

INSTITUTO POLITÉCNICO NACIONAL

CENTRO DE INVESTIGACIÓN EN COMPUTACIÓN

Laboratorio de Procesamiento Inteligente de Información Geoespacial

SIMILITUD SEMANTICA PARA LA IDENTIFICACIÓN DE INCONSISTENCIAS EN DATOS GEOESPACIALES CONCEPTUALIZADOS

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

PRESENTA JORGE VICTOR CARRERA TREJO

DR. MARCO ANTONIO MORENO IBARRA DR. MIGUEL JESUS TORRES RUIZ



INDICE

	Página
RESUMEN	Ι
ABSTRACT	III
AGRADECIMIENTOS	V
LISTA DE FIGURAS	XIII
LISTA DE TABLAS	XVII
Capítulo 1. Introducción	1
1.1 Planteamiento del Problema	4
1.2 Objetivos y metas	6
1.2.1 Objetivo General	6
1.2.2 Objetivos Particulares	6
1.2.3 Metas	7
1.3. Justificación	7
1.4 Organización de la tesis	8
Capítulo 2. Generalización	9
2.1. Generalización Manual	9
2.2. Generalización Automática	12
2.3. Modelos de Generalización	14
2.3.1. Modelo de Ratajski.	14
2.3.2. Modelo de Brassel y Weibel	15
2.3.3. Modelo de McMaster y Shea	16
2.4. Sistemas basados en conocimiento para la generalización	20
2.4.1. El Modelo de Filtrado Colaborativo en un Flujo de Trabajo	22
2.5. La generalización con sistemas multiagentes	22
2.5.1. El proyecto AGENT	22
2.5.2. El proyecto CartACom	24
2.6 Evaluación de la Calidad de la Generalización	25

2.6.1. Características cuantitativas
2.6.1.1. Evaluación de la calidad de un conjunto de dato
generalizado utilizando características cuantitativas
2.6.1.1.1. Medidas de calidad para la generalización de
polígonos
2.6.1.1.2. Medidas de Similitud Cuantitativa para Mapas
2.6.1.1.3. Evaluación de la calidad utilizando clases
2.6.2. Características Cualitativas
2.6.2.1. Evaluación de la calidad de un CDG utilizando
Características Cualitativas
2.6.2.1.1. Evaluación de la calidad en algoritmos de
generalización
2.6.2.1.2. Medidas de calidad utilizando ontologías
2.6.2.1.2.1. Modelo ontológico de las relaciones geográfica:
para mapas generalizados
2.6.2.1.2.2. Similitud Semántica entre sistemas de objeto
geográficos aplicada a la generalización de Datos Geo
espaciales
2.6.2.1.2.3. Evaluación de la calidad de la información
utilizando medidas de similitud semántica
2.6.2.1.2.3.1. Un modelo integral para la evaluación
cualitativa espacial utilizando medidas de similitud
2.6.2.1.2.3.2. Similitud de Escenas Espaciales
2.7. Comentarios del Capítulo
Capítulo 3. Ontologías y Similitud Semántica
3.1. Ontologías
3.1.1. Construcción de una ontología
3.1.2. Tipos de ontología
3.1.3. Metodologías de construcción de una ontología
3 1 3 1 Metodología de Uschold y King

	3.1.3.2. Methontology
	3.1.4. Geonto-Met
	3.1.4.1. Axiomas
	3.1.4.2. Relaciones
	3.1.4.3. Propiedades
	3.1.4.4. Habilidades
	3.1.4.5. Instancia
	3.1.4.6. Conceptos
	3.2. Relaciones Espaciales
	3.2.1. Relaciones Topológicas
	3.2.1.1. Modelos de Intersección
	3.2.1.1. Modelo 9-Intersección Dimensionalmente Extendido
	(DE-9IM)
	3.3. Medidas de Similitud Semántica
	3.3.1. Similitud de Relaciones Topológicas
	3.3.2. Similitud de Relaciones de Distancia
	3.3.3. Distancia semántica en analogía a la distancia espacial
	3.4. Comentarios del Capítulo
C	apítulo 4. Metodología
	4.1 Etapa de preprocesamiento
	4.1.1 Base de datos espacial generalizada
	4.1.2 Conceptualización del dominio
	4.1.2.1. Propiedades de los datos geográficos
	4.1.2.2. Caracterización de las relaciones topológicas
	4.1.2.3. Objetos geográficos que componen el dominio
	4.1.3 Construcción de la ontología
	4.2 Síntesis
	4.2.1 Medición de las relaciones topológicas
	4.2.1.1. Medición de la relación topológica para unidades
	habitacionales (RTD1)

4.2.1.2. Medición de la relación topológica para escuelas u
hospitales con respecto a las calles (RTD2)
4.2.1.3. Medición de la relación topológica para hospitales
escuelas con respecto a las manzanas (RTD3)
4.2.1.4. Medición de la relación topológica para hospitales
escuelas con respecto a las calles (RTC)
4.2.2 Construcción de la representación conceptual
4.3 Verificación
4.3.1 Esquemas de vecindad conceptual para relaciones topológicas
4.3.1.1. Esquema de vecindad conceptual disjunto-comparte
4.3.1.2. Esquema de vecindad conceptual disjunto-cruza
4.3.2 Verificación del CDG
4.3.2.1. Análisis a nivel micro
4.3.2.2. Análisis a nivel meso
4.3.2.3. Análisis a nivel macro
Capítulo 5. Pruebas y Resultados
5.1.1 Conjunto de Datos Fuente (CDF)
5.1.2 Conjunto de Datos Generalizado (CDG)
5.1.2 Implementación de la ontología
5.2 Síntesis
5.2.1 Medición de la relación topológica para unidades habitacionale
(RTD1)
5.2.2 Medición de la relación topológica entre escuelas y hospitale
con calles (RTD2)
5.2.3 Medición de la relación topológica entre hospitales y escuela
hospitales con manzanas (RTD3)
5.2.4 Medición de la relación topológica entre hospitales y escuela
con calles (RTC)
5.3 Verificación

Capítulo 6. Conclusiones, aportaciones y trabajo a futuro	121
6.1 Aportaciones	122
6.2 Trabajo a futuro	123
Referencias	125
Anexos	133
A. Diccionario de Datos	133
B. Archivo instancias.xml	138
C. Archivo relaciones.xml	139

LISTA DE FIGURAS

	Página
Figura 1.1. Inconsistencias generadas por la generalización	3
Figura 1.2. Conjunto de datos geográficos de la República Mexicana	4
Figura 1.3. Conjunto de datos geográficos de la República Mexicana	
generalizado	5
Figura 2.1. Eliminación	10
Figura 2.2. Desplazamiento	11
Figura 2.3. Agregación y Tipificación	11
Figura 2.4. Simplificación de Areas	11
Figura 2.5. Simplificación de Líneas	12
Figura 2.6. Un CDF, múltiples CDG's	14
Figura 2.7. Modelo de Bressel y Weibel	16
Figura 2.8. Modelo de McMaster y Shea	17
Figura 2.9. Operadores de Generalización de Shea y McMaster	19
Figura 2.10. Ciclo de vida de un agente [Barrault et al., 1991]	24
Figura 2.11. Comparación de dos mapas	29
Figura 2.12. Región Sippican, Carta Naútica a escala 1:40K	35
Figura 2.13. Jerarquización	35
Figura 2.14. Ontología Sippican	36
Figura 2.15. Diferentes relaciones entre los ríos y las líneas de elevación de	
contorno	37
Figura 2.16. Fragmento de la Ontología	38
Figura 2.17. Identificando una consistencia	38
Figura 2.18. Identificando una inconsistencia	39
Figura 2.19. Vecinos conceptuales de relaciones topológicas para dos objetos	
geográficos polígonos	41
Figura 2.20. Dos escenas similares y sus diferencias	42

Figura 3.1. Estructura conceptual de GEONTO-MET	52
Figura 3.2. Representación gráfica de conceptos	58
Figura 3.3. Overlap ó Intersect	59
Figura 3.4. Touch o Meet	60
Figura 3.5. Disjoint	60
Figura 3.6. Representación Gráfica de IAS	63
Figura 3.7. Diferentes nociones de similitud	64
Figura 3.8. Vecinos conceptuales de relaciones topológicas	67
Figura 3.9. Vecinos conceptuales de relaciones de distancia y los símbolos	
que representan	67
Figura 4.1. Metodología propuesta	73
Figura 4.2. Esquema general del sistema de generalización automática	74
Figura 4.3. Construcción de la ontología de la conceptualización del CDF y	
CDG	78
Figura 4.4. Implementación de las relaciones topológicas en la	
conceptualización del CDF y CDG	79
Figura 4.5. Esquema de vecindad conceptual disjoint-comparte	90
Figura 4.6. Esquema de vecindad conceptual <i>disjunto-cruza</i>	9
Figura 5.1. Conjunto de Datos Fuente (CDF)	98
Figura 5.2. Conjunto de Datos Generalizado (CDG)	100
Figura 5.3. Construcción de la ontología de la conceptualización del CDF y	
CDG	10
Figura 5.4. Implementación de las relaciones topológicas en la	
conceptualización del CDF y CDG	102
Figura 5.5. Los objetos geográficos se instancian en las clases a las que	
pertenecen de acuerdo a la conceptualización, como por ejemplo los edificios	104
Figura 5.6. Ontología diseñada con base en el dominio de trabajo	104
Figura 5.7. Instancias y relaciones en la ontología diseñada	10:
Figura 5.8. Unidades Habitacionales en el CDG	100

Figura 5.9. Tabla con las áreas de las Unidades Habitacionales en el CDG	107
Figura 5.10. Área de escuelas y hospitales en el CDG, y áreas de intersección	
en algunas calles en el CDG	108
Figura 5.11. Tabla con las areas de escuelas y hospitales en el CDG	110
Figura 5.12. Escuelas y hospitales en el CDG con respecto a algunas	
manzanas	111
Figura 5.13. Tabla con las áreas de las escuelas y hospitales, y áreas de	
intersección en algunas manzanas en el CDG	112
Figura 5.14. Distancia de cruce entre escuelas y hospitales con respecto a las	
calles en el CDG	114
Figura 5.15. Medición de las distancias de las áreas de las escuelas y	
hospitales con respecto a las calles en el CDG	114
Figura 5.16. Generación de archivos de relaciones e instancias con base en la	
ontología diseñada	116
Figura 5.17. Análisis a nivel <i>micro</i>	117
Figura 5.18. Análisis a nivel <i>meso</i>	119
Figura 5.19. Análisis a nivel <i>macro</i>	120

LISTA DE TABLAS

	Página
Tabla 2.1. Elementos básicos en el proceso de medición espacial	40
Tabla 3.1. Definición de componentes topológicos	61
Tabla 3.2. Matriz de patrones para relaciones topológicas en el modelo DE-	
9IM	61
Tabla 4.1. Caracterización de las relaciones topológicas utilizadas	76
Tabla 4.2. Clases de objetos geográficos	77
Tabla 4.3. Criterios para RTD1	82
Tabla 4.4. Criterios para RTD2	84
Tabla 4.5. Criterios para RTD3	85
Tabla 4.6. Criterios para RTC	86
Tabla 4.7. Rangos para definir cada uno de los conceptos de relaciones	
topológicas	88
Tabla 4.8. Grados de consistencia a nivel <i>micro</i>	93
Tabla 4.9. Grados de consistencia a nivel <i>meso</i>	94
Tabla 5.1. Relaciones Topológicas en el Conjunto de Datos Fuente (CDF)	98
Tabla 5.2. Cambios en el CDG con respecto al CDF	99
Tabla 5.3. Relaciones Topológicas en el Conjunto de Datos Generalizado	
(CDG)	100
Tabla 5.4. Áreas de las Unidades Habitacionales y Áreas de Intersección	106
Tabla 5.5. Agrupando Áreas de las Unidades Habitacionales y Áreas de	
Intersección	107
Tabla 5.6. Conceptualización con base en RTD1	107
Tabla 5.7. Áreas de las escuelas y hospitales y áreas sobre las calles	108
Tabla 5.8. Relación de los objetos de intersección	109
Tabla 5.9. Conceptualización con base en RTD2	110

Tabla 5.10. Áreas de las escuelas y hospitales, y áreas de intersección sobre	
algunas manzanas	112
Tabla 5.11. Relación de los objetos de intersección	113
Tabla 5.12. Conceptualización con base en RTD3	113
Tabla 5.13. Distancia de cruce de las áreas de las escuelas y hospitales sobre	
las calles	115
Tabla 5.14. Relación de los objetos de intersección	115
Tabla 5.15. Conceptualización con base en RTC	115

Resumen

En este trabajo de tesis, se presenta una metodología que permite valorar la consistencia de un conjunto de datos geográficos generalizados (CDG), con base en la comparación semántica de las relaciones espaciales que existen entre ellos.

Para este caso, se han utilizado mediciones de similitud semántica sobre un esquema de vecindad conceptual de relaciones espaciales. La metodología propuesta se compone de dos etapas básicas: Síntesis y Verificación. La primera etapa consiste en generar una representación conceptual de los datos geográficos, en dependencia al análisis espacial de los datos geoespaciales. La segunda etapa se enfoca en las comparaciones de las representaciones conceptuales de los datos fuente y generalizados. En esta etapa se definen tres niveles de análisis: el micro, meso y macro.

Asimismo, como caso de estudio, esta metodología fue aplicada principalmente a datos geoespaciales de la Ciudad de México, fundamentalmente en el contexto de la infraestructura vial y urbana.

Abstract

In this dissertation a methodology to assess the consistency on generalized geographic datasets (CDG) is described. This approch consists of the semantic comparison of spatial relations that are explicitly presented among geospatial data.

For this case, semantic similarity measurements are used over a schema of conceptual neighboorhood based on spatial relations. The proposed methodology is composed of two stages: Synthesis and Verification. The first stage aims to generate a conceptual representation of geospatial data, depending directly on the spatial analysis of the geographic objects. The second stage is focused on the comparisons of conceptual representations taken from source and generalized data. In this stage, three levels of analysis are defined: micro, meso and macro.

On the other hand, we proposed as a case study to apply this methodology in geographic datasets of México City, particularly in the context of urban and street infrastructure.

Capítulo 1. Introducción

Los Sistemas de Información Geográfica (GIS), analizan, manipulan y almacenan datos geográficos de diversas temáticas, organizadas de acuerdo con las necesidades de una aplicación en particular [Tomlinson, 2003]. Estos datos geográficos digitales son representaciones de entidades que existen en la realidad, los cuales son elaborados con diversos propósitos V utilizando las diversas tecnologías geoespaciales (Geoposicionamiento, Percepción Remota y cartografía automatizada). Por otro lado, la cartografía es el área encargada de realizar las representaciones del mundo real por medio de mapas, los cuales han sido usados por miles de años para representar la información sobre el entorno geográfico por lo que han tendido a muy diversas aplicaciones.

Los mapas son una representación gráfica de una realidad geográfica, donde dichas representaciones frecuentemente son generadas a partir de una base de datos geográfica. Las representaciones digitales del mundo real, pueden ser llamadas "Digital Geographic Model (DGM)" [Basaraner, 2007] y son utilizadas principalmente para el análisis espacial de datos, sin embargo, se pueden obtener representaciones gráficas digitales y analógicas (mapas en papel), donde estas representaciones son elaboradas con un propósito o temática en particular, el cual se vincula directamente con características como la escala, la cual se encuentra relacionada muy estrechamente, con el nivel de detalle en los datos y con una primitiva geométrica de representación asociada [Burrough, 1986]. Las aplicaciones de GIS tienen usos muy diversos, se estima que el 20% es usado por gobiernos, 15% en educación y el resto va desde la industria farmacéutica, oceanográfica y recreativo [Heywood et al 1998].

En ocasiones, es necesario modificar el nivel de detalle de los datos geográficos, de acuerdo con una escala o propósito en particular [Steiner, 1973][McMaster and Shea, 1992], con el fin de generar datos que se adapten a condiciones específicas, a este procedimiento se le

conoce como *generalización* [ICA, 1972]. En términos generales la generalización tiene como fin modificar el nivel de detalle sin perder la semántica de los datos generalizados, aumentando el valor cognitivo e interpretativo de los mismos. Esto implica, eliminar los detalles menos importantes de un conjunto de *datos fuente*, con el fin de resaltar los aspectos más característicos de acuerdo con un propósito en particular. Con lo que se genera un conjunto de datos generalizado. El proceso de generalización se puede realizar de forma manual y/o automática [Steiner, 1973][Moreno, 2007], siendo esta última muy popular por la gran diversidad de aplicaciones de GIS (Geoographic Information System). De forma general se realiza aplicando los denominados operadores de generalización [McMaster and Shea, 1992], que tratan de emular los procesos de manipulación que realiza un cartógrafo sobre los datos.

Sin embargo, la ejecución de este proceso se acompaña de algunos efectos en los datos, en particular para este trabajo nos centramos en errores originados por la modificación de las relaciones topológicas. Estas modificaciones representan inconsistencias entre el *conjunto datos fuente (CDF)* y el *conjunto de datos generalizado* (CDG); una inconsistencia puede interpretarse como la pérdida de la semántica de la representación generalizada [Moreno, 2007]. En la Figura 1.1, se muestran algunos efectos de la generalización, en el detalle a) se muestra el *CDF* donde se observa que todas las edificaciones se encuentran "*dentro de*" la tierra", mientras que en el detalle b) representa al conjunto de datos generalizado y existe una edificación que aparece "*dentro de*" agua, esto fue originado por la simplificación de la línea costera. Adicionalmente en la Figura 1.1 se observa la agrupación de las edificaciones, lo cual representa una transformación de los objetos geográficos. Como se puede observar, los errores son originados por la naturaleza propia de la generalización, que incluye transformaciones en la geometría, y que puede modificar las relaciones topológicas (por ejemplo, "*dentro de*"); generando inconsistencias entre los datos fuente y los datos generalizados

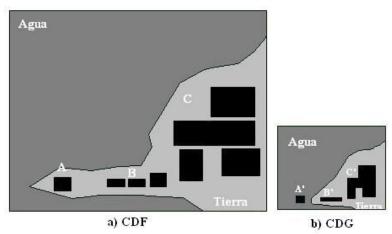


Fig. 1.1. Inconsistencias generadas por la generalización.

Los organismos nacionales encargados de la cartografía (*NMA*, *National Mapping Agency*), de diversos países generalizan datos geográficos en diferentes escalas [Stoter, 2005][Ruas, 2001], por ejemplo: en Dinamarca, se producen conjuntos de datos de escalas 1:50,000 a partir de conjuntos de datos a escala 1:10,000. Mientras que en Francia, se generan datos en escalas pequeñas a partir de datos con escala 1:25,000. De forma general debe cuidarse la pérdida de información originada por el procesamiento evaluando la calidad de los datos generalizados [USGS, 2007].

En años recientes se ha incorporado el uso de ontologías a la Computación, con el fin de representar el conocimiento sobre diferentes dominios. Este enfoque ha mostrado su utilidad y ha ganado popularidad, debido a que la forma de realizar el procesamiento a los datos es similar a la forma en que los seres humanos analizan variables cualitativas mediante su sistema cognitivo [Moreno, 2007]. En este caso, se propone la ontología para describir y representar la semántica del dominio de interés, compuesto por conceptos geográficos y relaciones espaciales. La ontología se genera a partir del esquema de base de datos de los conjuntos de datos, lo que describe las relaciones relevantes entre los objetos geográficos. Esto se utiliza para evaluar si la semántica de los datos generalizados ha sido preservada, dicha evaluación se realiza por medio de una medición de similitud semántica, la cual tiene como propósito identificar que tan cercanos son conceptualmente dos términos.

1.1 Planteamiento del problema

Consideremos el ejemplo de la Figura 1.2, donde tenemos un fragmento de la República Mexicana, mostrando, las temáticas: Carreteras (líneas verdes), Ciudades (puntos color negro) y Ríos (líneas azules). A partir de la visualización se observan las siguientes relaciones espaciales o topológicas para ésta.

- La ciudad de Durango, *pertenece* al estado de Durango.
- Una carretera *cruza* al sur de la ciudad de Durango.
- El Río San Pedro *cruza* al norte de la ciudad de Durango.

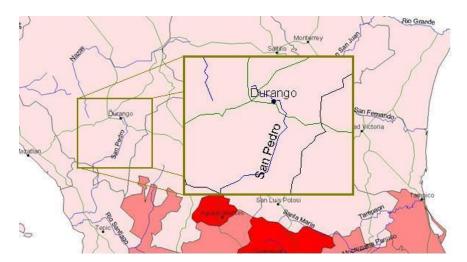


Figura 1.2. Conjunto de datos geográficos de la República Mexicana

Al someter al conjunto de datos geográficos de la Figura 1.2 a un proceso de generalización, se obtienen los datos que aparecen en la Figura 1.3.

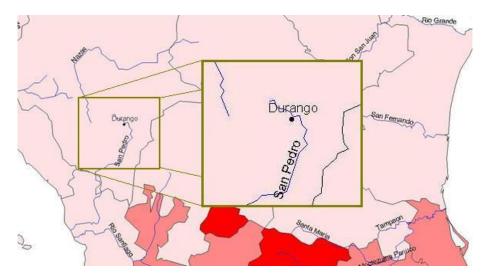


Figura 1.3. Conjunto de datos geográficos de la República Mexicana generalizado

De las relaciones espaciales que posee la Ciudad de Durango, podemos resaltar las relaciones que se conservan de acuerdo con el conjunto de datos fuente:

• La ciudad de Durango *pertenece* al estado de Durango.

Las relaciones que se eliminaron o sufrieron modificaciones con respecto a los datos fuente representan inconsistencias en los datos, en este caso son:

- Se eliminó la relación carretera *pasa por el sur* de la ciudad de Durango.
- Se modificó la relación donde el río San Pedro cruza al norte de la ciudad de Durango, ya que de acuerdo con el proceso de generalización, se observa que el Río San Pedro, pasa al norte y no cruza la ciudad de Durango como sucedía en el conjunto de datos original.

El enfoque propuesto se basa en identificar estas inconsistencias entre los conjuntos de datos. Por lo que, se hace necesario revisar las relaciones espaciales que guardan los distintos objetos geográficos que componen nuestra información. Para resolver este problema se propone evaluar al conjunto de datos geográficos fuente con al conjunto de datos generalizado, usando mediciones a los niveles *micro*, *macro* y *meso* [Ruas, 2000][Bard, 2003]. Para realizar esto, se deben emplear representaciones conceptuales (ontologías) que describan las relaciones existentes en los datos. La evaluación de la

consistencia se realiza utilizando medidas de similitud semántica, la cual tiene como objetivo comparar el cambio en las relaciones.

Para el caso de la presente tesis se ha considerado como caso de estudio la generalización de un conjunto de datos geográficos de un área urbana, en particular de la delegación Gustavo A. Madero del D.F. Los datos se constituyen por edificios, manzanas, hospitales, escuelas, calles y avenidas,

1.2 Objetivos y Metas

1.2.1 Objetivo General

Desarrollar una metodología para identificar las inconsistencias en conjuntos de datos geoespaciales generalizados, por medio de mediciones de similitud semántica aplicadas sobre una representación conceptual del dominio geográfico; con el fin de describir el grado de consistencia de los datos generalizados en diferentes niveles de análisis.

1.2.2 Objetivos Particulares

- Analizar el estado del arte en lo que respecta a la verificación de consistencia en datos generalizados.
- Analizar las propiedades de los datos geográficos, de acuerdo con el estado del arte en lo que respecta a la verificación de consistencia en datos generalizados.
- Desarrollar la conceptualización del caso de estudio.
- Diseñar e implementar una ontología para describir las propiedades, relaciones espaciales y entidades que existen en el dominio geográfico acotado al caso de estudio propuesto.
- Diseñar e implementar los métodos para extraer las relaciones espaciales entre objetos geográficos y representarlas explícitamente.

- Diseñar e implementar la medición de similitud semántica para evaluar la consistencia entre las relaciones espaciales existentes.
- Implementar un caso de estudio basado en la metodología propuesta para verificar conceptualmente el grado de similitud de los conjuntos de datos vectoriales.

1.2.3 Metas

Las metas alcanzadas en el desarrollo de esta tesis son:

- Un análisis exhaustivo del estado del arte en las diversas áreas relacionadas con este trabajo.
- Un análisis detallado de las diversas propiedades y relaciones que poseen los datos geoespaciales.
- Definición e implantación de las medidas de similitud entre dos representaciones conceptuales.
- Definición e implantación de representaciones conceptuales de objetos geoespaciales con base en propiedades y relaciones.
- Descripción e implantación de un caso de estudio.
- Definición e implantación de mediciones sobre datos geográficos para extraer las relaciones implícitas entre múltiples capas.
- Definición e implantación de mediciones sobre las relaciones topológicas que existen entre pares de objetos geográficos.

1.3 Justificación

Los datos obtenidos por medio de procedimientos automáticos de generalización se han hecho muy útiles, debido a que permiten adecuar el nivel de detalle de acuerdo con los requerimientos específicos. Sin embargo, tal y como se ha mencionado con anterioridad, frecuentemente aparecen inconsistencias que deben ser identificadas, ya que de no hacerlo pudiéramos utilizar datos llenos de inconsistencias. Por otro lado, el análisis de las

propiedades y relaciones espaciales que poseen los objetos geográficos permite el desarrollo de ontologías que pretenden capturar la semántica del dominio.

Consideramos que la utilización de ontologías para describir el dominio geográfico puede ser muy útil en la identificación de inconsistencias en datos geoespaciales generalizados, el principio del enfoque es verificar si la semántica se ha preservado en los datos generalizados. Para esto se utilizan mediciones de similitud semántica aplicadas sobre conceptos que describen a las relaciones espaciales.

1.4 Organización de la tesis

El resto de la tesis está organizada como sigue: el Capítulo 2, describe los trabajos relacionados con la generalización y los diferentes procedimientos para valorar la consistencia. En el Capítulo 3 se presentan los trabajos relacionados con el desarrollo de ontologías y similitud semántica. La metodología propuesta se presenta en el Capítulo 4, detallando el método propuesto y el caso de estudio. En el Capítulo 5, se muestran los resultados experimentales, mientras que las conclusiones y la propuesta de trabajos a futuro se presentan en el Capítulo 6.

Capítulo 2. Generalización

El término generalización procede de la palabra latina generalis¹; donde la raíz lingüística explica perfectamente la esencia de la generalización: la selección de lo principal, lo más importante. Esta operación se orienta hacia un objetivo teniendo en cuenta la realidad, sus rasgos principales y típicos, y además las particularidades características, de acuerdo con la asignación del mapa, con el tema y la escala [Moreno, 2007], por lo que también se puede definir como el proceso de derivar respecto a un propósito en particular, un conjunto de datos geográficos a una escala menor o mayor obteniendo así una resolución² menor o mayor [Basaraner, 2007].

La Asociación Internacional de Cartografía (ICA) define la generalización como "Selección y representación simplificada del detalle apropiado para la escala y/o el propósito del mapa" [ICA 1973].

La generalización se puede realizar en forma manual y/o automática.

2.1 Generalización Manual

La generalización manual es un proceso extremadamente laborioso y subjetivo, donde el resultado obtenido, depende de los métodos de generalización y el grado en que estos son aplicados. Este tipo de generalización es realizado por un usuario experto, quien determina el contenido y presentación de los datos generalizados con respecto a la información contenida en el conjunto de datos fuente, y que satisfaga los requerimientos del usuario. Además decide los métodos de generalización que se aplicarán.

¹ Relacionado a todo, general

El tamaño más pequeño de un objeto o característica, en el cual, dicho objeto o característica sea reconocible. [Goodchild, 1991]

El experto toma diversas decisiones, las cuales pueden ser muy complejas, ya que dependen del grado del conocimiento geográfico que tengan y que le permita crear representaciones consistentes [McMaster and Shea, 1992].

Los métodos de generalización que dicho usuario aplique, después de reducir la escala de un CDF, se han dividido en operadores [McMaster and Shea, 1992] con el objetivo de definir dicho proceso, estos operadores continuamente son refinados y formalizados, sin embargo en forma general se pueden considerar los siguientes [Brazile, 2000].

• Los objetos pueden ser eliminados, en la Figura 2.1a, se muestra un área urbana a escala 1:25000, mientras que en la Figura 2.1b tenemos la misma área urbana a una escala 1:50000, algunos objetos en la Figura 2.1a han sido eliminados.

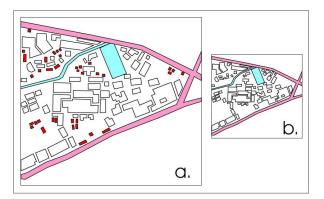


Figura 2.1. Eliminación

Los objetos pueden ser desplazados, en la Figura 2.2a se puede observar que los
edificios se encuentran muy cercanos entre sí, se define un umbral de separación de 0.3
mm, con este valor de umbral se mejora la legibilidad del mapa, dicho umbral se aplica
en la Figura 2.2b.

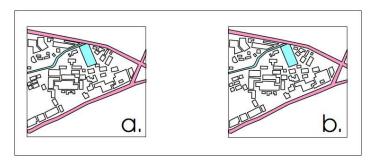


Figura 2.2. Desplazamiento

• Los elementos pueden ser agregados y tipificados, el mapa mostrado en la Figura 2.3a se encuentra a una escala 1:50000, dicha escala se reduce a 1:100000, el mapa resultante se muestra en la Figura 2.3b, como se observa este mapa no es claro, así que los edificios se agregan y tipifican en nuevos polígonos.

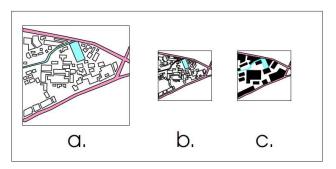


Figura 2.3. Agregación y Tipificación

• Las áreas pueden ser simplificadas, en la Figura 2.4a se muestra un mapa que contiene algunos lagos, a este mapa se le reduce la escala, el mapa resultante se muestra en la Figura 2.4b, estos lagos se simplifican en un solo lago, que es lo suficientemente largo para que se pueda mostrar en el mapa (ver la Figura 2.4c).

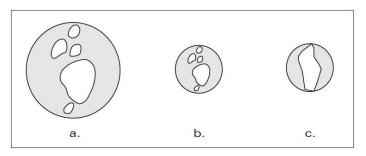


Figura 2.4. Simplificación de Áreas

• Las líneas pueden ser simplificadas, en la Figura 2.5a se muestra un mapa con una línea costera de China, a una escala de 1:40 millones, en la Figura 2.5b se eliminan las etiquetas y la línea costera, límites y ríos se simplifican apareciendo fusionados, mostrando solamente estos en el nuevo mapa.

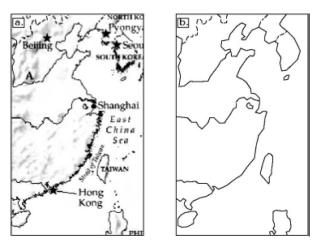


Figura 2.5. Simplificación de Líneas

- El detalle de los datos es resaltado.
- La simbología puede cambiar, de acuerdo con la identificación de los datos mediante símbolos que el experto defina.

2. 2 Generalización Automática

La generalización automática, contiene una gran cantidad de tareas, entre las que se incluyen, reducción del espacio de almacenamiento, manipulación de la escala, clasificación estadística y simbolización. Este tipo de generalización consiste en derivar de un conjunto fuente de datos geográficos, un nuevo conjunto de datos geográficos, a través de la aplicación de transformaciones espaciales y atributivas, a diferencia de la generalización manual, en forma automática. Además, permite reducir el espacio de almacenamiento utilizado, tipo y representación cartográfica del mapa o formato de la codificación del mapa consistente con su propósito y objetivo [McMaster and Shea, 1992].

Desde 1960, se han hecho esfuerzos para automatizar la generalización. Inicialmente, se realizaron algunos desarrollos enfocados a la generalización de grupos de objetos geográficos. Douglas y Peucker [Douglas and Peucker, 1973] y Deveau [Deveau, 1985], se enfocaron en el desarrollo de algoritmos para la generalización de datos lineales. Monmonier [Monmonier, 1983], Chrisman [Chrisman, 1983], Lichtner [Lichtner, 1978] y Töpfer [Töpfer and Pillewitzer, 1966] dirigieron sus investigaciones a la generalización de objetos puntuales y areales. Töpfer y Pillewizer [Töpfer and Pillewitzer, 1966], desarrollaron la *ley radical*, que define la cantidad de objetos que se presentarán en un conjunto de información geográfica, a una escala determinada. Aunque es empírica, se puede parametrizar de acuerdo con el fenómeno a cartografiar. Utilizando dicha ley, se puede estimar la densidad de elementos y el número de vértices³ a representar para una escala determinada. La *ley radical* se define, en la Ecuación 2.1.

$$n_T \approx n_S \sqrt{\frac{S_S}{S_T}}$$
 Ecuación 2.1

Donde:

n_s, es el número de objetos en la escala fuente,

S_s, es la escala fuente,

 S_{T} , es la escala destino,

n_T, es el número de objetos que debe tener el mapa destino.

Sin embargo, la generalización se vuelve un desafío real, tan pronto como las NMA's comienzan a elaborar bases de datos geográficas, utilizando procedimientos automáticos, en los 80's; es decir, a partir de un conjunto de datos base, se generan nuevos conjuntos de datos geográficos automáticamente, como se muestra en la Figura 2.6 [Ruas, 2001].

_

³ El número de vértices es considerado ya que se trata de representaciones vectoriales compuestas por nodos, vértices y arcos. En donde los vértices representan el nivel de detalle en cada línea.

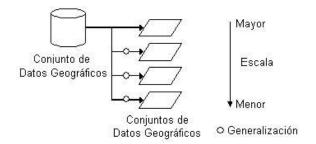


Figura 2.6. Un CDF, múltiples CDG's

2.3 Modelos de Generalización

Paralelamente a los intentos de automatización se realizaron investigaciones para definir e identificar con precisión los elementos que intervienen en la generalización, como resultado se obtuvieron modelos conceptuales. En esta sección se mencionan algunos de los cuales se consideran más relevantes.

2.3.1 Modelo de Ratajski

Este modelo fue desarrollado por el cartógrafo polaco Lech Ratajski [Ratajski, 1967], en el que se identifican dos componentes fundamentales de la generalización:

- Generalización cuantitativa, consiste en la reducción gradual del contenido o número de objetos del mapa en dependencia con el cambio de escala.
- Generalización cualitativa, la cual resulta de la transformación de formas elementales en formas más abstractas.

El argumento crítico de este modelo es el concepto denominado *punto de generalización*, este punto es alcanzado cuando la capacidad del mapa decrece al nivel donde es necesario un cambio en el método de representación⁴.

⁴ Esto implica transformaciones del tipo objeto areal en un objeto puntual o lineal

2.3.2 Modelo de Brassel y Weibel

Es uno de los modelos más detallados que se han desarrollado. Este modelo identifica cinco procesos en un ambiente digital, así que puede aplicarse específicamente en generalización automática. Los procesos en este modelo son [Brassel y Weibel, 1988].

- Reconocimiento de la estructura. Es el proceso encargado de identificar grupos de objetos y relaciones espaciales. Este proceso es controlado por los objetivos de la generalización (calidad de la base de datos, escala destino y reglas de comunicación).
- Proceso de identificación. Consiste en definir las funciones para generalizar y sus parámetros.
- Proceso de modelado. En este proceso se compilan las funciones de una biblioteca de procesos.
- Proceso de ejecución. Este proceso realiza la generalización, de acuerdo con las funciones definidas en las etapas anteriores.
- Visualización de datos. Este proceso se encarga de la presentación de los datos generados.

Adicionalmente, este modelo divide la generalización en dos tipos de generalización según su objetivo: *estadística* y *cartográfica*. La generalización estadística se define como un proceso de filtrado, que concierne a la compacidad de los datos y al análisis estadístico. Por otro lado, la generalización cartográfica se refiere a la modificación de la estructura del mapa.

Este modelo hace distinción entre el contenido y la estructura del conjunto de información geográfica fuente, lo cual se muestra en la Figura 2.7 (a, b y c) de los pasos mecánicos operacionales, mostrados en la Figura 2.7 (d y e).

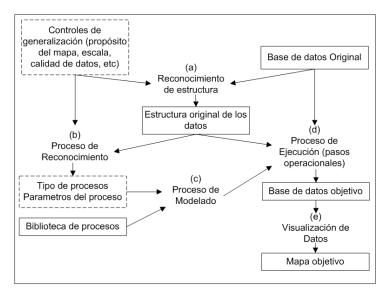


Figura 2.7. Modelo de Bressel y Weibel

2.3.3 Modelo de McMaster y Shea

Este modelo fue el primero que se elaboró considerando los objetivos filosóficos de la generalización y operadores de generalización para emular las técnicas de generalización manuales [Shea and McMaster, 1989]. El proceso de generalizar es descompuesto en tres áreas operacionales [McMaster and Shea, 1992].

- Consideración de los objetivos filosóficos de *por qué* generalizar.
- Evaluación cartométrica de las condiciones sobre *dónde* generalizar.
- Selección de las transformaciones de atributos y espaciales, las cuales proporcionan las técnicas sobre cómo generalizar.

Estas áreas operacionales se muestran en la Figura 2.8. El área operacional acerca de los objetivos filosóficos, involucra aspectos tales como la reducción de complejidad, el mantenimiento de la precisión espacial, de los atributos, de la calidad estética y de la jerarquía lógica del mapa. Considera además, aspectos dependientes de la aplicación como el propósito del mapa y la audiencia, así como también la escala apropiada que se requiere para mantener la claridad del mapa para un cierto nivel de detalle. En esta área se

contemplan los elementos computacionales que son importantes en el dominio digital. Dichos aspectos son el costo computacional de los algoritmos, su efectividad y los requerimientos de memoria y disco.

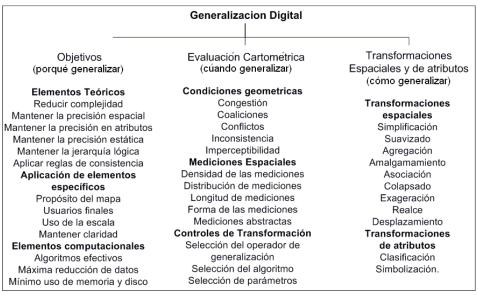


Figura 2.8. Modelo de McMaster y Shea

El área que involucra la evaluación cartométrica sobre dónde generalizar, la cual está dividida en tres partes: las condiciones geométricas, las mediciones espaciales y los controles de transformación. Las condiciones geométricas se definen con base en las transformaciones que pueden ocurrir al realizar la modificación de escala, estas medidas son congestión, fusión, conflicto, inconsistencia e imperceptibilidad. Por otro lado, las mediciones espaciales se utilizan para evaluar las relaciones entre los diferentes objetos, se utilizan mediciones de densidad, distribución, longitud, sinuosidad, forma, distancia y mediciones abstractas como simetría y homogeneidad.

El proceso de generalización está relacionado con la aplicación de *operadores de generalización*, que dependen en gran parte de la aplicación. Para obtener generalizaciones satisfactorias, es necesario aplicar los algoritmos correctos con los parámetros adecuados de manera que se realicen transformaciones correctas sobre los datos, por tal motivo este modelo contempla *controles de transformación*. Dichos controles se encargan de la parte de *donde* generalizar, que es muy importante, ya que los operadores pueden utilizarse de

diferentes formas (iterativamente, permutados o combinados), utilizando diferentes parámetros.

La tercer área operacional se refiere a las *transformaciones espaciales y de atributos*, que es la encargada de los procedimientos y funciones que realizan la generalización digital. Este componente sobre *cómo* se generalizará, es comúnmente realizado por operadores, desarrollados para tratar de emular a la generalización manual utilizando funciones matemáticas. Los operadores de generalización consisten de diez transformaciones espaciales: Simplificación, Suavizado, Agregación, Amalgamamiento, Asociación, Colapsado, Selección/Perfeccionamiento/Tipificación, Exageración, Realce y Desplazamiento; y dos transformaciones de atributos: Clasificación y Simbolización. Estos operadores se muestran en la Figura 2.9.

Tipo	Aplicado a:	Nombre	Descripción	Antes de aplicar	Después de aplicar
No Espacial	mapa	Reclasificación	Modifica la categoría a la que pertenece un objeto, así como la posibilidad de combinar objetos vecinos que pertenezcan a la misma clase		
Espacial d d		Agregación	Combina un objeto con otro de la misma (o similar) clase		
	Grupos de polígonos	Tipificación	Reduce la complejidad de un grupo de objetos, removiendo, desplazando, alargando y agregando objetos, manteniendo el arreglo típico de los objetos		
	polígonos grupos de polígonos	Desplazamiento	Implica el movimiento de un objeto, sin cambiar su forma		
		Exageración	Implica el incremento (decremento) local de un objeto, su forma es distorsionada		
	polígonos	Colapsado	Colapsa un polígono a una línea o un punto se colapsa a un píxel		
		Eliminación	Se elimina un objeto de los datos, y el espacio liberado se asigna a otras categorías		
		Realce	Incremento (decremento) global de un objeto		
	Borde de Polígonos Líneas	Simplificación	Reduce la granularidad de las líneas		
		Suavizado	Mejora la apariencia visual de las líneas		

Figura 2.9. Operadores de Generalización de Shea y McMaster

2.4 Sistemas basados en conocimiento para la generalización

A principios de los 90's, se describieron aplicaciones [Herbert et al., 1992, Joao, 1993] basadas en un sistema experto para asistir al usuario en las decisiones sobre la generalización. Estas decisiones se refieren a la selección del algoritmo y sus parámetros. El sistema asume que los usuarios no son expertos en cartografía, por lo que el sistema resulta de gran ayuda. Sin embargo, un verdadero sistema experto para generalización debe ser capaz de determinar si la tarea a realizar es posible o si la recomendación falla para obtener un resultado correcto. Un sistema experto debe ser capaz de generalizar con diferentes parámetros y evaluar dichos parámetros con base en el resultado son recomendables.

El éxito de un sistema experto está directamente ligado al conocimiento que contiene. Por lo que, cuanto mejor se define el problema, las reglas derivadas serán mejores y las soluciones que se alcancen serán más óptimas. Sin embargo, la generalización es un proceso muy complejo, además de existir poca formalización del conocimiento cartográfico [Fisher and Mackaness, 1987], [Weibel, 1991], [Ruas, 1995]. Se ha discutido [Armstrong, 1991][Weibel, 1995] sobre el conocimiento requerido para la generalización de un mapa, el cual se clasifica en tres grupos.

- Conocimiento Geométrico, por ejemplo el tamaño y forma de los objetos.
- Conocimiento Estructural, por ejemplo la jerarquía de las categorías representadas.
- Conocimiento Procedural, por ejemplo seleccionar los operadores adecuados en un conflicto determinado.

Las posibles fuentes de conocimiento cartográfico relacionadas con la generalización son: cartógrafos expertos, los mapas, libros de texto y guías de las instituciones encargadas de la confección de mapas (NMA's, INEGI⁵, etc). Se ha concluido [Weibel, 1995] que la adquisición de conocimiento cartográfico es el *cuello de botella* de los sistemas expertos para generalización. Con el fin de obtener *conocimiento cartográfico* se siguen las siguientes alternativas:

_

⁵ Instituto Nacional de Estadística, Geografía e Informática

- Entrevistas con cartógrafos expertos; es realizada por diversos métodos tales como entrevistas estructuradas, proponiendo casos críticos o problemas artificiales [McGraw and Harbison-Briggs, 1989], [Nickerson, 1991], [Schylberg, 1993], [Kilpeläinen, 2000].
- Ingeniería inversa, se basa en la recolección de conocimiento comparando objetos del mapa usando series de mapas a diversas escalas [Buttenfield and McMaster, 1991], [Leitner and Buttenfield, 1995], [Weibel 1995], [Edwardes and Mackaness, 2000].
- Trazo interactivo, genera reglas de la generalización interactivas realizadas por un cartógrafo experto [Weibel, 1991], [Weibel, 1995], [Reichenbacher, 1995a].
- Técnicas de *aprendizaje* (*machine learning*), tienen como propósito derivar reglas, a partir de un conjunto de ejemplos resueltos por un cartógrafo experto [Plazanet et al., 1998], [Mustière and Duchêne, 2001]. Se emplearon Redes Neuronales Artificiales como mecanismo de *aprendizaje* [Lagrange et al., 2000], [Sester and Brenner, 2000] y [Werschlein and Weibel, 1994].

A pesar de los intentos de adquisición de conocimiento cartográfico, se desarrollaron pocos sistemas expertos experimentales para generalización. Su éxito fue restringido a la solución de algunos problemas parciales bien definidos [Weibel, 1995], como los enfocados a la colocación de etiquetas [Cook and Jones, 1989], [Zoraster, 1991], [Freeman and Doerschler, 1992] o los enfocados a objetos lineales topológicamente estructurados [Nickerson 1988]. El conocimiento fue recolectado y empleado en estudios empíricos con la ayuda de los cartógrafos profesionales que juzgaban los ejemplos preparados según su corrección [Schylberg 1993] o que solucionaban algunos casos de estudio [Kilpeläinen 2000].

Hasta el momento, no hay pruebas de que pueda ser construido un sistema eficaz para generalización basado en reglas [Müller et al., 1995b]. Parece que los sistemas expertos no pueden por sí mismos resolver las necesidades de la generalización, ya que en este contexto el sistema debe adaptarse en forma dinámica y flexible a diversos conflictos. Para superar esta desventaja, la investigación de la generalización comenzó a desarrollarse en dirección de los sistemas multiagente.

2.4.1 El Modelo de Filtrado Colaborativo en un Flujo de Trabajo

En este modelo la generalización se realiza mediante un flujo de trabajo basado en servicios, posiblemente Web [Trouya, 2008], dichos servicios analizan los requerimientos de generalización, aplican los operadores de generalización o evalúan los resultados. Una técnica de aprendizaje de filtrado colaborativo define que operación de generalización (servicios) debe aplicarse según sea el caso, así, el sistema obtiene conocimiento de las distintas operaciones de generalización que se aplican en determinadas situaciones.

2.5 La generalización con sistemas multiagentes

En generalización, un sistema multiagente (MAS) puede ofrecer la funcionalidad necesaria para orquestar a los algoritmos, restricciones, estrategias, etc., en un sistema dinámico, flexible y automatizado. La aplicación de sistemas multiagentes a la generalización automática fue propuesta inicialmente por [Ruas, 1999] y fue enriquecido por el proyecto europeo AGENT [Barrault et al., 2001]. Este proyecto demostró las capacidades de un MAS para la generalización de mapas [Lamy et al., 1999], [Barrault et al., 2001], [Duchêne et al., 2002].

2.5.1 El proyecto AGENT

El proyecto del AGENT fue financiado por la Comisión Europea entre 1997 y 2000. Se investigó una solución de la generalización por medio de un MAS. En este proyecto cooperaban compañías privadas de GIS (Laser-Scan) e investigadores en generalización del Instituto Geográfico Nacional (IGN) de Francia y las Universidades de Zurich y Edimburgo. Todos ellos, apoyados en Inteligencia Artificial por el Instituto Politécnico de Grenoble.

Con el fin de describir más claramente este proyecto se presentarán algunos conceptos relevantes.

Entidades Geográficas: Son diseñadas como agentes, por ejemplo un edificio, un camino, un grupo de edificios, etc., pueden convertirse en un agente. En consideración de la definición de un agente proporcionada anteriormente, cada agente (entidad geográfica) en el prototipo de AGENT:

- Cumple una meta, la cual se formaliza por medio de restricciones.
- Posee sensores, modelados por mediciones, que evalúan la satisfacción de las restricciones y caracterizan el conflicto y los objetos implicados.
- Puede accionar los planes, para mejorar el cumplimento de las restricciones.

La generalización de un agente está ligada con un conjunto de restricciones, es decir, las necesidades de la generalización y la evaluación de la solución obtenida es dirigida por las restricciones. Considerando el conjunto de datos a generalizar se organiza jerárquicamente en *niveles de análisis* [Ruas, 1999][Ruas, 2000]. Usando dichos niveles, se facilita la caracterización de conflictos, la aplicación de algoritmos y se realiza la generalización de forma más eficiente. Así, todas las restricciones, operaciones, etc., están limitadas por un cierto nivel. El prototipo de AGENT fue construido en dos niveles [Barrault et al., 2001], [Duchêne, 2003].

- El nivel micro se encarga de las entidades geográficas individuales, tal como un edificio o un camino.
- El nivel meso se refiere a grupos de entidades geográficas (micro objetos), por ejemplo un bloque de ciudad integrado por edificios.

La generalización de un agente se realiza durante un ciclo de vida, que se observa en la Figura 2.10, es decir, considerando una secuencia de diversas etapas.

- 1. Caracterización del agente (es decir, evaluación de sus restricciones);
- 2. Si la generalización es necesaria (es decir, las restricciones no están satisfechas), se selecciona y se ejecuta para mejorar el cumplimento de las restricciones;

 De nuevo caracterizar las restricciones para decidir si la situación del agente mejoró.

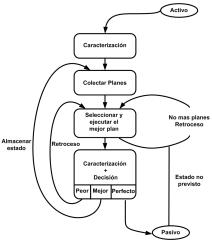


Figura 2.10. Ciclo de vida de un agente [Barrault et al., 1991]

El ciclo de vida de un agente termina cuando se alcanza un estado perfecto (todas las restricciones están satisfechas por completo) o no existe posibilidad de mejorar su estado. Una modificación del *algoritmo de Hill-climbing* se utiliza para comparar y buscar el mejor de los diferentes estados generados por el proceso iterativo de generalización. Este estado es entonces utilizado para actualizar el agente.

2.5.2 El Modelo CartACom

El modelo CartACom diseñado por Cécile Duchêne [Touya, 2008], realiza el proceso de generalización, al igual que el modelo AGENT, utilizando el paradigma de multiagentes, CartACom significa "cartografía con la comunicación de agentes", así el modelo se basa en la interacción transversal entre agentes.

Este modelo toma en cuenta las limitaciones de las relaciones entre dos objetos geográficos, los distintos agentes se comunican con sus vecinos y escogen el algoritmo apropiado de generalización, como desplazamiento o rotación.

CartACom se utiliza principalmente para la generalización de datos topográficos rurales.

2.6 Evaluación de la Calidad de la Generalización

La *calidad* de la información geográfica se puede definir, como la habilidad de dicha información para satisfacer las necesidades de los usuarios que la requieran [Podolskaya et al., 2007], semánticamente, en dicha *calidad* se consideran dos problemas principalmente, estos son, *dato a dato* o *aplicación a dato*, considerando las necesidades del usuario [Fischer 2009], por lo que el evaluar y asegurar dicha calidad debe ser un objetivo primordial en el proceso de generalización. Dicha evaluación, permitirá mejorar los resultados obtenidos y evitar errores, al proveer información acerca de la calidad de la información de los datos generalizados, sin embargo, el desarrollo de métodos para evaluar dichos resultados, ha recibido poca atención.

La evaluación de la calidad de los conjuntos de datos geográficos, es necesaria durante, antes y después del ciclo de la generalización [Weibel, 1995].

- Antes de la generalización, permite conocer el proceso de adquisición, por ejemplo si los datos de entrada son mapas en papel, se tendrá que realizar un proceso de digitalización, el estado de los datos que se generalizarán y la opinión de cartógrafos expertos acerca del proceso de generalización que se utilizará.
- Después de la generalización, permite comparar y clasificar los diferentes métodos de generalización, con base en sus resultados, y conocer la calidad de los datos obtenidos.
- Durante el proceso de generalización, se permite controlar el proceso de generalización evaluando sus resultados continuamente y obtener una solución óptima, utilizando diferentes parámetros en el proceso de generalización.

El proceso de evaluación, envuelve componentes cuantitativos, por ejemplo las violaciones a relaciones topológicas, y cualitativos, por ejemplo, claridad en la imagen de un mapa, estos componentes, se pueden representar como características cuantitativas y cualitativas.

2.6.1 Características Cuantitativas

Las características cuantitativas, son aquellas que indican alguna medida, dada por alguna especificación.

Algunas características cuantitativas son [Weibel, 1995].

- Medidas Globales, éstas indican el grado de generalización y si son constantes en todo el mapa, por ejemplo la densidad de objetos utilizando la Ley del Radical de Töpfer.
- Medidas Geométricas, nos indican dimensiones mínimas, grados de distorsión con respecto a las formas de los objetos geográficos.
- Medidas Topológicas, el propósito de estas medidas es identificar violaciones de relaciones topológicas, las cuales deben ser iguales con respecto a las que se encuentran en el conjunto de CDF.
- Medidas relacionadas con el software de generalización, son medidas que se relacionan con el desempeño adecuado del software de generalización, como son el tiempo de CPU, tiempo de horas de uso de los usuarios para un proceso de generalización en particular, tiempo de uso y costo de los equipos de generalización, tiempo de generalización de un mapa, etc.

2.6.1.1 Evaluación de la calidad de un conjunto de datos generalizados utilizando características cuantitativas

Uno de los primeros trabajos que se desarrollaron para evaluar la calidad de la información geográfica generalizada, utilizando características cuantitativas fueron los propuestos por:

 McMaster en 1987, diseña una variedad de características geométricas para analizar las distorsiones de las formas en términos de la diferencia de las líneas originales y las simplificadas [Goodchild et al., 1999].

- Mustiere en 1995, debido a que el método de McMaster, es válido sólo para líneas enteras, propone una serie de características que analizan las líneas cartográficas para posibles conflictos debido a la eliminación o agregación de símbolos en el proceso de generalización [Goodchild et al., 1999].
- Mark de Berg en 1995, verifica la consistencia topológica básicamente en líneas que se intersecten con sus líneas vecinas o puntos que causen alguna violación debido a la simplificación de líneas [Berg et al., 1995], [Goodchild et al., 1999].

Otros trabajos que proponen otras metodologías para evaluar la calidad de la información utilizando características cuantitativas, son los que se presentan a continuación.

2.6.1.1.1 Medidas de Calidad para la Generalización de Polígonos

Este trabajo presenta una propuesta para medir la calidad de la información en la generalización de polígonos, utilizando medidas propuestas por diversos autores. Como datos de prueba, utiliza una base de datos topográfica de información catastral de edificios [Podolskaya et al., 2007].

Identifica dos objetivos de la generalización para poder ser visualizados correctamente por parte del usuario, por un lado la cantidad de información se reduce en el conjunto de datos generalizado resultante, y por el otro, la información que contiene este conjunto debe ser similar a la información del conjunto de datos fuente que lo generó.

Para reducir los polígonos se identifican dos operaciones de generalización principalmente, agregación y eliminación.

La similitud entre el conjunto de datos fuente y generalizado se define con base en características cuantitativas, utilizando las siguientes medidas.

V_N, indica la reducción entre los vértices de los polígonos.

- V_P, el tamaño del objeto antes y después de la generalización, es decir, el valor de sus respectivos perímetros.
- V_{TF}, La similitud entre las formas, utilizando para ello la "Función Turning", la cual fue desarrollada por Frank y Ester [Frank and Ester, 2006] [Podolskaya et al., 2006]. Esta medida describe la forma de un objeto geográfico comparando el valor del perímetro contra el valor de la pendiente.
- V_A, El área de ambos polígonos.

Por último, se suman los valores de cada una de las medidas en una sola, utilizando la Ecuación 2.2.

$$V = C_{TF} * V_{TF} + C_N * V_N + C_A * V_A + C_F * V_F$$
 Ecuación 2.2

Donde:

$$C_{TF} + C_N + C_A + C_P = 1$$

 C_{TF} , C_N , C_A , C_F son pesos que se definen para cada una de las medidas utilizadas, estos valores se definen en forma arbitraria con base en la calidad visual del resultado.

2.6.1.1.2 Medida de Similitud Cuantitativa para Mapas

Este trabajo [Frank and Ester, 2006], se basa en la comparación de dos conjuntos de datos geográficos que se generan con respecto a un conjunto de datos fuente. Los resultados finales de cada comparación se evalúan con respecto a las preferencias de un usuario, considerando a éstas para realizar la generalización. Esto se muestra en la Figura 2.11.

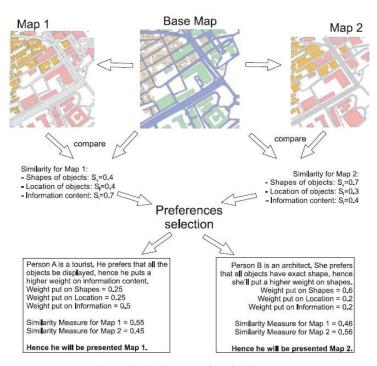


Figura 2.11. Comparación de dos mapas

Aquí, se presenta un método basado en tres medidas, que consideran cada una de ellas, operadores de generalización, dichas medidas son:

- Similitud de forma (SS), la cual se basa en los cambios en la forma del objeto geográfico, originado por los operadores de generalización, como la simplificación. A su vez, describe propiedades tales como el área, perímetro o longitud/angostura, y las utiliza para el cálculo de la similitud de forma, por medio de la "Función Turning" [Frank and Ester, 2006].
- Similitud de localización (LS), denota cambios en la posición del objeto geográfico con respecto a sus vecinos, originados por operadores de generalización como el desplazamiento. Para esto se utiliza la distancia entre el objeto geográfico con respecto a cada uno de sus vecinos, los cuales se identifican utilizando la adyacencia de celdas de Voronoi. A partir de esto se calcula la distancia, tomando en cuenta el centro de gravedad de cada uno de los objetos geográficos.

Un caso particular, que se puede presentar, es donde el objeto geográfico o alguno de sus vecinos se agregue o fusione con otro objeto. En este caso, el objeto resultante de la agregación no es el mismo con respecto a los objetos individuales, por lo que esta medida de similitud no se emplea.

 Similitud semántica en el contenido (SCS), la cual denota cambios en el contenido del mapa, al usar los operadores de generalización tales como la selección, exageración y agregación, fusión y tipificación.

Durante el proceso de generalización, muchos objetos pueden ser agregados o fusionados con otros, borrados, etc., esto genera pérdida de información, por lo cual este método considera métodos estadísticos para medir la cantidad de información dentro de un determinado conjunto de objetos geográficos, utilizando para ello regiones de Voronoi.

Por último, se suman los valores de cada una de las medidas en una sola, utilizando la Ecuación 2.3.

$$Quality(Map_A, Map_B) = w_1 * SS(Map_A, Map_B) + w_2 * LC(Map_A, Map_B) + w_3 * SCS(Map_A, Map_B)$$
Ecuación 2.3

Donde:

$$w_1 + w_2 + w_3 = 1,$$

 w_1, w_2, w_3 son pesos que se definen para cada tipo de usuario.

Un valor bajo en $Quality(Map_A,Map_B)$ indica que el mapa resultante ha tenido una gran pérdida de información.

2.6.1.1.3 Evaluación de la calidad utilizando clases

Basado en el Proyecto AGENT, este método, considera que el proceso de generalización es dividido en tres partes [Bard, 2003].

- La descripción y el análisis del contexto espacial de las características geográficas.
- El algoritmo de generalización, adaptado para este contexto.

 La evaluación de la generalización, es decir, la validación de los resultados arrojados por el proceso de generalización.

Este método se orienta en:

- Valorar si los datos preservados son pertinentes, es decir, que correspondan a las necesidades del usuario.
- Valorar si las modificaciones presentadas en los datos generalizados son aceptables, es decir, si los datos resultantes son correspondientes con la realidad.

Como parte del método, la información geográfica es representada a través de esquemas geográficos, los cuales se componen de clases de objetos geográficos que representan características de la realidad, por ejemplo, edificios, caminos, vegetación, etc. Para estas clases la generalización se puede definir como un proceso que cambia su estado, de un *estado inicial* (antes de la generalización) a un *estado final* (después de la generalización), estos estados se indican en cada clase. Se agregan dos clases para poder evaluar la calidad de la información.

Clase Caracterización, esta clase reporta la descripción de la información antes y después de la generalización, se encuentra asociada con cada una de las clases de los objetos geográficos. Contiene un método denominado Characterise() que se asocia a los valores de los atributos del objeto generalizado.

Los datos son caracterizados antes y después de la generalización, esto es, la caracterización consiste en la descripción de cada una de las propiedades de un objeto geográfico en particular, por ejemplo, geométricas, topológicas, etc., las cuales pueden ser asociadas con una o varias medidas, por ejemplo, área, perímetro, etc. Por lo que la caracterización requiere un análisis de la información geográfica a un nivel *micro, meso* y *macro* [Ruas, 2000],[Bard, 2003].

Clase Evaluación, reporta los valores de calidad de la información generalizada, se asocia con cada clase de objetos generalizados, y contiene los criterios de calidad para cada objeto geográfico y de forma general. Se definen dos métodos en esta clase:

- O Evaluation(), contiene una medida para cada criterio y calcula e interpreta la diferencia, con base en un umbral, el cual se aplica a los resultados obtenidos con respecto a datos de referencia, estos últimos son definidos por una caracterización inicial de los objetos.
- Agregation(), contiene una medida general para el objeto generalizado, este método calcula e interpreta los resultados obtenidos en tres niveles, a nivel de objeto geográfico, de clases y de mapa con todas las clases en general con base en valores de referencia.

Los distintos valores de referencia son definidos con base en criterios de diversos cartógrafos experimentados, definiéndose una jerarquía entre los criterios.

2.6.2 Características Cualitativas

La implementación de un procedimiento para evaluar características cualitativas, es complicado, por la dificultad de formular características cualitativas, debido, a que se deben de caracterizar rigurosa y consistentemente aspectos subjetivos de la generalización, por lo que en general, las características cualitativas se encuentran ligadas con algunas características cuantitativas, algunos ejemplos son:

- Linaje de los datos.
- Propósito de los datos.

2.6.2.1 Evaluación de la Calidad de un CDG utilizando Características Cualitativas

Uno de los primeros trabajos para evaluar la calidad de la información, fue propuesto por Ehrliholzer en 1995 [Goodchild et al., 1999], quien propone una metodología para integrar características cuantitativas con juicios cualitativos dados por expertos cartográficos, donde ambos resultados pueden ser integrados en una medición o colocados en un ranking de escala.

Otros trabajos que evalúan la calidad de la información geográfica, utilizando características cualitativas, con metodologías más concisas son los siguientes.

2.6.2.1.1 Evaluación de la calidad en algoritmos de generalización

Este trabajo [Filippovska, 2008] considera las distorsiones geométricas que se generan cuando se generaliza un objeto geográfico. Por ejemplo, que sufra un desplazamiento con respecto al objeto geográfico original, o bien que alguna línea que forme parte del polígono que conforma el objeto geográfico sufra un desplazamiento, en alguna dirección, lo cual se identifica cuando se comparan los datos generalizados con los datos originales.

Se basa en características cualitativas, ya que busca los valores mínimos en las respectivas medidas que utiliza, donde los objetos geográficos analizados sean visualmente correctos, es decir, donde las distorsiones no sean tan claras, las medidas utilizadas son:

- Identificando las diferencias directamente entre dos objetos geográficos utilizando la distancia de Hausdorff o la intersección del área.
- Describiendo las diferentes características de un objeto geográfico en particular, donde a cada una de las características se pueden representar como un valor numérico o bien como un vector que se puede comparar en un espacio de características con una función de distancia

2.6.2.1.2 Evaluación de la calidad utilizando ontologías

En esta sección se presentan algunos trabajos que evalúan la calidad de la información contenida en un conjunto de información geográfica generalizado, utilizando ontologías.

2.6.2.1.2.1 Modelo Ontológico de las relaciones geográficas para mapas generalizados

Un camino para enriquecer las bases de datos geográficas, es identificar subclases con características comunes de alguna entidad geográfica, por ejemplo, los caminos, se pueden dividir en carreteras, caminos locales, avenidas, etc. Estas clases incluyen nombres, atributos y relaciones topológicas. Estas bases de datos geográficas, pueden ser usadas para considerar esta información geográfica en aplicaciones con la misma temática y a diferentes escalas [Dutton, 2006].

Para describir los diferentes datos geográficos, sus propiedades y relaciones se utilizan ontologías, con las cuales se pueden realizar dichas descripciones, además de enlazar información geográfica de la temática o propósito de la información geográfica a la base de datos geográfica, enriqueciendo esta última, por lo que se pueden desarrollar ontologías que vinculen las distintas temáticas geográficas con sus representaciones cartográficas.

Para poder generar las ontologías, se deben especificar las propiedades semánticas y relaciones de los distintos objetos geográficos, por lo que se debe contar con herramientas de software, como por ejemplo Protégé, las cuales permitan representar:

- Objetos geográficos en forma jerárquica, junto con sus atributos.
- La posibilidad de compartir primitivas geométricas entre los objetivos geográficos.
- Relaciones entre los distintos objetos geográficos y entre ellos mismos.

Las distintas relaciones espaciales y semánticas son determinadas en forma analítica, así también, cuando estas relaciones pertenecen al CDF y se comparan con las relaciones del CDG, los resultados son analizados e interpretados también en forma analítica.

Por ejemplo, se considera la carta náutica mostrada en la Figura 2.12.

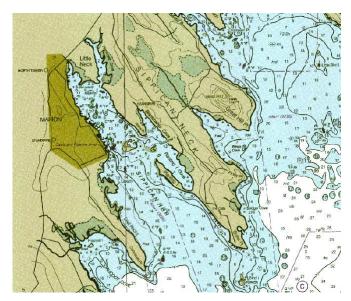


Figura 2.12. Región Sippican, Carta Náutica a escala 1:40K

Se genera una jerarquía, con los objetos geográficos marcados con puntos rojos (objetos geográficos tierra) y azules (objetos geográficos agua), agrupando mediante un óvalo café, aquellos que se encuentren relacionados, esto se muestra en la Figura 2.13.

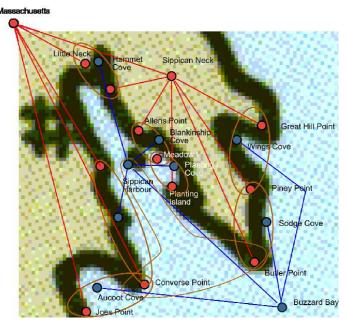


Figura 2.13. Jerarquización

Por último, se genera la ontología para la región Sippican, parte de la cual se muestra en la Figura 2.14.

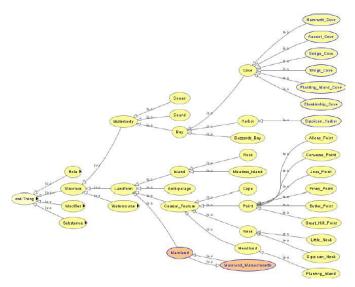


Figura 2.14. Ontología Sippican.

Para generar esta ontología, se consideraron los siguientes aspectos:

- Conceptos autónomos, que permitan representar las características intrínsecas, acordes con la realidad, procesos y roles.
- Refinación de conceptos, como particiones, por ejemplo pequeña, grande y largo.

Las clases son inferidas por un experto.

2.6.2.1.2.2 Similitud Semántica entre Sistemas de Objetos Geográficos aplicada a la Generalización de Datos Geoespaciales

Este trabajo [Moreno, 2007], presenta un método para verificar la consistencia de datos geoespaciales generalizados, utilizando una representación conceptual, siendo el caso de estudio la generalización de los ríos y las curvas de nivel.

Se basa en las siguientes etapas.

 Análisis, en la cual se extraen las particularidades que poseen los objetos geográficos, utilizando medidas cuantitativas, las cuales tienen como propósito caracterizar las relaciones entre los ríos y las curvas de nivel, como se muestra en la Figura 2.15, dependiendo de la forma en que cruzan y de la dirección del flujo. Esta caracterización se usa para identificar relaciones consistentes cuando un río cruza sólo una curva de nivel por un punto de convexidad, si un río cruza una línea de elevación en más de una ocasión se tiene una inconsistencia.

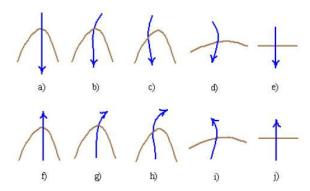


Figura 2.15. Diferentes relaciones entre los ríos y las líneas de elevación de contorno

Desde el punto de vista cartográfico, se presenta una inconsistencia cuando el río *pasa por* una convexidad máxima de la curva de nivel, como se muestra en la Figura 2.15a, por una convexidad, lo cual se puede observar en la Figura 2.15b, c y d o por una concavidad, lo cual se puede identificar en la Figura 2.15f, g y h, en los otros casos, el río pasa por una parte recta de la línea de elevación de contorno, como se observa en la Figura 2.15 e y j.

Entonces, se consideran seis diferentes relaciones: pass by maximum convexity, pass by almost maximum convexity, pass by convexitym, pass by straight, pass by almost maximum concavity y pass by maximum concavity, cada una de estas relaciones se caracteriza por un descriptor cuantitativo.

• *Sintesis*, en esta etapa se genera una representación conceptual (ontología) que cualitativa y explícitamente representa las seis diferentes relaciones entre los ríos y las líneas de elevación de contorno, obteniendo tuplas, llamadas descripciones semánticas, por ejemplo {River_1, PASS BY MAXIMUM CONVEXITY, Contour_8}, las cuales se pueden describir como descriptores cualitativos. Se muestra un fragmento de la ontología en la Figura 2.16.

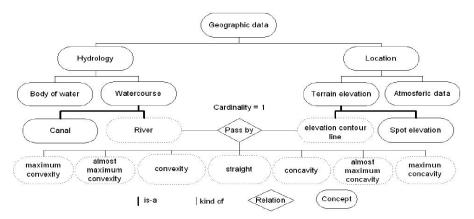


Figura 2.16. Fragmento de la Ontología

• Verificación, consiste de una comparación entre dos descripciones semánticas (del conjunto de datos fuente y generalizado), en donde se mide la distancia semántica (confusión) entre los conceptos de la ontología. A partir de la medición se definen tres conceptos: Igual, Desigual y Equivalente, los cuales caracterizan el estado de los datos generalizados. Igual y Equivalente representan consistencias, mientras que Desigual representa una inconsistencia.

Un ejemplo de consistencia se muestra en la Figura 2.17, del lado derecho se muestra el conjunto de datos fuente y del lado izquierdo el conjunto de datos generalizado, se puede ver que la curva de nivel y el río han mantenido formas similares en los datos generalizados, por lo que el río cruza a la línea de elevación de contorno por el mismo lado.

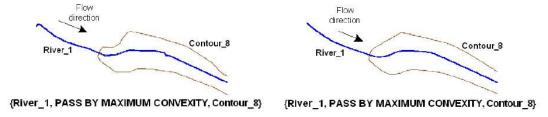


Figura 2.17. Identificando una consistencia

Por otro lado, un ejemplo de inconsistencia se muestra en la Figura 2.18, del lado derecho se muestra el CDF y del lado izquierdo el CDG, se puede observar que las formas de la curva de nivel y el río han cambiado en los datos generalizados, por lo que el río cruza a la línea de elevación de contorno en otro punto muy distinto al original.

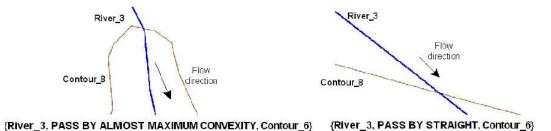


Figura 2.18. Identificando una inconsistencia

Este método no depende de coordenadas, escalas, unidades de medida, o proyección cartográfica.

2.6.2.1.3 Evaluación de la calidad de la información utilizando medidas de similitud semántica

En esta sección se presentan algunos trabajos que evalúan la calidad de la información contenida en un conjunto de información geográfica generalizado, utilizando medidas de similitud semántica.

2.6.2.1.3.1 Un modelo integral para la evaluación cualitativa espacial utilizando medidas de similitud

En [Fonseca and Li, 2006], se presenta un modelo basado en medidas de similitud semántica, denominado TDD (Topología-Dirección-Distancia), por medio del cual se mide la similitud entre dos escenas.

Este trabajo integra cuatro modelos de similitud conceptual, los cuales son:

- Modelo geométrico.
- Modelo de contraste de características.
- Modelo de transformación.
- Modelo de estructura de alineamiento.

Realiza la medición de coincidencias (C) y diferencias (D) entre dos escenas, la medida de similitud final (S) está dada por la Ecuación 2.4, la cual es una combinación de ambas.

$$S = C - D$$
 Ecuación 2.4

El modelo de estructura de alineación considera que en la comparación entre las dos escenas, cada uno de los objetos de una escena debe corresponder con cada uno de los objetos de la otra escena. Entonces, el proceso de comparación de similitud incluirá coincidencias, diferencias alineables y no-alineables de forma separada, por lo que el valor de **D** está dado por la Ecuación 2.5.

$$S = diferencias alineables + diferencias no - alineables$$
 Ecuación 2.5

Este modelo considera la similitud entre relaciones y Issimilitud de atributos, de acuerdo con la Tabla 2.1, utilizando diferentes pesos dependiendo de la evaluación de similitud que se requiera.

Nivel de Comparación Tipos de medidas de similitud Escena Relaciones Espaciales Topológica Dirección Métrica de distancia Distribución No-espaciales Atributo distancia Objeto Atributos Geométrica Tipos de objeto Temática Comparación de atributos

Tabla 2.1. Elementos básicos en el proceso de medición espacial.

Utilizando la Ecuación 2.5 y Tabla 2.1, además, considerando la escena y los objetos que la componen, se redefine la Ecuación 2.4, a un nivel de escena y objetos, lo cual se muestra en la Ecuación 2.6.

$$S = (C_{essena} + C_{objeto}) - (D_{escena} + D_{objeto})$$
 Ecuación 2.6

Incluyendo las medidas de similitud para las relaciones espaciales, como son topológicas, dirección y métrica de distancias, se obtiene una medida de similitud para la escena, por ejemplo, la Ecuación 2.7.

$$C_{escena} = (C_{topalógica} + C_{dirección} + C_{métricaDistancia})$$
 Ecuación 2.7

A un nivel de objeto, se pueden obtener medidas de similitud para coincidencias y diferencias, utilizando atributos geométricos y temáticos, como se muestra en la Ecuación 2.8.

$$C_{objeto} - (C_{geométrico} + C_{temático})$$
 Ecuación 2.8

En la Figura 2.19 se muestra el diagrama de vecinos conceptuales, con distancias conceptuales determinadas, para el cálculo de la similitud entre relaciones topológicas, para dos objetos polígonos, en el modelo.

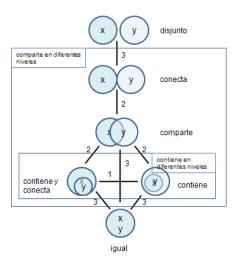


Figura 2.19. Vecinos conceptuales de relaciones topológicas para dos objetos geográficos polígonos.

2.6.2.1.3.2 Similitud de Escenas Espaciales

En [Bruns and Egenhofer, 1996], se presenta un modelo para la evaluación de la similitud espacial, basado en relaciones topológicas y relaciones de distancia, y sus vecinos conceptuales correspondientes.

El modelo utiliza el diseño de tres tipos de vecinos conceptuales para tres tipos de relaciones espaciales consideradas, las cuales son:

- Relaciones topológicas.
- Relaciones de distancia.
- Relaciones de dirección.

Dadas dos escenas con igual número de objetos espaciales, diferentes relaciones espaciales y considerando los esquemas de vecinos conceptuales, existe un número mínimo de cambios graduales, que transforman una escena en otra, el número de cambios graduales es considerado proporcional a la similitud entre las dos escenas. Por lo que entre más cambios se tengan, las escenas son menos similares.

La similitud está determinada por un proceso que, gradualmente reemplaza las relaciones espaciales por sus vecinos conceptuales, donde dicho proceso intenta construir una escena a partir de otra. Las escenas derivadas se clasifican con respecto al número de cambios, por lo que aquellas escenas que tienen menos cambios, son más similares que aquellas con más cambios.

En la Figura 2.20 se muestra un análisis de una escena, donde la única diferencia entre la Figura 2.20a y 2.20b, es que el objeto C, ha sido movido; para determinar el grado de similitud entre ambas escenas, se evalúan las relaciones espaciales y se cuentan sus diferencias.



Figura 2.20. Dos escenas similares y sus diferencias.

2.7 Comentarios del capítulo

El proceso de generalización es un proceso necesario para generar conjuntos de información geográfica, de acuerdo con alguna temática o escala en particular, la cual use alguna aplicación en específico, sin embargo, cuando se toman decisiones con base en esta información, es necesario contar con certeza que dicha información es de calidad.

Debido a que la información generada, se produce con base en operadores de generalización, éstos pueden generar diversas transformaciones en la información producida, por lo cual se han desarrollado algunos trabajos, que en algún sentido evalúan la calidad de los datos generados, tomando en cuenta características cuantitativas principalmente, y cualitativas con base en la temática de la que se trate.

El trabajo desarrollado en la presente tesis, se enfoca principalmente en la evaluación de la información con base en características cualitativas, tomando como referencia ciertas características cuantitativas, que a diferencia de los trabajos mostrados en este Capítulo, tiene como objetivo, informar al usuario en un lenguaje cotidiano si la información a utilizar en las aplicaciones, es consistente o no a la información de la cual se generó y que considerando su linaje, es correcta. Ciertamente, la evaluación de la calidad de la información no solamente se realiza al terminar el proceso de generalización, sino en cada una de las fases que dicho proceso involucre, o bien, no solamente tomando las características que el presente trabajo consideró. Así, este trabajo se fusionan los enfoques presentados como una alternativa más para evaluar la calidad de la información, pero a diferencia de los otros orientándose en el usuario final, ya que sin ser este un experto, el resultado presentado a él, pretende generar una idea global del estado del conjunto de datos generalizado con respecto al conjunto de datos fuente, sin ser específico devolviendo valores numéricos.

Capítulo 3. Ontologías y Similitud Semántica

En esta sección se presentan algunos trabajos relacionados con el estado del arte, en el diseño de ontologías, así como algunos enfocados en evaluar la similitud semántica en el contexto de los datos geográficos.

3.1 Ontologías

El término ontología, proviene de las raíces *ontos*, que significa *inicio* y *logos* o *tratado*, siendo un término tomado de la Filosofía, donde se le conceptualiza como "*una explicación sistemática del ser*" [Corcho et al., 2003].

Guarino and Giaretta, en 1995 [Corcho et al., 2003], proponen el uso de las palabras Ontología y ontología, siendo la primera, para su uso en Filosofía y la segunda en el sentido computacional e Ingeniería del Conocimiento.

Algunas definiciones generales, que intentan definir el término ontología, son las siguientes [Corcho et al., 2003].

- Para Neches y sus colegas, en 1991, "una ontología define los términos básicos y relaciones en un vocabulario, de un tema en particular, así como sus reglas para combinar términos y relaciones que definan extensiones del vocabulario".
- Gruber, en 1993, define una ontología como "una especificación explícita¹ de una conceptualización²".
- Tomando como base la definición de Gruber, Borst en 1997, la define como "una especificación formal de una conceptualización compartida³".

¹ El término "explícita", significa que el tipo de conceptos usados y las limitaciones de su uso son explíctamente definidos.

² El término "conceptualización", se refiere a un modelo abstracto de algún fenómeno en el mundo, que pueda ser identificado con los conceptos relevantes de dicho fenómeno [Corcho, 2003]

³ El término "formal"

- Studer, en 1998, combina las definiciones de Gruber y Borst en una sola, y define el término de ontología como "una especificación explícita y formal, de una conceptualización compartida".
- Guarino y Giaretta, en 1995, presentan siete definiciones, donde definen al término ontología, sin embargo, ellos proponen que una ontología es "una teoría lógica, la cual especifica parcialmente una conceptualización".
 - 1. Una disciplina filosófica.
 - 2. Un sistema conceptual informal.
 - 3. Una cuenta semántica formal, es decir, se concibe a la ontología como un marco conceptual a nivel semántico.
 - 4. Una especificación de una conceptualización.
 - 5. Una representación conceptual de un sistema dentro de una teoría lógica.
 - Caracterizada por propiedades formales específica.
 - Caracterizada sólo por sus propósitos específicos.
 - 6. Un vocabulario usado por una lógica teórica.
 - 7. Una (meta-nivel) especificación de una lógica teórica.

3.1.1 Construcción de una ontología

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

- Claridad y Objetividad. La ontología debe comunicar con eficacia el significado deseado de los términos definidos. Las definiciones deben ser objetivas y en un lenguaje natural.
- 2. *Coherencia*. La ontología debe permitir que las inferencias sean consistentes con las definiciones.
- Completa. La ontología debe expresar definiciones completas sobre definiciones parciales.
- 4. *Extensibilidad*. Una ontología se debe diseñar anticipando el uso de vocabulario compartido; es decir, se debe poder definir términos nuevos para aplicaciones

especiales basadas en el vocabulario existente, de manera que no se requieran revisar las definiciones existentes.

- 5. Minimizar el compromiso ontológico. Una ontología puede requerir un mínimo de compromiso ontológico, que sea suficiente para soportar las actividades relacionadas con compartir el conocimiento. Debido a que el compromiso ontológico está basado en el uso constante de vocabulario, se puede reducir al mínimo especificando una teoría más débil (que se permite en la mayoría de los modelos) y definiendo solamente esos términos que sean esenciales para la comunicación de conocimiento consistente.
- 6. *Principio de distinción ontológica*, lo cual significa que las clases en una ontología deben ser disjuntas, para lo cual se utiliza un criterio denominado de Identidad.
- 7. Diversificación de jerarquías, para incrementar la fortaleza de los mecanismos de herencia múltiple.
- 8. Modularidad para minimizar la unión entre módulos.
- 9. Minimizar la distancia semántica entre conceptos hermanos. Conceptos similares son agrupados y representados como subclases de una clase y pueden ser definidas usando las mismas primitivas, dnde los conceptos son poco similares y representados en la jerarquía.
- 10. Estandarizar nombres hasta donde sea posible.

3.1.2 Tipos de ontología

Los tipos más comunes de ontologías, de acuerdo con su nombre, dan una idea del conocimiento que incluyen [Corcho et al., 2003].

Ontologías que representan el conocimiento (Knowledge Representation Ontologies), capturan las primitivas de representación para formalizar el conocimiento en paradigmas que representen el conocimiento. El ejemplo más representativo de este tipo de ontologías es "the Frame-Ontology", la cual captura las primitivas de representación (clases, instancias, etc.), basadas en determinados lenguajes.

- Ontologías generales o comunes (General/Common Ontologies). Incluyen vocabularios relacionados con cosas, eventos, tiempo, espacio, funciones, causalidad, conductas, etc.
- Meta-ontologías (*Meta Ontologies*), también llamadas como ontologías genéricas o
 "Core Ontologies", las cuales se reutilizan a través de diversos dominios. El ejemplo
 más representativo puede ser una ontología de mereología, la cual incluye el término
 parte de (part-of).
- Ontología de Dominio (*Domain ontology*). Describe el vocabulario relacionado a un dominio genérico como medicina o automóviles y sus relaciones, de las actividades que tienen lugar en este dominio, y de las teorías y principios elementales que gobiernan dicho dominio.
- Ontología de Tarea (*Task ontology*). Describe un vocabulario sistematizado de los términos usados para resolver problemas asociados con tareas que pueden o no, pertenecer al mismo dominio. Estas ontologías proveen un conjunto de términos, que describen genéricamente como resolver un tipo de problemas.
- Ontología de Tarea de Dominio (*Domain-Task Ontologies*). Son ontologías de tarea reutilizables en un dominio en específico, pero no a través de otros dominios, por ejemplo, una ontología de tarea de dominio en el dominio medico, puede incluir términos relacionados con el calendario de una operación.
- Ontología de Aplicación (*Application ontology*), contienen el conocimiento necesario para modelar un dominio en particular.

3.1.3 Metodologías para construir una ontología

Se han desarrollado algunos trabajos, que proponen una serie de pasos y metodologías para el diseño de ontologías, entre los más representativos se citan los siguientes [Corcho et al., 2003].

 En 1990, Lenat y Guha publican algunos pasos generales y puntos importantes para el diseño de la ontología a utilizar dentro del proyecto CYC, proyecto relacionado con la inteligencia artificial.

- En 1995, debido a la experiencia obtenida con el diseño de una ontología empresarial, y el proyecto de ontología TOVE (Toronto Virtual Enterprise), ambas ontologías dentro del dominio empresarial, fueran propuestas y posteriormente refinadas las primeras directrices para el diseño de ontologías.
- En la 12°. Conferencia Europea para Inteligencia Artificial (ECAI'96), Bernaras, presenta un método para la construcción de una ontología en el dominio de las redes eléctricas como parte del proyecto Esprit KACTUS.
- En 1997, un nuevo método es propuesto para construir ontologías, basado en la ontología SENSUS.

3.1.3.1 Metodología de Uschold y King

Esta metodología se basa en la experiencia obtenida durante el diseño de la ontología Enterprise, aquí se modelan los siguientes pasos [Fernández-López, 1999].

- Identificar el propósito, cual es el objetivo, para el cual la ontología será construida.
- Construir la ontología, este paso se conforma por.
 - o Diseñar la ontología, lo cual significa.
 - Identificar los conceptos clave y relaciones en el dominio de interés.
 - Corregir definiciones ambiguas para conceptos y relaciones.
 - Identificar los términos que se refieran a conceptos y relaciones.
 - Codificar, representar el conocimiento adquirido en el paso anterior en un lenguaje formal.
 - o Integrar ontologías existentes, se refiere a utilizar ontologías ya existentes.
- Evaluación, se realiza un análisis técnico de la ontología, asociado con el software y documentación de la misma.
- Documentación, documentar cada paso del desarrollo y diseño de la ontología.

3.1.3.2 Methontology

La metodología Methontology [Fernández-López, 1999], fue diseñada en el Laboratorio de la Universidad Politécnica de Madrid, esta metodología permite la construcción de ontologías a un nivel de conocimiento e incluye:

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

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

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

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

3.1.4 GEONTO-MET

GEONTO – MET [Torres, 2008], es una metodología cuyo enfoque principal, se basa en proporcionar una técnica que permita conceptualizar el dominio geográfico, por medio de una ontología de este dominio. Con esta técnica se busca proporcionar un mecanismo alterno para la *integración* y *recuperación* de datos geoespaciales, los cuales puedan representarse a través de una descripción semántica basada en un *diagrama de descripción simbólica*, el cual se obtiene a partir de un *esquema conceptual*.

Por lo tanto, la descripción semántica está compuesta por relaciones y conceptos que envuelven diversos contextos geográficos de este dominio y se utiliza para describir explícitamente las propiedades mediante relaciones de los objetos geográficos, en donde estas relaciones son del tipo topológico y geométrico.

En la Figura 3.1 se muestra la *estructura conceptual* del método GEONTO-MET, el cual está compuesto de cuatro etapas fundamentales para llevar a cabo la integración y recuperación de objetos geoespaciales; estas etapas son:

• Conocimiento del dominio geográfico, este conocimiento está basado en diversas fuentes de información⁴, las cuales han sido generadas por diversos expertos de este dominio para llevar a cabo el análisis del entorno geográfico.

-

⁴ Las fuentes de información consideradas para esta tesis, son los datos recopilados por el INEGI (Instituto Nacional de Estadística, Geografía e Informática), tales como diccionarios de datos geoespaciales, metadatos de datos geoespaciales, mapas, entre otros.

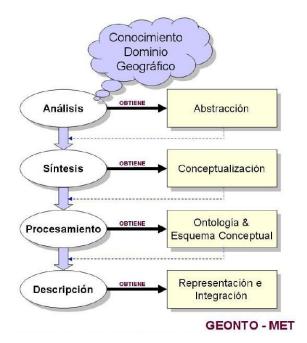


Figura 3.1. Estructura conceptual de GEONTO – MET

- Análisis, en esta etapa, se realiza una investigación exhaustiva sobre los objetos geográficos que se encuentran involucrados en el dominio geográfico, con el objeto de poder llevar a cabo un proceso de abstracción. Como resultado de esta etapa, se obtiene una abstracción completa del dominio geográfico, identificando todos aquellos objetos involucrados en este dominio.
- Síntesis, en esta etapa se realiza un análisis de los objetos geográficos involucrados en el dominio para obtener como resultado una clasificación esencial 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, con lo que se puede obtener una clasificación de los objetos geográficos. De igual forma, se cuenta con las propiedades o relaciones inmersas entre estos objetos; así como una jerarquía a partir de estas clases de entidades abstractas, en donde se puede observar que el nivel más superior de una entidad geográfica es el término "cosa⁵".

Por otra parte, es necesario restringir el dominio que puede tener una ontología del universo geográfico, por lo que, en esta etapa se realiza un análisis de los diversos

_

⁵ El término *cosa* puede referirse semánticamente a algo que puede ser un objeto, un pensamiento, acción o existencia referente a una entidad; es decir, la realidad que es la razón fundamental de las percepciones.

contextos que pueden presentar los objetos geográficos, de acuerdo con las fuentes de información analizadas para acotar el dominio⁶. 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. Estos significados pueden variar ligeramente, pero en esencia el vocabulario utilizado para describirlos es el mismo. Por lo tanto, la etapa de síntesis se encarga también, de generar un método para definir esquemas conceptuales, basándose en las características principales que describen a los objetos geográficos cuando éstos son representados en un modelo lógico. Como resultado de esta etapa, se obtiene una conceptualización del dominio geográfico.

• Procesamiento, es necesario 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 qué tipo de significado toman estos conceptos de acuerdo con el contexto en el que se encuentran dirigidos.

La etapa de *procesamiento* se encarga de discernir con respecto a los conceptos que se encuentran en la conceptualización del dominio geográfico y la interpretación de éstos en una región espacial específica. Por lo tanto, el objetivo de esta etapa es obtener los diversos significados que puede tener un concepto en diversos contextos hasta converger en la definición más abstracta del mismo. Como resultado de esta etapa se obtiene una ontología del dominio geográfico; así como una ontología de aplicación con su correspondiente esquema conceptual.

Descripción, por último, con base en las ontologías de dominio y aplicación; así
como las relaciones y contextos que son percibidos en este universo geográfico, esta
etapa se encarga de generar una descripción de los objetos involucrados en un
contexto particular.

_

⁶ La expresión "acotar el dominio" significa restringir los objetos geográficos que envuelven al universo de discurso geográfico, por medio de particiones basadas en diversos contextos, en donde el significado de los objetos geográficos involucrados puede variar, de acuerdo con el dominio de aplicación en que se encuentran situados o donde actúan estos objetos geográficos.

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, con un contexto determinado y dirigida hacia los objetos geográficos envueltos en este contexto.

De igual forma, se considera la definición implícita en las ontologías que pueden tener estos conceptos para recuperar *instancias* de los mismos con sus propios valores y atributos de estas instancias.

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 geográfico; siempre y cuando existan instancias de esos objetos geográficos en la base de datos.

3.1.4.1 Axiomas

Para la conceptualización del dominio geográfico, esta metodología hace uso de *axiomas*⁷ los cuales permiten validar el conocimiento, a través de expresiones lógicas que reflejen las propiedades de los objetos geográficos.

Estos axiomas son:

 En la Ecuación 3.1, se observa el primer axioma, el cual está compuesto por tres relaciones, denotadas en el conjunto A₁.

$$A_1 = \{es, tiene, hace\}, \rho \in A_1$$

$$\rho = es, tiene, hace$$
 Ecua

Ecuación 3.1

En donde el conjunto A1 representa los axiomas básicos que describen a las *relaciones* simples.

-

⁷ Un axioma es una verdad evidente sobre la cual descansa el resto del conocimiento o sobre la cual se construyen otros conocimientos. En matemáticas un axioma no es necesariamente es una verdad evidente, sino una expresión lógica utilizada en forma de una deducción para llegar a una conclusión. En matemáticas se distinguen dos tipos de axiomas: lógicos y no-lógicos

- o es, describe una relación de existencia de objetos geográficos.
- tiene, indica una relación de pertenencia de propiedades de los objetos geográficos o conceptos del 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.

- Área urbana es objeto geográfico artificial.
- Río *tiene* profundidad.

El segundo subconjunto de relaciones axiomáticas se denota por A_2 y está compuesto por términos simples que representan a las preposiciones, Ecuación 3.2.

$$A_{2} = \begin{cases} a. \ ante, bajo, con, ccntra, de, desde, en, entre, hacia, hasta, \\ para, por, según, sín, so, sobre, tras, junto_a, a_través_de \end{cases}, \pi \in A_{1}$$

$$\pi = preposiciones$$
 Ecuación 3.2

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

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

- Área urbana *hace* compartir *con* área verde urbana.
- Carretera *hace* conectar *con* municipio.
- Río *hace* cruzar *por* curva de nivel.

3.1.4.2 Relaciones

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

Las relaciones simples son todas aquellas que tienen la forma siguiente, $a\rho b \in R_s$, donde $a\rho b \in C$ y $\rho \in A_1$.

Las relaciones compuestas presentan una forma compleja, $\alpha \rho b \pi c \in R_c$, donde $\alpha, b, c \in C$, $\rho \in A_1$ y $\pi \in A_2$. Este tipo de relaciones se denotan como relaciones ternarias y están enfocadas principalmente a vincular relaciones axiomáticas entre la relación "hace" que pertenece a A_1 con elementos del conjunto A_2 .

3.1.4.3 Propiedades

Una propiedad se define como un concepto agregado o asociado, en donde la agregación se realiza a través de una relación de pertenencia ("tiene"). Las propiedades se encargan de describir las características que poseen o envuelven a un concepto.

Cabe señalar que el término de propiedad está relacionado a los *atributos* que contiene un concepto. Estas propiedades pueden describir características específicas de un del mismo. Sin embargo, se considera que una propiedad a su vez es un *concepto*, pero *agregado*, el cual a su vez puede contener un conjunto de valores que completan una partición.

3.1.4.4 Habilidades

Una *habilidad* se define como la forma en la cual se realizan las operaciones sobre los *conceptos estándar*. Este término se encuentra directamente asociado a la relación simple "*hace*".

3.1.4.5 Instancia

El término de instancia está relacionado directamente con los *individuos* que pertenecen a ciertos tipos de conceptos. Por tal motivo, para la conceptualización del dominio geográfico se define a una *instancia*, como una colección de habilidades y propiedades concretas, las cuales tienen una existencia única. Además, estos individuos únicamente pueden ser creados a partir conceptos estándar.

En otras palabras, una *instancia* puede considerarse como un concepto contenido en la jerarquía formada por la relación "es", cuyas propiedades y habilidades son siempre concretas.

Como características peculiares de una instancia se tienen las siguientes,

- Es un concepto que contiene propiedades concretas.
- Es un concepto que tiene habilidades concretas.
- Es un concepto que no tiene hijos en la jerarquía.
- Es un concepto que no tiene particiones completas; ya que son entidades dinámicas.
- Las instancias son individuos que no son clases.

3.1.4.6 Conceptos

Para la ontología de este dominio se define a un *concepto* como una colección de habilidades y propiedades abstractas, las cuales presentan una existencia.

Se identifican tres tipos de conceptos, los cuoles se muestran en la Figura 3.2, estos son,

- Conceptos de Relación, este tipo de conceptos están asociados a las acciones (verbos), que puede realizar un concepto, es decir, denotan una acción u operación sobre conceptos.
- Conceptos Estándar, este tipo de conceptos se relacionan directamente con los objetos abstractos (sustantivos), los cuales son elementos que pertenecen a una clase de entidades abstracta.
- Conceptos de Clase, este tipo de conceptos sirven para generar particiones entre los conceptos de relación y los conceptos estándar.

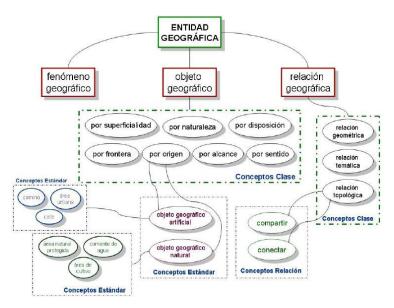


Figura 3.2. Representación gráfica de conceptos

3.2 Relaciones Espaciales

Las relaciones espaciales distinguen a la relación espacial de otra información, usualmente se codifican utilizando términos lingüísticos comunes del lenguaje humano cotidiano, que forman parte de una clase cerrada, lo cual significa que el número de elementos de la clase es pequeño, por lo que es limitado el número de categorías que se pueden formar, las relaciones espaciales son expresadas a través de verbos que describen trayectorias u otras acciones espaciales.

Matemáticamente, utilizando una métrica, se tiene un número infinito de relaciones espaciales posibles entre cualesquiera dos entidades. En las relaciones espaciales, predominan principalmente las relaciones topológicas, las relaciones espaciales que se agrupan en tres diferentes categorías. [Egenhofer and Franzosa, 1991]

- Relaciones Topológicas, las cuales son invariantes bajo ciertas transformaciones de los objetos de referencia [Plümer 1996] [Egenhofer and Herring, 1990].
- Relaciones Métricas, se definen en términos de distancia y dirección [Peuquet and Ci-Xiang, 1987].

 Relaciones de orden, se refieren al orden parcial y total de objetos espaciales [Kainz 1990] como las descritas por preposiciones, tales como *enfrente de*, *atrás de*, *arriba* de, y debajo de [Freeman 1975] [Chang et al., 1989]

3.2.1 Relaciones Topológicas

Las relaciones topológicas son un subconjunto de relaciones geométricas, que se preservan bajo transformaciones topológicas como son translación, rotación y escalamiento, este tipo de relaciones refieren a propiedades tales como: conectividad, adyacencia e intersección.

Este tipo de relaciones capturan la esencia de la configuración espacial entre los objetos geográficos de un conjunto de datos geográficos, por ejemplo, métricas refinadas [Egenhofer et al., 1994b].

Existen dos modelos para describir las relaciones topológicas:

- Los modelos de intersección.
- Esquemas basados en el cálculo de conexión de regiones.

3.2.1.1 Modelos de Intersección

Los modelos de intersección, desarrollados por Egenhofer formalizan las relaciones topológicas binarias entre dos objetos arbitrarios (A y B), dado que dichas relaciones son de naturaleza cualitativa, por lo que no pueden ser expresadas por medidas convencionales [Egenhofer and Al-Taha, 1992]. Dichos modelos se encuentran basados en la teoría de conjuntos de puntos [Egenhofer et al., 1994b].

Inicialmente se tienen dos objetos A y B, tal que, podemos usar términos como overlap (sobrepone), Figura 3.3a o intersect (intersecta). Para describir las relaciones entre estos objetos, se muestra la Figura 3.3.



Figura 3.3. Overlap ó Intersect

De la Figura 3.3b, podemos observar la relación que existe entre las fronteras (bordes o perímetro del objeto) y los interiores de ambos objetos. En otras palabras, las fronteras coinciden en dos puntos, parte de la frontera de cada objeto está en el interior del otro, y ambos interiores coinciden parcialmente. La relación topológica permanece en tanto sus características sean preservadas.

Sin embargo, si únicamente cambia una de las partes en común, como ocurre cuando las fronteras sólo coinciden en una sola parte y que los interiores no tengan nada en común, como se muestra en la Figura 3.4. En ese caso hablamos de una relación totalmente diferente, a la cual podemos denominar *touch* (toque) o *meet* (coinciden).



Figura 3.4. Touch o Meet

Por otro lado, si las fronteras de los objetos no tienen ningún punto en común y los interiores no coinciden en ningún punto, se trata de otra relación topológica la cual podemos denominar *disjoint* (disjunto), como se muestra en la Figura 3.5.



Figura 3.5. Disjoint

3.2.1.1.1 Modelo 9-Intersección Dimensionalmente Extendido (DE-9IM)

El modelo 9-Intersección Dimensionalmente Extendido (DE-9IM) [Strobl et. al., 2008] está basado en el modelo 9-Intersección [Egenhofer and Franzosa, 1991], considera los componentes topológicos de los objetos geográficos definidos por el interior(I), límite(B) y exterior(E) de los objetos geográficos, los cuales se describen en la Tabla 3.1. El modelo consiste en analizar las intersecciones de éstos, por medio de nueve relaciones, donde los valores de estas relaciones son -1, 0, 1 ó 2, correspondiendo el valor de -1 a una no-intersección (conjunto vacío).

Las relaciones de dos objetos, se describen por medio de una matriz de 3x3, la cual tiene la siguiente forma:

$$DE - 9IM(A, B) = \begin{bmatrix} \dim \left(I(A) \cap I(B) \right) & \dim \left(I(A) \cap B(B) \right) & \dim \left(I(A) \cap E(B) \right) \\ \dim \left(B(A) \cap I(B) \right) & \dim \left(B(A) \cap B(B) \right) & \dim \left(B(A) \cap E(B) \right) \\ \dim \left(E(A) \cap I(B) \right) & \dim \left(E(A) \cap B(B) \right) & \dim \left(E(A) \cap E(B) \right) \end{bmatrix}$$

Tabla 3.1 Definición de componentes topológicos

Tipo geométrico	interior(I)	límite(B)	exterior(E)	
Punto	El punto	Conjunto vacio	Puntos que no están en	
			el interior o límite	
Línea	Puntos que se	Los extremos del objeto	Puntos que no están en	
	encuentran entre los		el interior o límite	
	extremos.			
Área	Puntos dentro del	El contorno del	Puntos que no están en	
	polígono que se forma	polígono	el interior o límite	

Las relaciones espaciales descritas por el modelo son "Igual", "Disjunto", "Intersección", "Conecta", "Cruza", "Contiene" y "Sobrepone", que pueden ser especificadas mediante funciones booleanas, en forma de predicados topológicos, mostrados en la Tabla 3.2, que son utilizadas para verificar las relaciones espaciales entre dos objetos. El modelo DE-9IM, indica relaciones entre puntos, líneas y polígonos.

Tabla 3.2. Matriz de patrones para relaciones topológicas en el modelo DE-9IM

Predicado Topológico	Significado	Matriz Patrón
Igual	Las geometrías de los objetos son iguales.	$\begin{bmatrix} T & * & F \\ * & * & F \\ F & F & * \end{bmatrix}$
Disjunto	Las geometrías de los objetos no tienen puntos en común.	[F F *] F F * * * *]
Intersección	Las geometrías de los objetos tienen al menos un punto en común.	$\begin{bmatrix} T & * & * \\ * & * & * \\ * & * & * \end{bmatrix} or \begin{bmatrix} * & T & * \\ * & * & * \\ * & * & * \end{bmatrix}$ $or \begin{bmatrix} * & * & * \\ T & * & * \\ * & * & * \end{bmatrix} or \begin{bmatrix} * & * & * \\ * & T & * \\ * & * & * \end{bmatrix}$
Conecta	Las geometrías de los objetos tienen al menos un punto en común, en sus límites, pero no en su interior.	$\begin{bmatrix} F & T & * \\ * & * & * \\ * & * & * \end{bmatrix} \text{ or } \begin{bmatrix} F & * & * \\ * & T & * \\ * & * & * \end{bmatrix}$ $\text{or } \begin{bmatrix} F & * & * \\ T & * & * \\ * & * & * \end{bmatrix}$

Cruza	Las geometrías de los objetos comparten algunos, pero no todos sus puntos, y la dimensión de la intersección es menor que al menos una de las geometrías de los objetos.	[T * T] [0 * *]
Sobrepone	Las geometrías de los objetos comparten algunos, pero no todos sus puntos, y la dimensión de la intersección tienen la misma dimensión que alguna de las geometrías.	
Dentro de	La geometría de un objeto A se sitúa en el interior de la geometría de un objeto B.	$\begin{bmatrix} T & * & F \\ * & * & F \\ * & * & * \end{bmatrix}$
Contiene	La geometría de un objeto B se sitúa en el interior de la geometría de un objeto A.	$\begin{bmatrix} T & * & * \\ * & * & * \\ F & F & * \end{bmatrix}$

La matriz patrón representa el conjunto de los valores aceptados (p) para un predicado topológico en este modelo, de la geometría de dos objetos, y consiste de un conjunto de 9 valores, cada uno en una posición específica de la matriz, estos posibles valores son T, F, *, 0, 1 y 2, y su significado, para cualquier posición de la matriz es el siguiente:

```
p = T \rightarrow \dim(x) \ \epsilon(0,1,2), t. \ e.x = \emptyset
p = F \rightarrow \dim(x) = -1, i. \ e.x = \emptyset
p = * \rightarrow \dim(x) \ \epsilon(-1,0,1,2)
p = 0 \rightarrow \dim(x) = 0
p - 1 \rightarrow \dim(x) - 1
p = 2 \rightarrow \dim(x) = 2
```

Utilizando conceptos matemáticos de topología y métrica, los cuales tienen un profundo impacto en un modelado espacial, debido a que mientras que la topología se mantiene invariante en transformaciones que no consideran medidas métricas, un espacio métrico, se basa en la noción de distancia, por lo que topología y métrica, establecen una proximidad y además un espacio métrico genera una topología.

Por lo que, utilizando un modelado DE-9IM se pueden generar funciones booleanas, en forma de predicados topológicos, que se convierten en modelos de relaciones espaciales cualitativas mediante una combinación con propiedades métricas, las cuales permitan generar representaciones gráficas a partir de representaciones verbales, dando lugar a una métrica cualitativa de información [Egenhofer and Dube, 2009]. Con lo que se gana descriptividad al caracterizar la relación topológica

Por ejemplo, se puede definir una métrica, denominada *IAS*, como se muestra en la Ecuación 3.3, la cual indica la porción del interior del objeto *A*, que se intersecta con el interior del objeto *B*, como se muestra en la Figura 3.6.

$$IAS = \frac{area(A^{\circ} \cap B^{\circ})}{area(A)}$$
 Ecuación 3.3

В

Figura 3.6. Representación Gráfica de IAS

Combinando, los valores de la métrica IAS, con las distintas relaciones topológicas, se genera un predicado topológico, para la relación topológica disjunto, asociando dicha relación con un valor de IAS = 0, ya que $A^{\circ} \cap B^{\circ} = \emptyset$, por lo que el objeto A y el objeto B, son disjuntos sí y sólo sí IAS = 0 [Egenhofer et. al., 2009].

3.3 Medidas de Similitud Semántica

Las medidas de similitud semántica tienen una larga tradición en las ciencias cognitivas y especialmente en la psicología [Gentner et al., 1994, Goldstone et al., 2005] y son fundamentales en los procesos de categorización e inferencia humana [Medin et al., 1993], investigar sobre el proceso en el cual el ser humano realiza estas medidas de similitud, permite aplicar estas investigaciones en el ámbito de la computación, en las teorías de similitud computacionales, dentro de trabajos para recuperar y organizar información [Janowicz et al., 2008].

Los humanos, usan la similitud para almacenar y recuperar información, comparar nuevas situaciones con base en situaciones pasadas, además de clasificar el conocimiento y formarse nuevos conceptos basados en similitud. Mientras que las computadoras procesan la información mediante decisiones binarias de equivalencia o no-equivalencia, por lo que el problema de similitud en la computación no es un problema trivial. Dentro de GIScience, la similitud juega un rol primordial en muchas aplicaciones como los sistemas de soporte en decisiones espaciales, minería de datos o reconocimiento de patrones.

Dos nociones de similitud son encontradas en las medidas de similitud semántica, coincidencias y diferencias o distancia semántica, lo cual se muestra en la Figura 3.7.

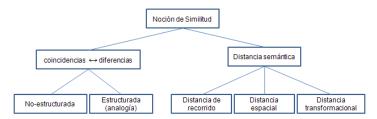


Figura 3.7. Diferentes nociones de similitud

En el caso de la noción de *coincidencias y diferencias*, entre dos representaciones conceptuales, toman un indicador de similitud, más coincidencias o menos diferencias, el más alto indica la similitud, utilizando medidas basadas en comparaciones estructuradas o no-estructuradas. Para aplicar la *distancia semántica* como una noción de la similitud semántica, todos los conceptos deben ser representados en una estructura común con alguna métrica específica, algunas medidas de similitud utilizan un espacio multidimensional como marco de referencia y una medida de distancia. La *distancia semántica* en un árbol o red estructurada que está definida por la longitud del camino más corto entre los nodos. La *distancia transformacional* es común en el uso de representaciones basadas en conjuntos de transformaciones, las cuales son compuestas y ejecutadas, unas después de otras, la distancia es medida por medio del número de transformaciones o su complejidad.

Algunas características de la similitud semántica son las siguientes:

- Propiedades de la similitud semántica, se pueden definir por tres axiomas [Schwering, 2008], estos axiomas han sido muy criticados, debido a que presuponen las métricas en las distancias conceptuales.
 - Minimalidad, este axioma indica que, si la distancia entre dos objetos es cero, por lo tanto, ambos conceptos son conceptualmente iguales, es decir, se trata del mismo concepto. Estos se puede ver en la Ecuación 3.4.

$$d(i,j) = 0 \implies i = j \land d(i,j) \ge 0$$
 Ecuación 3.4

Dónde t_i son conceptos y $d(t_i)$ la distancia conceptual entre ambos.

 Simetría, este axioma indica que la distancia entre dos conceptos es la misma partiendo de cualquiera de los dos conceptos y yendo al otro concepto. Lo que se muestra en la Ecuación 3.5.

$$d(i,j) = d(j,i)$$
 Ecuación 3.5

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

• Triángulo de desigualdad, este axioma indica que la distancia entre dos conceptos es menor o igual a la distancia entre ambos conceptos, vía un tercer concepto, lo que se muestra en la Ecuación 3.6.

$$d(i,j) + d(j,k) \ge d(i,k)$$
 Ecuación 3.6

Donde i, j, k son conceptos, d(i, j), d(j, k) y d(i, k) la distancia entre los conceptos.

- La similitud semántica depende del contexto. La comparación semántica entre dos objetos, no sólo depende de sus características, o a la clase que pertenezca, depende también de la motivación y aplicación del usuario [Medin et al. 1993][Janowicz, 2008].
- La similitud semántica depende de la representación. Las teorías de similitud están relacionadas directamente con un lenguaje de representación en particular, en una representación computacional, como puede ser una descripción lógica de los objetos, sin embargo, los objetos y clases a las que pertenecen, cambian con el tiempo, así como sus representaciones computacionales [Raubal 2008]. Por ejemplo,

podemos suponer que los ríos son objetos de agua lineales, y los lagos no lo son, sin embargo, en el caso de un desbordamiento del río, éste puede ser similar al lago [Kebler et al. 2007].

Similitud semántica, usabilidad y plausibilidad cognitiva. Comparando las medidas de clasificación computacionales con las medidas de similitud humanas, se define una estrategia para evaluar las medidas de similitud semántica, obteniendo clasificaciones para realizar comparaciones a nivel individual (objeto, a un nivel micro), clase (grupo de objetos, a un nivel meso) o escenas (a un nivel macro), usando una teoría computacional relacionada con clasificaciones de similitud humana [Smith et al., 2001] [Rodríguez et al., 2004].

Dentro del contexto de similitud semántica, se distinguen dos elementos principales dentro de una contextualización (*objetos* y *conceptos*). Un concepto "es una idea que caracteriza una idea en un conjunto o categoría de objetos" [Sloman et al., 1998], es decir, un concepto geoespacial describe la idea que caracteriza a una característica del tipo geográfico. Un objeto geoespacial refiere a una característica geográfica.

La semántica de objetos y conceptos geoespaciales tiene algunas características especiales, son descritos por propiedades tales como la forma, tamaño y localización, por otro lado, las relaciones, principalmente las relaciones topológicas, son muy importantes en la descripción semántica, ya que capturan la esencia de la información geográfica [Egenhofer, 1995b][Bruns and Egenhofer, 1996].

3.3.1 Similitud de Relaciones Topológicas

Las relaciones topológicas son atractivas para la evaluación de la similitud, cuando se generan pequeñas variaciones, *cambio gradual*, en un conjunto de información geográfica debido a un proceso de transformación, y estas variaciones modifican alguna relación topológica entre dos objetos en un conjunto de información geográfica. Por lo que, este nuevo conjunto puede ser similar o menos similar al conjunto de información al que se le aplicó el proceso de transformación.

El concepto de *cambio gradual* ha sido utilizado para crear el modelo de vecinos conceptuales de relaciones topológicas [Egenhofer and Al-Taha, 1992][Egenhofer and Mark, 1995a]. Este modelo facilita el ordenamiento de las relaciones topológicas, y permite determinar la similitud entre las relaciones. La Figura 3.8, muestra ocho relaciones topológicas para objetos simples del tipo área [Egenhofer and Franzosa, 1991]. Las relaciones conectadas por una línea en la figura representan vecinos conceptuales.

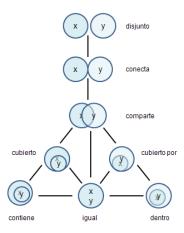


Figura 3.8. Vecinos conceptuales de relaciones topológicas.

La aplicación de un cambio gradual, espacialmente consistente, puede cambiar una relación en cualquiera de sus vecinos conceptuales. Utilizando este concepto, podemos concluir que "conecta es similar a comparte" o bien "conecta es más similar a comparte que a contiene" [Bruns and Egenhofer, 1996].

3.3.2 Similitud de Relaciones de Distancia

Las relaciones de distancia cualitativa, en general, son difíciles de definir para objetos espaciales, los conceptos y términos son altamente subjetivos y sensitivos a la escala de los datos espaciales que son considerados. La Figura 3.9 muestra un esquema de relaciones de distancia basado en el incremento de distancias.



Figura 3.9. Vecinos conceptuales de relaciones de distancia y los símbolos que representan.

Por lo que, la distancia entre dos objetos, puede variar entre *cero*, cuando dicha distancia no existe, es decir, es *cero*, y se puede incrementar dicha distancia, y ésta puede encontrase entre *muy cerca*, *cerca* y finalmente *lejos*. Así, tenemos que la relación de distancia entre dos objetos, los vecinos conceptuales se derivan por una relación de orden " < " (menor que) sobre la distancia de los objetos, por lo que, en el caso de la Figura 3.9, *muy cerca* es más similar a *cero* que *cerca* y *cerca* más similar que *lejos*, dado que *cero* < *muy cerca* < *cerca* < *lejos*.

3.3.3 Distancia semántica en analogía a la distancia espacial

La distancia semántica puede ser definida análogamente a la distancia espacial, es decir, la similitud puede ser medida como una función de la distancia espacial, como por ejemplo, la medida de distancia de Minkowski, la cual se muestra en la Ecuación 3.7.

$$d_{ij} = \left[\sum_{k=1}^{n} |x_{ik} - x_{jk}|^r \right]^{1/r}$$

Ecuación 3.7

Donde n es el número de dimensiones, x_{ik} es el valor de dimensión k para el estímulo i y x_{jk} es el valor de dimensión k para el estímulo j [Schwering, 2008].

3.4 Comentarios del capítulo

Las medidas de similitud semántica permiten clasificar valores numéricos, determinados por alguna fórmula, dentro de un rango de conceptos que pueden dar una idea conceptual, de algún estado de la información, sin embargo, al complementarlas con el uso de ontologías, nos permiten conocer más a detalle y en un lenguaje común, el estado del conjunto de información geográfica generalizada en el nivel que se requiera (*micro*, *meso* y *macro*), con respecto al conjunto de información fuente, ya que al conceptualizar dicha información, el nivel de comparación semántica (al mezclar ontologías y medidas de similitud semántica), se puede realizar un análisis multinivel, tomando en cuenta una gran cantidad de características, en el presente trabajo, estas características son las relaciones topológicas entre los distintos objetos geográficos, caracterizadas dentro de ciertas medidas de similitud semántica que han sido creadas para la evaluación de la calidad del conjunto de datos generalizado.

Capítulo 4. Metodología

En este capítulo se describe la metodología desarrollada para este trabajo, el cual se enfoca en la evaluación de la consistencia en datos geoespaciales generalizados. Como se ha mencionado, la metodología consiste en evaluar si la semántica de los datos generalizados ha sido preservada. En este caso, la semántica de los datos geoespaciales se representa en una ontología que por medio de conceptos se describen objetos geográficos y sus relaciones. A partir de esto, se generan *tuplas* denominadas *descripciones semánticas*, las cuales son comparadas por medio de una medición de similitud semántica de vecindad conceptual.

La metodología consiste de dos etapas principalmente, aunque se requieren algunos procedimientos adicionales para evaluar los conjuntos de datos. De forma general las etapas son:

- Síntesis. Tiene como propósito generar automáticamente la *representación* conceptual explícita del conjunto de datos fuente (CDF) y del generalizado (CDG), las cuales se denominan como descripciones semánticas. Dichas descripciones tienen la forma {O_i, **R**, O_j}, donde O_i y O_j son los identificadores de los objetos geográficos y **R** la relación espacial entre los objetos. Cada descripción semántica representa el significado de la relación topológica. Para identificar la relación que poseen los objetos geográficos se realiza análisis espacial, con lo que se obtiene un valor numérico que se mapea en la conceptualización del dominio.
- Verificación. En la verificación se evalúa la similitud semántica entre las descripciones semánticas de CDF y CDG con base en un esquema de vecindad conceptual. La idea es identificar si la relación espacial en los datos generalizados ha sido modificada, y evaluar si esto representa una inconsistencia. Para esto se ha considerado evaluar la consistencia en tres niveles de análisis micro, meso, y macro. El primero se refiere a l análisis uno a uno de las relaciones de un objeto, el segundo al conjunto de todas las relaciones que posee un objeto, mientras que el último se refiere a la totalidad de los objetos que existen en la representación. Cabe mencionar

que se definieron estos niveles de análisis para facilitar la interpretación de la identificación de inconsistencias.

Por otra parte, previamente a las etapas antes mencionadas, se requiere construir las bases de datos fuente y generalizadas, así como la construcción de la ontología que describe a cada una de las relaciones entre los objetos geográficos que constituyen el dominio. En las siguientes secciones se describirán de forma detallada cada una de las etapas que componen a la metodología.

Identificando si se ha preservado o no la semántica después de generalizar. Como resultado de la comparación, se pueden obtener tres conceptos: *Igual*, *Desigual* y *Equivalente*, donde *Igual* y *Equivalente* representan que la relación permanece consistente en los datos generalizados, mientras que *Desigual* representa una inconsistencia.

Este análisis see basa en una ontología que se usa para realizar descripciones del *CDF* y *CDG*. Como resultado de esta etapa se obtienen *tuplas*, denominadas descripciones semánticas, que representan a los objetos y a cada una de las relaciones topológicas del *CDF* mapeada a una relación topológica en el *CDG*; así como la generación de los esquemas de vecindad conceptual de las relaciones topológicas del CDF.

Frecuentemente, los modelos de datos geográficos representan en forma explícita, únicamente un conjunto de objetos básicos y sus propiedades (como su geometría y algunas relaciones topológicas). Sin embargo, en gran parte del mundo geográfico la semántica espacial aparece en las relaciones que poseen grupos de objetos geoespaciales [Worboys, 1996], por esta razón es que se ha considerado emplear una ontología.

Esto se realiza al comparar el *Conjunto de Datos Fuente* (*CDF*) con el *Conjunto de Datos Generalizado* (*CDG*), por medio de una medición de similitud semántica basada en la vecindad conceptual de los conceptos definidos en una ontología. Dicha ontología describe las relaciones espaciales entre los diferentes objetos que componen el dominio de interés.

Finalmente se muestra una lista donde se presentan a cada una de las *tuplas* indicando, si ésta es una (in)consistencia. El diagrama general de la metodología se muestra en la Figura 4.1.

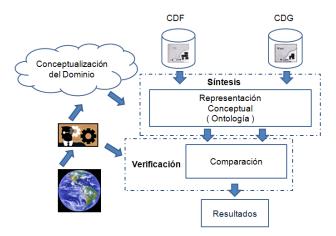


Fig. 4.1. Diagrama general de la metodología propuesta

4.1 Etapa de preprocesamiento

En esta sección se describen algunos componentes de la metodología que no forman parte del método propuesto, pero que son importantes para realizar el análisis aquí desarrollado.

4.1.1 Base de datos espacial generalizada

El método de generalización que se ha convenido usar es el definido en [Moreno, 2001], el cual se basa en el uso de los operadores de generalización propuestos en [McMaster y Shea, 1992] y se realiza de forma semi-automática. Este método está enfocado en la generalización de datos de áreas urbanas. El sistema se orienta totalmente hacia el enfoque tradicional de los GIS, donde las propiedades de los objetos son los atributos de la base de datos espacial y donde los criterios que definen la generalización son valores cuantitativos. La Figura 4.2, muestra un esquema general de generalización semi-automática vectorial para datos geográficos inmerso en un Sistema de Información Geográfica (GIS).

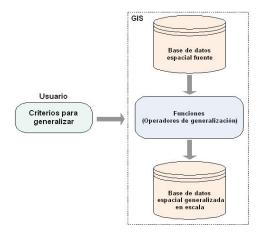


Figura. 4.2 Esquema general del sistema de generalización automática

Los componentes de este esquema son:

- Base de datos espacial fuente. Almacena los datos geográficos en formato vectorial, así como datos descriptivos.
- Funciones. Son un conjunto de procedimientos que involucran el uso de los
 operadores de generalización, que son parametrizados usando criterios
 previamente definidos. Los criterios especifican los intervalos a utilizar para la
 clasificación, selección, simplificación, etc., y son específicos para cada caso de
 estudio, por lo que se definen de acuerdo con el propósito de la generalización.
- Base de datos espacial generalizada en escala. En esta base de datos se almacenan los datos que se generalizaron automáticamente.
- *Criterios para generalizar*. Son los parámetros que se utilizan para los operadores de generalización y el usuario los define de acuerdo con el caso de estudio, escala, temática, etc.

Los datos generalizados y los criterios utilizados en este trabajo se presentan en el Capítulo 5 de Pruebas y Resultados.

4.1.2 Conceptualización del domino

La conceptualización del domino geográfico es muy importante para el desarrollo de esta tesis, ya que con base en ella se mide la similitud semántica y se identifica si se ha preservado la consistencia en los datos generalizados. En el contexto de este trabajo se

requiere de una conceptualización del entorno geográfico, por lo que describiremos algunas de las propiedades que poseen los objetos geográficos almacenados en bases de datos espaciales, con el fin de presentar los aspectos a considerar. Dicha conceptualización será la base para definir una ontología.

Primeramente, se analizan los objetos geográficos que componen nuestro caso de estudio, de acuerdo con el diccionario de datos del INEGI [INEGI, 1996][INEGI, 1993][INEGI, 1993b], para así conceptualizar a cada uno de los objetos geográficos, dentro de la ontología, indicando las clases a las que pertenecen y las relaciones topológicas, creando clases que permitan indicar dichas relaciones dentro de la ontología.

4.1.2.1 Propiedades de los datos geográficos

Las bases de datos espaciales almacenan datos descriptivos y geográficos. Los descriptivos están orientados a las características atributivas de los datos. Por otra parte, los datos geoespaciales o datos geográficos son instancias de los objetos geográficos que existen en la realidad. Históricamente, los datos geográficos se han dividido en dos tipos fundamentales: Vectoriales y Raster. Los *vectoriales*, se ocupan generalmente de los fenómenos discretos y se representan por un conjunto de primitivas geométricas (puntos, líneas, superficies o sólidos) agrupados en capas temáticas. Los *raster*, generalmente se ocupan de los fenómenos geográficos continuos sobre el espacio, almacenándose un conjunto de valores asociados con uno de los elementos en un arreglo regular de puntos o celdas. En el contexto de este trabajo, nos enfocaremos únicamente en las propiedades topológicas y temáticas.

• Propiedades topológicas, se refiere a las características de las figuras geométricas que permanecen invariantes si el espacio es deformado elástica y continuamente [OGC, 1999] (por ejemplo, al realizar un cambio de proyección cartográfica). En el contexto de la información geográfica, la topología es usada para describir la conectividad de un grafo n-dimensional, el que permanece invariante después de una transformación continua. La topología computacional proporciona información sobre la conectividad de primitivas geométricas que se derivan de la geometría subyacente. El INEGI propone únicamente las

- relaciones "comparte" y "conecta" para describir la topología en los mapas topográficos [INEGI, 1993]. En este caso se proponen las relaciones disjunto, comparte y conecta, las cuales se especificarán en la siguiente sección.
- Propiedades temáticas, se refiere a la categoría a la que pertenecen cada una de las capas temáticas, generalmente se encuentran muy relacionadas con la representación cartográfica. Para el caso de México, el INEGI establece los criterios de representación de cada objeto geográfico de acuerdo con la temática.

4.2.1.2 Caracterización de las relaciones topológicas

Las relaciones topológicas (*disjunto*, *comparte*, *cruza*) empleadas en la conceptualización son las que se especifican en la Tabla 4.1, en la cual se presenta el significado de cada una de ellas, así como el patrón de la matriz de intersección que la caracteriza de acuerdo con el modelo *DE-9IM* [Davis and Aquino, 2003].

Relación Significado **Ejemplo** Patrón para la matriz de intersección Disjunto Los objetos geográficos no tienen puntos en común. Las geometrías de los objetos **Comparte** geográficos tienen algunos puntos en común pero no todos. La dimensión de la intersección es igual a la de ambas geometrías. Cruza Las geometrías de los objetos geográficos tienen algunos puntos en común pero no todos. La dimensión de la intersección es menor que al menos una de las geometrías.

Tabla 4.1. Caracterización de las relaciones topológicas utilizadas.

Considerando que debemos garantizar que las relaciones topológicas sean caracterizadas de manera adecuada, se realizó una refinación métrica de cada una de ellas. Esto significa que al evaluar como predicado una relación topológica, no

únicamente se obtendrá Falso o Verdadero, sino diferentes casos de una relación. Por ejemplo, para la relación *comparte* se obtendrá "*compartir nada*", "compartir poco", "compartir mitad", "compartir casi completamente" y "compartir completamente". Los conceptos utilizados y sus restricciones se presentan en la sección Síntesis.

4.2.1.3 Objetos geográficos que componen el dominio

Los objetos geográficos que componen el dominio pertenecen a la carta topográfica del INEGI escala 1:50,000. Las clases de objetos geográficos en el conjunto de datos fuente y generalizado son las que aparecen en la Tabla 4.2. Adicionalmente se presenta el tipo de operación que se realizó para generalizar, dicha operación permite entender de forma conceptual la correcta jerarquización de los objetos, así como también la primitiva de representación utilizada en cada caso. Una descripción detallada de las clases de objetos aparece descrita en el Anexo 1 del presente documento.

Conjunto de datos fuente Operador de generalización Conjunto de datos generalizado Selección/eliminación. Unidades **Edificios (puntos)** Agrupación, habitacionales Tipificación (polígonos) Calles (líneas) Selección/eliminación Calles (líneas) Selección/eliminación, Manzana (polígonos) Manzana (polígonos) Agrupación Escuelas (puntos) Selección/eliminación, Escuelas (polígono) Exageración

Selección/eliminación,

Exageración

Tabla 4.2. Clases de objetos geográficos.

4.1.3 Construcción de la ontología

Hospitales (puntos)

Identificados los objetos geográficos y las relaciones topológicas del CDF y del CDG, se utilizó el software Protégé [Knublauch and Horridge, 2004] para construir la ontología por medio de la metodología GEONTO-MET [Torres, 2008]. A partir de las

Hospitales (polígonos)

clases y subclases de los distintos objetos geográficos del CDF y del CDG, como se muestra en la Figura 4.3.

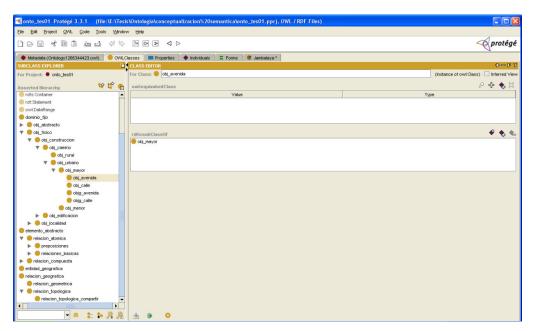


Figura 4.3. Construcción de la ontología de la conceptualización del CDF y CDG.

Dichas clases y subclases fueron determinadas con base en el diccionario de datos del INEGI [INEGI, 1996][INEGI, 1993][INEGI, 1993b] y la jerarquía de conceptos de WordNET [WordNET, 2010]. En este caso, se crearon dos clases principales que heredan de la clase principal *cosa*, denominadas *objeto físico* y *objeto abstracto*. En la primera, se incluyen los objetos geográficos de ambos conjuntos, es así que se crea la clase *construcción*, hija de la clase *objeto físico*, y padre de la clase *camino*, ésta a su vez, es padre de la clase *urbano* y ésta de la clase *mayor*. Por último, esta clase es padre de las clases *o_avenida*, *g_avenida*, *o_calle* y *g_calle*, que son las clases, a partir de las cuales se instanciarán los objetos geográficos *calles* y *avenidas*, del CDF (nombre de clase que empieza con o_) y CDG (nombre de la clase que empieza con g_) (Ver Figura 4.3).

Por otro lado, la clase *construcción* también es padre de la clase *edificación* y ésta a su vez, de las clases *o_educativo* y *g_educativo*, a partir de las cuales se instanciarán los objetos geográficos *escuelas* de ambos conjuntos de datos y de la clase *salud* que es padre de las clases *o_hospital* y *g_hospitales*, a partir de las cuales se instanciarán los objetos geográficos *hospitales* del CDF y CDG.

Los objetos geográficos edificios del CDF se instancian de la clase unidadhabitacional, hija de la clase área, que a su vez es hija de la clase región y ésta a su vez de la clase localidad y siendo esta última de la clase físico. Por último, la clase área es también padre de la clase g_unidadhabitacional, de la cual se instancian los objetos geográficos unidad habitacional del CDG.

Por otro lado, una vez identificadas las relaciones topológicas entre objetos, éstas son implementadas en la ontología, como se muestra en la Figura 4.4.

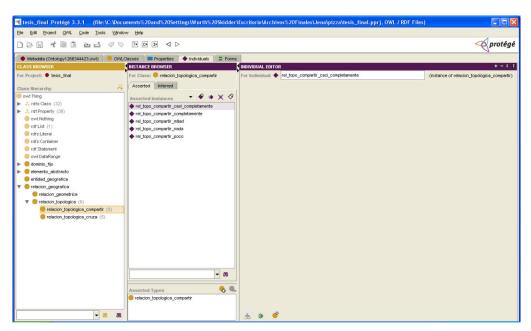


Figura 4.4. Implementación de las relaciones topológicas en la conceptualización del CDF y CDG.

Inicialmente se construye la clase *relación geográfica*, padre de la *clase relación topológica*, a partir de la cual se instancian las siguientes relaciones topológicas:

- Relación topológica comparte (RT comparte).
- Relación topológica disjunto (RT disjunto).
- Relación topológica cruza (RT cruza).

La clase *relación topológica* es a su vez padre de cada una de las clases conceptuales identificadas en el análisis para la conceptualización, es así, que las clases que son hijas de dicha relación son, relación *topo_comparte* y relación *topo_cruza* de las cuales se

instancian los conceptos indicados en los esquemas de vecindad conceptual que se utilizarán para evaluar la similitud entre las relaciones.

Como resultado de la construcción de la ontología se obtiene un archivo con extensión OWL, el cual contiene la conceptualización del dominio de los conjuntos de datos geográficos.

4.2 Síntesis

Esta etapa tiene como propósito generar la representación conceptual de los conjuntos de datos, esto se realiza instanciando las clases de la ontología previamente elaborada. Cada uno de los conceptos representados se obtienen al evaluar la relación topológica que existe entre ellos. Para relacionar los datos geográficos y las relaciones topológicas con las clases definidas en la ontología se realiza una función de mapeo, la cual relaciona las entidades geográficas con una clase.

4.2.1 Medición de las relaciones topológicas

Con el objeto de expresar mediante una descripción cualitativa cada una de las relaciones existentes en los conjuntos de datos geográficos, se realiza una medición. Esta medición tiene como propósito extraer la relación que aparece implícitamente representada en los datos, por su naturaleza esta medición arroja un valor numérico. Podemos ver esta medición como un mapeo.

El término de mapeo el que se define como: Sean A y B dos conjuntos arbitrarios. Se asigna para cada elemento $a \in A$ un elemento $m(a) \in B$ que se denomina mapa de A en B (A into B^I), simbólicamente es representado por $m: A \to B$. El elemento m(a) es llamado el valor de la función de m en a o la representación de a sobre b. Para cualquier subconjunto A' de A, m(A') es el conjunto $\{m(a): a \in A'\}$. El conjunto A se denomina dominio de m, el conjunto B se llama a con a. En general no se asume que todos los

-

¹ Se trata de una función inyectiva, cuando las elementos en el rango corresponden con elementos diferentes del rango.

valores de B que existen en $b \in B$ que no tienen una imagen² en ninguna a, es decir, m(a) es subconjunto de B. Si m(A)=B, el mapa se llama A sobre B (A onto $B^3)$ en vez de A en B.

Un mapeo se considera como una colección de pares (a, m(a)). Si A consiste de un número finito de elementos $a_1, a_2, a_3, \ldots, a_n$, el mapeo se puede representar como una tabla. En la mayor parte de las mediciones ocurre que cada elemento de B aparece más de una vez como imagen de A. En ese caso cada elemento $b \in m(A)$.

Considerando lo anterior, una medición se puede ver como el mapeo $m: A \to B$, donde A representa el conjunto de datos a ser medidos y B representa el conjunto de mediciones obtenidas, o dicho de otra forma, B representa el conjunto de A representa el conjunto de A

La topología por sí misma no permite distinguir entre diferentes casos de las relaciones topológicas. Por tal motivo, se requiere un procedimiento específico para extraer de la mejor forma posible la riqueza de la relación topológica. Siguiendo la premisa *topología importa, mientras que las métricas refinan* [Egenhofer y Mark, 1995] es necesario agregar métricas para extraer más detalles y evitar ambigüedades en la descripción de la relación. Por ello, definiremos *relaciones topológicas métricamente refinadas*, que engloban tanto aspectos topológicos como geométricos. Un factor crítico que refuerza lo anterior, es que los errores sobre las relaciones espaciales en la cognición humana son típicamente de métricas en vez de ser de naturaleza topológica [Tversky, 1981]

En este caso se desarrollaron las siguientes medidas para evaluar las relaciones topológicas:

- Medición de la RT disjunta para unidades habitacionales.
- Medición de la RT disjunta para hospitales y escuelas con respecto a las calles.

.

² Sean conjuntos X y Y, m el mapa $m: X \to Y$, y x sea algún miembro de X. Entonces la imagen de x sobre m, se define como m(x), y es el único miembro y de Y que asocia m con x. La imagen del mapeo m se denota por im(m) y es el rango de f, o de forma más precisa la imagen de su dominio.

³ Se trata de una función suprayectiva (también denominada sobreyectiva, epiyectiva, o suryectiva, o exhaustiva) en donde todo elemento del rango tiene una imagen de al menos un elemento del conjunto dominio.

- Medición de la RT disjunta para hospitales y escuelas con respecto a las manzanas.
- Medición de la RT cruza para hospitales y escuelas con respecto a las calles.

A continuación se detallan cada una de las mediciones desarrolladas.

4.2.1.1 Medición de la relación topológica para unidades habitacionales (RTD1)

Esta relación se denominará como RTD1 y se basa en calcular el porcentaje de área común entre dos unidades habitacionales (objetos área). RTD1 se especifica como sigue:

RTD1 =
$$(area(O_1^{\circ} \cap O_2^{\circ}) / area(O_1^{\circ}) + area(O_2^{\circ}))*100$$
 Ecuación 4.1

Donde:

- $area(O_1^o \cap O_2^o)$, es el área de intersección entre las dos unidades habitacionales;
- $area(O_1^o)$, representa el área interior del objeto O_1 ;
- $area(O_2^o)$ representa el área interior del objeto O_2 .

Como resultado de esta función podemos obtener rangos de valores que van desde 0 (cero) hasta 100. Estos valores pueden interpretarse como qué tan disjuntos son los objetos geográficos en cuestión. Si RTD1=0, significa que los objetos son totalmente disjuntos. Por otro lado, si RTD1=100 entonces significa que no están disjuntos sino que comparten su área interior totalmente.

La Tabla 4.3 describe las restricciones que contiene cada clase para esta relación. Esto representa la correspondencia que existe entre un valor numérico y un valor cualitativo.

Tabla 4.3. Criterios para RTD1.

RTD1 (%)	Concepto en la conceptualización	Representación Gráfica
90 < RTD1 ≤ 100	compartir_completamente	•

50 < RTD1 ≤ 90	compartir_casi_completamente	(Sy)
5 < <i>RTD</i> 1 ≤ 50	compartir_mitad	
0 < RTD1 < 5	compartir_poco	ху
0	compartir_nada	ху

4.2.1.2 Medición de la relación topológica para escuelas u hospitales con respecto a las calles (RTD2)

Esta relación se denominará como RTD2 y se basa en calcular el porcentaje de área de un objeto geográfico escuela u hospital (objetos área), O_1 , que se encuentra en un objeto geográfico calle (objeto área), O_2 . RTD2 se especifica como sigue:

RTD2 =
$$(area(O'_{1^0}) / area(O_{1^0}))*100$$
 Ecuación 4.2

Donde:

- $area(O'_{I}^{o})$, es el área interior del objeto O_{I} , escuela u hospital, que se encuentra en el objeto geográfico calle;
- $area(O_1^o)$ representa el área interior del objeto O_1 .

Como resultado de esta función podemos obtener rangos de valores que van desde 0 (cero) hasta 100. Estos pueden interpretarse como qué tan disjuntos son los objetos geográficos en cuestión. Si RTD2=0, significa que los objetos son totalmente disjuntos. Por otro lado, si RTD2=100 entonces significa que no están disjuntos sino que parte del área interior del objeto geográfico O_I se encuentra sobre el objeto geográfico calle.

La Tabla 4.4 describe las restricciones que contiene cada clase para esta relación. Esto representa la correspondencia que existe entre un valor numérico y un valor cualitativo.

RTD2 (%)
Concepto en la conceptualización
Representación Gráfica $55 < RTD2 \le 100$ compartir_completamente $20 < RTD2 \le 55$ compartir_casi_completamente $5 < RTD2 \le 20$ compartir_mitad $0 < RTD2 \le 5$ compartir_poco $0 < RTD2 \le 5$ compartir_nada $0 < RTD2 \le 5$ compartir_nada $0 < RTD2 \le 5$ compartir_nada

Tabla 4.4. Criterios para RTD2.

4.2.1.3 Medición de la relación topológica para hospitales y escuelas con respecto a las manzanas (RTD3)

Esta relación se denominará como RTD3 y se basa en calcular el porcentaje de área común entre un objeto geográfico escuela u hospital (objetos área), O_1 , y un objeto gráfico manzana (objeto área), O_2 . RTD3 se especifica como sigue:

RTD3 =
$$(area(0'_1) / area(0_1))*100$$
 Ecuación 4.3

Donde:

- $area(O'_{l}^{o})$, es el área interior del objeto O_{l} (escuela u hospital) que se intersecta con el área interior de un objeto geográfico manzana;
- $area(O_1^o)$ representa el área interior del objeto O_1 .

Como resultado de esta función podemos obtener rangos de valores que van desde 0 (cero) hasta 100. Estos pueden interpretarse como qué tan disjuntos son los objetos geográficos en cuestión. Si RTD3=0, significa que los objetos son totalmente disjuntos.

Por otro lado, si RTD3=100 entonces significa que no están disjuntos sino que el área interior del objeto geográfico O_1 se intersecta dentro del objeto geográfico O_2 .

La Tabla 4.5 describe las restricciones que contiene cada clase para esta relación. Esto representa la correspondencia que existe entre un valor numérico y un valor cualitativo.

 RTD3 (%)
 Concepto en la conceptualización
 Representación Gráfica

 $55 < RTD3 \le 100$ compartir_completamente

 $20 < RTD3 \le 55$ compartir_casi_completamente

 $5 < RTD3 \le 20$ compartir_mitad

 $0 < RTD3 \le 5$ compartir_poco

 $0 < RTD3 \le 5$ compartir_nada

 $0 < RTD3 \le 5$ compartir_nada

Tabla 4.5. Criterios para RTD3.

4.2.1.4 Medición de la relación topológica para hospitales y escuelas con respecto a las calles (RTC)

Esta relación se denominará como RTC y se basa en calcular la distancia entre un objeto geográfico escuela u hospital (objetos área) O_I , y el extremo opuesto de un objeto geográfico calle (objeto lineal), O_2 . RTC se especifica como sigue:

$$RTC = L_c - L_0$$
 Ecuación 4.4

Donde:

• L_c , es el ancho del objeto geográfico O_2 .

• L_o , es la longitud ocupada por el objeto geográfico O_I , en el objeto geográfico O_2 .

Como resultado de esta función podemos obtener rangos de valores que van desde 0 (cero) hasta 100. Estos pueden interpretarse que tanto cruza el objeto geográfico O_I al objeto geográfico O_2 . Si RTC= L_c , significa que el objeto geográfico O_I no cruza nada al objeto geográfico O_2 . Por otro lado, si RTD3=0 entonces significa que el objeto geográfico O_I cruza totalmente al objeto geográfico O_2 .

La Tabla 4.6 describe las restricciones que contiene cada clase para esta relación. Esto representa la correspondencia que existe entre un valor numérico y un valor cualitativo.

RTC (m)
Concepto en la conceptualización
Representación Gráfica

RTC > 8
cruza_nada $4 < RTC \le 8$ cruza_poco $2 < RTC \le 4$ cruza_mitad $1 < RTC \le 2$ cruza_casi_completamente $0 < RTC \le 1$ cruza_completamente

Tabla 4.6. Criterios para RTC.

4.2.2 Construcción de la representación conceptual

La representación conceptual de los datos geográficos tiene como propósito describir la semántica de la representación. En [Carnap, 1956] se explica cómo las clases se definen por las propiedades que son características de sus miembros, es decir, los objetos de una clase tienen propiedades que los distinguen de los objetos que no pertenecen a la clase. Esto implica que para cada uno de los descriptores cuantitativos siempre debe decidirse

si es miembro o no de una determinada clase. Se debe formular un criterio para cada clase con el fin de determinar a qué clase pertenece cada DC. Los criterios representan la *restricción* de pertenencia de un DC_i a una clase C_j ,. Así la *función de membresía de la clase* se define como en la Ecuación 4.5:

$$M[DC_i, C_j] = 1$$
 si DC_i si es miembro de C_j ; de otra forma $M[DC_i, C_j] = 0$ Ecuación 4.5

Para definir la clase a la que pertenece cada *DC* (*descriptor cuantitativo*) se utiliza la función *CLASE(DC)* que se define como en la Ecuación 4.6:

$$CLASE(DC) = \{C/M[DC, C]=1\}$$
 Ecuación 4.6

La extensión de la clase es el conjunto de todos los objetos que pertenecen a ella, como se muestra en la Ecuación 4.7:

$$EXT(C_i) = \{DC_i | M[DC_i, C_j] = 1\}$$
 Ecuación 4.7

La forma correcta de expresar la relación entre el DC y la clase es la sentencia: *el descriptor cualitativo* DC_i *es un elemento de la clase j*. Es posible que la clase que se defina no tenga elementos; en este caso la extensión es vacía $EXT(C) = \emptyset$. Considerar el caso de un estudio demográfico de *localidades* con clases compuestas por:

Poblados: restricción 100 < número de habitantes < 1,000

Cabecera Municipal restricción 1,000 < número de habitantes < 50,000

Ciudades restricción 50,000 < número de habitantes < 10,000,000

Mega ciudades restricción 10,000,000 < número de habitantes

Existen muchos países que tendrán mega ciudades, pero para la gran mayoría de los países esta clase no tendrá miembros, por lo que su extensión es vacía. En caso del ejemplo anterior, el número de habitantes representa los valores que poseen los descriptores cuantitativos.

Entonces, utilizando la función de membresia se puede identificar la clase a la que pertenece cada descriptor cuantitativo y de esa forma asignar ese descriptor a una clase definida en la conceptualización.

De tal forma que las descripciones semánticas deben tener la forma $\{O_i, R, O_j\}$ donde O_i , y O_j son los identificadores de los objetos geográficos y R la relación espacial entre los objetos. Esto pretende representar

En este caso las relaciones topológicas métricamente refinadas que se conceptualizan, se ajustan a los intervalos establecidos en la Tabla 4.7, en donde se especifican las particularidades de cada una.

Tabla 4.7. Rangos para definir cada uno de los conceptos de relaciones topológicas

Relación RTD1	Relación RTD2	Relación RTD3		Relación RTC	
Rango	Rango	Rango	Concepto	Rango	Concepto
90 < RTD1 ≤ 100	55 < <i>RTD</i> 2 ≤ 100	55 < <i>RT D</i> 3 ≤ 100	compartir_comple tamente	<i>RTC</i> > 8	cruza_nada
50 < RTD1 ≤ 90	20 < RT D2 ≤ 55	20 < RT D3 ≤ 55	compartir_casi_co	4 < RTC ≤ 8	cruza_poco
5 < RTD1 ≤ 50	5 < RTD2 ≤ 20	5 < RTD3 ≤ 20	compartir_mitad	2 < RTC ≤ 4	cruza_mitad
0 < RTD1 ≤ 5	0 < RT D2 ≤ 5	0 < <i>RT D3</i> ≤ 5	compartir_poco	1 < RTC ≤ 2	cruza_casi_com pletamente
0	0	0	compartir_nada	0 < RTC ≤ 1	cruza_completa mente

A partir de los conceptos definidos se realiza la verificación de la consistencia, la cual se detalla en la siguiente sección.

4.3. Verificación

Esta etapa tiene como propósito verificar la consistencia de los datos generalizados utilizando una medida de similitud semántica basada en un esquema de vecindad conceptual, tomando como base el conjunto de datos fuente. Con el fin de identificar si la relación espacial en los datos generalizados no *rompe* la semántica de la representación, e identifica si esto representa una inconsistencia.

Para mejorar la expresividad del análisis se ha considerado evaluar la consistencia en tres niveles *micro*, *meso*, y *macro*. El primero se refiere a una a una las relaciones de un objeto, el segundo al conjunto de todas las relaciones que posee un objeto, mientras que el último se refiere a la totalidad de los objetos que existen en la representación. Se definieron estos niveles de análisis para facilitar la interpretación de la identificación de inconsistencias.

4.3.1 Esquemas de vecindad conceptual para relaciones topológicas

Se crearon dos esquemas de vecindad conceptual [Bruns and Egenhofer, 1996][Fonseca and Li, 2006], tomando como referencia, principalmente la relación topológica disjunto (disjoint). Los esquemas de vecindad conceptual desarrollados son:

- Esquema de vecindad conceptual *disjoint comparte*.
- Esquema de vecindad conceptual *disjoint cruza*.

4.3.1.1 Esquema de vecindad conceptual disjunto-comparte

En la Figura 4.5 se muestra el esquema de vecindad conceptual *disjunto* – *comparte*, dicho esquema, fue diseñado tomando en cuenta lo siguiente. Considerando dos objetos geográficos con representación punto, línea o polígono, los cuales se encuentran totalmente separados, es decir, sus áreas ó limites no se encuentran intersectada(o)s, por lo que no comparten un espacio geográfico, por lo que la relación topológica entre ambos es *disjoint*. Es así que el esquema de vecindad conceptual tiene como fin describir el conjunto de relaciones que se encuentran entre *disjunto* y *comparte* es

representada por cinco casos (compartir_nada, compartir_poco, compartir_mitad, compartir_casi_completamente y compartir_completamente). Adicionalmente, en la Figura 4.6 aparece el peso que tiene el pasar de una relación a otra. Por conveniencia, se establecieron estos valores de forma arbitraria obteniendo con estos valores los mejores resultados, de tal manera que pasar de un extremo a otro del esquema tiene un valor de 1.

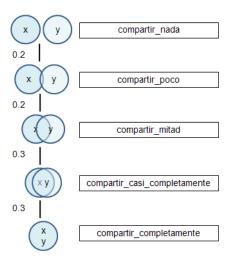


Fig. 4.5. Esquema de vecindad conceptual disjoint-comparte.

4.3.1.2 Esquema de vecindad conceptual disjunto- cruza

En la Figura 4.6 se muestra el esquema de vecindad conceptual *disjunto – cruza*, dicho esquema, fue diseñado tomando en cuenta lo siguiente.

Considerando dos objetos geográficos, los cuales se encuentran totalmente separados, es decir, el área del objeto con representación polígono no cruza al objeto con representación área, por lo que la relación topológica entre ambos es *disjunto*. Al momento de aplicar un proceso de generalización en el conjunto de datos geográficos al cual pertenecen, éstos pueden sufrir una transformación en sus áreas y pueden aumentar, por ejemplo, por un cambio de escala, por lo que ahora el área del objeto geográfico con representación polígono puede cruzar parcial o totalmente al objeto geográfico con representación área.

Es así que el esquema de vecindad conceptual tiene como fin describir el conjunto de relaciones que se encuentran entre *disjunto* y *cruza*, se caracteriza por cinco casos (*cruzar_nada*, *cruzar_poco*, *cruza_mitad*, *cruza_casi_completamente* y *cruza_completamente*). De forma similar al esquema anterior, en la Figura 4.7 aparece el peso que tiene el pasar de una relación a otra, los cuales se establecieron de forma arbitraria y pasar de un extremo a otro del esquema tiene un valor de 1.

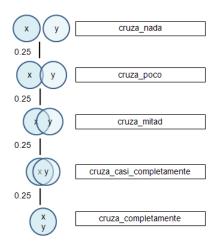


Fig. 4.6. Esquema de vecindad conceptual disjunto-cruza

4.3.2 Verificación del CDG

La verificación del CDG consiste en identificar con base en las tuplas obtenidas en la ontología, si se han preservado las relaciones topológicas en el *CDG* con respecto al *CDF* y generar así un valor de similitud entre el *CDG* y el *CDF*, el cual se define a partir de la verificación del *CDG* en distintos niveles, a nivel de RT's (*micro*), posteriormente a nivel de objeto geográfico (*meso*), y finalmente a un nivel de conjunto geográfico, es decir, que incluya todos los objetos geográficos contenidos en el conjunto de información geográfica (*macro*) [Ruas, 2000][Bard, 2003].

4.3.2.1 Análisis a nivel micro

A partir del mapeo definido entre las descripciones semánticas de *CDF* y *CDG*, se evalúa la similitud entre cada una de las RT's de un par de OG's en el CDF con respecto al par correspondiente de OG's en el CDG, es decir, se realiza una evaluación a un nivel

micro. Por lo que, la similitud se establece con base en los esquemas de vecindad conceptual que aparecen en las Figuras 4.5 y 4.6. Para lo cual se utilizan los pesos previamente definidos, los cuales se utilizan para definir la distancia conceptual entre dos conceptos de una manera cuantitativa. Esto con la finalidad de evaluar la similitud semántica entre pares de relaciones [Fonseca and Li, 2006] [Shwering, 2008].

Para evaluar la consistencia de los datos geográficos se utilizan las descripciones semánticas, las cuales tienen la forma (O_1, R_F, O_2) para los datos fuente, mientras que para datos generalizados (O_1, R_G, O_2) , de manera que O_1 y O_2 son objetos geográficos; R_F y R_G . Entonces la similitud semántica, entre ambas representaciones, se establece por la "distancia conceptual" entre los conceptos, utilizando la Ecuación 4.8.

$$Sim(R_F, R_G) = dif(R_F, R_G)$$
 Ecuación 4.8

Donde R_F es una relación topológica entre dos objetos geográficos que aparecen en el conjunto de datos fuente y R_G es una relación topológica en el conjunto de datos generalizado y dif (R_F , R_G), es la diferencia entre las relaciones siguiendo un esquema de vecindad conceptual. Por ejemplo, podemos suponer que R_F = $compartir_nada$ y R_G $compartir_mitad$, entonces $Sim(R_F, R_G) = 0.4$ (Ver Figura 4.5). De esta forma podemos decir que esta medición es simétrica.

Entonces la consistencia de los datos se analiza, al evaluar las relaciones topológicas entre pares de objetos geográficos del CDF y del CDG, para ello, se definen tres casos:

- *Igual*, como resultado de la similitud se obtiene $Sim(R_F, R_G)=0$, los conceptos son iguales. Se interpreta que la relación es consistente, entre ambas representaciones.
- *Desigual*, se define un umbral u, y hay relaciones que pudieran considerarse (in)consistentes, de acuerdo con lo siguiente. Si $u < Sim(R_F, R_G) < 1$, consideramos que la relación es d*esigual*, y representa una inconsistencia.

• *Equivalente*, Por otro lado, si $0 < Sim(R_F, R_G) \le u < 1$, consideramos que la relación es *equivalente*, y en cierto sentido, podemos decir que la semántica es preservada, y por lo tanto, representa que los datos son consistentes.

Estos conceptos reflejan de forma cualitativa el *grado de consistencia (CRT)* de los datos generalizados, y se utilizan para los siguientes niveles de análisis.

4.3.2.2 Análisis a nivel meso

A partir del análisis a nivel *micro*, se toman cada uno de los valores de *CRT*, de cada uno de los objetos geográficos en el *CDG* que son de interés, para así realizar un análisis semántico a nivel *meso*. Para esto se considera el *CRT* de cada una de las relaciones topológicas de cada objeto geográfico relacionado, y asignando a un valor cuantitativo, de acuerdo con la Tabla 4.8. Como resultado se obtiene el grado de *consistencia a nivel de objeto* (*GCO*).

Tabla 4.8. Grados de consistencia a nivel *micro*

CRT (nivel micro)	Valor numérico		
Igual	1		
Equivalente	0		
Desigual	-1		

Por lo que, el *GCO* está dado por la Ecuación 4.9, la cual indica la consistencia entre un objeto geográfico en el *CDG* con respecto a su correspondiente objeto geográfico en el *CDF*.

$$GCO\left(OGF_k,OGG_k\right) = \sum_{i=0}^{n} CRT_i - N$$
 Ecuación 4.9

Donde OGF_k indica el k-ésimo objeto geográfico en el CDF, OGG_k indica el k-ésimo objeto geográfico en el CDG, CRT_i indica el valor numérico del grado de consistencia de la i-ésima relación topológica en el CDG, i indica la i-ésima relación topológica en el CDG. Por último N y n indican el número total de relaciones topológicas en él. Por ejemplo, suponer un objeto geográfico en el CDG, que tiene i0 relaciones topológicas, entonces se detectaron que las i1 central i2 de éstas son iguales,

entonces $GCO(OGF_k, OGG_k) = (1+1+1)-3=0$, o bien, supongamos que para el mismo objeto se detectaron una consistencia Igual, una Equivalente y una Desigual, entonces $GCO(OGF_k, OGG_k) = (1+0-1)-3=-3$.

La consistencia *GCO* entre dos objetos geográficos, uno de ellos en el *CDF* y el otro en el *CDG*, se establece de la siguiente forma:

- Igual, si GCO (OGF_k, OGG_k) = 0, se considera que los objetos geográficos tienen las mismas relaciones en ambos conjuntos de datos, es decir, las relaciones topológicas del objeto geográfico del CDF no sufrieron ninguna modificación al momento de ser generalizado el conjunto de información geográfica.
- Equivalente, se define un umbral w, por lo que, los objetos geográficos en el CDG que pudieran considerarse consistentes, de acuerdo con lo siguiente: si w ≤ GCO (OGF_k, OGG_k) < 0, se interpreta como Equivalente. En cierto sentido podemos decir que la semántica a nivel objeto es preservada, y por lo tanto, representa que el objeto geográfico tiene relaciones consistentes.</p>
- Desigual. Si GCO (OGF_k, OGG_k) < w, consideramos que el objeto geográfico en el CDF es Desigual, y representa alguna inconsistencia. Entonces entre más negativo sea GCO significa que es más inconsistente.

4.3.2.3 Análisis a nivel macro

A partir del análisis hecho para cada uno de los objetos geográficos en el CDG, a un nivel *meso*, se toman los valores de GCO de cada uno de los objetos geográficos en el CDG, para poder así, realizar un análisis semántico del conjunto de datos (a un nivel *macro*). Considerando el valor cualitativo del grado de consistencia de cada una de los objetos geográficos en el CDG, y relacionando estos con un valor cuantitativo de acuerdo con la Tabla 4.9.

Tabla 4.9. Grados de consistencia a nivel meso

COG	Valor numérico
Igual	1
Equivalente	0
Desigual	-1

A este nivel se define el *grado de consistencia a nivel capa* (*GCC*), el cual se define por la Ecuación 4.10, la cual indica la consistencia entre el *CDG* con respecto al *CDF*.

$$GCC\left(CDF,CDG\right) = \sum_{i=0}^{t} GCO\left(OGF_{i},OGG_{i}\right) - T$$
 Ecuación 4.10

Donde OGF_i indica el i-ésimo objeto geográfico en el CDF, OGG_i indica el i-ésimo objeto geográfico en el CDG, $GCO\left(OGF_i, OGG_i\right)$ indica el valor numérico de la consistencia entre dos objetos geográficos uno de ellos en el CDF y el otro en el CDG. Por último T y t indican el número total de objetos geográficos en el CDG.

Por ejemplo, suponer que para un *CDG* que tiene 5 objetos geográficos, se detectaron que las consistencias a nivel objeto, *GCO* son para tres objetos geográficos *igual*, y para los dos restantes *Desigual*, entonces:

$$GCC(CDF, CDG) = (1+1+1-1)-5=-4$$

La consistencia a nivel conjunto de datos, entre el *CDF* y el *CDG*, se establece de la siguiente forma:

- Igual, si GCC(CDF, CDG) = 0, se considera que los conjuntos son iguales y por lo tanto se interpreta que el CDG es consistente con respecto al CDF.
- Desigual, se define un umbral z, por lo que, el CDG pudiera considerarse (in)consistente, de acuerdo con lo siguiente: si $GCC\left(CDF,CDG\right) < z$, consideramos que el CDG es Desigual, y representa una inconsistencia. Entonces entre más negativo sea GCC significa que es más inconsistente.

Equivalente, si z ≤ GCC (CDF, CDG) < 0, consideramos que el CDG es
Equivalente a CDF, y en cierto sentido podemos decir que la semántica a nivel
macro es preservada, y por lo tanto, representa que el CDG es consistente
respecto al CDF.

Considerando la Tabla 4.9, se realiza el cálculo del GCC, es decir, se realiza el análisis a nivel macro para los edificios en el CDF, con su correspondiente unidad habitacional en el CDG, utilizando la Ecuación 4.10 y considerando $COG(Equivalente) = 0 \ y \ COG(Desigual) = -1$, de acuerdo con la Tabla 4.8 tenemos que con seis objetos geográficos (E01, E02, E03, E04, E05, E06) se obtiene:

$$GCC(CDF, CDG) = 0 + 0 + 0 + 0 - 1 - 1 - 6 = -8$$

Considerando un valor de w = -6 y con base en el grado de similitud a nivel de conjunto de datos, se verifica la consistencia CCG del CDG con respecto al CDF, considerando únicamente los objetos geográficos, es Desigual, lo que indica que ambos conjuntos de datos no son consistentes.

Capítulo 5. Pruebas y Resultados

En esta sección se presentan los resultados experimentales obtenidos al realizar la implementación de la metodología propuesta. Para esto se implementó un software en *Java*, que utiliza la extensión *Jena* [Jena, 2009] para el manejo de la ontología. Como datos de prueba se utiliza un fragmento de la delegación Gustavo A. Madero del Distrito Federal, estos datos se muestran en la Figura 5.1.

5.1 Datos de prueba

Inicialmente, se analiza el conjunto de datos fuente (CDF), identificando principalmente los objetos geográficos que los componen y las relaciones topológicas entre ellos. Además, se presenta el conjunto de datos generalizado, indicando el proceso de generalización e identificando los objetos geográficos que lo componen, para que así se construya con base en el CDF y el CDG, una ontología que conceptualize el dominio en el que se aplicará la metodología desarrollada en la presente tesis.

5.1.1 Conjunto de datos fuente (CDF)

Se analizan los objetos geográficos que componen nuestro caso de estudio, de acuerdo con el diccionario de datos del INEGI [INEGI, 1996][INEGI, 1993][INEGI, 1993b], para así conceptualizar a cada uno de los objetos geográficos dentro de la ontología, indicando las clases a las que pertenecen y las relaciones topológicas, creando clases que permitan indicar dichas relaciones dentro de la ontología.

El CDF se muestra en la Figura 5.1, dónde se identifican los siguientes objetos geográficos y relaciones.

• Los objetos geográficos que componen las diferentes capas del CDF son: *manzanas*, *escuelas*, *vialidades* (*calles* y *avenidas*), *hospitales* y *edificios de departamentos*.

• Las relaciones topológicas que se utilizan son: *conecta*, *cruza*, *comparte* y *disjunta*. Estas relaciones topológicas relacionan a los distintos objetos geográficos del conjunto. En la Tabla 5.1 se muestran en forma general, las relaciones topológicas entre las distintas clases de objetos geográficos.

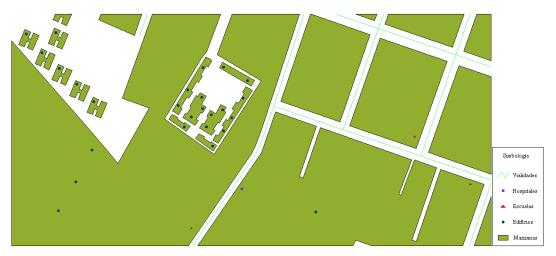


Figura 5.1. Conjunto de Datos Fuente (CDF)

Objeto Escuela Calle Iglesia Hospital Manzana Edificio Geográfico Escuela -----Disjunta Disjunta Disjunta Disjunta Comparte Disjunta Calle Conecta Conecta Disjunta Conecta Conecta Cruza Disjunta Avenida Conecta Disjunta Disjunta Disjunta Disjunta Disjunta -----Iglesia Disjunta Disjunta Disjunta Disjunta Disjunta Disjunta Hospital Disjunta Disjunta Disjunta Disjunta -----Comparte Disjunta Manzana Disjunta Disjunta Disjunta Disjunta Disjunta Disjunta Edificio Disjunta Disjunta Disjunta Disjunta Disjunta Comparte -----

Tabla 5.1. Relaciones Topológicas en el Conjunto de Datos Fuente (CDF)

En el Anexo A, se muestra el diccionario de datos de los objetos geográficos que componen el conjunto de datos fuente.

5.1.2 Conjunto de Datos Generalizado (CDG)

Siguiendo los procedimientos de generalización definidos, se generó el conjunto de datos generalizados. Algunas de las reglas que se utilizaron para generalizar los datos son:

- En la capa temática *edificios*, se agrupan los edificios por el atributo *dirección*, de la tabla de atributos.
- En la capa temática *manzanas*, se agrupan en una sola manzana aquellas manzanas que se encuentren rodeadas por las mismas calles y/o avenidas.
- En las capas temáticas de *escuelas* y *hospitales*, se realiza una transformación punto a polígono de los objetos geográficos, con el fin de exagerar su representación.

Debido a la aplicación de dichas reglas, el proceso de generalización genera algunos cambios en el CDG con respecto al CDF, los cuales se indican en la Tabla 5.2.

Objeto Geográfico Comentarios **CDG CDF** Manzana Manzanas Se utilizan los operadores de generalización agregación y tipificación. Las manzanas del CDF se agrupan en una sola manzana en el CDG. Se preserva el tipo de objeto geográfico como polígono. Unidades Habitacionales **Edificios** Se utilizan los operadores de generalización agregación, tipificación, exageración y realce. Los edificios del CDF se agrupan en objetos de unidad habitacional. Se origina transformación de objetos tipo punto, a objetos tipo área. Algunas relaciones espaciales del CDF se preservan, pero otras cambian. Escuelas y Hospitales Escuelas y Hospitales Se utilizan los operadores de generalización exageración y realce, sobre escuelas y hospitales del CDF. Esto genera nuevas relaciones topológicas como conectar y

Tabla 5.2. Cambios en el CDG con respecto al CDF.

compartir con algunas manzanas.

Como resultado del proceso de generalización se obtiene el CDG que aparece en la Figura 5.2, en donde se pueden identificar:

- Los objetos geográficos que componen al conjunto son: manzanas, escuelas, vialidades, hospitales y edificios de departamentos.
- Las relaciones topológicas que se caracterizan son: conecta, cruza, comparte y
 disjunta. Estas relaciones topológicas relacionan a los distintos objetos geográficos
 del conjunto, en la Tabla 5.3 se muestran en forma general, las relaciones
 topológicas entre los distintos objetos geográficos del conjunto de datos.

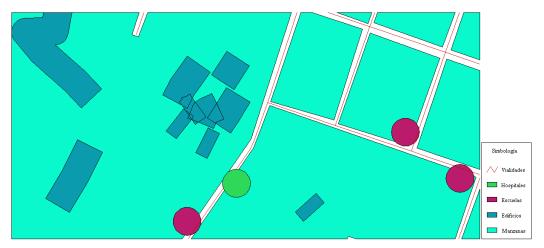


Figura 5.2. Conjunto de Datos Generalizado (CDG)

Tabla 5.3. Relaciones Topológicas Semánticas en el Conjunto de Datos Fuente (CDF)

Objeto Geográfico	Escuela	Calle	Avenida	Hospital	Manzana	Unidad Habitacional
Escuela		Comparte Cruza	Disjunta	Disjunta	Comparte Conecta	Disjunta
Calle	Conecta	Conecta	Cruza	Conecta	Conecta	Disjunta
Avenida	Disjunta	Disjunta		Disjunta	Conecta	Disjunta
Hospital	Disjunta	Comparte Cruza	Disjunta		Comparte Conecta	Disjunta
Manzana	Disjunta	Disjunta	Disjunta	Disjunta		Disjunta
Unidad Habitacional	Disjunta	Disjunta	Disjunta	Disjunta	Comparte	Comparte

5.1.3 Implementación de la ontología

Una vez que se tuvieron identificados los objetos geográficos y las relaciones topológicas del CDF, junto con los mismos elementos en el CDG, haciendo uso del software denominado Protégé [Knublauch and Horridge, 2004], se construye la ontología, iniciando con la creación de clases y subclases a las que pertenecen los distintos objetos geográficos del CDF y del CDG, como se muestra en la Figura 5.3.

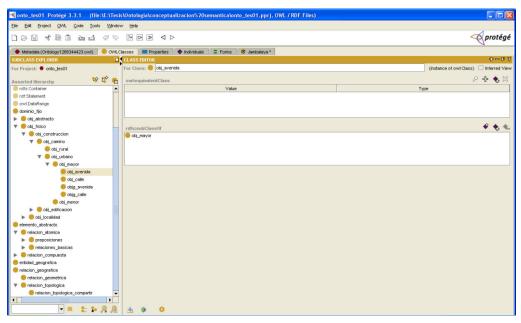


Figura 5.3. Construcción de la ontología de la conceptualización del CDF y CDG

Dichas clases y subclases fueron determinadas con base en el diccionario de datos del INEGI [INEGI, 1996][INEGI, 1993][INEGI, 1993b] y los conceptos devueltos por WordNET [WordNET, 2010]. En este caso, se crearon dos clases principales que heredan de la clase principal *cosa*, denominadas *objeto_físico* y *objeto_abstracto*, en la primera se conceptualizarán los objetos geográficos de ambos conjuntos, es así que se crea la clase *construcción*, hija de la clase *objeto_físico*, y padre de la clase *camino*, ésta a su vez es padre de la clase *urbano* y éste de la clase *mayor*, por último esta clase es padre de las clases *o_avenida*, *g_avenida*, *o_calle* y *g_calle*, que son las clases, a partir de las cuales se instanciarán los objetos geográficos *calles* y *avenidas*, del CDF (nombre de clase que empieza con o_) y CDG (nombre de la clase que empieza con g_).

Por otro lado, la clase *construcción* también es padre de la clase *edificación* y ésta a su vez de las clases *o_educativo* y *g_educativo*, a partir de las cuales se instanciarán los objetos geográficos *escuelas* de ambos conjuntos de datos y de la clase *salud* que es padre de las clases *o_hospital* y *g_hospitales*, a partir de las cuales se instanciarán los objetos geográficos *hospitales* del CDF y CDG.

Los objetos geográficos *edificios* del CDF se instancian de la clase *unidadhabitacional*, hija de la clase *área*, que a su vez es hija de la clase *región* y ésta a su vez de la clase *localidad* y esta última de la clase *físico*. Por último la clase *área* es también padre de la clase *g_unidadhabitacional*, de la cual se instancian los objetos geográficos *unidad habitacional* del CDG.

Por otro lado, una vez identificadas las relaciones topológicas en el CDF entre los objetos geográficos y las relaciones topológicas semánticamente equivalentes en el CDG, se lleva a cabo la implementación en la ontología, para ello, se utilizó la metodología GEONTO-MET [Torres, 2008], para la implementación de las relaciones, como se muestra en la Figura 5.4.

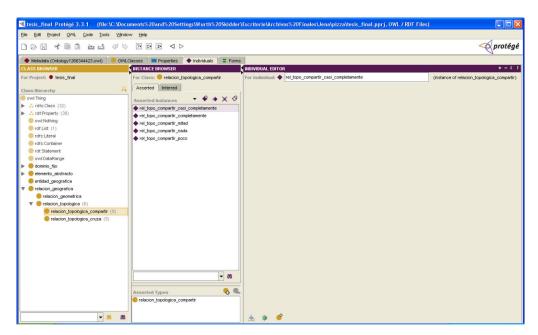


Figura 5.4. Implementación de las relaciones topológicas en la conceptualización del CDF y CDG

Inicialmente se construye las clase *relación_geográfica*, padre de la *clase relación_topológica*, a partir de la cual se instancian las siguientes relaciones topológicas:

- Relación topológica compartir.
- Relación topológica disjunta.
- Relación topológica cruza.
- Relación topológica conecta.
- Relación topológica equivale.

Esta última relación indica, que el objeto geográfico del CDF corresponde a algún objeto geográfico del CDG.

La clase *relación_topológica* es a su vez padre de cada una de las clases conceptuales identificadas en el análisis para la conceptualización, es así, que las clases que son hijas de dicha relación son, relación *topo_compartir* y relación *topo_cruza*, de las cuales se instancian los conceptos indicados en los esquemas de vecindad conceptual definidos anteriormente.

Ya que se ha conceptualizado el dominio de los conjuntos de datos geográficos CDF y CDG, se instancian los objetos geográficos de los conjuntos de datos geográficos en la ontología, como se muestra en la Figura 5.5.

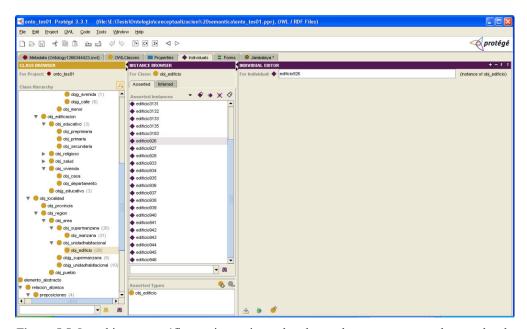


Figura 5.5. Los objetos geográficos se instancian en las clases a las que pertenecen de acuerdo a la conceptualización, como por ejemplo los edificios

Utilizando una vista "Nested/Tree Map" se muestra la ontología final, en la cual se conceptualiza el dominio de trabajo, esto se muestra en la Figura 5.6.

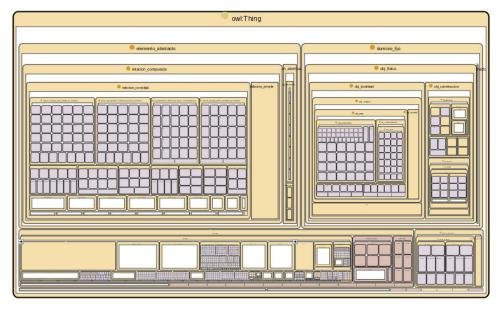


Figura 5.6. Ontología diseñada con base en el dominio de trabajo.

En la Figura 5.7 se muestra la ontología en un diagrama de árbol, donde se pueden observar parte de las instancias, que se refieren a los objetos geográficos de ambos

conjuntos de datos geográficos y las relaciones entre ellos, como parte de la conceptualización del dominio de pruebas.

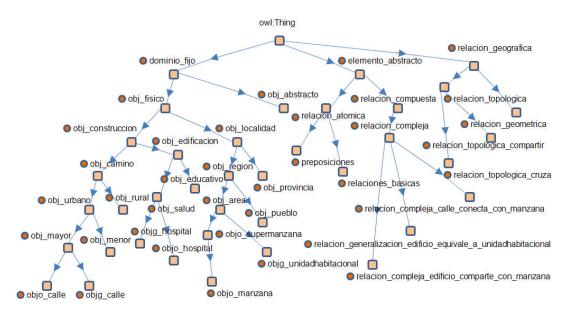


Figura 5.7. Instancias y relaciones en la ontología diseñada.

Como resultado de la etapa de síntesis se obtiene un archivo con extensión OWL, en un formato XML, el cual contiene la conceptualización del dominio de los conjuntos de datos geográficos en una ontología.

5.2 Síntesis

Realizando un análisis de los datos geográficos se lleva a cabo la identificación de las relaciones topológicas del *CDF* y *CDG*. A continuación se presentan algunos de los resultados obtenidos.

5.2.1 Medición de la relación topológica para unidades habitacionales (RTD1)

Para definir las relaciones topológicas semánticamente equivalentes similares del CDG a la relación RTD1 (Ver Sección 4.2.1.1) del CDF y CDG. En la Figura 5.8 se muestra el CDG, con los objetos geográficos *unidad habitacional* y las áreas que se intersectan.

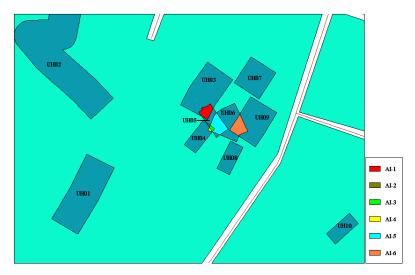


Figura 5.8. Unidades Habitacionales en el CDG

En la Tabla 5.4 se muestran las medidas de cada una de las áreas, para el cálculo de estas mediciones se utilizó el software ArcGIS 9.2 y la capa del conjunto de datos "unidad habitacional". Después de realizar el cálculo en forma automática, como se muestra en la Figura 5.9.

Tabla 5.4. Áreas de la Unidades Habitacionales y Áreas de Intersección

Objeto	$ \acute{\text{A}}\text{rea}(\emph{O}^{o}) \text{ en m}^{2} $	Objeto	$\acute{ ext{Area}}({ ext{O}_1}^o { ext{O}_2}^{ ext{O}})$ en
			m^2
UH01	2182.29	AI-1	115.19
UH02	4315.96	AI-2	5.52
UH03	1431.29	AI-3	13.28
UH04	416.82	AI-4	6.88
UH05	419.32	AI-5	196.57
UH06	830.39	AI-6	181.78
UH07	872.03		
UH08	397.80		
UH09	1166.61		
UH10	401.93		

Para realizar el cálculo se utilizó la función "Field Calculator" de ArcGIS 9.2, después de crear el campo "AREA", correspondiente a la tabla de atributos de la capa de datos fuente, utilizando el siguiente script, cuyos resultados aparecen en la Figura 5.9.

```
Dim Output as double
Dim pArea as Iarea
Set pArea = [shape]
Output = pArea.area
[AREA] = Output
```

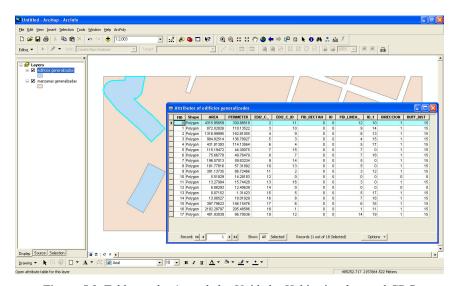


Figura 5.9. Tabla con las áreas de las Unidades Habitacionales en el CDG

A continuación, se identifican cada una de las regiones en las cuales se intersectan algunos de los objetos geográficos y se calcula cada una de estas áreas. En la Tabla 5.5 se muestran las unidades habitacionales que se intersectan y el área de intersección entre ellas.

Objetos 0 ₁ y 0 ₂	Area(O ^o) en m ²	Intersecciones	Area $({\rm O_1}^o {\buildrel \cap} {\rm O_2}^o)$ en m ²
UH03 – UH05	1850.61	AI1	115.19
UH04 – UH05	836.14	AI2, AI3	18.8
UH05 – UH06	1249.71	AI3, AI5	209.85
UH04 – UH06	1247.21	AI3, AI4	20.16

AI6

Tabla 5.5. Agrupando Áreas de la Unidades Habitacionales y Áreas de Intersección

A partir de las mediciones realizadas, se lleva a cabo el cálculo de RTD1, para los objetos geográficos y se hace la clasificación de acuerdo con los criterios de la Tabla

1997

UH06 - UH09

181.78

4.3. Los resultados de la clasificación a los objetos del conjunto de datos fuente aparecen en la Tabla 5.6.

Objetos Geográficos Concepto que representa RTD1 UH03 - UH05 6.22 compartir_mitad UH04 - UH05 2.25 compartir poco UH05 - UH06 16.79 compartir mitad UH04 – UH06 1.62 compartir poco UH06 - UH09 9.1 compartir mitad

Tabla 5.6. Conceptualización con base en RTD1

Este procedimiento se realiza tanto para el conjunto de datos fuente como para el conjunto de datos generalizado.

5.2.2 Medición de la relación topológica entre hospitales y escuelas con calles (RTD2)

Para definir las relaciones topológicas semánticamente equivalentes similares del CDG a la relación RTD2 (Ver Sección 4.2.1.2) del CDF y CDG. En la Figura 5.10 se muestra el CDG, con los objetos geográficos *escuelas, hospitales y calles* y las áreas de las *escuelas y hospitales* que se encuentran sobre las *calles*.

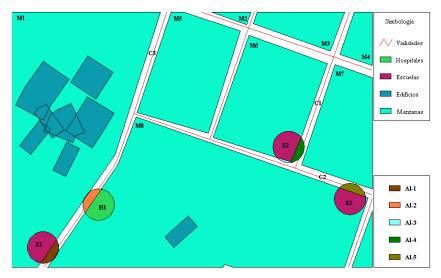


Figura 5.10. Area de escuelas y hospitales en el CDG con respecto a las calles

En la Tabla 5.7 se muestran las medidas de cada una de las áreas, para el cálculo de estas mediciones, se utilizó el software ArcGIS 9.2 y las capas del conjunto de datos "escuelas", "hospitales" y "avenidas y calles". Después de hacer el cálculo en forma automática, como se muestra en la Figura 5.11.

Tabla 5.7. Áreas de las Escuelas y Hospitales y Áreas sobres las calles

Objeto	Area(O ^o) en m ²	Objeto	Area(O ₁ ^o) en m ²
E1	703.28	AI-1	162.85
E2	703.28	AI-2	162.41
E3	703.28	AI-3	19.47
H1	703.28	AI-4	108.70
		AI-5	160.07

Para realizar el cálculo se utilizó la función "Field Calculator" de ArcGIS 9.2, después de crear el campo "AREA" en la tabla de atributos de la capa de datos fuente, utilizando el siguiente script, cuyos resultados aparecen en la Figura 5.11.

Dim Output as double
Dim pArea as Iarea
Set pArea = [shape]
Output = pArea.area
[AREA] = Output

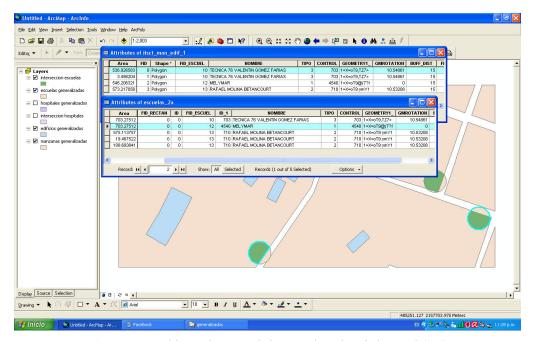


Figura 5.11. Tabla con las áreas de las escuelas y hospitales en el CDG

A continuación, se identifican cada una de las regiones de las *escuelas* y *hospitales*, las cuales se encuentran sobre las calles. En la Tabla 5.8 se muestran estas regiones, relacionándolas con las calles y escuela u hospital correspondiente.

Tabla 5.8. Relación de los objetos de Intersección

Objetos	Objetos de Intersección
H1 – C3	AI-2
E1 – C3	AI-1
E2 – C2	AI-3
E2 – C1	AI-4
E3 – C2	AI-5

A partir de las mediciones realizadas, se lleva a cabo el cálculo de RTD2, para los objetos geográficos y se hace la clasificación de acuerdo con los criterios de la Tabla 4.4. Los resultados de la clasificación a los objetos de conjunto de datos fuente aparecen en la Tabla 5.9.

Objetos Geográficos $0_1 ext{ y } 0_2$	RTD2	Concepto que representa
H1 – AI-2	23.1	compartir_casi_completamente
E1 – AI-1	23.1	compartir_casi_completamente
E2 – AI-3	2.8	compartir_poco
E2 – AI-4	15.5	compartir_mitad
E3 – AI-5	22.8	compartir_casi_completamente

Tabla 5.9. Conceptualización con base en RTD2

Este procedimiento se realiza tanto para el conjunto de datos fuente como para el conjunto de datos generalizado.

5.2.3 Medición de la relación topológica entre hospitales y escuelas con manzanas (RTD3)

Para definir las relaciones topológicas semánticamente equivalentes similares del CDG a la relación RTD3 (Ver Sección 4.2.1.3) del CDF y CDG. En la Figura 5.12 se muestra el CDG, con los objetos geográficos *escuelas, hospitales y calles* y las áreas de las *escuelas y hospitales* que se encuentran sobre algunas *manzanas*.

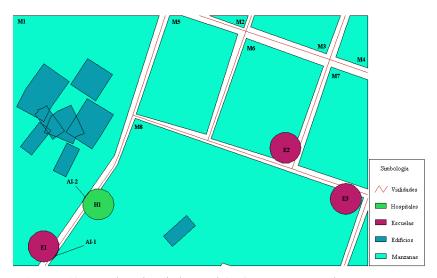


Figura 5.12. Escuelas y hospitales en el CDG con respecto a algunas manzanas.

En la Tabla 5.10 se muestran las medidas de cada una de las áreas, para el cálculo de estas mediciones, se utilizó el software ArcGIS 9.2 y las capas del conjunto de datos "escuelas", "hospitales" y "manzanas". Después de realizar el cálculo en forma automática, como se muestra en la Figura 5.13.

Tabla 5.10. Áreas de las Escuelas y Hospitales y Áreas de intersección sobre algunas manzanas

Objeto	Area (O^0) en m^2	Objeto	Area (O_1^0) en m ²
E1	703.28	AI-1	6.04
E2	703.28	AI-2	3.50
E3	703.28		
H1	703.28		

Para realizar el cálculo se utilizó la función "Field Calculator" de ArcGIS 9.2, después de crear el campo "AREA" en la tabla de atributos de la capa de datos fuente, utilizando el siguiente script, cuyos resultados aparecen en la Figura 5.13.

```
Dim Output as double
Dim pArea as Iarea
Set pArea = [shape]
Output = pArea.area
[AREA] = Output
```

8. Anexos

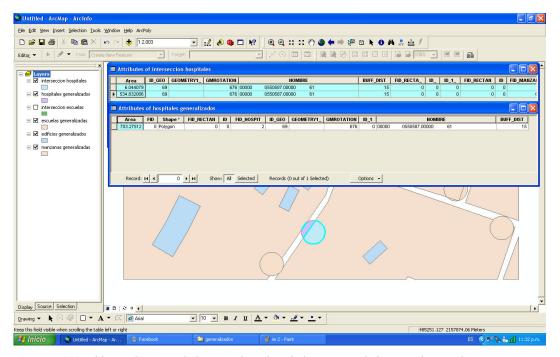


Figura 5.13. Tabla con las áreas de las escuelas y hospitales, y áreas de intersección en algunas manzanas en el CDG

A continuación, se identifican cada una de las regiones de las *escuelas* y *hospitales*, las cuales se encuentran sobre algunas manzanas. En la Tabla 5.11 se muestran estas regiones, relacionándolas con las manzanas y escuela u hospital correspondiente.

Tabla 5.11. Relación de los objetos de Intersección

Objetos	Objetos de Intersección
H1 – M1	AI-2
E1 – M8	AI-1

A partir de las mediciones realizadas, se hace el cálculo de RTD3, para los objetos geográficos y se lleva a cabo la clasificación de acuerdo con los criterios de la Tabla 4.5. Los resultados de la clasificación a los objetos de conjunto de datos fuente aparecen en la Tabla 5.12.

Tabla 5.12. Conceptualización con base en RTD3

Objetos Geográficos	RTD3	Concepto que representa
$0_1 \text{ y } 0_2$		
H1 – AI-2	0.86	compartir_poco
E1 – AI-1	0.50	compartir_poco

Este procedimiento se realiza tanto para el conjunto de datos fuente como para el conjunto de datos generalizado.

5.2.4 Medición de la relación topológica entre hospitales y escuelas con calles (RTC)

Para definir las relaciones topológicas semánticamente equivalentes similares del CDG a la relación RTC (Ver Sección 4.2.1.4) del CDF y CDG. En la Figura 5.14 se muestra el CDG, con los objetos geográficos *escuelas, hospitales y calles* y la distancia que cruza el área de las *escuelas u hospitales* en las *calles*.

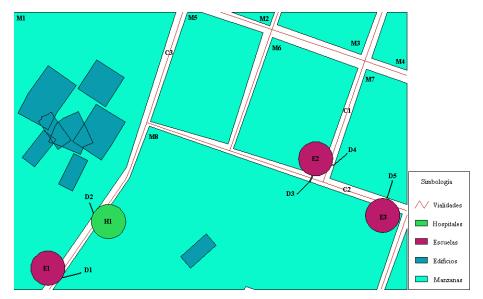


Figura 5.14. Distancia de cruza entre escuelas y hospitales con respecto a las calles en el CDG.

En la Tabla 5.13 se muestran las medidas de cada una de las distancias, para el cálculo de éstas, se utilizó el software ArcGIS 9.2 y las capas del conjunto de datos "escuelas", "hospitales" y "avenidas y calles". Después de realizar el cálculo en forma automática, como se muestra en la Figura 5.15.

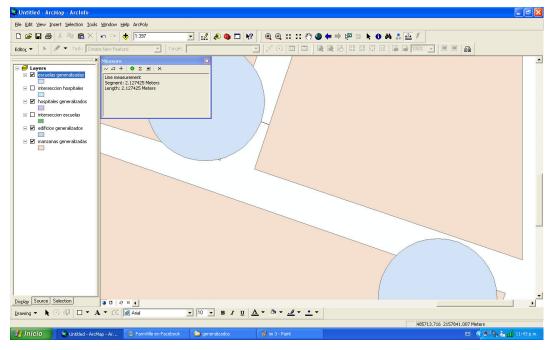


Figura 5.15. Medición de las distancias de las áreas de las escuelas y hospitales con respecto a las calles en el CDG

Tabla 5.13. Distancia de cruce de las Áreas de las Escuelas y Hospitales sobres las calles

Objeto	Distancia en m	Distancia	Distancia en m
C1 con respecto a E2	8.64	D1	8.28
C2 con respecto a E2	8.60	D2	8.33
C2 con respecto a E3	9.62	D4	6.17
C3 con respecto a H1	8.28	D3	2.69
C3 con respecto a E1	8.33	D5	8.77

A continuación, se identifican cada una de las distancias de cruce de las *escuelas* y *hospitales*, con respecto a algunas calles, en la Tabla 5.14 se muestran estas distancias, relacionándolas con las calles y escuela u hospital correspondiente.

Tabla 5.14. Relación de los objetos de Intersección

Objetos	Objetos de Intersección
H1 – C3	D1
E1 – C3	D2
E2 – C1	D4
E2 – C2	D3
E3 – C2	D5

A partir de las mediciones realizadas se lleva a cabo el cálculo de RTC, para los objetos geográficos y se hace la clasificación de acuerdo con los criterios de la Tabla 4.6. Los resultados de la clasificación a los objetos de conjunto de datos fuente aparecen en la Tabla 5.15.

Tabla 5.15. Conceptualización con base en RTC

Objetos Geográficos	RTC	Concepto que representa
$0_1 \text{ y } 0_2$		
H1 – C3	0	cruza_completamente
E1 – C3	0	cruza_completamente
E2 – C2	5.91	cruza_poco
E2 – C1	2.47	cruza_mitad
E3 – C2	0.85	cruza_completamente

Este procedimiento se realiza tanto para el conjunto de datos fuente como para el conjunto de datos generalizado.

5.3 Verificación

Después de realizar la conceptualización de los conjuntos de datos geográficos, se realizará la verificación de la similitud semántica entre ambos conjuntos, identificando si el *CDG* es consistente con respecto al *CDF*.

Inicialmente, se considera a la ontología contenida en un archivo OWL, con formato XML que ha sido obtenido en la etapa de síntesis. En el proceso de verificación de este archivo y haciendo uso del programa diseñado en Java y con la extensión Jena [Jena, 2009] para el manejo de ontologías y la extensión Xerces como parser *XML* [Xerces, 2010], se extraen del archivo OWL cada de uno de los objetos geográficos y las relaciones topológicas, almacenando en un archivo denominado **instancias.xml**, las instancias, objetos geográficos de ambos conjuntos de datos geográficos, parte de este archivo se muestra en el Anexo 2 y en el archivo denominado **relaciones.xml**, las relaciones topológicas, entre los objetos geográficos del CDF y CDG, parte del cual se muestra en el Anexo 3; este proceso se muestra en la Figura 5.16.



Figura 5.16. Generación de archivos de relaciones e instancias con base en la ontología diseñada

Ya que se obtuvieron los archivos, se integran como entrada, al proceso desarrollado en Java que realizará el análisis a un nivel *micro* de acuerdo con la Figura 5.17.

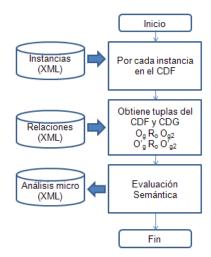


Figura 5.17. Análisis a nivel micro

En este proceso, se toma inicialmente el archivo xml que contiene las instancias de ambos conjuntos de datos geográficos, para que así, identificando aquellas instancias que pertenecen al *CDF*, y por cada instancia u objeto geográfico identificado, se obtienen, las tuplas con sus relaciones topológicas en el CDF, las cuales están contenidas en el archivo XML, denominado relaciones.

Por ejemplo, para el objeto geográfico "<calleg7698>calle7698</calleg7698>", encontrado en el archivo instancias.xml, se identifica la tupla, en el archivo relaciones.xml, como se muestra a continuación:

```
<rel_calle_cruza_avenida_01><_origen_>o_calle7698</_origen_><_accion_>a_rel_to
po_cruza</_accion_><_destino_>d_avenida5494</_destino_></rel_calle_cruza_aveni
da 01>
```

De esta tupla, podemos identificar la relación topológica "a_rel_topo_cruza" y el objeto geográfico "d_avenida5494", haciendo uso de esta información se puede determinar en el archivo **relaciones.xml** que el objeto geográfico "calle7698" es equivalente al objeto geográfico "calleg03", al obtener la relación de equivalencia.

```
<rel_calle_equivale_calle_03><_destino_>d_calleg03</_destino_><_accion_>a_rel_
topo_equivale</_accion_><_origen_>o_calle7698</_origen_></rel_calle_equivale_c
alle 03>
```

Y el objeto geográfico "avenida5494", es equivalente al objeto geográfico "avenidag01", de acuerdo con la siguiente relación de equivalencia:

```
<rel_avenida_equivale_avenida_01><_destino_>d_avenidag01</_destino_><_accion_>
a_rel_topo_equivale</_accion_><_origen_>o_avenida5494</_origen_></rel_avenida_
equivale avenida 01>
```

Ya que se han identificado las equivalencias de los objetos geográficos, se determina la relación topológica entre ambos objetos geográficos generalizados, la cual es "rel topo cruza completamente".

```
<rel_calle_cruza_avenida_02><_origen_>o_calleg03</_origen_><_destino_>d_avenid
ag01</_destino_><_accion_>a_rel_topo_cruza_completamente</_accion_></rel_calle
cruza avenida 02>
```

Posteriormente, ya identificados los objetos geográficos y las relaciones topológicas de a identificar si la ambos conjuntos, se procede relación "rel topo cruza completamente" del objeto geográfico "calleg03" es (in)consistente, con base en el esquema de vecindad conceptual mostrado en la Figura 4.6, utilizando un valor pequeño de umbral u = 0.2, el cual fue determinado arbitrariamente, con el objetivo de obtener relaciones topológicas que fuesen muy consistentes. Estas relaciones topológicas pertenecen a los objetos geográficos de las capas "unidades habitacionales", "escuelas" y "hospitales", con respecto a los objetos geográficos de las capas "unidades habitacionales", "manzanas" y "calles y avenidas" del conjunto de datos generalizado, este proceso se realiza para cada uno de los objetos geográficos en el CDG, considerando los esquemas de vecindad conceptual mostrados en las Figuras 4.5 y 4.6, y utilizando el mismo valor de umbral u.

Después de realizar el análisis a nivel *micro*, se realiza un análisis a un nivel *meso*, de acuerdo con la Figura 5.18, haciendo uso de la ecuación 4.9, con un valor de umbral de w = -8, el cual fue determinado en forma arbitraria, considerando que entre menos negativo fuese el valor de umbral, mejor consistencia entre los datos se determina, con base en la ecuación 4.9.

Así, se obtiene una lista como la que se muestra a continuación, con los distintos grupos de objetos geográficos en el *CDG*, indicando si éstos son (in)consistentes con los correspondientes grupos de objetos geográficos en el *CDF*.

Escuela#Igual
Calle#Igual
Edificio#Equivalente
Hospital#Igual
Avenida#Igual

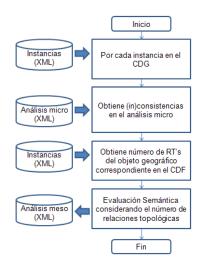


Figura 5.18. Análisis a nivel meso

Ya que se han identificado las (in)consistencias de cada uno de los objetos geográficos del CDG con respecto a sus correspondientes objetos geográficos en el CDF, se verifica si el conjunto de datos geográfico generalizado es (in)consistente con respecto al conjunto de datos geográfico fuente, realizando un análisis a un nivel *macro*, de acuerdo con la Figura 5.19, y haciendo uso de la ecuación 4.10, estableciendo un valor de umbral de z = -8, el cual fue determinado en forma arbitraria, considerando que entre menos negativo fuese el valor de umbral, mejor consistencia entre los datos se obtiene, de acuerdo con la ecuación 4.10.

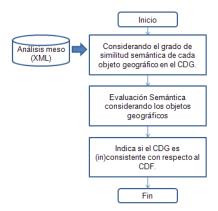


Figura 5.19. Análisis a nivel macro

Se determina así, que el conjunto de datos geográficos generalizado es *consistente* con el conjunto de datos geográficos fuente, arrojando el programa el siguiente resultado.

Mapa#Equivalente

Capítulo 6. Conclusiones y trabajo a futuro

El proceso de generalización se utiliza para modificar el nivel de detalle en datos geográficos; sin embargo, en numerosas ocasiones se generan inconsistencias. Por lo que se hace de gran importancia verificar que ambos conjuntos de datos sean consistentes entre sí, ya que de no ser así, cometeríamos muchos errores en el procesamiento subsecuente de los datos generalizados.

Con la metodología propuesta se incorporan diversos aspectos que resultan útiles para evaluar la calidad de los datos, el cual se basa en la representación de los objetos geográficos y sus relaciones que son caracterizadas por medio de una conceptualización. Esto se utiliza para comparar desde el punto de vista conceptual los datos generalizados con los datos fuente, por medio de la medición de similitud semántica, lo que representa una analogía con la forma en que lo realizaría un cartógrafo.

La utilización de medidas sobre los datos geográficos permite evaluar las relaciones topológicas, considerando detalles métricos, y obteniendo una caracterización adecuada. Con lo que se pueden utilizar los esquemas de vecindad conceptual para evaluar la similitud y/o proximidad conceptual entre las relaciones espaciales.

La valoración de la consistencia en diferentes niveles de detalle (*micro, meso y macro*) [Ruas, 2000][Bard, 2003], permite considerar aspectos que frecuentemente no son considerados. Con lo que se obtienen conceptos que describen la consistencia a diferente nivel de detalle, los cuales de definen con base en valores de umbral (u, w y z).

6.1 Aportaciones

Las aportaciones que genera este trabajo de tesis son las siguientes:

- Una metodología que permite identificar el grado de consistencia o inconsistencia de un conjunto de datos geográfico con respecto al conjunto de datos geográfico, a partir del cual fue generado, con base en un proceso de generalización.
- La caracterización de relaciones topológicas utilizando esquemas de vecindad conceptual, basados en medidas de similitud semántica, dependiendo de la relación topológica y del tipo de objetos geográficos que se relacionan con dicha relación topológica.
- La evaluación de la consistencia en los datos generalizados considerando diferentes niveles de detalle.
- Una metodología que puede formar parte de un esquema de control de calidad del conjunto de datos geográfico, no únicamente para datos generalizados.
- Elaboración del artículo "Detección de inconsistencias en datos geográficos generalizados con base en representaciones semánticas", presentado en el Congreso Internacional CIINDET 2010, en conjunto con el Dr. Marco A. Moreno Ibarra y el Dr. Miguel J. Torres Ruiz,.
- Elaboración del artículo "Verificación de la consistencia de objetos geográficos generalizados utilizando representaciones semánticas", presentado en el Simposium Internacional SELPER 2010, en conjunto con el Dr. Marco A. Moreno Ibarra y el Dr. Miguel J. Torres Ruiz.

6.3 Trabajo a futuro

El trabajo a futuro va encaminado a la caracterización de relaciones topológicas en otros dominios de trabajo diferentes al usado en el presente, y al refinamiento de otras y de las mismas relaciones topológicas utilizadas en esta tesis, pero con diferentes objetos geográficos a los utilizados.

Para así poder generar una base de relaciones topológicas caracterizadas en esquemas de vecindades conceptuales que permita ser utilizada en diversos proyectos, donde sea necesario conocer al detalle la calidad de los datos, siendo esta metodología parte de un esquema mayor de verificación de calidad de los datos geográficos. Así también, generar nuevas medidas de similitud semántica que permitan enriquecer este trabajo que dependan del dominio de nuestros conjuntos de datos.

El valor de los umbrales utilizados en este trabajo, debe ser determinado por expertos, haciendo uso de metodologías que dependan de los objetivos y áreas de uso de esta metodología. Para evitar que dichos valores se determinen de manera arbitraria, se debe desarrollar una técnica para ajustar los umbrales con base en el análisis de las preferencias del usuario.

Referencias

[Armstrong, 1991] Armstrong, M.P., "Knowledge classification and organization", in Buttenfield B.P. and McMaster R.B. (eds.) Map Generalization: Making Rules for Knowledge Representation. London, U.K. Longman, 86–102, 1991.

[Bard, 2003] Bard S., "Quality Assessment of Cartographic Generalisation", Transactions in GIS 8(1), 2003.

[Barrault et al., 2001] Barrault M., Regnauld N., Duchêne C., Haire K., Baeijs C., Demazeau Y., Hardy P., Mackaness W., Ruas A. and Weibel R., "Integrating multiagent, object-oriented and algorithmic techniques for improved automated map generalization", in Proceedings 20th International Cartographic Conference, Beijing, China, 2110–2116, 2001.

[Basaraner, 2007] Basaraner, M., "Model and Cartographic Generalisation in GIS: A Comparison", Modern technologies, education and professional practice in geodesy and related fields, Sofia, Bulgaria, 2007.

[Brassel and Weibel, 1988] Brassel, K., Weibel, R., "A Review and Conceptual Framework of Automated Map Generalization", International Journal of Geographic Information Systems, 2(3): 229-244, 1988.

[Berg et al., 1995] Berg M de, Krevel M van, Schirra S, "A new approach to subdivision simplification", ACSM/ASPRS Annual Convention and Exposition 4 (Proceedings AutoCarto 12): 79 – 88, 1995.

[Brazile, 2000] Brazile L. F., "Semantic Infrastructure and Methods to Support Quality Evaluation in Cartographic Generalization", PhD Thesis, Department of Geography, Zurich University, German, 2000.

[Bruns and Egenhofer, 1996] Bruns T. H., Egenhofer M., "Similarity of Spatial Scenes", 7th International Symposium on Spatial Data Handling, Delf, The Netherlands, 1996.

[Burrough, 1986] Burrough, P.A., "Principles of geographical information systems for land resources assessment", Oxford: Clarendon Press, 1986.

[Buttenfield and McMaster, 1991] Buttenfield B.P., McMaster R.B., Map Generalization. Making Rules for Knowledge Representation, London, UK: Longman, 1991.

[Chang et al., 1989] Chang, S.K., Jungert, E., Li, Y., "The Design of Pictorial Databases Based Upon the Theory of Symbolic Projections", Symposium on the Design and Implementation of Large Spatial Databases, Lecture Notes in Computer Science, Vol. 409, pages 303-323, Springer-Verlag, 1989.

[Chrisman, 1983] Chrisman, N. R., "Epsilon Filtering: A technique For Automated Scale Changing", 43rd annual ACSM Meeting, Washington D.C., 322-331, 1983.

[Cohn et al., 1997] Cohn, A., Bennett, B., Gooday, J., Gotts, N. "Qualitative spatial representation and reasoning with the region connection calculus", GeoInformatica, 1:275-316, 1997.

[Cook and Jones, 1989] Cook A. and Jones C., "Rule-based cartographic name placement with prolog", in Proceedings AutoCarto 9, Baltimore, USA, 231–240, 1989.

[Corcho et al., 2003] Corcho O., Fernández-López M., Gómez-Pérez A., "Methodologies, tools and languages for building ontologies. Where is their meeting point?", Data & Knowledge Engineering 46:41-64, 2003.

[Cohn et al., 1994] Cohn, A., Randell, D., Cui, Z.,"Taxonomies of logically defined qualitative spatial relations", Formal Ontology in Conceptual Analysis and Knowledge Representation, Kluwer, 1994.

[Cui et al., 1993] Cui, Z., Cohn, A.G., Randell, D.A., "Qualitative and Topological Relationships in Spatial Databases", Third Symposium on Large Spatial Databases, Lecture Notes in Computer Science 692:296-315, Singapore, 1993.

[Davis and Aquino, 2003] Davis, M., Aquino, J., "JTS Topology Suite – Technical Specifications", Vivid Solutions Victoria, British Columbia, 2003.

[Deveau, 1985] Deveau, T.J., "Reducing the number of points in a plane curve representation". Proceedings of the Seventh International Symposium on Automated Cartography, AUTO-CARTO VII. Falls Church, VA: American Society of photogrammetry and the American Congress on Surveying and Mapping. 152-160, 1985.

[Douglas and Peucker, 1973] Douglas, D., Peucker T., "Algorithms for the reduction of the number of points required to represent a digitized line or its caricature". The Canadian Cartographer, 10(2), 112-122, 1973.

[Duchêne et al., 2002] Duchêne C., "Coordinative agents for automated generalisation of rural areas", in Proceedings 5th Workshop on Progress in Automated Map Generalization, Paris, France, 2002.

[Edwardes and Mackaness, 2000] Edwardes A., Mackaness W., "Modelling knowledge for automated generalization of categorical maps - a constraint based approach", in Atkinson P. and Martin D. (eds.), GIS and GeoComputation (Innovations in GIS 7), London: Taylor&Francis, 161–173, 2000.

[Egenhofer and Herring, 1990] Egenhofer, M., Herring, J., "A Mathematical Framework for the Definition of topological Relationships", Fourth International Symposium on

Spatial Data Handling, pages 803-813, Zurich, Switzerland, 1990.

[Egenhofer and Franzosa, 1991] Egenhofer M., Franzosa R., "Point-Set Topological Spatial Relations", International Journal of Geographical Information Systems 5(2):161-174, 1991.

[Egenhofer and Al-Taha, 1992] Egenhofer M., Al-Taha K., "Reasoning About Gradual Changes of Topological Relationships", Theories and Methods of Spatio-Temporal Reasoning in Geographic Space, Lecture Notes in Computer Science 639:196-219, Springer-Verlag, Nueva York, U.S., 1992.

[Egenhofer et al., 1994b] Egenhofer, M., Mark, D., Herring, J., "Categorizing topological spatial relationships between point, line and area objects, The 9-intersections: formalism and its use for natural language spatial predicates, Technical report 94-1, National Center for Geographic Information and Analysis, 1994.

[Egenhofer and Mark, 1995a] Egenhofer M., Mark D., "Modeling Conceptual Neighborhoods of Topological Line-Region Relations", International Journal of Geographical Information Systems 9(5):555-565, 1995.

[Egenhofer and Dube, 2009] Egenhofer M., Dube M., "Topological Relations from Metric Refinements", 17th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, Seattle WA, 2009.

[Fernández-López, 1999] Fernández-López M., "Overview of Methodologies for Building Ontologies", Proceedings of the IJCAI-99 workshop on Ontologies and problem-Solving Methods (KRR5), Sweden, 1999.

[Filippovska, 2008] Filippovska Y., Walter, V., Fritsch, D., "Quality Evaluation of Generalization Algorithms", The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Vol. XXXVII, Part B2, Commission 2. ISPRS Congress, Beijing, China. P. 799ff. ISSN 1682-1750, 2008.

[Fisher and Mackaness, 1987] Fisher, P.F., Mackaness, W., "Are Cartographic Expert Systems Possible?", Proceedings AutoCarto 8, Baltimore, USA, 530–534, 1987.

[Fisher et al., 2009] Fisher P., Comber A., Wadsworth R., "What is in a name? Semantics, Uncertainty Standards and Data Quality", 6th International Symposium on Spatial Data Quality, St John's, Canada, July 2009.

[Fonseca and Li, 2006] Fonseca F., Li B., "TDD – A Comprehensive Model for Qualitative Spatial Similarity Assessment", Spatial Cognition & Computation: An Interdisciplinary Journal 6(1):31-62, Taylor & Francis, 2006.

[Frank and Ester, 2006] Frank R., Ester M., "A Quantitative Similarity Measure for Maps", Proc. 12th International Symposium on Spatial Data Handling, Vienna, Austria, 2006.

[Freeman 1975] Freeman, J., "The modeling of Spatial Relations", Computer Graphics and Image Processing, 4:156-171, 1975.

[Freeman and Doerschler, 1992] Freeman H.R., Doerschler J.S., "A rule-based system for dense map name placement", Communication of the ACM, 35(2), 68–79, 1992.

[ICA, 1973] International Cartographic Association: (1973) Multilingual Dictionary of Technical Terms in Cartography, Steiner, Wiesbaden.

[Gómez-Pérez, 1997] Gómez-Pérez Asunción, "Ontological Engineering: A State of The Art", Facultad de Informática, Universidad Politécnica de Madrid, 1997.

[Goodchild, 1991] Goodchild M., Issues of quality and uncertainty, In: Muller, J.-C. (ed.) Advances in Cartography, International Cartographic Association (ICA), London: Taylor&Francis, pp:113-139, 1991.

[Goodchild, 1999] Goodchild M. F., Paul A. L., Davild J. M., Geographical Information Systems: Principles and Technical Issues, John Wiley & Sons Inc, 2a Ed., pp 128, 150, 1999.

[Heywood et al., 1998] Heywood I., Cornelius S., Carver S., "An Introducction to Geographical Information Systems", Prentice Hall, Addison Wesley Longman Limited, pp 147, 178, 200, 1998.

[INEGI, 1993] Base de datos geográficos. Modelo de Datos Vectoriales, Instituto Nacional de Estadística Geografia e Informática, 1996.

[INEGI, 1993b] Base de datos geográficos. Acerca de los Modelos de Datos Vectoriales, Instituto Nacional de Estadística Geografía e Informática, 1996.

[INEGI, 1996] Base de datos geográficos. Diccionario de datos topográficos Escalas 1:50,000, Instituto Nacional de Estadística Geografía e Informática, 1996.

[Janowicz, 2008] Janowicz K., "Kinds of context and their impact on semantic similarity measurement", Proceedings on the 5th IEEE Workshop on Context Modeling and Reasoning (CoMoRea), Hong Kong, 2008.

[Janowicz et al., 2008] Janowicz K., Shwering A., Kuhn W., "Semantic Similarity Measurement and Geospatial Applications (editorial)", Special Issue on Semantic Similarity Measurement and Geospatial Applications, Transactions in GIS, 12(6): 651-659, 2008.

[Jena, 2009] Jena Onthology API, http://jena.sourceforge.net/ontology/

[Jennings et. al., 1998] Jennings N., Sycara K., Wooldridge M., "A roadmap of agent research and development", Autonomous Agents and Multi-Agent Systems, (1), 7–38, 1998.

[Joao, 1993] João, E.M., "Causes and Consequences of Map Generalisation", Ph.D. thesis, London School of Economics, UK, 1993.

[Kainz 1990] Kainz, W., "Spatial Relationships - Topology versus Order", Fourth International Symposium on Spatial Data Handling, pages 814-819, Zurich, Switzerland, 1990.

[Kebler et al., 2007] Kebler C., Raubal M., Janowicz K., "The effect of context on semantic similarity measurement", On the Move to Meaningful Internet Systems, Berlin, Springer Lecture Notes in Computer Science No. 4806:1274-84, 2007.

[Kilpeläinen, 2000] Kilpeläinen T., "Knowledge Acquisition for Generalization Rules", Cartography and Geographic Information Science, 27(1), 41–50, 2000.

[Knublauch and Horridge, 2004] Knublauch H., Horridge M., "A practical guide to building OWL Ontologies using the Protégé-OWL Plugin and CO-ODE Tools", 1st ed., Manchester University, 2004.

[Lagrange et al., 2000] Lagrange P., Landras B., Mustière S., "Machine Learning Techniques for Determing Parameters of Cartographic Generalisation Algorithms", in ISPRS (ed.), International Archives of Photogrammetry and Remote Sensing, 718–725, 2000.

[Lamy et al., 1999] Lamy S., Ruas A., Demazeau Y., Jackson M., Mackaness W. and Weibel R., "The Application of Agents in Automated Map Generalisation", in Proceedings 19th International Cartographic Conference, Ottawa, Canada, 1225–1234, 1999.

[Leitner and Buttenfield, 1995] Leitner M., Buttenfield B.P., "Acquisition of Procedural Cartographic Knowledge by Revers Engineering", Cartography and Geographic Information Systems, 22(3), 232–241, 1995.

[Lichtner, 1978] Lichtner, W., "Locational Characteristics and the sequence of computer Assisted Proces of Cartographic Generalization", Nachrichten aus dem Karten-und Vermessungwew, Reike II, 65-75, 1978.

[McGraw and Harbison-Briggs, 1989] McGraw K.L., Harbison-Briggs K., Knowledge acquisition, principles and guidelines, Englewood Cliffs, New Jersey: Prentice Hall, 1989.

[McMaster and Shea, 1992] McMaster, R.B., Shea, K.S., "Generalization in Digital Cartography", In: Association of American Geographers, 1992.

[Medin et al., 1993] Medin D., Goldstone R., Gentner D., "Respects for similarity", Psychological Review, 100(2): 254-278, 1993.

[Monmonier, 1983] Monmonier, M., "Raster-Mode Area Generalization of Land Use and Land Cover Maps. Cartographica", 20(4), 65-91, 1983.

[Moreno, 2007] Moreno I. M. A., "Similitud semántica entres sistemas de objetos geográficos aplicada a la generalización de datos geoespaciales", Tésis de Doctorado, Centro de Investigación en Computación. México, D.F., Octubre, 2007.

[Moreno, 2008] Moreno J. A., "Sistemas y Análisis de la Información Geográfica, Manual de autoaprendizaje con ArcGIS", Ed. Alfaomega Ra-Ma, 2d. Ed., 2008.

[Müller et al., 1995b] Müller J.C., Weibel R., Lagrange J.P., Salg'e F., "Generalization: State of the Art and Issues", in Muller J.C., Lagrange J.P. and Weibel R. (eds.), GIS and Generalization. Methodology and Practice. GISDATA 1, Taylor&Francis, 3–17, 1995.

[Mustière and Duchêne, 2001] Mustière S., Duchêne C., "Comparison of different approaches to combine generalization algorithms: GALBE, AGENT and CartoLearn", in Proceedings 4th Workshop on Progress in Automated Map Generalization, Beijing, China, URL http://www.geo.unizh.ch/ICA/, 2001.

[Nickerson, 1988] Nickerson B.G., "Automated cartographic generalization for linear features", Cartographica, 25(3), 15–66, 1988.

[Nickerson, 1991] Nickerson B.G., "Knowledge engineering for generalization", in Buttenfield B.P. and Mc-Master R.B. (eds.), Map Generalization: Making Rules for Knowledge Representation, London: Longman, 40–55, 1991.

[Peuquet and Ci-Xiang, 1987] Peuquet, D., Ci-Xiang, Z., "An algorithm to determine the directional relationship between arbitrarily-shaped polygons in the plane", Pattern Recognition 20(1): 65-74, 1987.

[Plazanet et al., 1998], Plazanet C., Bigolin M., Ruas A., "Experiments with Learning Techniques for Spatial Model Enrichment and Line Generalization", GeoInformatica, 2(4), 315–333, 1998.

[Plümer 1996] Plümer, L., "Achieving Integrity of Geometry and Topology in Geographical Information Systems", First International Conference on Geographics Information Systems in Urban Regional and Environmental Planing, Samos, Greece, pages 45-60, April, 1996.

[Podolskaya et al., 2007] Podolskaya E. S., Anders K. H., Sester M., "Quality Assessment for Polygon Generalization", Proceedings of the 5th International Symposium on Spatial Data Quality, Enschede, The Netherlands, 2007.

[Ratajski, 1967] Ratajski, L., "Phénomènes des points de généralization". International Yearbook of Cartography, 7, 143-151, 1967.

[Raubal, 2008] Raubal M., "Representing concepts in time", Spatial Cognition VI: Learning, Reasoning, and Talking about Space, Berlin, Springer Lecture Notes in Artificial Intelligence No. 5428:328-43, 2008.

[Reichenbacher, 1995] Reichenbacher T., "Eine Methode f'ur den Wissenserwerb in der kartographischen Generalisierung durch Interaktionsaufzeichnung und induktives Lernen", Master's thesis, Department of Geography, University of Zurich, 1995.

[Ruas, 1995] Ruas, A., "Multiple paradigms for automated map generalization: Geometry, topology, hierarchical partitioning and local triangulation", Proceedings AutoCarto 12, Charlotte, USA, 69–78, 1995.

[Ruas, 1999] Ruas A., "Modèle de généralization de donnèes géographiques `a base de contraints et d'autonomie", Ph.D. thesis, Université de Marne-la-Vallée, 1999. [Ruas, 2000] Ruas A., "The Roles of Meso Objects for Generalisation", 9th International Symposium on Spatial Data Handling, Beijing, China, 2000.

[Ruas, 2001] Ruas A., "Automating the generalization of geographical data: The Age of Maturity", The 20th International Cartographic Conference, Beijing, China, 2001.

[Schylberg, 1993] Schylberg L., "Computational Methods for Generalization of Cartographic Data in a Raster Environment", Ph.D. thesis, Royal Institute of Technology, Stockholm, 1993.

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

[Sester and Brenner, 2000] Sester M., Brenner C., "Typification based on Kohonen Feature Nets", in Proceedings GIScience 2000. The First International Conference on Geographic Information Science, Savannah, USA, 21–22, 2000.

[Shea y McMaster, 1989] Shea, K.S., McMaster, R.B., "Cartographic Generalization in a Digital Environment: When and How to Generalize", Proceedings, Ninth International Symposium on Computer-Asissted Cartography, AUTO-CARTO IX, Baltimore, Maryland. Falls Church, VA, American Society of Photogrammetry and Remote Sensing and American Congress on Surveying and Mapping, 56-67, 1989.

[Sloman et al., 1998] Sloman S. A., Love B. C., Woo-Kyung A., "Feature centrality and conceptual coherence", Cognitive Science, 22:189-228, 1998.

[Smith et al., 2001] Smith B., Mark, D.M., "Geographical categories: an ontological

investigation", International Journal of Geographical Information Science 15(7):591-612, Taylor & Francis, 2001.

[Steiner, 1976] Steiner W., "Multilingual Dictionary of Technical Terms in Cartography", International Cartographic Association, 1973.

[Strobl, 2008], Strobl C., "Dimensionally Extended Nine-Intersection Model (DE-9IM)", Encyclopedia of GIS 240-245, Springer, 2008.

[Stoter, 2005] Stoter J.E., "Generalisation within NMA's"in the 21st Century", Proceedings of the 22nd International Cartographic Conference: Mapping Approaches into a changing world, Coruña, Spain, 2005.

[Tomlinson, 2003] Tomlinson R. F., "Thinking about GIS: Geographic information systems planning for managers", Redlands, CA: ESRI Press, 2003.

[Töpfer and Pillewitzer, 1966] Töpfer, F., Pillewitzer, W., "The principles of selection". Cartographic Journal, 3, 10-16, 1966.

[Torres, 2008] Torres R. M. J., "Ontological representation based on semantic descriptors applied to geographic objects", Computación y Sistemas 12(3):356-371, 2008.

[Torsun, 1995] Torsun I.S., "Foundations of Intelligent Knowledge-Based Systems", Academic Press, London, UK, 1995.

[Touya, 2008] Touya G., "First thoughts for the orchestration of generalisation methods on heterogeneous landscapes", In workshop on generalisation and multiple representation, Montpellier, France, 2008.

[USGS, 2007] USGS, "Committee on Research Priorities for the USGS Center of Excellence for Geospatial Information Science", National Research Council(CB), Research Agenda for GIScience, Washington, DC, USA, National Academy Press, 2007.

[Waterman, 1986] Waterman D.A., "A Guide to Expert Systems, Reading", Massachussets, Addison-Wesley, USA, 1986.

[Weibel, 1991] Weibel, R., "Amplified intelligence and rule-based systems", In Buttenfield B.P. and Mc-Master R.B. (eds.). Map Generalization: Making Rules for Knowledge Representation. Longman, London, UK, 172–186, 1991.

[Weibel, 1995] Weibel, R., "Three essential building blocks for automated generalization", In Muller J.C., Lagrange J.P. and Weibel R. (eds.). GIS and Generalization, GISDATA 1, Taylor&Francis, 56–69, 1995.

[Weibel, 1995b] J.-C. Müller, J.-P. Lagrange, and R. Weibel. "GIS and Generalization: Methodology and Practice", Taylor & Francis: London, 1995.

[Werschlein and Weibel, 1994] Werschlein T., Weibel R., "Use of neural networks in line generalization", in Proceedings EGIS'94, Paris, France, 76–85, 1994.

[Wooldridge, 1999] Wooldridge M., "Intelligent Agents", in Weiss G. (ed.), Multiagent Systems. A modern Approach to Distributed Artifical Intelligence, The MIT Press, 27–78, 1999.

[Wordnet, 2010] http://wordnet.princeton.edu/

[Xerces, 2010] Apache Xerces Project, http://xerces.apache.org/

[Zoraster, 1991] Zoraster S., "Expert Systems and the Map Label Placement Problem", Cartographica, 28(1), 1–9, 1991.

Anexos

A. Diccionario de Datos

ESCUELA			
Descripción	Edificación acondicionada o construida para impartir instrucción en cualquiera de las		
	modalidades del Sistema Educativo Nacional.		
Atributos	Dominio Fijo		
	Tipo	Clasificación que se da a la escuela de acuerdo con el Sistema Educativo Nacional.	
	Dominio de Valores	 Preescolar. Edificación donde se imparten estudios previos a la primera enseñanza. Primaria. Edificación donde se imparte la primera enseñanza. Secundaria. Edificación donde se imparten estudios posteriores a la primera enseñanza de acuerdo con el Sistema Educativo Nacional. Medio Superior. Incluye Preparatorias, Colegios de Ciencias y Humanidades, Bachilleratos Tecnológicos, Centros de Estudios Científicos y Tecnológicos, Colegios de Bachilleres y CONALEP. Superior. Incluye universidades, institutos tecnológicos y escuela normal superior. Grupo de escuelas donde se imparte la enseñanza 	
		superior con potestad para otorgar títulos académicos.	
		Relaciones	
Datos	Escuela(P) conecta Calle(L)		
Originales	Escuela(A) conecta Calle(L)		
	Escuela(A) comparte Calle(L)		
	Escuela(A) comparte Hospital(A)		
	Escuela(A) comparte Manzana(A)		
	Escuela(A)) comparte Templo(A)	
Datos Generalizados	Escuela(A) Escuela(A) Escuela(A) Escuela(A) Escuela(A) Escuela(A)	conecta Calle(L)) conecta Calle(L)) comparte Calle(L)) comparte Hospital(A)) comparte Manzana(A)) comparte Templo(A)) cruza Calle(L)) conecta Manzana(P)	

CALLE			
Descripción	Vialidad de	finida para el tránsito vehicular o peatonal.	
Atributos	Dominio Fijo		
	Tipo	Clasificación que se da a la calle en función de la capacidad de flujo vehicular.	
	Dominio	- Primer orden. Boulevares, periféricos, ejes viales, viaductos y	
	de	avenidas principales.	
	Valores	 Segundo orden. Avenidas y calles de menor importancia que las definidas para calles de primer orden. Vialidad principal dentro de localidades pequeñas que sirven de enlace entre las carreteras que 	
		cruzan.	
		 Tercer orden. De tránsito local dentro de un área urbana. Cuarto orden. De tránsito local dentro de áreas diversas privada. 	
		cerrada, andadores.	
	<u> </u>	Relaciones	
	G 11 (T)		
Datos		necta Hospital(P)	
Originales		necta Centro comercial(P)	
		necta Estación del metro(P)	
		necta Instalación deportiva(P) necta Escuela(P)	
		necta Escuela(F)	
	Calle(L) comparte Escuela(A) Calle(L) conecta Manzana(A)		
	Calle(L) conecta Manzana(A) Calle(L) comparte Manzana(A)		
	Calle(L) comparte Manzana(A) Calle(L) conecta Templo(A)		
	Calle(L) comparte Templo(A)		
	Calle(L) conecta Edificación(P)		
	Calle(L) conecta Edificación(A)		
		mparte Edificación(A)	
		necta Calle(L)	
	Calle(L) cru	ıza Calle(L)	
Datos	Calle(L) con	necta Escuela(P)	
Generalizados	Calle(L) con	necta Escuela(A)	
		mparte Escuela(A)	
		necta Manzana(P)	
		necta Manzana(A)	
		mparte Manzana(A)	
		necta Templo(P)	
		necta Templo(A)	
		mparte Templo(A)	
		necta Edificación(P)	
		necta Edificación(A)	
		mparte Edificación(A)	
		necta Calle(L)	
	Calle(L) cru	iza Cane(L)	

MANZANA (PO	LÍGONO I	(IRBANO)	
Descripción		nitada por vialidades y/o rasgos culturales, cuyo perímetro puede recorrerse	
	en su totali		
	Extensión	territorial que está constituido por un grupo de viviendas, edificios, predios,	
	lotes o terrenos de uso habitacional, comercial, industrial o de servicios, se considera		
		nidad mínima de desagregación del Marco Geoestadístico, generalmente se	
		ear caminando en su totalidad y está delimitada por calles, andadores,	
		eredas, cercas, arroyos, barrancos o límites prediales.	
Atributos	Dominio F	ijo	
	Tipo	Definida con base en las dimensiones de la entidad y a las características de las edificaciones existentes en ellas.	
	Dominio	- Camellón. Parte que separa dos vías de comunicación.	
	de	- Glorieta. Lugar donde convergen varias calles.	
	Valores	- Manzana. Conjunto de edificaciones contiguas, delimitado por calles	
		por lo menos en un lado.	
		Relaciones	
Datos	Manzana(A	A) comparte Manzana(A)	
Originales		A) comparte Edificación(A)	
	Manzana(A	A) comparte Escuela(A)	
	Manzana(A	A) conecta Hospital(P)	
	Manzana(A	A) conecta Centro comercial(P)	
		A) conecta Escuela(P)	
		A) conecta Estación del metro(P)	
		A) conecta Instalación deportiva(P)	
		A) conecta Escuela(A)	
		A) conecta Manzana(A)	
		A) conecta Templo(A)	
		A) comparte Escuela(A)	
		A) comparte Manzana(A)	
	Manzana(A	A) comparte Templo(A)	
D	N (/		
Datos		A) comparte Manzana(A)	
Generalizados		A) comparte Edificación(A)	
		A) comparte Escuela(A)	
		A) conecta Hospital(P) A) conecta Centro comercial(P)	
	,	A) conecta Centro Confectat(F) A) conecta Escuela(P)	
		A) conecta Escueia(F) A) conecta Estación del metro(P)	
		A) conecta Estacton del metro(r) A) conecta Instalación deportiva(P)	
		A) conecta Histaración deportiva(1)	
		A) conecta Manzana(A)	
		A) conecta Templo(A)	
		A) comparte Escuela(A)	
		A) comparte Manzana(A)	
		A) comparte Templo(A)	
	(1	-,	

EDIFICACIÓN	(EDIFICIO	O DE DEPARTAMENTOS)		
Descripción	Construcción permanente que sirve para vivienda, usos culturales, de servicios, etc.			
Atributos	Dominio Fijo			
	Tipo	Tipo Definida con base en las dimensiones de la entidad y a las características		
		de uso.		
	Dominio	- Aduana y/o control de migración. Módulos extraurbanos donde se		
	de	lleva a cabo la revisión legal de personas y/o mercancías que llegan o		
	Valores	salen del país.		
		- Centro de asistencia médica. Para proporcionar servicios médicos.		
		- Escuela. Para la enseñanza oficial de nivel básico a medio superior.		
		- Estación del metro. Para el ascenso y descenso de pasajeros del		
		Sistema de Transporte Colectivo (Metro).		
		- Estación del tren ligero. Para el ascenso y descenso de pasajeros del		
		Sistema de Transporte Colectivo (Tren Ligero).		
		- Granja o Establo. Para la reproducción, cría industrial y explotación		
		de especies animales Pirámide. Construcción prehispánica.		
		 Firallide. Construcción prehispanica. Sitio histórico. Lugar donde ocurrió un suceso histórico de relevancia 		
		nacional.		
		- Templo. Para la celebración de un culto religioso.		
		- Otro. El valor del atributo es diferente de los valores citados		
		anteriormente.		
		Relaciones		
Datos	Edificació	n(A) comparte Manzana(A)		
Originales				
Datos	Edificación(A) comparte Manzana(A)			
Generalizados	Edificación(A) comparte Edificación(A)			
	Edificación(A) conecta Edificación(P)			
	Edificación(A) cruza Edificación(A)			
		Manzana(A) conecta Estación del metro(P)		
		Manzana(A) conecta Instalación deportiva(P)		
	Edificación(A) conecta Templo(P)			
	Edificación	n(A) comparte Templo(A)		

CENTRO DE A	ASISTENCIA MÉDICA (HOSPITAL)		
Descripción	Establecimientos dedicados a brindar servicios de diagnóstico, tratamiento y rehabilitación a pacientes.		
Atributos	Dominio Fijo		
	Tipo Clasificación que se da al centro de asistencia médica en función de los servicios que ofrece.		
	Dominio de Hospital. Establecimiento donde se realizan operaciones quirúrgicas y recuperación física de los pacientes. Valores - Unidad Médica. Local destinado a otorgar consulta a pacientes ambulatorios. - Unidad de Urgencias. Instalación para la atención inmediata del paciente en estado crítico de salud.		
	Relaciones		
Datos Originales	Hospital(P) conecta Calle(L) Hospital(A) conecta Calle(L) Hospital(A) comparte Calle(L) Hospital(A) comparte Hospital(A) Hospital(A) comparte Escuela(A) Hospital(A) comparte Manzana(A) Hospital(A) comparte Templo(A)		
Datos Generalizados	Hospital(P) conecta Calle(L) Hospital(A) conecta Calle(L) Hospital(A) comparte Calle(L) Hospital(A) comparte Hospital(A) Hospital(A) comparte Escuela(A) Hospital(A) comparte Manzana(A) Hospital(A) comparte Templo(A) Hospital(A) cruza Calle(L) Hospital(A) conecta Manzana(P)		

B. Archivo instancias.xml

```
<?xml version="1.0" encoding="ISO-8859-1"?>
<instancias>
<calleg01>calleg01</calleg01>
<calleg06>calleg06</calleg06>
<calleg03>calleg03</calleg03>
<unidadhabitg07>unidadhabitg07</unidadhabitg07>
<unidadhabitg03>unidadhabitg03</unidadhabitg03>
<unidadhabitg05>unidadhabitg05</unidadhabitg05>
<unidadhabitg10>unidadhabitg10</unidadhabitg10>
<unidadhabitg01>unidadhabitg01</unidadhabitg01>
<unidadhabitg08>unidadhabitg08</unidadhabitg08>
<unidadhabitg04>unidadhabitg04</unidadhabitg04>
<unidadhabitg06>unidadhabitg06</unidadhabitg06>
<unidadhabitg02>unidadhabitg02</unidadhabitg02>
<unidadhabitg09>unidadhabitg09</unidadhabitg09>
<escuela4540>escuela4540</escuela4540>
<escuela703>escuela703</escuela703>
<escuela710>escuela710</escuela710>
<edificio936>edificio936</edificio936>
<edificio927>edificio927</edificio927>
<edificio935>edificio935</edificio935>
<edificio3133>edificio3133</edificio3133>
<edificio3131>edificio3131</edificio3131>
<edificio3129>edificio3129</edificio3129>
<edificio3125>edificio3125</edificio3125>
<edificio3127>edificio3127</edificio3127>
<edificio939>edificio939</edificio939>
<edificio3132>edificio3132</edificio3132>
<edificio943>edificio943</edificio943>
<edificio934>edificio934</edificio934>
<hospital609>hospital609</hospital609>
<manzana16640>manzana16640</manzana16640>
<manzana4720>manzana4720</manzana4720>
<manzana9279>manzana9279</manzana9279>
<manzana2498>manzana2498</manzana2498>
<manzana9095>manzana9095</manzana9095>
<manzana4727>manzana4727</manzana4727>
<manzana4729>manzana4729</manzana4729>
<manzana4738>manzana4738</manzana4738>
<manzana9318>manzana9318</manzana9318>
<manzana16618>manzana16618</manzana16618>
<manzana9286>manzana9286</manzana9286>
<manzana4714>manzana4714</manzana4714>
<manzana16958>manzana16958
<avenida5494>avenida5494</avenida5494>
<escuelag01>escuelag01</escuelag01>
<escuelag03>escuelag03</escuelag03>
<escuelag02>escuelag02</escuelag02>
<manzana16640>manzana16640</manzana16640>
<manzana4720>manzana4720</manzana4720>
<manzana4729>manzana4729</manzana4729>
<manzana4738>manzana4738</manzana4738>
<manzana9318>manzana9318</manzana9318>
<manzana16618>manzana16618</manzana16618>
<manzana9286>manzana9286</manzana9286>
<manzana4714>manzana4714</manzana4714>
<manzana16620>manzana16620</manzana16620>
<manzana16553>manzana16553</manzana16553>
<manzana9301>manzana9301</manzana9301>
<manzana16617>manzana16617</manzana16617>
</instancias>
```

C. Archivo relaciones.xml

```
<?xml version="1.0" encoding="ISO-8859-1"?>
<relaciones>
<rel_manzana_equivale_manzana_29><_destino_>d_smanzanag01</_destino_><_origen_>o_manzana4737</_origen_><_accion_</pre>
>a rel topo equivale</a> accion ></rel manzana equivale manzana 29>
<rel_calle_cruza_avenida_02><_origen_>o_calleg03</_origen_><_destino_>d_avenidag01</_destino_><_accion_>a_rel_topo_cru
za_completamente</_accion_></rel_calle_cruza_avenida_02>
<rel calle cruza_avenida_01><_origen_>o_calle7698</_origen_><_accion_>a_rel_topo_cruza</_accion_><_destino_>d_avenida
5494</ destino ></rel calle cruza avenida 01>
<rel_calle_cruza_avenida_03><_destino_>d_avenida5494</_destino_><_accion_>a_rel_topo_cruza</_accion_><_origen_>o_calle
7705</_origen_></rel_calle_cruza_avenida_03>
\label{lem:condition} $$ \end{conjugate} $$ \end{
rigen >o calleg04</ origen ></rel calle cruza avenida 04>
<rel_calle_conecta_manzana_36><_destino_>d_smanzanag08</_destino_><_accion_>a_rel_topo_conecta</_accion_><_origen_>
o_calleg06</_origen_></rel_calle_conecta_manzana_36>
<rel calle conecta manzana 17>< accion >a rel topo conecta</ accion >< destino >d smanzanag08</ destino >< origen >
o calleg02</ origen ></rel calle conecta manzana 17>
<rel_calle_conecta_manzana_01><_origen_>o_calle7696</_origen_><_accion_>a_rel_topo_conecta</_accion_><_destino_>d_ma
nzana16557</_destino_></rel_calle_conecta_manzana_01>
<rel_calle_conecta_manzana_27><_origen_>o_calle7705</_origen_><_destino_>d_manzana9301<destino_>d_manzana9301
po conecta</ accion ></rel calle conecta manzana 27>
<rel_calle_conecta_manzana_24><_destino_>d_manzana2791</_destino_><_accion_>a_rel_topo_conecta</_accion_><_origen_>
o_calle7705</_origen_></rel_calle_conecta_manzana_24>
<rel_calle_conecta_manzana_34><_origen_>o_calleg04</_origen_><_destino_>d_smanzanag03</_destino_><_accion_>a_rel_top
o conecta</ accion ></rel calle conecta manzana 34>
<rel calle conecta manzana 15>< origen >o calleg03</ origen >< destino >d smanzanag03</ destino >< accion >a rel top
o_conecta</_accion_></rel_calle_conecta_manzana_15>
<rel_calle_conecta_manzana_09><_origen_>o_calle7689</_origen_><_accion_>a_rel_topo_conecta</_accion_><_destino_>d_ma
nzana16559</ destino_></rel_calle_conecta_manzana_09>
<rel_calle_conecta_manzana_31><_accion_>a_rel_topo_conecta</_accion_><_destino_>d_smanzanag06</_destino_><_origen_>
o calleg04</ origen ></rel calle conecta manzana 31>
<rel calle conecta manzana 25>< origen >o calle7696</ origen >< destino >d manzana16971</ destino >< accion >a rel t
opo_conecta</_accion_></rel_calle_conecta_manzana_25>
<rel_calle_conecta_manzana_22><_origen_>o_calle7698</_origen_>< accion_>a_rel_topo_conecta</_accion_><_destino_>d_ma
nzana9322</ destino ></rel calle conecta manzana 22>
<rel_calle_conecta_manzana_18><_destino_>d_smanzanag09</_destino_><_origen_>o_calleg06</_origen_><_accion_>a_rel_top
o_conecta</_accion_></rel_calle_conecta_manzana_18>
<rel calle conecta_manzana_13><_accion_>a_rel_topo_conecta/_accion_><_destino_>d_smanzanag07/_destino_><_origen_>
o_calleg04</_origen_></rel_calle_conecta_manzana_13>
<rel_calle_conecta_manzana_06><_destino_>d_manzana9286</_destino_><_accion_>a_rel_topo_conecta</_accion_><_origen_>
o_calle7698</_origen_></rel_calle_conecta_manzana_06>
<rel calle conecta manzana 23>< destino >d manzana9322</ destino >< origen >o calle7705</ origen >< accion >a rel to
po_conecta</_accion_></rel_calle_conecta_manzana_23>
<rel_calle_conecta_manzana_16><_accion_>a_rel_topo_conecta</_accion_><_origen_>o_calleg04</_origen_><_destino_>d_sma
nzanag04</_destino_></rel_calle_conecta_manzana_16>
<rel calle conecta_manzana_32><_accion_>a_rel_topo_conecta</_accion_><_origen_>o_calleg05</_origen_><_destino_>d_sma
nzanag06</ destino ></rel calle conecta manzana 32>
<rel_calle_conecta_manzana_11><_origen_>o_calleg02</_origen_><_destino_>d_smanzanag05</_destino_><_accion_>a_rel_top
o_conecta</_accion_></rel_calle_conecta_manzana_11>
<rel_calle_conecta_manzana_04><_destino_>d_manzana2791</_destino_><_origen_>o_calle5495</_origen_><_accion_>a_rel_to
po conecta</ accion ></rel calle conecta manzana 04>
<rel calle conecta manzana 03>< accion >a rel topo conecta</a> accion >< destino >d manzana9322</ destino >< origen >
o_calle5495</_origen_></rel_calle_conecta_manzana_03>
<rel_calle_conecta_manzana_33><_origen_>o_calleg05</_origen_><_destino_>d_smanzanag07</_destino_><_accion_>a_rel_top
o conecta</ accion ></rel_calle_conecta_manzana_33>
<rel calle conecta manzana 29>< accion >a rel topo conecta</ accion >< origen >o calleg03</ origen >< destino >d sma
nzanag05</ destino ></rel calle conecta manzana 29>
<rel_calle_equivale_calle_06><_accion_>a_rel_topo_equivale</_accion_><_destino_>d_calleg06</_destino_><_origen_>o_calle7
712</ origen ></rel calle equivale calle 06>
<rel calle equivale calle 01>< accion >a rel topo equivale
accion >< destino >d calleg01
destino >< origen >o calle7
689</ origen ></rel calle equivale calle 01>
</relaciones>
```