



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

REPOSITORIO SEMÁNTICO DE DATOS
ESPACIALES

TESIS QUE PARA OBTENER EL GRADO DE MAESTRÍA
EN CIENCIAS DE LA COMPUTACIÓN

PRESENTA:

JULIO CÉSAR VIZCARRA ROMERO

DIRECTORES DE TESIS:

DR. MIGUEL JESÚS TORRES RUIZ

DR. ROLANDO QUINTERO TÉLLEZ



México D. F. Noviembre, 2010

CONTENIDO

Resumen	I
Abstract	III
Agradecimientos	V
Índice	VII
Índice de figuras	X
Índice de tablas	XIII

CAPÍTULO 1

INTRODUCCIÓN.....	1
1.1 <i>Introducción.....</i>	<i>1</i>
1.2 <i>Descripción del problema</i>	<i>4</i>
1.3 <i>Objetivos.....</i>	<i>5</i>
1.3.1 <i>Objetivos particulares.....</i>	<i>5</i>
1.4 <i>Hipótesis</i>	<i>6</i>
1.5 <i>Justificación</i>	<i>7</i>
1.6 <i>Estructura de la tesis</i>	<i>8</i>

CAPÍTULO 2

ESTADO DEL ARTE	9
2.1 <i>Introducción</i>	<i>9</i>
2.2 <i>Enfoque general</i>	<i>9</i>
2.3 <i>Recuperación semántica</i>	<i>10</i>
2.3.1 <i>Mäoller, R., et al</i>	<i>10</i>
2.3.2 <i>Jones C. B., et al</i>	<i>11</i>
2.3.3 <i>Bernard L., et al.</i>	<i>12.</i>
2.3.4 <i>Soto J., et al.....</i>	<i>13</i>
2.2.7 <i>Kashyap, V. & Sheth A.</i>	<i>14</i>
2.2.9 <i>Jones , C. B., et al R.</i>	<i>16</i>
2.2.10 <i>Hiramatsu, K. & Reitsma, F.</i>	<i>17</i>
2.2.11 <i>Visser , U. & Stuckenschmidt , H.</i>	<i>18</i>
2.2.13 <i>Kemp, Z. et al.....</i>	<i>19</i>
2.4 <i>Integración semántica</i>	<i>20</i>

2.2.5 Guarino N.	20
2.2.6 Bishr, Y.& Kuhn , W.	21
2.2.8 Bernard L., et al, R	22
2.2.12 Manoj, P. & Ghosh, S.K.	23
2.5.Comentarios Generalesl.....	25

CAPÍTULO 3

MARCO TEÓRICO.....	26
3.1 Introducción	26
3.2 El formato shapefile	26
3.3 Ontologías	27
3.3.1 El enfoque de Gruber.....	29
3.2.1 El enfoque de Guarino	31
3.3.3 Lenguajes de ontologias.....	34
3.3.4 Editor de ontologías Protegé	38
3.3.5 Lenguaje de Consulta Sparql.....	39
3.3.6. Marco de trabajo Jena para el manejo de ontologías	41
3.3.7 Motores de inferencia para ontologías.....	45
3.4 Metadatos en Información Geográfica.....	53
3.5 Geonto.Met - Metodología para la construccion de ontologías.....	59
3.6 Algoritmo de Floyd-Warshall.....	60

CAPÍTULO 4

METODOLOGÍA.....	63
4.1 Introducción	63
4.2 Descripción general de la metodología.....	64
4.2.1 Conceptualización.....	67
4.2.2 Síntesis	76
4.2.3 Análisis	79
4.3 Modelado del sistema SemGsearch	85

CAPÍTULO 5

RESULTADOS EXPERIMENTALES	91
5.1 Introducción.....	91
5.2 Resultados etapa conceptualización	91
5.3 Resultados etapa síntesis.....	107
5.4 Resultados etapa análisis.....	112
5.4.1 Sintaxis de consulta	114
5.4.2 Recuperación de resultados	114
5.4.3 Despliegue de resultados	115
5.4.4 Interfaz de búsqueda	115

5.4.5 Interfaz de administración de SemGsearch	116
5.4.6 Búsqueda de consultas en SemGsearch.....	117
5.4.7 Opciones de recuperación en la búsqueda.....	123
5.4.8 Opciones del administrador en el sistema SemGsearch	125

CAPÍTULO 6

CONCLUSIONES Y TRABAJO A FUTURO	127
6.1 Conclusiones	127
6.2 Aportaciones del presente trabajo	128
6.2.1 Aportaciones científicas.....	128
6.2.2 Aportaciones tecnológicas	129
6.3 Trabajo a futuro	130

REFERENCIAS	131
--------------------------	------------

ANEXOS	135
Anexo I Divisiones de la República Mexicana	135
Anexo II Archivo de Metadatos	137

ÍNDICE DE FIGURAS

<i>Figura 1.1. Heterogeneidad semántica dentro de un grupo de usuarios y en el repositorio de datos geoespaciales..</i>	5
<i>Figura 1.2. Indexación semántica para búsquedas.....</i>	6
<i>Figura 1.3. Recuperación semántica en repositorios de información geoespacial.....</i>	7
<i>Figura 3.1. Tipos de ontologías de acuerdo con su nivel de dependencia para una tarea en particular. Las líneas representan relaciones de especialización...</i>	33
<i>Figura 3.2. Lenguajes de marcado de ontologías.....</i>	35
<i>Figura 3.3. Diagrama de paquetes de SPARQL.....</i>	40
<i>Figura 3.4. Interfaz de objetos de SPARQL.....</i>	41
<i>Figura 3.5. Modelo de consulta de SPARQL.....</i>	41
<i>Figura 3.6. Interfaz de Jena para comunicarse con razonadores.....</i>	43
<i>Figura 3.7. Modelo para el motor de inferencia de Jena.....</i>	43
<i>Figura 3.8. Herencia entre interfaces de Jena.....</i>	44
<i>Figura 3.9. Operación de un motor de razonamiento... ..</i>	47
<i>Figura 3.10. Uso de FaCT++ con Protégé.....</i>	49
<i>Figura 3.11. Estructura de RacerPro... ..</i>	50
<i>Figura 3.12. La interfaz gráfica RacerPorter.....</i>	51
<i>Figura 3.13. Estructura general de la FGDC.....</i>	54
<i>Figura 3.14. Firma de la estructura de FGDC.....</i>	55
<i>Figura 3.15. Sección de información de identificación de los datos (FGDC)....</i>	55
<i>Figura 3.16. Sección de información de calidad de los datos (FGDC)....</i>	56
<i>Figura 3.17. Sección de información de la organización de los datos geoespaciales (FGDC)....</i>	56
<i>Figura 3.18. Sección de información de la referencia espacial de los datos geoespaciales (FGDC)....</i>	57
<i>Figura 3.19. Sección de información de tipo de entidad y atributo de los datos geoespaciales (FGDC)....</i>	57
<i>Figura 3.20. Sección de información de la distribución de los datos geoespaciales (FGDC)....</i>	58
<i>Figura 3.21. Sección de información de referencia de los metadatos (FGDC).....</i>	58
<i>Figura 3.22. Secciones finales de la estructura de los metadatos (FGDC).....</i>	58
<i>Figura 3.23. Definición de ruta más corta utilizando el algoritmo Floyd-Warshall.....</i>	61
<i>Figura 4.1. Estructura conceptual de SemGSearch.....</i>	65
<i>Figura 4.2. Solución general conforme metodología... ..</i>	67
<i>Figura 4.3. Ontología General... ..</i>	68
<i>Figura 4.4. Dominios definidos en la conceptualización... ..</i>	69
<i>Figura 4.5. Relaciones semánticas en la ontología... ..</i>	70
<i>Figura 4.6. Fragmento de la ontologías con relaciones axiomáticas entre conceptos... ..</i>	74

Figura 4.7. Grafo con las distancias conceptuales obtenidas a partir de la ontología de la Figura 4.4.....	75
Figura 4.8. Esquema para generar las instancias de los datos geoespaciales.....	77
Figura 4.9. Etapa de síntesis en la ontología.....	78
Figura 4.10. Conceptos clave de cada dominio en la fase de búsqueda de términos.....	79
Figura 4.11. Diagrama funcional del proceso de búsqueda y recuperación semántica en el caso $k=0$	80
Figura 4.12. Distancia conceptual extendiendo el rango de búsqueda.....	82
Figura 4.13. Intersección de resultados correspondientes a los dominio espacial, temático y temporal.....	83
Figura 4.14. Resultado y proceso de la recuperación semántica para $k \geq 0$	84
Figura 4.15. Resultado final de la recuperación semántica.....	85
Figura 4.16. Caso de uso – Usuario que desea buscar y recuperar un objeto geoespacial 1.5.....	86
Figura 4.17. Caso de uso – Usuario administrador del sistema SemGSearch.....	87
Figura 4.18. Diagrama de secuencia para carga de dominios OWL/RDF del sistema SemGSearch.....	88
Figura 4.19. Diagrama de secuencia para agregar nuevo servidor, carga de metadatos y poblar ontología.....	88
Figura 4.20. Bases de datos que integran al sistema SemGSearch.....	88
Figura 4.21. Modelo entidad – relación de las bases de datos del sistema SemGSearch.....	89
Figura 4.22. Arquitectura del sistema SemGSearch.....	90
Figura 5.1. ontología general en visualizada en paquetes.....	92
Figura 5.2. ontología general en visualizada en forma de árbol.....	93
Figura 5.3. Partición de la ontología referente al dominio temático.....	94
Figura 5.4. Clase vías de comunicación con sus conceptos y relaciones.....	94
Figura 5.5. Clase área urbana con sus conceptos y relaciones.....	95
Figura 5.6. Clase edificación con sus conceptos y relaciones.....	95
Figura 5.7. Clase área verde con sus conceptos y relaciones.....	96
Figura 5.8. Partición de la ontología referente al dominio espacial.....	97
Figura 5.9. Clase división ecorregional con sus conceptos y relaciones.....	97
Figura 5.10. Clase división económica con sus conceptos y relaciones.....	98
Figura 5.11. Clase división zonas espaciales con sus conceptos y relaciones.....	98
Figura 5.12. Partición de la ontología referente al dominio temporal.....	99
Figura 5.13. Clase siglo XXI con sus conceptos y relaciones.....	99
Figura 5.14. Clase fecha con sus conceptos y relaciones.....	100
Figura 5.15. Partición de la ontología referente a los metadatos FGDC.....	100
Figura 5.16. Tablas de cada dominio de la ontología en la Base de Datos.....	101
Figura 5.17. Grafo DIS-C para el dominio espacial.....	103
Figura 5.18. Grafo DIS-C para el dominio espacial con 30 objetos.....	104
Figura 5.19. Interfaz para agregar servidores a SemGSearch.....	107
Figura 5.20. Enlace de metadatos con los dominios conceptualizados en la ontología.....	108

<i>Figura 5.21. Carga de archivo de metadatos desde el servidor.....</i>	<i>109</i>
<i>Figura 5.22. Archivos para cada metadato.....</i>	<i>110</i>
<i>Figura 5.23. Pantalla de inicio al sistema SemGSearch.....</i>	<i>115</i>
<i>Figura 5.24. Interfaz de búsqueda semántica en SemGSearch... ..</i>	<i>116</i>
<i>Figura 5.25. Interfaz de administración de SemGSearch... ..</i>	<i>117</i>
<i>Figura 5.26. Búsqueda semántica de concepto “calle”, “df”, “2007” con K=0..</i>	<i>117</i>
<i>Figura 5.27. Búsqueda semántica de concepto “calle”, “df”, “2007” con K=1..</i>	<i>118</i>
<i>Figura 5.28. Búsqueda semántica de concepto “calle”, “df”, “2007” con K=1..</i>	<i>119</i>
<i>Figura 5.29. Búsqueda semántica de concepto “calle”, “df”, “2007” con K=2..</i>	<i>120</i>
<i>Figura 5.30. Búsqueda semántica de concepto “boulevard”, “veracruz”, “2006” con K=0.....</i>	<i>121</i>
<i>Figura 5.31. Partición de la ontología para la división ecorregional, particularmente en selvas cálida húmedas.....</i>	<i>121</i>
<i>Figura 5.32. Búsqueda semántica de concepto “boulevard”, “veracruz”, “2006” con K=1.....</i>	<i>122</i>
<i>Figura 5.33. Búsqueda semántica de concepto “boulevard”, “veracruz”, “2006” con K=2.....</i>	<i>122</i>
<i>Figura 5.34. Descarga de los metadatos FGDC.....</i>	<i>123</i>
<i>Figura 5.35. Descarga de un archivo shapefile.....</i>	<i>124</i>
<i>Figura 5.36. Visualización de un archivo TIFF.....</i>	<i>124</i>
<i>Figura 5.37. Descarga de un archivo, en este caso Excel referente a un objeto geoespacial.....</i>	<i>125</i>
<i>Figura 5.38. Carga automática de dominios en SemGSearch.....</i>	<i>126</i>
<i>Figura 5.40. Ver metadatos de un servidor</i>	<i>126</i>

ÍNDICE DE TABLAS

<i>Tabla 3.1. Comparación de características entre razonadores.....</i>	<i>53</i>
<i>Tabla 3.2. Definición de términos de FGDC... ..</i>	<i>55</i>
<i>Tabla 3.3. Algoritmo Floyd-Warshall.....</i>	<i>62</i>
<i>Tabla 4.1. Algoritmo DIS-C.....</i>	<i>73</i>
<i>Tabla 4.2. Matriz generada por DIS-C, indicando el camino más corto.....</i>	<i>76</i>
<i>Tabla 4.3. Valores de crecimiento de los radios... ..</i>	<i>81</i>
<i>Tabla 4.4. Descripción de alto nivel del caso de uso usuario de búsqueda</i>	<i>86</i>
<i>Tabla 4.5. Descripción de alto nivel del caso de uso usuario administrador.....</i>	<i>87</i>
<i>Tabla 5.1. Tabla de objetos referentes al dominio espacial.....</i>	<i>102</i>
<i>Tabla 5.2. Tabla de tipo de objeto... ..</i>	<i>102</i>
<i>Tabla 5.3. Tabla de tipo de objeto... ..</i>	<i>103</i>
<i>Tabla 5.4. Valores del grafo DIS-C.....</i>	<i>105</i>
<i>Tabla 5.5. Tabla de distancias conceptuales más cortas entre todos los conceptos, aplicando el algoritmo de Floyd-Warshall.....</i>	<i>106</i>
<i>Tabla 5.6. Estructura del archivo de metadatos.....</i>	<i>109</i>
<i>Tabla 5.7. Dominios en el modelo persistente de base de datos.....</i>	<i>110</i>
<i>Tabla 5.8. Tabla de instancias del dominio metadato FGDC.....</i>	<i>111</i>
<i>Tabla 5.9. Definición de instancias en el dominio de metadatos... ..</i>	<i>111</i>
<i>Tabla 5.10. Relaciones vinculadas a cada dominio en la ontología.....</i>	<i>112</i>

Capítulo 1. Introducción

1.1. Introducción

En la actualidad, existe un gran volumen de datos recopilados acerca de la Tierra, los cuales son obtenidos por diferentes tecnologías, tales como GPS (Global Positioning System), imágenes satelitales, bases de datos geográficas, mapas en formato analógico, entre otras fuentes [1], y no solamente por los nuevos sistemas de información espaciales, sino también por las tecnologías de recolección de datos que cada vez son más sofisticadas [2]. Aunado a lo anterior, los datos geoespaciales son una parte importante para cualquier sistema con soporte de decisiones (DSS), éstos pueden ser considerados como elementos clave para la planificación y toma de decisiones en una variedad de aplicaciones. Este escenario ha suscitado en los últimos años el desarrollo de tecnologías para su integración y herramientas para su manejo y análisis [3].

En el transcurso de la historia, la cartografía y la elaboración de mapas ha constituido una herramienta de precisión, donde la matemática y las proyecciones de una esfera (o geode) sobre un plano han guiado con éxito la confección de mapas básicos y temáticos, así como las cartas náuticas; por lo que es posible afirmar que la confección de mapas y cartas por medios matemáticos exactos es una ciencia madura [4].

A pesar de lo anterior, entre los mapas y de forma general en las representaciones geoespaciales, no existe un común acuerdo en la representación de la semántica espacial [5]. Diferentes organizaciones dibujan con una exactitud determinadas líneas, puntos o polígonos sobre un plano; para representar ciudades, pozos de abasto de agua, o puntos de la red altimétrica, líneas de transmisión eléctrica, infraestructura vial. Sin embargo, no existe un consenso o acuerdo entre organizaciones o grupos de especialistas sobre el significado, semántica u ontología de estos trazos, lo que origina que para una organización una representación particular como es un “suelo”, para otra, es “una formación geológica”; lo que para una organización son “lagos artificiales”, para otra pueden ser “presas” . El problema ahora no está en cómo representar con exactitud un elemento geográfico, sino que dos representaciones espaciales o bases de datos geográficas representen lo mismo, o tengan una unidad semántica común, o que un usuario familiarizado con una nomenclatura (representación semántica) entienda otra información cartográfica hecha por otra organización u otro usuario [6]. Esta unión semántica es la base para la obtención de una verdadera interoperabilidad e intercambio de los datos geoespaciales entre diferentes

usuarios. Por lo tanto, los sistemas de información geográfica (SIG), al manejar y procesar estos datos no se encuentran exentos de esta problemática [7].

Dado que la mayoría de los SIG no son originalmente diseñados para trabajar en cooperación, surgen varios problemas de interoperabilidad, mientras que la integración de estas fuentes heterogéneas de datos geoespaciales no puede ser lograda en estos sistemas [8]. Esto se debe principalmente a que cada SIG proporciona especificaciones particulares para la representación de sus datos, tales como su propio formato y lenguaje de consulta específico [9].

Por otra parte, los recursos geográficos están diseñados para una variedad de propósitos, tal es el caso de las direcciones ortogonales en el diseño de los recursos geográficos, los cuales pueden afectar a la semántica de los datos geoespaciales y entorpecer su integración [10, 11, 12]. Estas discrepancias hacen la integración de los distintos recursos geográficos significativamente compleja. Las primeras investigaciones acerca de la integración en bases de datos, surgen a mediados de los años 80 del pasado siglo y se han venido desarrollando hasta la actualidad, en donde la interoperabilidad se ha transformado prácticamente en una ciencia de la integración [13]. Debido a la complejidad y riqueza de los datos geográficos y la dificultad de su representación, se producen una serie de problemas muy específicos para lograr la interoperabilidad en este tipo de datos [14]. Algunos autores han sugerido que estructuras sofisticadas como las ontologías son buenas candidatas para representar y resumir los datos geográficos, a la par de lograr que los mismos sean compartidos [15].

Las ontologías como una especificación de software fueron sugeridas desde 1991 [16]. El uso de ontologías en la construcción de sistemas de información es ampliamente discutida por Guarino [17] y específicamente en la construcción de SIG, en el trabajo de Frank [18]. Nuñez en [19] puntualiza que el primer paso para construir la nueva generación de SIG sería la construcción de una colección sistemática y específica de entidades geográficas, sus propiedades y relaciones. En años recientes, Fonseca, Egenhofer y Agouris en [20] introducen una nueva arquitectura SIG que puede facilitar la integración de manera flexible y se basa en el valor semántico de la información, sin tener en cuenta su representación. La solución propuesta es un SIG conducido por ontologías, el cual actúa como un sistema integrador independientemente del modelo. Un sistema de información conducido por ontologías evita la clasificación de los datos sobre la base de sus representaciones. El enfoque semántico basado en el concepto de entidades geográficas, habilita la integración "natural" de diferentes tipos de información, a través del uso de clases flexibles .

El éxito en la integración de la información y en el intercambio de datos entre distintos sistemas y diseños, es requerido para un acceso rápido y para la correcta interpretación de muchos tipos de información geoespacial. Por su naturaleza intrínseca, los datos geoespaciales son muy heterogéneos - no sólo se diferencian en la representación de los datos y de los métodos de almacenamiento, sino además también difieren en la forma de consultar los datos, encontrar y acceder a los mismos, lo cual origina que las tareas de integración e intercambio de información geoespacial sean una tarea crucial para la explotación de los datos geoespaciales en diferentes SIG.

Podemos concluir sin embargo, que toda esta información en algunos casos presenta diversas características, como pueden ser los formatos de desarrollo, aspectos relacionados con los sistemas de referencia, escala, así como heterogeneidad entre los términos definidos en las bases de datos [21].

El problema de la integración de fuentes de datos heterogéneas y en algunos casos no estructurados, conlleva a problemas relacionados directamente con la recuperación de la información geográfica, no obstante diversos métodos han sido propuestos utilizando como medio fundamental las ontologías [22]. En los últimos años, el problema de la precisión en la recuperación de información puede ser visto como consecuencia de la falta de significado o semántica [23], tanto en la Web como en bases de datos y repositorios heterogéneos no estructurados de información.

Por tanto, se puede constatar que el proceso para encontrar información que realmente sea útil en cualquier contexto, y bajo las especificaciones que se creen inherentes al perfil del usuario que realiza la búsqueda, representa un problema que puede ser solucionado parcialmente mediante el procesamiento semántico, ya que en la actualidad a diferentes conceptualizaciones de los dominios de los datos se obtienen consultas o se recupera información que en algunos casos es muy distinta a lo que se espera.

Dentro del área de las ciencias de la información geográfica (GIScience), la recuperación de información geoespacial, considera al atributo espacial de un objeto geográfico como un elemento importante y requerido en esta tarea, debido a que una parte recae directamente en la semántica y éste se considera como un factor delimitante y a su vez restrictivo en la búsqueda.

1.2. Descripción del problema

Hoy en día, diferentes áreas del conocimiento y diversos desarrollos tecnológicos presentan el problema de la *heterogeneidad e interoperabilidad semántica* de la información, aunado a la vez que entidades conceptualizadas entre personas son diferentes, debido al conocimiento particular sobre un dominio (expertos o usuarios comunes) [24]. La problemática a resolver radica en la diversificación de los datos geoespaciales en diferentes conceptualizaciones, modos de representación con diversos significados, lo cual incrementa en tiempo, esfuerzo y complejidad la búsqueda y recuperación. De acuerdo con lo anterior, el problema de la heterogeneidad semántica origina una ineficiente y poco confiable recuperación de información y además depende directamente de la *integración* exitosa de diversas fuentes de datos heterogéneas.

Por ejemplo, en la Figura 1.1 se muestra el problema relacionado con la integración de fuentes de datos heterogéneas. Se muestra a un grupo de usuarios que solicitan información en un repositorio, cada usuario requiere de un objeto geográfico, lo describe por medio de su conocimiento *a priori*, como cada uno lo conceptualiza pero no lo especifica como en realidad está almacenado; es decir, el usuario 1 intenta buscar *carretera pavimentada*, el usuario 2 *calle primaria*, y el usuario 3 *calle de primer orden*; en este caso, todos se refieren al objeto geográfico *avenida*, sin embargo, dentro del repositorio cada fuente de datos contiene a un solo objeto geográfico y éste se devuelve a cada usuario. De esta forma, se obtienen resultados parciales o en el peor de los casos no se encuentran objetos geográficos de forma directa que estén relacionados con el concepto que se está buscando en el repositorio.

Por lo tanto, esta propuesta radica en medir el grado de similitud conceptual entre todos los objetos geográficos que se encuentren en el repositorio o distribuidos en diferentes sitios, los cuales sean útiles, similares y/o iguales; así como de interés para el usuario y puedan ser recuperados como resultados alternativos a una petición original. Asimismo, en el ejemplo de la Figura 1.1 se puede observar la heterogeneidad semántica que se presenta entre los diversos usuarios y en las fuentes de datos.

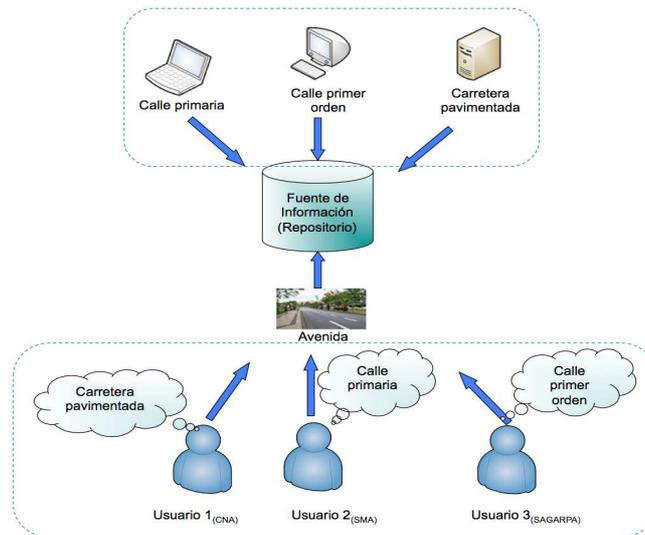


Figura 1.1. Heterogeneidad semántica dentro de un grupo de usuarios y en el repositorio de datos geoespaciales.

Actualmente, esta problemática se ve reflejada en la Secretaría del Medio Ambiente del Distrito Federal (SMA-DF), en donde es necesario realizar un cierto grado de integración para compartir y recuperar información geoespacial heterogénea por las diversas áreas de la Secretaría, utilizando para tal fin el procesamiento semántico en datos geoespaciales.

1.3. Objetivos

Construir una metodología de búsqueda, recuperación e integración semántica de datos geoespaciales en un repositorio, utilizando un conjunto de ontologías de aplicación para resolver problemas relacionados con la heterogeneidad semántica en fuentes de datos geoespaciales, para que los resultados pueden ser recuperados de forma inteligente y con un rango de relevancia en una aplicación Web, evitando siempre arrojar resultados vacíos.

1.3.1. Objetivos particulares

- Implementar un conjunto de ontologías de aplicación a partir del caso de estudio.
- Construir un motor de búsqueda para recuperar los datos espaciales con base en el conjunto de ontologías.
- Implementar una aplicación Web para desplegar resultados por rango de relevancia.
- Diseñar un repositorio semántico para administrar fuentes de datos heterogéneas.

1.4. Hipótesis

De acuerdo con la descripción del problema, se ha considerado como hipótesis que la recuperación de datos geospaciales se puede realizar de forma inteligente, considerando un proceso de síntesis semántica sobre los objetos geográficos, con base en diversos contextos en donde esta información puede variar su significado; sin embargo, en estos casos se trata del mismo objeto. Por lo tanto, las siguientes aseveraciones son propuestas para el desarrollo de la metodología:

1. Establecimiento y adecuación de metadatos para cada objeto geográfico, con la finalidad de obtener una especificación para realizar la recuperación. En este proceso se propone mejorar la indexación mediante la catalogación de los metadatos, y que éstos sean guardados en una base de datos espacial para su explotación por la ontología (ver Figura 1.4).

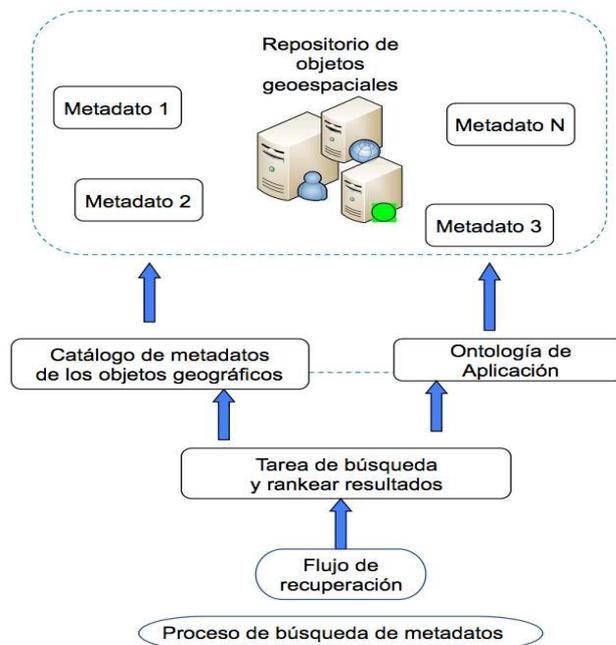


Figura 1.2. Indexación semántica para búsquedas.

2. Conceptualización e implementación de una ontología de aplicación para explotar los metadatos y convertirse en el elemento principal para la búsqueda y recuperación semántica. A partir de la ontología se crea una consulta, en donde las instancias que se obtengan serán mapeadas a la ontología para proceder a realizar la búsqueda en el proceso de recuperación, con el objetivo de proveer mayor eficiencia, ésta debe ejecutarse en la base de datos, lo cual devolverá objetos geográficos con una semántica similar (ver Figura 1.5).

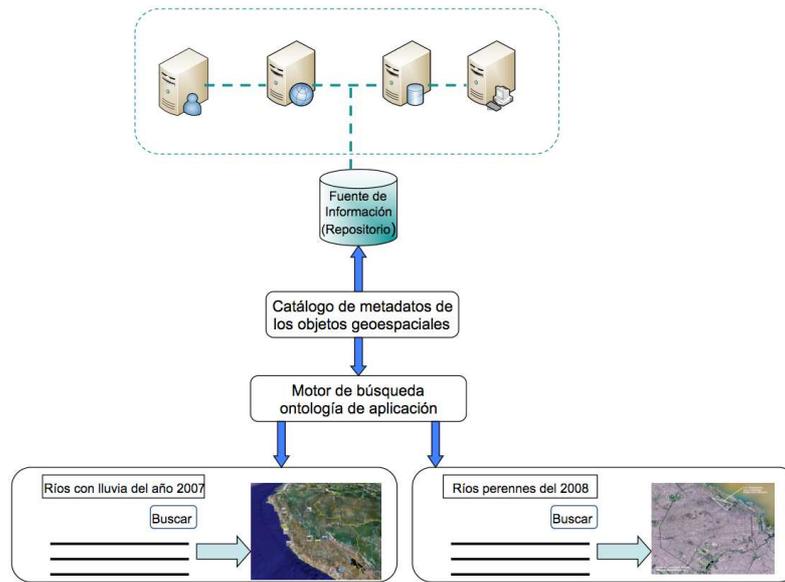


Figura 1.3. Recuperación semántica en repositorios de información geoespacial.

3. En la entrega de resultados se presentará la información geoespacial con una semántica similar; es decir, de acuerdo con la relevancia de resultados establecida por la similitud conceptual, obtenida de la búsqueda y recuperación. Los datos geoespaciales recuperados se desplegarán en una aplicación Web.
4. Determinar la mejor forma de explotación de los metadatos de los objetos geográficos para el desarrollo más eficiente de la ontología de aplicación.
5. Utilizar GEONTO-MET para construir la ontología de aplicación, capaz de conceptualizar de manera correcta los dominios de búsqueda. De esta forma se intenta mejorar la precisión de la recuperación semántica de datos geoespaciales.

1.5. Justificación

El presente trabajo de tesis presenta una solución al problema de la heterogeneidad semántica de la información geoespacial en repositorios no estructurados o fuentes de datos heterogéneas, proporcionando un mecanismo eficiente de *integración y recuperación semántica*, aplicada a datos geoespaciales. La característica principal radica en el uso de una ontología de aplicación, la cual provee un mejor resultado en la recuperación.

En este sentido, la SMA-DF presenta la problemática de que diferentes áreas necesitan información de otras y la recuperación es muy lenta, o en algunos casos ésta no se logra. Las ventajas de esta metodología de integración y recuperación semántica se mencionan a continuación:

- Con el desarrollo de una metodología de recuperación semántica se obtienen mejores resultados en las búsquedas y se puede explotar mejor la información entre diferentes usuarios, a través de un repositorio semántico, evitando principalmente devolver resultados nulos a los usuarios.
- Con el uso de una ontología de aplicación, se puede contribuir a resolver la heterogeneidad semántica y explotar los metadatos de la información geoespacial con que se cuenta.

1.6. Estructura de la tesis

El Capítulo 1 presenta la descripción del problema, los objetivos, la justificación, posibles hipótesis y el enfoque general de la investigación. En el Capítulo 2 se describen algunos trabajos, sobre esta temática y que ya han sido abordados y se apegan con el propósito de esta tesis. El Capítulo 3 describe el marco teórico, en donde se definen conceptos y herramientas importantes para el desarrollo de la tesis. El Capítulo 4 describe la metodología propuesta para la integración y recuperación semántica de objetos geográficos de diversas fuentes heterogéneas. El Capítulo 5 presenta los resultados experimentales, aplicando la metodología desarrollada, así como el proceso de análisis. Finalmente, el Capítulo 6 se presentan las conclusiones y el trabajo a futuro. Adicionalmente, se adjuntan a esta tesis los anexos correspondientes a la división espacial según CONABIO (Comisión Nacional para el Conocimiento y Uso de la Biodiversidad) y el archivo de metadatos desarrollado.

Capítulo 2. Estado del Arte

2.1. Introducción

En este capítulo se aborda el estado del arte relacionado con el trabajo de investigación de esta tesis. De igual forma, se exploran los resultados de diversas investigaciones y proyectos realizados por diversos científicos y académicos dentro del área de ciencias de la información geográfica.

2.2. Enfoque general

La recuperación de información, conocida también por *Information Retrieval (IR)*, es considerada ya, una ciencia, encargada de la búsqueda de información en documentos, búsqueda dentro de los mismos, búsqueda de metadatos que describan documentos, o también la búsqueda en bases de datos relacionales, ya sea a través de Internet, Intranet, para textos, imágenes, sonido o datos de otras características de manera pertinente y relevante [25].

La recuperación de información es un estudio interdisciplinario, el cual cubre tantas disciplinas que eso genera normalmente un conocimiento parcial desde tan solo una u otra perspectiva. Los sistemas automáticos de IR son utilizados para reducir la sobre *carga de información*.

Un proceso de recuperación de información inicia cuando un usuario introduce una consulta en el sistema. Las consultas son sentencias formales sobre una necesidad particular de información. En IR, una consulta no solo identifica un objeto único en una colección de datos, sino que un conjunto de varios objetos pueden coincidir con la consulta, tal vez, con diferente grado de *relevancia* [26]. Un objeto es una entidad que se representa por la información en una base de datos. Las consultas de los usuarios son mapeadas contra la información de la base de datos, dependiendo de la aplicación los objetos pueden ser de diversos tipos tales como: documentos de texto, imágenes, audio, video, mapas mentales entre otros.

La mayoría de los sistemas IR calculan una puntuación numérica sobre cómo cada objeto en la base de datos coincide en la consulta y establece un *ranking* de los objetos de acuerdo

con este valor. El valor más alto es el que se le presenta al usuario, por tanto, este proceso puede ser iterativo si se desea refinar la consulta.

De acuerdo con lo anterior, a continuación se presentan varios trabajos relacionados con los tópicos de investigación de recuperación de información, utilizando la componente semántica para recuperar datos, en algunos casos se describen trabajos orientados directamente en el contexto geoespacial.

2.3 Recuperación semántica

2.3.1. Mäoller, R., *et al.*

Este trabajo [27] muestra la problemática de cómo gradualmente en los sistemas de información se incrementa la cantidad de datos, haciéndose heterogénea y desorganizada una base de datos, con un ejemplo ilustrativo de usuarios consultando una base de datos para obtener la guía de televisión de su programa favorito. De este modo, un usuario desea una película con barcos de vela, la cual debe ser encontrada en todos los canales abiertos y privados en todos los horarios. En este caso se observa que cada usuario desea algún tipo de barco, pero la búsqueda depende de un criterio particular que puede ser la edad, lenguaje, entre otros, o bien solicitar directamente el título del programa o película.

La búsqueda se hace por medio de descriptores textuales, lo cual no se presta mucho a la flexibilidad en cuanto a la precisión que se logra obtener con descriptores conceptuales semánticos o lógicos. La solución propuesta es la construcción de un marco de trabajo de consultas sofisticadas, la cual presenta ventajas como tener mecanismos de inferencia para su proceso y además adecua mejor la información conceptualmente incompleta.

Así, la consulta sofisticada se realiza modelando los objetos con descripciones lógicas para una mejor clasificación, la cual será explotada por un lenguaje especializado por medio de consultas y operaciones sobre grupos de entidades de información.

A este lenguaje se le especificó una gramática, así como un orden de los conjuntos referidos a buscar, terminología específica para ayudar a definir mejor el dominio de búsqueda y hacerlo un poco más explícito sin perder la flexibilidad. Se generó una función de interpretación para que la terminología satisfaga los atributos del objeto o conjuntos de objetos especificados. Los servicios de inferencia se desarrollan a nivel de conjuntos y

explotando los descriptores lógicos para relacionar similitudes haciendo una unión entre ellos por atributos en común y dominios.

En resumen en este trabajo de investigación el método de recuperación se realizó extendiendo descriptores lógicos para agregar en la recuperación conceptos espaciales, de este modo se proporciona un lenguaje de consultas para hacer más explícitas y específicas las búsquedas sobre un repositorio de datos.

La ventaja del uso de los descriptores lógicos se refleja en el modelado del conjunto de inferencias entre los objetos del dominio y adecuando la información incompleta en el método de recuperación.

2.3.2. Jones, C.B., *et al.*

En este trabajo [28] se describen diversas formas de recuperación de información espacial, proponiendo un candidato de la siguiente manera:

- Por localización geográfica y por rango de cercanía con otros objetos similares de interés del usuario, por medio de relaciones espaciales que están dentro del mismo concepto.
- Por definición de un contexto global espacial a otros niveles de detalle.
- Por proximidad utilizando un algoritmo euclidiano estableciendo una razón para limitar la búsqueda por cercanía partiendo de centroides de los objetos espaciales.
- Por nombre, semántica, referencia geográfica y relaciones entre objetos. No aborda similitudes por formas y geometría.
- Por la implementación de una ontología del lugar para hacer inferencia, obteniendo medidas de distancia semántica.

De este modo, la primera opción de recuperación se basa en localización y proximidad y se usaron los nombres de lugar como descriptores para regiones del espacio, se tiene un enfoque en múltiples herencias geográficas en combinación con distancias euclidianas.

Para evitar almacenar esta información de proximidad y que ésta tenga que ser precalculada, con la finalidad de obtener estos valores se utilizaron Diagramas de Voronoi del conjunto de centroides, para referencias de objetos espaciales que se encuentran dentro de la región de interés y para las relaciones del conocimiento de la contención y la exclusión pudieron ser derivados de las relaciones topológicas almacenadas de “contenido” y la “conectividad” de la región.

Para la herencia en distancia se basaron en el método de súper clases no comunes [29], el cual clasifica en un contexto de métricas similares en combinación de relaciones de generalización, clasificación y atributos. Las regiones padres proveen una localización común para que después con la herencia se caractericen.

En resumen, el trabajo se enfoca en la recuperación de similitudes por proximidad basada en distancia euclidiana y de nombres o características semánticas clasificadas por regiones asociadas por tipos de lugar. La explotación de la ontología contribuye en gran medida para la recuperación conducida por nombres y además por proximidades semánticas de los objetos; es decir, objetos similares en su semántica se encuentran cerca de la región de interés en la recuperación de esas entidades.

2.3.3. Bernard, L., et al.

Este trabajo de investigación [30] es un ejemplo práctico para proporcionar información referente a cuestiones geográficas; pero varias instituciones o fuentes pueden proveer esta consulta definiendo de manera diferente sus datos y formatos. La intención es dar una solución al problema de la heterogeneidad semántica en la información geográfica en las búsquedas libres de contexto en catálogos de estructuras de datos espaciales.

Así, la solución propuesta fue realizar una extensión a las capacidades de consulta actuales, proporcionadas por la OGC (OpenGIS Consortium), en los catálogos de definición para la información geográfica; es decir, los metadatos. Por tanto, se procedió a realizar una mejora a lo anterior sobre ellos, aplicando una ontología para objetos o componentes en las búsquedas enfocadas a la Web.

Este proceso se dividió en dos fases: la primera de *descubrimiento* de objetos espaciales, donde la guía principal son los metadatos que definen el modo en que se creará la consulta.

También, fue necesario encontrar similitudes entre nombres (sinónimos) del metadato su significado (homónimo), los cuales son heterogéneos en distintas fuentes.

La segunda fase es la de *recuperación*, ya que se encuentra definido el objeto espacial por los metadatos, ahora se procede a encontrar similitudes entre el nombre y las características; es decir, encontrar la semántica propia de cada componente y compararla con otros para verificar si son similares.

A su vez, se implementó una ontología de tipo híbrida [31], en la cual independientemente de la fuente de los objetos se encuentra un vocabulario compartido en cada dominio. El esquema consiste de una fuente, la cual mapea el dominio por una ontología de aplicación; a su vez, entre todas ellas se encuentra un vocabulario compartido. Con ello los objetos de un dominio entienden conceptos comunes con otros de otro dominio diferente. Con el uso de este tipo de ontología el trabajo sufrió unas modificaciones, las cuales son que la información de los recursos no es anotada directamente; sino que se detallan los tipos de características de esa fuente y el dominio no solo se limita al de la ontología, sino a los conceptos fuera de ella del tipo de atributivo para lograr una mejor adecuación semántica de sus componentes. También se establecieron conceptos definidos para cubrir la heterogeneidad semántica en el descubrimiento y proveer una interpretación de tipos de características durante la recuperación de la información.

Como resumen el proyecto tomó como factores de solución a la heterogeneidad de los datos, el descubrimiento del tipo de metadatos que definen sus conceptos de la fuente, y la recuperación de los objetos por medio de la semántica entre ellos.

2.3.4. Soto, J., et al.

En este trabajo [32], a pesar que no se manejan datos espaciales, sino objetos de aprendizaje, se aborda también el problema de la semántica en la recuperación de estos objetos en los repositorios, un control y definición de metadatos y se propone una ontología para resolver la heterogeneidad mediante similitudes semánticas en su recuperación.

Todo lo anterior tiene como plataforma la Web, dando soporte a procesos y búsquedas automatizadas, como soporte a la delegación de tareas como apoyo a los agentes para generalizar sus objetos de aprendizaje, etc.

En la solución se implementó una ontología SLOR (Semantic Learning Object Repository), donde cualquier objeto enlazado a una actividad de aprendizaje se convierte en un objeto de aprendizaje, de modo que así se define mejor concepto del objeto y no se limita a una sola definición. Con esto se agrega una mejor flexibilidad para adoptar definiciones conceptuales normalizadas y se adecua a su propósito, relacionando valores semánticos. La ontología tiene la finalidad específica para la creación y administración de metadatos con el propósito de integración e intercambio entre varios usuarios, administradores y sistemas.

También se tiene la posibilidad de que el usuario final pueda consultar objetos en el repositorio, crear y manejar los descriptores de los metadatos; así como incluir en éstos especificaciones o anotaciones del propósito de los objetos, sin importar la conceptualización de quién explota los recursos, sea usuario o agente. Por tanto, se mejora la forma de hacer inferencia en las búsquedas. El proceso de búsqueda solicita instancias de las conceptualizaciones del modelo de la ontología y si ésta puede también enlazar otras ontologías externas. Con este proceso es posible hacer consultas complejas de conceptos abstractos o extensos, como solicitar por ejemplo, objetos en algún tiempo, evento social etc., es decir, no específica y más bien se refiere a conceptos generales que abarcan muchos elementos.

Como resumen este trabajo propone una nueva extensión a los repositorios actuales para poder manejar objetos y recuperarlos por su semántica y permite la unificación existente en los objetos, por medio de su conceptualización y no su sintaxis. Además se adecua a la Web semántica para su integración e intercambio, al integrar la ontología y abrir la puerta a más herramientas de explotación y agregar capacidades de recuperación. Una de sus principales ventajas fue el procesamiento autónomo de la meta-información y dar un ambiente colaborativo a agentes externos; así como el enlazado o mapeo a otras ontologías externas para el incremento de dominios de búsqueda y conceptualización.

2.3.5. Kashyap, V. & Sheth, A.

En este trabajo [33] se estudian los procesos de intercambio, así como compartir e integrar datos en especial para los GIS, por lo que se trabaja con la captura y representación de los metadatos en contextos y ontologías, con la finalidad de abstraer los detalles que representan los datos. La captura de información se realiza por dominio específico en sus metadatos; es decir, se utilizan los metadatos para la captura o llenado. Además se describe

la estructura y organización de los repositorios individualmente independientes de sus formatos de las múltiples fuentes de información.

Después de la fase anterior, es necesario ubicar el propósito de la consulta, utilizando los metadatos se construye el contexto y se identifican los datos relevantes en el mismo, se usan los términos (conceptos y roles) propios de la ontología que son definidos para inferir y además crear la descripción contextual o definir el contexto conceptualmente de la consulta; es decir, en realidad lo que se está buscando y a que entorno pertenece. Lo anterior es útil para cruzar varias ontologías de diferentes fuentes utilizando sus relaciones terminológicas

Para resolver el problema de diferente vocabulario en la representación de información similar, se explota el metadato/contexto usando los términos del dominio específico de la ontología para construir este nivel, el cual trata con la heterogeneidad, proponiendo relaciones terminológicas para lograr interoperabilidad semántica.

El artículo presenta la definición y clasificación de metadatos y muestra la importancia del cómo representar independientemente los detalles de los datos; es decir, los que no están asociados a la información, como por ejemplo la localización de un objeto. Después se especifica el contexto conceptualmente, mediante la descripción de los metadatos para hacer una representación y uso de ellos. La representación contextual brinda los beneficios de representación, razonamiento, mantenimiento y flexibilidad semántica, todo esto en los sistemas basados en conocimiento que modelan y representan el contexto.

Posteriormente se desarrolla un conjunto de inferencias, usando descripciones contextuales para habilitar la interoperabilidad semántica entre las bases de datos. Aquí es donde se crea un esquema para mapear estas descripciones contextuales usadas en las consultas, y así recuperar la información relevante, dando más importancia a ciertos parámetros, dependiendo donde se encuentra proyectada la consulta.

El esquema general de recuperación muestra un procesador de consultas que obtiene de la ontología su mapeo de contextos y los repositorios; todo lo anterior como un conjunto de nodos. Cada nodo se enlaza a otros por su procesador de consultas y todos ellos usan un componente llamado IRM (Inter Ontologies Terminological Relationships), nodo que se usa como unificador de nodos del cual el procesador de consultas gestiona la partición del usuario se expande la búsqueda a los demás nodos.

Como resumen el trabajo nos muestra una robusta estructura utilizando metadatos, identificando el contexto de las consultas y relaciones terminológicas para la interoperabilidad semántica. Además se proporciona una búsqueda distribuida en varias fuentes de datos, las cuales cuentan con su ontología que las describe y su mapeo en las bases de datos de los contextos.

2.3.6. Jones, C.B., et al.

El artículo [34] define una manera eficaz de hacer consultas geográficas, las cuales dependen de mapear la localización geográfica del usuario, de acuerdo con las relevancias de cada resultado. Para ello se necesita cierto nivel de semántica para poder llevarlo a cabo. Además, se exponen los actuales sistemas Web para brindar soporte a la información geográfica, del cual menciona (The Northern Light GeoSearch Tool from Vecinity (<http://www.northernlight.com/geosearch.html>)). Este proyecto permite al usuario encontrar lugares, dependiendo de este sistema se realiza una categoría de interés y es posible especificar un radio para encontrar otros lugares y poderlos mostrar en un mapa digital. A su vez, se explican las características del proyecto “Spirit”, el cual en su marco de trabajo presenta las siguientes propiedades:

- Ontología que modela la terminología geográfica.
- Técnicas de aprendizaje para la extracción de contexto de documentos Web y generación de metadatos para la obtención por contexto.
- Índices espaciales para acelerar la búsqueda.

El proyecto modela el vocabulario y la estructura espacial de los lugares mediante una ontología geográfica y además extiende los conceptos cuantitativamente y cualitativamente, por medio de las propiedades de un tesoro geográfico [35] para encontrar mejores semejanzas semánticas y poder *rankear* por nivel de interés los resultados.

En la interfaz de usuario de “Spirit” se proporciona información descrita explícitamente como la localización, relaciones espaciales y descriptores de datos. Se presenta un marco de trabajo donde se puede espacialmente extender su consulta en el mapa, seleccionar nombres de lugares, dibujar sobre el mapa regiones de interés, etc. La ontología creada elimina la ambigüedad, proporcionando una relevancia de la interpretación del sistema a información espacial imprecisa, en cuanto a dirección y proximidad.

Otro aporte del trabajo fue el desarrollo de técnicas para la extracción automática y codificación de metadatos de mapas digitales, explotando técnicas de geometría computacional, interpretación de imágenes y minería de datos geoespaciales. Como resumen, se explotaron debidamente los metadatos para la recuperación, además de agregar funcionalidades de localización espacial como proximidad y dirección. El proceso de automatización para la extracción de metadatos incrementó el dominio de búsqueda proporcionando mejores resultados.

2.3.7. Hiramatsu, K. & Reitsma, F.

Este trabajo [36] está relacionado con la Web semántica para encontrar información referenciada geográficamente en ella. El proyecto se enfoca en el desarrollo de una ontología geográfica escrita en OWL para describir a esta información y sus relaciones, además de crear dos herramientas para su explotación, una de ellas es un servicio Web que calcula las relaciones geográficas y la otra es un *plug-in* que permite crear archivos RDF2 de relaciones espaciales topológicas y de dirección.

Las ontologías permitieron relacionar información no espacial con otra que sí está geográficamente referenciada. De este modo, el proyecto creó dos ontologías, las cuales se encargan de describir las características geográficas de los objetos y sus relaciones geográficas las cuales fueron:

- Topológicas.
- Meteorológicas.
- De dirección.
- Distancia espacial.
- Distancia en tiempo.

Para hacer accesibles en Web las referencias geográficas, se implementó un repositorio y un mecanismo para calcular las relaciones dentro de un servicio Web. El primero almacena los enlaces de nombres y coordenadas para la recuperación de instancias espaciales por regiones geográficas. El segundo habilita al usuario las pruebas de relaciones basadas en coordenadas espaciales de los objetos geográficos, además como una ventaja, el formato de archivos que utiliza el *plug-in* GIS es el tipo *ShapeFile*.

Como resumen, el trabajo se enfoca en la recuperación de información referenciada geográficamente, vía la Web semántica, implementando una ontología, un repositorio de objetos, así como la creación de un *plug-in* GIS para el manejo de la información y un servicio Web encargado de establecer relaciones espaciales por sus coordenadas.

2.3.8. Visser, U. & Stuckenschmidt, H.

El trabajo [37] está enfocado en la recuperación de información geográfica, mediante un sistema de información llamado BUISY, el cual se considera inteligente porque utiliza una ontología para brindar un acceso por Internet e Intranet. Su principal enfoque se orienta en la explotación vía cómputo móvil.

El sistema de información utiliza información de impactos ambientales como los de agua subterránea en su adquisición y análisis, pero el aspecto más importante es la localización espacial de los objetos geográficos, mientras que la ontología automáticamente lanza una adquisición y una recuperación de información del sistema dada una localización. Este proceso se lleva a cabo mediante una comunicación inalámbrica, conectándose a un sistema de información geográfica y un sistema de posicionamiento global.

El sistema utiliza el modelo de un GIS basado en campos; es decir, el usuario está habilitado para acceder a ciertos recursos dependiendo del sensor GPS que indique su posición.

De esta forma, se habilitan y modifican ciertos objetos geográficos, éstos se vuelven disponibles y/o relevantes, los cuales se dividen en objetos geográficos y en dimensión temática. Otro problema que se resuelve es que se tiene diferentes fuentes de información por lo que se pueden usar diferentes conceptualizaciones y semánticos guardados en esos repositorios. La solución que se propone es la implementación de una ontología que tiene la capacidad de realizar tareas de recuperación inteligente, lo que ayuda a unificar semánticamente conceptos, los cuales pueden ser complejos en cuestión de varias interpretaciones del conocimiento, relacionando interrelaciones entre todos ellos.

Se utilizó el enfoque de Ontolingual para el desarrollo de herramientas, considerándose como un estándar, ya que habilita la traducción de la ontología específica en diferentes formatos. Para la recuperación se realizó una integración de fuentes de datos, el esquema fue el uso de la meta-información de las fuentes (metadatos) para la traducción de la

representación de una fuente en otra, obteniendo como resultado un modelo común que puede ser accedido por un método único.

Como resumen, la recuperación se realizó en diferentes fuentes, previendo una traducción de metadatos e integración de fuentes por medio de una ontología, la cual aportó la inteligencia de la búsqueda. La adquisición fue dependiente de la localización para registrar la posición en el sistema o dar un marco de trabajo, personalizando los recursos accesibles y la modificación de sus objetos geográficos relevantes a su interés.

2.3.9. Kemp, Z., et al.

En este trabajo [38] la línea de investigación es el control de áreas marinas como lo son las actividades de pesca, actividad industrial etc., y se enfoca en la integración, exploración y análisis interactivo de los datos en su sistema de aplicación. Su base de conocimiento captura semánticamente las componentes espacial, temporal y temática del dominio que se encuentra. El sistema ofrece los siguientes servicios de:

- Flexible extracción de fuentes heterogéneas, dependiendo de parámetros de usuario.
- Descubrimiento de datos en diferentes niveles de abstracción.
- Trabajo con los conjuntos de datos extraídos como es la reclasificación etc.
- Procesamiento de los datos para generar modelos.
- Visualización de la información en tablas, gráficas, etc.

La arquitectura empleada es una base de conocimiento para la búsqueda más avanzada. Está formada por metadatos, ontologías, taxonomías y especificación espacio-temporal, así como temática de los conceptos. Esta base proporciona servicios basados en conocimiento que interactúan con los servicios de análisis espacio-temporal y temático de la consulta y el servicio de recuperación de datos.

La petición de consulta se realiza en formato XML, la cual se transforma para proceder al módulo de análisis y recuperar los datos por localización, tiempo y temática para los científicos marinos. El razonamiento es realizado por los servicios basados en conocimiento sobre los dominios, mediante un repositorio de ontologías, las cuales describen las fuentes de datos con la definición de una capa de metadatos.

En resumen el proyecto se enfoca en la integración y recuperación semántica de diversas fuentes de datos con distintas estructuras, especificadas en los metadatos de cada repositorio. Se implementó un esquema de base conocimiento con sus servicios de entrada-salida para hacer un procesamiento de inferencia sobre las fuentes homogéneas en el dominio marino, con el propósito de crear un ambiente de trabajo basado en la consulta y recuperación semántica, procesamiento y explotación de datos a científicos especialistas.

Al agregar las propiedades espacio-temporales al procesamiento ontológico, se ha desarrollado una mejor integración y recuperación, agregando mayor precisión en los datos devueltos a la salida del sistema.

2.4 Integración semántica

2.4.1. Guarino, N.

Este artículo [39] proporciona un marco teórico para el diseño y uso de las ontologías, con el propósito de extracción y recuperación de información. La búsqueda debe tener un orden de relevancia. Se muestra el problema de la unión semántica de las búsquedas textuales en lenguaje natural, a texto definidas por el usuario. La manera en cómo las ontologías funcionan como mecanismo de unión en la recuperación; es decir, teóricamente describen el significado de vocabularios compartidos y su extracción desde el lenguaje natural.

En el artículo se describe el concepto de ontología como se realiza la conceptualización. A su vez, se define la ontología como una base particular de conocimiento y el modo en que se realizan inferencias, de lo que el usuario en su propia conceptualización desea recuperar, muestra los tipos de ontologías y su propósito. También se presenta el diseño herramientas de construcción de las mismas.

Con respecto a la recuperación se describe OntoSeek Project, el cual utiliza una ontología lingüística, haciendo la traducción del lenguaje natural que se usa para la búsqueda. En ella, se recupera y re-usan componentes de software orientado a objetos. Así mismo, en este proyecto se eligió usar la ontología SENSUS [40] para no partir de cero y solo hacer la traducción al lenguaje natural de los parámetros de la recuperación, esta ontología cuenta con 50,000 nodos para explotar la reusabilidad.

En primer lugar una codificación de la frase a buscar en términos de relaciones y conceptos de las palabras para posteriormente realizar la consulta en una base de datos de información previamente codificada, por un proceso de unión resultante de una ontología. Además se realiza un llenado de la estructura léxica de lo que se desea recuperar para lograr un mejor rendimiento de las preguntas y los datos. En el llenado se realiza un grafo léxico conceptual, en donde los nodos y arcos etiquetados con las palabras que forman la frase, son utilizados para poder hacer una traducción hacia la ontología. En el proceso de búsqueda la consulta es representada por un grafo de palabras, con la finalidad de ser comparado con la ontología y estos elementos del grafo son buscados en la base de datos en el orden definido previamente.

Como resumen, las ventajas de la recuperación léxica es una simple representación del lenguaje, pero semánticamente riguroso, una variada gama de términos para codificar y recuperar las frases o construcciones a buscar y además se agrega la posibilidad de comprobar la semántica conducida por una ontología. Este método se puede implementar en sistemas basados en esquemas con un número definido de descriptores; es decir, en cierta forma no son homogéneos los repositorios, pero además tienen la posibilidad de superar esto, debido a la flexibilidad con que se cuenta y así describir repositorios grandes heterogéneos.

2.4.2. Bishr, Y. & Kuhn, W.

Este artículo [41] muestra la forma en que la información geoespacial puede ser modelada bajo una ontología, lo cual es la base para brindar la capacidad a la recuperación de objetos espaciales, mediante su semántica. Este tipo de modelado proporciona más fundamentos cognitivos en los sistemas de información y minimiza en cierto grado los problemas de heterogeneidad semántica. Además se discute sobre el papel que juega el modelado de información y las teorías necesarias para el desarrollo de ontologías en aplicaciones geoespaciales.

El trabajo describe varios enfoques de autores tales como [17], [42] y [43], los cuales han desarrollado formas de identificar y unir objetos semánticamente similares en bases de datos, y se ha visto que el mapeo funciona en un *nivel extensional* en el conocimiento que describe. Por tanto, se necesita un *nivel intencional* para su correcto mapeo extensional en las bases de datos. Con lo anterior, varios modelos conceptuales que son intencionales por definición, no pueden contener suficiente conocimiento y así no son ricos semánticamente

para proveer una interpretación única a un objeto y hacer de alguna manera inferencia en ellos por un insuficiente significado de éste.

Así, el rol de las ontologías para compartir información es proveer un meta-lenguaje que puede ser usado para describir palabras de una realidad dada [17, 44, 45, 46], y además controlar la información a compartir, garantizando que no exista ambigüedad. La ontología proporciona una especificación al vocabulario, palabras en algún dominio y cubre múltiples y desconocidos sistemas de información; así se pueden construir modelos conceptuales que sirvan para diferentes propósitos pero con ontologías parecidas como su referencia. De este modo si los modelos conceptuales tienen un parecido fundamental en una ontología, ésta puede ser simplificada.

En resumen, el artículo describe un método conceptual, en donde la ontología debe ser enfocada a proveer de manera suficiente el conocimiento; de acuerdo con su intención. Esto debido a que una misma ontología puede ser interpretada de formas diferentes y crear instancias iguales, lo cual ayudará en la recuperación más consistente y no ambigua de los objetos geográficos.

2.4.3. Bernard, L., *et al.*

Este trabajo [47] describe cómo las infraestructuras de datos espaciales (SDI) y servicios distribuidos resuelven problemas relacionados con la heterogeneidad sintáctica más no con la semántica. Con base en lo anterior, los autores ha desarrollado el proyecto de investigación denominado: Interoperabilidad Semántica por medio de Geo-servicios (Semantic Interoperability by means of Geoservices).

Los autores proponen un caso de estudio orientado al manejo de desastres, el cual se encuentra implementado con una SDI para distribuir el procesamiento de la información geográfica, mostrando las debilidades de las SDI para la búsqueda por medio de su semántica. En este caso de estudio se argumenta que la adquisición procesamiento y análisis es vital para determinar que los problemas causados por la heterogeneidad semántica se deben a que se cuentan con muchas fuentes de datos, las cuales proveen información al manejador, tales como conceptos: nivel de agua, red vial, etc. En otras palabras, se pueden tener diferentes sintaxis, o dependiendo de su uso en el modelado se describen los eventos y los significados son adaptados a su aplicación.

En las SDIs se cuenta con la componente principal que es el (servicio GI), el cual debe lograr la interoperabilidad sintáctica, además de implementar la semántica. Asimismo, se tienen los catálogos y sus metadatos que también son fundamentales para permitir al usuario encontrar y acceder a sus recursos (datos y servicios) disponibles en los servidores. Además, se muestra que en el catálogo se describen los metadatos y operaciones sobre ellos para adquirir nuevos metadatos de otros proveedores de servicios. Con lo anterior se estandarizan los datos geográficos, por tanto, la búsqueda e interpretación se vuelve heterogénea; así como los vocabularios y el grupo de usuarios dirigido.

En el caso de estudio se realizó la búsqueda al proyectar los diferentes recursos (de percepción remota, bases de datos, servicios, entre otros) a una aplicación *web-mapping*. En la solución se establece que es necesario el dominio y los metadatos para poder tener un contexto de lo que se desea recuperar. Su enfoque fue el de búsqueda inteligente y traducción semántica de datos y metadatos. La búsqueda se realizó por medio de una ontología con términos, para que los datos de un término puedan ser traducidos de un vocabulario a otro y hacer un mapeo ontológico; es decir, que sean iguales en definición.

La ontología define relaciones de conceptos para mejores y más amplios resultados. La búsqueda inteligente se hace en dos partes: búsqueda por servicios y por datos, ambos dependen de los metadatos que también sean definidos por la ontología y sea más evidente la conceptualización. Del mismo que con los datos y servicios, los metadatos también son traducidos semánticamente por la ontología o en su caso se hace una fusión de ellos para hacer una extensión por significados en varias fuentes de datos.

Como resumen, el trabajo se enfoca en la búsqueda inteligente mediante ontologías; Además de una traducción semántica de datos, servicios y metadatos. Asimismo posibles fusiones (integración semántica) se proponen extender y mejorar las recuperación en infraestructuras de datos espaciales.

2.2.12. Manoj, P. & Ghosh, S.K.

Este trabajo [48] está orientado hacia una mejor integración, y tareas para compartir datos geoespaciales; así como la forma de poder consultar la información dentro del ámbito de la heterogeneidad semántica que poseen. Se enfoca principalmente en la integración de diversos repositorios de datos geoespaciales para aplicaciones geográficas, mediante una arquitectura orientada a servicios Web.

En la metodología se hace el descubrimiento y la recuperación por medio de una ontología central como información de metadatos que sirve como un servicio, la aplicación del proyecto se ejecuta en un entorno basado en servicios, uno Web y otro de mapas. Asimismo, se implementó la arquitectura SOA (Service Oriented Architecture) por métodos basados en servicios Web, debido a que muchas organizaciones integran y comparten información de diferentes tipos de actores que proveen datos como resultados de sus procesos. Además de la comunicación que necesitan entre ellos para lograr la colaboración, se agregó un gran rango de posibilidades contemplando el dominio geoespacial y se vuelve más importante por la emergente Web geoespacial semántica.

Para permitir a un servicio consumidor encontrar y acceder a recursos en diferentes repositorios se diseñó un catálogo de metadatos para construir una ontología híbrida. La semántica de cada fuente se describe por una ontología de aplicación propia, la cual colabora con un vocabulario compartido global, donde los miembros de un dominio le permite entender conceptos de otros. Este vocabulario es mapeado en una ontología de dominio, en un nivel superior que conforma la parte medular de la arquitectura SOA implementada. Para registrar información de la fuente, los metadatos fueron descritos por un XML para hacerse relativamente compatible con los estándares de la Web.

El proceso de búsqueda semántica se divide en dos servicios: el de *descubrimiento* y el de *recuperación*. En el primero se decide cual de las fuentes registradas debe ser usada para responder a la petición, buscando en cada una de sus ontologías de aplicación, a través del mapeo de conceptos en un vocabulario compartido para hacer una correspondencia al dominio. El proceso anterior es hecho por medio de un razonador terminológico, en este caso RACER (Reasoner for A-Boxes and Concept Expressions Renamed) [49]. En el segundo se cuenta con un módulo capaz de hacer las traducciones de la consulta del usuario a otras subconsultas, refiriéndose directamente al esquema donde están los datos del recurso. De esta forma, en la recuperación para este propósito se implementó una ontología central que es un módulo de procesamiento de la consulta que da el servicio tomando la función de intermediario, integrando y combinando datos entre repositorios. Esta ontología se encuentra sustentada en metadatos para interpretar los resultados de la búsqueda, en donde los datos fueron recuperados de diferentes catálogos distribuidos. En el esquema SOA la ontología central proporciona un servicio de descripción para devolver la consulta. Además publica el servicio para que pueda ser encontrado por los usuarios y se les envíe sus peticiones, en caso de no ser deseada se interactúa directamente sobre las fuentes de datos.

Como resumen, el trabajo se enfoca a la interoperabilidad, integración e intercambio de información semánticamente en un ambiente empresarial vía Web orientado a servicios. Su esquema de modularidad de ontología de aplicación para cada repositorio y uno de dominio, englobando a todas ofrece opciones de crecimiento para enriquecer los vocabularios, se puede hacer la recuperación distribuida mediante la ayuda de la ontología central, integrando metadatos y así interpretar la búsqueda.

2.5. Comentarios generales

En el estado del arte se han analizado y descrito varios enfoques que se han utilizado para la recuperación de información en distintas fuentes de datos, con el objetivo de resolver el problema de la heterogeneidad semántica.

Algunos de estos trabajos han explotado los metadatos, diseñando descriptores de la información contenida en los repositorios. Otros proyectos han propuesto el uso de las ontologías, pero el común de estas investigaciones radica en la búsqueda inteligente, de manera que se han desarrollado modelos para recuperar un objeto o concepto de varias fuentes.

La opción que mejores resultados ha proporcionado fueron aquellas que utilizan una representación conceptual, como es el caso de las ontologías, ya que los dominios han sido conceptualizados, describiendo explícitamente las relaciones entre conceptos y propiedades de éstos.

Asimismo, se observó la posibilidad de mejorar algunos aspectos de relaciones y extensiones ontológicas, con el fin de mejorar el nivel de especialización a cierta aplicación, utilizando sus propios metadatos para describir su información. Además se agregaron más componentes que ayudan a mejorar la inferencia y hacer las búsquedas de elementos de forma más precisa y por métodos cualitativos y cuantitativos.

Capítulo 3. Marco Teórico

3.1. Introducción

En este capítulo se describe el marco teórico del trabajo de investigación de la tesis, el cual se concentra en proporcionar las definiciones y teorías apropiadas que se utilizan como herramientas para el desarrollo de la tesis, tales como ontología, instancia, concepto, conceptualización, relación, clase, restricción, entre otras.

De igual forma, se describen un conjunto de definiciones y conceptos nuevos que han sido generados, como parte del marco conceptual del trabajo de tesis para la creación de ontologías de dominio. Asimismo, se puntualiza en un modelo formal para el desarrollo de estas ontologías; así como la definición del término esquemas conceptuales, con lo cual se sientan las bases necesarias para utilizar todas estas herramientas en la parte metodológica.

3.2. El formato Shapefile

El formato *Shapefile* (SHP) es un formato de archivo abierto de datos espaciales desarrollado por la compañía ESRI [50], quien crea y comercializa software para sistemas de información geográfica como Arc/Info o ArcGIS. Originalmente se creó para la utilización con su producto ArcView GIS, pero actualmente se ha convertido en formato estándar de facto para el intercambio de información geográfica entre sistemas de información y bases de datos geográficas. En otras palabras un shapefile es un formato vectorial de almacenamiento digital, donde se guarda la localización de los elementos geográficos y los atributos asociados a ellos. El formato carece de capacidad para almacenar información topológica.

Un shapefile [51] guarda la geometría e información atributiva para las características espaciales para un conjunto de datos. La geometría se almacena por coordenadas de vector. Como se mencionó anteriormente, los shapefiles no permiten almacenar explícitamente la estructura topológica de los datos geoespaciales. Sin embargo, tienen ventajas sobre otras fuentes de datos como la velocidad de dibujo y facilidades de edición.

Un shapefile está compuesto por varios archivos. El número mínimo requerido es de tres y tienen las extensiones siguientes:

- **.shp** - es el archivo que almacena las entidades geométricas de los objetos.
- **.shx** - es el archivo que almacena el índice de las entidades geométricas.
- **.dbf** - es el archivo que almacena la información de los atributos de los objetos en formato dBASE.

Además de estos tres archivos requeridos, opcionalmente se pueden utilizar otros para mejorar el funcionamiento en las operaciones de consulta a la base de datos, información sobre la proyección cartográfica, o almacenamiento de metadatos. Estos archivos son:

- **.sbn** y **.sbx** - Almacena el índice espacial de las entidades.
- **.fbn** y **.fbx** - Almacena el índice espacial de las entidades para los shapefiles que son inalterables (solo lectura).
- **.ain** y **.aih** - Almacena el índice de atributo de los campos activos en una tabla o el tema de la tabla de atributos.
- **.prj** - Es el archivo que guarda la información referida a sistema de coordenadas.
- **.shp.xml** - Almacena los metadatos del shapefile.

3.3. Ontologías

Una ontología define los términos que se usan para describir y representar un cierto dominio, entendido éste como un área específica de interés o un área de conocimiento [52]. Para que la Web semántica sea una realidad, la tecnología a utilizar son las ontologías. Por tanto, si la información es catalogada con las ontologías, al hacer las búsquedas de los recursos en la Web, tales como páginas, documentos, sonido o imagen, se utiliza el significado de las palabras como *entes*, en lugar de palabras clave.

Con la información organizada por medio de ontologías, serían mucho mejores la búsqueda y la integración de datos, debido al recurso de la interpretación del *significado*, que correspondería a los agentes de software.

Las ontologías son vocabularios comunes para los usuarios y las aplicaciones que pertenecen al campo de la Inteligencia Artificial (IA). Éstas agregan el conocimiento que cada persona utiliza mediante estas representaciones con las que “se representa y se entiende el mundo que nos rodea”. Las ontologías se consideran explícitas porque los significados son comúnmente conocidos y por lo tanto no es necesario tenerlas plasmadas

en un documento, además de que no están organizadas jerárquica o matemáticamente en el cerebro de las personas.

Al igual que los seres humanos, las máquinas carecen de ontologías explícitas, pero en su caso no comprenden el entorno y no pueden comunicarse entre sí. Si se pretende que las palabras sean procesadas por las máquinas, es necesario manejar las ontologías en forma explícita; es decir, desarrollarlas formalmente, mediante un vocabulario específico o en “una forma que sea legible para las máquinas”. En el desarrollo de ontologías explícitas debe contemplarse como mínimo un listado de términos, con el significado de cada uno. De esta manera dos sistemas de información podrían interactuar, gracias a la ausencia de problemas semánticos.

Debido a que las ontologías almacenan conocimiento, en una Web semántica sería posible extraer información automáticamente y procesarla, como lo haría un agente de software cuando un usuario tiene el propósito de buscar, por ejemplo, una impresora en razón de capacidad y calidad. Con las ontologías de carácter comercial, posteriormente, se tendría que escoger un establecimiento para su adquisición y elegir el precio más conveniente. En resumen, las ontologías:

- Favorecen la comunicación entre personas, organizaciones y aplicaciones.
- Permiten la interoperabilidad entre sistemas.
- Facilitan el razonamiento automático.
- Contribuyen a especificar los sistemas de software.

El área de la investigación referente a las ontologías ha tenido un incremento en estos últimos años, debido a que son muchas las áreas en donde es de gran utilidad el contar con una de ellas, y como muestra de ello se encuentra el área de la Inteligencia Artificial, la Lingüística Computacional y la Teoría de Base de Datos y las Ciencias de la Información Geográfica [17].

Para el desarrollo de una ontología, se parte teniendo como punto de inicio la base del conocimiento que el ser humano posee y con base en ésta se realiza un análisis más minucioso para tener una conceptualización aceptable de las cosas. Razón por la cual a continuación se hace referencia a algunos trabajos relacionados con ontologías.

3.3.1. El enfoque de Gruber

En el trabajo de [38], se expone que el conocimiento representado formalmente está basado sobre una conceptualización: los objetos, conceptos, y otras entidades que se asumen existen en alguna área de interés y las relaciones que mantienen entre ellas [53].

Una conceptualización de manera resumida, es la vista simplificada del mundo que se desea representar para algún propósito. Entonces a partir de esta aseveración se puede definir a una ontología como una *especificación explícita de una conceptualización*. El término es tomado de la Filosofía, donde una ontología es una *cuenta sistemática de la existencia*. Para el área de la Inteligencia Artificial, lo que “existe” es lo que puede ser representado. Entonces se reafirma la definición mencionada anteriormente donde se dice: *una ontología es una descripción formal de los conceptos y las relaciones entre estos*.

En resumen, una ontología es un sistema de representación del conocimiento que resulta de seleccionar un dominio o ámbito del conocimiento, y aplicar sobre él un método con el fin de obtener una representación formal de los conceptos que contiene y de las relaciones que existen entre dichos conceptos. Además, una ontología se construye con relación a un contexto de uso. En otras palabras, una ontología especifica una conceptualización o una forma de ver el mundo, por lo que cada ontología incorpora un punto de vista. Además, una ontología contiene definiciones que proveen del vocabulario para referirse a un dominio. Estas definiciones dependen del lenguaje que se utilice para describirlas. Todas las conceptualizaciones (definiciones, categorizaciones, jerarquías, propiedades, herencia, etc.) de una ontología pueden ser procesadas por una computadora. Los beneficios de utilizar ontologías se pueden resumir de la siguiente forma:

- Proporcionan una forma de representar y compartir el conocimiento utilizando un vocabulario común.
- Permiten usar un formato de intercambio de conocimiento.
- Proporcionan un protocolo específico de comunicación.
- Permiten una reutilización del conocimiento.

Para evaluar el diseño de una ontología se necesitan criterios objetivos que estén fundamentados en el propósito del resultado esperado y basado en notaciones *a priori* de naturaleza o verdad de los objetos implicados en una ontología. Un ejemplo de estos criterios dados por Gruber son los siguientes:

1. **Claridad.** Una ontología debe de tener el significado de los términos bien definidos, es decir, las definiciones deben ser objetivas. Las ontologías se diseñan para requerimientos computacionales o sociales, la definición debe ser independiente del contexto.
2. **Coherencia.** Una ontología debe de ser coherente, esto es que las inferencias sean consistentes con las definiciones. Al menos, los axiomas que se definen deben ser lógicamente consistentes. La coherencia también debe aplicarse a los conceptos que están definidos de manera informal. Por ejemplo, los descritos con lenguaje natural. Si una oración que se puede inferir a partir de los axiomas contradice una definición o un ejemplo dado informalmente, entonces la ontología es incoherente.
3. **Extensión.** Una ontología se debe diseñar anticipándose en el uso del vocabulario que pueda llegar a ocupar; es decir, debe definir términos nuevos para aplicaciones especiales basadas en el vocabulario existente, de manera que no se requiera revisar las definiciones existentes.
4. **Reducir las tendencias en la codificación.** La conceptualización se debe especificar al nivel del conocimiento sin depender de una codificación a nivel de símbolo en particular. Las tendencias de codificación deben ser reducidas al mínimo, ya que los agentes que comparten conocimiento pueden ser implantados en diferentes sistemas y estilos de representación.
5. **Reducir el compromiso ontológico.** Una ontología puede requerir un mínimo de compromiso ontológico, el cual sea suficiente para soportar las actividades relacionadas con compartir el conocimiento. Debido a que el compromiso ontológico está basado en el uso constante de vocabulario, se puede reducir al mínimo especificando una teoría más débil (que se permite en la mayoría de los modelos) y definiendo solamente esos términos que sean esenciales para la comunicación de conocimiento consistente.

Las posibles aplicaciones y usos de las ontologías pueden ser las siguientes, no siendo éstas las únicas:

- Repositorios para la organización del conocimiento.
- Servir de herramienta para la adquisición de información.

- Servir de herramientas de referencia en la construcción de sistemas de bases de conocimiento que aporten consistencia, fiabilidad y falta de ambigüedad a la hora de recuperar información.
- Permitir compartir conocimiento.
- Posibilitar el trabajo cooperativo al funcionar como soporte común de conocimiento entre organizaciones, comunidades científicas, etc.
- Hacer la integración de diferentes perspectivas de usuarios.
- Permitir el tratamiento ponderado del conocimiento para recuperar información de forma automatizada.
- Posibilitar la construcción automatizada de mapas conceptuales y mapas temáticos.
- Permitir la reutilización del conocimiento existente en nuevos sistemas.
- Permitir la interoperabilidad entre sistemas distintos.
- Establecer modelos normativos que permitan la creación de la semántica de un sistema y un modelo para extenderlo y transformarlo entre diferentes contextos.
- Servir de base para la construcción de lenguajes de representación del conocimiento.

3.3.2. El enfoque de Guarino

En [17] se definen algunos conceptos relacionados con ontología, compromiso ontológico y conceptualización. En algunos casos, el término “ontología” es el nombre extravagante que denota el resultado de actividades familiares como el análisis conceptual y el modelado del dominio. En el sentido filosófico, podemos referir a una ontología como un sistema particular de categorías relacionadas para una cierta visión del mundo. Como tal, este sistema no depende de un lenguaje en particular.

La ontología de Aristóteles es siempre la misma, independientemente del lenguaje utilizado para describirla. Por otra parte, en su uso más frecuente en Inteligencia Artificial, una ontología se refiere a un artefacto de la ingeniería, constituido por un vocabulario específico usado para describir cierta realidad, además de un conjunto de suposiciones explícitas con respecto al significado previsto de las palabras del vocabulario. Este sistema de suposiciones tiene generalmente la forma de una teoría lógica de primer orden, donde las palabras del vocabulario aparecen como nombres de predicados unarios o binarios, respectivamente, llamados conceptos y relaciones. En el caso más simple, una ontología describe una jerarquía de conceptos conectados por relaciones de pertenencia; en casos más sofisticados, se agregan los axiomas convenientes para expresar otras relaciones entre los conceptos y restringir la interpretación prevista.

Con el fin de solucionar el problema de la interpretación de la ontología, se utiliza una conceptualización, una palabra muy usada en el ámbito de la IA. Dos ontologías pueden usar diferente vocabulario (por ejemplo, español e inglés) y compartir la misma conceptualización.

Una conceptualización se ha definido como una estructura $\langle D, R \rangle$, donde D es un dominio y R es un conjunto de relaciones relevantes en D . Esta definición fue utilizada por [38], quién definió a una ontología como "una especificación de una conceptualización".

Resaltando estas aclaraciones, se puede proceder a refinar la definición de Gruber para hacer más clara la diferencia entre una ontología y una conceptualización: Una *ontología* es una teoría lógica que explica el significado previsto de un vocabulario formal; esto es, su *compromiso ontológico* para una *conceptualización* en particular del mundo. Los modelos previstos del lenguaje lógico utilizados como vocabulario son restringidos por su compromiso ontológico. Una ontología indirectamente refleja su compromiso y el de la conceptualización subyacente, mediante la aproximación de los modelos previstos.

Es importante enfatizar que una ontología es dependiente del lenguaje, mientras que una conceptualización es independiente del lenguaje. De hecho en la IA, el término "ontología" se colapsa en esos dos aspectos, pero una separación clara entre ellos llega a ser esencial para tratar los asuntos relacionados con compartir, fusionar, y traducir ontologías, que en general implican múltiples vocabularios y conceptualizaciones.

Se puede definir una ontología de manera que exista una clara distinción entre los términos ontología y conceptualización. Así podemos clasificar a las ontologías según su exactitud para caracterizar la conceptualización. Hay dos maneras posibles en que una ontología puede estar más cercana a una conceptualización: desarrollando una axiomatización más rica, y adoptando un dominio más rico y/o un sistema más rico de relaciones conceptuales relevantes. En el primer caso, la distancia entre el conjunto de modelos de la ontología y el conjunto de modelos previstos se reduce. En el segundo caso, es posible (al menos en principio) incluir en el conjunto de relaciones conceptuales relevantes aquellas relaciones que caractericen un estado del mundo, ampliando en el mismo tiempo el dominio para incluir las entidades involucradas con tales relaciones.

Otra posibilidad de incrementar la aproximación de una ontología a la conceptualización consiste en adoptar un modelo lógico, con lo cual se permite expresar coacciones a través de mundos, o solamente refinando mundos como objetos ordinarios del dominio. Una ontología fina logra una especificación más cercana del significado previsto de un vocabulario (y por lo tanto puede ser utilizado para establecer consenso en relación a compartir dicho vocabulario, o a una base de conocimiento que utilice ese vocabulario); pero puede ser difícil de desarrollar, debido al número de axiomas y la expresividad del lenguaje adoptado. Lo anterior se debe a la representación que cuenta con información muy detallada y se utiliza también el término ontología de bajo nivel (*low-level ontology*).

Una ontología tosca, puede consistir de un conjunto mínimo de axiomas escritos en un lenguaje de expresividad mínima, para soportar solamente un conjunto limitado de servicios específicos, previstos para ser compartidos entre los usuarios que están de acuerdo con la conceptualización subyacente. En este caso, la representación de información es general y se conoce también como ontología de alto nivel (*high-level ontology*).

Las ontologías también son clasificadas de acuerdo con su dependencia a una tarea específica o a un punto de vista y a su nivel de generalización, como se muestra en la Figura 3.1.

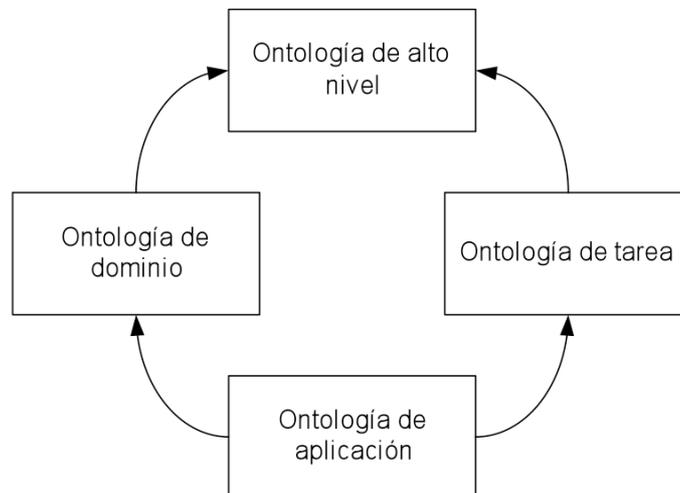


Figura 3.1. Tipos de ontologías de acuerdo con su nivel de dependencia para una tarea en particular. Las líneas representan relaciones de especialización.

Ontología de alto nivel (Top-level Ontology). Describen conceptos muy generales como espacio, tiempo, materia, objeto, eventos, acciones, etc., que son independientes de un

dominio o problema en particular; por lo tanto, parece razonable, por lo menos en teoría, tener ontologías a nivel superior unificadas para grandes comunidades de usuarios.

Ontología de Dominio (Domain ontology). Describe el vocabulario relacionado a un dominio genérico como medicina o automóviles.

Ontología de Tarea (Task ontology). Describe el vocabulario relacionado con una tarea en específico como diagnóstico o ventas, cada una especializando los términos introducidos en la ontología de nivel superior.

Ontología de Aplicación (Application ontology). Describe conceptos tanto de un dominio como de una tarea particular que frecuentemente son especializaciones de ambos. Estos conceptos corresponden a los roles desempeñados por las entidades del dominio mientras realizan cierta actividad, como unidad reemplazable o componente de repuesto. Asimismo, representan las necesidades del usuario con respecto a un uso específico, tal como una valoración de la contaminación en la Cd. de México [5].

3.3.3. Lenguajes de ontologías

En la década de los 90's, particularmente en el campo de la IA, surge un cúmulo de lenguajes para la creación de ontologías basados en paradigmas de representación del conocimiento. Se presentan lenguajes basados en marcos, redes semánticas y FOL (First Order Logic - Lógica de Primer Orden), para el modelado y el intercambio de conocimiento, los cuales pueden ser considerados como los precursores de los actuales lenguajes para la construcción de ontologías. Entre estos lenguajes [54] se tienen los siguientes:

- **KIF** (Formato para Intercambio de Conocimiento). Su modelo de conocimiento incluye lógica de primer orden, complementado con marcos. Es un lenguaje de bajo nivel pero de mayor poder expresivo a diferencia de sus sucesores como OKBC y Ontolingua. Fue diseñado para permitir transformaciones entre lenguajes de representación de conocimiento que fueran más especializados.
- **Ontolingua** (basado en KIF). Se convirtió en un estándar por la comunidad ontológica en la década de los noventa, incluye una perspectiva orientada a objetos, una extensión basada en marcos y una librería de ontologías.

- **LOOM.** Fue creado para construir bases de conocimiento en general y no ontologías. Está basado en lógica descriptiva, reglas de producción y provee clasificación automática.
- **OKBC** (protocolo para la Conectividad de Bases de Conocimiento Abierto). Este protocolo permite acceder a bases de conocimiento almacenadas en diferentes sistemas de representación de conocimiento y las cuales pueden estar basadas en diferentes paradigmas de representación de conocimiento.
- **OCML, Flogic** entre otros.

Con el auge de Internet, se generó un aumento de lenguajes para explotar las características de la Web. La Figura 3.2 presenta la relación que se establece, entre los lenguajes de marcado, orientados a mejorar la semántica de la Web. Se puede observar que todos ellos están basados en la capacidad sintáctica de XML.

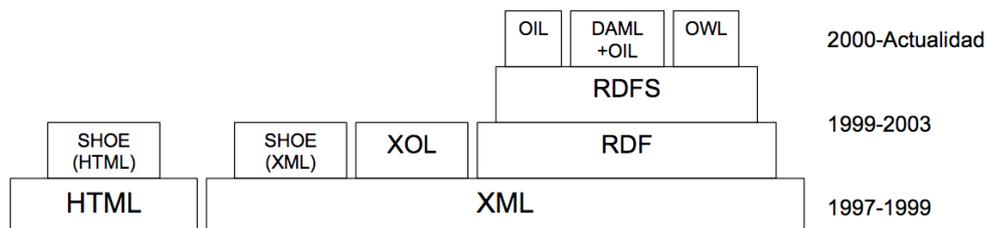


Figura 3.2. Lenguajes de marcado de ontologías.

- **SHOE** (Simple HTML Ontology Extension). Es un lenguaje de representación del conocimiento diseñado para la Web, basado en la asociación de contenido de páginas Web, vinculándolas con una o varias ontologías. Se trata de un anotador de contenido asociado a páginas Web, su modelo de conocimiento son los marcos y está basado en la sintaxis de XML, así como XOL y OML.
- **OML.** Fue parcialmente desarrollado con base en SHOE por la Universidad de Washington, y fue considerado una serialización de SHOE, puesto que comparten múltiples características. Su modelo de conocimiento son los grafos conceptuales.
- **XOL.** Lenguaje para intercambio de Ontologías basado en XML, fue desarrollado tomando como base OML y Ontolingua, por parte de investigadores del área de bioinformática en los Estados Unidos.

- **RDF** (Resource Description Framework). Es un marco para la descripción de recursos, su modelo de datos compuesto por: *Objetos* considerados como recursos definidas como expresiones RDF descritas mediante URI's y opcionalmente anclados a ID's. *Propiedades* que describen las características, relaciones y atributos o aspectos específicos de los recursos. *Instrucciones* que se utilizan para asignar un valor a una propiedad en un recurso específico. RDF se considera una base para procesar metadatos; proporciona interoperabilidad entre aplicaciones que intercambian información legible por máquina en la Web. RDF destaca por la facilidad para habilitar el procesamiento automatizado de los recursos Web. El modelo de datos RDF no hace ninguna aseveración sobre la estructura de un documento que contiene información RDF, permitiendo que las instrucciones puedan aparecer en cualquier orden dentro de una ontología. Tampoco provee primitivas de modelado para definir las relaciones entre propiedades y recursos. Esta limitación es solucionada mediante el lenguaje para describir vocabulario, conocido como RDF Schema.
- **RDF Schema**. Es un vocabulario utilizado para describir relaciones entre propiedades y clases de recursos RDF, con una semántica para la generalización y jerarquización tanto de propiedades como de clases. La combinación RDF y RDFS, es conocida como RDF(S). Las primitivas de RDFS son agrupadas en: clases, propiedades, contenedores de clases y propiedades, colecciones, vocabulario de refinación (transformación de algo abstracto en concreto) y propiedades de utilidad. RDF(S) provee las primitivas básicas, necesarias para modelar ontologías, existe un balance adecuado entre sus capacidades de expresividad y razonamiento. Su desarrollo se realizó buscando un núcleo estable que pudiera ser fácilmente extendido. RDFS es usado ampliamente por diferentes herramientas como Protégé, Mozzila y Amaya.
- **OWL** (Ontology Web Language) [55]. Es un lenguaje de marcado para publicar y compartir datos usando ontologías a través de la Web, añade más vocabulario para describir: clases (conceptos), propiedades, propiedades (de tipos de datos y de objetos), relaciones entre clases, individuos (instancias). Asimismo, contiene una serie de predicados predefinidos para la definición de ontologías tales como: clases equivalentes *SameAs* (igual que), definición de clases a partir de su descripción (enumeraciones {rojo, verde, azul}, restricciones {todos los individuos que tienen

más de dos hijos}, enunciados lógicos {persona que no es estudiante y tiene ojos azules}. OWL tiene mayor capacidad para expresar significado y semántica que XML, RDF, y RDF-S. De esta forma, OWL va más allá de estos lenguajes en su capacidad para representar contenido interpretable por una máquina en la Web. OWL proporciona tres sub-lenguajes, cada uno con nivel de expresividad mayor que el anterior, diseñados para ser utilizados por comunidades específicas de desarrolladores y usuarios.

- **OWL Lite.** Representa un pasaje para la migración desde otras taxonomías. Orientado a clasificación de jerarquías y restricciones simples. Se plantea que sea lo más simple posible para facilitar su desarrollo, a la vez que admite restricciones de cardinalidad, sólo permite establecer valores cardinales de 0 ó 1. OWL Lite proporciona una ruta rápida de migración para tesauros y otras taxonomías. OWL Lite tiene también una menor complejidad formal que OWL DL. En OWL Lite cada clase contiene: subclases, condiciones necesarias y suficientes para la clase, y expresiones de clase que incluyan el operador de intersección.
- **OWL DL.** Contiene los constructores del lenguaje pero con restricciones jerárquicas y de resolución (todos los cálculos se resuelven en un tiempo finito). OWL DL incluye todas las construcciones del lenguaje de OWL, pero sólo pueden ser usados bajo ciertas restricciones (por ejemplo, mientras una clase puede ser una subclase de otras muchas clases, una clase no puede ser una instancia de otra). OWL DL es denominado de esta forma debido a su correspondencia con la lógica de descripción (Description Logic, en inglés), un campo de investigación que estudia la lógica que compone la base formal de OWL. Se extienden las funciones de OWL LITE para incluir: aserciones que las instancias de clase no pueden compartir con la expresión de clase, definiciones de clase en forma extensiva, y expresiones de clase que incluyan operadores de unión, intersección o complemento.
- **OWL Full.** Posee un vocabulario completo interpretado más ampliamente que en OWL-DL, cuenta con un máximo poder de expresividad y libertad sintáctica y no ofrece garantías computacionales. OWL Full permite aumentar el significado del vocabulario preestablecido (RDF o OWL) de una ontología. Es poco probable que cualquier software de razonamiento sea capaz de obtener un razonamiento completo para cada característica de OWL Full. La elección entre OWL Lite y OWL DL depende de las necesidades de los usuarios sobre la expresividad de las

construcciones, proporcionando OWL DL las más expresivas. La elección entre OWL DL y OWL Full depende principalmente de las necesidades de los usuarios sobre los recursos del meta-modelado del esquema RDF (por ejemplo, definir clases de clases, o definir propiedades de clases). Cuando se usa OWL Full en comparación con OWL DL, el soporte en el razonamiento es menos predecible, ya que no existen en este momento implementaciones completas de OWL Full. Éste puede ser considerado como una extensión de RDF, mientras que OWL Lite y OWL DL pueden ser considerados como extensiones de una visión restringida de RDF. Cada documento OWL (Lite, DL, Full) es un documento RDF, y cada documento RDF es un documento de OWL Full, pero sólo algunos documentos RDF serán legalmente documentos OWL Lite o OWL DL. Por tal motivo, se debe tener precaución cuando un usuario quiera migrar un documento de RDF a OWL. Cuando se considere que la expresividad de OWL DL u OWL Lite es adecuada, han de tomarse algunas medidas para asegurar que el documento RDF original cumple con las restricciones adicionales impuestas por OWL DL y OWL Lite. OWL es un lenguaje de ontologías para la Web. Los lenguajes anteriores han sido utilizados para desarrollar herramientas y ontologías destinadas a comunidades específicas (especialmente para ciencias y aplicaciones específicas de comercio electrónico), no fueron definidos para ser compatibles con la arquitectura de la World Wide Web en general, y la Web Semántica en particular. OWL rectifica esto proporcionando un lenguaje que utiliza la conexión proporcionada por RDF para añadir las siguientes capacidades a las ontologías:

- Capacidad de ser distribuidas a través de varios sistemas.
- Escalable a las necesidades de la Web.
- Compatible con los estándares Web de accesibilidad e internacionalización.
- Abierto y extensible.

3.3.4. Editor de ontologías Protégé

Protégé [54] es una herramienta para el desarrollo de ontologías y sistemas basados en conocimiento, desarrollado en Java por la Universidad de Stanford. Las aplicaciones desarrolladas con Protégé son empleadas en la resolución de problemas y toma de decisiones en dominios particulares. La herramienta Protégé emplea una interfaz de usuario que facilita la creación de una estructura de *frames* (marcos) con clases, restricciones e

instancias de una forma integrada. Las características generales de Protégé se mencionan a continuación:

- Cuenta con una arquitectura de tres capas.
- La extensibilidad de la herramienta se basa en extensiones o *plug-ins*.
- Permite el almacenamiento de ontologías por medio de archivos OWL.
- Permite importar a lenguajes como RDF y XML.
- Permite exportar a lenguajes como XML, RDF, Flogic y CLIPS.
- La representación del conocimiento se basa en marcos, FOL y metaclasses.
- Proporciona un lenguaje de axiomas basado en PAL (Programmable Array Logic).
- Cuenta con un motor de inferencia nativo; así como el uso del motor REASONER.
- Soporte para el manejo de restricciones y revisión de inconsistencias.
- Mantiene la representación gráfica basada en una taxonomía.
- Cuenta con un mecanismo de poda gráfica, basado en restricciones y relaciones.
- Presenta un soporte para la generación de bibliotecas de ontologías.

3.3.5. Lenguaje de consulta SPARQL

En el marco de la recuperación y organización de la información, SPARQL (Protocol and RDF Query Language) [56] se define un lenguaje de recuperación para RDF/RDFS y también para OWL. Esta tecnología de consulta permite que los usuarios puedan centrarse en la información que requieren, sin tener en cuenta la tecnología de base de datos o el formato utilizado para almacenar a estos datos. Debido a que las consultas en el lenguaje SPARQL expresan objetivos de alto nivel, es fácil extenderlos a orígenes de datos inesperados, o incluso transferirlos a nuevas aplicaciones.

El lenguaje de recuperación SPARQL ha sido diseñado para un uso a escala de la Web, así permite hacer consultas sobre orígenes de datos distribuidos, independientemente del formato. Es más fácil crear una consulta sencilla y recuperar información en una sola consulta a través de diferentes almacenes de datos, que crear múltiples consultas, además de tener un costo menor y de ofrecer mejores resultados.

Debido a que SPARQL no está ligado a un formato de base de datos específico, puede ser utilizado para beneficiarse de la Web 2.0 y de la composición de éstos con otros recursos de la Web Semántica en las aplicaciones. Además, debido a que los orígenes de datos dispares

pueden no tener el mismo formato o compartir las mismas propiedades, SPARQL ha sido diseñado para consultar datos que no son uniformes.

La especificación de SPARQL define un lenguaje de consulta y un protocolo, y trabaja con el resto de las tecnologías esenciales del W3C de la Web Semántica: Infraestructura de Descripción de Recursos (RDF) para la representación de datos; RDF Schema; Lenguaje de Ontologías Web (OWL) para construcción de vocabularios. SPARQL también usa otros estándares del W3C existentes en las implementaciones de servicios Web, como Lenguaje de Descripción de Servicios Web (WSDL). En la Figura 3.3 se muestra el diagrama de paquetes de la especificación de SPARQL.

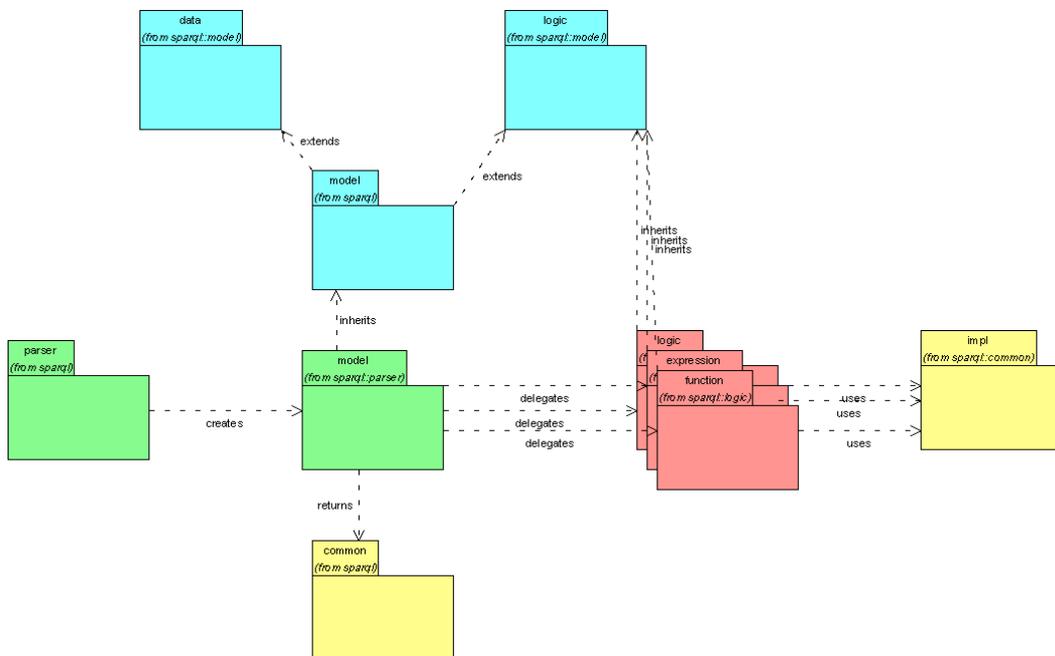


Figura 3.3. Diagrama de paquetes de SPARQL.

Por otra parte, en la Figura 3.4 se muestra la interfaz de objetos que maneja SPARQL. En ésta se puede observar que solo los objetos con los cuales interactúa un cliente pueden procesar una consulta. Estas interfaces definen las entradas y salidas de la biblioteca que sean las más semánticamente definidas para los objetos.

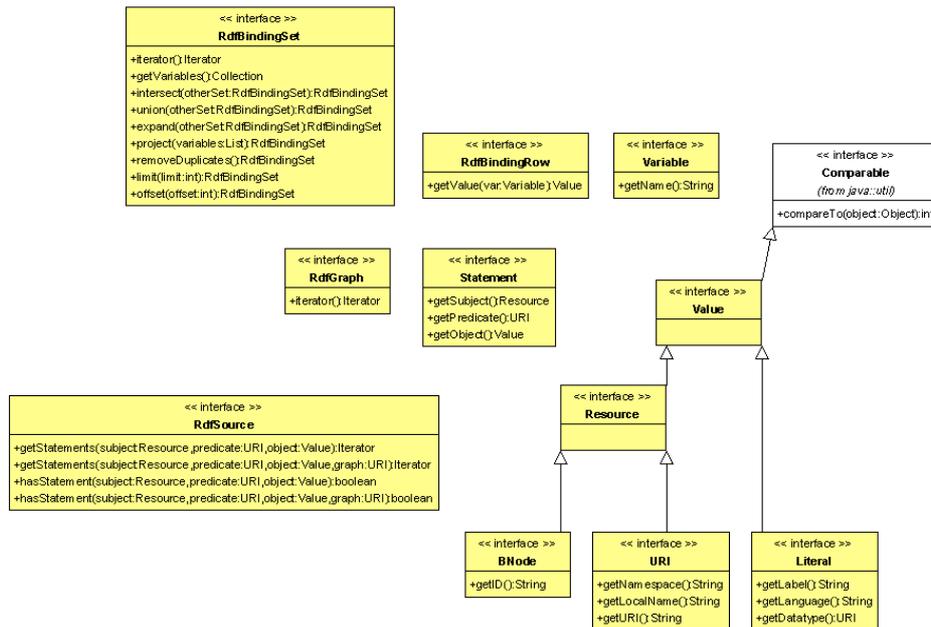


Figura 3.4. Interfaz de objetos de SPARQL.

El modelo de consulta de SPARQL prepara y evalúa las interfaces implementadas, así como todas aquellas interfaces que definen los métodos de acceso para la consulta de objetos. En la Figura 3.5 se muestra el modelo de consultas que maneja SPARQL.

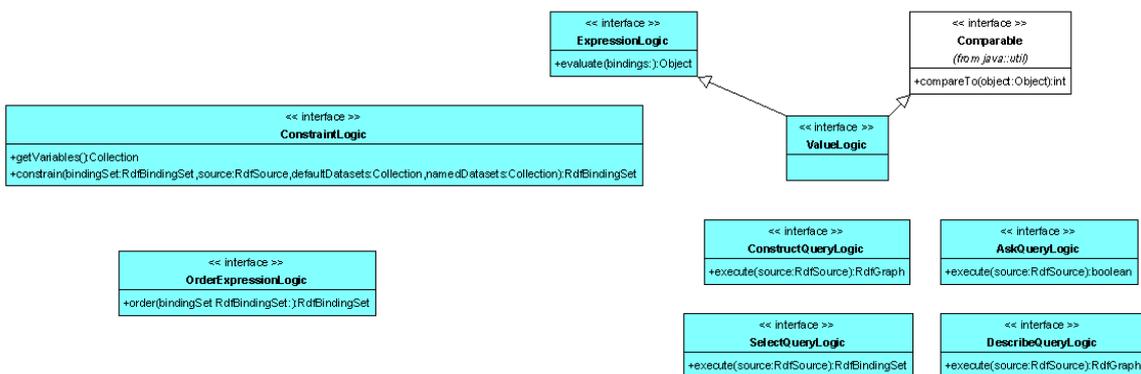


Figura 3.5. Modelo de consulta de SPARQL.

3.3.6. Marco de trabajo Jena para el manejo de ontologías

Es un marco de trabajo implementado en Java para construir aplicaciones de Web Semántica. Jena proporciona un entorno de programación de RDF, RDFS y OWL. Además de que utiliza SPARQL para la construcción de consultas e incluye un motor de inferencia basado en reglas. Las características más relevantes de Jena son las siguientes:

- API de RDF y OWL.
- Lectura y escritura de RDF en RDF/XML.
- Almacenamiento en memoria y persistente.
- Motor de consultas SPARQL.

Jena fue diseñado con la intención de dar soporte para OWL Full y para el razonamiento de casos no incluidos en el subconjunto de la sintaxis OWL DL. Jena permite trabajar con contenedores, además, incluye un convertidor sintáctico. Es capaz de procesar información RSS (*rdf site summary*), lo que la convierte en un módulo ideal para cualquier proceso destinado a ser utilizado en la Web.

Jena es una iniciativa de software abierto que cuenta con el apoyo de HP Labs Semantic Web Program. Su primera versión Jena 1 se centraba en grafos y su transformación a XML/RDF y en su siguiente versión (Jena 2) añade por encima una capa para tratamiento de ontologías y su transformación a OWL [55]. Jena permite además conectarse a razonadores por medio de DIG. Por tanto, Jena se puede considerar como una infraestructura para trabajar con la Web semántica programando con Java.

Para implementar el paso a OWL de la información disponible existen dos soluciones que son las más referenciadas a la hora de generar archivos OWL de forma automática cuando se dispone de una *TBox* ya definida:

1. Utilizar la API ofrecida por Protégé para el manejo de ontologías.
2. Utilizar la API de Jena.

Ambas API's están desarrolladas en Java, son *open source* y comparten muchas características comunes. La API de Protégé incluye mucha más complejidad, ya que permite realizar muchas más acciones. Para la interfaz con RDF/OWL la API de Protégé utiliza Jena para la transformación. Jena proporciona una API que permite leer y escribir en los formatos XML/RDF, N3 y N-Triples (formatos para grafos abreviados). Además dispone de un motor para poder buscar dentro de los documentos RDF por medio de RDQL (Lenguaje de Consulta de Jena).

Por último, ofrece también mecanismos para leer y escribir en formato OWL. Adicionalmente, Jena no se limita a sus propias soluciones, incluye componentes de tipo

plug-in e interfaces con otras soluciones externas, de manera de poder conectar a Jena con un razonador independiente si se requiere, por si los que incluye Jena no satisfacen las necesidades del usuario. En la Figura 3.6 se muestra el diagrama base del estándar de interfaz que define Jena para comunicarse con razonadores o motores de inferencia, el mismo es llamado DIG (*description logic reasoner interface*).

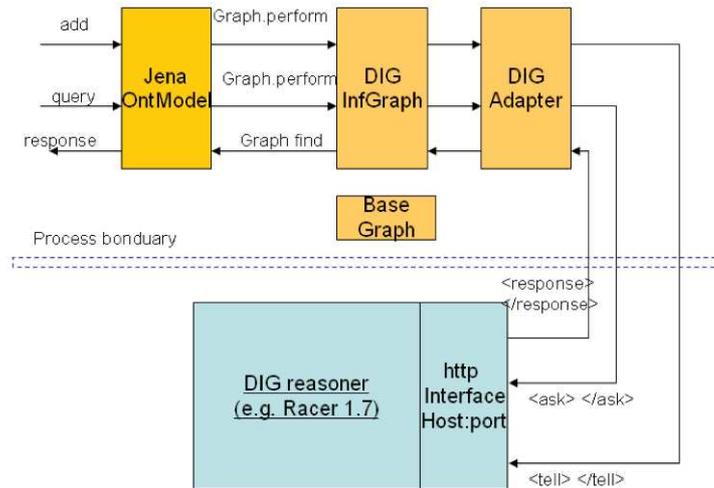


Figura 3.6. Interfaz de Jena para comunicarse con razonadores.

Por otra parte, Jena provee soporte para RDF y OWL, incluyendo el motor de SPARQL para consultas del tipo ARQ. Además, incluye una variedad de tipos de razonadores: *transitivo* basado en reglas RDFS, OWL y DAML y *genérico* para reglas. En la Figura 3.7 se muestra el modelo para el motor de inferencia de Jena.

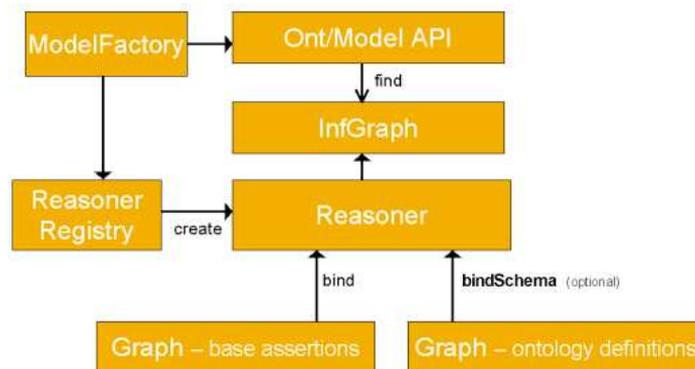


Figura 3.7. Modelo para el motor de inferencia de Jena.

Asimismo, Jena trabaja con la clase *Model* para acceder a la información del documento RDF. En el caso de ampliación del RDF a ontologías propiamente dichas (con clases,

instancias, propiedades, etc.), la clase *Model* se especializa, dando lugar a la clase *OntModel*; es decir, la ontología con la que está trabajando el sistema. Además trabaja con los conceptos (interfaces de Java para permitir polimorfismo), el cual se presenta en Figura 3.8.

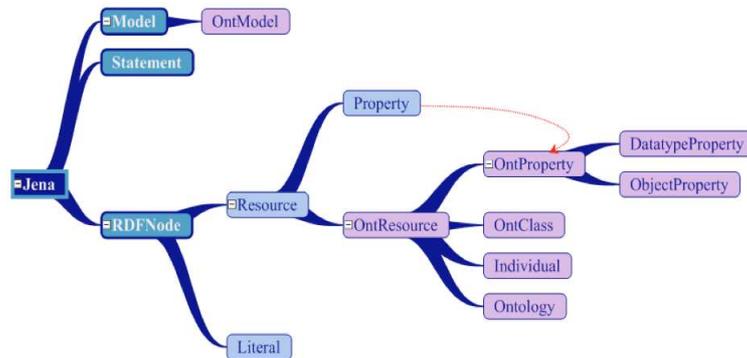


Figura 3.8. Herencia entre interfaces de Jena.

- **Resource.** Entendido como cualquier recurso definible en RDF. Es una interfaz de Java de forma que los siguientes conceptos son subclases de esta interfaz.
- **OntClass.** Representa una clase.
- **Individual.** Interfaz que representa una instancia de una clase.
- **Property.** Representa la propiedad de relación entre *Resources*. Se divide en *ObjectProperty* (para propiedades con un objeto de tipo instancia) y *DatatypeProperty* (para propiedades de tipo de datos).
- **Literal.** Es un *Resource* que representa un elemento de tipo de datos (representado por xsd:).
- **Statement.** Representa la frase que une un sujeto por medio de una propiedad con un objeto.

En la Figura 3.8 se observa la relación de herencia entre las distintas interfaces de Jena 2, siendo las de color azul las relativas a OWL y las de color morado las relacionadas con OWL propiamente dicho. Jena 2 aunque permite trabajar con ontologías y conectar razonadores que implementen una interfaz determinada, sigue estando muy orientado a RDF (procesamiento de nodos, enunciados, grafos, etc). Sin embargo, la importación es un punto poco importante, ya que la lógica de la ontología de clases (*TBox*) ha sido desarrollada por medio de otra aplicación y siempre será más rápido generar instancias usando RDF (lo que significa trabajar a un nivel de abstracción más bajo).

3.3.7. Motores de inferencia para ontologías

Las tareas de razonamiento son llevadas a cabo por los motores de inferencia también llamados *razonadores*, los cuales se definen como módulos de software que implementan algoritmos deductivos clásicos optimizados para el subconjunto de la **Lógica de Primer Orden** que delimita el nivel de la **Lógica de Descripciones** que aceptan. En este apartado se describirán los diferentes motores de inferencia existentes para trabajar en conjunto a una ontología, así como las características de cada uno.

Definición de razonador

Un *razonador* es básicamente, una herramienta que aprovecha la componente semántica de las ontologías, de forma que, aplicándose sobre éstas los razonadores o motores de inferencia puedan extraer el conocimiento que no está expresado de forma explícita en su construcción. El *motor de inferencia* será el que permita sacar conclusiones de la base de conocimiento y resolver el problema que se le ha planteado.

El motor de inferencia permite obtener nuevas sentencias (conclusiones) a partir de las sentencias de su base de conocimiento y determinar, por ejemplo, si una sentencia es deducible a partir de las sentencias de su base de conocimiento. También permite determinar si una determinada sentencia es imposible a partir de su base de conocimiento; es decir, lleva a una contradicción. Otro posible resultado es concluir que una sentencia no es imposible, pero no es deducible de la base de conocimiento. Estas herramientas son las que finalmente propician gran parte el potencial de las ontologías como herramientas de manejo de la información.

Es importante recordar que un lenguaje lógico es un lenguaje formal que se utiliza para representar información de la que se puedan sacar conclusiones, y dispone de una sintaxis (las sentencias válidas en el lenguaje) y una semántica (qué significan esas sentencias). Una **base de conocimiento** puede usar una ontología para especificar su estructura (tipos de entidades y relaciones) y su esquema de clasificación. Una ontología junto con un grupo de instancias de sus clases constituye una base de conocimiento.

¿Por qué emplear razonadores al trabajar con una ontología?

A la hora de definir una ontología se pueden encontrar tres partes principales:

- Clases e instancias que son de forma básica los objetos o elementos que la componen.
- Propiedades que representan las relaciones entre los anteriores.
- Reglas que se emplean para modelar el conocimiento que no puede recogerse a través de los elementos anteriores.

Si no se emplearan estos últimos elementos se estaría ante lo que se puede considerar como una ontología ligera, que no es más que una taxonomía o clasificación de elementos. La no utilización de reglas origina el problema de la poca capacidad expresiva del conocimiento.

De hecho, podría incluso establecerse una clasificación de las distintas tecnologías existentes para el tratamiento de la información, distinguiendo para cada una de ellas el nivel de semántica recogido. Así, por ejemplo las bases de datos relacionales, si bien resultan de gran utilidad, presentan un poder casi nulo en lo que se refiere a la **recolección de aspectos semánticos**.

En el caso particular de las ontologías, cuanto mayor sea el trabajo de definición de axiomas y relaciones, mayor será su potencial semántico. De esta forma, una ontología se convierte en algo más que en un establecimiento de categorías o subcategorías de conceptos. Es en este momento cuando el empleo de razonadores cobra gran interés.

Como se menciona en los párrafos anteriores el empleo de razonadores está motivado por el gran poder lógico o de deducción de conocimiento que presentan las ontologías. De esta forma, se puede decir que recogen más conocimiento del que realmente se presenta en un principio y éste puede extraerse, haciendo deducciones o respondiendo a consultas lógicas, mediante el empleo de razonadores.

Funcionamiento de un razonador

Todo este tema gira en torno al poder lógico que incluye una ontología, mejor dicho a la lógica descriptiva presente, cuyos fundamentos permitirán a un motor de lógica descriptiva hacer deducciones o responder consultas lógicas con base en los axiomas ya definidos en la ontología. El cómo este motor o razonador de lógica de descripción funciona, no es simple de explicar: Los razonadores trabajan con base en la *lógica de descripción*, ésta se refiere, a descripciones de conceptos usadas para describir un dominio y, por otro lado, a la semántica que puede recogerse en la lógica de predicados de primer orden. La lógica de

descripción (DL) se diseñó como una extensión de *frames* (marcos) y redes semánticas, los cuales no estaban equipados con semántica basada en la lógica. Lo nuevo que tiene la lógica de primer orden es que no está presente en la lógica proposicional es la **cuantificación**. La lógica de predicados de primer orden tiene suficiente poder expresivo para la formalización, considerándose aspectos como la cuantificación de los conceptos, por lo tanto, la realización de motores o razonadores basados en una lógica de descripción se hacen viables. En la figura 3.9 se muestra como opera un motor de razonamiento o inferencia empleando la lógica de descripción.

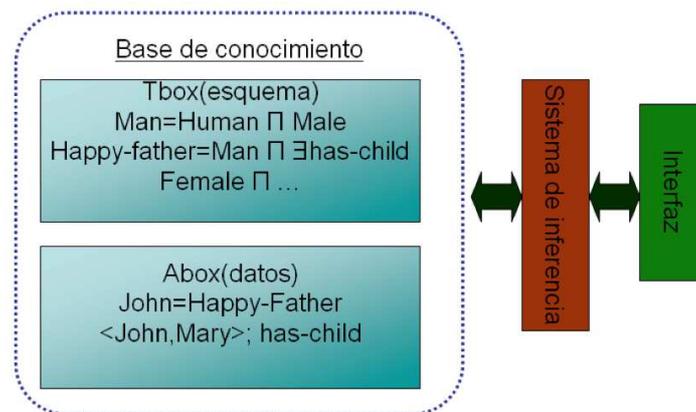


Figura 3.9. Operación de un motor de razonamiento.

Un **motor de razonamiento** basado en **lógica descriptiva**, asocia dos mecanismos internos en su entendimiento del conocimiento. El primero denominado *TBox* (caja terminológica) y un segundo llamado *ABox* (caja de aserciones). Esta separación es puramente operativa, ya que estas distinciones permiten a un razonador de lógica de descripción operar de mejor forma. En cuanto a los elementos mostrados que componen de forma básica un razonador se describe lo siguiente:

- *TBox* contiene sentencias describiendo conceptos jerárquicos; es decir, relaciones entre los conceptos.
- *ABox* contiene sentencias indicando a donde pertenecen los individuos en la jerarquía; es decir, relaciones entre individuos y conceptos.

En cuanto a la forma interna de razonamiento, este proceso es bastante complejo y se basa en la realización de inferencia que permiten con base en axiomas o reglas, realizar deducciones mayores o transitivas y de otros tipos.

Ventajas del trabajo con razonadores

En primer lugar es importante recordar que se parte de la idea de emplear las ontologías no como una forma más de clasificación de información, sino que se pretende explotar su potencial semántico, considerando la posibilidad de emplear su capacidad para el empleo de la lógica. Por tanto, y con base en los requerimientos anteriores, se hace necesario la indicación de reglas.

Una alternativa de recuperación de información son las consultas realizadas sobre una ontología, empleando para ello el lenguaje SPARQL. Sin embargo, cabe destacar que su propósito es distinto del perseguido, si se emplea un razonador. A través de las consultas se pretende obtener información previamente recogida de forma explícita, pero de forma que se acceda a la misma, de manera más rápida o eficaz. A diferencia de las anteriores, los razonadores pretenden obtener información nueva (nuevo conocimiento), empleando para ello la lógica.

El empleo de razonadores puede resultar muy útil en aquellos casos en que, una vez construida una ontología, se desea trabajar sobre la misma, bien comprobando su consistencia o llevando a cabo un proceso de razonamiento o inferencia sobre su información. Puede concluirse entonces que el empleo de razonadores es la mejor alternativa, dado que:

- No plantea problemas de escalabilidad y es más fácil de implementar para entornos tácticos y estratégicos.
- Permite inferir información de la recogida en una ontología.
- Permite comprobar la validez de la anterior.

En este punto es importante recalcar cual de los razonadores más utilizados para OWL (FaCT++, Racer, Pellet) y con interfaz DIG¹ (Description Logic Reasoner Interface) es el más adecuado para las ontologías que se utilizan y el esquema propuesto. Por esta razón se describirán las tres posibilidades. Por tanto, a continuación se detallan los tres razonadores y se elegirá el más adecuado para el propósito del presente trabajo.

Razonador FaCT++ 1.1.3

Desarrollado por la Universidad de Manchester, es la continuación del razonador *open source* para *TBoxes* FaCT (Fast Classification of Terminologies), es la versión antigua desarrollada en Lisp y la versión actual desarrollada en C++ para acelerar los algoritmos.

La nueva versión permite razonar con *ABoxes* y en las últimas versiones ha mejorado bastante su interfaz con DIG. FaCT++ está optimizado para la lógica de OWL DL. Los algoritmos utilizados son optimizaciones del algoritmo *Tableaux*.

Su instalación sobre Windows conlleva cierta dificultad al tener la necesidad de instalar la plataforma .NET 2.0, y no se ha encontrado información alguna sobre su instalación, manejo y mensajes de error mostrados por el razonador. No tiene el apoyo de ninguna comunidad de desarrolladores, por lo que se encuentra aún en una fase algo inestable.

Aunque no disponga de información de uso y ayuda, el servidor HTTP de DIG al menos muestra por su salida estándar bastante información relativa a las acciones que realiza y errores que ocurren. Su interfaz DIG permite ejecutarlo como servidor HTTP. FaCT ofrece además una arquitectura cliente-servidor basada en CORBA para acceder a las ontologías. En la Figura 3.10 se observa un ejemplo de utilización de FaCT con Protégé.

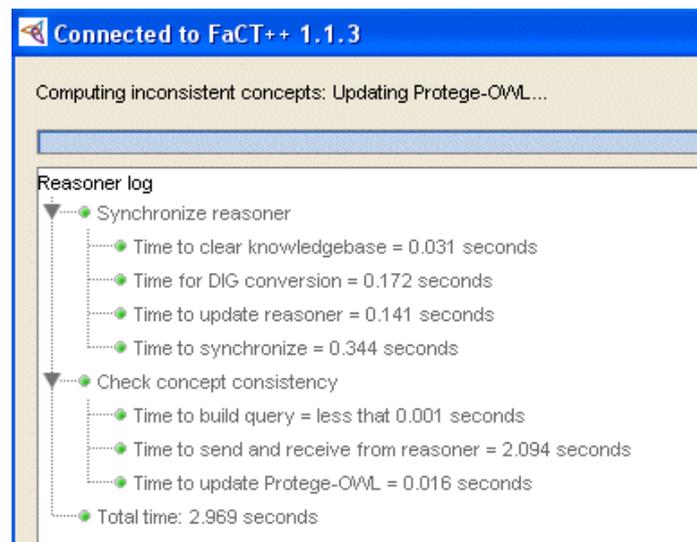


Figura 3.10. Uso de FaCT++ con Protégé.

Razonador Racer 1.9

Racer (Reasoner for ABoxes and Concept Expressions Renamed) fue uno de los primeros razonadores para *ABoxes* que apareció. Se desarrolló por las universidades alemanas de Hamburgo y actualmente es software propietario y de pago, aunque permite una descarga de prueba de tres meses. Racer 1.9 soporta OWL DL excepto para los nominales (clases definidas por una enumeración de sus miembros, que implementa como definiciones parciales) y para tipos de datos no estándar. Al contrario de las primeras versiones, por defecto no asume UNA (Unique Name Assumption), para ser compatible con OWL. Se basa en el algoritmo *Tableaux* con optimizaciones y cacheos de inferencias obtenidas.

Razonador RacerPro

Los orígenes de RacerPro corren parejos con los de las DL's. RacerPro se utiliza para manejar las ontologías basadas en OWL; es decir, como motor de razonamiento para los editores de ontologías. En concreto, RacerPro es el razonador por defecto de Protégé. Debe iniciarse previamente al editor, y presta su servicio a través del puerto de comunicación <http://localhost:8080/>. En la Figura 3.11 se observa la estructura de RacerPro.

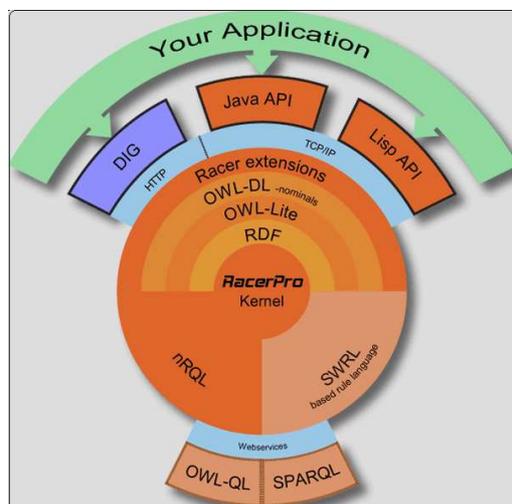


Figura 3.11. Estructura de RacerPro.

RacerPro es un razonador con más posibilidades que un simple razonador OWL, ya que incluye un cliente para hacer consultas en OWL-QL (RacerPorter), su propio lenguaje de consultas (nRQL, New Racerpro Query Language) e incluye una herramienta de cliente exclusiva RICE (Racer Interactive Client Environment), y una API en Java para poder

interactuar con él directamente (JRacer) [48]. Los servicios que presta para las ontologías escritas en OWL y RDF son los siguientes:

- Chequeo de consistencia de una ontología OWL y un conjunto de descripciones de datos.
- Encuentra relaciones de subclase implícitas.
- Encuentra sinónimos de los recursos (clases, instancias).
- Búsqueda incremental en las consultas.
- Implementa a DIG para interconectar la ontología con la interfaz.

RacerPro no tiene una interfaz de usuario. Su salida estándar es la consola, a través de ella se introducen comandos y se reciben los resultados. Las herramientas como Protégé lo utilizan como servidor de inferencias *back-end* pero no ofrecen funciones para controlar al propio razonador. Para solucionar este problema se ha creado RacerPorter, una herramienta gráfica de interfaz con el usuario. Para utilizarlo debe estar ejecutando RacerPro, al cual se conecta vía TCP/IP. En la Figura 3.12 se aprecia un ejemplo con las distintas funcionalidades de RacerPorter tales como cargar, visualizar y manejar las ontologías. Además se observa seleccionada la solapa **Taxonomy** (en la parte superior). En el panel central aparece el árbol gráfico de la taxonomía. En este ejemplo la consulta es por la **cajaT** y la **cajaA** y en el panel inferior aparecen las respuestas. La consulta se expresa de la siguiente forma: `ABOX-CONSISTENT? T (true)`, y `TBOX-COHERENT? T (true)`.

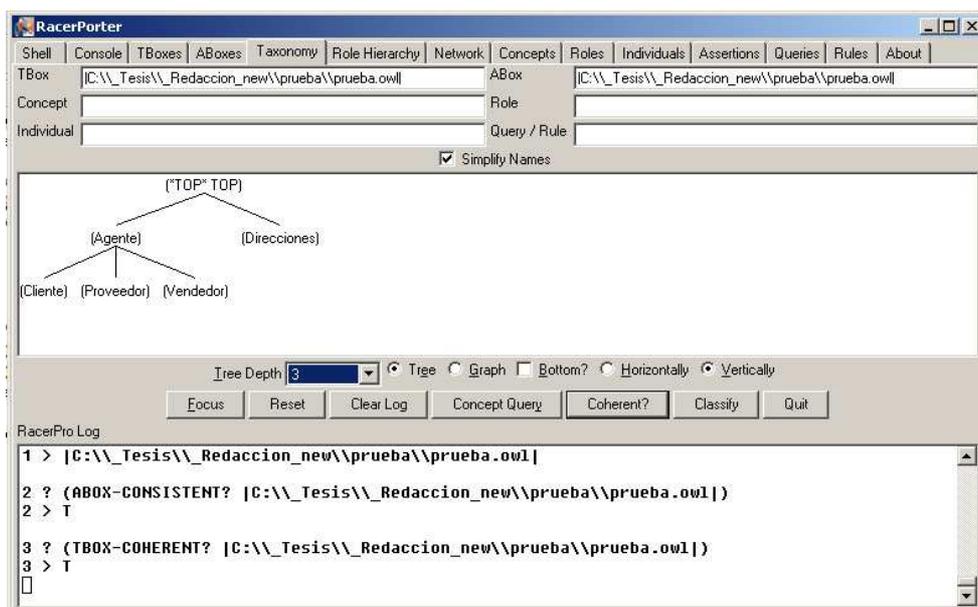


Figura 3.12. La interfaz gráfica RacerPorter.

Para las **TBox** soporta las siguientes utilidades basándose en la semántica del lenguaje de representación: consistencia de conceptos, determinación de padres e hijos (en las clases). Las utilidades sobre las **ABox** son las siguientes: chequeo de consistencia de la **ABox** respecto de la **TBox**; así como el chequeo de la instancias (a qué clase pertenecen).

Además, recupera las tuplas de aquellas instancias que satisfacen ciertas condiciones. Procesa los tipos directos de un individuo con respecto de una **ABox** y una **TBox**. Posee un lenguaje propio de consultas (**nRQL**), que soporta la negación como fallo, las restricciones numéricas, etc.

Pellet 1.3

Es un razonador *open source* desarrollado por Mindswap. Pellet es un razonador exclusivo para OWL DL y desarrollado en Java, el cual implementa la interfaz DIG y además RDQL para consultas sobre la información en RDF. Además, dispone de una API para poder ser utilizado directamente desde Java.

Como los otros dos sistemas, se basa en el algoritmo *Tableaux*. La implementación de DIG es la que presenta errores más aclaratorios para la corrección de ontologías y además es el razonador que más chequeos hace al principio, lo cual permite corregir errores en la ontología directamente. Además, al recibir una ontología, realiza optimizaciones internas, de forma que las siguientes consultas sean más rápidas. Es el único que se presenta como una Web, donde se puede enviar el archivo OWL para validarlo. Los servicios de razonamiento de Pellet se describen mediante las siguientes funcionalidades:

- **Comprobación de la consistencia.** Se encarga de comprobar que no existen contradicciones en la ontología. La semántica de OWL define una especificación formal para la definición de la consistencia en una ontología empleando Pellet. En terminología DL esta operación consiste en chequear la consistencia de la **ABox** respecto de la **TBox**.
- **Corrección de los conceptos.** Verifica si es posible que se definan instancias para una clase; si no es así, el definir una instancia para una clase provocaría que toda la ontología fuera inconsistente.
- **Clasificación.** Observa la relación entre cada clase y crea la jerarquía de clases completa.

- **Realización.** Encuentra las clases más específicas a las que pertenece una instancia; en otras palabras, determina la clase a la que pertenece cada uno de los individuos. Esta operación sólo se puede realizar después de la clasificación, puesto que los tipos directos se definen respecto de la jerarquía de clases.

Comparación de las características entre los razonadores

Como se puede observar en la Tabla 3.1, se muestra una tabla comparativa entre los razonadores, todos ellos, implementan la interfaz DIG, por lo que se podrá utilizar dicha interfaz a la hora de implementar un prototipo que pueda ir cambiando de razonador.

Tabla 3.1. Comparación de características entre razonadores.

Característica	FaCT++	Racer 1.9	Pellet 1.3
Evaluación Objetiva			
Multi-plataforma	No	No	Sí
Código abierto	Sí	No	Sí
Software libre	Sí	No	Sí
Implementación DIG	Sí	Sí	Sí
API	No	Sí	Sí
Uso en OWL	Total	Media	Total
Evaluación Subjetiva			
Fácil instalación	Medio (.NET 2)	Alto	Alto
Errores de inferencia	Medio	Bajo	Alto
Comunidad de desarrollo	No	No	Sí

3.3.8. Metadatos en la información geográfica

Los metadatos [57] son datos altamente estructurados que describen la información, el contenido, la calidad, la condición y otras características de los datos. En otras palabras, se define como la "*información sobre información*" o "*datos sobre los datos*". El uso principal de los metadatos se mencionan a continuación:

- Organizar y mantener el acervo del conjunto de datos de una organización.
- Proporcionar información para catálogos de datos y centros de distribución de metadatos (*clearing houses*).

- Proveer información necesaria para interpretar y procesar los datos transferidos por otra organización.

Un ejemplo de metadatos para datos geoespaciales es el siguiente:

- *Identificación.* Título, área incluida, temas, actualidad, restricciones, etc.
- *Calidad de los datos.* Precisión, nivel de completitud de los datos, linaje, etc.
- *Organización de los datos espaciales.* Modelo vector, modelo raster.
- *Referencia espacial.* Proyección, datum, sistemas de coordenadas, etc.
- *Entidad y atributos.* Información acerca de las entidades, atributos, dominio de valores de los atributos, etc.
- *Distribución.* Distribuidor, formatos, medios, estatus, precio, etc.
- *Referencia de los metadatos.* Nivel de actualización, institución que provee la información, persona responsable, etc.

Asimismo, el estándar FGDC (Federal Geographic Data Committee) es una especificación de la organización internacional de estándares (ISO) para datos geoespaciales [58], la estructura de éste se muestra en la Figura 3.13.

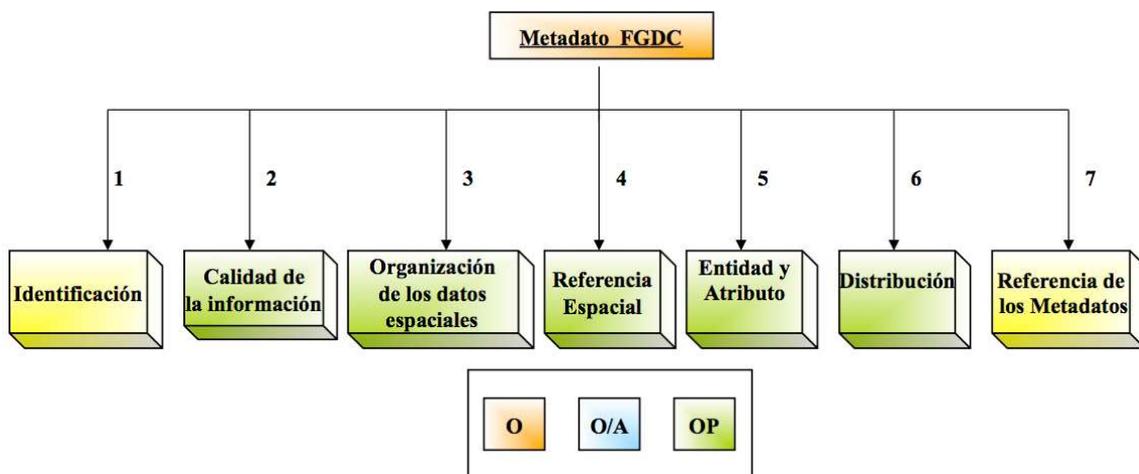


Figura 3.13. Estructura general de la FGDC.

Los metadatos están estructurados por un mínimo de elementos tales como: título, autor, fecha de creación, etc. Típicamente los elementos que conforman los metadatos están definidos por algún estándar, donde los usuarios que deseen compartir metadatos están de acuerdo con un significado preciso de cada elemento. La firma e interpretación de FGDC se muestra en la Figura 3.14 y la descripción de cada elemento se detalla en la Tabla 3.2.

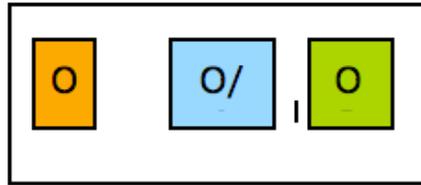


Figura 3.14. Firma de la estructura de FGDC.

Tabla 3.2. Definición de términos de FGDC.

Definición	Descripción
	Los elementos obligatorios deben proporcionarse. Si se desconoce la información para un elemento obligatorio se debe señalar "desconocido".
	Los elementos son obligatorios si son aplicables, y deben proporcionarse si los conjuntos de datos presentan las características definidas por el elemento.
	Los elementos opcionales deben ser proporcionados a discreción del proveedor.

En la Figura 3.15 se muestra la sección 1, correspondiente a la información de identificación de los metadatos.

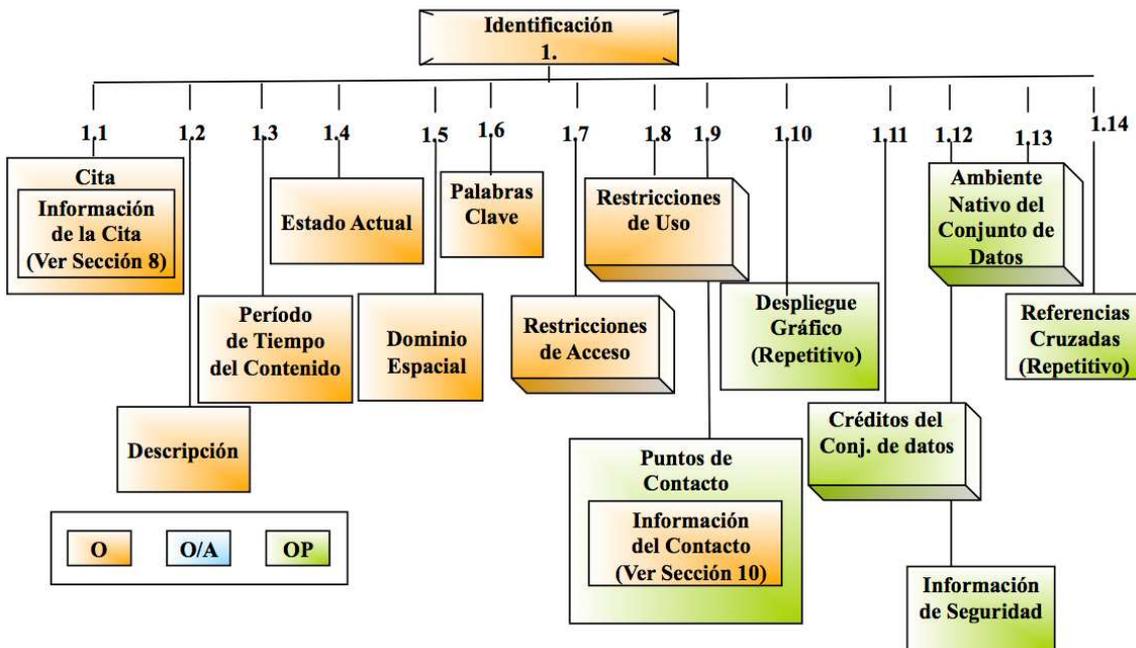


Figura 3.15. Sección de información de identificación de los datos (FGDC).

Asimismo, en la Figura 3.16 se muestra la sección 2, correspondiente a la información de calidad de los datos.

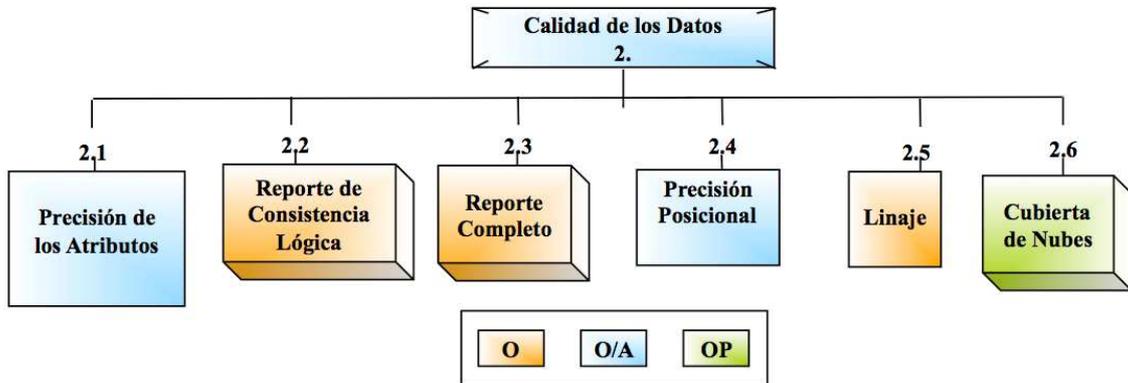


Figura 3.16. Sección de información de calidad de los datos (FGDC).

Por otra parte, en la Figura 3.17 se muestra la sección 3, correspondiente a la información de la organización de los datos geoespaciales.

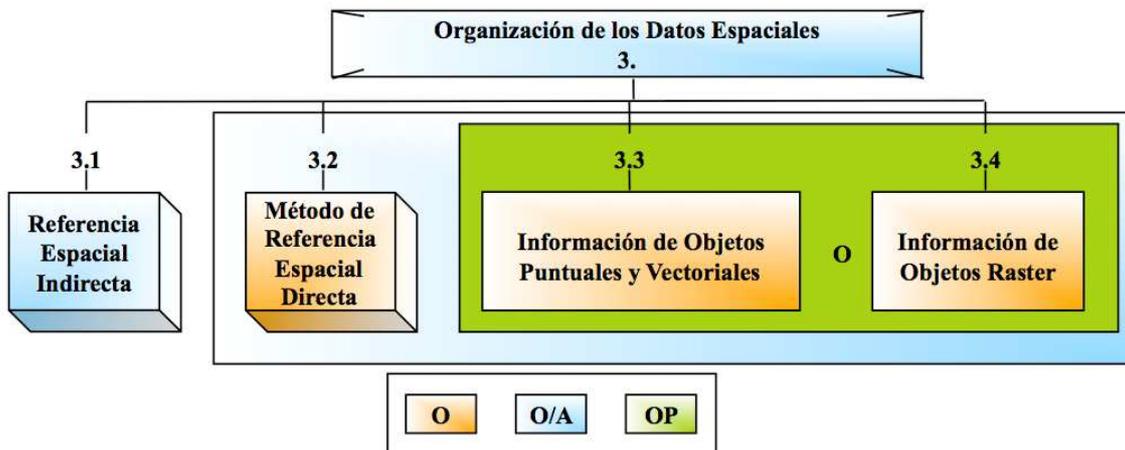


Figura 3.17. Sección de información de la organización de los datos geoespaciales (FGDC).

En la Figura 3.18 se muestra la sección 4, que describe la información correspondiente a la referencia espacial de los datos geoespaciales.

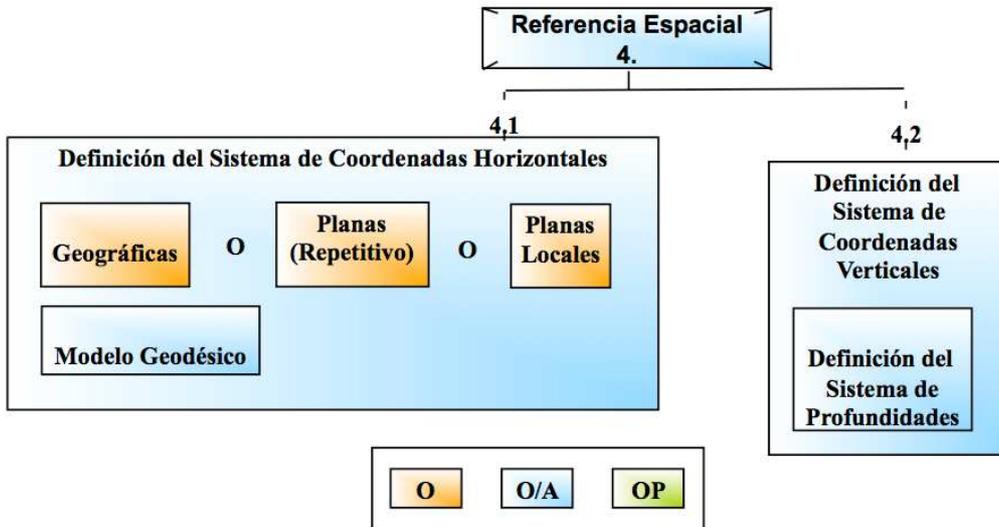


Figura 3.18. Sección de información de la referencia espacial de los datos geoespaciales (FGDC).

En la Figura 3.19 se muestra la sección 5, que describe la información del tipo de entidad y atributo de los datos geoespaciales.

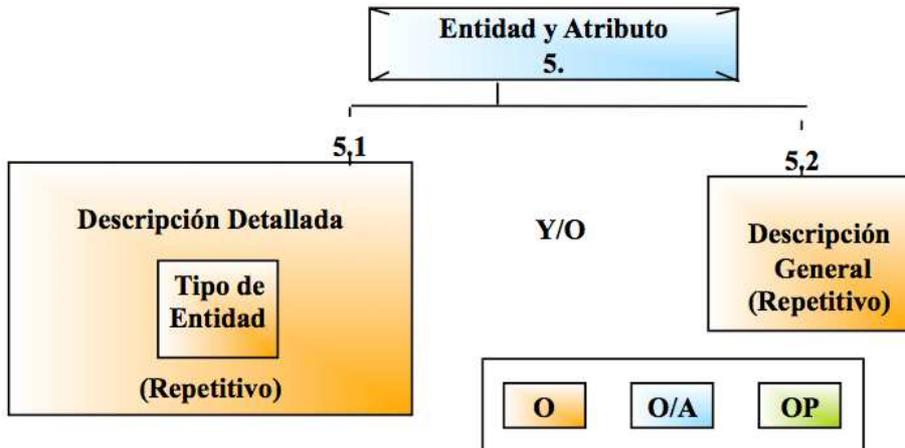


Figura 3.19. Sección de información de tipo de entidad y atributo de los datos geoespaciales (FGDC).

En la Figura 3.20 se muestra la sección 6, correspondiente a la distribución de los datos geoespaciales.

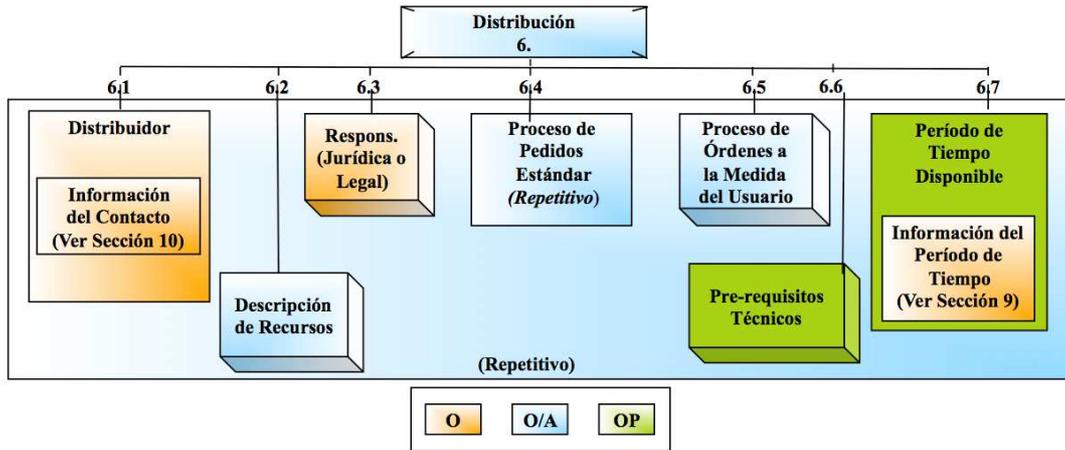


Figura 3.20. Sección de información de la distribución de los datos geospaciales (FGDC).

En la Figura 3.21 se muestra la sección 7, correspondiente a la información de referencia de los metadatos.

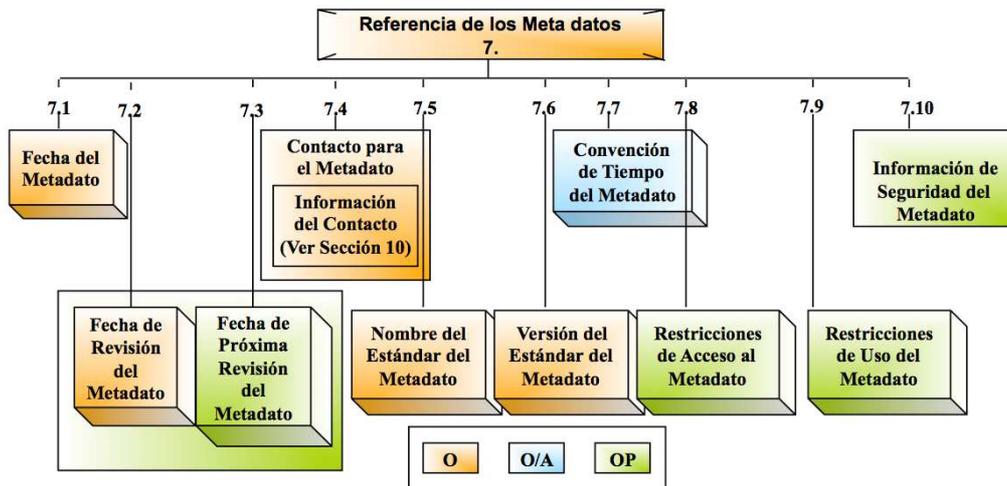


Figura 3.21. Sección de información de referencia de los metadatos (FGDC).

Finalmente, en la Figura 3.22 se muestra la secciones finales, correspondientes a la estructura de los metadatos.

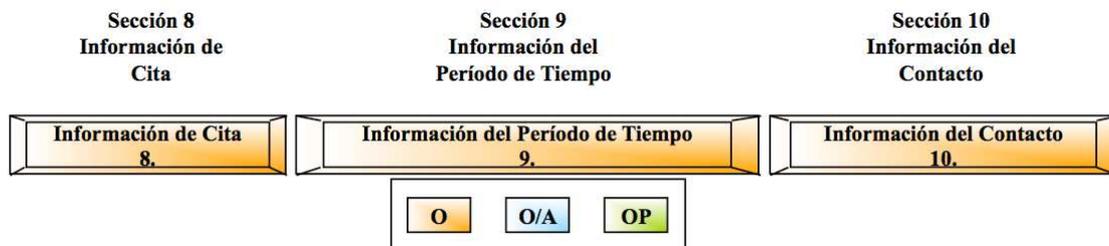


Figura 3.22. Secciones finales de la estructura de los metadatos (FGDC).

3.3.9. GEONTO-MET - Metodología para la construcción de ontologías

La metodología utilizada en este trabajo para la construcción de ontologías en la fase de conceptualización es GEONTO-MET, la cual se describe ampliamente en [59]. De acuerdo con el modelo orientado a objetos, una clase está compuesta por atributos y métodos, en donde una clase puede ser instanciada para crear un objeto. Este objeto posee todas las características de la clase y puede comunicarse con otra clase a través de algún tipo de relación, la cual permita mapearlo hacia un objeto de otra clase.

Retomando la concepción del modelado orientado a objetos, *Kaab-Ontology* es una ontología de dominio geográfico que relaciona conceptos, por medio de un conjunto de relaciones axiomáticas, en donde estas relaciones tienen un mapeo directo semejante con las clases del modelado orientado a objetos. En otras palabras, cuando se realiza el proceso de abstracción del mundo real, las entidades son representadas primeramente por medio de clases de entidades abstractas.

Una ontología del dominio geográfico, esencialmente intenta describir este universo; sin embargo, no existe una posibilidad real en la cual se pueda construir una ontología completa sobre este dominio.

De igual forma, es necesario construir una ontología del dominio geográfico con base en los requerimientos primordiales de expertos en el área, los cuales puedan proporcionar el conocimiento apropiado para casos de estudio particulares; en otras palabras, orientados a contextos específicos.

De acuerdo con lo anterior, es importante mencionar que con estas condiciones se proporcionan ventajas para incrementar el contenido de conceptos estándar que puede tener una ontología del dominio geográfico y de esta forma, solo restricciones lógicas, basadas en axiomas pueden restringir el crecimiento de una ontología del dominio geográfico.

Para *Kaab-Ontology* se proponen un conjunto de componentes o términos esenciales para conformar a esta estructura, y que además juegan un papel vital para representar la conceptualización del dominio geográfico.

Para la conceptualización del dominio geográfico se consideran un conjunto finito de relaciones, las cuales pueden clasificarse de dos tipos: relaciones *simples* y relaciones *compuestas*.

Las relaciones simples son todas aquellas que tienen la forma $apb \in R_S$, donde $a, b \in C$ y $\rho \in A_1$. Las relaciones compuestas presentan una forma compleja: $apb\pi c \in R_S$ donde $a, b, c \in C$; $\rho \in A_1$ y $\pi \in A_2$. Este tipo de relaciones se denotan como relaciones ternarias y están enfocadas principalmente a vincular relaciones axiomáticas entre la relación “*hace*” que pertenece a A_1 con elementos del conjunto A_2 . El conjunto de A_1 está definido por $A_1 = \{es, tiene, hace\}$, mientras que A_2 se define como $A_2 = \{preposiciones\}$.

En este sentido, la relación “*es*”, se considera una relación de existencia o de identidad, por medio de la cual se pueden caracterizar los conceptos. Esta relación involucra herencia de propiedades y habilidades. Además, permite jerarquizar los conceptos. En otras palabras, se puede formar una jerarquía de conceptos, a través de esta relación. Algunos sinónimos que puede tener esta relación pueden ser “*hijo_de*” y “*es_un*”.

Con respecto a la relación “*tiene*”, ésta es una relación de pertenencia (también llamada de agregación) o bien de uso (relacionada con asociación), mediante la cual se pueden definir las propiedades que describen y envuelven a un concepto.

Por último, la relación “*hace*”, es una relación de ejecución o de acción. Con esta relación se definen las habilidades, acciones u operaciones que están asociadas a un concepto estándar. Esta relación tiene como objetivo esencial mapear conceptos del tipo relacional con alguna relación axiomática del subconjunto A_2 para vincular el origen (un concepto estándar) con un destino (que en este caso es otro concepto estándar).

3.3.10. Algoritmo de Floyd-Warshall

El algoritmo de Floyd-Warshall [60] resuelve el problema de encontrar el camino más corto para todos los pares de vértices sobre un grafo dirigido $G=(V, E)$. Este método se ejecuta bajo un tiempo $\Theta(V^3)$. Además los pesos negativos de las aristas se pueden presentar, pero se asume que no hay ciclos negativos.

El algoritmo de Floyd-Warshall se basa en la estructura para el camino más corto de una matriz de multiplicación para todos los vértices que componen el grafo. Se consideran

vértices “intermedios”, aquellos en donde una ruta simple $p=\{v_1, v_2, \dots, v_l\}$ es cualquier vértice de p diferente a v_1 o v_l ; es decir, cualquier vértice en el conjunto $\{v_2, v_3, \dots, v_{l-1}\}$.

El algoritmo Floyd-Warshall se basa en la siguiente observación. Se asume que los vértices de G son $V=\{1, 2, \dots, n\}$, permitiendo considerar un subconjunto $\{1, 2, \dots, k\}$ de vértices para algún k . Para cualquier par de vértices $i, j \in V$, considerar todas las rutas de i a j , cuyos vértices intermedios son todos los trazos de $\{1, 2, \dots, k\}$, y pueda ser p la ruta de peso mínimo entre ellos (la ruta p es simple). El algoritmo Floyd-Warshall explora una relación entre la ruta p y la ruta más corta de i a j con todos los vértices intermedios en el conjunto $\{1, 2, \dots, k-1\}$. La relación depende si k no es un vértice intermedio de la ruta p . Por lo tanto:

- Si k no es un vértice intermedio de la ruta p , todos los vértices intermedios de la ruta p están en el conjunto $\{1, 2, \dots, k-1\}$. Por lo tanto, la ruta más corta del vértice i a j con todos los vértices intermedios en el conjunto $\{1, 2, \dots, k-1\}$ es también la ruta más corta de i a j con todos los vértices en el conjunto $\{1, 2, \dots, k\}$.
- Si k es un vértice intermedio de la ruta p , se interrumpe p dentro i, k, j como se muestra en la Figura 3.23. Entonces p_1 es la ruta más corta de i a k con todos los vértices intermedios en el conjunto $\{1, 2, \dots, k\}$. Esto se debe a que el vértice k no es un vértice intermedio de la ruta p_1 , al analizar p_1 se observa que es la ruta más corta de i a k con todos los vértices intermedios en el conjunto $\{1, 2, \dots, k-1\}$. De forma similar, p_2 es la ruta más corta del vértice k al vértice j con todos los vértices intermedios en el conjunto $\{1, 2, \dots, k-1\}$.

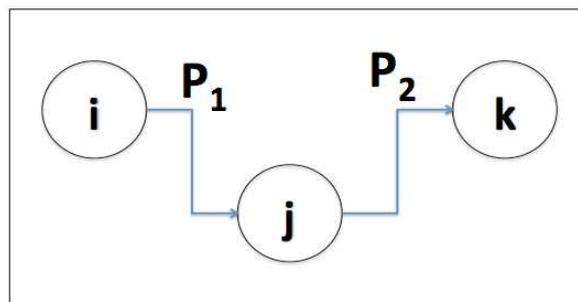


Figura 3.23. Definición de ruta más corta utilizando el algoritmo Floyd-Warshall.

De acuerdo con lo anterior, es necesario definir una función recursiva para determinar la ruta más corta. Por lo tanto, Sea $D_{ij}^{(k)}$ el peso de la ruta más corta del vértice i al vértice j

para los cuales todos los vértices intermedios están en el conjunto $\{1, 2, \dots, k\}$, entonces se

define la función de la siguiente manera:
$$D_{ij}^{(k)} = \begin{cases} w_{ij} & \text{si } k = 0 \\ \min(d_{ij}^{(k-1)}, d_{ik}^{(k-1)} + d_{kj}^{(k-1)}) & \text{si } k \geq 1 \end{cases}$$

Considerando la recurrencia de la función anterior, el siguiente procedimiento puede ser utilizado en orden incremental de k . La entrada es una matriz de $n \times n$ de pesos W . Entonces el procedimiento devuelve la matriz $D_{(n)}$ de las rutas más cortas con los pesos. A continuación, en la Tabla 3.3 se muestra el algoritmo de Floyd-Warshall utilizado.

Tabla 3.3. Algoritmo Floyd-Warshall.

FLOYD-WARSHALL(W):	
1	$n \leftarrow \text{rows}[W]$
2	$D_{(0)} \leftarrow W$
3	for $k \leftarrow 1$ to n
4	do for $i \leftarrow 1$ to n
5	do for $j \leftarrow 1$ to n
6	do $D_{ij}^{(k)} \leftarrow \text{MIN}(D_{ij}^{(k-1)}, D_{ik}^{(k-1)} + D_{kj}^{(k-1)})$
7	return $D_{(n)}$

Capítulo 4. Metodología

4.1. Introducción

En este capítulo se describe la metodología propuesta para llevar a cabo la integración y recuperación de datos geoespaciales. Hoy en día, la búsqueda y recuperación presentan diversos retos en la forma de cómo obtener de manera conceptual los datos geoespaciales para proporcionar un rango de objetos relacionados semánticamente, y que pertenezcan al mismo dominio. Con lo anterior, es posible ampliar la búsqueda, más allá de los mecanismos que ofrece el método sintáctico.

Así, una forma de lograr búsquedas semánticas, es procesando a nivel conceptual, los conjuntos de datos geoespaciales, considerando que un objeto geográfico puede describirse de muchas formas por el grado de conocimiento, abstracción e interpretación, por lo cual es necesario que sean definidos dentro de los dominios que se encuentren relacionados entre sí y con un consenso particular.

Para este fin, es necesario llevar a cabo una conceptualización del dominio en el cual se presenta el problema de integración; por lo tanto, la técnica utilizada para llevar a cabo este proceso es GEONTO-MET [59], la cual consiste esencialmente de una estructura base de trabajo para generar una ontología del dominio geográfico.

Esta estructura está compuesta básicamente por un conjunto de clases abstractas que permiten clasificar en forma jerárquica a los objetos geográficos, estableciendo sus relaciones correspondientes entre diversas clases abstractas y definiendo las restricciones y axiomas que restringen el contexto de los objetos geográficos.

Asimismo, las fuentes de información para la construcción de las ontologías fueron tomadas del INEGI y los metadatos definidos explícitamente en la representación conceptual se basan en el estándar FGDC.

Aunado a lo anterior, se ha propuesto la implementación del algoritmo DIS-C (distancia conceptual) para calcular la distancia conceptual entre los conceptos para la definición en la recuperación, cuando una consulta es realizada por el usuario. Con esto, los usuarios no obtendrán respuestas vacías, sino que si el concepto exacto no es encontrado en las fuentes, entonces obtienen como respuesta algún concepto que puede ser conceptualmente similar.

4.2. Descripción general de la metodología

La presente metodología se basa en proporcionar una técnica que permita integrar y recuperar semánticamente conjuntos de datos geoespaciales en fuentes de datos heterogéneas y en algunos casos no estructuradas.

Estas fuentes de datos, probablemente no se encuentran almacenadas en un servidor único, sino que pueden localizarse distribuidas dentro de una red, con lo cual la técnica propuesta permite establecer un mecanismo de integración basado en la conceptualización del dominio, y con la descripción de los metadatos referentes al contexto de la aplicación.

Con esta técnica se busca proporcionar un mecanismo alternativo para la integración y recuperación de datos geoespaciales, los cuales puedan representarse a través de una descripción semántica, descrita explícitamente en la ontología.

Cabe señalar, que para mostrar el funcionamiento de la metodología, se ha implementado la aplicación *SemGsearch*, la cual tiene como propósito buscar semánticamente objetos geográficos heterogéneos en una Intranet; además de integrar las fuentes de datos y devolver un *ranking* de resultados, extendiendo radios de búsqueda, mostrando la información de cada dato geoespacial por medio de sus metadatos. La lista de objetos recuperados son mostrados vía Web.

Por otra parte, a continuación, en la Figura 4.1 se muestra la estructura conceptual del método *SemGSearch*, el cual está compuesto de tres etapas fundamentales para llevar a cabo la integración y recuperación de objetos geoespaciales, mediante el uso de una ontología.

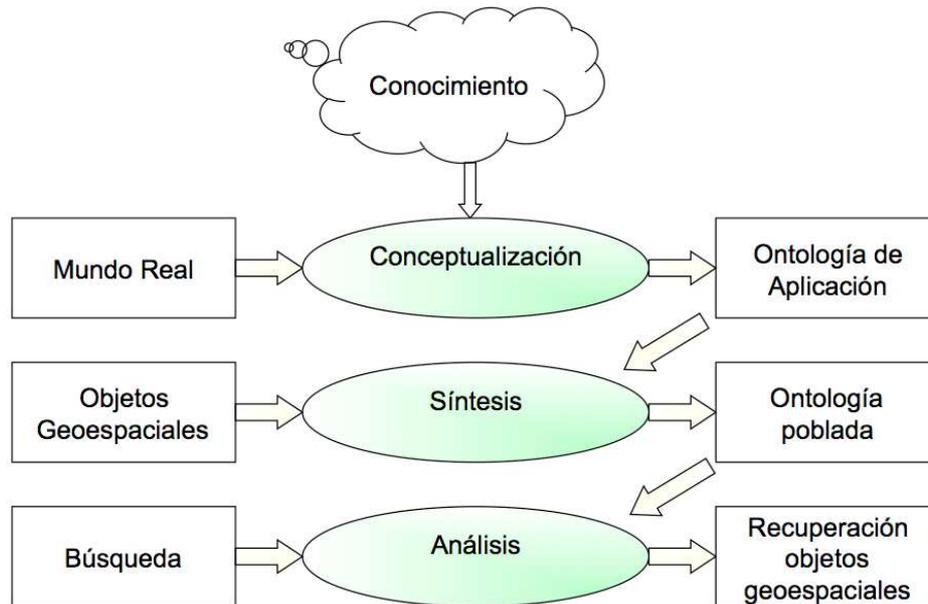


Figura 4.1. Estructura conceptual de *SemGSearch*.

Las etapas que componen a la metodología *SemGSearch* son las siguientes:

- **Conceptualización.** En esta etapa se lleva a cabo el proceso de la construcción de la base de conocimiento, representada directamente por una ontología. Esta base almacena la información de los dominios relacionados con la búsqueda.
- **Síntesis.** En esta etapa se realiza la *instanciación* de conceptos que representan objetos geográficos del dominio en cuestión. En otras palabras la etapa tiene como función poblar a la ontología.
- **Análisis.** En esta etapa se utilizan los metadatos descritos en las fuentes de datos, los cuales se encuentran localizados en los repositorios para dar paso a realizar la búsqueda sobre los dominios específicos de los conceptos y objetos relacionados en la petición.

De acuerdo con lo que se observa de la Figura 4.1, se puede decir que para describir las relaciones semánticas entre los objetos, es necesario partir del conocimiento del dominio y conceptualizar dichas relaciones, con ello se enriquecerá la ontología diseñada en la metodología.

En la etapa de conceptualización, la abstracción parte del mundo real, llevando a cabo la simplificación de las entidades mediante la metodología GEONTO-MET. Como resultado de la etapa se obtiene una ontología.

Con respecto a la etapa de síntesis, ésta presenta el esqueleto de la representación conceptual, la cual se encuentra lista para ser poblada con instancia del mundo real, con la finalidad de tener los objetos geográficos y este es el proceso en el cual la ontología es poblada con instancias del mundo real, para obtener objetos espaciales que contiene.

Finalmente, en la etapa de análisis se verifican los objetos geoespaciales identificados en la ontología como conceptos, para posteriormente hacer la búsqueda de los términos deseados en la consulta, mediante relaciones semánticas entre sí, por último la recuperación se lleva a cabo.

A continuación, se muestra la Figura 4.2, en la cual se describe la solución general conforme a las tres etapas propuestas en la metodología. En la etapa de *conceptualización* se toma la ontología de aplicación que contiene la base de conocimiento de los cuatro dominios descritos, para posteriormente utilizar un algoritmo de similitud semántica y construir un grafo que determine un valor de distancia conceptual a cada relación, según la metodología de GEONTO-MET de un concepto a otro. Finalmente, mediante el algoritmo de Floyd-Warshall se determina el valor más pequeño o costo del nodo hacia otro indicando, el valor de similitud conceptual.

En la etapa de *síntesis* se toman los datos geoespaciales que por medio del metadato FGDC se utilizan para poblar a la ontología; en este caso, una instancia corresponde a cada dato geoespacial. Posteriormente se realiza un proceso de enlazado de cada uno de estos objetos a los otros tres dominios: temático, espacial y temporal; por medio del campo palabras clave provisto por los metadatos.

Finalmente, en la etapa de análisis se realiza la búsqueda a nivel semántico, descomponiendo la consulta en conceptos para que por medio de rangos de búsqueda conceptual se recuperen los datos geoespaciales. Primeramente se realiza una búsqueda de conceptos exacta; es decir, con una distancia conceptual con $k=0$ y posteriormente, en caso de no existir datos para este valor, se extiende la búsqueda en cada dominio con un $k>0$, tomando como base la tabla resultante de Floyd Warshall en la conceptualización.

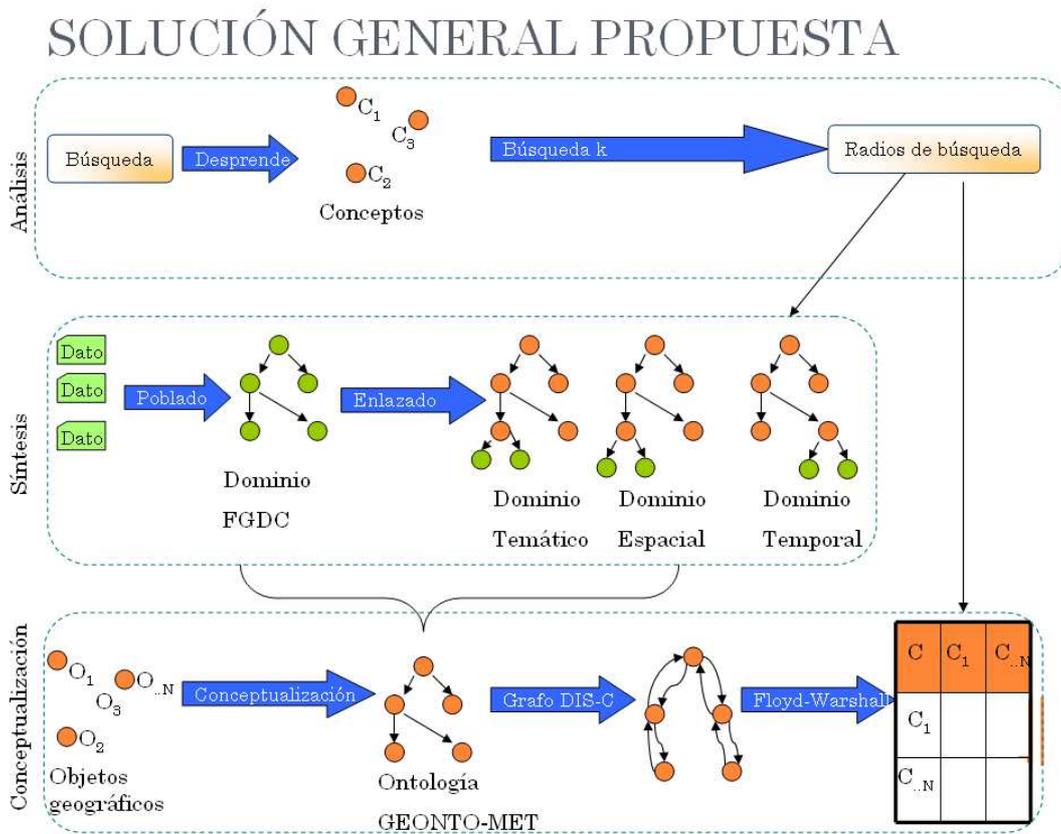


Figura 4.2. Solución general conforme metodología.

4.2.1. Etapa de conceptualización

En esta etapa se conceptualizan los tres dominios que definen a los objetos geográficos: el *temático* que se encarga de describir a los conceptos que representan a los objetos geográficos, así como sus componentes, propiedades y acciones que ejercen. Todos los elementos anteriores, describen particularmente a cada objeto geográfico de otro y provee características cualitativas.

En el dominio *espacial* se describe una cierta correspondencia entre los conceptos, por medio de su localización y así poder tener una referencia de contigüidad o cercanía entre ellos.

Finalmente, en el dominio *temporal* se describen los conceptos de tiempo que interpretan distintas nociones con respecto a una periodicidad, en donde la semántica de los objetos geográficos puede verse modificada o en su defecto desaparecer con respecto a un intervalo o línea de tiempo.

Cabe señalar que independientemente de los dominios definidos aquí, es posible llevar a cabo una extensión a un número N de dominios, dependiendo directamente de la información que proporcione la fuente.

Este hecho implica que a mayor cantidad de dominios descritos, mayor granularidad semántica se puede obtener de la ontología para refinar una consulta y hacerla aún más especializada; con lo cual se realiza un mapeo entre estos dominios y atributos relevantes en la búsqueda y recuperación.

En la Figura 4.3 se muestra la estructura de la ontología general, donde se definen los N dominios que pueden describir a un objeto geográfico en la fase de su conceptualización, el concepto raíz se refiere a cualquier cosa (*Thing*), del cual se hereda un objeto geográfico, y además se definen N dominios que pueden caracterizar a dicho objeto bajo un cierto ámbito o dominio.

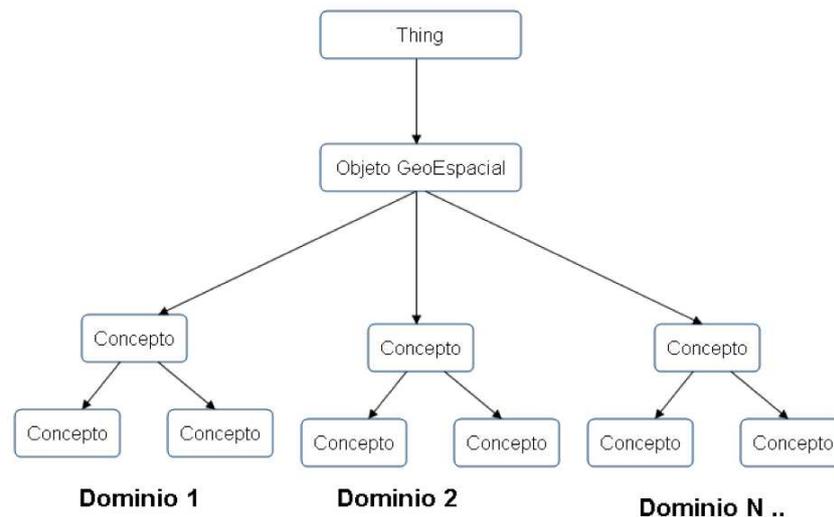


Figura 4.3. Ontología general.

Para el proceso de construcción de la ontología, después de un análisis del dominio o universo de discurso, se eligió incluir los dominios espacial, temporal y temático como características propias de un objeto geográfico, los cuales componen el motor de búsqueda o la base de conocimiento.

En esta base los conceptos que se indican en una consulta se pueden inferir con conceptos similares semánticamente, por medio de medidas de distancia conceptual, como es el caso de DIS-C. Para describir concretamente un objeto geográfico localizado en un determinado lugar del repositorio distribuido y con la intención de solucionar el problema de

heterogeneidad de los datos, se utilizó su metadato correspondiente a cada objeto geográfico, bajo la norma FGDC, ya que ésta define conceptos relevante de dicho objeto como son: *nombre*, *cita*, *descripción*, etc., pero para fines de búsqueda del presente trabajo se hace un énfasis en el *campo obligatorio* dentro de la especificación al campo *palabras clave*. Este campo define palabras que son características del objeto y son de gran utilidad por el conocimiento acertado y simplificado que muestran.

En la Figura 4.4 se muestran las jerarquías de conceptos generales de cada uno de los dominios definidos en este trabajo.

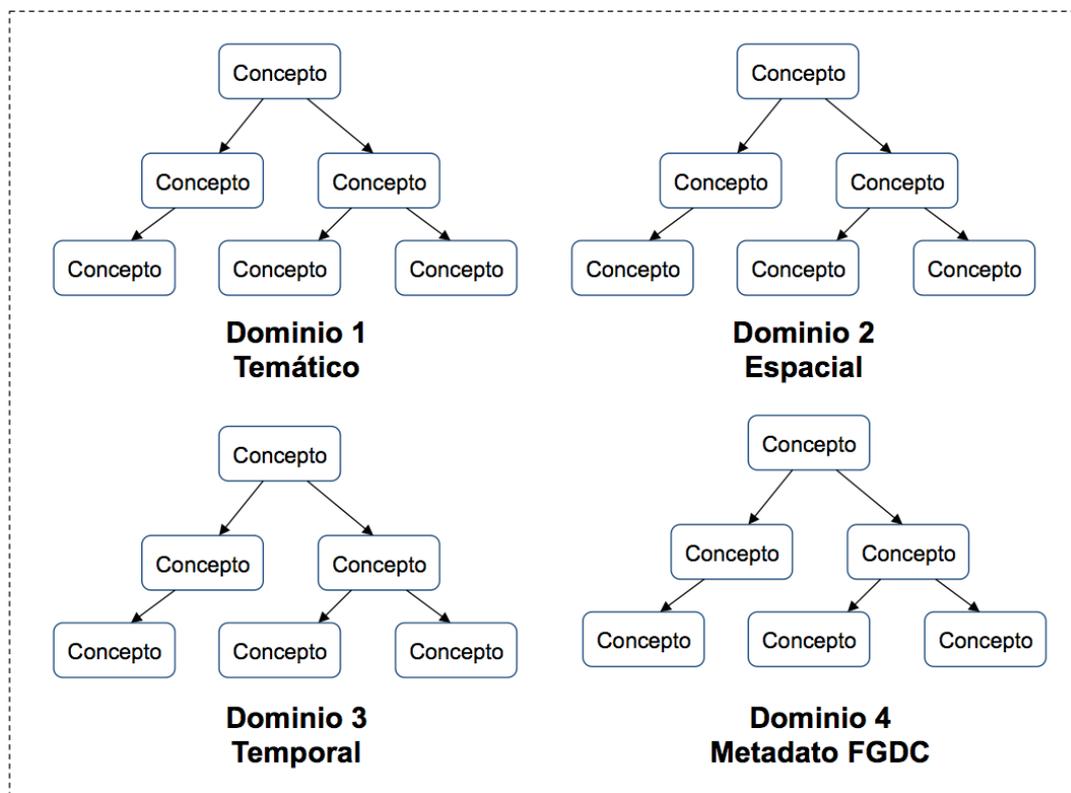


Figura 4.4. Dominios definidos en la conceptualización.

De acuerdo con la metodología GEONTO-MET, las relaciones semánticas que permiten describir a los conceptos de estos dominios son las siguientes:

- “*Es*” esta relación jerárquicamente especifica un concepto más especializado que el antecesor del cual se desprende por medio del mecanismo de herencia.
- “*Tiene*” esta relación describe conceptos componentes del padre, lo cual lo caracterizan y finalmente la relación

- “Ejecuta” la cual muestra acciones que realizan los conceptos. Adicionalmente, se agrega una lista de posibles términos que pueden ser similares con los que son conceptualizados, de este forma se sumarán a enriquecer el conocimiento, junto con las relaciones semánticas para realizar una mejor descripción.

En la Figura 4.5 se muestran las relaciones semánticas que utiliza GEONTO-MET para relacionar los conceptos de la ontología.

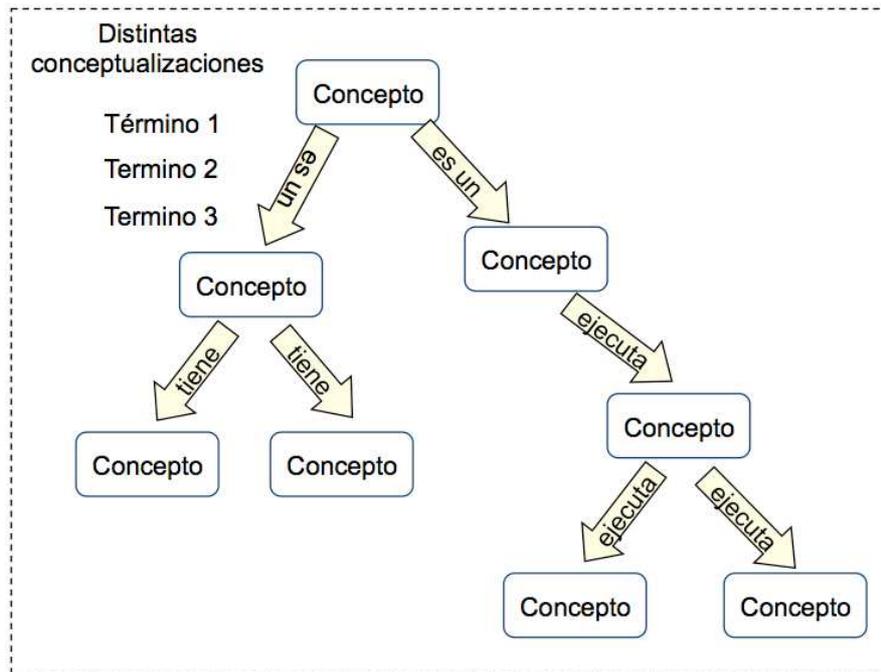


Figura 4.5. Relaciones semánticas en la ontología.

Por otra parte, ya que se tienen contemplados los dominios que tendrá la ontología y los tipos de relaciones entre sus conceptos, el proceso siguiente es la construcción del conocimiento en cada dominio, preferentemente por medio de algún estándar que tenga como base un consenso, por lo que para el dominio temático se realizó la conceptualización bajo el diccionario de INEGI [61]. En el dominio espacial se utilizó la Republica Mexicana bajo tres esquemas: el de división eco-regional según CONABIO [62], división por zonas espaciales, de acuerdo por CFE [63] y por zonas económicas.

Cabe destacar que por lo anterior, esta parte de la conceptualización realizada en la metodología del trabajo; la creación de los dominios y generación de la ontología es el único proceso manual, por la necesidad del conocimiento de algún estándar o conocimiento *a priori* de un sujeto sobre determinado dominio.

En cuanto para el dominio temporal, la conceptualización fue realizada de manera particular, conforme a como se describe y compone una fecha y sus distintas interpretaciones.

Cuando ya se tiene construida toda la ontología, el proceso siguiente es determinar si son similares o no dos conceptos, por lo cual se introduce a la metodología una medida de similitud conceptual, la cual dando prioridad al significado de los conceptos, si éstos son similares o diferentes sobre el carácter lexicográfico de ellos, se recupera el concepto más cercano conceptualmente al originalmente requerido.

Por lo anterior, uno de los objetivos de la metodología es determinar en que medida son similares semánticamente dos conceptos C_x , C_y entre ellos, siendo $C_x, C_y \in M$, siendo M , cualquier dominio de la ontología O . Por ejemplo, para el dominio temático dos conceptos geográficos son similares cuando su distancia conceptual es muy cercana y ésta es acotada por un contexto. Sea el caso del objeto geográfico “*avenida*” y “*calle pavimentada*”, en el caso del dominio espacial, particularmente en el país de México ¿qué tan similares son el estado de “*Veracruz*” y “*Chiapas*”? Finalmente hablando en el dominio temporal y de manera conceptual ¿qué tan cercanos son el año “2010”, “2009”, y “1996”?.

Para responder ante los anteriores cuestionamientos se introduce el término de *distancia conceptual* que se refiere hablando dentro de una ontología a “la distancia o peso w entre dos conceptos, tomando en cuenta la distancia entre particiones donde se localiza el concepto C_x al otro concepto C_y y el tipo de relaciones entre ellos.

Una vez que ya se ha definido el término distancia conceptual se procede a determinar el peso de cada relación entre los conceptos, por lo que se utiliza el algoritmo DIS-C, en donde un peso w se asigna para cada relación, definida por la metodología GEONTO-MET. Esta metodología cuenta con las siguientes relaciones axiomáticas: “*es*”, “*tiene*” y “*ejecuta*”; las cuales fueron utilizadas para definir la conceptualización del dominio.

Por otra parte, DIS-C genera un grafo fuertemente conexo al dar un peso del C_x al C_y y viceversa de C_y a C_x . Este proceso proporciona como resultado un grafo que define similitudes entre cualquier par de conceptos de la ontología.

Dicho lo anterior, la medición de distancia conceptual o similitud semántica se realiza en dos fases: la primera es la creación de un grafo de pesos “ w ”, entre todos los conceptos que componen a cada dominio, por medio del tipo de relación que los une, y la segunda es cuando se ha definido el grafo para determinar la menor distancia o ruta entre todos los conceptos o nodos. Esta distancia o peso “ w ” identifica la similitud semántica entre objetos.

De acuerdo con lo descrito anteriormente, en la generación del grafo se ejecuta para cada dominio de la ontología el algoritmo DIS-C, este algoritmo devuelve un grafo fuertemente conexo por cada dominio donde se especifican los valores de distancia conceptual o peso w entre todos los conceptos relacionados.

Como se mencionó anteriormente, el algoritmo DIS-C define la distancia conceptual Δ_c entre los pares de conceptos definidos explícitamente en una relación simple, tal como lo define GEONTO-MET [ref]. En otras palabras, esta relación simple no utiliza el conjunto de relaciones axiomáticas definidas en el conjunto A_2 que contiene al conjunto de preposiciones que le dan causalidad a la descripción ontológica y sirven para conectar un concepto del tipo estándar con otro. Para ello, DIS-C utiliza las siguientes restricciones.

1. Si $a(es)b \in R_R$,
 - a. $\Delta_c(a,b)=0$
 - b. $\Delta_c(b,a)=1$
2. Si $a(tiene)b \in R_R$,
 - a. $\Delta_c(a,b)=1$; $R_{(p)}$ es el número de ocurrencias de la propiedad $p = a(tiene)b$ en R_R , donde normalmente este valor es 1.
 - b. $\Delta_c(b,a) = card(P(a))$; $R_{(p)}$ es $P(a) = \{x | a(tiene)x \in R_R\}$ para cualquier concepto $x \in C$ y $R_{(p)}$ es el número de ocurrencias de la propiedad $p = a(tiene)b$ en R_R .
3. Si $a(hace)b \in R_R$,
 - a. $\Delta_c(a,b)=1$; $R_{(h)}$ es el número de ocurrencias de la habilidad $h = a(hace)b$ en R_R , normalmente este valor es 1.

- b. $\Delta_c(b,a) = \text{card}(H(a))$; $R_{(h)}$ es $H(a) = \{x \mid a(\text{tiene})x \in R_R\}$ para cualquier concepto $x \in C_c$ y $R_{(h)}$ es el número de ocurrencias de la habilidad $h = a(\text{tiene})b$ en R_R .

En la Tabla 4.1 se muestra el algoritmo para determinar la distancia conceptual entre los conceptos en la ontología.

Tabla 4.1. Algoritmo DIS-C.

DIS-C(O) :	
Entrada: Una ontología O desarrollada GEONTO-MET	
Salida: Una tabla T con distancias conceptuales entre conceptos de O	
1	$G_o \leftarrow$ grafo dirigido con pesos
2	$R_R \leftarrow$ relaciones en O
3	$C_d \leftarrow$ conceptos en O
4	$R(apb) \leftarrow$ Número de ocurrencias de la propiedad $p=(apb)$; $a, b \in C_d$ y $r \in R_R$
5	$\text{card}(P(a)) \leftarrow$ Número de relaciones de $P(a) = \{x \mid a(p)x \in R_R\}$ para cualquier concepto $x \in C_d$
6	for each $a, b \in C_d$ y $\rho \in R_R$ do
7	switch ρ do
8	case "es":
9	Add ($a, b, 0$) $\rightarrow G_o$
10	Add ($b, a, 1$) $\rightarrow G_o$
11	end
12	case "tiene":
13	Add $\left(a, b, \frac{1}{R(apb)}\right) \rightarrow G_o$
14	Add $\left(b, a, \frac{\text{card}(P(a))}{R(apb)}\right) \rightarrow G_o$
15	end
16	case "ejecuta":
17	Add $\left(a, b, \frac{1}{R(apb)}\right) \rightarrow G_o$
18	Add $\left(b, a, \frac{\text{card}(P(a))}{R(apb)}\right)$
19	end
20	end

21	end
----	-----

Para ejemplificar DIS-C en la Figura 4.6 se muestra un fragmento de la ontología con las relaciones axiomáticas descritas en GEONTO-MET, entre un conjunto de conceptos. Asimismo, en la Figura 4.7 se muestra el grafo de la distancia conceptual generado a partir de la Figura 4.6.

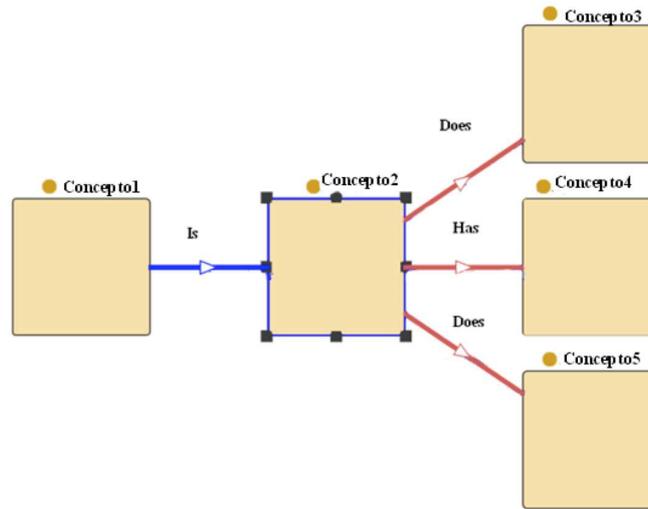


Figura 4.6. Fragmento de la ontologías con relaciones axiomáticas entre conceptos.

Sobre la ontología anterior, se aplica el algoritmo DIS-C, por cada tipo de relación se asigna un peso w , entre cualquier par de conceptos, un valor numérico además de asegurar que cada nodo del grafo tendrá un grado $G > 2$, por lo cual este proceso devolverá un grafo fuertemente conexo.

Para ejemplificar el procedimiento, se tiene el (concepto1, concepto2) relacionados bajo la relación “es”, de acuerdo con DIS-C, del “concepto1” al “concepto2”, la distancia conceptual o peso w será 1 y viceversa 0.

En el caso (concepto2, concepto4) relacionados bajo la relación “tiene”, del concepto1 al concepto2 el peso w es $\frac{1}{R(a\phi b)}$; es decir, se tiene 1/1 y viceversa $\frac{card(P(a))}{R(a\phi b)}$ es 1/1.

Finalmente los conceptos (concepto2, concepto3), bajo la relación “ejecuta”, el peso w del concepto1 al concepto2 es $\frac{1}{R(a\phi b)}$; es decir, se tiene 1/1 y viceversa $\frac{card(P(a))}{R(a\phi b)}$ es 2/1.

La Figura 4.7 muestra el grafo resultante del algoritmo DIS-C para la ontología de la Figura 4.6, desarrollada con la metodología GEONTO-MET.

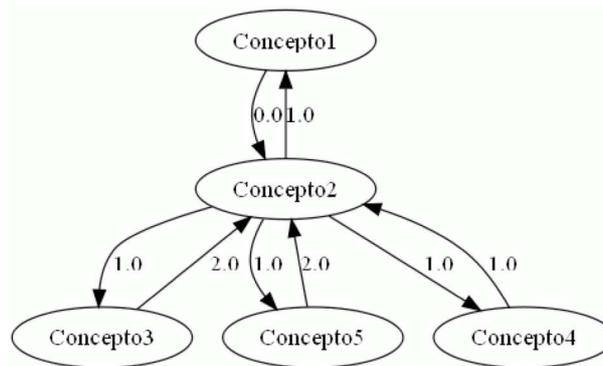


Figura 4.7. Grafo con las distancias conceptuales obtenidas a partir de la ontología de la Figura 4.5.

Continuando con la metodología en la etapa de conceptualización, una vez que se obtiene el grafo con los pesos w correspondientes a las distancias conceptuales, el proceso siguiente es determinar el menor peso w entre dos conceptos a , b que pertenecen a un determinado dominio en la ontología. Este proceso se lleva a cabo aplicando el algoritmo de Floyd-Warshall, el cual toma como entrada el grafo resultante del algoritmo DIS-C como una matriz pesos w entre un concepto C_x a otro C_y que pertenecen al mismo dominio.

Cabe señalar que se implementó el algoritmo Floyd-Warshall, debido a que por el número de conceptos contenidos en los dominios, éste resultó ser el más rápido en tiempo de ejecución que otros algoritmos que se encuentran disponibles. Esto se debe al uso de matrices de multiplicación, lo cual es sumamente útil para la aplicación *SemGSearch* que trabaja vía web, en donde el tiempo de respuesta es preponderante para la correcta operación del sistema.

Cuando se obtiene dicha matriz se ejecuta el algoritmo Floyd-Warshall que especificará el menor peso para todos los pares de conceptos relacionados entre ellos y el resultado será otra matriz, indicando el camino más corto o peso mínimo entre cualquier par de nodos C_x , C_y que pertenecen al grafo DIS-C. En la Tabla 4.2 se muestra la matriz que indica el peso mínimo entre un concepto C_x a un concepto C_y .

Tabla 4.2. Matriz generada por DIS-C, indicando el camino más corto.

Conceptos	C ₁	C ₂	C ₃	C ₄	C ₅
C ₁	0	0	1	1	1
C ₂	1	0	1	1	1
C ₃	3	2	0	3	3
C ₄	3	2	3	0	3
C ₅	2	1	2	2	0

4.2.2. Etapa de síntesis

En la etapa de síntesis se generan las instancias en el dominio de metadatos basado en FGDC, es decir; para cada objeto geoespacial se crea una instancia siguiendo la especificación como ejemplo su cita descripción tamaño y en espacial el campo obligatorio de palabras clave siendo este campo el que servirá de que se ligan estas palabras a los otros dominios conceptualizados en la etapa anterior, a saber, el dominio temporal, espacial y temático. En otras palabras, se crean objetos específicos que representan objetos geográficos y se utilizan la ontología alojar las instancias de estos objetos para posteriormente sean ligados a los tres dominios faltantes que caracterizan al objeto geográfico. Este proceso se realiza básicamente con las siguientes dos tareas:

1. Se crean instancias de objetos geoespaciales localizados en las fuentes de datos, utilizando para tal fin a los metadatos; es decir, para los campos que define la norma FGDC se llena con la información propia por objeto y se genera una instancia de todo el dominio de metadatos. Entonces, se generan N instancias del dominio metadatos. En la Figura 4.8 se muestra el esquema general que se implementa cuando se crean las instancias de los datos geoespaciales.

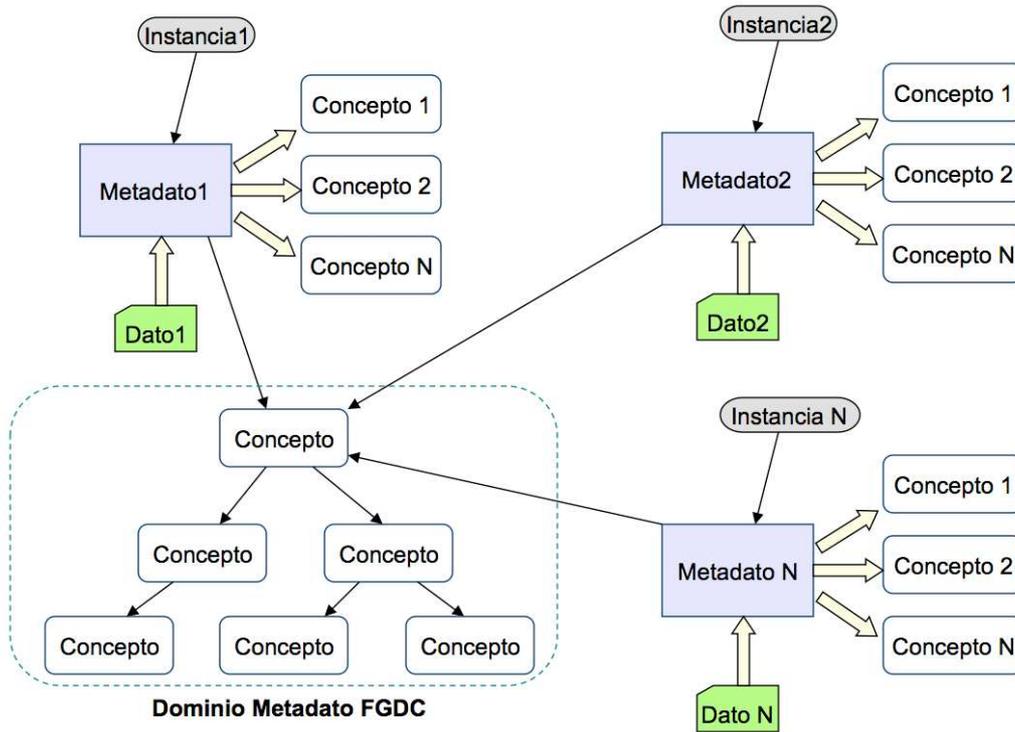


Figura 4.8. Esquema para generar las instancias de los datos geoespaciales.

2. A partir de las instancias u objetos geoespaciales generados anteriormente, se lleva a cabo el proceso de poblar el dominio de los metadatos y se crean relaciones de correspondencia de cada instancia en particular a los dominios definidos en la ontología de aplicación. Para cada una de las instancias n del dominio de metadatos estará ligada 1, 2, ..., M dominios en la ontología (para la presente metodología se contemplan solo tres: temático, espacial y temporal). En cada instancia del dominio metadatos se identifica el campo *palabras clave* y se extraen los conceptos para formar un vector de esta palabra, si se hace una búsqueda o *matching* localizando cada concepto en los tres dominios, al encontrarse coincidencia se realiza un enlace entre la instancia y el concepto al dominio, bajo la relación “es”.

En la Figura 4.9 se muestra la estructura que general de la etapa de síntesis.

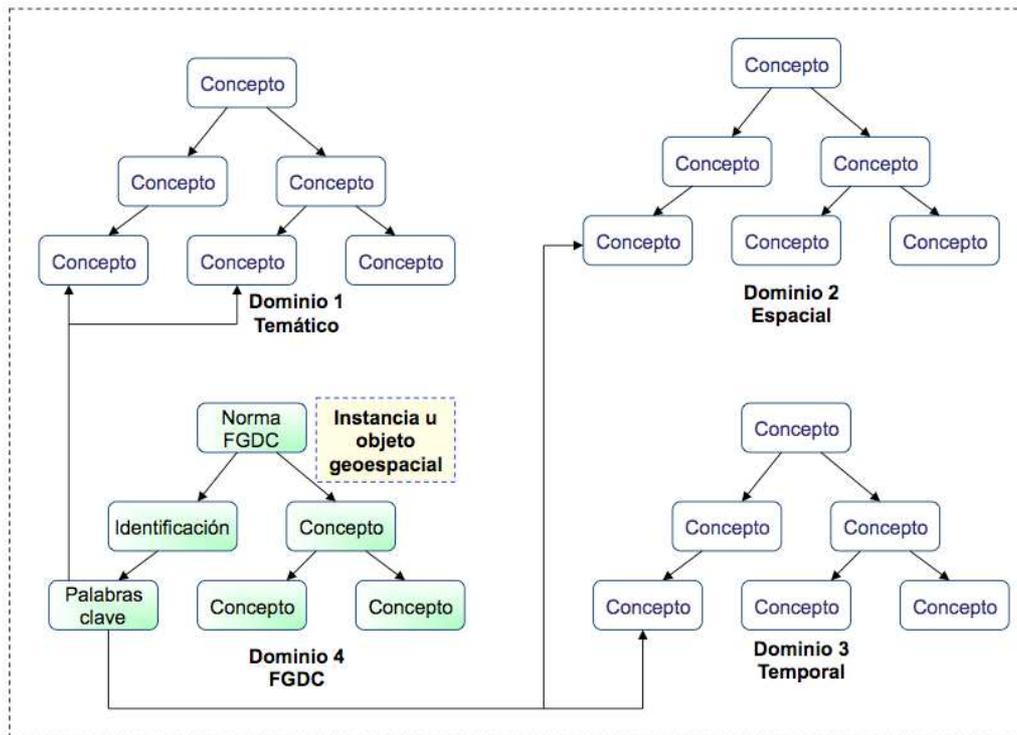


Figura 4.9. Etapa de síntesis en la ontología.

En términos generales, el procedimiento de la etapa de síntesis consiste de los siguientes pasos:

1. Se lee la información de un determinado objeto geográfico, por medio de un archivo nombrado como su metadato, se identifican los campos correspondientes a los conceptos dominio en la ontología FGDC y se crea una instancia para todo el dominio; es decir, el *metadato1* y sus atributos de identificación de descripción como palabras clave, las cuales conformarán la *instancia1* del dominio FGDC, y así sucesivamente para el *metadatos2, 3, ..., N*.
2. Posteriormente, en cada instancia del dominio FGDC se identifica el grupo de palabras clave p_1, p_2, p_3 , por cada p_x del objeto geoespacial generando un vector. Después, estos conceptos se buscan en cada dominio restante; es decir, se busca en el dominio temático, temporal y espacial; al encontrarse correspondencia del mismo concepto en ambos dominios, la actual instancia se vincula bajo la relación “es” al dominio encontrado. Este proceso se repite para todas las palabras clave.

3. Se realizan los pasos 1 y 2 para todos los objetos geográficos, leyendo sus metadatos correspondientes.

4.2.3. Etapa de análisis

La etapa de análisis se encarga de llevar a cabo las tareas de recuperación y búsqueda semántica; así como la visualización de los resultados recuperados. Esta etapa está compuesta de cuatro tareas básicas:

1. A partir de los términos buscados, se obtienen los conceptos clave para ser recuperados en los dominios definidos en la ontología. En esta fase se identifican los conceptos que componen la consulta uno por uno, para posteriormente incluirlos en un vector que define a los objetos geográficos que se están buscando. Dentro de los conceptos de la búsqueda se define por *grupos*, indicando a que dominio pertenecen; es decir, si son relacionados al dominio temático, espacial o temporal; esta clasificación ayuda a realizar en forma más precisa la búsqueda refiriendo un contexto en particular. En la Figura 4.10 se muestran los conceptos (*palabras clave*) que deben definirse para cada dominio, para que posteriormente sean analizados y se puedan referir a objetos geográficos por sus metadatos e iniciar el proceso de análisis semántico.

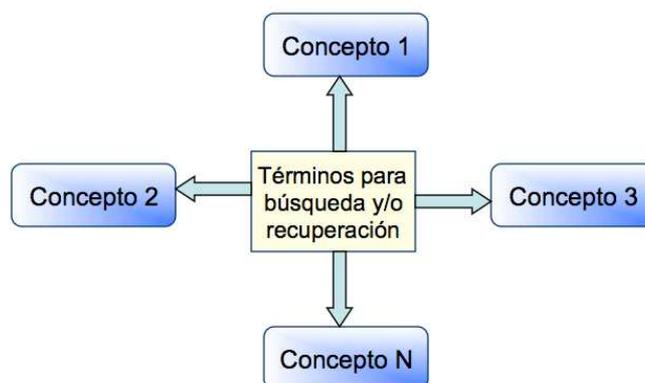


Figura 4.10. Conceptos clave de cada dominio en la fase de búsqueda de términos.

2. Con el vector resultante de la fase anterior, se procede a buscar cada concepto de cada grupo de la consulta con su respectivo dominio; es decir, con tres dominios que en la presente metodología caracterizan al objeto espacial, temático, y temporal. Cabe señalar que en cada grupo se debe al menos tener una concordancia de un concepto de la

consulta con otro en el dominio ontología, para la posterior fase de crecimiento de radios de *distancia conceptual*, sino por defecto se sitúa en alguno léxicamente parecido a alguno del grupo o se hace una concordancia o *matching* al elemento más general del dominio; es decir, a la raíz “ *objeto geográfico*”. Este concepto es la raíz de cada dominio en la ontología. Lo anterior se realiza con el fin de evitar un resultado nulo al realizar la intersección de los dominios (ver Figura 4.12). Posteriormente, ya que se tiene una concordancia de conceptos de la consulta con conceptos de la ontología, se procede a revisar en el dominio de los metadatos, cuales objetos geográficos tienen a estos conceptos de la ontología, consultando las *palabras clave* de cada instancia y este grupo resultante del dominio formará parte en el proceso de recuperación para ser devueltos al usuario que gestionó la consulta. Cabe señalar que este proceso de búsqueda es inverso al de síntesis o población de la ontología, propuesta en este trabajo en la etapa anterior, además en este caso se busca por término exacto; es decir, para una distancia conceptual $k=0$; esta distancia es la más cercana a los términos de la consulta (valor de respuesta exacto). En la Figura 4.11 se muestra el diagrama funcional del proceso de búsqueda y recuperación semántica.

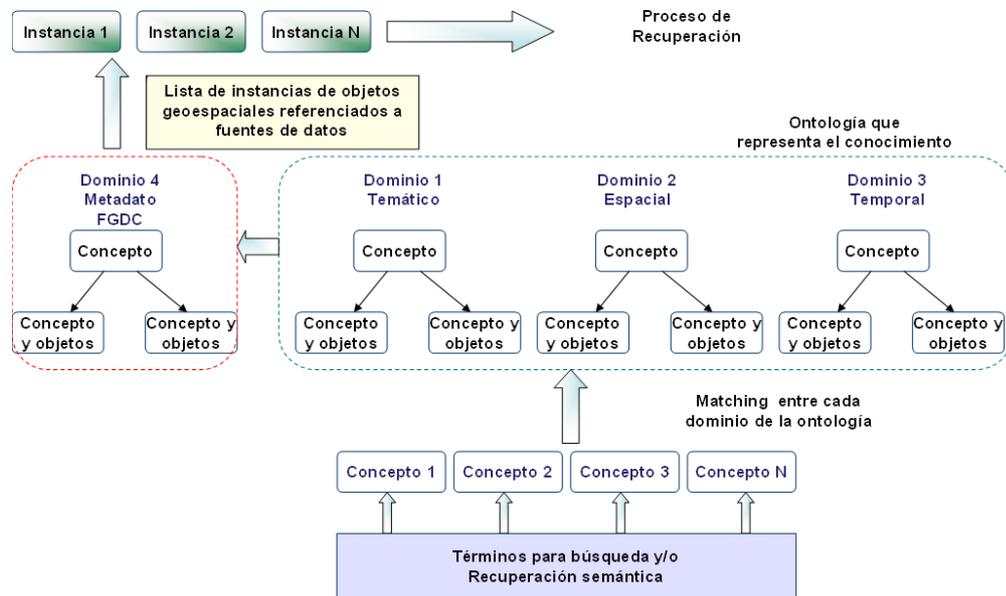


Figura 4.11. Diagrama funcional del proceso de búsqueda y recuperación semántica en el caso $k=0$.

3. El paso siguiente cuando ya se encontró la instancia con la distancia conceptual más cercana; es decir, el término exacto o distancia $k = 0$, se continua extendiendo una para $k \geq 0$. Cabe mencionar que se estima el criterio “ ≥ 0 ” debido a que en DIS-C la relación “es” se tienen los hijos de un concepto con un peso w o valor de similitud de $k = 0$, por lo que no son términos exactos sintácticamente como en la fase anterior, pero se tiene el mismo peso y se debe contemplar este caso en el crecimiento de la similitud k . Para lograr lo anterior se identifica un concepto de la fase anterior, se consulta la matriz resultante del algoritmo de Floyd-Warshall de la etapa de conceptualización; se identifica este concepto y se obtienen los valores numéricos de similitud semántica con otros conceptos.

Por tanto, es posible incrementar el valor de k con este vector y crecer los radios de búsqueda para un criterio específico, con la finalidad de obtener un rango con un mayor número de resultados, los cuales serán proporcionales a la distancia conceptual que se abarque.

En la Tabla 4.3 se muestran los valores de crecimiento para los radios con la finalidad de crecer la búsqueda semántica, en caso de no encontrar un concepto “exacto” en la petición de la consulta. En esta tabla se busca el concepto C_x anteriormente encontrado para $k = 0$ y se extraen los valores numéricos con los pesos de similitud semántica para los demás C_y que son conceptos relacionados con el mismo. Por ejemplo, en color verde se indica que se encontró el concepto C_{x1} y obtiene su vector con valores numéricos a los otros conceptos $C_{y1}, C_{y2}, \dots, C_{yn}$.

Tabla 4.3. Valores de crecimiento de los radios.

Conceptos	Cy ₁	Cy ₂	Cy ₃	Cy ₄	Cy _N
Cx ₁	valor	valor	valor	valor	valor
Cx ₂	valor	valor	valor	valor	valor
Cx ₃	valor	valor	valor	valor	valor
Cx ₄	valor	valor	valor	valor	valor
Cx _N	valor	valor	valor	valor	valor

El procedimiento para expandir los radios de búsqueda se realiza como sigue a continuación. Además, se muestra en la Figura 4.12 la extensión de la distancia conceptual utilizando un rango de búsqueda.

- En primera instancia para un criterio φ de búsqueda se analizan los conceptos relacionados semánticamente de modo directo; es decir, los conceptos más cercanos con la distancia conceptual de valor 1.
- Para un criterio φ de búsqueda se analizan los conceptos que están relacionados semánticamente por una distancia K .
- Si se desea expandir el radio de búsqueda para un criterio φ se analizan los conceptos relacionados semánticamente para una distancia $K + 1$.

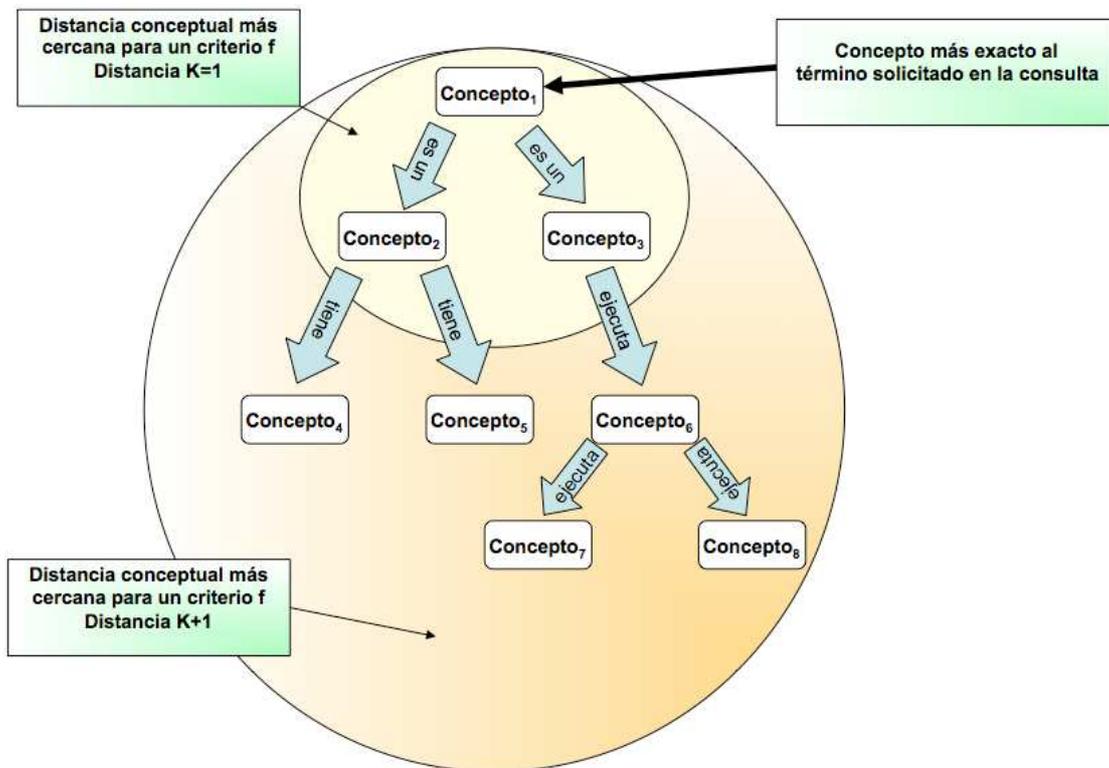


Figura 4.12. Distancia conceptual extendiendo el rango de búsqueda.

Finalmente, cuando ya se han crecido los radios para valores $k \geq 0$ y se aplicó este procedimiento para cada uno de los dominios espacial, temporal y temático; se devuelve una lista de conceptos que son comunes entre los tres dominios, por medio de la intersección entre los conjuntos de conceptos geográficos. En la Figura 4.13 se muestra la intersección referente a los dominios mencionados.

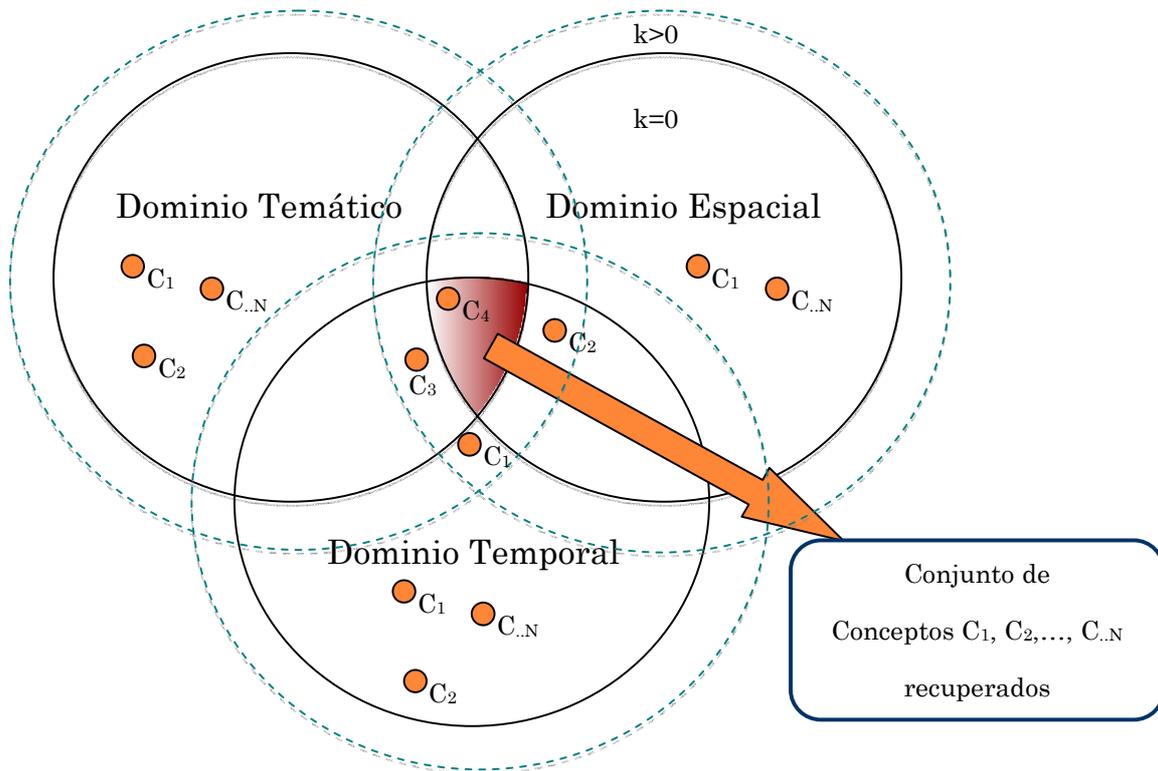


Figura 4.13. Intersección de resultados correspondientes a los dominio espacial, temático y temporal.

Con esta lista de conceptos comunes obtenida de la intersección entre los tres dominios, ahora se procede a buscar los conceptos en el dominio de Metadatos FGDC, esto con la finalidad de obtener las instancias que serán devueltas en forma de una lista de recuperación. En otras palabras, a partir del conjunto de conceptos, resultado de la búsqueda, se devuelven las instancias que son los objetos geográficos en los diferentes repositorios, los cuales posteriormente serán recuperados. En la Figura 4.14 se muestra el proceso de recuperación semántica.

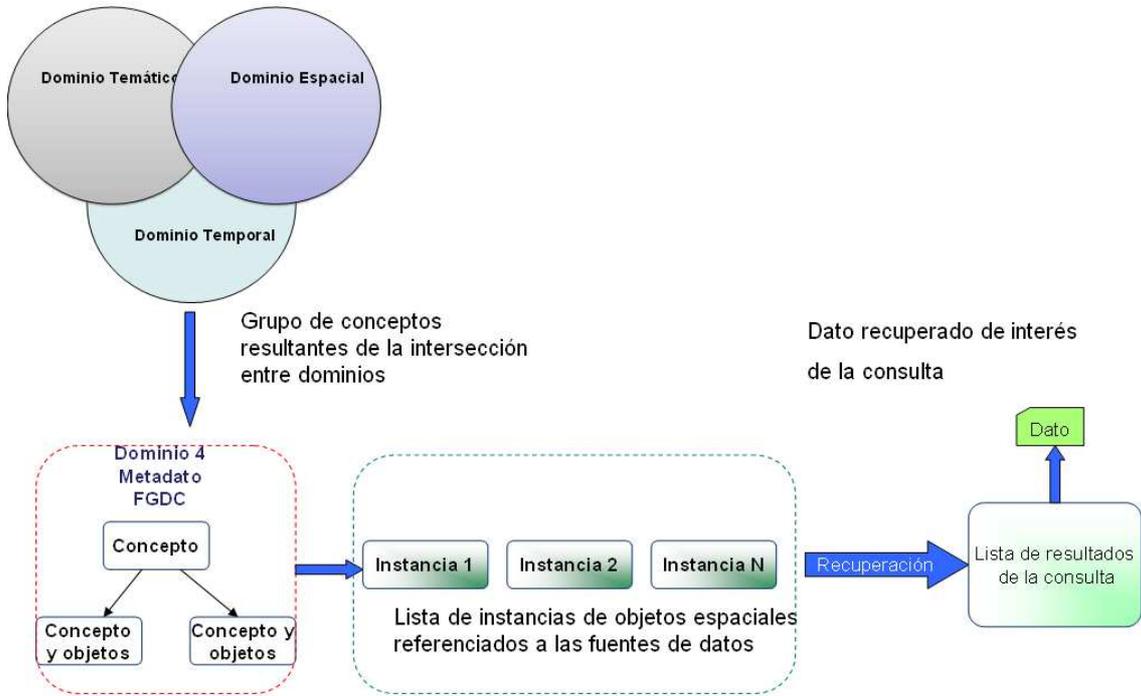


Figura 4.14. Resultado y proceso de la recuperación semántica para $k \geq 0$.

4. La última fase de la búsqueda consiste en devolver todo el conjunto de instancias de las fases anteriores; es decir, del conjunto de conceptos de entrada en la consulta, es necesario devolver las instancias para $k = 0$ (fase 2) que representa un valor exacto y $k \geq 0$ (fase 3). Por otra parte, ya en la presentación de resultados se muestran las ubicaciones para poder recuperar a los objetos geográficos dentro los diferentes repositorios o lugares donde se localizan estos objetos, los cuales posteriormente serán recuperados. Además se provee de información sobre sus características directamente extraídas de la especificación planteada en el dominio FDGC de estas instancias, esta información puede ser su nombre, descripción, fecha de creación, etc. En la Figura 4.15 se muestra el proceso final de recuperación semántica.

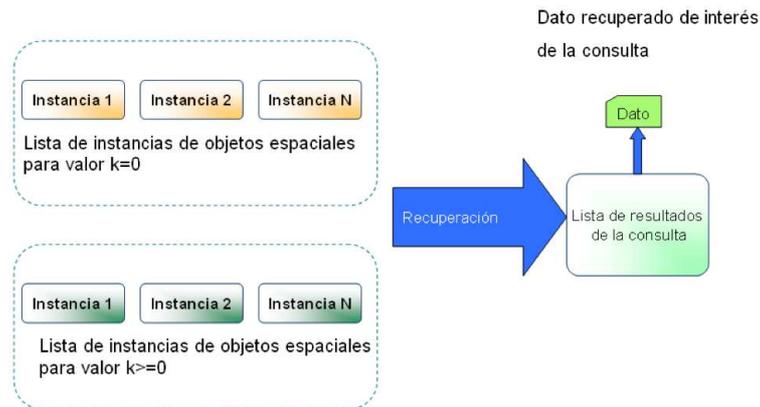


Figura 4.15. Resultado final de la recuperación semántica.

4.3. Modelado del sistema *SemGSearch*

El modelado del sistema de recuperación semántica se muestra a continuación en las siguientes figuras, de las cuales se pueden observar los casos de uso, los diagramas de secuencia, el modelo entidad – relación del modelo persistente, así como la arquitectura del sistema *SemGSearch*. En las Figuras 4.16 y 4.17 se muestran los diagramas de caso de uso para un usuario que desea realizar una búsqueda y para un usuario administrador, así mismo, en las Figuras 4.18 y 4.19 se describe con diagramas de secuencia el proceso de carga de dominios en el sistema *SemGSearch* y el proceso de agregar un nuevo servidor y poblar la ontología.

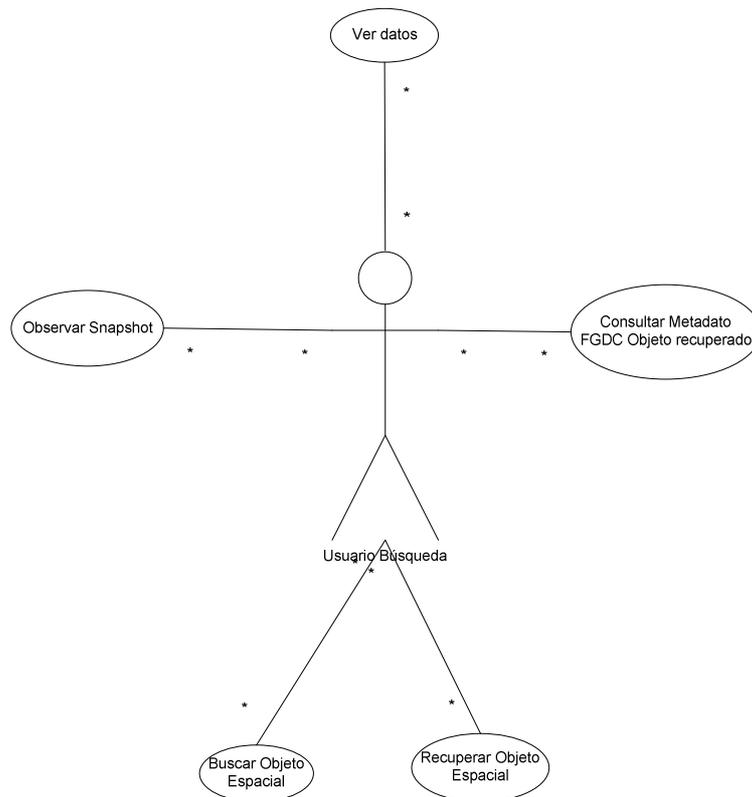


Figura 4.16. Caso de uso – Usuario que desea buscar y recuperar un objeto geoespacial.

Tabla 4.4. Descripción de alto nivel del caso de uso usuario de búsqueda

<Identificador>	<Nombre descriptivo>
Ver datos	Ver datos propios del objeto recuperado como nombre, descripción propia, tamaño fecha servidor donde se encuentra
Observar Snapshot	Observar Snapshot de cada Objeto recuperado
Buscar Objeto Espacial	Ver características de los datos espaciales por medio de su metadatos FGDC
Recuperar Objeto Espacial	Recuperar objeto espacial con posibilidad de descarga o visualización dependiendo del tipo de archivo

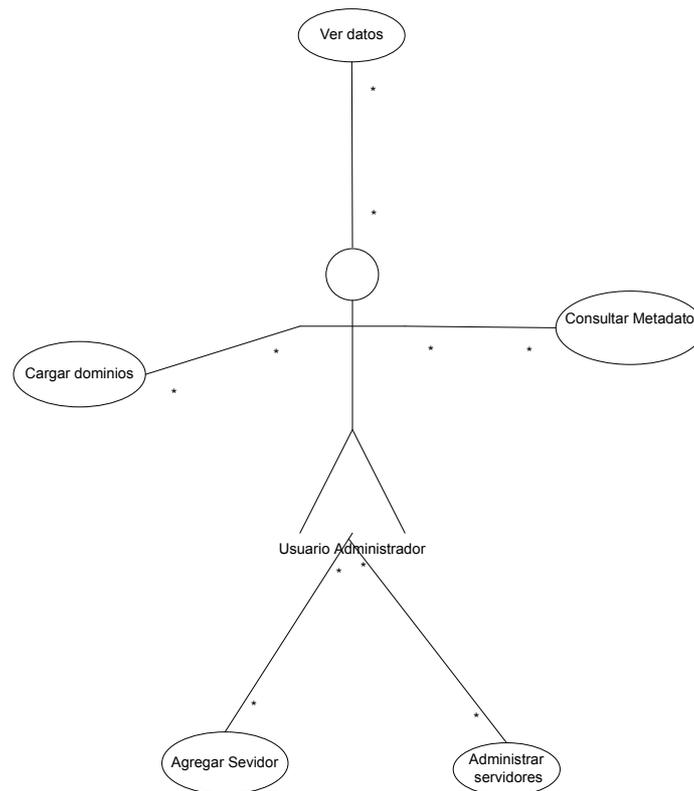


Figura 4.17. Caso de uso – Usuario administrador del sistema *SemGSearch*.

Tabla 4.5. Descripción de alto nivel del caso de uso usuario administrador.

<Identificador>	<Nombre descriptivo>
Ver datos	Ver datos propios del objeto almacenado como nombre, descripción propia, tamaño fecha servidor donde se encuentra
Cargar Dominios	Cargar dominios Espacial temático y temporal por medio del archivo OWL/RDF
Consultar metadatos	Consultar Metadato FGDC Objeto almacenado
Agregar Servidor	Agregar nuevo servidor y cargar metadatos
Administrar Servidores	Consultar lista de servidores , actualizar metadatos en el servidor o eliminar servidor



Figura 4.18. Diagrama de secuencia para carga de dominios OWL/RDF del sistema *SemGSearch*.

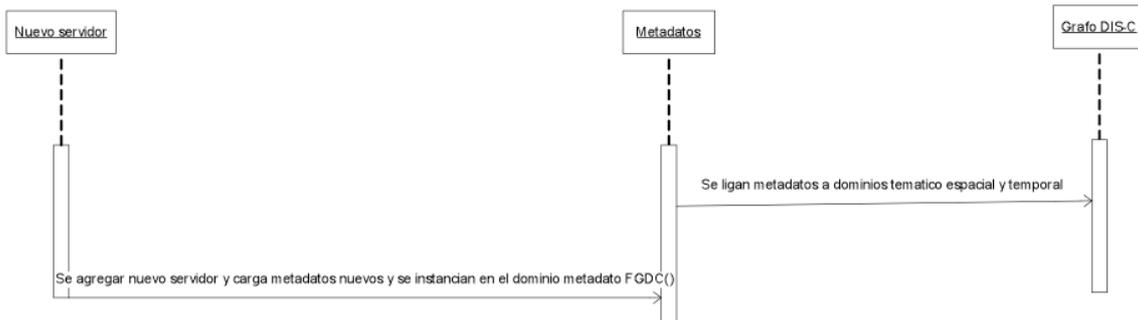


Figura 4.19. Diagrama de secuencia para agregar nuevo servidor, carga de metadatos y poblar ontología.

En el sistema se encuentran cuatros bases de datos, las cuales están relacionadas con respecto a los dominios: Metadato FGDC, temporal, espacial y temático; las cuales fueron convertidas directamente de la ontología a un modelo persistente de base de datos e implementadas en MySQL. Asimismo, se ha implementado una base de datos adicional para la administración del sistema. En la Figura 4.20 se muestran las bases de datos del sistema *SemGSearch*.

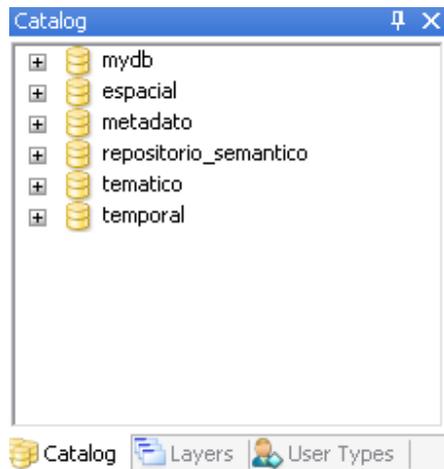


Figura 4.20. Bases de datos que integran al sistema *SemGSearch*.

Por otra parte, el modelo entidad – relación de las bases de datos se muestra en la Figura 4.21.

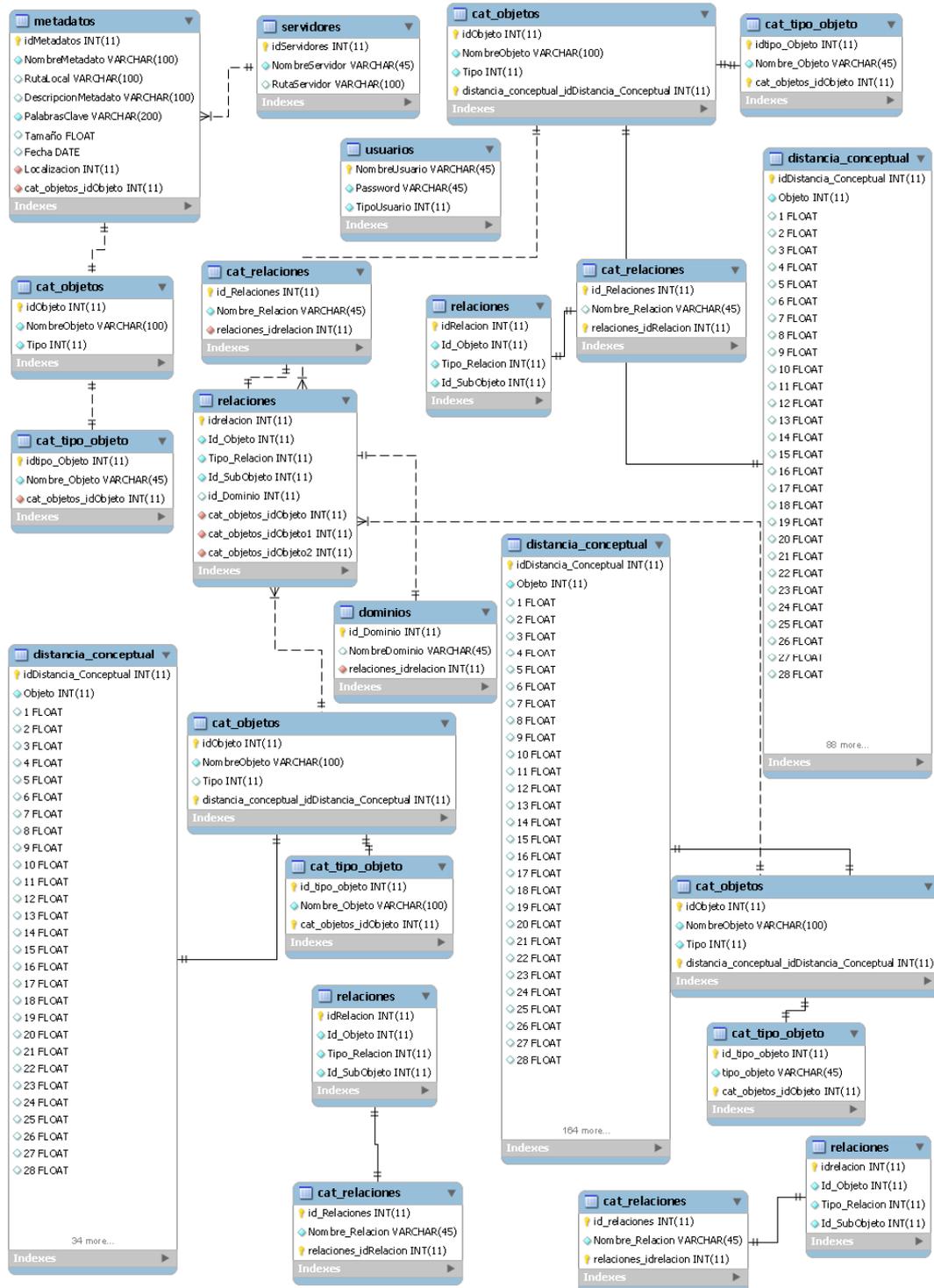


Figura 4.21. Modelo entidad – relación de las bases de datos del sistema SemGSearch.

De igual forma, en la Figura 4.22 se presenta la arquitectura del sistema *SemGSearch*, la cual se diseñó a través de un modelo empresarial de tres capas, por medio de la tecnología EJB (Enterprise Java Beans).

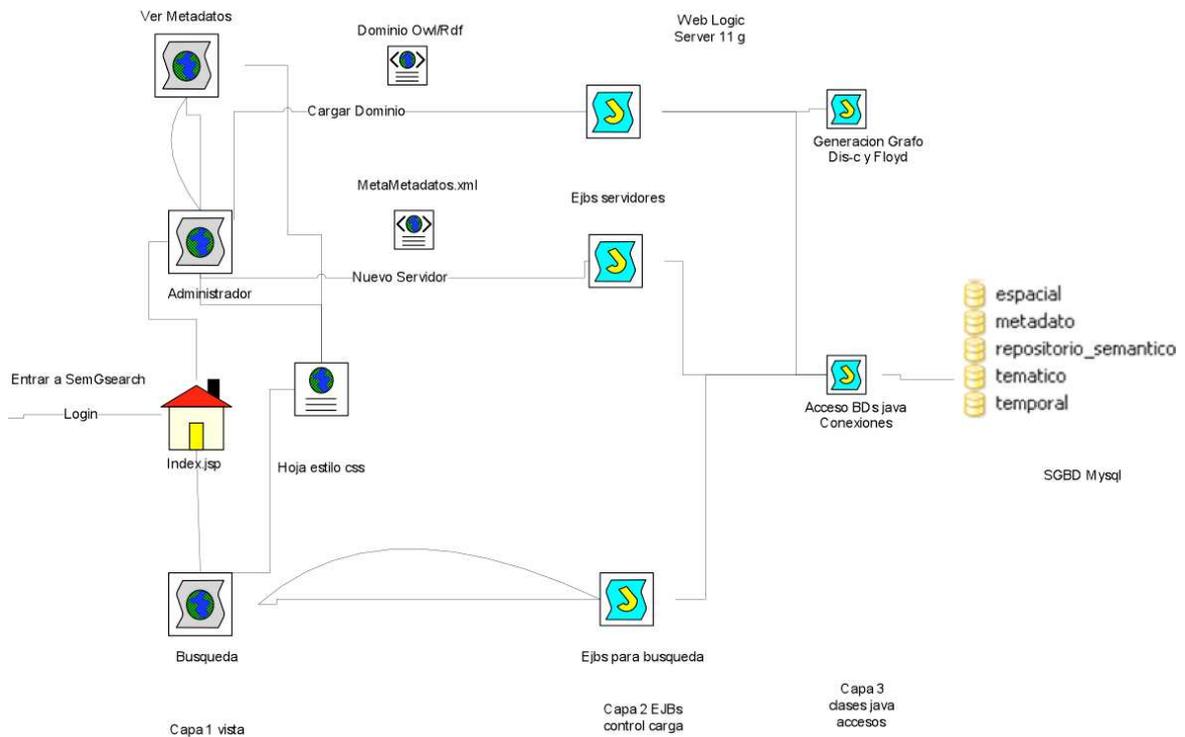


Figura 4.22. Arquitectura del sistema *SemGSearch*.

Capítulo 5. Resultados Experimentales

5.1. Introducción

En esta sección se presentan los resultados obtenidos en esta tesis. Los resultados están orientados básicamente en la metodología de recuperación semántica en fuentes de datos heterogéneas, así como la implementación de la ontología considerando los tres dominios y la especificación FGDC referente a los metadatos

De igual forma, se presenta como casos de estudio, el análisis de los dominios temáticos de:

- Áreas urbanas
- Vías de comunicación terrestres
- Recursos forestales

En el *dominio espacial* se contempla la República Mexicana con distintas clasificaciones (ver Anexo 1); es decir, la división ecorregional, según la CONABIO (Comisión Nacional para el Conocimiento y Uso de la Biodiversidad), división zona espacial según CFE (Comisión Federal de Electricidad) y por zona económica. Para el *dominio temporal* se utilizan las fechas con variantes respectivas a las fuentes de datos. Finalmente, el *dominio temático* está descrito por la información referente a los metadatos de las instituciones antes mencionadas.

5.2. Resultados etapa de conceptualización

Los dominios mencionados anteriormente, han sido conceptualizados en una ontología; así mismo, los metadatos de la FGDC han sido definidos en la misma ontología para describir a la información geoespacial. Esta ontología ha sido diseñada bajo la metodología GEONTO-MET e implementada en Protégé.

La conceptualización del dominio temático ha sido considerada con la fuente del diccionario de INEGI. Además con el fin de enriquecer el dominio se agregaron más conceptos, los cuales fueron tomados directamente de la norma de la Secretaría del Medio Ambiente del Distrito Federal.

A continuación se muestran en Figuras 5.1 y 5.2 la estructura general de la ontología con los cuatro dominios mencionados anteriormente. Posteriormente se mostrará en forma individual a cada uno de estos dominios. Con un recuadro en color morado se representa el dominio temático, en color rojo el temporal, en color azul el espacial y en color verde el dominio de los Metadatos FGDC.

La ontología cuenta con 450 conceptos distribuidos en los cuatro dominios, siendo el de mayor número, el dominio temático; posteriormente espacial, el dominio de Metadatos y temporal sucesivamente. Los conceptos son relacionados con alrededor de 150 relaciones del tipo “es”, 50 del tipo “tiene” y finalmente del tipo “ejecuta” aproximadamente 40. Esta ontología se muestra en forma de tipo de paquetes.

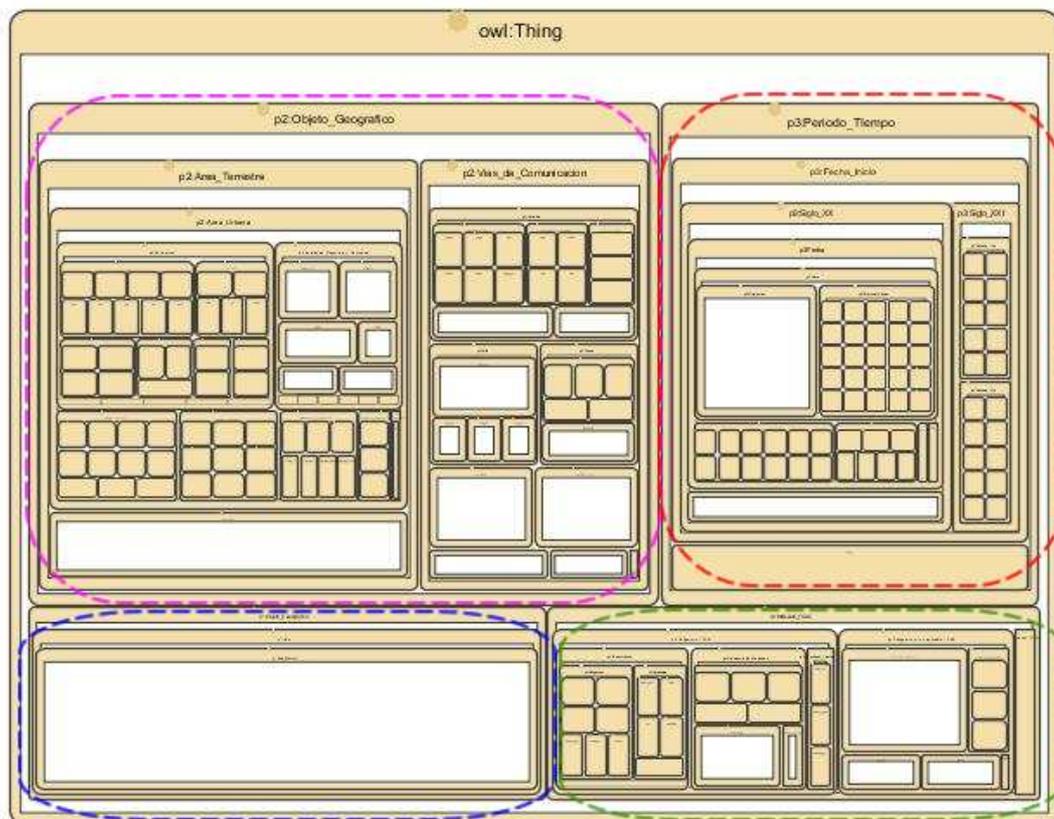


Figura 5.1. Ontología general visualizada en paquetes.

A continuación en la Figura 5.2 se muestra a la ontología mediante una estructura arbórea, la cual es difícil de apreciar por la cantidad de conceptos y relaciones descritas explícitamente en la misma.

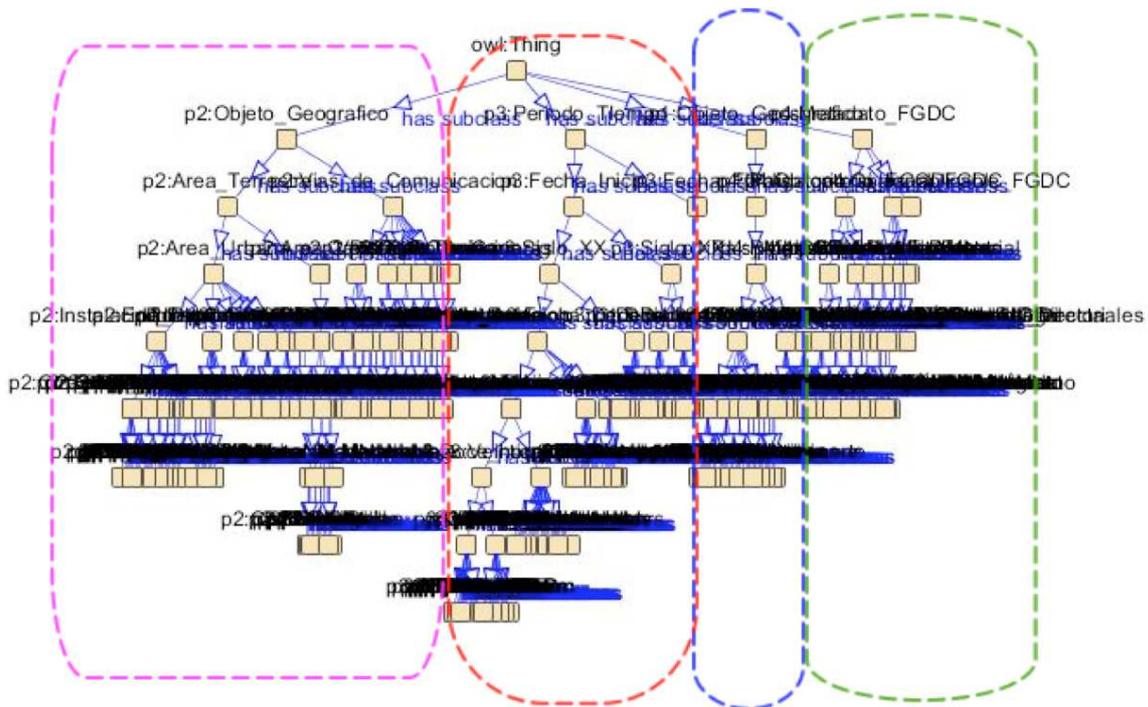


Figura 5.2. Ontología general visualizada en forma de árbol.

A partir de las figuras anteriores se procede a mostrar con mayor detalle cada dominio, iniciando por la partición correspondiente al dominio temático, la cual se muestra en la Figura 5.3. En este dominio la estructura principal tiene como base el diccionario de datos del INEGI bajo la relación “*es*”. Posteriormente, para incrementar la base de conocimiento se agregaron conceptos en las últimas particiones de la ontología, además de unir estos conceptos con las relaciones “*tiene*” y “*ejecuta*”.

De este dominio podemos describir que de un objeto geográfico se desprende un “objeto terrestre”, este último puede ser un “área urbana”, compuesta por “edificación” de muchos tipos como son: “escuelas”, “aduanas”, “edificios”, “hospitales”, etc. Todos estos conceptos han sido descritos en las particiones de “área urbana”, el “objeto terrestre” también puede ser un “área verde”, compuesto de “franjas de bosques”, “arbustos”, “plantas” etc., cada uno de estos conceptos con su taxonomía correspondiente, a decir, el “área urbana” y “área verde” son conectados por “vías de comunicación” como son “calles”, “puentes”, etc.

Todo el esquema anterior se construyó bajo la relación “*es*” y algunas relaciones como “*tiene*”. En las últimas particiones se relacionaron conceptos nuevos no incluidos en el diccionario del INEGI, como por ejemplo en una “aduana” se “*ejecuta*” “exportaciones” y “*tiene*” “agentes aduanales”.

En la Figura 5.7 se muestra la clase de área verde para la partición del dominio temático.

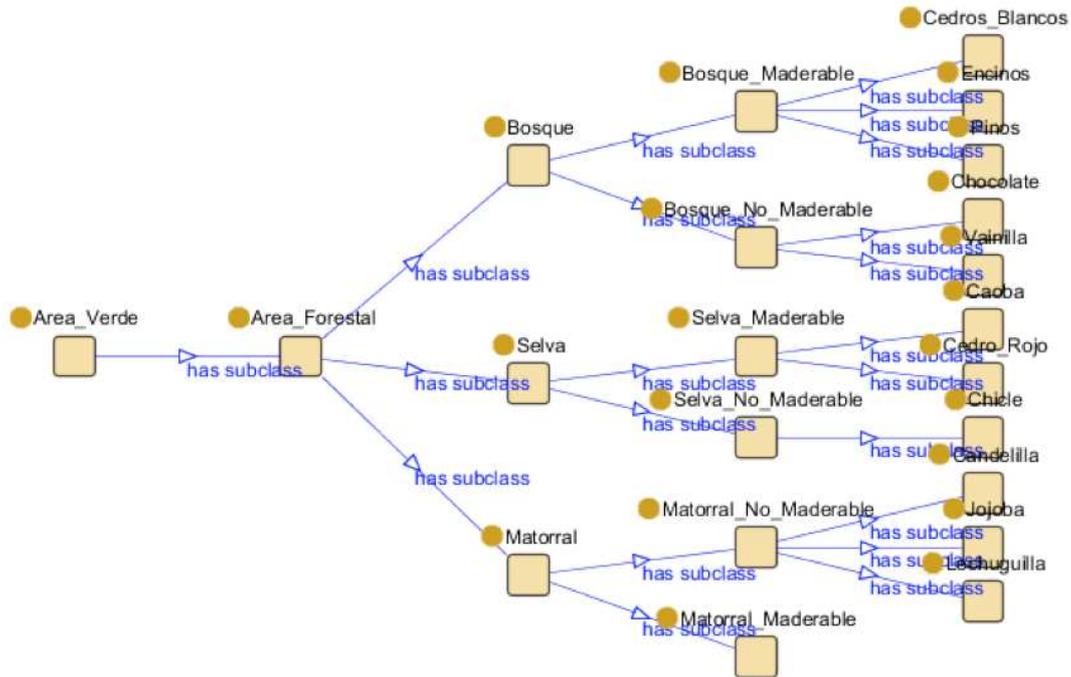


Figura 5.7. Clase área verde con sus conceptos y relaciones.

Ya que se tiene descrito el dominio temático del mismo modo se conceptualiza el dominio Espacial. En este caso, la República Mexicana fue dividida bajo los perfiles de CONABIO (ecorregional) CFE (económico) y por zona territorial; Estas divisiones se encuentran en el Anexo 1.

El esquema comienza a partir de un objeto geográfico que “es” un concepto “país” y se especifica al de “México”. Este último se relaciona bajo la relación “tiene”, una división ecorregional por zonas espaciales y económica, define a cada una de las anteriores divisiones en la siguiente subpartición. Por último y en la parte final de este dominio de la ontología todos los conceptos resultantes de las divisiones se relacionan con un estado de la República Mexicana. Cabe señalar que de acuerdo con lo anterior, un estado puede tener una o más clasificaciones; por ejemplo, el estado de Chiapas está relacionado con Veracruz por medio de la división ecorregional, porque poseen “selvas calido-húmedas”, al mismo tiempo que Chiapas se relaciona con Tabasco bajo el esquema de “zona espacial sureste”. Al tener varios criterios para definir una cercanía conceptual en cuanto al ámbito espacial, ayuda a hacer más flexible y dar más gama de posibilidades en la recuperación. En la Figura 5.8 se muestra la partición de la ontología correspondiente al dominio espacial.

En la Figura 5.10 se muestra la clase de división económica para la partición del dominio espacial.

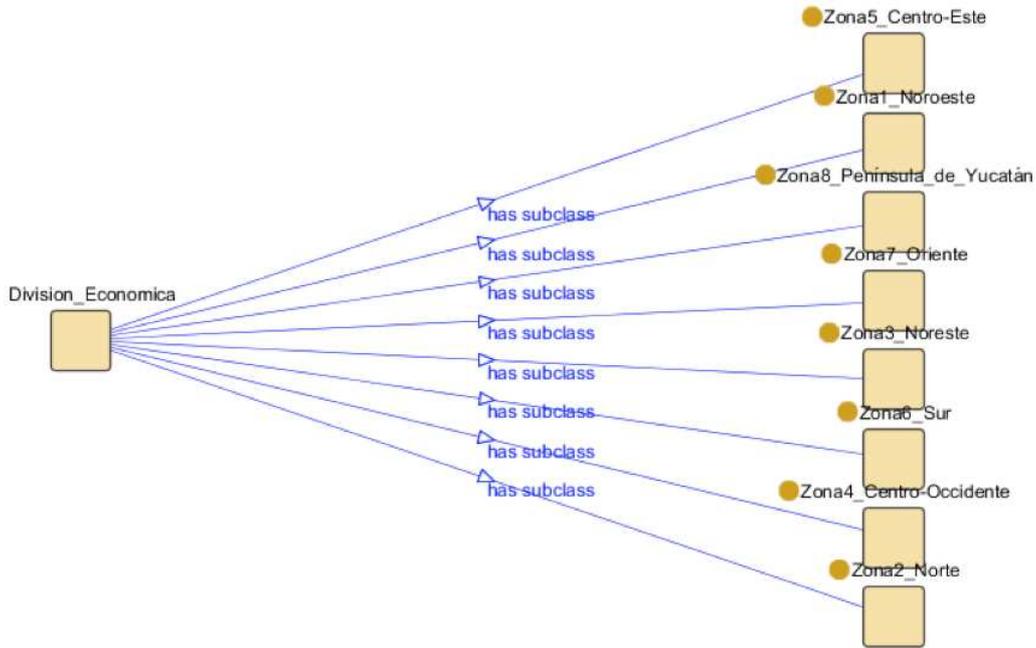


Figura 5.10. Clase división económica con sus conceptos y relaciones.

En la Figura 5.11 se muestra la clase de división territorial o por zona espacial para la partición del dominio espacial.

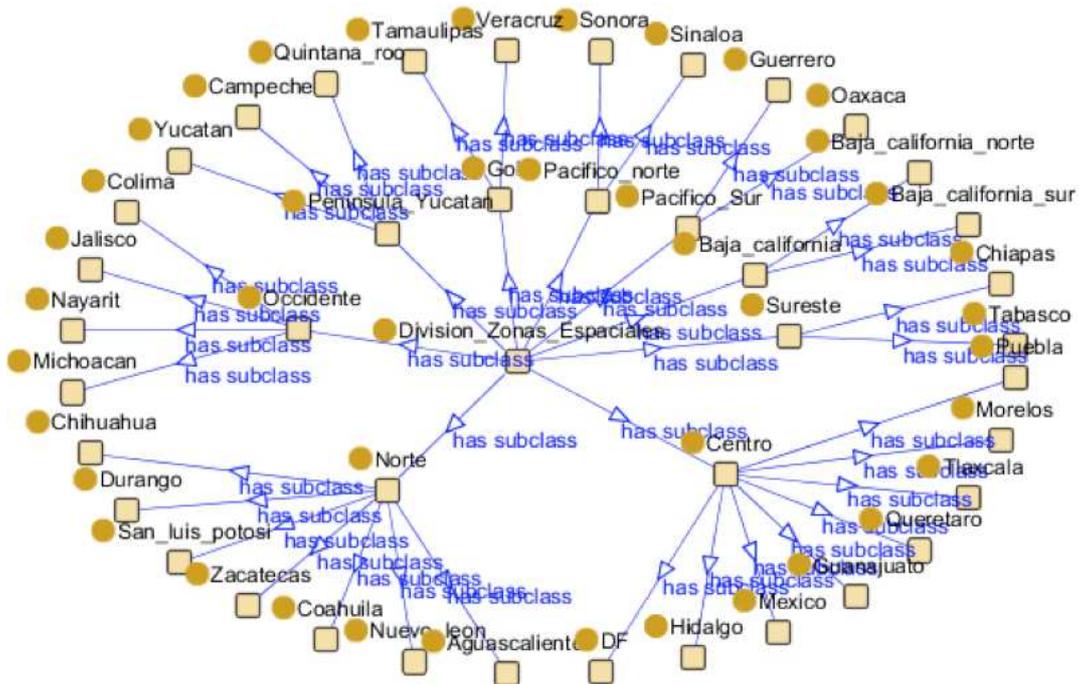


Figura 5.11. Clase división zonas espaciales con sus conceptos y relaciones.

Para la conceptualización del dominio temporal, éste fue conceptualizado sin previa clasificación, por lo que se llevó a cabo mediante el conocimiento *a priori*. De esta forma, la división fue clasificada por décadas. En la Figura 5.12 se muestra la partición del dominio temporal de la ontología de manera general.

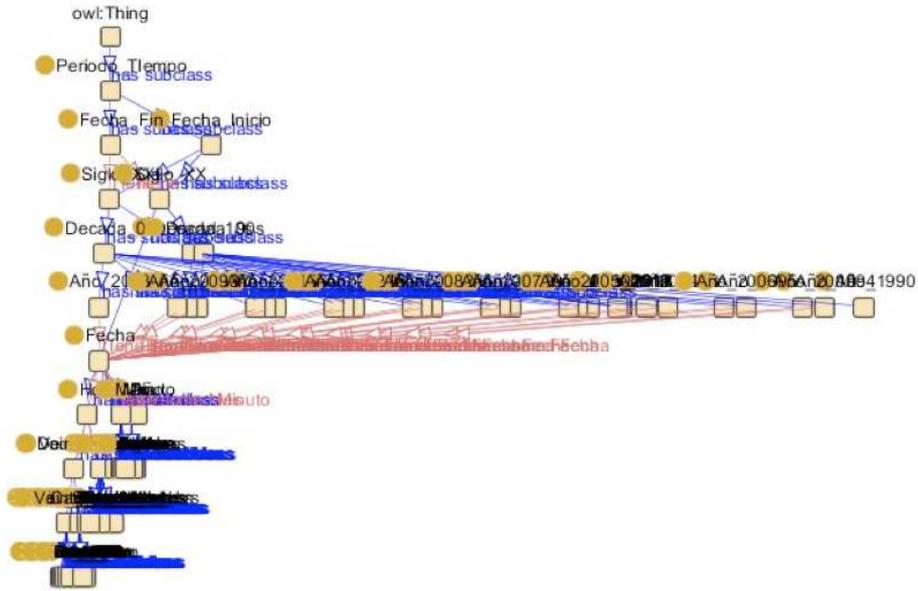


Figura 5.12. Partición de la ontología referente al dominio temporal.

En las Figuras 5.13 se muestra la clase de década, referente al dominio temporal, para este caso se considera la clase siglo XXI.

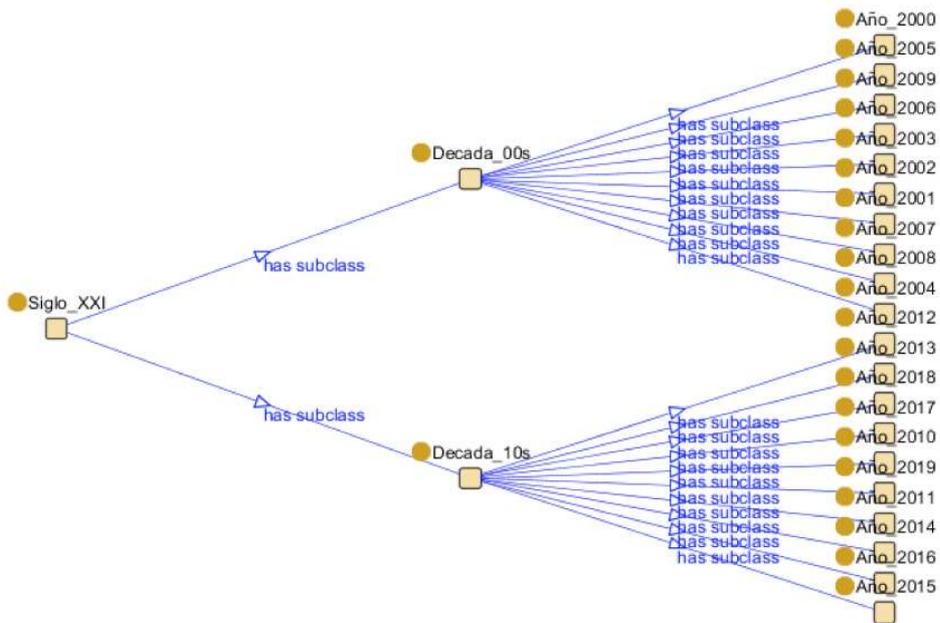


Figura 5.13. Clase siglo XXI con sus conceptos y relaciones.

Posteriormente, en la Figura 5.14 se muestra la clase fecha, la cual se utiliza para definir un periodo de tiempo específico para cada instancia.

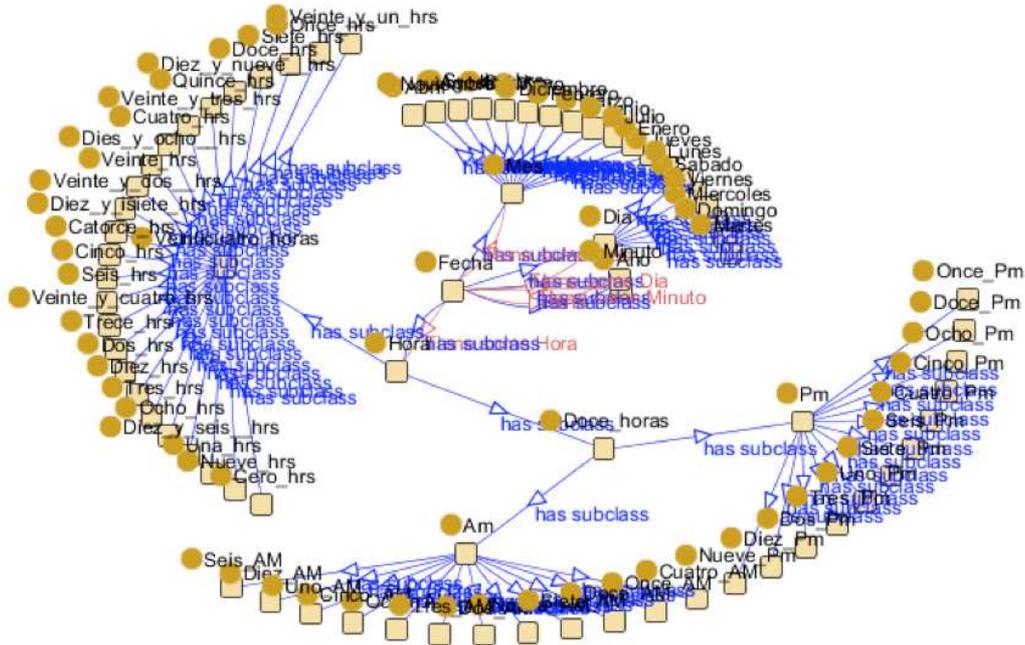


Figura 5.14. Clase fecha con sus conceptos y relaciones.

Finalmente, la ontología cuenta con una partición adicional, la cual se encarga de describir a los metadatos. Para esta conceptualización se tomó como fuente de información a la especificación FGDC. En la Figura 5.15 se muestra la partición de la clase Metadato FGDC con sus conceptos y relaciones.

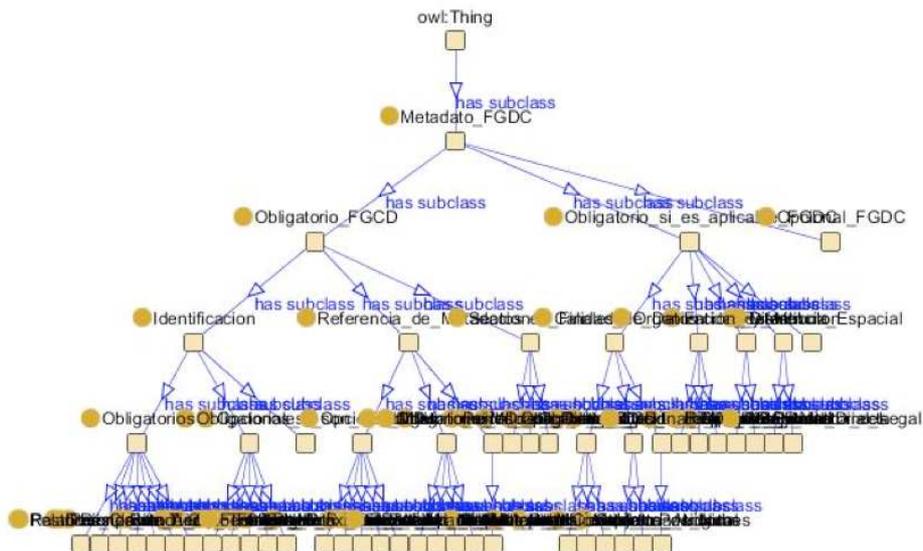


Figura 5.15. Partición de la ontología referente a los metadatos FGDC.

El proceso de conceptualización de los cuatro dominios de la ontología se realizó de forma manual, con el fin de que un experto con conocimientos sobre los dominios aquí descritos proporcionen las bases para construir la ontología en la herramienta Protégé. Como resultado de estos archivos en formato OWL para que sean tomados en el proceso siguiente es convertir la ontología en un *modelo persistente de datos*; de este modo, se podrá almacenar la estructura de la ontología y en la base de datos al agregar un nuevo servidor y por ende nuevos objetos geográficos, con sus respectivos metadatos, con la finalidad para que éstos sean instanciados y almacenados en estos dominios.

Para ello, cada uno de los dominios se crea una base de datos y sus características son convertidas en tablas. Como ejemplo, se describe al dominio espacial. Por lo tanto, en este caso cada dominio se almacena en cuatro tablas, a saber:

1. Tabla de objetos.
2. Tabla de tipo de objeto (clase o instancia).
3. Tabla de relaciones definidas por GEONTO-MET.
4. Tabla de distancias conceptuales.

Además se ha implementado una base de datos adicional para la administración del sistema *SemGsearch*, la cual ha sido nombrada “repositorio semántico” para la administración de los usuarios y servidores registrados, así como los metadatos en cada servidor. En la Figura 5.16 se muestra el sistema de base de datos con las tablas anteriormente descritas que lo componen.



Figura 5.16. Tablas de cada dominio de la ontología en el repositorio semántico.

En la Tabla 5.1 se muestra la tabla correspondiente a los objetos del dominio espacial que han sido previamente conceptualizados y transformados a un modelo persistente.

Tabla 5.1. Tabla de objetos referentes al dominio espacial.

idObjeto	NombreObjeto	Tipo
1	Zona7_Oriente	1
2	Guerrero	1
3	Sureste	1
4	DF	1
5	Zacatecas	1
6	Peninsula_Yucatan	1
7	San_Luis_Potosi	1
8	Jalisco	1
9	Tabasco	1
10	Norte	1
11	Michoacan	1
12	Division_Zonas_Espaciales	1
13	Sonora	1
14	Pais	1
15	Puebla	1
16	California_Mediterranea	1
17	Elevaciones_Semiaridas_Meridionales	1
18	Zona8_Peninsula_de_Yucatan	1
19	Zona6_Sur	1
20	Campeche	1
21	Zona2_Norte	1
22	Golfo	1
se	Grandes_planicies	1
24	Zona3_Noreste	1
25	Tlaxcala	1

En la Tabla 5.2 se muestra la tabla correspondiente al tipo de objeto de la ontología; en este caso si se trata de una clase o bien de una instancia.

Tabla 5.2. Tabla de tipo de objeto.

idtipo_Objeto	Nombre_Objeto
1	Clase
2	Instancia

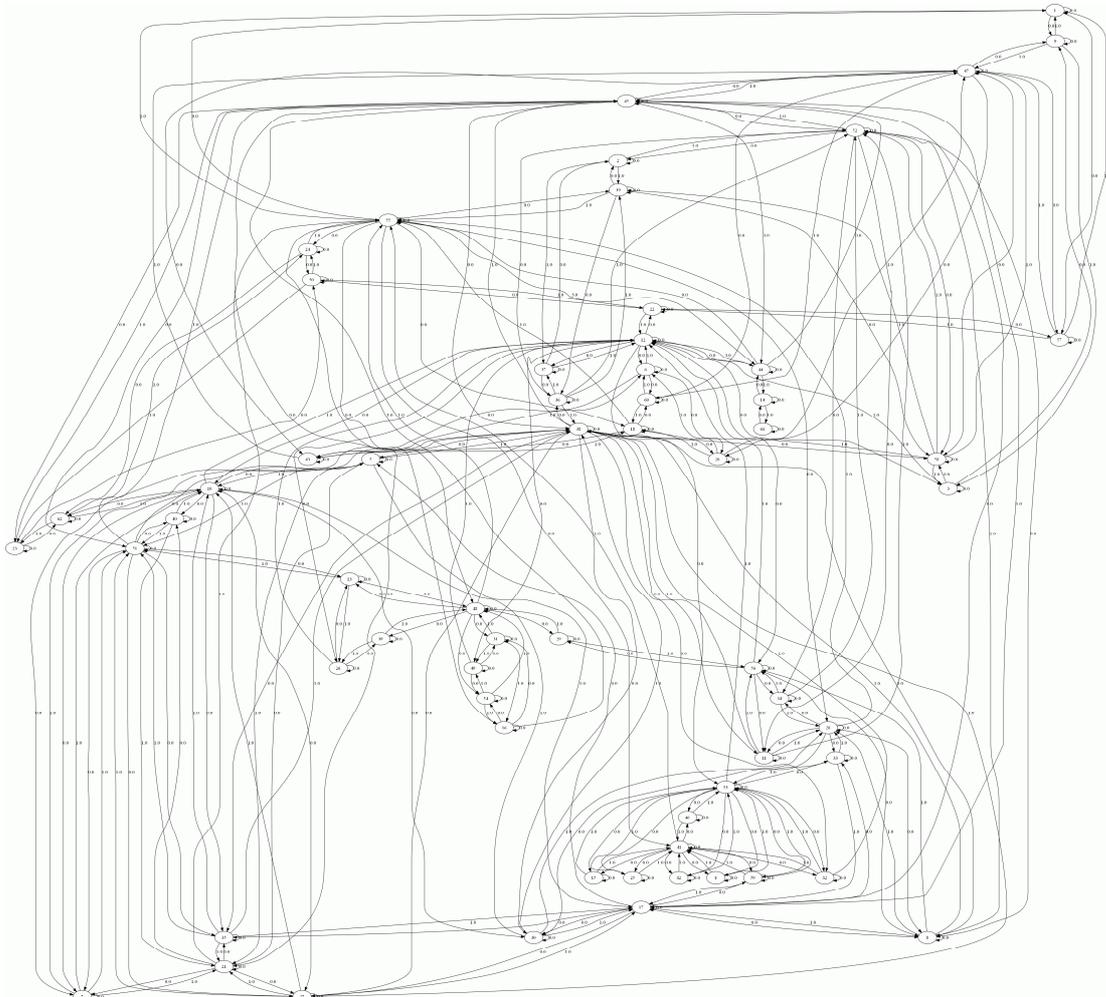
En la Tabla 5.3 se muestran las relaciones definidas en la ontología diseñada, descritas en la metodología GEONTO-MET como relaciones axiomáticas y que representan al dominio conceptualizado.

Tabla 5.3. Tabla de tipo de objeto.

Id_Relaciones	Nombre_Relacion
1	Es
2	Tiene
3	Ejecuta

Para generar la cuarta tabla, referente a las distancias conceptuales entre los conceptos de la ontología, se realizan los siguientes procesos:

Primeramente, se aplica el algoritmo de DIS-C para generar un grafo con pesos ponderados, en donde esta ponderación se basa directamente en la distancia conceptual que se obtiene entre los conceptos. En la Figura 5.17 se muestra el grafo completo para los 62 objetos contemplados en el dominio espacial, esto con la finalidad de observar que se cuenta con un grafo fuertemente conexo al aplicar el algoritmo DIS-C.

**Figura 5.17.** Grafo DIS-C para el dominio espacial.

Por lo tanto, con la finalidad de conocer y visualizar los pesos de la representación anterior, en la Figura 5.18 se muestra un grafo reducido con únicamente 30 objetos del dominio espacial.

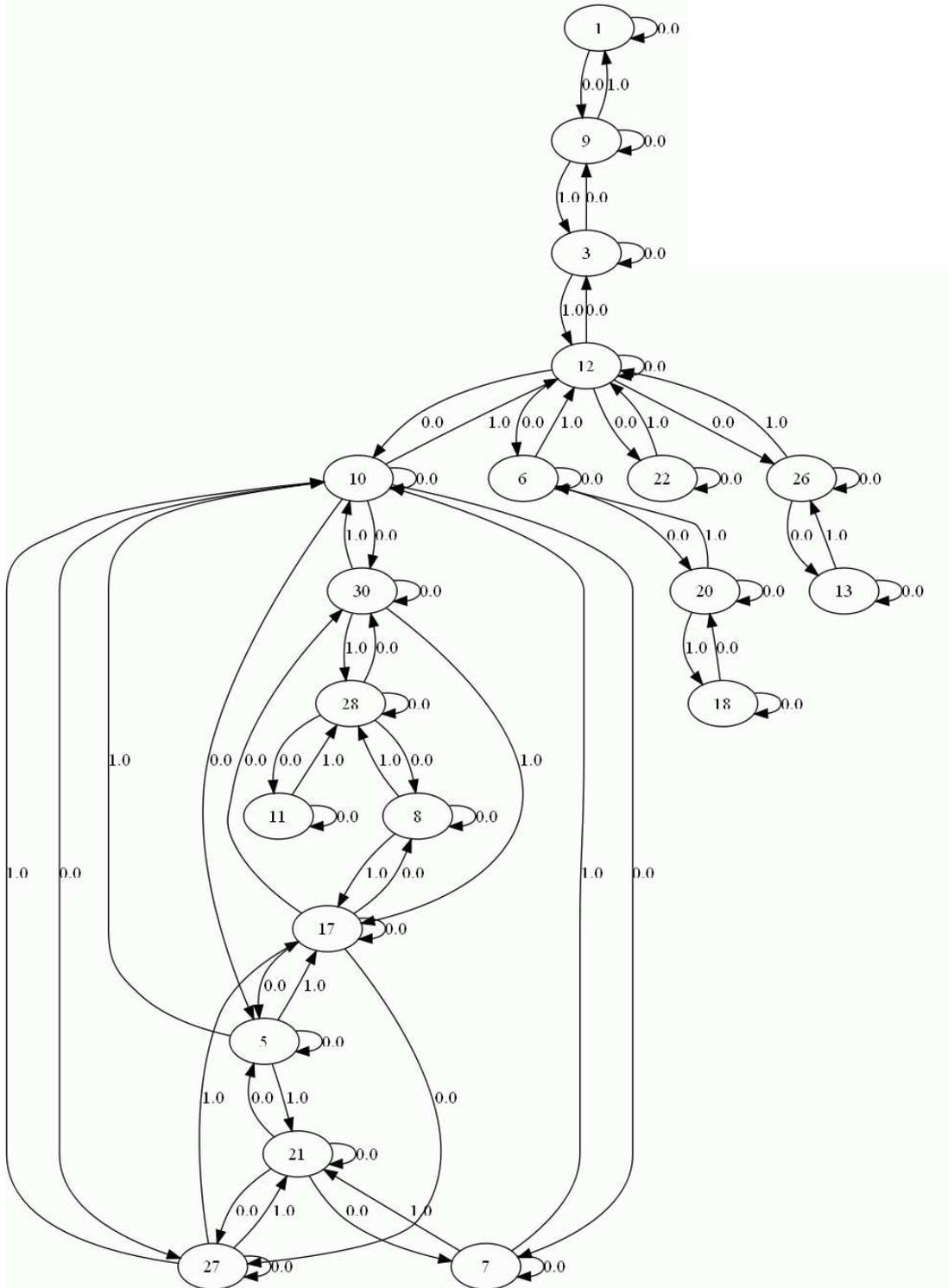


Figura 5.18. Grafo DIS-C para el dominio espacial con 30 objetos.

Posteriormente, el grafo es almacenado en la base de datos por medio de una tabla donde un objeto C_x se relaciona con otro C_y , donde C_x, C_y son conceptos de un dominio de la ontología. En esta tabla se indican los valores numéricos de peso w , según DIS-C. Además se indica por medio de un valor numérico alto los conceptos que no están relacionados y simulan estos el valor de ∞ para que se indique de esta forma en el siguiente proceso la ejecución del algoritmo de Floyd-Warshall, el cual obtendrá los caminos conceptualmente más cortos entre los conceptos que representan a los objetos del dominio espacial. Para el caso únicamente de 13 objetos los valores que se obtienen se muestran en la Tabla 5.4.

Tabla 5.4. Valores del grafo DIS-C.

Objetos	1	2	3	4	5	6	7	8	9	10	11	12	13
1	0	10000	10000	10000	10000	10000	10000	10000	0	10000	10000	10000	10000
2	10000	0	10000	10000	10000	10000	10000	10000	10000	10000	10000	10000	10000
3	10000	10000	0	10000	10000	10000	10000	10000	0	10000	10000	1	10000
4	10000	10000	10000	0	10000	10000	10000	10000	10000	10000	10000	10000	10000
5	10000	10000	10000	10000	0	10000	10000	10000	10000	1	10000	10000	10000
6	10000	10000	10000	10000	10000	0	10000	10000	10000	10000	10000	1	10000
7	10000	10000	10000	10000	10000	10000	0	10000	10000	1	10000	10000	10000
8	10000	10000	10000	10000	10000	10000	10000	0	10000	10000	10000	10000	10000
9	1	10000	1	10000	10000	10000	10000	10000	0	10000	10000	10000	10000
10	10000	10000	10000	10000	0	10000	0	10000	10000	0	10000	1	10000
11	10000	10000	10000	10000	10000	10000	10000	10000	10000	10000	0	10000	10000
12	10000	10000	0	10000	10000	0	10000	10000	10000	0	10000	0	10000
13	10000	10000	10000	10000	10000	10000	10000	10000	10000	10000	10000	10000	0
14	10000	10000	10000	10000	10000	10000	10000	10000	10000	10000	10000	10000	10000

Finalmente, es necesario obtener el camino más corto para cada concepto (objeto) de la ontología, por lo cual se aplica el algoritmo de Floyd-Warshall. La matriz resultante del algoritmo es guardada en tabla de la base de datos para que sea consulta en la fase de recuperación, se muestra los valores numéricos para un concepto C_x relacionado con otro C_y que tienen un peso w como medida de distancia conceptual, donde C_x, C_y son conceptos de un dominio de la ontología. Por ejemplo se muestra en color verde el vector del concepto C_1 con sus respectivas distancias conceptuales a los otros conceptos C_1, C_2, \dots, C_n .

En este caso, en la Tabla 5.5 se muestra el valor de cada camino para 25 objetos que pertenecen al dominio espacial.

Tabla 5.5. Tabla de distancias conceptuales más cortas entre todos los conceptos, aplicando el algoritmo de Floyd-Warshall.

Id Distancia Conceptual	Concepto	c ₁	c ₂	c ₃	c ₄	c ₅	c ₆	c ₇	c ₈	c ₉	c ₁₀	c ₁₁	c ₁₂	c ₁₃	c ₁₄	c ₁₅	c ₁₆	c ₁₇	c ₁₈	c ₁₉
1	1	0	1	1	1	1	2	1	1	0	2	1	2	1	5	1	2	2	1	1
2	2	2	0	2	2	2	2	2	1	2	2	1	2	2	6	2	2	2	2	1
3	3	1	1	0	1	1	1	1	1	0	1	1	1	1	5	1	2	2	2	1
4	4	2	2	2	0	2	2	2	2	2	2	2	2	2	6	1	3	2	2	2
5	5	2	2	2	2	0	2	1	1	2	1	1	2	1	6	2	2	1	2	2
6	6	2	1	1	1	1	0	1	1	1	1	1	1	1	5	1	2	2	1	2
7	7	2	2	2	2	1	2	0	1	2	1	1	2	1	6	2	2	2	2	2
8	8	2	1	2	2	1	2	1	0	2	2	1	2	2	6	2	2	1	2	2
9	9	1	2	1	2	2	2	2	2	0	2	2	2	2	6	2	2	2	2	2
10	10	2	1	1	1	0	1	0	1	1	0	1	1	1	5	1	2	1	2	2
11	11	2	1	2	2	1	2	1	1	2	2	0	2	2	6	2	2	2	2	2
12	12	1	0	0	0	0	0	0	0	0	0	0	0	0	4	0	1	1	1	1
13	13	2	2	2	2	1	2	1	2	2	2	2	2	0	6	2	2	2	2	2
14	14	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
15	15	2	2	2	1	2	2	2	2	2	2	2	2	2	6	0	3	2	2	2
16	16	2	1	2	2	1	2	1	1	1	2	1	2	1	5	2	0	1	2	2
17	17	2	1	2	1	0	2	1	0	1	1	1	2	1	5	1	1	0	2	2
18	18	1	1	2	1	1	1	1	1	1	2	1	2	1	5	1	2	2	0	1
19	19	1	0	1	1	1	2	1	1	1	2	1	2	1	5	1	2	2	1	0

Entonces, funcionalmente y sobre el sistema *SemGsearch* lo anteriormente descrito, se realiza de forma automática, tomando un archivo por cada dominio desarrollado en Protegé en formato OWL/RDF para realizar la conversión de la ontología a un modelo persistente de base de datos, la generación de su grafo DIS-C, la resolución de menor distancias conceptuales por Floyd-Washall y como resultado la matriz o tabla de distancias conceptuales entre los conceptos de cada dominio.

Este proceso quien tiene los privilegios para realizarlo es el usuario administrador, depositando cada archivo de un dominio de la conceptualización a una carpeta en localizada en *c://OWIDominios/*. Posteriormente, se cargan los dominios tanto espacial, temático y temporal utilizando la interfaz que ha sido desarrollada para este proceso y que se muestra en la Figura 5.19.

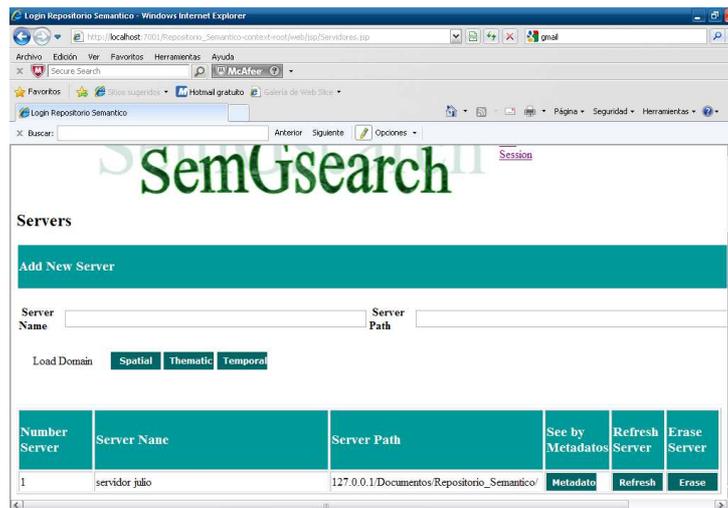


Figura 5.19. Interfaz para agregar servidores a SemGSearch.

Cabe señalar que la carga de los dominios no necesariamente se debe realizar en un mismo servidor, puede darse el caso que algún dominio se encuentre almacenado en algún otro equipo; por lo cual este sistema es distribuido y no necesariamente trabaja en un ambiente *standalone*.

5.3. Resultados etapa de síntesis

En esta etapa, la ontología con los dominios espacial, temático y temporal es poblada con cada uno de los datos localizados en los diversos repositorios. El proceso se realiza mediante la creación de instancias, por medio de sus respectivos metadatos.

Este proceso de crear instancias se realiza ingresando el nombre del servidor y la IP (Internet Protocol) dentro de la Intranet y la ruta del archivo que describe a los metadatos que serán compartidos; el archivo de metadatos ha sido llamado *meta_metadatos.xml* y fue construido bajo la estructura de XML. En este caso, los metadatos serán instancias en la ontología de metadatos.

En la etapa de síntesis, por cada objeto espacial localizado en el repositorio se crea una instancia en el dominio de Metadatos bajo la especificación FGDC, después se lee su respectivo metadato, específicamente las *palabras clave* para crear un grupo de conceptos, de los cuales se crean instancias enlazadas a los otros tres dominios conceptualizados; es decir, se enlaza cada metadato a los dominios espacial, temporal y temático. Por ejemplo, en la Figura 5.20 se muestra el enlace de cada metadato a los dominios.

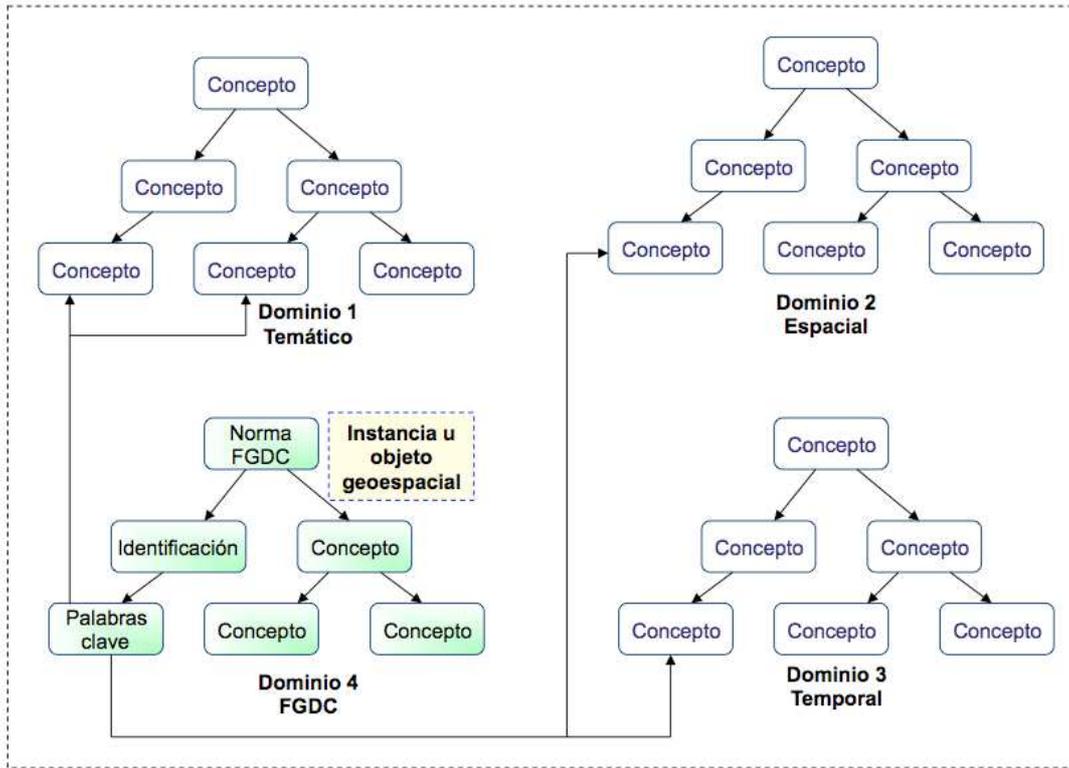


Figura 5.20. Enlace de metadatos con los dominios conceptualizados en la ontología.

Posteriormente, en el sistema se da de alta el servidor, ingresando el nombre del servidor y la ruta en la Intranet para localizar el archivo de *meta_metadatos.xml*; esa carpeta compartida alojará en cada subcarpeta local cada objeto geográfico, el cual se compone del archivo que se va a compartir, visualizando una imagen previa del contenido (snapshot); así como su descripción por metadato FGDC. En la Figura 5.21 se muestra la interfaz para realizar esta carga de datos.



Figura 5.21. Carga de archivo de metadatos desde el servidor.

Por lo tanto, en esta ruta se encuentra el archivo *meta_metadatos.xml*, el cual está compuesto por la estructura descrita en la Tabla 5.6, la cual muestra únicamente cinco metadatos. Asimismo, en el Anexo 2 se muestran todos los elementos del archivo de metadatos, con la finalidad de observar los resultados de las búsquedas posteriores.

Tabla 5.6. Estructura del archivo de metadatos.

```
<?xml version="1.0" encoding="ISO-8859-1" ?>
<Metadatos>
  <Metadato Numero="1">
    <NombreMetadato>boulevard_pue.zip</NombreMetadato>
    <RutaLocal>Metadato_1</RutaLocal>
    <DescripcionMetadato>Shape Boulevard en avenida principal
puebla</DescripcionMetadato>
    <PalabrasClave>boulevard,puebla,año_2008</PalabrasClave>
    <Tamaño>4567</Tamaño>
    <Fecha>2008/01/14</Fecha>
  </Metadato >
  <Metadato Numero="2">
    <NombreMetadato>calles_mor.zip</NombreMetadato>
    <RutaLocal>Metadato_2</RutaLocal>
    <DescripcionMetadato>Shape plano de calles Mexico morelos
</DescripcionMetadato>
    <PalabrasClave>calle,morelos,año_2008</PalabrasClave>
    <Tamaño>2048</Tamaño>
    <Fecha>2008-09-13</Fecha>
  </Metadato >
  <Metadato Numero="3">
    <NombreMetadato>carreteras_mor.zip</NombreMetadato>
    <RutaLocal>Metadato_3</RutaLocal>
    <DescripcionMetadato>Shape Red de carreteras en morelos ano
2008</DescripcionMetadato>
    <PalabrasClave>carretera,morelos,año_2008</PalabrasClave>
    <Tamaño>4567</Tamaño>
    <Fecha>2008/01/14</Fecha>
  </Metadato >
  <Metadato Numero="4">
    <NombreMetadato>viaducto_df.zip</NombreMetadato>
    <RutaLocal>Metadato_4</RutaLocal>
```

```

<DescripcionMetadato>Shape Viaductos en la capital de
Mexico</DescripcionMetadato>
<PalabrasClave>Viaducto,DF,año_2007</PalabrasClave>
<Tamaño>3048</Tamaño>
<Fecha>2007-09-13</Fecha>
</Metadato >
<Metadato Numero="5">
<NombreMetadato>Periferico_chi.zip</NombreMetadato>
<RutaLocal>Metadato_5</RutaLocal>
<DescripcionMetadato>Shape Periferico en chiapas 2009</DescripcionMetadato>
<PalabrasClave>Periferico,chiapas,año_2009</PalabrasClave>
<Tamaño>4567</Tamaño>
<Fecha>2008/01/14</Fecha>
</Metadato >
</Metadatos>
    
```

En la ruta del archivo *meta_metadatos.xml* se localizan, por cada metadato una subcarpeta ubicada en la raíz de la carpeta compartida y en cada subcarpeta se almacenan los tipos de datos que son utilizados para la recuperación semántica (ver Figura 5.22).

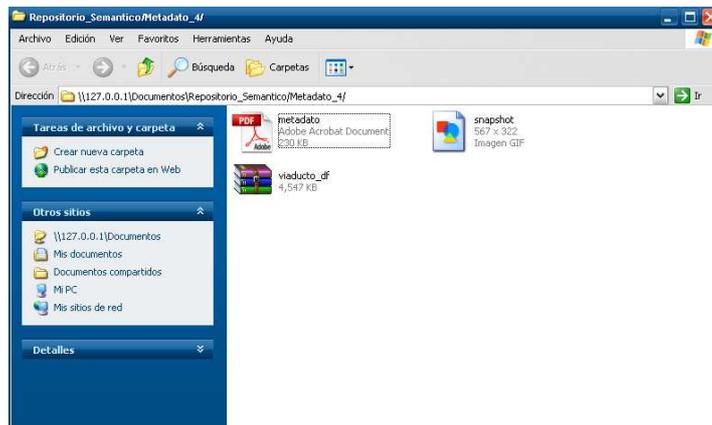


Figura 5.22. Archivos para cada metadato.

Para el proceso de captura, es necesario dar de alta al nuevo servidor en la base de datos. Posteriormente se carga cada metadato, mapeando cada palabra clave a su determinado dominio espacial temporal o temático según corresponda. Los dominios contemplados son el local con valor de “0” que corresponde al dominio del metadato; es decir, a la norma FGDC y los valores “1”, “2” y “3” para cada dominio (ver Tabla 5.7), por lo tanto, estos valores se utilizan para vincular las instancias del dominio de metadatos.

Tabla 5.7. Dominios en el modelo persistente de base de datos.

id_Dominio	NombreDominio
1	Espacial
2	Temático
3	Temporal

En la Tabla 5.8 se describen algunos metadatos o instancias del dominio metadato FGDC, almacenados en la base de datos.

Tabla 5.8. Tabla de instancias del dominio metadato FGDC.

Nombre_Metadato	Ruta_Local	Descripción_Metadato	Palabras_Clave	Tamaño	Fecha
boulevard_pue.zip	Metadato_1/	Shape Boulevard en avenida principal Puebla	boulevard,puebla,año_2008	4567	14/01/2008
calles_mor.zip	Metadato_2/	Shape plano de calles México Morelos	calle,morelos,año_2008	2048	13/09/2008
carreteras_mor.zip	Metadato_3/	Shape Red de carreteras en Morelos año 2008	carretera,morelos,año_2008	4567	14/01/2008
viaducto_df.zip	Metadato_4/	Shape Viaductos en la capital de México	Viaducto,DF,año_2007	3048	13/09/2007
Periferico_chi.zip	Metadato_5/	Shape Periférico en Chiapas 2009	Periferico,chiapas,año_2009	4567	14/01/2008
vereda_tla.zip	Metadato_6/	Shape veredas rurales orillas de Tlaxcala	vereda,tlaxcala,año_2007	2048	13/09/2007
calle_enlaze_pue.zip	Metadato_7/	Shape calles que enlazan a carreteras de Puebla	calle_enlace_carretera,puebla,año_2006	5567	14/01/2006
calle_morelos.zip	Metadato_8/	Shape Calles de Morelos	calle,morelos,año_2007	2048	13/04/2007
punte_guerrero.zip	Metadato_9/	Shape línea de conexión puentes en Guerrero 2007	punte,guerrero,año_2007	4567	14/03/2007

Una restricción que se utiliza para identificar entre elementos que son *conceptos* y los que son *instancias*, es necesario insertar en el dominio de metadatos FGDC el valor de “2”, tal como se muestra en la Tabla 5.9.

Tabla 5.9. Definición de instancias en el dominio de metadatos.

idObjeto	NombreObjeto	Tipo
60	Secciones_Finales	1
61	Cubierta_de_Nubes	1
62	Distribuidor	1
88		1 2
89		2 2
90		3 2
91		4 2
92		5 2
93		6 2
94		7 2
95		8 2
96		9 2

Por último en la Tabla 5.10 se muestran las relaciones que se vinculan a cada dominio de acuerdo con las palabras clave para cada metadato.

Tabla 5.10. Relaciones vinculadas a cada dominio en la ontología.

Id_relación	Id_Objeto	Tipo_Relación	Id_SubObjeto	Id_Dominio
55	46	1	59	0
56	13	1	60	0
57	56	1	61	0
58	6	1	62	0
110	1	1	59	2
111	1	1	15	1
112	1	1	81	3
113	2	1	41	2
114	2	1	42	1
115	2	1	81	3
116	3	1	35	2
117	3	1	42	1
118	3	1	81	3
119	4	1	13	2
120	4	1	4	1
121	4	1	58	3
122	5	1	126	2
123	5	1	58	1
124	5	1	96	3
125	6	1	150	2
126	6	1	25	1
127	6	1	58	3

De esta forma, se realiza el proceso de poblar la ontología en el dominio del metadato y ligado a los tres restantes: temático, temporal y espacial; con este proceso se concluye con la fase de síntesis de la metodología .

5.3. Resultados etapa de análisis

En la etapa de análisis se lleva a cabo el proceso de recuperación semántica, así como las tareas de visualización de los datos geoespaciales recuperados. En el sistema *SemGsearch* se realiza la petición de una consulta, acotando que tipo de objetos geográficos se buscarán. Estos conceptos introducidos se agrupan conforme a un dominio específico, por medio de una sintaxis de consulta y se genera un vector para el proceso semántico de análisis.

El paso siguiente con este vector es buscar cada concepto en su correspondiente dominio almacenado en la base de datos por el modelo persistente; es decir, el de la base de datos temático, espacial y temporal; en caso de que sí se encuentre o exista una correspondencia

léxica exacta o similar de este concepto, entonces se redirige el proceso a la base de datos de metadatos y como resultado se recuperan los objetos geográficos con estos metadatos, esto se cumple para el criterio con una distancia conceptual $k = 0$.

Posteriormente cuando ya se tiene el concepto base $k = 0$, se consultan las tablas de distancias conceptuales generadas por Floyd-Warshall para obtener conceptos y continuar el proceso de redirigirse a la base de datos de metadatos y recuperar más objetos geográficos para valores $k \geq 0$.

Al término de este proceso, y ya con los objetos geográficos recuperados y agrupados bajo el dominio temático, espacial y temporal, se consultan sus *palabras clave* y se realiza una intersección de cada grupo obtenido, con la finalidad de hacer más precisa la búsqueda con conceptos comunes y se devuelven al usuario.

Lo anterior se repite y se termina indicando que valor de k se solicita; es decir, al principio solo se muestran objetos geográficos recuperados con un valor de $k = 0$ y si se desean observar más resultados la siguiente visualización será para un valor conceptual de $k+1$.

Finalmente en el proceso de visualización con la lista de objetos resultantes de la intersección, éstos se enlistan por medio de un *ranking*, mostrando sus características descritas en el estándar FGDC, definidas y almacenadas en la base de datos del dominio metadatos; en otras palabras, se enlistan los objetos geográficos mostrando su nombre, tamaño, descripción, fecha de creación etc. Además de estos datos se muestra un *snapshot* del objeto en cuestión para su mejor visualización.

Cabe señalar que la metodología fue diseñada para evitar resultados nulos; en el sentido de que sino se encuentra una coincidencia exacta en la búsqueda de conceptos, se muestran los siguientes conceptos (datos geoespaciales) parecidos semánticamente. No obstante, en la implementación del *SemGSearch* se realiza el cálculo de la distancia conceptual, con fines de mostrar el crecimiento de radios para obtener siempre resultados, lo cual puede tener una semejanza directa con los sistemas *question answering*. Por lo tanto, el valor de K ; se incrementa en $K > 0$ para encontrar los conceptos geográficos que se asemejan a la consulta original.

5.3.1. Sintaxis de la consulta

De acuerdo con lo descrito anteriormente, es necesario especificar mediante una sintaxis de consulta a los grupos de conceptos y a que dominios pertenecen, con la finalidad de proporcionar más exactitud, reducir el tiempo de procesamiento y de búsqueda. Por lo tanto, se construye un vector de búsqueda en dos fases:

- Primeramente, concepto por concepto, separando cada uno de ellos por medio de comas “,” y respondiendo a las términos “que” objetos geográficos se están buscando, y “donde” se podrían localizar espacialmente o en el ámbito que se desenvuelven y “cuando” para indicar alguna referencia temporal.
- Posteriormente respondiendo a las preguntas anteriores, se generan grupos que indican una correspondencia de conceptos a una pregunta o dominio, esto se lleva a cabo por separadores “,” y se generan *tres segmentos*, el primero de ellos corresponde al dominio “que” o temático, el siguiente al “donde” o espacial y el tercero al temporal o “cuando”.

Por ejemplo: Se solicita recuperar las calles, túneles, hospitales del estado de Puebla y Veracruz, en el año 2006. En este caso la estructura de la consulta quedaría construida de la siguiente forma:

Calles, túneles, hospitales ; Puebla, Veracruz ; 2006

De la forma anterior, se mejora la exactitud de una búsqueda de los términos en su respectivo dominio, devolviendo como resultado el conjunto de objetos geográficos más cercanos conceptualmente a lo buscado.

5.3.2. Recuperación de resultados

En este apartado se inicia el proceso de recuperación que sigue la secuencia inicial de iniciar una sesión como usuario que desea realizar una consulta o el usuario para administrar el *SemGsearch*.

Si se accesa como usuario búsqueda, se tendrán habilitadas las opciones de consultas únicamente, ingresando su vector de consulta para ser procesada y sobre la lista de objetos geográficos se recuperarán los elementos de interés.

A continuación en la Figura 5.23 se muestra la pantalla inicial al sistema de recuperación semántica *SemGSearch*.



Figura 5.23. Pantalla de inicio al sistema *SemGSearch*.

5.3.3. Despliegue de resultados

La lógica para el despliegue de resultados se realiza con respecto a saltos de medida conceptual k ; es decir, se enlistan primeramente objetos geográficos que conforme a lo consultado están a una distancia semántica de $k=0$; si se desean conocer más resultados el siguiente paso es indicar al final de la página el incremento del valor de k en un valor de $k+1$; y así sucesivamente.

De estas listas la de objetos geográficos devueltos a partir de la búsqueda semántica, se muestra la información de los metadatos de los datos recuperados tales como: tamaño del archivo, localización geoespacial, localización del servidor y fecha. Además de los atributos anteriores se muestra un *snapshot* de cada objeto geográfico para su mejor visualización.

5.3.4. Interfaz de búsqueda

A continuación, en la Figura 5.24 se muestra la interfaz de acceso para recuperar semánticamente un objeto geoespacial en la Intranet, por medio de una consulta, en la cual se indican los conceptos relevantes a la misma.



Figura 5.24. Interfaz de búsqueda semántica en *SemGSearch*.

5.3.5. Interfaz de administración de SemGSearch

La interfaz de acceso al usuario administrador del sistema con las opciones de agregar nuevo servidor, cargar un determinado dominio espacial temático o temporal, y administrar los servidores de datos de alta en la Intranet, se muestra en la Figura 5.25.

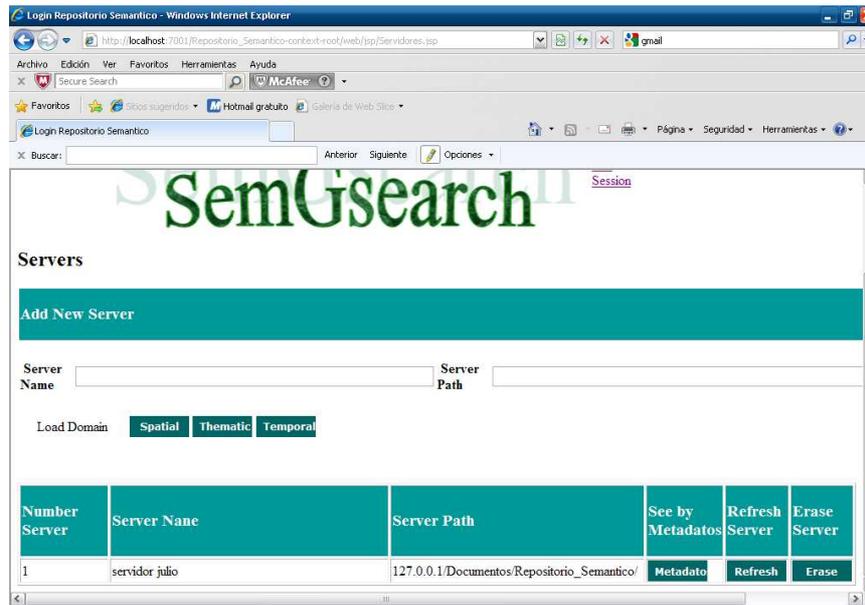


Figura 5.25. Interfaz de administración de SemGSearch.

5.3.6. Búsqueda de consultas en SemGSearch

La recuperación por dominio temático en un valor conceptual $K = 0$, para este caso no se localiza “calle” en los metadatos del servidor, entonces lo más cercano es “viaducto” y con una localización espacial exacta. En la Figura 5.26 se muestran los datos recuperados con valor $K = 0$.



Figura 5.26. Búsqueda semántica de concepto “calle”; “df”; “2007” con $K=0$.

Para el caso, cuando $K = 1$ se recuperan semánticamente otros objetos geospaciales similares, en el dominio temático a “calle”, para el dominio espacial cercanos al “DF”, y como dominio temporal cercanos a “2007”. En la Figura 5.27 y 5.28 se muestran los datos recuperados con el valor $K = 1$.

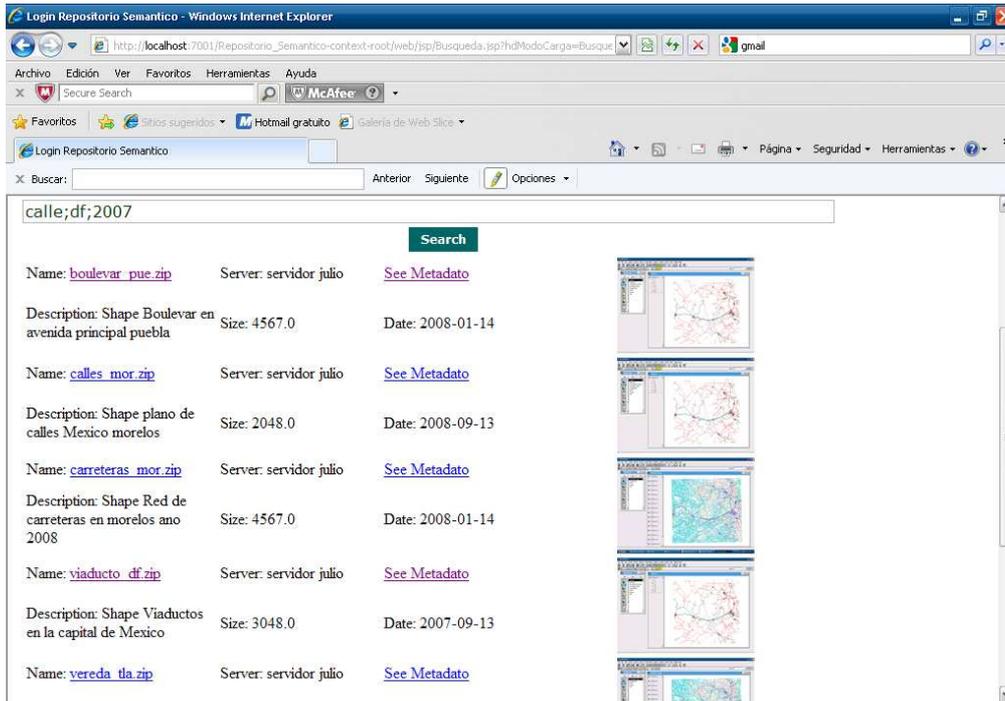


Figura 5.27. Búsqueda semántica de concepto “calle”; “df”; “2007” con $K=1$.

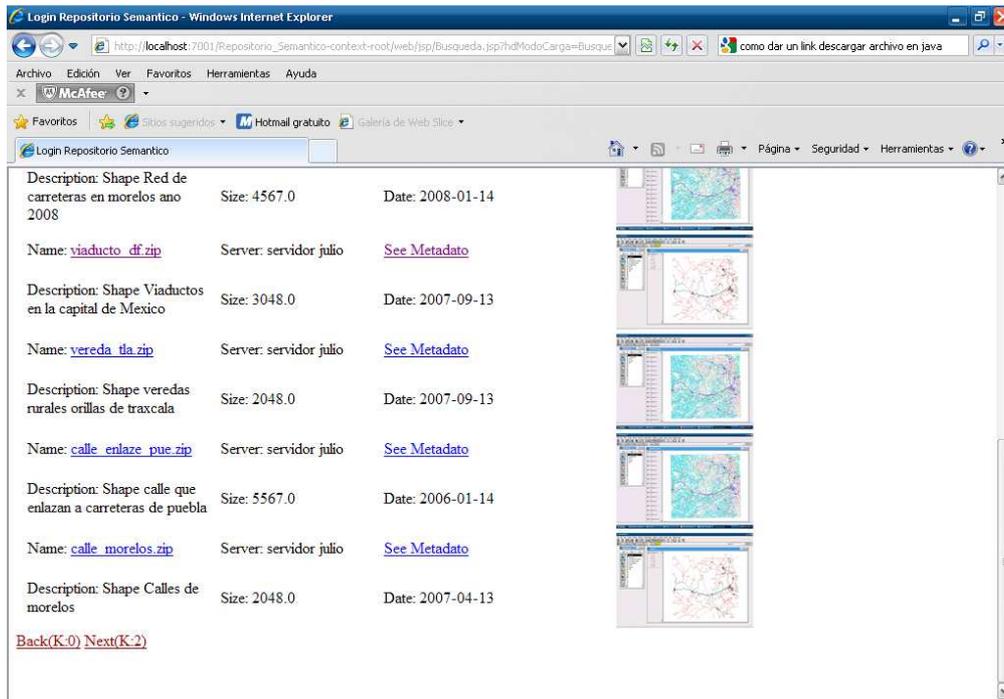


Figura 5.28. Búsqueda semántica de concepto “calle”; “df”; “2007” con $K=1$.

Para un valor de $K = 2$, el radio de intersección continua extendiéndose entre los conjuntos de datos de los dominios; por lo tanto, el sistema semánticamente recuperará los datos geoespaciales, tomando como base la distancia conceptual K con respecto a la consulta original. Este mismo procedimiento se utiliza para extender el dominio de búsqueda en $K + 1$. En la Figura 5.29 se muestra el resultado de la búsqueda semántica con un valor de $K = 2$.

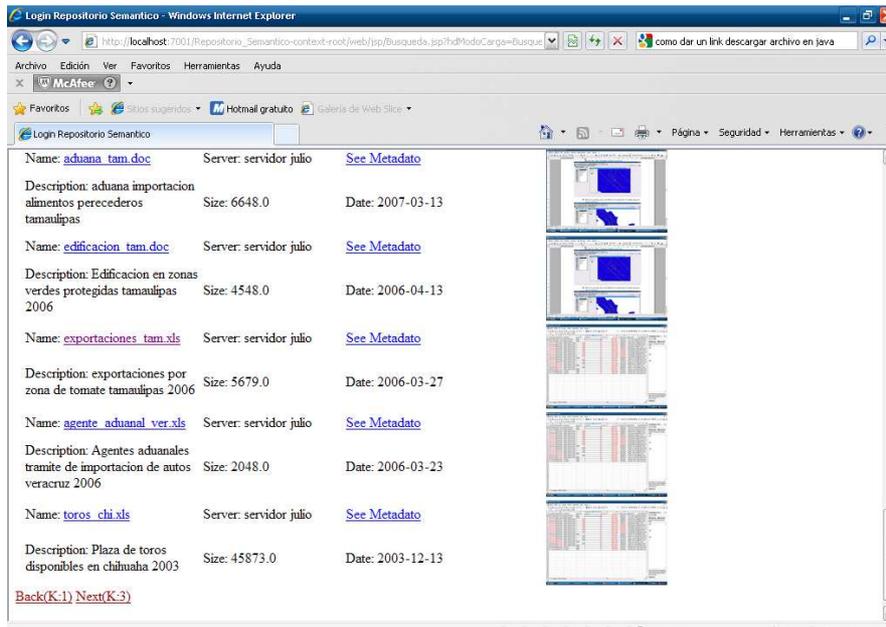


Figura 5.29. Búsqueda semántica de concepto “calle”; “df”; “2007” con $K=2$.

A continuación se muestra una consulta en la cual se desea recuperar un objeto geoespacial, en este caso “boulevard” en el estado de “Veracruz”, y posteriormente se efectúa el acercamiento espacial, evitando que se tengan resultados vacíos; es decir, para esta consulta no se tiene para un $K = 0$, un término exacto o hijos de esos conceptos. Para un $K = 1$ en “Veracruz” y en los metadatos localizados en un servidor se encuentra el objeto geoespacial “Chiapas”, el cual es vecino a “Veracruz” con respecto a su localización espacial. Por lo tanto, para una distancia conceptual, donde $K = 0$ no se tienen datos (ver Figura 5.30), sin embargo, en la Figura se observa que hay datos para un valor con $K = 1$.



Figura 5.30. Búsqueda semántica de concepto “boulevard”; “veracruz”; “2006” con $K=0$.

De acuerdo con lo anterior, para un valor $K = 1$, se realiza un acercamiento con base en la *zona espacial*, en donde no se encontró ninguna coincidencia por *zona económica*, ni por *cercanía espacial*, pero se tiene una correspondencia a la división ecorregional, que en este caso es: *selvas calido húmedas*, en la cual se tienen clasificados juntos a los estados de “Chiapas” y “Veracruz”. En la Figura 5.31 se muestra una partición de la ontología en donde se puede observar lo anteriormente descrito.

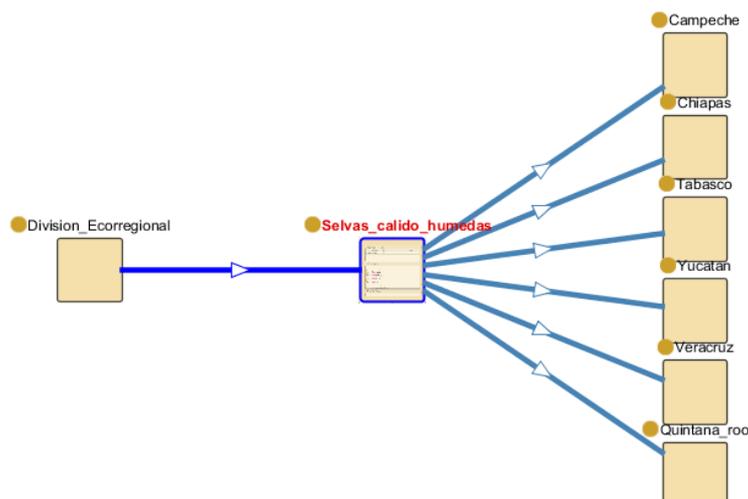


Figura 5.31. Partición de la ontología para la división ecorregional, particularmente en selvas cálido húmedas.

Por lo tanto, en la Figura 5.32 se muestra que no se tienen resultados vacíos para la consulta de “boulevard” en “Veracruz” con un valor de $K = 1$, y en la Figura 5.33 se muestran los resultados con un valor de $K = 2$.



Figura 5.32. Búsqueda semántica de concepto “boulevard”; “veracruz”; “2006” con $K=1$.

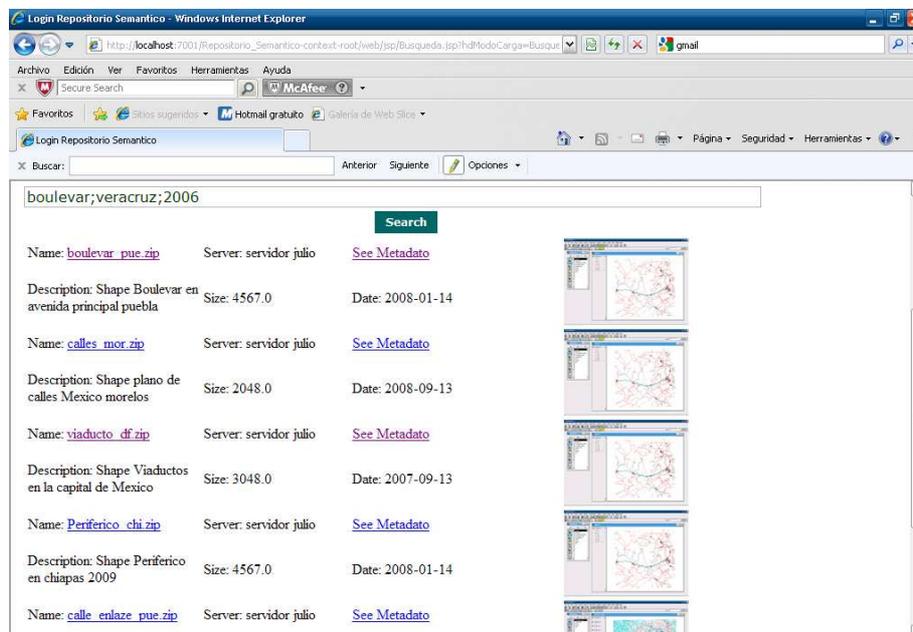


Figura 5.33. Búsqueda semántica de concepto “boulevard”; “veracruz”; “2006” con $K=2$.

Finalmente, en el acercamiento sobre el dominio temporal, éste se realiza bajo la misma década, en la cual todos los metadatos registrados se localizan sobre el mismo dominio.

5.3.7. Opciones de recuperación en la búsqueda

Aquí se muestran las opciones que tiene un usuario de cómo recuperar los objetos espaciales de su interés de forma general, mediante dos formas: la primera si la información no está completa se extraen de los metadatos que se muestran en la lista devuelta de la consulta semántica y puede ir directamente al archivo de metadatos para revisar la especificación. La segunda es directamente descargar el objeto geográfico de su interés si es un archivo *shapefile*, éste estará empaquetado en un formato de compresión ZIP, así mismo, si es una imagen se mostrará en otra ventana y si es cualquier otro archivo se habilitará la función de descarga, proporcionando el enlace correspondiente.

Para la consulta de los metadatos, éstos se pueden descargar directamente del sistema *SemGSearch*. En la Figura 5.34 se muestra el archivo en formato PDF de los metadatos de la especificación FGDC para alguna consulta en particular.

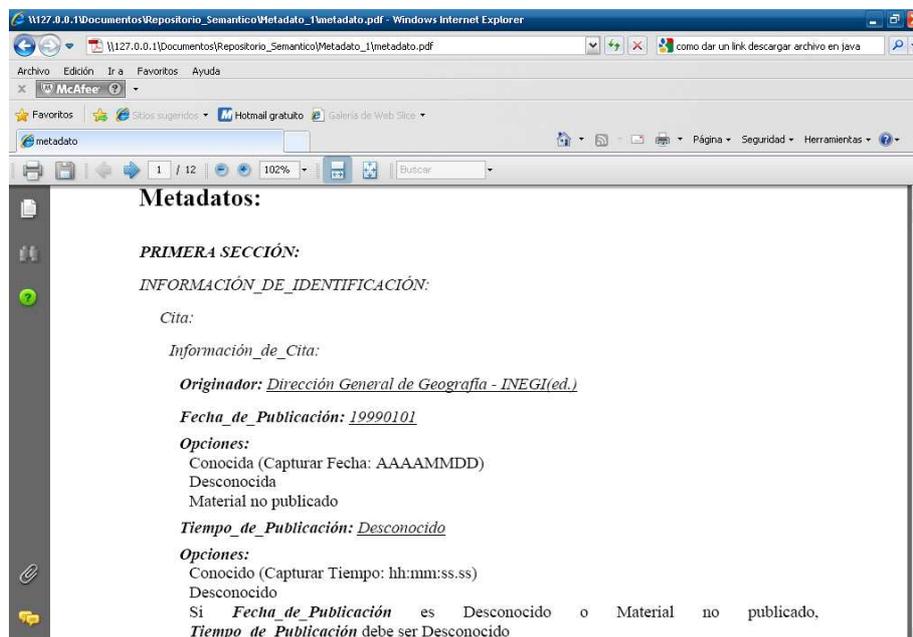


Figura 5.34. Descarga de los metadatos FGDC.

Asimismo, para la descarga de los objetos geoespaciales, éstos pueden ser descargados directamente del sistema *SemGSearch*. En la Figura 5.35 se muestra el conjunto de datos geoespaciales en formato *shapefile* que se pueden descargar directamente de la aplicación.

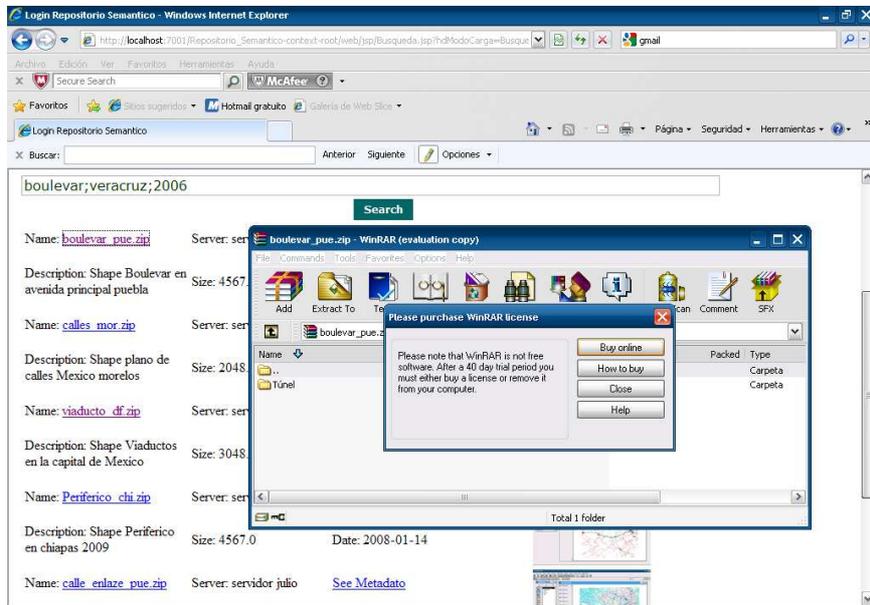


Figura 5.35. Descarga de un archivo *shapefile*.

Para la visualización de los datos geoespaciales, éstos pueden ser visualizados directamente en cualquier aplicación de formatos de imagen, como puede ser el TIFF. En la Figura 5.36 se muestra la visualización de este tipo de datos que pueden ser recuperados.

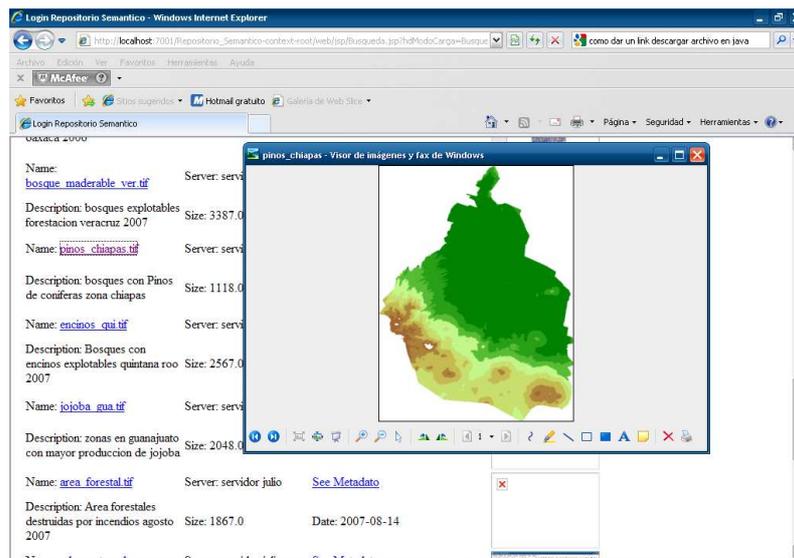


Figura 5.36. Visualización de un archivo TIFF.

Finalmente, los archivos recuperados semánticamente pueden ser descargados directamente a la máquina que sea cliente del servicio de la Intranet. En la Figura 5.37 se muestra esta operación.

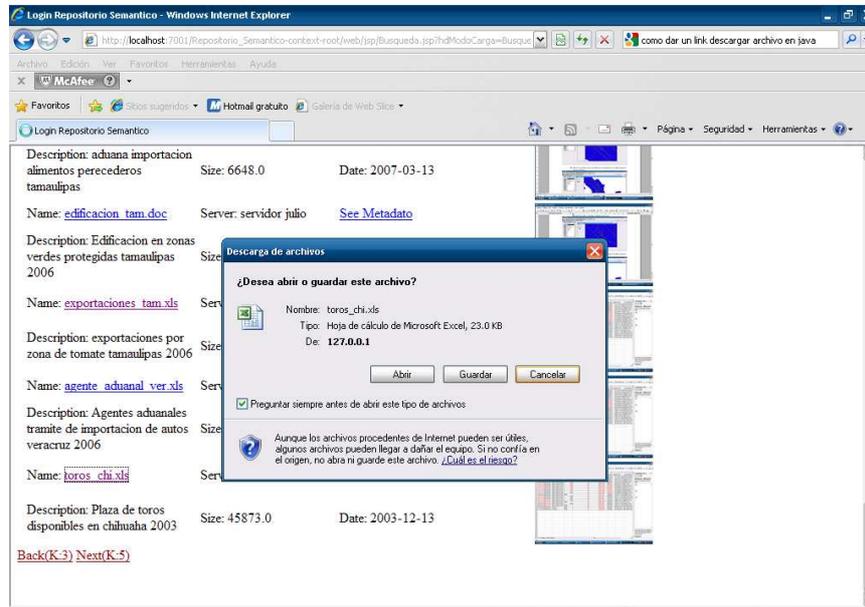


Figura 5.37. Descarga de un archivo, en este caso Excel referente a un objeto geoespacial.

5.3.8. Opciones del administrador en el sistema *SemGSearch*

En la interfaz del sistema, es posible agregar un servidor nuevo, del cual se puedan obtener a través de un archivo *meta_metadatos.xml*, todos aquellos metadatos que se desean compartir. De igual forma, éstos se cargan para poder recuperar en la búsqueda conjuntos de objetos geoespaciales; así como actualizar los datos que se desean compartir y eliminar de un servidor.

Por otra parte, también se ofrece la posibilidad de cargar una nueva conceptualización para realizar el proceso de recuperación semántica. En otras palabras, una ontología de cualquier otro contexto puede ser almacenada, solo que ésta debe cumplir con el estándar de desarrollo de la metodología GEONTO-MET para su definición, y debe contener los dominios espacial, temporal y temático en formato OWL/RDF, lo cual indica que la ontología debe ser implementada en la herramienta Protégé. En la Figura 5.38 se muestra la interfaz de carga automática para los dominios mencionados y definidos en una ontología.

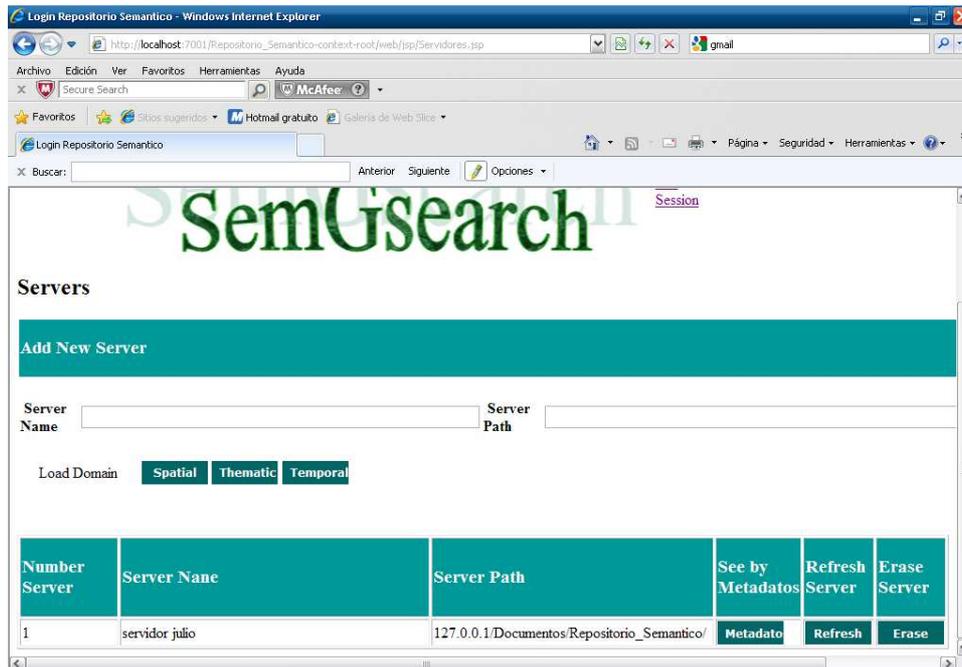


Figura 5.38. Carga automática de dominios en SemGSearch.

Por otra parte, los metadatos también pueden ser consultados directamente de todos los servidores y se puede llevar a cabo el proceso de carga de la especificación FGDC. En la Figura 5.39 se muestra la interfaz para realizar esta tarea.

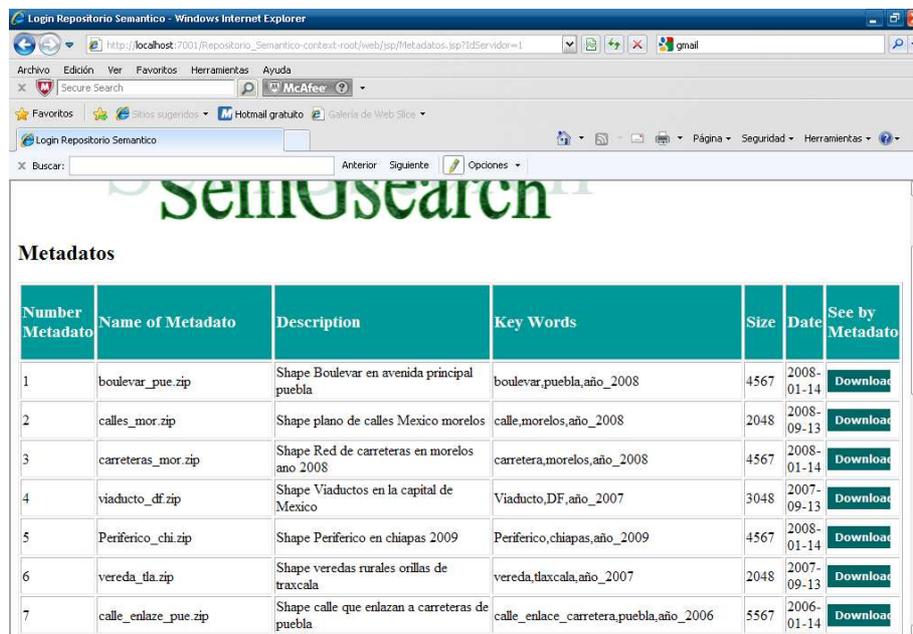


Figura 5.39. Ver metadatos de un servidor.

Capítulo 6. Conclusiones Y Trabajo a Futuro

6.1. Conclusiones

En el presente trabajo de tesis se ha desarrollado una metodología enfocada en las tareas de integración semántica de diversas fuentes de datos heterogéneas, las cuales pueden estar localizadas en diferentes servidores dentro de una Intranet.

De igual forma, el objetivo principal de este trabajo es poder compartir información geoespacial, a través de un mecanismo de recuperación semántica, utilizando para ello una representación conceptual, en este caso una ontología, la cual permite conceptualizar tres dominios que forman una parte esencial de las características implícitas de los objetos geográficos, como es el dominio temporal, espacial y temático.

Asimismo, se ha realizado la incorporación de un dominio adicional, en este caso los metadatos de los objetos geográficos, tomando como fuente principal de información el estándar de FGDC, el cual ha sido incorporado también a la ontología.

La metodología desarrollada cuenta con un mecanismo que permite medir la distancia conceptual, aplicando el algoritmo DIS-C. Este algoritmo se encarga de medir que tan cerca se encuentran conceptualmente dos conceptos, lo cual permite que el sistema pueda resolver consultas de acuerdo con un valor de distancia conceptual K , lo que permite evitar proporcionar al usuario respuestas vacías.

Esta medida conceptual trabaja en estos momentos para todas aquellas ontologías que han sido diseñadas con la metodología de GEONTO-MET, considerando básicamente un conjunto de tres relaciones axiomáticas y otro conjunto de preposiciones para darle causalidad al significado de una descripción.

El proceso de recuperación semántica consiste básicamente en llevar a cabo una intersección de los tres dominios, con base en los objetos encontrados de la consulta, con lo cual se garantiza la precisión del resultado, puesto que se ha calculado la distancia conceptual antes de la intersección. La recuperación semántica puede incrementar su granularidad de acuerdo con el crecimiento de los radios en un orden de $K+I$, lo que permite tener diversos resultados que han sido procesados semánticamente.

Cabe señalar que estos resultados son obtenidos directamente de la ontología, la cual contiene el conjunto de instancias pertenecientes a las clases y conceptos de la representación conceptual, la cual ha sido convertida en un modelo persistente de base de datos.

Por otra parte, la búsqueda de información geoespacial vía un enfoque semántico es una técnica importante que brinda precisión en el proceso, contando como base de un conocimiento del dominio, lo más granular posible donde concuerda el contexto de las palabras que describen a un objeto geoespacial para proporcionar resultados sobre el objeto buscado que pueden ser similares y de interés.

Esta metodología propuesta para la recuperación semántica en una Intranet es una solución para reducir el esfuerzo y tiempo en la búsqueda, debido a la subjetividad en la creación de los metadatos, lo cual se realiza de manera rápida y lo más precisa con respecto a un objeto geoespacial que se está buscando y permitiendo además, evitar los resultados vacíos en una consulta.

Por otro lado, como el proceso de carga de los dominios es automático resulta más rápido y sencillo agregar nuevo conocimiento al sistema; lo cual implica dotar a la aplicación de un conocimiento más refinado y completo de un contexto en particular, mejorando la gama y número de conceptos.

Con respecto a las pruebas o resultados experimentales, se puede observar que el nivel de precisión es proporcional al nivel de conocimiento detallado explícitamente en las ontologías que describen a los dominios espacial temático y temporal; por lo que con una conceptualización más profunda, será más precisa la búsqueda.

6.2. Aportaciones científicas

La metodología aquí desarrollada es de gran utilidad para la integración de fuentes datos heterogéneas, debido a que unifica el conocimiento usando una ontología de aplicación; es decir, se crea un conjunto de conceptos similares semánticamente, lo que resulta de gran utilidad cuando se desea compartir datos e intercambiarlos entre sistemas de información geográfica.

En la recuperación al aportar un nivel de relevancia con respecto a la consulta dada, se provee una mejor gama de resultados afines a lo que se está buscando.

En la etapa de conceptualización, la incorporación de los tres dominios espacial, temático y temporal permiten describir las características de un objeto geoespacial teniendo en cuenta que son los más usados al realizar una consulta y suelen ser los de mayor importancia. En cuanto al dominio FGDC, éste describe metadatos de forma estructurada con respecto a los archivos localizados en diversas fuentes, y permiten integrar y extraer información relevante de los objetos a ser compartidos.

Al tener conceptualizados los dominios, se utiliza el algoritmo DIS-C para establecer los valores numéricos de similitud de un concepto a otro, obteniendo un grafo sobre el cual el algoritmo de Floyd Warshall calcula la menor distancia conceptual. Estos algoritmos permiten describir la similitud semántica dando como resultado de este proceso la tabla de distancias conceptuales entre conceptos.

En la etapa de síntesis, se realiza el poblado de la ontología, convirtiendo los dominios a un modelo persistente de base de datos, utilizando los metadatos de los archivos geoespaciales. Con el uso de este modelo, es posible utilizar cualquiera de las técnicas ya ampliamente conocidas en la teoría de bases de datos.

En la etapa de análisis, se llevan a cabo las tareas de búsqueda y recuperación, con la opción de extender los radios de búsqueda y proveer los resultados más cercanos semánticamente a lo deseado.

6.3. Aportaciones Tecnológicas

En el desarrollo del sistema *SemGSearch* se aporta una arquitectura robusta y flexible por medio de J2EE poniendo a disposición los EJB (Enterprise Java Beans) para la aplicación Web. Asimismo, se enlistan las principales aportaciones tecnológicas.

- Se implementó una ontología en Protégé, a través del lenguaje OWL para representar el conocimiento de los dominios.
- Se implementó el algoritmo DIS-C, así como el de Floyd-Warshall para el cálculo de la distancia conceptual.

- Se diseñó e implementó el sistema Web *SemGSearch*, cuyas principales características son la administración de las fuentes de datos y las interfaces de búsqueda entre las mismas.

6.3. Trabajo a futuro

El trabajo a futuro sobre esta temática se enfoca principalmente en los siguientes puntos:

- Incrementar el nivel de conocimiento de los tres dominios; en otras palabras, aumentar la resolución y granularidad de conocimiento de los mismos.
- Ofrecer la posibilidad de conceptualizar e incrementar en más de tres dominios que describan a los objetos geográficos, con la finalidad de clasificar jerárquicamente e intersectar un mayor número de palabras clave de cada metadato. En este caso, podría llevarse a cabo la adición de un dominio meteorológico que permita realizar de manera más exhaustiva una recuperación semántica temporal.
- Agregar un gama de operaciones entre los conjuntos resultantes en la búsqueda al recuperar semánticamente por los tres dominios: temático espacial y temporal.
- Probar otros resultados para la construcción del grafo conceptual, de tal forma que se puedan medir distancias conceptuales entre las entidades de la ontología.
- Enfocar esta metodología hacia la recuperación semántica en un ambiente de Web semántica, en donde la Internet sea la fuente proveedora de datos heterogéneos y en ciertos casos no estructurados.

Referencias

1. Aguilera Arilla, M.J, Azcárate Luxán, M.V. González Yanci, M.P.. Fuentes, tratamiento y representación de la Información Geográfica. Universidad Nacional de Educación a Distancia (UNED).Madrid.2003.
2. Freitas, H.; Janissek-Muniz, R.; Moscarola, J. Dinámica del proceso de recolección a análisis de datos vía Web. Consejo Latinoamericano de Escuelas de Administración, Santiago do Chile.2005.
3. Heywood, I., Cornelius, S. y Carver, S., An Introduction to Geographical Information Systems, Longman. Singapore. Prentice Hall, 1998.
4. Khun, S.T., La estructura de las revoluciones científicas. Fondo de Cultura Económica, 2002.
5. Moreno Ibarra Marco Antonio. Similitud semántica entre sistemas de objetos geográficos aplicada a la generalización de datos geoespaciales.. Centro de investigación en computación, IPN México. 2007
6. Isla José Luis, Gutiérrez Francisco Luis, *et al.* Descripción de Patrones de Organización y su Modelado con AMENITIES. Departamento de Lenguajes y Sistemas Informáticos Universidad de Cádiz..Departamento de Lenguajes y Sistemas Informáticos Universidad de Granada. 2004
7. Vilches Blázquez Luis Manuel, Ramos Gargantilla José Ángel, Corcho Oscar, Capdevila Subirana Joan. Hacia una armonización semántica de la información geográfica. Ontology Engineering Group. Departamento de Inteligencia Artificial Facultad de Informática. Universidad Politécnica de Madrid Instituto Geográfico Nacional. Servicio Regional en Cataluña.2009
8. Buccella Agustina, Cechich Alejandra, & Pablo Fillottrani GIISCO Research Group. Integración de Sistemas de Información Geográfica . Departamento de Ciencias de la Computación Universidad Nacional del Comahue Neuquen, Argentina. Departamento de Ciencias e Ingeniería de la Computación. Universidad Nacional del Sur. Bahia Blanca, Argentina.2007
9. Y. Bishr. Overcoming the Semantic and Other Barriers to GIS Interoperability. International Journal of Geographical Information Science, 12:299–314, 1998.
10. Cruz F. Isabel, Rajendran Afsheen & William Sunna.Handling Semantic Heterogeneities Using Declarative Agreements. In the proceedings of ACM GIS'02, McLean, Virginia, USA, 2002.
11. E. Leclercq, D. Benslimane, and K. Yetongnon. ISIS, A Semantic Mediation Model and an Agent Based Architecture for GIS Interoperability. In International Database Engineering and Applications Symposium (IDEAS), pages 87–91, 1999.
12. Egenhofer J. Max. Toward the Semantic Geospatial Web. In the Proceedings of ACM GIS'02, November 8-9,McLean, Virginia, USA.2002
13. Wiederhold, G. Mediation to Deal with Heterogeneous Data Sources, in Interoperating Geographic Information Systems -, INTEROP'99, Zurich, Switzerland, vol. 1580, Lecture Notes in Computer Science, , pp. 1-16. 1999
14. Estándares de Datos Geográficos e interoperabilidad GIS. Grupo EP. Versión. Departamento Marketing.
15. Gómez A Héctor F. Ontología Para sistemas de información geográfica.Universidad Nacional de Educación a Distancia. Universidad Técnica Particular de Loja. Sistemas de Información Geográfica.2004

16. Gruber, T., Role of Common Ontology in Achieving. Sharable, Reusable Knowledge Bases. Principles of Knowledge. Representation and Reasoning, Cambridge, MA, 1991, pp. 601-602. 1991.
17. Guarino, N., Formal Ontology and Information Systems, in Formal Ontology in Information Systems Ed. Amsterdam, Netherlands: IOS Press, 1998.
18. Frank, A.U.Spatial Ontology: A Geographical Point of View, in Spatial and Temporal Reasoning, O. Stock, Ed. Dordrecht, The Netherlands: Kluwer Academic Publishers, 1997, pp. 135-153.1997
19. Nuñez, J.Geographic Space as a Set of Concrete Geographical Entities, in Cognitive and Linguistic Aspects of Geographic Space, D. Mark and A. Frank (Eds). Norwell, MA: Kluwer Academic Publishers, 1991, pp. 9-33.
20. Fonseca, F., M. Egenhofer, P. Agouris, (2002): Using Ontologies for Integrated Geographic Information System. Transaction in GIS 6(3), 2002.
21. Korth, Henry F.,Silberschatz, Abraham, Sudarshan, S. Fundamentos de base de datos Editorial McGraw-Hill.2002
22. Fonseca Frederico, EgenHofer Max,¿ & A. Karla. Borges.Ontologías e interoperabilidad Semántica entre SIGs. National Center for Geographic Information and Analysis Department of Spatial Information Science and Engineering. University of Maine, Orono, USA. 2000
23. Peis Redondo Eduardo, Hassan Montero Yusef .Herrera Viedma Enrique, Herrera,juan. Carlos. Ontologías, meta datos y agentes: recuperación semántica de la información. Dept. Biblioteconomía y Documentación Fac. Biblioteconomía y Documentación. Campus de la Cartuja, Universidad de Granada – España.2003
24. J. Cañas Alberto, M. Ford Kenneth, Coffey John *et al* . Herramientas para construir y compartir modelos de conocimiento. Institute for Human and Machine Cognition University of West Florida11000 University Parkway Pensacola.2000
25. García Figuero la, L. C., Zazo Rodríguez A. F. y Alonso Berrocal, J. L. La interacción con el usuario en los sistemas de Recuperación de Información: realimentación por relevancia. "Scire", 8 (1), 87-94.2002
26. Sperber, Dan, Wilson, Deirdre.La relevancia: comunicación y procesos cognitivos , Fondo Xavier Clavigero, S.J. - ITESO (Mexico).2004
27. MÄoller Ralf, Haarslev Volker, Neumann Bernd. Semantics-based information retrieval. University of Hamburg, Computer Science Department, Hamburg, Germany.1998
28. B. Jones Christopher, Harith Alani and Douglas Tudhope. Geographical Information Retrieval with Ontologies of Place. Department of Computer Science, Cardiff University Queens Buildings, United Kingdom Department of Electronics and Computer Science, University of Southampton School of Computing, University of Glamorgan Ontology-Based Discovery and Retrieval of Geographic Information in Spatial Data Infrastructures.2001
29. Spanouddakis y Constantopoulos. Similarity for analogical software reuse: A computational model.1994
30. L. Bernard, U. Einspanier, S. Haubrock, *et al*. Ontology-Based Discovery and Retrieval of Geographic Information in Spatial Data Infrastructures. Institute for Geoinformatics (IfGI), Münster.2004
31. Visser & Stuckenschmidt. Interoperability in GIS-enabling technologies. 2002

32. Soto Jesús, García Elisa. Semantic learning object repositories .Departamento de Inteligencia artificial, Universidad Pontificia de Salamanca. Y Salvador Sanchez-Alonso Departamento de Ciencias de la Computación, Universidad de Alcalá.2007
33. Vipul Kashyap and Amit Sheth. Semantic Heterogeneity in Global Information Systems.The Role of Metadata_ Context and Ontologies. LSDIS, Dept_ of Computer Science, Univ_ of Georgia, Athens Dept of Computer Science, Rutgers University, New Brunswick.1996
34. B. Jones Christopher. Spatial Information Retrieval and Geographical Ontologies.An Overview of the SPIRIT Project Christopher B. Jones Cardiff University, Department of Computer Science.2002.
35. The Thesaurus Harpring, P., Proper words in proper places: The Thesaurus of Geographic Names. MDA Information, 1997, 2(3): 5-12.
36. Kaoru Hiramatsu, Femke Reitsma GeoReferencing the Semantic Web: ontology-based markup of geographically referenced information. NTT Communication Science Laboratories, 2-4, Hikaridai, Seika-cho, Soraku-gun, Kyoto, JAPAN.Maryland Information and Network Dynamics Laboratory, University of Maryland, College.Park, USA.2004
37. Ubbo Visser & heiner Stuckenschmidt.Intelligent, Location-Dependetn acquisition and environmental information Intelligent Systems Group, Center for Computing Technologies, University of Bremen, Germany.1999.
38. Zarine Kemp, Lei Tan ,et al. Interoperability for Geospatial Analysis: a semantics and ontologybased approach. Computing Laboratory, University of Kent Canterbury, Kent.2007
39. Guarino Nicola.Semantic Matching: Formal Ontological Distinctions for Information Organization, Extraction, and Integration. LADSEB-CNR, National Research Council, Padova.1997
40. Knight and Luk .ontología SENSUS, 1994
41. Yaser Bishr, Werner Kuhn,.Ontology-Based Modelling of Geospatial Information Institute for Geoinformatics, University of Münster. Münster, Germany.2000
42. Sheth, A., Gala, S., & Navathe, S., On Automatic Reasoning for Schema Integration.International Journal of Intelligent and Cooperative Information Systems.1993.
43. Bishr. Overcoming the semantic and other barriers to GIS interoperability International Journal of Geographical Information 1998 - Taylor & Francis 1997
44. Gruber. Toward Principles for the Design of Ontologies Used for Knowledge Sharing Revision: August 23, Thomas R. Gruber Stanford Knowledge Systems Laboratory.1993
45. Bateman JA. Upper Modeling: organizing knowledge for natural language processing USC/Information Sciences Institute Marina del Rey, CA, U.S.A.1990
46. Burg. The impact of linguistics on conceptual models: consistency and understandability J. F. M. Burg, y R. P. van de Riet Department of Computer Science, Vrije Universiteit, de Boelelaan, Amsterdam, The Netherlands.1997
47. L. Bernard, U. Einspanier, S. Haubrock, S. *et al.* Ontologies for Intelligent Search and Semantic Translation in Spatial Data Infrastructures. Delphi InformationsMusterManagement (DELPHI IMM), Potsdam. Institute for Geoinformatics (IfGI), Münster. Center for Computing Technologies (TZI), Bremen.2003
48. Manoj Paul and S. K. Ghosh. An Approach for Service Oriented Discovery and Retrieval of Spatial Data. School of Information Technology Indian Institute of Technology, Kharagpur.2006
49. Nipkow, Tobias, Haarslev, Racer System Description. Volker. University of Hamburgo, Computer Science Department. Hamburgo Germany.2001

50. ESRI. Environmental Systems Research Institute <http://www.esri.com/> fecha de consulta:01-11-2010
51. Shapefile Technical Description. ESR ESRI White Paper—July 1998. <http://www.esri.com/library/whitepapers/pdfs/shapefile.pdf> fecha de consulta:01-11-2010
52. World Wide Web Consortium (W3C) <http://www.w3c.es/Divulgacion/Guiasbreves/WebSemantica> fecha de consulta: 01/11/08
53. Genereseth, MR & Nilsson, NJ Logical Foundations of Artificial Intelligence, Morgan-Kaufmann, 1987
54. Protégé The Protégé Ontology Editor and Knowledge Acquisition System <http://protege.stanford.edu/overview/> fecha de consulta: 01/11/2010
55. Deborah L. McGuinness. OWL Web Ontology Language Overview Latest version: <http://www.w3.org/TR/owl-features/> (Knowledge Systems Laboratory, Stanford University) Frank van Harmelen (Vrije Universiteit, Amsterdam) 2010
56. Prud'hommeaux, E., Seaborne, A.: Sparql Query Language for RDF. W3C Working Draft, 15 January 2008.
57. Duval, Erik. Metadata standards: What, who & why. (Departement computer wetens chappen Katholieke Universiteit Leuven, Belgium.2002
58. FGDC. Federal Geographic Data Committee <http://www.fgdc.gov/metadata/geospatial-metadata-standards> fecha de consulta: 01/11/08
59. Torres M., Representación ontológica basada en descriptores semánticos aplicada a objetos geográficos. Centro de investigación en computación.Instituto Politécnico nacional. México, DF.2007
60. Thomas H. Cormen, Charles E. Leiserson, Ronald L. Rivest, Clifford Stein Introduction to Algorithms, Second Edition The MIT Press Cambridge , Massachusetts London, England McGraw-Hill Book Company Boston Burr Ridge , IL Dubuque , IA Madison , WI New York San Francisco St. Louis Montréal Toronto
61. INEGI, “Diccionario de Datos Topográficos, Escala 1:50,000”,Instituto Nacional de Estadística y Geografía (INEGI)<http://www.inegi.org.mx/> fecha de consulta : 01/11/2010
62. Comisión nacional para el conocimiento y uso de la biodiversidad www.conabio.gob.mx/ fecha de consulta : 01/11/2010
63. CFE .Comisión Federal de Electricidad www.cfe.gob.mx/ fecha de consulta : 01/11/2010

Anexo I. Divisiones de la República Mexicana

En este anexo se presentan las tres divisiones del país, de acuerdo con diversas instituciones y que fueron utilizadas como parte de la conceptualización de los dominios que forman parte de la ontología. En la Figura A.1 se muestra la división espacial según CONABIO.

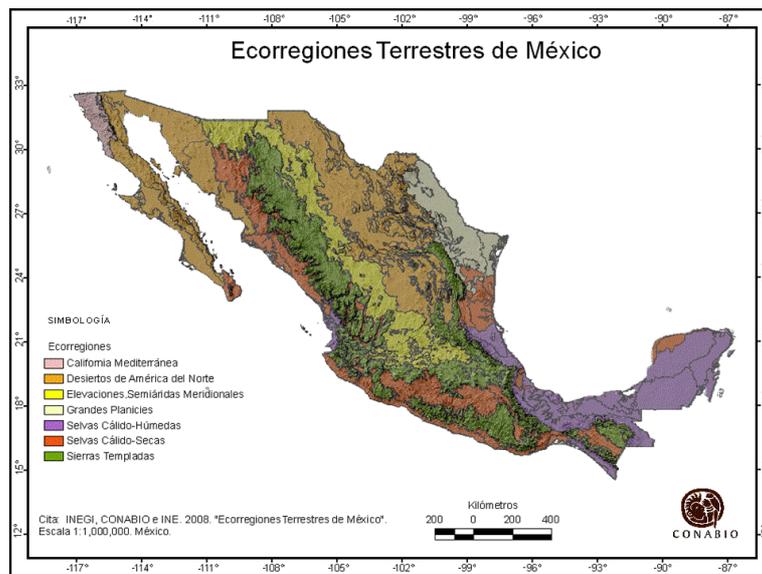


Figura A.1. División espacial basada en ecorregiones (CONABIO).

En la Figura A.2 se muestra la división espacial según la CFE.



Figura A.2. División espacial por regiones realizada por la CFE.

Finalmente, la Tabla A.1 muestra las zonas económicas en las cuales se divide la República Mexicana.

Tabla A.1. División por zonas económicas de México.

Zona 1	Zona 2	Zona 3	Zona 4
<u>Noroeste</u>	<u>Norte</u>	<u>Noreste</u>	<u>Centro-Occidente</u>
Baja California Baja California Sur Sonora Sinaloa Nayarit	Chihuahua Coahuila Durango Zacatecas San Luis Potosí	Nuevo León Tamaulipas	Jalisco Aguascalientes Colima Michoacán Guanajuato
Zona 5	Zona 6	Zona 7	Zona 8
<u>Centro-Este</u>	<u>Sur</u>	<u>Oriente</u>	<u>Península de Yucatán</u>

<ul style="list-style-type: none"> • Querétaro • Estado de México • D.F. • Morelos • Hidalgo • Tlaxcala • Puebla. 	<ul style="list-style-type: none"> • Guerrero • Oaxaca • Chiapas 	<ul style="list-style-type: none"> • Veracruz • Tabasco 	<ul style="list-style-type: none"> • Campeche • Yucatán • Quintana Roo
--	---	---	---

Anexo II. Archivo de metadatos

En este anexo se presenta el archivo de metadatos utilizado para las pruebas de búsqueda y recuperación semántica, utilizando el sistema *SemGSearch*.

Meta_Metadatos.xml
<pre> <?xml version="1.0" encoding="ISO-8859-1" ?> <Metadatos> <Metadato Numero="1"> <NombreMetadato>boulevard_pue.zip</NombreMetadato> <RutaLocal>Metadato_1</RutaLocal> <DescripcionMetadato>Shape Boulevard en avenida principal puebla</DescripcionMetadato> <PalabrasClave>boulevard,puebla,año_2008</PalabrasClave> <Tamaño>4567</Tamaño> <Fecha>2008/01/14</Fecha> </Metadato > <Metadato Numero="2"> <NombreMetadato>calles_mor.zip</NombreMetadato> <RutaLocal>Metadato_2</RutaLocal> <DescripcionMetadato>Shape plano de calles Mexico morelos </DescripcionMetadato> <PalabrasClave>calle,morelos,año_2008</PalabrasClave> <Tamaño>2048</Tamaño> <Fecha>2008-09-13</Fecha> </Metadato > <Metadato Numero="3"> <NombreMetadato>carreteras_mor.zip</NombreMetadato> <RutaLocal>Metadato_3</RutaLocal> <DescripcionMetadato>Shape Red de carreteras en morelos ano 2008</DescripcionMetadato> <PalabrasClave>carretera,morelos,año_2008</PalabrasClave> <Tamaño>4567</Tamaño> <Fecha>2008/01/14</Fecha> </Metadato > <Metadato Numero="4"> <NombreMetadato>viaducto_df.zip</NombreMetadato> <RutaLocal>Metadato_4</RutaLocal> <DescripcionMetadato>Shape Viaductos en la capital de Mexico</DescripcionMetadato> <PalabrasClave>Viaducto,DF,año_2007</PalabrasClave> <Tamaño>3048</Tamaño> <Fecha>2007-09-13</Fecha> </Metadato > <Metadato Numero="5"> <NombreMetadato>Periferico_chi.zip</NombreMetadato> <RutaLocal>Metadato_5</RutaLocal> <DescripcionMetadato>Shape Periferico en chiapas 2009</DescripcionMetadato> <PalabrasClave>Periferico,chiapas,año_2009</PalabrasClave> <Tamaño>4567</Tamaño> <Fecha>2008/01/14</Fecha> </Metadato > <Metadato Numero="6"> <NombreMetadato>vereda_tla.zip</NombreMetadato> <RutaLocal>Metadato_6</RutaLocal> <DescripcionMetadato>veredas rurales orillas de traxcala</DescripcionMetadato> <PalabrasClave>vereda,tlaxcala,año_2007</PalabrasClave> <Tamaño>2048</Tamaño> <Fecha>2007-09-13</Fecha> </pre>

```

</Metadato >
<Metadato Numero="7">
  <NombreMetadato>calle_enlaze_pue.zip</NombreMetadato>
  <RutaLocal>Metadato_7/</RutaLocal>
  <DescripcionMetadato>calle que enlazan a carreteras de puebla</DescripcionMetadato>
  <PalabrasClave>calle_enlaze_carretera,puebla,año_2006</PalabrasClave>
  <Tamaño>5567</Tamaño>
  <Fecha>2006/01/14</Fecha>
</Metadato >
<Metadato Numero="8">
  <NombreMetadato>calle_morelos.zip</NombreMetadato>
  <RutaLocal>Metadato_8/</RutaLocal>
  <DescripcionMetadato>Shape Calles de morelos </DescripcionMetadato>
  <PalabrasClave>calle,morelos,año_2007</PalabrasClave>
  <Tamaño>2048</Tamaño>
  <Fecha>2007-04-13</Fecha>
</Metadato >
<Metadato Numero="9">
  <NombreMetadato>puente_guerrero.zip</NombreMetadato>
  <RutaLocal>Metadato_9/</RutaLocal>
  <DescripcionMetadato>shape linea de conexion puentes en guerrero
2007</DescripcionMetadato>
  <PalabrasClave>puente,guerrero,año_2007</PalabrasClave>
  <Tamaño>4567</Tamaño>
  <Fecha>2007/03/14</Fecha>
</Metadato >
<Metadato Numero="10">
  <NombreMetadato>autodromo_gua.xls</NombreMetadato>
  <RutaLocal>Metadato_10/</RutaLocal>
  <DescripcionMetadato>Shape autodromos existente en guadalajara</DescripcionMetadato>
  <PalabrasClave>autodromo,guadalajara,año_2006</PalabrasClave>
  <Tamaño>2048</Tamaño>
  <Fecha>2006-05-13</Fecha>
</Metadato >
<Metadato Numero="11">
  <NombreMetadato>aulas_tol.zip</NombreMetadato>
  <RutaLocal>Metadato_11/</RutaLocal>
  <DescripcionMetadato>Shape Escuelas primaria publicas con pocas aulas en
2007</DescripcionMetadato>
  <PalabrasClave>aulas,tluca,año_2007</PalabrasClave>
  <Tamaño>6767</Tamaño>
  <Fecha>2007/01/23</Fecha>
</Metadato >
<Metadato Numero="12">
  <NombreMetadato>velodromo_son.zip</NombreMetadato>
  <RutaLocal>Metadato_12/</RutaLocal>
  <DescripcionMetadato>Velodromo existentes en sonora 2006</DescripcionMetadato>
  <PalabrasClave>velodromo,sonora,año_2006</PalabrasClave>
  <Tamaño>2848</Tamaño>
  <Fecha>2006-11-13</Fecha>
</Metadato >
<Metadato Numero="13">
  <NombreMetadato>maestros_hidalgo.zip</NombreMetadato>
  <RutaLocal>Metadato_13/</RutaLocal>
  <DescripcionMetadato>Maestros inconformes salario por zona hidalgo
2007</DescripcionMetadato>
  <PalabrasClave>maestros,hidalgo,año_2007</PalabrasClave>
  <Tamaño>4567</Tamaño>
  <Fecha>2008/01/14</Fecha>
</Metadato >
<Metadato Numero="14">
  <NombreMetadato>autos_oax.shp</NombreMetadato>
  <RutaLocal>Metadato_14/</RutaLocal>
  <DescripcionMetadato>venta de autos en oaxaca 2006</DescripcionMetadato>
  <PalabrasClave>autos,oaxaca,año_2006</PalabrasClave>
  <Tamaño>2048</Tamaño>
  <Fecha>2006-05-22</Fecha>
</Metadato >
<Metadato Numero="15">
  <NombreMetadato>bosque_maderable_ver.tif</NombreMetadato>
  <RutaLocal>Metadato_15/</RutaLocal>
  <DescripcionMetadato>bosques explotables forestacion veracruz
2007</DescripcionMetadato>
  <PalabrasClave>bosque_maderable,veracruz,año_2007</PalabrasClave>
  <Tamaño>3387</Tamaño>
  <Fecha>2007/03/04</Fecha>
</Metadato >
<Metadato Numero="16">
  <NombreMetadato>pinos_chiapas.tif</NombreMetadato>
  <RutaLocal>Metadato_16/</RutaLocal>
  <DescripcionMetadato>bosques con Pinos de coniferas zona chiapas </DescripcionMetadato>
  <PalabrasClave>pinos,chiapas,año_2006</PalabrasClave>
  <Tamaño>1118</Tamaño>
  <Fecha>2006-04-23</Fecha>
</Metadato >
<Metadato Numero="17">
  <NombreMetadato>encinos_qui.tif</NombreMetadato>
  <RutaLocal>Metadato_17/</RutaLocal>
  <DescripcionMetadato>Bosques con encinos explotables quintana roo
2007</DescripcionMetadato>
  <PalabrasClave>encinos,matorral,quintana_roo,año_2007</PalabrasClave>
  <Tamaño>2567</Tamaño>

```

```

        <Fecha>2007/01/01</Fecha>
    </Metadato >
    <Metadato Numero="18">
        <NombreMetadato>jojoba_gua.tif</NombreMetadato>
        <RutaLocal>Metadato_18</RutaLocal>
        <DescripcionMetadato>zonas en guanajuato con mayor produccion de
jojoba</DescripcionMetadato>
        <PalabrasClave>jojoba,guanajuato,año_2010</PalabrasClave>
        <Tamaño>2048</Tamaño>
        <Fecha>2010-01-23</Fecha>
    </Metadato >
    <Metadato Numero="19">
        <NombreMetadato>area_forestal.tif</NombreMetadato>
        <RutaLocal>Metadato_19</RutaLocal>
        <DescripcionMetadato>Area forestales destruidas por incendios agosto
2007</DescripcionMetadato>
        <PalabrasClave>area_forestal,tamaulipas,año_2007</PalabrasClave>
        <Tamaño>1867</Tamaño>
        <Fecha>2007/08/14</Fecha>
    </Metadato >
    <Metadato Numero="20">
        <NombreMetadato>vias_de_comunicacion_gua.doc</NombreMetadato>
        <RutaLocal>Metadato_20</RutaLocal>
        <DescripcionMetadato>Conjunto de vías de comunicacion vehiculares guadalajara
2009</DescripcionMetadato>
        <PalabrasClave>vias_de_comunicacion,guadalajara,año_2009</PalabrasClave>
        <Tamaño>8953</Tamaño>
        <Fecha>2009-12-13</Fecha>
    </Metadato >
    <Metadato Numero="21">
        <NombreMetadato>aduana_tam.doc</NombreMetadato>
        <RutaLocal>Metadato_21</RutaLocal>
        <DescripcionMetadato>aduana importacion alimentos perecederos
tamaulipas</DescripcionMetadato>
        <PalabrasClave>aduana,tamaulipas,año_2007</PalabrasClave>
        <Tamaño>6648</Tamaño>
        <Fecha>2007-03-13</Fecha>
    </Metadato >
    <Metadato Numero="22">
        <NombreMetadato>edificacion_tam.doc</NombreMetadato>
        <RutaLocal>Metadato_22</RutaLocal>
        <DescripcionMetadato>Edificacion en zonas verdes protegidas tamaulipas 2006
</DescripcionMetadato>
        <PalabrasClave>Edificacion,tamaulipas,año_2006</PalabrasClave>
        <Tamaño>4548</Tamaño>
        <Fecha>2006-04-13</Fecha>
    </Metadato >
    <Metadato Numero="23">
        <NombreMetadato>exportaciones_tam.xls</NombreMetadato>
        <RutaLocal>Metadato_23</RutaLocal>
        <DescripcionMetadato>exportaciones por zona de tomate tamaulipas
2006</DescripcionMetadato>
        <PalabrasClave>exportaciones,tamaulipas,año_2006</PalabrasClave>
        <Tamaño>5679</Tamaño>
        <Fecha>2006-03-27</Fecha>
    </Metadato >
    <Metadato Numero="24">
        <NombreMetadato>agente_aduanal_ver.xls</NombreMetadato>
        <RutaLocal>Metadato_24</RutaLocal>
        <DescripcionMetadato>Agentes aduanales tramite de importacion de autos veracruz
2006</DescripcionMetadato>
        <PalabrasClave>agente_aduanal,veracruz,año_2006</PalabrasClave>
        <Tamaño>2048</Tamaño>
        <Fecha>2006-03-23</Fecha>
    </Metadato >
    <Metadato Numero="25">
        <NombreMetadato>toros_chi.xls</NombreMetadato>
        <RutaLocal>Metadato_25</RutaLocal>
        <DescripcionMetadato>Plaza de toros disponibles en chihuahua 2003</DescripcionMetadato>
        <PalabrasClave>toros,chihuahua,año_2003</PalabrasClave>
        <Tamaño>45873</Tamaño>
        <Fecha>2003-12-13</Fecha>
    </Metadato >
</Metadatos>

```