>

</dominio\_fijo>

= <dominio\_variable>

= <identificador\_de\_carretera>

<w1>Un número secuencial que se incrementa con cada ocurrencia</w1>

= <dominio\_de\_valores>

<w1>En el intervalo de 1 a N</w1>

</dominio\_de\_valores>

</identificador\_de\_carretera>

= <Número\_de\_carretera>

<w1>Número oficial asignado a la carretera.</w1>

= <dominio\_de\_valores>

<w1>En el intervalo de 1 a N</w1>

= <Restricciones\_del\_atributo>

<a1>Ninguno : El atributo carece de valor para alguna ocurrencia de la entidad.</a1>

<a2>No aplicable (N/A) : Indica que este atributo no es válido, cuando el valor del atributo Administración de la carretera es otro o bien cuando tiene la restricción desconocido.</a2>

</Restricciones\_del\_atributo>

</dominio\_de\_valores>

</Número\_de\_carretera>

= <Jurisdiccion>

<w1>Abreviatura del nombre de la entidad federativa que ejerce autoridad jurídica sobre la carretera.</w1>

= <dominio\_de\_valores>

<j1>Aguascalientes (Ags.)</j1>

<j2>Guerrero (Gro.)</j2>

<j3>Quintana Roo (Q.R.)</j3>

<j4>Baja California (B.C.)</j4>

<j5>Hidalgo (Hgo.)</j5>  
<j6>San Luis Potosí (S.L.P.)</j6>  
<j7>Baja California Sur (B.C.S.)</j7>  
<j8>Jalisco (Jal.)</j8>  
<j9>Sinaloa (Sin.)</j9>  
<j10>Campeche (Camp.)</j10>  
<j11>México, estado de (Edo. Mex.)</j11>  
<j12>Sonora (Son.)</j12>  
<j13>Coahuila (Coah.)</j13>  
<j14>Michoacán (Mich.)</j14>  
<j15>Tabasco (Tab.)</j15>  
<j16>Colima (Col.)</j16>  
<j17>Morelos (Mor.)</j17>  
<j18>Tamaulipas (Tamps.)</j18>  
<j19>Chiapas (Chis.)</j19>  
<j20>Nayarit (Nay.)</j20>  
<j21>Tlaxcala (Tlax.)</j21>  
<j22>Chihuahua (Chih.)</j22>  
<j23>Nuevo León (N.L.)</j23>  
<j24>Veracruz (Ver.)</j24>  
<j25>Distrito Federal (D.F.)</j25>  
<j26>Oaxaca (Oax.)</j26>  
<j27>Yucatán (Yuc.)</j27>  
<j28>Durango (Dgo.)</j28>  
<j29>Puebla (Pue.)</j29>  
<j30>Zacatecas (Zac.)</j30>  
<j31>Guanajuato (Gto.)</j31>  
<j32>Querétaro (Qro.)</j32>  
<j33>Federación\* (Fed.)</j33>

= <Restricciones\_del\_atributo>

<a1>No aplicable (N/A) : Indica que este atributo no es válido, cuando el valor del atributo Administración de la carretera es otro o bien cuando tiene la restricción desconocido.</a1>

</Restricciones\_del\_atributo>

</dominio\_de\_valores>

</Jurisdiccion>

</dominio\_variable>

</atributos>

= <restricciones\_de\_integridad>

= <COMBINACIONES\_AUTORIZADAS\_DE\_VALORES\_DE\_ATRIBUTOS>

- <c1>Pavimentada, un carril, cuota, concesionada, en operación.</c1>
- <c2>Pavimentada, un carril, cuota, estatal, en operación.</c2>
- <c3>Pavimentada, un carril, cuota, federal, en operación.</c3>
- <c4>Pavimentada, un carril, libre, estatal, en operación.</c4>
- <c5>Pavimentada, un carril, libre, federal, en operación.</c5>
- <c6>Pavimentada, un carril, libre, desconocido, en operación.</c6>
- <c7>Pavimentada, un carril, libre, otro, en operación.</c7>
- <c8>Pavimentada, dos carriles, cuota, concesionada, en operación.</c8>
- <c9>Pavimentada, dos carriles, cuota, estatal, en operación.</c9>
- <c10>Pavimentada, dos carriles, cuota, federal, en operación.</c10>
- <c11>Pavimentada, dos carriles, libre, desconocido, en operación.</c11>
- <c12>Pavimentada, dos carriles, libre, estatal, en operación.</c12>
- <c13>Pavimentada, dos carriles, libre, federal, en operación.</c13>
- <c14>Pavimentada, dos carriles, libre, otro, en operación.</c14>
- <c15>Pavimentada, tres carriles, cuota, concesionada, en operación.</c15>
- <c16>Pavimentada, tres carriles, cuota, estatal, en operación.</c16>
- <c17>Pavimentada, tres carriles, cuota, federal, en operación.</c17>
- <c18>Pavimentada, tres carriles, libre, estatal, en operación.</c18>
- <c19>Pavimentada, tres carriles, libre, federal, en operación.</c19>
- <c20>Pavimentada, cuatro carriles, cuota, concesionada, en operación.</c20>
- <c21>Pavimentada, cuatro carriles, cuota, estatal, en operación.</c21>
- <c22>Pavimentada, cuatro carriles, cuota, federal, en operación.</c22>
- <c23>Pavimentada, cuatro carriles, libre, estatal, en operación.</c23>
- <c24>Pavimentada, cuatro carriles, libre, federal, en operación.</c24>
- <c25>Pavimentada, cinco carriles, cuota, concesionada, en operación.</c25>
- <c26>Pavimentada, cinco carriles, cuota, estatal, en operación.</c26>
- <c27>Pavimentada, cinco carriles, cuota, federal, en operación.</c27>
- <c28>Pavimentada, cinco carriles, libre, estatal, en operación.</c28>
- <c29>Pavimentada, cinco carriles, libre, federal, en operación.</c29>
- <c30>Pavimentada, seis carriles, cuota, concesionada, en operación.</c30>
- <c31>Pavimentada, seis carriles, cuota, estatal, en operación.</c31>
- <c32>Pavimentada, seis carriles, cuota, federal, en operación.</c32>
- <c33>Pavimentada, seis carriles, libre, estatal, en operación.</c33>

<c34>Pavimentada, seis carriles, libre, federal. en operación.</c34>  
<c35>Pavimentada, más de seis carriles, cuota, concesionada, en  
operación.</c35>  
<c36>Pavimentada, más de seis carriles, cuota, estatal, en operación.</c36>  
<c37>Pavimentada, más de seis carriles, cuota, federal, en operación.</c37>  
<c38>Pavimentada, más de seis carriles, libre, estatal, en operación.</c38>  
<c39>Pavimentada, más de seis carriles, libre, federal, en operación.</c39>  
<c40>Pavimentada, N/A, restringido, N/A, N/A.</c40>  
<c41>Pavimentada, N/A, N/A, N/A, en construcción.</c41>  
<c42>Pavimentada, N/A, N/A, N/A, en operación.</c42>  
<c43>Pavimentada, N/A, N/A, N/A, fuera de uso.</c43>  
<c44>Terracería, un carril, N/A, N/A, en operación.</c44>  
<c45>Terracería, dos carriles, N/A, N/A, en operación.</c45>  
<c46>Terracería, N/A, restringido, N/A, N/A.</c46>  
<c47>Terracería, N/A, N/A, N/A, fuera de uso.</c47>

</COMBINACIONES\_AUTORIZADAS\_DE\_VALORES\_DE\_ATRIBUTOS>

= <CALIFICADOR\_AUTORIZADO\_DE\_REPRESENTACIONES\_GEOMETRICAS>

<a1>Definida.</a1>

<a2>Aproximada</a2>

</CALIFICADOR\_AUTORIZADO\_DE\_REPRESENTACIONES\_GEOMETRICAS>

</restricciones\_de\_integridad>

= <relaciones>

<r1>Conecta Aeropuerto (P)</r1>

<r2>Conecta Banco de material (P)</r2>

<r3>Conecta Caseta de peaje (P)</r3>

<r4>Conecta Cementerio (P)</r4>

<r5>Conecta Mina (P)</r5>

<r6>Conecta Pozo de explotación (P)</r6>

<r7>Conecta Ruta de embarcación (P)</r7>

<r8>Conecta Vado (P)</r8>

<r9>Conecta Calle (L)</r9>

<r11>Conecta Carretera (L)</r11>

<r12>Conecta Ruta de embarcación (L)</r12>

<r13>Conecta Vía férrea (L)</r13>

<r14>Conecta Aeropuerto (A)</r14>

<r15>Conecta Área urbana (A)</r15>

<r16>Conecta Cementerio (A)</r16>  
<r17>Conecta Instalación diversa (A)</r17>  
<r18>Conecta Instalación industrial (A)</r18>  
<r19>Comparte Bordo (L)</r19>  
<r20>Comparte Instalación portuaria (L)</r20>  
<r21>Comparte Presa (L)</r21>  
<r22>Comparte Puente (L)</r22>  
<r23>Comparte Túnel (L)</r23>  
<r24>Comparte Vado (L)</r24>  
<r25>Comparte Aeropuerto (A)</r25>  
<r26>Comparte Área natural protegida (A)</r26>  
<r27>Comparte Área urbana (A)</r27>  
<r28>Comparte Cementerio (A)</r28>  
<r29>Comparte Instalación diversa (A)</r29>  
<r30>Comparte Instalación industrial (A)</r30>  
<r31>Comparte Área urbana (A)</r31>

</relaciones>

</Carretera>

</elemento>

= <elemento nombre="Caseta\_de\_peaje">

= <Caseta\_de\_peaje>

= <atributos>

= <dominio\_fijo>

<a1>Ninguno</a1>

</dominio\_fijo>

= <dominio\_variable>

= <identificador\_de\_caseta\_de\_peaje>

<w1>Un número secuencial que se incrementa con cada ocurrencia</w1>

= <dominio\_de\_valores>

<w1>En el intervalo de 1 a N</w1>

</dominio\_de\_valores>

</identificador\_de\_caseta\_de\_peaje>

</dominio\_variable>

</atributos>

= <restricciones\_de\_integridad>

= <COMBINACIONES\_AUTORIZADAS\_DE\_VALORES\_DE\_ATRIBUTOS>

```
<c1>Ninguna</c1>
  </COMBINACIONES_AUTORIZADAS_DE_VALORES_DE_ATRIBUTOS>
= <CALIFICADOR_AUTORIZADO_DE_REPRESENTACIONES_GEOMETRICAS>
  <a1>Definida</a1>
  </CALIFICADOR_AUTORIZADO_DE_REPRESENTACIONES_GEOMETRICAS>
  </restricciones_de_integridad>
= <relaciones>
  <r1>Conecta Carretera (L)</r1>
  </relaciones>
  </Caseta_de_peaje>
  </elemento>
= <elemento nombre="Cementerio">
= <Cementerio>
= <atributos>
= <dominio_fijo>
  <a1>Ninguno</a1>
  </dominio_fijo>
= <dominio_variable>
= <identificador_de_cementerio>
  <w1>Un número secuencial que se incrementa con cada ocurrencia</w1>
= <dominio_de_valores>
  <w1>En el intervalo de 1 a N</w1>
  </dominio_de_valores>
  </identificador_de_cementerio>
  </dominio_variable>
  </atributos>
= <restricciones_de_integridad>
= <COMBINACIONES_AUTORIZADAS_DE_VALORES_DE_ATRIBUTOS>
  <c1>Ninguna</c1>
  </COMBINACIONES_AUTORIZADAS_DE_VALORES_DE_ATRIBUTOS>
= <CALIFICADOR_AUTORIZADO_DE_REPRESENTACIONES_GEOMETRICAS>
  <a1>Aproximada</a1>
  <a2>Definida</a2>
  <a3>Virtual</a3>
  </CALIFICADOR_AUTORIZADO_DE_REPRESENTACIONES_GEOMETRICAS>
  </restricciones_de_integridad>
```

```
= <relaciones>
  <r1>Conecta Camino (L)</r1>
  <r2>Conecta Carretera (L)</r2>
  <r3>Conecta Camino (L)</r3>
  <r4>Conecta Carretera (L)</r4>
  <r5>Comparte Calle (L)</r5>
  <r6>Comparte Camino (L)</r6>
  <r7>Comparte Carretera (L)</r7>
  <r8>Comparte Área de cultivo (A)</r8>
  <r9>Comparte Área verde urbana (A)</r9>
  <r10>Comparte Área urbana (A)</r10>
  </relaciones>
  </Cementerio>
  </elemento>
= <elemento nombre="Conducto">
= <Conducto>
= <atributos>
= <dominio_fijo>
= <Número_de_conductos>
  <a1>Un conducto.</a1>
  <a2>Dos conductos.</a2>
  <a3>Tres conductos.</a3>
= <Restricciones_del_atributo>
  <a1>No aplicable (N/A) : Indica que este atributo no es válido, cuando el
    valor del atributo Condición del Conducto es Fuera de Uso y el Tipo de
    conducto es Otro.</a1>
  </Restricciones_del_atributo>
  </Número_de_conductos>
= <Tipo_de_conducto>
  <a1>Pemex. Para petróleo y sus derivados</a1>
  <a2>Otro. El valor del atributo es diferente del citado anteriormente</a2>
  </Tipo_de_conducto>
= <Relacion_del_conducto_con_el_suelo>
  <a1>Subterráneo : Bajo la superficie del terreno.</a1>
  <a2>Superficial : Sobre la superficie del terreno.</a2>
= <Restricciones_del_atributo>
```

**<a1>No aplicable (N/A) : Indica que este atributo no es válido, cuando el valor del atributo Condición del Conducto es Fuera de Uso y el Tipo de conducto es Otro.</a1>**

</Restricciones\_del\_atributo>

</Relacion\_del\_conducto\_con\_el\_suelo>

= <Condicion\_del\_conducto>

**<a1>En operación : Que está en servicio o puede usarse</a1>**

**<a2>Fuera de uso : Que no está en uso</a2>**

</Condicion\_del\_conducto>

</dominio\_fijo>

= <dominio\_variable>

= <identificador\_del\_conducto>

**<w1>Un número secuencial que se incrementa con cada ocurrencia</w1>**

= <dominio\_de\_valores>

**<w1>En el intervalo de 1 a N</w1>**

</dominio\_de\_valores>

</identificador\_del\_conducto>

</dominio\_variable>

</atributos>

= <restricciones\_de\_integridad>

= <COMBINACIONES\_AUTORIZADAS\_DE\_VALORES\_DE\_ATRIBUTOS>

**<c1>Un conducto, Pemex, subterráneo, en operación.</c1>**

**<c2>Un conducto, Pemex, superficial, en operación.</c2>**

**<c3>Un conducto, otro, subterráneo, en operación.</c3>**

**<c4>Un conducto, otro, superficial, en operación.</c4>**

**<c5>Dos conductos, Pemex, subterráneo, en operación.</c5>**

**<c6>Dos conductos, Pemex, superficial, en operación.</c6>**

**<c7>Dos conductos, otro, subterráneo, en operación.</c7>**

**<c8>Dos conductos, otro, superficial, en operación.</c8>**

**<c9>Tres conductos, Pemex, subterráneo, en operación.</c9>**

**<c10>Tres conductos, Pemex, superficial, en operación.</c10>**

**<c11>Tres conductos, otro, subterráneo, en operación.</c11>**

**<c12>Tres conductos, otro, superficial, en operación.</c12>**

**<c13>Un conducto, Pemex, subterráneo, fuera de uso.</c13>**

**<c14>Un conducto, Pemex, superficial, fuera de uso.</c14>**

**<c15>Dos conductos, Pemex, subterráneo, fuera de uso.</c15>**

**<c16>Dos conductos, Pemex, superficial, fuera de uso.</c16>**

<c17>Tres conductos, Pemex, subterráneo, fuera de uso.</c17>

<c18>Tres conductos, Pemex, superficial, fuera de uso.</c18>

<c19>N/A, otro, N/A, fuera de uso.</c19>

</COMBINACIONES\_AUTORIZADAS\_DE\_VALORES\_DE\_ATRIBUTOS>

= <CALIFICADOR\_AUTORIZADO\_DE\_REPRESENTACIONES\_GEOMETRICAS>

<a1>Definida.</a1>

<a2>Aproximada.</a2>

</CALIFICADOR\_AUTORIZADO\_DE\_REPRESENTACIONES\_GEOMETRICAS>

</restricciones\_de\_integridad>

= <relaciones>

<r1>Conecta Pozo de explotación (P)</r1>

<r2>Conecta Instalación de bombeo (P)</r2>

<r3>Conecta Conducto (L)</r3>

<r4>Conecta Instalación de bombeo (A)</r4>

</relaciones>

</Conducto>

</elemento>

= <elemento nombre="Corriente\_de\_agua">

= <Corriente\_de\_agua>

= <atributos>

= <dominio\_fijo>

= <Condicion\_de\_la\_corriente\_de\_agua>

<a1>Intermitente : Con presencia de agua en determinadas épocas del año.</a1>

<a2>Perenne : Con presencia de agua permanentemente.</a2>

</Condicion\_de\_la\_corriente\_de\_agua>

</dominio\_fijo>

= <dominio\_variable>

= <identificador\_de\_la\_corriente\_de\_agua>

<w1>Un número secuencial que se incrementa con cada ocurrencia</w1>

= <dominio\_de\_valores>

<w1>En el intervalo de 1 a N</w1>

</dominio\_de\_valores>

</identificador\_de\_la\_corriente\_de\_agua>

</dominio\_variable>

</atributos>

```
= <restricciones_de_integridad>
= <COMBINACIONES_AUTORIZADAS_DE_VALORES_DE_ATRIBUTOS>
  <c1>Intermitente.</c1>
  <c2>Perenne.</c2>
  </COMBINACIONES_AUTORIZADAS_DE_VALORES_DE_ATRIBUTOS>
= <CALIFICADOR_AUTORIZADO_DE_REPRESENTACIONES_GEOMETRICAS>
  <a1>Definida</a1>
  </CALIFICADOR_AUTORIZADO_DE_REPRESENTACIONES_GEOMETRICAS>
</restricciones_de_integridad>
= <relaciones>
  <r1>Conecta Corriente que desaparece (P)</r1>
  <r2>Conecta Entrada a gruta (P)</r2>
  <r3>Conecta Manantial (P)</r3>
  <r4>Conecta Planta generadora (P)</r4>
  <r5>Conecta Rápido (P)</r5>
  <r6>Conecta Ruta de embarcación (P)</r6>
  <r7>Conecta Salto de agua (P)</r7>
  <r8>Conecta Bordo (L)</r8>
  <r9>Conecta Canal (L)</r9>
  <r10>Conecta Corriente de agua (L)</r10>
  <r11>Conecta Presa (L)</r11>
  <r12>Conecta Canal (A)</r12>
  <r13>Conecta Cuerpo de agua (A)</r13>
  <r14>Comparte Límite (L)</r14>
  <r15>Comparte Área de cultivo (A)</r15>
  <r16>Comparte Área urbana (A)</r16>
  </relaciones>
  </Corriente_de_agua>
</elemento>
= <elemento nombre="Corriente_que_desaparece">
= <Corriente_que_desaparece>
= <atributos>
= <dominio_fijo>
  <a1>Ninguno</a1>
  </dominio_fijo>
= <dominio_variable>
```

```

= <identificador_de_corriente_que_desaparece>
  <w1>Un número secuencial que se incrementa con cada ocurrencia</w1>
= <dominio_de_valores>
  <w1>En el intervalo de 1 a N</w1>
  </dominio_de_valores>
  </identificador_de_corriente_que_desaparece>
  </dominio_variable>
  </atributos>
= <restricciones_de_integridad>
= <COMBINACIONES_AUTORIZADAS_DE_VALORES_DE_ATRIBUTOS>
  <c1>Ninguna</c1>
  </COMBINACIONES_AUTORIZADAS_DE_VALORES_DE_ATRIBUTOS>
= <CALIFICADOR_AUTORIZADO_DE_REPRESENTACIONES_GEOMETRICAS>
  <a1>Definida</a1>
  </CALIFICADOR_AUTORIZADO_DE_REPRESENTACIONES_GEOMETRICAS>
  </restricciones_de_integridad>
= <relaciones>
  <r1>Conecta Corriente de agua (L)</r1>
  </relaciones>
  </Corriente_que_desaparece>
  </elemento>
= <elemento nombre="Cuerpo_de_agua">
= <Cuerpo_de_agua>
= <atributos>
= <dominio_fijo>
= <Condicion_del_cuerpo_de_agua>
  <a1>Intermitente : Con presencia de agua en determinadas épocas del
    año.</a1>
  <a2>Perenne : Con presencia de agua permanentemente.</a2>
  </Condicion_del_cuerpo_de_agua>
  </dominio_fijo>
= <dominio_variable>
= <identificador_de_cuerpo_de_agua>
  <w1>Un número secuencial que se incrementa con cada ocurrencia</w1>
= <dominio_de_valores>
  <w1>En el intervalo de 1 a N</w1>

```

</dominio\_de\_valores>  
</identificador\_de\_cuerpo\_de\_agua>  
</dominio\_variable>  
</atributos>

= <restricciones\_de\_integridad>

= <COMBINACIONES\_AUTORIZADAS\_DE\_VALORES\_DE\_ATRIBUTOS>

<c1>Intermitente.</c1>

<c2>Perenne.</c2>

</COMBINACIONES\_AUTORIZADAS\_DE\_VALORES\_DE\_ATRIBUTOS>

= <CALIFICADOR\_AUTORIZADO\_DE\_REPRESENTACIONES\_GEOMETRICAS>

<a1>Definida</a1>

<a2>Aproximada.</a2>

<a3>Virtual</a3>

</CALIFICADOR\_AUTORIZADO\_DE\_REPRESENTACIONES\_GEOMETRICAS>

</restricciones\_de\_integridad>

= <relaciones>

<r1>Conecta Acueducto (L)</r1>

<r2>Conecta Canal (L)</r2>

<r3>Conecta Corriente de agua (L)</r3>

<r4>Conecta Curva de nivel (L)</r4>

<r5>Conecta Instalación portuaria (L)</r5>

<r6>Conecta Rápido (L)</r6>

<r7>Conecta Ruta de embarcación (L)</r7>

<r8>Conecta Salto de agua (L)</r8>

<r9>Comparte Bordo (L)</r9>

<r10>Comparte Instalación portuaria (L)</r10>

<r11>Comparte Límite (L)</r11>

<r12>Comparte Presa (L)</r12>

<r13>Comparte Área de cultivo (A)</r13>

<r14>Comparte Área urbana (A)</r14>

<r15>Comparte Área verde urbana (A)</r15>

<r16>Comparte Canal (A)</r16>

<r17>Comparte Cuerpo de agua (A)</r17>

<r18>Comparte Instalación portuaria</r18>

<r19>Comparte Malpaís (A)</r19>

<r20>Comparte Pantano (A)</r20>

<r21>Comparte Salina (A)</r21>

<r22>Comparte Terreno sujeto a inundación (A)</r22>

<r23>Comparte Vegetación densa (A)</r23>

<r24>Comparte Zona arenosa (A)</r24>

</relaciones>

</Cuerpo\_de\_agua>

</elemento>

= <elemento nombre="Curva\_de\_nivel">

= <Curva\_de\_nivel>

= <atributos>

= <dominio\_fijo>

= <Tipo\_de\_curva\_de\_nivel>

<a1>Depresión : Para representar un hundimiento en el terreno donde no hay salida del drenaje.</a1>

<a2>Otro : El valor del atributo es diferente de los valores citados anteriormente.</a2>

</Tipo\_de\_curva\_de\_nivel>

</dominio\_fijo>

= <dominio\_variable>

= <identificador\_de\_curva\_de\_nivel>

<w1>Un número secuencial que se incrementa con cada ocurrencia</w1>

= <dominio\_de\_valores>

<w1>En el intervalo de 1 a N</w1>

</dominio\_de\_valores>

</identificador\_de\_curva\_de\_nivel>

= <elevacion\_de\_curva\_de\_nivel>

<w1>Valor en metros de la elevación.</w1>

= <dominio\_de\_valores>

<w1>-10 mayor o igual que valor menor o igual que 5610</w1>

</dominio\_de\_valores>

</elevacion\_de\_curva\_de\_nivel>

</dominio\_variable>

</atributos>

= <restricciones\_de\_integridad>

= <COMBINACIONES\_AUTORIZADAS\_DE\_VALORES\_DE\_ATRIBUTOS>

<c1>Depresion</c1>

<c2>Otro</c2>

```

    </COMBINACIONES_AUTORIZADAS_DE_VALORES_DE_ATRIBUTOS>
  = <CALIFICADOR_AUTORIZADO_DE_REPRESENTACIONES_GEOMETRICAS>
    <a1>Definida</a1>
    <a2>Aproximada</a2>
    </CALIFICADOR_AUTORIZADO_DE_REPRESENTACIONES_GEOMETRICAS>
    </restricciones_de_integridad>
  = <relaciones>
    <r1>Conecta Curva de nivel (L)</r1>
    <r2>Conecta Cuerpo de agua (A)</r2>
    </relaciones>
    </Curva_de_nivel>
    </elemento>
  = <elemento nombre="Deposito_de_desechos">
  = <Deposito_de_desechos>
  = <atributos>
  = <dominio_fijo>
  = <Tipo_de_deposito_de_desechos>
    <a1>Jale o terrero: Depósito para desecho de minas e industria.</a1>
    <a2>Relleno sanitario : Depósito que se va rellinando en forma sistemática
      con capas de basura.</a2>
    <a3>Tiradero : Depósito a cielo abierto.</a3>
    </Tipo_de_deposito_de_desechos>
    </dominio_fijo>
  = <dominio_variable>
  = <identificador_de_deposito_de_desechos>
    <w1>Un número secuencial que se incrementa con cada ocurrencia</w1>
  = <dominio_de_valores>
    <w1>En el intervalo de 1 a N</w1>
    </dominio_de_valores>
    </identificador_de_deposito_de_desechos>
    </dominio_variable>
    </atributos>
  = <restricciones_de_integridad>
  = <COMBINACIONES_AUTORIZADAS_DE_VALORES_DE_ATRIBUTOS>
    <c1>Jale.</c1>
    <c2>Relleno sanitario.</c2>

```

```
<c3>Tiradero.</c3>
  </COMBINACIONES_AUTORIZADAS_DE_VALORES_DE_ATRIBUTOS>
= <CALIFICADOR_AUTORIZADO_DE_REPRESENTACIONES_GEOMETRICAS>
  <a1>Aproximada</a1>
  </CALIFICADOR_AUTORIZADO_DE_REPRESENTACIONES_GEOMETRICAS>
  </restricciones_de_integridad>
= <relaciones>
  <r1>Ninguna</r1>
  </relaciones>
  </Deposito_de_desechos>
  </elemento>
= <elemento nombre="Edificacion">
= <Edificacion>
= <atributos>
= <dominio_fijo>
= <Funcion_de_la_edificacion>
  <a1>Aduana y/o control de migración : Módulos extraurbanos donde se lleva
    a cabo la revisión legal de personas y/o mercancías que llegan o salen del
    país.</a1>
  <a2>Centro de asistencia médica : Para proporcionar servicios médicos.</a2>
  <a3>Escuela : Para la enseñanza oficial de nivel básico a medio
    superior.</a3>
  <a4>Estación del metro : Para el ascenso y descenso de pasajeros del
    Sistema de Transporte Colectivo (Metro).</a4>
  <a5>Estación del tren ligero : Para el ascenso y descenso de pasajeros del
    Sistema de Transporte Colectivo (Tren Ligero).</a5>
  <a6>Granja o Establo : Para la reproducción, cría industrial y explotación de
    especies animales.</a6>
  <a7>Monumento u obelisco : Escultura de carácter conmemorativo.</a7>
  <a8>Pirámide : Construcción prehispánica.</a8>
  <a9>Sitio histórico : Lugar donde ocurrió un suceso histórico de relevancia
    nacional.</a9>
  <a10>Templo : Para la celebración de un culto religioso.</a10>
  <a11>Otro : El valor del atributo es diferente de los valores citados
    anteriormente.</a11>
  </Funcion_de_la_edificacion>
  </dominio_fijo>
```

```

= <dominio_variable>
= <identificador_de_edificacion>
  <w1>Un número secuencial que se incrementa con cada ocurrencia</w1>
= <dominio_de_valores>
  <w1>En el intervalo de 1 a N</w1>
  </dominio_de_valores>
  </identificador_de_edificacion>
  </dominio_variable>
  </atributos>
= <restricciones_de_integridad>
= <COMBINACIONES_AUTORIZADAS_DE_VALORES_DE_ATRIBUTOS>
  <c1>Aduana y/o control de migración.</c1>
  <c2>Centro de asistencia médica.</c2>
  <c3>Escuela.</c3>
  <c4>Estación del metro.</c4>
  <c5>Estación del tren ligero.</c5>
  <c6>Granja o establo.</c6>
  <c7>Monumento u obelisco.</c7>
  <c8>Pirámide.</c8>
  <c9>Sitio histórico.</c9>
  <c10>Templo.</c10>
  <c11>Otro.</c11>
  </COMBINACIONES_AUTORIZADAS_DE_VALORES_DE_ATRIBUTOS>
= <CALIFICADOR_AUTORIZADO_DE_REPRESENTACIONES_GEOMETRICAS>
  <a1>Definida.</a1>
  <a2>Virtual.</a2>
  </CALIFICADOR_AUTORIZADO_DE_REPRESENTACIONES_GEOMETRICAS>
  </restricciones_de_integridad>
= <relaciones>
  <r1>Ninguna</r1>
  </relaciones>
  </Edificacion>
  </elemento>
= <elemento nombre="Entrada_a_gruta">
= <Entrada_a_gruta>
= <atributos>

```

= <dominio\_fijo>  
 <a1>**Ninguno**</a1>  
 </dominio\_fijo>

= <dominio\_variable>

= <identificador\_de\_entrada\_a\_gruta>  
 <w1>**Un número secuencial que se incrementa con cada ocurrencia**</w1>

= <dominio\_de\_valores>  
 <w1>**En el intervalo de 1 a N**</w1>  
 </dominio\_de\_valores>  
 </identificador\_de\_entrada\_a\_gruta>  
 </dominio\_variable>  
 </atributos>

= <restricciones\_de\_integridad>

= <COMBINACIONES\_AUTORIZADAS\_DE\_VALORES\_DE\_ATRIBUTOS>  
 <c1>**Ninguna.**</c1>  
 </COMBINACIONES\_AUTORIZADAS\_DE\_VALORES\_DE\_ATRIBUTOS>

= <CALIFICADOR\_AUTORIZADO\_DE\_REPRESENTACIONES\_GEOMETRICAS>  
 <a1>**Aproximada.**</a1>  
 </CALIFICADOR\_AUTORIZADO\_DE\_REPRESENTACIONES\_GEOMETRICAS>  
 </restricciones\_de\_integridad>

= <relaciones>  
 <r1>**Conecta Corriente de agua (L)**</r1>  
 </relaciones>  
 </Entrada\_a\_gruta>  
 </elemento>

= <elemento nombre="Estanque">

= <Estanque>

= <atributos>

= <dominio\_fijo>

= <Tipo\_de\_estanque>  
 <a1>**Estanque acuícola : Utilizado para reproducir y criar especies acuícolas.**</a1>  
 <a2>**Estanque de sedimentación : Utilizado en el reciclaje de aguas residuales.**</a2>  
 <a3>**Estanque regulador : Utilizado en un sistema de conducción de agua.**</a3>

```

<a4>Otro : El valor del atributo es diferente de los valores citados
    anteriormente.</a4>
</Tipo_de_estanque>
</dominio_fijo>
= <dominio_variable>
= <identificador_de_estanque>
  <w1>Un número secuencial que se incrementa con cada ocurrencia</w1>
= <dominio_de_valores>
  <w1>En el intervalo de 1 a N</w1>
  </dominio_de_valores>
  </identificador_de_estanque>
  </dominio_variable>
  </atributos>
= <restricciones_de_integridad>
= <COMBINACIONES_AUTORIZADAS_DE_VALORES_DE_ATRIBUTOS>
  <c1>Estanque acuícola.</c1>
  <c2>Estanque de sedimentación.</c2>
  <c3>Estanque regulador.</c3>
  <c4>Otro.</c4>
  </COMBINACIONES_AUTORIZADAS_DE_VALORES_DE_ATRIBUTOS>
= <CALIFICADOR_AUTORIZADO_DE_REPRESENTACIONES_GEOMETRICAS>
  <a1>Definida.</a1>
  <a2>Virtual.</a2>
  </CALIFICADOR_AUTORIZADO_DE_REPRESENTACIONES_GEOMETRICAS>
  </restricciones_de_integridad>
= <relaciones>
  <r1>Conecta Acueducto (L)</r1>
  <r2>Conecta Canal (L)</r2>
  <r3>Conecta Separador (L)</r3>
  <r4>Comparte Canal (A)</r4>
  </relaciones>
  </Estanque>
  </elemento>
= <elemento nombre="Estructura_eleveda">
= <Estructura_eleveda>
= <atributos>

```

```

- <dominio_fijo>
- <Tipo_de_estructura_elevada>
  <a1>Caseta forestal : Para vigilancia y detección de incendios en
    bosques.</a1>
  <a2>Silo : Para el almacenamiento de granos.</a2>
  </Tipo_de_estructura_elevada>
  </dominio_fijo>
- <dominio_variable>
- <identificador_de_estructura_elevada>
  <w1>Un número secuencial que se incrementa con cada ocurrencia</w1>
- <dominio_de_valores>
  <w1>En el intervalo de 1 a N</w1>
  </dominio_de_valores>
  </identificador_de_estructura_elevada>
  </dominio_variable>
  </atributos>
- <restricciones_de_integridad>
- <COMBINACIONES_AUTORIZADAS_DE_VALORES_DE_ATRIBUTOS>
  <c1>Caseta forestal.</c1>
  <c2>Silo</c2>
  </COMBINACIONES_AUTORIZADAS_DE_VALORES_DE_ATRIBUTOS>
- <CALIFICADOR_AUTORIZADO_DE_REPRESENTACIONES_GEOMETRICAS>
  <a1>Aproximada.</a1>
  <a2>Definida.</a2>
  </CALIFICADOR_AUTORIZADO_DE_REPRESENTACIONES_GEOMETRICAS>
  </restricciones_de_integridad>
- <relaciones>
  <r1>Ninguna</r1>
  </relaciones>
  </Estructura_elevada>
  </elemento>
- <elemento nombre="Fango">
- <Fango>
- <atributos>
- <dominio_fijo>
  <a1>Ninguno</a1>

```

```

    </dominio_fijo>
= <dominio_variable>
= <identificador_de_fango>
    <w1>Un número secuencial que se incrementa con cada ocurrencia</w1>
= <dominio_de_valores>
    <w1>En el intervalo de 1 a N</w1>
    </dominio_de_valores>
    </identificador_de_fango>
    </dominio_variable>
    </atributos>
= <restricciones_de_integridad>
= <COMBINACIONES_AUTORIZADAS_DE_VALORES_DE_ATRIBUTOS>
    <c1>Ninguna.</c1>
    </COMBINACIONES_AUTORIZADAS_DE_VALORES_DE_ATRIBUTOS>
= <CALIFICADOR_AUTORIZADO_DE_REPRESENTACIONES_GEOMETRICAS>
    <a1>Definida.</a1>
    <a2>Virtual.</a2>
    </CALIFICADOR_AUTORIZADO_DE_REPRESENTACIONES_GEOMETRICAS>
    </restricciones_de_integridad>
= <relaciones>
    <r1>Ninguna</r1>
    </relaciones>
    </Fango>
    </elemento>
= <elemento nombre="Faro_Radiofaro_Vor">
= <Faro_Radiofaro_Vor>
= <atributos>
= <dominio_fijo>
= <Tipo_de_Faro_Radiofaro_Vor>
    <a1>Faro : Emisor de ondas luminosas para la orientación en la navegación
        marítima.</a1>
    <a2>Radiofaro : Emisor de ondas electromagnéticas para la orientación en la
        navegación marítima o aérea.</a2>
    <a3>Vor : Emisor de ondas electromagnéticas para la orientación en la
        navegación aérea.</a3>
    </Tipo_de_Faro_Radiofaro_Vor>
    </dominio_fijo>

```

```

= <dominio_variable>
= <identificador_de_Faro_Radiofaro_Vor>
  <w1>Un número secuencial que se incrementa con cada ocurrencia</w1>
= <dominio_de_valores>
  <w1>En el intervalo de 1 a N</w1>
  </dominio_de_valores>
  </identificador_de_Faro_Radiofaro_Vor>
  </dominio_variable>
  </atributos>
= <restricciones_de_integridad>
= <COMBINACIONES_AUTORIZADAS_DE_VALORES_DE_ATRIBUTOS>
  <c1>Faro</c1>
  <c2>Radiofaro.</c2>
  <c3>Vor.</c3>
  </COMBINACIONES_AUTORIZADAS_DE_VALORES_DE_ATRIBUTOS>
= <CALIFICADOR_AUTORIZADO_DE_REPRESENTACIONES_GEOMETRICAS>
  <a1>Aproximada.</a1>
  <a2>Definida.</a2>
  </CALIFICADOR_AUTORIZADO_DE_REPRESENTACIONES_GEOMETRICAS>
  </restricciones_de_integridad>
= <relaciones>
  <r1>Ninguna</r1>
  </relaciones>
  </Faro_Radiofaro_Vor>
  </elemento>
= <elemento nombre="Instalación_de_bombeo">
= <Instalación_de_bombeo>
= <atributos>
= <dominio_fijo>
  <a1>Ninguno</a1>
  </dominio_fijo>
= <dominio_variable>
= <identificador_de_Instalación_de_bombeo>
  <w1>Un número secuencial que se incrementa con cada ocurrencia</w1>
= <dominio_de_valores>
  <w1>En el intervalo de 1 a N</w1>

```

```

</dominio_de_valores>
</identificador_de_Instalación_de_bombeo>
</dominio_variable>
</atributos>
= <restricciones_de_integridad>
= <COMBINACIONES_AUTORIZADAS_DE_VALORES_DE_ATRIBUTOS>
  <c1>Ninguna.</c1>
  </COMBINACIONES_AUTORIZADAS_DE_VALORES_DE_ATRIBUTOS>
= <CALIFICADOR_AUTORIZADO_DE_REPRESENTACIONES_GEOMETRICAS>
  <a1>Definida.</a1>
  <a2>Virtual.</a2>
  </CALIFICADOR_AUTORIZADO_DE_REPRESENTACIONES_GEOMETRICAS>
  </restricciones_de_integridad>
= <relaciones>
  <r1>Conecta Acueducto (L)</r1>
  <r2>Conecta Canal (L)</r2>
  <r3>Conecta Conducto (L)</r3>
  <r4>Conecta Acueducto (L)</r4>
  <r5>Conecta Conducto (L)</r5>
  <r6>Comparte Canal (L)</r6>
  <r7>Comparte Canal (A)</r7>
  </relaciones>
  </Instalación_de_bombeo>
  </elemento>
= <elemento nombre="Instalacion_de_comunicacion">
= <Instalacion_de_comunicacion>
= <atributos>
= <dominio_fijo>
= <Tipo_de_Instalacion_de_comunicacion>
  <a1>Antena de microondas de telefonía : Estación transmisora y receptora de
  la Red de Telefonía.</a1>
  <a2>Antena de radio : Para transmisión o retransmisión de
  radiodifusoras.</a2>
  <a3>Antena de televisión : Para transmisión o retransmisión de empresas
  televisivas.</a3>
  <a4>Estación terrestre de telecomunicaciones : Estación transmisora y
  receptora de señales para telecomunicaciones.</a4>

```

**<a5>Repetidora de fibra óptica : Para transmisión y recepción de la red de telefonía.</a5>**

**<a6>Torre de microondas : Estación transmisora y receptora de la Red Nacional de Microondas.</a6>**

**<a7>Otro : El valor del atributo es diferente de los valores citados anteriormente.</a7>**

**</Tipo\_de\_Instalacion\_de\_comunicacion>**

**</dominio\_fijo>**

**= <dominio\_variable>**

**= <identificador\_de\_Instalacion\_de\_comunicacion>**

**<w1>Un número secuencial que se incrementa con cada ocurrencia</w1>**

**= <dominio\_de\_valores>**

**<w1>En el intervalo de 1 a N</w1>**

**</dominio\_de\_valores>**

**</identificador\_de\_Instalacion\_de\_comunicacion>**

**</dominio\_variable>**

**</atributos>**

**= <restricciones\_de\_integridad>**

**= <COMBINACIONES\_AUTORIZADAS\_DE\_VALORES\_DE\_ATRIBUTOS>**

**<c1>Antena de microondas de telefonía.</c1>**

**<c2>Antena de radio.</c2>**

**<c3>Antena de televisión.</c3>**

**<c4>Estación terrestre de telecomunicaciones.</c4>**

**<c5>Repetidora de fibra óptica.</c5>**

**<c6>Torre de microondas.</c6>**

**<c7>Otro.</c7>**

**</COMBINACIONES\_AUTORIZADAS\_DE\_VALORES\_DE\_ATRIBUTOS>**

**= <CALIFICADOR\_AUTORIZADO\_DE\_REPRESENTACIONES\_GEOMETRICAS>**

**<a1>Aproximada.</a1>**

**<a2>Definida</a2>**

**</CALIFICADOR\_AUTORIZADO\_DE\_REPRESENTACIONES\_GEOMETRICAS>**

**</restricciones\_de\_integridad>**

**= <relaciones>**

**<r1>Conecta Línea de comunicación (L)</r1>**

**</relaciones>**

**</Instalacion\_de\_comunicacion>**

**</elemento>**

```

= <elemento nombre="Instalacion_deportiva_o_recreativa">
= <Instalacion_deportiva_o_recreativa>
= <atributos>
= <dominio_fijo>
= <Tipo_de_Instalacion_deportiva_o_recreativa>
  <a1>Autódromo : Para carreras de automóviles.</a1>
  <a2>Balneario : Lugar de esparcimiento que cuenta con albercas.</a2>
  <a3>Campo de Golf : Para la práctica del golf.</a3>
  <a4>Campo de Tiro : Para la práctica de tiro.</a4>
  <a5>Diverso : Para la práctica de deportes diferentes.</a5>
  <a6>Estadio : Para la celebración de competencias deportivas.</a6>
  <a7>Galgódromo : Para carrera de galgos.</a7>
  <a8>Hipódromo : Para competencias de carrera de caballos.</a8>
  <a7>Lienzo charro : Para la práctica de la charrería.</a7>
  <a8>Plaza de toros : Para corridas de toros.</a8>
  <a9>Velódromo : Para carreras de bicicletas.</a9>
  <a10>Otro : El valor del atributo es diferente de los valores citados
    anteriormente.</a10>
  </Tipo_de_Instalacion_deportiva_o_recreativa>
  </dominio_fijo>
= <dominio_variable>
= <identificador_de_Instalacion_deportiva_o_recreativa>
  <w1>Un número secuencial que se incrementa con cada ocurrencia</w1>
= <dominio_de_valores>
  <w1>En el intervalo de 1 a N</w1>
  </dominio_de_valores>
  </identificador_de_Instalacion_deportiva_o_recreativa>
  </dominio_variable>
  </atributos>
= <restricciones_de_integridad>
= <COMBINACIONES_AUTORIZADAS_DE_VALORES_DE_ATRIBUTOS>
  <c1>Autódromo*.</c1>
  <c2>Balneario.</c2>
  <c3>Campo de Golf.</c3>
  <c4>Campo de tiro.</c4>
  <c5>Diverso.</c5>

```

<c6>Estadio.</c6>

<c7>Galgódromo \*.</c7>

<c8>Hipódromo \*.</c8>

<c9>Lienzo charro \*\*.</c9>

<c10>Plaza de toros \*\*.</c10>

<c11>Velódromo.</c11>

<c12>Otro.</c12>

</COMBINACIONES\_AUTORIZADAS\_DE\_VALORES\_DE\_ATRIBUTOS>

= <CALIFICADOR\_AUTORIZADO\_DE\_REPRESENTACIONES\_GEOMETRICAS>

<a1>Aproximada.</a1>

<a2>Definida.</a2>

<a3>Virtual</a3>

</CALIFICADOR\_AUTORIZADO\_DE\_REPRESENTACIONES\_GEOMETRICAS>

</restricciones\_de\_integridad>

= <relaciones>

<r1>Comparte Área urbana (A)</r1>

<r2>Comparte Calle (L)</r2>

<r3>Comparte Carretera (L)</r3>

<r4>Comparte Instalación deportiva o recreativa (A)</r4>

<r5>Comparte Vía férrea (L)</r5>

</relaciones>

</Instalacion\_deportiva\_o\_recreativa>

</elemento>

= <elemento nombre="Instalacion\_Diversa">

= <Instalacion\_Diversa>

= <atributos>

= <dominio\_fijo>

= <Tipo\_de\_Instalacion\_Diversa>

<a1>Almacenes de depósito : Para almacenar mercancías (sólo los de ANDSA o los rurales que sean significativos a la escala).</a1>

<a2>Central de autobuses : Para la administración y servicios propios del transporte de pasajeros por autobús</a2>

<a3>Centro de estudios superiores : Para la enseñanza superior.</a3>

<a4>Centro de investigación : Centros oficiales para la investigación.</a4>

<a5>Complejo médico : Para proporcionar servicios médicos (prevención, hospitalización, recuperación, etc.), en cualquier especialidad .</a5>

**<a6>Estación de ferrocarril : Para la administración y servicios propios del transporte de pasajeros y/o carga por tren.</a6>**

**<a7>Observatorio astronómico : Para la observación y estudio de los astros.</a7>**

**<a8>Planta desalinadora : Para la potabilización de agua salada.</a8>**

**<a9>Planta de tratamiento de agua : Para la potabilización o reciclaje de aguas fluviales o residuales.</a9>**

**<a10>Reclusorio : Para el confinamiento y la rehabilitación de convictos.</a10>**

**<a11>Otro : El valor del atributo es diferente de los valores citados anteriormente.</a11>**

**</Tipo\_de\_Instalacion\_Diversa>**

**</dominio\_fijo>**

**= <dominio\_variable>**

**= <identificador\_de\_Instalacion\_Diversa>**

**<w1>Un número secuencial que se incrementa con cada ocurrencia</w1>**

**= <dominio\_de\_valores>**

**<w1>En el intervalo de 1 a N</w1>**

**</dominio\_de\_valores>**

**</identificador\_de\_Instalacion\_Diversa>**

**</dominio\_variable>**

**</atributos>**

**= <restricciones\_de\_integridad>**

**= <COMBINACIONES\_AUTORIZADAS\_DE\_VALORES\_DE\_ATRIBUTOS>**

**<c1>Almacenes de depósito.</c1>**

**<c2>Central de autobuses.</c2>**

**<c3>Centro de estudios superiores.</c3>**

**<c4>Centro de investigación.</c4>**

**<c5>Complejo médico.</c5>**

**<c6>Estación de ferrocarril \*.</c6>**

**<c7>Observatorio astronómico.</c7>**

**<c8>Planta desalinadora.</c8>**

**<c9>Planta de tratamiento de agua.</c9>**

**<c10>Reclusorio.</c10>**

**<c11>Otro.</c11>**

**</COMBINACIONES\_AUTORIZADAS\_DE\_VALORES\_DE\_ATRIBUTOS>**

**= <CALIFICADOR\_AUTORIZADO\_DE\_REPRESENTACIONES\_GEOMETRICAS>**

<a1>Aproximada.</a1>

<a2>Definida.</a2>

<a3>Virtual.</a3>

</CALIFICADOR\_AUTORIZADO\_DE\_REPRESENTACIONES\_GEOMETRICAS>

</restricciones\_de\_integridad>

= <relaciones>

<r1>Conecta Vía férrea (L)</r1>

<r2>Conecta Acueducto (L)</r2>

<r3>Conecta Calle (L)</r3>

<r4>Conecta Canal (L)</r4>

<r5>Conecta Carretera (L)</r5>

<r6>Comparte Calle (L)</r6>

<r7>Comparte Carretera (L)</r7>

<r8>Comparte Vía férrea (L)</r8>

<r9>Comparte Área de cultivo (A)</r9>

<r10>Comparte Área urbana (A)</r10>

<r11>Comparte Área verde urbana (A)</r11>

<r12>Comparte Canal (A)</r12>

</relaciones>

</Instalacion\_Diversa>

</elemento>

= <elemento nombre="Instalación\_industrial">

= <Instalación\_industrial>

= <atributos>

= <dominio\_fijo>

= <Tipo\_de\_instalacion\_industrial>

<w1>Aserradero : Para la explotación y tratamiento de la madera.</w1>

<w2>Fundidora : Complejo industrial para la fundición de metales.</w2>

<w3>Planta automotriz : Para el armado de automóviles y/o camiones.</w3>

<w4>Planta azufrera : Para la obtención de compuestos de azufre.</w4>

<w5>Planta cementera : Para la producción de cemento.</w5>

<w6>Planta petroquímica : Para la obtención de derivados del petróleo diferentes de los hidrocarburos.</w6>

<w7>Refinería : Para la obtención de hidrocarburos.</w7>

<w8>Zona industrial : De propósito múltiple.</w8>

<w9>Otro : El valor del atributo es diferente de los valores citados anteriormente.</w9>

```

    </Tipo_de_instalacion_industrial>
    </dominio_fijo>
= <dominio_variable>
= <identificador_de_instalacion_industrial>
    <w1>Un número secuencial que se incrementa con cada ocurrencia</w1>
= <dominio_de_valores>
    <w1>En el intervalo de 1 a N</w1>
    </dominio_de_valores>
    </identificador_de_instalacion_industrial>
    </dominio_variable>
    </atributos>
= <restricciones_de_integridad>
= <COMBINACIONES_AUTORIZADAS_DE_VALORES_DE_ATRIBUTOS>
    <c1>Aserradero.</c1>
    <c2>Fundidora.</c2>
    <c3>Planta automotriz.</c3>
    <c4>Planta azufrera.</c4>
    <c5>Planta cementera.</c5>
    <c6>Planta petroquímica.</c6>
    <c7>Refinería.</c7>
    <c8>Zona industrial.</c8>
    <c9>Otro.</c9>
    </COMBINACIONES_AUTORIZADAS_DE_VALORES_DE_ATRIBUTOS>
= <CALIFICADOR_AUTORIZADO_DE_REPRESENTACIONES_GEOMETRICAS>
    <a1>Aproximada.</a1>
    <a2>Definida.</a2>
    <a3>Virtual.</a3>
    </CALIFICADOR_AUTORIZADO_DE_REPRESENTACIONES_GEOMETRICAS>
    </restricciones_de_integridad>
= <relaciones>
    <r1>Conecta Carretera (L)</r1>
    <r2>Conecta Vía férrea (L)</r2>
    <r3>Comparte Calle (L)</r3>
    <r4>Comparte Carretera (L)</r4>
    <r5>Comparte Vía férrea (L)</r5>
    <r6>Comparte Área de cultivo (A)</r6>

```

<r7>Comparte Área verde urbana (A)</r7>  
<r8>Comparte Instalación industrial (A)</r8>  
<r9>Comparte Instalación deportiva o recreativa (A)</r9>  
</relaciones>  
</Instalación\_industrial>  
</elemento>

= <elemento nombre="Instalacion\_portuaria">  
= <Instalacion\_portuaria>  
= <atributos>  
= <dominio\_fijo>  
= <Tipo\_de\_Instalacion\_portuaria>  
<w1>Malecón : Para salvaguardar de la crecida de las aguas marítimas o  
fluviales.</w1>  
<w2>Muelle o embarcadero : Para atracar las embarcaciones.</w2>  
<w3>Rompeolas o escollera : Para salvaguardar los puertos de el  
oleaje.</w3>  
</Tipo\_de\_Instalacion\_portuaria>  
</dominio\_fijo>  
= <dominio\_variable>  
= <identificador\_de\_Instalaciones\_portuarias>  
<w1>Un número secuencial que se incrementa con cada ocurrencia</w1>  
= <dominio\_de\_valores>  
<w1>En el intervalo de 1 a N</w1>  
</dominio\_de\_valores>  
</identificador\_de\_Instalaciones\_portuarias>  
</dominio\_variable>  
</atributos>  
= <restricciones\_de\_integridad>  
= <COMBINACIONES\_AUTORIZADAS\_DE\_VALORES\_DE\_ATRIBUTOS>  
<c1>Malecón</c1>  
<c2>Muelle o embarcadero.</c2>  
<c3>Rompeolas o escollera.</c3>  
</COMBINACIONES\_AUTORIZADAS\_DE\_VALORES\_DE\_ATRIBUTOS>  
= <CALIFICADOR\_AUTORIZADO\_DE\_REPRESENTACIONES\_GEOMETRICAS>  
<a1>Definida.</a1>  
<a2>Virtual.</a2>  
</CALIFICADOR\_AUTORIZADO\_DE\_REPRESENTACIONES\_GEOMETRICAS>

</restricciones\_de\_integridad>

= <relaciones>

<r1>Conecta Instalación portuarias (L)</r1>

<r2>Conecta Cuerpo de agua (A)</r2>

<r3>Conecta Instalación portuaria (A)</r3>

<r4>Comparte Carretera (L)</r4>

<r5>Comparte Cuerpo de agua (A)</r5>

<r6>Comparte Área urbana (A)</r6>

<r7>Conecta Instalación portuaria (L)</r7>

<r8>Comparte Calle (L)</r8>

<r9>Comparte Vía férrea (L)</r9>

<r10>Comparte Área urbana (A)</r10>

<r11>Comparte Cuerpo de agua (A)</r11>

</relaciones>

</Instalacion\_portuaria>

</elemento>

= <elemento nombre="Limite">

= <Limite>

= <atributos>

= <dominio\_fijo>

= <Tipo\_de\_Limite>

<w1>Internacional : Línea común que divide dos países.</w1>

<w2>Estatad : Línea común que divide dos estados.\*</w2>

<w3>Municipal : Línea común que divide dos municipios.\*</w3>

</Tipo\_de\_Limite>

</dominio\_fijo>

= <dominio\_variable>

= <identificador\_de\_Limite>

<w1>Un número secuencial que se incrementa con cada ocurrencia</w1>

= <dominio\_de\_valores>

<w1>En el intervalo de 1 a N</w1>

</dominio\_de\_valores>

</identificador\_de\_Limite>

</dominio\_variable>

</atributos>

= <restricciones\_de\_integridad>

```
= <COMBINACIONES_AUTORIZADAS_DE_VALORES_DE_ATRIBUTOS>
  <c1>Internacional.</c1>
  </COMBINACIONES_AUTORIZADAS_DE_VALORES_DE_ATRIBUTOS>
= <CALIFICADOR_AUTORIZADO_DE_REPRESENTACIONES_GEOMETRICAS>
  <a1>Aproximada **.</a1>
  <a2>Definida</a2>
  </CALIFICADOR_AUTORIZADO_DE_REPRESENTACIONES_GEOMETRICAS>
  </restricciones_de_integridad>
= <relaciones>
  <r1>Conecta Mojonera (P)</r1>
  <r2>Comparte Área de cultivo (A)</r2>
  <r3>Comparte Área urbana (A)</r3>
  <r4>Comparte Corriente de agua (L)</r4>
  <r5>Comparte Cuerpo de agua (A)</r5>
  <r6>Comparte Malpaís (A)</r6>
  <r7>Comparte Pantano (A)</r7>
  <r8>Comparte Vegetación densa (A)</r8>
  <r9>Comparte Zona arenosa (A)</r9>
  </relaciones>
  </Limite>
  </elemento>
= <elemento nombre="Lindero">
= <Lindero>
= <atributos>
= <dominio_fijo>
  <w1>Ninguno</w1>
  </dominio_fijo>
= <dominio_variable>
= <identificador_de_Lindero>
  <w1>Un número secuencial que se incrementa con cada ocurrencia</w1>
= <dominio_de_valores>
  <w1>En el intervalo de 1 a N</w1>
  </dominio_de_valores>
  </identificador_de_Lindero>
  </dominio_variable>
  </atributos>
```

```

= <restricciones_de_integridad>
= <COMBINACIONES_AUTORIZADAS_DE_VALORES_DE_ATRIBUTOS>
  <c1>Ninguna.</c1>
  </COMBINACIONES_AUTORIZADAS_DE_VALORES_DE_ATRIBUTOS>
= <CALIFICADOR_AUTORIZADO_DE_REPRESENTACIONES_GEOMETRICAS>
  <a1>Definida.</a1>
  </CALIFICADOR_AUTORIZADO_DE_REPRESENTACIONES_GEOMETRICAS>
  </restricciones_de_integridad>
= <relaciones>
  <r1>Ninguna</r1>
  </relaciones>
  </Lindero>
  </elemento>
= <elemento nombre="Linea_de_comunicacion">
= <Linea_de_comunicacion>
= <atributos>
= <dominio_fijo>
= <Tipo_de_Linea_de_comunicacion>
  <w1>Telefónica : Para la comunicación vía telefónica.</w1>
  <w2>Telegráfica : Para la comunicación vía telegráfica.</w2>
  </Tipo_de_Linea_de_comunicacion>
= <Relacion_con_el_suelo>
  <w1>Aéreo</w1>
  <w2>Subterráneo</w2>
  </Relacion_con_el_suelo>
= <Condicion_de_la_linea_de_comunicacion>
  <w1>En operación : Que está en servicio.</w1>
  <w2>Fuera de Uso : Que no está en uso (Carece totalmente de
  mantenimiento).</w2>
  </Condicion_de_la_linea_de_comunicacion>
  </dominio_fijo>
= <dominio_variable>
= <identificador_de_Linea_de_comunicacion>
  <w1>Un número secuencial que se incrementa con cada ocurrencia</w1>
= <dominio_de_valores>
  <w1>En el intervalo de 1 a N</w1>

```

```

    </dominio_de_valores>
    </identificador_de_Linea_de_comunicacion>
    </dominio_variable>
    </atributos>
= <restricciones_de_integridad>
= <COMBINACIONES_AUTORIZADAS_DE_VALORES_DE_ATRIBUTOS>
  <c1>Aéreo, telefónica, en operación.</c1>
  <c2>Subterráneo, telefónica, en operación.</c2>
  <c3>Aéreo, telegráfica, en operación.</c3>
  <c4>Aéreo, telefónica, fuera de uso.</c4>
  <c5>Subterráneo, telefónica, fuera de uso.</c5>
  <c6>Aéreo, telegráfica, fuera de uso.</c6>
  </COMBINACIONES_AUTORIZADAS_DE_VALORES_DE_ATRIBUTOS>
= <CALIFICADOR_AUTORIZADO_DE_REPRESENTACIONES_GEOMETRICAS>
  <a1>Aproximada</a1>
  </CALIFICADOR_AUTORIZADO_DE_REPRESENTACIONES_GEOMETRICAS>
  </restricciones_de_integridad>
= <relaciones>
  <r1>Conecta Instalación de comunicación</r1>
  <r2>Conecta Línea de comunicación (L)</r2>
  <r3>Comparte Línea de comunicación (L)</r3>
  <r4>Comparte Línea de transmisión (L)</r4>
  </relaciones>
  </Linea_de_comunicacion>
  </elemento>
= <elemento nombre="Linea_de_transmision">
= <Linea_de_transmision>
= <atributos>
= <dominio_fijo>
= <Número_de_alineamientos_de_soporte>
  <w1>Una línea de torres</w1>
  <w2>Dos líneas de torres</w2>
  <w3>Más de dos líneas de torres</w3>
  <w4>Una línea de postería sencilla</w4>
  <w5>Dos líneas de postería sencilla</w5>
  <w6>Más de dos líneas de postería sencilla</w6>

```

**<w7>Una línea de postería doble (H)</w7>**

**<w8>Dos líneas de postería doble (H)</w8>**

**<w9>Más de dos líneas de postería doble (H)</w9>**

**= <Restricciones\_del\_atributo>**

**<a1>Desconocido : El valor del atributo es alguno de los citados en el dominio, pero no se sabe actualmente cual de ellos es para alguna(s) Línea(s) de transmisión.</a1>**

**<a2>No aplicable (N/A) : Indica que este atributo no es válido, cuando el valor del atributo Condición de la Línea de Transmisión es Fuera de uso.</a2>**

**</Restricciones\_del\_atributo>**

**</Número\_de\_alineamientos\_de\_soporte>**

**= <Condicion\_de\_la\_linea\_de\_transmision>**

**<w1>En operación : Que está en servicio.</w1>**

**<w2>Fuera de Uso : Que no está en uso.</w2>**

**</Condicion\_de\_la\_linea\_de\_transmision>**

**</dominio\_fijo>**

**= <dominio\_variable>**

**= <identificador\_de\_Linea\_de\_transmision>**

**<w1>Un número secuencial que se incrementa con cada ocurrencia</w1>**

**= <dominio\_de\_valores>**

**<w1>En el intervalo de 1 a N</w1>**

**</dominio\_de\_valores>**

**</identificador\_de\_Linea\_de\_transmision>**

**</dominio\_variable>**

**</atributos>**

**= <restricciones\_de\_integridad>**

**= <COMBINACIONES\_AUTORIZADAS\_DE\_VALORES\_DE\_ATRIBUTOS>**

**<c1>Una línea en torres de acero, en operación.</c1>**

**<c2>Dos líneas en torres de acero, en operación.</c2>**

**<c3>Más de dos líneas en torres de acero, en operación.</c3>**

**<c4>Una línea en postería sencilla, en operación.</c4>**

**<c5>Dos líneas en postería sencilla, en operación.</c5>**

**<c6>Más de dos líneas en postería sencilla, en operación.</c6>**

**<c7>Una línea en postería doble (H), en operación.</c7>**

**<c8>Dos líneas en postería doble (H), en operación.</c8>**

**<c9>Más de dos líneas en postería doble (H), en operación.</c9>**

<c10>Desconocido, en operación.</c10>

<c11>N/A, fuera de uso.</c11>

</COMBINACIONES\_AUTORIZADAS\_DE\_VALORES\_DE\_ATRIBUTOS>

= <CALIFICADOR\_AUTORIZADO\_DE\_REPRESENTACIONES\_GEOMETRICAS>

<a1>Aproximada.</a1>

<a2>Definida.</a2>

</CALIFICADOR\_AUTORIZADO\_DE\_REPRESENTACIONES\_GEOMETRICAS>

</restricciones\_de\_integridad>

= <relaciones>

<r1>Conecta Planta generadora (P)</r1>

<r2>Conecta Subestación eléctrica (P)</r2>

<r3>Conecta Línea de transmisión (L)</r3>

<r4>Conecta Planta generadora (A)</r4>

<r5>Conecta Subestación eléctrica (A)</r5>

<r6>Comparte Línea de comunicación (L)</r6>

</relaciones>

</Linea\_de\_transmision>

</elemento>

= <elemento nombre="Lumbrera">

= <Lumbrera>

= <atributos>

= <dominio\_fijo>

<w1>Ninguno</w1>

</dominio\_fijo>

= <dominio\_variable>

= <identificador\_de\_Lumbrera>

<w1>Un número secuencial que se incrementa con cada ocurrencia</w1>

= <dominio\_de\_valores>

<w1>En el intervalo de 1 a N</w1>

</dominio\_de\_valores>

</identificador\_de\_Lumbrera>

= <Número\_de\_lumbrera>

<w1>Número asignado a cada lumbrera.</w1>

= <dominio\_de\_valores>

<w1>1 mayor o igual que valor menor o igual que 99</w1>

</dominio\_de\_valores>

```

- <restricciones_del_atributo>
  <w1>Desconocido : El valor del atributo es alguno de los citados en el
    dominio, pero no se sabe actualmente cual de ellos es para alguna(s)
    Lumbrera(s).</w1>
  </restricciones_del_atributo>
  </Número_de_lumbrera>
  </dominio_variable>
  </atributos>
- <restricciones_de_integridad>
- <COMBINACIONES_AUTORIZADAS_DE_VALORES_DE_ATRIBUTOS>
  <c1>Ninguna.</c1>
  </COMBINACIONES_AUTORIZADAS_DE_VALORES_DE_ATRIBUTOS>
- <CALIFICADOR_AUTORIZADO_DE_REPRESENTACIONES_GEOMETRICAS>
  <a1>Definida</a1>
  </CALIFICADOR_AUTORIZADO_DE_REPRESENTACIONES_GEOMETRICAS>
  </restricciones_de_integridad>
- <relaciones>
  <r1>Ninguna</r1>
  </relaciones>
  </Lumbrera>
  </elemento>
- <elemento nombre="Malpais">
- <Malpais>
- <atributos>
- <dominio_fijo>
  <w1>Ninguno</w1>
  </dominio_fijo>
- <dominio_variable>
- <identificador_de_Malpais>
  <w1>Un número secuencial que se incrementa con cada ocurrencia</w1>
- <dominio_de_valores>
  <w1>En el intervalo de 1 a N</w1>
  </dominio_de_valores>
  </identificador_de_Malpais>
  </dominio_variable>
  </atributos>
- <restricciones_de_integridad>

```

```
= <COMBINACIONES_AUTORIZADAS_DE_VALORES_DE_ATRIBUTOS>
  <c1>Ninguna.</c1>
  </COMBINACIONES_AUTORIZADAS_DE_VALORES_DE_ATRIBUTOS>
= <CALIFICADOR_AUTORIZADO_DE_REPRESENTACIONES_GEOMETRICAS>
  <a1>Definida.</a1>
  <a2>Virtual.</a2>
  </CALIFICADOR_AUTORIZADO_DE_REPRESENTACIONES_GEOMETRICAS>
  </restricciones_de_integridad>
= <relaciones>
  <r1>Comparte Cuerpo de agua (A)</r1>
  <r2>Comparte Terreno sujeto a inundación</r2>
  <r3>Comparte Vegetación densa (A)</r3>
  <r4>Comparte Zona arenosa (A)</r4>
  </relaciones>
  </Malpais>
  </elemento>
= <elemento nombre="Mamantial">
= <Mamantial>
= <atributos>
= <dominio_fijo>
  <w1>Ninguno</w1>
  </dominio_fijo>
= <dominio_variable>
= <identificador_de_Malpais>
  <w1>Un número secuencial que se incrementa con cada ocurrencia</w1>
= <dominio_de_valores>
  <w1>En el intervalo de 1 a N</w1>
  </dominio_de_valores>
  </identificador_de_Malpais>
  </dominio_variable>
  </atributos>
= <restricciones_de_integridad>
= <COMBINACIONES_AUTORIZADAS_DE_VALORES_DE_ATRIBUTOS>
  <c1>Ninguna.</c1>
  </COMBINACIONES_AUTORIZADAS_DE_VALORES_DE_ATRIBUTOS>
= <CALIFICADOR_AUTORIZADO_DE_REPRESENTACIONES_GEOMETRICAS>
```

```

<a1>Aproximada.</a1>
<a2>Definida.</a2>
  </CALIFICADOR_AUTORIZADO_DE_REPRESENTACIONES_GEOMETRICAS>
  </restricciones_de_integridad>
= <relaciones>
  <r1>Conecta Acueducto (L)</r1>
  <r2>Conecta Canal (L)</r2>
  <r3>Conecta Corriente de agua (L)</r3>
  </relaciones>
  </Mamantial>
  </elemento>
= <elemento nombre="Mina">
= <Mina>
= <atributos>
= <dominio_fijo>
= <Condicion_de_la_mina>
  <w1>En operación : Que está en servicio o puede usarse</w1>
  <w2>Fuera de uso : Que no está en uso</w2>
  </Condicion_de_la_mina>
  </dominio_fijo>
= <dominio_variable>
= <identificador_de_mina>
  <w1>Un número secuencial que se incrementa con cada ocurrencia</w1>
= <dominio_de_valores>
  <w1>En el intervalo de 1 a N</w1>
  </dominio_de_valores>
  </identificador_de_mina>
  </dominio_variable>
  </atributos>
= <restricciones_de_integridad>
= <COMBINACIONES_AUTORIZADAS_DE_VALORES_DE_ATRIBUTOS>
  <c1>En operación.</c1>
  <c2>Fuera de uso.</c2>
  </COMBINACIONES_AUTORIZADAS_DE_VALORES_DE_ATRIBUTOS>
= <CALIFICADOR_AUTORIZADO_DE_REPRESENTACIONES_GEOMETRICAS>
  <a1>Aproximada</a1>

```

```

    </CALIFICADOR_AUTORIZADO_DE_REPRESENTACIONES_GEOMETRICAS>
    </restricciones_de_integridad>
= <relaciones>
    <r1>Conecta Camino (L)</r1>
    <r2>Conecta Carretera (L)</r2>
    </relaciones>
    </Mina>
    </elemento>
= <elemento nombre="Mojonera">
= <Mojonera>
= <atributos>
= <dominio_fijo>
    <w1>Ninguno</w1>
    </dominio_fijo>
= <dominio_variable>
= <identificador_de_Mojonera>
    <w1>Un número secuencial que se incrementa con cada ocurrencia</w1>
= <dominio_de_valores>
    <w1>En el intervalo de 1 a N</w1>
    </dominio_de_valores>
    </identificador_de_Mojonera>
= <Número_de_Mojonera>
    <w1>Número asignado a cada mojonera.</w1>
= <dominio_de_valores>
    <w1>En el intervalo de 1 a N</w1>
    <w2>S/N . Valor para las ocurrencias que no tienen número.</w2>
    </dominio_de_valores>
    </Número_de_Mojonera>
    </dominio_variable>
    </atributos>
= <restricciones_de_integridad>
= <COMBINACIONES_AUTORIZADAS_DE_VALORES_DE_ATRIBUTOS>
    <c1>Ninguna</c1>
    </COMBINACIONES_AUTORIZADAS_DE_VALORES_DE_ATRIBUTOS>
= <CALIFICADOR_AUTORIZADO_DE_REPRESENTACIONES_GEOMETRICAS>
    <a1>Definida.</a1>

```

```

    </CALIFICADOR_AUTORIZADO_DE_REPRESENTACIONES_GEOMETRICAS>
    </restricciones_de_integridad>
= <relaciones>
  <r1>Conecta Límite (L)</r1>
  </relaciones>
  </Mojonera>
  </elemento>
= <elemento nombre="Muro_de_contencion">
= <Muro_de_contencion>
= <atributos>
= <dominio_fijo>
  <w1>Ninguno</w1>
  </dominio_fijo>
= <dominio_variable>
= <identificador_de_Muro_de_contencion>
  <w1>Un número secuencial que se incrementa con cada ocurrencia</w1>
= <dominio_de_valores>
  <w1>En el intervalo de 1 a N</w1>
  </dominio_de_valores>
  </identificador_de_Muro_de_contencion>
  </dominio_variable>
  </atributos>
= <restricciones_de_integridad>
= <COMBINACIONES_AUTORIZADAS_DE_VALORES_DE_ATRIBUTOS>
  <c1>Ninguna</c1>
  </COMBINACIONES_AUTORIZADAS_DE_VALORES_DE_ATRIBUTOS>
= <CALIFICADOR_AUTORIZADO_DE_REPRESENTACIONES_GEOMETRICAS>
  <a1>Definida</a1>
  </CALIFICADOR_AUTORIZADO_DE_REPRESENTACIONES_GEOMETRICAS>
  </restricciones_de_integridad>
= <relaciones>
  <r1>Ninguna</r1>
  </relaciones>
  </Muro_de_contencion>
  </elemento>
= <elemento nombre="Nieve_perpetua">

```

```
= <Nieve_perpetua>
= <atributos>
= <dominio_fijo>
  <w1>Ninguno</w1>
  </dominio_fijo>
= <dominio_variable>
= <identificador_de_Nieve_perpetua>
  <w1>Un número secuencial que se incrementa con cada ocurrencia</w1>
= <dominio_de_valores>
  <w1>En el intervalo de 1 a N</w1>
  </dominio_de_valores>
  </identificador_de_Nieve_perpetua>
  </dominio_variable>
  </atributos>
= <restricciones_de_integridad>
= <COMBINACIONES_AUTORIZADAS_DE_VALORES_DE_ATRIBUTOS>
  <c1>Ninguna.</c1>
  </COMBINACIONES_AUTORIZADAS_DE_VALORES_DE_ATRIBUTOS>
= <CALIFICADOR_AUTORIZADO_DE_REPRESENTACIONES_GEOMETRICAS>
  <a1>Definida.</a1>
  <a2>Virtual.</a2>
  </CALIFICADOR_AUTORIZADO_DE_REPRESENTACIONES_GEOMETRICAS>
  </restricciones_de_integridad>
= <relaciones>
  <r1>Ninguna</r1>
  </relaciones>
  </Nieve_perpetua>
  </elemento>
= <elemento nombre="Pantano">
= <Pantano>
= <atributos>
= <dominio_fijo>
  <w1>Ninguna</w1>
  </dominio_fijo>
= <dominio_variable>
= <identificador_de_Pantano>
```

<w1>Un número secuencial que se incrementa con cada ocurrencia</w1>

= <dominio\_de\_valores>

<w1>En el intervalo de 1 a N</w1>

</dominio\_de\_valores>

</identificador\_de\_Pantano>

</dominio\_variable>

</atributos>

= <restricciones\_de\_integridad>

= <COMBINACIONES\_AUTORIZADAS\_DE\_VALORES\_DE\_ATRIBUTOS>

<c1>Ninguna</c1>

</COMBINACIONES\_AUTORIZADAS\_DE\_VALORES\_DE\_ATRIBUTOS>

= <CALIFICADOR\_AUTORIZADO\_DE\_REPRESENTACIONES\_GEOMETRICAS>

<a1>Definida.</a1>

<a2>Virtual.</a2>

</CALIFICADOR\_AUTORIZADO\_DE\_REPRESENTACIONES\_GEOMETRICAS>

</restricciones\_de\_integridad>

= <relaciones>

<r1>Comparte Cuerpo de agua (A)</r1>

<r2>Comparte Límite (L)</r2>

<r3>Comparte Terreno sujeto a inundación</r3>

<r4>Comparte Vegetación densa (A)</r4>

</relaciones>

</Pantano>

</elemento>

= <elemento nombre="Pista\_de\_aviacion">

= <Pista\_de\_aviacion>

= <atributos>

= <dominio\_fijo>

= <Tipo\_de\_Pista>

<w1>Pavimentada : Tiene revestimiento de asfalto.</w1>

<w2>Terracería : Sin revestimiento alguno.</w2>

= <restricciones\_del\_atributo>

<w1>No aplicable (N/A): Indica que este atributo no es válido, cuando el valor del atributo Condición de la Pista es Fuera de uso o En construcción.</w1>

</restricciones\_del\_atributo>

</Tipo\_de\_Pista>

```

=> <Condicion_de_la_pista>
  <w1>En construcción : Que está en proceso de construcción.</w1>
  <w2>En operación : Que está en servicio o puede usarse.</w2>
  <w3>Fuera de uso : Que está abandonada.</w3>
  </Condicion_de_la_pista>
  </dominio_fijo>
=> <dominio_variable>
=> <identificador_de_Pista_de_aviacion>
  <w1>Un número secuencial que se incrementa con cada ocurrencia</w1>
=> <dominio_de_valores>
  <w1>En el intervalo de 1 a N</w1>
  </dominio_de_valores>
  </identificador_de_Pista_de_aviacion>
  </dominio_variable>
  </atributos>
=> <restricciones_de_integridad>
=> <COMBINACIONES_AUTORIZADAS_DE_VALORES_DE_ATRIBUTOS>
  <c1>Pavimentada, en operación.</c1>
  <c2>Terracería, en operación.</c2>
  <c3>N/A, en construcción.</c3>
  <c4>N/A, fuera de uso.</c4>
  </COMBINACIONES_AUTORIZADAS_DE_VALORES_DE_ATRIBUTOS>
=> <CALIFICADOR_AUTORIZADO_DE_REPRESENTACIONES_GEOMETRICAS>
  <a1>Definida.</a1>
  <a2>Virtual.</a2>
  </CALIFICADOR_AUTORIZADO_DE_REPRESENTACIONES_GEOMETRICAS>
  </restricciones_de_integridad>
=> <relaciones>
  <r1>Conecta Pista de aviación (L)</r1>
  </relaciones>
  </Pista_de_aviacion>
  </elemento>
=> <elemento nombre="Pista_de_carreras">
=> <Pista_de_carreras>
=> <atributos>
=> <dominio_fijo>

```

```

<w1>Ninguno</w1>
  </dominio_fijo>
= <dominio_variable>
= <identificador_de_Pista_de_carreras>
  <w1>Un número secuencial que se incrementa con cada ocurrencia</w1>
= <dominio_de_valores>
  <w1>En el intervalo de 1 a N</w1>
  </dominio_de_valores>
  </identificador_de_Pista_de_carreras>
  </dominio_variable>
  </atributos>
= <restricciones_de_integridad>
= <COMBINACIONES_AUTORIZADAS_DE_VALORES_DE_ATRIBUTOS>
  <c1>Ninguna</c1>
  </COMBINACIONES_AUTORIZADAS_DE_VALORES_DE_ATRIBUTOS>
= <CALIFICADOR_AUTORIZADO_DE_REPRESENTACIONES_GEOMETRICAS>
  <a1>Definida.</a1>
  </CALIFICADOR_AUTORIZADO_DE_REPRESENTACIONES_GEOMETRICAS>
  </restricciones_de_integridad>
= <relaciones>
  <r1>Conecta Pista de carreras (L)</r1>
  </relaciones>
  </Pista_de_carreras>
  </elemento>
= <elemento nombre="Planta_Generadora">
= <Planta_Generadora>
= <atributos>
= <dominio_fijo>
= <Tipo_de_Planta_Generadora>
  <w1>Eoloeléctrica : Por acción del viento.</w1>
  <w2>Geotérmica : Por acción del vapor de agua generado en el interior de la
    tierra.</w2>
  <w3>Hidroeléctrica : Por acción de la fuerza hidráulica.</w3>
  <w4>Nucleoeléctrica : Por acción del vapor de agua generado mediante el
    uso de energía nuclear.</w4>
  <w5>Termoeléctrica : Por acción del vapor de agua.</w5>

```

**<w6>Otro : El valor del atributo es diferente de los valores citados anteriormente.</w6>**

**</Tipo\_de\_Planta\_Generadora>**

**= <Condicion\_de\_la\_planta\_generadora>**

**<w1>En construcción : Que está en un proceso de construcción.</w1>**

**<w2>En operación : Que está en servicio o puede usarse.</w2>**

**</Condicion\_de\_la\_planta\_generadora>**

**</dominio\_fijo>**

**= <dominio\_variable>**

**= <identificador\_de\_Planta\_Generadora>**

**<w1>Un número secuencial que se incrementa con cada ocurrencia</w1>**

**= <dominio\_de\_valores>**

**<w1>En el intervalo de 1 a N</w1>**

**</dominio\_de\_valores>**

**</identificador\_de\_Planta\_Generadora>**

**</dominio\_variable>**

**</atributos>**

**= <restricciones\_de\_integridad>**

**= <COMBINACIONES\_AUTORIZADAS\_DE\_VALORES\_DE\_ATRIBUTOS>**

**<c1>Eoloeléctrica, en construcción.</c1>**

**<c2>Geotérmica, en construcción.</c2>**

**<c3>Hidroeléctrica, en construcción.</c3>**

**<c4>Nucleoeléctrica, en construcción.</c4>**

**<c5>Termoeléctrica, en construcción.</c5>**

**<c6>Otro, en construcción.</c6>**

**<c7>Eoloeléctrica, en operación.</c7>**

**<c8>Geotérmica, en operación.</c8>**

**<c9>Hidroeléctrica, en operación.</c9>**

**<c10>Nucleoeléctrica, en operación.</c10>**

**<c11>Termoeléctrica, en operación.</c11>**

**<c12>Otro, en operación.</c12>**

**</COMBINACIONES\_AUTORIZADAS\_DE\_VALORES\_DE\_ATRIBUTOS>**

**= <CALIFICADOR\_AUTORIZADO\_DE\_REPRESENTACIONES\_GEOMETRICAS>**

**<a1>Definida</a1>**

**<a2>Aproximada.</a2>**

**<a3>Virtual</a3>**

```

    </CALIFICADOR_AUTORIZADO_DE_REPRESENTACIONES_GEOMETRICAS>
    </restricciones_de_integridad>
= <relaciones>
    <r1>Conecta Corriente de agua (L)</r1>
    <r2>Conecta Línea de transmisión</r2>
    <r3>Conecta Línea de transmisión</r3>
    <r4>Comparte Presa (L)</r4>
    </relaciones>
    </Planta_Generadora>
    </elemento>
= <elemento nombre="Pozo_de_explotacion">
= <Pozo_de_explotacion>
= <atributos>
= <dominio_fijo>
= <Tipo_de_Pozo_de_explotacion>
    <w1>Petróleo</w1>
    <w2>Gas</w2>
    </Tipo_de_Pozo_de_explotacion>
    </dominio_fijo>
= <dominio_variable>
= <identificador_de_Pozo_de_explotacion>
    <w1>Un número secuencial que se incrementa con cada ocurrencia</w1>
= <dominio_de_valores>
    <w1>En el intervalo de 1 a N</w1>
    </dominio_de_valores>
    </identificador_de_Pozo_de_explotacion>
    </dominio_variable>
    </atributos>
= <restricciones_de_integridad>
= <COMBINACIONES_AUTORIZADAS_DE_VALORES_DE_ATRIBUTOS>
    <c1>Petróleo.</c1>
    <c2>Gas.</c2>
    </COMBINACIONES_AUTORIZADAS_DE_VALORES_DE_ATRIBUTOS>
= <CALIFICADOR_AUTORIZADO_DE_REPRESENTACIONES_GEOMETRICAS>
    <a1>Definida.</a1>
    </CALIFICADOR_AUTORIZADO_DE_REPRESENTACIONES_GEOMETRICAS>

```

```

    </restricciones_de_integridad>
= <relaciones>
  <r1>Conecta Camino (L)</r1>
  <r2>Conecta Carretera (L)</r2>
  <r3>Conecta Conducto (L)</r3>
  </relaciones>
  </Pozo_de_explotacion>
  </elemento>
= <elemento nombre="Presa">
= <Presa>
= <atributos>
= <dominio_fijo>
= <Condicion_de_la_Presa>
  <w1>En construcción : Que está en un proceso de construcción.</w1>
  <w2>En operación : Que está en servicio o puede usarse.</w2>
  <w3>Fuera de uso: Que está abandonada.</w3>
  </Condicion_de_la_Presa>
  </dominio_fijo>
= <dominio_variable>
= <identificador_de_>
  <w1>Un número secuencial que se incrementa con cada ocurrencia</w1>
= <dominio_de_valores>
  <w1>En el intervalo de 1 a N</w1>
  </dominio_de_valores>
  </identificador_de_>
  </dominio_variable>
  </atributos>
= <restricciones_de_integridad>
= <COMBINACIONES_AUTORIZADAS_DE_VALORES_DE_ATRIBUTOS>
  <c1>En construcción.</c1>
  <c2>En operación.</c2>
  <c3>Fuera de uso.</c3>
  </COMBINACIONES_AUTORIZADAS_DE_VALORES_DE_ATRIBUTOS>
= <CALIFICADOR_AUTORIZADO_DE_REPRESENTACIONES_GEOMETRICAS>
  <a1>Definida.</a1>
  </CALIFICADOR_AUTORIZADO_DE_REPRESENTACIONES_GEOMETRICAS>

```

</restricciones\_de\_integridad>

= <relaciones>

<r1>Conecta Camino (L)</r1>

<r2>Conecta Corriente de agua (L)</r2>

<r3>Comparte Camino (L)</r3>

<r4>Comparte Carretera (L)</r4>

<r5>Comparte Cuerpo de agua (A)</r5>

<r6>Comparte Planta generadora (A)</r6>

</relaciones>

</Presa>

</elemento>

= <elemento nombre="Puente">

= <Puente>

= <atributos>

= <dominio\_fijo>

= <Tipo\_de\_Puente>

<w1>Colgante : Suspendido por cables de acero, generalmente para uso de personas y animales.</w1>

<w2>Levadizo : Que tiene una plataforma movable.</w2>

<w3>Para canal : Usado para soportar un cauce artificial.</w3>

<w4>Peatonal : Para uso de peatones.</w4>

<w5>Otro : El valor del atributo es diferente de los valores citados anteriormente.</w5>

= <restricciones\_del\_atributo>

<w1>No aplicable (N/A): Indica que este atributo no es válido, cuando el valor del atributo Condición del Puente es Fuera de uso o En construcción</w1>

</restricciones\_del\_atributo>

</Tipo\_de\_Puente>

= <Condicion\_del\_Puente>

<w1>En construcción : Que está en un proceso de construcción.</w1>

<w2>En operación : Que está en servicio o puede usarse.</w2>

<w3>Fuera de uso : Que está fuera de uso.</w3>

</Condicion\_del\_Puente>

</dominio\_fijo>

= <dominio\_variable>

= <identificador\_de\_Puente>

<w1>Un número secuencial que se incrementa con cada ocurrencia</w1>

= <dominio\_de\_valores>

<w1>En el intervalo de 1 a N</w1>

</dominio\_de\_valores>

</identificador\_de\_Puente>

</dominio\_variable>

</atributos>

= <restricciones\_de\_integridad>

= <COMBINACIONES\_AUTORIZADAS\_DE\_VALORES\_DE\_ATRIBUTOS>

<c1>Colgante, en operación.</c1>

<c2>Levadizo, en operación.</c2>

<c3>Para canal, en operación.</c3>

<c4>Peatonal, en operación.</c4>

<c5>Otro, en operación.</c5>

<c6>N/A, en construcción.</c6>

<c7>N/A, fuera de uso.</c7>

</COMBINACIONES\_AUTORIZADAS\_DE\_VALORES\_DE\_ATRIBUTOS>

= <CALIFICADOR\_AUTORIZADO\_DE\_REPRESENTACIONES\_GEOMETRICAS>

<a1>Definida.</a1>

</CALIFICADOR\_AUTORIZADO\_DE\_REPRESENTACIONES\_GEOMETRICAS>

</restricciones\_de\_integridad>

= <relaciones>

<r1>Comparte Calle (L)</r1>

<r2>Comparte Camino (L)</r2>

<r3>Comparte Canal (L)</r3>

<r4>Comparte Carretera (L)</r4>

<r5>Comparte Vía férrea (L)</r5>

</relaciones>

</Puente>

</elemento>

= <elemento nombre="Punto\_acotado">

= <Punto\_acotado>

= <atributos>

= <dominio\_fijo>

<w1>Ninguno</w1>

</dominio\_fijo>

= <dominio\_variable>  
= <identificador\_de\_Punto\_acotado>  
  <w1>**Un número secuencial que se incrementa con cada ocurrencia**</w1>  
= <dominio\_de\_valores>  
  <w1>**En el intervalo de 1 a N**</w1>  
  </dominio\_de\_valores>  
  </identificador\_de\_Punto\_acotado>  
= <elevacion\_de\_Punto\_acotado>  
  <w1>**Valor en metros de la elevación del punto.**</w1>  
= <dominio\_de\_valores>  
  <w1>**-20 mayor o igual que valor menor o igual que 5750**</w1>  
  </dominio\_de\_valores>  
  </elevacion\_de\_Punto\_acotado>  
  </dominio\_variable>  
  </atributos>  
= <restricciones\_de\_integridad>  
= <COMBINACIONES\_AUTORIZADAS\_DE\_VALORES\_DE\_ATRIBUTOS>  
  <c1>**Ninguna.**</c1>  
  </COMBINACIONES\_AUTORIZADAS\_DE\_VALORES\_DE\_ATRIBUTOS>  
= <CALIFICADOR\_AUTORIZADO\_DE\_REPRESENTACIONES\_GEOMETRICAS>  
  <a1>**Definida.**</a1>  
  </CALIFICADOR\_AUTORIZADO\_DE\_REPRESENTACIONES\_GEOMETRICAS>  
  </restricciones\_de\_integridad>  
= <relaciones>  
  <r1>**Ninguna**</r1>  
  </relaciones>  
  </Punto\_acotado>  
  </elemento>  
= <elemento nombre="Rapido">  
= <Rapido>  
= <atributos>  
= <dominio\_fijo>  
  <w1>**Ninguno**</w1>  
  </dominio\_fijo>  
= <dominio\_variable>  
= <identificador\_de\_Rapido>

<w1>Un número secuencial que se incrementa con cada ocurrencia</w1>

= <dominio\_de\_valores>

<w1>En el intervalo de 1 a N</w1>

</dominio\_de\_valores>

</identificador\_de\_Rapido>

</dominio\_variable>

</atributos>

= <restricciones\_de\_integridad>

= <COMBINACIONES\_AUTORIZADAS\_DE\_VALORES\_DE\_ATRIBUTOS>

<c1>Ninguna.</c1>

</COMBINACIONES\_AUTORIZADAS\_DE\_VALORES\_DE\_ATRIBUTOS>

= <CALIFICADOR\_AUTORIZADO\_DE\_REPRESENTACIONES\_GEOMETRICAS>

<a1>Aproximada</a1>

</CALIFICADOR\_AUTORIZADO\_DE\_REPRESENTACIONES\_GEOMETRICAS>

</restricciones\_de\_integridad>

= <relaciones>

<r1>Conecta Corriente de agua (L)</r1>

<r2>Conecta Cuerpo de agua (A)</r2>

</relaciones>

</Rapido>

</elemento>

= <elemento nombre="Rasgo\_arqueologico">

= <Rasgo\_arqueologico>

= <atributos>

= <dominio\_fijo>

= <Tipo\_de\_Rasgo\_arqueologico>

<w1>Pinturas rupestres \* : Donde existen pinturas sobre rocas, realizadas por culturas autóctonas.</w1>

<w2>Sitio arqueológico : Donde se localizan vestigios culturales prehispanicos.</w2>

<w3>Zona arqueológica : Donde existen edificaciones y diversos elementos prehispanicos.</w3>

</Tipo\_de\_Rasgo\_arqueologico>

</dominio\_fijo>

= <dominio\_variable>

= <identificador\_de\_Rasgo\_arqueologico>

<w1>Un número secuencial que se incrementa con cada ocurrencia</w1>

```

- <dominio_de_valores>
  <w1>En el intervalo de 1 a N</w1>
  </dominio_de_valores>
  </identificador_de_Rasgo_arqueologico>
  </dominio_variable>
  </atributos>
- <restricciones_de_integridad>
- <COMBINACIONES_AUTORIZADAS_DE_VALORES_DE_ATRIBUTOS>
  <c1>Pinturas rupestres.</c1>
  <c2>Sitio arqueológico.</c2>
  <c3>Zona arqueológica.</c3>
  </COMBINACIONES_AUTORIZADAS_DE_VALORES_DE_ATRIBUTOS>
- <CALIFICADOR_AUTORIZADO_DE_REPRESENTACIONES_GEOMETRICAS>
  <a1>Definida.</a1>
  <a2>Aproximada **.</a2>
  <a3>Virtual.</a3>
  </CALIFICADOR_AUTORIZADO_DE_REPRESENTACIONES_GEOMETRICAS>
  </restricciones_de_integridad>
- <relaciones>
  <r1>Conecta Camino (L)</r1>
  <r2>Conecta Carretera (L)</r2>
  <r3>Conecta Camino (L)</r3>
  <r4>Conecta Carretera (L)</r4>
  <r5>Comparte Camino (L)</r5>
  <r6>Comparte Área de cultivo (A)</r6>
  <r7>Comparte Área urbana (A)</r7>
  <r8>Comparte Carretera (L)</r8>
  <r9>Comparte Vegetación densa (A)</r9>
  </relaciones>
  </Rasgo_arqueologico>
  </elemento>
- <elemento nombre="Roca">
- <Roca>
- <atributos>
- <dominio_fijo>
  <w1>Ninguno</w1>

```

```

    </dominio_fijo>
= <dominio_variable>
= <identificador_de_Roca>
    <w1>Un número secuencial que se incrementa con cada ocurrencia</w1>
= <dominio_de_valores>
    <w1>En el intervalo de 1 a N</w1>
    </dominio_de_valores>
    </identificador_de_Roca>
    </dominio_variable>
    </atributos>
= <restricciones_de_integridad>
= <COMBINACIONES_AUTORIZADAS_DE_VALORES_DE_ATRIBUTOS>
    <c1>Ninguna.</c1>
    </COMBINACIONES_AUTORIZADAS_DE_VALORES_DE_ATRIBUTOS>
= <CALIFICADOR_AUTORIZADO_DE_REPRESENTACIONES_GEOMETRICAS>
    <a1>Aproximada.</a1>
    <a2>Definida.</a2>
    </CALIFICADOR_AUTORIZADO_DE_REPRESENTACIONES_GEOMETRICAS>
    </restricciones_de_integridad>
= <relaciones>
    <r1>Ninguna</r1>
    </relaciones>
    </Roca>
    </elemento>
= <elemento nombre="Ruta_de_embarcacion">
= <Ruta_de_embarcacion>
= <atributos>
= <dominio_fijo>
= <Tipo_de_Ruta_de_embarcacion>
    <w1>Se indica que tipo de embarcaciones transitan por la ruta.</w1>
= <dominio_de_valores>
    <w1>Chalana : Embarcación de fondo plano que se utiliza para el transporte
    de pasajeros y automóviles de una orilla a otra de un río.</w1>
    <w2>Panga : Embarcación de fondo plano que se utiliza para el transporte de
    pasajeros y automóviles de una orilla a otra de un río, ésta embarcación es
    de dimensiones menores que la chalana.</w2>

```

<w3>Transbordador : Barco acondicionado para transportar automóviles,  
 vagones, pasajeros, etc. de un puerto a otro.</w3>  
 </dominio\_de\_valores>  
 </Tipo\_de\_Ruta\_de\_embarcacion>  
 </dominio\_fijo>  
 = <dominio\_variable>  
 = <identificador\_de\_Ruta\_de\_embarcacion>  
 <w1>Un número secuencial que se incrementa con cada ocurrencia</w1>  
 = <dominio\_de\_valores>  
 <w1>En el intervalo de 1 a N</w1>  
 </dominio\_de\_valores>  
 </identificador\_de\_Ruta\_de\_embarcacion>  
 = <ruta\_que\_cubre>  
 <w1>Especifica los puntos que enlaza la ruta de embarcación de tipo  
 transbordador.</w1>  
 = <dominio\_de\_valores>  
 <w1>Alfabeto, signos de puntuación y números.</w1>  
 </dominio\_de\_valores>  
 </ruta\_que\_cubre>  
 </dominio\_variable>  
 </atributos>  
 = <restricciones\_de\_integridad>  
 = <COMBINACIONES\_AUTORIZADAS\_DE\_VALORES\_DE\_ATRIBUTOS>  
 <c1>Chalana.</c1>  
 <c2>Panga.</c2>  
 <c3>Transbordador.</c3>  
 </COMBINACIONES\_AUTORIZADAS\_DE\_VALORES\_DE\_ATRIBUTOS>  
 = <CALIFICADOR\_AUTORIZADO\_DE\_REPRESENTACIONES\_GEOMETRICAS>  
 <a1>Definida \*\*.</a1>  
 <a2>Aproximada \*\*\*.</a2>  
 </CALIFICADOR\_AUTORIZADO\_DE\_REPRESENTACIONES\_GEOMETRICAS>  
 </restricciones\_de\_integridad>  
 = <relaciones>  
 <r1>Conecta Camino (L)</r1>  
 <r2>Conecta Carretera (L)</r2>  
 <r3>Conecta Corriente de agua (L)</r3>  
 <r4>Conecta Camino (L)</r4>

```

<r5>Conecta Carretera (L)</r5>
  </relaciones>
  </Ruta_de_embarcacion>
</elemento>
= <elemento nombre="Ruta_de_funicular_teleferico">
= <Ruta_de_funicular_teleferico>
= <atributos>
= <dominio_fijo>
= <Tipo_de_Ruta>
  <w1>Aérea : La empleada por el teleférico.</w1>
  <w2>Terrestre : La empleada por el funicular.</w2>
  </Tipo_de_Ruta>
  </dominio_fijo>
= <dominio_variable>
= <identificador_de_Ruta_de_funicular_teleferico>
  <w1>Un número secuencial que se incrementa con cada ocurrencia</w1>
= <dominio_de_valores>
  <w1>En el intervalo de 1 a N</w1>
  </dominio_de_valores>
  </identificador_de_Ruta_de_funicular_teleferico>
  </dominio_variable>
  </atributos>
= <restricciones_de_integridad>
= <COMBINACIONES_AUTORIZADAS_DE_VALORES_DE_ATRIBUTOS>
  <c1>Aérea.</c1>
  <c2>Terrestre.</c2>
  </COMBINACIONES_AUTORIZADAS_DE_VALORES_DE_ATRIBUTOS>
= <CALIFICADOR_AUTORIZADO_DE_REPRESENTACIONES_GEOMETRICAS>
  <a1>Definida.</a1>
  </CALIFICADOR_AUTORIZADO_DE_REPRESENTACIONES_GEOMETRICAS>
  </restricciones_de_integridad>
= <relaciones>
  <r1>Ninguna</r1>
  </relaciones>
  </Ruta_de_funicular_teleferico>
</elemento>

```

```

- <elemento nombre="Salina">
- <Salina>
- <atributos>
- <dominio_fijo>
- <Tipo_de_Salina>
  <w1>Artificial : Construida o acondicionada expreso para la explotación de
    sal.</w1>
  <w2>Natural : Donde existe un yacimiento de sal susceptible de ser
    explotado.</w2>
  </Tipo_de_Salina>
  </dominio_fijo>
- <dominio_variable>
- <identificador_de_Salina>
  <w1>Un número secuencial que se incrementa con cada ocurrencia</w1>
- <dominio_de_valores>
  <w1>En el intervalo de 1 a N</w1>
  </dominio_de_valores>
  </identificador_de_Salina>
  </dominio_variable>
  </atributos>
- <restricciones_de_integridad>
- <COMBINACIONES_AUTORIZADAS_DE_VALORES_DE_ATRIBUTOS>
  <c1>Artificial.</c1>
  <c2>Natural.</c2>
  </COMBINACIONES_AUTORIZADAS_DE_VALORES_DE_ATRIBUTOS>
- <CALIFICADOR_AUTORIZADO_DE_REPRESENTACIONES_GEOMETRICAS>
  <a1>Definida.</a1>
  <a2>Virtual</a2>
  </CALIFICADOR_AUTORIZADO_DE_REPRESENTACIONES_GEOMETRICAS>
  </restricciones_de_integridad>
- <relaciones>
  <r1>Conecta Separador (L) </r1>
  <r2>Comparte Cuerpo de agua (A) </r2>
  <r3>Comparte Terreno sujeto a inundación</r3>
  <r4>Comparte Zona arenosa (A) </r4>
  </relaciones>
</Salina>

```

```

    </elemento>
= <elemento nombre="Salto_de_agua">
= <Salto_de_agua>
= <atributos>
= <dominio_fijo>
  <w1>Ninguno</w1>
  </dominio_fijo>
= <dominio_variable>
= <identificador_de_Salto_de_agua>
  <w1>Un número secuencial que se incrementa con cada ocurrencia</w1>
= <dominio_de_valores>
  <w1>En el intervalo de 1 a N</w1>
  </dominio_de_valores>
  </identificador_de_Salto_de_agua>
  </dominio_variable>
  </atributos>
= <restricciones_de_integridad>
= <COMBINACIONES_AUTORIZADAS_DE_VALORES_DE_ATRIBUTOS>
  <c1>Ninguna</c1>
  </COMBINACIONES_AUTORIZADAS_DE_VALORES_DE_ATRIBUTOS>
= <CALIFICADOR_AUTORIZADO_DE_REPRESENTACIONES_GEOMETRICAS>
  <a1>Definida</a1>
  </CALIFICADOR_AUTORIZADO_DE_REPRESENTACIONES_GEOMETRICAS>
  </restricciones_de_integridad>
= <relaciones>
  <r1>Conecta Corriente de agua (L)</r1>
  <r2>Conecta Cuerpo de agua (A)</r2>
  </relaciones>
  </Salto_de_agua>
  </elemento>
= <elemento nombre="Separador">
= <Separador>
= <atributos>
= <dominio_fijo>
= <Tipo_de_Separador>
  <w1>Estanque acuícola : División dentro de un estanque acuícola.</w1>

```

```

<w2>Estanque de sedimentación : División dentro de un estanque de
sedimentación.</w2>
<w3>Salina artificial : División dentro de una salina artificial.</w3>
  </Tipo_de_Separador>
  </dominio_fijo>
= <dominio_variable>
= <identificador_de_Separador>
  <w1>Un número secuencial que se incrementa con cada ocurrencia</w1>
= <dominio_de_valores>
  <w1>En el intervalo de 1 a N</w1>
  </dominio_de_valores>
  </identificador_de_Separador>
  </dominio_variable>
  </atributos>
= <restricciones_de_integridad>
= <COMBINACIONES_AUTORIZADAS_DE_VALORES_DE_ATRIBUTOS>
  <c1>Estanque acuícola.</c1>
  <c2>Estanque de sedimentación.</c2>
  <c3>Salina artificial.</c3>
  </COMBINACIONES_AUTORIZADAS_DE_VALORES_DE_ATRIBUTOS>
= <CALIFICADOR_AUTORIZADO_DE_REPRESENTACIONES_GEOMETRICAS>
  <a1>Definida</a1>
  </CALIFICADOR_AUTORIZADO_DE_REPRESENTACIONES_GEOMETRICAS>
  </restricciones_de_integridad>
= <relaciones>
  <r1>Conecta Estanque (A)</r1>
  <r2>Conecta Salina (A)</r2>
  </relaciones>
  </Separador>
  </elemento>
= <elemento nombre="Subestacion_electrica">
= <Subestacion_electrica>
= <atributos>
= <dominio_fijo>
= <Condición_de_la_Subestacion_electrica>
  <w1>En construcción : Que está en un proceso de construcción.</w1>

```

<w2>En operación : Que está en servicio o puede usarse.</w2>

<w3>Fuera de uso : Que no está en uso.</w3>

</Condición\_de\_la\_Subestacion\_electrica>

</dominio\_fijo>

= <dominio\_variable>

= <identificador\_de\_Subestacion\_electrica>

<w1>Un número secuencial que se incrementa con cada ocurrencia</w1>

= <dominio\_de\_valores>

<w1>En el intervalo de 1 a N</w1>

</dominio\_de\_valores>

</identificador\_de\_Subestacion\_electrica>

</dominio\_variable>

</atributos>

= <restricciones\_de\_integridad>

= <COMBINACIONES\_AUTORIZADAS\_DE\_VALORES\_DE\_ATRIBUTOS>

<c1>En construcción.</c1>

<c2>En operación.</c2>

<c3>Fuera de uso.</c3>

</COMBINACIONES\_AUTORIZADAS\_DE\_VALORES\_DE\_ATRIBUTOS>

= <CALIFICADOR\_AUTORIZADO\_DE\_REPRESENTACIONES\_GEOMETRICAS>

<a1>Definida.</a1>

<a2>Aproximada \*.</a2>

<a3>Virtual.</a3>

</CALIFICADOR\_AUTORIZADO\_DE\_REPRESENTACIONES\_GEOMETRICAS>

</restricciones\_de\_integridad>

= <relaciones>

<r1>Conecta Línea de transmisión (L)</r1>

<r2>Conecta Línea de transmisión (L)</r2>

</relaciones>

</Subestacion\_electrica>

</elemento>

= <elemento nombre="Tanque">

= <Tanque>

= <atributos>

= <dominio\_fijo>

= <Contenido\_del\_Tanque>

```

<w1>Combustible : Para petróleo y sus derivados.</w1>
<w2>Otro : El valor del atributo es diferente de los valores citados
    anteriormente.</w2>
</Contenido_del_Tanque>
</dominio_fijo>
= <dominio_variable>
= <identificador_de_Tanque>
    <w1>Un número secuencial que se incrementa con cada ocurrencia</w1>
= <dominio_de_valores>
    <w1>En el intervalo de 1 a N</w1>
    </dominio_de_valores>
    </identificador_de_Tanque>
    </dominio_variable>
    </atributos>
= <restricciones_de_integridad>
= <COMBINACIONES_AUTORIZADAS_DE_VALORES_DE_ATRIBUTOS>
    <c1>Combustible.</c1>
    <c2>Otro.</c2>
    </COMBINACIONES_AUTORIZADAS_DE_VALORES_DE_ATRIBUTOS>
= <CALIFICADOR_AUTORIZADO_DE_REPRESENTACIONES_GEOMETRICAS>
    <a1>Definida.</a1>
    <a2>Virtual.</a2>
    </CALIFICADOR_AUTORIZADO_DE_REPRESENTACIONES_GEOMETRICAS>
    </restricciones_de_integridad>
= <relaciones>
    <r1>Ninguna</r1>
    </relaciones>
    </Tanque>
    </elemento>
= <elemento nombre="Tanque_de_agua">
= <Tanque_de_agua>
= <atributos>
= <dominio_fijo>
= <Tipo_de_Tanque_de_agua>
    <w1>Caja de agua : Ubicado a nivel del suelo y empleado para el
        abastecimiento de una localidad y/o en una red de distribución para el
        almacenamiento y regulación.</w1>

```

```

<w2>Tanque elevado ** : Empleado para el abastecimiento de una localidad,
soportado por una estructura.</w2>
</Tipo_de_Tanque_de_agua>
</dominio_fijo>
= <dominio_variable>
= <identificador_de_Tanque_de_agua>
<w1>Un número secuencial que se incrementa con cada ocurrencia</w1>
= <dominio_de_valores>
<w1>En el intervalo de 1 a N</w1>
</dominio_de_valores>
</identificador_de_Tanque_de_agua>
</dominio_variable>
</atributos>
= <restricciones_de_integridad>
= <COMBINACIONES_AUTORIZADAS_DE_VALORES_DE_ATRIBUTOS>
<c1>Caja de agua.</c1>
<c2>Tanque elevado.</c2>
</COMBINACIONES_AUTORIZADAS_DE_VALORES_DE_ATRIBUTOS>
= <CALIFICADOR_AUTORIZADO_DE_REPRESENTACIONES_GEOMETRICAS>
<a1>Definida.</a1>
<a2>Virtual.</a2>
</CALIFICADOR_AUTORIZADO_DE_REPRESENTACIONES_GEOMETRICAS>
</restricciones_de_integridad>
= <relaciones>
<r1>Conecta Acueducto (L)</r1>
<r2>Conecta Canal (L)</r2>
<r3>Conecta Acueducto (L)</r3>
<r4>Conecta Canal (L)</r4>
<r5>Comparte Canal (A)</r5>
</relaciones>
</Tanque_de_agua>
</elemento>
= <elemento nombre="Terreno_sujeto_a_inundacion">
= <Terreno_sujeto_a_inundacion>
= <atributos>
= <dominio_fijo>

```

```

<w1>Ninguno</w1>
  </dominio_fijo>
= <dominio_variable>
= <identificador_de_Terreno_sujeto_a_inundacion>
  <w1>Un número secuencial que se incrementa con cada ocurrencia</w1>
= <dominio_de_valores>
  <w1>En el intervalo de 1 a N</w1>
  </dominio_de_valores>
  </identificador_de_Terreno_sujeto_a_inundacion>
  </dominio_variable>
  </atributos>
= <restricciones_de_integridad>
= <COMBINACIONES_AUTORIZADAS_DE_VALORES_DE_ATRIBUTOS>
  <c1>Ninguna.</c1>
  </COMBINACIONES_AUTORIZADAS_DE_VALORES_DE_ATRIBUTOS>
= <CALIFICADOR_AUTORIZADO_DE_REPRESENTACIONES_GEOMETRICAS>
  <a1>Definida.</a1>
  <a2>Virtual.</a2>
  </CALIFICADOR_AUTORIZADO_DE_REPRESENTACIONES_GEOMETRICAS>
  </restricciones_de_integridad>
= <relaciones>
  <r1>Comparte Área natural protegida (A)</r1>
  <r2>Comparte Cuerpo de agua (A)</r2>
  <r3>Comparte Límite (L)</r3>
  <r4>Comparte Malpaís (A)</r4>
  <r5>Comparte Pantano (A)</r5>
  <r6>Comparte Salina (A)</r6>
  <r7>Comparte Zona arenosa (A)</r7>
  </relaciones>
  </Terreno_sujeto_a_inundacion>
  </elemento>
= <elemento nombre="Tunel">
= <Tunel>
= <atributos>
= <dominio_fijo>
= <Condicion_del_Tunel>

```

<w1>En construcción : Que está en un proceso de construcción.</w1>

<w2>En operación : Que está en servicio o puede usarse.</w2>

</Condicion\_del\_Tunel>

</dominio\_fijo>

= <dominio\_variable>

= <identificador\_de\_>

<w1>Un número secuencial que se incrementa con cada ocurrencia</w1>

= <dominio\_de\_valores>

<w1>En el intervalo de 1 a N</w1>

</dominio\_de\_valores>

</identificador\_de\_>

</dominio\_variable>

</atributos>

= <restricciones\_de\_integridad>

= <COMBINACIONES\_AUTORIZADAS\_DE\_VALORES\_DE\_ATRIBUTOS>

<c1>En construcción.</c1>

<c2>En operación.</c2>

</COMBINACIONES\_AUTORIZADAS\_DE\_VALORES\_DE\_ATRIBUTOS>

= <CALIFICADOR\_AUTORIZADO\_DE\_REPRESENTACIONES\_GEOMETRICAS>

<a1>Aproximada.</a1>

</CALIFICADOR\_AUTORIZADO\_DE\_REPRESENTACIONES\_GEOMETRICAS>

</restricciones\_de\_integridad>

= <relaciones>

<r1>Conecta Túnel (L)</r1>

<r2>Comparte Acueducto (L)</r2>

<r3>Comparte Calle (L)</r3>

<r4>Comparte Canal (L)</r4>

<r5>Comparte Carretera (L)</r5>

<r6>Comparte Vía férrea (L)</r6>

</relaciones>

</Tunel>

</elemento>

= <elemento nombre="Vado">

= <Vado>

= <atributos>

= <dominio\_fijo>

```

<w1>Ninguno</w1>
  </dominio_fijo>
= <dominio_variable>
= <identificador_de_Vado>
  <w1>Un número secuencial que se incrementa con cada ocurrencia</w1>
= <dominio_de_valores>
  <w1>En el intervalo de 1 a N</w1>
  </dominio_de_valores>
  </identificador_de_Vado>
  </dominio_variable>
  </atributos>
= <restricciones_de_integridad>
= <COMBINACIONES_AUTORIZADAS_DE_VALORES_DE_ATRIBUTOS>
  <c1>Ninguna</c1>
  </COMBINACIONES_AUTORIZADAS_DE_VALORES_DE_ATRIBUTOS>
= <CALIFICADOR_AUTORIZADO_DE_REPRESENTACIONES_GEOMETRICAS>
  <a1>Aproximada.</a1>
  <a2>Definida</a2>
  </CALIFICADOR_AUTORIZADO_DE_REPRESENTACIONES_GEOMETRICAS>
  </restricciones_de_integridad>
= <relaciones>
  <r1>Conecta Carretera (L)</r1>
  <r2>Comparte Carretera (L)</r2>
  </relaciones>
  </Vado>
  </elemento>
= <elemento nombre="Vegetacion_densa">
= <Vegetacion_densa>
= <atributos>
= <dominio_fijo>
  <w1>Ninguno</w1>
  </dominio_fijo>
= <dominio_variable>
= <identificador_de_Vegetacion_densa>
  <w1>Un número secuencial que se incrementa con cada ocurrencia</w1>
= <dominio_de_valores>

```

```

<w1>En el intervalo de 1 a N</w1>
  </dominio_de_valores>
  </identificador_de_Vegetacion_densa>
  </dominio_variable>
  </atributos>
= <restricciones_de_integridad>
= <COMBINACIONES_AUTORIZADAS_DE_VALORES_DE_ATRIBUTOS>
  <c1>Ninguna.</c1>
  </COMBINACIONES_AUTORIZADAS_DE_VALORES_DE_ATRIBUTOS>
= <CALIFICADOR_AUTORIZADO_DE_REPRESENTACIONES_GEOMETRICAS>
  <a1>Definida.</a1>
  <a2>Virtual.</a2>
  </CALIFICADOR_AUTORIZADO_DE_REPRESENTACIONES_GEOMETRICAS>
  </restricciones_de_integridad>
= <relaciones>
  <r1>Comparte Área de cultivo (A)</r1>
  <r2>Comparte Área urbana (A)</r2>
  <r3>Comparte Cuerpo de agua (A)</r3>
  <r4>Comparte Límite (L)</r4>
  <r5>Comparte Malpaís (A)</r5>
  <r6>Comparte Pantano (A)</r6>
  <r7>Comparte Rasgo arqueológico (A)</r7>
  <r8>Comparte Zona arenosa (A)</r8>
  </relaciones>
  </Vegetacion_densa>
  </elemento>
= <elemento nombre="Via_ferrea">
= <Via_ferrea>
= <atributos>
= <dominio_fijo>
= <Tipo_de_Via_ferrea>
  <w1>Vía sencilla : De una sola vía.</w1>
  <w2>Vía doble: De dos vías paralelas.</w2>
= <Restricciones_del_atributo>
  <w1>No aplicable (N/A): Indica que este atributo no es válido, cuando el
    valor del atributo Condición de la Vía es En construcción o Fuera de
    uso</w1>

```

</Restricciones\_del\_atributo>  
</Tipo\_de\_Via\_ferrea>  
= <Condicion\_de\_la\_Via\_ferrea>  
<w1>En construcción : Que está en un proceso de construcción.</w1>  
<w2>En operación : Que está en servicio o puede usarse.</w2>  
<w3>Fuera de uso : Que no está en uso.</w3>  
</Condicion\_de\_la\_Via\_ferrea>  
</dominio\_fijo>  
= <dominio\_variable>  
= <identificador\_de\_Via\_ferrea>  
<w1>Un número secuencial que se incrementa con cada ocurrencia</w1>  
= <dominio\_de\_valores>  
<w1>En el intervalo de 1 a N</w1>  
</dominio\_de\_valores>  
</identificador\_de\_Via\_ferrea>  
</dominio\_variable>  
</atributos>  
= <restricciones\_de\_integridad>  
= <COMBINACIONES\_AUTORIZADAS\_DE\_VALORES\_DE\_ATRIBUTOS>  
<c1>Vía sencilla , en operación.</c1>  
<c2>Doble vía, en operación.</c2>  
<c3>N/A, en construcción.</c3>  
<c4>N/A, fuera de uso.</c4>  
</COMBINACIONES\_AUTORIZADAS\_DE\_VALORES\_DE\_ATRIBUTOS>  
= <CALIFICADOR\_AUTORIZADO\_DE\_REPRESENTACIONES\_GEOMETRICAS>  
<a1>Definida.</a1>  
<a2>Aproximada.</a2>  
</CALIFICADOR\_AUTORIZADO\_DE\_REPRESENTACIONES\_GEOMETRICAS>  
</restricciones\_de\_integridad>  
= <relaciones>  
<r1>Conecta Calle (L) </r1>  
<r2>Conecta Carretera (L) </r2>  
<r3>Conecta Vía férrea (L) </r3>  
<r4>Conecta Instalación diversa (A) </r4>  
<r5>Conecta Instalación industrial (A) </r5>  
<r6>Conecta Área urbana (A) </r6>

<r7>Comparte Instalación portuaria (L)</r7>  
<r8>Comparte Túnel (L)</r8>  
<r9>Comparte Puente (L)</r9>  
<r10>Comparte Área natural protegida (A)</r10>  
<r11>Comparte Área urbana (A)</r11>  
<r12>Comparte Instalación diversa (A)</r12>  
<r13>Comparte Instalación industrial (A)</r13>

</relaciones>

</Via\_ferrea>

</elemento>

= <elemento nombre="Zona\_arenosa">

= <Zona\_arenosa>

= <atributos>

= <dominio\_fijo>

= <Tipo\_de\_Zona\_arenosa>

<w1>Arena : En terreno sensiblemente plano.</w1>

<w2>Dunas : En terreno con presencia de montículos.</w2>

</Tipo\_de\_Zona\_arenosa>

</dominio\_fijo>

= <dominio\_variable>

= <identificador\_de\_Zona\_arenosa>

<w1>Un número secuencial que se incrementa con cada ocurrencia</w1>

= <dominio\_de\_valores>

<w1>En el intervalo de 1 a N</w1>

</dominio\_de\_valores>

</identificador\_de\_Zona\_arenosa>

</dominio\_variable>

</atributos>

= <restricciones\_de\_integridad>

= <COMBINACIONES\_AUTORIZADAS\_DE\_VALORES\_DE\_ATRIBUTOS>

<c1>Arena.</c1>

<c2>Dunas</c2>

</COMBINACIONES\_AUTORIZADAS\_DE\_VALORES\_DE\_ATRIBUTOS>

= <CALIFICADOR\_AUTORIZADO\_DE\_REPRESENTACIONES\_GEOMETRICAS>

<a1>Definida.</a1>

<a2>Virtual</a2>

</CALIFICADOR\_AUTORIZADO\_DE\_REPRESENTACIONES\_GEOMETRICAS>

</restricciones\_de\_integridad>

= <relaciones>

<r1>Comparte Canal (L)</r1>

<r2>Comparte Límite (L)</r2>

<r3>Comparte Cuerpo de agua (A)</r3>

<r4>Comparte Malpaís (A)</r4>

<r5>Comparte Salina (A)</r5>

<r6>Comparte Terreno sujeto a inundación</r6>

<r7>Comparte Vegetación densa (A)</r7>

<r8>Comparte Zona arenosa (A)</r8>

</relaciones>

</Zona\_arenosa>

</elemento>

</Diccionario1-50000>

## ANEXO B. Código Fuente

En esta sección se despliegan los códigos utilizados por el sistema *iGIR*, comenzando por el código del algoritmo *OntoExplora* explicado en la sección 4.5. Finalizamos con el mecanismo de ponderación integral *iRank*. También se muestran los códigos fuentes que se utilizaron para extraer información de sitios Web. En particular se usa el lenguaje de programación java, el lenguaje de programación *ruby*, el lenguaje javascript, y la interfaz de programación de *Google Maps*. Adicionalmente, se muestran los códigos que permiten el procesamiento de los formatos de datos XML, KML y de *TopologyFiles*.

## B.1. Generador de KML

```
<% @page contentType="text/html"%>
<% @page pageEncoding="UTF-8"%>
<!DOCTYPE HTML PUBLIC "-//W3C//DTD HTML 4.01 Transitional//EN"
"http://www.w3.org/TR/html4/loose.dtd">

<html>
<head>
<meta http-equiv="Content-Type" content="text/html; charset=UTF-8">
<title>JSP Page</title>
<link href="styles.css" rel="stylesheet" type="text/css" />
</head>
<body>

<% @ page import="java.io.*" %>
<%

String latitud = null;
String longitud = null;

latitud = request.getParameter("latitud");
longitud = request.getParameter("longitud");

String name="prueba1";
String lat = "-122.0822035425683";
String lon = "37.42228990140251";
lon = latitud;
lat = longitud;

String[] arch;
arch = new String [11];

arch[0] = "<?xml version=\"1.0\" encoding=\"UTF-8\"?>";
arch[1] = "<kml xmlns=\"http://earth.google.com/kml/2.1\">";
arch[2] = "<Placemark>";
arch[3] = "<name>Simple placemark</name>";
arch[4] = "<description>Attached to the ground. Intelligently places itself " +
"at the height of the underlying terrain.</description> ";
arch[5] = "<Point>";
arch[6] = "<coordinates>"+lat+","+lon+",0</coordinates>";
arch[7] = "</Point>";
arch[8] = "</Placemark>";
arch[9] = "</kml>";

String nameOfFile = "C:/Documents and Settings/Administrador/Escritorio/congresos, pifi y SIP 07/ss y
pifi/iGIR/build/web/impar.kml";

try {
    PrintWriter pw = new PrintWriter(new FileOutputStream(nameOfFile));
        for(int i = 0; i<10; i++)
            pw.println(arch[i]);
        //clean up
        pw.close();
    } catch(IOException e) {
        out.println(e.getMessage());
    }
}
%>

<h1>Generador de KMLS</h1>
<br>

<%
out.print("<a href=impar.kml>Mostrar Localizacion</a>");

%>

%>
</body>
</html>
```

## B.2. Procesador de nombres usando Geonames

```
<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.0 Strict//EN" "http://www.w3.org/TR/xhtml1/DTD/xhtml1-strict.dtd">
<html xmlns="http://www.w3.org/1999/xhtml">
<head>
<script type="text/javascript" src="jsr_class.js"></script>

</head>

<%
/**
 * Passing Java Script Value to Jsp can be done in this way
 *
 * 1. Assign the Java Script value to a hidden text box in the form
 * 2. On Submit of that form, get that value and assign to Jsp variable
 */

String latitud = null;
String longitud = null;

if(request.getParameter("submit") != null){

    latitud = request.getParameter("latitud");
    longitud = request.getParameter("longitud");

    out.println("latitud : " + latitud);
    out.println("longitud : " + longitud);
}
%>
<body>
<script>

// this function will be called by our JSON callback
// the parameter jData will contain an array with geonames objects
function getLocation(jData) {
    if (jData == null) {
        // There was a problem parsing search results
        return;
    }

    var html = "";
    var geonames = jData.geonames;
    for (i=0;i< geonames.length;i++) {
        var name = geonames[i];
        // we create a simple html list with the geonames objects
        // the link will call the center() javascript method with lat/Ing as parameter
        html = html + '<a href="javascript:center('+ name.lat +','+ name.lng + ');">' + name.name + '</a><br>';
    }
    document.getElementById('resultDiv').innerHTML = html;
}

// centers the google map on the lat/Ing
function center(lat,lng){
    document.getElementById('info').innerHTML="Las coordenadas son <br>latitud:'<br>longitud:'<br>";
    document.index.latitud.value = lat;
    document.index.longitud.value = lng;
}

// calls the geonames JSON webservice with the search term
function search() {
    request = 'http://ws.geonames.org/searchJSON?q=' + encodeURIComponent(document.getElementById('q').value) + '&maxRows=10&callback=getLocation';

    // Create a new script object
    // (implementation of this class is in /export/jsr_class.js)
    aObj = new XMLHttpRequest(request);
    // Build the script tag
    aObj.buildScriptTag();
    // Execute (add) the script tag
    aObj.addScriptTag();
}

</script>

<div id="info"></div>
<form id="searchForm" onsubmit="search(); return false;" style="display:inline">
    <input id="q" type="text" name="q">
    <input type="submit" value="search">
</form>
<small>

<div id="resultDiv"></div>

<form name="index" onSubmit="genkml.jsp" method="get" action="genkml.jsp">
<input type="hidden" name="latitud"><br>
<input type="hidden" name="longitud"><br>
<input type="hidden" name="name"><br>
<input type="submit" name="submit" value="Generar KML">
</form>
</body>

</html>
```

## B.3. Pagina Índice

```
<% @page contentType="text/html"%>
<% @page pageEncoding="UTF-8"%>
<%--
The taglib directive below imports the JSTL library. If you uncomment it,
you must also add the JSTL library to the project. The Add Library... action
on Libraries node in Projects view can be used to add the JSTL 1.1 library.
--%>
<%--
<% @taglib uri="http://java.sun.com/jsp/jstl/core" prefix="c"%>
--%>

<!DOCTYPE HTML PUBLIC "-//W3C//DTD HTML 4.01 Transitional//EN"
"http://www.w3.org/TR/html4/loose.dtd">

<html>
<head>
    <meta http-equiv="Content-Type" content="text/html; charset=UTF-8" />
    <title>CIC-IPN MIGUEL FELIX MATA</title>
    <style type="text/css">
    body{
        font:75%/150% "Trebuchet MS", "Lucida Grande", "Bitstream Vera Sans", Arial, Helvetica, sans-serif;
        color:#666666;
        margin:60px;
    }
    </style>
    <script type="text/javascript" src="prettyForms.js"></script>
    <link rel="stylesheet" href="prettyForms.css" type="text/css" media="screen" />
</head>
<body onload="prettyForms()">
<% @ page import="java.io.*" %>
<% @ page import="javax.xml.parsers.DocumentBuilderFactory" %>
<% @ page import="javax.xml.xpath.*" %>
<% @ page import="org.w3c.dom.Document" %>
<% @ page import="org.w3c.dom.Element" %>
<% @ page import="org.xml.sax.InputSource" %>
<%
String status = "0";
status = request.getParameter("status");
%>

<% if(status == "0" || status == null){ %>
<h1>Busqueda en el diccionario de datos</h1>
<hr>

<form name="form1" method="get" action="index.jsp" target="fram2">
<p>
Elemento: <input type="text" name="elemen">
</p><br />
<p>
<input type="hidden" name="status" value="1">
<input type="submit" name="busqueda" value="buscar">
</p>
</form>
<% }%>

<%
String elemen = request.getParameter("elemen");

String aux = status;
String prb = status;
if(status != null){
    %>
<form action="page2.jsp" method="get" name="resultados" id="resultados" target="fram3">
<select name="elemen" onChange="MM_jumpMenu('parent',this,0)">

<%
try {
XPath xpath = XPathFactory.newInstance().newXPath();
DocumentBuilderFactory factory = DocumentBuilderFactory.newInstance();
Document XMLDoc = factory.newDocumentBuilder().parse(new InputSource(new FileInputStream("C:\\DBF GIR\\dico.xml")));
Element element = XMLDoc.getDocumentElement();
XPathExpression exp;
exp = xpath.compile("count(Diccionario1-50000/elemento/@nombre)");
int number = new Integer((String) exp.evaluate(element, XPathConstants.STRING)).intValue();

String sentence,res;
for(int i = number -1 ; i>=0 ; i--){
    sentence = "/Diccionario1-50000/elemento[last()-"+i+"]/@nombre";
    exp = xpath.compile(sentence);
    res = (String) exp.evaluate(element, XPathConstants.STRING);
    if( res.contains(elemen)){
```

## B.4. Recuperación Ontológica

```
<% @page contentType="text/html"%>
<% @page pageEncoding="UTF-8"%>

<!DOCTYPE HTML PUBLIC "-//W3C//DTD HTML 4.01 Transitional//EN"
"http://www.w3.org/TR/html4/loose.dtd">
<html>
<head>

    <meta http-equiv="Content-Type" content="text/html; charset=UTF-8">
    <title>JSP Page</title>
</head>
<body>
<% @ page import="java.io.*" %>
<% @ page import="javax.xml.parsers.DocumentBuilderFactory" %>
<% @ page import="javax.xml.xpath.*" %>
<% @ page import="org.w3c.dom.Document" %>
<% @ page import="org.w3c.dom.Element" %>
<% @ page import="org.xml.sax.InputSource" %>
<%
String status = "0";
status = request.getParameter("status");
%>

<% if(status == "0" || status == null){ %>
<h1>Busqueda en el diccionario de datos</h1>
<hr>

<form name="form1" method="get" action="index.jsp">
<p>
Elemento: <input type="text" name="elemen">
</p>
<p>

    <input type="hidden" name="status" value ="1">
    <input type="submit" name="busqueda" value="buscar">
</p>
</form>
<% }%>

<%
String elemen = request.getParameter("elemen");

String aux = status;
String prb = status;
if(status != null){

    %>
<form action="page2.jsp" method="get" name="resultados" id="resultados">
<select name="elemen" onChange="MM_jumpMenu('parent',this,0)">
```

## B.5. Recuperación Geográfica

```
<% @page contentType="text/html"%>
<% @page pageEncoding="UTF-8"%>
<%--
The taglib directive below imports the JSTL library. If you uncomment it,
you must also add the JSTL library to the project. The Add Library... action
on Libraries node in Projects view can be used to add the JSTL 1.1 library.
--%>
<%--
<% @taglib uri="http://java.sun.com/jsp/jstl/core" prefix="c"%>
--%>

<!DOCTYPE HTML PUBLIC "-//W3C//DTD HTML 4.01 Transitional//EN"
"http://www.w3.org/TR/html4/loose.dtd">

<html>
<head>
<meta http-equiv="Content-Type" content="text/html; charset=UTF-8">
<style type="text/css">
body{
font:75%/150% "Trebuchet MS", "Lucida Grande", "Bitstream Vera Sans", Arial, Helvetica, sans-
serif;
color:#666666;
margin:60px;
}
</style>
<link rel="stylesheet" href="lista.css" type="text/css" media="screen" />
</head>
<body>
<% @ page import="com.svcon.jdbf.*" %>
<% @ page import="java.io.*" %>
<% @ page import="javax.xml.parsers.DocumentBuilderFactory" %>
<% @ page import="javax.xml.xpath.*" %>
<% @ page import="org.w3c.dom.Document" %>
<% @ page import="org.w3c.dom.Element" %>
<% @ page import="org.xml.sax.InputSource" %>

<% !public String[] fun1(String so){
String [] salida = new String[3];
String min = so.toLowerCase();
int lon;
min = min.replace("Ã¡", "a");
min = min.replace("Ã©", "e");
min = min.replace("Ã­", "i");
min = min.replace("Ã³", "o");
min = min.replace("Ãº", "u");
String [] cads;
cads = min.split(" ");
lon = cads.length;
String camb = "";
for(int i = 1; i< lon-1; i++){
if(i!=lon-2)
camb += cads[i]+"_";
else
camb +=cads[i];
}
salida[0]=cads[0];
salida[1]=camb;
salida[2]=cads[lon-1];
return salida;
}%>

<%
String elemen = request.getParameter("elemen");
out.print(elemen+"<hr><br>");
```



## B.7. Página Principal

```
<% @page contentType="text/html"%>
<% @page pageEncoding="UTF-8"%>

<!DOCTYPE HTML PUBLIC "-//W3C//DTD HTML 4.01 Transitional//EN"
"http://www.w3.org/TR/html4/loose.dtd">

<!DOCTYPE HTML PUBLIC "-//W3C//DTD HTML 4.01 Frameset//EN"
"http://www.w3.org/TR/html4/frameset.dtd">
<html>
<head>
<meta http-equiv="Content-Type" content="text/html; charset=iso-8859-1">
<title>Untitled Document</title>
</head>

<frameset rows="25%,*" cols="*" frameborder="NO" border="0" framespacing="0">
  <frame src="enc.jsp" name="enc" scrolling="Y" noresize>
  <frameset rows="*" cols="25%,*" framespacing="0" frameborder="NO" border="0">
    <frame src="index.jsp" name="fram1" scrolling="NO" noresize>
    <frameset rows="*" cols="*,30%" framespacing="0" frameborder="NO" border="0">
      <frame src="blanco.html" name="fram2">
      <frameset rows="*" cols="*,10%" framespacing="0" frameborder="NO"
border="0">
        <frame src="blanco.html" name="fram3" scrolling="YES" noresize>
        <frame src="blanco.html" name="fram4" scrolling="YES" noresize>
      </frameset>
    </frameset>
  </frameset>
</frameset>
</frameset>
<noframes><body>
</body></noframes>
</html>
```

## B.8. Código Ruby para extraer contenido de Wikipedia (1)

```
require "rexml/document"
require "FileUtils"
require "open-uri"
include REXML

if File.exists?("wikifiles")
  puts " ya existe el fichero"
else
  FileUtils.mkdir 'wikifiles'
end

doc = Document.new File.new("dico.xml")
doc.elements.each("*/") do |i|
  i.elements.each do |e|
    nombre = e.attributes["nombre"]
    filtered = File.new "wikifiles/"+e.attributes["nombre"]+".txt", "w"
    wikifile = open("http://es.wikipedia.org/wiki/"+e.attributes["nombre"]).read.split("\n")
    filtered.write wikifile
    filtered.close
  end
end
```

## B.9. Código Ruby para extraer contenido de Wikipedia (2)

```
require "open-uri"

puts "Downloading descriptor file..."
#gvp_id = ARGV[0].gsub(/D/, "")
gvp_doc = open("http://wapedia.mobi/es/aeropuerto").read.split("\n")
puts gvp_doc
#gvp_doc.find { |x| x =~ /^url:(.*)$/ }
#gvp_movie_url = $1.gsub(/&/, '\\&').gsub(/\?/, '\\?')

#gvp_doc.find { |x| x =~ /^title:(.*)$/ }
#gvp_title = $1

#puts "Downloading \"#{gvp_title}\"..."
#puts "wget -O #{(gvp_title + ".avi").inspect} #{gvp_movie_url}"
```

## B.10. Código Ruby para extraer contenido de Wikipedia (3)

```
require "rexml/document"
require "FileUtils"
require "open-uri"
require 'rubygems'
require 'hpricot'
include REXML
if File.exists?("wikifiles")

puts " ya existe el fichero"
else
FileUtils.mkdir 'wikifiles'
end
if File.exists?("testi.txt")
  File.delete("testi.txt")
end
doc = Document.new File.new("dico.xml")
doc.elements.each("*/") do |i|
  i.elements.each do |e|
    nombre = e.attributes["nombre"]
    nombre = nombre.sub(/[á]/, 'a')
    nombre = nombre.sub(/[é]/, 'e')
    nombre = nombre.sub(/[í]/, 'i')
    nombre = nombre.sub(/[ó]/, 'o')
    nombre = nombre.sub(/[ú]/, 'u')
    filtered = File.new "testi.txt", "a"
    filtered.write nombre+"\n"
    filtered.close

  end
end
File.open("testi.txt", "r") do |f|
  f.each_line { |line|
    line = line.sub(/[\n]/, "")
    filtered = File.new "wikifiles/"+line+".htm", "w"
    wikifile = open("http://es.wikipedia.org/wiki/"+line).read.split("\n")
    puts line
    filtered.write wikifile
    filtered.close

    doc = open("wikifiles/"+line+".htm") { |f| Hpricot(f) }
    elements = doc.search("/html/body//p")
    #puts elements

    filtered = File.new "wikifiles/"+line+"3"+".htm", "w"
    nombre = (doc/"#toc").to_html
    nombre = nombre.sub(/[á]/, 'a')
    nombre = nombre.sub(/[é]/, 'e')
    nombre = nombre.sub(/[í]/, 'i')
    nombre = nombre.sub(/[ó]/, 'o')
    nombre = nombre.sub(/[ú]/, 'u')
    filtered.write nombre
    filtered.close
```

## B.11. Código Ruby para extraer contenido de Wikipedia (4)

```
require 'rubygems'
require 'hpricot'

doc = open("wikifiles/acueducto.htm") { |f| Hpricot(f) }
elements = doc.search("/html/body//p")
#puts elements
puts (doc/"#bodyContent").to_html

filtered = File.new "prb.htm", "w"
nombre = (doc/"#toc").to_html
  nombre = nombre.sub(/[á]/, 'a')
  nombre = nombre.sub(/[é]/, 'e')
  nombre = nombre.sub(/[í]/, 'i')
  nombre = nombre.sub(/[ó]/, 'o')
  nombre = nombre.sub(/[ú]/, 'u')
filtered.write nombre
filtered.close
```