



Universitat d'Alacant
Universidad de Alicante

Uso de la similitud semántica para la
recuperación de información geoespacial

Neili Machado García



Tesis

Doctorales

www.eltallerdigital.com

UNIVERSIDAD de ALICANTE



Universitat d'Alacant
Universidad de Alicante

**Departamento de Lenguajes y Sistemas
Informáticos
Universidad de Alicante**



Uso de la similitud semántica para la recuperación de información geoespacial.

Tesis Doctoral

Autor: Ing. Neili Machado García, MSc.
Director: Ing. Andrés Montoyo Guijarro, Dr.C.

Enero 2015



Universitat d'Alacant
Universidad de Alicante

Dedicatoria



Universitat d'Alacant
Universidad de Alicante

A mi familia por su apoyo incondicional.

A Claudia, por ser mi tesoro más preciado.



Universitat d'Alacant
Universidad de Alicante

Agradecimientos

Este trabajo de tesis lo dedico a todas las personas que han estado conmigo durante todo este tiempo, en especial:

A Claudia, por saber valorar mi esfuerzo a pesar de su corta edad.

A Alfredo, por su apoyo en los largos días de trabajo, por siempre estar ahí para mí.

A mis padres, por su amor y cariño, por ser parte de todos mis logros, por no dejarme titubear, sin ustedes nunca lo hubiera logrado.

A mi hermana, mi Tata, por ser mi otra versión de mi misma. A mi sobrina preferida, Patricia, por sus ocurrencias.

A mis ángeles que están en el cielo y la que tengo a mi lado, mis abuelitos, por su inmenso amor.

A mis tíos y mis primos que siempre están ahí para mí, apoyándome a seguir adelante.

A mi familia neopazina por saber que siempre puedo contar con ellos.

A mi tutor Dr.C. Andrés Montoyo, por sus conocimientos, sus consejos, sus revisiones y comentarios para culminar esta tesis.

A Mercedes, por ser mi amiga, por no dejarme dar un paso atrás, por sentir mi trabajo como suyo.

A Mingelkis, por ser mi prima, mi amiga, por prestarme su hombro cada vez que necesito recostarme.

A Alexander, por ser mi amigo en las distintas etapas de nuestras vidas.

A Iriarte, porque me abrió las puertas en el mundo de la investigación.

A Izarys, Haliuska, Margarita, por ser mis amigas incondicionales.

A Balmaseda, porque cuando creía que no podía confié en mi trabajo más que yo misma.

A todos mis amigos del Departamento de Informática de la UNAFI; a Aponte, a Lilibeth, a Yusny y a Reydi por todas las sendas que hemos transitado juntos. ¡Sí se puede! A Yadira, por siempre poder contar con ella.

A mis amigos de todos los tiempos.

A todos los profesores que a través de los años me han formado, en especial, a Ariel que siempre confié en mí.

A mi familia mexicana, a Any y Beto, por no dejarme sentir sola cuando estaba tan lejos de mi casa.

A la Facultad de Ciencias Técnicas de la UNAFI; a Antihus, a Annia, al Consejo Científico, a Elisa y a Mariela por apoyar este trabajo.

Al Centro de Investigación en Computación del Instituto Politécnico Nacional, en especial, a Dr.C. Marco Moreno, a Dr.C. Miguel Torres y a Dr.C. Giovanni Gasmán, por todo el conocimiento que me aportaron en mi estancia allí.

Al Departamento de Lenguajes y Sistemas Informáticos de la Universidad de Alicante; a Jesús Peral, por siempre poder contar con él.

A los estudiantes que he tutorado, Mingkai, Amaury, Yanet, Yaimi, Yean Michel, Eduardo, que forman parte de estos resultados.

A la Universidad Agraria de La Habana por todo lo que ha aportado en mí; a Pablo Hernández (Pablito).

A todos los profesores y amigos de la CUAE que me acompañaron y ayudaron a formarme como ingeniera y permitirme cumplir mi sueño, en especial, a Martica, Sepúlveda y Rosete, que dejaron huellas inolvidables.

Gracias a todos los que directa e indirectamente hicieron posible estos pasos de crecimiento en el largo camino de la investigación.



Universitat d'Alacant
Universidad de Alicante

*El éxito no se logra sólo con cualidades especiales.
Es sobre todo un trabajo de constancia, de método y de organización.*
J.P. Sergent



Universitat d'Alacant
Universidad de Alicante

Índice general

1. Introducción	1
1.1 Contextualización y motivación.....	1
1.2 Planteamiento del problema	7
1.3 Objetivos.....	9
1.4 Alcance e interés de la investigación.	10
1.5 Organización de la tesis.....	12
2. Estado del arte.	13
2.1 Contextualización y motivación.....	13
2.1.1 Modelos de recuperación de información.	16
2.1.2 Expansión de consultas.....	21
2.1.3 Evaluación en la recuperación de información.....	22
2.2 Sistemas de información geográfica.	24
2.2.1 El modelo de datos vectoriales.....	26
2.2.2 Relaciones espaciales.....	28
2.2.3 Comparación entre los RI y los SIG.....	33
2.3 Recuperación de información geográfica.	35
2.3.1 Las relaciones espaciales en la RIG.....	37
2.4 Recuperación semántica de información geográfica.	39
2.4.1 Uso de las ontologías en el dominio espacial.	41
2.4.2 Similitud semántica entre datos geoespaciales.	47
2.5 Conclusiones	66
3. Estudio de la medida de similitud óptima.....	67
3.1 Selección de una medida de similitud.	67
3.2 Distancia Conceptual DIS-C.	68
3.3 Modelo de similitud espacial Topología-Dirección-Distancia	73

3.3.1 Relaciones espaciales.....	75
3.3.2 Relaciones direccionales.....	76
3.3.3 Relación de distancia.....	78
3.4 Integración de criterios.....	78
3.5 Conclusiones	81
4. Propuesta y diseño de la ontología geoespacial.....	83
4.1 Metodología GEONTO-MET	84
4.2 Diseño de la ontología	86
4.2.1 Modelado del dominio geográfico	92
4.2.2 Modelado del dominio de aplicación	100
4.3 Conclusiones	102
5. Metodología de recuperación de información geoespacial.	105
5.1 Arquitectura general.....	105
5.2 Conceptualización del dominio.	107
5.3 Procesamiento multicriterio.....	109
5.3.1 Procesamiento de la consulta.....	109
5.3.2 Recuperación de objetos geográficos.....	111
5.4 Interfaces de usuarios.....	119
5.4.1 Visualización de los resultados.....	119
5.5 Conclusiones	120
6. Resultados experimentales.	121
6.1 Discusión sobre los principales componentes.....	121
6.2 Ontología espacial como estructura de indexación.	123
6.3 Recuperación semántica de información geoespacial.....	125
6.4 Evaluación de los resultados.....	130
6.5 Conclusiones	133
7. Conclusiones y trabajos futuros.	135

7.1 Líneas de trabajo futuro.	137
8. Publicaciones relacionadas con este trabajo de tesis.	139
9. Referencias bibliográficas	141
10. Anexos	151
Anexo 1. Mapa de suelos del municipio San José de Las Lajas.	151
Anexo 2. Mapa de suelos de la provincia Mayabeque.	152
Anexo 3. Mapa de suelos de la provincia Pinar del Río.	153
Anexo 4. Plataforma de experimentación.	154
Anexo 5. Clasificación de los suelos según Nueva Versión de Clasificación Genética de los Suelos de Cuba.....	155
Anexo 6. Ontología geoespacial.	161



Universitat d'Alacant
Universidad de Alicante



Universitat d'Alacant
Universidad de Alicante

Índice de figuras

Figura 2.1. Esquema del proceso de recuperación de información.	15
Figura 2.2. Comparación de similitud en el modelo vectorial	18
Figura 2.3. Representación de entidades geográficas mediante el modelo vectorial.	27
Figura 2.4. Componentes del modelo vectorial.....	28
Figura 2.5. Una matriz 3x3 que indica la intersección del interior, los límites y el exterior describe formalmente las relaciones topológicas.....	29
Figura 2.6. Relaciones topológicas entre los objetos A y B.	30
Figura 2.7. Grafo del vecindario conceptual del Modelo de las 9-Intersecciones....	31
Figura 2.8. Grafo de vecindario conceptual para las relaciones de distancia cualitativas (Bruns and Egenhofer, 1996).....	32
Figura 2.9. Marco de referencia absoluto y relativo.	32
Figura 2.10. Diferentes enfoques de medidas de similitud.....	48
Figura 3.1. Jerarquía de seres vivos (Levachkine and Guzmán-Arenas, 2007).	70
Figura 3.2. Entorno de red conceptual de las relaciones topológicas: (a) dos polígonos; (b) una línea y un polígono; (c) un punto y un polígono; (d) dos líneas; (e) dos puntos; (f) un punto y una línea (Li and Fonseca, 2006).	75
Figura 3.3. (a)Vecindario direccional 9-direcciones; (b) ejemplo; (c) listado ordenado de los ejemplo de (b).	76
Figura 3.4. (a) Vecindario direccional p/2-direcciones; (b) ejemplo; (c) listado ordenado de los ejemplo de (b)	77
Figura 3.5. Red de distancia.....	78
Figura 3.6. Espacio para la comparación de criterios.	79
Figura 3.7. Comparación de objetos en el espacio para la comparación de criterios.	80
Figura 4.1. Estructura conceptual de GEONTO-MET (Torres Ruiz, 2007).....	84
Figura 4.2. Jerarquía básica de la ontología.....	93
Figura 4.3. Jerarquía de clases de “ <i>elemento_abstracto</i> ”	96
Figura 4.4. <i>Slot</i> de relaciones que realizan el mapeo de las relaciones axiomáticas.	97
Figura 4.5. Mapeo de una relación ternaria.	98

Figura 4.6. Jerarquía de la clase “ <i>datos</i> ”	101
Figura 4.7. Fragmento de la jerarquía de clases que componen la ontología.....	102
Figura 5.1. Arquitectura general para la recuperación semántica de información geográfica.	106
Figura 5.2. Diagrama de la fase de procesamiento de la consulta.....	109
Figura 5.3. Extracción de relaciones espaciales.....	110
Figura 5.4. Diagrama general de la recuperación de instancias de conceptos geográficos y su representación en el <i>web-mapping</i>	112
Figura 5.5. Detalles de la implementación del modelo TDD.	116
Figura 5.6. Implementación de las relaciones espaciales en PCP.....	118
Figura 6.1. Región espacial que describe el escenario de trabajo.	123
Figura 6.2. Visualización de las capas de datos	126
Figura 6.3. Mapas con las regiones recuperadas según una consulta determinada.	127
Figura 6.4. Precisión contra cobertura para la consulta “Regiones que comparten límite con suelo Aluvial”.....	132
Figura 6.5 Precisión contra cobertura en la tres consultas diferentes.....	133

Índice de tablas

Tabla 2.1. La semántica de las relaciones espaciales generalizadas se define a través de una definición formal de relaciones espaciales entre instancias de una misma clase.	39
Tabla 3.1. Valor de similitud de la jerarquía de los seres vivos utilizando <i>Confusion</i>	71
Tabla 3.2. Valor de similitud de la jerarquía de los seres vivos utilizando Cantidad de información.....	71
Tabla 3.3. Valor de similitud de la jerarquía de los seres vivos utilizando <i>Distance</i>	72
Tabla 3.4. Valor de similitud de la jerarquía de los seres vivos utilizando DIS-C	72
Tabla 3.5. Correlación de resultados.	73
Tabla 4.1. Clases esenciales que componen la ontología.	94
Tabla 5.1. Asociación entre términos identificados y operadores espaciales.....	110
Tabla 5.2. Clases de características simples soportadas por la API JTS, de acuerdo con la especificación de elementos simples de la OGC	114
Tabla 5.3. Relaciones topológicas soportadas por la API JTS.....	117
Tabla 6.1. Términos o palabras asociadas a relaciones espaciales.	127
Tabla 6.2. Comparaciones entre cantidad de actividades a realizar por el usuario para indicar la información que desea recuperar utilizando el método propuesto y la herramienta Quantum GIS 1.8.0.....	130



Universitat d'Alacant
Universidad de Alicante

1. Introducción

En este capítulo se presentan las generalidades en el campo de la recuperación semántica de información geográfica; así como los objetivos de la misma. Además, se describe el enfoque de la investigación, la justificación del trabajo, el planteamiento general del problema; con el cual se introduce la hipótesis del mismo y los objetivos del presente trabajo de investigación. Finalmente, se puntualiza el alcance e interés del tema de investigación y se detalla en forma general la organización del documento.

1.1 Contextualización y motivación.

La Web Semántica, (Berners-Lee et al., 2001) al igual que la Web, constituye un espacio distribuido y heterogéneo. En este espacio compartir y reutilizar información se convierte en un tema de gran importancia ya que los usuarios requieren acceder a la información recopilada para satisfacer sus propósitos. El principio fundamental de la Web Semántica es la idea de contar con datos definidos y vinculados de tal manera que se puede utilizar para la detección más eficaz, la automatización, la integración y la reutilización a través de diversas aplicaciones (Fensel et al., 2002). Con la idea de una Web Semántica en la que las máquinas pueden entender, procesar y razonar acerca de los recursos para apoyar mejor y más cómodo a los seres humanos, las técnicas de organización y búsqueda de la información geográfica han cobrado gran importancia.

La idea básica detrás de la Web Semántica es la creación de una infraestructura que permita la automatización de las máquinas con un mínimo de intervención humana. Las aplicaciones sobre la Web Semántica pueden acceder a diversas fuentes de datos para recuperar la información pertinente, y luego unirlos para presentar los resultados coherentes y precisos para los usuarios de las aplicaciones. La heterogeneidad en los datos almacenados en los sistemas de información constituye una de las principales dificultades dentro de este campo y

por tanto, un área de investigación creciente dentro de diferentes dominios de aplicación. (Buccella et al., 2009)

En el dominio de la información geográfica el desarrollo de tecnologías para capturar información sobre la Tierra y el incremento en la distribución de los Sistemas de Información Geográfica (SIG) en la Web ha generado el almacenamiento de gran cantidad de información geoespacial. La razón fundamental para utilizar un SIG es la gestión de información espacial. Lo más característico de un SIG es su capacidad de generar nueva información a partir de un conjunto previo de datos mediante su manipulación y reelaboración. El sistema permite separar la información en diferentes capas temáticas y las almacena independientemente, permitiendo trabajar con ellas y facilitando al profesional la posibilidad de relacionar la información existente a través de la topología de los objetos, con el fin de generar otra nueva en algunas situaciones. Otra característica importante de los SIG es que están dirigidos a especialistas con conocimiento en este tipo de información.

Como un intento de llevar la Web Semántica al dominio geográfico y lograr representaciones del conocimiento geográfico en estructuras semánticas más fáciles de manipular por cualquier tipo de usuario surge la Web Semántica Geoespacial. Según (Garea and Gil, 2007) más del 80% de la información que se encuentra en la Web se considera información geográfica, además, dentro de los textos u otro tipo de información también se incluyen referencias geográfica (por ejemplo, en la noticias generalmente se hace mención al lugar donde ocurre el evento). El tener en cuentas las referencias geográficas proporciona un valor añadido a los sistemas de recuperación de información (RI) clásicos. La importancia de las referencias geográficas reside en las características especiales que tienen debido a su naturaleza espacial. Pero la complejidad y diversidad de la información geográfica (Smith and Mark, 1998), formada a través de las asociaciones cognitivas y las referencias espacio-temporales, impone el desarrollo de métodos específicos para las necesidades de la disciplina geográfica (Rodríguez, 2000).

El campo de la RI y el campo de los SIG han sido dos áreas de investigación que han avanzado de forma independiente a lo largo de los años. Por un lado, las estructuras de indexación y técnicas propuestas desde el campo de la RI no tienen en cuenta la naturaleza espacial de las referencias geográficas. Por otro lado, las

estructuras de indexación espacial no son directamente aplicables en sistemas de RI. Sin embargo, los usuarios demandan cada vez más servicios que les permitan situar la información en su contexto espacial e incluso acceder a esa información mediante consultas que tengan en cuenta las características especiales de la información espacial.

Esta demanda ha provocado que muchos investigadores de cada área empiecen a prestar atención a la otra, a esto se le suman las nuevas tendencias de la Web Semántica, dando lugar a un nuevo campo de investigación denominado Recuperación Semántica de Información Geográfica. El objetivo de este nuevo campo es “proponer arquitecturas de sistemas, estructuras de indexación y otros componentes que permitan desarrollar sistemas mediante los cuales los usuarios puedan recuperar y compartir información relevante tanto temática como geográficamente” (Delgado and Capote, 2009). Pero la información geográfica tiene unas características especiales que determinan una serie de diferencias entre un SIG y un sistema de información tradicional.

Dentro de las características particulares que tiene la información geográfica se encuentran las relaciones espaciales, éstas establecen la ubicación en el espacio de algún objeto geográfico del mundo real con respecto a otro. Describen la configuración geoespacial de los objetos modeladas a través de regiones, líneas y puntos. Las relaciones espaciales se basan en conceptos matemáticos. Basado en sus diferentes propiedades matemática, la literatura (Egenhofer and Mark, 1995a), (Mark, 1999), (Papadias and Kavouras, 1994) identifica tres tipos de relaciones espaciales: las relaciones topológicas, de distancia y de dirección.

Las diferencias entre un SIG y un sistema de RI se reflejan en todos los niveles del sistema desde el almacenamiento de datos, donde se necesitan nuevos tipos de datos y estructuras de acceso a ellos, hasta la presentación de información al usuario, donde se necesitan opciones de visualización mucho más avanzadas que en los sistemas tradicionales. Egenhofer destaca la necesidad de una forma simple canónica para plantear las consultas de los datos geoespaciales y los métodos para evaluar la semántica de las fuentes de este tipo de datos para ajustarlos a las consultas geoespaciales (Egenhofer, 2002). Estos métodos que propone Egenhofer requieren que los sistemas cuenten con representaciones especiales de los datos geográficos, con estructuras eficientes para el acceso a ellas, con algoritmos

específicos de consulta y con interfaces de usuario adaptadas para permitir la interacción con ellas.

Con respecto al acceso a la información Fonseca (Fonseca, 2001) propone utilizar como estructura para el desarrollo de aplicaciones geográficas a las ontologías para la integración de datos espaciales. Esta propuesta fundamenta la integración de ontologías con enlaces a fuentes de información geográfica por medio de su significado. La recuperación por conceptos permite además de que los sistemas puedan identificar mejor los datos, procesarlos en consecuencia y realizar análisis más eficientes con independencia del especialista, lo que constituiría una gran ventaja porque la información no solo será comprendida por las personas sino también por las máquinas incidiendo favorablemente sobre la reutilización del conocimiento, haciendo al sistema más potente y eficaz.

Las ontologías se presentan precisamente como una alternativa potente para resolver muchos de estos problemas de heterogeneidad de las fuentes de información, permitiendo además desambiguar los términos de búsquedas porque en ellas el conocimiento se especifica a través de conceptos (Montoyo, 2008). Prueba de ello es la tendencia que se manifiesta en numerosas investigaciones en el área del procesamiento semántico de la información geográfica tales como: (Aparício et al., 2005), (Fonseca et al., 2002a), (Aparício et al., 2006), (Kavouras and Kokla, 2001) (Giger and Najar, 2003), (Gómez-Pérez, 2002), (Dolbear et al., 2005), (Rodríguez and Egenhofer, 2003), (Doan and Halevy, 2004), (Chen et al., 2009) y (Schwering and Raubal, 2005) que abordan esta problemática. Estas soluciones se basan principalmente en el uso de ontologías como mecanismo de representación del conocimiento.

Los conceptos en las ontologías están vinculados y regidos mediante relaciones y axiomas que enriquecen el contexto que conceptualizan. La definición más consolidada de ontología es la propuesta por Gruber (1993) y extendida por Studer et. al (1998) que la describen como “una especificación explícita y formal sobre una conceptualización compartida”. La interpretación de esta definición es que las ontologías definen conceptos y relaciones de algún dominio, de forma compartida y consensuada; y que esta conceptualización debe ser representada de una manera formal, legible y utilizable por los ordenadores.

En (Vilches et al., 2006) podemos ver otras características de las ontologías como la posibilidad de interactuar semánticamente entre ellas, lo cual constituye un aspecto de avance, ya que posibilita obtener información desde otras fuentes. Esto significa que pudiera obtenerse datos desde otros sistemas incluso independientemente del lenguaje que tengan, dado que las consultas serían interpretables en un ámbito donde la información contenida en una ontología estaría en correspondencia con otra representación tal vez de la misma información, lo que es muy útil en el campo de la información geográfica. En (Rodríguez et al., 2006) se señalan 3 ventajas muy significativas sobre las otras formas de organización de la componente semántica.

- 1) Reducen la confusión semántica. Disminuye la ambigüedad terminológica al tener en cuenta sinónimos y polisemias, repercutiendo sobre la comunicación. Un sencillo ejemplo de la reducción de términos usando sinónimos con el uso de tesauros lo constituye la consulta: *Obtenga los hospitales próximos al municipio San José*. Aquí próximo nos indica cercano, inmediato, contiguo. Con cualquiera de estas palabras que se hubiese formulado la pregunta el sistema debiera ser capaz de identificar qué es lo mismo y responder igual.
- 2) Permite la posibilidad de reutilizar el conocimiento. Esta ventaja es la referida a la interacción entre distintas ontologías del dominio geográfico, ya que resultado del propio desarrollo de ontologías hay diversas formas de conceptualizar una misma realidad.
- 3) La traducción e intersección semántica a través de comparaciones utilizados para describir correspondencias entre fenómenos (ej. río, river, rivière y fleuve) y entre diferentes ontologías (ej. ontología de fenómenos hidrográficos y ontología de las ciudades).

Las ontologías han sido analizadas en la Geociencia como un procedimiento de estandarización que facilita la comunicación entre diferentes fuentes de información (Chandrasekaran et al., 1999), (Smith, 1999), (Fonseca et al., 2002b). En la integración de las fuentes de información distribuidas la ambigüedad en la interpretación de los conceptos, también conocida como heterogeneidad semántica, constituye uno de los principales obstáculos para la realización de este proceso. La heterogeneidad generalmente se reduce en dos pasos: (i) a través de la

comparación de los conceptos para determinar la alineación y (ii) ejecutando la alineación de acuerdo a las necesidades de la aplicación (ej. integración de esquemas, integración de datos, preguntas-respuestas) (Gal and Shvaiko, 2010).

Un enfoque de integración de datos consiste en la alineación de los conceptos que describen los datos en las aplicaciones con una ontología global. Una vez establecida la alineación los usuarios pueden consultar cientos de fuentes de información a través de una sola consulta que oculta la heterogeneidad subyacente. Una propuesta para crear una ontología común es crear una base de conocimiento en términos de una ontología global e independiente del dominio. Un ejemplo de esto lo constituye Cyc (Lenat and Guha, 1990), (Lenat et al., 1995) compuesta por aproximadamente 40000 objetos. Utilizando Cyc una entidad de un recurso de información es mapeado en los conceptos de la ontología global a través de un conjunto de axiomas de articulación (Collet et al., 1991). Otro ejemplo lo constituye DBpedia¹, este es un proyecto para la extracción de datos de Wikipedia, en la base de datos se describen 3.640.000 entidades, almacenadas en formato RDF (Resource Description Framework), las cuales pueden ser accedidas realizando consultas a la base de datos a través de SPARQL. Otro enfoque consiste en alinear las ontologías existentes para obtener una ontología integrada. OBSERVER (Mena et al. 1996) es un sistema que facilita la interoperabilidad entre ontologías existentes basado en las relaciones terminológicas (ej. sinónimos, hipónimos, hiperónimos) que conectan los términos en las diferentes ontologías. (Kashyap & Sheth 1998).

Entre las principales técnicas utilizadas en los sistemas de recuperación, toma de decisiones, la minería de datos y el reconocimiento de patrones, las medidas de similitud juegan un rol fundamental (Schwering, 2006). Se necesitan adecuadas mediciones de similitud para procesar toda la información disponible y recuperar solamente la que es relevante para la consulta.

Debido a las raíces interdisciplinarias que tiene la Recuperación de Información Geográfica (RIG), díganse los SIG y los sistemas de RI, esta hereda los componentes que tradicionalmente forman parte de esos sistemas. Dentro de estos componentes podemos citar los modelos para calcular la similitud semántica entre conceptos y la similitud espacial entre objetos geográficos. En este trabajo se han

¹ <http://dbpedia.org>

organizado las medidas de similitud en tres grupos, la similitud semántica las cuales calculan la similitud desde el punto de vista de los significados de los términos y son heredadas fundamentalmente de los sistemas de RI, la similitud espacial que calculan la similitud teniendo en cuenta las relaciones espaciales entre objetos geográficos utilizadas mayormente en los SIG, y por último, la similitud semántica espacial, en las cuales se calcula la similitud teniendo en cuenta las representaciones conceptuales tanto de los objetos como de las relaciones espaciales que existen entre ellos. La similitud semántica espacial es fundamental para el procesamiento semántico de datos geoespaciales (Quintero et al., 2012). Establece el grado de interoperabilidad semántica entre los datos o los diferentes SIG y constituye la base para la recuperación y la integración de información geoespacial (Janowicz et al., 2008).

Valorar la similitud es un proceso que requiere que dos “objetos” sean descompuestos en los elementos en los que son iguales y en los que son diferentes. Este tipo de valoración es parte del pensamiento lógico de los seres humanos, es intuitiva y subjetiva por lo que se dificulta su representación como un modelo matemático estricto (Tversky, 1977). En los SIG la similitud es particularmente importante debido a la dificultad para obtener representaciones de los fenómenos geográficos y a la variedad de formalizaciones que existen de las propiedades espaciales tales como su forma, localización y relaciones espaciales (Leake et al., 2002).

Para el procesamiento semántico del conocimiento de los datos geoespaciales almacenados en las ontologías es fundamental el cálculo de la similitud semántica, la cual es esencial en el procesamiento de las consultas de datos de los usuarios y es la base para la recuperación e integración de información semántica (Schwering, 2008). La comparación, a través de medidas de similitud, de los conceptos que describen el significado de los datos es un enfoque que apunta a convertirse en una solución eficiente para el procesamiento de los datos geográficos desde el punto de vista semántico (Larín, 2013).

1.2 Planteamiento del problema

Actualmente, la investigación dentro de la recuperación semántica de información geográfica enfrenta diversos retos como la extensión del enfoque sintáctico al

semántico y espacial. Se han realizado trabajos en estos sentidos como el presentado por (Schwering, 2006) en el cual se presenta un método que representa las relaciones espaciales a nivel conceptual y calcula las distancias semántica entre los conceptos. A pesar de ser un intento para tratar las relaciones espaciales se queda a nivel general. Por su parte, (Li and Fonseca, 2006) proponen una medida de similitud que integra cuatro modelos (geométrico, de características, de transformación y el de alineación estructurada) para calcular la similitud entre escenas espaciales aplica el orden de prioridad topología, dirección, distancia. La valoración se hace a nivel geométrico y temático pero se deja el nivel semántico propuesto para trabajos futuros.

Otro de los retos de la recuperación de información geográfica está relacionado con la determinación de intención del usuario expresada en las consultas. (Renteria, 2009) propone un modelo de recuperación, que integra un criterio semántico con criterios geoespaciales y en el que se propone como trabajos futuros complementar el modelo de recuperación con un módulo de Procesamiento de Lenguaje Natural, para procesar consultas y proponer componentes de análisis topológico a través de la implementación de medidas de similitud semántica.

Al determinar la similitud semántica entre conceptos geoespaciales es importante tener en cuenta sus relaciones espaciales en el proceso de cálculo. Todos los objetos geoespaciales tienen una posición en el espacio con respecto a algún sistema de referencia espacial y por lo tanto, una autocorrelación espacial. La similitud espacial es difícil de tratar debido a las numerosas limitaciones de las propiedades espaciales y de la complejidad de las relaciones espaciales (Li and Fonseca, 2006).

Todas estas dificultades se ven afectadas además por la heterogeneidad de las fuentes de información, lo cual, dificulta la extracción de términos geográficos desde fuentes de datos estructuradas y un reto aún mayor desde fuentes no estructuradas teniendo en cuenta criterios espaciales. Como consecuencia de esto se identifica la necesidad de estructuras de representación del conocimiento que permita mejorar los procesos de recuperación

La mayoría de las representaciones semánticas de los conceptos espaciales se basan en características o descriptores de la información que tienen un carácter ambiguo en la explicación de lo que representan estas características. Gran parte

de los modelos más empleados para determinar la similitud semántica se aplican a dominios generales, no se enfocan en las propiedades particulares de los conceptos en el dominio espacial (Schwering, 2006) y utilizan una ontología compartida que relaciona los conceptos semánticamente. Los modelos basados en las relaciones semánticas generalmente incluyen dos tipos de relaciones: la sinonimia (equivalencia) y la hiponimia (es-un). En el dominio espacial es importante tener en cuenta la relación de meronimia (es parte de) que proporciona una mejor representación de las interrelaciones entre los conceptos espaciales. (Larín, 2013) desarrolla un método para la representación semántica multidimensional de objetos geoespaciales integrados en un entorno de Sistemas de Información Geográfica Gobernados por Ontologías (SIGGO). No obstante, no considera las relaciones topológicas.

El análisis de esta problemática ubica como **problema científico** la siguiente pregunta: ¿Será posible consultar datos geoespaciales de forma simple y obtener resultados satisfactorios mejorando la recuperación de la información geoespacial mediante la integración de la similitud semántica con la similitud espacial?

La determinación del problema científico situó como **objeto de estudio** las mediciones de la similitud semántica de la información geoespacial en el campo de la recuperación semántica de información geográfica.

Este trabajo parte de la **hipótesis** de que una metodología de recuperación de información que aplique medidas similitud semántica que integren las relaciones espaciales en la evaluación de la distancia conceptual dentro de una ontología, contribuye a mejorar la gestión semántica de la información geoespacial.

1.3 Objetivos.

Como respuesta al problema señalado el **objetivo general** consiste en diseñar una metodología de recuperación de información geoespacial, en la que se apliquen medidas de similitud semántica integrando las relaciones conceptuales y espaciales asociadas a los objetos geográficos representadas en una ontología.

Para la solución del problema científico y dar cumplimiento al objetivo general, se plantearon los siguientes **objetivos específicos**:

1. Revisar el estado actual de los sistemas de recuperación semántica de información geográfica.
2. Identificar y analizar las relaciones semánticas y espaciales que existen entre los objetos geográficos.
3. Diseñar una ontología que describa las principales propiedades y relaciones de los objetos en el dominio geográfico en la que se mapeen² información de fuentes de datos heterogéneas.
4. Analizar las potencialidades de las medidas de similitud semántica para la gestión de información geoespacial.
5. Seleccionar la medida de similitud semántica más apropiada para la gestión de información.
6. Estudiar la aplicación de la similitud semántica a la ontología.
7. Definir una metodología para la recuperación de información geoespacial orientada a personas no expertas.
8. Validar la metodología con la estrategia de recuperación propuesta.

1.4 Alcance e interés de la investigación.

Como se ha comentado anteriormente el campo de la Recuperación de Información Geográfica (RIG) es reciente, sin embargo, este campo está atrayendo el interés de muchos investigadores. Como prueba de este interés se puede citar el *Workshop on Geographic Information Retrieval* que se celebra anualmente desde el año 2004 en congresos tan prestigiosos como ACM SIGIR ("SIGIR", 2014) o ACM CIKM ("CIKM", 2008). Los informes elaborados como resumen de estos *workshops* identifican, entre otras líneas de investigación fundamentales en recuperación de información geográfica, las siguientes: arquitecturas para motores de búsqueda geográficos, indexación espacial de documentos e imágenes, extracción del contexto espacial de documentos y otros conjuntos de datos y diseño, construcción, mantenimiento y métodos de acceso para ontologías espaciales, *gazetteers* y tesauros geográficos.

Durante los últimos años se han propuesto nuevas arquitecturas de sistemas, estructuras de indexación y otros componentes que, si bien han supuesto

² El término mapeo de información se refiere a una comparación entre los elementos que describen dos datos representados por distintos modelos.

considerables avances en el estado del arte, dejan todavía muchos temas abiertos en todas estas líneas de investigación.

Entre **los aportes prácticos** esperados con este trabajo se encuentran:

- El desarrollo de una metodología de recuperación en la cual la principal aportación sobre otras propuestas es que, extiende los criterios de recuperación de información hacia criterios geográficos y de semántica espacial, es decir, combina medidas semánticas con análisis de las relaciones espaciales en la cuales se aplica el orden de topología, dirección y distancia.
- La metodología se define a través de una arquitectura para sistemas completos de recuperación. Esta arquitectura es genérica y está definida desde un punto de vista global, abarcando todas las tareas necesarias para la construcción del repositorio de documentos y su posterior explotación mediante consultas realizadas por los usuarios. Tiene un carácter modular y extensible que le permite adaptarse a nuevos resultados que puedan surgir relacionados con este tipo de sistemas.
- La aplicación de una ontología de dominio que permite simular el rol de un experto al que se le puede preguntar acerca de un tópico y esperar respuestas acertadas o relacionadas.
- La integración de fuentes de datos heterogéneas con la cual se enriquece el grado de conocimiento que se tiene de los objetos geográficos, ejemplo: a partir de los análisis espaciales podemos conocer si un río atraviesa una región de suelo y podemos visualizar estos objetos en un mapa geoespacial, conceptualmente se puede definir una clasificación para dicho tipo de suelo o las características que presentan y mostrar imágenes relacionadas con estos objetos geográficos.
- La aplicación de técnicas de procesamiento del lenguaje natural a la solicitud como la expansión de la consulta basada en la ontología facilita a los usuarios no especializados recuperar resultados relevantes sin tener que conocer cómo expresar la consulta.
- El desarrollo de un prototipo de sistema siguiendo la metodología para validar los componentes propuestos

1.5 Organización de la tesis.

El presente documento se encuentra estructurado en siete capítulos. A continuación de la introducción se presenta el Capítulo 2, donde se abordan conceptos y definiciones relacionados con la fundamentación y utilización de la similitud semántica y su aplicación en el campo de la recuperación de información geográfica. Se identifican y formalizan las relaciones espaciales que describen el comportamiento de los objetos geográficos. Se analizan revisiones bibliográficas que demuestran el desarrollo de las aplicaciones de recuperación que manipulan objetos geoespaciales teniendo en cuenta su semántica y las relaciones espaciales demostrando la actualidad de la investigación. De las propuestas revisadas se identifican las fortalezas y debilidades para definir las acciones a considerar.

En el Capítulo 3 se analizan las medidas de similitud semántica y espacial realizando una comparación entre ellas para determinar la más adecuada para su aplicación en la metodología. En el Capítulo 4 se presenta el diseño de la ontología desarrollada como parte de esta investigación. Inicialmente se realiza la conceptualización del dominio geográfico, en el cual, se identifican las propiedades y relaciones que se modelaran en el cuerpo de la ontología.

En el Capítulo 5 se describen cada una de las fases que incluye la metodología de recuperación semántica de objetos geográficos. Inicialmente se explica la estructura organizativa de la información con la que se va a trabajar. Posteriormente, se presenta el procesamiento que se le realiza a la solicitud del usuario para identificar cuáles son los elementos relevantes de la misma para lanzar la búsqueda y finalmente se explica cómo se realiza la recuperación y ponderación de cada uno de la información recuperadas.

En el Capítulo 6 se presentan las pruebas experimentales y resultados de la aplicación de la metodología realizando las evaluaciones correspondientes. Seguido de este capítulo se emiten las conclusiones generales del trabajo de investigación, donde se recogen las principales aportaciones y finalmente se muestran las producciones científicas derivadas de esta investigación.

2. Estado del arte.

En este capítulo se realiza un estudio del estado del arte de la recuperación semántica de información geográfica y de los campos de los cuales surge el mismo: la recuperación de información y los sistemas de información geográfica. Inicialmente, en la sección 2.1 se estudia la recuperación de información. Posteriormente, en la sección 2.2 se presentan las características fundamentales de los sistemas de información geográfica. Estas secciones tienen como objetivo presentar los conceptos básicos de cada campo e introducir de esta forma el estudio del estado del arte en el campo de la recuperación de información geográfica, sección 2.3. Posteriormente, en la sección 2.4 se presenta la recuperación de la información geográfica con enfoques semánticos, además, se explican diferentes modelos de mediciones de similitud semántica en las que se incorporan elementos espaciales como componente fundamental en este campo de investigación. Finalmente, se presentan en la sección 2.5 las conclusiones del capítulo.

2.1 Contextualización y motivación.

La Recuperación de Información (RI) basa sus procesos en los sistemas de información. Al hablar de los sistemas de información, es conveniente mencionar que, para que éstos den soporte a los procesos de recuperación se debe hablar de la cantidad y calidad de la información; pero un sistema de información no solo toma importancia por el volumen de información que contiene, sino también por la facilidad que proporciona el sistema para acceder a la información contenida, también por la representatividad y finalmente por el grado de satisfacción que la organización de esta información brinda al usuario.

Uno de los principales problemas que existe a la hora de tratar la recuperación de información es el desconocimiento de la gente no relacionada con el campo acerca de qué abarca este concepto. Se han publicado muchos recursos que abordan este tema y proporcionan definiciones aclarando qué es exactamente la recuperación de información.

Manning et. al (2008) realizan una definición muy intuitiva de este concepto: “la RI se ocupa de encontrar material (normalmente documentos) de una naturaleza no estructurada (normalmente texto) que se encuentra en grandes colecciones (normalmente almacenada de forma digital) y satisface una necesidad de información”. Esta definición hace referencia a una de las principales confusiones al hablar del tema: la naturaleza no estructurada del material a buscar. Esto deja fuera del ámbito de la RI temas como la búsqueda en bases de datos, donde la información se encuentra estructurada en un conjunto de campos, las búsquedas son exactas y los sistemas deben satisfacer ciertos requisitos de control de concurrencia, recuperación, etc. Además, la definición es lo suficientemente general como para no dejar fuera otros temas que se encuentran en el campo de la RI y que, en contra de un pensamiento bastante generalizado, no es solamente recuperación de documentos. La lingüística computacional (computational linguistic) (Mitkov, 2003), la búsqueda de respuestas (*question answering*) (Vicedo and Ferrández, 2000), la clasificación de textos (automated text categorization) (Cowie and Wilks, 2000) y la recuperación de información multilingüe (*cross language information retrieval*) (Martínez, 2004) son algunos de los temas más conocidos que se encuentran dentro de este campo. Una de las principales fuentes de confusión es la popularidad que han alcanzado los motores de búsqueda en la web, como Google, Yahoo, Lycos, etc. Esta popularidad provoca inevitablemente que los no expertos en el tema limiten el concepto RI al ámbito de la búsqueda en web.

La idea fundamental que se encuentra bajo la mayoría de los sistemas de este tipo es la similitud. Según la definición de Manning, el objetivo de estos sistemas consiste en recuperar material que satisfaga una necesidad de información. La necesidad de información suele estar expresada con una consulta y, por tanto, recuperar material que satisfaga esa necesidad se puede traducir en encontrar material que sea lo más parecido posible a los términos clave indicados en la consulta generalmente ordenadas por relevancia. Existe también el enfoque de exploración, en el cual el usuario se encarga de explorar visualmente sin tener que expresar previamente a través de una consulta cuáles son sus necesidades de información; es decir, el usuario reconoce lo que está buscando, sistemas de

ejemplo que incluyen esta forma de recuperación en un modo general son e-bay³, Mercadolibre⁴, entre otros.

El enfoque de esta investigación va dirigido a la recuperación de información que, mediante la utilización de medidas para determinar similitudes, se acerca en el mayor grado posible, a satisfacer las necesidades de los usuarios.

En la Figura 2.1 se presenta un esquema simplificado del proceso de recuperación de información. En la parte superior de la figura se representan las necesidades de información de los usuarios y la información con la que se alimenta el sistema. Estos conceptos no son lo suficientemente concretos como para ser manejados por sistemas de información. Por este motivo, el esquema indica que deben ser representados de manera más concreta en consultas y objetos indexados respectivamente. Mediante la *comparación* de las consultas con los objetos indexados se pueden obtener los *objetos recuperados* que se le presentarán al usuario.

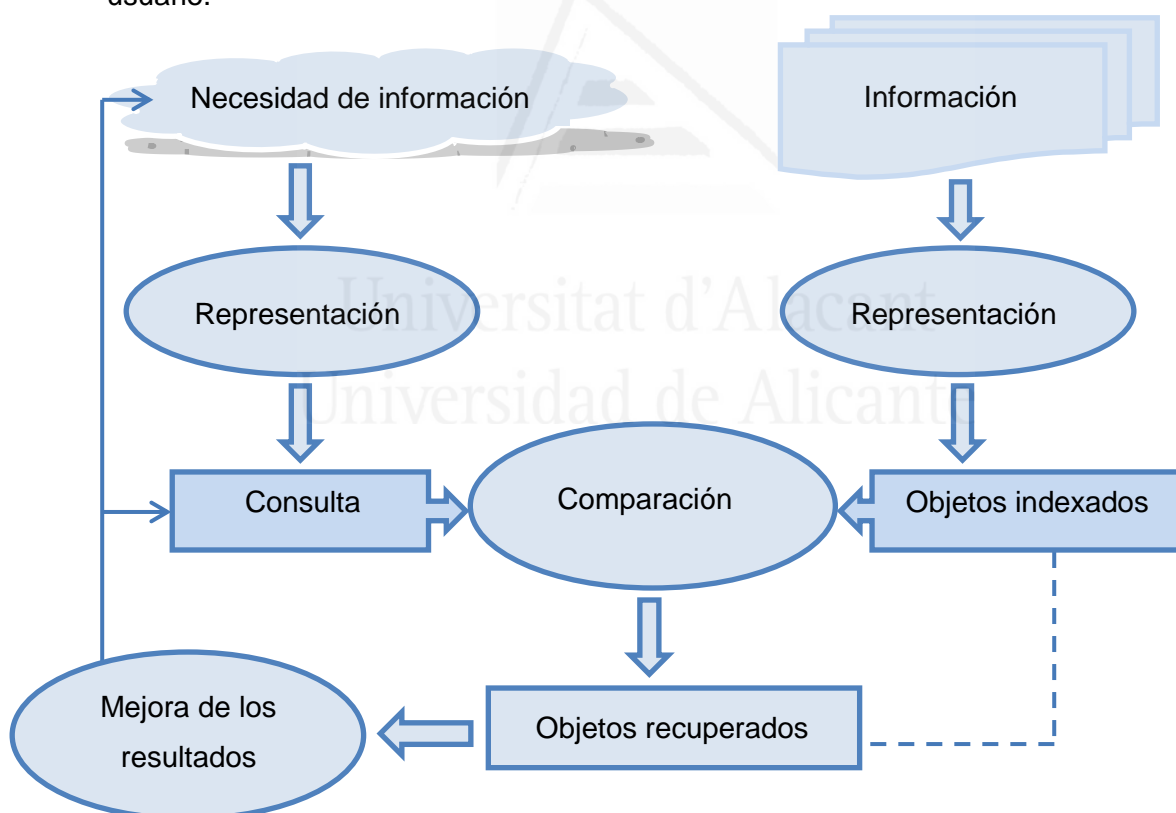


Figura 2.1. Esquema del proceso de recuperación de información.

³ <http://www.ebay.com/>

⁴ <http://www.mercadolibre.com/>

El esquema contempla también una tarea de mejora de los resultados por medio de la cual se puede especificar más la necesidad de información o la consulta para obtener resultados que se ajusten mejor a las necesidades del usuario. Esta tarea la puede llevar a cabo tanto el usuario como el propio sistema de manera automática.

Este esquema, aunque simplificado, representa las tareas más importantes que intervienen en la recuperación de información. Las tareas representadas en la Figura 2.1 con formas elípticas, son las que diferencian los distintos métodos o modelos de recuperación de información. Es decir, la manera de representar la información y las consultas, y la forma de medir la similitud entre ambos son algunas de las características distintivas de los modelos de recuperación de información. En la sección 2.1.1 se presentan algunos ejemplos de estos modelos como el booleano, el probabilístico y el vectorial. Estos modelos necesitan unas estructuras físicas que den soporte al proceso de acceso a la información de manera eficiente. Además, en la sección 2.1.2 se describe la expansión de consultas, una técnica muy habitual que se encuentra enmarcada dentro de la tarea de mejora de los resultados del sistema. Finalmente, aunque es una tarea que no se encuentra representada en el esquema de la figura, se habla de la evaluación de este tipo de sistemas. Esta tarea, destinada fundamentalmente a determinar si los sistemas de RI funcionan correctamente y de si un cambio determinado mejora su eficacia y/o eficiencia, se describe en la sección 2.1.3. Aunque el campo es mucho más amplio estas son las tareas más importantes en lo que se refiere al alcance de esta tesis.

2.1.1 Modelos de recuperación de información.

Los modelos de recuperación tienen que proporcionar de manera implícita o explícita una definición de relevancia para representar tanto documentos como consultas y establecer como se realiza la comparación de esas representaciones (Seco, 2009). Además, pueden describir el proceso computacional, el proceso humano y las variables que intervienen en el proceso global.

En esta sección se analizan tres modelos (el modelo booleano, el modelo probabilístico y el modelo vectorial), estos modelos representan tanto los documentos como las consultas mediante un conjunto de términos clave (*keywords*)

que se emplean como términos de indexación. La principal diferencia entre ellos reside en la forma de representar dichos términos y en la forma de relacionarlos. Otra característica que diferencia los distintos modelos de recuperación de información consiste en cómo realizan la comprobación de si un documento se ajusta a una consulta. Si el documento se tiene que corresponder exactamente con la consulta se trata de un modelo de búsqueda exacto (*exact match*), mientras que si el documento se puede corresponder sólo de manera parcial con la consulta se trata de un modelo de mejor ajuste (*best match*) (Muñoz, 2004). En los modelos del primer grupo el resultado de una consulta es un conjunto (no existe orden entre los elementos) donde se encuentran los elementos que satisfacen la búsqueda. Por el contrario, en los modelos del segundo grupo el resultado es una lista ordenada según lo bien que se ajustan a la consulta. Entre las ventajas de los modelos del primer grupo están el ser muy eficientes, previsibles, fáciles de explicar y que funcionan bien cuando el usuario es capaz de expresar exactamente lo que necesita (aunque esta es también su principal desventaja ya que es muy complicado que eso suceda). El modelo booleano es un ejemplo de este tipo. Los modelos de mejor ajuste, como el probabilístico o el vectorial, son más sencillos de emplear y, por tanto, significativamente más eficaces; además, existen implementaciones optimizadas que proporcionan una eficiencia similar.

2.1.1.1 Modelos booleanos.

El modelo booleano es uno de los modelos de recuperación de información más sencillos de utilizar. Está basado en la teoría de conjuntos y el álgebra de Boole. En este modelo las consultas son expresiones lógicas en las que los operandos son propiedades de los documentos. Los operadores booleanos puros son la conjunción, la disyunción y la negación (AND, OR, NOT). Además, la mayoría de los sistemas incluyen operadores de proximidad y expresiones regulares sencillas. En general, en estos sistemas el usuario refina la consulta hasta que el número de documentos relevantes que se recupera es razonable. Por tanto, el usuario está acostumbrado a realizar un proceso de retroalimentación por relevancia (*relevance feedback*).

La idea en la que está basado este modelo es que un término clave puede estar presente o ausente en un documento y, por tanto, sólo serán relevantes aquellos documentos que contengan los términos clave que se indica en la consulta.

Siguiendo estrictamente esta definición del modelo, el resultado de una consulta no contendría documentos que pueden ser relevantes a pesar de no encajar a la perfección con la consulta. Este grave inconveniente provocó que la mayoría de las implementaciones no sigan exactamente la definición sino que, basándose en esa idea, devuelven documentos que no se ajustan perfectamente a la consulta, aunque si en alguna medida, y ordenan el resultado asignándole pesos a los operadores booleanos, modelo booleano extendido.

La principal ventaja de este modelo es su sencillez y su eficiencia. No obstante, su principal inconveniente es la dificultad para formular exactamente las necesidades de información de los usuarios empleando el álgebra de Boole y términos clave. Otro inconveniente es que sólo obtiene los documentos que se ajustan perfectamente a la consulta y no proporciona nivel de relevancia a los resultados. Aunque este inconveniente se ha tratado de resolver en muchas implementaciones, la eficacia del modelo es inferior a la de otras alternativas.

2.1.1.2 Modelos vectoriales.

La idea fundamental en la que se basa el modelo vectorial (Salton and Lesk, 1968) es considerar que tanto los términos claves con respecto a un documento como las consultas se pueden representar a través de un vector en un espacio de k -dimensional. Por tanto, para evaluar la similitud entre un documento y una consulta; simplemente hay que realizar una comparación de los vectores que los representan.

Uno de los métodos más habituales para comparar el grado de similitud es calcular el coseno del ángulo que forman ambos vectores (Salton et al., 1975).

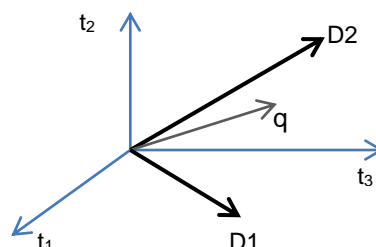


Figura 2.2. Comparación de similitud en el modelo vectorial

En la Figura 2.2 se presenta un ejemplo en el que se muestran los vectores representativos de dos documentos ($D1$ y $D2$) y de la consulta (q). Intuitivamente, el documento $D2$ es más relevante para la consulta q que el documento $D1$. La fórmula del coseno no hace más que trasladar esta idea intuitiva a la formalización de los vectores. El ángulo que forman los vectores de $D2$ y q es menor que el que forman los vectores de $D1$ y q , en consecuencia, el coseno del primer ángulo es mayor que el del segundo.

La cuestión que queda pendiente es cómo se construyen los vectores para representar los documentos y las consultas. En ambos casos, la construcción se realiza a partir de los vectores de los términos clave que los componen (por ejemplo, sumando dichos vectores). Estos vectores de términos clave tienen unos coeficientes (longitudes y pesos) que representan entre otros factores la presencia del término y su importancia. Por ejemplo, en la Figura 2.2 el vector del documento $D1$ se construye a partir de la suma de los vectores que representan la importancia de los términos clave $t1$, $t2$ y $t3$ en ese documento. La construcción de los vectores que representan a $D2$ y a q se realiza de manera análoga.

La definición de cómo se tienen que establecer los coeficientes que representan la importancia de cada término en el documento varía según el modelo utilizado, aunque, el método $tf * idf$ (Kaszkiel et al. 1999) (o alguna de sus variantes) es el más empleado. Este método se basa en que los pesos están determinados por la frecuencia de aparición del término clave en el documento (el parámetro tf) y la frecuencia inversa de aparición del término en el conjunto total de documentos (el parámetro idf). Intuitivamente, el factor tf captura la idea de que un término clave es más importante en un documento donde se menciona muchas veces que en otro donde se menciona menos. El factor idf , en cambio, captura la idea de que un término clave que aparece en muchos documentos tiene menor poder a la hora de discriminar documentos relevantes que otro que aparece en menos. Para intentar independizar los resultados del tamaño de los documentos, se suele normalizar las frecuencias tf .

2.1.1.3 Modelos probabilísticos.

El modelo probabilístico intenta resolver el problema de la recuperación de información a través del cálculo de probabilidades. Este modelo define la

recuperación como un proceso de clasificación (Robertson and Jones, 1976), forma tal que para cada consulta existen dos clases: la clase de los documentos relevantes y la clase de los documentos no relevantes. Por tanto, dado un documento D y una consulta determinada se puede calcular con qué probabilidad pertenece el documento a cada una de esas clases. Si la probabilidad de que pertenezca a la clase de los documentos relevantes es mayor que la probabilidad de que pertenezca a la clase de los no relevantes (es decir, $P(D|R) > P(D|NR)$) el documento será relevante para la consulta. El objetivo es localizar aquellos documentos que maximicen la probabilidad de pertenecer al conjunto de documentos relevantes. La medida de similitud se define según la ecuación 2.1.

$$sim(q, D) = \frac{P(D|R)}{P(D|NR)}. \quad (2.1)$$

Donde:

- $P(D|R)$: es la probabilidad de que el documento se halle en el conjunto de documentos relevantes.
- $P(D|NR)$: es la probabilidad de que el documento se halle en el conjunto de documentos no relevantes.

Las diferentes maneras en las que se pueden estimar estas probabilidades dan lugar a los diferentes modelos probabilísticos. Este modelo asume que la relevancia de un documento para una consulta es independiente del resto de documentos de la colección.

Este modelo necesita un método para calcular las probabilidades iniciales. Aunque existen numerosas alternativas para realizar este cálculo, la más habitual posiblemente sea emplear las frecuencias de los términos clave en cada documento y la frecuencia en el total de los documentos. Además, el modelo es iterativo, tras una primera aproximación donde se obtiene un subconjunto inicial se emplean los documentos en dicho subconjunto para refinar la búsqueda, recalculando las probabilidades. Este proceso se repite hasta obtener las probabilidades definitivas.

2.1.2 Expansión de consultas.

La expansión de consultas (*query expansion*) (Baeza-Yates and Ribeiro-Neto, 1999) se enmarca en la tarea de mejora de los resultados que se muestra en el esquema simplificado de la RI en la sección 2.1. Esta tarea se centra en la necesidad de mejora de las consultas realizadas por los usuarios para lidiar con problemas derivados de la sinonimia. Este término se refiere al empleo de diferentes palabras para referirse al mismo concepto y tiene un impacto muy grande en el correcto funcionamiento de los sistemas de recuperación de información.

La idea básica para mejorar los resultados reside en refinar la búsqueda realizada originalmente por el usuario con más información relacionada con los términos buscados. Según como se obtenga dicha información para refinar las búsquedas se puede hablar de distintos métodos de mejora de los resultados. La principal clasificación de estos métodos los divide en métodos basados en tesauros, métodos de retroalimentación (*relevance feedback*), métodos de análisis local y métodos de análisis global (Manning et al., 2008).

Un ejemplo típico de expansión de consultas consiste en emplear un tesoro, por ejemplo WordNet⁵ (Miller et al., 1990), para extender la consulta automáticamente con términos que pueden estar relacionados con los conceptos buscados. Un tesoro se puede definir como un diccionario que contiene listados de palabras o términos empleados para representar conceptos. Con el uso de los tesauros se puede analizar sinónimos (salino-salado), hiperónimos (halomórfico-salino), hipónimos (rendzina-húmico). Se pueden añadir variantes morfológicas o derivaciones de los términos. Es importante señalar que estas mejoras pueden traer problemas cuando se añaden sinónimos con sentidos diferentes al original. Además de emplear tesauros, también se ha propuesto el empleo de ontologías para realizar la expansión de consultas. Aunque no es el objetivo de esta sección explicar las aproximaciones para expansión de consultas basadas en ontologías, se introduce este concepto ya que la estructura de indexación que se utiliza en la metodología propuesta en esta basada en una ontología.

En la retroalimentación el usuario decide que documentos de los primeros 10 o 20 presentados son relevantes. Se añaden sus palabras claves (las de mayor

⁵ <http://www.cogsci.princeton.edu/~wn/>

frecuencia) a las de la pregunta y se vuelve a lanzar la pregunta. La ventaja de este método consiste en que se evita al usuario la reformulación de la pregunta. Los problemas radican en la eficiencia puesto que hay que lanzar nuevamente la pregunta. El número de términos añadidos se puede volver demasiado grande y necesita la intervención del usuario.

Al igual que la retroalimentación se realiza el método de análisis local solo que de forma automática sobre los documentos más relevantes. En el método de análisis global se establecen relaciones de co-ocurrencia entre palabras. En la fase de indexación se crea la matriz de co-ocurrencia y de forma automática, se añaden a la pregunta las palabras que suelen aparecer conjuntamente.

2.1.3 Evaluación en la recuperación de información.

La evaluación de los sistemas desarrollados dentro del campo ha sido uno de los temas a los que más atención se le ha prestado desde los orígenes de la RI. Dentro de este tema se abordan cuestiones relacionadas con las medidas acerca de lo buenos que son los sistemas desarrollados (si cumplen su función de manera eficaz y eficiente), si las mejoras aplicadas sobre un sistema realmente mejoran sus resultados, etc.

Algunos de los aspectos que se pueden evaluar en un sistema de RI son la velocidad a la hora de obtener resultados, la presentación de los documentos o la habilidad para encontrar documentos relevantes. Sin embargo, con el paso del tiempo la evaluación de la eficacia de la recuperación se ha ido imponiendo como el aspecto más importante a evaluar en estos sistemas. Además, se han propuesto diferentes funciones para medir esta eficacia. Aunque existen otras medidas que podríamos citar en este apartado sin duda, la eficacia (precisión y cobertura), la eficiencia (temporal y espacial) y la satisfacción del usuario (interfaz, modo de presentación de los resultados, etc). En (Baeza-Yates and Ribeiro-Neto, 1999) y (Manning et al., 2008) se puede encontrar una buena descripción de estas medidas.

Generalmente todas estas medidas se basan en el concepto de relevancia. Como no es fácil proporcionar una definición de relevancia completamente satisfactoria diferentes modelos proporcionan su propia definición de relevancia. Intuitivamente, decidir si un documento es relevante o no depende de factores como quién juzga

esa relevancia, qué criterio sigue para hacerlo (los seres humanos no suelen ser muy consistentes en su criterio), etc. Por estos motivos, con colecciones reales de documentos es imposible conocer el conjunto de documentos relevantes para cada consulta y los modelos definen este concepto en términos como la similitud entre los documentos y las consultas, la probabilidad de que un documento pertenezca a la clase de los documentos relevantes, etc.

Una vez que todos los modelos realizan su propia definición de relevancia es posible valorar los sistemas con funciones como la precisión y la cobertura (*recall*). Es conveniente efectuar la evaluación sobre colecciones de pruebas, las cuales hayan sido previamente evaluadas. Sobre las que se dispongan resultados de otros investigadores y deben estar formadas por documentos, preguntas y documentos relevantes para cada pregunta.

Existen distintas forma de construir las colecciones de pruebas: la tradicional, en la cual se estudia la relevancia de cada documento para cada pregunta y la segunda en la que solo se estudia la relevancia de los documentos devueltos. Esto tiene como ventaja que permite trabajar con colecciones más grandes pero tiene la desventaja de que no se sabe la relevancia de todos los documentos.

La precisión calcula la proporción de documentos relevantes que se han recuperado dentro del conjunto total de los documentos que se han recuperado. La ecuación 2.2 define la fórmula de la precisión empleando la notación de conjuntos del álgebra de Boole.

$$precision = \frac{|relevantes \cap recuperados|}{|recuperados|} \quad (2.2)$$

La cobertura calcula la proporción de documentos relevantes que se han recuperados dentro del conjunto total de los documentos relevantes que existen en la colección. La ecuación 2.3 define la fórmula de la cobertura empleando la notación de conjuntos del algebra de Boole.

$$cobertura = \frac{|relevantes \cap recuperados|}{|relevantes|} \quad (2.3)$$

Intuitivamente la precisión mide que el sistema no recupere muchos documentos que en realidad no son relevantes y la cobertura mide que el sistema recupere una

buena proporción de los documentos relevantes. Un buen sistema de RI debe presentar buenas medidas tanto de precisión como de cobertura ya que una cobra sentido en función de la otra. Por ejemplo, se podría alcanzar una cobertura del 100% recuperando todos los documentos ya que en el conjunto resultado están incluidos todos los documentos relevantes. Sin embargo, la precisión en ese caso sería muy baja ya que también se están recuperando muchos documentos no relevantes.

Existen otras medidas evaluación que se enfocan en la complejidad del sistema en cuanto al espacio ocupado por la información indexada o el tiempo de respuesta del sistema. Estas medidas no son las más utilizadas en investigación aunque si son fundamentales en los sistemas comerciales.

2.2 Sistemas de información geográfica.

Otro campo que ha recibido mucha atención en los últimos años es el de los sistemas de información geográfica (Worboys, 1992). Las mejoras recientes en el hardware han hecho posible que la implementación de este tipo de sistemas sea abordable por muchas organizaciones. Además, se ha llevado a cabo un importante esfuerzo colaborativo por parte de dos organismos internacionales (ISO (ISO/IEC, 2002) y el Open Geospatial Consortium (OGC, 2003) para definir estándares y especificaciones para la interoperabilidad de los sistemas. Este esfuerzo ha hecho posible que muchas organizaciones públicas estén trabajando en la construcción de infraestructuras de datos espaciales (“Global Spatial Data Infrastructure Association”, 2012) que les permitan compartir su información geográfica.

Al igual que la propia geografía, es difícil definir el término sistema de información geográfica (SIG) ya que engloba la integración de áreas muy diversas. Podría definirse a priori como un poderoso conjunto de herramientas para recoger, almacenar, recuperar, transformar y mostrar datos espaciales del mundo real (Burrough, 1988). Por esto no existe una única definición de SIG totalmente consensuada. En la literatura podemos encontrar muchas definiciones diferentes del término SIG y cada una de ellas considera la funcionalidad de estos sistemas desde diferentes perspectivas. En todas ellas podemos identificar tres aspectos diferentes:

- Un SIG es una colección de herramientas informáticas para realizar análisis y simulaciones geográficas (Laurini and Thompson, 1992,), (Burrough and McDonnell, 1998), (Rigaux et al., 2001).
- Un SIG está basado en un conjunto de estructuras de datos y algoritmos para recuperar y manipular información geográfica (Worboys, 2004),(Rigaux et al., 2001).
- Un SIG es una utilidad que ayuda a tomar decisiones en tareas relacionadas con la geografía (Longley et al., 2001), (Harmon and Anderson, 2003).

Combinando estos tres aspectos se podría describir un SIG como un “sistema informático para modelar, capturar, almacenar, manipular, consultar, recuperar, analizar y visualizar información de manera eficiente, donde parte de esa información es de naturaleza geográfica”.

Una definición de SIG bastante aceptada es la redactada por el NCGIA (National Centre of Geographic Information and Analysis): “Un SIG es un sistema de hardware, software y procedimientos elaborados para facilitar la obtención, gestión, manipulación, análisis, modelado, representación y salida de datos espacialmente referenciados, para resolver problemas complejos de planificación y gestión” (NCGIA, 1990).

Lo más característico de un SIG es su capacidad de generar nueva información a partir de un conjunto previo de datos mediante su manipulación y reelaboración. La razón fundamental para utilizar un SIG es la gestión de información espacial. El sistema permite separar la información en diferentes capas temáticas y las almacena independientemente, permitiendo trabajar con ellas de manera rápida y sencilla, y facilitando al profesional la posibilidad de relacionar la información existente a través de la topología de los objetos, con el fin de generar otra nueva en algunas situaciones.

Los modelos más recurridos a la hora de representar información espacial son el vectorial y el *raster*. Esta investigación se concentra en uno de ellos (vectorial) y estudia sus potencialidades en aras de incorporar parte de este modelo en la solución propuesta.

2.2.1 El modelo de datos vectoriales.

En la actualidad existe una diversidad de modelos que representan las propiedades geométricas y las relaciones topológicas con que se caracterizan los datos espaciales. El modelo vectorial es junto al *raster*, uno de los modelos más utilizados por los SIG debido a que brinda buen soporte para representar los fenómenos del mundo geográfico. Su utilización se basa sobre la estructura de representación en correspondencia con las características del tipo de objeto. En el caso vectorial, permitiendo representar más adecuadamente a los objetos que poseen una geometría definida, o sea, a los objetos asociados a fenómenos discretos que puedan ser descritos mediante puntos, líneas o polígonos, mientras que el modelo *raster* corresponde a objetos que pueden representarse mediante rejillas de valores temáticos, o sea, se asocia a los objetos que ostentan una continuidad espacial como pueden ser los modelos digitales del terreno. A pesar que se menciona el formato *raster* por su alto grado utilización en nuestros días no es objetivo de este estudio hacer una comparación entre ambos modelos, sino concentrar la atención en el modelo vectorial.

En este modelo se representan las características geográficas de los objetos a través de datos vectoriales de forma parecida a como lo hacen los mapas, estableciendo que cada tipo de representación geométrica puede estar definida en dos o tres dimensiones y en un sistema de coordenadas determinado. Los vectores posibilitan el almacenamiento explícito de la localización de los nodos y la topología definiendo a las entidades solamente por sus límites y los arcos conectados, pero solo recopila aquellos puntos que delimitan a las entidades y no considera el espacio fuera de éstas. Al almacenar los datos sólo de los elementos digitalizados requiere menos memoria para su almacenamiento y posterior tratamiento.

Los elementos básicos que lo componen son los puntos, líneas o arcos y los polígonos, como muestra la Figura 2.3. A continuación se presentan una serie de conceptos afines (Mendoza, L., 2008), (Rodríguez, 2008):

Punto. Abstracción de un objeto de cero dimensiones representado por un par de coordenadas X, Y. Normalmente un punto representa una entidad geográfica demasiado pequeña para ser representada como una línea o como una superficie; por ejemplo, la localización de un edificio en una escala de mapa pequeña, o la

localización de un área a la que una instalación da servicio en una escala de mapa media.

Línea. Conjunto de pares de coordenadas ordenados que representan la forma de entidades geográficas demasiado finas para ser visualizadas como superficies a la escala dada (ej.: curvas de nivel, ejes de calles o ríos), o entidades lineales sin área (límites administrativos). Una línea es sinónimo de arco.

Polígono. Entidad utilizada para representar superficies (ej.: una ciudad, una parcela o un lago). Un polígono se define por las líneas que forman su contorno y por un punto interno que lo identifica. Los polígonos tienen atributos que describen al elemento geográfico que representan.

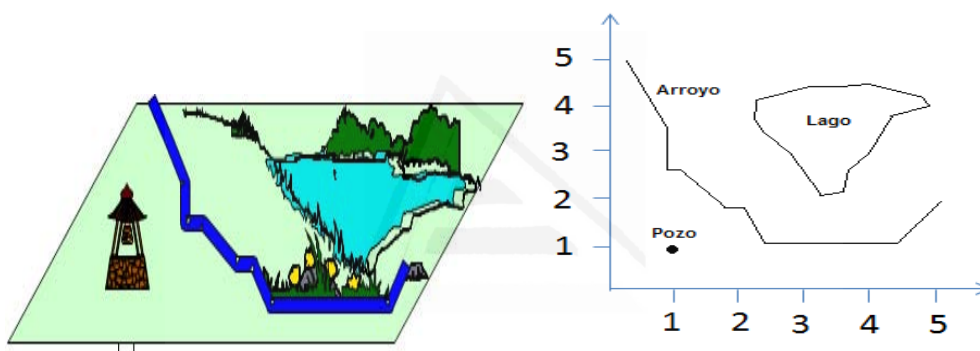


Figura 2.3. Representación de entidades geográficas mediante el modelo vectorial.

En la Figura 2.3 se muestra una representación de entidades geográficas del mundo real a través de los elementos básicos que componen la representación del modelos vectorial, es decir, puntos, líneas o arcos y polígonos.

En la literatura pueden encontrarse diferentes nombres o expresiones para los componentes geométricos, como se indica en la Figura 2.4. Las siguientes convenciones son usadas para evitar ambigüedades:

- Un punto en un grafo es llamado nodo.
- Un par de nodos conectados puede ser llamado arista o borde.
- Una cadena de aristas puede ser llamada arco o arcos.
- Un segmento de grafo es una cadena de arcos que no contiene nodos y más de dos arcos están conectados.

- Una combinación de arcos que forman un circuito cerrado puede ser llamado polígono.
- Un polígono encierra un segmento de área.
- Un segmento de área que no es intersectado por un polígono es llamado cara.

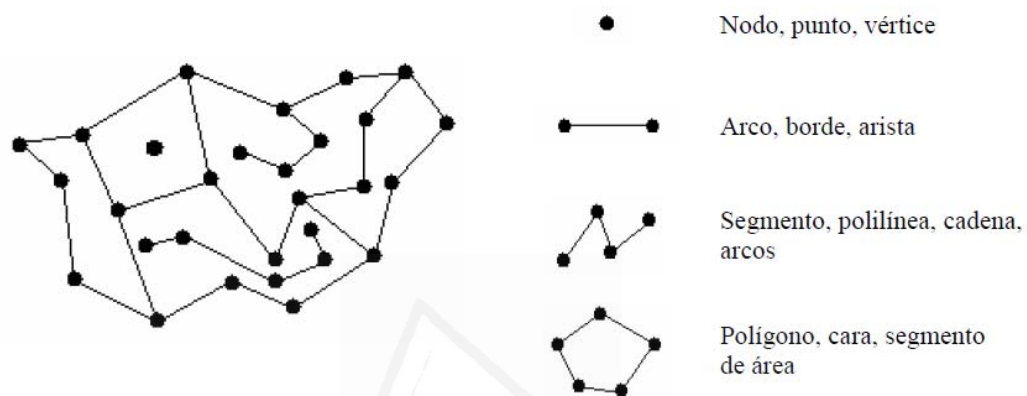


Figura 2.4. Componentes del modelo vectorial.

Un SIG puede reconocer y analizar las relaciones espaciales que existen en la información geográfica almacenada. Estas relaciones topológicas permiten realizar modelizaciones y análisis espaciales complejos. Así, por ejemplo, el SIG puede discernir la parcela o parcelas catastrales que son atravesadas por una línea de alta tensión, o bien saber qué agrupación de líneas forman una determinada carretera.

En el ámbito de los SIG se entiende como topología a las relaciones espaciales entre los diferentes elementos gráficos (topología de nodo/punto, topología de red/arco/línea, topología de polígono) y su posición en el mapa (proximidad, inclusión, conectividad, etc). Estas relaciones, que para el ser humano pueden ser obvias a simple vista, el software las debe establecer mediante un lenguaje y unas reglas de geometría matemática.

2.2.2 Relaciones espaciales.

Las relaciones espaciales establecen la ubicación en el espacio de algún objeto geográfico del mundo real con respecto a otro. Describen la configuración

geoespacial modeladas a través de regiones, líneas y puntos en un SIG se basan en conceptos matemáticos. Basado en sus diferentes propiedades matemática la literatura identifica tres tipos de relaciones espaciales: las relaciones topológicas, de distancia y de dirección (Egenhofer and Mark, 1995a), (Mark, 1999), (Papadias and Kavouras, 1994).

Relaciones topológicas. Las relaciones topológicas describen las posiciones en el espacio de los objetos geográficos, las cuales son invariantes bajo continuas transformaciones, tales como rotación, traslación, escalamiento, reflexión. Dos ejemplos de relaciones topológicas básicas son Coinciden y Contenido_en.

Egenhofer y sus colegas desarrollaron una categorización de las relaciones topológicas binarias denominado como el Modelo de las 9-Intersecciones. Este modelo describe formalmente las propiedades de las relaciones topológicas mediante una matriz de 3x3 (ver Figura 2.5), la cual indica si el interior, el exterior y los límites se interceptan o no. Se identificaron ocho relaciones diferentes entre dos regiones y diecinueve entre líneas y regiones en el grafo del vecindario conceptual (Egenhofer and Franzosa, 1991), (Egenhofer and Al-Taha, 1992), (Egenhofer and Mark, 1995b).

$$\begin{array}{c}
 \text{interior} \\
 \text{boundary} \\
 \text{exterior}
 \end{array}
 \begin{pmatrix}
 \text{interior} & \text{boundary} & \text{exterior} \\
 x_{1,1} & x_{1,2} & x_{1,3} \\
 x_{2,1} & x_{2,2} & x_{2,3} \\
 x_{3,1} & x_{3,2} & x_{3,3}
 \end{pmatrix}
 \quad x = \{\emptyset, -\emptyset\}$$

Figura 2.5. Una matriz 3x3 que indica la intersección del interior, los límites y el exterior describe formalmente las relaciones topológicas.

Las relaciones topológicas definidas por Egenhofer y sus colegas (Egenhofer and Herring, 1991), (Egenhofer, 1993), (Egenhofer and Franzosa, 1995) se definen de la siguiente forma:

Sea:

- A, B : Regiones pertenecientes a los objetos A y B respectivamente.

- $R(A, B)$: Define las relaciones existentes entre las regiones A y B dada por las siguientes expresiones:
 - A^0 : Referente al interior de la región del objeto A .
 - ∂A : Referente al contorno de la región del objeto A .
 - A^- : Referente al exterior de la región del objeto A .
 - B^0 : Referente al interior de la región del objeto B .
 - ∂B : Referente al contorno de la región del objeto B .
 - B^- : Referente al exterior de la región del objeto B .

Entonces es posible modelar las relaciones espaciales entre A y B a través de la ecuación 2.4, donde cada una de intersecciones puede tomar un valor de vacío (\emptyset) o no vacío ($\sim\emptyset$).

$$R(A, B) = \left\{ \begin{array}{lll} A^0 \cap B^0 = (\emptyset, \sim\emptyset) & A^0 \cap \partial B = (\emptyset, \sim\emptyset) & A^0 \cap B^- = (\emptyset, \sim\emptyset) \\ \partial A \cap B^0 = (\emptyset, \sim\emptyset) & \partial A \cap \partial B = (\emptyset, \sim\emptyset) & \partial A \cap B^- = (\emptyset, \sim\emptyset) \\ A^- \cap B^0 = (\emptyset, \sim\emptyset) & A^- \cap \partial B = (\emptyset, \sim\emptyset) & A^- \cap B^- = (\emptyset, \sim\emptyset) \end{array} \right\} \quad (2.4)$$

En la Figura 2.6 se representan las relaciones del tipo Separados, Contiene_a, Contenido_en, Iguales, Adyacentes, Cubre_a, Cubierto_por y Superposición modelas a través de la matriz de 3x3.

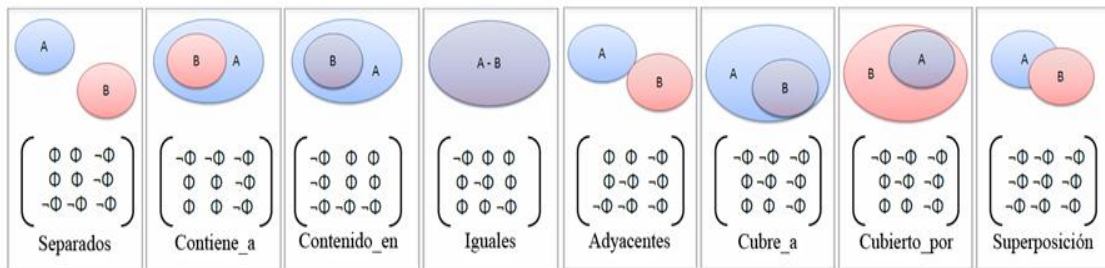


Figura 2.6. Relaciones topológicas entre los objetos A y B .

Las conexiones entre las relaciones dentro de un grafo del vecindario conceptual implican un cambio gradual de una configuración a otra, la Figura 2.7 visualiza las diferentes relaciones que se pueden establecer de acuerdo a los tipos de datos que se utilicen para representar los objetos geográficos.

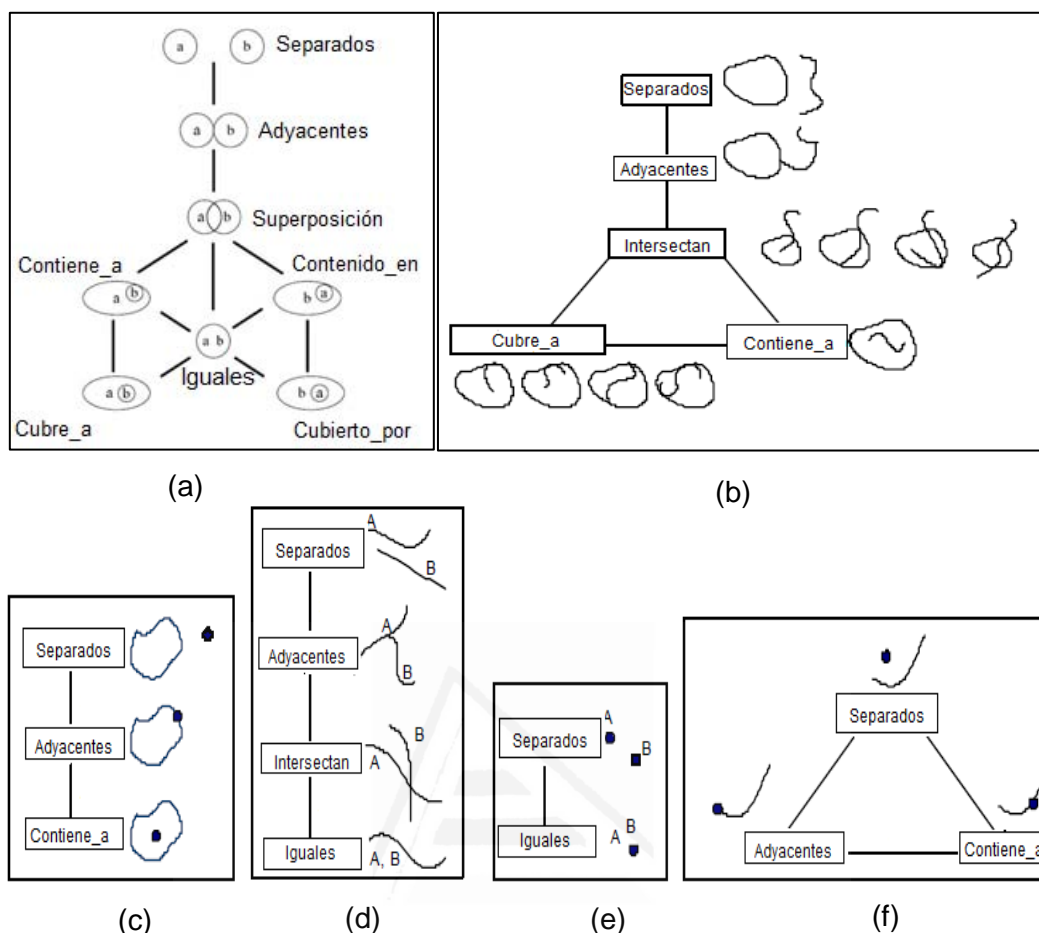


Figura 2.7. Grafo del vecindario conceptual del Modelo de las 9-Intersecciones.

En la Figura 2.7 (a) se muestran las relaciones entre polígonos, en (b) entre líneas y polígonos, en (c) entre polígonos y puntos, en (d) entre líneas, en (e) entre puntos y en (f) entre puntos y líneas.

Relaciones de distancia. Otras relaciones espaciales se centran en la distancia entre las características geográficas. Las relaciones de distancia son invariantes a las transformaciones de rotación, traslación, pero no bajo el escalamiento. La medida cuantitativa de las relaciones de distancia se calcula a partir de la distancia entre dos objetos geográficos basándose en un sistema métrico mientras que determinar cualitativamente las relaciones de distancia es muy subjetivo y dependiente de la escala. Hernández defiende las relaciones de distancia cualitativas para especificar los conceptos geoespaciales porque los conceptos espaciales cognitivos son cualitativos por naturaleza (Hernández, 1991). La precisión no está presente en los modelos cognitivos espaciales y las relaciones cuantitativas pueden provocar la producción de una precisión artificial que no existe.

Además, Hernández propone una representación relativa de las configuraciones mediante las relaciones comparativas.

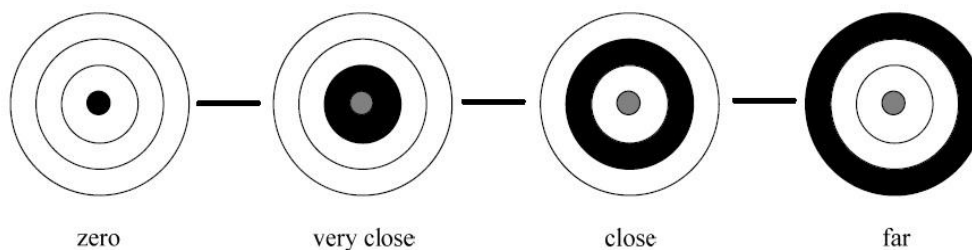


Figura 2.8. Grafo de vecindario conceptual para las relaciones de distancia cualitativas (Bruns and Egenhofer, 1996).

En el contexto de la medición de la similitud de las escenas espaciales Bruns y Egenhofer define las relaciones de distancia *iguales* (zero), *muy cerca* (very close), *cerca* (close) y *lejos* (far) (ver Figura 2.8).

Relaciones de dirección. Las relaciones de dirección describen la orientación de los objetos geográficos unos con respecto a otros en el espacio (Freksa, 1992). Estas relaciones son invariantes en la traslación y escalamiento pero no en la rotación.

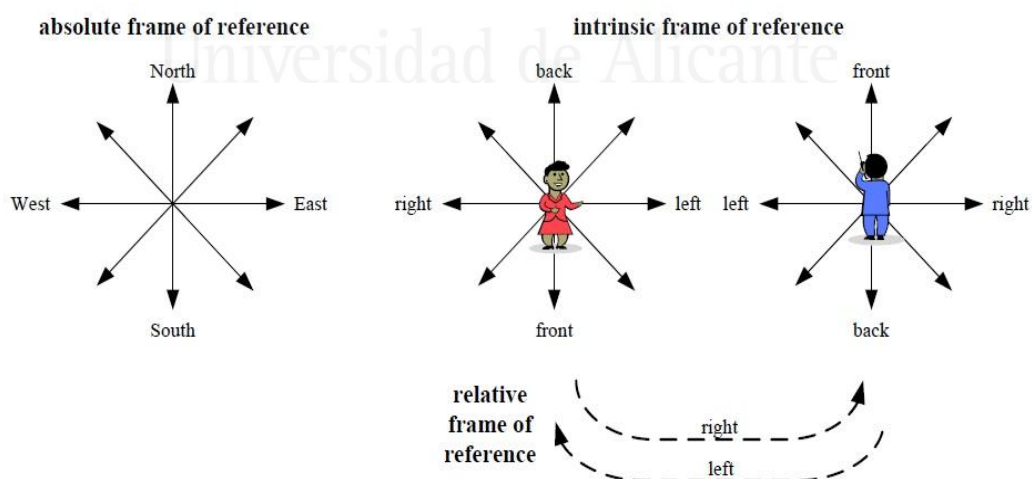


Figura 2.9. Marco de referencia absoluto y relativo.

Las direcciones dependen del punto de referencia, puede definirse un sistema de referencia absoluto o uno relativo (ver Figura 2.9). El absoluto tiene direcciones fijas

como las proporcionadas por la gravedad o las direcciones cardinales (Frank, 1996). El relativo está centrado en el objeto, en la Figura 2.9 las personas especifican las posiciones con respecto a su propio punto de vista. El sistema de referencia relativo está centrado en el observador, según la Figura 2.9 se observa que el hombre está a la derecha de la mujer y la mujer está a la izquierda del hombre.(Coventry and Garrod, 2004).

2.2.3 Comparación entre los RI y los SIG.

Según la literatura existen diferencias considerables entre los sistemas de información de propósito general y los sistemas de información geográfica. Estas diferencias se deben fundamentalmente a las características especiales de la información geográfica. Entre estas características especiales podemos destacar las siguientes:

- **Grandes volúmenes de datos y estructura intrínseca compleja.** Debido a esto adquiere especial importancia la definición de procedimientos eficientes de manipulación y transmisión de la información. Además, también determina una importancia destacada de los metadatos encargados de describir la información geográfica gestionada en los SIG.
- **Número ilimitado de tipos posibles de análisis por parte de los usuarios finales.** Por tanto, los sistemas deben proporcionar una funcionalidad primitiva lo más amplia posible y ofrecer la posibilidad de extensión por parte de los desarrolladores para la funcionalidad avanzada específica.
- **Necesidad de técnicas de análisis y transformaciones especiales.** Algunos ejemplos de estas técnicas son los algoritmos de búsqueda de rutas, análisis del terreno o la interpolación de datos geográficos. Estas técnicas requieren que los sistemas cuenten con representaciones especiales de los datos geográficos, con estructuras eficientes para el acceso a ellas, con algoritmos específicos de consulta y con interfaces de usuario adaptadas para permitir la interacción con ellas (por ejemplo, mediante herramientas de dibujo de objetos geográficos que puedan ser empleados como parámetros en consultas).

- **Necesidad de opciones de visualización avanzadas.** A diferencia de los sistemas de información tradicionales, donde se manejan fundamentalmente tipos de datos numéricos y textuales, la información geográfica es más compleja y necesita ser presentada al usuario con el nivel de detalle apropiado a diferentes escalas. Por ejemplo, el límite de una ciudad se puede representar como un polígono cuando se visualiza de cerca o como un punto cuando el nivel de detalle requerido es menor.
- **Estrecha relación con la componente tiempo.** Desde hace años se reconoce que hay una relación muy estrecha entre espacio y tiempo. Esta relación es muy importante para el desarrollo de SIG que soporten información geográfica dinámica.

A modo de resumen, la información geográfica tiene unas características especiales que determinan una serie de diferencias entre un SIG y un sistema de información tradicional. Estas diferencias se reflejan en todos los niveles del sistema desde el almacenamiento de datos, donde se necesitan nuevos tipos de datos y estructuras de acceso a ellos, hasta la presentación de información al usuario, donde se necesitan opciones de visualización mucho más avanzadas que en los sistemas tradicionales.

Los SIG presentan dificultades (Larín, 2013) de las cuales algunas están relacionados con los incapacidad de los usuarios de localizar sus necesidades y se mencionan a continuación:

- Hay una escasez crítica de expertos en los SIG que conozcan a fondo la tecnología, los métodos de análisis espacial, el diseño de aplicaciones, etc. Hay pocos lugares de capacitación e instructores, faltan investigaciones y financiación, e incluso falta la conciencia cabal del problema que representan estas carencias.
- Existen pocas soluciones para enfrentar la enorme afluencia de datos, que aumenta a un ritmo exponencial; hay que encontrar mejores métodos para almacenar, archivar, estructurar y procesar todos esos datos. ¿Cómo evaluar si los datos adquiridos son realmente necesarios?
- Muchos datos son reunidos por grupos que no tienen interés en aprovechar sus aspectos espaciales, por lo que no proporcionan los datos en un formato apropiado para el análisis espacial.

- Numerosos sistemas aún carecen de interfaces que faciliten su uso y permitan un mayor acceso por parte de usuarios no expertos en los sistemas. Estas personas pueden tener claro lo que quieren obtener del SIG, pero no tienen ganas o tiempo de leer extensos manuales de instrucción o aprender las complejidades del funcionamiento del ordenador. En muchos casos, esto puede haber generado en los usuarios una resistencia a la nueva tecnología.

2.3 Recuperación de información geográfica.

La Recuperación de Información Geográfica (RIG) es un tema interdisciplinario de rápido desarrollo académico y comercial; surge de la confluencia de los Sistemas de Información Geográfica (SIG) con la Recuperación de Información (RI). Esta incluye todas las áreas de investigación y componentes que tradicionalmente forman el núcleo de la RI, como son: los sistemas y motores de búsqueda, pero además con un énfasis en la información procedente de datos del dominio geográfico (Bucher et al., 2005). Se puede concluir entonces que, la RIG es la encargada de proporcionar accesos a fuentes de información geográfica y espacial, incluyendo tanto, procesos de ponderación, almacenamiento e indexado (Sanderson, et al., 2004), es decir, se preocupa por la recuperación de información en donde se procesan datos por medio de alguna noción de la relevancia geográfica. Esta necesidad no está ligada tan sólo a la web sino que también aparece en bibliotecas digitales, sistemas de gestión documental y en cualquier sistema de información en general que gestione información con un componente espacial.

Aunque una de las principales tendencias en RIG es heredada de las áreas de procesamiento de lenguaje natural, al definir estructuras de indexación y técnicas para almacenar y recuperar documentos de manera eficiente empleando tanto las referencias textuales, como las referencias geográficas contenidas en el texto; no obstante, existe un amplio espectro de enfoques para resolver las tareas en RIG, los cuales van desde aproximaciones simples de recuperación de información sin indexación de términos geográficos, a modelos derivados de técnicas complejas de inteligencia artificial. Algunas de las técnicas usadas en la actualidad incluyen extracción de entidades geográficas, análisis semántico, bases de conocimiento

geográfico (conjuntos de reglas, ontologías,...), adquisición de conocimiento, grafos y técnicas de expansión y flexibilización de consultas.

A pesar de la naturaleza geográfica de parte de la información, común entre los SIG y los sistemas RIG, existen dos diferencias fundamentales en los requisitos de ambos tipos de sistemas que se deben tener en cuenta. En primer lugar, **la componente espacial** de las consultas en sistemas RIG es mucho más sencilla que las consultas que se pueden realizar en los SIG y que involucran relaciones espaciales complejas. Por ejemplo, una consulta típica en un sistema RIG puede ser “suelos ferralíticos en San José” y, por tanto, la única relación espacial que hay que comprobar es si el ámbito espacial del documento se encuentra dentro de la geometría asociada con San José o no. Por otro lado, una consulta sencilla en SIG puede ser “suelos ferralíticos en regiones adyacentes a San José”. En este caso, habría que encontrar el lugar donde se encuentra San José y luego buscar que regiones adyacentes tienen suelos ferralíticos.

La segunda diferencia se encuentra en **las consultas típicas** de ambos sistemas. Mientras que en los SIG las consultas son parecidas a las de las bases de datos tradicionales, donde de manera categórica los objetos pertenecen al resultado o no, en los sistemas RIG son parecidas a las de los sistemas de RI, donde los objetos pueden pertenecer al resultado con una cierta probabilidad. Siguiendo con las consultas del ejemplo, en el caso de los sistemas RIG, suelos ferralíticos en Madrugá (un municipio próximo a San José aunque no adyacente) pueden ser relevantes para el usuario (aunque con menor importancia que los situados en las afueras de San José), mientras que en el caso de los SIG sólo suelos ferralíticos en regiones adyacentes a San José forman parte del resultado de la consulta.

De manera similar a lo que ocurre con la RI, donde muchas veces se restringe su alcance a la recuperación de documentos en la web, una confusión habitual a la hora de definir el alcance de la recuperación de información geográfica es restringirlo a lo que hacen los motores de búsqueda local como Google Maps⁶, Yahoo Mapas⁷, Microsoft Live Search⁸. Aunque podemos considerar estos sistemas dentro del ámbito de la recuperación de información geográfica, ya que cumplen la definición al recuperar información relevante tanto temática como geográficamente,

⁶ <http://maps.google.es/>

⁷ <http://espanol.maps.yahoo.com/>

⁸ <http://maps.live.com/>

están posiblemente más próximos al campo de los SIG. Su funcionalidad principal se basa en el almacenamiento estructurado (por ejemplo, en tablas de bases de datos) de negocios, organizaciones, puntos de interés y elementos pertenecientes a otras categorías prefijadas. Para cada uno de los elementos se almacena también su situación geográfica y esto permite que los usuarios puedan realizar búsquedas que las tengan en cuenta. Para completar los resultados, pueden buscar la parte textual de las consultas en sus motores de búsqueda en web tradicionales.

Las relaciones espaciales entre los objetos geoespaciales son frecuentemente tan importantes como las características geográficas (Papadias and Kavouras, 1994). Son muy importantes en la descripción semántica de los objetos geográficos debido a que todos los objetos geográficos tienen una localización en la tierra por lo que ellos automáticamente tienen una relación espacial con respecto a todos los otros objetos geográficos. Establecer relaciones espaciales entre clases no siempre es posible debido al carácter abstracto de las mismas pero existen ciertas características que pueden ser generalizadas a todas las instancias.

2.3.1 Las relaciones espaciales en la RIG.

Como se explica en la sección 2.2.2 todos los objetos geoespaciales poseen una localización en el espacio, por lo tanto, todo par de objetos geoespaciales están relacionados espacialmente.

Las relaciones espaciales formalizadas implican semánticas con fundamentos matemáticos no ambiguos. Las relaciones espaciales en lenguaje natural son complicadas y a menudo más complejas que las relaciones formales. A continuación se presentan diferentes tipos de relaciones espaciales formalizadas y de manera resumida como estructurar las relaciones en lenguaje natural. Finalmente se explica un modelo computacional propuesto por Shariff, Egenhofer y Mark (Shariff, et al., 1998) para representar las relaciones espaciales en lenguaje natural a través de las relaciones espaciales formalizadas.

El lenguaje natural y la percepción juegan un papel fundamental en la forma en la que las personas estructuran el espacio (Talmy, 1983), (Talmy, 1996). Cuando una persona describe una escena espacial no describe cada detalle sino que generaliza. Generalmente resalta algunas propiedades espaciales y otras las

ignora. El lenguaje natural ayuda a describir una escena solamente por su estructura y las propiedades espaciales que se consideran importantes en el contexto en particular (Tversky and P. U. Lee, 1998).

Las relaciones espaciales en lenguaje natural tienen semánticas complejas y generalmente incluyen más de una relación espacial. Las relaciones entre dos características geográficas a menudo son más complejas de describir a través de relaciones topológicas, de distancia o de dirección que a través de una combinación de ellos. Generalmente se utilizan las preposiciones para describir las relaciones espaciales. Las personas no pueden determinar la distancia exacta entre la presa y la carretera que pasa cerca de ella pero establecen una relación espacial utilizando una representación como, “la presa está cerca de la carretera” o “la distancia entre la presa y la carretera es más o menos 70 metros.”

Las relaciones espaciales se pueden describir matemáticamente o a nivel conceptual, estas dos representaciones son diferentes (Schwering, 2006): las relaciones espaciales matemáticamente se calculan utilizando unidades de medidas. Estas representaciones no siempre coinciden con la forma en la que las personas las entienden individualmente. Las relaciones espaciales formalizadas se utilizan en los SIG para especificar las relaciones del mundo real. Las relaciones a nivel conceptual generalmente son inexactas y se describen en la mayoría de los casos en lenguaje natural.

Las relaciones espaciales en lenguaje natural están influenciadas por la perspectiva del observador (Egenhofer and Mark, 1995a), (Golledge et al., 1969), (Kosslyn et al., 1978). La relación que existe realmente no siempre es igual a la relación conceptual percibida por el observador.

(Donnelly and Bittner, 2005) desarrollaron una teoría formal para definir las relaciones espaciales y proponen axiomas para el razonamiento espacial de ellas. Para el proceso de generalización de las relaciones espaciales entre las instancias, Donnelly y Bittner visualizan el universo como un conjunto de individuos y definen cinco tipos de relaciones espaciales entre ellos que se muestran en la Tabla 2.1.

Tabla 2.1. La semántica de las relaciones espaciales generalizadas se define a través de una definición formal de relaciones espaciales entre instancias de una misma clase.

Tipo de relación	Definición formal entre relaciones espaciales
$R_{some}(A,B)$	$\exists x \exists y (Instance(x,A) \& Instance(y,B) \& R_{x,y})$
$R_{all-1}(A,B)$	$\forall x (Instance(x,A) \rightarrow \exists y (Instance(y,B) \& R_{xy}))$
$R_{all-2}(A,B)$	$\forall y (Instance(y,B) \rightarrow \exists x (Instance(x,A) \& R_{xy}))$
$R_{all-12}(A,B)$	$R_{all-1}(A,B) \& R_{all-2}(A,B)$
$R_{all-all}(A,B)$	$\forall x \forall y (Instance(x,A) \& Instance(y,B) \rightarrow R_{xy})$

La relación R_{some} es la más simple porque se cumple con que al menos una de las instancias de cada clase esté relacionada. Las relaciones R_{all-1} y R_{all-2} son las más utilizadas en las relaciones espaciales a nivel conceptual: todas las instancias de un universo están relacionadas con al menos una instancia del otro universo. La relación $cercaDe(río, mar)$ significa que todos los ríos están cerca de un mar pero no todos los mares están cerca de un río. La relación R_{all-12} requiere que se cumplan las relaciones R_{all-1} y R_{all-2} al mismo tiempo. Por ejemplo, todos los suelos tienen un tipo de horizonte y todos los horizontes están asociados a los tipos de suelos. $R_{all-all}$ requiere que todas las instancias de cada universo tengan relaciones espaciales entre ellas; esta relación no es común.

La teoría formal de las relaciones espaciales puede ser aplicada directamente a las relaciones espaciales a nivel conceptual, sustituyendo los individuos por los objetos para definir las relaciones. Las relaciones espaciales a nivel conceptual son una generalización de las relaciones entre las instancias de un concepto.

2.4 Recuperación semántica de información geográfica.

Los métodos semánticos proponen volverse un nuevo paradigma en la recuperación de información geográfica. Según (Kuhn, 2005) muchos lenguajes relevantes en el contexto de los sistemas de información (lenguajes de programación, lenguajes de consulta, lenguajes de modelado de flujo de trabajo, etc.) manifiestan como los seres humanos conceptualizan las cosas con el

propósito de representarlas y manipularlas en las máquinas. Sin embargo, añadir significado a las expresiones del lenguaje es un fenómeno conceptual, ya que el lenguaje natural evoca símbolos y expresiones en nuestra mente que son usadas para dar significado a esos conceptos. Lo mismo ocurre en la RIG. La semántica geoespacial puede comprender y convertir el contenido del SIG y capturar este conocimiento en teorías formales. Hoy en día contamos con varias estructuras para representar el conocimiento, siendo las ontologías las más eficientes para abarcar la mayor cantidad posible de soluciones a los problemas que conciernen al dominio geográfico.

Existen varios trabajos que comparan y evalúan las propuestas de integración de datos geográficos y semántica. En (Visser, 2004) se introduce el sistema BUSTER (Bremen University Semantic Translator for Enhanced Retrieval) que propone una arquitectura formada por dos subsistemas: el primero tiene como objetivo resolver los problemas de recuperación de información mediante búsquedas inteligentes y el segundo con el objetivo de resolver la heterogeneidad semántica propone una forma de integración de la información.

En (Kavouras et al., 2003) se presenta una metodología para explorar e identificar información semántica a partir de una categorización provista por geo-ontologías, esta metodología propone medidas sobre qué ontologías pueden ser integradas o asociadas mediante categorías.

En (Sotnykova et al., 2005) se propone una metodología para la integración de esquemas conceptuales espacio-temporales mediante modelos conceptuales y Lógica Descriptiva (LD). En (Hakimpour, 2003) se introduce una arquitectura y una metodología basada en LD para crear un sistema integrado de información geográfica. Las ontologías fueron propuestas en (Hess and Lochpe, 2004) como mediadoras en la integración semántica de datos geográficos.

En el sistema GioNis (Stoimenov et al., 2006) se define una propuesta de ontología híbrida basada en una arquitectura semántica que combinada con LD permite descubrir correspondencias entre conceptos de diferentes ontologías.

En su propuesta (Li and Fonseca, 2006) proponen una medida de similitud que integra cuatro modelos, el modelo geométrico, el modelo de características, el modelo de transformación y el modelo de alineación estructurada para calcular las

igualdades y las diferencias entre escenas espaciales y a nivel de capa. Aplica el orden de prioridad topología, dirección, distancia y se disminuyen los costos de transformación. Ambas características son implementadas a través de la aplicación de los pesos.

Un enfoque de integración de las relaciones espaciales y mediciones de similitud semántica entre diferentes conceptos geoespaciales considerando que las relaciones espaciales son partes fundamentales de la descripción semántica de los geo-datos se presentan en (Schwering, 2006). Se seleccionan un conjunto de relaciones espaciales formalizadas en lenguaje natural según el modelo computacional de (Shariff, et al., 1998). El trabajo está enfocado en la medición de las distancias semánticas en el nivel conceptual.

Muchos de los trabajos que se llevan a cabo dentro del este campo de investigación utilizan las ontologías como estructuras de representación del conocimiento, explotando descripciones semánticas explícitas de los contenidos para facilitar la localización, compartición e integración de la información. El concepto de ontología ha atraído una atención creciente en la comunidad de las ciencias de la información debido a su capacidad para lograr una representación del conocimiento compartido.

2.4.1 Uso de las ontologías en el dominio espacial.

En el mundo de la informática y las comunicaciones una ontología es una representación conceptual de un dominio determinado que facilita el intercambio de información entre diferentes sistemas. Una ontología proporciona los significados que describen explícitamente la conceptualización del conocimiento representado en una base de conocimientos que las máquinas puedan entender, es por esto que el vocabulario debe ser definido con gran precisión permitiendo diferenciar términos y referenciarlos de manera precisa. Las ontologías no solo incluyen definiciones de conceptos básicos del dominio sino también las relaciones entre ellos. Las ontologías proporcionan la vía para representar conocimiento.

Existen varias definiciones de las ontologías dentro del campo de la inteligencia artificial, se puede decir que precisamente las definiciones dadas en esta área de las ciencias técnicas fue lo que comenzó a popularizar su uso dentro de los sistemas informáticos. La definición más consolidada de ontología es la propuesta

por Gruber (1993) y extendida por Studer et. al (1998) que la describe como “una especificación explícita y formal sobre una conceptualización consensuada”. La interpretación de esta definición es que las ontologías definen conceptos, propiedades, relaciones, funciones, restricciones y axiomas de forma explícita, compartida y consensuada; y que esta conceptualización debe ser representada de una manera formal, legible y utilizable por los ordenadores.

Guarino dice que “una ontología es un conjunto de axiomas lógicos diseñados para alcanzar el significado intencional de un vocabulario, o sea, un comportamiento ontológico hacia una conceptualización específica del mundo. Una ontología refleja indirectamente su compromiso y su conceptualización subyacente” (Guarino, 1998). Smith, más enfocado en sistemas de información, dice que “una ontología puede verse como un diccionario de términos o un vocabulario común compartido por diferentes comunidades de sistemas de información” (Smith, 1999).

Fonseca las define informalmente como “acuerdos sobre conceptualizaciones compartidas” (Fonseca et al., 2002b). Wiederhold dice que los acuerdos son representados como ontologías, una por área de estudio (Wiederhold, 1999). Las ontologías contienen objetos, propiedades de los objetos y posibles relaciones entre los objetos de un dominio específico del conocimiento (Chandrasekaran et al., 1999).

A diferencia de los glosarios y tesauros, cuentan con restricciones lógicas que especifican qué son los términos, como están relacionados con otros y si pueden o no estar relacionados. En una ontología también se especifican reglas para combinar términos y sus relaciones extendiendo así el vocabulario. En la ontología los significados de los términos no son ambiguos, lo que los hace semánticamente independientes del usuario y del contexto. Se puede traducir los términos de una ontología de un lenguaje a otro y conceptualmente no hay cambios. Las ontologías proveen un vocabulario y un entendimiento común procesable por el ordenador del área o dominio que describe (Larín, 2013).

Por otra parte, las ontologías cuentan con una taxonomía de forma tal que es procesable y entendible por el ordenador. Una taxonomía es una clasificación o categorización jerárquica de entidades de un dominio concreto (Gómez-Pérez, 1999). No obstante, una ontología es más que una taxonomía ya que es una

especificación más completa de un dominio específico. Una característica taxonómica fundamental de las ontologías son las relaciones subclases-de (*subclass-of*) las cuales son formalmente especificadas.

De igual modo, las ontologías identifican clases de objetos, sus relaciones y las jerarquías de conceptos dentro de un dominio específico, pero además se especifican de forma elaborada a través de lenguajes de representación ontológica.

El propósito fundamental de una ontología no es servir de vocabulario o taxonomía sino el compartimiento y reusabilidad del conocimiento entre aplicaciones. Cada ontología provee una descripción de los conceptos y sus relaciones dentro de un dominio determinado, lo cual puede ser compartido y reutilizado entre diferentes agentes inteligentes y aplicaciones. Las ontologías son muy usadas dentro del campo de la inteligencia artificial y la representación del conocimiento con diferentes propósitos como el razonamiento inductivo, la clasificación y diferentes técnicas de resolución de problemas.

Para que la especificación de un dominio sea considerada una ontología debe presentar dos tipos de componentes: los elementos y las relaciones entre estos (Pedraza-Jiménez et al., 2007).

Varios autores (Gruber, 1995), (Uschold and Gruninger, 1996), (Studer et al., 1998), (Chandrasekaran et al., 1999), (Arpírez et al., 2001) coinciden que los principales componentes de una ontología son los siguientes:

- **Conceptos:** Son las ideas básicas reflejadas en términos que se intentan formalizar. Los conceptos pueden ser clases de objetos, métodos, planes, estrategias, procesos de razonamientos, etc. Las entidades del “mundo real” se pueden categorizar en grupos o conjuntos de objetos con similares características. Las clases constituyen el núcleo de una ontología y describen los conceptos de un dominio.
- **Individuos, instancias o ejemplares:** Son representaciones de los conceptos o elementos particulares de una clase. Ejemplos de instancias de la clase “Células” son “Cistolito” y “Esclereidas”. Hay que señalar que a veces resulta difícil distinguir entre individuos y clases (Torres Ruiz, 2007).
- **Relaciones:** Éstas representan la interacción y enlace entre los conceptos de un dominio. Suelen formar la taxonomía del dominio. Por ejemplo,

subclase-de, *parte-de*, *conectado-a*. etc. Las relaciones también son utilizadas para describir las propiedades de los objetos. Las entidades que pertenecen a una clase poseen atributos determinados; por ejemplo, tienen un nombre, un color o un peso. Por tanto, las propiedades consisten en pares de atributo/valor que sirven para describir las características relevantes de las entidades que forman las clases. Las relaciones entre los elementos de una ontología son las siguientes:

- **Clase – Individuo:** permiten asociar individuos o instancias a una clase. Por ejemplo, “Esclereida” es una instancia de la clase “Célula”. Estas asociaciones se expresan mediante la relación “es un” (“is-a”).
- **Clase – Propiedad:** las clases tienen un conjunto de propiedades, que son denominadas restricciones pues sirven tanto para definir la clase como para delimitar la pertenencia de los individuos a la clase. Por ejemplo, la clase “Célula” posee la propiedad “tener un núcleo”.
- **Clase – Subclase:** las clases pueden tener subclases, Por ejemplo, “aves” y “reptiles” son subclases de “Especies”. Esta relación se expresa también como “es un”.
- **Funciones:** Son un tipo concreto de relación donde se identifica un elemento mediante el cálculo de una función que considera varios elementos de la ontología. Por ejemplo, pueden aparecer funciones como: *asignar-fecha*, *categorización-clase*, etc.
- **Reglas de restricción o axiomas:** Son teoremas que se declaran sobre relaciones que deben cumplir los elementos de la ontología. Por ejemplo: “Si *A* y *B* son de la clase *C*, entonces *A* no es subclase de *B*”, Los axiomas junto con la herencia de conceptos, permiten *inferir conocimiento* que no esté indicado explícitamente en la taxonomía de conceptos.

Además de las anteriores, en una ontología también se dan otras clases de relaciones siguiendo otros enfoques. En concreto, se suelen contemplar las relaciones entre conceptos (clases) de sinonimia, antonimia, hiponimia y meronimia. Algunas de estas relaciones son similares a las que se contemplan en los tesauros. También debe tenerse en cuenta que las relaciones clase-subclase y clase-individuo son la base de taxonomías y tesauros, y de aquí la tendencia a confundir los tres sistemas. Sin embargo, las ontologías están formalizadas, es decir, están expresadas mediante una rigurosa lógica formal, que permite que sean procesadas

por aplicaciones informáticas y además contemplan un conjunto más amplio de relaciones (Guarino, 1998).

La Web Semántica es un área pujante en la confluencia de la Inteligencia artificial y las tecnologías web, que propone la introducción de descripciones semánticas explícitas de los contenidos en la web para facilitar su localización, compartición e integración (Berners-Lee et al., 2001). El concepto de ontología ha atraído una atención creciente en la comunidad de las ciencias de la información debido a su capacidad para lograr una representación del conocimiento compartido. El uso de ontologías en la información geográfica tiende a ser diferente dependiendo de la perspectiva y objetivos de los usuarios (Winter, 2000).

Las principales aplicaciones de las ontologías en la ciencia de la información geográfica según (Agarwal, 2005) son:

1. ontología para la generación de conocimiento,
2. ontología para la especificación de dominio, y
3. ontología para el desarrollo de sistemas de información.

Las ontologías se presentan precisamente como una alternativa potente para resolver muchos de estos problemas de heterogeneidad de las fuentes de información, permitiendo además desambiguar los términos de búsquedas porque en ellas el conocimiento se especifica a través de conceptos. Estos a su vez, están vinculados y regidos mediante relaciones y axiomas que enriquecen el contexto que conceptualizan. En (Vilches et al., 2006) podemos ver otras ventajas de las ontologías. El uso de ontologías permite centrarse en el significado de los datos en lugar de su sintaxis o estructura propiciando así una buena interoperabilidad entre los diferentes sistemas. Además, permiten la integración de datos geográficos de una manera mucho más abstracta en la que el conocimiento juega un rol fundamental.

La recuperación por conceptos permite además de que los sistemas puedan identificar mejor los datos, procesarlos en consecuencia y realizar análisis más eficientes con independencia del especialista, lo que constituiría una gran ventaja porque la información no solo será comprendida por las personas sino también por las máquinas incidiendo favorablemente sobre la reutilización del conocimiento, haciendo al sistema más potente y eficaz. También la posibilidad de interactuar

semánticamente con otras ontologías constituye un aspecto de avance, ya que posibilita obtener información desde otras fuentes. Esto significaría que pudiera obtenerse datos desde otros sistemas incluso independientemente del lenguaje que tengan, dado que las consultas serían interpretables en un ámbito donde la información contenida en una ontología estaría en correspondencia con otra representación tal vez de la misma información, lo que es muy útil en el campo de la información geográfica.

En (Rodríguez et al., 2006) se señalan tres ventajas muy significativas que suponen preeminencias sobre las otras formas de organización de la componte semántica.

1. Reducen la confusión semántica. Disminuye la ambigüedad terminológica al tener en cuenta sinónimos y polisemias, repercutiendo sobre la comunicación. Un sencillo ejemplo de la reducción de términos usando sinónimos con el uso de tesauros lo constituye la consulta: Obtenga los *hospitales próximos al municipio San José*. Aquí próximo nos indica cercano, inmediato, contiguo. Con cualquiera de estas palabras que se hubiese formulado la pregunta el sistema debiera ser capaz de identificar que es lo mismo y responder igual.
2. Permite la posibilidad de reutilizar el conocimiento. Esta ventaja es la referida a la interacción entre distintas ontologías del dominio geográfico, ya que resultado del propio desarrollo de ontologías hay diversas formas de conceptualizar una misma realidad.
3. La traducción e intersección semántica a través de comparaciones (cotejos) utilizados para describir correspondencias entre fenómenos (ej.: río, river, rivière y fleuve) y entre diferentes ontologías (ej.: ontología de fenómenos hidrográficos y ontología de las ciudades).

Con la utilización de ontologías se dota a los sistemas de la capacidad de interactuar por sí solos en consecuencia de la información que procesen convirtiéndolos en herramientas capaces de buscar e inferir conocimiento. Las ontologías tienen una gran importancia en la creación y uso de las normas de intercambio de datos, así como en la solución de problemas derivados de la heterogeneidad y poca interoperabilidad de los datos geográficos. También pueden ser usadas como una alternativa para representar los datos y, de forma explícita, el conocimiento acerca de ellos (Ehrig, 2007). Para el procesamiento semántico del

conocimiento de los datos geoespaciales almacenados en las ontologías es fundamental el cálculo de la similitud semántica, la cual es esencial en el procesamiento de las consultas de datos de los usuarios y es la base para la recuperación e integración de información semántica (Schwering, 2008).

2.4.2 Similitud semántica entre datos geoespaciales.

La similitud constituye un elemento básico para la inferencia analógica. (Markman, 1997). Las personas utilizan la similitud para almacenar y recuperar información comparando nuevas situaciones con experiencias similares ocurridas en el pasado también para el aprendizaje y la formación de conceptos (Schwering, 2008). Mientras que la comparación de igualdad entre dos objetos puede ser calculada por las computadoras de forma rápida y precisa la similitud es un problema complejo de calcular. Aun siendo difícil de determinar, la similitud juega un papel fundamental en muchas aplicaciones como sistemas de recuperación o de toma de decisiones. Lo mismo ocurre con la similitud espacial en los procesos de recuperación e integración de información espacial (Rodríguez and Egenhofer, 2003).

Debido a las raíces interdisciplinarias que tiene la RIG, díganse los SIG y los sistemas de RI, esta hereda los componentes que tradicionalmente forman parte de esos sistemas. Dentro de estos componentes podemos citar los modelos para calcular la similitud. Ahora bien, si dentro de estos modelos incorporamos métodos semánticos pues ya entonces se puede hacer referencia a la Recuperación Semántica de Información Geográfica que se aborda en esta sección.

En este trabajo se han organizado las medidas de similitud en tres grupos (ver Figura 2.10), la similitud semántica las cuales calculan la similitud desde el punto de vista de los significados de los términos y son heredadas fundamentalmente de los sistemas de RI, la similitud espacial que calculan la similitud teniendo en cuenta las relaciones espaciales entre objetos geográficos utilizadas mayormente en los SIG, y por último, la similitud semántica espacial, en las cuales se calcula la similitud teniendo en cuenta las representaciones conceptuales tanto de los objetos como de las relaciones espaciales que existen entre ellos y constituyen la base de la recuperación semántica dentro del campo de las Geociencias.

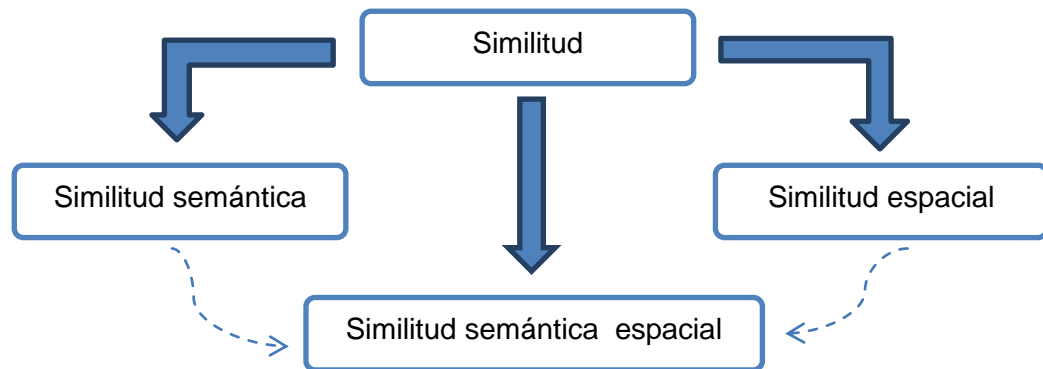


Figura 2.10. Diferentes enfoques de medidas de similitud.

2.4.2.1 *Similitud semántica.*

La similitud entre conceptos busca cuantificar el grado de cercanía al utilizar las relaciones semánticas (ej.: WordNet). En la literatura se encuentran diferentes modelos que permiten calcular la similitud y utilizan diferentes representaciones para entender la semántica de la información que necesitan manipular. Por lo general, toman una par de palabras de entrada y devuelven un valor que indica el grado de similitud que existe entre ambos conceptos. A continuación se ejemplifican algunos de estos modelos.

- La distancia de Levenshtein o distancia de edición (*edit distance*) (Levenshtein, 1966), fue creada en 1965 por el científico ruso Vladimir Levenshtein. La idea consiste en determinar el número mínimo de operaciones requeridas para transformar una cadena de caracteres en otra, estas operaciones son: inserción, eliminación o sustitución de un carácter. Por ejemplo, la distancia de Levenshtein entre los términos “hotel” y “hostal” es de dos porque se necesitan al menos dos ediciones elementales para cambiar un término en el otro término.
 - 1) hotel (?) hostal
 - 2) hostel (?) hostal (inserción de ‘s’ antes de t)
 - 3) hostel (?) hostal (sustitución de ‘e’ por ‘a’)
 - 4) hostal (?) hostal

- Uno de los algoritmos destacados en el cálculo de la similitud semántica es el propuesto en (Resnik, 1993), el cual propone que la similitud entre dos conceptos C_1 y C_2 de una estructura taxonómica, puede ser obtenida mediante la ecuación 2.5:

$$\text{sim}(C_1, C_2) = \max_{C \in S(C_1, C_2)} (-\log p(C)) \quad (2.5)$$

Donde:

- $S(C_1, C_2)$ representa el conjunto de conceptos de los cuales tanto C_1 como C_2 descienden.
- $p(C)$ es la probabilidad de encontrar una instancia de tipo C .

El término concepto hace referencia al conjunto de términos que apuntan a una misma idea. Ahora bien, para estimar la probabilidad de un concepto C , Resnik utiliza la frecuencia de aparición de los términos de ese concepto en el Brown Corpus of American. Posteriormente propone variantes en el cálculo de la probabilidad de C , dependiendo el método totalmente de la taxonomía (Resnik, 1995), (Resnik, 1999).

En (Resnik, 1995) se introduce el término de **contenido de información**, que es una medida de la especificación de un concepto determinado, y está definida, según la ecuación 2.6, en base a su probabilidad de ocurrencia en un corpus extenso.

$$IC(C) = -\log(p(C)) \quad (2.6)$$

Dado un corpus, $p(C)$ es la probabilidad de encontrar una instancia de tipo C . El valor para $p(C)$ es mayor en conceptos listados en la parte superior de la jerarquía y llega a su máximo valor para el concepto que se encuentra en la cima (si la jerarquía tiene una única cima, entonces el valor para este concepto es uno). Resnik define una medida de similitud semántica entre dos palabras al utilizar el Lowest Common Subsumer (*LCS*). El *LCS* es el primer concepto de la red semántica que contiene a los dos conceptos, es decir, el primer nodo común para el que existe un camino desde el concepto C_1 y el concepto C_2 . En la ecuación 2.7 se muestra esta medida.

$$\text{sim}(C_1, C_2) = IC(LCS(C_1, C_2)) \quad (2.7)$$

- El uso de la probabilidad también es el fundamento de otros trabajos, por ejemplo el trabajo propuesto por Lin (Lin, 1998), en el cual se obtiene de forma teórica la ecuación 2.8 para el cálculo de la similitud entre dos conceptos C_1 y C_2 en una taxonomía como la de WordNet.

$$\text{sim}(C_1, C_2) = \frac{2 * \log p(C_3)}{\log p(C_1) + \log p(C_2)} \quad (2.8)$$

Donde C_3 es el nodo padre común a C_1 y C_2 directo, o más próximo. Bajo este enfoque, Lin pretende dar una definición universal de similitud en términos de la teoría de la información. Para llegar a dicha definición Lin parte de tres premisas intuitivas:

- 1) La similitud entre A y B se relaciona con su entorno. Entre más características en común compartan, más similares son.
 - 2) La similitud entre A y B está relacionada con las diferencias que existen entre ellos. Entre más diferencias hayan, menos similares son.
 - 3) La máxima similitud entre A y B se alcanza cuando A y B son idénticos.
- Una propuesta más para medir la similitud es presentada por Leacock y Chodorow (Leacock and Chodorow, 1998), estos dos autores consideran que la similitud entre C_1 y C_2 puede ser obtenida con la ecuación 2.9.

$$\text{sim}(C_1, C_2) = \log\left(\frac{\text{len}(C_1, C_2)}{2 * MAX}\right) \quad (2.9)$$

Donde:

- $\text{len}(C_1, C_2)$ cuantifica el número de saltos entre C_1 y C_2 , y
 - MAX es el número de saltos entre el nodo raíz y los nodos hoja de la taxonomía.
- De igual modo, otro destacado trabajo de investigación usando WordNet es el de Hirts y St-Onge (Hirts and St-Onge, 1998). En este trabajo se emplean cadenas léxicas como representación de contexto para la detección y

correlación de nodos entre conceptos. Se alejan un poco del enfoque probabilístico y cuantitativo (conteo de nodos entre conceptos), sustituyéndolo por la idea de cadenas léxicas para poder obtener la similitud.

Después de analizar destacados trabajos que han abordado parte del problema relacionado con el presente trabajo de investigación, se observa la importancia de WordNet en la definición de las medidas de similitud, ya que gran número de las propuestas que existen en la actualidad van dirigidas al uso de estructuras taxonómicas, pero debido a que en la presente investigación se manejan una serie de términos o palabras muy técnicas se decidió no utilizar WordNet debido a que este no contaba en gran medida con los términos técnicos que se manejan en la ontología diseñada en el capítulo 4.

2.4.2.1.1 Modelo de correspondencia basada en Grafos.

La idea principal de este trabajo (Graph Matching Ontology-GMO) (Jian et al. 2005) es medir la similitud entre dos entidades (de dos ontologías), considerando la acumulación de similitudes de sentencias involucradas (tripletas). Las entidades a comparar se consideran dentro de la misma estructura: sujeto, predicado y objeto. Por otro lado, GMO también toma en cuenta la acumulación de las similitudes de entidades vecinas respecto a las dos entidades que están siendo comparadas.

Como entrada GMO, acepta un conjunto de pares de entidades relacionadas, que son encontradas previamente por otros algoritmos de similitud (*matching*). A la salida GMO proporciona parejas de entidades adicionales al comparar la similitud estructural. Las similitudes que considera GMO son principalmente la similitud entre propiedades y la similitud entre superclases

Similitud entre propiedades.

Esta medida tiene el propósito de comparar la distancia entre el conjunto de propiedades de dos clases, considerando la simetría de la similitud; es decir que la *Similitud_propiedades*("hotel", "botel") sea la misma que *Similitud_propiedades*("botel", "hotel")

El objetivo es identificar si cada propiedad del conjunto de propiedades de una clase, coincide con otra propiedad del conjunto de propiedades de otra clase Para

realizar dicha comparación entre cada propiedad, se utilizan las etiquetas de ambas propiedades (o nombres) como entrada a una medida de similitud léxica, en este caso se ocupa la distancia de Levenshtein.

Similitud entre superclases.

Esta medida se enfoca en la similitud entre superclases (o similitud entre conceptos ascendentes) a partir de dos clases iniciales. La similitud entre algún par de superclases puede influir para establecer una correspondencia entre dos clases. Para calcular esta medida de similitud es necesario calcular previamente la similitud léxica y de propiedades de todas las parejas de clases.

A continuación se define la similitud entre superclases a través de la ecuación 2.10, dado C_1, C_2 una pareja de subclases:

$$Similitud_{superclases}(C_1, C_2) = Similitud_{parcial}(SuperClases(C_1), SuperClases(C_2)) \quad (2.10)$$

La ecuación 2.11 expresa la similitud parcial: dado una pareja de superclases obtenida desde C_1, C_2 :

$$Similitud_{parcial}(C_{a1}, C_{a2}) = \frac{Similitud_{lexica}(C_{a1}, C_{a2}) + Similitud_{propiedades}(C_{a1}, C_{a2})}{2} \quad (2.11)$$

La idea intuitiva de la similitud parcial es obtener un valor que integre preliminarmente la similitud léxica y entre propiedades de forma equitativa. El objetivo es comparar parejas de superclases y seleccionar la pareja de superclases con mayor similitud entre sí, a fin de encontrar de manera aproximada la pareja de superclases en donde coincidan o converjan las clases C_1, C_2 .

2.4.2.2 Similitud espacial

En la Ciencia de la Información Geográfica o Geociencia (GIScience), la similitud juega un papel muy importante; es necesaria para resolver consultas vagas, consultas en lenguaje natural y es la base para la recuperación e integración semántica (Schwering, 2008).

Los Sistemas de Información Geográfica dependen de la similitud espacial entre escenas espaciales para recuperar información, proporcionar interconexión entre diferentes fuentes de información y clasificar objetos o fenómenos espaciales. A diferencia de la similitud entre documentos, la cual trabaja con la comparación de las palabras, la similitud espacial depende de diferentes elementos como las relaciones y la distribución espacial, los atributos geométricos y temáticos y las relaciones semánticas (Li and Fonseca, 2006).

La similitud espacial es difícil de determinar por las restricciones de las propiedades espaciales y la complejidad de las relaciones espaciales. Debido a que las relaciones espaciales, principalmente la topología, la dirección y la distancia, capturan lo fundamental de la estructura de las escenas (Bruns and Egenhofer, 1996), la mayoría de las investigaciones de la similitud entre escenas espaciales se enfocan en las relaciones espaciales.

El cómo se encuentran relacionados los objetos en el espacio juega un rol fundamental en el manejo de la información espacial, dado que dichas relaciones espaciales se encuentran estrechamente vinculadas a la forma en que percibimos las configuraciones espaciales, el cómo razonamos sobre ellas y también el cómo las describimos (Egenhofer and Mark, 1995a).

Las relaciones espaciales representan un alto nivel de abstracción en comparación con la información posicional (Bruns and Egenhofer, 1996). Estas relaciones pueden ser representadas por primitivas cualitativas, como la conectividad (Randell et al., 1992), y también pueden expresarse con respecto a un marco de referencia, como en el ejemplo: "A se encuentra a la izquierda de B" (Hernández, 1994).

Las relaciones espaciales capturan la estructura de las escenas espaciales por lo que muchas de las investigaciones para calcular similitud se basan en la comparación de las relaciones espaciales. (Bruns and Egenhofer, 1996).

A pesar que se ha realizado investigación en la psicología para entender el proceso cognitivo, no existe una teoría común en la medición de la similitud semántica. A continuación, con base en Schwering (2008), se presenta un análisis de diferentes modelos para mediciones de similitud semántica en el contexto de la geociencia, en cada uno se analiza su forma de representar el conocimiento, su medición de la similitud, sus propiedades métricas, sus requerimientos y suposiciones, se

muestran algunos ejemplos y finalmente se presenta su evaluación para la medición de similitud.

2.4.2.2.1 Modelo geométrico

Los modelos geométricos fueron los primeros utilizados en la psicología para explotar la analogía entre el espacio y las mediciones de similitud (Attneave, 1950), (Torgerson, 1958), (Torgerson, 1965), (Schwering, 2008): en estos modelos se representan los conceptos como puntos en un espacio multidimensional y la similitud se calcula mediante el vector distancia entre dos puntos correspondientes en ese espacio (Tversky, 1977), (Thomas and Mareschal, 1997), (Markman, 2001), (Nedas and Egenhofer, 2003), (Goldstone, 2004).

Representación del conocimiento

Están basados en la noción de espacios vectoriales multidimensionales. Cada dimensión se utiliza para describir propiedades de objetos y conceptos. La mayoría de los modelos geométricos se enfocan en modelar únicamente objetos.

Mediciones de la similitud

La distancia semántica es análoga a la distancia espacial. La similitud es medida como una función de la distancia espacial. Las mediciones más comúnmente utilizadas son las de Minkowski (Schwering, 2008).

Propiedades métricas

La propiedad fundamental de los modelos geométricos y a su vez la más criticada es que estos modelos asumen los axiomas métricos de Minimalidad, Simetría y Desigualdad Triangular (Tversky, 1977), (Thomas and Mareschal, 1997), (Goldstone, 2004). Esta propiedad no coincide con la forma en la que las personas determinan la similitud.

- Minimalidad: $d(a, b) \geq d(a, a)$ (2.12)

$$d(a, b) = 0 \rightarrow a = b \wedge d(a, b) \geq 0 \quad (2.13)$$

Este axioma (ecuación 2.12) establece que si la distancia espacial entre dos conceptos es cero, entonces los conceptos son iguales (ecuación 2.13). Se

deriva de este axioma que la máxima similitud existe en la comparación de un concepto con el mismo. Sin embargo, este axioma no siempre se cumple debido a que la similitud no es la misma para todos los objetos idénticos. Por ejemplo, si los objetos idénticos son complejos tienen más características similares que si los objetos son simples (Takane and Sergent, 1983).

- Simetría: $d(a, b) = d(b, a)$ (2.14)

La ecuación 2.14 establece que la distancia, y por lo tanto la similitud semántica, es la misma entre un concepto a otro.

- Desigualdad triangular $d(a, b) + d(b, a) \geq d(a, c)$ (2.15)

Establece que la distancia entre dos conceptos es siempre menor o igual que la distancia entre dos conceptos por medio de un tercer concepto. Este axioma (ecuación 2.15) no se cumple cuando las comparaciones se realizan en diferentes dimensiones (por ejemplo, balón-luna-lámpara). Por lo mismo la similitud en dimensiones diferentes no es transitiva.

Requerimientos y suposiciones

Los modelos geométricos de similitud semántica deben asumir las siguientes suposiciones:

- Independencia del elemento de representación: se asume que las propiedades son independientes entre sí.
- Solubilidad: el conjunto de propiedades usadas para describir un concepto debe ser lo suficientemente rico y representativo para la conceptualización. Un conjunto de propiedades que no refleje la conceptualización humana no podrá dar buenos resultados de similitud.
- Constante de los elementos representativos: la equivalencia de intervalos en las dimensiones debe preservarse a través de las dimensiones, esto es, las dimensiones debe estar normalizadas.
- Complejidad de la representación: al agregar más información a la descripción de la distancia semántica, ésta sólo puede incrementarse y disminuir la similitud. Así, los modelos geométricos son adecuados sólo para la comparación de conceptos con idéntico número de dimensiones.

Ejemplos de los modelos geométricos

Las representaciones más comunes son los espacios conceptuales (Grdenfors, 2000). Los espacios conceptuales representan la información a un nivel conceptual y están formados por un conjunto de dimensiones de cualidad. Las dimensiones son conectadas a cualidades perceptibles por el sistema sensorial humano. Los objetos son representados como puntos y los conceptos como regiones k-dimensionales.

Evaluación de la medición de similitud entre datos geoespaciales

Los objetos y conceptos son representados del mismo modo. La medición de similitud sólo se puede realizar entre puntos en el espacio conceptual. Por este motivo es difícil realizar mediciones entre conceptos (ya que involucran varios objetos), y se proponen diferentes soluciones, como reducir un concepto a un solo punto, o tomar el promedio de diferentes distancias. Las propiedades son representadas como dimensiones cualitativas pero no hay una representación de las relaciones entre conceptos.

2.4.2.2.2 Modelo de características

El modelo basado en las características asume que los objetos son representados como colecciones de propiedades y la similitud entre ellos está basada en la comparación de estas (Tversky, 1977), (Markman, 2001), (Goldstone, 2004). Mientras en los modelos geométricos las propiedades son dimensiones, en estos modelos son valores booleanos, las propiedades o atributos están o no asociados a un concepto. Se basan en la suposición de que dos conceptos con el mismo atributo son similares en algún aspecto. Las medidas de similitud de modelos de características asumen que la similitud de conceptos aumenta mientras más atributos tienen en común.

Las características iguales en la comparación de dos elementos es la intersección de los conjuntos de características de cada uno de ellos. En este modelo la similitud aumenta con el aumento de los conjuntos de características iguales y disminuye con el aumento de los conjuntos de características diferentes. (Markman, 1993). La similitud se expresa como una función lineal de las características iguales y diferentes. Tversky (1977) plantea que las características iguales tienden a incrementar más la similitud de lo que las características diferentes tienden a disminuirlas.

Representación del conocimiento

El modelo de propiedades tiene una representación del conocimiento con base en la teoría de conjuntos: los objetos y propiedades que tiene un determinado concepto son representados en conjuntos de atributos sin estructura. Al igual que las dimensiones en el modelo geométrico, las características pueden representar variables nominales, ordinales o intervalos.

Existen atributos de adición o sustitución. Para los atributos de adición no es necesario revisar el resto de atributos, mientras que los de sustitución tienen una permanencia en el conjunto restringida, esto es, un atributo sustituirá a otro y viceversa, pero no podrán estar presentes los dos atributos en un mismo conjunto. Los atributos de sustitución pueden verse como colecciones de atributos, donde un objeto podrá tener sólo un atributo de la colección.

Medición de la similitud

Se realiza mediante un modelo de correspondencia de características. Este modelo establece que la similitud no es necesariamente métrica. Los conceptos son representados como colecciones de características. Al representar cada concepto con diferentes conjuntos de características, se les puede aplicar operaciones elementales para estimar similitudes y diferencias.

(Tversky, 1977) propone una medición de la similitud entre conceptos a y b , basada en la teoría de conjuntos, en la cual se calcula la similitud en función de sus atributos similares y diferentes:

$$s(a, b) = f(A \cap B, A - B, B - A) \quad (2.16)$$

Al basarse en lógica de conjuntos, este modelo no soporta correspondencias parciales.

Propiedades métricas

Este modelo es no-métrico. La función tiene como componentes la diferencia de A y B , la diferencia de B y A y su similitud, siendo A y B dos objetos a compararse. Y dependiendo del peso de estos componentes, la función de similitud $s(a, b)$ (ecuación 2.16) es no-métrica.

Tversky probó empíricamente que los tres axiomas métricos no son consistentes en el establecimiento de la similitud humana, dijo que: el axioma de minimalidad es problemático, la simetría es aparentemente falsa y que la desigualdad triangular apenas es convincente.

Requerimientos y suposiciones

- Independencia entre los elementos: el grado en el que una característica compartida por dos conceptos afecta su similitud no debe depender de otra característica.
- Solubilidad: el conjunto de características debe ser lo suficiente rico y representativo. Esto es, si no se tiene un adecuado conjunto de características, los resultados serán también inadecuados.
- Constante de los elementos representativos: los intervalos entre características se asumen equivalentes.

Ejemplos del modelo de características

El modelo de características más conocido es el modelo de contraste y proporción de Tversky, que se basa en la función de similitud como una versión normalizada de la diferencia entre atributos similares y distintos.

La función de similitud puede determinarse por la cardinalidad de los conjuntos o con base en la importancia de algunos atributos.

El *Matching distance similarity measurement* (MDSM) de (Rodríguez and Egenhofer, 2004) es un método para la medición de la similitud entre conceptos geoespaciales que combina dos estrategias: la similitud de características (feature matching) y el cálculo de la distancia semántica. Al tomar en consideración las propiedades cognitivas, este modelo busca representar un método para la medición del grado de interoperabilidad. En su propuesta, un proceso de similitud de propiedades y el cálculo de la distancia semántica, provee una estrategia para crear un modelo que captura la similitud. En la práctica confirman, como dice Tversky, que la similitud no siempre es simétrica (Tversky, 1977). Aunque MDSM utiliza datos espaciales considerando la distancia semántica y la similitud de propiedades, no considera la componente espacial de los datos.

Evaluación de la medición de similitud entre datos geoespaciales

Objetos y conceptos: Tversky aplicó este modelo sólo para objetos.

Propiedades y relaciones espaciales: En un modelo de propiedades es imposible relacionar dos objetos de un modo estructurado. En estos casos la relación es expresada en términos de atributos compuestos (por ejemplo, `junto_a`). Pero hay que recordar que no se detectan correspondencias parciales, por lo que no habrá similitud entre `junto_a_Río` y `junto_a_CuerpoDeAgua`.

2.4.2.2.3 Modelo de red

Este modelo está basado en la teoría de grafos para evaluar la similitud y se basa en redes semánticas para la representación del conocimiento.

Representación del conocimiento

Las redes semánticas están compuestas por nodos y aristas etiquetadas. Los nodos representan unidades de conocimiento: objetos, conceptos o propiedades; las aristas conectan nodos y representan relaciones entre ellos de forma explícita.

Una red semántica necesita una terminología estándar

A pesar que los modelos de red siempre tienen la misma estructura, se diferencian en que dependiendo de su implementación, algunos modelos permiten sólo relaciones taxonómicas, otros de hiponimia y paronimia, y otras permiten todo tipo de relaciones. Algunas redes semánticas restringen la dirección de la relación y asignan pesos para modelar su importancia.

Medición de la similitud

Como el modelo geométrico, los modelos de red basan su medición de similitud en la noción de distancia. Es aquí donde se aplican algoritmos para grafos que calculan el camino más corto (Cormen et al., 2001) como: Búsqueda en anchura (en inglés BFS - *Breadth First Search*), Búsqueda en profundidad (en inglés DFS - *Depth First Search*) o algoritmos que calculan la distancia de acuerdo a pesos como el algoritmo de Floyd-Warshall (Rosen, 2003).

Propiedades métricas

Las medidas de similitud en teorías basadas en grafos son métricas, si la distancia entre conceptos se mide sin importar la dirección de las aristas. Cuando se considera la dirección de las aristas el cómputo resulta en una similitud no métrica: la similitud es asimétrica, pero se cumplen los axiomas de inequidad triangular y de la similitud con uno mismo es igual a cero.

Requerimientos y suposiciones

Solubilidad: La similitud entre dos conceptos sólo puede medirse si existe un camino entre ellos.

Constante de los elementos representativos: se asume que cada relación es relevante y que tienen la misma influencia en el valor de la similitud.

Ejemplo del modelo de red

En (Rada et al., 1989) se propone una medida denominada *Distance*, la cual utiliza de relaciones taxonómicas y relaciones de asociación. Mide la distancia entre dos nodos o conjuntos de nodos. Supone que el juicio humano de similitud es métrico y que la asimetría de la similitud entre conceptos no se deriva de la asimetría propia de la similitud sino de una asimetría entre conceptos, como por ejemplo una categorización difusa.

Por otra parte, (Resnik, 1995) propone una medición de similitud basada en la noción de contexto de información. En enfoques como *Distance* las relaciones representan distancias uniformes, lo que no es cierto en taxonomías reales. La medición de Resnik con base en el contenido supera estas deficiencias. La probabilidad de un concepto se incrementa mientras más alto está en la jerarquía.

Su enfoque es simétrico y transitivo, pero no cumple que la similitud entre un concepto y sí mismo es cero, esto se cumple sólo para el concepto en la cima de la jerarquía.

Evaluación de la medición de similitud entre datos geoespaciales

Objetos y conceptos: este modelo fue pensado en modelar sólo conceptos pero funcionan también con objetos. En cuanto a las propiedades y relaciones espaciales, la fortaleza del modelo está en la representación de sus relaciones.

2.4.2.2.4 Modelo de alineación

Como en el modelo de propiedades, el modelo de alineación utiliza las similitudes y diferencias como nociones de similitud, pero contempla también la estructura relacional donde las propiedades y relaciones se encuentran. El modelo de alineación plantea que la similitud no solo depende de la comparación de las características iguales y diferentes sino también de la alineación de estas características (Markman, 1993); (Goldstone, 1998) (Goldstone, 2004).

En la comparación de dos elementos, las partes de un objeto tienen que estar alineados en correspondencia con las partes del otro (Goldstone, 1994). El proceso de comparación de la similitud devuelve el conjunto de igualdades, el conjunto de diferencias alineadas y el conjunto de las diferencias no alineadas. (Markman and Gentner, 1993), (Medin et al., 1993), (Gentner and Markman, 1994), (Markman and Gentner, 1996).

Representación del conocimiento

Se realiza en una forma estructurada adoptando el marco de trabajo de alineación estructural de Gentner (Gentner and Markman, 1995). Los objetos son representados por sus propiedades incorporadas en un sistema de relaciones. Como parte central de este modelo hay relaciones de alineación que indican la analogía estructural de dos elementos, propiedades o relaciones, pertenecientes a dos objetos diferentes.

Las relaciones de alineación deben ser estructuralmente consistentes y sistemáticas. Para ser estructuralmente consistentes deben cumplir que cada elemento corresponda cuando más a un elemento (correspondencia uno a uno) y que los argumentos correspondientes a cada par de correspondencias tengan también correspondencia. La alineación se llama sistemática si existe una estructura más profunda de atributos y relaciones correspondientes.

Medición de la similitud

Consiste en revisar la similitud entre las estructuras relacionales, mientras el modelo geométrico y de atributos revisa únicamente las similitudes de atributos, el modelo de alineación también considera si estos elementos se alinean o no. Así, las relaciones alineables incrementan la similitud más que las relaciones no alineables.

Requerimientos y suposiciones

- Independencia de los elementos de representación.
- Estructura homogénea: las ventajas de este modelo sólo son evidentes cuando se trabaja en una estructura homogénea. Las reglas de alineación necesitan una estructura uniforme para trabajar de manera automática.
- Solubilidad.
- Constante de los elementos representativos: este modelo asume que cada elemento tiene el mismo peso en la similitud.

Ejemplos del modelo de alineación

El Modelo Similitud como Activación y Mapeo Interactivo (SIAM) se utiliza para medir la similitud entre escenas espaciales, fue propuesto por Goldstone y Medin (Goldstone and Medin, 1994). La similitud entre escenas espaciales se mide en un proceso de aprendizaje iterativo basado en redes neuronales. SIAM calcula todas las posibles alineaciones de relaciones y atributos de manera iterativa hasta que las alineaciones son consistentes y puede determinarse la similitud.

Evaluación de la medición de similitud entre datos geoespaciales

- Objetos y conceptos: SIAM fue definido para medir la similitud entre objetos en escenarios espaciales. La regla de correspondencia uno a uno es difícil de lograr en la similitud entre conceptos si son descritos con diferente nivel de granularidad. Se podría interpretar que un concepto corresponde a un elemento a más detalle de otro concepto.
- Propiedades y relaciones espaciales: SIAM soporta relaciones jerárquicas así como relaciones espaciales. Las propiedades son también consideradas en la medición de la similitud.

2.4.2.2.5 Modelo transformación

El modelo de transformación calcula la similitud utilizando la distancia de transformación (Imai, 1977), (Hahn and Chater, 1997), (Goldstone, 2004). El concepto de distancia de transformación entre dos objetos es el menor número de operaciones necesarias para transformar un objeto en el otro (Goldstone, 2004).

El cómputo de la similitud se realiza de manera diferente que en el resto de los modelos, aquí se definen las transformaciones necesarias para distorsionar un concepto en otro y la similitud está definida en términos del número de transformaciones necesarias para hacer los conceptos iguales.

Representación del conocimiento

Para este modelo, se plasma en la representación un conjunto de transformaciones que pueden ser aplicadas para transformar un concepto. Este conjunto de transformaciones depende de la naturaleza de los conceptos. A pesar de que las transformaciones no tienen que ser perceptivas, la identificación de éstas no es una tarea fácil. Y para aplicar el modelo de transformación se necesita que se defina antes un conjunto finito de transformaciones.

Medición de la similitud

La base de la medición de similitud es el número de transformaciones necesarias para convertir un concepto en otro. Se asume que la similitud decrece de forma monotonica cuando el número de transformaciones aumenta (Imai, 1977), (Wiener-Ehrlich et al., 1980). En (Hahn et al., 2003) se propone el uso de la complejidad de Kolmogorov para calcular la similitud.

Propiedades métricas

El modelo de transformación es asimétrico pero los axiomas de minimalidad e inequidad triangular se conservan (Hahn and Chater, 1997). La similitud entre un concepto y el mismo es máxima, ya que no se necesita de ninguna transformación. Hahn y Chater dicen que la similitud es asimétrica ya que las transformaciones inversas pueden no ser de la misma complejidad.

Requerimientos y suposiciones

Solubilidad: El modelo debe contener al menos todas las transformaciones para llegar de un concepto a otro.

Comparación de elementos de representación: para considerar las transformaciones como un método para medir la similitud, cada transformación debe afectar en mismo grado la similitud, esto es, cada transformación debe ser de

la misma complejidad. La complejidad de Kolmogorov hace posible la comparación entre dos transformaciones de diferente complejidad.

Complejidad de la representación: la transformación de un concepto en otro de la misma granularidad requerirá una transformación simple, mientras que transformaciones entre conceptos de diferente granularidad requerirán transformaciones más complejas.

Ejemplos del modelo de transformación

Este modelo se ha aplicado principalmente en estímulos perceptivos, como cadenas alfabéticas (Wiener-Ehrlich et al., 1980), cadenas de alveolos (Imai, 1977) o complejos geométricos (Hahn et al., 2003). Las transformaciones se enfocan hasta la fecha en atributos perceptibles únicamente. Algunas de las transformaciones usadas para modificar las cadenas son espejo, reversa o agregar símbolos. Transformaciones para modificar el arreglo geométrico son rotación, traslación y dilatación (Goldstone and Son, 2005).

Evaluación de la medición de similitud entre datos geoespaciales

Objetos y conceptos: Hasta ahora se ha aplicado a objetos simples y con una similitud perceptual, pero en teoría se puede aplicar a conceptos geoespaciales y con una similitud conceptual.

Propiedades y relaciones espaciales: Las transformaciones se han aplicado a conceptos y objetos únicamente, pero se puede aplicar también a relaciones. Para aplicar el modelo a relaciones es necesario definir el tipo de transformaciones posibles dentro del marco de trabajo.

2.4.2.3 Similitud semántica espacial.

Como se ha visto hasta ahora a pesar del intento de clasificación de las medidas de similitud en modelos semánticos que se aplican a objetos simples y con una similitud perceptual, que en teoría se puede aplicar a conceptos geoespaciales y con una similitud conceptual, y modelos que determinen similitud entre relaciones espaciales aplicados solo a las propiedades y relaciones espaciales los modelos se presentan con alto grado de integración entre ellos.

Existen muchos enfoques que surgen para los sistemas clásicos de RI y se han adaptado al cálculo de la similitud entre relaciones espaciales. Otros han ido surgiendo orientados específicamente para la recuperación semántica de este tipo de información que tiene sus características particulares. A continuación se presentan algunos de los modelos de similitud más utilizados en la recuperación semántica de información geográfica.

El enfoque del Vecino más Cercano está basado en el modelo de transformación, el cual determina la similitud según la distancia entre dos conceptos en una red calculando el camino más corto entre dos nodos (Rada et al., 1989),(Budanitsky, 1999). Este enfoque ha sido ampliamente utilizado en el cálculo de la similitud entre relaciones espaciales.

(Egenhofer and Al-Taha, 1992) le realizaron cambios graduales a las relaciones topológicas basados en el Modelo de la 9-intersección de Egenhofer. Ellos trabajan con el vecindario conceptual de las relaciones topológicas, cuando se produce algún cambio como una traslación o rotación, el proceso se describe como una secuencia de movimientos sobre la red conceptual. A cada movimiento se le asigna un valor y dependiendo de la cantidad de movimientos que ocurren se determina el valor de la similitud.

En (Papadias and Dellis, 1997) se extiende este modelo en un espacio dimensional superior para representar la similitud de las relaciones espaciales en base a topología, dirección y distancia métrica.

En (Chang and Lee, 1991) se presenta una red de vecindario conceptual de 169 relaciones espaciales posibles entre rectángulos aplicando las relaciones de intervalo 1-D de Allen y las proyecciones ortogonales. (Bruns and Egenhofer, 1996) representan la similitud de las relaciones espaciales a partir del grafo de (Chang and Lee, 1991) y la distancia del modelo de vecindario conceptual; describiendo el proceso de medición de la similitud como la transformación de una escena en otra a través de una secuencia de cambios graduales de las relaciones espaciales. La cantidad de cambios es lo que se compara, las escenas que requieran menor cantidad de cambios son las más similares.

El algoritmo DIS-C (Quintero et al., 2012), evalúa la distancia conceptual entre conceptos de una ontología con base en la ponderación automática de las relaciones entre los conceptos. Se basa en la analogía de una ontología a un grafo fuertemente conexo. El grafo define las relaciones entre conceptos y cada relación

tiene asignada una ponderación que se considera para valorar la proximidad entre conceptos.

2.5 Conclusiones

En esta sección se presentó una panorámica general del estado del arte en el área de la recuperación semántica de información geográfica la cual por ser multidisciplinaria involucra diferentes líneas de investigación, la recuperación de información y los sistemas de información geográfica. Se comentan y discuten diferentes enfoques y técnicas presentados en algunos de los trabajos más relevantes que se han desarrollado. Sin embargo, se identifican varios problemas en las alternativas propuestas y sobre algunas de estas se basan las aportaciones de esta investigación.

El estado de arte muestra que los trabajos desarrollados enfocados a la recuperación de información geográfica, consideran la explotación de las relaciones geométricas y topológicas de forma parcial o en otros casos sin profundizar en su contenido semántico. Por otra parte, se encuentran trabajos que señalan y enfatizan la importancia de las ontologías, sin embargo, no la utilizan como mecanismo que conduzca el proceso de recuperación.

Otro punto importante identificado en esta investigación lo constituye considerar la semántica del análisis espacial que se aplica a los datos geográficos, ya que a través de un análisis espacial se puede obtener nuevos datos. Además, se determina la importancia de considerar las relaciones de topología, de dirección y de distancia en la recuperación por ser características principales en la descripción de los datos geográficos. Muchas investigaciones están encaminadas a encontrar la manera de codificar formalmente las relaciones geográficas.

Uno de los principales objetivos de un sistema de recuperación de información geográfica es resolver consultas que combinen la componente textual y la componente espacial. Las consultas en la búsqueda de información geográfica generalmente consisten en expresiones con alto contenido de semántica espacial (*cerca, alrededor, en, etc.*). Realizar una interpretación automatizada de esta semántica espacial es determinante para obtener los resultados relevantes.

3. Estudio de la medida de similitud óptima.

En este capítulo se presenta una valoración de diferentes mediciones de similitud semántica en el contexto geoespacial y se propone un modelo de integración de similitud semántica espacial (MISE), el cual establece una combinación de dos enfoques, el algoritmo Distancia Conceptual (DIS-C) (Quintero et al., 2012) y el modelo Topología-Dirección-Distancia (TDD) (Li and Fonseca, 2006), complementando ambos enfoques para mejorar la relevancia de la información recuperada.

3.1 Selección de una medida de similitud.

En el área de la recuperación de información geográfica la similitud semántica se calcula trabajando con objetos y conceptos geoespaciales. Un concepto es una idea que caracteriza un conjunto de objetos. (Sloman et al., 1998). Un objeto geoespacial se refiere a una instancia ubicada en el espacio geográfico. Las medidas de similitud deben trabajar con objetos y conceptos para medir la interoperabilidad entre fuentes de datos geoespaciales.

La semántica de objetos geoespaciales es compleja y tiene algunas características particulares: es descrita por propiedades como forma, tamaño y ubicación. Además, que las relaciones, y en particular las relaciones espaciales juegan un papel principal en la descripción semántica (Schwering, 2008).

Determinar el “mejor” enfoque de medición de la similitud semántica o cual proporciona los mejores resultados no es posible porque depende de muchos factores (Goldstone and Son, 2005):

- La similitud entre dos conceptos depende de los otros conceptos con los que esté relacionado. Por ejemplo, la similitud entre dos conceptos disminuye cuando el número de conceptos iguales aumenta.
- La similitud depende del contexto, la cual juega un rol protagónico en las mediciones. Algunos modelos representan el contexto a través de factores denominado *pesos* que se establecen entre las relaciones. La importancia de las propiedades puede variar dependiendo de la tarea en la que se utilice.
- La valoración de los resultados obtenidos también depende si los valora un experto o no en la materia.

En el campo de la recuperación de información geográfica la medida de similitud establece la relevancia de una información en dependencia de la consulta realizada. Se tiene que establecer una comparación directa de la consulta con la información a recuperar, solamente los criterios especificados en esta consulta son relevantes para determinar si los datos son relevantes. Sin embargo, las medidas de similitud tienen que incluir propiedades y relaciones espaciales en el cálculo de las mismas por estas son esenciales en la semántica de los datos geoespaciales. Los sistemas de recuperación de información comparan dos conceptualizaciones: la conceptualización del usuario representada a través de la consulta y la conceptualización de las fuentes de datos.

En la sección 2.4.2 del Capítulo 1 se presenta una panorámica de las mediciones de la similitud, ninguna de estas medidas cumple con todos los requerimientos de la recuperación semántica de información geográfica. En este capítulo se presenta como una combinación de dos enfoques, el algoritmo DIS-C (Quintero et al., 2012) con el modelo TDD (Li and Fonseca, 2006), estos enfoques se complementan y mejoran la relevancia de la información recuperada.

3.2 Distancia Conceptual DIS-C.

Mucho de los modelos presentados en la sección 2.4.2, solo funcionan con ontologías de relaciones taxonómicas lo que limita su aplicación. El algoritmo DIS-C no tiene esa limitante por lo que es aplicable a cualquier tipo de ontología porque utiliza un modelo de red. Este algoritmo se encarga de determinar el valor de

distancia conceptual a cada uno de los tipos de relaciones en la ontología. Esto se logra al convertir esta última en un grafo dirigido y ponderado en el que cada concepto se convierte en un nodo y cada relación se convierte en un par de aristas. Esto es aplicable de manera directa a las relaciones binarias pero además, es posible aplicarlo a relaciones n-arias. Una vez que se tiene el grafo, es posible aplicar diferentes técnicas de teoría de grafos. El paso natural consiste en calcular las distancias mínimas, con lo que se puede encontrar la distancia entre conceptos que no se encuentran directamente relacionados.

Algo significativo que se maneja es el cálculo de la distancia conceptual, partiendo de la distancia que existe entre conceptos directamente relacionados; la cual es asignada a-priori por quien ha creado la ontología. El algoritmo que se propone para la propagación de la distancia conceptual establece que una relación debe tener una distancia conceptual asociada, tanto en el sentido de la relación, como en el sentido inverso; con esta información se genera un grafo fuertemente conexo en el que cada concepto es un vértice y cada relación se representa con dos aristas (una en el sentido directo y otra en el sentido inverso). Utilizando un algoritmo de Floyd-Warshall, se propagan las distancias locales para poder determinar la distancia entre cualquier par de conceptos.

Así, los pasos a seguir para calcular la distancia conceptual en DIS-C son los siguientes:

1. Para cada tipo de relación, se debe asignar la distancia conceptual que genera la existencia de dicha relación. Además, se debe asignar una ponderación en cada sentido de la relación; por ejemplo, si tenemos la relación “es” y que “gato es animal”, se debe establecer la distancia de “gato” a “animal” y de “animal” a “gato”. A manera de ilustrar se puede establecer que $distancia(gato, animal) = 1$ y $distancia(animal, gato) = 0$. Más formalmente, sea $K(C, \mathfrak{R}, R)$ una conceptualización donde C es el conjunto de conceptos, \mathfrak{R} es el conjunto de tipos de relaciones y R es el conjunto de relaciones en la conceptualización. Se puede decir que para cada tipo de relación $\rho \in \mathfrak{R}$, se deben establecer los valores de δ^ρ , para la relación en el sentido normal, y de $\bar{\delta}^\rho$, para la relación en el sentido inverso. En el ejemplo anterior $\delta^{es} = 0$ y $\bar{\delta}^{es} = 1$.

2. Se crea el grafo $G_K(V, A)$ para la conceptualización. Primero, se agrega cada concepto $c \in C$ como un vértice del grafo G_K , es decir que $V = C$. Ahora para cada relación $a\rho b \in R$, donde $a, b \in C$ y $\rho \in \mathfrak{R}$; se añaden las aristas (a, b, δ^ρ) y $(b, a, \bar{\delta}^\rho)$.

3. Se calculan los caminos mínimos entre cada par de vértices a través del algoritmo de Floyd-Warshall, los cuales están mapeados directamente a los conceptos en la ontología; por lo que el resultado es la distancia conceptual diseminada a todos los conceptos en una conceptualización K .

A continuación se realiza una comparación de los resultados de DIS-C con métodos de similares características, es decir, métodos que utilicen el modelo de red, como es el caso de **Confusion** (Levachkine and Guzmán-Arenas, 2007), **Cantidad de información** (Resnik, 1999) y **Distance** (Radat al., 1989). La comparación se realiza tomando como referencia los resultados presentados en (Levachkine and Guzmán-Arenas, 2007), donde se utiliza una jerarquía de seres de vivos (ver Figura 3.1).

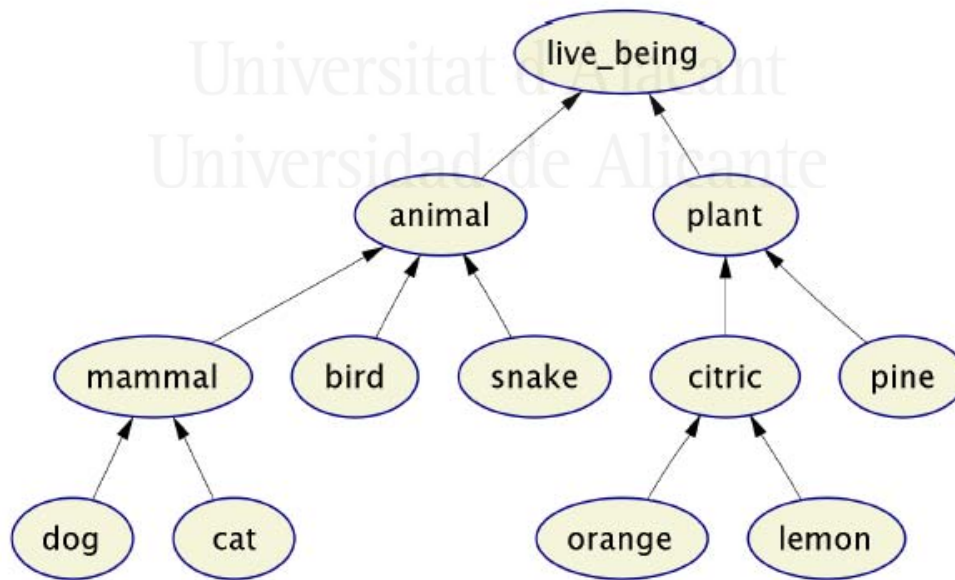


Figura 3.1. Jerarquía de seres vivos (Levachkine and Guzmán-Arenas, 2007).

En las Tabla 3.1, Tabla 3.2, Tabla 3.3 y Tabla 3.4 se presentan los resultados del valor de similitud de la jerarquía de la Figura 3.1 utilizando **Confusión**, **Cantidad de**

información, **Distance** y **DIS-C**, respectivamente. Asimismo, en la Tabla 3.5 se presenta la coorelación de los resultados obtenidos con los diferentes métodos, donde se puede apreciar que en general los resultados obtenidos con DIS-C se correlacionan fuertemente con los resultados de los otros métodos; de hecho, tiene el promedio de coorelación más alto de todos.

Tabla 3.1. Valor de similitud de la jerarquía de los seres vivos utilizando *Confusion*.

<i>Confusion</i>	live_being	animal	plant	mammal	bird	snake	citric	pine	cat	dog	lemon	orange
live_being	0	1	1	2	2	2	2	2	3	3	3	3
animal	0	0	1	1	1	1	2	2	2	2	3	3
plant	0	1	0	2	2	2	1	1	3	3	2	2
mammal	0	0	1	0	1	1	2	2	1	1	3	3
bird	0	0	1	1	0	1	2	2	2	2	3	3
snake	0	0	1	1	0	1	2	2	2	2	3	3
citric	0	1	0	2	2	2	0	1	3	3	1	1
pine	0	1	0	2	2	2	1	0	3	3	2	2
cat	0	0	1	0	1	1	2	2	0	1	3	3
dog	0	0	1	0	1	1	2	2	1	0	3	3
lemon	0	1	0	2	2	2	0	1	3	3	0	1
orange	0	1	0	2	2	2	0	1	3	3	1	0

Tabla 3.2. Valor de similitud de la jerarquía de los seres vivos utilizando Cantidad de información.

Resnik	live_being	animal	plant	mammal	bird	snake	citric	pine	cat	dog	lemon	orange
live_being	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
animal	0.0	0.3	0.0	0.3	0.3	0.3	0.0	0.0	0.3	0.3	0.0	0.0
plant	0.0	0.0	0.4	0.0	0.0	0.0	0.4	0.4	0.0	0.0	0.4	0.4
mammal	0.0	0.3	0.0	0.6	0.3	0.3	0.0	0.0	0.6	0.6	0.0	0.0
bird	0.0	0.3	0.0	0.3	1.1	0.3	0.0	0.0	0.3	0.3	0.0	0.0
snake	0.0	0.3	0.0	0.3	0.3	1.1	0.0	0.0	0.3	0.3	0.0	0.0
citric	0.0	0.0	0.4	0.0	0.0	0.0	0.6	0.4	0.0	0.0	0.6	0.6
pine	0.0	0.0	0.4	0.0	0.0	0.0	0.4	1.1	0.0	0.0	0.4	0.4

cat	0.0	0.3	0.0	0.6	0.3	0.3	0.0	0.0	1.1	0.6	0.0	0.0
dog	0.0	0.3	0.0	0.6	0.3	0.3	0.0	0.0	0.6	1.1	0.0	0.0
lemon	0.0	0.0	0.4	0.0	0.0	0.0	0.6	0.4	0.0	0.0	1.1	0.6
orange	0.0	0.0	0.4	0.0	0.0	0.0	0.6	0.4	0.0	0.0	0.6	1.1

 Tabla 3.3. Valor de similitud de la jerarquía de los seres vivos utilizando *Distance*

RADA	live_being	animal	plant	mammal	bird	snake	citric	pine	cat	dog	lemon	orange
live_being	6	5	5	4	4	4	4	4	3	3	3	3
Animal	5	6	4	5	5	5	3	3	4	4	2	2
plant	5	4	6	3	3	3	5	5	2	2	4	4
mammal	4	5	3	6	4	4	2	2	5	5	1	1
bird	4	5	3	4	6	4	2	2	3	3	1	1
snake	4	5	3	4	4	6	2	2	3	3	1	1
citric	4	3	5	2	2	2	6	4	1	1	5	5
pine	4	3	5	2	2	2	4	6	1	1	3	3
cat	3	4	2	5	3	3	1	1	6	4	0	0
dog	3	4	2	5	3	3	1	1	4	6	0	0
lemon	3	2	4	1	1	1	5	3	0	0	6	4
orange	3	2	4	1	1	1	5	3	0	0	4	6

Tabla 3.4. Valor de similitud de la jerarquía de los seres vivos utilizando DIS-C

DISC	live_being	animal	plant	mammal	bird	snake	citric	pine	cat	dog	lemon	orange
live_being	0.00	0.73	0.73	1.47	1.47	1.47	1.47	1.47	2.20	2.20	2.20	2.20
animal	0.26	0.00	0.99	0.73	0.73	0.73	1.73	1.73	1.47	1.47	2.46	2.46
plant	0.26	0.99	0.00	1.73	1.73	1.73	0.73	0.73	2.46	2.46	1.47	1.47
mammal	0.51	0.26	1.25	0.00	0.99	0.99	1.98	1.98	0.73	0.73	2.72	2.72
bird	0.51	0.26	1.25	0.99	0.00	0.99	1.98	1.98	1.73	1.73	2.72	2.72
snake	0.51	0.26	1.25	0.99	0.99	0.00	1.98	1.98	1.73	1.73	2.72	2.72
citric	0.51	1.25	0.26	1.98	1.98	1.98	0.00	0.99	2.72	2.72	0.73	0.73
pine	0.51	1.25	0.26	1.98	1.98	1.98	0.99	0.00	2.72	2.72	1.73	1.73
cat	0.77	0.51	1.50	0.26	1.25	1.25	2.24	2.24	0.00	0.99	2.97	2.97
dog	0.77	0.51	1.50	0.26	1.25	1.25	2.24	2.24	0.99	0.00	2.97	2.97
lemon	0.77	1.50	0.51	2.24	2.24	2.24	0.26	1.25	2.97	2.97	0.00	0.99

orange	0.77	1.50	0.51	2.24	2.24	2.24	0.26	1.25	2.97	2.97	0.99	0.00
--------	------	------	------	------	------	------	------	------	------	------	------	------

Tabla 3.5. Correlación de resultados.

	DISC	Confusion	Resnik	Rada
DISC	1	0.9546	0.6405	0.9360
Confusion	0.9546	1	0.5397	0.7885
Resnik	0.6405	0.5397	1	0.6845
Rada	0.9360	0.7885	0.6845	1
Promedio	0.8827	0.8207	0.7161	0.8522

La correlación con *Confusion* es muy alta, de hecho si se redondean los valores obtenidos con DIS-C se obtienen cerca del 80% de valores idénticos a los de *Confusion*. Esto parece indicar que DIS-C proporciona una mayor precisión en la estimación de la diferencia entre dos conceptos y que al mismo tiempo es compatible con los resultados de *Confusion*. Otro aspecto interesante es que DIS-C se corelaciona fuertemente tanto con *Confusion* (95%) y con *Distance* (94%), pero sin embargo, la corelacion entre ellos no es del mismo orden (78%). Esto indica que DIS-C arroja resultados que confluyen entre estos dos métodos, y que proporciona una medida acorde a ambas visiones. Basado en estos resultados se selecciona el algoritmo DIS-C.

3.3 Modelo de similitud espacial Topología-Dirección-

Distancia

El modelo Topología-Dirección-Distancia (TDD) (Li and Fonseca, 2006) se enfoca a escenas con dos objetos para realizar el análisis de similitud. Este modelo calcula el conjunto de características iguales entre los objetos geográficos, C , y el conjunto de las diferencias alineadas y no alineadas, D . La similitud final S (ecuación 3.1) se calcula como una combinación de ambas.

$$S = C - D. \quad (3.1)$$

Donde:

$$C = C_{topologica} + C_{direccionales} + C_{distancia} \quad (3.2)$$

$$D = |\alpha, \beta| * \left| \frac{\text{diferencia alineadas}}{\text{diferencias no alineadas}} \right| \quad (3.3)$$

Donde: α, β son los pesos

Los valores de los pesos pueden variar en dependencia del contexto pero estos valores por defecto son $\alpha = 1$ y $\beta = 1$ o $\beta = 0$. El valor 0 significa que las diferencia no alineadas no existen.

El modelo TDD para la comparación se basa en el modelo de características alineadas, donde las partes de un objeto deben de ser alineadas o colocadas con las partes de otro. Se enfoca en la similitud entre las relaciones considerando que la similitud entre las relaciones tienen diferentes impactos sobre la valoración de las igualdades y las diferencias (Tversky, 1977). Dependiendo del contexto; las diferencias entre las relaciones serán más o menos relevantes en la evaluación de la similitud. Enfocado en este problema la solución propuesta en el modelo TDD es tener diferentes pesos asociados a cada atributo y a cada relación. De este modo los usuarios del modelo TDD colocan los pesos de forma interactiva así ellos reflejan la situación particular en el cual la medida de similitud está siendo hecha.

Las relaciones en las que se basa el modelo son las topológicas, de dirección y de distancia. Otra contribución del modelo TDD incluye un orden de prioridad en la aplicación (Topología \rightarrow Dirección \rightarrow Distancia) en la valoración de la similitud espacial. La estrategia adoptada para ordenar la prioridad en la aplicación del modelo TDD es colocando los costos de los pesos de transformación sobre los tres tipos diferentes de relación espacial. Cada tipo de relación espacial tiene su propio vecindario conceptual.

$$\text{peso (topología)} > \text{peso(dirección)} > \text{peso(distancia)} \quad (3.4)$$

Por defecto: $\text{peso (topología)}:\text{peso(dirección)}:\text{peso(distancia)} = 3:2:1$

3.3.1 Relaciones espaciales.

En la Figura 3.2 se muestra la red de las relaciones topológicas que se pueden establecer entre los diferentes tipos de objetos geospaciales. Los objetos geospaciales, como se explica en la sección 2.2.1, se pueden representar a través de puntos, líneas y polígonos. En la Figura 3.2 (a) se visualizan las relaciones topológicas entre polígonos y se establecen los pesos entre las relaciones. Los seis nodos pueden ser divididos en tres grupos: separados, superpuestos (en diferentes niveles) e iguales. En la Figura 3.2 (b) se visualizan las relaciones para un polígono y una línea, en (c) un polígono y un punto, en (d) entre dos líneas, en (e) entre dos puntos y en (f) entre una línea y un punto.

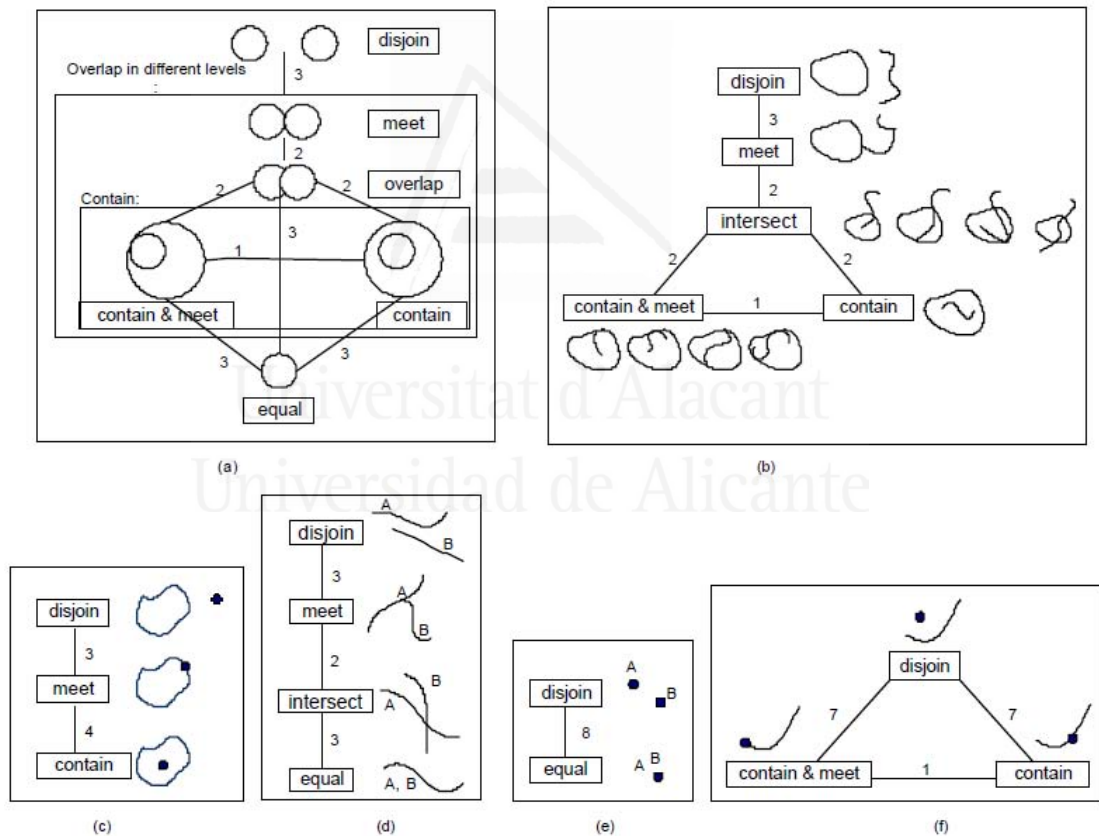


Figura 3.2. Entorno de red conceptual de las relaciones topológicas: (a) dos polígonos; (b) una línea y un polígono; (c) un punto y un polígono; (d) dos líneas; (e) dos puntos; (f) un punto y una línea (Li and Fonseca, 2006).

El enfoque de vecindario conceptual aplica el modelo de transformación para calcular la similitud de las relaciones a través de la distancia de transformación. Esta distancia se calcula como el número de arcos que hay que atravesar para ir de un estado a otro. Cuantos menos arcos hay que pasar mayor es la similitud. Se aplica un costo de transformación que está representado en la Figura 3.2.

3.3.2 Relaciones direccionales.

La relación de dirección en el modelo TDD trabaja sobre el sistema de 9-direcciones

$\{\text{norte, noroeste, oeste, suroeste, sur, sureste, este, noreste, igual}\}$ para representar las direcciones. La correspondencia del vecindario conceptual se representa en la Figura 3.3.

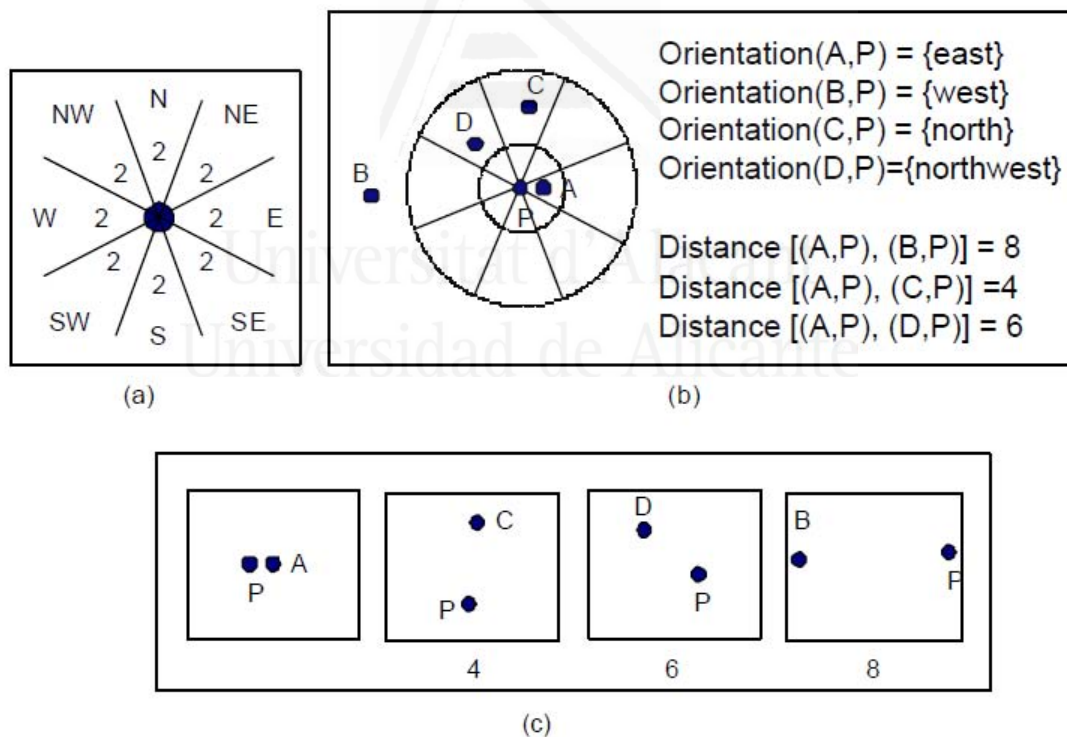


Figura 3.3. (a)Vecindario direccional 9-direcciones; (b) ejemplo; (c) listado ordenado de los ejemplo de (b).

(Egenhofer and Franzosa, 1991) proponen una variante para el vecindario direccional llamado p/2-direcciones y en este trabajo se reemplaza la red de la

Figura 3.3 (a) por el de la Figura 3.4 (a). Hay cinco nodos en la direccional p/2-direcciones:

$\{igual\}$ $\{este, oeste\}$ $\{noreste, suroeste\}$ $\{norte, sur\}$ y $\{noroeste, sureste\}$ (ver Figura 3.4 (a)). Los costos de transformación también se encuentran representados, el costo de transformación de un nodo a su vecino es 2. El costo de transformación dentro de un nodo es 1.

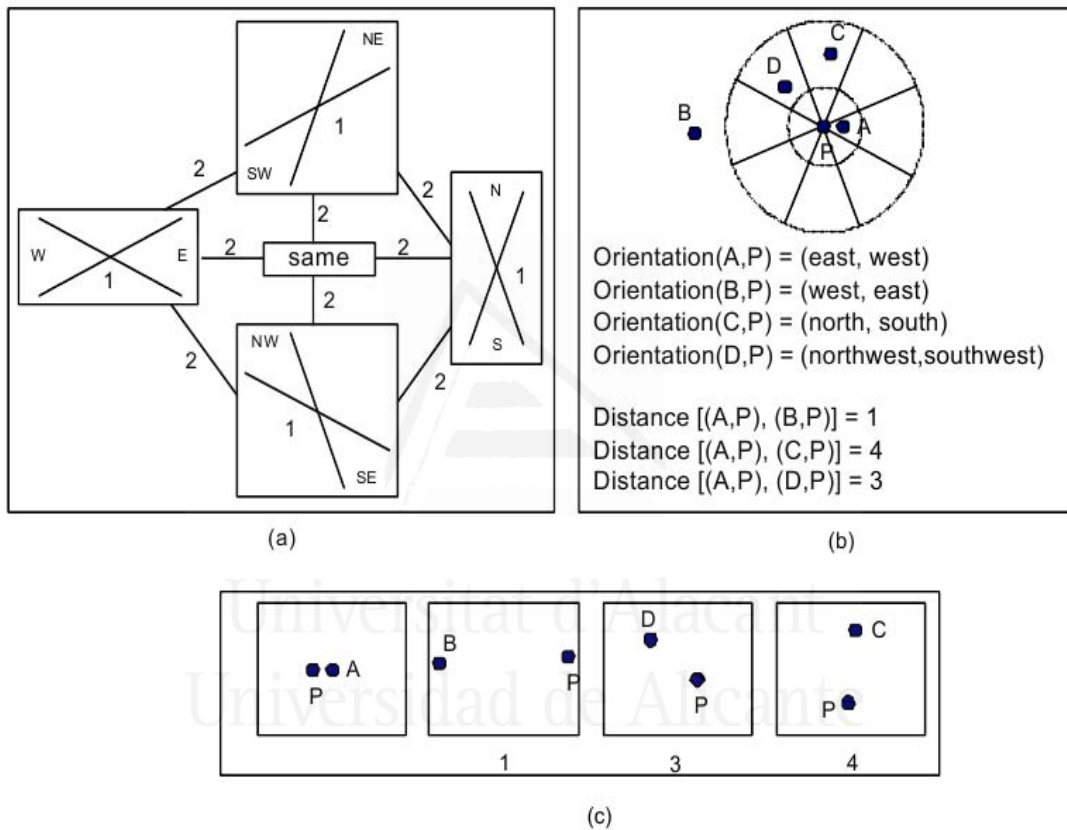


Figura 3.4. (a) Vecindario direccional p/2-direcciones; (b) ejemplo; (c) listado ordenado de los ejemplo de (b)

Por ejemplo, para calcular la similitud entre una relación direccional (*este oeste*) y una (*oeste, este*) se calculan las transformaciones que hay que realizar con su costo. En la Figura 3.4 (b) se muestra un ejemplo para el cálculo de la transformación. La $Distance[(A,P), (B,P)] = 1$ porque (A,P) y (B,P) están ubicados en el mismo nodo $\{este, oeste\}$ aunque tenga diferente dirección. La $Distance[(A,P), (C,P)] = 4$ porque hay dos arcos entre los nodos donde (A,P) y (C,P) están localizados. Para el caso de (A,P) y (D,P) es más complicado. Hay

solamente un arco entre los nodos donde se encuentran ubicados; sin embargo, el ángulo entre \overrightarrow{AP} y \overrightarrow{DP} es mayor, es necesario igualar los ángulos, al realizar esta operación cambiamos de nodo, realizando la primera transformación y después está también la transformación dentro del mismo nodo por lo que la distancia es $2 + 1 = 3$.

3.3.3 Relación de distancia.

Para determinar la similitud en cuanto a distancia se utiliza el modelo de 4-granularidad {igual (*equal*), cerca (*near*), lejos (*medium*), muy lejos (*far*)} que se presenta en la Figura 3.5.

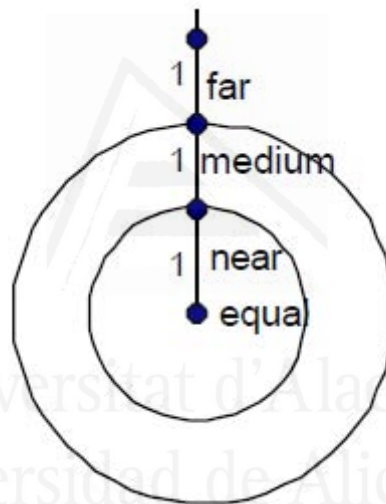


Figura 3.5. Red de distancia

En la Figura 3.5 se presentan cuatro nodos y tres arcos en la red de distancia. Cada arco tiene asociado como costo de transformación 1. Si se calcula la distancia entre la relación cerca y muy lejos, el costo de transformación es $1 + 1 = 2$.

3.4 Integración de criterios

Debido a que se tienen definidos dos criterios en la valoración de la similitud, es necesario, establecer la forma como se integrarán. Con este objetivo se propone un espacio en analogía con los espacios métricos multidimensionales.

El principal objetivo de este espacio es servir como mecanismo para combinar el análisis semántico con el análisis espacial, ya que si estos se complementan mutuamente, se obtienen resultados más satisfactorios, es decir, resultados con la precisión y flexibilidad de los procedimientos de análisis de similitud semántica y con la refinación de las operaciones geoespaciales.

El espacio para la integración de criterios debe cumplir las siguientes propiedades y definiciones:

- 1) Cada dimensión es usada para mapear un criterio de recuperación.
- 2) Cada criterio o dimensión es independiente de los otros criterios.
- 3) Cada dimensión es configurada de tal forma que los valores más cercanos al origen, son los más apropiados para satisfacer una consulta, como se ilustra en la Figura 3.6.
- 4) Cada dimensión tienen un grado de relevancia, determinado por un peso ω_i , estos pesos se le pueden solicitar al usuario cuando realice la consulta, que configure el nivel de importancia que él asigna a cada criterio en su consulta, o en su lugar tomar valores por defecto.



Figura 3.6. Espacio para la comparación de criterios.

En la Figura 3.6 se visualiza que los valores más cercanos a cero satisfacen mejor la consulta del usuario. Para un objeto O_i que satisface una consulta, será representado como una tupla de valores $O_i(\varepsilon, \sigma)$ en el espacio. Con las siguientes definiciones:

- El valor de la similitud conceptual ε se define como el valor de similitud obtenido al aplicar el algoritmo DIS-C: si $\varepsilon \rightarrow 0$ entonces el objeto O_i es muy similar al buscado, si $\varepsilon = 0$ entonces el objeto satisface exactamente a la consulta.

- El valor de la similitud conceptual σ se define como el valor de similitud obtenido al aplicar el algoritmo TDD: si $\sigma \rightarrow 0$ entonces el objeto O_i es muy similar, en cuanto a sus relaciones espaciales, al buscado, si $\sigma = 0$ entonces el objeto satisface exactamente a la consulta.

El valor de relevancia R , de cada resultado, es decir, cada punto en el espacio para la integración de criterios, es calculado como señala la ecuación 3.5:

$$R_w(c_i, w_i) = 1 - \left(\frac{\sum_{i=1}^n (c_i)^{w_i}}{n} \right) \quad (3.1)$$

Donde c_i son los criterios de recuperación, w_i los pesos asignados a cada criterio y n el número de criterios a utilizar, en este caso $n = 2$.

La ecuación 3.5 permite determinar los resultados que más se aproximan o satisfacen la consulta. De manera tal que, si $R_w \rightarrow 1$ entonces el resultado satisface totalmente la consulta, un ejemplo del espacio para la integración se muestra en la Figura 3.7.

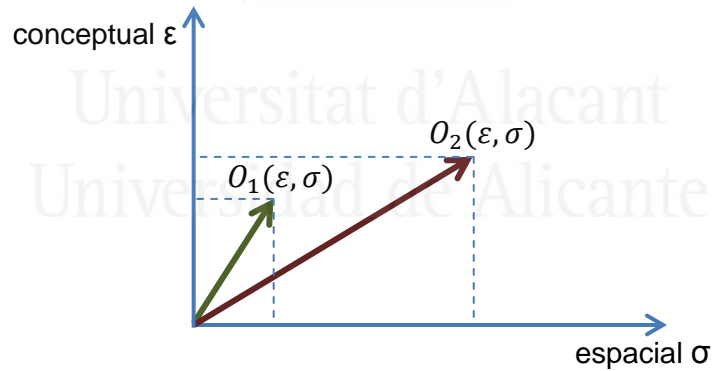


Figura 3.7. Comparación de objetos en el espacio para la comparación de criterios.

Según la Figura 3.7 el objeto $O_1(\varepsilon, \sigma)$ es más satisfactorio por estar más próximo al origen. En este ejemplo, $\sigma_1 < \sigma_2$ y $\varepsilon_1 < \varepsilon_2$.

3.5 Conclusiones

En este capítulo se presentó un modelo mediante el cual se puede calcular la distancia semántica entre regiones conceptuales e incluye las relaciones espaciales en la medición de esta similitud. En la sección 3.1 se describe el algoritmo DIS-C mediante el cual se calcula la distancia semántica entre conceptos, en la sección 3.2 se explica el método TDD el cual se enfoca más en las relaciones espaciales.

Finalmente, se explica en la sección 3.4 como se realiza la integración de los criterios de similitud utilizados (Machado et al., 2014c). Con este método se obtiene una combinación de DIS-C con el modelo TDD. Por lo tanto, se combina el modelo geométrico, de características, el de transformación y el de alineación como ya propone TDD pero además le incorpora el componente semántico a través de la aplicación de DIS-C.



Universitat d'Alacant
Universidad de Alicante



Universitat d'Alacant
Universidad de Alicante

4. Propuesta y diseño de la ontología geoespacial

En este capítulo se describe la metodología para la representación de datos geoespaciales. El principal objetivo es proporcionar descripciones semánticas que representen las propiedades y relaciones que describen el comportamiento de los objetos geográficos a través de una ontología geográfica (GeoOntología). Esta decisión se basa en el hecho de que no hay disponibilidad de GeoOntologías en español para los datos geográficos utilizados en esta investigación.

Con la concepción de esta ontología se espera mejorar las tareas de integración e intercambio de geo-información, para proporcionar soluciones con respecto a la ambigüedad de términos en el desarrollo de los SIG. Las ontologías imponen una estructura sobre el dominio y restringen las posibles interpretaciones de los términos (Weigand, 1997).

Las ontologías son estructuras que pueden crecer, integrarse con otras ontologías y reutilizarse en la construcción de ontologías de otros dominios. Lo más importante a resaltar es su potencial semántico que le da el hecho de incluir axiomas y de esta manera da la posibilidad de emplear la lógica y poder inferir conocimientos implícitos en los conocimientos formalizados.

Esta ontología ha sido desarrollada aplicando la metodología GEONTO-MET (Torres Ruiz, 2007). Esta metodología permite conceptualizar el dominio geográfico proporcionando un mecanismo para la integración y recuperación de datos geoespaciales, representados a través de una descripción semántica que se obtiene a partir de un esquema conceptual. Por lo tanto, la descripción semántica está compuesta por relaciones y conceptos que se utilizan para describir explícitamente las propiedades de los objetos geográficos; donde estas relaciones son del tipo topológico y geométrico.

4.1 Metodología GEONTO-MET

GEONTO-MET (Torres Ruiz, 2007) es una metodología orientada a formalizar la conceptualización del dominio geográfico. Su principal objetivo es proporcionar descripciones semánticas, las cuales representen las propiedades y relaciones que describen el comportamiento de los objetos geográficos. Está compuesta por cuatro etapas: **análisis**, **síntesis**, **procesamiento** y **descripción** (Ver Figura 4.1). Se parte del **conocimiento del dominio geográfico**, basado en diversas fuentes de información generadas por expertos del dominio.

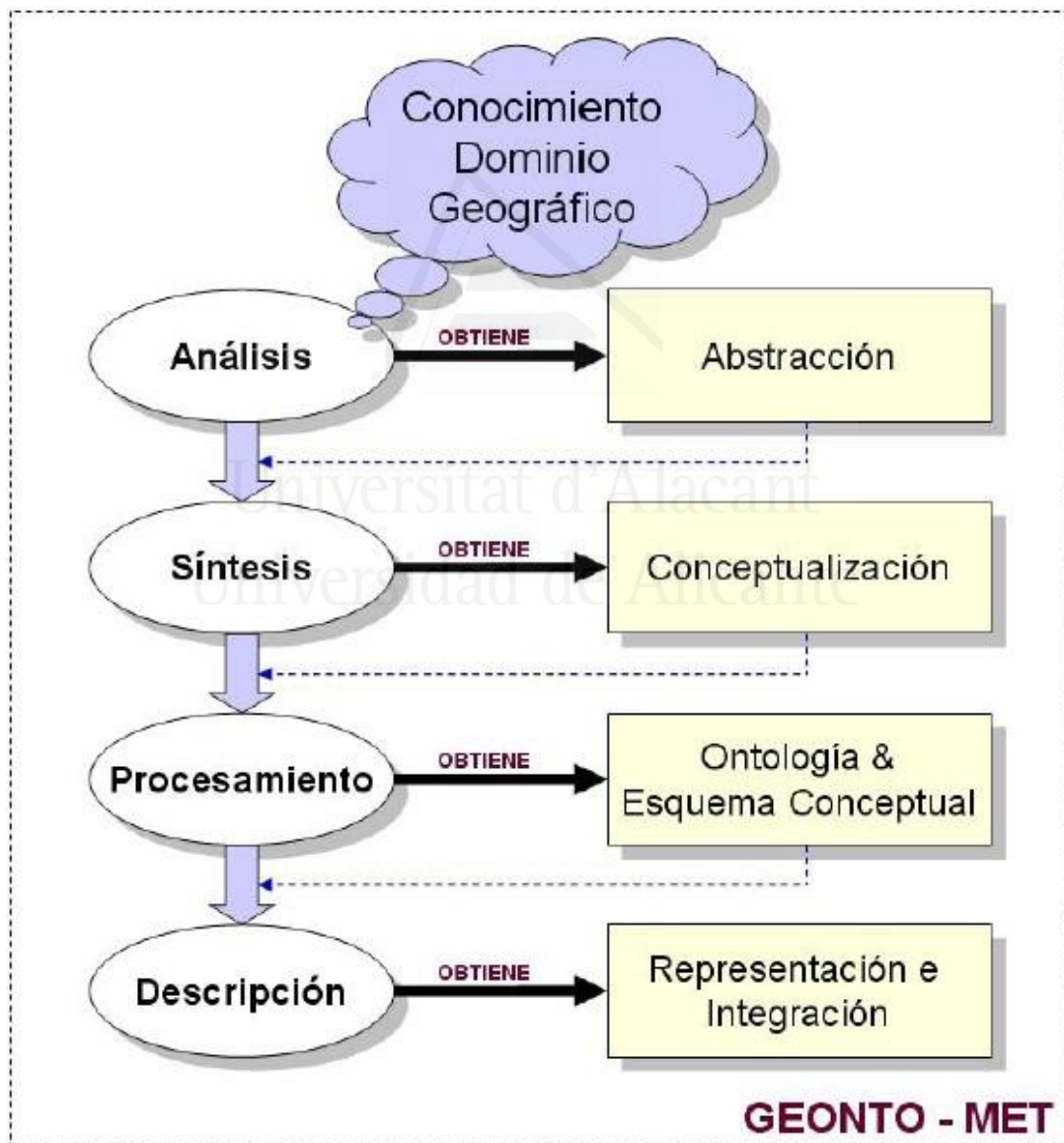


Figura 4.1. Estructura conceptual de GEONTO-MET (Torres Ruiz, 2007)

En la etapa de **análisis** se realiza una investigación exhaustiva sobre los objetos geográficos que se encuentran involucrados en el dominio geográfico para proporcionar un modelo de abstracción en el que quedan identificados todos aquellos objetos, de acuerdo con la forma en que los expertos del dominio perciben estas entidades. Las entidades abstractas permiten simplificar la forma en la que pueden ser organizados, clasificados y almacenados jerárquicamente los conceptos, lo cual se realiza por medio de clases que reflejan la naturaleza o características que posee cada concepto geográfico.

Una vez obtenida la abstracción del dominio geográfico en sentido general se pasa a la etapa de **síntesis**. En esta etapa se realiza un análisis de los objetos geográficos involucrados para obtener como resultado una clasificación de estos, definiendo un conjunto de *clases de entidades abstractas*. Estas clases permiten describir en términos generales y basándose en las características implícitas de los objetos geográficos a estas entidades. Además, se cuenta con las propiedades y relaciones entre estos objetos; así como una jerarquía a partir de estas clases de entidades abstractas.

Dentro de esta etapa se hace necesario restringir el dominio que puede tener la ontología dentro del universo geográfico, de acuerdo con las fuentes de información analizadas para acotar el dominio, es decir, definir el significado de los objetos geográficos por medio de particiones basadas en diferentes *contextos*, donde el significado de los objetos geográficos involucrados puede variar, de acuerdo al dominio de aplicación en que se encuentran situados o donde actúan estos objetos geográficos. El contexto de un objeto geográfico juega un rol preponderante, puesto que éste presenta una dependencia directa para asignar un significado a los objetos geográficos. Como resultado de esta etapa, se obtiene una *conceptualización* del dominio geográfico.

El objetivo de la etapa de **procesamiento** es realizar una interacción directa y dependiente entre sí de las clases de entidades abstractas de la ontología con el esquema conceptual. La interacción consiste en verificar que conceptos (objetos geográficos) de la ontología de dominio se encuentran conceptualizados y pueden ser mapeados como instancias de conceptos al esquema conceptual; es decir, qué objetos geográficos se encuentran presentes en diversos contextos y que tipo de significado toman estos conceptos de acuerdo al contexto en el que se encuentran

dirigidos. Como resultado de esta etapa se obtiene una ontología del dominio geográfico; así como una ontología de aplicación con su correspondiente esquema conceptual.

Por último, con base en las ontologías de dominio y de aplicación; así como las relaciones y contextos que son percibidos en este universo geográfico, la etapa de **descripción** se encarga de generar una descripción de los objetos en un contexto particular. El objetivo de esta descripción es solucionar en forma parcial problemas de integración y recuperación de datos geoespaciales en diversas bases de datos geográficas.

La forma de llevar a cabo esta descripción está basada principalmente en consultas específicas, considerando la definición implícita en las ontologías que pueden tener estos conceptos para recuperar *instancias* de los mismos con sus propios valores y atributos.

Esta descripción proporciona como resultado la definición del concepto al que pertenece el objeto geográfico, junto con todas y cada una de las instancias que pertenecen a ese objeto; siempre y cuando existan instancias de los mismos en las bases de datos.

En las secciones posteriores se describe el funcionamiento de la metodología para generar la ontología de dominio geográfico y la ontología de aplicación proporcionando algunos ejemplos de fragmentos de las mismas.

4.2 Diseño de la ontología

Para iniciar con la definición del desarrollo de la ontología, se requiere definir el dominio; es decir, es necesario responder a un conjunto de preguntas o cuestionario de competencia, los cuales permiten describir la **conceptualización** del mismo. Entre las preguntas esenciales que se formularon se encuentran las siguientes:

- ¿Cuál es el dominio que la ontología cubrirá?
 - Dentro del dominio geográfico, la clasificación y degradación de los suelos de Cuba.
- ¿Para qué se utilizará la ontología?

- Para conceptualizar el dominio geográfico, y de éste poder generar ontologías que permitan recuperar datos geoespaciales.
- ¿Para qué tipos de preguntas la información de la ontología proporcionará respuestas?
 - ¿Cuáles son los tipos genéticos de suelos en Cuba?
 - ¿Cómo se encuentran agrupados los tipos de suelos?
 - ¿Cuáles son los subtipos de un determinado tipo de suelos?
 - ¿Cuáles son las características de los distintos tipos de suelos?
 - ¿Cuáles son los tipos de suelos que se encuentran en el Municipio San José de las Lajas?
 - ¿Se encuentra un suelo Ferralítico Rojo junto a un suelo Pardo?
 - ¿Qué es la degradación de los suelos?
 - ¿Cuáles son las principales causas de la degradación de los suelos?
 - ¿A qué se le denomina deterioro biológico de los suelos?
 - ¿A qué se le denomina deterioro físico de los suelos?
 - ¿A qué se le denomina deterioro químico de los suelos?
 - ¿Cómo se puede solucionar el déficit de materia orgánica de los suelos?
 - ¿Existen prácticas de conservación para suelos con niveles elevados de acidez?
 - ¿Qué es la erosión de los suelos?
 - ¿Cuántos tipos de erosión existen?
 - ¿Cómo se denomina el proceso mediante el cual las partículas de suelo son arrastradas producto de la acción del viento?
 - ¿Qué es la erosión hídrica?
 - ¿Qué se entiende por deforestación?
 - ¿Cómo se denomina el proceso mediante el cual las áreas de pastoreo son utilizadas de forma indiscriminada sin considerar la masa de ganado por superficie?
 - ¿Qué es la compactación de los suelos?
 - ¿A qué le llamamos lixiviación de los suelos?
 - ¿Cuándo ocurre sellado de los suelos?
 - ¿Cuándo podemos decir que el suelo es ácido?
 - ¿Cuáles son las causas que provocan la acidez en los suelos?

- ¿En qué se fundamenta el fenómeno de la alcalinización?
- ¿Cuándo podemos decir que un suelo está contaminado?
- ¿Cómo podemos saber cuándo estamos en presencia de un suelo salino?
- ¿A qué llamamos Cuenca Hidrográfica?
- ¿Qué consecuencias pudiera tener la deforestación indiscriminada?
- ¿Para qué se emplea la rotación de cultivos en la agricultura?
- ¿Qué es el cultivo en franjas?
- ¿A qué se le denomina Fajas Buffer?
- ¿Para qué se utiliza el Encalado de los suelos?
- ¿Cuándo se llevará a cabo el mantenimiento de la ontología?
 - El mantenimiento de la ontología se realiza con en los cambios que puedan presentar los estándares del Open Geospatial Consortium (OGC) (OGC, 2003) y de la norma ISO 19115⁹.

Para el desarrollo de esta tesis se consideraron como fuentes de información cuatro componentes fundamentales:

1. La especificación ISO 19115 orientada a metadatos e información geográfica.
2. Documentos especializados con las definiciones de los términos que representan los objetos geográficos con los que se trabajaron.
3. Manual para la aplicación de la Nueva versión de clasificación genética de los suelos de Cuba.
4. Criterio del especialista geográfico (manejo de terminología especializada de un experto para el dominio espacial).

Esta ontología está compuesta por un conjunto de clases de entidades abstractas, las cuales posteriormente agrupan a un conjunto de conceptos. Por lo tanto, para la ontología de este dominio se define a un concepto¹⁰ como una colección de habilidades y propiedades abstractas, las cuales presentan existencia única. Existen tres tipos de conceptos que pueden ser:

⁹ La norma ISO 19115 es un estándar de metadatos para información geográfica, la cual ha sido establecida por el TC 211 on Geographic Information/Geomatics.

¹⁰ De acuerdo con la definición de WordNet, un concepto puede definirse como una concepción o construcción de algo abstracto o de una idea inferida o derivada de la realidad o de instancias específicas.

- Conceptos relación: Están asociados a las acciones (verbos) que pueden realizar un concepto. Se definen como elementos que denotan una acción u operación sobre conceptos. Al conjunto de los conceptos relación se denominan como C_R .
- Conceptos estándar: Se relacionan directamente con los objetos abstractos (sustantivos), los cuales se definen como elementos que pertenecen a una clase de entidades abstractas. Al conjunto de los conceptos estándar se denominan como C_E .
- Conceptos Clase: Sirven para generar particiones de C_R y C_E . A este tipo de conceptos se denotan por C_L .

De esta manera el conjunto C se expresa como $C = C_R \cup C_E \cup C_L$, tal que $C_E \cap C_R = \emptyset$, $C_E \cap C_L = \emptyset$ y $C_L \cap C_R = \emptyset$; es decir que C_L , C_R y C_E son conjuntos disjuntos y no existen excepciones para ningún caso.

Es importante señalar que las relaciones básicas que se consideran con respecto a los objetos geográficos son topológicas, de dirección y de distancia. No obstante, en el esqueleto de la ontología se agregan relaciones de pertenencia y del tipo temático. Asimismo, las relaciones axiomáticas de la metodología pueden considerarse como hipónimos. En este caso, se dice que:

- **A** es – un hipónimo de **B**, si existe una relación es – un de **A** a **B**. Es decir, se consideran hipónimos si **A** es subtipo de **B**, o bien si **A** es una clase de **B**.

Por

ejemplo:

(**ferralítico_rojo_compactado** es_subtipo_de **ferralítico_rojo**), por lo que (**ferralítico_rojo_compactado** es_hipónimo_de **ferralítico_rojo**) ; ya que comparten propiedades tales como el perfil¹¹ y el color rojo.

- En el sentido inverso de la relación
(**ferralítico_rojo** es_hiperónimo_de **ferralítico_rojo_compactado**).

Con base en la metodología propuesta para conceptualizar el dominio geográfico, se han generado dos conceptualizaciones diferentes: (1) referente al dominio de los objetos geográficos y (2) referentes a los diferentes perfiles de suelos y su

¹¹ El perfil del suelo es la ordenación vertical de todos los horizontes de suelo.

degradación. Por cuestiones de simplificación estas ontologías serán denominadas O_G y O_S respectivamente.

Ambas ontologías deben estar relacionadas, una con respecto a la otra y viceversa. Por lo tanto, deben existir relaciones entre los conceptos de O_G con los conceptos de O_S . En este caso, se puede decir que el concepto $a \in O_S$, el concepto $b \in G(a)$ y el concepto $b \in O_G$. Por lo tanto, $\forall a \exists c \exists b(es)c \in R_R$, esto indica que todos los conceptos de O_S se encuentran relacionados mediante un mecanismo de herencia con algún concepto de O_G .

Donde:

- $G(a)$ es la genealogía de un concepto $a \in C$, la cual se define como el conjunto de conceptos que tienen una relación de existencia; es decir, la genealogía de a es el conjunto de los padres, abuelos, bisabuelos, etc., del concepto a ; a través de los cuales se define la existencia del mismo. Esto puede expresarse de la siguiente forma:
Sea $a, b \in C$, entonces la genealogía $G(a)$ está dada de manera recursiva por $G(a) = \{b | a(es)b \in R_V\} \cup_{j=1}^n G(x_j)$.
- R_R es el conjunto de relaciones que existen en una conceptualización real o concreta.

Por otro parte, se denota a R_V como el conjunto de relaciones válidas. Entonces, R_V se considera el conjunto de todas aquellas relaciones, tales que cumplen con las restricciones definidas para conceptualizar el dominio geográfico, lo cual puede expresarse como: $R_V = (r_{v_1}, r_{v_2}, r_{v_3}, \dots, r_{v_{1n}}) \in A_2$.

Para realizar el proceso de clasificación jerárquico de los conceptos dentro de las clases, GEONTO-MET define dos conjuntos de axiomas esenciales que permiten evitar inconsistencias en la representación del concepto; así como una serie de relaciones simples y complejas que permiten asociar a los conceptos geográficos con otras clases de cualquier tipo.

Los axiomas en la ontología especifican las definiciones de los términos y las restricciones de su interpretación. En este sentido es importante entender el

significado del uso de los axiomas, un axioma debe definir la semántica o significado de todos y cada uno de los conceptos de las ontologías. Como se observa en la ecuación 4.1 el primer axioma consta de tres relaciones, las cuales se denotan como A_1 .

$$A_1 = \{es, tiene, hace\}, \rho \in A_1 \quad (4.1)$$

$$\rho \in es, tiene, hace$$

Donde el conjunto A_1 representa los axiomas básicos que describen a las relaciones simples:

“*es*” describe una relación de **existencia** de objetos geográficos.

“*tiene*” indica una relación de **pertenencia** de propiedades de los objetos geográficos o conceptos de tipo relación.

“*hace*” es una relación de **operación**, la cual describe acciones y eventos de los objetos geográficos.

Algunos ejemplos para el uso de este axioma se describen a continuación:

- “**polígono de suelo** *es* **objeto geográfico natural**”
- “**suelo** *tiene* **textura**”
- “**ferralítico amarillento lixiviado** *hace* compartir con **pardo**”

El segundo subconjunto de relaciones axiomáticas se denota por A_2 y está compuesto por términos simples que representan a las preposiciones (ver ecuación 4.2)

$$A_2 = \left\{ a, ante, bajo, con, contra, de, desde, en, entre, hacia, hasta, para, por, según, sín, sobre, tras, junto_a, a_través_de \right\}, \pi \in A_2 \quad (4.2)$$

$$\pi = preposiciones$$

En este caso, el conjunto A_2 se utiliza para definir las restricciones de la relación de operación *hace*. Los elementos de este conjunto son *preposiciones*, las cuales se utilizan para mapear conceptos del tipo *relación* con conceptos *estándar*.

Los conceptos *relación* C_R denotan una acción u operación (*verbos*) sobre conceptos. Los conceptos *estándar* C_E son elementos que pertenecen a una clase de entidades abstracta (sustantivos).

Es importante señalar que el número de relaciones posibles que puede contener la ontología está dado por la ecuación 4.3

$$N_{rp} = n_{\alpha_1}n(n-1) + n_{\alpha_2}n(n-2) \quad (4.3)$$

Donde:

$n = \text{card}(C)$ es el número de conceptos que tiene el conjunto R_V

$n_{\alpha_1} = \text{card}(A_1)$ es el número de relaciones axiomáticas del conjunto A_1

$n_{\alpha_2} = \text{card}(A_2)$ es el número de relaciones axiomáticas del conjunto A_2

4.2.1 Modelado del dominio geográfico

En esta sección se presentan los elementos que componen a la ontología. El editor utilizado para el diseño fue Protégé¹² versión 4.1 y además se utilizó la interfaz Jena para acceder al editor de ontologías y recuperar conceptos para vincularlos con los datos geoespaciales. Esta ontología está compuesta por un conjunto de clases de entidades abstractas, las cuales posteriormente agrupan a un conjunto de *conceptos* que pueden ser del tipo *relación*, *clase* y *estándar*.

En la Figura 4.2 se muestra la jerarquía básica de la ontología. Se puede observar que los conceptos que la componen son considerados clases dentro de la misma ontología. El concepto o clase padre es “OWL: Thing”, el cual a través de una relación “*es*” permite generar clases esenciales para conceptualizar los objetos del dominio geográfico. La clase “*entidad_geografica*” está compuesta por tres conceptos del tipo clase que envuelven al dominio geográfico, tales como “*fenomeno:geografico*”, “*relacion_geografica*” y “*objeto_geografico*”.

¹² <http://protege.stanford.edu/>

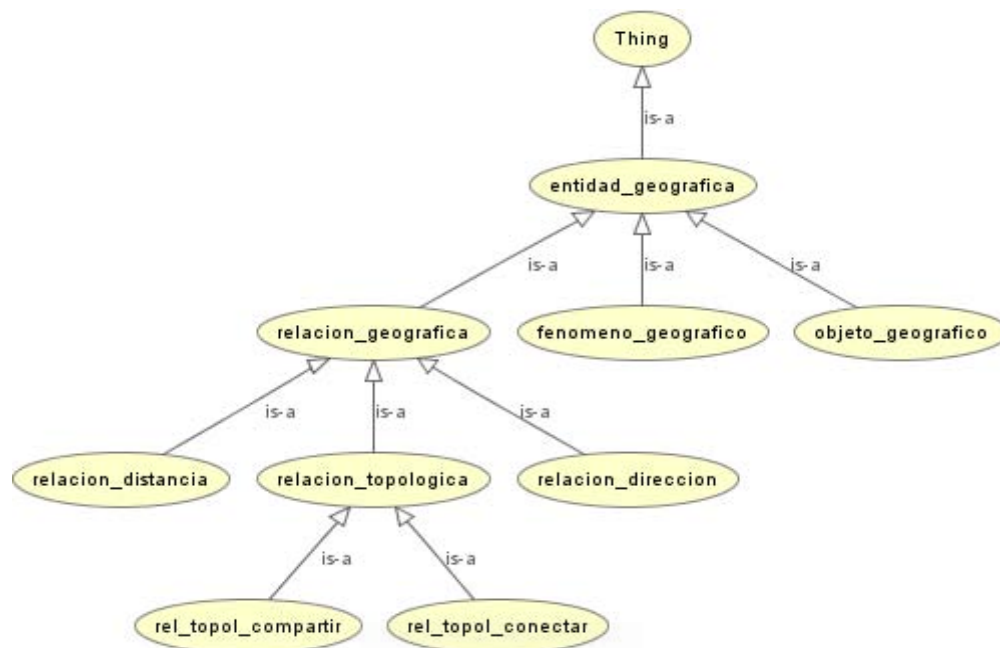


Figura 4.2. Jerarquía básica de la ontología.

La relación de existencia “es” juega un rol vital en la herencia o especialización de otras clases. De acuerdo a la metodología una clase solo puede tener una relación de existencia para evitar las herencias redundantes y restringir aquellos términos que sirven para clasificar a los conceptos a los conceptos del tipo estándar.

Además, se puede observar en la Figura 4.2 que la clase “*relacion_geografica*” está compuesta por tres clases abstractas “*relacion_direccion*”, “*relacion_distancia*” y “*relacion_topologica*”, la cual a su vez está compuesta por un conjunto de conceptos de tipo relación, como es el caso de “*rel_topol_compartir*” y “*rel_topol_conectar*”.

Para el caso de estudio desarrollado en esta tesis solo se consideran los dos conceptos del tipo relación representados en la Figura 4.2. No obstante, esta metodología permite agregar un mayor número de conceptos del tipo relación siempre y cuando éstas sean requeridas por el usuario. En la Tabla 4.1 se proporciona una definición de las clases esenciales que componen la ontología.

Tabla 4.1. Clases esenciales que componen la ontología.

Entidad básica	Definición
Entidad geográfica	Elemento que ocupa una posición en el espacio, puede considerarse como una clase genérica con un registro de localización. Los elementos que componen una entidad son fenómeno geográfico, objeto geográfico y relación geográfica.
Clase esencial	Definición
Fenómeno geográfico	Es un factor esporádico que puede alterar el ambiente, como lo puede ser un huracán, una tormenta tropical o un terremoto.
Objeto geográfico	Es una abstracción de un fenómeno del mundo real, se considera un elemento asociado con una localización relativa a la Tierra. Estos objetos pueden ocurrir en dos niveles: instancias y tipos. Una representación del mundo real puede ser pensada como un conjunto de objetos.
Relación geográfica	Son vínculos que permiten definir el comportamiento que presentan los objetos geográficos con respecto a otros objetos. Además, permiten describir restricciones de los objetos geográficos al ser analizados y desplegados. Las relaciones son los elementos esenciales que describen y dan significado al contexto de los objetos geográficos.
Clase abstracta	Definición
Relación topológica	Son relaciones entre objetos geográficos que son caracterizados por ser invariantes a transformaciones geométricas del espacio, tales como traslación, rotación y escalamiento. Se utilizan para describir la conectividad de las primitivas geométricas que pueden derivarse de la geometría fundamental
Relación de dirección	Describen la orientación de características geográficas espaciales. Estas relaciones son invariantes en la traslación y

	escalamiento pero no en la rotación.
Relación de distancia	Se centran en la distancia entre las características geográficas. Son invariantes a las transformaciones de rotación, traslación, pero no bajo el escalamiento. Se determinan de forma cuantitativa o cualitativa determinando la distancia entre dos características geográficas.

Es importante hacer notar que la semántica de los objetos geográficos se encuentra realmente en las relaciones entre los mismos. No obstante, el complemento del vocabulario representado por un lenguaje de objetos geográficos, permite interpretar en forma adecuada el conjunto de relaciones permisibles R_R que pueden existir entre éstos.

Para este trabajo se consideran las relaciones tipificadas por el INEGI (Instituto Nacional de Estadística, Geografía e Informática) en sus diccionarios de datos (INEGI, 1996), las cuales han sido conceptualizadas por expertos de dominio y se consideran como la base para vincular diversos objetos geográficos entre sí. Además, estas relaciones permiten identificar el comportamiento que pueden tener los objetos geográficos, acordes a un contexto específico en el que se encuentran inmersos.

En algunas ocasiones estas relaciones pueden considerarse como implícitas; sin embargo, en la mayoría de los casos estas relaciones deben ser explícitas para tener en forma clara la interpretación de éstas y dar un significado concreto¹³ a la conceptualización dada por los expertos.

Como se mencionó anteriormente, las relaciones concebidas en la investigación son: “*compartir*” y “*conectar*”. Estas relaciones han sido definidas como *conceptos tipo relación* y pertenecen a la clase “*relación_topológica*”. Para describir este tipo de relación se utiliza un mapeo basado en una *relación ternaria*, la cual se denomina **relación compuesta**. Este tipo de relación se describe de la siguiente forma: $apb\pi c \in R_C$, donde $a, b, c \in C$; $p \in A_1$ y $\pi \in A_2$.

¹³ Significa no tener ambigüedad en la interpretación y representación explícita de las relaciones.

Para este caso, “*relacion_topologica*” cumple con una condición necesaria y suficiente para describir las relaciones de este tipo entre diversos objetos geográficos. Está vinculada por medio de la relación axiomática “hace” la cual pertenece al conjunto A_1 y se vincula con los elementos del conjunto A_2 . La implantación de esta relación se basó en los siguientes hechos:

Existe una clase definida como “*elemento abstracto*” (ver Figura 4.3), la cual hereda clases por medio de la relación de existencia “es”, obteniendo así las clases “*relacion_atomica*” y “*relacion_compuesta*”. La clase “*relacion_atomica*” está compuesta por las clases “*preposiciones*” y “*relacion_basica*”. A su vez, cada una de estas clases contiene conceptos del tipo relación que se encargan de poblar las clases citadas anteriormente. Estas clases se pueblan con la relaciones axiomáticas A_1 y A_2 respectivamente.

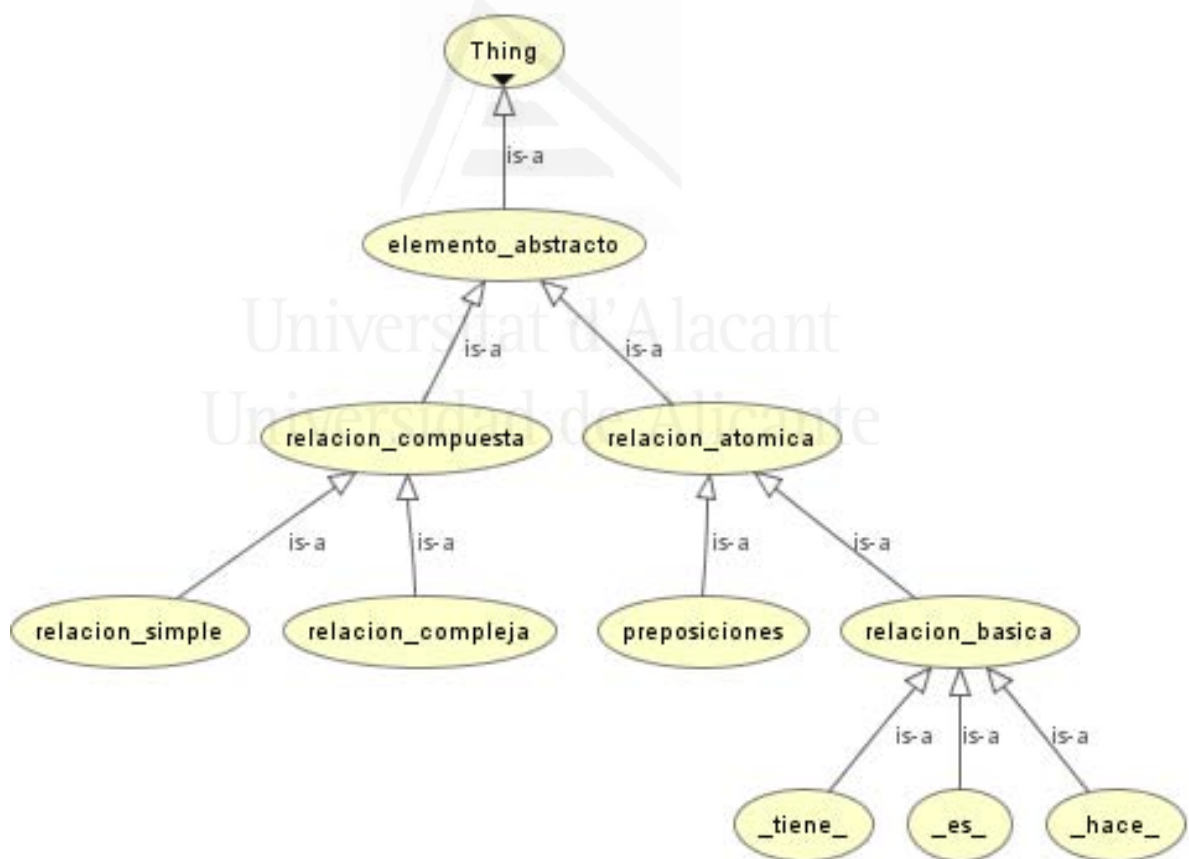


Figura 4.3. Jerarquía de clases de “*elemento_abstracto*”

Para implementar las *relaciones complejas* o *ternarias* se ha diseñado un *slot* denominado “*propiedades_de_relacion*”, el cual se muestra en la Figura 4.4. Este *slot*¹⁴ contiene objetos del tipo relación los cuales se mencionan a continuación:

- “_origen_”. Pertenece al dominio “*relación_compleja*” y el rango¹⁵ sobre el cual actúa es “*entidad_geografica*”.
- “_preposicion_”. Pertenece al dominio “*relacion_compleja*” y el rango sobre el cual actúa es “*entidad_geografica*” y “*elemento_abstracto*”.
- “_accion_”. Pertenece al dominio “*relacion_compleja*” y el rango sobre el cual actúa es. “*relacion_geografica*”.
- _destino_. Pertenece al dominio “*relacion_compleja*” y el rango sobre el cual actúa es. “*entidad_geografica*”.
- _relacion_. Pertenece al dominio “*relacion_compleja*” y el rango sobre el cual actúa es. “*entidad_geografica*” y “*elemento_abstracto*”.

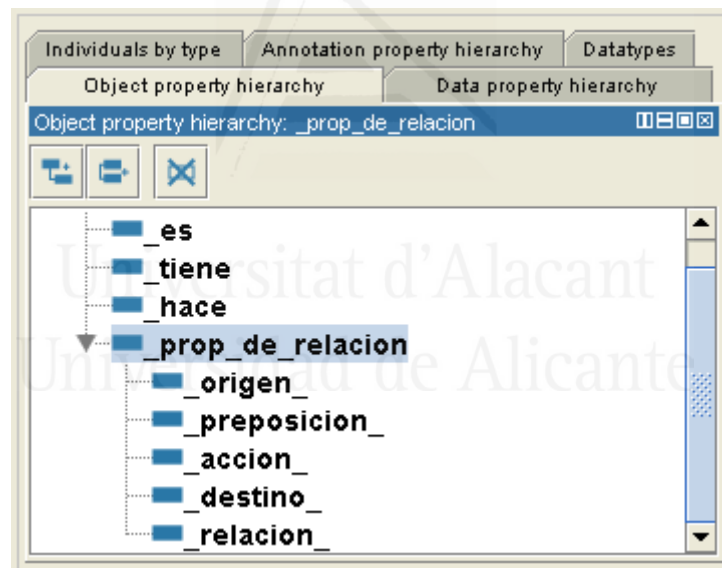


Figura 4.4. *Slot* de relaciones que realizan el mapeo de las relaciones axiomáticas.

Por cuestiones de implementación la relación ternaria es almacenada como un individuo con el objetivo de describir este tipo de relación. Si se tiene como ejemplo “**polígono suelo comparte con presa**”. La forma de describir este tipo de relación

¹⁴ Un *slot* en Protégé es equivalente a una propiedad. También se conocen como roles en lógica descriptiva y relaciones en UML y otras nociones orientadas a objetos.

¹⁵ El rango indica el alcance que tiene un objeto hacia una clase; es decir, hace referencia de todos los conceptos que pueden ser alcanzados para relacionarse con un objeto del tipo relación. Este rango se define en el momento de establecer las restricciones de una clase en Protégé.

entre los tres tipos de conceptos: polígono suelo (**concepto estándar**), comparte (**concepto relación**) con (**relación axiomática**) y presa (**concepto estándar**) es cumpliendo las siguientes restricciones:

“**polígono_suelo_comparte_con_presa**” se encuentra definida en la clase “relación_compleja”, con lo cual:

$\forall_origen_ \Rightarrow poligono_suelo,$

$\forall_relacion_ \Rightarrow_hace_,$

$\forall_accion_ \Rightarrow rel_topol_compartir,$

$\forall_preposicion_ \Rightarrow_con_,$

$\forall_destino_ \Rightarrow presa$

De esta forma, se realiza el mapeo entre objetos geográficos utilizando una relación ternaria. En la Figura 4.5 se muestra un ejemplo de la relación compleja descrita anteriormente.

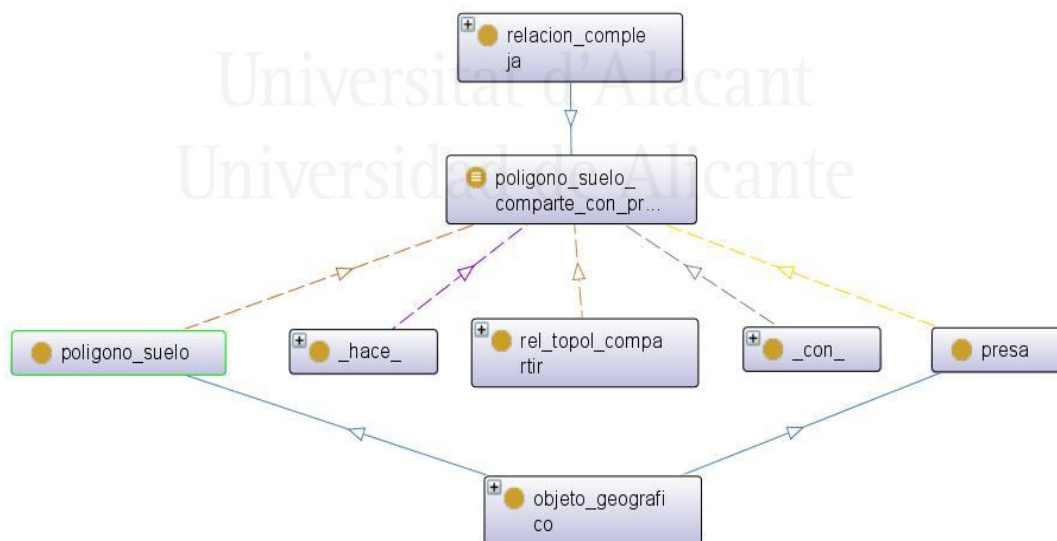


Figura 4.5. Mapeo de una relación ternaria.

En la Figura 4.5 se puede observar que la clase “*relación_compleja*” tiene como subclase a “*polígono_suelo_comparte_con_presa*”, esta relación cuenta con dos conceptos estándar, dos relaciones axiomáticas y un concepto del tipo relación.

Todos los enlaces se describen con líneas de diferentes colores y se puede observar además, la representación de una relación compuesta $apbpc \in R_C$.

Por lo tanto, si se desea describir ontológicamente las relaciones de instancias pertenecientes a conceptos estándar y relación, se utiliza como ejemplo la instancia “*relación_ternaria_SP*”, la cual está basada en el caso anterior, “*poligono_suelo_comparte_con_presa*”. Entonces, si se tiene como instancia del concepto “*poligono_suelo*” a “*poligono_suelo_4*” y como instancia de “*presa*” a “*presa_mampostón*”, estos son individuos descritos por medio de objetos de tipo relación.

Las propiedades que poseen los objetos geográficos almacenados en las bases de datos espaciales constituyen la conceptualización que se modeló en la ontología. Las bases de datos espaciales almacenan datos descriptivos y datos espaciales. Los descriptivos están orientados a las características atributivas de los datos, tales como áreas de los objetos, indicadores, entre otros. Los espaciales poseen diversas propiedades entre las que se encuentran las topológicas, de dirección y de distancia (Egenhofer and Mark, 1995a), (Mark, 1999), (Papadias and Kavouras, 1994).

Las relaciones espaciales establecen la ubicación en el espacio de alguna propiedad geográfica del mundo real con respecto a otra propiedad, describen la configuración geoespacial. Las relaciones espaciales entre características geográficas son frecuentemente tan importantes como las características geográficas (Papadias and Kavouras, 1994).

Las relaciones espaciales son muy importantes en la descripción semántica de los objetos geográficos: debido a que todos los objetos geográficos tienen una localización en la tierra ellos automáticamente tienen una relación espacial con respecto a todos los otros objetos geográficos. En este trabajo se manejan las relaciones espaciales a nivel conceptual. Establecer relaciones espaciales entre clases no siempre es posible debido al carácter abstracto de las mismas pero existen ciertas características que pueden ser generalizadas a todas las instancias (Schwering, 2006).

Es importante restringir el dominio que puede tener la ontología dentro del universo geográfico de acuerdo con las fuentes analizadas para acotar el dominio, es decir,

restringir los objetos geográficos basados en diversos contextos en los cuales el significado de los objetos puede variar. Este contexto depende del dominio de aplicación en que se encuentran situados o donde actúan estos objetos geográficos.

La ontología geográfica se construyó de forma semiautomática, es decir, manualmente y utilizando programas de cómputos. Esta decisión se basa en el hecho de que no hay disponibilidad de GeoOntologías en español para los datos geográficos utilizados en esta investigación.

La ontología del contexto Suelo está compuesta por un conjunto de conceptos de tipo clase y relación. Está basada en la Nueva Versión de Clasificación Genética de los Suelos de Cuba (Hernández et al., 1999). Según la metodología utilizada para conceptualizar el dominio geográfico se propone una estructura para clasificar y ordenar estos datos. Por otra parte, la conceptualización del dominio de aplicación está relacionada intrínsecamente en el contexto del caso de estudio, el cual en este trabajo está enfocado a aspectos topográficos y de clasificación de los suelos cubanos.

4.2.2 Modelado del dominio de aplicación

La ontología de aplicación se desarrolla acerca del contexto de los Suelos de Cuba (Hernández et al., 1999). Está compuesta por un conjunto de conceptos del tipo clase, estándar y relación. A continuación se describen los términos esenciales de esta ontología.

La base de la clasificación se apoya en la génesis de los suelos. Los horizontes principales sirven para clasificar los **Agrupamientos o Grupos de Suelos** y dentro de cada uno se separan los **Tipos genéticos**. Dentro de cada Tipo se establecen los Subtipos de suelos teniendo en cuenta también los horizontes normales, pero en su mayor parte por las características de diagnóstico.

El **tipo genético** es la unidad taxonómica básica de la clasificación y comprende aquellos suelos que presentan una manifestación clara de un mismo proceso de formación.

El **subtipo de suelos** refleja las formas cualitativas de manifestarse el proceso de formación de los suelos y puede constituir, además, transiciones entre los tipos.

El **género** está determinado por las características químico mineralógicas del suelo, relacionado con la roca madre o con la intensidad del lavado de las sustancias.

En el Anexo 5 se presentan las categorías de clasificación que se utilizaron en el diseño de la ontología. Dentro de los conceptos de la ontología, además, existe una clase denominada “*datos*” que se encarga de contener clases de conceptos que derivan en instancias, los cuales permiten clasificar a otros conceptos que pertenecen y están asociados a las clases de entidades abstractas.

Las instancias contenidas en la clase “*datos*” pueden ser consideradas como **atributos** o **propiedades** particulares de conceptos estándar. Esta clase hereda directamente de la superclase “**OWL:Thing**” (ver Figura 4.6).

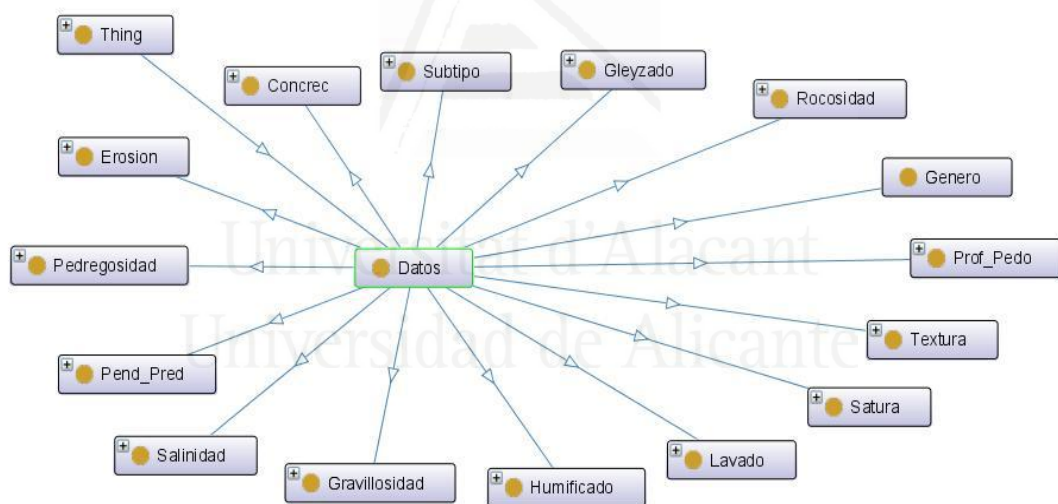


Figura 4.6. Jerarquía de la clase “*datos*”.

Las instancias de estas clases son propiedades o atributos que permiten calificar a los conceptos estándar con tipos más articulares, enriqueciendo la semántica de los mismos; puesto que se tiene mayor granularidad en la descripción del concepto al especializar éste o contando con mayor información al respecto.

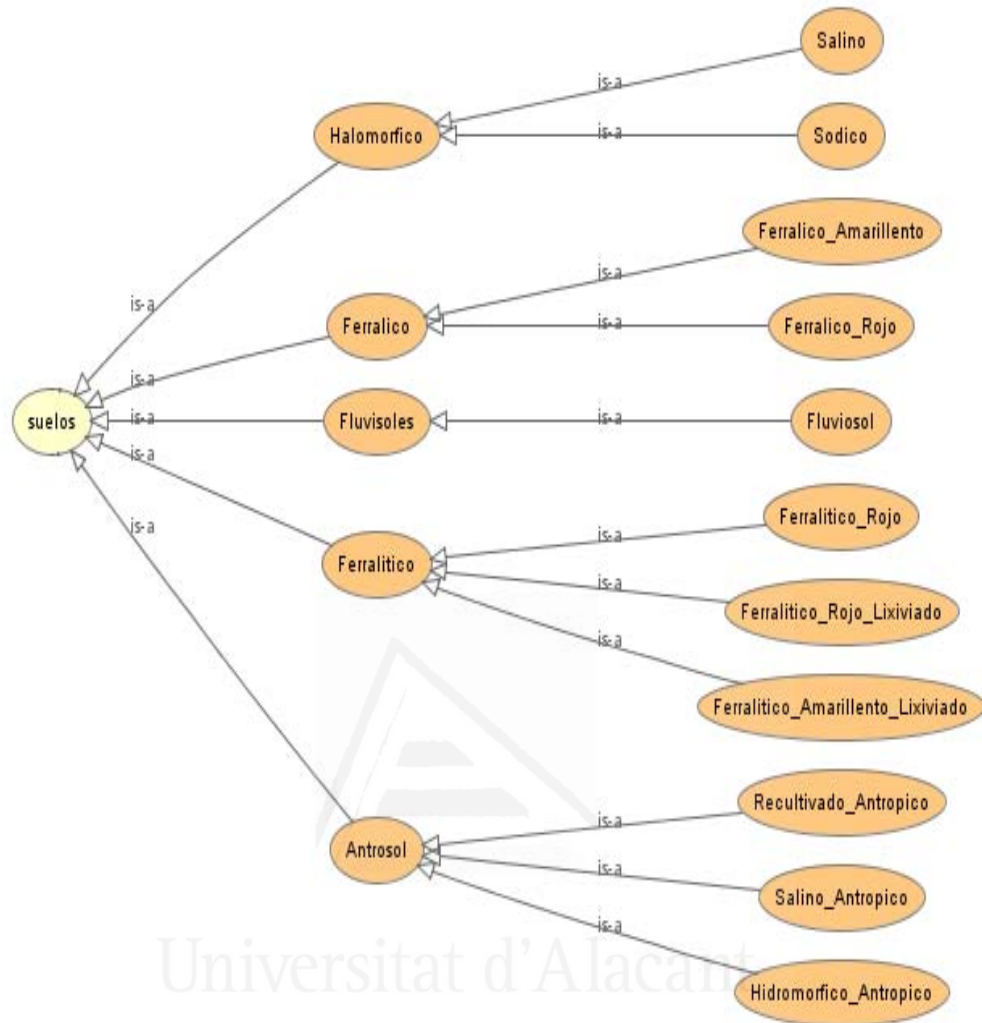


Figura 4.7. Fragmento de la jerarquía de clases que componen la ontología.

Como se puede apreciar en la Figura 4.7, La jerarquía de clases está compuesta por la clase Suelos la cual genera clases de conceptos que agrupan a un conjunto de subclases que representan los Subtipos.

4.3 Conclusiones

En este capítulo se ha descrito el diseño general de la ontología utilizada en esta investigación a través de la metodología GEONTO-MET, la cual está compuesta por cuatro etapas: análisis, síntesis, procesamiento y descripción.

Los conceptos y relaciones de esta ontología se fundamentan en la descripción de los objetos geográficos y las relaciones espaciales que se establecen entre ellos y se contextualizan sobre la base de la Nueva Versión de Clasificación Genética de los Suelos de Cuba. En Cuba los antecedentes el uso del enfoque semántico en la gestión de los datos espaciales no ha sido explotado, sólo algunas técnicas han sido aplicadas indistintamente en diversas instituciones aunque no bajo la denominación de “semántica”, sino como parte de la Geomática, por lo que la contextualización de este dominio constituye uno de los aportes fundamentales de esta investigación (Machado et al., 2014a).

Es importante señalar que la conceptualización representada en la ontología delimita lo que es común a los objetos geoespaciales del contexto de los suelos de Cuba para que pueda ser extensible y aplicada a otros dominios. Una de las bondades de esta conceptualización es que permite clasificar objetos geoespaciales según las relaciones explícitas que los distinguen. Además, los conceptos geográficos que integran la ontología contienen propiedades que describen características implícitas de éstos. Con esta ontología (ver Anexo 6) se pretende que personas no expertas en el dominio geográfico y de los suelos puedan interpretar y entender esta conceptualización en términos generales.

El uso de ontologías posibilita que diversos grupos e instituciones puedan integrar información geoespacial y compartirla entre sí, utilizando un mecanismo de orden semántico para tal efecto. Con esta integración se mejoran los procesos de recuperación semántica y se facilita la búsqueda dentro de fuentes de datos heterogéneas (Machado et al., 2013).



Universitat d'Alacant
Universidad de Alicante

5. Metodología de recuperación de información geoespacial.

En este capítulo se describe la metodología desarrollada en el presente trabajo de investigación. En particular, se representa la metodología a través de una arquitectura genérica para la recuperación semántica de información geográfica desde diferentes fuentes de datos heterogéneas. Inicialmente en la sección 5.1 se ofrece una visión general de la arquitectura y de los distintos subsistemas que la componen. La arquitectura está compuesta por cuatro bloques que agrupan respectivamente los subsistemas que intervienen en el flujo de trabajo para la conceptualización del dominio (sección 5.2), el procesamiento e interacción con el usuario (sección 5.3), la recuperación de información geográfica (sección 5.4) y el ordenamiento de los resultados (sección 5.5). Por último, en la sección 5.6 se resumen las conclusiones más importantes acerca de la propuesta descrita.

5.1 Arquitectura general.

Según el estudio del estado del arte (Capítulo 2) se pudo concluir que uno de los principales objetivos de un sistema de recuperación de información geográfica es resolver consultas que combinan una componente textual y una componente espacial. En ambos casos es importante la interpretación de su semántica. Por tanto, una metodología para la recuperación de la información geográfica debe estar centrada en una estructura que permita un procesamiento semántico para resolver dichas consultas.

La metodología propuesta se describe a través de una arquitectura que integra tres capas: la conceptualización del dominio, que se encarga de la gestión de los datos, el procesamiento multicriterio, que se encarga del procesado de la información y las interfaces de usuarios, que se encarga de la presentación de la información al usuario (Ver Figura 5.1). En la parte inferior de la figura se visualizan los elementos a partir de los cuales se inicia el proceso de conceptualizan; los diferentes medios

de almacenamiento de la información y los especialistas en el campo de la información geográfica.

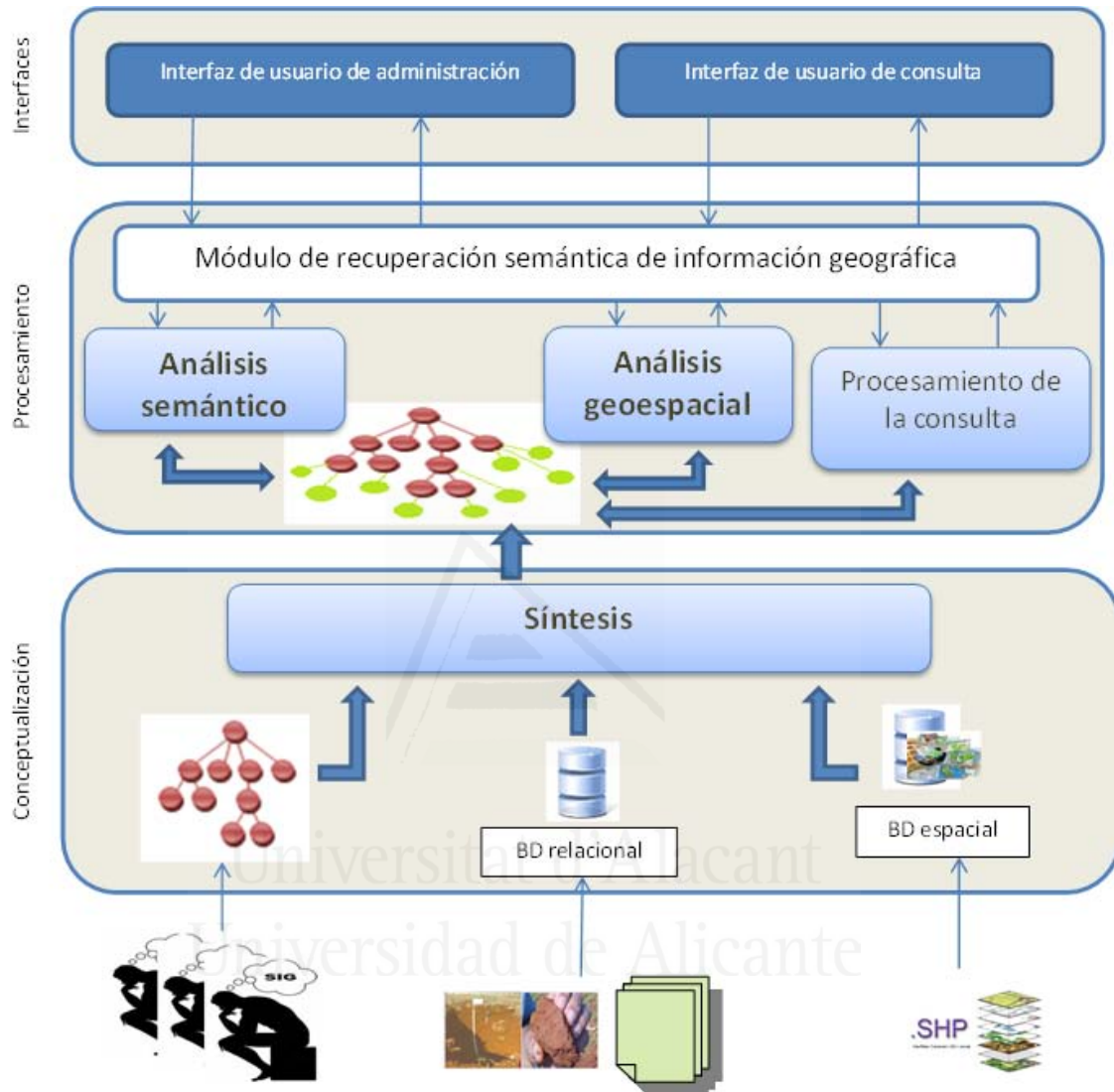


Figura 5.1. Arquitectura general para la recuperación semántica de información geográfica.

En la sección 5.2 se describe la capa de conceptualización del dominio. Esta capa incluye el diseño de la ontología como estructura de indexación, el cual ha sido descrito en el Capítulo 4, mientras que en esta sección se limita a explicar cómo se integra en la metodología. Además, en esta capa se encuentra el proceso de síntesis, en el cual se realiza la instanciación de la ontología desde las diferentes fuentes de datos.

La capa de procesamiento multicriterio se describe en la sección 5.3, dentro de esta capa se encuentran en la mitad izquierda el análisis semántico y el análisis geoespacial. A la derecha se encuentra el procesamiento de la consulta (sección 5.3.1), el cual se encarga de recibir las consultas y convertirlas en una representación semántica equivalente. Todos estos procesos trabajan con la ontología instanciada. Sobre estos procesos se encuentra el módulo de recuperación semántica de información geográfica (sección 5.3.2) que es el encargado de coordinar las tareas anteriores para dar respuesta a las solicitudes de los usuarios.

Finalmente, en la capa superior de la arquitectura se encuentran las interfaces que permiten la interacción de los usuarios (sección 5.4). La interfaz de administración destinada a realizar el mantenimiento de la ontología y del sistema en general y la interfaz de usuario de consulta, destinada a los usuarios para realizar consultas e interactuar con los resultados.

5.2 Conceptualización del dominio.

Una recuperación de información eficiente necesita un modelo con una representación estructurada de la semántica, un modelo en el que se representen semánticamente las propiedades y las relaciones. El objetivo de esta investigación no es describir todo el espacio geográfico (ya que sería una tarea muy costosa en tiempo y esfuerzo) sino sólo para el ámbito determinado en que se va a implementar la metodología. Para que esta aportación sea realmente útil, es importante aclarar que la arquitectura, está diseñada para ser fácilmente adaptable a cualquier dominio de aplicación.

En la Figura 5.1 se presenta como base de la arquitectura la capa de Conceptualización del dominio de trabajo describiéndola a través de una jerarquía de objetos con sus relaciones y características. El diseño de la ontología que se utiliza en esta implementación de la metodología se describe en Capítulo 4. El uso de las ontologías, como mecanismos para representar conocimiento de un dominio concreto, puede ayudar a los sistemas de recuperación a focalizar las consultas de los usuarios y a llegar a sitios donde la comparación sintáctica entre cadenas de caracteres no es suficiente.

Uno de los requisitos de la metodología es que debe ser genérica para soportar la indexación de distintos tipos de documentos. Estos documentos pueden ser diferentes no sólo en cuanto a ser almacenados empleando formatos de archivos diferentes (texto plano, XML, HTML, .jpg, .shp, etc.), sino también en cuanto a sus esquemas de contenido. Un esquema de contenido podría tener un conjunto de atributos de interés para el sistema (por ejemplo, identificador del documento, autor y texto del documento), mientras que otro esquema podría tener un conjunto de atributos diferentes (por ejemplo, identificador de los objetos, propiedades, coordenadas geográficas).

Para poder indexar distintos tipos de documentos se han definido dos abstracciones que se representan como dos conjunto de campos; una abstracción se almacena en una base de datos relacional y la otra en una base de datos espacial. Cada uno de estos campos contiene un valor extraído de los documentos. Cada campo almacenado es indexado en la ontología y puede ser recuperado mediante consultas.

Para soportar diferentes tipos de documentos y diferentes formatos de archivo, la metodología propone la abstracción de documentos como una interfaz de programación que se puede extender mediante implementaciones particulares para diferentes configuraciones de formatos de archivo y esquemas de documento.

En la Síntesis se instancia la ontología con los datos almacenados en la base de datos espacial PostGis y los documentos e imágenes que están almacenados en una base de datos relacional. En este proceso se establece entre diferentes fuentes de datos con características diferentes en cuanto al almacenamiento de la información. Con esto se garantiza que se puedan establecer recuperación sobre datos generados por diferentes especialistas y en diferentes formatos. Esta instanciación de la ontología se realiza de forma semiautomática a través del editor de ontologías Protégé. Los conceptos se describen a través de las propiedades dentro del espacio conceptual. La similitud se calcula a través de la comparación de cada una de las propiedades. Los conceptos se relacionan a través de los arcos en la red.

La integración de datos espaciales en ontologías debe realizarse mediante un proceso previo de anotación de los datos. La forma más simple de integrar datos espaciales a ontologías es almacenando completamente el dato en la ontología, es

decir, los *bytes*¹⁶ que representan al dato espacial estarán embebidos en la estructura ontológica en algún apartado previsto para ello. Pero no basta con embeber el dato íntegramente en la ontología sino que es necesario hacer una correspondencia entre los objetos geográficos presentes en el dato y los conceptos vinculados con ellos.

5.3 Procesamiento multicriterio.

Dentro de la capa de procesamiento se realizan una serie de análisis multicriterios dentro de los que se encuentra el procesamiento de la consulta, el análisis semántico a través de la aplicación de la medida de similitud semántica DIS-C y el análisis espacial mediante la aplicación de medidas de similitud espacial TDD. Todo esto se hace para generar los resultados esperados en el proceso de recuperación de la información. A continuación se presenta una explicación de cada uno de ellos.

5.3.1 Procesamiento de la consulta

El procesamiento de las consultas (ver Figura 5.2) es el componente de la arquitectura que se encarga de analizar las consulta de los usuarios con el objetivo de enriquecerla, convirtiendo la consulta sintáctica en una representación semántica equivalente, a partir de las relaciones entre los conceptos de una ontología y de las reglas que se le aplican.

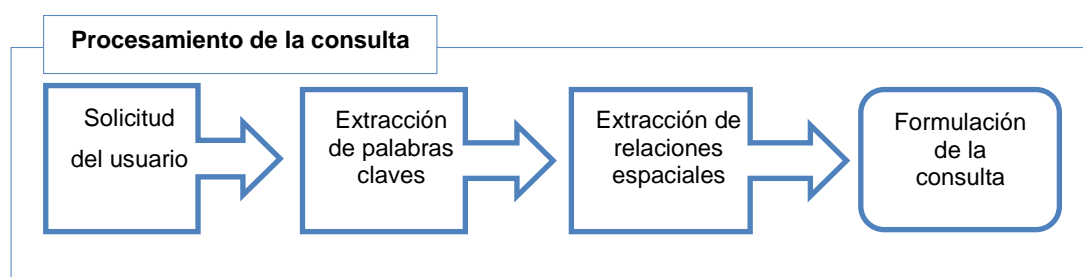


Figura 5.2. Diagrama de la fase de procesamiento de la consulta.

El primer paso que sufre la cadena de búsqueda introducida por el usuario consiste en identificar las entidades y posteriormente simplificar y normalizar las palabras de la consulta, eliminando palabras no relevantes (*stop words*), singularizando las palabras, etc. Para este paso se utiliza el analizador morfológico Freeling (Atserias,

¹⁶ Un byte es una secuencia contigua de bits binarios en un flujo de datos serie.

et al., 2006). Aquí hay una excepción con las palabras de paradas porque las preposiciones son fundamentales para identificar las relaciones espaciales.

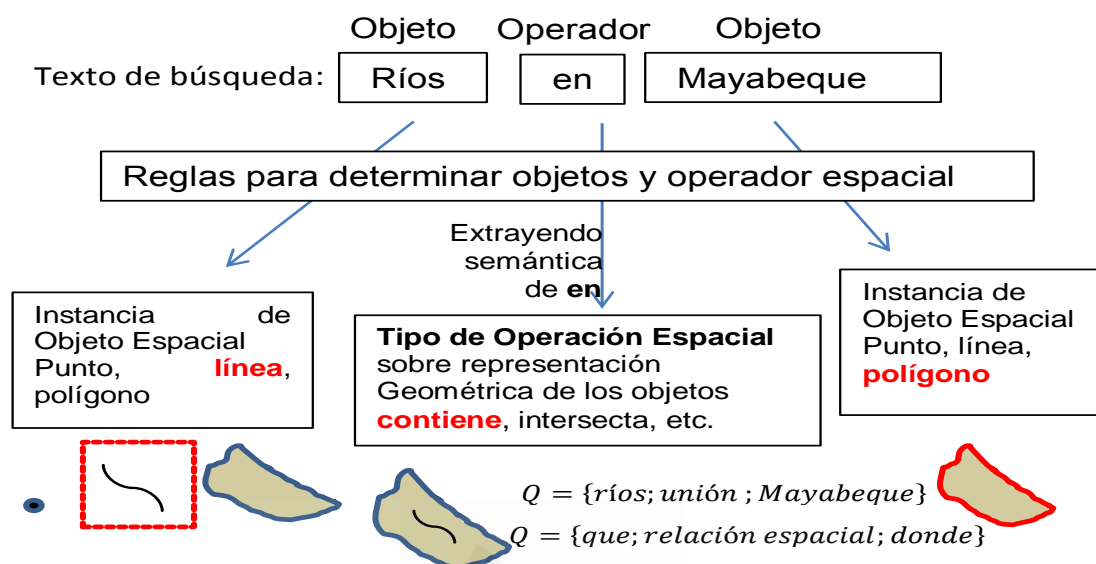


Figura 5.3. Extracción de relaciones espaciales.

El segundo paso tiene por objetivo reconocer los objetos y los operadores espaciales que los relacionan (ver Figura 5.3). Para ello se emplean determinadas reglas en las que se establecen los términos asociados a los operadores espaciales que identifican las relaciones con las que trabaja la metodología. En la Tabla 5.1 se muestran ejemplos de asociación con la relación espacial.

Tabla 5.1. Asociación entre términos identificados y operadores espaciales.

Ejemplo de uso de la relación espacial	Relación espacial	Operación o relación espacial asociada
Jaruco está «junto» a Santa Cruz	Junto	Conecta
La presa Mampostón está «en» San José	En	Contenido_en
El río Mayabeque «intersecta» el municipio Guines	Intersecta	Intersección
El río Jaruco «cruza» el municipio Jaruco	Cruza	Intersección

Finalmente, las consultas se transforman en tripletas $\{\text{que, relación espacial, donde}\}$ donde *que* y *donde* identifican objetos que se encuentran relacionados. También se lanzan consultas simples donde solo se realiza la búsqueda por los conceptos.

Con estos términos se activan en la ontología los conceptos subyacentes. Para esto cada concepto en la ontología cuenta con un conjunto de sinónimos, cada uno de los términos relevantes de la consulta se localizan en la ontología, utilizando el conjunto de sinónimos.

5.3.2 Recuperación de objetos geográficos.

Para llevar a cabo la recuperación de la información geográfica se propone la API (en inglés *Application Program Interface*) de Jena¹⁷. Esta API es una estructura de Java que se utiliza para construir aplicaciones de la Web Semántica. A su vez, proporciona un ambiente integrado de programación para RDF, RDFS, OWL y SPARQL. Además, incluye un motor de inferencia basado en reglas.

Esta API es utilizada para extraer instancias de conceptos de la ontología. La implementación de esta ontología fue realizada en OWL; por lo cual, se genera un modelo persistente de esta conceptualización para establecer el vínculo entre los dos componentes.

La recuperación de las instancias se ha implementado con SPARQL, el cual es un lenguaje de consulta que ofrece un protocolo de acceso a metadatos de una estructura OWL o RDF. Con el puente que se genera entre Jena y Protégé se lleva a cabo la recuperación de las instancias de los conceptos geográficos, donde se extrae la siguiente información de la ontología en OWL:

- Recurso: todo aquello que se puede describir por una expresión RDF.
- Propiedad: una característica, un atributo o relación utilizada para describir un recurso.
- Literal: un tipo de dato simple.
- Cadena: un recurso junto con una propiedad y con un valor asociado.

En la Figura 5.4 se muestra el diagrama general del proceso utilizado para la recuperación de instancias de conceptos geográficos y la representación espacial de los elementos recuperados por la consulta a través de Geoserver.

¹⁷ <http://jena.apache.org/index.html>

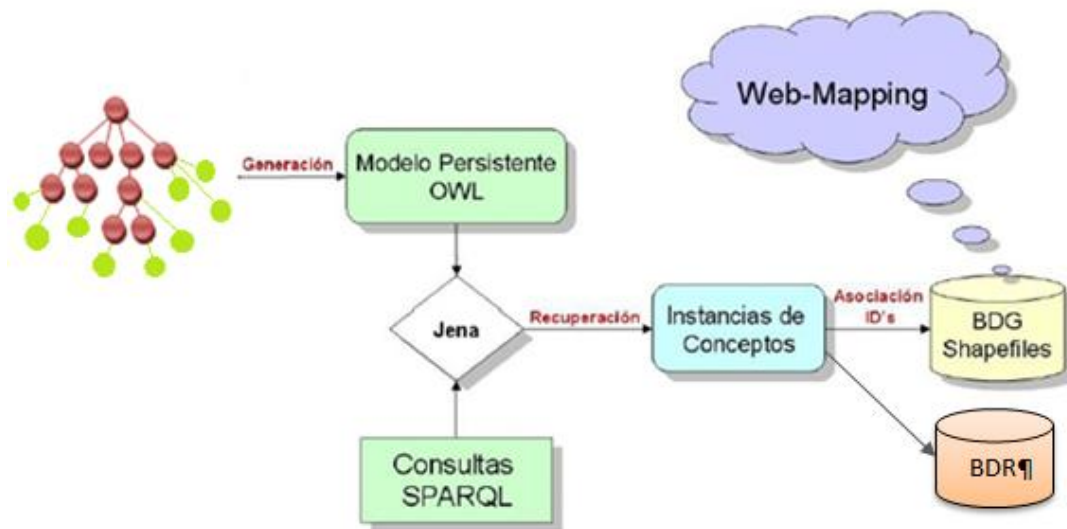


Figura 5.4. Diagrama general de la recuperación de instancias de conceptos geográficos y su representación en el *web-mapping*.

De acuerdo con la Figura 5.4, el proceso inicia con la transformación de la ontología de dominio geográfico a un modelo persistente en OWL. Esta tarea se realiza mediante el uso del editor Protégé y consiste en transformar las clases de entidades abstractas y los conceptos que heredan de estas clases a un formato tabular, en el cual por medio de SPARQL se puede tener acceso a todos los elementos de la ontología.

Mediante la interfaz de Jena, es posible acceder al modelo persistente para obtener los recursos OWL. Para esto se crea un modelo vacío para definir los recursos que serán almacenados. Este modelo que se crea es una extensión del OWL, el cual proporciona capacidades adicionales para manejar los datos de la ontología. Posteriormente, se carga en memoria el esquema definido utilizando el operador nativo de Jena.

A través de las consultas SPARQL se establece la conexión directa al modelo persistente identificando los conceptos de interés en la recuperación. La ventaja de este mecanismo de recuperación de datos es que el modelo de consulta es orientado a datos, por lo cual a través del significado de un vocabulario y representado en un lenguaje, es posible extraer semánticamente instancias de conceptos relacionados al contexto de la ontología.

Se dan 3 posibles clasificaciones para las coincidencias obtenidas como resultados del algoritmo en esta fase:

- atómica (encuentra solo uno de los elementos que conforman la tripleta).
- parcial (encuentra parejas de los elementos que conforman la tripleta).
- completa (encuentra todos los elementos que conforman a una consulta de tipo tripleta).

5.3.2.1 Análisis semántico.

Posteriormente se activan los términos similares a los términos de búsqueda. Para identificar los conceptos similares en la ontología se utiliza el algoritmo de DIS-C (sección 3.2).

El algoritmo DIS-C funciona con todo tipo de ontologías porque no depende de las relaciones taxonómicas. La idea principal es determinar el valor de distancia conceptual a cada uno de los tipos de relaciones en la ontología. Esto se logra al convertir esta última en un grafo dirigido y ponderado en el que cada concepto se convierte en un nodo y cada relación se convierte en un par de aristas. Esto es aplicable de manera directa a las relaciones binarias, pero también es posible aplicarlo a relaciones n-arias. Una vez que se tiene el grafo, se aplican técnicas de teoría de grafos. El paso natural consiste en calcular las distancias mínimas, con lo que se puede encontrar la distancia entre conceptos que no se encuentran directamente relacionados.

Una vez ejecutadas las consultas se recurre al algoritmo DIS-C y se determinan los conceptos más cercanos al deseado. Ejemplo de “ríos”, con “cuerpos de agua” como resultado más alejado y “lagos” como más próximo.

5.3.2.2 Análisis espacial.

Con los objetos espaciales recuperados se realiza un análisis de los mismos. Para esto a cada objeto le corresponde una cadena de texto estándar, la cual es propuesta por la OGC (Open Geospatial Consortium - por sus siglas en inglés) para el intercambio de información geométrica denominada WKT (Well Known Text – por sus siglas en inglés). Esta cadena se encarga de enumerar todos los nodos y

vértices que conforman un objeto espacial. Esta cadena enumera todos los nodos y vértices que conforman un objeto espacial.

Las cadenas de texto WKT contienen un tipo de dato “primitivo” en las bases de datos extendidas espacialmente, como es el caso de PostGIS (Postgress + Extensiones GIS), por lo que se pueden almacenar en un campo dentro de cualquier tabla de la base de datos. Mediante estas cadenas se crean y representan los siguientes tipos de geometría, los cuales se asocian con los nombres considerados por la API JTS en la Tabla 5.2

Tabla 5.2. Clases de características simples soportadas por la API JTS, de acuerdo con la especificación de elementos simples de la OGC

Tipo	Descripción
<i>Point</i>	Un punto es implementado como una coordenada simple.
<i>LineString</i>	Es implementado como un arreglo de coordenadas.
<i>LinearRing</i>	Contiene n coordenadas y es implementado con un arreglo de coordenadas que contiene $n+1$ “ <i>Points</i> ” y $coord[0] = coord[n]$.
<i>Polygon</i>	Es implementado como un “ <i>LinearRing</i> ” simple para el borde y un arreglo de “ <i>LinearRings</i> ” para los hoyos. El borde tiene orientación en el sentido de las manecillas del reloj y los hoyos el sentido inverso.
<i>MultiPoint</i>	Hereda la implementación de “ <i>GeometryCollection</i> ” pero contiene solo “ <i>Point</i> ”.
<i>MultiLineString</i>	Hereda la implementación de “ <i>GeometryCollection</i> ” pero contiene solo “ <i>LineStrings</i> ”.
<i>MultiPolygon</i>	Hereda la implementación de “ <i>GeometryCollection</i> ” pero contiene solo “ <i>Polygons</i> ”.
<i>Geometry</i>	Es una clase abstracta a la que pertenecen todas las clases de objetos mencionados; es decir, todos son una geometría.
<i>GeometryCollection</i>	Es implementado como un arreglo de objetos “ <i>Geometry</i> ”. La

	dimensión de un “ <i>GeometryCollection</i> ” heterogéneo es la dimensión máxima de sus elementos.
--	--

Una vez que se tiene la cadena WKT, la cual se recupera según la siguiente instrucción:

```
Geometry Ob1 = new WKTRReader().read("POLYGON(120 0, 120 10, 160 10, 160 30, 250 30, 250 0, 120 0)");
```

Después de haber recuperado los objetos geográficos, se puede invocar a todos los métodos de análisis topológico y geométrico que soporta JTS, con el fin de calcular la similitud espacial implementando el modelo de TDD (ver sección 3.3).

Si se considera cada tripleta como una escena, entonces se tiene la tripleta forma de la consulta del usuario y las recuperadas. El modelo TDD se implementa como una extensión del Plano de Coordenadas Paralelas (PCP-por sus siglas en inglés) (Inselberg, 1985). PCP tiene ejes paralelos los cuales representan los elementos a analizar de cada objeto. Cada objeto espacial en una escena es representado como una línea que atraviesa los ejes. Una ventaja del PCP es que puede visualizar varios atributos simultáneamente. Cuanto más similares son dos objetos más cerca se representan sus líneas en el PCP. Generalmente, PCP se utiliza para visualizar variables numéricas y sus relaciones. Sin embargo, en el análisis de la similitud espacial en vez de variables numéricas se utilizan variables cualitativas, como las relaciones topológicas, de dirección y de distancia. Además, se considera que todos los objetos son alineables, es decir, que para cada eje todos los objetos tienen valores mediante los cuales son comparados.

Sin embargo, en la evaluación de la similitud espacial se pueden presentar situaciones en las que los objetos no se pueden alinear en cuanto a sus valores. Por ejemplo, en el caso en que dos escenas espaciales tengan la misma relación espacial pero diferentes tipos de objetos espaciales. Una escena puede contener dos regiones disjuntas (polígonos) mientras la otra contiene dos ríos que se intersectan (líneas). Como las regiones y los ríos tienen diferentes atributos pueden ser comparados en cuanto a las relaciones espaciales pero no en cuanto a los atributos.

En esta implementación, PCP se utiliza solo para comparar las relaciones espaciales, los atributos se comparan antes de acuerdo a las descripciones semánticas de la ontología. Se trabaja con 3 ejes para manejar las relaciones. Estos representan las relaciones topológicas, de dirección y de distancia. Los objetos espaciales se representan como líneas a través del panel de visualización como se muestra en la Figura 5.5, donde cada línea representa las relaciones espaciales de los objetos identificados. La similitud entre ellos se observa a través de la distancia entre las líneas que los representan.

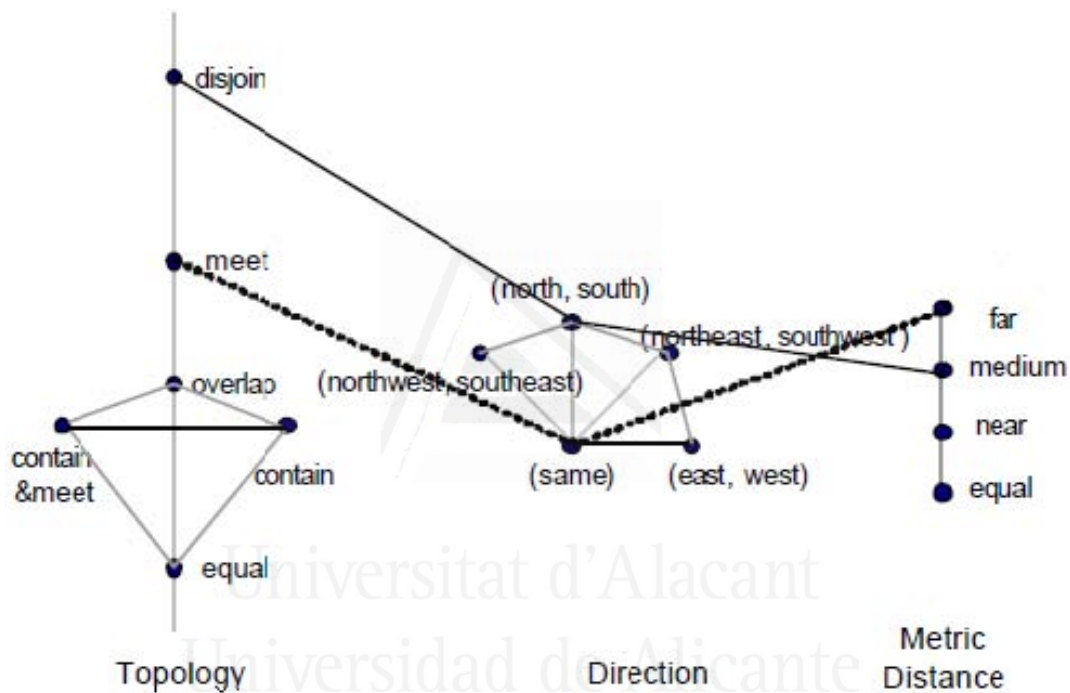


Figura 5.5. Detalles de la implementación del modelo TDD.

Para visualizar las relaciones espaciales cualitativas, el eje vertical del PCP tradicional se reemplaza en esta implementación por una red del vecindario conceptual. En el eje de las relaciones topológicas, se representan seis nodos para los diferentes tipos de relaciones topológicas: igual (*equal*), separado (*disjoin*), adyacentes (*meet*), superposición (*overlap*), contiene (*contain*) y cubre (*contain&meet*). Las longitudes de los arcos en el eje entre los nodos reflejan el peso definido en la red topológica. La relación entre el peso de cada arco y la longitud métrica de cada arco en el eje se describe como:

$$\frac{\text{peso 1}}{\text{peso 2}} = \frac{\text{longitud del arco 1}}{\text{longitud del arco 2}} \quad (5.1)$$

Una excepción es el arco de contiene a *cubre* en este caso el peso es 1, de *superposición* a los otros nodos es 2 y de *igual* a otros nodos es 3. Sin embargo, esta relación de peso no puede ser geoméricamente representada debido a la regla que, en un triángulo, la suma de los dos arcos es mayor que el tercer arco.

El análisis de las relaciones topológicas utiliza los métodos que soporta la API JTS a partir de la especificación de la OGC del DE-9IN (*Dimensionally-Extended 9 Intersection Matrix* –por sus siglas en inglés). Las relaciones soportadas se describen en la Tabla 5.3.

Tabla 5.3. Relaciones topológicas soportadas por la API JTS.

Método	Descripción
igual (equal),	La geometría es topológicamente equivalente.
separados (disjoin),	La geometría no tiene ningún punto en común.
adyacentes (meet),	La geometría tiene al menos un punto en común (el inverso de separado)
superpone (overlap)	La geometría comparte solo algunos puntos en común y la intersección tiene la misma dimensión que las geometrías mismas.
contiene (contain)	La geometría B se encuentra en el interior de la geometría A.
intersecta (intersect)	La geometría tiene al menos un punto de borde en común, pero no tiene puntos interiores.

De esta forma se calculan todas las relaciones y se establece cuál de ellas se cumple para dos geometría de objetos espaciales cualesquiera.

En el eje de la relación de dirección las cinco relaciones son *iguales*, *este*, *oeste*, *noreste*, *sureste*, *norte*, *sur*, *noroeste* y/o *suroeste*. El peso puesto para cada arco es 2. La figura formada por las cuatro líneas y los cinco puntos representa la relación direccional definida. Es posible establecer dichas relaciones con JTS porque se pueden considerar las coordenadas (x,y) del centroide de un objeto espacial

directamente o de su polígono convexo. Después de comparar dichas coordenadas, y asumiendo que la región espacial en estudio está al norte orientado hacia “arriba”, es fácil comparar la coordenada X de dos objetos geográficos y determinar la que tenga un valor mayor (más a la derecha) estará “al este” de dicho objeto (qué entonces está al oeste).

De la misma forma, se pueden comparar las coordenadas Y de dos objetos geográficos y establecer el que tenga la coordenada con mayor valor (hacia arriba) entonces estará la norte del otro objeto (que se encuentra al sur).

El eje de la relación de distancia representa geométricamente la distancia métrica con una línea y cuatro puntos que representan las cuatro relaciones de distancia: *igual, cerca, medio y lejos*.

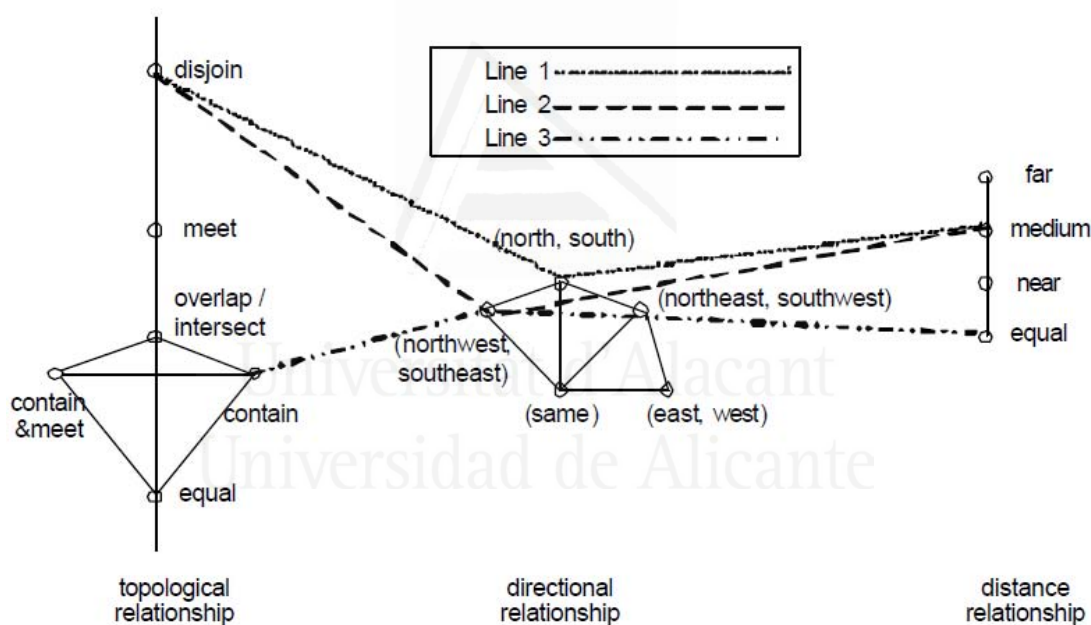


Figura 5.6. Implementación de las relaciones espaciales en PCP.

Cuando se comparan objetos espaciales, la similitud entre ellos se observa a través de la distancia entre las líneas que los representan. En la Figura 5.6. se observa que la *line 2* es más similar a la *line 1* que a la *line 3*. Como el peso está ubicado en las relaciones topológicas, direccionales y de distancia métrica y la longitud de cada arco coincide con este peso la visualización de la similitud espacial es consecuente con la literatura (Renz, 2002).

Las imágenes recuperadas están relacionadas semánticamente con los conceptos de la base de conocimientos y en los mapas se detalla su ubicación espacial. Al usuario se le muestran las imágenes del suelo que buscaba, o alguno con características similares, con su respectiva ubicación en el mapa. Si desea ver la imagen ampliada, la selecciona y además se le brinda la descripción de la imagen. Si existen documentos asociados a estos conceptos se recuperan de igual forma.

5.4 Interfaces de usuarios

La interfaz de usuario es un componente que no puede faltar en los sistemas de recuperación de información. Como metodología genérica se le dejan los detalles de implementación a los desarrolladores. Solo tiene como requerimientos el permitirle a los usuarios que introducir las consultas y los componentes de visualización de los mapas.

5.4.1 Visualización de los resultados

Aunque resolver correctamente las consultas es el objetivo principal de todo sistema de recuperación de información, presentar de manera amigable para el usuario los resultados también es una cualidad muy importante. Esta metodología contempla la inclusión de un *Web Map Service* (WMS). Este componente tiene dos objetivos principales. El primer objetivo es proporcionar los mapas para la interfaz de usuario. Esto permite que el usuario pueda navegar por la cartografía que proporciona el sistema. El segundo objetivo es la creación de representaciones cartográficas de los resultados de las consultas. A diferencia de los sistemas de recuperación de información tradicionales, en este tipo de sistemas los documentos son relevantes para una consulta porque están geo-referenciados en un lugar determinado. Por tanto, realizar la representación cartográfica de los lugares donde están geo-referenciados los documentos o imágenes relevantes aporta un valor añadido a estos sistemas.

El servicio WMS es un estándar definido por el OGC (OGC, 2003) para proporcionar mapas con datos geo-referenciados. Este servicio define un protocolo sencillo sobre HTTP para la realización de peticiones y la obtención de los mapas. De forma simplificada, en las peticiones se especifica la capa, o capas, en las que se

encuentra la información geográfica y el área de interés; y la respuesta contiene una imagen, o un conjunto de imágenes, que se pueden visualizar en cualquier navegador web.

Este servicio es posiblemente uno de los más utilizados en el campo de los sistemas de información geográfica. Por este motivo, existen una gran cantidad de implementaciones, tanto de pago como en software libre, que proporcionan este servicio. Además, también existen muchos clientes que permiten incluir mapas en un entorno web de forma sencilla. Estos mapas se generan mediante consultas a servicios WMS. En esta tesis se utiliza Geoserver (Geoserver Community, 2013). Geoserver es uno de los WMS de referencia dentro del mundo del software libre. Está escrito en Java y tiene una comunidad de desarrolladores muy activa.

5.5 Conclusiones

En este capítulo se ha presentado la metodología de recuperación semántica propuesta en esta investigación. En la sección 5.1 se muestra una visión general de la propuesta y se presenta la arquitectura en la cual se basa la metodología, la cual está compuesta por tres capas: conceptualización, procesamiento e interfaces. En primer lugar, en la capa de conceptualización (ver sección 5.2) se incluyen los componentes involucrados en el flujo de trabajo que permite indexar los documentos, el diseño de la ontología y la instanciación o síntesis de la misma desde diferentes fuentes de datos. Se utilizan dos fuentes de datos, una para los datos geográficos clásicos y otra para todo tipo de documento que pueda contener dentro algún tipo de información geo-referenciada.

En segundo lugar, la capa de procesamiento (ver sección 5.3) agrupa una serie de procesos independientes que se pueden emplear desde otros componentes de la arquitectura. Dentro de estos componentes se encuentran el procesamiento de la consulta (sección 5.3.1), y la aplicación de las medidas de similitud y espacial. La última capa es la de interfaces de usuario (ver sección 5.5) donde se definen las dos interfaces que permiten la interacción con el sistema tanto para su administración como para la realización de consultas.

6. Resultados experimentales.

En este capítulo se presentan los resultados obtenidos con la implantación de la metodología propuesta aplicada a un caso de estudio. Se muestra el proceso de recuperación a partir de varias consultas y la secuencia de pasos aplicados de acuerdo a la metodología. Para la validación de la metodología se implementó un prototipo de sistema que implementa las funcionalidades necesarias para la recuperación semántica de información geográfica desde fuentes de datos heterogéneas (Machado et al., 2013). Se muestran ejemplos de consultas que incluyen relaciones espaciales, destacándose que para cada consulta se puede recuperar distintas opciones como objetos geoespaciales, imágenes que contengan los conceptos solicitados en la consulta, algunas de las propiedades de los objetos o la relación espacial indicada en la consulta. Para evaluar la metodología de recuperación de datos geoespaciales se aplica una estrategia de triangulación de las búsquedas.

6.1 Discusión sobre los principales componentes.

Los datos con los que se han trabajado proceden de los mapas de suelos de la región de San José de Las Lajas en la provincia de Mayabeque (ver Anexo 1), el mapa de la propia provincia de Mayabeque (ver Anexo 2) y el de la provincia de Pinar del Río, región occidental de Cuba (ver Anexo 3). También existen imágenes que visualizan las características de estos suelos.

El suelo se selecciona como caso de estudio debido a que es el elemento fundamental de la superficie de la Tierra que sirve de sustento a la vida, su uso racional y sostenible en la agricultura es una cuestión de importancia estratégica por lo que su zonificación agroecológica constituye una herramienta de suma utilidad para la planificación adecuada de las labores agrícolas a realizar. La aplicación de fertilizantes requiere de este instrumento para una correcta aplicación de las dosis que realmente necesita el suelo sin detrimento de sus propiedades físicas y químicas a la par de lograr el necesario aumento de su fertilidad.

Para el caso específico de los mapas se encuentran almacenados en el formato vectorial *ESRI Shapefile*¹⁸. El *ESRI Shapefile* es un formato vectorial que almacena la ubicación de los objetos geográficos y los atributos asociados a ellos pero carece de la capacidad para el almacenamiento de la semántica de esta información. De manera general un *Shapefile* está compuesto por varios archivos, pero deben existir al menos los tres principales. Estos tienen las siguientes extensiones:

- “*.shp”: Este es el archivo principal, en él se almacena las entidades geométricas de los objetos (Componente Espacial).
- “*.dbf”: Este es el archivo que almacena la información de los atributos de los objetos (Componente Temática).
- “*.shx”: Este es el archivo que almacena el índice de las entidades geométricas actuando como vínculo entre la componente espacial y la temática para cada uno de los datos almacenados.

Para conceptualizar y describir el dominio geográfico basándose en la metodología propuesta se utiliza la estructura lógica de la ontología propuesta en el Capítulo 4. La población de la ontología se utiliza para determinados contextos. A continuación se presentan un escenario, como caso de prueba, sobre el que se aplica la metodología propuesta

Escenario: Se tiene una región espacial del municipio San José de Las Lajas, que se encuentra en formato ESRI Shapefile en una base de datos geográfica. Esta región contiene diferentes tipos de temáticas, las cuales están compuestas por diferentes capas; en donde cada capa contiene objetos geográficos representados por primitivas geoespaciales del modelo vectorial. Se puede observar en la Figura 6.1 que la región está compuesta por polígonos de tipos de suelos (Ps), por perfiles (Pf) y por elementos hidrológicos (Hd). Cada capa temática se representa por una leyenda, la cual por medio de símbolos describe cada uno de los objetos geográficos del mundo real. La región está compuesta por 21 objetos de tipo polígono, 1 objeto de tipo línea y 19 de tipo punto.

¹⁸ El format ESRI Shapefile es un formato propietario abierto de datos geográficos desarrollado por la compañía ESRI, quien desarrolla y comercializa software para Sistemas de Información Geográfica. Originalmente se creó para la utilización con ArcView GIS, pero actualmente se ha convertido en formato estándar por la importancia que los productos ESRI tienen en el mercado SIG.

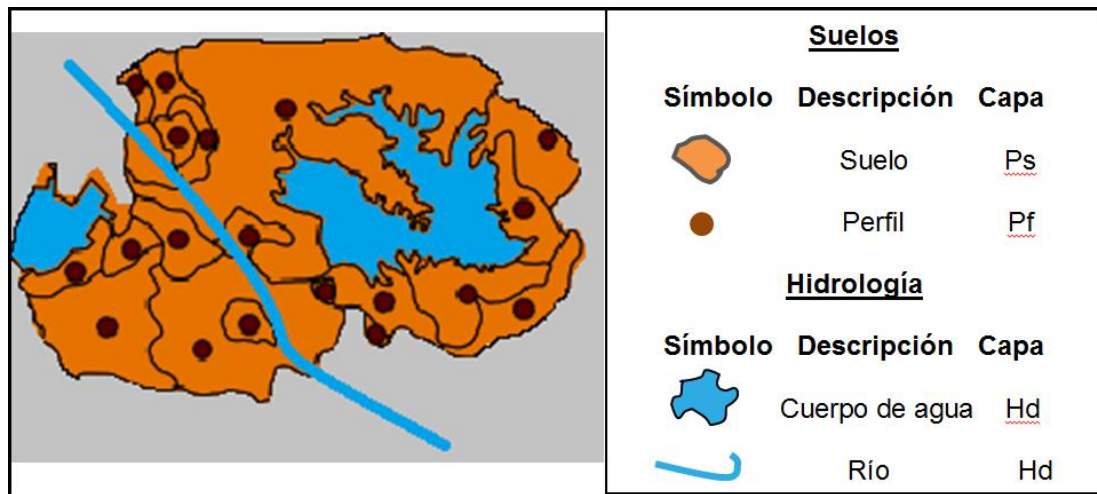


Figura 6.1. Región espacial que describe el escenario de trabajo.

Además, se cuenta con 26 imágenes de los diferentes tipos de suelos que están representados en esos polígonos. Estas imágenes son utilizadas en la carrera de Ingeniería Agronómica, en la cual un alto porcentaje de sus asignaturas emplea imágenes como apoyo a los diferentes procesos de identificación y diagnóstico que se estudian en esta carrera. Como parte de esta investigación se diseñó una herramienta (Machado and Montoyo, 2007) que extrae dichas imágenes de la base de datos y las convierte en objetos de aprendizaje (Machado et al., 2014d) catalogados según el estándar IMS-CP.

De los 19 polígonos de suelos representados en el mapa nueve son Ferralíticos rojos, cinco son Pardos, dos son Fersialíticos, uno es Aluvial y tres son Húmicos. Las imágenes son 26 y se corresponden con los siguientes tipos: dos de Alíticos, dos de Ferríticos rojos, cinco de Ferralítico rojo, una de Ferrálico rojo, una de Fersialítico, cuatro de Pardos, una de Húmico, una de Vertisol, dos de Salino, dos de Sódicos, una de Histosol y cuatro de Gley nodular. A continuación se explica la forma en que se realiza el mapeo del escenario en la ontología definida en el Capítulo 4.

6.2 Ontología espacial como estructura de indexación.

En el diseño de una estructura de indexación en el ámbito de los sistemas de recuperación de información geográfica existen dos requisitos fundamentales que se deben cumplir (Seco, 2009). En primer lugar, la estructura debe proporcionar

una descripción apropiada de la naturaleza jerárquica del espacio geográfico para permitir la asociación de información con los distintos elementos singulares que forman la jerarquía. En segundo lugar, la estructura debe resolver de forma eficiente consultas textuales puras, consultas espaciales puras y consultas que especifiquen tanto información espacial como información textual.

A continuación se explica la forma en que se realiza el mapeo entre el escenario de trabajo a la ontología. En este trabajo se describe una ontología que conceptualiza las características geográficas de los suelos de la zona occidental de Cuba siguiendo la Nueva Versión de Clasificación Genética de los Suelos. La ontología propuesta en el Capítulo 4 modela todo este espacio conceptual, por lo que permite instanciar todos los objetos que se encuentran dentro de los mapas utilizados así como las relaciones que existen entre ellos. Además, tiene instancias de los documentos y archivos con extensión “.jpg” que se encuentran almacenados en una base de datos relacional.

Cada una de las muestras de las clases a partir de las cuales son clasificados estos datos están representadas por cada uno de los conceptos presentes en la ontología del dominio de los suelos, por ejemplo, el concepto Suelos en la ontología representa la clase Suelos y es a su vez la única muestra existente de esa clase y contiene las siguientes características:

- Clase Suelos: Id, Nombre, Grupo, Tipo, Textura, Erosión, Saturación, Humificación, Graviliosidad, Pedregosidad, Rocosidad, Acidez y Salinidad.

Una vez procesado los datos se procede a la vinculación de estos con la ontología. Este vínculo complementa y enriquece, al verse como un todo la información semántica existentes. Por tanto es necesario extraer los vectores de características comunes para cada dato y cada clase. Si se encuentran instancias de conceptos en la ontología, se buscan las propiedades de éstas. Estas propiedades están asociadas a las instancias, por lo que el mapeo de éstas llena los atributos del esquema conceptual.

Una vez que todas las instancias han sido definidas en el esquema conceptual, es necesario seleccionar un par de instancias que representan objetos geográficos, con el objetivo de verificar si la relación entre ellos existe; entonces si se cumple esta condición una parte de la descripción para este par de instancias se genera.

Por ejemplo, de acuerdo con la Figura 6.1, la cual está compuesta de varios objetos geográficos, en este caso los objetos en la capa reflejan una relación de existencia “es”: “*R1 es ObjetoLineal*”. Además, el concepto relación “*Intersecta*” está vinculado con *R1* y varios de los polígonos de suelos; donde *R1* es un objeto lineal e instancia del concepto “río”. A su vez, la relación “*Intersecta*” esta generalizada como una relación topológica, que dentro de la ontología hereda de la clase “*relación_geográfica*”.

Para determinar las relaciones entre los objetos geográficos, se calculan los WKT de cada una de las instancias que tienen una representación espacial en la base de datos geográfica. Este análisis se realiza utilizando la API JTS, donde el objetivo esencial es verificar las relaciones topológicas y geométricas.

De esta forma se instancian tanto las imágenes provenientes de la base de datos relacional como los objetos geográficos de la base de datos geográfica. La clasificación de las capas de datos fue realizada y todos los objetos pertenecientes a cada una de las capas temáticas fueron correctamente clasificados. Este experimento demuestra cómo es posible realizar la representación semántica de objetos geoespaciales.

6.3 Recuperación semántica de información geoespacial.

En esta sección se describen algunos resultados obtenidos con respecto a la recuperación de información geográfica utilizando la metodología de recuperación propuesta (Machado et al., 2014b). Se implementó un mecanismo de consulta (recuperación de instancias), con el objetivo de acceder a una ontología OWL, a través de Jena.

Se implementó una plataforma de experimentación¹⁹ (ver Anexo 4) con las funcionalidades básicas existentes en los SIG convencionales accesibles a través de la barra de herramientas, véase Figura 6.2, estas son:

- Carga y visualización de capas de datos.

¹⁹ <http://sigsem.unah.edu.cu/>

- Interacción con el mapa (Paneo, Zoom, Restablecer).
- Consulta a los datos.

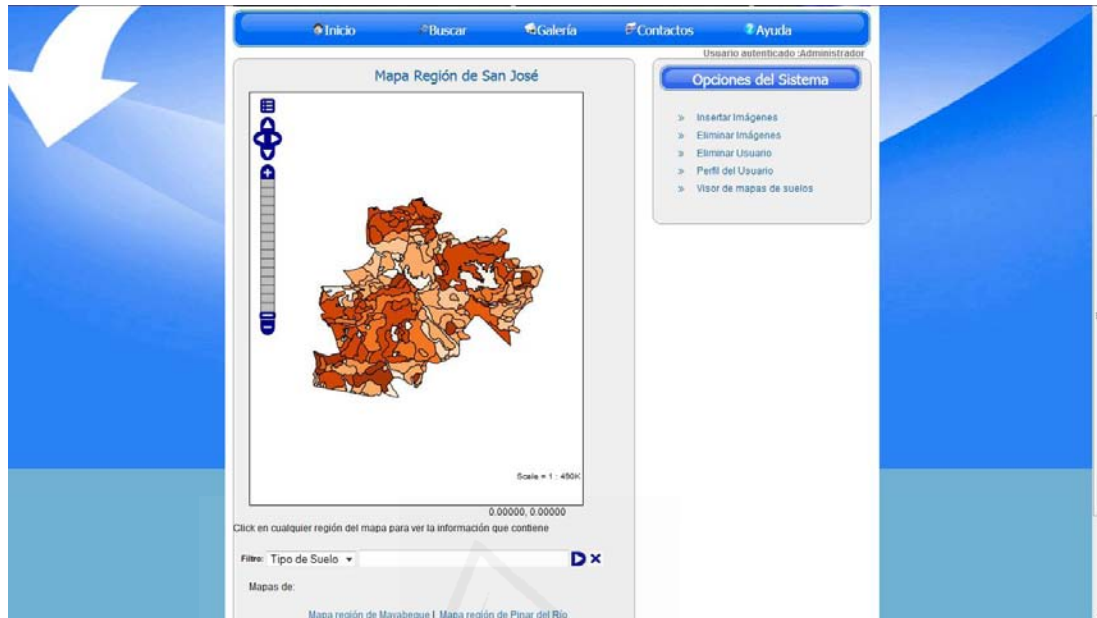


Figura 6.2. Visualización de las capas de datos

La plataforma también posee un visor de mapas en el cual se visualizan los datos cargados a través de su componente espacial. El sistema de gestión resultante permite realizar búsquedas de información de los diferentes tipos de suelos. Al seleccionar la opción buscar, por ejemplo, si se lanza una búsqueda de determinado tipo de suelo “ferralítico rojo”, se recupera un conjunto de imágenes que se relacionan con la búsqueda y un mapa de la región en el cual se visualiza la localización de este tipo de suelo. En la Figura 6.3 se muestra la representación visual de los objetos recuperados.

Además, de mostrarse todas las imágenes relacionadas con la solicitud si el usuario lo desea al dar clic en alguna de estas imágenes esta se mostrará con una vista más amplia, mostrando la descripción de la misma y los conceptos de la ontología que tienen relación semántica con la misma.

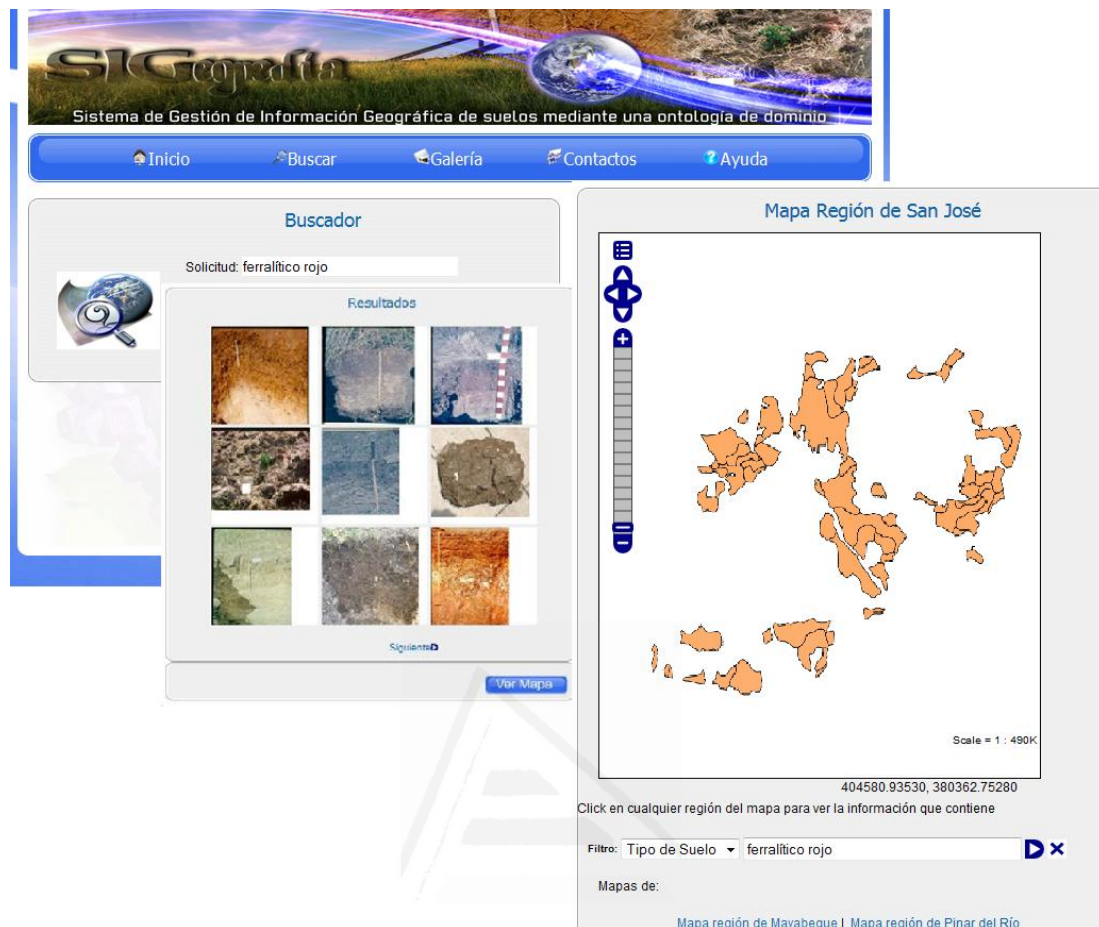


Figura 6.3. Mapas con las regiones recuperadas según una consulta determinada.

En la Tabla 6.1 se presentan ejemplo de como el sistema una vez que procesa la solicitud del usuario determina a través de cual operador, el cual identifica a la relación espacial, se deben relacionar los objetos que se están buscando.

Tabla 6.1. Términos o palabras asociadas a relaciones espaciales.

Ejemplo de uso de la relación espacial	Relación espacial	Operación o relación espacial asociada
Jaruco está «junto» a Santa Cruz	Junto	Conecta
El río Mayabeque «intersecta» el municipio Guines	Intersecta	Intersección

El río Jaruco «cruza» el municipio Jaruco	Cruza	Intersección
---	-------	--------------

Con la implementación de la metodología propuesta se logra un ahorro de tiempo en la realización de los análisis espaciales. Esto es debido a que se eliminan una serie de pasos que conllevan a un gasto de tiempo. Además, se le simplifica el trabajo al usuario y se evita que cometa errores en la recuperación ya que:

- La búsqueda de las capas de datos con las que se desea trabajar se encuentran conceptualizadas en la ontología generada.
- No tiene que generar una consulta en lenguaje SQL, porque puede plantear su búsqueda en lenguaje natural.

Ejemplos de consultas que el sistema automáticamente no podría dar respuestas utilizando los métodos convencionales serían las siguientes:

“¿Qué ríos se encuentran en el municipio San José?”

“¿Cuáles son las regiones de suelos que rodean la presa Mampostón?”

Este tipo de consulta no puede ser obtenida a través de lenguaje SQL ya que tanto en la base de datos como en el archivo *ShapeFile* no existe ninguna tabla nombrada con los términos utilizados en la consulta. Por tanto, a través del uso de la ontología el sistema puede entender fácilmente lo que se desea recuperar. Además, las recuperaciones se realizan sin la necesidad de la construcción de una consulta en lenguaje SQL por lo que es posible expandir su utilización a usuarios no expertos.

Sea la cadena de texto introducida por el usuario:

$Qstr = \text{“¿Qué ríos se encuentran en el municipio San José?”}$

El primer paso que sufre la cadena de búsqueda introducida por el usuario tiene como objetivo identificar las entidades, para esto existe un fichero de entidades identificadas, el cual se actualiza con los objetos mapeados en la ontología cada vez que se agrega un nuevo objeto de tipo entidad.

$Qstr = \text{“¿Qué ríos se encuentran en el ?”, municipio San José}$

El segundo paso consiste en simplificar y normalizar las palabras de la consulta, eliminando palabras no relevantes (stop words), singularizando las palabras, etc. Teniendo como excepción que no se eliminan las preposiciones.

$$Q_{str} = \text{rio}, \quad \text{en}, \quad \text{municipio San José}$$

Posteriormente se identifica la relación espacial <<en>> asociandola el operador relacional *Contenido_en*. De esta forma se transforma la consulta en una triplete de la forma {*que, relación espacial, donde*}

$$Q = \{\text{rio}, \quad \text{Contenido_en}, \quad \text{municipio San José}\}$$

Con la triplete se conforma la consulta SPARQL y se lanza la búsqueda. A partir de aquí se reconocen y activan en la ontología los conceptos subyacentes a los términos de búsqueda introducidos por el usuario para realizar el proceso de recuperación.

En la Tabla 6.2 se presentan los resultados comparativos en cuanto a eficiencia de los métodos propuesto Estos resultados ponen de manifiesto como las relaciones existentes entre los objetos, en este caso las relaciones topológicas, resultan útiles en tareas como la determinación automática de la clasificación de los suelos. Otro aspecto a destacar es que las tareas de recuperación de datos utilizando la estructura y los métodos propuestos resultan ser más simples y rápidas que los métodos convencionales para el caso de usuarios no expertos en aplicaciones SIG, ya que posibilita el sistema al navegar por la ontología realiza distintos tipos de análisis sin la necesidad de que el usuario conozca como se realiza el proceso, a diferencia de las herramientas de análisis de los SIG. Los resultados demuestran que la recuperación de la información, para los usuarios no expertos, mediante el método propuesto aumenta la calidad de los análisis pues reduce los posibles errores en las operaciones que deben realizar, por ejemplo, al crear la consulta.

Tabla 6.2. Comparaciones entre cantidad de actividades a realizar por el usuario para indicar la información que desea recuperar utilizando el método propuesto y la herramienta Quantum GIS 1.8.0²⁰.

Pasos Metodológicos	Selección de regiones por tipo de suelo	Cantidad de actividades
Operaciones en el SIG	<ul style="list-style-type: none"> • Selección manual de la capa. • Conformación manual de la sentencia SQL. • Selección automática de los objetos. 	3
Operación en la plataforma	<ul style="list-style-type: none"> • Conformación manual del criterio de búsqueda en lenguaje natural.. 	1

Estos resultados muestran que el uso de la ontología como una capa intermedia entre el sistema de recuperación y las fuentes de datos permite el manejo de los objetos a través de su representación conceptual (abstracción). Dando la posibilidad de que el sistema disponga de herramientas para el procesamiento semántico de los datos geoespaciales. Esto permite que los diferentes tipos de análisis realizados sobre estos tipos de datos puedan ser realizados tanto por especialistas como por usuarios no expertos en el uso de SIG.

6.4 Evaluación de los resultados

La evaluación del modelo se realiza en base a los valores de la precisión y de la cobertura (*recall*). La curva de precisión contra cobertura es una buena medida para valorar la efectividad de un sistema de recomendación (Salton and McGill, 1986). Un sistema de recuperación que consiga los valores más cercanos al punto (1.0,1.0) en el sistema de coordenadas obtiene los mejores resultados.

En el análisis presentado en esta investigación primero se valora el modelo de recuperación utilizando solamente la medida de similitud semántica (DIS-C), después se aplica el análisis semántico con el espacial (MISE), integrándolo como se propone en la Sección 3.4. Finalmente, se presentan los resultados lanzando diferentes consultas.

²⁰ <http://www.qgis.org>

Las consultas utilizadas fueron formuladas por cinco especialistas del Departamento de Riego, Drenaje y Ciencias del Suelo de la UNAH, los cuales fueron responsables además de evaluar los resultados obtenidos en la plataforma de prueba. Del grupo de consultas formuladas se seleccionaron tres como casos de pruebas, las cuales se enuncian a continuación:

- 1) Regiones de suelos Ferralíticos rojos.
- 2) Regiones de suelos intersectados por un río.
- 3) Regiones que comparten límite con suelo Aluvial.

Este grupo de especialista analizó además que de las 213 instancias mapeadas en la ontología 26 eran relevantes para la primera consulta, 18 para la segunda consulta y 21 para la tercera.

La evaluación en función de la precisión y la cobertura en el escenario de prueba demuestran que integrar la similitud espacial mejora los resultados de la recuperación. En la Figura 6.4 se muestran los valores de precisión y cobertura aplicando la metodología solo con la medida de similitud DIS-C, línea azul, y aplicándola con la integración de los elementos espaciales (MISE), línea roja, para la consulta “Regiones que comparten límite con suelo Aluvial”.

Las relaciones juegan un rol importante en la representación de la información geográfica y por eso influyen en la medición de la similitud. Una comparación general entre ambos modelos muestra que el modelo MISE es más efectivo que DIS-C en la recuperación de la información.

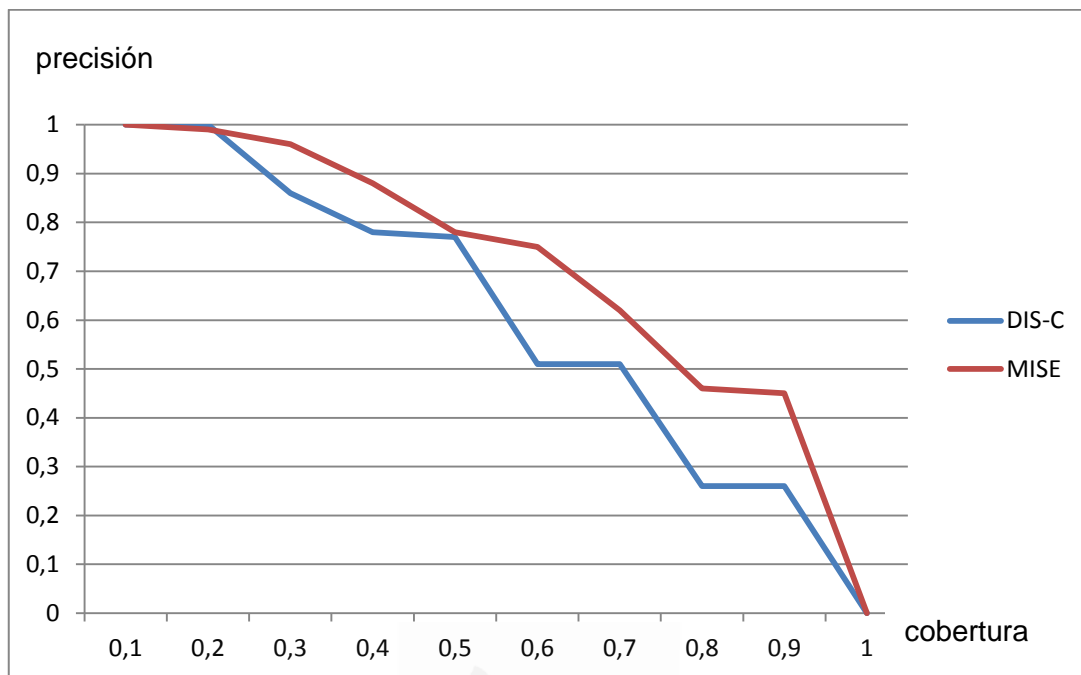


Figura 6.4. Precisión contra cobertura para la consulta “Regiones que comparten límite con suelo Aluvial”.

Un buen sistema de recuperación debe proporcionar buenos resultados para cualquier consulta que se le realice. Sin embargo, la importancia de los conceptos en la definición de la consulta varía y se influencia por las relaciones espaciales entre un concepto y otro. En la definición de la consulta “Regiones de suelos intersectados por un río.” La relación espacial “*intersecta*” es esencial. En la Figura 6.5 Precisión contra cobertura en la tres consultas diferentes se muestra las curvas de precisión contra cobertura para las consultas:

C1 = “Regiones de suelos Ferralíticos rojos”

C2 = “Regiones de suelos intersectados por un río”

C3 = “Regiones que comparten límite con suelo Aluvial”

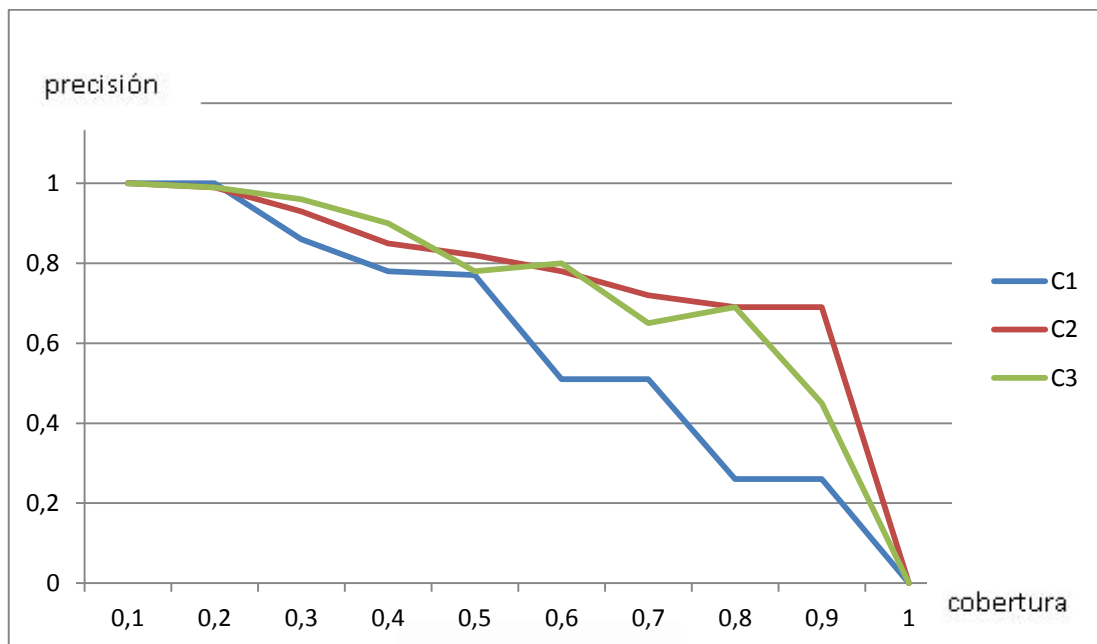


Figura 6.5 Precisión contra cobertura en la tres consultas diferentes.

En la Figura 6.5 Precisión contra cobertura en la tres consultas diferentes. se puede observar que la eficiencia del modelo de recuperación es buena independientemente de los conceptos de la consulta que se utilicen. Las tres curvas se encuentran por encima de la diagonal. Las curvas no se diferencian grandemente pero se encuentran mejores resultados en las consultas que involucran relaciones espaciales.

6.5 Conclusiones

Se demostró que el uso del tipo de ontología propuesto combinado con técnicas de recuperación de información permite enfrentar las tareas de detección y recuperación de la información mediante su aplicación sobre objetos geoespaciales de forma automática aumentando la efectividad de este proceso. Quedó demostrado además que el uso de la ontología en la representación semántica de estos objetos permite la combinación de rasgos de bajo nivel con abstracciones semánticas disminuyendo así la brecha semántica existente entre Hombre-Máquina. Además, abre la posibilidad de que la recuperación de los objetos de interés sea sobre la base del procesamiento de consultas realizadas en lenguaje

natural habilitando el procesamiento semántico de objetos en las aplicaciones de este tipo.

La metodología propuesta se basa en un procesamiento semántico que permite localizar un determinado tipo de suelo con determinadas características. Además de devolver la ubicación del suelo recupera imágenes de estos suelos permitiéndole al usuario una visualización de las características y propiedades de los mismos. Aplica medidas de similitud semánticas y emplea un vocabulario estructurado que incluye la información geográfica referida al Recurso Suelo pero los resultados, desde el punto de vista metodológico, pueden ser válidos para cualquier otro tipo de Información Geográfica.



Universitat d'Alacant
Universidad de Alicante

7. Conclusiones y trabajos futuros.

Como resultado de esta Tesis doctoral, se le dió cumplimiento a su objetivo principal con el desarrollo de una metodología general para la recuperación de información que aplica un modelo integrado que combina mediciones de similitud entre las relaciones conceptuales y espaciales mediante las cuales se pueden describir los objetos geoespaciales, la cual constituye la aportación más importante de este trabajo de investigación.

El estudio del estado de arte denotó la relevancia de los sistemas de recuperación de información geográfica. Este campo, tiene su origen en dos más consolidados como son la recuperación de información y los sistemas de información geográfica. Su objetivo principal, y también parte de los objetivos de este trabajo de tesis, es la indexación de información a través del uso de una ontología, teniendo en cuenta tanto las características temáticas como las geográficas, haciendo énfasis en las relaciones espaciales.

Los modelos de datos geográficos generalmente representan explícitamente un conjunto de objetos básicos, su geometría y propiedades. Sin embargo, en muchos ambientes geográficos, la semántica aparece en las relaciones que enlazan estos objetos. No obstante, muchas de estas relaciones no son explícitamente representadas en modelos de datos geoespaciales, ya que aparecen solo implícitamente en el despliegue de las bases de datos geográficas; por lo cual es de suma importancia contar con descripciones semánticas que permitan realizar este tipo de representaciones entre objetos geográficos. Identificar el conjunto de propiedades y relaciones que poseen implícitamente los datos geoespaciales es otra de las aportaciones obtenidas en este trabajo.

Con base en los elementos característicos identificados de los objetos geoespaciales se define una ontología geoespacial aplicada al dominio de los suelos cubanos, la cual constituye una estructura de indexación que tiene en cuenta tanto el ámbito textual como el espacial de la información geográfica. Esta ontología constituye otra de las aportaciones principales de este trabajo. La ontología

propuesta para sistemas de RIG combina el espacio geográfico con el dominio de los suelos de Cuba.

Se presentó un enfoque de integración en las medidas de similitud mediante la utilización de la ontología. Este enfoque se basa en la integración de medidas de similitud semántica con similitud en el análisis geoespacial. Los resultados obtenidos muestran que es posible la integración de criterios en un solo modelo utilizando el enfoque semántico geoespacial.

La estructura propuesta permite resolver consultas clásicas, como pueden ser las consultas textuales puras, pero también nuevos tipos de consultas que combinan aspectos textuales y espaciales. Utilizar la ontología como estructura de indexación presenta una ventaja cualitativa muy importante sobre otras alternativas para sistemas de RIG que se han propuesto recientemente. Esta ventaja es que realiza expansión de los términos de consulta de manera implícita. Realizar la expansión de la consulta de manera implícita le facilita a los usuarios no expertos, hacia los cuales está orientada esta metodología de recuperación, obtener los resultados esperados. La representación conceptual le permite manipular la información geográfica sin considerar aspectos que normalmente se han utilizado en enfoques tradicionales como unidades de medida, escala, sistemas de referencia, etc. Así que esto difiere de los SIG tradicionales que se basan en las primitivas geométricas de representación.

Todos estos elementos se presentan en una arquitectura completa para sistemas de RIG. Esta arquitectura define todos los componentes necesarios para que la información indexada pueda ser consultada de manera cómoda por los usuarios, y presentada de manera clara y organizada. Los resultados de los experimentos realizados demuestran que la estructura presenta buena relación entre la precisión y la cobertura de las búsquedas.

El prototipo de la arquitectura ha permitido valorar el grado de consecución de los objetivos y el cumplimiento de los requisitos impuestos a la arquitectura. Las evaluaciones de los revisores de los congresos nacionales e internacionales en los que se han publicado los resultados obtenidos han servido también de validación externa del trabajo realizado.

7.1 Líneas de trabajo futuro.

Se pueden realizar mejoras de la estructura de indexación propuesta para la recuperación de información geográfica. Una de las posibles mejoras es la inclusión de nuevas relaciones. Este tipo de relaciones se pueden representar fácilmente en la ontología del espacio geográfico ya que esta fue diseñada con el requisito de que este tipo de cambios se puedan incorporar de manera sencilla. Además, se puede analizar el uso de otras ontologías de aplicación para mapearlas con la propuesta y ampliar el campo de aplicación.

Los componentes que forman la arquitectura (o el prototipo desarrollado en base a la misma) son frutos de trabajos de investigación reciente y, por tanto, constituyen líneas abiertas de investigación.

Los problemas derivados de la ambigüedad de los nombres de lugar hacen que las técnicas de resolución de topónimos no alcancen todavía unos resultados comparables a la geo-referenciación manual. Por tanto, el desarrollo de componentes de desambiguación de topónimos que tengan en cuenta el contexto está atrayendo el interés de muchos investigadores del área.

Aplicar nuevas técnicas de Procesamiento del Lenguaje Natural en el procesamiento de la consulta.



Universitat d'Alacant
Universidad de Alicante

8. Publicaciones relacionadas con este trabajo de tesis.

1. Exportador de Imágenes para un Repositorio de Objetos de Aprendizaje. Evento UCIENCIA 2007. ISBN 978-959-286-005-6
2. Método de recomendación de objetos de aprendizaje en la carrera de medicina veterinaria utilizando ontologías. Evento ENCE 2009. ISBN 978-959-18-0498-3
3. Recomendación de objetos de aprendizaje almacenados en repositorios lor@server según las preferencias del usuario. Octava Conferencia Iberoamericana de Ingeniería e Innovación Tecnológica, CIIT 2009. Orlando, Florida, EUA.
4. Sistema de recomendación de recursos digitales utilizando ontologías. Anuario científico "Ciencia en la UNAH 2010", ISBN: 978-959-16-1163-5.
5. Método de gestión y recomendación de imágenes con información geográfica. Décima conferencia Iberoamericana en Sistemas, Cibernética e Informática, CISC 2011. Orlando, Florida, EUA.
6. Sistema de gestión semántica de información geográfica. Oncena Edición de la Semana Tecnológica en el Centro Internacional de Formación Postal (CIFP). Noviembre, 2011. ISSN 2076-9792.
7. Gestión semántica de información geográfica. Anuario científico "Ciencia en la UNAH 2012". ISBN 978-959-16-2176-8
8. Recuperación de objetos geoespaciales utilizando medidas de similitud semántica. Congreso Internacional COMPUMAT 2013. ISBN 978-959-286-022-3.

9. Recuperación de información geográfica de fuentes de datos heterogéneas conducida por ontologías. AGRING 2013.
10. Representación ontológica para la interpretación semántica de datos geoespaciales en el dominio suelo. Twelfth Latin American and Caribbean Conference for Engineering and Technology (LACCEI'2014) "Excellence in Engineering To Enhance a Country's Productivity" 2014 Guayaquil, Ecuador.
11. *Similitud semántica en la recuperación de información geográfica para la toma de decisiones*. Novena Conferencia Ibérica de Sistemas y Tecnologías de Información CISTI 2014. Barcelona, España,

Los artículos fueron publicados en libro y en CD con ISBN y además están disponibles en la biblioteca digital XPLore del IEEE, EBSCO y SCOPUS.

12. *Similitud semántica en la interpretación de información geoespacial*. Primera conferencia científica internacional de la Universidad de las Ciencias Informáticas. UCIENCIA 2014
13. *Semantic and spatial similarity in geographic information retrieval*. ICAI'14 - The 2014 International Conference on Artificial Intelligence. WORLDCOMP'14. Las Vegas, USA.
14. *Methodology to build learning objects using semantic similarity measures*. Methodology to build learning objects using semantic similarity measures. EDULEARN 2014. Barcelona, España.
15. *Recuperación de objetos geoespaciales utilizando medidas de similitud semántica*. *Revista Cubana de Ciencias Informáticas* Vol. 8, No. 2, Mes Abril, 2014 ISSN: 2227-1899 | RNPS: 2301 (Grupo 2).

9. Referencias bibliográficas

- Agarwal, P., 2005. Ontological considerations in GIScience 501–536.
- Aparício, A. d. S., Farias, O.L.M. d., Santos, N. d., 2005. Applying Ontologies in the Integration of Heterogeneous Relational Databases.
- Aparício, A. d. S., Farias, O.L.M. d., Santos, N. d., 2006. Integration of Heterogeneous Databases and Ontologies.
- Arpirez, J.C., Corcho, O., Fernández-López, M., Gómez-Pérez, A., 2001. WebODE: a scalable workbench for ontological engineering., in: Gil, Y., Musen, M., Shavlik, J. (Eds.), Presented at the First International Conference on Knowledge Capture, K-CAP'01, ACM Press, Victoria, Canada, pp. 6–13.
- Atserias, J., Casas, B., Comelles, E., González, M., Padró, L., Padró, M., 2006. FreeLing 1.3: Syntactic and semantic services in an open-source NLP library. Presented at the Fifth International Conference on Language Resources and Evaluation.
- Attneave, F., 1950. Dimensions of similarity. *American Journal of Psychology* 63, 516–56.
- Baeza-Yates, R., Ribeiro-Neto, B., 1999. *Modern Information Retrieval*. Addison Wesley.
- Berners-Lee, T., Hendler, J., Lassila, O., 2001. *The Semantic Web*. Scientific American.
- Bruns, H.T., Egenhofer, M. J., 1996. Similarity of Spatial Scenes. *Seventh International Symposium on Spatial Data Handling (SDH '96)* 4A, 31–42.
- Buccella, A., Cechich, A., Fillottrani, P., 2009. Ontology-driven geographic information integration - A Survey of current Approaches. *Computer & Geosciences* 35, 710–723.
- Bucher, B.P., Clough, H., Joho, R., 2005. Geographic IR Systems: Requirements and Evaluation. Presented at the 22nd International Cartographic Conference.
- Budanitsky, A., 1999. *Lexical Semantic Relatedness and Its Application.*, Natural Language Processing. Computer Systems Research Group, University of Toronto, Technical Report.
- Burrough, P., 1988. *Principles of Geographical Information Systems for land resources assessment*. Oxford University Press, Oxford. 80.
- Burrough, P., McDonnell, R., 1998. *Principles of Geographical Information Systems.*, in: Oxford University Press, Oxford.
- Chandrasekaran, B., Johnson, T.R., Benjamins, V.R., 1999. Ontologies: what are they? Why do we need them? *IEEE Intelligent Systems and Their Applications: Special Issue on Ontologies* 14, 20–26.
- Chang, C., Lee, S., 1991. Retrieval of Similarity Pictures on Pictorial Databases. *Pattern Recognition*. 24(7), 675–680.
- Chen, Y., Garcia, E.K., Gupta, M.R., Rahimi, A., Cazzanti, L., 2009. Similarity-based Classification - Concepts and Algorithms. *Journal of Machine Learning Research* 10, 747–776.

- Collet, Huhns, M., Shen, W., 1991. Resource Integration Using a Large Knowledge Base. *Carnot. Computer* 24(12), 55–62.
- Conference on Knowledge Management. [WWW Document], 2008. .
- Cormen, T.H., Leiserson, C.E., Rivest, R.L., Stein, C., 2001. Introduction to Algorithms, Second ed. MIT Press and McGraw-Hill.
- Coventry, K.R., Garrod, S.C., 2004. Saying, seeing, and acting: The psychological semantics of spatial prepositions. Hove, UK, Psychology Press.
- Cowie, J., Wilks, Y., 2000. Information Extraction. *The Handbook of Natural Language Processing*.
- Delgado, T., Capote, J.L. (Eds.), 2009. Semántica espacial y descubrimiento de conocimiento para el desarrollo sostenible., Proyecto CYTED-IDEDES. Evaluación y potenciación de Infraestructuras de datos espaciales para el desarrollo sostenible en América Latina y el Caribe. CUJAE, La Habana, Cuba.
- Doan, A., Halevy, A.Y., 2004. Semantic Integration Research in the Database Community - A Brief Survey. American Association for Artificial Intelligence.
- Dolbear, C., Goodwin, J., Mizen, H., Ritchie, J., 2005. Semantic interoperability between topographic data and a flood defence ontology 15.
- Donnelly, M., Bittner, T., 2005. Spatial relations between classes of individuals. 4th Conference On Spatial Information Theory (COSIT05), Ellicottville, New York, Springer.
- Egenhofer, M., Al-Taha, K., 1992. Reasoning About Gradual Changes of Topological Relationships. *Theories and Methods of Spatio-Temporal Reasoning in Geographic Space.*, Pisa, Italy.
- Egenhofer, M., Franzosa, R., 1995. On the equivalence of topological relations. Presented at the International journal of geographical information systems., pp. 133–152.
- Egenhofer, M.J., 1993. A model for detailed binary topological relationships. *Geomatica* 47.
- Egenhofer, M.J., 2002. Toward the Semantic Geospatial Web. *The Tenth ACM International Symposium on Advances in Geographic Information Systems*, McLean, Virginia 1–4.
- Egenhofer, M.J., Franzosa, R., 1991. Point-set topological spatial relations. *International Journal of Geographical Information Systems*. 5(2), 161–174.
- Egenhofer, M.J., Herring, J.R., 1991. Categorizing Binary Topological Relations Between Regions, Lines, and Points in Geographic Databases. Department of Surveying Engineering, University of Maine.
- Egenhofer, M.J., Mark, D.M., 1995a. Naive geography. *International Conference on Spatial Information Theory (COSIT95)* Berlin, Germany, Springer, 1–15.
- Egenhofer, M.J., Mark, D.M., 1995b. Modeling conceptual neighborhoods of topological line-region relations. *International Journal of Geographical Information Systems*. 9(5), 555–565.
- Ehrig, M., 2007. Ontology Alignment-Bridging the Semantic Gap, *SEMANTIC WEB AND BEYOND Computing for Human Experience*.
- Fensel, D., Wahlster, W., Lieberman, L., Hendler, J., 2002. *Spinning the Semantic Web: Bringing the World Wide Web to its Full Potential*. MIT Press, Cambridge, MA.
- Fonseca, F., 2001. *Ontology-Driven Geographic Information Systems*.
- Fonseca, F., Davis, C., Câmara, G., 2002a. Bridging Ontologies and Conceptual Schemas in Geographic Information Integration.

- Fonseca, F.T., Egenhofer, M.J., Agouris, P., Camara, G., 2002b. Using ontologies for integrated geographic information systems. *Transactions in GIS* 6, 231–257.
- Frank, A., 1996. Qualitative spatial reasoning: cardinal directions as an example. *International Journal of Geographical Information Systems*. 10(3), 269–290.
- Freksa, C., 1992. Using Orientation Information for Qualitative Spatial Reasoning. *Theories and Methods of Spatial Temporal Reasoning in Geographic Space.*, Springer-Verlag, New York. 162–178.
- Gal, A., Shvaiko, P., 2010. *Advances in Ontology Matching*.
- Garea, E., Gil, J., 2007. Los Sistemas de Información Geográfica Gobernados por Ontologías como Herramienta para la Interpretación semántica de la Información espacial y su integración a la IDERC., in: CD, *Memorias De Informatica 2007*. Presented at the V Congreso Internacional de Geomática., La Habana, Cuba.
- Gentner, D., Markman, A.B., 1994. Structural alignment in comparison: No difference without similarity. *Psychological Science*. 5(3), 152–158.
- Gentner, D., Markman, B., 1995. Similarity is like analogy: Structural alignment in comparison. *Similarity in Language, Thought and Perception*. C. Cacciari. Brussels, Belgium. Brepols, 111–147.
- Geoserver Community, 2013. GeoServer WMS. [WWW Document]. URL <http://geoserver.org>
- Giger, C., Najar, C., 2003. Ontology-based integration of data and metadata BT -. 6th AGILE Conference on Geographic Information Science 586 – 594.
- Global Spatial Data Infrastructure Association., 2012. . GSDI.
- Goldstone, R.L., 1994. Similarity, Interactive Activation, and Mapping. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. 20(1), 3–28.
- Goldstone, R.L., 1998. Hanging Together: A Connectionist Model of Similarity. *Localist Connectionist Approaches to Human Cognition.*, Lawrence Erlbaum Associates, Mahwah, NJ, 283–325.
- Goldstone, R.L., 2004. Similarity., *Mit Encyclopedia of the Cognitive Sciences*. MIT Press.
- Goldstone, R.L., Medin, D.L., 1994. Similarity, Interactive Activation and Mapping: An overview. *Advances in Connectionist and Neural Computation Theory*. J. Barnden and H. K. New Jersey, Ablex. Vol. 2: Analogical Connections., 321–362.
- Goldstone, R.L., Son, J., 2005. Similarity., *Cambridge Hand- book of Thinking and Reasoning*. Cambridge University Press: 1336., Cambridge.
- Golledge, R., Briggs, R., Demko, D., 1969. The configuration of distances in intra-urban space. *Proceedings of the Association of American Geographers*, Washington, DC:Association. 60–65.
- Gómez-Pérez, A., 1999. *Ontological Engineering: a state of the Art*.
- Gómez-Pérez, A., 2002. OntoWeb. Ontology-based information exchange for knowledge management and electronic commerce IST-2000-29243.
- Grdenfors, P., 2000. *Conceptual Spaces: The Geometry of Thought*. MIT Press. Cambridge, Massachusetts, EE. UU.
- Gruber, T., 1995. Toward Principles for the Design of Ontologies Used for Knowledge Sharing. *International Journal of Human and Computer Studies* 43, 907–928.
- Gruber, T.R., 1993. A translation approach to portable ontologies. *Knowledge Acquisition* 5, 199 – 220.

- Guarino, N., 1998. Formal Ontology and Information Systems, in: *Formal Ontology in Information Systems*. Amsterdam, Netherlands, pp. 3–15.
- Hahn, U., Chater, N., 1997. Concepts and Similarity. *Knowledge, Concepts, and Categories*. Psychology Press/MIT Press, Hove, UK. 43–92.
- Hahn, U., Chater, N., Richardson, L.B., 2003. Similarity as transformation. *Cognition* 87, 132.
- Hakimpour, F., 2003. Using ontologies to resolve semantic heterogeneity for integrating spatial database schemata. (Doctoral).
- Harmon, J.E., Anderson, S.J., 2003. The Design and Implementation of Geographic Information Systems. Presented at the John Wiley & Sons.
- Hernández, A., Morales, M., Ascanio, M.O., Morell, F., 1999. Manual para la aplicación de la Nueva Versión de Clasificación Genética de los Suelos de Cuba. 193.
- Hernández, D., 1991. Relative representation of spatial knowledge: The 2-C case. *Cognitive and Linguistic Aspects of Geographic Space*. D. M. Mark and A. U. Frank. Dordrecht, The Netherlands, Kluwer Academic Publishers: 373–385.
- Hernández, D., 1994. *Qualitative Representation of Spatial Knowledge*. Berlin, Springer-Verlag.
- Hess, G.N., Lochpe, C., 2004. Ontology-driven resolution of semantic heterogeneities in gdb conceptual schemas. Presented at the Proceedings of the GEOINFO'04: VI Brazilian Symposium on GeoInformatics., Campos do Jordão, Brazil., pp. 247–263.
- Hirts, G., St-Onge, D., 1998. Lexical chains as representation of context for the detection and correction of malapropisms. Presented at the WordNet: An Electronic Lexical Database., MIT Press., pp. 265–283.
- Imai, S., 1977. Pattern Similarity and Cognitive Transformations. *Acta Psychologica* 41, 443–447.
- INEGI, 1996. *Diccionario de Datos Topográficos, Escala 1:50,000*.
- Inselberg, A., 1985. The Plane with Parallel Coordinates. *The Visual Computer* 1, 69–97.
- ISO/IEC, 2002. *Geographic Information – Reference Model*. International Standard 19101.
- Janowicz, K., Raubal, M., Schwering, A., Kuhn, W., 2008. Semantic Similarity Measurement and Geospatial Applications. *Workshop on Semantic Similarity Measurement and Geospatial Applications*.
- Kavouras, M., Kokla, M., 2001. Ontology-based fusion of Geographic Databases.
- Kavouras, M., Kokla, M., Tomai, E., 2003. Comparing categories among geographic ontologies. *Computers&Geosciences*, (special issue) 31, 145–154.
- Kosslyn, S.M., Ball, T.M., Reiser, B.J., 1978. Visual images preserve metric spatial information: evidence from studies of image scanning. *Journal of Experimental Psychology*. 4(1), 47–60.
- Kuhn, W., 2005. Geospatial Semantics: Why, of What, and How? *Data Semantics III* 1–24.
- Larín, R., 2013. Nuevo tipo de ontología para la representación semántica de objetos geoespaciales.
- Laurini, R., Thompson, D., 1992. Fundamentals of spatial informations systems., in: *The APIC Series 37 - Academic Press*.
- Leacock, C., Chodorow, M., 1998. Combining local context and WordNet similarity for word sense identification. Presented at the WordNet: An Electronic Lexical Database, MIT Press, pp. 265–283.

- Leake, D.B., Maguitman, A., Cañas, A., 2002. Assessing Conceptual Similarity to Support Concept Mapping. American Association for Artificial Intelligence.
- Lenat, Guha, R., 1990. Building Large Knowledge Based Systems: Representation and Inference in the Cyc Project. Addison-Wesley Publishing Company, Reading, MA.
- Lenat, Miller, G., Yokoi, T., 1995. Cyc, WordNet, and EDR: Critiques and Responses. Communications of the ACM. 38(11), 45–48.
- Levachkine, S., Guzmán-Arenas, A., 2007. Hierarchy as a new data type for qualitative variables. Expert Systems with Applications 32, 899–910.
- Levenshtein, I.V., 1966. Binary codes capable of correcting deletions, insertions, and reversals. Soviet Physics Doklady 10, 707–710.
- Li, B., Fonseca, F., 2006. TDD - A Comprehensive Model for Qualitative Spatial Similarity Assessment. Spatial Cognition and Computation 6(1), 31–62.
- Lin, D., 1998. An information-theoretic definition of similarity. Presented at the 15th International Conference on Machine Learning, pp. 296–304.
- Longley, P., Goodchild, M., Maguire, D., Rhind, D., 2001. Geographic Information Systems and Science. Presented at the John Wiley & Sons.
- Machado, N., Balmaseda, C., Montoyo, A., 2014a. Representación ontológica para la interpretación semántica de datos geoespaciales en el dominio suelo. Presented at the Twelfth Latin American and Caribbean Conference for Engineering and Technology (LACCEI'2014) "Excellence in Engineering To Enhance a Country's Productivity," Guayaquil, Ecuador.
- Machado, N., Gonzáles, L., Balmaseda, C., Montoyo, A., 2014b. Recuperación de objetos geoespaciales utilizando medidas de similitud semántica. Revista Cubana de Ciencias Informáticas. 8.
- Machado, N., Marrero, Y., Balmaseda, C., Montoyo, A., 2014c. Semantic and spatial similarity in geographic information retrieval. Presented at the ICAI'14 - The 2014 International Conference on Artificial Intelligence. WORLDCOMP'14, Las Vegas, USA.
- Machado, N., Montoyo, A., 2007. Exportador de Imágenes para un Repositorio de Objetos de Aprendizaje. Presented at the Evento UCIENCIA 2007, La Habana, Cuba.
- Machado, N., Rodríguez, I., Machado, M., Balmaseda, C., 2013. Recuperación de información geográfica de fuentes de datos heterogéneas conducida por ontologías. Presented at the AGRING 2013, Mayabeque, Cuba.
- Machado, N., Rodríguez, I., Montoyo, A., 2014d. Methodology to build learning objects using semantic similarity measures. Presented at the 6th annual International Conference on Education and New Learning Technologies. EDULEARN'14, Barcelona, España.
- Manning, C.D., Raghavan, P., Schütze, H., 2008. Introduction to Information Retrieval., Cambridge University Press.
- Mark, D., 1999. Spatial representation: A cognitive view. Geographical Information Systems: Principles and Applications. D. J. Maguire, M. F. Goodchild, D. W. Rhind and P. Longley. 81–89.
- Markman, A.B., 1993. Structural Alignment During Similarity Comparisons. Cognitive Psychology. 25, 431–467.
- Markman, A.B., 1997. Constraints on Analogical Inference. Cognitive Science. 21(4), 373–418.
- Markman, A.B., 2001. Thinking. Annual Review of Psychology. 52, 223–247.
- Markman, A.B., Gentner, D., 1993. Splitting the Differences: A Structural Alignment View of Similarity. Journal of Memory & Language. 32, 517–535.

- Markman, A.B., Gentner, D., 1996. Commonalities and Differences in Similarity Comparisons. *Memory & Cognition*. 24, 235–249.
- Martínez, F., 2004. El problema de la fusión de colecciones en la recuperación de información multilingüe y distribuida: cálculo de la relevancia documental en dos pasos. (Doctoral).
- Medin, D.L., Goldstone, R.L., Gentner, D., 1993. Respects for Similarity. *Psychological Review*. 100(2), 254–278.
- Mendoza, L., 2008. Anotación Semántica de Datos Vectoriales en Sistemas de Información Geográfica. Implementación en gvSIG.
- Miller, G.A., Beckwith, R., Fellbaum, C., Gross, D., Miller, K.J., 1990. Introduction to WordNet: An On-Line Lexical Database. *International Journal of Lexicography*. 3, 235–244.
- Mitkov, R. (Ed.), 2003. *The Oxford Handbook of Computational Linguistics.*, Oxford University Press. ed.
- Montoyo, A., 2008. Uso y diseño de ontologías en procesamiento del lenguaje natural y la web semántica.
- Muñoz, R., 2004. Aplicación de técnicas semánticas a la Recuperación de Documentos Transcritos. (Memoria de suficiencia investigadora).
- NCGIA, 1990. Introduction to GIS. National Center for Geographic Information and Analysis. University of California 1.
- Nedas, K., Egenhofer, M., 2003. Spatial Similarity Queries with Logical Operators. 8th International Symposium, SSTD 2003.
- OGC, 2003. OpenGIS Reference Model. OpenGIS Project Document 03-040.
- Papadias, D., Dellis, V., 1997. Relation-Based Similarity. Fifth ACM Workshop on Advances in Geographic Information Systems, Las Vegas, NV.
- Papadias, D., Kavouras, M., 1994. Acquiring, representing and processing spatial relations. *Proceedings of the 6th International Symposium on Spatial Data Handling*, Edinburgh, U.K., Taylor Francis. 631–643.
- Pedraza-Jiménez, R., Codina, L., Rovira, C., 2007. Web semántica y ontologías en el procesamiento de la información documental. Presented at the El profesional de la información, pp. 569–578.
- Quintero, R., Moreno-Ibarra, M., Torres, M., Menchaca-Mendez, R., Guzman, G., 2012. DIS-C: Conceptual Distance.
- Rada, R., Mili, H., Bicknell, E., Blettner, M., 1989. Development and Application of a Metric on Semantic Nets. *IEEE transactions on systems, man and cybernetics* 19, 17–30.
- Randell, D., Cui, Z., Cohn, A., 1992. A Spatial Logic Based on Regions and Connection. *Principles of Knowledge Representation and Reasoning, KR '92.*, Cambridge, MA, St. Charles, IL: Morgan Kaufmann 165–176.
- Renteria, W., 2009. Recuperación controlada de información cualitativa desde repositorios de datos.
- Renz, J., 2002. *Qualitative Spatial Representation and Reasoning. Qualitative Spatial Reasoning with Topological Information.* Springer-Verlag Berlin Heidelberg.
- Resnik, P., 1993. Selection and Information: A Class-Based Approach to Lexical Relationships. (Doctoral).
- Resnik, P., 1995. Using information content to evaluate semantic similarity in a taxonomy. 14th International Joint Conference on Artificial Intelligence (IJCAI), Montreal, Canada. 448–453.

- Resnik, P., 1999. Semantic similarity in a taxonomy: An information-based measure and its application to problems of ambiguity in natural language. *Journal of Artificial Intelligence Research* 11, 95–130.
- Rigaux, P., Scholl, M., Voisard, A., 2001. *Spatial Databases With Application To GIS*. Presented at the Academic Press.
- Robertson, S.E., Jones, S.K., 1976. Relevance weighting of search terms. *Journal of the American Society for Information Science* 27, 129–146.
- Rodríguez, A., 2008. *Curso Bases de Datos Espaciales*.
- Rodríguez, A., Egenhofer, M.J., 2004. Comparing geospatial entity classes: An asymmetric and context-dependent similarity measure. *International Journal of Geographical Information Science*. 18(3), 229–256.
- Rodríguez, A.F., Vilches, L.M., Bernabé, M.A., 2006. *Ingeniería Ontológica: El camino hacia la mejora del acceso a la información geográfica en el entorno web*. Presented at the Congreso JIDEE 2006. III Jornadas de la IDEE, Castellón, España.
- Rodríguez, M.A., 2000. Assessing semantic similarity among spatial entity classes.
- Rodríguez, M.A., Egenhofer, M.J., 2003. Determining semantic similarity among entity classes from different ontologies. *IEEE Transactions on Knowledge and Data Engineering* 15(2), 442–456.
- Rosen, K.H., 2003. *Discrete Mathematics and Its Applications.*, 5ª Edición. ed. Addison Wesley.
- Salton, G., Lesk, M., 1968. Computer Evaluation of Indexing and Text Processing. *ACM* 15, 8–36.
- Salton, G., McGill, M.J., 1986. *Introduction to modern information retrieval*, 3rd Ed.
- Salton, G., Wong, A., Yang, C.S., 1975. A Vector Space Model for Automatic Indexing. *Communications of the ACM*. 18, 613–620.
- Sanderson, M., Järvelin, K., Allan, J., 2004. SIGIR 2004, in: Bruza, P. (Ed.), Presented at the 27th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, Sheffield, UK.
- Schwering, A., 2006. Semantic Similarity Measurement including Spatial Relations for Semantic Information Retrieval of Geo-Spatial Data.
- Schwering, A., 2008. Approaches to Semantic Similarity Measurement for Geo-Spatial Data - A Survey. *Transactions in GIS* 12(1), 5–29.
- Schwering, A., Raubal, M., 2005. *Spatial Relations for Semantic Similarity Measurement*. Ordnance Survey.
- Seco, D., 2009. *Técnicas de indexación y recuperación de documentos utilizando referencias geográficas y textuales*. (Doctoral).
- Shariff, A.R., Egenhofer, M.J., Mark, D.M., 1998. Natural-language spatial relations between linear and areal objects: The topology and metric of English language terms. *International Journal of Geographical Information Science*. 12(3), 215–246.
- Sloman, S.A., Love, B.C., Woo-Kyoung, A., 1998. Feature centrality and conceptual coherence. *Cognitive Science*. 22, 189228.
- Smith, B., 1999. *Ontology: philosophical and computational*.
- Smith, B., Mark, D.M., 1998. Ontology and geographic kinds. *Proceedings of the 8th International Symposium on Spatial Data Handling (SDH'98)*, Vancouver: International Geographical Union 308–320.
- Sotnykova, A., Vangenot, C., Cullot, N., Bennacer, N., Aufaure, M., 2005. Semantic mappings in description logics for spatio-temporal data-base schema integration. *Journal on Data Semantics III* 143–167.

- Special Interest Group on Information Retrieval. [WWW Document], 2014. . URL <http://www.sigir.org/>
- Stoimenov, L., Stanimirovic, A., Djordjevic-Kajan, S., 2006. Discovering mappings between ontologies in semantic integration process., in: Proceedings of the AGILE'06: 9th Conference on Geographic Information Science. Visegrád, Hungary, pp. 213–219.
- Studer, S., Benjamins, V.R., Fensel, D., 1998. Knowledge Engineering: Principles and Methods. *Data and Knowledge Engineering* 25, 161–197.
- Takane, Y., Sergeant, J., 1983. Multidimensional scaling models for reaction times and same-different judgement. *Psychometrika* 48, 393–423.
- Talmy, L., 1983. How language structures space. *Spatial orientation: Theory research and application*. H. L. Pick and L. P. Acredole. New York, USA, Plenum Press. 225–282.
- Talmy, L., 1996. Fictive motion and change in language and perception. *Language and space*. P. Bloom, M. A. Peterson, L. Nadel and M. Garrett. Cambridge, MIT Press. 211–276.
- Thomas, M., Mareschal, D., 1997. Connectionism and Psychological Notions of Similarity. *Proceedings of the 19th Annual Conference of the Cognitive Science Society*.
- Torgerson, W.S., 1958. *Theory and Methods of Scaling*. New York, John Wiley and Sons.
- Torgerson, W.S., 1965. Multidimensional scaling of similarity. *Psychometrika* 30, 379–93.
- Torres Ruiz, M.J., 2007. Representación ontológica basada en descriptores semánticos aplicada a objetos geográficos.
- Tversky, A., 1977. Features of Similarity. *Psychological Review*. 84(4), 327–352.
- Tversky, B., P. U. Lee, 1998. How space structures language. *Spatial cognition: An interdisciplinary approach to representation and processing of spatial knowledge*, Springer. 157–175.
- Uschold, M., Gruninger, M., 1996. Ontologies: principles, methods and applications. *Knowledge Engineering Review* 11, 93–155.
- Vicedo, J.L., Ferrández, A., 2000. A semantic approach to Question Answering systems. Presented at the TREC-9.
- Vilches, L.M., Rodríguez, A.F., Bernabé, M.A., 2006. Aplicaciones de la ingeniería ontológica a la gestión y análisis de la información geográfica. Presented at the El acceso a la información espacial y las nuevas tecnologías geográficas., pp. 545–556.
- Visser, U., 2004. Intelligent information integration for the semantic web. *Lecture Notes in Computer Science* vol.3159, 150pp.
- Weigand, H., 1997. Multilingual Ontology-Based Lexicon for News Filtering –The TREVI Project. K. Mahesh.
- Wiederhold, G., 1999. Mediation to Deal with Heterogeneous Data Sources., in: Vckovski, A., Brassel, K., Schek, H.J. (Eds.), *Lecture Notes in Computer Science*. Presented at the Interoperating Geographic Information Systems - Second International Conference, INTEROP'99, Berlin: Springer-Verlag, Zurich, Switzerland, pp. 1–16.
- Wiener-Ehrlich, W.K., Bart, W.M., Millward, R., 1980. An analysis of generative representation systems. *Journal of Mathematical Psychology*. 21, 21946.
- Winter, S., 2000. Ontology in Geographical Information Science. Presented at the Euro Conference on Ontology and Epistemology for Spatial Data Standards - EURESCO.

- Worboys, M., 1992. A Geometric Model for Planar Geographical Objects. *International Journal of Geographical Information Systems*. 6(5), 353–372.
- Worboys, M.F., 2004. GIS: A Computing Perspective. CRC, 2004. Presented at the CRC.
- World Wide Web Consortium, (W3C). Available: <http://www.w3.org/>



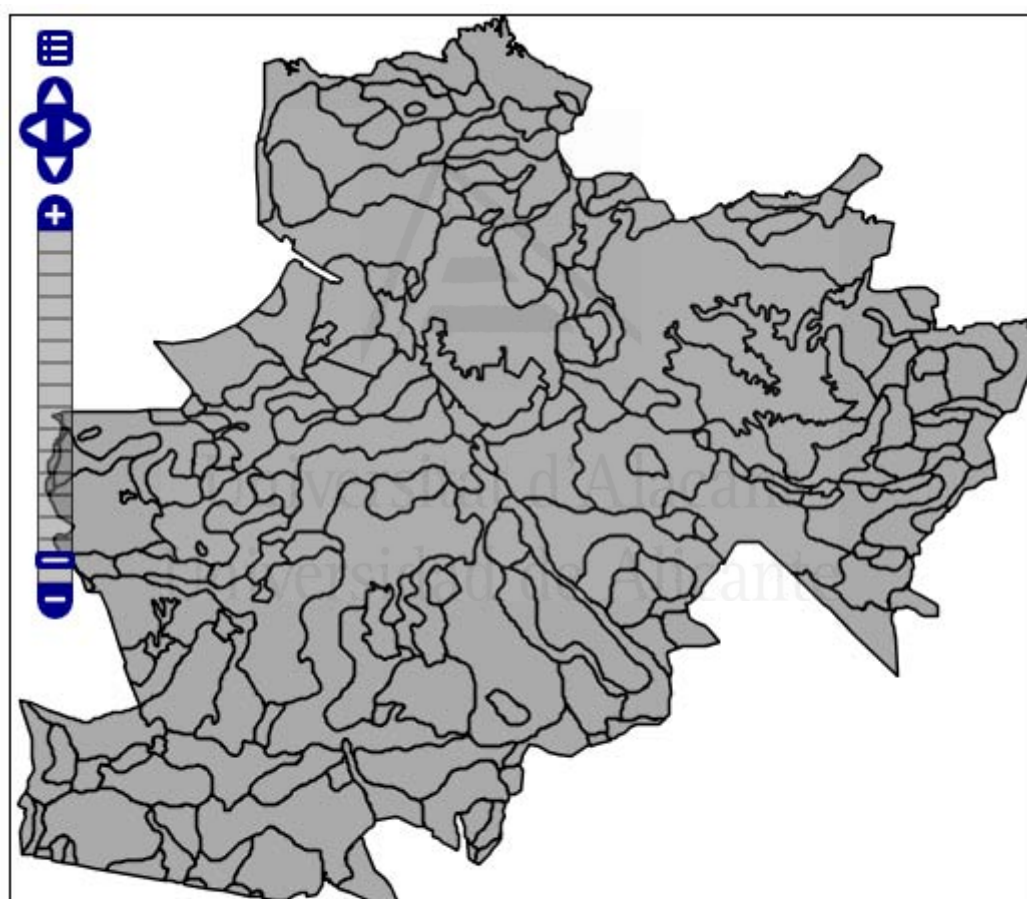
Universitat d'Alacant
Universidad de Alicante



Universitat d'Alacant
Universidad de Alicante

10. Anexos

Anexo 1. Mapa de suelos del municipio San José de Las Lajas.



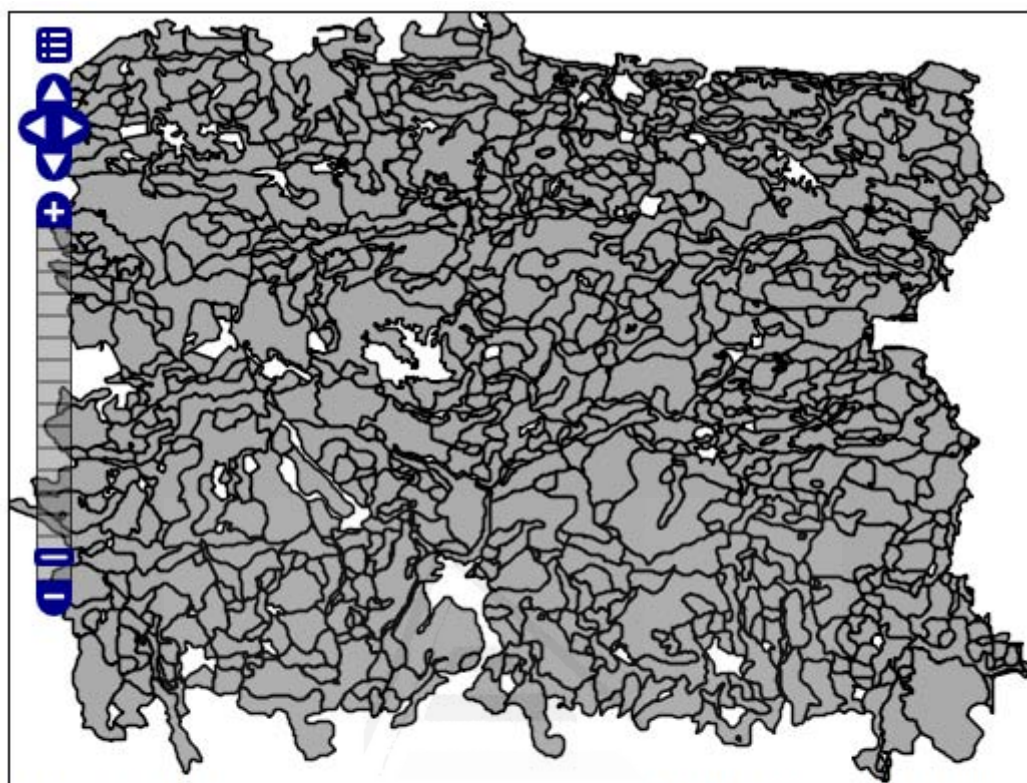
Scale = 1 : 245K

395054.34496, 350257.77769

SUELOS_SANJOSE

FID	CENTER ID	CLAVE	SUB_TIPO	M_BASAL	SATURA
suelos_sanjose.137	197	58	37841b255	Típico	Roca_ultrabásica Carbonatado

Anexo 2. Mapa de suelos de la provincia Mayabeque.



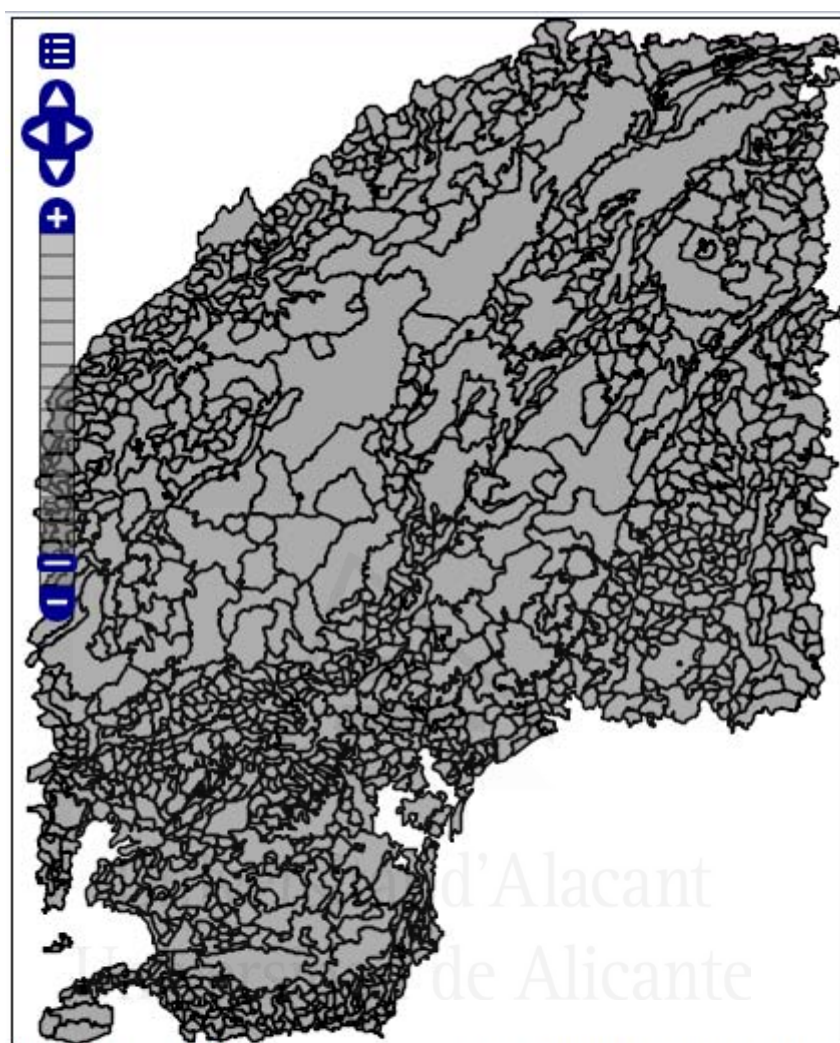
Scale = 1 : 492K

391925.59178, 351255.25077

SUELOS_MAYABEQUE

FID	ID	CLAVE	SUPERFICIE	TIPOID	TIPODES
suelos_mayabeque.483	84.0	3784Ia121	1881.67	2	Ferralítico_Rojo

Anexo 3. Mapa de suelos de la provincia Pinar del Río.



Scale = 1 : 595K

198739.57291, 278512.30322

SUELOS_PINAR

FID	TIPODES
suelos_pinar.721	Esquelético
suelos_pinar.722	Ferralítico_Cuarcítico_Amarrillo_Rojo_Lixiviado

Anexo 4. Plataforma de experimentación.



The screenshot displays the homepage of the SIGeografía platform. At the top, a banner features the title 'SIGeografía' and a subtitle 'Sistema de Gestión de Información Geográfica de suelos mediante una ontología de dominio'. Below the banner is a navigation bar with links: Inicio, Buscar, Galería, Contactos, and Ayuda. The main content area is divided into two sections. The left section, titled 'Bienvenido al Sistema!', contains a welcome message and a paragraph explaining the system's capabilities. The right section, titled 'Opciones del Sistema', includes a link to 'Visor de mapas de suelos' and a login form with fields for 'Usuario:' and 'Contraseña:', followed by an 'Aceptar' button. A footer at the bottom states 'Todos los Derechos Reservados a UNAH © 2014. Desarrollada y diseñada por Neili Machado García'.

¿Eres nuevo aquí? [Regístrate](#)

SIGeografía

Sistema de Gestión de Información Geográfica de suelos mediante una ontología de dominio

[Inicio](#) [Buscar](#) [Galería](#) [Contactos](#) [Ayuda](#)

Bienvenido al Sistema!

Este sistema brinda la posibilidad a todos los usuarios que naveguen por él, de hacer búsquedas de imágenes de suelos según sus necesidades e interés, además si estás logueado se almacenarán según la búsqueda realizada tus preferencias en el perfil. Tú también puedes formar parte de él, [registrándote](#) y así podrás disfrutar de todas las posibilidades que aquí se te brindan. Si tienes dudas y quieres informarte solo accede desde aquí a nuestra [ayuda](#) en la cual se explica cómo navegar y qué hacer durante la navegación.

Opciones del Sistema

» [Visor de mapas de suelos](#)

Iniciar Sesión

Usuario:

Contraseña:

[Aceptar](#)

Todos los Derechos Reservados a UNAH © 2014. Desarrollada y diseñada por Neili Machado García

Universitat d'Alacant
Universidad de Alicante

Anexo 5. Clasificación de los suelos según Nueva Versión de Clasificación Genética de los Suelos de Cuba.

Agrupamiento	Tipo Genético	Subtipo	Género
Suelos Alíticos	Alítico de Baja Actividad Arcillosa Rojo	- Típico - Ócrico - Nodular ferruginoso	- Corteza de intemperismo antigua - Eluvio de rocas básicas - Ferrítico
	Alítico de Baja Actividad Arcillosa Rojo Amarillento	- Típico - Ócrico - Gléyico	- Corteza de intemperismo antigua - Eluvio de rocas básicas
	Alítico de Baja Actividad Arcillosa Amarillento	- Típico - Nodular ferruginoso - Petroférrico - Gléyico - Húmico - Arénico	- Depósitos binarios - Cuarcítico - Eluvio de esquistos
	Alítico de Alta Actividad Arcillosa Rojo Amarillento	- Típico - Ócrico	- Corteza de intemperismo antigua - Eluvio de rocas madres
	Alítico de Alta Actividad Arcillosa Amarillento	- Típico - Nodular ferruginoso - Petroférrico - Gléyico - Húmico - Arénico	- Depósitos binarios - Cuarcítico - Eluvio de esquistos
Suelos Ferríticos	Ferrítico Rojo Oscuro	- Típico - Hidratado - Petroférrico - Húmico	- Corteza de intemperismo antigua - Eluvio de rocas ultrabásicas
	Ferrítico Amarillo	- Típico - Húmico	- Corteza de intemperismo antigua

Suelos Ferralíticos	Ferralítico Rojo	<ul style="list-style-type: none"> - Típico - Compactado - Hidratado - Nodular ferruginoso - Húmico 	<ul style="list-style-type: none"> - Eútrico - Dístrico
		<ul style="list-style-type: none"> - Gléyico 	
	Ferralítico Rojo Lixiviado	<ul style="list-style-type: none"> - Típico - Nodular ferruginoso - Petroférrico - Húmico - Arénico - Gléyico 	<ul style="list-style-type: none"> - Eútrico - Dístrico - Cuarcítico
	Ferralítico Amarillento Lixiviado	<ul style="list-style-type: none"> - Típico - Petroférrico - Gléyico - Gleýico-Petroférrico - Húmico - Húmico-Petroférrico - Arénico - Arénico-Húmico 	<ul style="list-style-type: none"> - Eútrico - Dístrico - Depósitos binarios - Cuarcítico
Suelos Ferrálicos	Ferrálico Rojo	<ul style="list-style-type: none"> - Típico - Compactado - Húmico - Hidratado 	<ul style="list-style-type: none"> - Eútrico - Dístrico
	Ferrálico Amarillento	<ul style="list-style-type: none"> - Típico - Lixivado - Gléyico - Húmico - Arénico - Arénico-Húmico - Nodular ferruginoso 	<ul style="list-style-type: none"> - Eútrico - Dístrico - Cuarcítico
Suelos Fersialíticos	Fersialítico Pardo Rojizo	<ul style="list-style-type: none"> - Mullido - Ócrico - Lixiviado 	<ul style="list-style-type: none"> - Eútrico - Dístrico - Ferromagnésico - Lítico - Paralítico - Con carbonatos

	Fersialítico Rojo	- Mullido - Ócrico - Lixiviado	- Eútrico - Dístrico - Lítico - Paralítico - Con carbonatos
	Fersialítico Amarillento	- Mullido - Ócrico	- Eútrico - Dístrico - Lítico - Paralítico
Suelos Pardos Sialíticos	Pardo	- Mullido - Ócrico - Cálcico - Gléyico - Gleyzoso	- Carbonatado - Medianamente lavado - Lavado - Sin carbonatos
		- Vértico - Gleyzoso y Vértico	- Ferromagnesial - Lítico - Paralítico
	Pardo Grisáceo	- Mullido - Ócrico - Nodular ferruginoso - Gléyico	- Eútrico - Dístrico - Cuarcítico- feldespático - Lítico - Paralítico
Suelos Húmicos Sailíticos	Húmico Calcimórfico	- Típico - Gléyico - Vértico - Gleyzoso - Gleyzoso y Vértico	- Carbonatado - Medianamente lavado - Lavado - Lítico - Paralítico
	Rendzina	- Negra - Roja	- Carbonatado y lítico - Carbonatado y paralítico
Vertisoles	Vertisol Pélico	- Típico - Mullido - Cálcico - Gléyico - Gléyico en profundidad - Gleyzoso	- Carbonatado - Medianamente lavado - Lavado - Sin carbonatos - Salinizado - Sódico
	Vertisol Crómico	- Típico - Mullido - Cálcico - Nodular ferruginoso - Gléyico - Gléyico en	- Carbonatado - Medianamente lavado - Lavado - Sin carbonatos - Salinizado - Sódico

		profundidad	
Suelos Hidromórficos	Gley Vértico	- Típico - Mullido - Cálcico - Crómico - Crómico- Nodular ferruginoso	- Carbonatado - Medianamente lavado - Lavado - Sin carbonatos - Salinizado - Sódico
	Gley Húmico	- Típico - Mullido - Cálcico - Nodular ferruginoso -Turboso	- Carbonatado - Medianamente lavado - Lavado - Sin carbonatos - Salinizado - Sódico
	Gley Nodular	- Típico	- Eútrico
	Ferruginoso	- Álbico - Petroférrico - Álbico- Petroférrico - Húmico - Húmico- Petroférrico - Arénico - Arénico-Húmico - Arénico- Petroférrico	- Dístrico - Cuarcítico - Ferrálico - Alítico - Fersialítico - Salinizado - Sódico
Suelos Halomórficos	Salino	- Típico - Mullido - Cálcico - Gléyico - Gléyico en profundidad - Salino en profundidad	- Carbonatado - Medianamente lavado - Lavado - Por el tipo de salinidad - Sódico
	Sódico	- Típico - Mullido - Gléyico - Sódico en profundidad - Arénico	- Carbonatado - Medianamente lavado - Lavado - Salinizado

Fluvisoles	Fluvisol	<ul style="list-style-type: none"> - Típico - Diferenciado - Mullido - Cálcico - Gléyico - Gléyico en profundidad - Vértico - Arénico 	<ul style="list-style-type: none"> - Eútrico - Dístrico - Carbonatado - Medianamente lavado - Lavado - Salinizado - Sódico - Cuarácítico
Histosoles	Histosol Fíbrico	<ul style="list-style-type: none"> - Típico - Térrico - Hídrico - Salinizado - Límnico - Sulfídico - Sulfático 	<ul style="list-style-type: none"> - Por el grado de lavado de los carbonatos - Por el grado de salinidad - Lítico - Paralítico
	Histosol Mésico	<ul style="list-style-type: none"> - Típico - Térrico - Hídrico - Salinizado - Límnico - Sulfídico - Sulfático 	<ul style="list-style-type: none"> - Por el grado de lavado de los carbonatos - Por el grado de salinidad - Lítico - Paralítico
	Histosol Sáprico	<ul style="list-style-type: none"> - Típico 	<ul style="list-style-type: none"> - Por el grado de
		<ul style="list-style-type: none"> - Térrico - Hídrico - Salinizado - Límnico - Sulfídico - Sulfático 	<ul style="list-style-type: none"> lavado de los carbonatos - Por el grado de salinidad - Lítico - Paralítico
Suelos Poco Evolucionados	Arenosol	<ul style="list-style-type: none"> - Típico - Húmico - Petrocálcico - Gléyico 	<ul style="list-style-type: none"> - Cuarácítico - Carbonatado
	Lithosol	<ul style="list-style-type: none"> - Eútrico - Dístrico 	<ul style="list-style-type: none"> - Por el material de origen
	Protorrencia	<ul style="list-style-type: none"> - Negra - Roja 	<ul style="list-style-type: none"> - Lítico
Antrosoles	Salino Antrópico	<ul style="list-style-type: none"> - Mullido - Cálcico - Gléyico - Gléyico en profundidad - Vértico 	<ul style="list-style-type: none"> - Carbonatado - Medianamente lavado - Lavado - Por el tipo de salinidad - Sódico

	Hidromórfico Antrópico	- Mullido - Cálcico - Gléyico - Gléyico en profundidad - Vértico	
	Recultivado Antrópico	- Turboso - Residual	



Universitat d'Alacant
Universidad de Alicante

Anexo 6. Ontología geoespacial.

En este anexo se muestra un fragmento de la ontología generada según lo que se explica en el Propuesta y diseño de la ontología geoespacial.

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

<!DOCTYPE rdf:RDF [
  <!ENTITY owl "http://www.w3.org/2002/07/owl#" >
  <!ENTITY xsd "http://www.w3.org/2001/XMLSchema#" >
  <!ENTITY owl2xml "http://www.w3.org/2006/12/owl2-xml#" >
  <!ENTITY rdfs "http://www.w3.org/2000/01/rdf-schema#" >
  <!ENTITY rdf "http://www.w3.org/1999/02/22-rdf-syntax-ns#"
>
  <!ENTITY Ontology1357643418249
"http://www.semanticweb.org/ontologies/2013/0/Ontology13576434
18249.owl#" >
    <!ENTITY Pedregosidad_moderadamente
"http://www.semanticweb.org/ontologies/2013/0/Ontology13576434
18249.owl#Pedregosidad_moderadamente:" >
]>

<rdf:RDF
xmlns="http://www.semanticweb.org/ontologies/2013/0/Ontology13
57643418249.owl#"

xml:base="http://www.semanticweb.org/ontologies/2013/0/Ontolog
y1357643418249.owl"
xmlns:owl2xml="http://www.w3.org/2006/12/owl2-xml#"

xmlns:Ontology1357643418249="http://www.semanticweb.org/ontolo
gies/2013/0/Ontology1357643418249.owl#"
xmlns:xsd="http://www.w3.org/2001/XMLSchema#"
xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
xmlns:owl="http://www.w3.org/2002/07/owl#">
<owl:Ontology rdf:about=""/>

<!--

////////////////////////////////////
////////////////////////////////////
//
// Object Properties
//

////////////////////////////////////
////////////////////////////////////
```

```
<!--
http://www.semanticweb.org/ontologies/2013/0/Ontology135764341
8249.owl#concrecion -->
```

```
<owl:ObjectProperty rdf:about="#concrecion">
  <rdfs:subPropertyOf rdf:resource="#datos"/>
</owl:ObjectProperty>
```

```
<!--
http://www.semanticweb.org/ontologies/2013/0/Ontology135764341
8249.owl#erosion -->
```

```
<owl:ObjectProperty rdf:about="#erosion">
  <rdfs:subPropertyOf rdf:resource="#datos"/>
</owl:ObjectProperty>
```

```
<!--
http://www.semanticweb.org/ontologies/2013/0/Ontology135764341
8249.owl#gleyzado -->
```

```
<owl:ObjectProperty rdf:about="#gleyzado">
  <rdfs:subPropertyOf rdf:resource="#datos"/>
</owl:ObjectProperty>
```

```
<!--
http://www.semanticweb.org/ontologies/2013/0/Ontology135764341
8249.owl#salinidad -->
```

```
<owl:ObjectProperty rdf:about="#salinidad">
  <rdfs:subPropertyOf rdf:resource="#datos"/>
</owl:ObjectProperty>
```

```
<!--
http://www.semanticweb.org/ontologies/2013/0/Ontology135764341
8249.owl#satura -->
```

```
<owl:ObjectProperty rdf:about="#satura">
  <rdfs:subPropertyOf rdf:resource="#datos"/>
</owl:ObjectProperty>
////////////////////////////////////
////////////////////////////////////
//
```

```

// Classes
//

////////////////////////////////////
////////////////////////////////////
-->

<!--
http://www.semanticweb.org/ontologies/2013/0/Ontology135764341
8249.owl#Alitico -->

<owl:Class rdf:about="#Alitico">
  <rdfs:subClassOf rdf:resource="#suelos"/>
</owl:Class>

<!--
http://www.semanticweb.org/ontologies/2013/0/Ontology135764341
8249.owl#Alta_Actividad_Arcillosa_Amarillento -->

<owl:Class
rdf:about="#Alta_Actividad_Arcillosa_Amarillento">
  <rdfs:subClassOf rdf:resource="#Alitico"/>
</owl:Class>

<!--
http://www.semanticweb.org/ontologies/2013/0/Ontology135764341
8249.owl#Alta_Actividad_Arcillosa_Rojo_Amarillento -->

<owl:Class
rdf:about="#Alta_Actividad_Arcillosa_Rojo_Amarillento">
  <rdfs:subClassOf rdf:resource="#Alitico"/>
</owl:Class>

<!--
http://www.semanticweb.org/ontologies/2013/0/Ontology135764341
8249.owl#Antrosol -->

<owl:Class rdf:about="#Antrosol">
  <rdfs:subClassOf rdf:resource="#suelos"/>
</owl:Class>

```



```

<!--
http://www.semanticweb.org/ontologies/2013/0/Ontology135764341
8249.owl#Arenosol -->

```

```

<owl:Class rdf:about="#Arenosol">
  <rdfs:subClassOf rdf:resource="#Poco_Evolucionado"/>
</owl:Class>

```

```

<!--
http://www.semanticweb.org/ontologies/2013/0/Ontology135764341
8249.owl#Baja_Actividad_Arcillosa_Amarillento -->

```

```

<owl:Class
rdf:about="#Baja_Actividad_Arcillosa_Amarillento">
  <rdfs:subClassOf rdf:resource="#Alitico"/>
</owl:Class>

```

```

<!--
http://www.semanticweb.org/ontologies/2013/0/Ontology135764341
8249.owl#Baja_Actividad_Arcillosa_Rojo -->

```

```

<owl:Class rdf:about="#Baja_Actividad_Arcillosa_Rojo">
  <rdfs:subClassOf rdf:resource="#Alitico"/>
</owl:Class>

```

```

<!--
http://www.semanticweb.org/ontologies/2013/0/Ontology135764341
8249.owl#Baja_Actividad_Arcillosa_Rojo_Amarillento -->

```

```

<owl:Class
rdf:about="#Baja_Actividad_Arcillosa_Rojo_Amarillento">
  <rdfs:subClassOf rdf:resource="#Alitico"/>
</owl:Class>

```

```

<!--
http://www.semanticweb.org/ontologies/2013/0/Ontology135764341
8249.owl#Ferralico -->

```

```

<owl:Class rdf:about="#Ferralico">
  <rdfs:subClassOf rdf:resource="#suelos"/>
</owl:Class>

```

```

<!--
http://www.semanticweb.org/ontologies/2013/0/Ontology135764341
8249.owl#Ferralico_Amarillento -->

```

```

<owl:Class rdf:about="#Ferralico_Amarillento">
  <rdfs:subClassOf rdf:resource="#Ferralico"/>
</owl:Class>

```

```

<!--
http://www.semanticweb.org/ontologies/2013/0/Ontology135764341
8249.owl#Ferralico_Rojo -->

```

```

<owl:Class rdf:about="#Ferralico_Rojo">
  <rdfs:subClassOf rdf:resource="#Ferralico"/>
</owl:Class>

```

```

<!--
http://www.semanticweb.org/ontologies/2013/0/Ontology135764341
8249.owl#Ferralitico -->

```

```

<owl:Class rdf:about="#Ferralitico">
  <rdfs:subClassOf rdf:resource="#suelos"/>
</owl:Class>

```

```

<!--
http://www.semanticweb.org/ontologies/2013/0/Ontology135764341
8249.owl#Ferralitico_Rojo -->

```

```

<owl:Class rdf:about="#Ferralitico_Rojo">
  <rdfs:subClassOf rdf:resource="#Ferralitico"/>
</owl:Class>

```

```

<!--
http://www.semanticweb.org/ontologies/2013/0/Ontology135764341
8249.owl#Gley_Humico -->

```

```

<owl:Class rdf:about="#Gley_Humico">
  <rdfs:subClassOf rdf:resource="#Hidromorfico"/>
</owl:Class>

```

```
<!--
http://www.semanticweb.org/ontologies/2013/0/Ontology135764341
8249.owl#Gley_Nodular_Ferruginoso -->
```

```
<owl:Class rdf:about="#Gley_Nodular_Ferruginoso">
  <rdfs:subClassOf rdf:resource="#Hidromorfico"/>
</owl:Class>
```

```
<!--
http://www.semanticweb.org/ontologies/2013/0/Ontology135764341
8249.owl#Gley_Vertico -->
```

```
<owl:Class rdf:about="#Gley_Vertico">
  <rdfs:subClassOf rdf:resource="#Hidromorfico"/>
</owl:Class>
```

```
<!--
http://www.semanticweb.org/ontologies/2013/0/Ontology135764341
8249.owl#Pardo -->
```

```
<owl:Class rdf:about="#Pardo">
  <rdfs:subClassOf rdf:resource="#Pardo_Sialitico"/>
</owl:Class>
```

```
<!--
http://www.semanticweb.org/ontologies/2013/0/Ontology135764341
8249.owl#Pardo_Gris&#225;ceo -->
```

```
<owl:Class rdf:about="#Pardo_Gris&#225;ceo">
  <rdfs:subClassOf rdf:resource="#Pardo_Sialitico"/>
</owl:Class>
```

```
<!--
http://www.semanticweb.org/ontologies/2013/0/Ontology135764341
8249.owl#Pardo_Sialitico -->
```

```
<owl:Class rdf:about="#Pardo_Sialitico">
  <rdfs:subClassOf rdf:resource="#suelos"/>
</owl:Class>
```

```
////////////////////////////////////
////////////////////////////////////
```

```

//
// Individuals
//

////////////////////////////////////
////////////////////////////////////
-->

<!--
http://www.semanticweb.org/ontologies/2013/0/Ontology135764341
8249.owl#1 -->

<owl:Thing rdf:about="#1">
  <rdf:type rdf:resource="#Pardo"/>
  <Profefec>15</Profefec>
  <concrecion rdf:resource="#Concrec_Sin_Concreciones"/>
  <erosion rdf:resource="#Erosion_Poca"/>
  <gravillosidad
rdf:resource="#Gravillosidad_Sin_Gravillosidad"/>
  <humificado rdf:resource="#Humificado_Humificado"/>
  <lavado rdf:resource="#Lavado_Carbonatado"/>
  <m_basal rdf:resource="#MBasal_Roca_ignea_basica"/>
  <pend_pred
rdf:resource="#PendPred_Ligeramente_Ondulado"/>
  <prof_pedo rdf:resource="#ProfPedo_Poco_Profundo"/>
  <rocosidad rdf:resource="#Rocosidad_sin_rocosidad"/>
  <salinidad rdf:resource="#Salinidad_sin_salinidad"/>
  <satura rdf:resource="#Satura_Carbonatado"/>
  <subtipo rdf:resource="#Subtipo_Tipico"/>
  <textura
rdf:resource="#Textura_Arcilla_Montmorillonitica"/>
</owl:Thing>

<!--
http://www.semanticweb.org/ontologies/2013/0/Ontology135764341
8249.owl#10 -->

<owl:Thing rdf:about="#10">
  <rdf:type rdf:resource="#Pardo"/>
  <Profefec>16</Profefec>
  <concrecion rdf:resource="#Concrec_Sin_Concreciones"/>
  <erosion rdf:resource="#Erosion_Poca"/>
  <gleyzado rdf:resource="#Gleyzado_No_Gleyzado"/>
  <gravillosidad rdf:resource="#Gravillosidad_Mediana"/>
  <humificado rdf:resource="#Humificado_Humificado"/>
  <lavado rdf:resource="#Lavado_Carbonatado"/>
  <m_basal rdf:resource="#MBasal_Roca_ignea_basica"/>
  <pend_pred rdf:resource="#PendPred_Ondulado"/>
  <prof_pedo rdf:resource="#ProfPedo_Poco_Profundo"/>

```

```

        <rocosidad rdf:resource="#Rocosidad_sin_rocosidad"/>
        <salinidad rdf:resource="#Salinidad_sin_salinidad"/>
        <satura rdf:resource="#Satura_Carbonatado"/>
        <subtipo rdf:resource="#Subtipo_Tipico"/>
        <textura
rdf:resource="#Textura_Arcilla_Montmorillonitica"/>
    </owl:Thing>

<!--
http://www.semanticweb.org/ontologies/2013/0/Ontology135764341
8249.owl#104 -->

    <owl:Thing rdf:about="#104">
        <rdf:type rdf:resource="#Pardo"/>
        <Profefec>43</Profefec>
        <concrecion rdf:resource="#Concrec_Sin_Concreciones"/>
        <erosion rdf:resource="#Erosion_Poca"/>
        <gleyzado rdf:resource="#Gleyzado_No_Gleyzado"/>
        <gravillosidad
rdf:resource="#Gravillosidad_Muy_Poca"/>
        <humificado
rdf:resource="#Humificado_Medianamente_Humificado"/>
        <lavado rdf:resource="#Lavado_Carbonatado"/>
        <m_basal rdf:resource="#MBasal_Roca_ignea_basica"/>
        <pedregosidad
rdf:resource="#Pedregosidad_Muy_pedregoso"/>
        <pend_pred
rdf:resource="#PendPred_Ligeramente_Ondulado"/>
        <prof_pedo
rdf:resource="#ProfPedo_Medianamente_Profundo"/>
        <rocosidad rdf:resource="#Rocosidad_sin_rocosidad"/>
        <salinidad rdf:resource="#Salinidad_sin_salinidad"/>
        <satura rdf:resource="#Satura_Carbonatado"/>
        <subtipo rdf:resource="#Subtipo_Tipico"/>
        <textura
rdf:resource="#Textura_Arcilla_Montmorillonitica"/>
    </owl:Thing>

<!--
http://www.semanticweb.org/ontologies/2013/0/Ontology135764341
8249.owl#11 -->

    <owl:Thing rdf:about="#11">
        <Profefec>40</Profefec>
        <concrecion rdf:resource="#Concrec_Sin_Concreciones"/>
        <erosion rdf:resource="#Erosion_Poca"/>
        <gleyzado rdf:resource="#Gleyzado_No_Gleyzado"/>

```

```

        <gravillosidad
rdf:resource="#Gravillosidad_Sin_Gravillosidad"/>
        <humificado rdf:resource="#Humificado_Humificado"/>
        <lavado rdf:resource="#Lavado_Carbonatado"/>
        <m_basal rdf:resource="#MBasal_Roca_ultrabasica"/>
        <pend_pred rdf:resource="#PendPred_Ondulado"/>
        <prof_pedo
rdf:resource="#ProfPedo_Medianamente_Profundo"/>
        <rocosidad rdf:resource="#Rocosidad_sin_rocosidad"/>
        <salinidad rdf:resource="#Salinidad_sin_salinidad"/>
        <satura rdf:resource="#Satura_Carbonatado"/>
        <subtipo rdf:resource="#Subtipo_Tipico"/>
        <textura
rdf:resource="#Textura_Arcilla_Montmorillonitica"/>
    </owl:Thing>

    <!--
http://www.semanticweb.org/ontologies/2013/0/Ontology135764341
8249.owl#110 -->

    <owl:Thing rdf:about="#110">
        <rdf:type rdf:resource="#Fersialitico_Pardo_Rojizo"/>
        <Profefec>46</Profefec>
        <concrecion rdf:resource="#Concrec_Sin_Concreciones"/>
        <erosion rdf:resource="#Erosion_Poca"/>
        <gleyzado rdf:resource="#Gleyzado_No_Gleyzado"/>
        <gravillosidad
rdf:resource="#Gravillosidad_Sin_Gravillosidad"/>
        <humificado
rdf:resource="#Humificado_Medianamente_Humificado"/>
        <m_basal
rdf:resource="#MBasal_materiales_transportados_arcillosos_gene
ralmente"/>
        <pedregosidad
rdf:resource="#Pedregosidad_Sin_pedregosidad"/>
        <pend_pred
rdf:resource="#PendPred_Ligeramente_Ondulado"/>
        <prof_pedo
rdf:resource="#ProfPedo_Medianamente_Profundo"/>
        <rocosidad rdf:resource="#Rocosidad_sin_rocosidad"/>
        <salinidad rdf:resource="#Salinidad_sin_salinidad"/>
        <satura rdf:resource="#Satura_Saturado"/>
        <subtipo rdf:resource="#Subtipo_Tipico"/>
        <textura rdf:resource="#Textura_Arcilla"/>
    </owl:Thing>

```

```

<!--
http://www.semanticweb.org/ontologies/2013/0/Ontology135764341
8249.owl#119 -->

```

```

<owl:Thing rdf:about="#119"/>

```

```

<!--
http://www.semanticweb.org/ontologies/2013/0/Ontology135764341
8249.owl#12 -->

```

```

<owl:Thing rdf:about="#12">
  <rdf:type
rdf:resource="#Ferralitico_Amarillento_Lixiviado"/>
  <Profefec>67</Profefec>
  <concrecion rdf:resource="#Concrec_Sin_Concreciones"/>
  <erosion rdf:resource="#Erosion_Poca"/>
  <gleyzado rdf:resource="#Gleyzado_No_Gleyzado"/>
  <gravillosidad
rdf:resource="#Gravillosidad_Sin_Gravillosidad"/>
  <humificado
rdf:resource="#Humificado_Medianamente_Humificado"/>
  <m_basal
rdf:resource="#MBasal_materiales_transportados_arcillosos_gene
ralmente"/>
  <pend_pred rdf:resource="#PendPred_Casi_llano"/>
  <prof_pedo rdf:resource="#ProfPedo_Profundo"/>
  <rocosidad rdf:resource="#Rocosidad_sin_rocosidad"/>
  <salinidad rdf:resource="#Salinidad_sin_salinidad"/>
  <subtipo rdf:resource="#Subtipo_Tipico"/>
  <textura rdf:resource="#Textura_Arcilla_Caolinitica"/>
</owl:Thing>

```

```

<!--
http://www.semanticweb.org/ontologies/2013/0/Ontology135764341
8249.owl#123 -->

```

```

<owl:Thing rdf:about="#123">
  <rdf:type
rdf:resource="#Ferralitico_Amarillento_Lixiviado"/>
  <Profefec>88</Profefec>
  <concrecion rdf:resource="#Concrec_Sin_Concreciones"/>
  <erosion rdf:resource="#Erosion_Poca"/>
  <gleyzado rdf:resource="#Gleyzado_No_Gleyzado"/>
  <gravillosidad
rdf:resource="#Gravillosidad_Sin_Gravillosidad"/>
  <humificado
rdf:resource="#Humificado_Medianamente_Humificado"/>

```

```

        <m_basal
rdf:resource="#MBasal_materiales_transportados_arcillosos_gene
ralmente"/>
        <pedregosidad
rdf:resource="#Pedregosidad_Sin_pedregosidad"/>
        <pend_pred rdf:resource="#PendPred_Casi_llano"/>
        <prof_pedo rdf:resource="#ProfPedo_Profundo"/>
        <rocosidad rdf:resource="#Rocosidad_sin_rocosidad"/>
        <salinidad rdf:resource="#Salinidad_sin_salinidad"/>
        <satura
rdf:resource="#Satura_Medianamente_desaturado"/>
        <subtipo rdf:resource="#Subtipo_Tipico"/>
        <textura rdf:resource="#Textura_Arcilla_Caolinitica"/>
</owl:Thing>

```

```

<!--
http://www.semanticweb.org/ontologies/2013/0/Ontology135764341
8249.owl#14 -->

```

```

<owl:Thing rdf:about="#14">
    <rdf:type rdf:resource="#Ferralitico_Rojo"/>
    <Profefec>35</Profefec>
    <concrecion rdf:resource="#Concrec_Sin_Concreciones"/>
    <erosion rdf:resource="#Erosion_Poca"/>
    <gleyzado rdf:resource="#Gleyzado_No_Gleyzado"/>
    <gravillosidad
rdf:resource="#Gravillosidad_Sin_Gravillosidad"/>
    <humificado
rdf:resource="#Humificado_Medianamente_Humificado"/>
    <m_basal
rdf:resource="#MBasal_materiales_transportados_arcillosos_gene
ralmente"/>
    <pend_pred rdf:resource="#PendPred_Casi_llano"/>
    <prof_pedo
rdf:resource="#ProfPedo_Medianamente_Profundo"/>
    <rocosidad rdf:resource="#Rocosidad_sin_rocosidad"/>
    <salinidad rdf:resource="#Salinidad_sin_salinidad"/>
    <satura rdf:resource="#Satura_Saturado"/>
    <subtipo rdf:resource="#Subtipo_Tipico"/>
    <textura rdf:resource="#Textura_Arcilla_Caolinitica"/>
</owl:Thing>

```

```

<!--
http://www.semanticweb.org/ontologies/2013/0/Ontology135764341
8249.owl#140 -->

```

```

<owl:Thing rdf:about="#140">
    <rdf:type rdf:resource="#Ferralitico_Rojo"/>

```



```

    <Profefec>32</Profefec>
    <concrecion rdf:resource="#Concrec_Sin_Concreciones"/>
    <erosion rdf:resource="#Erosion_Poca"/>
    <gleyzado rdf:resource="#Gleyzado_No_Gleyzado"/>
    <gravillosidad
rdf:resource="#Gravillosidad_Sin_Gravillosidad"/>
    <humificado
rdf:resource="#Humificado_Medianamente_Humificado"/>
    <m_basal
rdf:resource="#MBasal_materiales_transportados_arcillosos_gene
ralmente"/>
    <pedregosidad
rdf:resource="#Pedregosidad_Sin_pedregosidad"/>
    <pend_pred rdf:resource="#PendPred_Casi_llano"/>
    <prof_pedo
rdf:resource="#ProfPedo_Medianamente_Profundo"/>
    <rocosidad rdf:resource="#Rocosidad_sin_rocosidad"/>
    <salinidad rdf:resource="#Salinidad_sin_salinidad"/>
    <satura
rdf:resource="#Satura_Medianamente_desaturado"/>
    <subtipo rdf:resource="#Subtipo_Tipico"/>
    <textura rdf:resource="#Textura_Arcilla_Caolinitica"/>
</owl:Thing>

```

```

<!--
http://www.semanticweb.org/ontologies/2013/0/Ontology135764341
8249.owl#148 -->

```

```

<owl:Thing rdf:about="#148">
    <rdf:type rdf:resource="#Rendzina"/>
    <Profefec>37</Profefec>
    <concrecion rdf:resource="#Concrec_Sin_Concreciones"/>
    <erosion rdf:resource="#Erosion_Poca"/>
    <gleyzado rdf:resource="#Gleyzado_No_Gleyzado"/>
    <gravillosidad rdf:resource="#Gravillosidad_Mediana"/>
    <humificado rdf:resource="#Humificado_Humificado"/>
    <m_basal
rdf:resource="#MBasal_materiales_transportados_arcillosos_gene
ralmente"/>
    <pedregosidad
rdf:resource="#Pedregosidad_Sin_pedregosidad"/>
    <pend_pred rdf:resource="#PendPred_Casi_llano"/>
    <prof_pedo
rdf:resource="#ProfPedo_Medianamente_Profundo"/>
    <rocosidad rdf:resource="#Rocosidad_sin_rocosidad"/>
    <salinidad rdf:resource="#Salinidad_sin_salinidad"/>
    <satura rdf:resource="#Satura_Saturado"/>
    <subtipo rdf:resource="#Subtipo_Tipico"/>
    <textura rdf:resource="#Textura_Arcilla"/>
</owl:Thing>

```

```
<!--
http://www.semanticweb.org/ontologies/2013/0/Ontology135764341
8249.owl#155 -->
```

```
<owl:Thing rdf:about="#155">
  <rdf:type rdf:resource="#Rendzina"/>
  <Profefec>23</Profefec>
  <concrecion rdf:resource="#Concrec_Sin_Concreciones"/>
  <erosion rdf:resource="#Erosion_Poca"/>
  <gleyzado rdf:resource="#Gleyzado_No_Gleyzado"/>
  <gravillosidad rdf:resource="#Gravillosidad_Mediana"/>
  <humificado rdf:resource="#Humificado_Humificado"/>
  <gleyzado rdf:resource="#Lavado_Carbonatado"/>
  <m_basal
rdf:resource="#MBasal_materiales_transportados_arcillosos_gene
ralmente"/>
  <pedregosidad
rdf:resource="#Pedregosidad_Sin_pedregosidad"/>
  <pend_pred
rdf:resource="#PendPred_Ligeramente_Ondulado"/>
  <prof_pedo
rdf:resource="#ProfPedo_Medianamente_Profundo"/>
  <rocosidad rdf:resource="#Rocosidad_sin_rocosidad"/>
  <salinidad rdf:resource="#Salinidad_sin_salinidad"/>
  <satura rdf:resource="#Satura_Carbonatado"/>
  <subtipo rdf:resource="#Subtipo_Tipico"/>
  <textura rdf:resource="#Textura_Arcilla"/>
</owl:Thing>
```

```
<!--
http://www.semanticweb.org/ontologies/2013/0/Ontology135764341
8249.owl#201 -->
```

```
<owl:Thing rdf:about="#201">
  <rdf:type rdf:resource="#Pardo"/>
  <Profefec>15</Profefec>
  <concrecion rdf:resource="#Concrec_Sin_Concreciones"/>
  <erosion rdf:resource="#Erosion_Mediana"/>
  <gleyzado rdf:resource="#Gleyzado_No_Gleyzado"/>
  <gravillosidad rdf:resource="#Gravillosidad_Mediana"/>
  <humificado
rdf:resource="#Humificado_Poco_Humificado"/>
  <lavado rdf:resource="#Lavado_Carbonatado"/>
  <m_basal rdf:resource="#MBasal_Roca_ultrabasica"/>
  <pedregosidad
rdf:resource="#Pedregosidad_Sin_pedregosidad"/>
```

```

        <pend_pred
rdf:resource="#PendPred_Fuertemente_ondulado"/>
        <prof_pedo rdf:resource="#ProfPedo_Poco_Profundo"/>
        <rocosidad rdf:resource="#Rocosisad_sin_rocosidad"/>
        <salinidad rdf:resource="#Salinidad_sin_salinidad"/>
        <satura rdf:resource="#Satura_Carbonatado"/>
        <subtipo rdf:resource="#Subtipo_Tipico"/>
        <textura
rdf:resource="#Textura_Arcilla_Montmorillonitica"/>
    </owl:Thing>

<!--
http://www.semanticweb.org/ontologies/2013/0/Ontology135764341
8249.owl#202 -->

    <owl:Thing rdf:about="#202">
        <rdf:type rdf:resource="#Ferralitico_Rojo"/>
        <Profefec>108</Profefec>
        <concrecion rdf:resource="#Concrec_Sin_Concreciones"/>
        <erosion rdf:resource="#Erosion_Poca"/>
        <gleyzado rdf:resource="#Gleyzado_No_Gleyzado"/>
        <gravillosidad
rdf:resource="#Gravillosidad_Sin_Gravillosidad"/>
        <humificado
rdf:resource="#Humificado_Medianamente_Humificado"/>
        <m_basal
rdf:resource="#MBasal_materiales_transportados_arcillosos_gene
ralmente"/>
        <pedregosidad
rdf:resource="#Pedregosidad_Sin_pedregosidad"/>
        <pend_pred rdf:resource="#PendPred_Casi_llano"/>
        <prof_pedo
rdf:resource="#ProfPedo_Medianamente_Profundo"/>
        <rocosidad rdf:resource="#Rocosisad_sin_rocosidad"/>
        <salinidad rdf:resource="#Salinidad_sin_salinidad"/>
        <satura
rdf:resource="#Satura_Medianamente_desaturado"/>
        <subtipo rdf:resource="#Subtipo_Tipico"/>
        <textura rdf:resource="#Textura_Arcilla_Caolinitica"/>
    </owl:Thing>

<!--
http://www.semanticweb.org/ontologies/2013/0/Ontology135764341
8249.owl#22 -->

    <owl:Thing rdf:about="#22">
        <rdf:type rdf:resource="#Fersialitico_Pardo_Rojizo"/>
        <Profefec>15</Profefec>
        <concrecion rdf:resource="#Concrec_Sin_Concreciones"/>

```

```

        <erosion rdf:resource="#Erosion_Poca"/>
        <gleyzado rdf:resource="#Gleyzado_No_Gleyzado"/>
        <gravillosidad
rdf:resource="#Gravillosidad_Sin_Gravillosidad"/>
        <humificado
rdf:resource="#Humificado_Medianamente_Humificado"/>
        <m_basal
rdf:resource="#MBasal_materiales_transportados_arcillosos_gene
ralmente"/>
        <pend_pred
rdf:resource="#PendPred_Ligeramente_Ondulado"/>
        <prof_pedo rdf:resource="#ProfPedo_Poco_Profundo"/>
        <rocosidad rdf:resource="#Rocosidad_sin_rocosidad"/>
        <salinidad rdf:resource="#Salinidad_sin_salinidad"/>
        <satura rdf:resource="#Satura_Saturado"/>
        <subtipo rdf:resource="#Subtipo_Tipico"/>
        <textura rdf:resource="#Textura_Arcilla"/>
    </owl:Thing>

    <!--
http://www.semanticweb.org/ontologies/2013/0/Ontology135764341
8249.owl#30 -->

    <owl:Thing rdf:about="#30">
        <rdf:type rdf:resource="#Rendzina"/>
        <Profefec>30</Profefec>
        <concrecion rdf:resource="#Concrec_Sin_Concreciones"/>
        <erosion rdf:resource="#Erosion_Poca"/>
        <gleyzado rdf:resource="#Gleyzado_No_Gleyzado"/>
        <gravillosidad rdf:resource="#Gravillosidad_Mediana"/>
        <humificado rdf:resource="#Humificado_Humificado"/>
        <lavado rdf:resource="#Lavado_Carbonatado"/>
        <m_basal
rdf:resource="#MBasal_materiales_transportados_arcillosos_gene
ralmente"/>
        <pend_pred rdf:resource="#PendPred_Casi_llano"/>
        <prof_pedo
rdf:resource="#ProfPedo_Medianamente_Profundo"/>
        <rocosidad
rdf:resource="#Rocosidad_moderadamente_rocoso"/>
        <salinidad rdf:resource="#Salinidad_sin_salinidad"/>
        <satura
rdf:resource="#Satura_Medianamente_desaturado"/>
        <subtipo rdf:resource="#Subtipo_Tipico"/>
        <textura rdf:resource="#Textura_Arcilla"/>
    </owl:Thing>

```

```
<!--
http://www.semanticweb.org/ontologies/2013/0/Ontology135764341
8249.owl#42 -->
```

```
<owl:Thing rdf:about="#42">
  <rdf:type rdf:resource="#Fersialitico_Pardo_Rojizo"/>
  <Profefec>40</Profefec>
  <concrecion rdf:resource="#Concrec_Sin_Concreciones"/>
  <erosion rdf:resource="#Erosion_Poca"/>
  <gleyzado rdf:resource="#Gleyzado_No_Gleyzado"/>
  <gravillosidad rdf:resource="#Gravillosidad_Fuerte"/>
  <humificado
rdf:resource="#Humificado_Medianamente_Humificado"/>
  <m_basal
rdf:resource="#MBasal_materiales_transportados_arcillosos_gene
ralmente"/>
  <pend_pred
rdf:resource="#PendPred_Fuertemente_ondulado"/>
  <rocosidad rdf:resource="#Rocosidad_sin_rocosidad"/>
  <salinidad rdf:resource="#Salinidad_sin_salinidad"/>
  <satura rdf:resource="#Satura_Saturado"/>
  <subtipo rdf:resource="#Subtipo_Tipico"/>
  <textura rdf:resource="#Textura_Arcilla"/>
</owl:Thing>
```

```
<!--
http://www.semanticweb.org/ontologies/2013/0/Ontology135764341
8249.owl#56 -->
```

```
<owl:Thing rdf:about="#56">
  <rdf:type rdf:resource="#Ferralico_Rojo"/>
  <Profefec>15</Profefec>
  <concrecion rdf:resource="#Concrec_Sin_Concreciones"/>
  <erosion rdf:resource="#Erosion_Poca"/>
  <gleyzado rdf:resource="#Gleyzado_No_Gleyzado"/>
  <gravillosidad
rdf:resource="#Gravillosidad_Sin_Gravillosidad"/>
  <humificado
rdf:resource="#Humificado_Medianamente_Humificado"/>
  <m_basal
rdf:resource="#MBasal_materiales_transportados_arcillosos_gene
ralmente"/>
  <pend_pred rdf:resource="#PendPred_Casi_llano"/>
  <prof_pedo rdf:resource="#ProfPedo_Poco_Profundo"/>
  <rocosidad rdf:resource="#Rocosidad_sin_rocosidad"/>
  <salinidad rdf:resource="#Salinidad_sin_salinidad"/>
  <satura rdf:resource="#Satura_Saturado"/>
  <subtipo rdf:resource="#Subtipo_Tipico"/>
  <textura rdf:resource="#Textura_Arcilla_Caolinitica"/>
</owl:Thing>
```

```
<!--
http://www.semanticweb.org/ontologies/2013/0/Ontology135764341
8249.owl#57 -->
```

```
<owl:Thing rdf:about="#57">
  <rdf:type rdf:resource="#Ferralico_Rojo"/>
  <Profefec>28</Profefec>
  <concrecion rdf:resource="#Concrec_Sin_Concreciones"/>
  <erosion rdf:resource="#Erosion_Poca"/>
  <gleyzado rdf:resource="#Gleyzado_No_Gleyzado"/>
  <gravillosidad
rdf:resource="#Gravillosidad_Muy_Poca"/>
  <humificado
rdf:resource="#Humificado_Medianamente_Humificado"/>
  <m_basal
rdf:resource="#MBasal_materiales_transportados_arcillosos_gene
ralmente"/>
  <pend_pred rdf:resource="#PendPred_Casi_llano"/>
  <prof_pedo
rdf:resource="#ProfPedo_Medianamente_Profundo"/>
  <rocosidad rdf:resource="#Rocosidad_sin_rocosidad"/>
  <salinidad rdf:resource="#Salinidad_sin_salinidad"/>
  <satura rdf:resource="#Satura_Saturado"/>
  <subtipo rdf:resource="#Subtipo_Tipico"/>
  <textura rdf:resource="#Textura_Arcilla_Caolinitica"/>
</owl:Thing>
```

```
<!--
http://www.semanticweb.org/ontologies/2013/0/Ontology135764341
8249.owl#58 -->
```

```
<owl:Thing rdf:about="#58">
  <rdf:type rdf:resource="#Ferralico_Rojo"/>
  <Profefec>172</Profefec>
  <concrecion rdf:resource="#Concrec_Sin_Concreciones"/>
  <erosion rdf:resource="#Erosion_Poca"/>
  <gleyzado rdf:resource="#Gleyzado_No_Gleyzado"/>
  <humificado
rdf:resource="#Humificado_Medianamente_Humificado"/>
  <m_basal
rdf:resource="#MBasal_materiales_transportados_arcillosos_gene
ralmente"/>
  <pend_pred rdf:resource="#PendPred_Casi_llano"/>
  <prof_pedo rdf:resource="#ProfPedo_Muy_profundo"/>
  <rocosidad rdf:resource="#Rocosidad_sin_rocosidad"/>
  <salinidad rdf:resource="#Salinidad_sin_salinidad"/>
  <satura rdf:resource="#Satura_Saturado"/>
```

```

        <subtipo rdf:resource="#Subtipo_Antropico"/>
        <textura rdf:resource="#Textura_Arcilla_Caolinitica"/>
    </owl:Thing>

<!--
http://www.semanticweb.org/ontologies/2013/0/Ontology135764341
8249.owl#65 -->

    <owl:Thing rdf:about="#65">
        <rdf:type rdf:resource="#Pardo"/>
        <Profefec>47</Profefec>
        <concrecion rdf:resource="#Concrec_Sin_Concreciones"/>
        <erosion rdf:resource="#Erosion_Poca"/>
        <gleyzado rdf:resource="#Gleyzado_No_Gleyzado"/>
        <gravillosidad
rdf:resource="#Gravillosidad_Sin_Gravillosidad"/>
        <humificado
rdf:resource="#Humificado_Medianamente_Humificado"/>
        <lavado rdf:resource="#Lavado_Medianamente_Lavado"/>
        <m_basal rdf:resource="#MBasal_Roca_ultrabasica"/>
        <pend_pred rdf:resource="#PendPred_Casi_llano"/>
        <prof_pedo rdf:resource="#ProfPedo_Profundo"/>
        <rocosidad rdf:resource="#Rocosidad_sin_rocosidad"/>
        <salinidad rdf:resource="#Salinidad_sin_salinidad"/>
        <satura rdf:resource="#Satura_Carbonatado"/>
        <subtipo rdf:resource="#Subtipo_Tipico"/>
        <textura
rdf:resource="#Textura_Arcilla_Montmorillonitica"/>
    </owl:Thing>

<!--
http://www.semanticweb.org/ontologies/2013/0/Ontology135764341
8249.owl#66 -->

    <Pardo rdf:about="#66">
        <rdf:type rdf:resource="&owl;Thing"/>
    </Pardo>

<!--
http://www.semanticweb.org/ontologies/2013/0/Ontology135764341
8249.owl#67 -->

    <owl:Thing rdf:about="#67">
        <rdf:type rdf:resource="#Pardo"/>
        <Profefec>47</Profefec>
        <concrecion rdf:resource="#Concrec_Sin_Concreciones"/>

```

```

        <erosion rdf:resource="#Erosion_Mediana"/>
        <gleyzado rdf:resource="#Gleyzado_No_Gleyzado"/>
        <gravillosidad
rdf:resource="#Gravillosidad_Sin_Gravillosidad"/>
        <humificado
rdf:resource="#Humificado_Poco_Humificado"/>
        <lavado rdf:resource="#Lavado_Carbonatado"/>
        <m_basal
rdf:resource="#MBasal_Eluvios_de_esquistos_micaceos_con_venas_
de_cuarzo"/>
        <pend_pred rdf:resource="#PendPred_Ondulado"/>
        <prof_pedo rdf:resource="#ProfPedo_Profundo"/>
        <rocosidad rdf:resource="#Rocosidad_sin_rocosidad"/>
        <salinidad rdf:resource="#Salinidad_sin_salinidad"/>
        <satura rdf:resource="#Satura_Carbonatado"/>
        <subtipo rdf:resource="#Subtipo_Tipico"/>
        <textura
rdf:resource="#Textura_Arcilla_Montmorillonitica"/>
    </owl:Thing>

```

```

    <!--
http://www.semanticweb.org/ontologies/2013/0/Ontology135764341
8249.owl#68 -->

```

```

    <owl:Thing rdf:about="#68"/>

```

```

    <!--
http://www.semanticweb.org/ontologies/2013/0/Ontology135764341
8249.owl#69 -->

```

```

    <owl:Thing rdf:about="#69">
        <rdf:type rdf:resource="#Ferralitico_Rojo"/>
        <Profefec>69</Profefec>
        <concrecion rdf:resource="#Concrec_Sin_Concreciones"/>
        <erosion rdf:resource="#Erosion_Poca"/>
        <gleyzado rdf:resource="#Gleyzado_No_Gleyzado"/>
        <gravillosidad
rdf:resource="#Gravillosidad_Sin_Gravillosidad"/>
        <humificado
rdf:resource="#Humificado_Medianamente_Humificado"/>
        <m_basal
rdf:resource="#MBasal_materiales_transportados_arcillosos_gene
ralmente"/>
        <pend_pred rdf:resource="#PendPred_Casi_llano"/>
        <prof_pedo rdf:resource="#ProfPedo_Profundo"/>
        <rocosidad rdf:resource="#Rocosidad_sin_rocosidad"/>
        <salinidad rdf:resource="#Salinidad_sin_salinidad"/>
        <satura rdf:resource="#Satura_Saturado"/>
    </owl:Thing>

```



```

        <subtipo rdf:resource="#Subtipo_Tipico"/>
        <textura rdf:resource="#Textura_Arcilla_Caolinitica"/>
    </owl:Thing>

    <!--
    http://www.semanticweb.org/ontologies/2013/0/Ontology135764341
    8249.owl#7 -->

    <Ferralitico_Rojo rdf:about="#7">
        <rdf:type rdf:resource="#owl:Thing"/>
        <Profefec>15</Profefec>
        <concrecion rdf:resource="#Concrec_Sin_Concreciones"/>
        <erosion rdf:resource="#Erosion_Poca"/>
        <gleyzado rdf:resource="#Gleyzado_No_Gleyzado"/>
        <gravillosidad
rdf:resource="#Gravillosidad_Sin_Gravillosidad"/>
        <humificado
rdf:resource="#Humificado_Medianamente_Humificado"/>
        <m_basal
rdf:resource="#MBasal_materiales_transportados_arcillosos_gene
ralmente"/>
        <pend_pred
rdf:resource="#PendPred_Ligeramente_Ondulado"/>
        <prof_pedo rdf:resource="#ProfPedo_Poco_Profundo"/>
        <rocosidad rdf:resource="#Rocosidad_rocoso"/>
        <salinidad rdf:resource="#Salinidad_sin_salinidad"/>
        <satura rdf:resource="#Satura_Saturado"/>
        <subtipo rdf:resource="#Subtipo_Tipico"/>
        <textura rdf:resource="#Textura_Arcilla_Caolinitica"/>
    </Ferralitico_Rojo>

    <!--
    http://www.semanticweb.org/ontologies/2013/0/Ontology135764341
    8249.owl#70 -->

    <owl:Thing rdf:about="#70">
        <rdf:type rdf:resource="#Pardo"/>
        <Profefec>25</Profefec>
        <concrecion rdf:resource="#Concrec_Sin_Concreciones"/>
        <erosion rdf:resource="#Erosion_Poca"/>
        <gleyzado rdf:resource="#Gleyzado_No_Gleyzado"/>
        <gravillosidad
rdf:resource="#Gravillosidad_Sin_Gravillosidad"/>
        <humificado
rdf:resource="#Humificado_Poco_Humificado"/>
        <lavado rdf:resource="#Lavado_Carbonatado"/>
        <m_basal rdf:resource="#MBasal_Roca_ignea_basica"/>

```

```

        <pend_pred
rdf:resource="#PendPred_Ligeramente_Ondulado"/>
        <prof_pedo
rdf:resource="#ProfPedo_Medianamente_Profundo"/>
        <rocosidad rdf:resource="#Rocosisad_sin_rocosidad"/>
        <salinidad rdf:resource="#Salinidad_sin_salinidad"/>
        <satura rdf:resource="#Satura_Carbonatado"/>
        <subtipo rdf:resource="#Subtipo_Tipico"/>
        <textura
rdf:resource="#Textura_Arcilla_Montmorillonitica"/>
    </owl:Thing>

<!--
http://www.semanticweb.org/ontologies/2013/0/Ontology135764341
8249.owl#76 -->

    <owl:Thing rdf:about="#76">
        <rdf:type rdf:resource="#Fersialitico_Pardo_Rojizo"/>
        <Profefec>16</Profefec>
        <concrecion rdf:resource="#Concrec_Sin_Concreciones"/>
        <gleyzado rdf:resource="#Gleyzado_No_Gleyzado"/>
        <gravillosidad rdf:resource="#Gravillosidad_Mediana"/>
        <humificado
rdf:resource="#Humificado_Medianamente_Humificado"/>
        <m_basal
rdf:resource="#MBasal_materiales_transportados_arcillosos_gene
ralmente"/>
        <pend_pred rdf:resource="#PendPred_Alomado"/>
        <prof_pedo rdf:resource="#ProfPedo_Poco_Profundo"/>
        <rocosidad rdf:resource="#Rocosisad_rocoso"/>
        <salinidad rdf:resource="#Salinidad_sin_salinidad"/>
        <satura rdf:resource="#Satura_Saturado"/>
        <subtipo rdf:resource="#Subtipo_Tipico"/>
        <textura rdf:resource="#Textura_Arcilla"/>
    </owl:Thing>

<!--
http://www.semanticweb.org/ontologies/2013/0/Ontology135764341
8249.owl#77 -->

    <owl:Thing rdf:about="#77">
        <rdf:type rdf:resource="#Fersialitico_Pardo_Rojizo"/>
        <Profefec>31</Profefec>
        <concrecion rdf:resource="#Concrec_Sin_Concreciones"/>
        <erosion rdf:resource="#Erosion_Poca"/>
        <gleyzado rdf:resource="#Gleyzado_No_Gleyzado"/>

```

```

        <gravillosidad
rdf:resource="#Gravillosidad_Sin_Gravillosidad"/>
        <humificado rdf:resource="#Humificado_Humificado"/>
        <m_basal
rdf:resource="#MBasal_materiales_transportados_arcillosos_gene
ralmente"/>
        <pend_pred
rdf:resource="#PendPred_Ligeramente_Ondulado"/>
        <prof_pedo
rdf:resource="#ProfPedo_Medianamente_Profundo"/>
        <rocosidad rdf:resource="#Rocosidad_sin_rocosidad"/>
        <salinidad rdf:resource="#Salinidad_sin_salinidad"/>
        <satura rdf:resource="#Satura_Saturado"/>
        <subtipo rdf:resource="#Subtipo_Tipico"/>
        <textura rdf:resource="#Textura_Arcilla"/>
</owl:Thing>

```

```

<!--
http://www.semanticweb.org/ontologies/2013/0/Ontology135764341
8249.owl#86 -->

```

```

<owl:Thing rdf:about="#86">
    <rdf:type rdf:resource="#Pardo"/>
    <Profefec>40</Profefec>
    <concrecion rdf:resource="#Concrec_Sin_Concreciones"/>
    <erosion rdf:resource="#Erosion_Poca"/>
    <gleyzado rdf:resource="#Gleyzado_No_Gleyzado"/>
    <gravillosidad
rdf:resource="#Gravillosidad_Sin_Gravillosidad"/>
    <humificado
rdf:resource="#Humificado_Medianamente_Humificado"/>
    <m_basal rdf:resource="#MBasal_Roca_ignea_basica"/>
    <pend_pred rdf:resource="#PendPred_Ondulado"/>
    <prof_pedo
rdf:resource="#ProfPedo_Medianamente_Profundo"/>
    <rocosidad rdf:resource="#Rocosidad_sin_rocosidad"/>
    <salinidad rdf:resource="#Salinidad_sin_salinidad"/>
    <satura rdf:resource="#Satura_Carbonatado"/>
    <subtipo rdf:resource="#Subtipo_Tipico"/>
    <textura
rdf:resource="#Textura_Arcilla_Montmorillonitica"/>
</owl:Thing>

```

```

<!--
http://www.semanticweb.org/ontologies/2013/0/Ontology135764341
8249.owl#99 -->

```

```

<owl:Thing rdf:about="#99">

```

```

    <rdf:type rdf:resource="#Pardo"/>
    <Profefec>27</Profefec>
    <concrecion rdf:resource="#Concrec_Sin_Concreciones"/>
    <erosion rdf:resource="#Erosion_Poca"/>
    <gleyzado rdf:resource="#Gleyzado_No_Gleyzado"/>
    <gravillosidad
rdf:resource="#Gravillosidad_Sin_Gravillosidad"/>
    <humificado
rdf:resource="#Humificado_Medianamente_Humificado"/>
    <lavado rdf:resource="#Lavado_Muy_Lavado"/>
    <m_basal rdf:resource="#MBasal_Roca_ultrabasica"/>
    <pedregosidad
rdf:resource="#Pedregosidad_Sin_pedregosidad"/>
    <pend_pred rdf:resource="#PendPred_Casi_llano"/>
    <prof_pedo
rdf:resource="#ProfPedo_Medianamente_Profundo"/>
    <rocosidad rdf:resource="#Rocosidad_sin_rocosidad"/>
    <salinidad rdf:resource="#Salinidad_sin_salinidad"/>
    <satura rdf:resource="#Satura_Saturado"/>
    <subtipo rdf:resource="#Subtipo_Tipico"/>
    <textura
rdf:resource="#Textura_Arcilla_Montmorillonitica"/>
  </owl:Thing>

<!-- Generated by the OWL API (version 2.2.1.1138)
http://owlapi.sourceforge.net -->

```