



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



**Similitud semántica entre sistemas de objetos geográficos  
aplicada a la generalización de datos geoespaciales**

**T E S I S**

QUE PARA OBTENER EL GRADO DE

**DOCTOR EN CIENCIAS DE LA COMPUTACIÓN**

PRESENTA:

**MARCO ANTONIO MORENO IBARRA**

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

MÉXICO, D.F.

Octubre 2007

# AGRADECIMIENTOS

Gracias a mi Mamá, por ser un ejemplo cada día, por su esfuerzo, por dejarme ser, por escucharme, por comprenderme, por alentarme, por tu sonrisa y sobre todo por ser mi mamá.

Gracias a mis abuelos Manuela† y Roberto, por enseñarme a valorar las cosas realmente importantes de la vida, por cuidar de mi y por tanto cariño.

Gracias a mis abuelos Carolina y Rafael†, por ser tan buenos conmigo y enseñarme una parte de mí que no conocía.

Gracias a mi tíos Betón, Can, Chio, Tere, Rich, Ara, Marthita, Jesús†, Bernardo, Paty y Benjamín, a mis queridos primos Yaya, Pau, Ara, Tere, Beto, Berni, Chucho, Wendy, Rocío, Saúl, Chomín, Manuel y Hugo, a mis sobrinos Santiago, Ariadna y Sebastián. A todos, gracias por estar ahí, por estar pendientes, por crecer juntos y ser una de las partes más importantes de mi vida. Los quiero.

Gracias Rosario, por estar cerca de mí y por el enorme valor de un abrazo y un beso en el momento preciso. Te adoro.

Gracias al Dr. Serguei Levachkine por sus invaluable consejos, por compartir sus ideas y conocimientos cada día, por su paciencia, por su ánimo, por las enseñanzas, por su preocupación diaria por mejorar el trabajo, por los regaños oportunos, por las pláticas interminables, por hacer un grupo que se consolida cada día más, por las oportunidades y por guiarme en el camino de la investigación. Mi admiración y agradecimiento infinito.

Gracias a mis Sinodales, Dr. Felipe Riverón, Dr. Menchaca García, Dr. Velázquez, Dr. Pogrebnyak y Dr. Guevara por sus comentarios, soporte y ayuda para llevar a buen término esta tesis. Tomaré sus consejos en cuenta cada día.

Gracias a mis queridos hermanos bonobos del PIIG, Miguel, Rolando y Giovanni, por compartir tantos momentos, las risas imparables en el GIScience, los viajes a SAGARPA, la tarjeta telefónica en Inglaterra, las desveladas, los CIARP, por las bromas, por los chistes, las canciones, por los apodos, los seminarios y las clases. Pero sobre todo gracias por su amistad, por la ayuda diaria y por hacer equipo. Los quiero entrañablemente.

Gracias a Citlalli, Omar, Chad, Melissa, Rolas Menchaca, Reyna, Epifanio, Doña Reyna, Inés, Grace y Malú por su amistad y por estar ahí en el momento más oportuno. En especial quisiera agradecer a mi hermana Nic y a mi cuñado Erick por estar cerca en los momentos importantes.

Gracias a todos los PIIGs. Benjamín, Edgar, Félix, Gerardo, Jonathan, Karina, Miguel, Paola, Roberto, Victor y Virginia, por siempre colaborar conmigo, por su amistad y apoyo.

Gracias a las personas cuyos consejos han influido positivamente mi vida académica, Dr. Viniegra, Dr. Polshkov†, Dr. Fonseca, Dr. Chema y Dr. Díaz de León.

Gracias a todas las personas que trabajan en el CIC que de alguna u otra forma em han ayudado, en especial quiero agracer el soporte de Martha Durán, Elda y Carmen.

Gracias al Centro de Investigación en Computación y al Instituto Politécnico Nacional, por haber facilitado los medios finalizar la presente tesis.

Gracias a TODOS.

# ÍNDICE

---

RESUMEN .....	I
ABSTRACT .....	III
AGRADECIMIENTOS .....	V
ÍNDICE .....	IX
ÍNDICE DE FIGURAS .....	XIII
ÍNDICE DE TABLAS .....	XVII
<b>CAPÍTULO 1 INTRODUCCIÓN.....</b>	<b>1</b>
1.1 OBJETIVO.....	2
1.2 METAS.....	3
1.3 HIPÓTESIS .....	4
1.4 ALCANCES .....	5
1.5 ORGANIZACIÓN DEL DOCUMENTO.....	6
<b>CAPÍTULO 2 ESTADO DEL ARTE .....</b>	<b>7</b>
2.1 GENERALIZACIÓN CARTOGRÁFICA.....	7
2.1.1 <i>Primeros intentos de automatización</i> .....	8
2.1.2 <i>Modelos de Generalización</i> .....	8
2.1.2.1 Modelo de Ratajski .....	9
2.1.2.2 Modelo de Brassel and Weibel.....	9
2.1.2.3 Modelo de McMaster y Shea.....	11
2.1.3 <i>Los sistemas basados en conocimiento para la Generalización</i> .....	14
2.1.4 <i>La Generalización con sistemas multiagentes</i> .....	17
2.1.4.1 El proyecto AGENT .....	18
2.1.5 <i>Otros tópicos en la Generalización</i> .....	20
2.1.5.1 Algoritmos de la generalización.....	21
2.1.5.2 Manipulación de los efectos secundarios .....	21

2.1.5.3 Semántica en la Generalización.....	22
2.1.4.4 Restricciones.....	23
2.1.6 <i>Ontologías en la Generalización</i> .....	24
2.2 REPRESENTACIÓN DEL CONOCIMIENTO CARTOGRÁFICO .....	26
2.2.1 <i>Semiótica</i> .....	27
2.2.2 <i>Mapas de Aspectos</i> .....	27
2.3 ONTOLOGÍAS .....	31
2.3.1 <i>Thomas R. Gruber</i> .....	31
2.3.2 <i>Nicola Guarino</i> .....	34
2.4 DISCUSIÓN SOBRE EL ESTADO DEL ARTE .....	42
<b>CAPÍTULO 3 MARCO TEÓRICO.....</b>	<b>45</b>
3.1 TOPOLOGÍA GEOESPACIAL.....	45
3.1.1 <i>Estructura vectorial</i> .....	46
3.1.2 <i>Representación de los datos vectoriales</i> .....	48
3.1.3 <i>Conceptos básicos de la teoría de grafos para analizar las relaciones entre los objetos geográficos</i> .....	48
3.1.3.1 Relaciones entre los elementos de un grafo plano.....	49
3.1.3.1.1 Relaciones entre nodos y arcos.....	50
3.1.3.1.2 Relaciones entre arcos y caras.....	50
3.1.3.1.3 Relaciones entre nodos y caras.....	51
3.1.3.1.4 Relaciones entre cara - cara.....	51
3.1.4. <i>Descripción de objetos geográficos usando el modelo vectorial</i> .....	52
3.1.4.1 Relaciones topológicas en datos vectoriales.....	53
3.1.5 <i>Componentes topológicos</i> .....	54
3.1.6 <i>Modelo de 9-intersección</i> .....	56
3.2 MEDIDAS SOBRE DATOS GEOESPACIALES.....	58
3.2.1 <i>Selección de la medida adecuada</i> .....	59
3.2.2 <i>Tipos de Medidas</i> .....	60
3.2.3 <i>Clasificación de medidas</i> .....	61
3.3 JERARQUÍAS Y TEORÍA DE LA CONFUSIÓN .....	62
3.3.1 <i>Concepto de Jerarquía</i> .....	62
3.3.1.1 Jerarquía simple, ordenada, porcentual y mixta .....	65
3.3.2 <i>Teoría de la Confusión</i> .....	66
3.3.2.1 Confusión de usar r en vez de s para jerarquías simples .....	66
3.3.2.2 Confusión de usar r en vez de s, para jerarquías ordenadas .....	66

3.3.2.3 Confusión de usar <i>r</i> en vez de <i>s</i> , para jerarquías porcentuales .....	67
3.3.2.4 Confusión de usar <i>r</i> en vez de <i>s</i> , para jerarquías mixtas.....	67
3.3.2.5 El conjunto de valores que son iguales a otros, dada una confusión.....	68
3.3.2.5.1 Consultas usando la confusión .....	68
3.3.2.5.2 Objetos idénticos, muy similares, algo similares. ....	68
<b>CAPÍTULO 4 METODOLOGÍA .....</b>	<b>71</b>
4.1 LA CONCEPTUALIZACIÓN DEL DOMINIO TOPOGRÁFICO POR UN EXPERTO .....	75
4.1.1 <i>Propiedades de los datos geográficos</i> .....	75
4.1.1.1 Propiedades geométricas .....	75
4.1.1.2 Propiedades topológicas .....	76
4.1.1.3 Propiedades temáticas .....	78
4.1.2 <i>Diseño de la ontología</i> .....	79
4.1.2.1 Objetivo de la ontología .....	80
4.1.2.2 Análisis de las componentes del dominio a conceptualizar .....	82
4.1.2.2.1 Ríos y redes hidrológicas.....	82
4.1.2.2.1.1 Clases de ríos.....	84
4.1.2.2.1.2 Características de las redes hidrológicas .....	85
4.1.2.2.1.3 Representación de las redes hidrológicas.....	86
4.1.2.2.2 Curvas de nivel .....	87
4.1.2.2.3 Relación entre las curvas de nivel y los ríos.....	89
4.1.2.3 Términos de la ontología .....	90
4.1.2.4 Las relaciones entre las clases .....	93
4.1.2.5 Propiedades de las clases.....	96
4.1.2.6 Mantenimiento de la ontología.....	99
4.2 NORMALIZACIÓN .....	102
4.2.1 <i>Consistencia de la base de datos</i> .....	102
4.2.2 <i>Asignación automática de identificadores</i> .....	105
4.3 ANÁLISIS .....	107
4.3.1 <i>Medición de la Sinuosidad</i> .....	109
4.3.2 <i>Medición de la relación &gt;Pasa-por&lt;</i> .....	111
4.4 SÍNTESIS .....	114
4.4.1 <i>Generación de la tabla de propiedades conceptualmente representadas para la sinuosidad de los ríos</i> .....	119
4.4.2 <i>Generación de la tabla de relaciones conceptualmente representadas para la relación &gt;Pasa-por&lt; entre los ríos y las curvas de nivel</i> .....	120

4.5 EL SISTEMA DE GENERALIZACIÓN .....	122
4.6 VERIFICACIÓN .....	125
4.6.1 <i>Medición de la confusión entre relaciones</i> .....	127
4.6.2 <i>Medición de la confusión entre propiedades</i> .....	128
4.6.3 <i>Casos de verificación de (in)consistencias</i> .....	129
4.6.3.1 Igual.....	129
4.6.3.2 Desigual.....	130
4.6.3.3 Equivalente.....	131
4.6.4 <i>Verificación de la consistencia entre tablas de propiedades conceptualmente representadas</i> .....	133
4.6.5 <i>Verificación de la consistencia entre tablas de relaciones conceptualmente representadas</i> .....	134
4.6.6 <i>Medición de la similitud de la Sinuosidad entre dos representaciones conceptuales</i> .....	136
4.6.7 <i>Verificación de la consistencia de la relación Pasa-por representada conceptualmente</i> .....	137
4.6.8 <i>Medición de la similitud a nivel global</i> .....	138
4.6.8.1 Error acumulado .....	139
4.6.8.2 Error Medio .....	139
4.6.8.3 <i>La medición de la similitud y su relación con las preferencias de los usuarios</i> .....	140
<b>CAPÍTULO 5 PRUEBAS Y RESULTADOS .....</b>	<b>143</b>
5.1 EL SISTEMA DE GENERALIZACIÓN .....	144
5.1.1 <i>Generalización de las Redes Hidrológicas</i> .....	144
5.1.2 <i>Generalización de las curvas de nivel</i> .....	152
5.2 GENERACIÓN DE LA BASE DE DATOS ESPACIAL PARCIALMENTE CONCEPTUALIZADA .	154
5.3 VERIFICACIÓN DE LA CONSISTENCIA DE LOS DATOS GENERALIZADOS .....	159
5.4 VERIFICACIÓN DE LA CONSISTENCIA A NIVEL GLOBAL .....	165
<b>CAPÍTULO 6 CONCLUSIONES Y TRABAJO FUTURO .....</b>	<b>171</b>
6.1 CONCLUSIONES.....	171
6.2 CONTRIBUCIONES .....	174
6.3 TRABAJO FUTURO .....	177
<b>REFERENCIAS .....</b>	<b>179</b>
<b>ANEXO I CÓDIGO FUENTE .....</b>	<b>189</b>

# ÍNDICE DE FIGURAS

---

## CAPÍTULO 2 ESTADO DEL ARTE

Figura 2.1 Modelo de Bressel y Weibel .....	10
Figura 2.2 Modelo de McMaster y Shea .....	11
Figura 2.3 Operadores de Generalización de Shea y McMaster .....	13
Figura 2.4 Componentes básicos de un sistema experto .....	14
Figura 2.5 Interacción de agentes .....	17
Figura 2.6 Ciclo de vida de un agente .....	20
Figura 2.7 Simplificación de líneas usando DMin .....	26
Figura 2.8 Fragmento de la red del Metro de Hamburgo .....	29
Figura 2.9 Ejemplo de estructura jerárquica para la representación de la correspondencia.	30
Figura 2.10 Ejemplo de jerarquía de precedencia pictórica .....	30
Figura 2.11 Los <i>modelos previstos</i> de un lenguaje reflejan el compromiso de la conceptualización. Una ontología indirectamente refleja el compromiso (y la conceptualización) aproximando este conjunto de modelos previstos.....	39
Figura 2.12 Tipos de ontologías de acuerdo a su nivel de dependencia para una tarea en particular (Las líneas representan relaciones de especialización) .....	41

## CAPÍTULO 3 MARCO TEÓRICO

Figura 3.1 Componentes del modelo vectorial.....	47
Figura 3.2 Datos geográficos en un GIS .....	48
Figura 3.3. Ejemplo de un grafo.....	49
Figura 3.4 Conexión geométrica de dos caras.....	51
Figura 3.5 Topología de conjunto de puntos en $\mathbb{R}^2$ .....	55
Figura 3.6 Matriz de 9-intersección.....	57
Figura 3.7 Relaciones topológicamente similares.....	57
Figura 3.8 Clasificación de medidas para datos geográficos.....	62

## CAPÍTULO 4 METODOLOGÍA

Figura 4.1. Esquema de la metodología a desarrollar en esta tesis .....	74
Figura 4.2 Relación Conecta .....	77
Figura 4.3 Relación Comparte.....	78



Figura 4.4 Características de representación del INEGI para las corrientes de agua en escala 1:50,000 .....	79
Figura 4.5 Esquema simplificado de una cuenca .....	83
Figura 4.6 Tipos de drenaje .....	84
Figura 4.7 Configuraciones de representación para redes hidrológicas. a) línea-línea, b) área-área, c) mixta .....	86
Figura 4.8 Curvas de nivel.....	87
Figura 4.9 Propiedades de las curvas de nivel; a) Todas las curvas de nivel forman áreas cerradas, b) Superposición de curvas de nivel; c). Las curvas de nivel no se bifurcan; d) Representación de una corriente de agua relacionada con las curvas de nivel.....	88
Figura 4.10 Comparación entre curvas de nivel a diferentes escalas .....	89
Figura 4.11 Relación entre los ríos y las curvas de nivel .....	89
Figura 4.12 Casos de relaciones entre las curvas de nivel y los ríos.....	90
Figura 4.13 Jerarquía de conceptos con las relaciones es-un y tipo-de .....	93
Figura 4.14 Representación gráfica de los casos de la relación <i>&gt;Pasa-por&lt;</i> ; a) pasa por máxima convexidad, b) casi pasa por máxima convexidad, c) pasa por convexidad, d) pasa por recta, e) pasa por concavidad, f) casi pasa por máxima concavidad, g) pasa por máxima concavidad.....	94
Figura 4.15 Representación jerárquica de la relación <i>&gt;Pasa-por&lt;</i> .....	95
Figura 4.16 Representación de la ontología, al considerar la relación <i>&gt;Pasa-por&lt;</i> .....	95
Figura 4.17 Representación jerárquica de la propiedad <i>&lt;Sinuosidad&gt;</i> .....	98
Figura 4.18 Representación de la ontología, al considerar la relación <i>&gt;Pasa-por&lt;</i> y la propiedad <i>&lt;Sinuosidad&gt;</i> .....	98
Figura 4.19 Casos generales para verificar consistencia durante la etapa de normalización .....	104
Figura 4.20 Identificación de la dirección del flujo de una red hidrológica. Las Líneas con flechas indican la dirección del flujo, las áreas verde oscuras representan la zona con menor elevación.....	104
Figura 4.21 Clasificación jerarquizada por longitud (CLAJER).....	105
Figura 4.22 Asignación de identificadores a los objetos en las redes hidrológicas.....	106
Figura 4.23 Asignación de identificadores a las curvas de nivel.....	106
Figura 4.24 Sinuosidad de un río.....	110
Figura 4.25 Componentes para medir la Sinuosidad en un río.....	111

Figura 4.26 Análisis topológico para identificar el grado de concavidad/convexidad por el que pasa un río a través de una curva de nivel .....	112
Figura 4.27 Determinación de la concavidad/convexidad en una curva de nivel .....	113
Figura 4.28 Equivalencia entre la representación gráfica de la relación <i>&gt;Pasa-por&lt;</i> entre un río y una curvas de nivel con respecto al valor de DCP. Las líneas punteadas representan ríos, las líneas sólidas representan curvas de nivel.....	114
Figura 4.29 Esquema general del sistema de generalización automática.....	123
Figura 4.30 Datos generados por el sistema de generalización automático .....	124
Figura 4.31 Esquema general de la etapa de verificación .....	126
Figura 4.32 Representación gráfica de “igual” para la relación “ <i>está al W</i> ” .....	129
Figura 4.33 Medición de $conf(r,s)=0$ , sobre una jerarquía de <i>ubicación relativa</i> .....	130
Figura 4.34 Representación gráfica de “desigual” .....	130
Figura 4.35 Medición de $conf(r,s)= 4/16$ , sobre una jerarquía de la <i>ubicación relativa</i> ....	131
Figura 4.36 Representación gráfica de “ <i>equivalente</i> ” .....	131
Figura 4.37 Medición de $conf(r,s)= 1/15$ , con $u=1/15$ , sobre una jerarquía de la <i>ubicación relativa</i> .....	132

## CAPÍTULO 5 PRUEBAS Y RESULTADOS

Figura 5.1 Corrección automática del flujo para una red hidrológica (las puntas de las flechas indican la dirección del flujo). a) Antes de la corrección automática. b) Después de la corrección automática.....	146
Figura 5.2 Corrección interactiva del flujo para una red hidrológica.....	147
Figura 5.3 Clasificación por subsistemas, donde se muestran los identificadores para cada subsistema.....	148
Figura 5.4 Clasificación por la función de cada arco en la red hidrológica .....	148
Figura 5.5. Clasificación de la red hidrológica usando CLAJER.....	149
Figura 5.6 Asignación de identificadores, usando CLAJER .....	150
Figura 5.7 Capa de los ríos generalizada.....	151
Figura 5.8 Fragmento de la tabla <i>Rios-1.aat</i> y su representación gráfica .....	152
Figura 5.9 Generalización de las curvas de nivel .....	154
Figura 5.10 Capa combinada de los ríos y las curvas de nivel.....	156
Figura 5.11 Fragmentos donde se identificó el <i>AAR</i> .....	157
Figura 5.12 Extracción de arcos que participan en la relación <i>&gt;Pasa-por&lt;</i> .....	157

Figura 5.13 Datos generalizados e identificación visual de inconsistencias .....	160
Figura 5.14 Verificación desigual de consistencia en datos generalizados, caso igual.....	163
Figura 5.15 Verificación desigual de consistencia en datos generalizados, caso desigual	164
Figura 5.16 Verificación desigual de consistencia en datos generalizados, caso equivalente .....	165
Figura 5.17 Datos fuente y datos generalizados.....	166
Figura 5.18 Interface en ArcView para visualización de resultados .....	168

# ÍNDICE DE TABLAS

---

## CAPÍTULO 4 METODOLOGÍA

Tabla 4.1. Conceptos que conforman la ontología .....	92
Tabla 4.2. Propiedades de las clases de la ontología .....	96
Tabla 4.3. Mediciones a utilizar .....	109
Tabla 4.4. Restricciones para pertenecer a las subclases <Sinuosidad>. ....	119
Tabla 4.5. Restricciones para pertenecer a las subclases de >Pasa-Por< .....	121
Tabla 4.6. El error medio y su relación con los requerimientos de diversos usuarios.....	142

## CAPÍTULO 5 PRUEBAS Y RESULTADOS

Tabla 5.1. Tabla de relaciones conceptualmente representadas de datos generalizados .....	160
Tabla 5.2. Tabla de relaciones conceptualmente representadas después de identificar inconsistencias sobre los datos generalizados.....	161
Tabla 5.3. Tabla de relaciones conceptualmente representadas de los datos fuente .....	163
Tabla 5.4. Tabla de relaciones semánticamente consistentes.....	167

# CAPÍTULO 1

## INTRODUCCIÓN

Hoy en día, las ontologías juegan un papel muy importante en el desarrollo de las ciencias geoespaciales, ya que permiten describir explícitamente las propiedades de los objetos geográficos, por medio de las relaciones que existen entre diferentes conceptos, representan objetos geoespaciales que van desde nombres de las ciudades hasta una red hidrológica, entre otras aplicaciones. Cada día el uso de las denominadas *ontologías espaciales* se expande hacia los campos que tradicionalmente fueron tratados por el procesamiento numérico, por ejemplo la cartografía digital y los Sistemas de Información Geográfica (GIS).

Por otro lado, la cartografía digital es la base para desarrollar los GIS, los cuales proporcionan mecanismos para almacenar, manipular, construir y analizar datos geográficos. Sin embargo, existen ciertos problemas relacionados con el análisis, manipulación y representación de datos geográficos que no han podido resolverse en su totalidad utilizando los mecanismos computacionales tradicionales (por ejemplo, la generalización). Esto se debe a que la representación (representación cartográfica) requiere de la experiencia que posea un usuario experto en este campo y sustenta las tareas de elaboración, análisis e interpretación de los datos geoespaciales. Los cartógrafos y usuarios de datos geográficos hacen uso de recursos visuales tales como color, forma y patrones para comunicar la información sobre los objetos geográficos y las relaciones espaciales que existen entre ellos. Están orientados a representar la geometría (incluyendo la topología) de los objetos o sistemas de objetos geográficos y actualmente están bien desarrollados desde el punto de vista computacional. Sin embargo, en los sistemas computacionales enfocados a la manipulación de datos geoespaciales modernos se requiere ir más allá, es decir, las diferentes descripciones e interpretaciones deben converger en un punto común. Este

problema es de mucha importancia debido a que hay una gran variedad de “geoproductores”, es decir, las entidades que producen la información geoespacial en diferentes formatos, bajo diferentes estándares, etc. Éstas raras veces se ponen de acuerdo en la interpretación de ¿qué es? por ejemplo un “área verde”. Por lo tanto, se requieren mecanismos automáticos capaces de interpretar las discrepancias entre diferentes formas de representar la información geoespacial. En esta investigación se pretende generar automáticamente descripciones basadas en una representación conceptual con el fin de soportar la verificación de generalizaciones realizadas a datos geoespaciales. Todo esto sin depender de factores que tradicionalmente se han empleado para manipular datos geográficos, como la escala, sistema de referencia, etc. Para este fin, las ontologías son sumamente útiles, ya que añaden un componente semántico (las relaciones entre diferentes conceptos) que normalmente no es considerado en enfoques tradicionales y en sí misma representa un consenso entre un grupo o comunidad, lo que origina que las interpretaciones sean consistentes.

Como caso de estudio de la presente tesis se ha considerado la generalización cartográfica de las redes hidrológicas y curvas de nivel, que consiste en modificar el nivel de detalle en los datos geográficos, preservando su estructura. En otras palabras, la generalización tiene como fin modificar el nivel de detalle en los datos sin que estos pierdan la semántica, lo que significa que sean consistentes después de generalizar. Para verificar la consistencia se medirá la similitud semántica con base en representaciones conceptuales de los datos geoespaciales antes y después de generalizar. La ontología empleada será diseñada con el propósito específico de describir ciertas relaciones y propiedades muy específicas del caso de estudio y será diseñada con base en otras conceptualizaciones y los resultados de un análisis exhaustivo de las temáticas involucradas.

## **1.1 Objetivo**

Diseñar una metodología enfocada a la medición de la similitud semántica entre conjuntos de datos geoespaciales basada en una ontología del dominio geográfico, la cual se utilizará para generar una representación conceptual de los objetos geográficos y se enfoca

principalmente en aspectos geométricos y topológicos con aplicación directa a la evaluación de la generalización automática de las redes hidrológicas y curvas de nivel.

## 1.2 Metas

Las metas a alcanzadas por el desarrollo de la presente tesis son:

- Un análisis exhaustivo del estado del arte en las diversas áreas relacionadas con esta tesis.
- Un análisis detallado de las diversas propiedades y relaciones que poseen las redes hidrológicas y curvas de nivel, con el fin de soportar las diversas etapas dentro del desarrollo de esta tesis tales como definir la conceptualización, análisis, verificación, entre otras.
- Un análisis de las mediciones sobre datos geográficos con el propósito de identificar qué mediciones se requieren para este trabajo.
- Definición de una conceptualización del entorno geográfico enfocada a la generalización de las redes hidrológicas y curvas de nivel, con base en otras conceptualizaciones existentes.
- Definición e implantación de mediciones sobre datos geográficos para extraer las propiedades implícitas que poseen los mismos.
- Definición e implantación de mediciones sobre relaciones topológicas que existen entre pares de objetos geográficos.
- Definición e implantación del mapeo entre propiedades y relaciones de los datos geoespaciales representadas en forma numérica con propiedades y relaciones representadas en forma conceptual.

- Definición e implantación de representaciones conceptuales de objetos geoespaciales con base en propiedades y relaciones.
- Definición e implantación de las medidas de similitud entre dos diferentes representaciones conceptuales.
- Descripción detallada e implantación del caso de estudio enfocado hacia la generalización de las redes hidrológicas y curvas de nivel.

### 1.3 Hipótesis

Las hipótesis planteadas para esta tesis son:

- Con base en conceptualizaciones existentes, puede definirse una nueva conceptualización de propósito específico para soportar las diversas fases para realizar la medición de la similitud semántica entre dos diferentes conjuntos de datos.
- El análisis por medio de mediciones a los objetos geoespaciales puede ser empleado para extraer las propiedades implícitas de los datos geográficos.
- Con base en las mediciones realizadas sobre los datos geoespaciales se puede definir e implantar una base de datos conceptual compuesta por tablas, las cuales permitan representar mediante conceptos los aspectos descriptivos, geométricos, y topológicos. A su vez, las tablas se componen por *tuplas* denominadas *descripciones semánticas* de datos geoespaciales.
- Dicha descripción es una aproximación a la semántica que tiene la base de datos espacial, la cual facilita la interpretación y verificación de la estructura de un conjunto de datos geoespaciales.
- Las descripciones se basan en un conjunto de relaciones que presentan los objetos geográficos.



- Dichas descripciones pueden tener un alto contenido semántico, ya que representan las peculiaridades de un objeto en específico o bien de un grupo de objetos que están relacionados espacialmente, facilitando su interpretación automática.
- El alto contenido semántico de las *tuplas* de la representación conceptual (descripciones semánticas), permite emplearlas en tareas cartográficas complejas como la generalización automática, con el objetivo de evaluar la calidad en datos generalizados e identificación de errores originados por el proceso automático.
- Con base en una ontología se puede verificar la consistencia semántica de un conjunto de datos geográficos

## 1.4 Alcances

Esta sección ha sido incluida para describir desde un punto de vista global, los tópicos que se desarrollaron y aquellos que hizo falta cubrir.

La metodología propuesta se basa en una conceptualización del dominio topográfico, particularmente enfocándose a curvas de nivel y ríos. Dicha metodología está orientada a evaluar la similitud semántica de datos generalizados con respecto a los datos fuente. Para identificar la similitud, fueron desarrolladas medidas a nivel local y global. A nivel local se consideran una a una las relaciones existentes en las temáticas involucradas. Por otro lado, en las medidas a nivel global se consideran las relaciones en conjunto que permiten identificar si los datos cubren los requerimientos de ciertos grupos de usuarios.

La aplicación fue desarrollada en Arc/INFO, el cual es un software comercial para implementar Sistemas de Información Geográfica. La programación fue realizada en el lenguaje de programación de Arc/INFO (AML<sup>1</sup>). Fue elegido este entorno de desarrollo debido a que proporciona las funciones necesarias para los fines de la presente tesis, entre

---

<sup>1</sup> Lenguaje de Macros de ARC, en inglés Arc Macro Language,

las que podemos señalar, edición geométrica y topológica, operaciones de análisis espacial, además de que permite generar tablas de atributos para los datos vectoriales.

El conjunto de datos empleados fue del tipo vectorial y elaborado por el Instituto Nacional de Estadística, Geografía e Informática. Estos se encuentran georeferenciados utilizando la proyección Cónica Conforme de Lambert. Para el caso de las redes hidrológicas, nos enfocamos hacia aquellas compuestas únicamente por objetos línea, mientras que para las curvas de nivel, se seleccionaron aquellas que aseguraran consistencia topológica. En específico se utilizaron datos del estado de Puebla, por cumplir con las características adecuadas.

## **1.5 Organización del documento**

En el Capítulo 2 se presenta el estado de arte del área de investigación dentro de la cual se encuentra la presente tesis. Por otro lado, los fundamentos sobre Topología geoespacial, Medidas sobre datos geoespaciales, Jerarquías y Teoría de la Confusión son presentados en el Capítulo 3. El Capítulo 4 es la parte más importante de este documento. Es ahí donde se presenta la metodología propuesta. El Capítulo incluye la descripción detallada de cada una de las fases que componen el método propuesto. Los resultados experimentales, así como la discusión de los mismos, se presentan en el Capítulo 5; y en el Capítulo final, el 6, se exponen las conclusiones y recomendaciones para el trabajo futuro.

# CAPÍTULO 2

## ESTADO DEL ARTE

El desarrollo de la presente investigación involucra diferentes áreas de la Geocomputación. A continuación se describen algunos de los trabajos que se consideran los más relevantes en las áreas involucradas. En este caso se describirán aquellos relacionados con la generalización, la representación del conocimiento espacial y las ontologías espaciales.

### 2.1 Generalización cartográfica

El término *generalización* procede de la palabra latina *generalis*<sup>2</sup>. La raíz lingüística explica perfectamente la esencia de la generalización: la selección de lo principal, lo más importante orientado hacia un objetivo y su generalización, teniendo en cuenta la realidad en sus rasgos principales y típicos, y además las particularidades características, de acuerdo con la asignación de la base de datos geográfica, con el tema y la escala.

La Asociación Internacional de Cartografía (ICA) define la generalización cartográfica como “Selección y simplificación de los detalles de un mapa de acuerdo a la escala y/o el propósito del mismo” [ICA, 1973].

La generalización aumenta el valor cognoscitivo del mapa, ya que facilita el descubrimiento de las peculiaridades y rasgos generales que son propios de los fenómenos que se cartografían y contribuye a establecer las regularidades entre dichos fenómenos.

---

<sup>2</sup> Relacionado con todo, general.

### 2.1.1 Primeros intentos de automatización

Desde la década de los 60`s, tanto en la industria como en la academia, lucharon contra las dificultades para automatizar la generalización. Muchos de los involucrados en esta investigación no estaban convencidos de que se pudiera automatizar la generalización de forma completa; sin embargo, se hicieron algunos desarrollos enfocados a la generalización de grupos de objetos geográficos. Douglas y Peucker [Douglas y Peucker, 1973] y Deveau [Deveau, 1985], por mencionar algunos, se enfocaron a la generalización de datos lineales y desarrollaron algoritmos para este fin. Otros investigadores como Monmonier [Monmonier, 1983], Chrisman [Chrisman, 1983], Lichtner [Lichtner, 1979] y Töpfer [Töpfer y Pillewitzer, 1966] enfocaron sus investigaciones a la generalización de objetos punto y área. [Töpfer y Pillewitzer, 1966] desarrollaron la *ley radical*, que define la cantidad de objetos geográficos que se deben presentar en una escala determinada. Aunque dicha ley es empírica, se puede adecuar de acuerdo al fenómeno a cartografiar. Utilizándola, se puede estimar la densidad de elementos y el número de vértices<sup>3</sup> a representar para una escala determinada. La *ley radical* se define como:

$$n_T \approx n_S \sqrt{S_S / S_T},$$

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.

### 2.1.2 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 y como resultado se obtuvieron algunos modelos conceptuales. En esta sección se mencionarán algunos que se consideran los más relevantes.

<sup>3</sup> 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.

### 2.1.2.1 Modelo de Ratajski

Fue desarrollado por el cartógrafo polaco Lech Ratajski [Ratajski, 1967]. En él se identifican dos componentes fundamentales de la generalización:

- Generalización cuantitativa, que consiste en la reducción gradual del contenido del mapa en dependencia del 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 representación<sup>4</sup>.

### 2.1.2.2 Modelo de Brassel and Weibel

Es uno de los modelos más detallados que se han desarrollado. En él se identifican cinco procesos en un ambiente digital, así que se aplica específicamente en la generalización automática. De acuerdo a [Brassel y Weibel, 1988] los procesos que se siguen durante la generalización son:

- *Reconocimiento de la estructura,*
- *Proceso de identificación,*
- *Proceso de modelado,*
- *Proceso de ejecución y*
- *Visualización de datos.*

*El Reconocimiento de la estructura* es el proceso encargado de identificar grupos de objetos y sus relaciones espaciales. Este proceso es controlado por los objetivos de la generalización (calidad de la base de datos, escala destino y reglas de comunicación). El *proceso de identificación* identifica las funciones a generalizar y sus parámetros. En el

---

<sup>4</sup> Esto implica transformaciones de un objeto primeramente representado como área, en un nuevo objeto representado como punto o línea.

*proceso de modelado* se compilan las funciones en una biblioteca de procesos, y la generalización se realiza en el *proceso de ejecución*. La *visualización de datos* 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, relacionado con 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.

La principal contribución de este modelo es la distinción de los pasos que caracterizan el contenido y estructura de la base de datos fuente (pasos a, b, y c) de los pasos mecánicos operacionales (d y e). (Ver Figura 2.1)

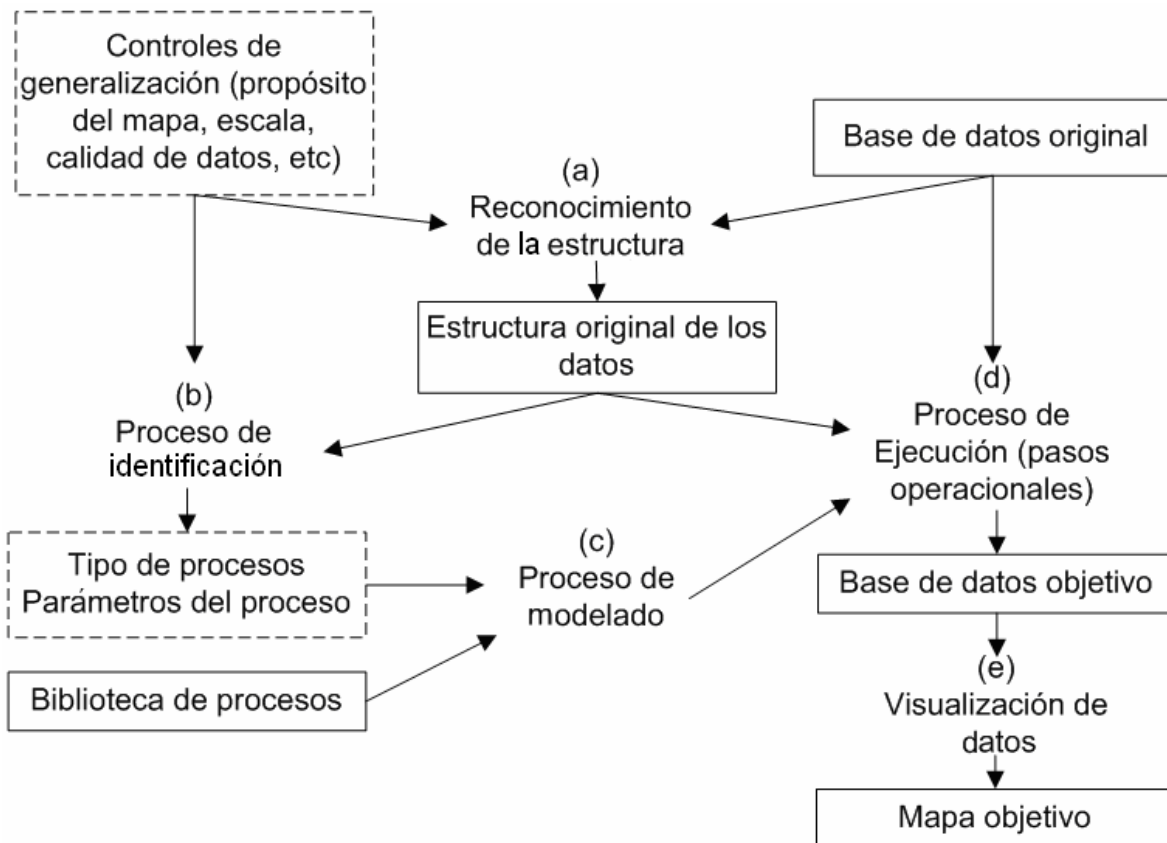


Figura 2.1 Modelo de Bressel y Weibel

### 2.1.2.3 Modelo de McMaster y Shea

Este modelo fue el primero que se elaboró considerando los objetivos filosóficos de la generalización y los operadores de generalización para emular las técnicas de generalización manual [Shea y McMaster, 1989]. En dicho modelo la generalización fue descompuesta en tres áreas operacionales:

1. Consideración de los objetivos filosóficos de por qué generalizar,
2. Evaluación cartométrica de las condiciones sobre dónde generalizar, y
3. Selección de las transformaciones de atributos y espaciales, las cuales proporcionan las técnicas sobre cómo generalizar [McMaster y Shea, 1992] (Ver Figura 2.2).

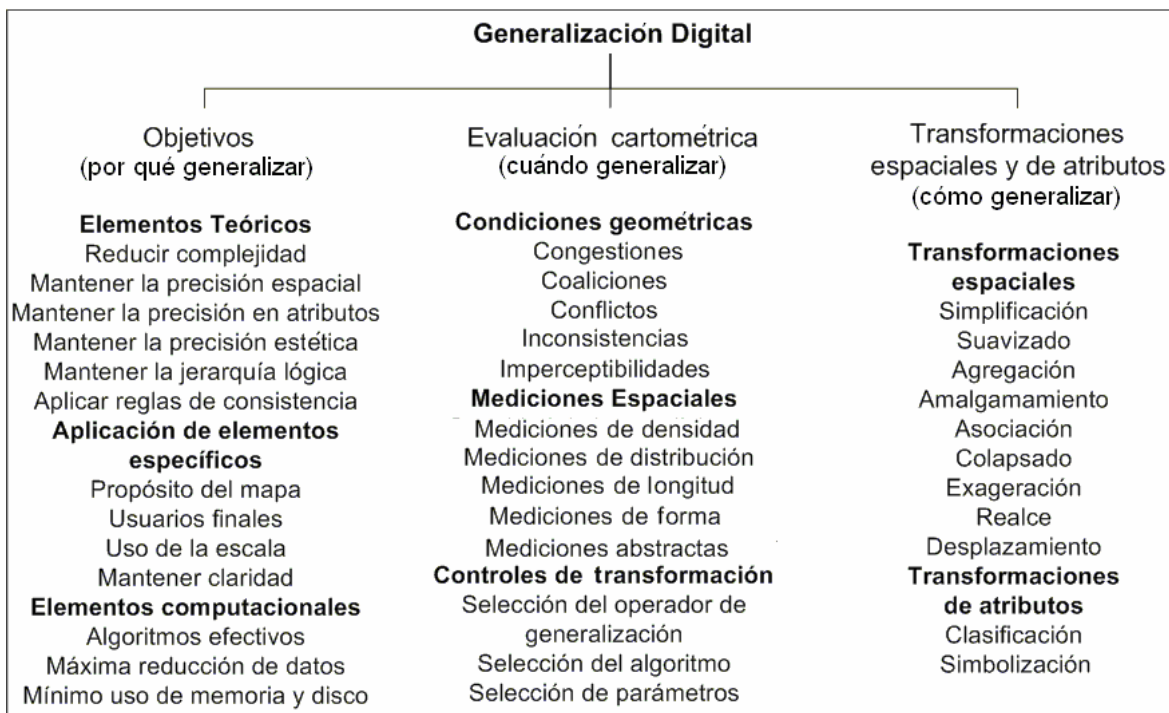


Figura 2.2 Modelo de McMaster y Shea

El área operacional acerca de los objetivos filosóficos involucra aspectos tales como: la reducción de complejidad, el mantenimiento de la precisión espacial y de los atributos, la calidad estética y de la jerarquía lógica del mapa. Además, se consideran los aspectos dependientes de la aplicación, tales 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 operacional 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.

El área que involucra la evaluación cartométrica sobre *cuándo* generalizar 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. Para esto se definen mediciones espaciales tales como: congestión, fusión, conflicto, inconsistencia e imperceptibilidad. Adicionalmente, las mediciones espaciales se utilizan para evaluar las relaciones entre los diferentes objetos, como: 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 los *controles de transformación*. Dichos controles se encargan de la parte de *dónde* generalizar, lo que es muy importante, ya que los operadores pueden utilizarse de diferentes formas (iterativamente, permutados o combinados) utilizando diferentes parámetros.

La tercera 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 las funciones matemáticas. Los operadores de generalización consisten de 10 transformaciones espaciales: Simplificación, Suavizado, Agregación, Amalgamamiento, Asociación, Colapsado, Selección/Perfeccionamiento/Tipificación, Exageración, Realce y Desplazamiento; y dos Transformaciones de atributos; Clasificación y Simbolización (Ver Figura 2.3).



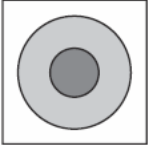
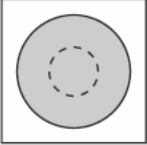
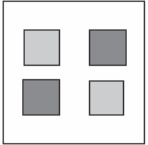
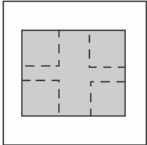
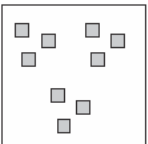
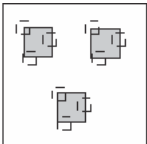
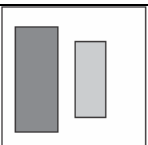
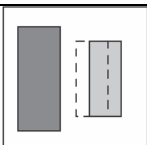
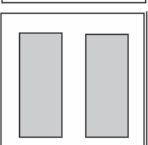
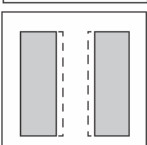
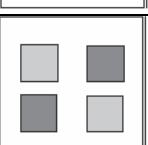
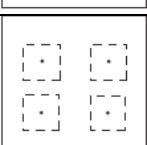
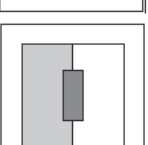
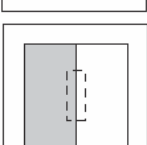
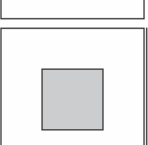
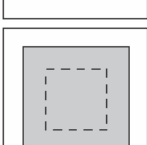
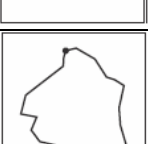
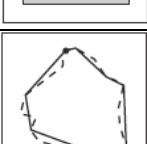
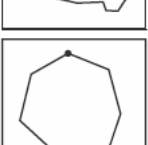
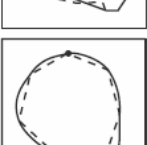


Tipo	Aplicado a:	Nombre	Descripción	Antes de aplicar	Después de aplicar
No Espacial	mapa	Reclasificación	Modifica la categoría a la que pertenece un objeto, así como la posibilidad de combinar objetos vecinos que pertenezcan a la misma clase		
		Agregación	Combina un objeto con otro de la misma (o similar) clase		
Espacial	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		
		Desplazamiento	Implica el movimiento de un objeto, sin cambiar su forma		
	polígonos	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	Simplificación	Reduce la granularidad de las líneas		
líneas		Suavizado	Mejora la apariencia visual de las líneas		

Figura 2.3 Operadores de Generalización de Shea y McMaster

### 2.1.3 Los sistemas basados en conocimiento para la Generalización

Los sistemas expertos pretenden emular el comportamiento de un experto humano en la resolución de problemas. Se encargan de la decisión de dónde y cuándo realizar una acción específica. Los componentes básicos de un sistema experto son (Ver Figura 2.4 ):

- La *Base de Conocimientos* contiene el *conocimiento formalizado* (estructurado) para llegar a una solución dado un problema. Este conocimiento es representado comúnmente por *reglas* que proporcionan descripciones formales de recomendaciones, directivas, estrategias, etc. [Waterman, 1986]. Las reglas son expresadas por la sintaxis: ‘Si *condición* entonces *acción*’; por ejemplo:
  1. Si (objeto área  $\leq$  1000) entonces (alargar objeto)
  2. Si (distancia entre objetos  $\leq$  200) entonces (desplazar objetos)
- Los *Hechos* (Base de Hechos) describen el problema a ser resuelto por el sistema experto, esto es, se utiliza como base para la toma de decisión del sistema.
- El *Motor de Inferencia* aplica las reglas de la base de conocimientos con los hechos. En un ciclo para buscar coincidencias (*matching*) de los hechos con las condiciones de las reglas, para seleccionar una regla que hace *matching* con los hechos y ejecuta la acción correspondiente [Torsun, 1995]. Se repite este ciclo hasta que se alcanza una solución al problema.

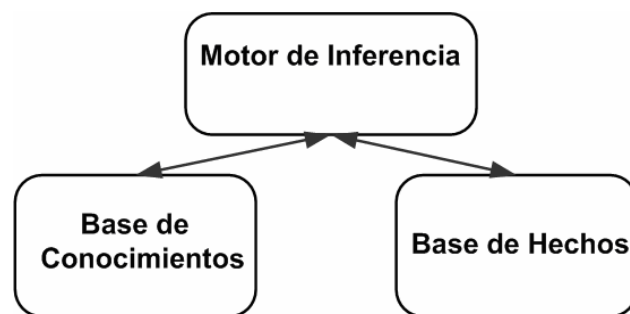


Figura 2.4 Componentes básicos de un sistema experto

A principios de los 90's, [Herbert y Joao, 1992] y [Joao, 1993] describieron aplicaciones basadas en un sistema experto para asistir al usuario en las decisiones sobre la generalización. Dichas decisiones se refieren a la selección del algoritmo y sus parámetros. El sistema asume que los usuarios no son expertos en cartografía y en este caso el sistema resulta de gran ayuda. En teoría, esto puede ser mejor que utilizar los métodos estándares de generalización con parámetros por omisión, y establecer en lo posible el conjunto de reglas que cobren todas las necesidades de la generalización. Desafortunadamente, esto puede no ser suficiente, ya que un verdadero sistema experto para la generalización debe ser capaz de determinar si la tarea a realizar es imposible o si la recomendación falla y no se obtiene un resultado correcto. La propuesta de un conjunto de parámetros es simplemente un avance muy pequeño. Un verdadero sistema experto (que sea realmente útil) debe ser capaz de experimentar con diferentes parámetros y entonces evaluar si es válida la recomendación.

El éxito de un sistema experto depende directamente del conocimiento que contiene. Entonces, cuanto mejor se define el problema, se pueden derivar mejores reglas, lo que origina que se obtengan mejores soluciones. Sin embargo, la generalización es altamente compleja y existe poca formalización del conocimiento cartográfico [Fisher y Mackaness, 1987], [Weibel, 1991], [Ruas, 1995]. En [Armstrong, 1991] y [Weibel, 1995] se discute sobre el conocimiento requerido para la generalización de un mapa; como resumen podemos decir que lo clasificaron en tres grupos:

- Conocimiento geométrico (por ejemplo, tamaño y forma de los objetos)
- Conocimiento estructural (por ejemplo, jerarquía de las categorías representadas)
- Conocimiento procedural (por ejemplo, para 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<sup>5</sup>, INEGI<sup>6</sup>). En [Weibel, 1995] se concluye que la adquisición

---

<sup>5</sup> National Mapping Agencies.

<sup>6</sup> Instituto Nacional de Estadística, Geografía e Informática.

de conocimiento cartográfico es el *cuello de botella* de los sistemas expertos para la generalización. Con el fin de obtener *conocimientos cartográficos* se siguieron las siguientes alternativas:

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

A pesar de los intentos de adquisición de conocimiento cartográfico, se desarrollaron pocos sistemas expertos experimentales para la 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 y Jones, 1989], [Zoraster, 1991], [Freeman y Doerschler, 1992] o los enfocados a objetos lineales topológicamente estructurados [Nickerson, 1991]. El conocimiento fue recolectado y empleado en estudios empíricos con la ayuda de los cartógrafos profesionales que juzgaban los ejemplos preparados según su experiencia [Schylberg, 1993] o bien que solucionaban algunos casos de estudio [Kilpeläinen, 1995], [Kilpeläinen, 2000].

*Hasta ahora, no hay pruebas de que pueda ser construido un sistema eficaz para la generalización basado en reglas* [Müller et al., 1995]. 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 de forma dinámica y flexible a diversos conflictos. Para superar

esta desventaja, la investigación de la generalización comenzó a desarrollarse en la dirección de los sistemas multiagente.

### 2.1.4 La Generalización con sistemas multiagentes

Un Agente es un sistema computacional situado en algún ambiente, capaz de ejecutar una acción autónoma para resolver sus objetivos” [Wooldridge, 1999]. Es decir, un agente autónomo percibe su ambiente y por otra parte el agente modifica su ambiente con sus acciones (Ver Figura 2.5a). Por lo tanto, un agente puede adaptarse dinámicamente a un ambiente que cambia. Se considera como inteligente si funciona flexiblemente y racionalmente acorde a su ambiente y capacidades [Weiss, 1999].

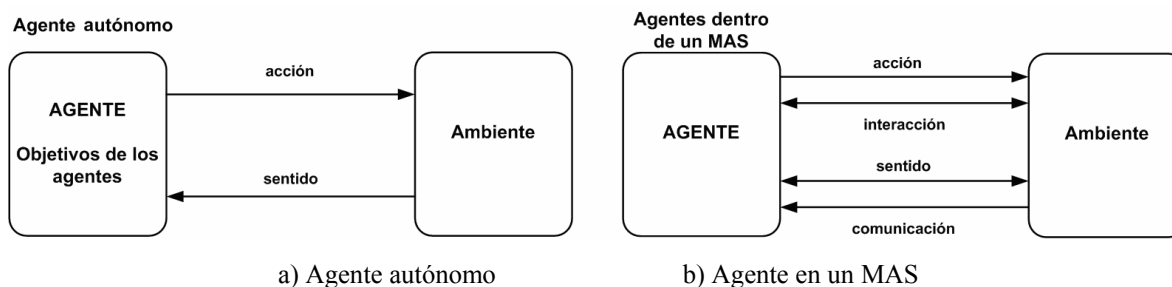


Figura 2.5 Interacción de agente

Un sistema multiagente (MAS) es un método de inteligencia artificial distribuida [O'Hare y Jennings, 1996], [Weiss, 1999]. MAS se refiere a la "solución cooperativa de problemas por un grupo descentralizado de procesos o agentes" [Torsun, 1995]. Un MAS se define como un sistema "en el cual diversos agentes inteligentes interactúan persiguiendo metas o ejecutando un conjunto de tareas" [Weiss, 1999]. En [Jennings et al., 1998] especifica que un MAS tiene las siguientes características:

- Cada agente tiene metas, capacidades específicas y conocimientos individuales. Así pues, los agentes se complementan uno a otro en su funcionalidad y capacidad;
- El sistema de control y los datos se descentralizan, es decir, los problemas se descomponen en *subproblemas*,
- El cómputo es asíncrono.

Además, en un MAS un agente puede comunicarse y cooperar con otros agentes (probablemente en el mismo ambiente) para satisfacer sus propios objetivos relacionados con un *subproblema*, y así, solucionar un problema de alto nivel (objetivos del MAS) (Ver Figura 2.5b). Algunos ejemplos de las áreas donde la tecnología de MAS ha probado ser útil:

- *E-Commerce*: donde los agentes negocian a nombre de sus usuarios [Chávez y Maes, 1996];
- Recuperación de datos en la Internet: tal como búsqueda y filtrado de información [Bigus y Bigus, 2001];
- Modelación de datos geográficos: en procesos como el análisis del crecimiento urbano [Loibl y Totzer, 2002].

En lo que se refiere a la generalización, un MAS puede ofrecer la metodología necesaria para orquestar los algoritmos, restricciones, estrategias, etc. en un sistema dinámico, flexible y automatizado. La aplicación de sistemas multiagentes a la generalización digital se propuso primeramente en la tesis doctoral de Anne Ruas publicada en 1999 [Ruas, 1999] y fue enriquecido por el proyecto europeo AGENT [Barrault et al., 2001]. El proyecto AGENT demostró las capacidades de un MAS para la generalización de mapas [Lamy et al., 1999], [Barrault et al., 2001], [Duchêne et al., 2002], el cual se basa en restricciones, niveles de análisis y agentes.

#### **2.1.4.1 El proyecto AGENT**

El proyecto 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 [Barrault et al., 2001] [AGENT, 2000]. 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. Considerando la definición de *agente* proporcionada anteriormente, cada agente (entidad geográfica) en el prototipo AGENT, tiene las siguientes características:

- 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 cumplimiento de las restricciones.

La generalización de un agente está ligada a un conjunto de restricciones, es decir, las necesidades de la generalización y la evaluación de la solución que se obtiene son dirigidas por las restricciones. Según [Ruas, 1999] y [Ruas, 2000], el conjunto de datos a generalizar se organiza jerárquicamente en niveles de análisis. Usando dichos niveles, se facilita la caracterización de los conflictos, la aplicación de los algoritmos y se realiza la generalización de forma más eficiente. Así, todas las restricciones, operaciones, etc., están limitadas por un cierto nivel. El prototipo AGENT fue construido en dos niveles [Barrault et al., 2001], [Duchêne et al., 2002]:

- El nivel micro se encarga de las entidades geográficas individuales, tal como un edificio o un camino;
- El nivel mezo, 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, 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 cumplimiento de las restricciones;
3. De nuevo caracterizar las restricciones para decidir si la situación del agente mejoró (Ver Figura 2.6).

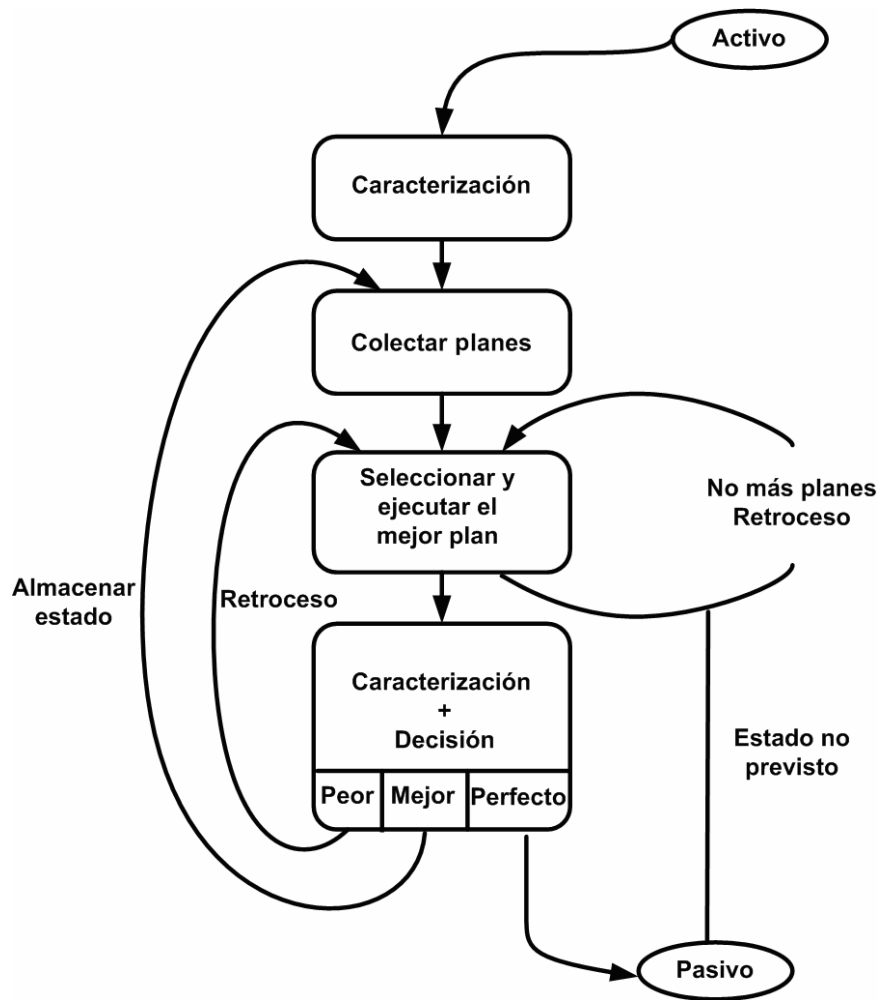


Figura 2.6 Ciclo de vida de un agente

El ciclo de vida de un agente termina cuando se alcanza un estado perfecto (cuando todas las restricciones están satisfechas por completo) o no existe la posibilidad de mejorar su estado. Se utilizó una modificación del *algoritmo de Hill-climbing* para comparar y buscar el mejor de los diferentes estados generados por el proceso iterativo de generalización [Regnauld, 2001]. Este estado es entonces utilizado para actualizar el agente.

### 2.1.5 Otros tópicos en la Generalización

En esta subsección se detallarán algunos aspectos relacionados con otros tópicos que han sido desarrollados para la generalización automática.



### **2.1.5.1 Algoritmos de la generalización**

*Un algoritmo de generalización es una construcción matemática formal que soluciona un problema de generalización modificando la geometría o las propiedades de un objeto* [Bader et al., 1999]. Los algoritmos representan implementaciones concretas de los operadores de generalización. La estructura específica de las capas de datos y la automatización deseada del proceso, determinan algunos requisitos de los algoritmos. Un algoritmo de generalización adecuado debe:

- No requerir entradas interactivas;
- Dar apoyo general a la generalización, es decir, varios conflictos se pueden solucionar simultáneamente;
- Asegurar la consistencia del conjunto de datos, según la definición de una capa de datos, es decir, cada localización en el mapa debe pertenecer a solamente una categoría;
- Preservar las características de la capa de datos (por ejemplo, la distribución, las alineaciones de los objetos, o la forma específica del objeto, etc.);
- Que pueda manejar diversos tipos de datos y escalar los rangos.

### **2.1.5.2 Manipulación de los efectos secundarios**

Debido al carácter de las capas de datos, la transformación de un objeto individual requiere siempre la adaptación subsecuente con todos los objetos adyacentes. Es decir, todas las modificaciones de un objeto generalizado se deben propagar a la capa. Estas adaptaciones requeridas se llaman los efectos secundarios. La dirección de los efectos secundarios fue investigada con respecto a la generalización de las redes de caminos [Nickerson, 1988], [Bader, 2001] y de polígonos disjuntos tales como edificios [Burghardt, 2000], [Højholt, 2000].

Los efectos secundarios se pueden tratar como sub-tareas independientes del proceso de generalización, es decir, los cambios deben propagarse a los objetos vecinos una vez que se

termine la transformación, pueden existir desplazamientos. Los cuales, pueden generar nuevos conflictos, es decir inconsistencias en el mapa [Bader y Barrault, 2000]. Al contrario, [Burghardt, 2000], [Bader, 2001] y [Harrie, 1999], hicieron uso de técnicas que permiten el tratamiento de las transformaciones y las propagaciones en sentido global, es decir, ambos procesos se realizan en paralelo y proporcionan una solución común.

### 2.1.5.3 Semántica en la Generalización

Al generalizar los datos espaciales, la semántica de los datos juega un papel importante [Le Men, 1996]. La generalización permite la transformación semántica y controla la toma de decisión para las operaciones geométricas. Sin embargo, los componentes espaciales y semánticos de una capa de datos están ligados íntimamente, y cualquier tratamiento de uno aislando el otro, tendrá el riesgo de lograr una representación errónea del fenómeno cartográfico [Mark et al., 1989]. Entonces, el conocimiento sobre el esquema primario de clasificación de un experto sobre el tema de un mapa (y a veces incluso de los datos auxiliares) es imprescindible para producir no solamente una “buena vista” de los datos, sino también una agrupación significativa de las clases de datos y la generalización del conjunto de datos [Imhof, 1972]. La semántica de la generalización es importante por diversas razones:

- Ayuda a seleccionar el operador adecuado para un conflicto particular. Por ejemplo, un objeto aislado de una categoría rara puede ser acentuado más bien que eliminado, o un grupo significativo 'demasiado pequeño' puede agregarse en vez de quitarse.
- La semántica controla la reducción del número de clases (por ejemplo, la combinación de clases en *super clases*, es decir, la combinación del *bosque de hojas caducas* con el *bosque de coníferas* en una clase *bosque*). Más allá de cualquier duda, la reclasificación es el operador más importante para reducir la complejidad de una capa de datos [Jaakkola, 1998], especialmente al realizar cambios drásticos en la escala (por ejemplo, cambio de 1:25,000 a 1:200,000). Además, la combinación geométrica de objetos adyacentes de la misma categoría (que resulta de la reclasificación) implica la solución de conflictos geométricos potenciales, tales como polígonos que sean demasiado pequeños o demasiado estrechos [Spiess, 1990].

- Determina el objeto vecino más probable para llenar el “hueco” que existe después de que la eliminación de un objeto se identifica con respecto a la semántica. La compatibilidad de dos objetos depende, además de sus características geométricas, principalmente de su similitud semántica. [Yaolin, 2002] propusieron una medida de semejanza semántica para la generalización de la base de datos. El relleno del “hueco” se puede lograr alternativamente mediante una solución geométrica que distribuye el área del objeto eliminado entre los objetos adyacentes [Bader, 1997], [Bader et al., 1999].
- La semántica interviene en la resistencia del límite de un polígono a la deformación. Es decir, la modificación de los límites denominados *suaves* (por ejemplo, el contorno del área de un talud), se prefieren en relación a la de los límites denominados *duros* (por ejemplo, el contorno de un lago).
- Valida una solución usando criterios de integridad semántica, por ejemplo, un lago en el mar no es aceptable.

#### **2.1.4.4 Restricciones**

En 1991, Kate Beard [Beard, 1991] utilizó restricciones (*constraints*) en vez de reglas de generalización. Definió una restricción como “condición similar al predicado en una regla de producción. La distinción es que una restricción no está limitada a una acción” [Beard, 1991]. Alternativamente, una restricción se puede ver como una especificación de diseño para el cual un objeto, grupo de objetos, o el mapa, deben adherirse [Weibel y Dutton, 1998]. Las aplicaciones basadas en restricciones para la generalización fue resaltado por [Ruas y Plazanet, 1996], [Weibel, 1996], [Ruas, 1999], [Lamy et al., 1999], [Barrault et al., 2001] y [Galanda, 2003]. Dado que las restricciones definidas para un objeto contradicen frecuentemente una solución de la generalización, hay siempre una negociación entre las diversas restricciones y su satisfacción. Por ejemplo, es imposible evitar la distorsión de la forma de un polígono, si el contorno del polígono es ilegible en la escala destino; debido a otros detalles se requiere alguna simplificación para hacerlo legible.

[Edwardes y Mackaness, 2000] y [Weibel, 1996] discutieron acerca de la generalización en un nivel conceptual y de diseño. Las restricciones de la generalización se ligan íntimamente con las medidas que ayudan a calcular el cumplimiento de una restricción.

### 2.1.6 Ontologías en la Generalización

En [Kulik et al., 2005] se presenta un algoritmo que se basa en la inclusión de información ontológica orientada a la generalización digital de objetos lineales, el cual puede adaptarse para diversos usuarios y tareas. El algoritmo propuesto se denomina *DMin* (*decimation-min- $\epsilon$* <sup>7</sup>), el cual se enfoca hacia la minimización del error originado por la simplificación de los objetos espaciales. El algoritmo se basa en una *función de ponderación* que contiene dos componentes: una geométrica y otra semántica. Adicionalmente, el diseño del algoritmo incluye la preservación de la consistencia topológica y la geometría de los objetos.

Un aspecto crucial de *DMin* es la asignación de pesos en cada vértice de todas las líneas que componen la capa temática. Como se mencionó anteriormente, la ponderación tiene dos componentes: una semántica y otra geométrica. *DMin* usa la *función de ponderación* para decidir qué puntos serán borrados cuando la línea es simplificada. Se utiliza la técnica *min- $\epsilon$*  [Cromley, 1991], que consiste en eliminar durante cada iteración el vértice con la menor importancia (la ponderación más pequeña) en cuyo caso no se requiere preestablecer un límite para el error. El número total de vértices borrados puede ser especificado por el usuario. Por ejemplo, si se requiere una compresión de 10:1, entonces el algoritmo debe remover el 90% de los puntos que tienen menor importancia, tanto en términos de la geometría como de la semántica. En el contexto de las aplicaciones móviles, se puede definir un nivel de compresión que permita ajustarse a las limitaciones de ancho de banda o de resolución de la pantalla.

El componente geométrico de *DMin* mide el impacto del vértice en la forma de la línea a la que pertenece. El peso en ese caso frecuentemente depende de la longitud de los segmentos a los cuales pertenece y del ángulo que forman. En este caso la *función de ponderación* es ternaria  $f: \mathbb{R}^3 \rightarrow \mathbb{R}$ . La función es  $f(s_1, s_2, \alpha) \mapsto s_1 \cdot s_2 \cdot \alpha$  y depende de los segmentos involucrados y el ángulo que forman los segmentos entre sí. Un aspecto clave de *DMin* es la *adaptabilidad*, ya que puede utilizarse para diferentes requerimientos cartográficos o

<sup>7</sup> Cualquier algoritmo de simplificación introduce una desviación de las líneas originales, el error  $\epsilon$ , de la simplificación. [Cromley, 1991]

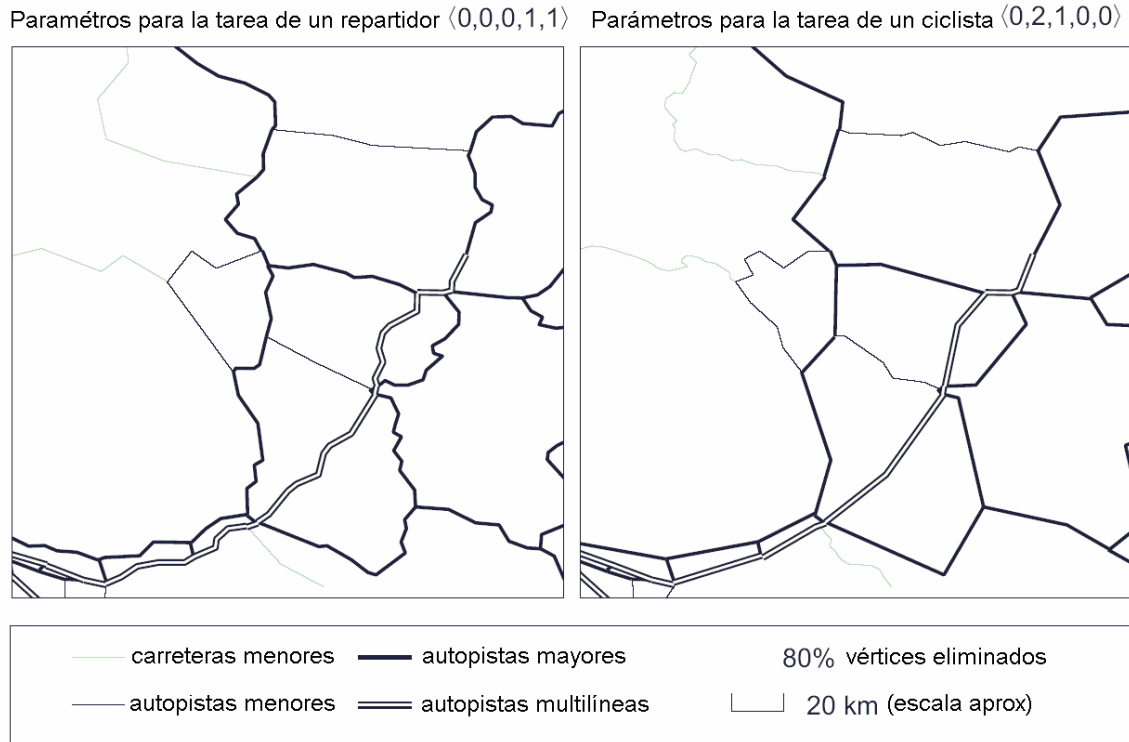
psicológicos. El algoritmo adopta la norma  $L^2$  para asegurar la compatibilidad con los desarrollos existentes<sup>8</sup>.

En el artículo se propone una *función de ponderación* basada en un conjunto de ponderaciones (denominada tupla de ponderación) definidos por el usuario. Los pesos se asignan según la clase de objetos en el mapa. El componente semántico depende del caso de estudio, en este caso, una red de transporte. La red de transporte se compone por cinco clases: rutas de ferry, carreteras menores, autopistas menores, autopistas mayores y autopistas multilínea, que corresponden a valores en la tupla  $\langle a, b, c, d, e \rangle$ . Para el caso de estudio, se define la siguiente *función de ponderación*  $w_s(v) \mapsto k^x$ , en donde  $k$  es constante y  $x$  es el peso definido por el usuario para cada clase de objeto al que el vértice  $v$  pertenece. Así, los valores para cada elemento de la tupla se pueden manipular para requerimientos particulares. Por ejemplo, en el caso de un repartidor, que considera más importantes las autopistas multi-líneas y las autopistas mayores, se modela con la tupla  $\langle 0, 0, 0, 1, 1 \rangle$ . Por otro lado para requerimientos turísticos se puede utilizar la tupla  $\langle 1, 0, 1, 1, 0 \rangle$  (para los turistas son de menor importancia las autopistas multilíneas y las carreteras menores que las rutas de ferry y autopistas menores). De forma similar, para los ciclistas que desean evitar el tráfico pesado, pueden considerar de gran relevancia las carreteras menores y las autopistas menores, por lo que se puede usar la siguiente tupla  $\langle 0, 2, 1, 0, 0 \rangle$ .

Para incorporar la información semántica al proceso de simplificación, es necesario combinar la función de ponderación semántica con la función de ponderación geométrica. En la Figura 2.7 se muestra un mapa en donde se realizó la simplificación para el caso de dos diferentes requerimientos, para un repartidor y un ciclista. En cada uno de los casos se puede apreciar que el nivel de detalle es preservado de diferente forma en relación a los requerimientos de cada usuario.

---

<sup>8</sup> La norma  $L^2$  es un estándar usado en la geometría para cuantificar la discrepancia entre dos curvas, en términos del área formada por las curvas. [Hoffman y Singh, 1997]



**Figura 2.7** Simplificación de líneas usando DMin

En resumen, podemos decir que este enfoque posee dos características principales:

- *Componente Geométrico*, que a diferencia de propuestas anteriores para la generalización, en este caso no existen umbrales fijos para la generalización.
- *Componente Semántico*, que considera la relevancia de las propiedades del mapa para el usuario. Un punto importante en este algoritmo es la asignación de pesos a cada uno de los vértices de los arcos que se almacenan en la base de datos geográfica. El nivel de detalle en el mapa que se genera es adaptado automáticamente considerando la semántica de las propiedades representadas, y eso lo definen los requerimientos del usuario.

## 2.2 Representación del conocimiento cartográfico

En esta sección se presentan dos enfoques relacionados con la representación del conocimiento cartográfico.

### 2.2.1 Semiótica

La semiótica, como teoría de símbolos y sus significados, se ocupa de la comunicación humana por medio de los sistemas organizados de los significados [Eco, 1976]. Según [Wikipedia, 2007], la semiótica se define como el estudio de los signos. Un signo (del griego *semeion*) es todo lo que se refiere a otra cosa (referente), es la materia prima del pensamiento y por lo tanto de la comunicación. Por consiguiente, las representaciones cartográficas se pueden considerar como sistemas de comunicación que emplean signos cartográficos dentro de un sistema de significados, el mapa.

En la teoría semiótica, un signo consta de dos componentes [Head, 1991]:

1. Una extensión física que depende de la forma de comunicación (por ejemplo, un sonido, un gesto, o una marca en un pedazo de papel). La extensión física generalmente se llama la *expresión* del signo.
2. Cada signo tiene un contenido, es decir, su *significado* o *concepto mental*. Con el fin de usar signos dentro de un sistema de significados, su contenido y sus expresiones tienen que ser ligadas mediante un código común a las personas involucradas en el proceso de comunicación.

Con respecto a la cartografía, es obvio que los dos componentes del término semiótico “signo” tiene, por un lado, las entidades cartográficas representadas en el mapa y por otro lado, el conocimiento geográfico que les corresponde. Para la cartografía, como en el proceso de lectura del mapa, hay siempre dos casos de conocimiento implicados:

1. El mapa con sus entidades cartográficas.
2. Los conceptos mentales del cartógrafo y/o del lector del mapa.

### 2.2.2 Mapas de Aspectos

Este enfoque se basa en el hecho de que los mapas están diseñados para representar solo ciertas partes del conocimiento acerca de las entidades geográficas. Es decir, la selección de una parte o aspecto de un mapa que nos lleva a la noción de *mapas de aspectos* (*aspect*

*maps*). Un *mapa de aspectos* es una descripción formal de un mapa que permite distinguir entre las partes pretendidas o representadas de información y aquellas incorporadas por la representación pictórica originadas por la visualización de la base de datos [Barkowsky, 1997].

Un *mapa de aspectos*, es descrito por la tupla  $\langle E, R, M \rangle$  donde E es el conjunto de entidades que contiene el mapa, R es el conjunto de relaciones que tienen dichas entidades y contiene la posición de las mismas (entidades lineales o áreas), y M es la referencia al meta-conocimiento sobre el contenido pictórico descrito en E y R.

Para entender un mapa se requiere de cierto conocimiento adicional sobre la forma en que los aspectos representados en el mapa deben interpretarse. De acuerdo a este planteamiento, la meta-conocimiento (M) representa la liga entre las relaciones espaciales con el conocimiento espacial existente en la representación pictórica. Entonces, la meta-conocimiento lleva la representación pictórica a una representación espacial de acuerdo al entendimiento de lo que es el término “mapa geográfico”.

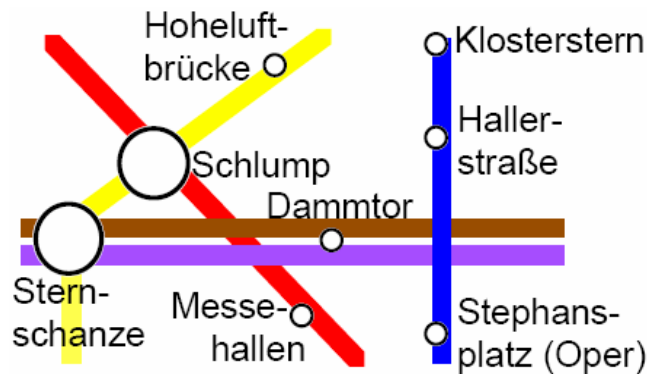
El meta-conocimiento puede entenderse como el conocimiento involucrado entre cada una de las referencias posibles de las entidades cartográficas con las relaciones geográficas de los objetos representados. Por lo tanto, es posible realizar cualquier combinación entre las formas *puramente descriptivas* y *puramente simbólicas* de las representaciones cartográficas, y así tomar en cuenta las diversas formas de modificar las entidades durante la generalización cartográfica.

Para ilustrar esto, se consideró el siguiente ejemplo en donde se demuestran algunas diferencias extremas en la forma en que se trata el contenido del mapa. La Figura 2.8 muestra un fragmento de la red del metro de Hamburgo. En el caso de esta representación el mapa está relajado geoméricamente de su base topográfica.

Con base en la Figura 2.8, se utilizaron tres casos para ilustrar cómo diversos aspectos pueden considerarse de diferentes formas:



1. Las conexiones representadas entre las estaciones se pueden tomar como una representación literal de la conectividad.
2. La orientación de las estaciones no debe tomarse como la representación exacta de la orientación, pero puede ser que sean asignadas a un cierto patrón tosco de orientación, por ejemplo, una clasificación cualitativa 8-sector, y



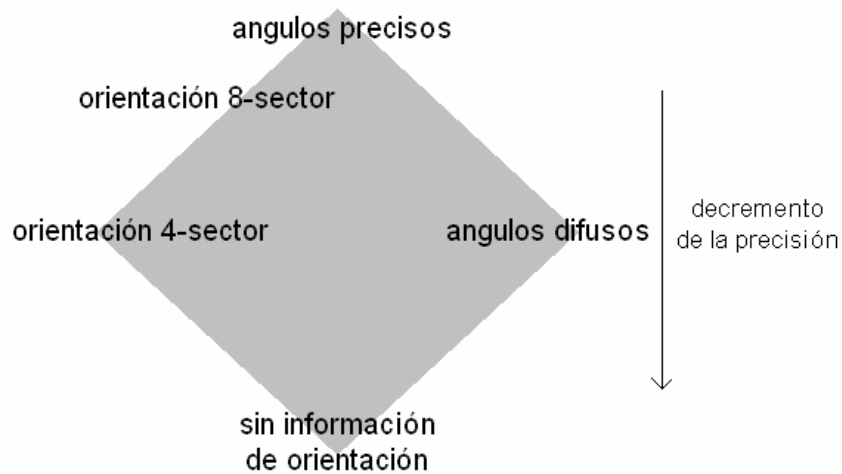
**Figura 2.8** Fragmento de la red del Metro de Hamburgo

3. Las distancias entre los símbolos de la estación en el mapa no reflejan las distancias reales entre las estaciones, por lo que no es posible realizar comparaciones cualitativas entre ellas.

Basándose en este ejemplo, se justifica que en el caso de haber empleado correctamente el meta-conocimiento, se puede tener una interpretación incorrecta del mapa. Para el caso del ejemplo puede ser que existan diferencias entre las orientaciones que poseen las estaciones.

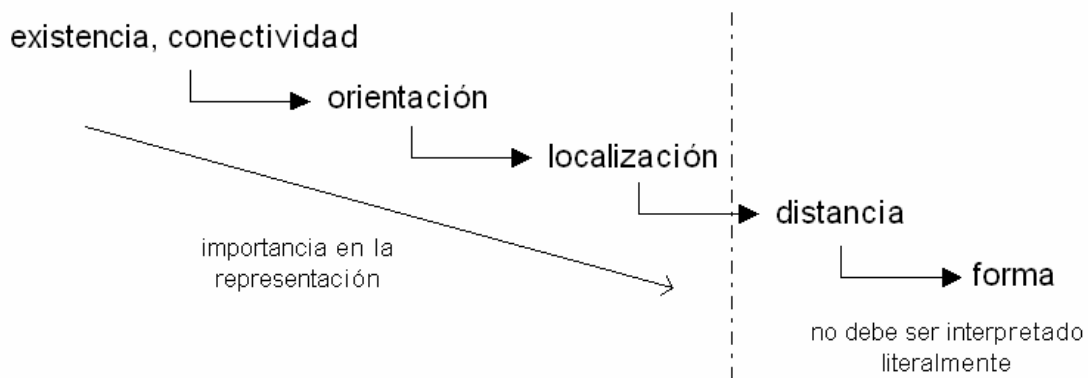
Por esta razón, se propone que los componentes del meta-conocimiento sean organizados jerárquicamente, de manera que se establezca una correspondencia entre la representación del mapa y la relación espacial en el mundo real y que pueda ser ordenada jerárquicamente en dependencia de la “fuerza” de su correspondencia en la representación. Para el caso de la orientación se propone la Figura 2.9, en donde se representa de forma gráfica la jerarquización de los niveles de precisión en relación con la realidad. En la Figura 2.9, se muestra la jerarquización de un aspecto de los mapas. Sin embargo, es necesario

representar otros aspectos que poseen los mapas. Por esta razón se propuso utilizar una jerarquía de precedencia pictórica del mapa (Ver Figura 2.10).



**Figura 2.9** Ejemplo de estructura jerárquica para la representación de la correspondencia

De la Figura 2.10, podemos ver que aparece como de mayor importancia la existencia y la conectividad (que se refiere a una relación topológica), que son importantes para describir la red del Metro (Ver Figura 2.8). Asimismo, como de menor importancia aparece la forma, que no es tan importante para representar la red del Metro.



**Figura 2.10** Ejemplo de jerarquía de precedencia pictórica

Un aspecto que está colocado en una posición baja dentro de la jerarquía, significa que puede ser debilitado, es decir, que se puede hacer más flexible su representación si se requiere representar un aspecto de prioridad más alta.

## 2.3 Ontologías

Algunas definiciones generales sobre Ontología:

- “Una ontología es el estudio metafísico<sup>9</sup> de la naturaleza del ser y su existencia” [Word Net, 2007].
- “Campo de la metafísica que investiga y explica la naturaleza, las características y las relaciones esenciales de todos los seres, como tales, o los principios y las causas del ser” [Merriam-Webster, 2007].
- “Un informe detallado de la Existencia”. [Hiperdictionary, 2006].
- “Del griego οντος, genitivo del participio del verbo ειμι, ser, estar y λογος, ciencia, estudio, teoría) es una disciplina que se suele identificar con la *Metafísica general* o bien indica una de las ramas de ésta que estudia lo que es en tanto que es y existe. La ontología se ocupa de la definición del ser y de establecer las categorías fundamentales o modos generales de ser de las cosas a partir del estudio de sus propiedades”. [Wikipedia, 2007]

En informática el término ontología se refiere a al intento de formular un exhaustivo y riguroso esquema conceptual dentro de un dominio dado, con la finalidad de facilitar la comunicación y la compartición de la información entre diferentes sistemas. Aunque toma su nombre por analogía, ésta es la diferencia con el significado filosófico de la palabra ontología [Wikipedia, 2007]. A continuación, se describen algunas de los principales trabajos relacionados con la formalización de las ontologías computacionales.

### 2.3.1 Thomas R. Gruber

En 1993, Thomas Gruber publicó un artículo en donde se presentan a las ontologías como un medio para reutilizar y compartir el conocimiento entre las aplicaciones [Gruber, 1993]. Desde este punto de vista, las ontologías formales son vistas como artefactos diseñados para propósitos específicos. Además, se define el rol de las ontologías en las actividades

---

<sup>9</sup> Metafísica. El estudio filosófico del ser y del saber.

relacionadas con compartir el conocimiento y se describen algunos criterios para desarrollar las ontologías.

De acuerdo con [Gruber, 1993], una *ontología* es una especificación explícita de una conceptualización<sup>10</sup>. Cuando el conocimiento sobre un dominio se representa en un formalismo declarativo, el sistema de objetos que puede ser representado se llama *universo de discurso*. Este sistema de objetos y las relaciones descriptibles entre ellos, se reflejan en el vocabulario de representación con el que un programa representa el conocimiento. En este contexto, es posible describir la ontología de un programa definiendo el conjunto de términos representativos. En dicha ontología, las definiciones asocian nombres de entidades del universo de discurso (por ejemplo, clases, relaciones, funciones, u otros objetos) a describir por medio de texto en lenguaje natural (*human-readable*) que representan los nombres y axiomas formales que restringen la interpretación y el uso de términos bien formados. Formalmente, una ontología es la declaración de una teoría lógica. Las ontologías se comparan frecuentemente con jerarquías taxonómicas de clases por las definiciones de clase y las relaciones de subsunción (*subsumption*, de una categoría general a una particular), pero las ontologías no necesitan ser limitadas a esas formas.

Comúnmente se usan ontologías para describir los *compromisos ontológicos* para un conjunto de agentes, de modo que puedan comunicarse sobre un dominio de discurso sin necesariamente utilizar una teoría global compartida. Se dice que un agente está *comprometido* con una ontología, si sus acciones observables son consistentes con las definiciones de la ontología. La idea de los compromisos ontológicos se basa en la perspectiva del Nivel de Conocimiento [Newell, 1982]. El nivel de conocimiento es el nivel de descripción del conocimiento de un agente, el cual es independiente de la representación a nivel de símbolo usado por el agente. El conocimiento es atribuido a los agentes observando sus acciones; un agente *sabe* algo, si actúa racionalmente *como si* tuviera la información para alcanzar sus metas. Las "acciones" de los agentes pueden verse a través

---

<sup>10</sup> Según [Genesereth y Nilsson, 1987], el cuerpo del conocimiento representado formalmente es basado en una conceptualización: son los objetos, conceptos y otras entidades que se asume existen en un área de interés, así como las relaciones que existen entre ellas. Una conceptualización es una vista abstracta simplificada del mundo que se quiere representar para algún propósito. Cualquier base de conocimiento, sistema basado en el conocimiento o conocimiento a nivel de agente, está comprometido con alguna conceptualización, explícita o implícita.

de una interfaz funcional [Levesque, 1984], donde un cliente interactúa con un agente haciendo aseveraciones lógicas (*dice*) y con el planteamiento de preguntas (*pide*).

Pragmáticamente, una ontología común define el vocabulario con el cual los agentes pueden intercambiar consultas y aseveraciones. Los *compromisos ontológicos* son acuerdos para utilizar vocabulario compartido en forma coherente y consistente. Los agentes que comparten un vocabulario no necesitan compartir una base de conocimientos; cada uno sabe cosas que el otro no, y un agente que confía en la ontología no requiere contestar a todas las preguntas que se puedan formular en el vocabulario compartido. En conclusión, podemos decir que el *compromiso ontológico* es una garantía de consistencia, pero no por completo, con respecto a consultas y aseveraciones usando el vocabulario de la ontología.

Cuando se elige cómo representar algo en la ontología se toman decisiones. Para guiar y evaluar los diseños, se necesita de criterios imparciales basados en el propósito del diseño, nociones a priori de la naturaleza, o de la verdad. Los criterios de diseño propuestos por Gruber son:

1. *Claridad*. La ontología debe comunicar con eficacia el significado deseado de los términos definidos. Las definiciones deben ser objetivas. Las ontologías se realizan para requerimientos computacionales o sociales; la definición debe ser independiente del contexto.
2. *Coherencia*. Una ontología debe ser coherente, esto es, debe aprobar las inferencias que sean consistentes con las definiciones. Al menos, los axiomas que se definen deben ser lógicamente consistentes. La coherencia debe aplicarse también a los conceptos que se definen informalmente, por ejemplo, los descritos con lenguaje natural. Si una oración que se puede inferir a partir de los axiomas contradice una definición o un ejemplo dado informalmente, entonces la ontología es incoherente [Enderton, 1972].
3. *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 requiera revisar las definiciones existentes.

4. *Minimizar las tendencias en la codificación.* La conceptualización se debe especificar al nivel del conocimiento sin depender de una codificación a nivel de símbolo en particular. Las tendencias de la codificación deben ser reducidas al mínimo, ya que los agentes que comparten conocimiento pueden ser implantados en diferentes sistemas y estilos de representación.
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 un vocabulario, se puede reducir al mínimo especificando una teoría más débil (que se permite en la mayoría de los modelos) y definiendo solamente aquellos términos que sean esenciales para la comunicación de conocimiento consistente.

### 2.3.2 Nicola Guarino

En [Guarino, 1995], se describe la introducción formal de los principios básicos para la ingeniería de conocimiento, con el fin de explorar las relaciones entre la ontología y la representación del conocimiento. Se argumenta que en la “vista de modelado” para adquirir conocimiento propuesta por [Clancey, 1993], se debe establecer una correspondencia entre la base de conocimientos y dos subsistemas:

1. El comportamiento del agente (la *expertiz* en el problema a resolver) y
2. Su ambiente (el dominio del problema)

Las metodologías existentes para modelar el conocimiento tienden a enfocarse hacia el comportamiento del agente y consideran que el conocimiento sobre el dominio del problema es muy dependiente de una tarea particular; de hecho, los investigadores en Inteligencia Artificial (IA) parecieron estar más interesados en la naturaleza del

razonamiento que en la naturaleza del mundo verdadero. Las ontologías son bases del conocimiento independientes de la tarea y son adecuadas para la integración a gran escala.

Además, se realiza una comparación de la dicotomía<sup>11</sup> entre razonamiento y representación para la distinción filosófica entre epistemología y ontología. Epistemología se puede definir como "el campo de la filosofía que se ocupa de la naturaleza y de las fuentes del conocimiento" [Nutter, 1987]. Usualmente, el conocimiento se compone de *proposiciones*, que estructuradas *formalmente* son la fuente del nuevo conocimiento. El aspecto deductivo parece ser esencial para la epistemología: el estudio de la "naturaleza" del conocimiento se limita al significado superficial (es decir, la forma), debido a que es motivado principalmente por el estudio de procesos de inferencia. Por otro lado, la ontología se puede ver como el estudio de la organización y naturaleza del mundo independientemente de la forma de nuestro conocimiento sobre él. La *ontología formal* se ha definido como "el desarrollo sistemático, formal, axiomático de la lógica de todas las formas y modos de ser" [Cocchiarella, 1991]. La interpretación del término *ontología formal* sigue siendo una cuestión de discusión [Poli, 1996], pues considera ambos significados del adjetivo "formal": por un lado, es sinónimo de "riguroso", mientras que por otro lado significa "relacionado con las *formas* de ser". Por lo tanto, *ontología formal* no se refiere únicamente a la existencia de ciertos detalles específicos, sino a la descripción rigurosa de sus formas. En la práctica, la *ontología formal* se puede interpretar como la teoría de *distinciones a priori*:

- entre las entidades del mundo (objetos físicos, eventos, regiones, cantidades ...);
- entre categorías a meta-nivel usadas para modelar el mundo (conceptos, características, calidades, estados, roles, partes...).

En [Guarino, 1998] se definen formalmente algunos conceptos como ontología, compromiso ontológico y conceptualización

---

<sup>11</sup> División en dos partes de alguna cosa. Se aplica este término al método de clasificación en que las divisiones y subdivisiones sólo tienen dos partes.

En el sentido filosófico, podemos referir a una ontología como un sistema particular de categorías para una cierta visión del mundo. Como tal, este sistema no depende de un lenguaje en particular: La ontología de Aristóteles es siempre igual, independientemente del lenguaje usado para describirla. Por otra parte, en su uso más frecuente en la IA, una ontología se refiere a un artefacto de la ingeniería, constituido por un vocabulario específico usado para describir la realidad, además de un sistema de suposiciones explícitas con respecto al significado previsto de las palabras del vocabulario. Este sistema de suposiciones tiene generalmente la forma de una teoría lógica de primer orden, donde las palabras del vocabulario aparecen como nombres de predicados unarios o binarios, respectivamente, llamados conceptos y relaciones. En el caso más simple, una ontología describe una jerarquía de conceptos conectados por relaciones de *subsunción*; en casos más sofisticados, se agregan los axiomas convenientes para expresar otras relaciones entre los conceptos y restringir la interpretación prevista.

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

Una conceptualización se ha definido como una estructura  $\langle D, \mathbf{R} \rangle$ , donde  $D$  es un dominio y  $\mathbf{R}$  es un conjunto de relaciones relevantes en  $D$ . Esta definición fue utilizada por Gruber, quien definió la ontología como "especificación de una conceptualización" [Gruber, 1993]. El problema con la noción de [Genesereth y Nilsson., 1987] (utilizada también por Gruber) sobre conceptualización es que se refiere a las relaciones matemáticas ordinarias en  $D$ , como relaciones *extensionales*. Estas relaciones reflejan una situación particular: por ejemplo, las relaciones que pueden existir entre un conjunto de bloques sobre una mesa. En este caso se requiere el significado de las relaciones, independientemente de la situación, por lo que se habla de *relaciones intencionales* o *relaciones conceptuales*, reservando el término relación para las relaciones matemáticas ordinarias.

Una manera estándar de representar las *relaciones conceptuales*, es considerarlas como funciones que mapean de los mundos posibles con conjuntos de relaciones. Mientras que



las relaciones ordinarias se definen en un cierto dominio, las relaciones conceptuales se definen en un *espacio de dominio* (*space domain*). El espacio de dominio se define como una estructura  $\langle D, W \rangle$ , donde  $D$  es el dominio y  $W$  es un conjunto de estados de las relaciones del dominio (también llamado los mundos posibles). Por ejemplo,  $D$  puede ser un conjunto de bloques en una mesa y  $W$  puede ser el sistema de todos los arreglos espaciales posibles de los bloques. Dado un espacio de dominio  $\langle D, W \rangle$ , definiremos una relación conceptual  $\rho^n$  de cardinalidad  $n$  sobre  $\langle D, W \rangle$  y como función a  $\rho^n : W \rightarrow 2^{D^n}$  de  $W$  sobre el conjunto de todas las relaciones  $n$ -arias (ordinarias) sobre  $D$ . Para una relación conceptual genérica  $\rho$ , el conjunto  $E\rho = \{\rho(w) \mid w \in W\}$  contendrá las extensiones admisibles de  $\rho$ . Una conceptualización para  $D$  se puede definir ahora como la tripleta ordenada  $\mathbf{C} = \langle D, W, \mathfrak{R} \rangle$ , donde  $\mathfrak{R}$  es un conjunto de relaciones conceptuales en el espacio de dominio  $\langle D, W \rangle$ . Por lo tanto, se puede decir que una conceptualización es un sistema de relaciones conceptuales definidas sobre un espacio de dominio.

Considerando la estructura  $\langle D, \mathbf{R} \rangle$  definida por [Gruber, 1993], que se refiere a un mundo en particular (o situación), se denominará *estructura del mundo*. Es fácil ver que una conceptualización contiene muchas estructuras del mundo, una para cada mundo: serán llamadas las *estructuras previstas del mundo* según tal conceptualización. Sea  $\mathbf{C} = \langle D, W, \mathfrak{R} \rangle$  una conceptualización. Para cada mundo posible  $w \in W$ , la estructura pretendida de  $w$ , considerando que  $\mathbf{C}$  es la estructura  $S_{wC} = \langle D, \mathbf{R}_{wC} \rangle$ , donde  $\mathbf{R}_{wC} = \{\rho(w) \mid \rho \in \mathfrak{R}\}$ , es el conjunto de las extensiones (relacionadas a  $w$ ) de los elementos de  $\mathfrak{R}$ . Se denota con  $S_C$  el conjunto  $\{S_{wC} \mid w \in W\}$  de todas las estructuras previstas del mundo de  $\mathbf{C}$ .

Considerando el lenguaje Lógico  $\mathbf{L}$ , con un vocabulario  $V$ . Reacomodando la definición estándar, podemos definir un modelo para  $\mathbf{L}$  con la estructura  $\langle \mathbf{S}, I \rangle$ , donde  $\mathbf{S} = \langle D, \mathbf{R} \rangle$  es la estructura del mundo y por otro lado,  $I: V \rightarrow D \cup \mathbf{R}$  es la *función de interpretación* que asigna elementos de  $D$  a símbolos constantes de  $V$ , y elementos de  $\mathbf{R}$  a símbolos predicado de  $V$ . Un modelo fija una interpretación *extensional* particular del lenguaje. Análogamente, podemos fijar una *interpretación intencional* por medio de una estructura  $\langle \mathbf{C}, \mathfrak{I} \rangle$ , en donde  $\mathbf{C} = \langle D, W, \mathfrak{R} \rangle$  es una conceptualización y una función  $\mathfrak{I} = V \rightarrow D \cup \mathfrak{R}$  que asigna

elementos de  $D$  a símbolos constantes de  $V$  y elementos de  $\mathfrak{R}$  como predicados a símbolos de  $V$ . Denominaremos a esta interpretación intencional como *compromiso ontológico* para  $L$ . Si  $K = \langle C, \mathfrak{I} \rangle$  es un *compromiso ontológico* para  $L$ , se dice que  $L$  compromete a  $C$  por medio de  $K$ , mientras que  $C$  es la conceptualización subyacente de  $K$ . La expresión *compromiso ontológico* en ocasiones se ha utilizado para denotar el resultado del compromiso en sí mismo, es decir, en nuestra terminología, la conceptualización subyacente.

Dado un lenguaje  $L$  con vocabulario  $V$ , y un *compromiso ontológico*  $K = \langle C, \mathfrak{I} \rangle$ , para  $L$ , un modelo  $\langle S, I \rangle$ , será *compatible* con  $K$  si:

- 1)  $S \in S_c$  ;
- 2) para cada constante  $c$ ,  $I(c) = \mathfrak{I}(c)$  ;
- 3) Existe un mundo  $w$  tal que, para cada símbolo  $p$  del predicado,  $I$  mapea el predicado a una extensión admisible de  $\mathfrak{I}(p)$ , es decir, existe una relación conceptual  $\rho$  tal que  $\mathfrak{I}(p) = \rho \wedge \rho(w) = I(p)$ .

El conjunto  $I_K(L)$  de todos los modelos de  $L$  que sean compatibles con  $K$  será llamado el sistema de *modelos previstos* (intended models) de  $L$  de acuerdo a  $K$ .

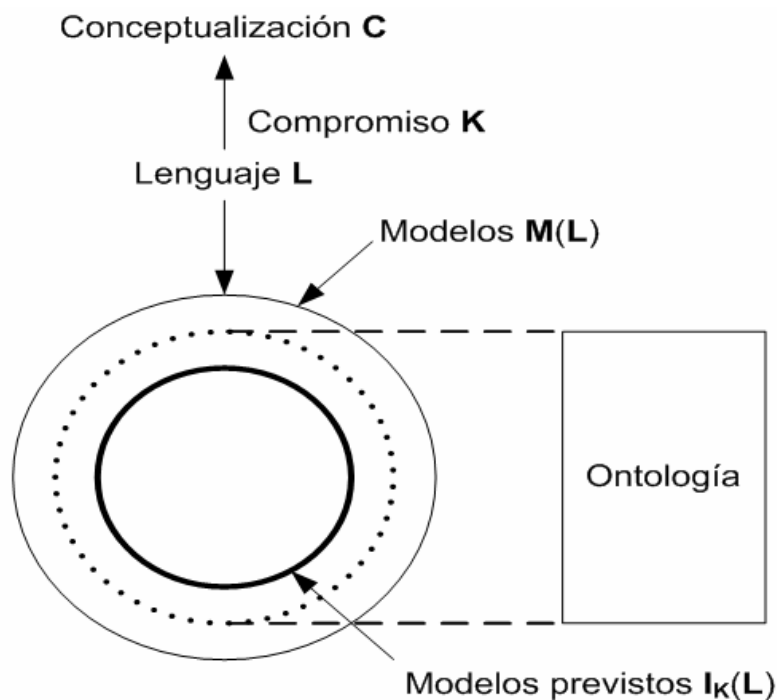
En general, no habrá forma de reconstruir el compromiso ontológico de un lenguaje a partir de un conjunto de sus modelos previstos, dado que un modelo no refleja necesariamente un mundo particular; de hecho, debido a que las relaciones relevantes consideradas pueden no ser suficientes para caracterizar totalmente una situación, un modelo puede describir realmente una situación común para *muchos* estados. Esto significa que es imposible reconstruir la correspondencia entre los mundos y las relaciones *extensionales* establecidas por la conceptualización subyacente. Un conjunto de modelos previstos es, por lo tanto, solamente una caracterización débil de una conceptualización: apenas excluye algunas interpretaciones absurdas, sin realmente describir el "significado" del vocabulario.

El papel de una ontología es considerado como un sistema de axiomas lógicos diseñados para explicar el *significado previsto* de un vocabulario. Dado un lenguaje  $L$  con el

compromiso ontológico  $\mathbf{K}$ , una ontología para  $\mathbf{L}$  es un conjunto de axiomas diseñados de manera que el conjunto de modelos se aproxima tanto como es posible al sistema de modelos previstos de  $\mathbf{L}$  según  $\mathbf{K}$  (Ver Figura 2.11).

En general, no es fácil (ni siempre conveniente) encontrar el sistema correcto de axiomas de modo que una ontología pueda admitir otros modelos previstos. Por lo tanto, una ontología puede “especificar” una conceptualización solamente en una manera muy indirecta, debido a que:

- 1) Puede aproximar solamente a un sistema de modelos previstos;
- 2) El sistema de modelos previstos es solamente una caracterización débil de una conceptualización.



**Figura 2.11** Los *modelos previstos* de un lenguaje reflejan el compromiso de la conceptualización. Una ontología indirectamente refleja el compromiso (y la conceptualización) aproximando este conjunto de modelos previstos

Se puede decir que una ontología  $\mathbf{O}$  para un lenguaje  $\mathbf{L}$  aproxima una conceptualización  $\mathbf{C}$  si existe un compromiso ontológico  $\mathbf{K} = \langle \mathbf{C}, \mathfrak{S} \rangle$  tales que los modelos previstos de  $\mathbf{L}$  que comprometen a  $\mathbf{K}$  están incluidos en los modelos de  $\mathbf{O}$ . Una ontología compromete a  $\mathbf{C}$  si:

- 1) Se ha diseñado con el propósito de caracterizar a **C**;
- 2) Aproxima a **C**. Un lenguaje **L** compromete una ontología **O** si compromete a una cierta conceptualización **C** tal que **O** está de acuerdo con **C**.

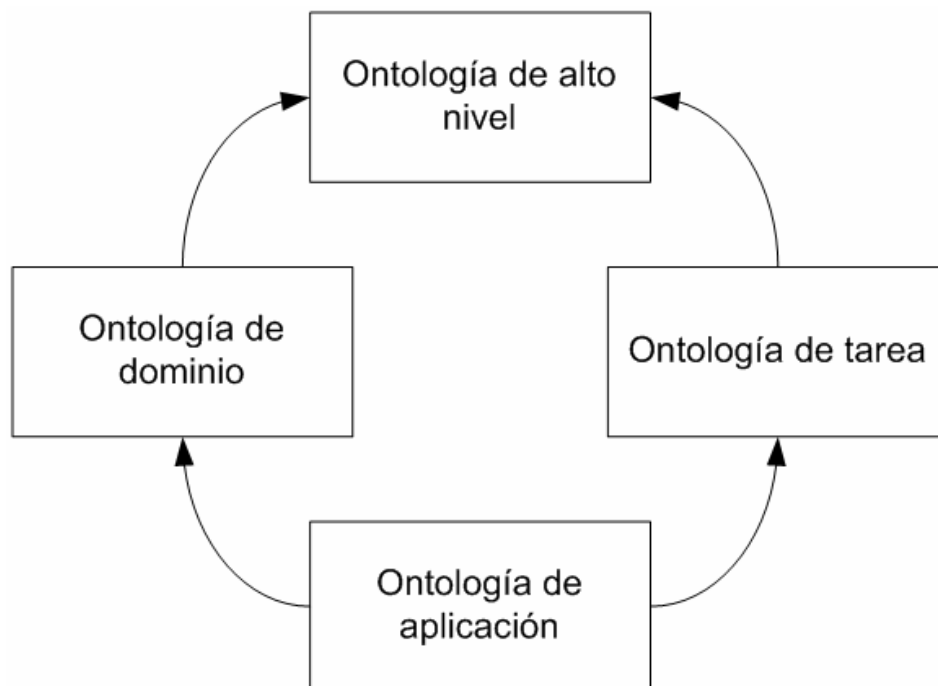
Las relaciones entre vocabulario, conceptualización, compromiso ontológico y ontología se ilustran en la Figura 2.11. Es importante destacar que una ontología es *dependiente del lenguaje*, mientras que una conceptualización es *independiente del lenguaje*. De hecho en la IA, el término "ontología" se colapsa en esos dos aspectos, pero una separación clara entre ellos llega a ser esencial para tratar los asuntos relacionados con compartir, fusionar, y traducir ontologías, que en general implican múltiples vocabularios y conceptualizaciones. Considerando lo anterior, podemos definir una ontología de manera que exista una clara distinción entre los términos ontología y conceptualización. Así una ontología se define como: *Una teoría lógica que explica el significado previsto de un vocabulario formal, es decir, su compromiso ontológico para una conceptualización en particular del mundo. Los modelos previstos del lenguaje usados como vocabulario son restringidos por su compromiso ontológico. Una ontología indirectamente refleja su compromiso (y el de la conceptualización subyacente) aproximando los modelos previstos.*

Podemos clasificar a las ontologías según su exactitud para caracterizar la conceptualización. Hay dos maneras posibles en que una ontología puede estar más cercana a una conceptualización: desarrollando una axiomatización más rica, y adoptando un dominio más rico y/o un sistema más rico de *relaciones conceptuales relevantes*. En el primer caso, la distancia entre el conjunto de modelos de la ontología y el conjunto de modelos previstos se reduce. En el segundo caso, es posible (al menos en principio) incluir en el conjunto de *relaciones conceptuales relevantes* aquellas relaciones que caractericen un estado del mundo, ampliando al mismo tiempo el dominio para incluir las entidades involucradas con tales relaciones; por ejemplo, en la caja de los bloques, podemos considerar la localización espacial de un bloque como una relación conceptual relevante, incluyendo las localizaciones en el dominio, y considerando una relación *sobre*( $x, y$ ) como totalmente derivable de las localizaciones de  $x$  y  $y$ . Debido a que cada modelo lleva la información referente al estado del mundo al que se refiere, la conceptualización

subyacente se puede reconstruir a partir del conjunto de sus modelos previstos. En este caso, si una ontología es axiomatizada de manera que tenga exactamente los mismos modelos, entonces sería una ontología "perfecta".

Una *ontología fina* consigue una especificación más cercana del significado previsto de un vocabulario (y por lo tanto puede ser utilizado para establecer consenso en relación a compartir dicho vocabulario, o una base de conocimientos que utilice ese vocabulario), pero puede ser difícil de desarrollar, debido al número de axiomas y la expresividad del lenguaje adoptado. En la representación que tiene información muy detallada se utiliza también el término *ontología de bajo nivel (low-level ontology)* [Fonseca, 2001]. Una *ontología tosca*, puede consistir de un conjunto mínimo de axiomas escritos en un lenguaje de expresividad mínima, para soportar solamente un conjunto limitado de servicios específicos, previstos para ser compartido entre los usuarios que están de acuerdo con la conceptualización subyacente. Es una representación de información general y se conoce también como *ontología de alto nivel (high-level ontology)* [Fonseca, 2001].

Las ontologías también son clasificadas de acuerdo a su dependencia de una tarea específica o de un punto de vista. (Ver Figura 2.12)



**Figura 2.12** Tipos de ontologías de acuerdo a su nivel de dependencia para una tarea en particular (Las líneas representan relaciones de especialización)

- Ontología de alto nivel (*Top-level Ontology*). Describe conceptos muy generales, tales como espacio, tiempo, materia, objeto, eventos, acciones, etc., que son independientes de un dominio o problema en particular; por lo tanto, parece razonable, por lo menos en teoría, tener ontologías a nivel superior unificadas para grandes comunidades de usuarios. Por ejemplo, la teoría que describe la relación entre las partes y el todo, denominada mereotopología [Smith, 1995].
- Ontología de Dominio (*Domain Ontology*). Describe el vocabulario relacionado con un dominio genérico, tales como medicina o automóviles, especializando los términos introducidos en la ontología de nivel superior.
- Ontología de Tarea (*Task Ontology*). Describe el vocabulario relacionado con una tarea en específico, tales como un diagnóstico o ventas, especializando los términos introducidos en la ontología de nivel superior.
- Ontología de Aplicación (*Application Ontology*). Describe conceptos tanto de un dominio como de una tarea en particular, que frecuentemente son especializaciones de ambos. Estos conceptos corresponden a los roles desempeñados por las entidades del dominio mientras realizan cierta actividad, como unidad reemplazable o componente de repuesto. Representan las necesidades del usuario con respecto a un uso específico, tal como la valoración de la contaminación en la Cd. de México.

## 2.4 Discusión sobre el estado del Arte

Los trabajos que en esta sección se han presentado son de diversas áreas. En la parte que corresponde a la generalización podemos resaltar algunos puntos:

- Los modelos: Estos modelos tienen como fin descomponer en etapas funcionales el proceso de generalización. Pretenden ser muy semánticos, ya que por sí misma la generalización tiene un gran componente semántico. Sin embargo, la generalización no pudo resolverse en su totalidad empleando dichos modelos.

- La Automatización: Básicamente es procesamiento numérico, que resolvió muchos de los problemas de índole geométrico; sin embargo, estos originaron inconsistencias en la base de datos.
- Conocimiento espacial en generalización: En esta etapa se comenzaron a desarrollar aplicaciones que colocaban un toque semántico al procesamiento de los datos, sin embargo, no se trataron cuestiones relacionadas al contexto de los datos, simplemente se desarrollaron mecanismos generales. Sin embargo, se sentaron las bases de lo que hoy se desarrolla en el área de representación del conocimiento espacial.
- Ontologías para generalización: A nuestro conocimiento el trabajo [Kulik et al., 2005] es el primero enfocado a generalizar que incluye una ontología para generalización. Éste trata aspectos sobre el propósito del mapa. Sin embargo, sólo está enfocado al procesamiento de redes de transporte.

En la etapa relacionada con la Representación de Conocimiento Espacial se presentaron dos enfoques, el primero de ellos, orientado a la semiótica, que se refiere en gran parte al significado de símbolos en los datos espaciales. Por otro lado, el enfoque de los mapas de aspectos propone una estructura compuesta por el conjunto de entidades que contiene el mapa, el conjunto de relaciones que tienen dichas entidades y contiene la posición de las entidades así como también la referencia al meta-conocimiento, lo cual se parece mucho a una ontología. Además, se presenta una estructura jerárquica donde se describen los diversos ángulos que pueden existir en una representación pictórica. Asimismo, para considerar de forma general las propiedades de la representación, presenta una jerarquía de precedencia pictórica, en donde aparecen de forma explícita los detalles de mapa de la red del metro, lo que representa, que con este enfoque se pueden considerar los requerimientos de la aplicación.

El resto del estado del arte está enfocado a describir el panorama en general de las ontologías, presentando los aspectos más relevantes, como conceptualización, compromiso ontológico, dominio, relaciones relevantes, etc.

Como se ha visto en esta sección, la evolución del procesamiento de cartografía digital necesita enfocarse a ciertos aspectos de índole semántica, ya que por su naturaleza necesita

incluirse. De acuerdo a la Sección 2.3, podemos ver a una ontología como la descripción de un dominio en particular. Desde ese punto de vista, la ontología nos permite modelar el dominio geográfico de manera que nos sea útil en la solución de problemas que normalmente se han abordado con un enfoque numérico. Así mismo, la ontología puede desarrollarse con un propósito en específico según se requiera, cubriendo un dominio muy particular o uno muy general.

Adicionalmente, la ontología representa el conocimiento que tiene un experto o un grupo de expertos sobre un dominio. Considerando que el caso de estudio de esta tesis está enfocado a la generalización. Y como se ha presentado en este capítulo, la generalización es un proceso que depende en gran parte del usuario experto. El cual se encarga de la definición de los criterios de representación, para decir dónde y cómo generalizar; por ende, también se encarga de definir las reglas de consistencia en los datos geográficos generalizados. Podemos decir que la ontología es una alternativa viable para definir las reglas de consistencia que deben cumplir los datos generalizados. Es decir, las ontologías permiten definir una convergencia entre representaciones, descripciones e interpretaciones de los datos geográficos. Ya que una ontología tiene las siguientes características: define con claridad los términos de un dominio, describe sus detalles, relaciones entre entidades, es consistente, reduce la codificación y puede ser extendida tanto como se requiera.



# CAPÍTULO 3

## MARCO TEÓRICO

En este capítulo se describirán algunos de los términos que se utilizarán durante el desarrollo de la metodología. Primeramente se describirá el concepto de topología geoespacial, así como también algunos conceptos básicos del modelo vectorial, con el fin de sustentar el desarrollo de la metodología. Posteriormente, se describirán conceptos fundamentales relacionados con las mediciones en los datos geográficos las que serán utilizadas para extraer las propiedades implícitas de los datos. Finalmente, se definirán los aspectos básicos sobre la Teoría de la Confusión y jerarquías, los cuales serán utilizados para medir la similitud entre datos geográficos.

### 3.1 Topología Geoespacial

La topología geoespacial es la aplicación de la topología matemática en los problemas geoespaciales. La topología tiene su origen en el hecho de que algunos problemas geométricos no dependen de la forma exacta de los objetos implicados, sino se enfoca a “la manera en que se conectan” [Wikipedia, 2007]. Uno de los primeros artículos en topología fue la demostración, por Leonhard Euler, consistente en encontrar una ruta óptima que atravesase los siete puentes de la ciudad de Königsberg (ahora Kaliningrad) cruzando exactamente una vez por cada uno de ellos. El resultado no depende de las longitudes de los puentes, ni de la distancia, sino solamente de características de conectividad: ¿qué puentes están conectados con cuáles islas?. El problema de los siete puentes de Königsberg ahora es famoso en matemáticas, y tratado por la rama de las matemáticas como teoría de grafos.

La Topología es una disciplina matemática que estudia las propiedades de los espacios topológicos<sup>12</sup> y las funciones continuas [Wikipedia, 2007]. La Topología se interesa por conceptos tales como número de agujeros, conectividad, compacidad, adyacencia, etc.

La Geometría es un aspecto que poseen las representaciones de datos geográficos, que cambia cuando la información es transformada de un sistema de referencia o sistema de coordenadas a otro. Por otro lado, la topología trata con las características geométricas que permanecen invariantes si el espacio es deformado, por ejemplo, cuando los datos geográficos son transformados de un sistema de coordenadas a otro. En el contexto de los datos geográficos, la Topología es comúnmente usada para describir la conectividad de un grafo<sup>13</sup> de  $n$  dimensiones, una propiedad que es invariante bajo transformaciones continuas del grafo [OGC, 2005].

Podemos decir que la topología se refiere a las relaciones que existen entre los objetos geográficos, es decir, si están conectados, si son adyacentes, etc. Consecuentemente, las relaciones topológicas son las relaciones entre los objetos espaciales y son caracterizadas por ser invariantes a transformaciones geométricas del espacio, tales como traslación, rotación y escalamiento [Egenhofer, 1989].

### 3.1.1 Estructura vectorial

Los elementos geométricos de la estructura espacial en una representación vectorial pueden ser interpretados como los componentes de un grafo (Ver Figura 3.1). Un grafo consiste de nodos conectados por arcos o bordes. Con estos elementos, se pueden formar cadenas, las cuales pueden ser interpretadas como polilíneas en un sentido geométrico. Las polilíneas pueden ser combinadas para formar anillos, circuitos cerrados o polígonos que encierran caras o segmentos de área.

---

<sup>12</sup> El espacio topológico es la noción base de la topología elemental, dominio que sólo depende de la teoría de conjuntos (no está construido a partir de otra cosa), y que tiene consecuencias importantes en el campo del análisis porque permite definir rigurosamente la continuidad y los límites.

<sup>13</sup> En el modelo vectorial, los datos son modelados como si se tratara de un grafo.

En la literatura pueden encontrarse diferentes nombres o expresiones para los componentes geométricos de un grafo, como se indican en la Figura 3.1. Las siguientes convenciones son usadas para evitar ambigüedades:

- Un punto en un grafo es llamado *nodo*.
- Un par de nodos conectados puede ser llamado *arista* o *borde*.
- Una cadena de aristas puede ser llamada *arco* o *arcos*.
- Un segmento de grafo es una cadena de arcos que no contiene nodos y más de dos arcos están conectados.
- Una combinación de arcos que forman un circuito cerrado puede ser llamado *polígono*.
- Un polígono encierra un *segmento de área*.
- Un segmento de área que no es intersectado por un polígono es llamado *cara*.

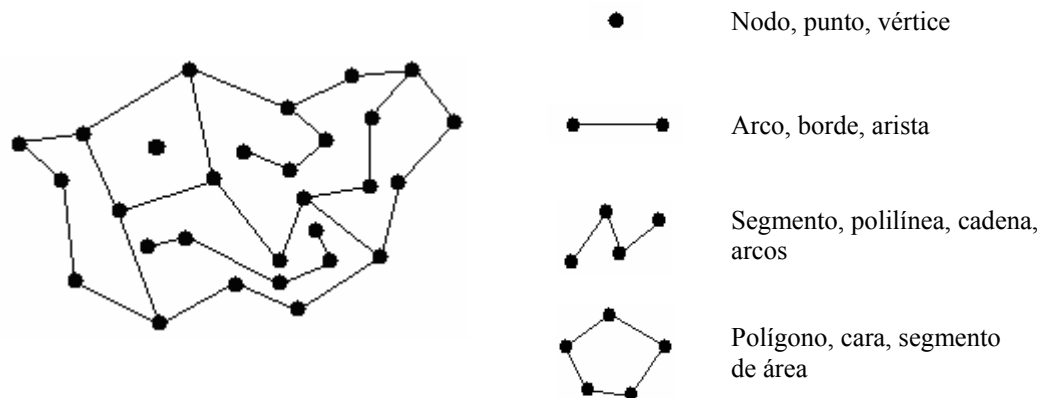


Figura 3.1 Componentes del modelo vectorial.

Considerando la geometría de los objetos geográficos o espaciales, éstos pueden clasificarse en tres tipos<sup>14</sup>: *objetos punto*, *objetos línea* y *objetos área*. Los objetos punto no tienen dimensión espacial, únicamente tienen una posición y pueden ser representados por nodos. Los objetos línea tienen una dimensión espacial, ya que tienen longitud y forma, y pueden ser representados por arcos. Los objetos área tienen dos dimensiones espaciales y pueden ser representados por polígonos.

<sup>14</sup> Primitivas geométricas de representación

### 3.1.2 Representación de los datos vectoriales

Existen tres conjuntos de datos que se almacenan para los objetos espaciales en un GIS:

- Identificadores para cada objeto espacial.
- Atributos descriptivos.
- Geometría de los datos espaciales.

Éstos deben estar ligados entre sus diferentes tipos de datos; de manera que las relaciones entre las capas de datos (datos espaciales) y los datos geométricos puedan ser analizadas. Dichos datos son conectados a través de los identificadores de los objetos espaciales (Ver Figura 3.2). Las relaciones topológicas entre los objetos espaciales pueden ser manipuladas en una geometría estructurada vectorialmente.

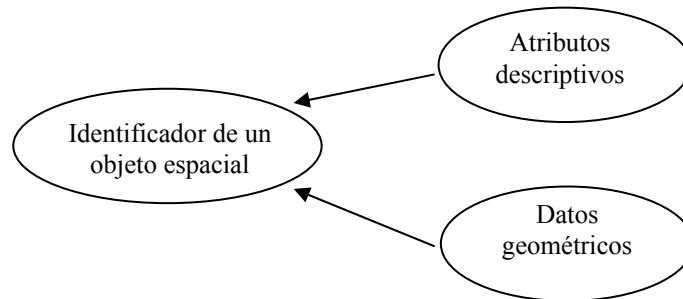


Figura 3.2 Datos geográficos en un GIS

### 3.1.3 Conceptos básicos de la teoría de grafos para analizar las relaciones entre los objetos geográficos

Para el desarrollo de la presente investigación es muy importante definir algunos conceptos de la teoría de grafos, ya que dicha teoría es la base para analizar representaciones vectoriales de los datos geográficos. Un grafo consiste de dos conjuntos:

- Un conjunto de nodos  $N = \{n_1, n_2, \dots, n_N\}$
- Un conjunto de arcos  $E = \{e_1, e_2, \dots, e_E\}$

Un arco es la conexión de dos nodos:

$$e_i = \{n_p, n_q\} \text{ con } e_i \in E \text{ y } n_p, n_q \in N$$

Dos nodos son *adyacentes* si están conectados por un arco. La secuencia de los nodos es de importancia, ya que:

$$\{n_p, n_q\} \neq \{n_q, n_p\}$$

Entonces los arcos son *dirigidos*, esto es:

$$e_i = \{n_p, n_q\}$$

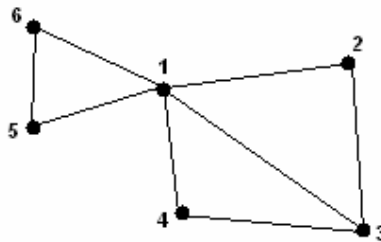
donde:

$n_p$  es el nodo inicial de  $e_i$

$n_q$  es el nodo final de  $e_i$

La información de posición está dada por las coordenadas de los nodos ( $n_p(x_1, y_1)$  y  $n_q(x_2, y_2)$ ). La Figura 3.3 muestra un ejemplo de lo que es un grafo  $G$ . Un grafo  $G$  está dado por un conjunto de nodos  $N$  y un conjunto de arcos  $E$ :

$$G = \{N, E\}$$



$$G_1 = \{N_1, E_1\}$$

$$N_1 = \{1(x,y), 2(x,y), 3(x,y), 4(x,y), 5(x,y), 6(x,y)\}$$

$$E_1 = \{\{1, 2\}, \{1, 3\}, \{1, 4\}, \{1, 5\}, \{1, 6\}, \{2, 3\}, \{3, 4\}, \{5, 6\}\}$$

**Figura 3.3.** Ejemplo de un grafo

Adicionalmente a estos conceptos se tienen los siguientes:

- Dos *arcos son adyacentes* si tienen un nodo en común.
- Si un nodo  $n_p$  ocurre en  $m$  arcos, entonces el *grado de un nodo*(DEGREE) es  $DEGREE(n_p) = m$ , por lo tanto si  $DEGREE(n_p) = 0$ , entonces el nodo no ocurre en ningún arco.

### 3.1.3.1 Relaciones entre los elementos de un grafo plano

La formalización de conceptos presentado en [Molenaar, 1989] [Molenaar et al., 1994] [Molenaar, 1998] está basada en el uso de seis tipos de entidades:

- Tres tipos geométricos: *nodos*, *bordes* y *caras*.
- Tres tipos de objetos geométricos: *objetos punto*, *objetos línea* y *objetos área*.

### 3.1.3.1.1 Relaciones entre nodos y arcos

Se pueden definir las siguientes relaciones entre los elementos geométricos de un grafo plano:

- Un arco  $e_i$  tiene un nodo  $n_j$  como el nodo inicio  $\rightarrow \text{Begin}[e_i, n_j] = 1$
- Un arco  $e_i$  tiene un nodo  $n_k$  como el nodo final  $\rightarrow \text{End}[e_i, n_j] = 1$

Los arcos deben tener nodos inicial y final distintos. Si  $n_j$  es un nodo del arco  $e_i$ , éste puede ser encontrado a través de la siguiente función:

$$N[e_i, n_j] = \text{Begin}[e_i, n_j] + \text{End}[e_i, n_j]$$

Si  $N[e_i, n_j] = 1$  entonces  $n_j$  es un nodo de  $e_i$ ; si  $N = 0$  entonces el nodo no es parte del arco.

El grado de un nodo puede ser encontrado a través de la siguiente función:

$$\text{DEGREE}(n_j) = \sum_i (N[e_i, n_j])$$

### 3.1.3.1.2 Relaciones entre arcos y caras

Cada arco tendrá una cara exactamente a su lado izquierdo y exactamente una a su lado derecho. Esta relación puede ser expresada por la siguiente función:

- Un arco  $e_i$  tiene un objeto área  $f_a$  a su lado izquierdo  $\Rightarrow \text{Le}[e_i, f_a] = 1$ , para cualquier  $f_b \neq f_a \rightarrow \text{Le}[e_i, f_b] = 0$
- Un arco  $e_i$  tiene un objeto área  $f_a$  a su lado derecho  $\Rightarrow \text{Ri}[e_i, f_a] = 1$ , para cualquier  $f_b \neq f_a \rightarrow \text{Ri}[e_i, f_b] = 0$

Si un arco  $e_i$  es parte del límite de un objeto área  $f_a$  entonces sólo una de las funciones  $Ri$  y  $Le$  es igual a 1, pero no ambas. Esto se define con la siguiente función:

$$B[e_i, f_a] = \text{Le}[e_i, f_a] + \text{Ri}[e_i, f_a]$$

Entonces si  $e_i$  es parte del límite de  $f_a$ , el valor de  $B[e_i, f_a] = 1$ . El límite de un objeto área  $f_a$  es:

$$\partial f_a = \{N_a, E_a\}$$

$$E_a = \{e \mid B[e, f_a] = 1\}$$

$N_a$  contiene los nodos de los arcos de  $E_a$

Los arcos que conforman la frontera de una cara forman un *polígono*. Tal que para cualquier arco  $e_i$  que es parte de un polígono en un grafo plano hay solo dos caras, tal que  $B[e_i, f] = 1$ , y para todas las otras caras  $B[e_i, f_a] = 0$ .

### 3.1.3.1.3 Relaciones entre nodos y caras

Si el grado de un nodo  $n_i$  es  $DEGREE[n_i] = 0$ , entonces no está relacionado con algún arco. En tal caso debe de estar dentro de una cara  $f_a$ . Esta relación es expresada con la siguiente función:

$$ISIN[n_i, f_a] = 1$$

Si  $DEGREE[n_i] \neq 0$ , entonces existe algún arco  $e_j$  tal que  $N[e_j, n_i] \neq 0$ .

### 3.1.3.1.4 Relaciones entre cara - cara

Las caras de un grafo plano no deben traslaparse, tal que sólo pueden estar relacionadas por la relación de adyacencia. Para dos caras  $f_1$  con límite  $\partial f_1 = \{N_1, E_1\}$  y  $f_2$  con  $\partial f_2 = \{N_2, E_2\}$  hay cuatro posibilidades para la intersección de sus límites (Ver Figura 3.4):

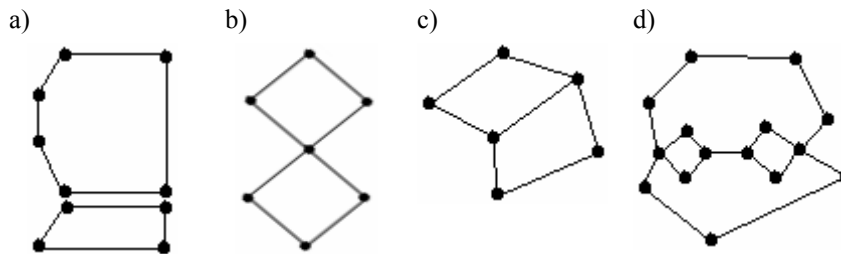


Figura 3.4 Conexión geométrica de dos caras

- No existe conexión geométrica si las caras no tienen arcos o nodos en común (Ver Figura 3.4a), por lo tanto:

$$N_1 \cap N_2 = \emptyset \quad \text{y} \quad E_1 \cap E_2 = \emptyset$$

- Las caras se tocan si tienen uno o más nodos en común, pero no tienen arcos en común (Ver Figura 3.4b), por lo tanto:

$$N_1 \cap N_2 \neq \emptyset \quad \text{y} \quad E_1 \cap E_2 = \emptyset$$

- Dos caras son *adyacentes* si tienen uno o más arcos en común (Ver Figura 3.4c), en tal caso los objetos también tienen nodos en común; por lo tanto, es suficiente decir:

$$E_1 \cap E_2 \neq \emptyset$$

- Las caras pueden tener una combinación de relaciones, tales como *toca* y *adyacente* (Ver Figura 3.4d). Esto significa que tienen nodos en común que no pertenecen a los arcos en común.

### 3.1.4. Descripción de objetos geográficos usando el modelo vectorial

La geometría de datos vectoriales se define a través de conjuntos de elementos geométricos; es decir, la geometría de un conjunto de datos vectorial  $M$  puede ser representada por:

$$GEOM(M) = \{G_M, F_M\} = \{N_M, E_M, F_M\},$$

donde:  $G_M$  es el grafo que representa a los datos vectoriales

$N_M$  es el conjunto de todos los nodos de la capa

$E_M$  es el conjunto de todos los arcos de la capa

$F_M$  es el conjunto de todas las caras formadas por  $G_M$

La totalidad de objetos representados en un conjunto de datos vectorial  $M$  es llamado *universo de discurso* y es denotado por  $U_M$ . Este universo contiene tres subconjuntos:

$$U_M = \{PO_M, LO_M, AO_M\},$$

donde:  $PO_M$  es el conjunto de todos los objetos punto de  $M$

$LO_M$  es el conjunto de todos los objetos línea de  $M$

$AO_M$  es el conjunto de todos los objetos área de  $M$

Existen dos categorías principales de modelos de datos para estructurar los datos geográficos, éstos son el modelo orientado a campos y el modelo orientado a objetos.



En el *modelo orientado a objetos* los datos son organizados por objetos; es decir, a través de los identificadores de la temática de datos y de los datos geométricos los objetos son relacionados, y pueden ser accedidos por los identificadores.

En el *modelo orientado a campos*, los atributos del terreno son especificados para cada posición, debido a que el valor que posea cada atributo depende de la posición. El caso de la posición se define por una celda raster, y el valor del atributo es asignado al área cubierta por la celda.

### 3.1.4.1 Relaciones topológicas en datos vectoriales

La descripción geométrica de objetos espaciales está dada en términos de nodos, bordes y caras. Los componentes topológicos de cada objeto espacial son subconjuntos de estos elementos geométricos. El conjunto universo de puntos está definido por el conjunto de datos, por ejemplo,  $\mathbb{R}^2$ , son representados por  $Y$ .

Los componentes topológicos están dados para objetos elementales. Los objetos complejos pueden ser construidos a partir de los objetos elementales, de acuerdo con reglas que deben ser especificadas. Con estas reglas debe ser posible encontrar los componentes topológicos de los objetos complejos. La geometría de un objeto punto  $O_p$  es representada por un nodo que puede ser caracterizado como un *conjunto punto* con dimensión 0. De esta forma tenemos:

$$\begin{array}{ll}
 \textit{límite relativo} & \partial O_p = \{n_p | Repr[n_p, O_p] = 1\} \\
 \textit{interior} & (\textit{no está definido}) \\
 \textit{exterior} & \bar{O}_p = Y - \partial O_p \\
 & Y \approx \mathbb{R}^2
 \end{array}$$

La geometría de un objeto línea  $O_l$  puede ser representada por una cadena o polilínea que es un conjunto de puntos con una dimensión = 1. Los componentes topológicos son:

$$\begin{aligned}
 \text{límite relativo} & \quad \partial O_l = \{n_i \in N_l : DEGREE_l(n_i) = 1\} \\
 \text{interior relativo} & \quad \circ O_l = G_l - \partial l = \{N_l - \partial O_l, E_l\} \\
 \text{exterior} & \quad \bar{O}_l = Y - (\partial O_l \cup \circ O_l) \\
 & \quad Y \approx \mathbb{R}^2
 \end{aligned}$$

La geometría de un objeto área  $O_a$  puede ser representada por un segmento que consiste de un conjunto contiguo de caras, nodos y arcos relacionados. La geometría de un objeto área puede ser representada por  $GEOM(O_a) = \{N_a, E_a, F_a\}$ . Los componentes topológicos son:

$$\begin{aligned}
 \text{límite} & \quad \partial O_a = \{N_{\partial O_a}, E_{\partial O_a}\} \\
 & \quad N_{\partial O_a} = \{n_j \mid n_j \in e_i, e_i \in E_{\partial O_a}\} \\
 & \quad E_{\partial O_a} = \{e_i \mid B[e_i, O_a] = 1\} \\
 \text{interior} & \quad \circ O_a = \{N_{\circ O_a}, E_{\circ O_a}, F_a\} \\
 & \quad E_{\circ O_a} = \{e_i \mid B[e_i, O_a] = 2\} \\
 & \quad N_{\circ O_a} = \{n_j \mid n_j \in e_i, e_i \in E_{\circ O_a}\} \\
 & \quad F_a = \{f_i \mid Part_{22}[f_i, O_a] = 1\} \\
 \text{exterior} & \quad \bar{O}_a = Y - (\partial O_a \cup \circ O_a) \\
 & \quad Y \approx \mathbb{R}^2
 \end{aligned}$$

Los nodos y arcos relacionados con el interior de un objeto área pueden ser encontrados a través de las caras que definen el interior del objeto área. La descripción geométrica de un objeto  $O_i$  se describe:

$$GEOM(O_i) = \{N_{O_i}, E_{O_i}, F_{O_i}\}$$

### 3.1.5 Componentes topológicos

A continuación, se describirán los componentes topológicos de los objetos geográficos, con el fin de formalizar las relaciones topológicas entre dichos objetos, empleando los modelos

basados en intersecciones. Se considera un conjunto de puntos  $S_I$  mostrado en la Figura 3.5 donde se pueden identificar tres tipos de puntos en este conjunto.

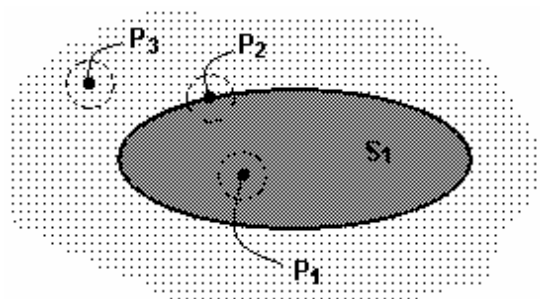


Figura 3.5 Topología de conjunto de puntos en  $\mathbb{R}^2$

Los componentes topológicos se describen como sigue:

El primer tipo representado por  $P_1$ . Para el conjunto de puntos  $U_P$  que únicamente contiene puntos que pertenecen a  $S_I$ , es decir,  $U_P \subset S_I$ . Estos puntos son llamados *puntos interiores* de  $S_I$ . Entonces el interior de un conjunto  $S$  es:

$$INTERIOR(S) = S^\circ = \{P \in S \mid \exists U_P \Rightarrow U_P \subset S\}$$

Por lo tanto, el interior de  $S$  es el conjunto de todos los puntos de  $S$  que están contenidos en  $S$ .

El segundo tipo representado por  $P_3$ . Los puntos de este tipo no pertenecen a  $S_I$  y el conjunto de puntos  $U$  son todos los puntos que no pertenecen a  $S_I$ , es decir,  $U_P \cap S_I = \emptyset$ . Estos puntos son llamados *puntos exteriores* de  $S_I$ . Entonces el exterior de  $S$  es:

$$EXTERIOR(S) = S^c = \{P \notin S \mid \exists U_P \Rightarrow U_P \cap S = \emptyset\}$$

Por lo tanto, el exterior de  $S$  es el conjunto de todos los puntos que no pertenecen a  $S$  y que no contienen ningún punto de  $S$ .

El tercer tipo representado por  $P_2$ , los puntos de este tipo son los puntos  $U_P$  que contienen puntos que pertenecen a  $S_I$  y puntos que no pertenecen al interior de  $S_I$  (es decir, estos puntos pertenecen al conjunto complemento  $COMPLEMENT(S_I) = (R_2 - S_I)$ ). Estos puntos son llamados *puntos frontera*. Entonces la frontera de un conjunto  $S$  es:

$$FRONTIER(S) = {}^F S = \{P \in R_2 \mid \forall U_P \Rightarrow U_P \cap S \neq \emptyset \\ \text{and } U_P \cap COMPLEMENT(S) \neq \emptyset\}$$

Consecuentemente hay puntos que satisfacen la definición de puntos frontera, que sin embargo difieren del caso del punto  $P_2$  de la Figura 3.4, puesto que contienen puntos del *EXTERIOR* ( $S_I$ ). Puntos como el caso de  $P_2$  son llamados *puntos límite* de  $S_I$ . Entonces el límite de un conjunto  $S$  es:

$$BOUNDARY(S) = \partial S = \{P \in {}^F S \mid \forall U_P \Rightarrow U_P \cap S \neq \emptyset \text{ and } U_P \cap \bar{S} \neq \emptyset\}$$

Esto significa que la relación entre  $S$  y su exterior sólo puede encontrarse a través de sus límites. Los puntos frontera que no pertenecen al límite se dice que están *dentro* de  $S$ .

### 3.1.6 Modelo de 9-intersección

Una vez que se han definido los componentes del modelo vectorial, podemos presentar el modelo que nos servirá de base para describir el conjunto de relaciones que se proponen en esta investigación. El modelo de 9-intersección fue propuesto por [Egenhofer, 1990] [Egenhofer y Franzosa, 1991] [Egenhofer et al., 1994] para formalizar relaciones topológicas binarias entre dos objetos arbitrarios. En dicho modelo, un objeto  $O_i$  es modelado como un conjunto de puntos compuesto por tres componentes topológicos (Ver Sección 3.1.5):

- El límite de  $O_i$  representado por  $\partial O_i$
- El interior de  $O_i$  representado por  $O_i^\circ$
- El exterior de  $O_i$  representado por  $O_i^-$

El modelo se basa en la evaluación del conjunto intersección ( $\cap$ ) para cada uno de los 9 elementos de una matriz de 3 por 3 entre dos conjuntos de puntos arbitrarios  $O_1$  y  $O_2$ ; cada elemento presenta una intersección entre un componente (frontera, interior o exterior) de  $O_1$  y un componente de  $O_2$  (Ver Figura 3.6).

$$\begin{bmatrix} \partial O_1 \cap \partial O_2 & \partial O_1 \cap O_2^\circ & \partial O_1 \cap O_2^- \\ O_1^\circ \cap \partial O_2 & O_1^\circ \cap O_2^\circ & O_1^\circ \cap O_2^- \\ O_1^- \cap \partial O_2 & O_1^- \cap O_2^\circ & O_1^- \cap O_2^- \end{bmatrix}$$

Figura 3.6 Matriz de 9-intersección

Esta matriz es llamada *matriz de intersección para componentes topológicos de objetos* o simplemente *matriz de intersección*. Cada elemento de esta matriz puede evaluarse como *vacío* ( $\emptyset$ ) o *no vacío* ( $\neg\emptyset$ ). Esto nos da un total de  $2^9$  relaciones binarias candidatas mutuamente excluyentes entre dos objetos arbitrarios. No todas las 512 relaciones topológicas candidatas son posibles. Por ejemplo, cualquiera de las 512 relaciones candidatas entre dos objetos punto A y B en donde  $\partial A \cap \partial B = \neg\emptyset$  y  $\partial A \cap B^- = \neg\emptyset$  son relaciones topológicamente imposibles, ya que esto significaría que A y B coinciden y no coinciden al mismo tiempo. Sin embargo, algunas de las relaciones candidatas son inconsistentes con el modelo de datos. Estas relaciones inconsistentes deben ser identificadas y removidas de la lista de relaciones candidatas para procurar un conjunto consistente de relaciones topológicas mutuamente exclusivas entre cualesquiera dos objetos. En [Egenhofer et al., 1994] se muestra una estrategia de pre-procesamiento que puede ser desarrollada para verificar la consistencia en el tráfico de consultas con relaciones topológicas complejas entre conjuntos de más de dos objetos. El conjunto de relaciones que son topológicamente consistentes puede ser reducido por la combinación de relaciones que son topológicamente similares [Kufoniyi, 1995]. Dos relaciones son topológicamente similares si éstas son relaciones inversas, esto es que si es posible cambiar los A en B y los B en A. Estas relaciones inversas sólo pueden ocurrir entre objetos del mismo tipo: punto/punto, línea/línea y área/área. Por ejemplo, en la Figura 3.7 se muestran dos relaciones similares entre dos objetos área.

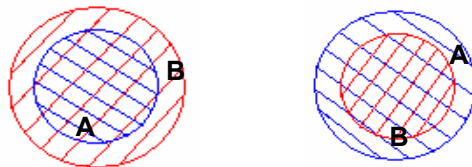


Figura 3.7 Relaciones topológicamente similares

Las relaciones se representan por un código binario que pueda entonces ser convertido a un número decimal. Para obtener los códigos binarios en *tuplas* de 9 elementos, ésta se define como sigue:

$$\partial O_1 \cap \partial O_2 \quad O_1^\circ \cap O_2^\circ \quad O_1^\circ \cap \partial O_2 \quad \partial O_1 \cap O_2^\circ \quad \partial O_1^- \cap O_2^- \quad \partial O_1 \cap O_2^- \quad O_1^\circ \cap O_2^- \quad O_1^- \cap \partial O_2$$

Para distinguir la codificación, una intersección vacía será interpretada como un bit con valor 0 y no vacía con un bit con valor 1. Entonces la representación decimal de la cadena de bits representa a la relación. Así la relación r511 (111 111 111 en binario) representa una tupla, donde todas las 9 intersecciones son no vacías. De la misma forma, la relación r000 (000 000 000) representa la relación en que todas las 9 intersecciones son vacías.

### 3.2 Medidas sobre datos geoespaciales

Una medida es un procedimiento para calcular mediciones<sup>15</sup>, que se utiliza para evaluar las características de una entidad geográfica y en el caso de la generalización se utiliza determinar la necesidad y/o el éxito del proceso. Una medida puede consistir de una fórmula o puede implicar un algoritmo complejo que incluye, el cómputo de estructuras de datos auxiliares o representaciones adicionales.

Una *medida* es un método que no cambia el estado de los objetos geoespaciales, pero se utiliza para caracterizarlo [Bader et al., 1999]. Las *medidas* se utilizan frecuentemente para diversos propósitos, tales como:

- Ayudar a detectar estructuras significativas en los datos geoespaciales, tales como, grupos de los objetos (por ejemplo, alineaciones) o características espaciales (por ejemplo, las formas de objetos y distribución de objetos).
- Agrupar aquellas estructuras, representadas por grupos de objetos, es decir, agrupar situaciones semejantes bajo un mismo concepto.
- Encontrar las peculiaridades de un objeto (por ejemplo, que tan sinuoso o largo).

<sup>15</sup> Medición: Valor numérico asignado a una observación (medida), que refleja la magnitud o la cantidad de una característica [Davis, 1986].

Principalmente las medidas se utilizan para identificar la forma de entidades. Los GIS es solamente una de las áreas donde es importante la forma de los objetos. Se utilizan en visión por computadora, matemáticas, geometría, estadística, geología, geografía y la ciencia cognitiva; en todos los casos buscan medir la naturaleza de las entidades (o el grupo de entidades) y la construcción de nuevas representaciones. Sin embargo, el uso y metas en cada disciplina, son diferentes, por lo que la demanda de las medidas cambia. Por ejemplo, en el campo de la visión por computadora, las técnicas se basan en la necesidad de reconocer y representar formas gráficamente. La forma áspera de objetos tienen que ser detectados y el resultado debe ser estable si el ruido fue agregado. Para el propósito de esta tesis, tenemos los objetos geográficos y estamos interesados en las diferencias que tienen entre ellos. Además de eso, la visión por computadora trata principalmente de representaciones *raster*, mientras que en este trabajo nos centramos principalmente en las representaciones vectoriales.

### 3.2.1 Selección de la medida adecuada

Según [AGENT, 1999], la definición teórica de una buena medida puede ser: "una medida es discriminatoria de la característica observada de un objeto". Sin embargo, esto implicaría que cada una de las características no se relaciona con las otras. Así, esta última suposición y la enunciación exacta de las características necesarias hacen más duro diseñar una medida adecuada.

En [AGENT, 1999] se mencionan siete características que deben cumplir las medidas:

Desde el punto de vista teórico:

1. *Robustez*: Las variaciones pequeñas no deben originar diferencias muy grandes en las mediciones efectuadas. Y es un prerequisite para el punto siguiente
2. *Separabilidad*: Una medida debe capturar tanto como sea posible el aspecto que se está midiendo.

Desde el punto de vista de los procesos:

3. *Invariante de la persona*: Una buena medida debe dar lugar a la misma medición, independientemente de la persona que la aplica. Esto es realizable solamente si la medida es invariante bajo diversas opciones de uso.
4. *Independiente de la representación*: La representación de un objeto geográfico no debe influenciar una medida. Los mismos objetos representados de diferente forma pueden tener como resultado la misma medición.

Desde el punto de vista del usuario:

5. *Facilidad para calcular*: Una buena medida debe ser fácil de calcular
6. *Facilidad de uso*: Una buena medida tiene un conjunto de parámetros limitado, documentado y fácil de utilizar.
7. *Facilidad de Interpretación*: Una buena medida debe ser fácil de interpretar. Por lo tanto, para medidas muy similares una medición similar debe obtenerse. Y el resultado debe estar en alguna escala

### 3.2.2 Tipos de Medidas

[McMaster, 1986] fue uno de los primeros en tratar las medidas en cartografía digital. En una publicación subsecuente, [McMaster y Shea, 1992] proponen un grupo de 7 tipos de medidas: medidas de densidad, medidas de distribución, medidas de longitud y sinuosidad, medidas de forma, medidas de distancia, medidas de Gestalt y medidas abstractas. Por otro lado el Instituto Geográfico Nacional de Francia, propone una clasificación más detallada usando 18 categorías, que parecen ser construidas principalmente para clasificar medidas de caminos [AGENT, 1999].

Debido a que el proceso de medidas es un acto de cuantificar el carácter/propiedad de objetos geográficos, está claro que las medidas se pueden también clasificar usando cualquier clase para la clasificación del carácter de los datos. Por ejemplo, utilizar el tamaño, la distribución, el patrón, la contigüidad, la vecindad, la forma, la escala y la orientación de la clase.

En [Ruas y Plazanet, 1996] se diferenciaron significativamente la posición y la forma. Estos componentes no sólo son fundamentales para describir la calidad de datos



geográficos, sino cruciales para determinar la situación inicial y de los aspectos semánticos esenciales para una distribución particular de entidades cartográficas.

Otra distinción común se hace entre patrón y forma. Mientras que la forma se enfoca a objetos individuales, el patrón describe la distribución geográfica de un grupo de objetos. Por ejemplo, [Campbell, 1993] hace la distinción entre forma, patrón y la organización de los objetos geográficos.

### 3.2.3 Clasificación de medidas

Las medidas sobre objetos geográficos definen las maneras para las caracterizaciones que pueden ser dadas de manera lógica a un objeto (entidad). Por ejemplo la sinuosidad, la longitud y orientación de un río pueden ser calculadas.

La Figura 3.8 muestra una clasificación de medidas para la descripción geométrica. Dicha descripción puede ser obtenida analizando *posición y geometría, forma* o la *topología*. Sin embargo, la geometría que es obtenida por medio de la *posición y geometría y topología* es implícita porque sólo se puede medir la posición y orientación de objetos, como también las relaciones topológicas existentes entre objetos geográficos. La descripción de la geometría por su forma es la más óptima ya que puede representarse explícitamente, porque podemos obtener características específicas de un objeto geográfico como el tamaño, la sinuosidad/complejidad, el alargamiento/extenuidad, así como también la compactibilidad así como otros aspectos importantes.

Las medidas de tamaño están relacionadas con algunas características como: área, perímetro, longitud, y altura de una curva que es obtenida por un par de puntos. La medida de sinuosidad: es la condición de ser curvo o curvando en forma o movimiento. La excentricidad una constante que describe la forma de una sección cónica. Es igual a la proporción de la distancia de un punto fijo con respecto a cualquier otro punto sobre la curva.

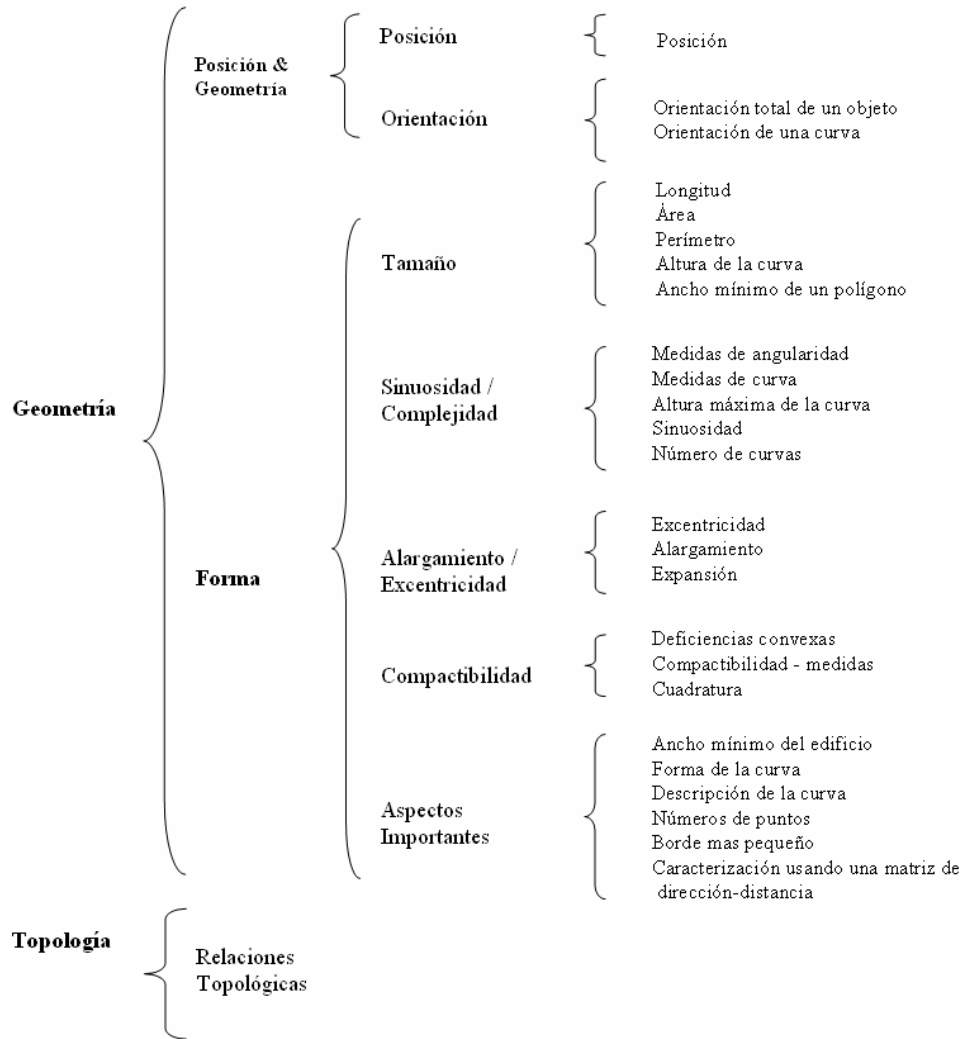


Figura 3.8 Clasificación de las medidas para datos geográficos

### 3.3 Jerarquías y Teoría de la Confusión

En esta sección se presentarán los conceptos básicos de jerarquías así como los fundamentos de Teoría de Confusión.

#### 3.3.1 Concepto de Jerarquía

En [Levachkine y Guzman-Arenas, 2007] aparecen las siguientes definiciones formales relacionadas con el concepto de jerarquía.

**Definición** *Conjunto de elementos*: Un conjunto  $E$  cuyos elementos se definen explícitamente. ♦

*Ejemplo 3.1*: {rojo, azul, verde, naranja, amarillo}.

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

*Ejemplo 3.2*: {muy frío, frío, templado, tibio, caliente, muy caliente}.

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

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

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

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

**Definición** *Partición*:  $P$  es una partición del conjunto  $E$ , si  $P$  y un conjunto exclusivo son cubrimiento de  $E$ .

**Definición** *Variable cualitativa*: Una variable uni-valuada que toma valores simbólicos. ♦

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

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

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

*Ejemplo*, suponga una **Jerarquía H<sub>1</sub>**, Para  $E = \{\text{Canadá, USA, México, Cuba, Puerto\_Rico, Jamaica, Guatemala, Honduras, Costa\_Rica}\}$ , la jerarquía  $H_1$  es  $\{\text{Norteamérica, Islas\_Caribe, América\_Central}\}$ , donde Norteamérica  $\propto \{\text{Canadá, USA, México}\}$ ; Islas\_Caribe  $\propto \{\text{Islas\_Hispano-parlantes, English\_Speaking\_Island}\}$ ; Islas\_Hispano-parlantes  $\propto \{\text{Cuba, Puerto\_Rico}\}$ ; English\_Speaking\_Island  $\propto \{\text{Jamaica}\}$ ; América\_Central  $\propto \{\text{Guatemala, Honduras, Costa\_Rica}\}$ <sup>16</sup>.

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

**Definición Variable Jerárquica:** Una variable jerárquica es una variable cualitativa cuyos valores pertenecen a la jerarquía (El tipo de dato de una variable jerárquica es jerarquía<sup>17</sup>).

♦

*Ejemplo:* lugar\_de\_origen toma valores de  $H_1$ <sup>18</sup>.

**Definición** Se utilizará la siguiente notación: a) **padre\_de** ( $v$ ). En un árbol que representa una jerarquía, el nodo **padre\_de** un nodo es aquel del cual “cuelga”; b) los **hijos\_de** ( $v$ ) son los valores que cuelgan de  $v$ . Los nodos con el mismo padre son **hermanos**; c) **abuelo de**,

<sup>16</sup> El elemento *otro* se añade por omisión, donde se necesita completar la particiones de la jerarquía.

<sup>17</sup> Las variables jerárquicas son uni-valuadas. Así, un valor para lugar\_de\_origen puede ser Norteamérica or México, pero no  $\{\text{Canada, USA, Mexico}\}$ , aunque Norteamérica  $\propto \{\text{Canada, USA, Mexico}\}$ . Cuando un árbol representa una jerarquía, la partición de cada nodo se muestra como descendientes (subconjunto) de cada nodo.

<sup>18</sup> También escribiremos una jerarquía como  $H_1$  por lo que:  $\{\text{Norte\_America} \propto \{\text{Canada USA Mexico}\}$  Islas Caribe  $\propto \{\text{Islas\_Hispano\_parlantes} \propto \{\text{Cuba Puerto\_Rico}\}$  English\_Speaking\_Island  $\propto \{\text{Jamaica}\}$  } América\_Central  $\propto \{\text{Guatemala Honduras Costa\_Rica}\}$  }, algunas veces se omite el símbolo  $\propto$  por simplicidad.

**hermano\_de, tío\_de, ascendentes, descendientes...** son definidos, cuando ellos existen; d)  
La **raíz** es el nodo que no tiene padre. ♦

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

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

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

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

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

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

*Ejemplo* objeto {diminuto, pequeño, mediano, grande}\*<sup>19</sup>.

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

*Ejemplo:* Continente Americano (740M) {Norteamérica (430M) {USA (300M), Canadá (30M), México (100M)} América Central (10M), Sudamérica (300M)}<sup>20</sup>.

**Definición Jerarquía Mixta:** Una jerarquía mixta combina los casos anteriores ♦

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

---

<sup>19</sup> Notación: Un \* es colocado al final de la partición para representar que se trata de una partición ordenada.

<sup>20</sup> El tamaño de cada conjunto escribe entre paréntesis después del conjunto. En el caso del ejemplo se trata del número de habitantes.

### 3.3.2 Teoría de la Confusión

Es un enfoque para modelar la similitud entre objetos de una o de diversas jerarquías, se basa en una medida asimétrica y dependiente del contexto denominada *confusión* (de usar un valor cualitativo en vez del valor esperado o correcto) [Levachkine y Guzmán-Arenas, 2004]. El termino confusión fue introducido para diferenciar este enfoque de otros que se orientan a medir distancias entre conceptos (por ejemplo, simetría, medidas independientes del contexto, cercanía, similitud). La asimetría de la confusión se da por su definición y su dependencia al contexto por la estructura jerárquica. El concepto de confusión permite definir la cercanía a la que un objeto cumple un predicado como también derivar otras operaciones y propiedades entre valores jerárquicos:

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

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

$$\text{conf}(r, r) = \text{conf}(r, s) = 0, \text{ donde } s \text{ es cualquier ascendente de } r; \quad (\text{regla 1})$$

$$\text{conf}(r, s) = 1 + \text{conf}(r, \text{padre\_de}(s)). \quad \blacklozenge \quad (\text{regla 2})$$

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

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

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

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

- $\text{conf}^{\circ}(r, s) = 1 + \text{conf}^{\circ}(r, \text{padre\_de}(s))$ . ♦

Esto es como *conf* para las *jerarquías constituidas por conjuntos*, excepto que allí el error entre dos hermanos es 1, y aquí es un número  $\leq 1$ . Por ejemplo, Temperatura={congelado, frío, normal, tibio, caliente, ardiendo}; de esta lista,  $\text{conf}^{\circ}(\text{congelado}, \text{normal}) = 1/3$ , mientras que  $\text{conf}^{\circ}(\text{congelado}, \text{ardiendo}) = 5/5 = 1$ .

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

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

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

- $\text{conf}^{\circ\circ}(r, r) = \text{conf}^{\circ\circ}(r, s) = 0$ , cuando *s* es cualquier ascendente de *r*;
- $\text{conf}^{\circ\circ}(r, s) = 1 - \text{proporción relativa de } s \text{ en } r$ . ♦<sup>21</sup> (regla 4)

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

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

- Aplicar (regla 1) para la ruta *ascendente* de *r* a *s*;
- En la ruta descendente, usar (regla 3) en vez de (regla 2), si *p* es un conjunto ordenado<sup>22</sup>; o usar regla (regla 4) en vez de (regla 2), cuando los tamaños de *p* y *q* son conocidos. ♦

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

---

<sup>21</sup> El número de elementos de *E* que están tanto en *r* como en *s* / número de elementos de *E* que también están en *r* = población relativa o porcentaje o proporción de *s* en *r*. Notar que esta definición toma en cuenta el contenido de información.

<sup>22</sup> *p* y *q* son dos elementos consecutivos en la ruta de *r* a *s*, donde *q* sigue inmediatamente después de *p*. Esto es,  $r \rightarrow \dots \rightarrow p \rightarrow q \rightarrow \dots \rightarrow s$ .

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

**Definición** Un valor  $u$  es igual al valor  $v$ , dentro de una confusión dada  $\varepsilon$ , escrita  $u =_{\varepsilon} v$ , Ssi<sup>23</sup>  $\text{conf}(u, v) \leq \varepsilon$  (Esto significa que el valor  $u$  puede ser usado en vez de  $v$ , con un error  $\varepsilon$ )<sup>24</sup>. ♦

#### 3.3.2.5.1 Consultas usando la confusión

Objetos que poseen muchas propiedades (o variables), algunos de ellos probablemente como variables jerárquicas, pueden ser almacenados como renglones en una base de datos relacional. Se extiende la noción de consulta en tablas con variables jerárquicas, definiendo el conjunto de objetos que satisfacen un predicado  $P$  dada una confusión  $\varepsilon$ .

**Definición**  $P$  se cumple para el objeto  $o$  con una confusión  $\varepsilon$ , o  $P$  se cumple para  $o$  con  $\varepsilon$ ,

1. cuando  $P$  está formado por variables no-jerárquicas, Ssi  $P$  es verdadero para  $o$ ;
2. cuando  $pr$  (propiedad) es una variable jerárquica y  $P$  es de la forma  $(pr = c)$ , ssi para el valor  $v$  de la propiedad  $pr$  en el objeto  $o$ ,  $v =_{\varepsilon} c$  (si el valor  $v$  del objeto puede ser usado en vez de  $c$  con confusión  $\varepsilon$ );
3. cuando  $P$  es de la forma  $P_1 \vee P_2$ , ssi  $P_1$  se cumple para  $o$  con  $\varepsilon$  o  $P_2$  se cumple para  $o$  con  $\varepsilon$ ;
4. cuando  $P$  es de la forma  $P_1 \wedge P_2$ , ssi  $P_1$  se cumple para  $o$  con  $\varepsilon$  y  $P_2$  se cumple para  $o$  con  $\varepsilon$ ;
5. cuando  $P$  es de la forma  $\neg P_1$ , ssi  $P_1$  no se cumple para  $o$  con  $\varepsilon$ . ♦

#### 3.3.2.5.2 Objetos idénticos, muy similares, algo similares.

Los objetos son entidades descritas por pares  $k$  (propiedades, valores), que en la notación se hará referencia como pares (variable, valor) [Levachkine y Guzmán-Arenas, 2007]. En bases de datos son denominados como pares (relación, atributo). Un objeto  $o$  con  $k$  pares (variable, valor) se describe como  $(o (v_1 a_1) (v_2 a_2) \dots (v_k a_k))$ .

<sup>23</sup> De aquí en adelante Si y sólo si, será representado por ssi

<sup>24</sup> Notar que  $=_{\varepsilon}$  no es ni *simétrico* ni *transitivo*.



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

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

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

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

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

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

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

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

**Definición** En general,  $o'$  es *similar<sub>n</sub>* a  $o$ , si  $\sum_i \text{conf}(a'_i, a_i) = n^{25}$ . ♦

---

<sup>25</sup> Estas relaciones no son simétricas.

En este capítulo se presentaron algunos fundamentos de topología geoespacial, mediciones sobre datos geográficos, jerarquías y teoría de confusión. Todo esto con el fin de dar soporte al desarrollo de la presente tesis. En un sentido más amplio podemos decir que la topología geoespacial y mediciones sobre datos geográficos son cuestiones relacionadas al procesamiento tradicional que se ha dado a los datos geográficos y por otro lado jerarquías y teoría de confusión forman parte del enfoque que tiene este trabajo, orientado al análisis de la semántica de los datos geoespaciales. En donde con base en una conceptualización (una jerarquía) de dominio es posible medir la similitud entre valores cualitativos que describen una situación en particular.

# CAPÍTULO 4

## METODOLOGÍA

En este capítulo se describirá la metodología desarrollada para este trabajo, la cual se enfoca hacia evaluar la similitud semántica entre datos geoespaciales a diferente nivel de detalle con base en una conceptualización del entorno geográfico. En este contexto la similitud semántica está definida con base en la consistencia de los datos. Según la definición de Gruber, una ontología constituye “la especificación formal y explícita de una conceptualización compartida”. De la definición anterior; *conceptualización* se refiere a un modelo abstracto de algún fenómeno del mundo del que se identifican los conceptos que son relevantes; *explícita* hace referencia a la necesidad de especificar de forma consciente los distintos conceptos que conforman una ontología; *formal* indica que la especificación debe representarse por medio de un lenguaje de representación formalizado; *compartida* refleja que una ontología debe, en el mejor de los casos, dar cuenta de conocimiento aceptado (como mínimo, por el grupo de personas que deben usarla).

Actualmente, los modelos de datos geográficos generalmente son representados explícitamente por un conjunto de objetos básicos y sus propiedades (como su geometría y su topología entre otros). En gran parte del mundo geográfico la semántica aparece en las relaciones que poseen grupos de objetos geoespaciales [Worboys, 1996] [Papadias y Theodoris, 1997] [Ruas, 1999] [Mark, 1999]. Sin embargo, las relaciones no representan *fielmente* la semántica de los objetos geográficos, existen otros componentes (como la geometría) que deben ser considerados para lograr capturar de mejor forma la semántica de los datos.

En el contexto de esta investigación, las *descripciones semánticas* están orientadas a encapsular el conjunto de propiedades y relaciones que implícitamente poseen conjuntos de datos geográficos. La organización de los datos geográficos se realiza por medio de capas temáticas, donde cada capa temática posee propiedades particulares y en algunos casos las relaciones con otras temáticas son muy peculiares. Por ello se pretende aprovechar las *descripciones semánticas* para representar dichas peculiaridades de forma muy concreta, con el fin de evitar ambigüedades.

La metodología propuesta está basada fundamentalmente en cuatro etapas funcionales: Normalización, Análisis, Síntesis y Verificación. La Figura 4.1 presenta el esquema general de la metodología a desarrollar. Como se puede ver, la conceptualización juega un rol muy importante, ya que de ella depende en gran parte del desarrollo de la metodología propuesta. Las etapas Normalización, Análisis, Síntesis y Verificación forman parte integral de la metodología, así como la conceptualización. La etapa de Procesamiento, en el caso de esta tesis, es la generalización, pero bien podría ser otra. Cada una de las bases de datos conceptualizadas almacena las representaciones conceptuales de los datos geoespaciales.

El método propuesto consta de cuatro etapas que son: *Normalización, Análisis, Síntesis y Verificación*.

- *Normalización*: Consiste en asegurar la consistencia topológica de la base de datos espacial que contiene los datos fuente y asignar automáticamente identificadores alfanuméricos.
- *Análisis*: En esta etapa se realizan mediciones sobre los datos geográficos, con el fin de extraer las propiedades que éstos poseen implícitamente. Principalmente nos enfocamos a la geometría y las relaciones topológicas entre objetos geográficos, representadas en la conceptualización definida a priori. Básicamente se identifican las relaciones topológicas y posteriormente cada relación es refinada utilizando métricas (*relación topológicamente refinada con métricas*) para extraer sus detalles finos, lo que origina

que se disminuyan las ambigüedades. Para cada relación se obtiene un valor numérico denominado *descriptor cuantitativo*.

- *Síntesis*: Tiene como propósito generar automáticamente una *representación conceptual explícita* de las relaciones denominada *base de datos parcialmente conceptualizada*. En ella las tablas están compuestas por tuplas denominadas *descripciones semánticas* y tienen la forma  $\{O_i, \mathbf{R}, O_j\}$  donde  $O_i$  y  $O_j$  son los identificadores de los objetos geográficos y  $\mathbf{R}$  representa el concepto (relación), el cual se obtiene mapeando el *descriptor cuantitativo* en la conceptualización. Cada descripción representa el significado de la relación topológica.
- *Verificación*: Los pasos anteriores se aplican a un caso de estudio (generalización). Se verifica que los datos generalizados sean semánticamente consistentes. Esto se realiza comparando las *descripciones semánticas* de los *datos fuente* con las *descripciones semánticas* de los *datos generalizados* automáticamente. Particularmente se emplean las *invariantes semánticas*, las cuales son relaciones que no deben cambiar después de realizar un procesamiento sobre los datos y de ellas depende la consistencia de los datos y por ende la semántica. Para comprobar las *invariantes semánticas* se utiliza la conceptualización (estructura jerárquica) y por medio de la Teoría de la Confusión se mide la similitud entre las descripciones semánticas (medición a nivel local). Como resultado de la medición de la similitud se obtienen *distancias semánticas* entre los conceptos (relaciones), la cual nos permite definir nuevos conceptos de similitud (*igual, equivalente, desigual*) que representan la correspondencia de la relación en los datos fuente con respecto a la relación en los datos generalizados. Esta etapa se incluyen también mediciones de similitud a nivel global, que representan la similitud entre los datos fuente y los datos generalizados. Una vez terminada la verificación las descripciones semánticas que se evaluaron como Igual y Equivalente se almacenarán en una denominada *base de datos conceptual semánticamente consistente*.

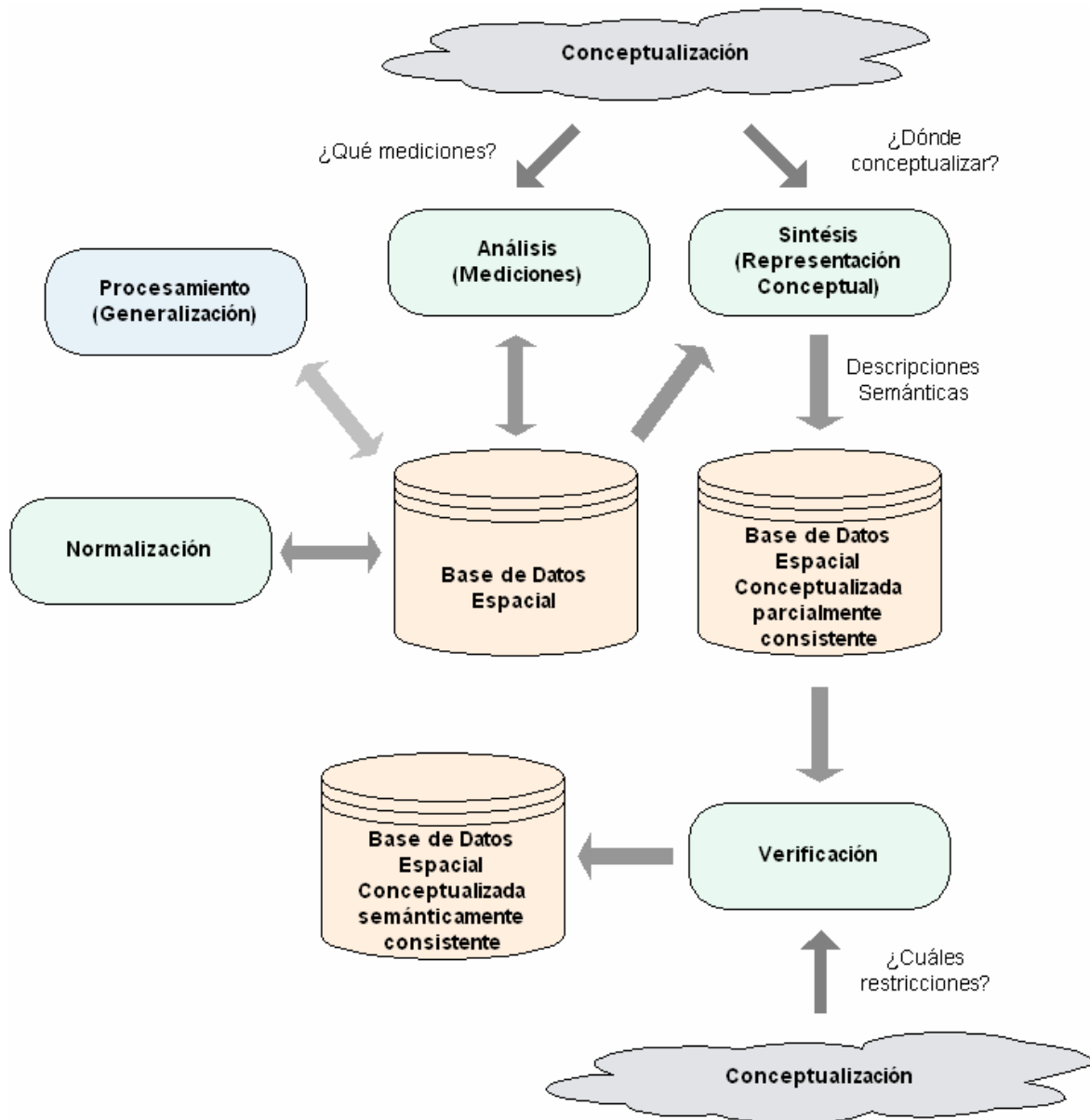


Figura 4.1. Esquema de la metodología a desarrollar en esta tesis

A diferencia de los enfoques tradicionales de procesamiento de datos geográficos, este trabajo incorpora un componente semántico (las relaciones entre conceptos) lo que hace posible verificar si dos representaciones con diferente nivel de detalle son semánticamente equivalentes. La representación conceptual presenta algunas ventajas con respecto a los enfoques tradicionales ya que no depende de la escala, sistema de referencia, dimensión, formato, etc.

## 4.1 La conceptualización del dominio Topográfico por un experto

La conceptualización del entorno geográfico es muy importante para el desarrollo de esta tesis, ya que con base en ella se medirá la similitud semántica. 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 que se deben cubrir. Dicha conceptualización será la base para definir una ontología.

### 4.1.1 Propiedades de los datos geográficos

Las bases de datos espaciales almacenan datos descriptivos y espaciales. Los descriptivos están orientados a las características atributivas de los datos, tales como área de los objetos, identificadores, entre otros. 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 datos *vectoriales*, se ocupan generalmente de los fenómenos discretos. Las características de un fenómeno en el mundo real se representan por un conjunto de primitivas geométricas (puntos, líneas, superficies o sólidos) agrupados en capas temáticas<sup>26</sup>. Otras características se pueden almacenar como atributos. Los datos *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 general, los datos espaciales poseen muy diversas propiedades que cubren diversos aspectos. En el contexto de este trabajo, nos enfocaremos a describir algunas de estas propiedades, entre las que se encuentran las geométricas, topológicas, temáticas, cartométricas y lógicas

#### 4.1.1.1 Propiedades geométricas

La geometría se encarga de analizar las relaciones, propiedades y mediciones de sólidos, superficies, líneas y ángulos. En el contexto de este trabajo la geometría proporciona las métricas para describir cuantitativamente las características geográficas. En México, el

---

<sup>26</sup> Para el caso de este documento se usarán indistintamente los términos capa temática, temática y cobertura.

organismo encargado de definir las reglas para la elaboración de datos geográficos es el *Instituto Nacional de Geografía, Estadística e Informática (INEGI)*. En lo que respecta al INEGI, describe el tipo de primitiva de representación que debe emplearse para cada temática y las dimensiones mínimas de los objetos que pertenecen a una temática en ciertas escalas. Esta información viene representada para mapas topográficos en los diccionarios a diferentes escalas. [INEGI, 1995] [INEGI, 1996]. Adicionalmente, se describe si la forma del objeto se representa *virtual*<sup>27</sup>, *aproximado*<sup>28</sup> o *definido*<sup>29</sup> [INEGI, 1993]. Esto es muy útil para identificar la precisión que deben tener cada una de las temáticas.

En la literatura existen muy diversos artículos relacionados con la identificación de las propiedades geométricas; por esta razón se ha considerado utilizar algunas propiedades como las propuestas por [McMaster y Shea, 1992] que evalúa las propiedades geométricas midiendo densidad, distribución, longitud, forma y sinuosidad de los objetos geográficos. Adicionalmente, puede considerar emplear el análisis de medidas del proyecto AGENT [AGENT, 1999], en el cual se describen y analizan diversas propiedades geométricas de los datos espaciales.

#### 4.1.1.2 Propiedades topológicas

La topología 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].

<sup>27</sup> Virtual significa que el fenómeno geográfico no puede percibirse sobre la superficie, por ejemplo: un canal subterráneo.

<sup>28</sup> Aproximado significa que su posición es calculada con base en otros procedimientos, por ejemplo: una curva de nivel.

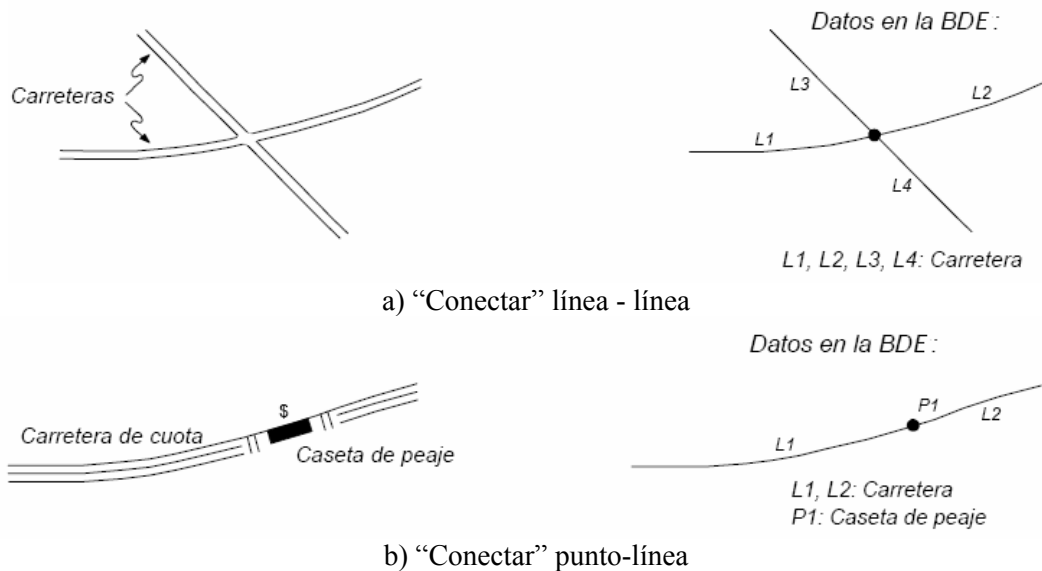
<sup>29</sup> Definido significa que su posición puede determinarse con precisión; por ejemplo, una carretera.



La relación *comparte* de acuerdo a INEGI exige lo siguiente:

- Que exista una intersección planimétrica entre los rasgos geográficos involucrados.
- Que la *base de datos espacial* (BDE) sustente esta relación.

Además, se define que esta relación requiere que las ocurrencias de las entidades involucradas en la relación, tengan las mismas coordenadas de sus ocurrencias de representación geométrica en el punto de conexión. (Ver Figura 4.2).

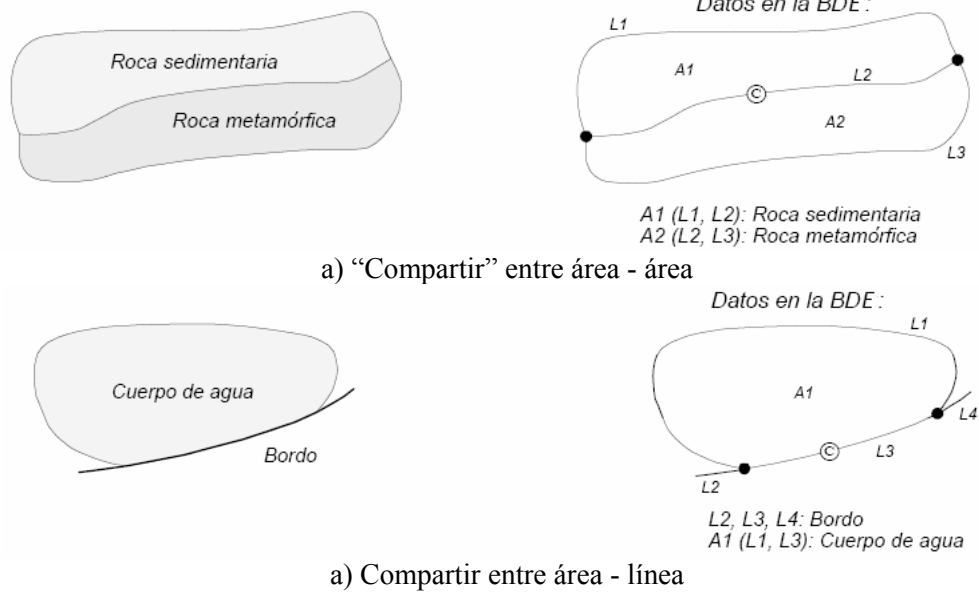


**Figura 4.2** Relación Conecta

INEGI define que existe una relación "comparte" cuando se satisfacen las siguientes condiciones:

- Que los rasgos geográficos percibidos como líneas o áreas sean parcial o totalmente contiguos o coincidentes.
- Que la base de datos espacial sustente esta relación.

Asimismo, esta relación requiere que las ocurrencias de las entidades involucradas en la relación *compartir* tengan las mismas coordenadas en sus ocurrencias de representación geométrica lineal (Ver Figura 4.3 ).



**Figura 4.3** Relación Comparte

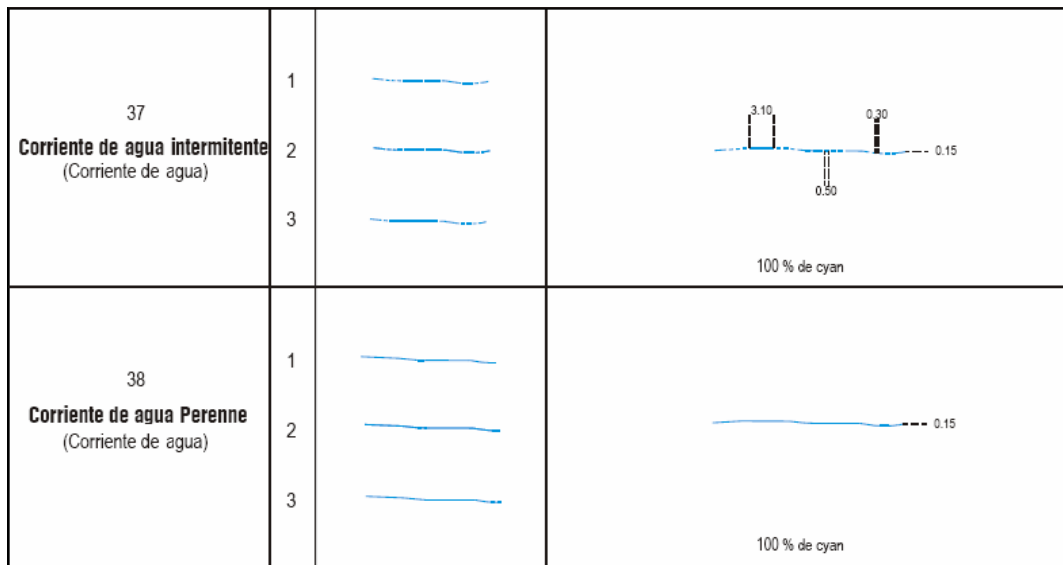
Dado que estas relaciones no captan en su totalidad la riqueza de la representación, es necesario enriquecer el conjunto de relaciones con el propósito de describir con mayor precisión el entorno geográfico.

#### 4.1.1.3 Propiedades temáticas

Las propiedades temáticas representan las características de la representación cartográfica de los datos geográficos. En este grupo de propiedades se ha considerado incluir aquellos atributos que den una idea de la jerarquía de los objetos. Por ejemplo, el orden<sup>30</sup> en el caso de los ríos o el nivel administrativo<sup>31</sup> en el caso de los poblados. Adicionalmente, es necesario incluir aquellas propiedades relacionadas con la representación cartográfica para representar adecuadamente los datos, por ejemplo, los ríos de color azul, con cierto tipo de línea y con determinado ancho. La Figura 4.4, muestra los diversos tipos de línea que son empleados para representar diversos componentes de la red hidrológica (corriente de agua intermitente y corriente de agua perenne); cada una de ellas utiliza una representación particular; para el caso de México el INEGI establece los criterios de representación de cada objeto geográfico.

<sup>30</sup> Orden es un método para clasificar jerárquicamente los elementos que componen una red hidrológica. También se conoce como ordenamiento de los cauces de una cuenca; existen diversos métodos, entre los que podemos mencionar: Horton (1945), Strahler (1957), Moreno (2001).

<sup>31</sup> Se refiere a la importancia administrativa que tiene cada población, por ejemplo, para el caso de México, existen capitales estatales y cabeceras municipales, que son representadas de diferente forma.



**Figura 4.4** Características de representación del INEGI para las corrientes de agua en escala 1:50,000

Estas características de representación cartográfica ayudan a percibir en gran parte la semántica de ciertos objetos, ya que en representaciones visuales dicha simbología permite identificar los objetos y saber ciertas peculiaridades inherentes a los datos.

De acuerdo al *INEGI*, dichas propiedades están especificadas en los Diccionarios de Datos [INEGI, 1995], donde se especifican los atributos que debe tener cada una de las temáticas que son descritas para los mapas topográficos. Hay ciertas temáticas en donde existen atributos que representan la jerarquía de los objetos, por ejemplo, para el caso de las *curvas de nivel* se clasifican en curvas maestras, ordinarias, auxiliares y aproximadas. Para los propósitos de este trabajo, se requieren propiedades que permitan representar explícitamente atributos en forma jerárquica que no se consideran en las normas del INEGI (por ejemplo, el orden de un río).

#### 4.1.2 Diseño de la ontología

Para llevar a cabo el diseño de la ontología nos basaremos en la descripción del dominio topográfico<sup>32</sup> por un usuario experto. Se considerarán algunos aspectos descritos en [Noy y

<sup>32</sup> Se denomina como dominio topográfico porque incluye a las capas temáticas involucradas: ríos y curvas de nivel.

McGuinness, 2001], el cual se basa en los editores de ontologías Ontolingua<sup>33</sup>, Chimaera<sup>34</sup> y Protégé<sup>35</sup>. Este último es en el que más se enfoca. Si bien [Noy y McGuinness, 2001], no describen en detalle todos los aspectos de diseño de las ontologías, proporciona una guía muy útil para desarrollarlas.

#### 4.1.2.1 Objetivo de la ontología

Para comenzar a definir el desarrollo de la ontología se requiere definir el dominio<sup>36</sup>, es decir, se debe responder primero a varias preguntas básicas:

- ¿Cuál es el dominio que cubrirá la ontología?
- ¿Para qué usaremos la ontología?
- ¿Para qué tipos de preguntas la información en la ontología deberá proveer respuestas?
- ¿Quién usará y mantendrá la ontología?

En el caso de este trabajo se definirá como dominio la topografía y particularmente nos enfocaremos al análisis de las relaciones entre los ríos y las curvas de nivel. Adicionalmente a las *relaciones jerárquicas* (de un nivel general a uno particular) propias de la ontología, nos interesa representar la relación entre los ríos y las curvas de nivel (objetos con la misma generalidad). Adicionalmente, se utilizarán conceptos que sean subclases de dicha relación para describirla de mejor forma. El concepto que define la relación depende tanto de la topología como de la geometría y se requiere identificar igualdad, equivalencia y desigualdad, así como también una medida global de similitud. La ontología diseñada se utilizará con fines de investigación, por lo que su uso y mantenimiento dependerá de grupos de investigación.

Básicamente, la ontología debe proporcionar respuesta a las siguientes preguntas:

- ¿Por qué parte de la *curva de nivel x* pasa el *río y*?
- ¿Se relaciona el *río x* con la *curva de nivel y*?
- ¿Con cuáles curvas de nivel se relaciona el *río x*?

<sup>33</sup> <http://www.ksl.stanford.edu/software/ontolingua/>

<sup>34</sup> <http://www-ksl.stanford.edu/software/chimaera/>

<sup>35</sup> <http://protege.stanford.edu/>

<sup>36</sup> Desde este punto el término conceptualización y ontología hacen referencia a la misma estructura, donde la conceptualización puede ser expresada en lenguaje natural y la ontología se define en términos formales.

- ¿Qué características tiene el *río x*?

Con el propósito de construir la ontología se requiere analizar otras conceptualizaciones existentes, ya que en sí mismas son acuerdos de un grupo de expertos que comparte su definición; a continuación se describirán las que se consultaron:

- La Especificación para Codificación de Datos Ambientales (EDCS) [ISO, 2005] es un estándar que proporciona mecanismos para desambiguar conceptos específicos utilizados para modelar conceptos ambientales. Se presentan descripciones de los conceptos y aspectos tales como clasificaciones, atributos, unidades, escala de las unidades y esquemas organizacionales. Se agrupan los fenómenos ambientales en categorías entre las que se incluyen conceptos abstractos, animales, hidrología, fisiografía, vegetación, localización y condiciones oceanográficas. Esta especificación resulta muy útil debido a que se pueden identificar y estructurar jerárquicamente los conceptos que se utilizarán.
- Diccionarios de Datos del INEGI a escalas 1:50,000 [INEGI, 1996] y 1:250,000, [INEGI, 1995] y la referencia del modelo vectorial [INEGI, 1993]. En dichos documentos se especifica la descripción de conceptos, restricciones y se definen algunas relaciones entre los datos. La utilización de estos documentos proporciona la base para identificar qué objetos existen en las Bases de Datos Espaciales generadas en México, así como también considera las relaciones definidas en los conceptos. Desafortunadamente, en estos documentos solamente se proporciona un mecanismo para organizar jerárquicamente los datos (por primitiva de representación, punto, línea, área).
- Word Net [WordNet, 2007] es una base de datos léxica del inglés, donde sustantivos, verbos, adjetivos y adverbios son agrupados en conjuntos de sinónimos cognitivos (*synsets*). Los *synsets* son interligados por medio de relaciones semánticas y léxicas. Esta base de datos se usa para identificar cómo son las relaciones entre los conceptos generales con los conceptos específicos (organización jerárquica).
- Wikipedia es un proyecto de enciclopedia basado en Web de contenido multilingüe [Wikipedia, 2007], que se utiliza para verificar las definiciones de los conceptos, ya que

se presentan definiciones extensas que permite identificar sinónimos y palabras que describan a los conceptos en otros idiomas.

Para construir la ontología vale la pena considerar lo que otra persona (organización) ha hecho y verificar si podemos refinar y extender recursos existentes que sean acordes con nuestro dominio y tarea particular. El uso de ontologías existentes puede ser un requerimiento, ya que pretendemos que nuestra metodología permita la interacción con otras aplicaciones que se basan en ontologías particulares.

Adicionalmente a esto es necesario realizar un análisis de los componentes que se requieren representar en la ontología, así como también identificar sus propiedades y relaciones.

#### **4.1.2.2 Análisis de las componentes del dominio a conceptualizar**

En esta sección se describirán brevemente los componentes del dominio de la conceptualización (ríos y curvas de nivel) que servirá como base para identificar los conceptos que se emplearán. Se presentará un listado de las características que poseen con el fin de identificar sus propiedades, relaciones y restricciones.

##### **4.1.2.2.1 Ríos y redes hidrológicas**

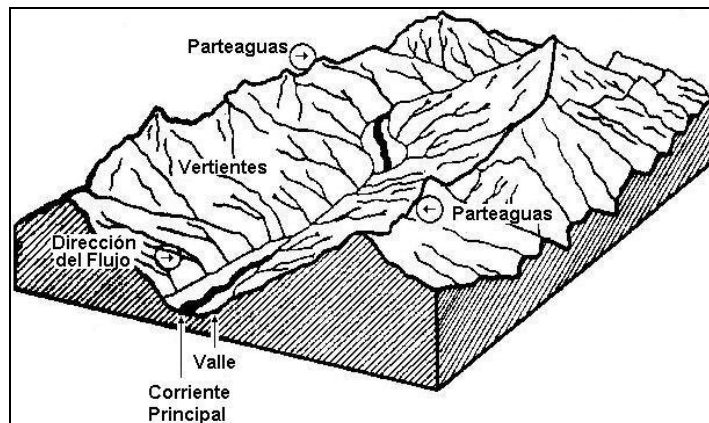
Los ríos y corrientes de agua forman parte de las cuencas hidrológicas, las cuales son áreas surcadas por un sistema de corrientes formadas por escurrimientos producto de las precipitaciones que fluyen hacia un cauce común, según las variaciones topográficas del terreno. La red hidrológica o sistema de drenaje es el patrón formado por corrientes, ríos y lagos en una cuenca en particular.

La cuenca está delimitada por los puntos de mayor elevación latitudinal que constituyen fronteras entre cuencas y subcuencas contiguas. A la unión de dichos puntos se les conoce como parteaguas la que reúne en un punto de salida el drenaje de las aguas que pueden formar grandes ríos, arroyos o corrientes (Ver Figura 4.5)

Las partes que constituyen una cuenca son:

- Parteaguas,

- Vertientes,
- Valle o cuenca baja y
- Red hidrológica, de avenamiento o drenaje



**Figura 4.5** Esquema simplificado de una cuenca

El parteaguas es la línea altimétrica de mayor elevación que delimita orográficamente a cuencas vecinas. Este lindero real marca las variaciones de la conducción del drenaje superficial que por efectos de la pendiente confluyen hacia la parte baja en ambas vertientes. Con fines prácticos, el parteaguas se cierra artificialmente en el estrangulamiento natural que se forma donde se delimita la montaña del valle y a este lugar generalmente se le llama boquilla, porque aquí es donde se marca el volumen de los escurrimientos.

Las vertientes son las áreas de captación y se constituyen como las zonas más estratégicas de la cuenca, dado que en ellas la susceptibilidad del fenómeno es altamente significativa. Las vertientes tienen diversas exposiciones topográficas, las que se pueden presentar como laderas, declives, depresiones, planicies, cauces, barrancos, taludes y cantiles.

El valle o cuenca baja, es la zona de menor altitud y donde generalmente se encuentran los cultivos agrícolas y los asentamientos humanos, aquí la conjunción de las corrientes tributarias han formado un río.

La red hidrológica es la disposición de los cauces y lechos por donde de manera superficial corre el agua excedente producto de la precipitación hacia un depósito natural o artificial.

La red consta de una corriente principal y un sistema de corrientes tributarias de menor importancia.

La configuración de las redes es el producto de las influencias que tienen sobre ellas los suelos, las rocas, el grado de fracturación y la topografía. Los estudios realizados con los sistemas de drenaje han permitido diferenciar cierto número de estos, basados exclusivamente en su forma, lo cual ha dado lugar a una clasificación de los patrones de drenaje y sus modificaciones.

El significado de cada patrón de drenaje tiene relación con la geología, aunque es imposible señalar en una forma precisa la correlación geológica para cada sistema de drenaje. Las seis formas más comunes son: Dendrítico, Rastrillo, Radial, Paralelo, Anular y Rectangular (Ver Figura 4.6).

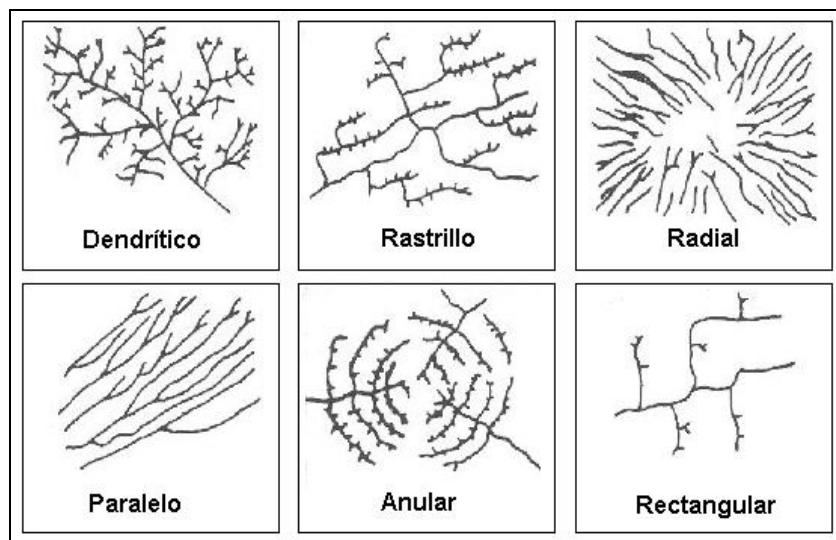


Figura 4.6 Tipos de drenaje

#### 4.1.2.2.1.1 Clases de ríos

Todos los ríos se clasifican en tres clases dependiendo de las características físicas y condiciones climáticas de la cuenca; así las corrientes pueden ser *efímeras*, *intermitentes* y *perennes*.

Las primeras son aquellas que sólo conducen agua cuando llueve e inmediatamente después, es decir, sólo capta *escurrimientos superficiales*. Las intermitentes son aquellas



que conducen por su cauce agua la mayor parte del tiempo, principalmente en *época de lluvias*; su aportación cesa cuando el *nivel freático*<sup>37</sup> desciende por debajo del fondo del cauce. Y por último, las perennes son aquellas que contienen agua todo el tiempo, ya que en la época de estiaje son abastecidos por las *aguas freáticas*, debido a que el nivel de éstas permanece por encima del fondo del cauce.

#### 4.1.2.2.1.2 Características de las redes hidrológicas

La red hidrológica se compone de una corriente o cauce principal y una serie de afluentes y para ordenar las corrientes se utilizan clasificaciones. Dichas clasificaciones consideran a los tributarios, para indicar la corriente principal que indica la extensión de la red de las corrientes en una cuenca. La clasificación debe efectuarse considerando tanto corrientes perennes como intermitentes.

La topografía<sup>38</sup> de una cuenca define en sí la longitud de los tributarios, ya que estos son indicadores de la pendiente. Generalmente, las áreas escarpadas y con buen drenaje, tienen numerosas corrientes pequeñas y efímeras, contrariamente a las áreas planas con suelos profundos y permeables que tienen corrientes largas y generalmente perennes.

Considerando lo anterior el drenaje presenta las siguientes características específicas:

- El drenaje siempre sigue una sola dirección.
- Un tributario se puede conectar a otros tributarios.
- Los tributarios se conectan también con otros objetos como cuerpos de agua.
- Todos los tributarios y escurrimientos tienen generalmente una salida.
- Cualquier sistema de drenaje está delimitado por un parteaguas.
- Los parteaguas que delimitan a un conjunto de tributarios, que tienen conexión entre sí, forman un área cerrada.
- Los sistemas que forman las redes hidrológicas generalmente tienen una salida.
- Las redes hidrológicas presentan configuraciones específicas, originadas por las características del suelo.

---

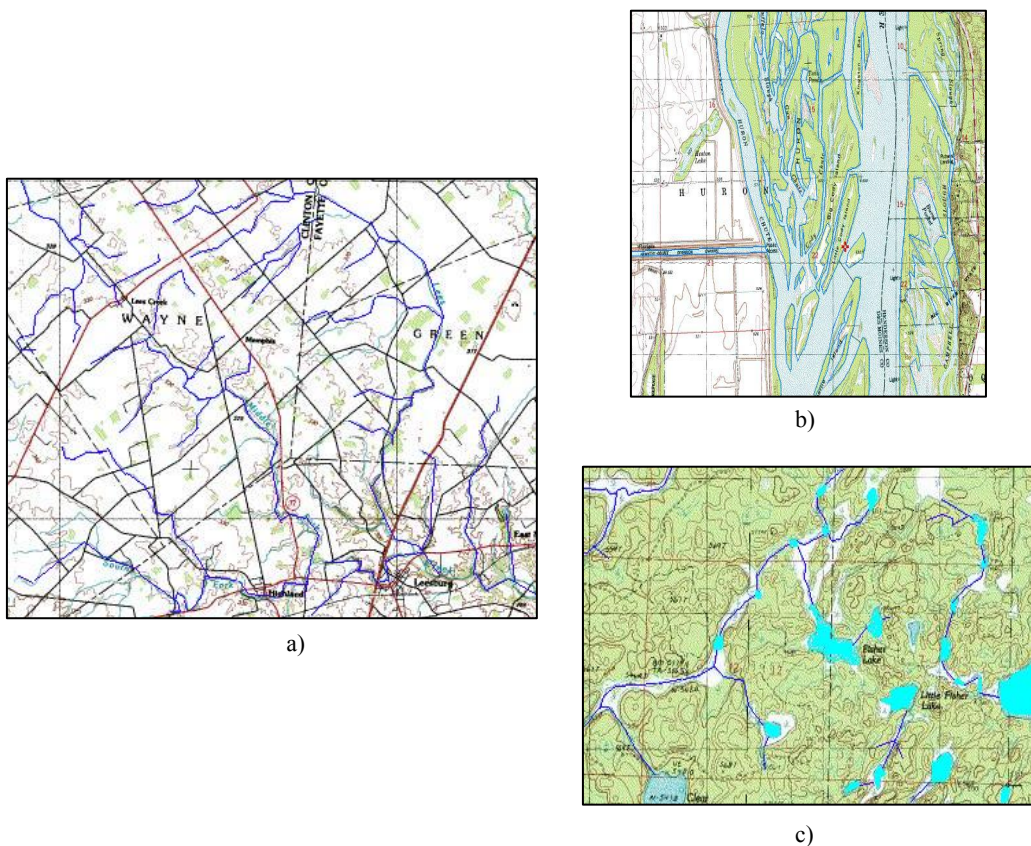
<sup>37</sup> Altura que alcanza la capa acuífera subterránea más superficial.

<sup>38</sup> La topografía es la ciencia que estudia el conjunto de procedimientos para determinar la posición de un punto sobre la superficie de la Tierra, tanto en planimetría como en altimetría.

- La altura donde comienza cualquier tributario es mayor que la altura donde se encuentra la salida de la cuenca.
- El drenaje depende de la topografía.

#### 4.1.2.2.1.3 Representación de las redes hidrológicas

Debido a las reglas de representación de los datos geográficos, las redes hidrológicas pueden representarse usando las primitivas geométricas (punto, línea y área), según corresponda. La representación de la red depende de la zona de estudio y de la escala a la que se encuentren los datos geográficos. El INEGI presenta las características que deben cumplir las representaciones a escala 1:50,000 y 1:250,000 [INEGI, 1996] [INEGI, 1995]. Partiendo de esa base proponemos tres configuraciones: línea-línea, área-área, o mixta. Línea-línea, se refiere a configuraciones compuestas únicamente por líneas; área-área, configuraciones que se componen de áreas; mixtas, configuraciones compuestas tanto por líneas como por áreas (Ver Figura 4.7).

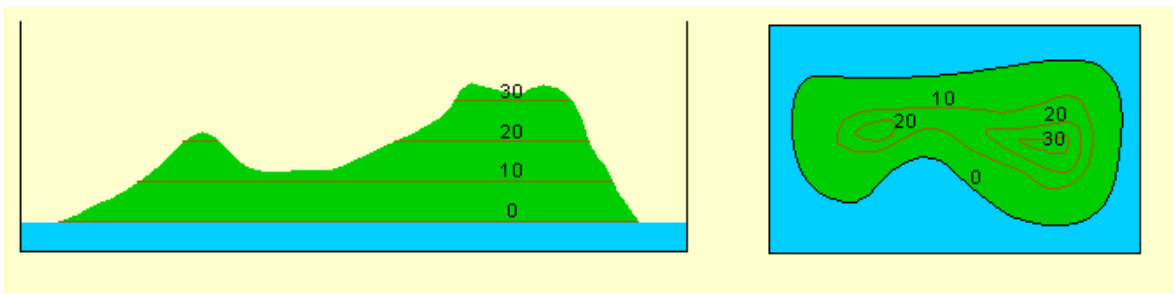


**Figura 4.7** Configuraciones de representación para redes hidrológicas. a) línea-línea, b) área-área, c) mixta

#### 4.1.2.2.2 Curvas de nivel

Las curvas de nivel son líneas imaginarias que unen puntos del terreno de igual cota o elevación. Por medio de ellas se representa planimétrica y altimétricamente el terreno, construyéndolas a una distancia vertical constante, denominada equidistante. Por facilidad, las alturas de las curvas de nivel, lo mismo que su equidistancia, serán siempre cantidades enteras, dependiendo del objeto, escala y de las características del relieve en estudio; las curvas se pueden espaciar cada metro, 2, 5, 10, 20, 50, 100m, etc.

Para una mejor comprensión de los conceptos anteriores, supondremos que el terreno se rebana en cada una de las curvas de nivel propuestas, originando así una superficie de nivel. Si proyectamos en una planta y sobre un mismo plano cada una de las superficies, tendremos un plano que nos representará la configuración del terreno (Ver Figura 4.8).

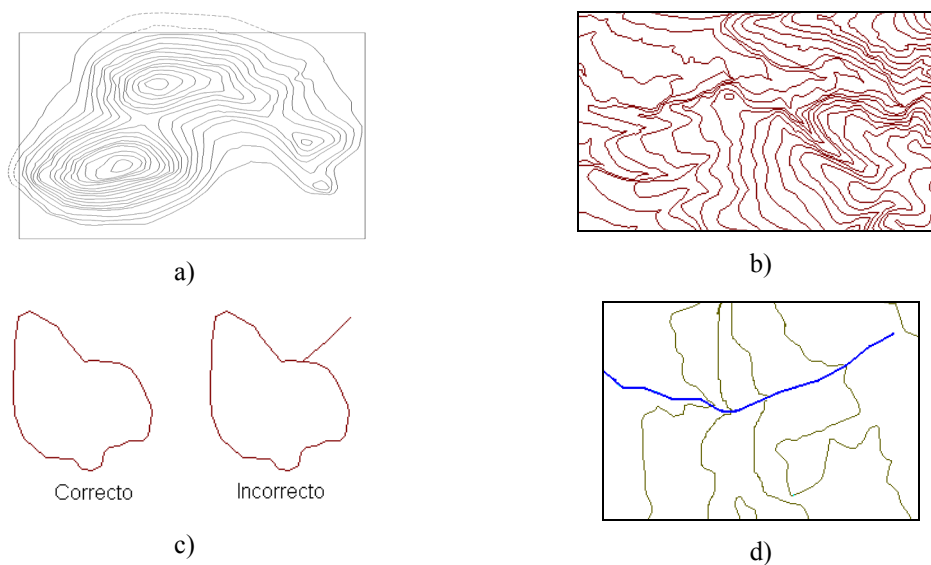


**Figura 4.8** Curvas de nivel

Podemos decir que las curvas de nivel guardan características comunes, de entre las que podemos enumerar:

- Por construcción y definición, todos los puntos de una misma curva tienen la misma elevación.
- Todas las curvas de nivel son cerradas, aunque en los límites planos no se puedan observar (Ver Figura 4.9a).
- Cuando existe una pendiente uniforme en el terreno, la distancia horizontal entre las curvas es la misma. Cuando la pendiente aumenta, crece la proximidad entre las curvas, y si comienzan a separarse será porque la pendiente comenzó a suavizarse.

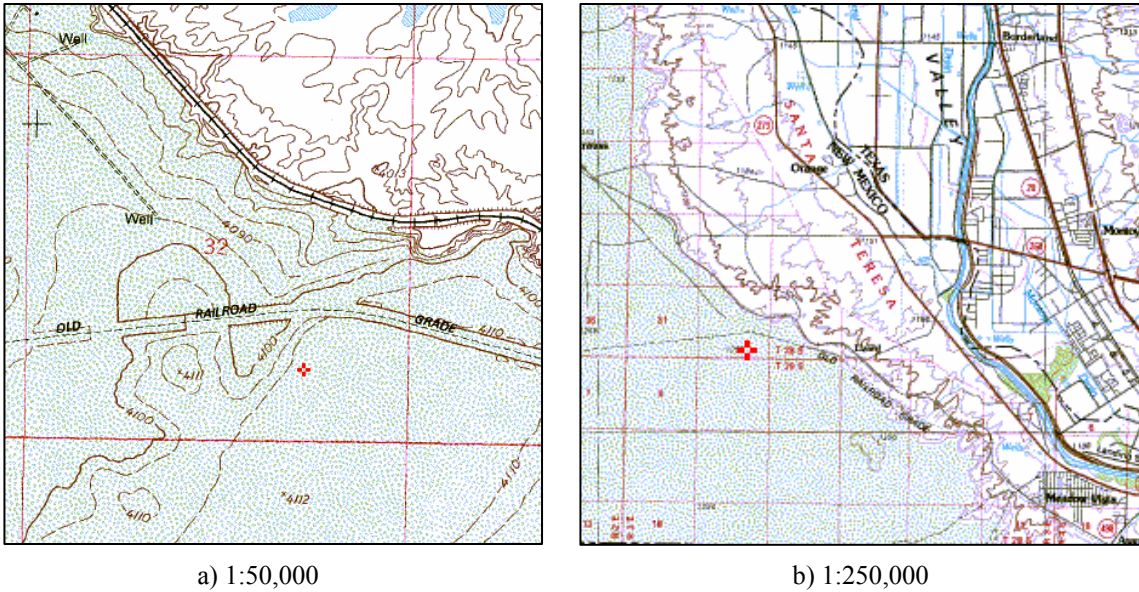
- Cuando dos o más curvas se juntan, es que en realidad están superpuestas, pues cada una está a diferente elevación (Ver Figura 4.9b).
- Si una serie de curvas cerradas son más o menos concéntricas podrían indicar una elevación o una depresión.
- Una curva de nivel nunca se ramifica o bifurca (Ver Figura 4.9c).
- Un parteaguas hará que las curvas descendan, las corten perpendicularmente y suban por el otro lado, es decir, presentan concavidad en la parte más alta.
- La existencia de una corriente o flujo hace que las curvas de nivel corran hacia arriba, crucen el fondo del cauce perpendicularmente a él y descendan por el otro lado, es decir, presentan una convexidad contraria al sentido de la corriente (Ver Figura 4.9d).



**Figura 4.9** Propiedades de las curvas de nivel; a) Todas las curvas de nivel forman áreas cerradas; b) Superposición de las curvas de nivel; c). Las curvas de nivel no se bifurcan; d) Representación de una corriente de agua relacionada con las curvas de nivel

En [INEGI, 1996] e [INEGI, 1995] se describen las características que deben tener las curvas a escala 1:50,000 y 1:250,000. Por ejemplo, a escala 1:50,000 se encuentran con una separación de 10 m entre cada una de ellas. Existe un grupo dentro de las curvas que son denominadas maestras, las cuales se encuentran en los múltiplos de 50; dicho de otra forma, las curvas maestras tienen valores de 50, 100, 150, 200, etc. A diferencia de las curvas de nivel a escala 1:50,000, las curvas a 1:250,000 tienen sus curvas maestras en múltiplos de 500 m.

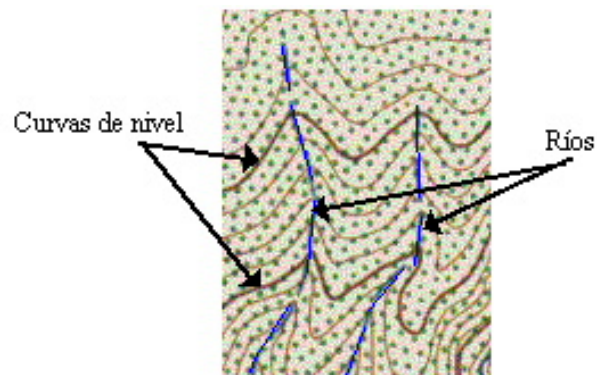
La Figura 4.10 muestra las diferencias que existen entre las representaciones a diferentes escalas, donde se pueden ver las mismas en a) el número de intervalos entre curvas (10 m), b) el nivel de detalle de las curvas de nivel.



**Figura 4.10** Comparación entre curvas de nivel a diferentes escalas

#### 4.1.2.2.3 Relación entre las curvas de nivel y los ríos

Como se ha mencionado anteriormente, las redes hidrológicas y las curvas de nivel tienen una relación muy estrecha de acuerdo a la propiedad enunciada anteriormente: *la existencia de una corriente o flujo, hace que las líneas corran hacia arriba, crucen el cauce perpendicularmente a él y descieran por el otro lado, es decir, presentan una convexidad contraria al sentido de la corriente.* Es decir, el flujo de la red hidrológica sigue el relieve del terreno y pasa perpendicularmente a las curvas (Ver Figura 4.11).



**Figura 4.11** Relación entre los ríos y las curvas de nivel

Desde el punto de vista cartográfico, el río debe pasar por la máxima convexidad<sup>39</sup> de la curva de nivel (Ver Figura 4.12a). No obstante, este no es el único caso *consistente*, pues existen situaciones donde pasa por algún punto de la convexidad que también son consistentes (Ver Figura 4.12b, c y d).

Adicionalmente existen casos donde el río no se enfrenta a una convexidad, por lo que el río *pasa por* una concavidad y esta situación representa que los datos son inconsistentes (Ver Figura 4.12f, g y h). Cuando la representación presenta una curva que no posee una concavidad ni una convexidad no se puede identificar si es consistente a menos que se evalúe la dirección del flujo (Ver Figura 4.12e y j).

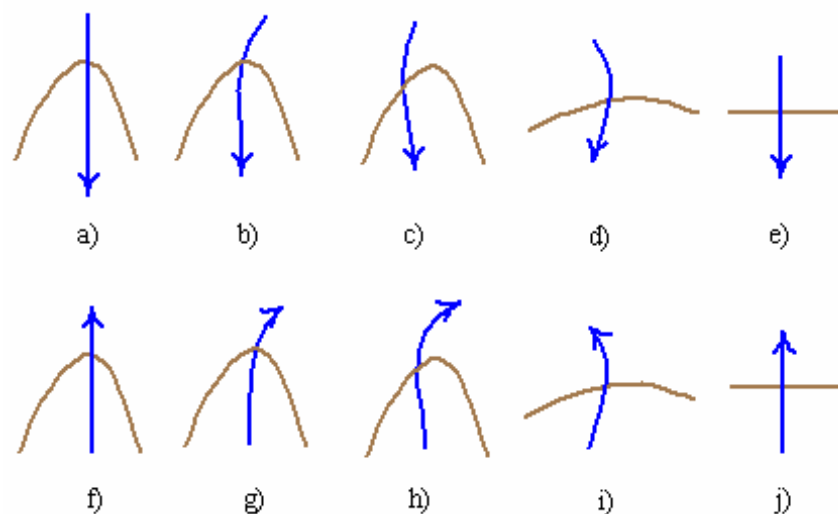


Figura 4.12 Casos de relaciones entre las curvas de nivel y los ríos

#### 4.1.2.3 Términos de la ontología

Una vez que se ha descrito el dominio, se deben definir de forma general los términos que componen la ontología. Esto se debe a que en la descripción del dominio aparecen implícitamente los términos necesarios para definir la ontología. Dichos términos deben encontrarse en los objetivos de la ontología (Ver Sección 4.1.2.1). Para definir los términos que conformarán la ontología se puede usar un proceso de desarrollo de *arriba-hacia-abajo* (*top-down*)<sup>40</sup>, de *abajo-hacia-arriba* (*bottom-up*)<sup>41</sup> o bien un desarrollo *combinado*

<sup>39</sup> En este caso la convexidad se define considerando la dirección del flujo.

<sup>40</sup> Comienza con la definición de los conceptos más generales en el dominio y la subsecuente especialización de los mismos.

utilizando tanto *top-down* como *button-up*, en donde primeramente se definen los conceptos más sobresalientes y luego los generalizamos y especializamos apropiadamente. La ontología se construirá a partir de los conceptos que queremos relacionar (río y curva de nivel) y el concepto raíz (entorno geográfico), por lo que utilizamos un desarrollo *combinado*.

Hay que recordar que la estructura que se defina depende de los posibles usos de la ontología, por lo que se construirá con base en el nivel de detalle necesario para la aplicación, preferencias personales, y algunas veces requerimientos de compatibilidad con otros modelos.

Para los propósitos de este trabajo la ontología estará compuesta por:

- *Clases* (conceptos) que se representarán de la forma <CLASE>
- *Propiedades* (slots) que se representan por (*propiedad*)
- *Relaciones* que se representan por >Relación<
- *Restricciones para propiedades* representadas por -*restricción*-
- *Restricciones para los conceptos* representadas por /*restricción*/

La Tabla 4.1 muestra las clases que componen la ontología. Se consideraron esas clases ya que no es necesario que la ontología contenga la información necesaria sobre el dominio, enfocándose a una aplicación en particular.

De la Tabla 4.1 se pueden ver clases definidas en negritas, las cuales significan que pueden tener instancias (por ejemplo, para la clase <RÍO> puede existir una *instancia* que represente al río Bravo). El resto de las clases se denominan como *abstractas*, ya que son aquellas que no tienen instancias directas. Las clases <CANAL>, <CUERPO-DE-AGUA>, <PUNTO-DE-ELEVACIÓN>, <DATO-ATMOSFÉRICO> se agregaron con el fin de dar consistencia a la ontología.

---

<sup>41</sup> Comienza con la definición de los conceptos más específicos en la jerarquía, con el subsecuente agrupamiento de esas clases en conceptos más generales.

**Tabla 4.1.** Conceptos que conforman la ontología

Clases	Descripción	Sinónimos
<RIO>	<CURSO-DE-AGUA> natural	corriente de agua, <i>river</i>
<CURSO-DE-AGUA>	Canal artificial o natural para el transporte de agua	flujo de agua, <i>watercourse</i>
<CANAL>	<CURSO-DE-AGUA> hecho por el hombre usado para transportación, irrigación o movimiento de agua	Canal
<CUERPO-DE-AGUA>	Extensión de agua limitada por tierra	cuerpo de agua, laguna, lago, cuerpo de agua, <i>waterbody</i>
<HIDROLOGÍA>	Se refiere al estudio del movimiento, distribución espacial y temporal en la tierra	hidrología, <i>hydrology</i>
<CURVA-DE -NIVEL>	Isolínea con <ELEVACIÓN> constante	curva de nivel, <i>elevation contour line.</i>
<ELEVACION-DEL-TERRENO>	La distancia sobre un punto de referencia (msnm <sup>42</sup> ) en la superficie de la tierra.	elevación del terreno, elevación, altitud, <i>elevation</i>
<PUNTO-DE-ELEVACIÓN>	entidad puntual con <ELEVACIÓN>	punto de elevación, punto acotado, <i>spot elevation</i>
<UBICACIÓN>	Datos que se encuentran relacionados con la especificación de la posición de un punto	ubicación, <i>location</i>
<DATO-ATMOSFÉRICO>	Localizado por encima de la superficie terrestre (en la atmósfera)	dato atmosférico
<DATO-AMBIENTAL>	Dato que representa a un fenómeno ambiental, que puede ser de índole geográfico	dato ambiental <sup>43</sup> , dato geográfico, <i>environmental data</i>

Adicionalmente en la Tabla 4.1 se presentan los sinónimos con el propósito de evitar ambigüedades en el entendimiento de la ontología. Los sinónimos en ontologías tienen el mismo sentido que en gramática, se refieren a palabras que en determinados contextos tienen un significado similar o idéntico.

<sup>42</sup> msnm, metros sobre el nivel del mar

<sup>43</sup> Se consideró dato ambiental, debido a que como base para definir los términos de la ontología se utilizó [ISO, 2005]



#### 4.1.2.4 Las relaciones entre las clases

Con base en la Tabla 4.1 se definen las relaciones *tipo-de* (*kind-of*) y *es-un* (*is-a*). La relación *es-un* representa un *hipónimo*<sup>44</sup>. Se dice que X es un *hipónimo* de Y si existe una relación *es-un* entre X y Y. Esto es, si X es subtipo de Y, o expresado de otra forma, si X es una clase de o género de Y. Por ejemplo, Río es un subtipo de Curso de agua, por lo tanto Río es *hipónimo* de curso de agua, ya que comparten propiedades como: flujo y dirección, pero Río es un curso de agua natural). En el sentido inverso de la relación Y es hiperónimo de X (Curso de agua es *hiperónimo* de Río).

Por otro lado, la relación *tipo-de* se dice que es una *meronimia*<sup>45</sup>. Se dice que X es *merónimo* de Y si X forma parte de Y. Por ejemplo Curso de agua forma parte de Hidrología, por lo que Curso de agua es *merónimo* de Hidrología y en sentido inverso Hidrología es *holónimo* de Curso de agua. Hasta este momento con base en las clases definidas se puede formar una taxonomía de clases como se ve en la Figura 4.13, en donde se presentan las relaciones entre los conceptos definidos.

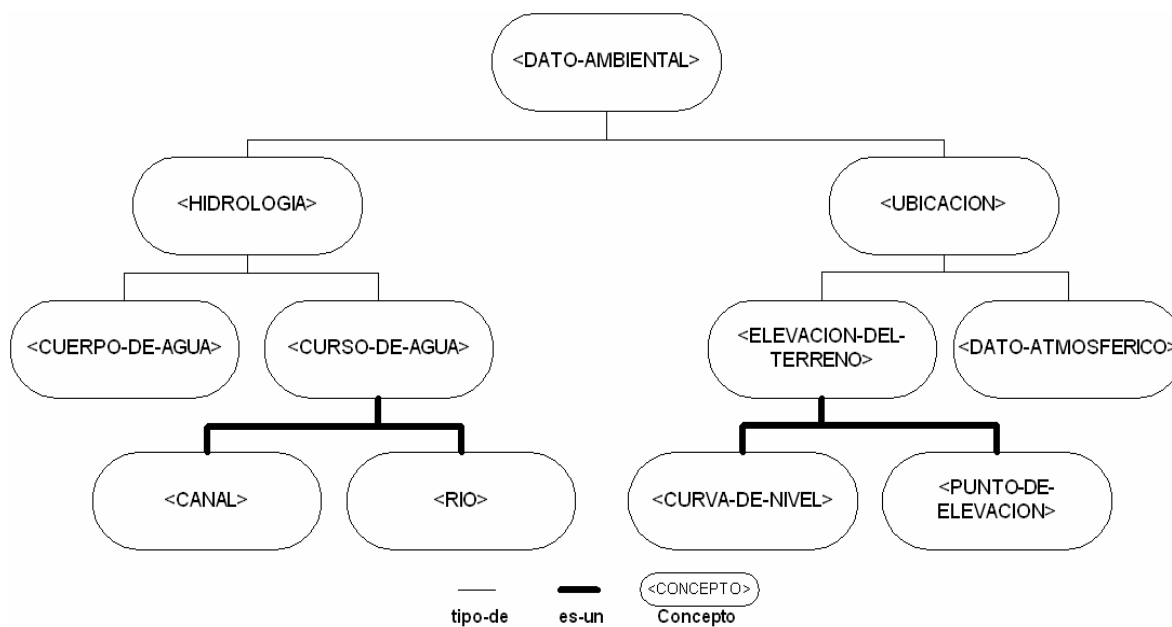
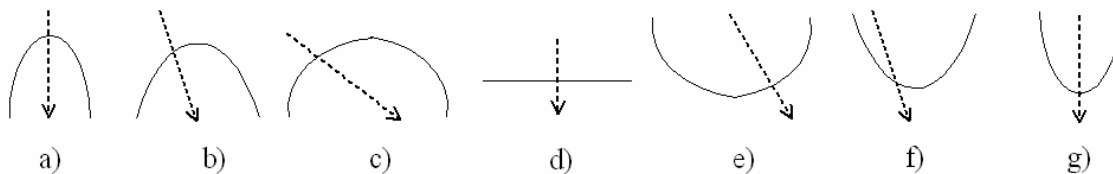


Figura 4.13 Jerarquía de conceptos con las relaciones es-un y tipo-de

<sup>44</sup> En semántica lingüística, se denomina *hipónimo* (del griego: *υπονόμιον*, que literalmente significa 'pocos nombres') a aquella palabra que posee todos los rasgos semánticos de otra más general, su *hiperónimo*, pero que añade en su definición otros rasgos semánticos que la diferencian de la segunda.

<sup>45</sup> La meronimia es una relación semántica no-simétrica entre los significados de dos palabras. Se denomina *merónimo* a la palabra cuyo significado constituye una parte del significado total de otra palabra, denominada *holónimo*.

Debido a que nuestra ontología se enfoca hacia la descripción de las relaciones entre los ríos y las curvas de nivel y en este momento la conceptualización tiene forma de estructura taxonómica, necesitamos agregar una relación que se denomina *>Pasa-por<* entre las clases *<RIO>* y *<CURVA-DE-NIVEL>*. Debido a que esta relación es muy importante necesitamos que sea lo suficientemente descriptiva para representar la forma en que pasa el río por la curva de nivel. Para esto, hemos considerado emplear las siguientes subclases; *<pasa-por-máxima-convexidad>*, *<casi-pasa-por-máxima-convexidad>*, *<pasa-por-convexidad>*, *<pasa-por-recta>*, *<pasa-por-concavidad>*, *<casi-pasa-por-máxima-concavidad>* y *<pasa-por- máxima-concavidad>* (Ver Figura 4.14).



**Figura 4.14** Representación gráfica de los casos de la relación *>Pasa-por<*; a) pasa por máxima convexidad, b) casi pasa por máxima convexidad, c) pasa por convexidad, d) pasa por recta, e) pasa por concavidad, f) casi pasa por máxima concavidad, g) pasa por máxima concavidad

En este caso, las propiedades que contenga la relación *>Pasa-por<* serán heredadas por sus subclases y podrán ser instanciadas. Esta relación tiene como restricción que solamente debe existir exactamente una relación entre objetos instanciados de las clases *<RIOS>* y *<CURVA- DE-NIVEL>*.

Desde el punto de vista de las propiedades de los datos geográficos, *>Pasa-por<* se refiere a una relación topológica entre dos entidades representadas por la primitiva de representación línea. En secciones posteriores se describirá en detalle cómo es que será obtenida sobre los datos, pero en este momento diremos que se trata de una *relación topológica refinada por métricas*, la cual se enfoca a la caracterización de cada uno de los casos presentados en la Figura 4.14.

Debido a que en ontologías las relaciones son también conceptos, la relación *>Pasa-por<* puede especificarse como el nodo padre de los conceptos que la describen, por lo que tendremos la estructura de la Figura 4.15, dónde la relación entre el padre y sus hijos es *tipo-de*, por lo que podemos decir que *<máxima-concavidad>* es un *tipo de relación*

>Pasa-por<. Adicionalmente, agregando conceptos que especialicen la >Pasa-por<, ganamos expresividad en la descripción de la ontología, ya que utilizando *pasa por máxima convexidad* es más descriptivo que decir simplemente *pasa por*.

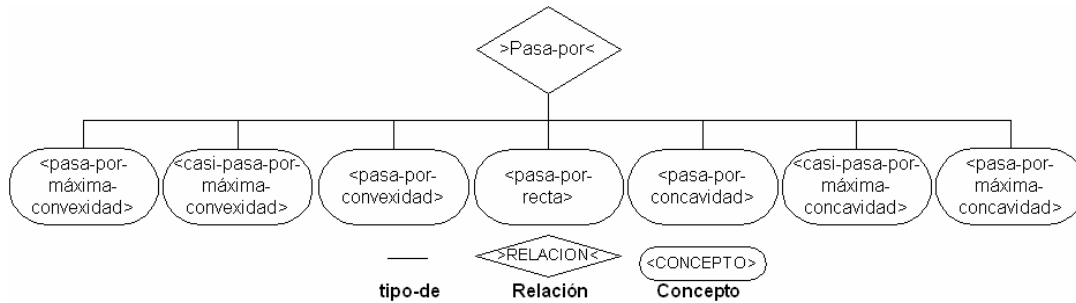


Figura 4.15 Representación jerárquica de la relación >Pasa-por<

Los conceptos hijo de >Pasa-por<, siguen una relación de orden. Dicha relación coloca en un extremo el concepto que se puede denominar como el ideal (*pasa por máxima convexidad*) y termina con el peor de los casos o menos consistente (*pasa por máxima concavidad*). Adicionalmente, esta relación posee una restricción: un río solo debe relacionarse una vez con una curva. Si el río *pasara* más de una vez por una curva de nivel representa que la relación es inconsistente y no cumple con las propiedades definidas en la Sección 4.1.2.2.3. Considerando lo anterior, diremos que >Pasa-por< tiene como restricción */cardinalidad = 1/*, es decir, cada río puede pasar por una curva solamente una vez. Al incluir la relación >Pasa-por< la representación gráfica de la ontología tendría la forma de la Figura 4.16.

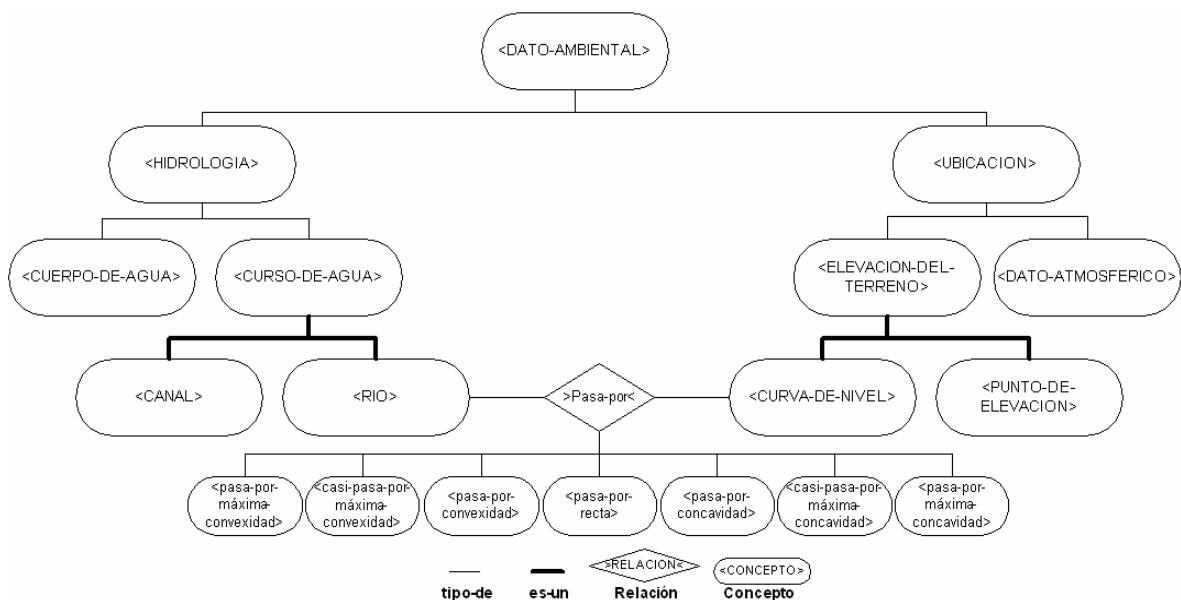


Figura 4.16 Representación de la ontología, al considerar la relación >Pasa-por<

### 4.1.2.5 Propiedades de las clases

En esta subsección describiremos las propiedades que deben tener cada una de las clases que componen la ontología. Esto se debe a que las clases por sí mismas no poseen suficiente información para responder a las preguntas planteadas en la Sección 4.1.2.1. Se consideran únicamente las clases que tendrán instancias, es decir las propiedades de las clases abstractas no son consideradas (Ver Tabla 4.2).

**Tabla 4.2.** Propiedades de las clases de la ontología

Clase	Propiedad	Descripción	Tipo	Formato	Rango
<RIO>	<i>Identificador</i> (IdRio)	Identificador único para cada objeto de la clase <RIO>	Alfanumérico	R-n	n: 1,2 ..., N
	<i>Nombre</i> (NomRio)	Nombre de cada objeto de la clase <RIO>	Nominal	Cadena[30]	Cadena de caracteres de tamaño 30
	<b>Tipo</b> (Tipo)	Indica si el río es temporal o perenne	Alfanumérico	Cadena[16]	<i>río-temporal</i> , <i>río-perenne</i>
	<b>Sinuosidad</b> (SinN)	Valor real que representa numéricamente la sinuosidad de cada objeto de la clase <RIO>	Real	r.r	r: 0, ..., R
	<b>Sinuosidad</b> (SinC)	Cadena de caracteres que representa conceptualmente la sinuosidad de la clase <RIO>	Cadena	Cadena[15]	Cadena de caracteres de tamaño 15
<CURVA-DE- NIVEL>	<i>Identificador</i> (IdCurva)	Identificador único para cada objeto de la clase <CURVA DE NIVEL>	Alfanumérico	C-n	n: 1,2 ..., N
	<i>Altura</i> (Alt)	Elevación sobre el msnm	Entero	N	n: 1,2 ..., N
>Pasa -por<	<i>Número de relación</i> (NumRel)	Número de relación	Entero	N	n: 1,2 ..., N
	<i>Dirección</i> (Direc)	Identificador que indica el número de relación que tiene un río siguiendo el flujo en sentido descendente <sup>46</sup>	Entero	N	n: 1,2 ..., N

<sup>46</sup> Significa que siguiendo el flujo del río desde su parte más alta, la primera relación >Pasa-por< entre el río y una curva de nivel será *Direc* = 1; la siguiente relación siguiendo el flujo será *Direc* = 2 , y así sucesivamente.

Hay que recordar que se consideran propiedades que sean útiles; propiedades adicionales podrían considerarse si la ontología estuviera enfocada hacia una aplicación distinta a la verificación de las relaciones entre las curvas de nivel y los ríos.

De la Tabla 4.2 se definen dos tipos de *propiedades*: *intrínsecas* y *extrínsecas*. Propiedades extrínsecas (en cursivas) se refiere a propiedades no esenciales que dependen de factores externos (por ejemplo, el nombre de un río o de una ciudad). Propiedades intrínsecas (en negritas) se refieren a propiedades internas o esenciales de los objetos que representan las clases (por ejemplo, el PH del agua de un río, la sinuosidad del río o el número de habitantes de una población).

Al igual que en el caso de las relaciones que se pueden definir en la ontología como conceptos (por ejemplo, la relación *>Pasa-por<*), también es posible representar las propiedades como conceptos. El propósito de esto es mejorar la expresividad de la ontología.

En el caso de esta tesis, incorporaremos a la ontología el concepto Sinuosidad, que se refiere a una *propiedad intrínseca* de los ríos. De esta forma diremos que *<Río> >Tiene< <Sinuosidad>*. Considerando lo anterior el concepto *<Sinuosidad>* tendrá como hijos a los siguientes conceptos: *<NadaSinuoso>*, *<PocoSinuoso>*, *<Sinuoso>*, *<MuySinuoso>* y *<DemasiadoSinuoso>*, los cuales podrán ser instanciados. La relación entre *<Río>* y sus conceptos hijos es *>tipo-de<*, por lo que *<NadaSinuoso>* es un *>tipo-de< <Sinuosidad>*.

Es importante señalar, que tanto los conceptos hijos de la relación *>Pasa-por<* como los de *<Sinuosidad>*, están ordenados por una relación de orden (Ver Sección 3.3.1.1). Entonces, podemos decir que el fragmento de la ontología que forman las clases *>Pasa-por<* y *<Sinuosidad>* y sus subclases, representa una *jerarquía ordenada* (Ver Figura 4.17).

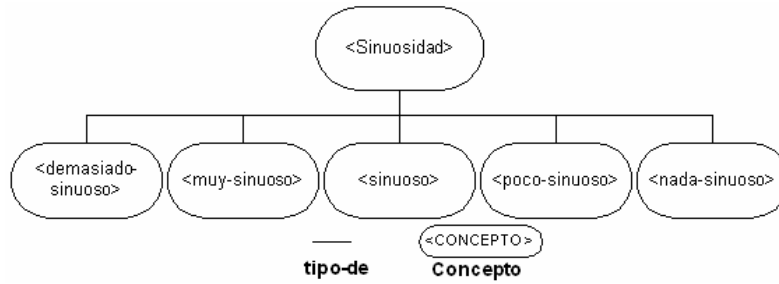


Figura 4.17 Representación jerárquica de la propiedad <Sinuosidad>

Así, la ontología definida tendrá la representación gráfica de la Figura 4.18. En dicha Figura aparecen la relación >Pasa-por< y la propiedad <Sinuosidad>, las relaciones *es-un* (línea gruesa), *tipo-de* (línea normal) y *tiene* (línea punteada). En caso de >Pasa-por<, se modelará como si fuera una clase. Asimismo, se presentan las clases (conceptos) abstractas (contorno normal) y las clases que pueden tener instancias (contorno punteado). Adicionalmente se agregó el concepto <Longitud>, que se refiere a la longitud del río, con el fin de dar consistencia a la ontología.

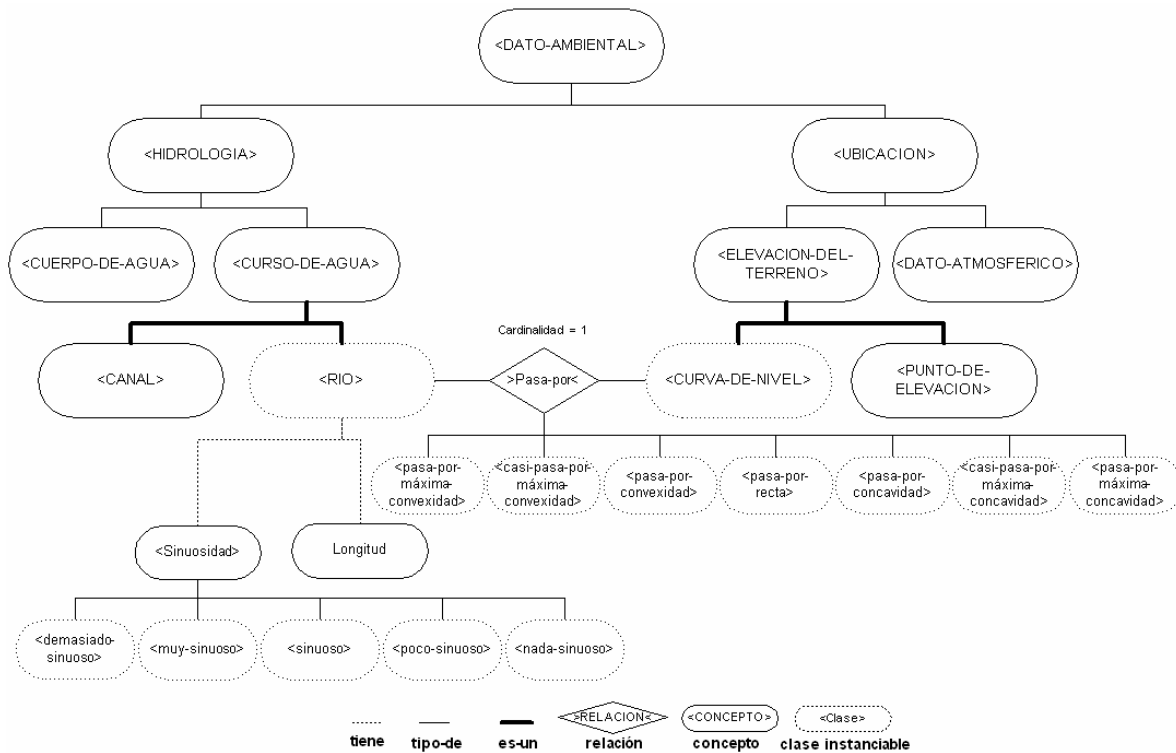


Figura 4.18 Representación de la ontología, al considerar la relación >Pasa-por< y la propiedad <Sinuosidad>

Por conveniencia, las restricciones para las subclases de <Sinuosidad> y >Pasa-Por< se presentarán en las Secciones 4.4.1. y 4.4.2 respectivamente. Esto se debe a que es ahí donde se describen las restricciones que se deben cumplir para pertenecer a las subclases <Sinuosidad> y >Pasa-Por<.

#### **4.1.2.6 Mantenimiento de la ontología**

Como se mencionó anteriormente esta ontología podrá ser enriquecida agregando nuevas clases, propiedades y relaciones. Para este propósito es necesario realizar un análisis de los objetos que se desean representar en la ontología.

Naturalmente, si se desean agregar más clases a la ontología hay que seguir el desarrollo *combinado*; definiendo conceptos generales y específicos. Con este fin se pueden consultar los documentos citados en la Sección 4.1.1. Los cuales son Especificación para la Codificación de Datos Ambientales (EDCS) [ISO, 2005], Diccionarios de Datos del INEGI a escalas 1:50,000 [INEGI, 1996] y 1:250,000, [INEGI, 1995] y Word Net [WordNet, 2007].

Un análisis de los objetos geográficos a describir permitirá identificar las relaciones que pueden existir entre las clases. Entre las relaciones que se pueden utilizar se encuentra el conjunto de relaciones espaciales: Topológicas (*intersecta con, cruza por, al lado de, dentro de, rodeado por, etc.*), de Dirección (*al norte de, al este de, etc.*) y de proximidad (*cerca de, muy lejos de, etc.*). Debido al enfoque que se le dio a la ontología descrita, estas relaciones, en la mayor parte de los casos, no forman parte de la estructura jerárquica, ya que son en sentido “horizontal”. Por ello es que se recomienda que sean modeladas en el mismo sentido que la relación >Pasa-por<, la que en sí misma representa una clase dentro de la ontología.

Entre los aspectos a tomar en cuenta se encuentra la definición de las relaciones *es-un* y *tipo-de*. Debe considerarse que: *Una subclase de una clase representa un concepto que es un “tipo de” concepto que la superclase representa.* Esto implica que debemos tener cuidado al definir o nombrar las clases. Éstas deben ser ya sea en singular o plural. Por

ejemplo, al definir una clase *Lagos* y una clase *Lago* como una subclase de *Lagos*, debido a que *Lago* no es un tipo de *Lagos*, esto puede originar un error en la interpretación de la ontología. Considerando lo anterior y con el fin de evitar ambigüedades en la ontología, todas las clases deben ser nombradas ya sea en singular o en plural.

Un aspecto importante a seguir al nombrar las clases es que cada clase represente un concepto y no las palabras que denotan esos conceptos, por lo que el nombre de una clase puede cambiar si se elige una terminología diferente. Por ejemplo, podemos crear una clase denominada *Curso de agua* que podrá ser cambiada por *Corriente de agua*. En términos prácticos los sinónimos no representan clases diferentes, por lo que se pueden poner términos en diversos idiomas o regionalismos que representen al mismo concepto.

Al agregar clases se deben evitar los ciclos. Se dice que hay un ciclo en una jerarquía cuando una clase A tiene una subclase B y al mismo tiempo B es una superclase de A. La creación de un ciclo como ese en una jerarquía equivale a declarar que las clases A y B son equivalentes: todas las instancias de A son instancias de B y todas las instancias de B son también instancias de A. Asimismo, puesto que B es una subclase de A, todas las instancias de B deben ser instancias de la clase A. Ya que A es una subclase de B, todas las instancias de A deben también ser instancias de la clase B.

Hay que evitar que se agreguen clases y se definan clases *hermanas* que no estén al mismo nivel de generalidad, lo que equivale en un libro de texto a que los capítulos no tengan la misma generalidad. Por lo anterior, se recomienda que las clases hermanas tengan la misma generalidad. Sin embargo, los conceptos en la raíz de la jerarquía (los cuales son a menudo representados como subclases directas de alguna clase muy general, como Thing (Cosa) representan divisiones principales del dominio y no tienen que ser conceptos similares.

No existen reglas para definir el número de subclases directas que una clase debería tener. Sin embargo, en algunas ontologías bien estructuradas tienen entre dos y una docena de subclases directas [Noy y McGuinness, 2001]. Considerando lo anterior, se pueden seguir las siguientes reglas: *Si una clase tiene solamente una subclase directa, puede existir un*



*problema de modelación, o si no la ontología no está completa. Si hay más de una docena de subclases para una clase dada, entonces pueden requerirse categorías intermedias adicionales.*

Cuando se agrega una clase, ya sea porque se requiere describir ciertos objetos que no aparecen en la ontología<sup>47</sup>, o bien se requiere especializar la ontología<sup>48</sup>, hay que considerar lo siguiente: Las subclases de una clase usualmente (1) tienen propiedades adicionales que la superclase no tiene, o (2) tiene diferentes restricciones que la superclase, o (3) participan en relaciones diferentes que la de las superclases. En otras palabras, cuando introducimos una nueva clase en la jerarquía, usualmente solo lo hacemos cuando hay algo que podamos decir acerca de esta clase que no podamos decir acerca de la superclase. En la práctica, cada subclase debe tener nuevas *propiedades (slots)* añadidos a ésta, o tener nuevos valores definidos para el *slot*, o bien se modifican las restricciones de las propiedades heredadas. Sin embargo, puede ser útil crear nuevas clases aún cuando no introduzcan nuevas propiedades. Por ejemplo, supongamos que la ontología se utilizará en la descripción detallada de los componentes de de redes hidrológicas; en este caso se puede incluir una clasificación de corrientes de agua muy especializada. Esta clasificación puede ser solo eso: una jerarquía de términos sin propiedades (o con el mismo conjunto de propiedades). En ese caso, es útil organizar los términos en la jerarquía en lugar de organizarlos en una lista plana porque (1) permitirá una exploración y navegación más fácil y (2) ayudará al especialista en la elección de un nivel de generalidad del término apropiado para una situación específica. Otra razón para introducir nuevas clases sin nuevas propiedades, es para modelar conceptos entre los cuales los expertos del dominio comúnmente hacen una distinción aun cuando no hayamos decidido modelar la distinción en sí.

Cuando modelamos un dominio, frecuentemente necesitamos decidir si modelar una distinción específica tal como río perenne o intermitente, como un valor de propiedad o como un conjunto de clases; nuevamente, depende del alcance del dominio y de la

---

<sup>47</sup> Tomando en cuenta la ontología que se definió anteriormente, suponer que se desean agregar vías de comunicación.

<sup>48</sup> Tomando en cuenta la ontología que se definió anteriormente, suponer que se desean definir los tipos de ríos perennes e intermitentes.

aplicación. Si la distinción es importante en el dominio y pensamos en los objetos con diferentes valores para la distinción como diferentes tipos de objetos, entonces deberíamos crear una nueva clase para la distinción.

Por último, para agregar clases a la ontología hay que considerar que no deberá contener toda la información posible del dominio: no se requiere especializar la aplicación más de lo necesario.

En las secciones anteriores se describió la ontología que se emplea en esta tesis, la cual está enfocada hacia la descripción de la forma en la que se relacionan los ríos y las curvas de nivel. Para tal propósito se han definido los conceptos que representan de forma explícita la totalidad de las relaciones existentes en el dominio. Debido a esto, es que se pretende utilizar para comparar dos conjuntos de datos y evaluar su similitud semántica. Adicionalmente, la ontología puede responder a preguntas específicas como: ¿Con cuáles curvas de nivel se relaciona el *río x*?, ¿Se relaciona el *río x* con la *curva de nivel y*?, etc.

## 4.2 Normalización

Esta etapa consiste en verificar la consistencia topológica de la base de datos espacial, así como en asignar identificadores únicos a cada uno de los objetos geográficos. Es importante esta etapa ya que de aquí depende en buena parte el éxito de las etapas siguientes (Análisis y Síntesis). Debido a que los datos que se utilicen deben ser topológicamente consistentes y deben poseer algunas características definidas en esta sección, se describen algunos aspectos que se consideran esenciales; otros aspectos dependen de los datos que se utilicen.

### 4.2.1 Consistencia de la base de datos

La *consistencia* se refiere a la ausencia de contradicciones lógicas. Una *inconsistencia* en un conjunto de datos es la violación de alguna de las restricciones de consistencia. Una restricción topológica es definida como la asociación de dos objetos geográficos, una relación topológica entre éstos y una especificación que puede ser la restricción *prohibida*,

por ejemplo, *prohibir* a un río (una línea) cruzar una construcción (polígono). Entonces, si se presentara el caso en donde un río cruza una construcción, esta relación violaría la restricción antes definida y por lo tanto es una inconsistencia [Ubeda y Egenhofer, 1997].

Básicamente, consideramos que para este trabajo nos debemos enfocar hacia:

- Relaciones topológicas entre objetos que pertenecen a la misma capa temática: (intersecciones, que representan conexiones entre objetos, en este caso verificar que los ríos que deban estar conectados entre sí). Así como también verificar que los arcos que forman una curva de nivel, se encuentren conectados.
- Dirección de los objetos: nos enfocamos hacia redes hidrológicas, por lo que la dirección del flujo es muy relevante.

Sin embargo, existen aplicaciones donde la forma de los objetos geográficos es importante, así que no debemos descartar la posibilidad de considerar otros casos.

Existen diversos procedimientos automáticos, semi-automáticos y manuales para verificar la consistencia como los presentados en [ESRI, 1999]. Proponemos los siguientes puntos para mejorar la consistencia de los datos geográficos<sup>49</sup>.

1. Verificar las conexiones entre los objetos línea, revisar si realmente se trata de un error o la representación es correcta<sup>50</sup> (Ver Figura 4.19a). La relación a verificar en términos de los componentes topológicos es  $n_k \in \partial O_i \cap O_j^\circ$ , donde  $n_k$  es cualquier nodo de un objeto lineal;  $\partial O_i$  es el límite de un objeto lineal;  $O_j^\circ$  es el interior de un objeto área.
2. Si dos objetos línea de la misma temática se intersectan, debe generarse un nodo en la intersección (Ver Figura 4.19c), este caso es complementario del anterior.
3. Verificar que arcos que representen objetos área realmente se representen por superficies cerradas<sup>51</sup> (Ver Figura 4.19b). En términos de los componentes topológicos

---

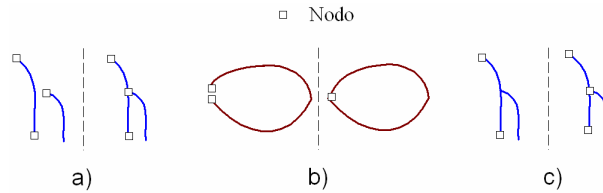
<sup>49</sup> Estas recomendaciones se dan en el contexto de este trabajo donde las temáticas son curvas de nivel y redes hidrológicas.

<sup>50</sup> En la literatura de ESRI se les conoce como Undershoot y Overshoot respectivamente.

<sup>51</sup> Es el caso de la curva de nivel, que si bien no será representada como un área, por definición describe una superficie cerrada.

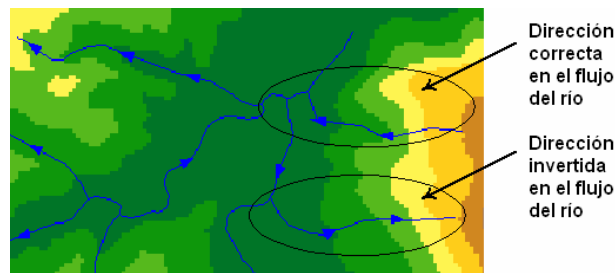
la relación a verificar puede representarse por  $n_k \in O_i^\circ \cap O_j^\circ$ ; donde  $O_i^\circ$  y  $O_j^\circ$  representan los límites de un objeto línea.

4. En el caso de curvas de nivel no deben existir bifurcaciones ni intersecciones, en este caso hay que evitar que  $n_k \in \partial O_i \cap O_j^\circ$ .



**Figura 4.19** Casos generales para verificar consistencia durante la etapa de normalización

Es importante verificar el flujo en objetos geográficos tales como ríos, ya que como se mencionó con anterioridad, la dirección del flujo se utiliza para detectar la consistencia de las relaciones con las curvas de nivel, pues un río con la dirección de flujo de una altura menor a una mayor, semánticamente es incorrecto. Por tal motivo, se propone utilizar un método automático como el propuesto en [Moreno, 2001] para verificar la dirección del flujo. Éste se basa en un modelo digital de elevaciones, para identificar la dirección correcta del flujo (Ver Figura 4.20)



**Figura 4.20** Identificación de la dirección del flujo de una red hidrológica. Las líneas con flechas indican la dirección del flujo, las áreas verde oscuras representan la zona con menor elevación

En términos formales del modelo vectorial debe revisarse para cada arco que la relación  $DIR\text{-}rio [arco_i] = 1$ , que se define como:

$$Dir\text{-}rio [arco_i] = 1 \text{ si } Elevacion(n_p) > Elevacion(n_q)$$

$$Dir\text{-}rio [arco_i] = 1 \text{ si } Elevacion(n_p) = Elevacion(n_q)$$

$$Dir\text{-}rio [arco_i] = 0 \text{ si } Elevacion(n_p) < Elevacion(n_q)$$

Donde:

$Elevacion(n_p)$ , es la elevación en el nodo inicial del arco,

$Elevacion(n_q)$ , es la elevación del nodo final del arco.

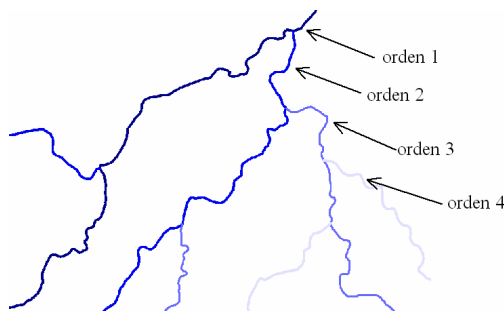
Para corregir la dirección del flujo se realiza lo siguiente<sup>52</sup>:

Donde  $Dir-rio [arco_i] = 0$ , hacer  $n_p := n_q$  y  $n_q := n_p$

Lo que representa que se invierte la dirección del flujo. En los casos en donde  $Elevacion(n_p) = Elevacion(n_q)$  y  $Dir-rio [arco_i] = 1$ , por omisión se considera que el río tiene la dirección correcta, para evitar errores en estos caso se utiliza un procedimiento interactivo donde el usuario define cual es la dirección correcta.

#### 4.2.2 Asignación automática de identificadores

La asignación de identificadores tiene como propósito que cada río puede estar compuesto de uno o más arcos<sup>53</sup>. Para realizar la identificación de qué arcos deben formar un sólo objeto proponemos usar la *Clasificación jerarquizada por longitud (CLAJER)* [Moreno, 2001], basada en la longitud de cada red hidrológica. Consiste en encontrar la ruta más larga desde cualquiera de las entradas hasta la salida<sup>54</sup>. La ruta más larga se clasifica como de primer orden. Los ríos de segundo orden son los ríos restantes más largos, que terminan en el río de primer orden. Los ríos de tercer orden serán aquellos que tengan una longitud mayor y que terminen en el río de segundo orden, y así sucesivamente (Ver Figura 4.21).



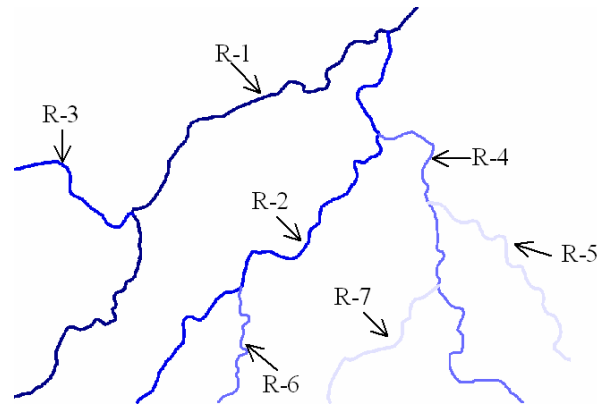
**Figura 4.21** Clasificación jerarquizada por longitud (CLAJER)

<sup>52</sup> Los términos de los datos geográficos representados con estructuras vectoriales con topología arco-nodo como las de las coberturas de Arc/Info  $n_p$  y  $n_q$ , se refieren a TNODE y FNODE respectivamente [ESRI, 1999].

<sup>53</sup> Los referimos como arcos ya que se trata del modelo vectorial.

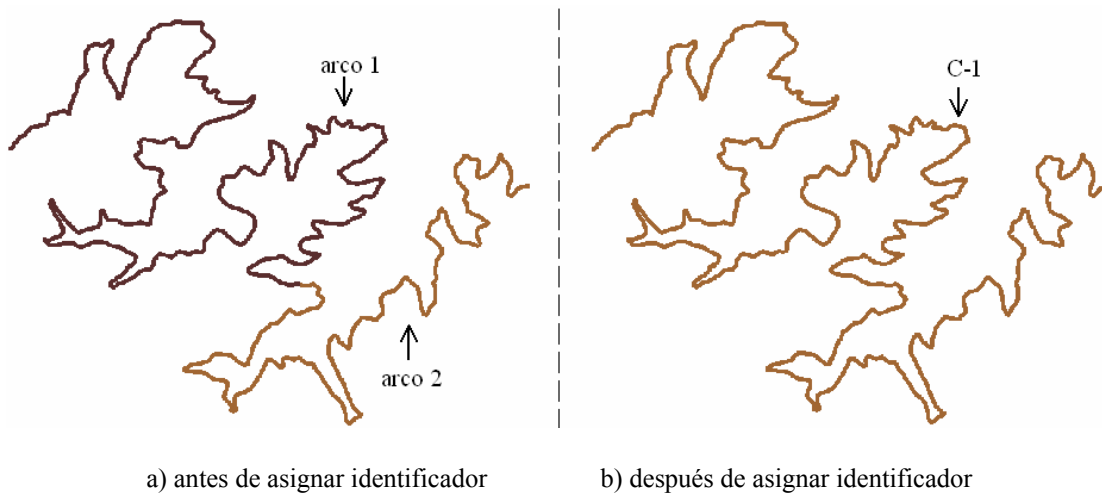
<sup>54</sup> La salida representa el punto más bajo de la red hidrológica; por ende las entradas representan los puntos con elevación mayor que en las salidas.

Así, cada uno de los ríos identificados por su orden se le asigna un identificador de la forma R-n, donde R se refiere a un río y  $n$  es un valor entero que va de 1 hasta  $N$ . Así los identificadores se asignan como en la Figura 4.22.



**Figura 4.22** Asignación de identificadores a los objetos en las redes hidrológicas

Para asignar un identificador único a las curvas de nivel se requiere utilizar los nodos que componen las curvas. Esto significa que para los arcos que tengan nodos comunes<sup>55</sup> representa que se trata de la misma curva (Ver Figura 4.23). Es posible realizar este proceso ya que en esta etapa se considera que todas las curvas de nivel poseen consistencia topológica, por lo que ya cumplen las reglas básicas que las definen (Ver Sección 4.2.1).



**Figura 4.23** Asignación de identificadores a las curvas de nivel

<sup>55</sup> No importando si se trata de nodo inicial o nodo final.

### 4.3 Análisis

Esta etapa se enfoca hacia la extracción y cuantificación de las propiedades que los datos geográficos poseen implícitamente. El mecanismo que se utiliza para realizar la cuantificación se basa en *medidas* (Ver Sección 3.2). Una *medida* es un método que no cambia el estado de los objetos geográficos, pero que se utiliza para caracterizarlo [Bader et al., 1999]. El concepto de medida tradicionalmente se ha aplicado a la generalización y por su naturaleza se ha considerado emplearlo en las descripciones de los datos geográficos. Se utiliza frecuentemente para diversos propósitos, tales como:

- Ayudar a detectar estructuras significativas en conjuntos de datos, tales como, grupos de objetos del mapa (por ejemplo, alineaciones) o características espaciales (por ejemplo, las formas y distribuciones de los objetos), que se deben preservar o al menos considerar durante el proceso de generalización.
- Conceptualizar aquellas estructuras representadas por grupos de objetos; pueden conceptualizarse al agrupar situaciones semejantes bajo un mismo concepto.
- Detectar conflictos, es decir, definir la necesidad de la generalización. En una aplicación basada en restricciones para la generalización, corresponde a la evaluación de las restricciones definidas para los objetos de la base de datos espacial.
- Seleccionar un operador de generalización apropiado en relación con la estructura semántica y espacial de los datos. Por ejemplo, polígonos aislados de la misma categoría pueden ser agrupados, mientras que los polígonos aislados de otras categorías se pueden desplazar en caso de un conflicto de proximidad.
- Definir los parámetros de los algoritmos de generalización; se derivan considerando la severidad de un conflicto.
- Evaluar la calidad de un resultado de la generalización comparando con el estado anterior, para decidir si el algoritmo seleccionado solucionó un conflicto. Con respecto a restricciones, este paso examina el cambio (mejora o empeora) del cumplimiento de la restricción.

En el contexto de esta investigación las medidas se utilizan para cuantificar las propiedades de los objetos espaciales. Como resultado de las medidas se obtiene un valor numérico denominado *descriptor cuantitativo*. Este descriptor representa en forma numérica una propiedad. Sin embargo, un número en sí mismo no proporciona una idea clara de lo que representa, al menos en el caso de la mayor parte de las personas. Por todo lo anterior, en la Sección 4.1.2.5 se detalla cómo se añade significado a cada propiedad.

Para definir la medición nos apoyamos en el concepto 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  $B$ )*<sup>56</sup>, simbólicamente es representado por  $m: A \rightarrow B$ . 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 *rango* de  $m$ . En general no se asume que todos los valores de  $B$  que existen en  $b \in B$  que no tienen una imagen<sup>57</sup> en ninguna  $a$ , es decir  $m(a)$  es subconjunto de  $B$ . Si  $m(A)=B$  el mapa se llama  *$A$  sobre  $B$  ( $A$  onto  $B$ )*<sup>58</sup> 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  $a_1, a_2, a_3, \dots, a_n$ , 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  $m: A \rightarrow B$ , 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 *descriptores cuantitativos*.

<sup>56</sup> Se trata de una función inyectiva, cuando las imágenes en el rango corresponden con elementos diferentes del rango.

<sup>57</sup> Sean conjuntos  $X$  y  $Y$ ,  $m$  el mapa  $m: X \rightarrow Y$ , 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.

<sup>58</sup> 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.



Es conveniente señalar que las mediciones pueden estar enfocadas a diferentes aspectos de los datos geográficos, los cuales pueden ser:

- *Relaciones entre objetos geográficos.* Las relaciones entre objetos geográficos se clasifican en tres grupos: Relaciones de Proximidad (por ejemplo, *cerca, lejos, muy lejos*), Dirección cardinal (por ejemplo, *al Este de, al Norte de*) y Topológicas (por ejemplo, *conectado, adyacente, pasa por*).
- *Propiedades geométricas de los objetos geográficos.* Se refiere a características geométricas que poseen los datos geográficos, las cuales serán cuantificadas mediante una medida.

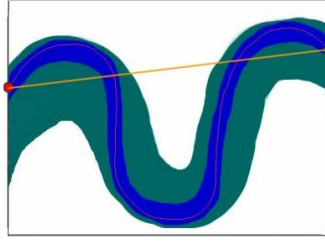
De acuerdo al caso de estudio de este trabajo, nos enfocaremos principalmente en la definición de dos mediciones que tienen que ver con aspectos geométricos y/o topológicos. Dichos aspectos están señalados en la conceptualización definida a priori. Las mediciones se presentan de forma general en la Tabla 4.3.

**Tabla 4.3.** Mediciones a utilizar

<b>Objetos</b>	<b>Medida</b>	<b>Tipo</b>
Ríos	Sinuosidad	Geométrica/Forma
Ríos y curvas de nivel	Medición de la relación Pasa por	Topológica refinada con métrica

### 4.3.1 Medición de la Sinuosidad

En general la sinuosidad se refiere a la cualidad o condición de ser sinuoso, y sinuoso significa curvo, ondulante, por lo que podemos decir que se refiere a cuan curva es una línea. Desde el punto de vista de la hidrología, Sinuosidad es la proporción de la longitud del río con respecto al valle que forma (Ver Figura 4.24).



**Figura 4.24** Sinuosidad de un río

Esta característica es muy peculiar en los ríos, ya que la sinuosidad define la apariencia del cauce. Comparando con distintas escalas puede apreciarse las diferencias que tiene la sinuosidad para el mismo río.

Desde el punto de vista de este trabajo se ha definido la sinuosidad de un río como:

$$Sin(O_{rio}) = \frac{\sum_1^k \sqrt{(v_{i_x} - v_{i-1_x})^2 + (v_{i_y} - v_{i-1_y})^2}}{\sqrt{(n_{q_x} - n_{p_x})^2 + (n_{q_y} - n_{p_y})^2}},$$

Donde:

$k$ , es el número de vértices que tiene la representación vectorial del objeto río,

$v_{i_x}$ ,  $v_{i_y}$ ,  $v_{i-1_x}$ ,  $v_{i-1_y}$  Son las coordenadas de los vértices del objeto río (Ver Figura 4.25),

$n_{p_x}$ ,  $n_{p_y}$  son las coordenadas del nodo inicial del objeto río,

$n_{q_x}$ ,  $n_{q_y}$  son las coordenadas del nodo final del objeto río.

Si se trata de un  $O_{rio}$  totalmente recto la sinuosidad será igual a 1.

Hay que señalar que en este caso no se calcula la sinuosidad de un arco sino la sinuosidad de un objeto río que puede estar compuesto por uno o más arcos. En este caso el valor de  $Sin(O_{rio})$  será denominado como *Descriptor Cualitativo de la Sinuosidad (DCS)*.

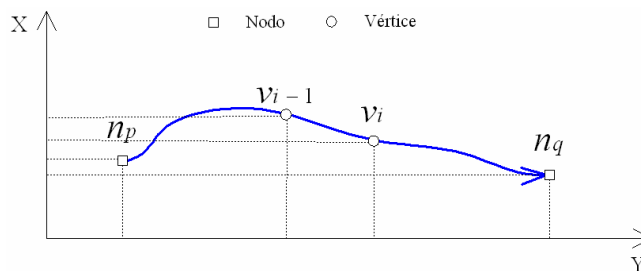


Figura 4.25 Componentes para medir la sinuosidad de un río

### 4.3.2 Medición de la relación >Pasa-por<

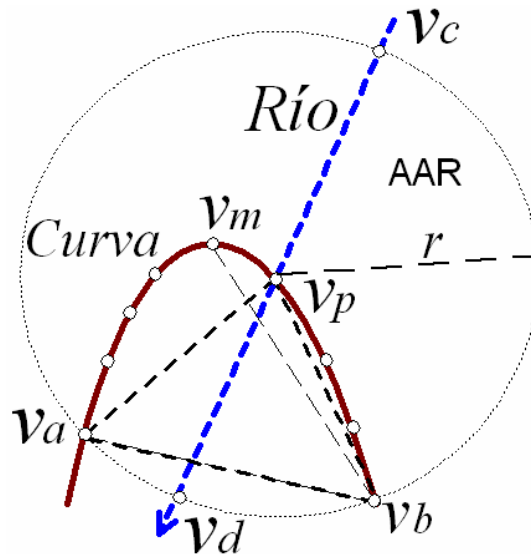
Como se mencionó en secciones anteriores la relación topológica >Pasa-por< será usada para describir la forma en que se relacionan ríos y las curvas de nivel. En el sentido más básico de la relación podría describirse como:  $n_k \in O_i^\circ \cap O_j^\circ$ , donde  $O_i^\circ$  y  $O_j^\circ$  son los interiores de dos objetos río y  $n_k$  es un nodo común de los objetos. Considerando los casos definidos en la conceptualización (Sección 4.1.2.2.3) no es posible identificar cada uno de ellos.

Esto se debe a que el modelo de 9-intersección no permite distinguir esos casos. 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 una *Relación Topológica refinada con métricas*, que conjunta 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]

Existen en la literatura definiciones de las relaciones topológicas refinadas con métricas como la presentada en [Nedas et al., 2007]. Para describir la relación *Pasa por* se utiliza el *Descriptor Cualitativo de la relación Pasa por entre ríos y curvas de nivel (DCP)*.

Dicho descriptor se calcula como sigue:

1. Se identifica como  $v_p$  la intersección del río ( $O_R$ ) con la curva de nivel ( $O_C$ )
2. Se define un área circular denominada *área de análisis de la relación (AAR)* con radio  $r$  y centro en  $v_p$ . (Ver Figura 4.26)
3. A partir de *AAR*, se identifican dos puntos de intersección entre la curva de nivel y *AAR*, los cuales se denominan como  $v_a$  y  $v_b$  ( $v_a, v_b \in O_C \cap \partial AAR$ ). Además se identifican 2 puntos de intersección entre el río y *AAR* que se denominan como  $v_c$  y  $v_d$  ( $v_c, v_d \in O_C \cap \partial AAR$ ).
4. Se busca un vértice sobre la curva de nivel, que forme el área más grande junto con  $v_a$  y  $v_b$ . A dicho vértice se le llama  $v_m$ . Se calcula  $A_P = \text{area}(v_a, v_b, v_p)^{59}$  y  $A_M = \text{area}(v_a, v_b, v_m)$ .



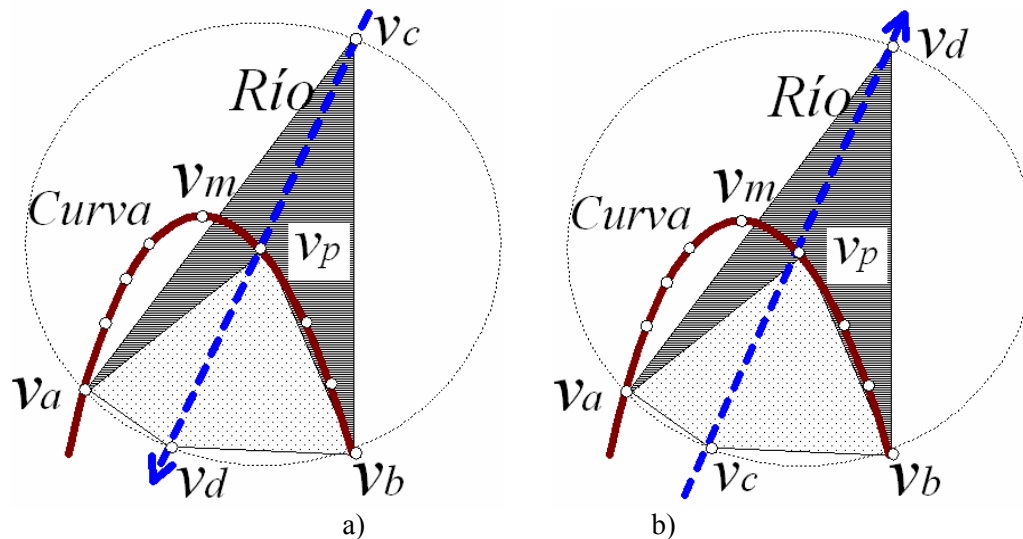
**Figura 4.26** Análisis topológico para identificar el grado de concavidad/convexidad por el que pasa un río a través de una curva de nivel

5. Se identifica concavidad o convexidad. Se calcula  $P_A = \text{perimetro}(v_a, v_b, v_p, v_c)^{60}$  y  $P_D = \text{perimetro}(v_a, v_b, v_p, v_d)$ 
  - Si  $P_A > P_D$ , representa que  $O_R$  pasa por una convexidad en  $O_C$ , entonces  $DCP = A_P / A_M (I)$  (Ver Figura 4.27a)

<sup>59</sup>  $\text{area}(v_a, v_b, v_p)$  calcula el área compuesta por 3 puntos. Para realizar este procedimiento recomendamos utilizar la fórmula de Herón que se basa en las distancias que existen entre dichos puntos, lo que facilita el cálculo.

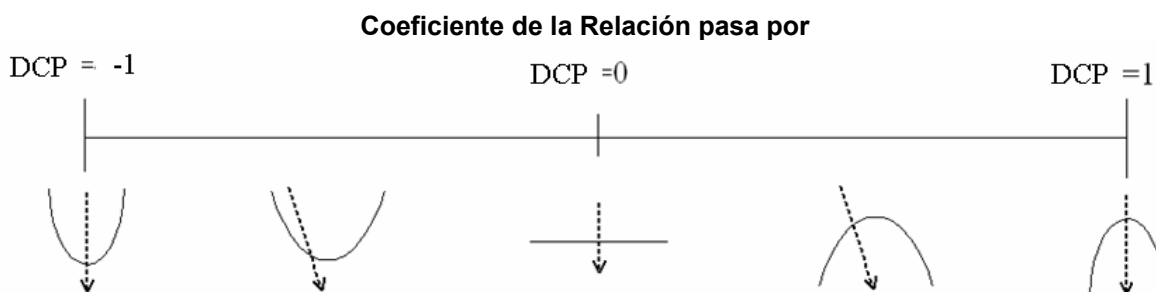
<sup>60</sup>  $\text{perimetro}(v_a, v_b, v_p, v_c)$  calcula el perímetro compuesto por 4 puntos

- Si  $P_A > P_D$ , representa que  $O_R$  pasa por una concavidad en  $O_C$ , entonces  $DCP = A_P / A_M (-1)$  (Ver Figura 4.27b)
- Si  $P_A = P_D$ , representa que  $O_R$  pasa por un parte recta de  $O_C$ , entonces  $DCP = 0$



**Figura 4.27** Determinación de la concavidad/convexidad en una curva de nivel  
De la forma en que se definió  $DCP$ , si el río pasa por la máxima convexidad de la curva entonces  $DCP = 1$ ; si el río pasa por la máxima concavidad de la curva de nivel será  $DCP = -1$  y en caso de que no exista ni concavidad ni convexidad,  $DCP = 0$  (es decir, que la curva de nivel es recta). Así que  $DCP$  tomará valores entre  $[-1 \leq 0 \leq 1]$ . Los valores de  $DCP$  se pueden interpretar como sigue (Figura 4.28):

- Si  $DCP = 1$ , significa que el río pasa exactamente por la máxima convexidad de la curva de nivel.
- Si  $1 > DCP > 0$ , significa que el río pasa por alguna máxima convexidad de la curva de nivel.
- Si  $DCP = 0$ , significa que el río pasa por una parte recta de la curva de nivel.
- Si  $0 > DCP > -1$ , significa que el río pasa por alguna máxima concavidad de la curva de nivel.
- Si  $DCP = -1$ , significa que el río pasa exactamente por la máxima concavidad de la curva de nivel.



**Figura 4.28** Equivalencia entre la representación gráfica de la relación >Pasa-por< entre un río y una curva de nivel con respecto al valor de DCP. Las líneas punteadas representan ríos, las líneas sólidas representan curvas de nivel

## 4.4 Síntesis

El objetivo de esta etapa es generar automáticamente una *representación conceptual explícita* de las propiedades que fueron extraídas de los datos y que están representadas por *descriptores cuantitativos (DC)*. Cada grupo de *DC* se clasifica en intervalos, de manera que cada intervalo representa una clase<sup>61</sup> o grupo que representa un concepto, por lo que la clasificación pretende agrupar valores en grupos significativos. Cada concepto pretende representar de forma explícita el significado que tiene una determinada propiedad y/o relación en cada objeto o grupos de objetos. Para evitar ambigüedades y redundancias, cada concepto se describió en la ontología definida a priori. Así el proceso de Síntesis se divide en dos etapas:

1. Clasificación de los *descriptores cuantitativos* de cada propiedad y/o relación en clases de acuerdo a su valor.
2. Asignación de los conceptos de acuerdo a los intervalos de clasificación de los *DC* de las mediciones y generación de la *base de datos espacial parcialmente conceptualizada*.

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

<sup>61</sup> En este caso se refiere a las clases que se definieron en la conceptualización.

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:

$$M[DC_i, C_j] = 1 \text{ si } DC_i \text{ es miembro de } C_j ; \text{ de otra forma } M[DC_i, C_j] = 0$$

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

$$CLASE(DC) = \{C | M[DC, C] = 1\}$$

La extensión de la clase es el conjunto de todos los objetos que pertenecen a ella:

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

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*<sup>62</sup> con clases compuestas por:

Pobladors:	<i>restricción</i> $100 < \text{número de habitantes} < 1,000$
Cabecera Municipal	<i>restricción</i> $1,000 < \text{número de habitantes} < 50,000$
Ciudades	<i>restricción</i> $50,000 < \text{número de habitantes} < 10,000,000$
Mega ciudades	<i>restricción</i> $10,000,000 < \text{número de habitantes}$

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 valores que poseen los *descriptores cuantitativos*.

Una vez que se determinó la pertenencia de un  $DC$  a una clase, el valor numérico será reemplazado por una etiqueta que describe la clase a la que pertenece. El propósito de esto, es representar cuantitativamente un *descriptor cuantitativo*. Desde otro punto de vista,

---

<sup>62</sup> El INEGI utiliza el término *localidad* para hacer referencia a pueblos, ciudades y en general para cualquier asentamiento humano con mas de 500 habitantes

representa la representación explícita de propiedades y/o relaciones que poseen los datos geográficos. Por ejemplo, suponer que tenemos la localidad *Morelia* con 500,000<sup>63</sup> habitantes, representa que el DC del número de habitantes es igual a 500,000. Entonces, *Morelia* es miembro de la clase Ciudades, por lo que el valor 500,000 será sustituido por Ciudad.

Una vez que se asignaron clases a los *descriptores cuantitativos*, se requiere una *representación conceptual* que describa explícitamente las propiedades y relaciones que fueron medidas. Para este propósito se define una *representación conceptual explícita* denominada *base de datos espacial parcialmente conceptualizada (BDEPC)*. Como en la ontología pueden representarse tanto propiedades como relaciones por medio de conceptos, la *BDEPC* estará compuesta por tablas de dos tipos.

- *Tabla relaciones conceptualmente representadas* ( $T_R$ ), la que almacena las relaciones entre dos objetos geográficos.
- *Tabla de propiedades conceptualmente representadas* ( $T_P$ ), la que almacena propiedades de objetos geográficos.

Las tablas  $T_R$  y  $T_P$  se definen como sigue:

Una tabla  $T_R$  de la *BDEPC* se define como:  $D_1$  que representa el conjunto de objetos que se relacionan (en este caso, los objetos de la clase <RIO>) y  $D_2$  que representa el conjunto de conceptos que describen una relación identificada utilizando una medición (en este caso, la relación >Pasa-por<) y  $D_3$  que representa el conjunto de los objetos relacionados con los objetos de  $D_1$  (en este caso, los objetos de la clase <CURVAS-DE-NIVEL>). Considerando lo anterior, la tabla  $T_R$  se puede representar como un subconjunto de  $D_1 \times D_2 \times D_3$ . Entonces, la Tabla  $T_R$  se puede describir como  $T_R(O_1, R, O_2)$  donde  $O_1 \in D_1$ ,  $R \in D_2$  y  $O_2 \in D_3$ . Las *tuplas* de  $T_R$  se especificarán como  $T_R \{O_1, R, O_2\}$ . Hay que señalar que los valores que tenga  $R$ , serán etiquetas que representan una clase. Es importante mencionar que en este caso  $R$  puede ser cualquier tipo de relación entre objetos geográficos, es decir, puede especificar una *relación de distancia* (*A cerca de B*), *dirección cardinal* (*A está al*

<sup>63</sup> Suponer que 500,000 habitantes fue obtenido por medio de una medición.



*Norte de B*) o bien una relación topológica (como en el caso de esta tesis, donde se utiliza una relación topológica refinada con métricas). Adicionalmente, en este caso las propiedades de las clases también son almacenadas, quedando las tablas de la forma  $T_R (O_1, R, O_2, P_1, P_2, \dots, P_n)$ , donde  $P_1, P_2, \dots, P_n$  son las propiedades de la clase que representa  $R$ .

Una tabla  $T_P$  de la *BDEPC* se define como:  $D_4$  que representa el conjunto de objetos que tienen una determinada propiedad, la cual fue medida y representada mediante un DC y se desea describir explícitamente (en este caso, los objetos de la clase <RIO>),  $D_5$  que representa las relación entre el objeto y la propiedad (en este caso, la relación *tiene*) y  $D_6$  que representa el conjunto de conceptos que describen una propiedad (en este caso, <Sinuosidad>). Considerando lo anterior, la tabla  $T_P$  se puede representar como un subconjunto de  $D_4 \times D_5 \times D_6$ . Entonces, la Tabla  $T_P$  se puede describir como  $T_P (O_1, R, P_1)$  donde  $O_1 \in D_1, R \in D_2$  y  $P_1 \in D_3$ . Las *tuplas* de  $T_P$  se especificarán como  $T_{P_i} \{O_1, R, P_1\}$ . Es importante mencionar que en el caso de esta tabla pudiera suceder que todas la tuplas tengan el mismo valor para  $R$ . Con base en lo anterior, podemos decir que  $R$  podría omitirse, ya que frecuentemente las propiedades de un objeto se especifican por la relación *tiene*. Así como en el caso anterior, los valores que tenga  $P$ , serán etiquetas que representan una clase. Para especificar a objetos geográficos se recomienda utilizar los identificadores de objetos (Ver Sección 4.2.2).

A las *tuplas* que componen a las tablas  $T_P$  y  $T_R$  se le denomina como *descripciones semánticas*. Esta denominación está motivada en que cada *tupla* representa la semántica de la ontología, por lo que, cada *descripción semántica* representa el significado de una relación. Donde la relación puede ser entre pares de objetos ( $T_R (O_1, R, O_2)$ ) o bien entre un objeto y una propiedad ( $T_P (O_1, R, P_1)$ ).

Es muy importante señalar que tanto  $T_P (O_1, R, P_1)$  como  $T_R (O_1, R, O_2)$ , representan a las clases que pueden ser instanciadas dentro de la ontología, tales como: <RIO>, <CURVA-DE - NIVEL>, <passa-por-máxima - convexidad>, <casi-pasa-por-máxima-convexidad>, <passa-por-convexidad>, <passa-por - recta>, <passa-por-concavidad>, <casi-pasa-por-máxima-

*concauidad*>, <*pasa-por - máxima-concauidad*>, <*NadaSinuoso*>, <*PocoSinuoso*>, <*Sinuoso*>, <*MuySinuoso*> y <*DemasiadoSinuoso*>.

Por su estructura la *BDEPC* tiene algunas ventajas con respecto a otros enfoques para representar los datos geográficos, ya que representa explícitamente propiedades y relaciones que éstos poseen. Debido a que se utiliza una Base de Datos su estructura es compacta y las consultas para encontrar alguna relación o propiedad se realiza utilizando lenguajes de consulta, por lo que puede implementarse en una gran cantidad de manejadores, aunque éstos no sean orientados a manipular información geoespacial, por lo que no depende del formato. Adicionalmente, con este enfoque se maneja información geográfica sin considerar aspectos que normalmente se han utilizado en enfoques tradicionales como: pares de coordenadas, unidades de medida, ángulos, escala, sistema de referencia, geoide, elipsoide, etc. Además no importa la dimensión que tenga el objeto, su representación conceptual siempre será igual.

Aspectos como la geometría (métricas) y topología se representan explícitamente por medio de conceptos, por lo que esta representación pretende acercarse a la conceptualización que una persona realiza sobre una situación en particular, por ejemplo considerando, la tupla {México, *es vecino de*, USA} se acerca bastante a la forma en la que una persona definiría una relación y ubicaría a los objetos geográficos en el espacio sin necesidad de coordenadas. No hay que olvidar que debe existir un consenso de la forma en que un grupo de personas describe e interpreta una relación o propiedad. Esta es una de las razones por las que proponemos una ontología, ya que en sí misma representa un consenso.

Adicionalmente, este enfoque permite la representación explícita de propiedades que tienen los datos implícitamente, como la Sinuosidad. Así mismo, el enfoque tradicional de GIS se basa en el uso de capas temáticas y en gran parte de los sistemas comerciales se manipula una temática a la vez. En el enfoque de *representación conceptual* de datos geográficos, es posible integrar en una representación compacta dos temáticas, donde no se pierden las características propias de cada temática, debido a que cada concepto (relación) se asigna con base en un *descriptor cuantitativo* que se mapea a una conceptualización de un dominio

en específico, donde se preservan las características que van heredando de las categorías superiores, por lo que cada concepto o clase en la representación conceptual no pierde las peculiaridades que lo definen.

En las siguientes secciones se describirán los casos de estudio desarrollados para esta investigación.

#### 4.4.1 Generación de la tabla de propiedades conceptualmente representadas para la sinuosidad de los ríos

Los criterios para clasificar la sinuosidad para ríos se presentan en la Tabla 4.4 y representan las restricciones que tienen los objetos para pertenecer a la clase <Sinuosidad>. Los valores fueron obtenidos por [Verastegui, 2007], con base en [Rosgen, 1964], utilizando resultados experimentales obtenidos por medio de encuestas a usuarios de datos geográficos y GIS.

Estos criterios o restricciones, se utilizan para identificar la clase a la que pertenece cada *Descriptor Cualitativo de la Sinuosidad (DCS)* obtenida por la medición de  $Sin(O_{rio})$ .

**Tabla 4.4.** Restricciones para pertenecer a las subclases <Sinuosidad>

Clase	Criterio
<NadaSinuoso>	[1.00 < DCS < 1.34]
<PocoSinuoso>	[1.35 < DCS < 1.5]
<Sinuoso>	[1.60 < DCS < 2.0]
<MuySinuoso>	[2.1 DCS < 2.6]
<DemasiadoSinuoso>	[2.7 < DCS ]

Considerando la Tabla 4.4, la definición de las clases para cada *DCS* se define como:

$$CLASE(DCS) = \{ \langle NadaSinuoso \rangle \mid M[DCS, \langle NadaSinuoso \rangle] = 1 \}$$

$$M[DCS, \langle NadaSinuoso \rangle] = 1 \text{ si y solo si } [1.00 < DCS < 1.34]$$

$$CLASE(DCS) = \{ \langle PocoSinuoso \rangle \mid M[DCS, \langle PocoSinuoso \rangle] = 1 \}$$

$$\begin{aligned}
 &M[DCS, \langle PocoSinuoso \rangle] = 1 \text{ si y solo si } [1.35 < DCS < 1.5] \\
 &CLASE(DCS) = \{ \langle Sinuoso \rangle \mid M[DCS, \langle Sinuoso \rangle] = 1 \} \\
 &M[DCS, \langle Sinuoso \rangle] = 1 \text{ si y solo si } [1.60 < DCS < 2.0] \\
 &CLASE(DCS) = \{ \langle MuySinuoso \rangle \mid M[DCS, \langle MuySinuoso \rangle] = 1 \} \\
 &M[DCS, \langle MuySinuoso \rangle] = 1 \text{ si y solo si } [2.20 < DCS < 2.6] \\
 &CLASE(DCS) = \{ \langle DemasiadoSinuoso \rangle \mid M[DCS, \langle DemasiadoSinuoso \rangle] = 1 \} \\
 &M[DCS, \langle DemasiadoSinuoso \rangle] = 1 \text{ si y solo si } [2.7 < DCS]
 \end{aligned}$$

Donde,  $DCS$  representa numéricamente la sinuosidad del  $O_{rio\ i}$  y en dependencia del valor se define la clase a la que cada  $DCS$  pertenece, las cuales pueden ser  $\langle NadaSinuoso \rangle$ ,  $\langle PocoSinuoso \rangle$ ,  $\langle Sinuoso \rangle$ ,  $\langle MuySinuoso \rangle$  y  $\langle DemasiadoSinuoso \rangle$ . Por esto, las etiquetas que se asignarán para cada objeto, reemplazando al  $DCS$  son:  $NadaSinuoso$ ,  $PocoSinuoso$ ,  $Sinuoso$ ,  $MuySinuoso$  y  $DemasiadoSinuoso$ .

De esta forma, la tabla  $T_P$  de la  $BDEPC$ , que se genera para esta propiedad tendrá descripciones semánticas de la forma  $T_{P_i} \{Río, Tiene, Sinuosidad\}$ .

#### 4.4.2 Generación de la tabla de relaciones conceptualmente representadas para la relación >Pasa-por< entre los ríos y las curvas de nivel

Las restricciones para definir la membresía de los *descriptores cuantitativos de la medición de la relación >Pasa-por<* ( $DCP$ ), se definieron con base en la experiencia del autor. Fueron el resultado de un análisis profundo de diversas representaciones que involucran curvas de nivel y ríos (mapas topográficos y geológicos del INEGI a diversas escalas). Esto es debido a que en mi opinión, en la literatura no existen valores para clasificar dicha relación (Ver Tabla 4.5).

**Tabla 4.5.** Restricciones para pertenecer a las subclases de >Pasa-Por<

Clase	Criterio
<pasa-por-máxima-convexidad>	$[0.95 < DCP < 1]$
<pasa-casi-por-máxima-convexidad>	$[0.85 < DCP < 0.94]$
<pasa-por-convexidad>	$[0.35 < DCP < 0.84]$
<pasa-por-recta>	$[-0.34 < DCP < 0.34]$
<pasa-por-concavidad>	$[-0.84 < DCP < -0.35]$
<pasa-casi-por-máxima-concavidad>	$[-0.94 < DCP < -0.85]$
<pasa-por-máxima-concavidad>	$[-1 < DCP < -0.95]$

Considerando la Tabla 4.5, la definición de las clases para cada *DCP* se define como:

$$CLASE (DCP) = \{ \langle \text{pasa-por-máxima-convexidad} \rangle \mid M [DCP, \langle \text{pasa-por-máxima-convexidad} \rangle] = 1 \}$$

$$M [DCP, \langle \text{pasa-por-máxima-convexidad} \rangle] = 1 \text{ si y solo si } [0.95 < DCP < 1]$$

$$CLASE (DCP) = \{ \langle \text{pasa-casi-por-máxima-convexidad} \rangle \mid M [DCP, \langle \text{pasa-casi-por-máxima-convexidad} \rangle] = 1 \}$$

$$M [DCP, \langle \text{pasa-casi-por-máxima-convexidad} \rangle] = 1 \text{ si y solo si } [0.85 < DCP < 0.94]$$

$$CLASE (DCP) = \{ \langle \text{pasa-por-convexidad} \rangle \mid M [DCP, \langle \text{pasa-por-convexidad} \rangle] = 1 \}$$

$$M [DCP, \langle \text{pasa-por-convexidad} \rangle] = 1 \text{ si y solo si } [0.35 < DCP < 0.84]$$

$$CLASE (DCP) = \{ \langle \text{pasa-por-recta} \rangle \mid M [DCP, \langle \text{pasa-por-recta} \rangle] = 1 \}$$

$$M [DCP, \langle \text{pasa-por-recta} \rangle] = 1 \text{ si y solo si } [-0.34 < DCP < 0.34]$$

$$CLASE (DCP) = \{ \langle \text{pasa-por-concavidad} \rangle \mid M [DCP, \langle \text{pasa-por-concavidad} \rangle] = 1 \}$$

$$M [DCP, \langle \text{pasa-por-concavidad} \rangle] = 1 \text{ si y solo si } [-0.84 < DCP < -0.35]$$

$$CLASE(DCP) = \{ \langle \text{pasa-casi-por-maxima-concavidad} \rangle \mid M [DCP, \langle \text{pasa-casi-por-maxima-concavidad} \rangle] = 1 \}$$

$$M [DCP, \langle \text{pasa-casi-por-maxima-concavidad} \rangle] = 1 \text{ si y solo } [-0.94 < DCP < -0.85]$$

$$CLASE(DCP) = \{ \langle \text{pasa-por-maxima-concavidad} \rangle \mid M [DCP, \langle \text{pasa-por-maxima-concavidad} \rangle] = 1 \}$$

$$M [DCP, \langle \text{pasa-por-maxima-concavidad} \rangle] = 1 \text{ si y solo } [-1 < DCP < -0.95]$$

Donde, DCP, representa el coeficiente de la relación  $\text{>Pasa-por<}$  entre un objeto río y un objeto curva de nivel. En dependencia del valor de DCP se define la membresía a las clases  $\text{<pasa-por-máxima - convexidad>}$ ,  $\text{<casi-pasa-por-máxima-convexidad>}$ ,  $\text{<pasa-por-convexidad>}$ ,  $\text{<pasa-por - recta>}$ ,  $\text{<pasa-por-concavidad>}$ ,  $\text{<casi-pasa-por-máxima-concavidad>}$  y  $\text{<pasa-por- máxima-concavidad>}$ . Por esto, las etiquetas que se asignarán a cada objeto, reemplazando a DCP son: PasaPorMáximaConvexidad, PasaCasiPorMáximaConvexidad, PasaPorConvexidad, PasaPorRecta, PasaPorConcavidad, PasaCasiPorMáximaConcavidad, PasaPorMáximaConcavidad.

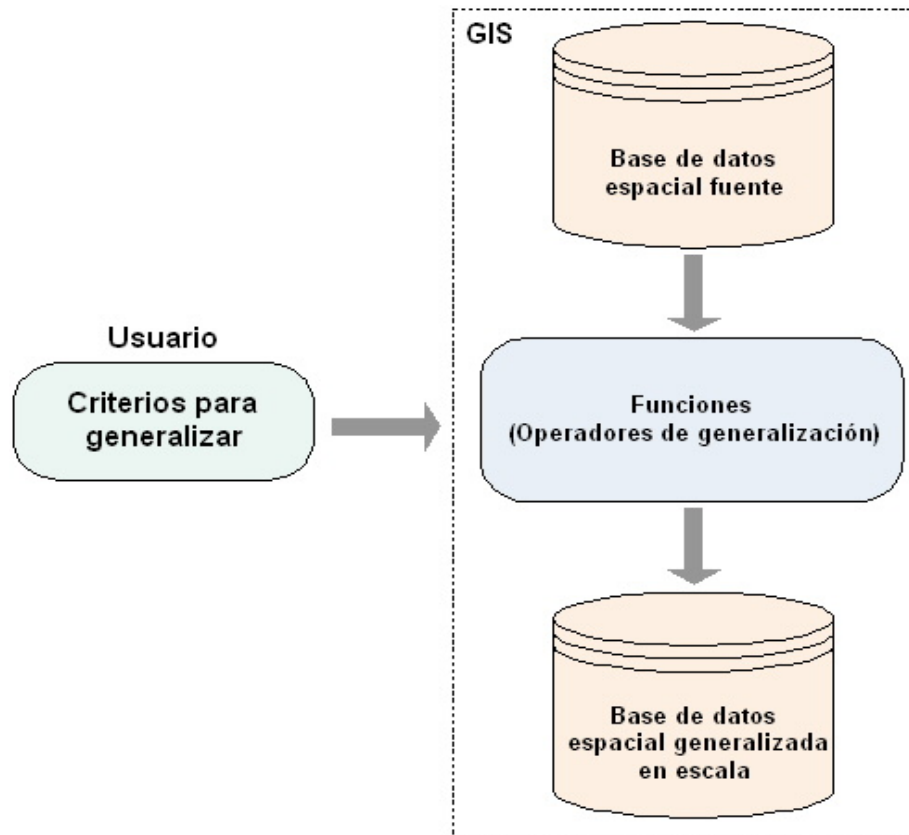
De esta forma, la tabla  $T_R$  de la *BDEPC* que se genera para esta relación, tendrá *descripciones semánticas* de la forma  $T_R \text{ } i\{Río, PasaPor, CurvaDeNivel\}$ . Agregando las propiedades esta tabla queda de la forma  $T_R (Río, PasaPor, CurvaDeNivel, Numrel, Direc)$ ; donde *Numrel*, es el número de relación y *Direc*, indica el número de relación que tiene un río siguiendo el flujo en sentido descendente<sup>64</sup>(Ver Tabla 4.2)

## 4.5 El sistema de generalización

El sistema de generalización que se ha convenido usar es el definido en [Moreno, 2001], [Moreno et al., 2002] y [Moreno y Levachkine, 2005] el cual se basa en el uso de los operadores de generalización propuestos en [McMaster y Shea, 1992]. Este sistema está enfocado a la generalización de redes hidrológicas y curvas de nivel. El sistema se orienta totalmente hacia el enfoque tradicional de los GIS<sup>65</sup>, 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 4.29 muestra un esquema general de generalización automática vectorial para datos topográficos que forma parte de un Sistema de Información Geográfica (GIS).

<sup>64</sup> Significa que siguiendo el flujo del río desde su parte más alta, será *Direc* = 1 la primera relación  $\text{>Pasa-por<}$  entre el río y una curva de nivel; la siguiente relación siguiendo el flujo será *Direc* = 2 , y así sucesivamente.

<sup>65</sup> Se refiere a que como usan métricas sobre las unidades de los datos, se considera un sistema de georeferencia, etc.



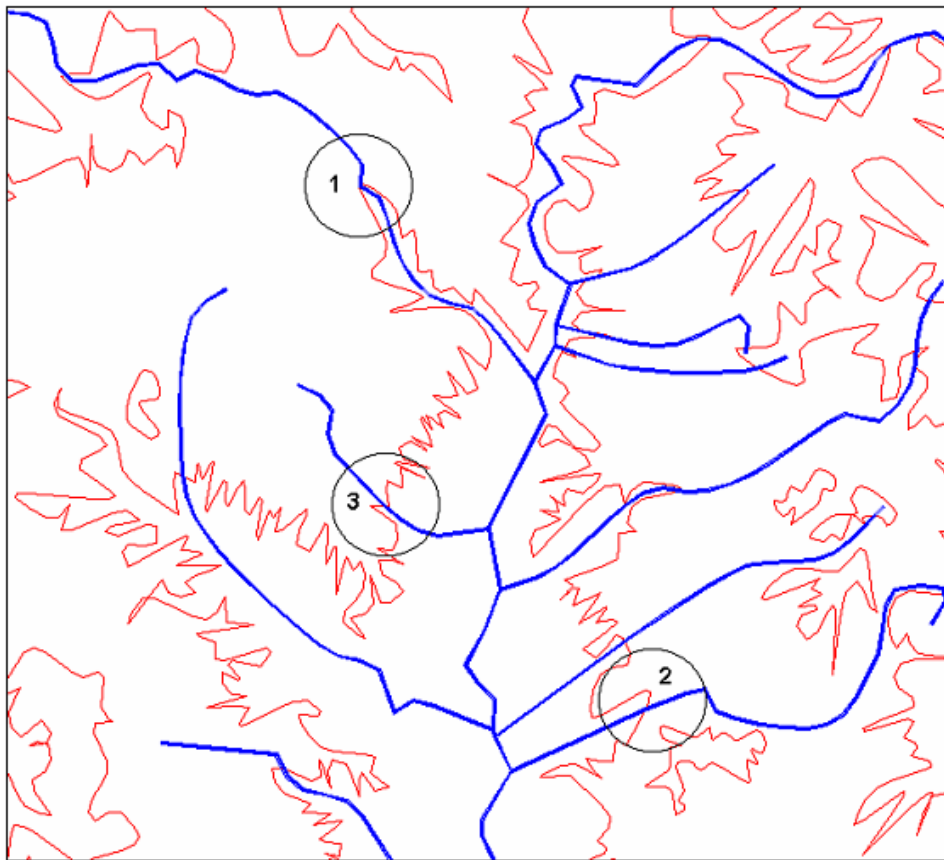
**Figura 4.29** Esquema general del sistema de generalización automática

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. 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 al propósito de la generalización.
- *Base de datos espacial generalizada en escala.* En esta base de datos se almacenan los datos que se generalizaron automáticamente.

Este sistema presenta un funcionamiento aceptable en cuanto a la estructura de la red hidrológica; sin embargo, se generan algunos errores considerables que rompen la

consistencia de la representación. En la Figura 4.30, se presentan tres casos: Caso 1, donde se preserva la relación entre la curva y la red hidrológica; Caso 2, la relación entre la curva y el río no es precisamente la mejor; y Caso 3, donde la relación rompe totalmente la consistencia. En resumen, este sistema realiza la generalización de forma automática; sin embargo genera algunas inconsistencias en la representación de los datos y debido a esto se eligió para emplearlo en la medición de la similitud semántica.



**Figura 4.30** Datos generados por el sistema de generalización automático

El conjunto de datos generalizados que se almacene en la *Base de datos espacial generalizada en escala* será representado también como tablas  $T_R$  en la *BDEPC*, y será utilizado en la siguiente etapa (Verificación).



## 4.6 Verificación

Esta etapa se enfoca a verificar la consistencia de datos generalizados por medio de *descripciones semánticas*. Para esto, los datos que se utilizarán durante la etapa de Verificación son:

- *Descripciones semánticas de los datos fuente*, es decir, las *descripciones semánticas* de los datos sin generalizar, que serán representados por una Tabla de la *BDEPC* denominada  $T_{RF}$ .
- *Descripciones semánticas de los datos generalizados*, es decir, las *descripciones semánticas* de los datos generalizados, que serán representados por una Tabla del *BDEPC* denominada  $T_{RG}$ .

Hay que hacer notar que tanto  $T_{RG}$  como  $T_{RF}$  deben tener la misma estructura es decir  $T_{R1}(O_1, R, O_2)$  y  $T_{R2}(O_1, R, O_2)$ , o bien  $T_{P1}(O_1, R, P_2)$  y  $T_{P2}(O_1, R, P_2)$ ,

Con base en lo anterior, diremos que se medirá la similitud entre las tuplas de  $T_{RG}$  y  $T_{RF}$ . Para esto, haremos uso de la Teoría de la Confusión [Levachkine y Guzman-Arenas, 2004], la cual se basa en una medida asimétrica y dependiente del contexto denominada *confusión* (Ver Sección 3.3.2). Se denomina *confusión* para diferenciarse de otras medidas como cercanía o similitud.

Debido a que la ontología cubre aspectos muy generales y se requiere medir la similitud entre cosas muy específicas, se realiza la medición de la confusión sobre un fragmento de la ontología. Para dicho fragmento se requiere que sea una jerarquía. En los casos que se han definido en la ontología definida a priori, las propiedades y relaciones están en forma de una jerarquía ordenada (por ejemplo, la jerarquía formada por la relación *>Pasa-por<* y sus subclases) (Ver Sección 3.3.1.1), es decir, las clases se encuentran definidas con base en una relación de orden. Por esto, la similitud entre propiedades y relaciones se calculará usando la confusión en jerarquías ordenadas.

El proceso de verificación básicamente consiste en comparar las *descripciones semánticas* de datos fuente ( $T_{RF}$ ) y de datos generalizados ( $T_{RG}$ ). Como resultado de la verificación se obtienen tres posibles resultados: Igual, Equivalente y Desigual. Igual y Equivalente son considerados casos consistentes. Las *descripciones semánticas* que son consistentes se almacenan en tablas de la *Base de Datos Espacial Conceptual Semánticamente Consistente (BDECSC)* (Ver Figura 4.31).

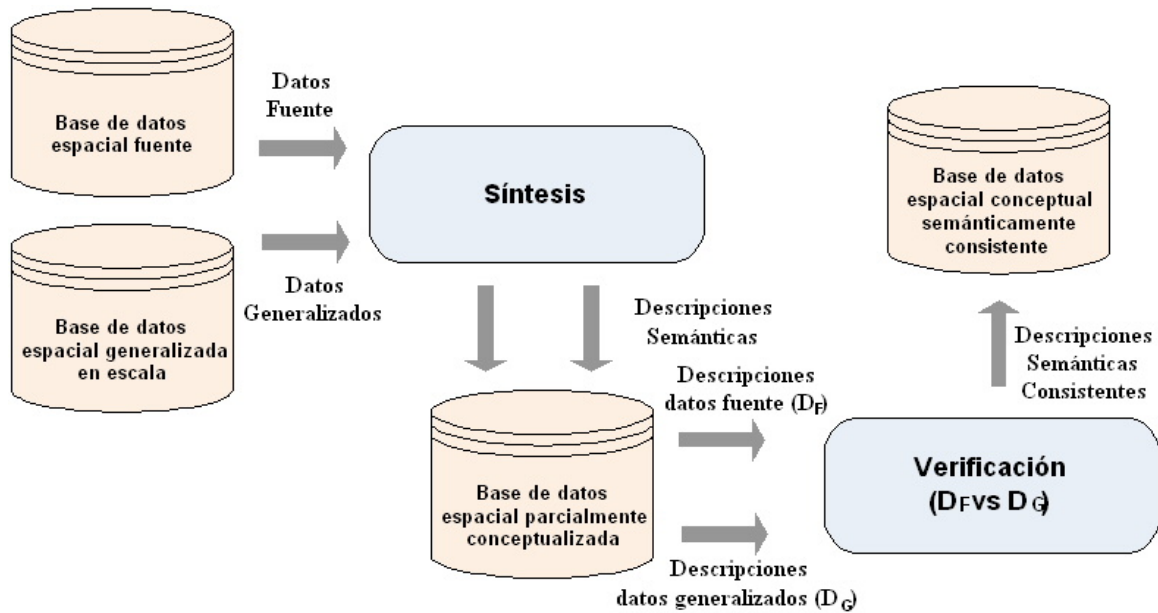


Figura 4.31 Esquema general de la etapa de verificación

Existen dos tipos de tablas que pueden generarse como resultado de la etapa de Verificación.

- *Tabla de Relaciones semánticamente consistentes ( $T_{RC}$ )*, almacena las relaciones consistentes entre objetos geográficos
- *Tabla de Propiedades semánticamente consistentes ( $T_{PC}$ )*, almacena propiedades consistentes de los objetos geográficos.

Una tabla  $T_{RC}$  en la *BDECSC* se define como:  $D_1$  que representa los objetos semánticamente consistentes que se relacionan con otros objetos;  $D_2$  que representa el dominio de todos los conceptos que describen una relación topológica refinada con métricas o una propiedad consistente y  $D_3$  que representa el dominio de los objetos

relacionados con  $D_1$ . Entonces, la Tabla  $T_{RC}$  se puede describir como  $T_{RC}(O_1, R, O_2)$  donde  $O_1 \in D_1, R \in D_2$  y  $O_2 \in D_3$ .

Una tabla  $T_{PC}$  en la *BDECSC* se define como:  $D_1$  que representa los objetos cuyas propiedades son semánticamente consistentes,  $D_2$  que representa las relación entre el objeto y la propiedad (en este caso, la relación *tiene*) y  $D_3$  que representa el conjunto de conceptos que describen una propiedad (en este caso, <Sinuosidad>). Entonces, la Tabla  $T_C$  se puede describir como  $T_{PC}(O_1, R, P_1)$  donde  $O_1 \in D_1, R \in D_2$  y  $P_1 \in D_3$ .

Existe un grupo de propiedades y relaciones que no deben cambiar después de efectuar algún procesamiento (por ejemplo, generalizar), lo que representa que son consistentes después del procesamiento. A este grupo de propiedades se les denomina *invariantes semánticas*. Por tal motivo, las *tuplas* que forman parte de las tablas de la *BDECSC* se les denominará *invariantes semánticas*.

#### 4.6.1 Medición de la confusión entre relaciones

Con el propósito de medir la similitud de una relación entre datos geográficos que pertenecen a dos conjuntos de datos representados mediante *descripciones semánticas*, se presentan dos tablas,  $T_{RF}(O_1, R, O_2)$  y  $T_{RG}(O_1, R, O_2)$ , las cuales se utilizan para verificar la semántica de una relación en datos generalizados; usando la confusión ( $conf_R(r,s)$ , de tener una relación  $r$  en vez de una relación  $s$ ), se define como:

Sea  $Co$  un conjunto de clases ordenadas por una relación de orden  $\{c_i, c_{i+1}, \dots, c_n\}$ , que comparten la misma clase *padre*, los cuales forman una *jerarquía ordenada*  $H$  representados en un vector de clases  $V[c_i, c_{i+1}, \dots, c_n]$ , donde  $i: 1, 2, \dots, n$ ; Siendo  $n$  la *cardinalidad* de  $V[]$ , de manera tal que  $V[1]=c_1$  y  $V[n]=c_n$ . La posición de la clase  $c_i$  en  $V[]$  se define con la función  $Pos(c_i)$  de manera que  $Pos(c_i) = i$  y  $Pos(c_{i+1}) = i+1$  y por lo tanto  $V[i]=c_i$  y  $V[i+1]=c_{i+1}$ , de tal forma que la *distancia entre dos conceptos* representados en  $V[]$  se define por la función  $Dist(c_j, c_k) = |Pos(c_j) - Pos(c_k)|$ , donde  $j$  y  $k \leq n$ ,  $Dist(c_1, c_n) = n$ . Hacer  $s = c_j$  y  $r = c_k$ , así que la  $conf_R(r,s)$  de tener  $r$  en vez de  $s$  en  $Co$ , se define como  $conf(r,s) = (Dist(c_j, c_k)) / (n-1)$ , donde  $c_j$ , representa el valor que tiene una relación después de

realizar un procesamiento a los datos ( $c_j \in T_{RG}$ ) y  $c_k$  representa el valor que tiene una relación antes de realizar un procesamiento ( $c_k \in T_{RF}$ ). Si  $r$  y  $s$  son iguales  $conf_R(r,s) = 0$ .

De esta forma la confusión  $conf_R(r,s)$  estará entre 0 y 1. Con el objetivo de ser más descriptivos, se denominará *distancia semántica* al valor numérico obtenido de la medición de la *confusión*, que representa la similitud entre objetos. El contexto de la  $conf_R(r,s)$  está definido por el padre de las clases ordenadas  $Co$  (por ejemplo, >Pasa-por<).

#### 4.6.2 Medición de la confusión entre propiedades

Suponga que se presentan dos tablas,  $T_{RF}(O_1, R, O_2)$  y  $T_{RG}(O_1, R, O_2)$ , que son representaciones conceptuales de una propiedad de los objetos geográficos de dos capas distintas, las cuales se utilizan para verificar la semántica de una propiedad en datos generalizados, usando la confusión ( $conf_P(r,s)$ ). La confusión de que un objeto tenga una propiedad  $r$  en vez de una propiedad  $s$  se define como:

Sea  $Co$  un conjunto de clases ordenadas por una relación de orden  $\{c_i, c_{i+1}, \dots, c_n\}$ , que comparten la misma clase *padre*, los cuales forman una *jerarquía ordenada*  $H$  representados en un vector de clases  $V[c_i, c_{i+1}, \dots, c_n]$ , donde  $i: 1, 2, \dots, n$ ; Siendo  $n$  la *cardinalidad* de  $V[]$ , de manera tal que  $V[1]=c_1$  y  $V[n]=c_n$ . La posición de la clase  $c_i$  en  $V[]$  se define con la función  $Pos(c_i)$  de manera que  $Pos(c_i) = i$  y  $Pos(c_{i+1}) = i+1$  y por lo tanto  $V[i]=c_i$  y  $V[i+1]=c_{i+1}$ . De tal forma que la *distancia entre dos conceptos* representados en  $V[]$  se define por la función  $Dist(c_j, c_k) = |Pos(c_j) - Pos(c_k)|$ , donde  $j$  y  $k \leq n$ ,  $Dist(c_1, c_n) = n-1$ . Hacer  $s = c_j$  y  $r = c_k$ , así que la  $conf_P(r,s)$  de tener  $r$  en vez de  $s$  en  $Co$ , se define como  $conf_P(r,s) = (Dist(c_j, c_k)) / (n-1)$ , donde  $c_j$ , representa el valor que tiene una propiedad después de realizar un procesamiento a los datos ( $c_j \in T_{PG}$ ) y  $c_k$  representa el valor que tiene una propiedad antes de realizar un procesamiento ( $c_k \in T_{PF}$ ). Si  $r$  y  $s$  son iguales  $conf_P(r,s) = 0$ .

De esta forma la confusión  $conf_P(r,s)$  estará entre 0 y 1. Con base en la *distancia semántica*, se definen tres posibles casos que se pueden obtener como resultado de la medición de la  $conf(r,s)$ . Al igual que el caso anterior, el contexto de la  $conf_P(r,s)$  esta definido por el padre de las clases ordenadas  $Co$  (por ejemplo, <Sinuosidad>).

### 4.6.3 Casos de verificación de (in)consistencias

Con base en  $conf(r,s)$ , se definen tres casos de verificación de inconsistencias: Igual, Desigual y Equivalente, los cuales se detallan en las siguientes subsecciones. En esta sección se utilizará la tabla  $T_F$  para hacer referencia a las descripciones de datos fuente y la tabla  $T_G$  para hacer referencia a las descripciones de datos generalizados. Tanto  $T_F$  como  $T_G$  se refieren a *tablas de relaciones conceptualmente representadas* ( $T_R$ ).

#### 4.6.3.1 Igual

En este caso la *distancia semántica* = 0 ( $conf(r,s)=0$ ) . Este caso representa que ambos conceptos (relación y/o propiedad) son exactamente iguales. Debido a esta igualdad, se dice que es una invariante semántica y por lo tanto la relación o propiedad es consistente. Dicho de otra forma, es el mismo valor en la propiedad o relación representada en  $T_F$  y  $T_G$ . Este caso se representa por el símbolo =.

Considere la Figura 4.32, donde se puede ver la representación gráfica y *descripciones semánticas* de la relación entre dos objetos geográficos antes y después de generalizar. Las *descripciones semánticas* tienen la forma  $T_{R_i}\{Objeto_1, R, Objeto_2\}$ , por tratarse de una relación (Ver sección 4.4). En este caso, se requiere verificar la relación  $R$  entre las tuplas  $T_{F_i}\{B, esta\ al\ W, A\}$  y  $T_{G_i}\{B, está\ al\ W, A\}$ , donde  $R$  es igual a *esta al W* para ambos casos. Tomando como referencia a la jerarquía ordenada de la Figura 4.33, tenemos que  $r=está\ al\ W$  y  $s=está\ al\ W$ , así que  $conf_R(r,s)=0$ . Entonces,  $T_{F_i} = T_{G_i}$ . Esto significa que la semántica de la relación fue preservada.

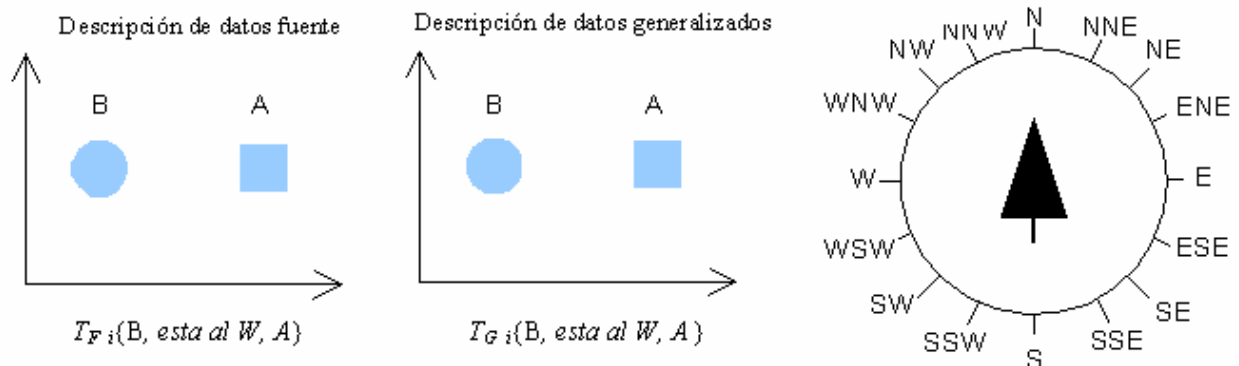


Figura 4.32 Representación gráfica de “igual” para la relación “está al W”

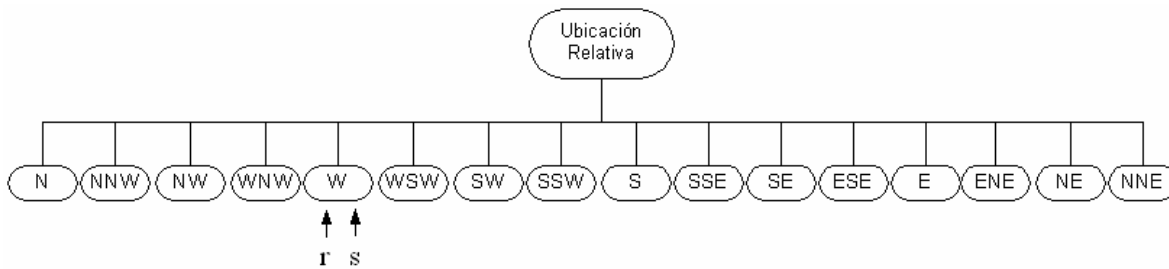


Figura 4.33 Medición de  $conf(r,s)=0$ , sobre una jerarquía de *ubicación relativa*<sup>66</sup>

### 4.6.3.2 Desigual

*Desigual* significa  $0 < distancia\ semántica \leq 1$  ( $conf(r,s) > 0$ ), los conceptos (relaciones o propiedades) son diferentes en  $T_F$  y  $T_G$ . En este caso, la relación y/o propiedad es inconsistente, la semántica no fue preservada y no representa una *invariante semántica*. Este caso se representa por el símbolo  $\neq$ .

Considere la Figura 4.34, donde de forma similar al ejemplo del caso anterior se presentan las *descripciones semánticas*  $T_{F_i}\{B, está\ al\ W, A\}$  y  $T_{G_i}\{B, está\ al\ S, A\}$ , por lo que para verificar la relación usando confusión tenemos,  $r = está\ al\ S$  y  $s = está\ al\ W$  (Ver Figura 4.35). Así, la  $conf_R(r,s) = (Dist(c_j,c_k))/n-1 = (|5-9|)/(16-1) = 4/16$ . Recordar que este contexto  $conf_R(r,s)$  mide la confusión de tener una relación  $r$  en vez de una relación  $s$ . Entonces,  $T_{F_i}\{B, está\ al\ W, A\} \neq T_{G_i}\{B, está\ al\ S, A\}$ . Esto significa que la semántica de la relación no fue preservada, por lo tanto es inconsistente.

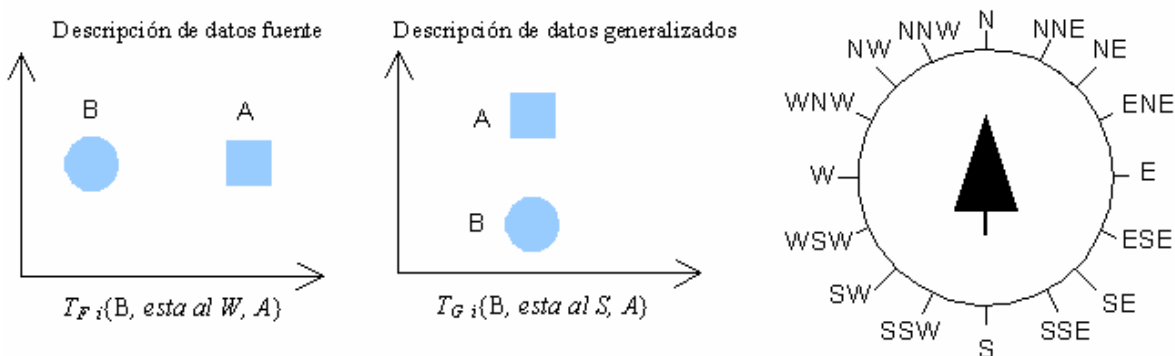


Figura 4.34 Representación gráfica de “desigual”

<sup>66</sup> La jerarquía presentada se utiliza únicamente como ejemplo. Y no puede aplicarse tal y como aparece en casos prácticos

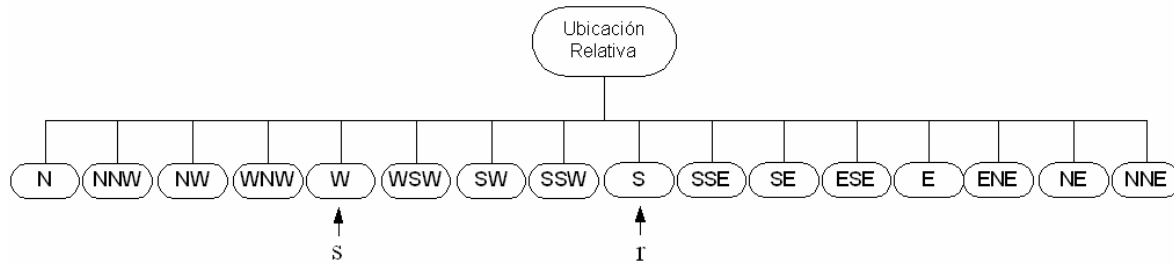


Figura 4.35 Medición de  $conf(r,s) = 4/16$ , sobre una jerarquía de la *ubicación relativa*

### 4.6.3.3 Equivalente

El último caso es *Equivalente*. Éste se considera, debido a que existen algunas aplicaciones que requieren de un cierto margen de tolerancia para definir si una relación o propiedad puede considerarse como consistente. Debido a esa razón, hemos considerado definir la relación de equivalencia con base en un umbral  $u$ . El umbral  $u$ , define un intervalo a partir del cual se definen las *distancias semánticas* que representan *equivalencias*. Así que  $u$  debe ser un valor  $0 < u < 1$ . Entonces, cualquier *distancia semántica*  $\leq u$ , se considera como *equivalente* y representa una *invariante semántica*. Se representa por el símbolo  $\cong$ . Por otro lado, si la *distancia semántica*  $> u$ , se considera como *desigual*.

Considere la Figura 4.36, donde aparecen  $T_{Fi}\{B, \text{está al } W, A\}$  y  $T_{Gi}\{B, \text{está al } WNW, A\}$ . Por lo que,  $r = \text{está al } S$  y  $s = \text{está al } W$ . Considerando la Figura 4.37, la  $conf_R(r,s) = (15 - 4) / (16 - 1) = 1/15$ . Entonces, si  $u = 1/15$ , esta relación se considera como equivalente ( $T_{Fi}\{B, \text{está al } W, A\} \cong T_{Gi}\{B, \text{está al } WNW, A\}$ ). Si bien,  $r$  y  $s$  son distintos para este caso, el uso de un umbral permite definir si se considera consistente o no.

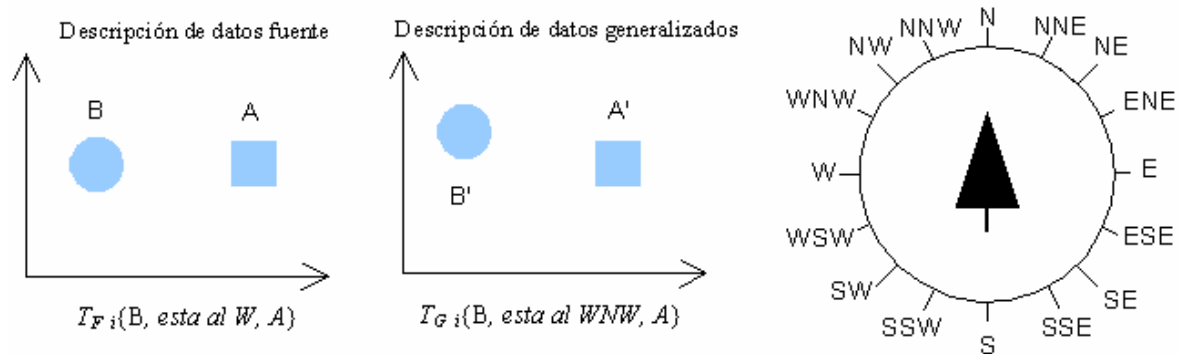
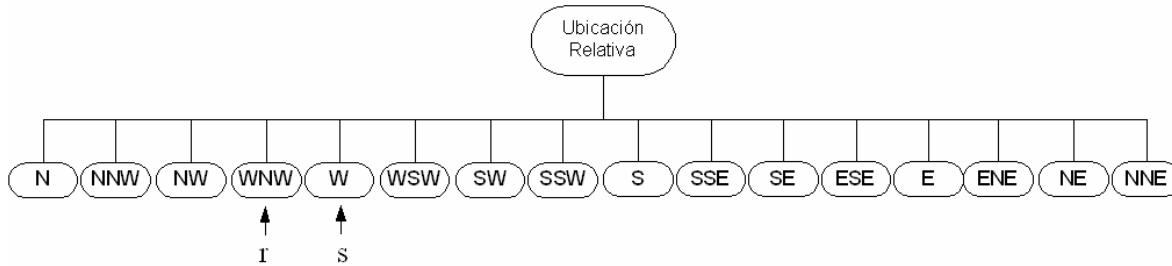


Figura 4.36 Representación gráfica de “equivalente”



**Figura 4.37** Medición de  $conf(r,s) = 1/15$ , con  $u=1/15$ , sobre una jerarquía de la ubicación relativa

Sin embargo, se debe tener cuidado al elegir el tamaño de  $u$ , debido a que pudieran cometerse errores considerables al definir como consistentes algunas relaciones que no tienen sentido. Por ejemplo, considerar que para la jerarquía de la Figura 4.34, se definió un  $u = 8/16$ ; en ese caso obtendremos que N es equivalente a S, lo que representa un error muy grave, si consideramos que se trata de un cambio después de un procesamiento. Aceptar esto, semánticamente representaría a decir que da lo mismo que el objeto esté al norte que al sur.

La definición de un  $u$  depende en gran parte del contexto. Consideremos el caso de la jerarquía de ubicación relativa, aplicado a puntos que se usan para definir la traza urbana por medio de un levantamiento topográfico. Si consideramos que esta aplicación debe tener la mayor precisión posible, por ejemplo, al definir un error de  $1/16$ , podría originar graves errores en la aplicación, debido a que se requiere que las relaciones sean iguales.

Adicionalmente, el valor de  $u$  no depende de la granularidad<sup>67</sup>. Pudiera pensarse que el valor de  $u$  depende de la *granularidad* si consideramos 2 jerarquías del mismo tipo. Consideremos una con *granularidad* pequeña (por ejemplo, una jerarquía ordenada  $H \propto \{N, W, E, N\}$ ) con  $u = 1/4$ , por otro lado, considerar una jerarquía con  $u = 1/4$  sobre una jerarquía con *granularidad* grande (por ejemplo, una jerarquía como la de la figura 4.34), podría suceder que con base en la *granularidad* se tuviera la creencia de que se pueden tener distintas clases de consistencia, aunque en realidad es exactamente lo mismo (la distancia semántica entre N y E sobre las jerarquía  $H$  y la presentada en la Figura 4.34 son exactamente iguales).

<sup>67</sup> En este caso en número de clases hermanas.



#### 4.6.4 Verificación de la consistencia entre tablas de propiedades conceptualmente representadas

Una vez que se ha definido cómo se mide la confusión para *tablas de propiedades conceptualmente representadas* y los casos posibles de la verificación de consistencia, pasaremos a describir la forma en la que se identifica la consistencia de propiedades mediante representaciones conceptuales.

Se utilizarán dos tablas de *propiedades conceptualmente representadas*  $T_{PF}$  (*Objeto, Relación, Propiedad*) y  $T_{PG}$  (*Objeto, Relación, Propiedad*). Recordar que la confusión para este caso, se define como  $conf_P(r, s)$ , donde  $r$  es *Propiedad* de  $T_{PG}$  y  $s$  es *Propiedad* de  $T_{PF}$ . Y se lee como “la confusión de que un objeto tenga una propiedad  $r$  en vez de una propiedad  $s$ ”.

En general, el procedimiento para medir la consistencia entre representaciones conceptuales de propiedades se realiza por medio de una comparación de las descripciones semánticas de de  $T_{PG}$  con las de  $T_{PF}$ . Para medir la consistencia entre representaciones conceptuales de propiedades se debe seguir el siguiente procedimiento:

```

Definir  $i := Cardinalidad [T_{PG}]$  que representa el número de tuplas de  $T_{PG}$ ,
Definir umbral  $u$ 
Definir Cardinalidad del conjunto de clases ordenadas  $Co$ , que definen
la Propiedad
Definir el vector de clases  $V [c_i, c_{i+1}, \dots, c_n]$ 
Definir una tabla  $T_{PC}$  (tabla de propiedades semánticamente
consistentes) en la BDESC.
 $O_k$  es una variable que almacena el identificador de un objeto
 $r$  y  $s$  son variables que almacenan los valores para calcular la
similitud
Repetir desde 0 hasta  $i$ 
  Seleccionar una tupla  $T_{PG} i \{Objeto, Relación, Propiedad\}$ 
    Hacer  $r := Propiedad$  de  $T_{PG} i$ 
    Hacer  $O_k := Objeto$  de  $T_{PG} i$ 
  Seleccionar una tupla  $T_{PF} i \{Objeto, Relación, Propiedad\}$  donde
 $O_k = Objeto$ 

```

Si *Cardinalidad [selección sobre  $T_{PF} i$ ]=1*  
 Hacer *s:= Propiedad de  $T_{PF} i$*   
 Calcular *conf<sub>P</sub> (r, s)*  
 Si *conf<sub>P</sub> (r, s) = 0* entonces  
 $T_{PF} i = T_{PG} i$   
 copiar la  $T_{PG} i\{Objeto, Relación, Propiedad\}$  en  $T_c$   
 Si *conf<sub>P</sub> (r, s) < u* entonces  
 $T_{PF} i \cong T_{PG} i$   
 copiar la  $T_{PG} i\{Objeto, Relación, Propiedad\}$  en  $T_c$   
 Si *conf<sub>P</sub> (r, s) > u* entonces  
 $T_{PF} i \neq T_{PG} i$   
 identificar la causa de la inconsistencia, ésta depende del caso de estudio

#### 4.6.5 Verificación de la consistencia entre tablas de relaciones conceptualmente representadas

Una vez que se ha definido cómo se mide la confusión para *tablas de relaciones conceptualmente representadas* y los casos posibles de la verificación de consistencia, pasaremos a describir la forma en la que se identifica la consistencia de una relación mediante representaciones conceptuales.

La consistencia la evaluaremos en dos sentidos, el primero, identificar que las relaciones existentes en los datos generalizados no estén duplicadas; a este tipo de verificación de consistencia le llamaremos consistencia topológica. Por otro lado, identificaremos la consistencia considerando aspectos geométricos, para lo que utilizaremos una medición de confusión.

Se utilizarán dos tablas de *relaciones conceptualmente representadas*  $T_{RF}$  (*Objeto1, Relación, Objeto2*) y  $T_{RG}$  (*Objeto1, Relación, Objeto2*). Recordar que la confusión para este caso se define como  $conf_R(r, s)$ , donde  $r$  es *Relación* de  $T_{PG}$  y  $s$  es *Relación* de  $T_{PF}$ . Y se lee como “la confusión de que dos objetos estén relacionados por la relación  $r$  en vez de la relación  $s$ ”.

En general, el procedimiento para medir la consistencia entre representaciones conceptuales de relaciones se realiza por medio de una comparación de las descripciones semánticas de  $T_{RG}$  con las de  $T_{RF}$ . Para medir la consistencia entre representaciones conceptuales de relaciones se debe seguir el siguiente procedimiento:

```

Definir  $i := \text{Cardinalidad } [T_{RG}]$  que representa el número de tuplas de  $T_{RG}$ 
Definir umbral  $u$ 
Definir Cardinalidad del conjunto de clases ordenadas  $Co$ , que definen
la Relación
Definir el vector de clases  $V [c_i, c_{i+1}, \dots, c_n]$ 
Definir una tabla  $T_{RC}$  (tabla de relaciones semánticamente consistentes)
en la BDESC.
 $O_k, O_l$  son variables que almacenan el identificador de un objeto
Repetir desde 0 hasta  $i$ 
  Seleccionar una tupla  $T_{RG} i \{Objeto1, Relación, Objeto2\}$ 
  Hacer  $r := Relación$  de  $T_{RG} i$ 
  Hacer  $O_k := Objeto1$  de  $T_{RG} i$ 
  Hacer  $O_l := Objeto2$  de  $T_{RG} i$ 
  Seleccionar tuplas  $T_{RF} i \{Objeto1, Relación, Objeto2\}$  donde
     $O_k = Objeto1 \wedge O_l = Objeto2$ 
    Si Cardinalidad [selección sobre  $T_{RF} i$ ] = 0
      Significa que la relación que está en  $T_{RG} i$  no está en
       $T_{RF} i$ , entonces es una inconsistencia
    Si Cardinalidad [selección sobre  $T_{RF} i$ ] = 1
      Hacer  $s := Relación$  de  $T_{RF} i$ 
      Calcular  $conf_R (r, s)$ 
    Si Cardinalidad [selección sobre  $T_{RF} i$ ] > 1
      Usar atributo para identificar la tupla topológicamente
      consistente
        A la relación topológicamente consistente
          Hacer  $s := Relación$  de  $T_{RF} i$ 
          Calcular  $conf_R (r, s)$ 
          Las tuplas adicionales son incorrectas
          identificar la causa de que las relaciones
          son inconsistentes.
    Si  $conf_R (r, s) = 0$  entonces
       $T_{RF} i = T_{RG} i$ 
      Copiar la  $T_{RG} i \{Objeto1, Relación, Objeto2\}$  en  $T_c$ 

```

Si  $conf_R(r, s) < u$  entonces  
 $T_{RF} i \cong T_{RG} i$   
copiar la  $T_{RG} i \{Objeto1, Relación, Objeto2\}$  en  $T_C$   
Si  $conf_R(r, s) > u$  entonces  
 $T_{RF} i \neq T_{RG} i$   
identificar la causa de la inconsistencia, esto depende del caso de estudio

#### 4.6.6 Medición de la similitud de la Sinuosidad entre dos representaciones conceptuales.

La Sinuosidad es una propiedad, por lo que por analogía con la Sección 4.4.1, el procedimiento a seguir es el siguiente.

Definir  $i := Cardinalidad[T_B]$  que representa el número de tuplas de  $T_B$ ,  
Definir umbral  $u = 1$   
Definir Cardinalidad del conjunto de clases ordenadas  $Co=5$   
Definir el vector de clases  $V[c_i, c_{i+1}, \dots, c_n] = V[NadaSinuoso, PocoSinuoso, Sinuoso, MuySinuoso, MuySinuoso]$   
Definir una tabla  $T_C$  en la BDESC.  
 $O_k$ , es una variable que almacena el identificador de un objeto  
 $r$  y  $s$  son variables que almacenan los valores para calcular la similitud  
Repetir desde 0 hasta  $i$   
    Seleccionar una tupla  $T_{Bi} \{O_{rio}, Tiene, Sinuosidad\}$   
        Hacer  $r := Sinuosidad \in T_{Bi}$   
        Hacer  $O_k := O_{rio} \ i \in T_{Bi}$   
    Seleccionar una tupla  $T_{BAi} \{O_{rio}, Tiene, Sinuosidad\}$  donde  $O_k = O_{rio} \ i$   
        Si  $Cardinalidad [selección sobre T_{Ai}] = 1$   
            Hacer  $s := Sinuosidad \in T_{Ai}$   
            Calcular  $conf_{AP}(r, s)$   
            Si  $conf_A(r, s) = 0$  copiar la  $T_{Bi} \{O_{rio}, Tiene, Sinuosidad\}$   
            en  $T_C$

Si  $conf_A(r,s) < u$  copiar la  $T_{Bi}$   $\{O_{rio}, Tiene, Sinuosidad\}$   
 en  $T_C$   
 Si  $conf_A(r,s) > u$  el error es originado por la  
 simplificación en la línea, se recomienda modificar  
 criterio

#### 4.6.7 Verificación de la consistencia de la relación *Pasa-por* representada conceptualmente.

La verificación de consistencia de la relación  $\langle Pasa-por \rangle$ , usando una representación conceptual se define como:

$T_{RFP}$  representa la tabla que contiene las relaciones  $\langle Pasa-por \rangle$  antes de generalizar

$T_{RGP}$  representa la tabla que contiene las relaciones  $\langle Pasa-por \rangle$  generalizadas

$i := Cardinalidad[T_{RGP}] = 7$ , es el número de tuplas de  $T_{RG}$

$u = 1/7$ , se propone un umbral de dos para este caso

$Cardinalidad$  de  $Co$ , es igual a 7, es decir el número de subclases de  $\langle Pasa-Por \rangle$

Vector de clases  $V[c_i, c_{i+1}, \dots, c_n] = V[PasaPorMáximaConvexidad, PasaCasiPorMáximaConvexidad, PasaPorConvexidad, PasaPorRecta, PasaPorConcavidad, PasaCasiPorMáximaConcavidad, PasaPorMáximaConcavidad]$

Tabla  $T_{RCP}$  (tabla de relaciones semánticamente consistentes de las relaciones  $\langle Pasa-por \rangle$ ) en la *BDESC*.

$O_k, O_l$  son variables que almacenan el identificador de un objeto

Repetir desde 0 hasta  $i$

    Seleccionar una tupla  $T_{RGP}$   $i\{Río, PasaPor, Curva, Direc, Numrel\}$

        Hacer  $r := PasaPor$  de  $T_{RGP}$   $i$

        Hacer  $O_k := Río$  de  $T_{RGP}$   $i$

        Hacer  $O_l := Curva$  de  $T_{RGP}$   $i$

    Seleccionar tuplas  $T_{RFP}$   $i\{Río, PasaPor, Curva\}$  donde  $O_k = Río \wedge O_l = Curva$

        Si  $Cardinalidad$  [selección sobre  $T_{RF}$   $i$ ] = 0

            Significa que la relación que está en  $T_{RGP}$   $i$  no está en  $T_{RFP}$   $i$ , entonces es una inconsistencia

Si *Cardinalidad [selección sobre  $T_{RFP} i$ ] = 1*  
 Hacer  $s := PasaPor$  de  $T_{RFP} i$   
 Calcular  $conf_R (r, s)$

Si *Cardinalidad [selección sobre  $T_{RFP} i$ ] > 1*  
 Usar *identificador de relación y atributo Direc*, para  
 identificar la *topológicamente consistente (Direc = 1,*  
*es topológicamente consistente)*  
 A la relación *topológicamente consistente*

Hacer  $s = PasaPor$  de  $T_{RFP} i$   
 Calcular  $conf_R (r, s)$   
 Seleccionar  
*Las tuplas adicionales son generadas por errores en la*  
*simplificación de líneas durante la generalización, se*  
*recomienda modificar el criterio de los objetos*  
*involucrados.*

Si  $conf_R (r, s) = 0$  entonces  
 $T_{RFP} i = T_{RGP} i$   
 copiar la  $T_{RGP} i \{Objeto1, Relación, Objeto2\}$  en  $T_{CP}$

Si  $conf_R (r, s) < u$  entonces  
 $T_{RFP} i \cong T_{RGP} i$   
 copiar la  $T_{RG} i \{Objeto1, Relación, Objeto2\}$  en  $T_{CP}$

Si  $conf_R (r, s) > u$  entonces  
 $T_{RFP} i \neq T_{RGP} i$   
*el error es originado por la simplificación en la*  
*línea, se recomienda modificar criterio*

#### 4.6.8 Medición de la similitud a nivel global

Como se presentó en la Sección 4.6.3, se ha considerado medir las inconsistencias que existen en cada una de las relaciones de los objetos río y curva de nivel. Dichas mediciones representan la similitud que existe entre una relación existente en los datos fuente con la misma relación en los datos generalizados. Esto representa una medición a nivel local. Con el objeto de medir la similitud a diferentes niveles de detalle, se desarrollaron un conjunto de medidas que se utilizan para evaluar la similitud a nivel global, es decir, considerando el conjunto de relaciones que poseen las capas temáticas involucradas.

#### 4.6.8.1 Error acumulado

El error acumulado ( $e_a$ ) se representa por la sumatoria de todas las confusiones obtenidas al comparar la descripción conceptual de los datos fuente con la descripción conceptual de los datos generalizados. La siguiente ecuación describe el error acumulado:

$$e_a = \sum_{i=1}^n \text{conf}_R(r_i, s_i).$$

Donde:

$n$ , es el número de relaciones a evaluar,

$\text{conf}_R(r_i, s_i)$ , es la confusión para la relación  $i$ .

El error acumulado se interpreta como sigue: Si  $e_a = 0$ , representa que las relaciones existentes en los datos fuente son exactamente iguales a las relaciones en los datos generalizados. De otra forma, entre más grande sea el valor de  $e_a$ , entonces los datos fuente y los datos generalizados son más disímiles. Ahora, consideremos dos casos; 1) se tienen 10 relaciones, de las cuales todas tienen  $\text{conf}_R(r, s) = 1/7$ , y supongamos también que dicho valor representa que las relaciones son equivalentes, es decir, todas las relaciones son *semánticamente consistentes*. En este caso  $e_a = 10/7$ . 2) considerar 10 relaciones, de las cuales 5 tienen  $\text{conf}_R(r, s) = 2/7$ , y supongamos que dichas relaciones se evalúan como desiguales, lo que representa que la mitad de ellas son *semánticamente inconsistentes*. En este caso  $e_a = 10/7$ . Si bien los dos casos, tienen  $e_a = 10/7$ , el primer caso es semánticamente consistente, mientras que en el segundo caso la mitad de las relaciones son inconsistentes. Entonces, podemos decir que no es posible caracterizar con facilidad los valores de  $e_a$ . Debido a lo anterior, se generó una medición denominada *error medio*.

#### 4.6.8.2 Error Medio

El error medio ( $e_m$ ), representa el promedio de las confusiones obtenidas durante el proceso de verificación de la consistencia. El error medio se describe matemáticamente como:

$$e_m = \frac{\sum_{i=1}^n \text{conf}_R(r_i, s_i)}{n} = \frac{e_a}{n}$$

Donde:

$n$ , es el número de relaciones a evaluar

$conf_R(r_i, s_i)$ , es la confusión para la relación  $i$

A diferencia de  $e_a$ , este error puede ser fácilmente caracterizado e interpretado, debido a que todos los valores posibles de  $e_m$  están entre 0 y 1. La interpretación del error medio es como sigue: Si  $e_m = 0$ , representa que las relaciones existentes en los datos fuente son exactamente iguales a las relaciones en los datos generalizados y por lo tanto que los datos son consistentes. Si  $e_m = 1$ , representa que las relaciones existentes en los datos fuente son totalmente desiguales a las relaciones en los datos generalizados, lo que significa que los datos son inconsistentes. Entonces valores de  $e_m$  cercanos a cero, se interpretan como consistentes y valores de  $e_m$  cercanos a uno, se interpretan como inconsistentes.

Utilizando el *error medio*, podemos definir conceptos que describan la similitud entre los datos fuente y los datos generalizados. Los valores numéricos obtenidos en la medición de  $e_m$  se interpretan como sigue:

- Si  $e_m = 0$ , los datos generalizados son *idénticos* a los datos fuente.
- Si  $0.01 \leq e_m \leq 0.10$ , los datos generalizados son *muy similares* a los datos fuente.
- Si  $0.11 \leq e_m \leq 0.35$ , los datos generalizados son *similares* a los datos fuente.
- Si  $0.36 \leq e_m \leq 0.75$ , los datos generalizados son *algo similares* a los datos fuente.
- Si  $0.76 \leq e_m \leq 1.0$ , los datos generalizados son *nada similares* a los datos fuente.

Los valores para identificar la similitud entre los datos fuente y datos generalizados fueron obtenidos por un análisis estadístico sobre diversos conjuntos de datos. Aunque los valores son susceptibles a ajustarse en dependencia de los requerimientos del problema. Adicionalmente, se ha considerado utilizar los valores del error medio para identificar que grupos de usuarios pueden utilizar satisfactoriamente los datos.

#### **4.6.8.3 La medición de la similitud y su relación con las preferencias de los usuarios**

Los usuarios de datos geográficos tienen muy diversas necesidades. Consideramos que la medición de la similitud entre dos conjuntos de datos puede ser útil para identificar si los



datos generalizados cumplen con los requerimientos de un grupo de usuarios. Hemos dividido a los usuarios de datos geográficos en tres categorías, que son:

- **Usuario Experto:** Este grupo incluye a hidrólogos, cartógrafos, topógrafos, geógrafos y a todos aquellos usuarios que requieran que los datos geográficos que utilicen sean consistentes, tanto en el aspecto geométrico como en el topológico. Esto se debe a que realizan ciertos análisis sobre los datos geográficos por lo que requieren de datos precisos. Tienen un amplio dominio teórico-metodológico sobre el área de su interés. Este grupo de usuarios incluye a los diseñadores de Sistemas de Información Geográfica, enfocados a realizar análisis topológico y geométrico.
- **Usuario Intermedio:** Se refiere a aquellos usuarios que utilizan la información geográfica, pero que no requieren de gran precisión en los datos. Generalmente, en este grupo se encontrarán los diseñadores de Sistemas de Información Geográfica, que desarrollan aplicaciones donde no se requiere garantizar la consistencia en los datos. En particular nos referimos a las aplicaciones enfocadas a la visualización y algunos análisis geométricos.
- **Usuario Básico:** En este grupo se encuentran los usuarios que utilizan los datos simplemente como referencia, sin considerar en detalle los aspectos de consistencia geométrica ni topológica. En este grupo se encuentran los usuarios que no realizan procesamiento sobre los datos y simplemente los utilizan para visualizaciones y referencia.

Considerando los tipos de usuarios, proponemos que según los requerimientos de un *usuario experto*, éste debe utilizar los datos que fueron evaluados como *idénticos* al utilizar el *error medio* ( $e_m$ ). Asimismo, un *usuario intermedio* puede utilizar datos en los que los  $e_m$  fueron tipificados como *idénticos* o *muy similares*. Por último, el usuario básico no tiene problema en utilizar datos *idénticos*, *muy similares* y *similares* (Ver Tabla 4.6).

**Tabla 4.6.** El error medio y su relación con los requerimientos de diversos usuarios

		error medio ( $e_m$ )				
		<i>idénticos</i>	<i>muy similares</i>	<i>similares</i>	<i>algo similares</i>	<i>nada similares</i>
<b>Tipos de Usuarios</b>	<i>Usuario Experto</i>					
	<i>Usuario Intermedio</i>					
	<i>Usuario Básico</i>					

Como se ha visto a lo largo de este capítulo, el método propuesto está sustentado en representaciones conceptuales. Dichas representaciones son generadas a partir de relaciones y propiedades que se definieron en una conceptualización del entorno geográfico. A partir de métodos tradicionales, mediante las mediciones sobre datos geográficos, se extrajeron tanto propiedades y relaciones implícitamente representados por conjuntos de datos vectoriales. Con base en las mediciones, se genera una representación conceptual que es cualitativa y explícita, donde un concepto o clase definido en la conceptualización representa una propiedad o relación que se almacena en una base de datos.

La representación conceptual se utiliza para identificar la similitud semántica entre datos fuente y datos procesados. La similitud se mide en dos niveles: local y global. A nivel local se mide relación por relación usando la confusión y la media para definir tres resultados: Igual, Desigual y Equivalente. Por otro lado, a nivel global, se consideran las mediciones de confusión de todas las relaciones existentes, donde utilizando el error medio se puede definir qué tan similares son los datos generalizados a los datos fuente. Se generan conceptos para expresar si los datos generalizados son *idénticos*, *muy similares*, *similares*, *algo similares* o *nada similares* a los datos fuente. Esto sirve para identificar si los datos cumplen los requerimientos de ciertos tipos de usuarios.

# CAPÍTULO 5

## PRUEBAS Y RESULTADOS

En este capítulo se presentan los resultados obtenidos con la implantación de la metodología propuesta, aplicados al caso de estudio de la generalización de redes hidrológicas. Primeramente se presentarán algunos resultados originados por el sistema de generalización y posteriormente serán mostrados los resultados obtenidos de la representación conceptual que involucra la medición de similitud. El sistema de generalización propuesto fue implantado en el SIG ARC/Info Desktop V. 8.1.2, usando codificación en AML (Arc Macro Language; ver Anexo I).

Los datos utilizados son provenientes de las cartas topográficas a escala 1:50,000 del INEGI. Dichos datos fueron vectorizados directamente por INEGI, con proyección Cónica Conforme de Lambert. Originalmente se encontraban en formato *shapefile*<sup>68</sup> y fueron transformados a *coberturas* de ARC/Info, para poder emplearse en los algoritmos desarrollados. Adicionalmente, los datos tuvieron que editarse de forma manual para corregir algunos errores topológicos y geométricos, con el fin de cubrir los requisitos de los algoritmos.

Se eligió ARC/Info debido a la versatilidad que posee para la manipular de datos geográficos. La región que se presentará como caso de estudio fue elegida aleatoriamente: en el caso de los datos que se presentan, corresponden al estado de Puebla.

---

<sup>68</sup> El formato ESRI Shapefile (SHP) es un formato propietario abierto de datos geográficos desarrollado por la compañía ESRI, quien desarrolla y comercializa software para Sistemas de Información Geográfica. Originalmente se creó para la utilización con ArcView GIS, pero actualmente se ha convertido en formato estándar por la importancia que los productos ESRI tienen en el mercado SIG.

## 5.1 El sistema de Generalización

El sistema de generalización se describió en la Sección 4.5 y está basado en los operadores de generalización definidos en [McMaster y Shea, 1992]. El sistema tiene como fin generar datos en la escala 1:250,000. Los resultados obtenidos de la implantación de la metodología se presentarán en esta sección.

Cabe señalar que algunos de los puntos indicados en la etapa de Normalización para los datos fuente se realizan durante las clasificaciones preliminares a la generalización.

### 5.1.1 Generalización de las Redes Hidrológicas

ARC/Info utiliza un formato de archivo denominado cobertura<sup>69</sup>; a las capas de datos geográficos en esta sección se les denominará coberturas o capas, donde cada capa de ríos se compone por los arcos que conforman la red hidrológica. La base de datos que posee cada cobertura que almacena los datos fuente posee la siguiente estructura:

*Rios.AAT (FNODE#, TNODE#, LPOLY#, RPOLY#, LENGTH, RIOS-#, RÍOS-ID, ID, CLAVE)*

Donde:

<i>FNODE#</i>	Número de nodo inicial de cada arco,
<i>TNODE#</i>	Número de nodo final de cada arco,
<i>LPOLY#</i>	Número interno del polígono de la izquierda,
<i>RPOLY#</i>	Número interno del polígono de la derecha,
<i>LENGTH</i>	Longitud de cada arco, medido en unidades de la cobertura,
<i>RIOS#</i>	El número de registro,
<i>RIOS-ID</i>	Identificador del usuario,
<i>ID</i>	Identificador de cada arco, definido por INEGI,
<i>CLAVE</i>	Identificador de la carta topográfica de INEGI a la que pertenece dicho arco.

De acuerdo a la Sección 4.2, se requiere verificar la consistencia en los datos, por lo que se realiza previo a la generalización. Primeramente se asignan alturas a los nodos de cada arco de la capa de ríos. Esto se realiza por medio de una rejilla vectorial regular, generada a

<sup>69</sup> Se denominan ARC/Info Coverages o simplemente coberturas.

partir de un modelo digital de elevación con resolución de 25m. Mediante operaciones de *análisis espacial*<sup>70</sup>, *intersección*<sup>71</sup> (*Intersect*) y *delimitación de zonas*<sup>72</sup> (*Buffer*) son extraídas las elevaciones para cada nodo a partir de la rejilla vectorial (Ver sección 4.2.1). Se genera una tabla con la siguiente estructura:

*Rios-AN.AAT (FNODE#, TNODE#, LPOLY#, RPOLY#, LENGTH, RIOS-#, RÍOS-ID, ID, CLAVE, I-Altura, F-Altura)*<sup>73</sup>

Donde:

*I-Altura*<sup>74</sup>      Altura del nodo inicial de cada arco,  
*F-Altura*        Altura del nodo final de cada arco.

Una vez asignada las alturas en cada uno de los nodos que poseen los arcos que componen la red hidrológica, se verifica y corrige automáticamente la dirección del flujo (Ver Figura 5.1). La corrección se hace considerando que el flujo del río debe ir de la parte más alta hacia la parte más baja.

---

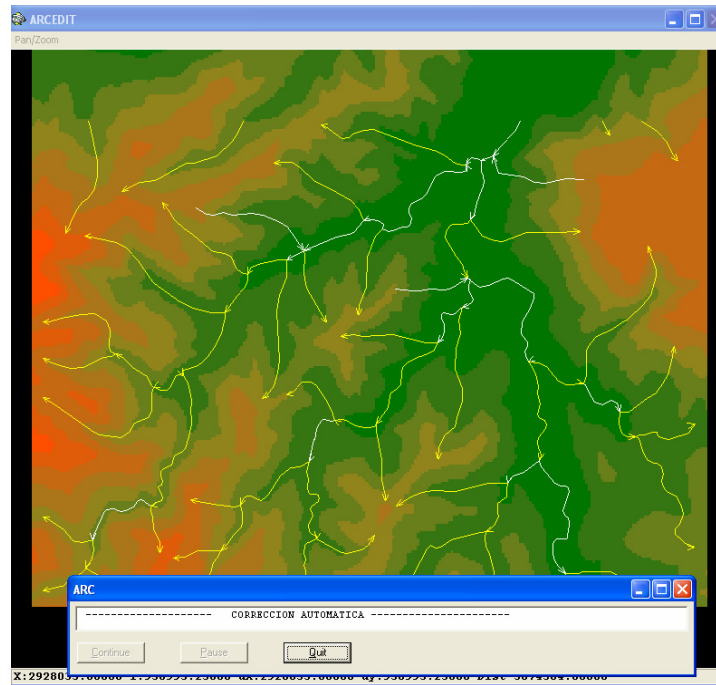
<sup>70</sup> Análisis Espacial en el contexto de los GIS, se refiere a los métodos para extraer de los datos existentes, aquellos que cumplan determinadas condiciones. Suelen incluir funciones de superposición topológica, intersecciones, uniones, entre otros. Generalmente se extrae o se generan nuevos datos a partir de la información de un grupo de objetos geográficos.

<sup>71</sup> Genera la intersección geométrica entre dos capas de datos geográficos, cuyo resultado se almacena en una nueva capa.

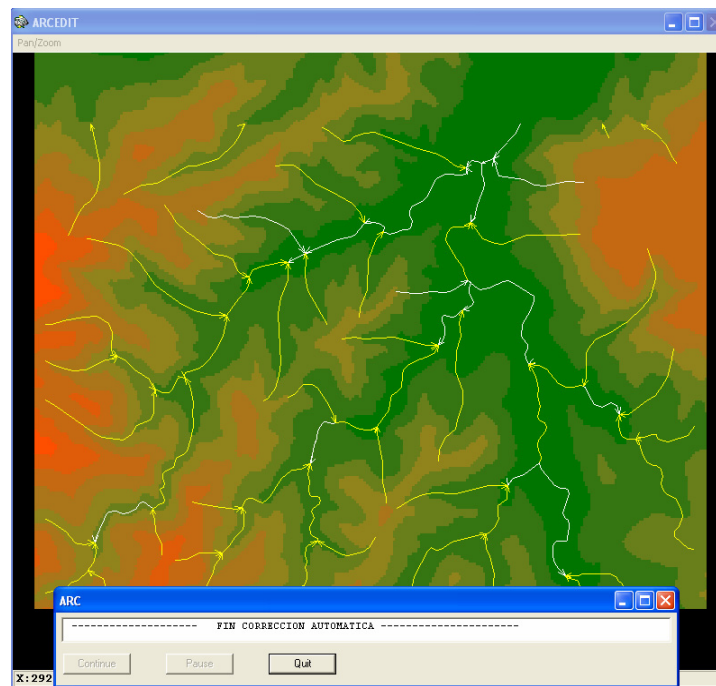
<sup>72</sup> Consiste en generar uno o más polígonos con forma envolvente trazado alrededor de un objeto geográfico (punto, línea o polígono). Generalmente, los parámetros para crear un buffer son el elemento o elementos que sirven de esqueleto para el trazado y una distancia de envoltura. El parámetro introducido puede ser también un área en vez de una distancia. No debe confundirse este concepto con el de equidistancia, puesto que en el concepto de buffer se garantiza que todos los puntos de la envolvente estén como mínimo a la distancia especificada.

<sup>73</sup> Los atributos definidos en Mayúsculas, representa que fueron definidos por el INEGI .

<sup>74</sup> Por convención, los atributos adicionales a los definidos por INEGI se escribirán en mayúsculas y el minúsculas, dichos atributos serán añadidos durante cada una de las etapas que serán presentadas a los largo del capítulo.



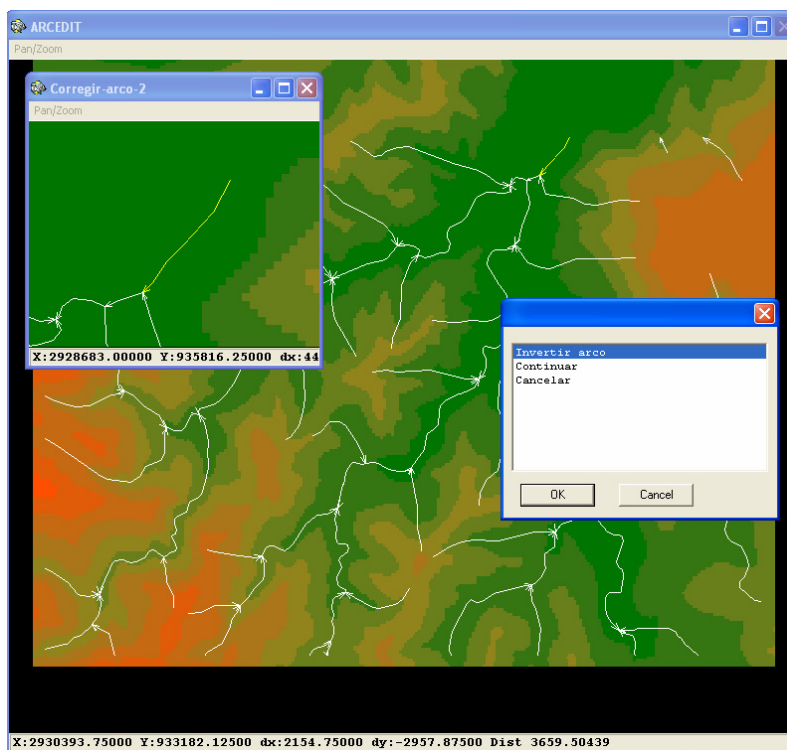
a)



b)

**Figura 5.1** Corrección automática del flujo para una red hidrológica (las puntas de las flechas indican la dirección del flujo). a) Antes de la corrección automática. b) Después de la corrección automática

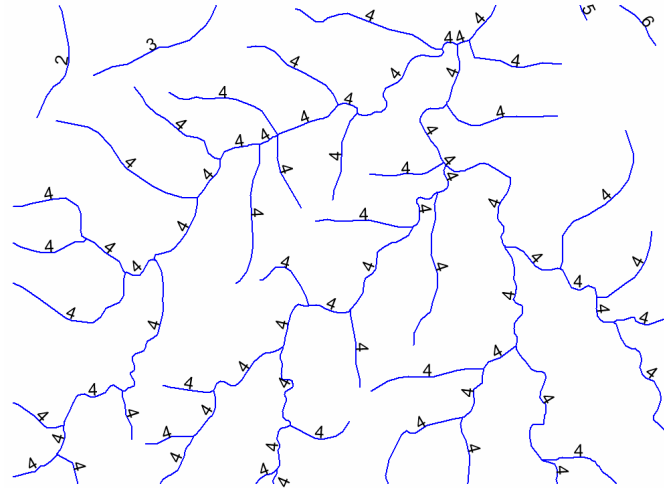
Debido a que se generan casos donde la altura para el nodo inicial y final es la misma (originado por la resolución del Modelo Digital de Elevación (DEM)), estos casos deben ser resueltos de forma interactiva (Ver Figura 5.2), donde el usuario utiliza el DEM como base para determinar si la dirección es correcta o no.



**Figura 5.2** Corrección interactiva del flujo para una red hidrológica

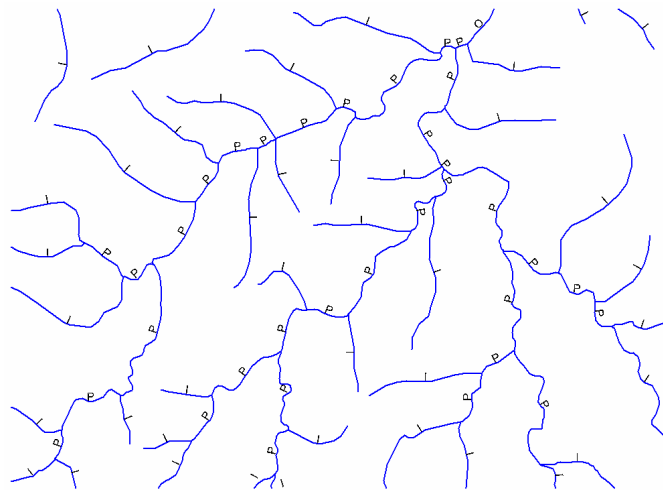
En los casos en los que se aplicó este método se corrigieron automáticamente alrededor del 80% de los arcos, pero este número depende del caso de estudio y de la forma del relieve.

Posteriormente se realizan clasificaciones sobre los datos que auxiliarán el procesamiento subsecuente. La primera clasificación es por subsistemas, la que se enfoca a colocar un identificador único a cada uno de los arcos que forman la red hidrológica que tenga alguna posible conexión (Ver Figura 5.3). Así, un subsistema estará formado por los arcos de la capa de ríos que tengan alguna posible conexión.



**Figura 5.3** Clasificación por subsistemas, donde se muestran los identificadores para cada subsistema

Posteriormente, los arcos de la red hidrológica son clasificados por la función que desempeñan en la red. Los posibles atributos de esta clasificación son entrada (I), salida (O) e intermedio (P) (Ver Figura 5.4).



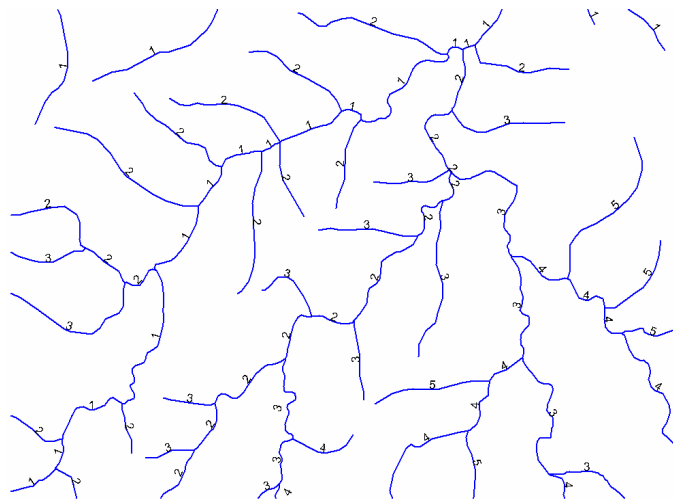
**Figura 5.4** Clasificación por la función de cada arco en la red hidrológica

La última clasificación a emplear es la *Clasificación Jerarquizada por Longitud (CLAJER)* (Ver Sección 4.2.2), la cual resulta muy útil, ya que se emplea tanto para generar los identificadores para cada arco que compone la red hidrológica, como también para cuantificar los objetos que conforman la red por medio de un atributo denominado *orden*. CLAJER [Moreno, 2001] tiene como propósito destacar las corrientes principales (las de mayor longitud) que conforman una red hidrológica. Este tipo de clasificaciones son muy



utilizadas por los hidrólogos para cuantificar los componentes de una red hidrológica, y puede tener diversas aplicaciones tales como la elaboración de cartas de hidrología superficial y de riesgos en cuencas.

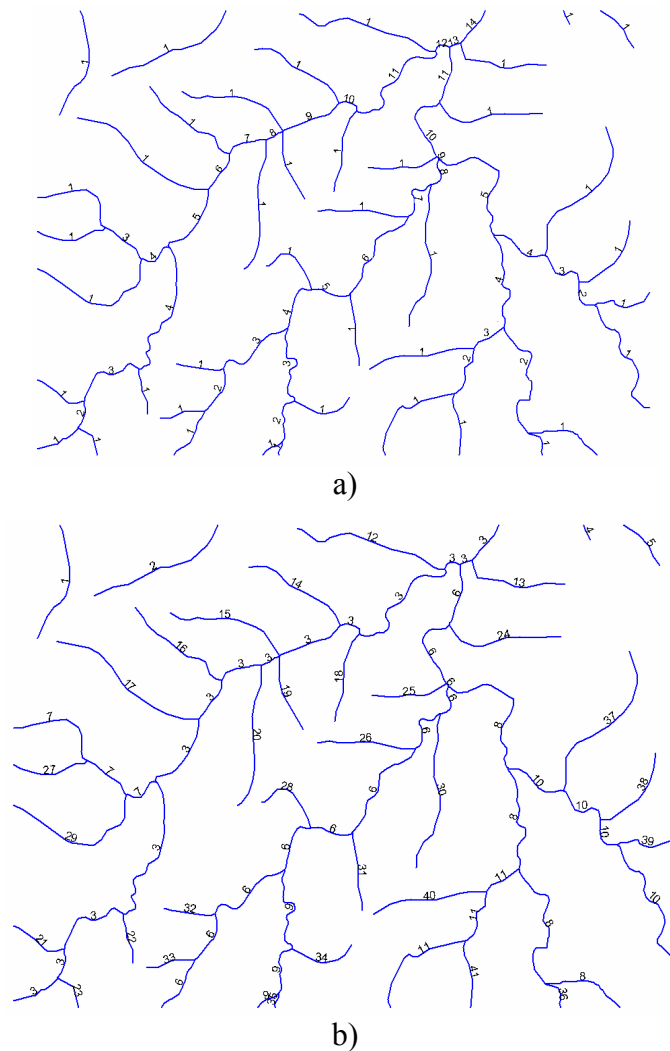
Para identificar la importancia jerárquica de un arco en la red, se utiliza un atributo denominado *orden*. El orden de mayor importancia es 1. El método de clasificación se basa en encontrar los arcos que siguiendo el flujo a contracorriente formen la corriente más larga, comenzando desde el punto más bajo (salida) hasta algún punto de mayor elevación (entrada). La serie de arcos que formen la corriente de mayor longitud, se les asignará *orden* = 1. Los de *orden* = 2, serán los más largos que se conecten a la corriente de *orden* = 1 y así sucesivamente (Ver Figura 5.5). Este proceso puede repetirse hasta clasificar todos los arcos de un red, sin embargo, en algunas aplicaciones, como un cambio de escala drástico (de 1:50,000 a 1:1,000,000) solo se requieren los de *orden* 1 y 2. Esta función forma parte del conjunto de algoritmos desarrollados para esta tesis.



**Figura 5.5.** Clasificación de la red hidrológica usando CLAJER

Adicionalmente durante esta clasificación se genera un identificador para cada arco, el cual define la dirección del flujo. Comenzando con 1 para a todos los arcos entrada hasta encontrar el arco salida. (Ver Figura 5.6a). Adicionalmente, durante esta clasificación se asigna un identificador único a cada uno de los ríos, detectados al realizar la CLAJER. Se requiere añadir un identificador adicional a los proporcionados por INEGI, denominado *Id-*

R. Esto se debe a que dicho identificador no hace referencia a un río dentro de la red hidrológica, sino a un arco (Ver Figura 5.6b).



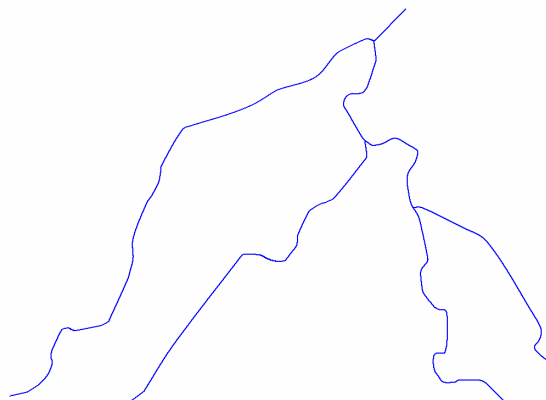
**Figura 5.6** Asignación de identificadores usando CLAJER

Una vez terminadas las clasificaciones se procede a modificar la escala en los datos, lo que se realiza haciendo uso de los operadores de generalización selección/eliminación y simplificación<sup>75</sup>. La selección/eliminación se realiza sobre los ríos con longitud menor a 2,500m<sup>76</sup>.

<sup>75</sup> Ver Sección 2.1.2.3 Modelo de McMaster y Shea.

<sup>76</sup> Este valor procede de [INEGI, 1995].

Por otro lado, la simplificación se hace por medio del algoritmo de Douglas – Peucker [Douglas y Peucker, 1973]. El resultado de la generalización de la red hidrológica aparece en la Figura 5.7.



**Figura 5.7** Capa de los ríos generalizada

La estructura que tiene la tabla de la capa de ríos al final de la generalización (*Rios-100.AAT*) tiene la siguiente estructura:

*Rios-100.AAT* (*FNODE#*, *TNODE#*, *LPOLY#*, *RPOLY#*, *LENGTH*, *RIOS-100#*, *RIOS-100-ID*, *ID*, *CLAVE*, *I-Altura*, *F-Altura*, *Función*, *Orden*, *Id-R*, *Subsis*, *Direc*)

Donde:

<i>FNODE#</i>	Número de nodo inicial de cada arco,
<i>TNODE#</i>	Número de nodo final de cada arco,
<i>LPOLY#</i>	Número interno del polígono de la izquierda,
<i>RPOLY#</i>	Número interno del polígono de la derecha,
<i>LENGTH</i>	Longitud de cada arco, medido en unidades de la cobertura,
<i>RIOS-100#</i>	Número de registro,
<i>RIOS-100-ID</i>	Identificador del usuario,
<i>ID</i>	Identificador de cada arco, definido por INEGI,
<i>CLAVE</i>	Identificador de la carta topográfica de INEGI a la que pertenece dicho arco,
<i>Función</i>	Función de cada arco de la red hidrológica (Entrada, Intermedio y Salida),
<i>Subsis</i>	Define la conectividad entre arcos de la red hidrológica que tienen conexión,
<i>Orden</i>	Define el valor de clasificación obtenido por CLAJER,
<i>Id-R</i>	Identificador único para cada río.
<i>Direc</i>	Número de relación que tiene un río siguiendo el flujo descendente.

Debido a que durante las etapas de Análisis, Síntesis y Verificación se utilizarán algunos de los atributos que se agregaron durante la generalización, los datos usados en dicha etapas tendrán la misma estructura que los datos generalizados. Estas capas se nombrarán como *Ríos-I*. La Figura 5.8 muestra un fragmento de la tabla *Ríos-I.aat* y su representación gráfica

<i>FnodeA</i>	<i>TnodeA</i>	<i>LpolyA</i>	<i>RpolyA</i>	<i>Length</i>	<i>R3-1A</i>	<i>R3-7A</i>	<i>R3-pacA</i>	<i>R3-pac-id</i>	<i>Faltura</i>	<i>I-faltura</i>	<i>Id</i>	<i>Clave</i>	<i>Fc</i>
25	24	0	0	205.27270	11	11127	66	11127	260	260	1930413	f14d74	
26	25	0	0	456.31253	12	11128	12	11128	280	260	1930414	f14d74	
31	26	0	0	461.80090	13	11129	15	11129	300	280	1930415	f14d74	
41	31	0	0	851.71699	14	11231	22	11231	320	300	1930517	f14d74	
60	41	0	0	1583.39684	15	11234	40	11234	400	320	1930520	f14d74	
65	60	0	0	876.27147	16	11236	74	11236	400	400	1930522	f14d74	
70	65	0	0	301.25037	17	11335	50	11335	420	400	1930621	f14d74	
73	70	0	0	529.94628	18	11338	52	11338	440	420	1930624	f14d74	
2	1	0	0	634.22589	19	11024	0	0	240	240	1930310	f14d74	
5	2	0	0	838.70709	20	11021	0	0	260	240	1930307	f14d74	
6	5	0	0	41.01159	21	11028	0	0	260	260	1930314	f14d74	
8	6	0	0	354.04884	22	11027	0	0	260	260	1930313	f14d74	
11	8	0	0	558.23225	23	11026	0	0	260	260	1930312	f14d74	
9	13	0	0	1054.06787	24	11329	0	0	558	380	1930615	f14d74	
13	18	0	0	531.86053	25	11233	0	0	380	340	1930519	f14d74	
18	16	0	0	400.47456	26	11232	0	0	340	320	1930518	f14d74	

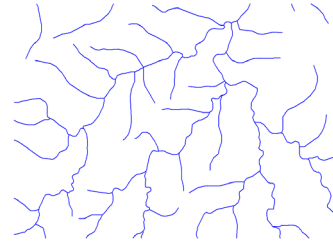


Figura 5.8 Fragmento de la tabla *Ríos-I.aat* y su representación gráfica

### 5.1.2 Generalización de las curvas de nivel

Al igual que en el caso de las redes hidrológicas en el caso de curvas de nivel se utilizaron datos del INEGI. La estructura que tienen las tablas que almacenan las curvas de nivel tiene la siguiente estructura.

*Curvas.AAT (FNODE#, TNODE#, LPOLY#, RPOLY#, LENGTH, CURVAS-#, CURVAS-ID, ID, CLAVE, ELEVATION)*

Donde:

- FNODE#* Número de nodo inicial de cada arco,
- TNODE#* Número de nodo final de cada arco,
- LPOLY#* Número interno del polígono de la izquierda,
- RPOLY#* Número interno del polígono de la derecha,
- LENGTH* Longitud de cada arco, medido en unidades de la cobertura,
- CURVAS#* Número de registro,
- CURVAS-ID* Identificador del usuario,
- ID* Identificador de cada arco, definido por INEGI,
- CLAVE* Identificador de la carta topográfica de INEGI a la que pertenece dicho arco,
- ELEVATION* Elevación de cada curva.

Al igual que el caso de redes hidrológicas se requiere añadir un identificador único a cada una de las curvas. Esto se debe a que el identificador de INEGI se refiere solo a un arco y no a la curva en su totalidad. Para realizar esta operación se hace uso de funciones de *análisis espacial*.

Las coberturas de las curvas de nivel son generalizadas empleando los operadores de generalización selección/eliminación y simplificación. La selección se realiza sobre las curvas de nivel con elevación diferente a múltiplos de 100m<sup>77</sup>. Después son simplificadas utilizando el algoritmo de Douglas – Peucker (Ver Figura 5.9). Las tablas de las coberturas que almacenan las curvas de nivel generalizadas (*Curvas-I.aat*) tienen la siguiente estructura:

*Curvas-100.AAT (FNODE#, TNODE#, LPOLY#, RPOLY#, LENGTH, CURVAS-#,  
CURVAS-ID, ID, CLAVE, ELEVATION, Id-C)*

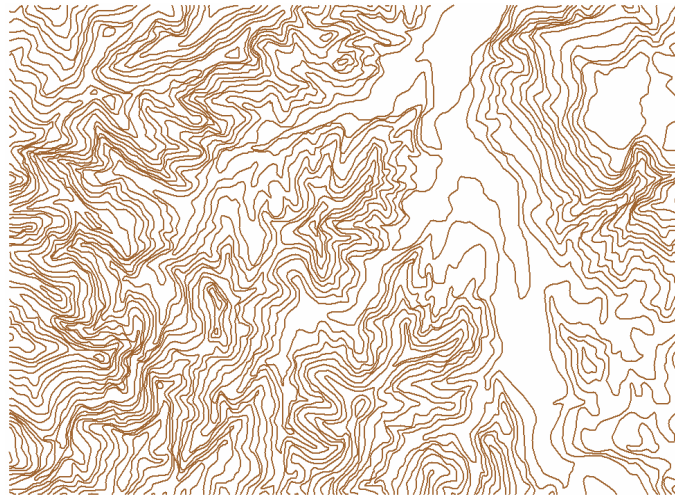
Donde:

<i>FNODE#</i>	Número de nodo inicial de cada arco,
<i>TNODE#</i>	Número de nodo final de cada arco,
<i>LPOLY#</i>	Número interno del polígono de la izquierda,
<i>RPOLY#</i>	Número interno del polígono de la derecha,
<i>LENGTH</i>	Longitud de cada arco, medido en unidades de la cobertura,
<i>CURVAS-100#</i>	El número de registro,
<i>CURVAS-100-ID-</i>	Identificador del usuario,
<i>ID</i>	Identificador de cada arco, definido por INEGI,
<i>CLAVE</i>	Identificador de la carta topográfica de INEGI a la que pertenece dicho arco,
<i>ELEVATION</i>	Elevación de cada curva.
<i>Id-C</i>	Identificador único para cada curva de nivel.

Al igual que en el caso de las redes hidrológicas, se utilizarán los atributos que se agregaron durante la generalización. Esto implica que los datos fuente tendrán la misma estructura y se denominarán *Curvas-I*.

---

<sup>77</sup> Se considera 100m, por las especificaciones del INEGI



a) antes de generalizar



b) después de generalizar

**Figura 5.9** Generalización de las curvas de nivel

## 5.2 Generación de la base de datos espacial parcialmente conceptualizada

La implantación de la *base de datos espacial parcialmente conceptualizada*, está basada en las mediciones realizadas a los datos geográficos. En este caso presentaremos el procedimiento de implantación de la *medición de la relación >Pasa-por<*. Posteriormente describiremos la generación de la *tabla relaciones conceptualmente representadas* para la Relación Pasa-Por. Únicamente describiremos el caso de los datos fuente, ya que el procedimiento para datos generalizados es exactamente el mismo. La única diferencia es en

el nombre de las capas, que este caso se denominan como *Curvas-100* y *Rios-100* en vez de *Curvas-1* y *Rios-1*.

Ya que por definición los Sistemas de Información Geográfica almacenan solo una temática a la vez, se requiere combinar las capas de ríos y de curvas de nivel para realizar la *medición de la relación Pasa-por* (función *append*<sup>78</sup>). El resultado de la combinación será agregado a la capa denominada como RC-N. Esta combinación es posible debido a que ambas capas están representadas como líneas. Recordemos que las tablas de atributos de las capas mencionadas tienen la forma.

*Curvas-1.AAT* (FNODE#, TNODE#, LPOLY#, RPOLY#, LENGTH, CURVAS-#, CURVAS-ID, ID, CLAVE, ELEVATION, Id-C)

*Rios-1.AAT* (FNODE#, TNODE#, LPOLY#, RPOLY#, LENGTH, RIOS#, RIOS-ID, ID, CLAVE, I-Altura, F-Altura, Función, Orden, Id-R, Subsis, Direc)

Con el fin de diferenciar a los *objetos río* de los *objetos curva de nivel* en la capa resultante se utilizan los atributos *Id-R* e *Id-C* (*identificadores de objetos*) son utilizados para definir el nombre de cada objeto. Esto se realiza agregando el atributo *Nom*. Para cada tupla de las tablas *Rios1.AAT* y *Curvas1.AAT* se realiza la concatenación de los prefijos *Curva\_* y *Río\_* (según corresponda) con el valor de *Id-R* e *Id-C* (por ejemplo, *Río\_1*). Para la capa de curvas de nivel, adicionalmente, se concatena el valor de *Nom* con el valor de *ELEVATION* (por ejemplo, *Curva\_1(1500)*).

Además, es agregado el atributo *Direc* a la capa de curvas. Para distinguir entre objetos curva de nivel y ríos, se agrega el atributo *Tipo*. Para cada registro *Rios1.AAT* y *Curvas1.AAT*, se asignan los valores *R* y *C* respectivamente. Esto se realiza con el fin de que esos atributos se preserven en la capa *RC-I-N*, ya que por definición de la función *append*, únicamente se preservarán los atributos que tengan en común las tablas que serán combinadas. Las tablas *Rios1.AAT* y *Curvas1.AAT* quedan de la siguiente forma:

*Curvas-1.AAT* (FNODE#, TNODE#, LPOLY#, RPOLY#, LENGTH, CURVAS-#, CURVAS-1-ID, ID, CLAVE, ELEVATION, *Nom*, *Direc*)

---

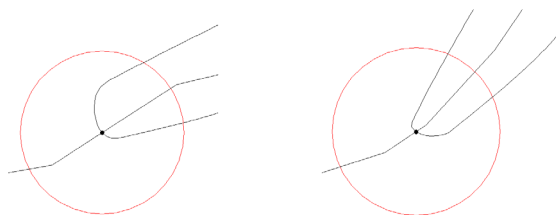
<sup>78</sup> Es proporcionada por ARC/INFO, para mayores referencias consultar, ARCDoc (ayuda en línea de ARC/INFO).





de arcos son una instancia de >Pasa-Por<. Asimismo, se agrega un identificador único para cada relación denominado *Numrel*, el cual es un valor entero secuencial que comienza en 1.

Sobre los nodos identificados con *Inter = I*, se realiza un análisis *Buffer*. Esto se hace para delimitar el *área de análisis de la relación (AAR)* (Ver Sección 4.3.2). La Figura 5.11 muestra el *AAR* que se identificó con base en un análisis *Buffer*. Hay que señalar que en este momento contamos con dos capas: la capa que almacena el Buffer y la capa que contiene la combinación de los ríos y las curvas de nivel (*RC-I-N.AAT*).



**Figura 5.11** Fragmentos donde se identificó el *AAR*

Una vez que se identificó el *AAR* con base en las intersecciones, se procede a extraer los arcos de *RC-I-N.AAT* que están dentro de una *AAR*, es decir, que forman parte de una relación *Pasa-por*. Esto se realiza por medio de la función *Clip (recortar o extraer)*<sup>79</sup>. El resultado de la función *Clip* se almacena en una nueva capa denominada *CLIP-RC-I-N*. La Figura 5.12 muestra los arcos que fueron extraídos usando *Clip*.



**Figura 5.12** Extracción de arcos que participan en la relación >Pasa-por<

La estructura de la tabla *CLIP-RC-I-N.AAT* es como sigue:

*CLIP-RC-I-N.AAT (FNODE#, TNODE#, LPOLY#, RPOLY#, LENGTH, CLIP-RC-I-N#, CLIP-RC-I-N-ID, Nom, Tipo, Direc, Inter, Numrel)*

<sup>79</sup> Es proporcionada por ARC/INFO, Para mayores referencias consultar, ARCDoc (ayuda en línea de ARC/INFO) y se refiere a un tipo de análisis espacial que tiene como propósito la extracción de una subclase de datos espaciales dentro de un grupo de datos mayor, mediante la selección única de los datos localizados dentro de unos límites prefijados. Al área encerrada por esos límites se la denomina CLIP WINDOW. Como resultado se genera una nueva capa con los datos extraídos.

Una vez extraídos los arcos que conforman las relaciones entre los objetos *río* y los objetos *curva de nivel*, se procede a calcular el *descriptor de la relación Pasa por (DCP)* sobre la capa *CLIP-RC-1-N.AAT*. Dicho coeficiente se obtiene considerando los nodos y vértices que contienen los arcos que representan una relación (Ver Sección 4.3.2). Con el propósito de almacenar el valor de *DCP*, se agrega un atributo denominado *Clasif*; el cual será utilizado para clasificar los valores de *DCP* en grupos que representan las subclases de la relación Pasa Por (<*pasa-por-máxima-convexidad*>, <*casi-pasa-por-máxima-convexidad*>, <*pasa-por-convexidad*>, <*pasa-por-recta*>, <*pasa-por-concavidad*>, <*casi-pasa-por-máxima-concavidad*> y <*pasa-por- máxima-concavidad*>). Para realizar la representación explícita de la relación >Pasa-por<, se agrega el atributo *Cpto*, el cual asignará los valores PasaPorMáximaConvexidad, PasaCasiPorMáximaConvexidad, PasaPorConvexidad, PasaPorRecta, PasaPorConcavidad, PasaCasiPorMáximaConcavidad y PasaPorMáximaConcavidad según corresponda.

*CLIP-RC-1-N.AAT (FNODE#, TNODE#, LPOLY#, RPOLY#, LENGTH, CLIP-1-N#, CLIP-1-N-ID, Nom, Tipo, Direc, Inter, Numrel, Cpto)*

En este momento es cuando se genera la *tabla relaciones conceptualmente representadas* de la relación Pasa por ( $T_{RP}$ ). El primer paso es generar una tabla de la forma  $T_R (O_1, R, O_2)$ , para esto se agrega un atributo denominado *Nom-C* que almacenará el nombre de los objetos curva. Posteriormente se debe seguir el siguiente procedimiento:

```

Reg, variable que almacena un número de registro
Nc, es una variable que almacena el nombre de un objeto curva
Repetir desde i = 1 hasta el valor mas grande en Numrel
    SELECCIONAR Numrel = i
    RESELECCIONAR80 Tipo = R //arcos tipo río que forman parte de la
relación
    Hacer Reg := el número de registro del primer arco seleccionado
    BORRAR-SELECCION
    SELECCIONAR Numrel = i
    RESELECCIONAR Tipo = C //arcos tipo curva que forman parte de la
relación

```

<sup>80</sup> RESELECCIONAR, Significa hacer una selección sobre los registros ya seleccionados.

```
Hacer Nc := el valor del atributo Nom en el primer registro
seleccionado
Para el registro = Reg,
    hacer Nom-C := NC
Eliminar los registros que no tengan valores para Nom-C
```

Se debe renombrar el Atributo *Nom* como *Nom-R*. Los resultados de este procedimiento serán almacenados en una tabla denominada *CP-RC-I-N*, la cual posee la siguiente estructura:

*CP-RC-I-N.AAT* (*FNODE#*, *TNODE#*, *LPOLY#*, *RPOLY#*, *LENGTH*, *CP-I-N#*, *CP-I-N-ID*, *Nom-R*, *Cpto*, *Nom-C*, *Numrel*, *Tipo*, *Direc*)

De la estructura de *CP-RC-I-N.AAT* se puede ver que ya tiene la estructura  $T_R (O_1, R, O_2)$ , donde *Nom-R* representa a  $O_1$ , *Cpto* representa a  $R$  y *Nom-R* representa a  $O_2$ . Por otro lado, los atributos, *Numrel*, *Tipo*, *Direc*, representan las propiedades de la clase *>Pasa-por<*. El resto de los atributos no serán considerados, simplemente deben dejarse para manejar esta tabla dentro de ARC/INFO.

Para el caso de una  $T_{RP}$  de datos generalizados, se genera una capa denominada *CP-RC-100-N*, la cual tendrá una tabla de atributos con estructura:

*CP-RC-100-N.AAT* (*FNODE#*, *TNODE#*, *LPOLY#*, *RPOLY#*, *LENGTH*, *CP-100-N#*, *CP-100-N-ID*, *Nom-R*, *Cpto*, *Nom-C*, *Numrel*, *Tipo*, *Direc*)

### 5.3 Verificación de la Consistencia de los datos generalizados

Una vez generada la *tabla de relaciones conceptualmente representadas* de la relación *>Pasa-por<* ( $T_{RP}$ ) se procede a verificar la consistencia de la relaciones entre las capas de ríos y de curvas de nivel. El primer paso para encontrar las inconsistencias se hace verificando los datos generalizados. La Figura 5.13, presenta la representación gráfica de las capas de datos generalizados automáticamente.

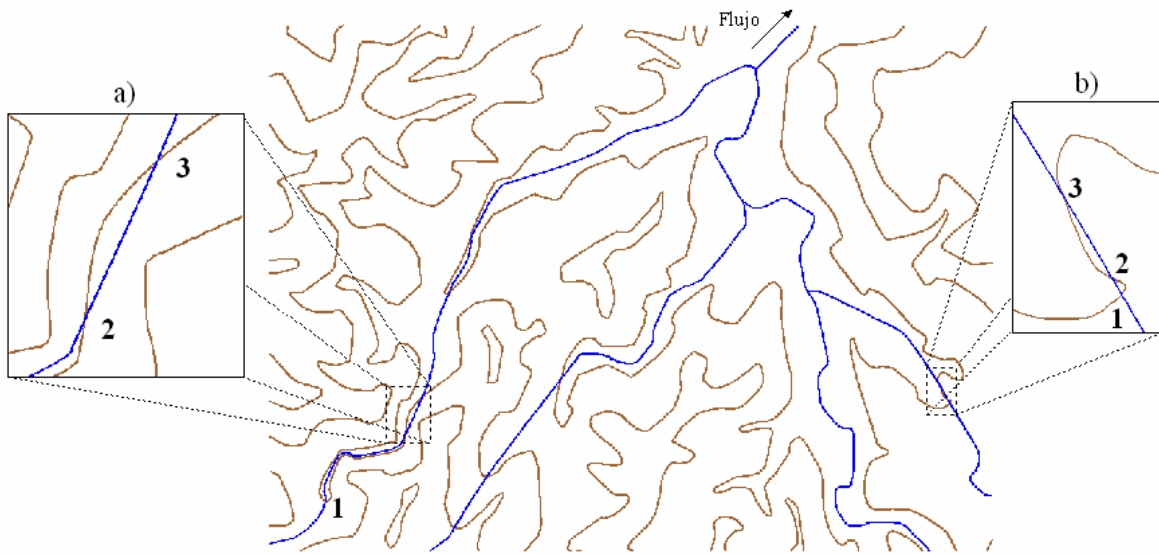


Figura 5.13 Datos generalizados e identificación visual de inconsistencias

De la Figura 5.13, se pueden identificar dos casos en donde se generaron relaciones adicionales que representan inconsistencias. Estos casos son los identificados con 2 y 3. Son inconsistencias debido a que por definición un río y una curva solo deben intersectarse una vez. Desde el punto de vista de la representación cartográfica, representa que el flujo del río baja y sube y después vuelve a bajar, por lo que representa una situación absurda. Las inconsistencias son detectadas automáticamente utilizando la *tabla de relaciones conceptualmente representadas* de los datos generalizados, es decir *CP-RC-100-N.AAT* (Ver Tabla 5.1 ).

Tabla 5.1 Tabla de relaciones conceptualmente representadas de datos generalizados

CP_RC_100#	Nom-R	Cpto	Nom-C	Direc	Numrel	
1	RIO_3	PASA CASI POR MAXIMA CONVEXIDAD	Curva_6	1	1	
2	RIO_6	PASA POR RECTA	Curva_6	1	2	
3	<b>RIO_10</b>	<b>PASA POR RECTA</b>	<b>Curva_20</b>	<b>3</b>	<b>3</b>	X
4	RIO_3	<i>PASA POR RECTA</i>	Curva_5	3	4	Y
5	<b>RIO_10</b>	<b>PASA CASI POR CONVEXIDAD</b>	<b>Curva_20</b>	<b>2</b>	<b>5</b>	X
6	<b>RIO_10</b>	<b>PASA CASI POR MAXIMA CONVEXIDAD</b>	<b>Curva_20</b>	<b>1</b>	<b>6</b>	X
7	RIO_3	<i>PASA POR RECTA</i>	Curva_5	2	7	Y
8	RIO_6	PASA CASI POR MAXIMA CONVEXIDAD	Curva_5	1	8	
9	RIO_3	<i>PASA POR CONVEXIDAD</i>	Curva_5	1	9	Y
10	RIO_6	PASA POR MAXIMA CONVEXIDAD	Curva_13	1	10	

De la Tabla se pueden ver los registros de la  $T_{RP}$ <sup>81</sup> que aparecen en la Figura 5.13. Los marcados con *X* y en cursivas, son los que corresponden al detalle a) de la Figura 5.13; por otro lado los marcados con **Y** y en negritas, corresponden al detalle b) de la Figura 5.13. Para distinguir las relaciones inconsistentes se utiliza la Tabla.5.2 .

**Tabla 5.2** Tabla de relaciones conceptualmente representadas después de identificar inconsistencias sobre los datos generalizados

CP_RC_100#	Nom-R	Cpto	Nom-C	Direc	Numrel
1	RIO_3	PASA CASI POR MAXIMA CONVEXIDAD	Curva_6	1	1
2	RIO_6	PASA POR RECTA	Curva_6	1	2
6	RIO_10	PASA CASI POR MAXIMA CONVEXIDAD	Curva_20	1	6
8	RIO_6	PASA CASI POR MAXIMA CONVEXIDAD	Curva_5	1	8
9	RIO_3	PASA CASI POR CONVEXIDAD	Curva_5	1	9
10	RIO_6	PASA POR MAXIMA CONVEXIDAD	Curva_13	1	10

Se seleccionan los registros que tengan valores iguales para *Nom-C* y *Nom-R*. Con base en esto, si tenemos que dos o más registros tienen valores iguales para *Nom-C* y *Nom-R*, el registro con el menor valor en *Direc*, representa las relaciones consistentes. El resto de los registros con *Nom-C* y *Nom-R* iguales representan inconsistencias, ya que por definición una curva de nivel y un río deben intersectarse exactamente una vez. De manera que los errores con inconsistencias, no serán incluidos en los siguientes pasos del procedimiento, por lo que tendremos la  $T_{RFP}$  que será utilizada.

Ahora, se procederá a mostrar cómo se identifican las inconsistencias usando la Teoría de la Confusión. Para esto se utilizará la  $T_{RGP}$  de los datos fuente (*CP-RC-I-N.AAT*) (Ver Tabla 5.3).

<sup>81</sup> Se eliminaron algunos atributos en esta visualización, ya que son innecesarios.

**Tabla 5.3** Tabla de relaciones conceptualmente representadas de los datos fuente

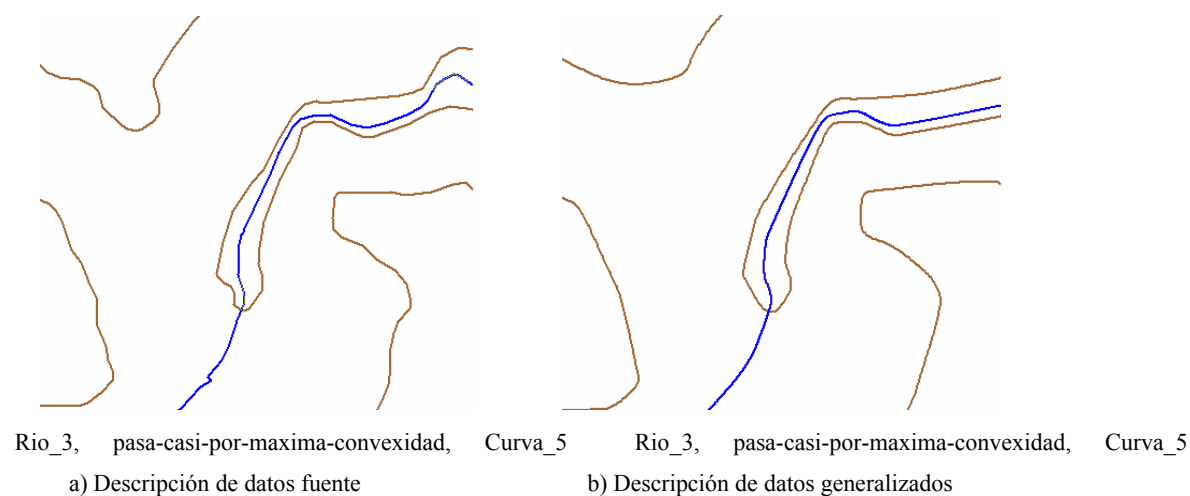
CP_RC_1#	Nom-R	Cpto	Nom-C	Direc	Numrel
1	RIO_4	PASA CASI POR MAXIMA CONVEXIDAD	Curva_21	1	1
2	RIO_5	PASA CASI POR MAXIMA CONVEXIDAD	Curva_22	1	2
3	RIO_1	PASA POR CONVEXIDAD	Curva_1	1	3
4	RIO_2	PASA POR MAXIMA CONVEXIDAD	Curva_4	1	4
5	RIO_2	PASA CASI POR MAXIMA CONVEXIDAD	Curva_3	1	5
6	RIO_1	PASA POR MAXIMA CONVEXIDAD	Curva_2	1	6
7	RIO_7	PASA POR RECTA	Curva_3	1	7
8	RIO_7	PASA POR MAXIMA CONVEXIDAD	Curva_5	1	8
9	RIO_3	PASA POR MAXIMA CONVEXIDAD	Curva_6	1	9
10	RIO_10	PASA CASI POR MAXIMA CONVEXIDAD	Curva_20	1	10
11	RIO_6	PASA POR CONVEXIDAD	Curva_6	1	11
12	RIO_9	PASA CASI POR MAXIMA CONVEXIDAD	Curva_6	1	12
13	RIO_6	PASA CASI POR MAXIMA CONVEXIDAD	Curva_5	1	13
14	RIO_11	PASA POR MAXIMA CONVEXIDAD	Curva_6	1	14
15	RIO_3	PASA CASI POR MAXIMA CONVEXIDAD	Curva_5	1	15
16	RIO_6	PASA POR MAXIMA CONVEXIDAD	Curva_13	1	16
17	RIO_4	PASA CASI POR MAXIMA CONVEXIDAD	Curva_21	1	17
18	RIO_5	PASA CASI POR MAXIMA CONVEXIDAD	Curva_22	1	18

Es importante mencionar que la tabla  $T_{RGP}$  tiene menos registros que  $T_{RFP}$ , debido a que se trata de datos generalizados. Hay que recordar que para realizar la verificación se hace uso del los conceptos definidos en la ontología, en particular de las subclases de  $\langle Pasa-Por \rangle$  (Ver Sección 4.6.7). La medición que soporta la verificación es  $conf_R(r,s)$ , la confusión de tener una relación  $r$  en vez de una relación  $s$ . A continuación se presentará la verificación de algunas relaciones.

Consideremos la Figura 5.14, y supongamos un umbral  $u = 1/7$ . La verificación se basa en comparar dos registros que relacionan a los mismos objetos donde medimos la confusión

entre un registro de la tabla  $T_{RGP}$  {Rio\_3, pasa-casi-por-maxima-convexidad, Curva\_5} y un registro de la tabla  $T_{RFP}$  {Rio\_3, pasa-casi-por-maxima-convexidad, Curva\_5}.

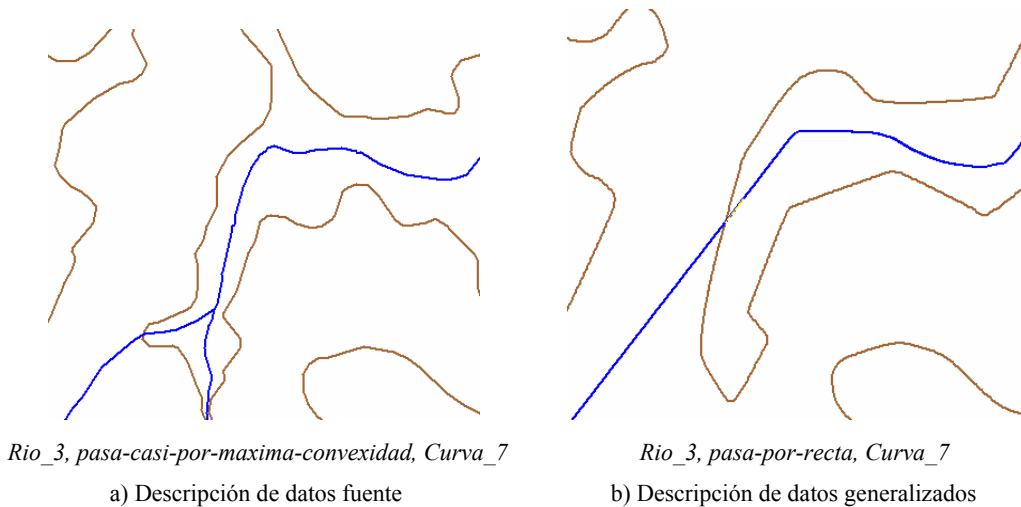
La confusión se mide como  $conf_R(r, s)$ , donde  $r = \text{pasa-por-recta}$  y  $s = \text{pasa-casi-por-maxima-convexidad}$ . En este caso,  $r$  equivale a la posición 2 ( $c_j = 2$ ) dentro del vector de clases  $V[]$ . Por otro lado,  $s$  equivale a la posición 2 ( $c_k = 2$ ) dentro de  $V[]$ . De manera que la  $conf_P(r,s) = conf_P(2,2) = 0$  (por definición). El considerar lo anterior, significa que la distancia semántica entre *pasa-casi-por-maxima-convexidad* y *pasa-casi-por-maxima-convexidad* es 0 (Ver Sección 4.6.3.1).



**Figura 5.14** Verificación desigual de la consistencia en datos generalizados, caso igual

La Figura 5.15 presenta el caso de una relación que se verificará. Recordar que en este caso el umbral  $u$  vale  $1/6$ . La verificación se basa en comparar dos registros que relacionan a los mismos objetos donde medimos la confusión entre un registro de la tabla  $T_{RGP}$  {Rio\_3, pasa-casi-por-maxima-convexidad, Curva\_7(600)} y un registro de la tabla  $T_{RFP}$  {Rio\_3, pasa-por-recta, Curva\_7(600)}. Entonces se mide la  $conf_R(r, s)$ , donde  $r = \text{pasa-por-recta}$  y  $s = \text{pasa-casi-por-maxima-convexidad}$ . En este caso,  $r$  equivale a la posición 2 ( $c_j = 2$ ) dentro del vector de clases  $V[]$ . Por otro lado,  $s$ , equivale a la posición 4 ( $c_k = 4$ ) dentro de  $V[]$ , de manera que la  $conf_P(r,s) = (Dist(2, 4) / 7 - 1) = 2/6$ .

Según lo anterior significa que la distancia semántica entre *pasa-por-recta* y *pasa-casi-por-maxima-convexidad* es de  $2/6$ , debido a que  $u = 1/6$  representa que al verificar los casos de la Figura 5.15 obtendremos que son *desiguales* (Ver Sección 4.6.3.2).

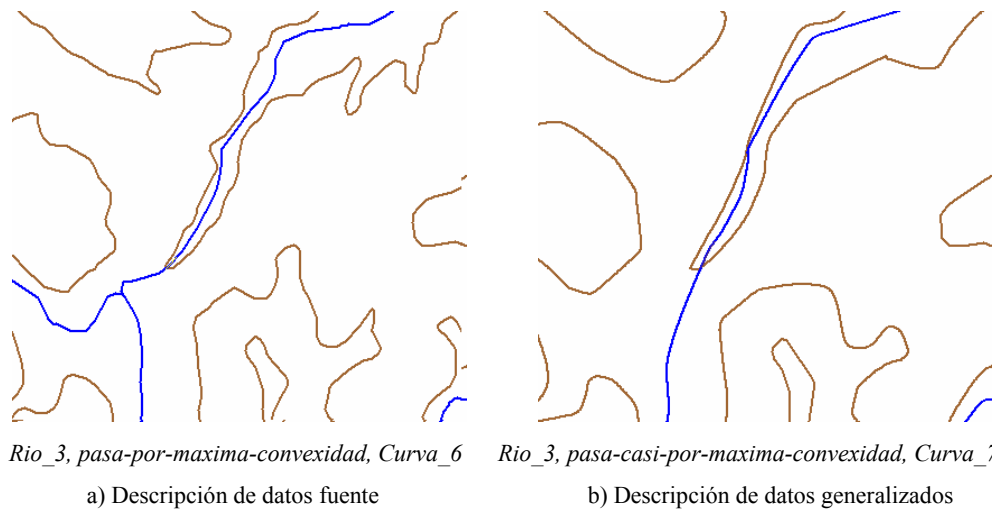


**Figura 5.15** Verificación desigual de consistencia en datos generalizados, caso desigual

La Figura 5.16 presenta el caso de una relación que se verificará cuando el umbral  $u$  vale  $1/6$ . La verificación se basa en la comparación de dos registros que relacionan a los mismos objetos y medición de la confusión, que es un registro de la tabla  $T_{RGP}$  {Rio\_3, pasa-por-maxima-convexidad, Curva\_6} y un registro de la tabla  $T_{RFP}$  {Rio\_3, pasa-casi-por-maxima-convexidad, Curva\_6}. La  $conf_R(r, s)$ , significa que  $r = pasa-casi-por-maxima-convexidad$  y  $s = pasa-por-maxima-convexidad$ . En este caso,  $r$  equivale a la posición 2 ( $c_j = 2$ ) dentro del vector de clases  $V[j]$ . Por otro lado,  $s$ , equivale a la posición 1 ( $c_k = 1$ ) dentro de  $V[j]$ . De manera que la  $conf_P(r,s) = 1/6$ .

Según lo anterior, significa que la distancia semántica entre *pasa—por-maxima-convexidad* y *pasa-casi-por-maxima-convexidad* es de  $1/6$ , debido a que  $u = 1/6$ , representa que al verificar los casos de la Figura 5.16 obtendremos que son *equivalentes* (Ver Sección 4.6.3.3).



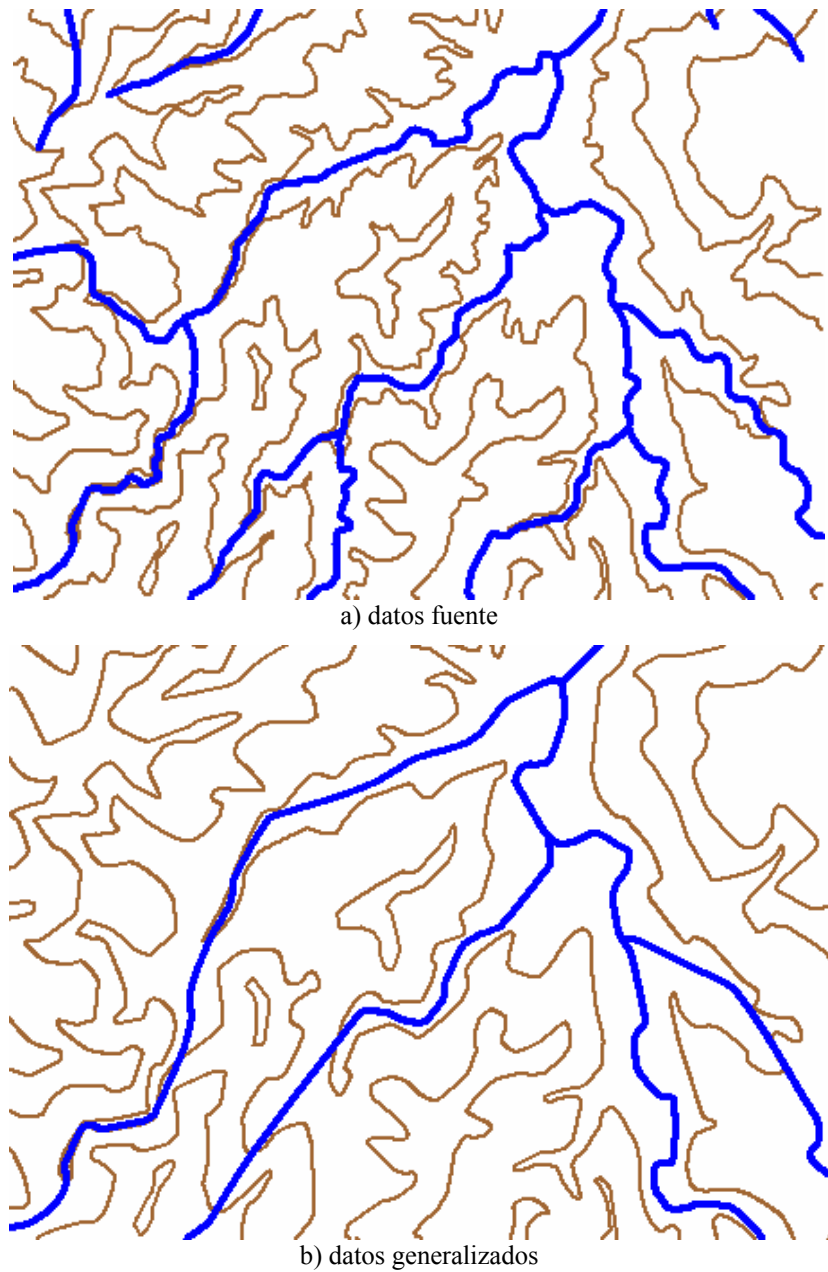


**Figura 5.16** Verificación desigual de consistencia en datos generalizados, caso equivalente

De esta forma, es como se evalúan los datos generalizados usando las representaciones conceptuales de las relaciones entre objetos geográficos. Es muy importante resaltar que se realiza la verificación sin el uso de coordenadas ni sistema de referencia. Estos quedan implícitamente representados en la representación conceptual. Si bien el procesamiento de los datos durante la etapa de análisis puede resultar un tanto exhaustivo, la etapa de verificación se reduce a consultas sobre una base de datos. Es decir realizar consultas sobre las tablas  $T_{RGP}$  y  $T_{RFP}$ , y de esta forma medir de la confusión, para evaluar la similitud a nivel local.

## 5. 4 Verificación de la consistencia a nivel global

En esta sección presentaremos los resultados de las mediciones de consistencia a nivel global, las cuales resultan muy útiles para identificar el tipo de usuarios que pueden utilizar los datos generalizados automáticamente. La Figura 5.17, muestra la representación gráfica de un conjunto de datos antes y después de generalizar, los cuales serán utilizados para medir el error acumulado y el error medio.



**Figura 5.17** Datos fuente y datos generalizados

Para calcular la similitud del conjunto de datos se hará uso de la tabla semánticamente consistente (Tabla 5.4), que tiene una estructura de la forma:

$$SC(SC\_RC\_\#, Nom-R, Cpto-r, Cpto-s, Nom-C, Conf)$$

Donde:

<i>SC_RC_#</i>	Número de registro,
<i>Nom-R</i>	Nombre del objeto río,
<i>Cpto-r</i>	Concepto que describe la relación en los datos generalizados,
<i>Cpto-s</i>	Concepto que describe la relación en los datos fuente,

*Nom-C* Nombre del objeto curva de nivel,  
*Conf* Medición de la confusión,

Las *relaciones semánticamente consistentes* que se utilizarán como ejemplo se presenta en la Tabla 5.4.

**Tabla 5.4** Tabla de relaciones semánticamente consistentes

SC_RC_#	Nom-R	Cpto-r	Cpto-s	Nom-C	Conf
1	RIO_3	PASA CASI POR MAXIMA CONVEXIDAD	PASA POR MAXIMA CONVEXIDAD	Curva_6	1/6
2	RIO_6	PASA POR RECTA	PASA POR CONVEXIDAD	Curva_6	1/6
3	RIO_10	PASA CASI POR MAXIMA CONVEXIDAD	PASA CASI POR MAXIMA CONVEXIDA	Curva_20	0
4	RIO_6	PASA CASI POR MAXIMA CONVEXIDAD	PASA CASI POR MAXIMA CONVEXIDA	Curva_5	0
5	RIO_3	PASA POR CONVEXIDAD	PASA CASI POR MAXIMA CONVEXIDA	Curva_5	1/6
6	RIO_6	PASA POR MAXIMA CONVEXIDAD	PASA CASI POR MAXIMA CONVEXIDA	Curva_13	1/6

El error acumulado ( $e_a$ ) (Ver Sección 4.6.8.1), se obtiene como sigue:

$$e_a = \sum_{i=1}^n \text{conf}_R(r_i, s_i) = 1/6 + 1/6 + 0 + 0 + 1/6 + 1/6 = 4/6$$

Si bien el valor del error acumulado es 4/6, en términos generales podemos decir que si el valor no es mayor que 1, los datos son hasta cierto punto aceptables, pero esto no representa un argumento suficiente para decidir si los datos tienen cierta consistencia. Debido a eso, se propone el uso del error medio ( $e_m$ ). Para este caso de estudio  $e_m$  se calcula como sigue:

$$e_m = \frac{\sum_{i=1}^n \text{conf}_R(r_i, s_i)}{n} = \frac{1/6 + 1/6 + 0 + 0 + 1/6 + 1/6}{6} = \frac{4/6}{6} = 0.111$$

Considerando el valor de  $e_m = 0.111$  y de acuerdo a lo que se definió en la Sección 4.6.8.2, tenemos que si  $0.11 < e_m < 0.35$ , los datos generalizados son *similares* a los datos fuente, al menos en cuanto a sus relaciones consistentes se refiere.

Considerando los requerimientos de diferentes usuarios podemos decir que dichos datos podrían ser utilizados por lo que clasificamos como *Usuario Básico* (Ver Tabla 4.6), el cual utiliza los datos generalizados únicamente para visualización sin realizar análisis espacial,

ya que los resultados podrían ser afectados por las inconsistencias. Este grupo excluye a los usuarios encargados de elaborar la cartografía.

Con el propósito de visualizar los resultados se desarrolló una vista en el Software ArcView<sup>82</sup> (Ver Figura 5.18). Proporciona mecanismos para poder visualizar los resultados de cada una de las operaciones que se deben desarrollar para obtener un resultado. Desafortunadamente, los datos deben transformarse a *shapefiles* para manipularse adecuadamente en la interfaz.

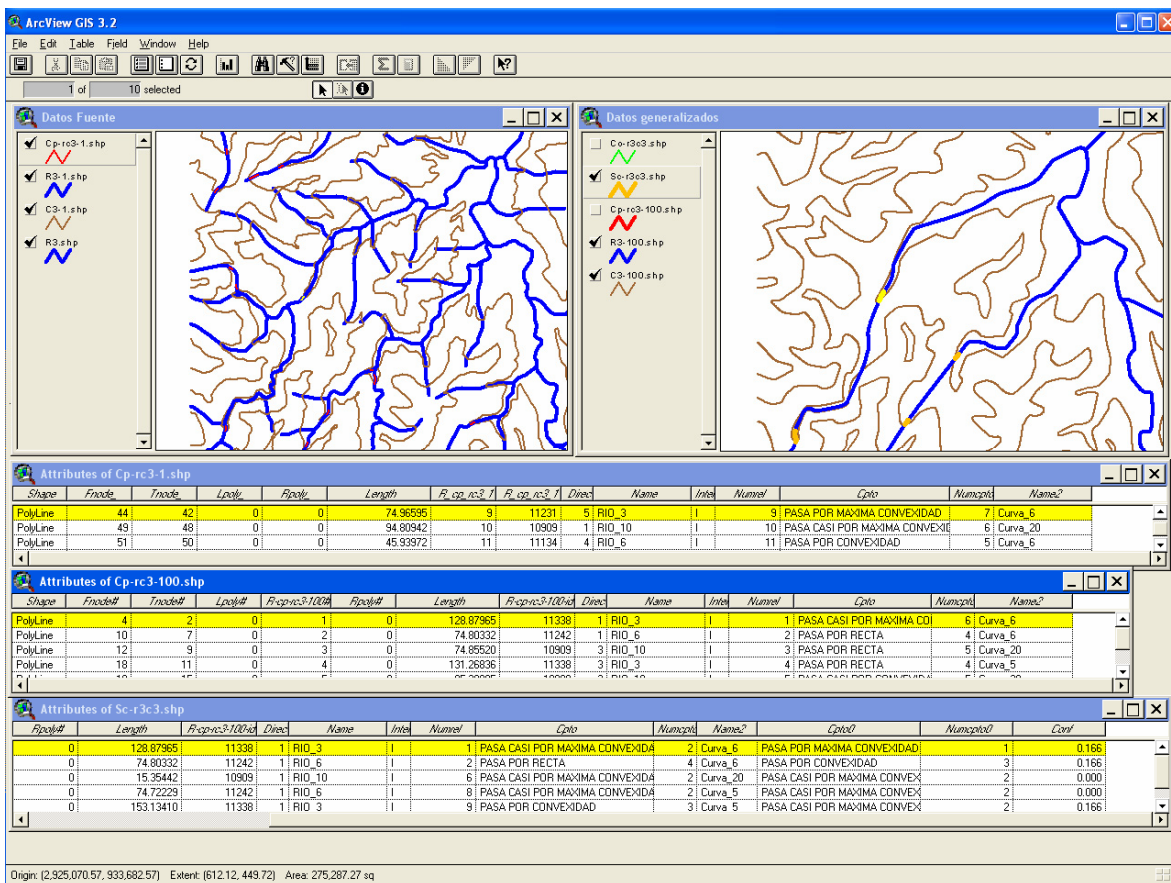


Figura 5.18. Interface en ArcView para visualización de los resultados

En la Figura 5.18 se pueden ver tanto las representaciones gráficas y conceptuales como las de los datos geográficos, lo que permite identificar claramente las ventajas de la metodología propuesta.

<sup>82</sup> Sistema de Información Geográfica de escritorio desarrollado por la compañía ESRI.

Una vez que hemos terminado de presentar los resultados para cada una de las etapas de la metodología propuesta, podemos decir que las mediciones a los datos geoespaciales son una parte importante, ya que las mediciones extraen las relaciones implícitamente representadas. Pero más importante resulta la generación de la representación conceptual y su manipulación para identificar las inconsistencias. Por lo anterior, consideramos que esta metodología puede ser aplicada a otros casos de estudio, en los que probablemente se requiere de modificar las mediciones, añadir relaciones y clases a la ontología, así como también definir los umbrales para definir los casos que se considerarán consistentes.

Podemos ver que la identificación de inconsistencias se basa únicamente en la representación conceptual, lo que representa una diferencia importante con los trabajos previos que utilizan valores numéricos. Adicionalmente, las medidas a nivel global deben ser refinadas para destacar otros aspectos de los datos, tales como el número de relaciones que desaparecen al generalizar los datos fuente.



# CAPÍTULO 6

## CONCLUSIONES Y TRABAJO

### FUTURO

Esta tesis está enfocada hacia la verificación de la consistencia de datos generalizados automáticamente, utilizando una representación conceptual generada a partir de una ontología que se genera con base en una conceptualización del dominio de interés. Adicionalmente, la representación conceptual no depende de la escala, por lo que puede aplicarse para identificar inconsistencias en los datos generalizados. La metodología propuesta se basa en un Sistema de Información Geográfica que provee los mecanismos de análisis espacial, manipulación y visualización de la base de datos espacial. El caso de estudio considerado fue la generalización de redes hidrológicas y curvas de nivel.

#### 6.1 Conclusiones

Las conclusiones del presente trabajo con:

- La descripción de un dominio por parte de un experto representa una conceptualización. Dicha descripción se puede realizar con base en un análisis exhaustivo de las peculiaridades de las redes hidrológicas y las curvas de nivel, la cual permitió caracterizar las relaciones entre ríos y curvas de nivel.
- Utilizando como base una conceptualización, se identificaron los términos relevantes del dominio, sus relaciones, restricciones y propiedades, lo que se utilizó para definir una ontología, apegada a los lineamientos que comúnmente son considerados, además de que se usaron conceptualizaciones existentes. Esto es porque la ontología representa

el conocimiento que tiene un experto o un grupo de expertos sobre un dominio en particular.

- Frecuentemente, las propiedades representadas en ontologías no se expresan como clases, sino como atributos de la clase. En este caso, las propiedades se definen como clases, mejorando la expresividad de la ontología.
- Las relaciones entre los objetos geográficos también fueron modeladas como clases, al igual que las propiedades, con lo que ganamos expresividad en la ontología.
- Se construyó la ontología considerando que es útil para modelar el dominio geográfico, ya que nos resulta útil en la solución de problemas que normalmente se han abordado con un enfoque numérico.
- Como la ontología se diseñó expresamente para un caso en particular, se definieron algunas recomendaciones para extender la ontología. Los aspectos descritos se enfocan a la agregación de nuevas clases, propiedades y relaciones.
- El uso de mediciones permite extraer las propiedades y relaciones que implícitamente poseen los datos geográficos; en este caso se desarrollaron dos medidas: Medición de la relación *>Pasa-por<* y la medición de Sinuosidad. De las mediciones depende en gran parte el éxito de la metodología propuesta. La ontología nos ayuda a identificar qué se debe usar.
- Se utilizó el procesamiento tradicional de los datos geográficos basado en el procesamiento numérico y sintáctico, que sirve de base para realizar el procesamiento semántico de los datos geoespaciales.
- Fue desarrollada una medida específica para medir la relación *>Pasa-por<*, la cual se basa en los componentes geométricos de los datos geoespaciales.



- Con base en las mediciones se obtiene un valor cuantitativo que nos permite identificar una etiqueta que represente cualitativamente a una propiedad o relación. Esto se hace considerando las restricciones que se han definido para pertenecer a una determinada clase.
- Utilizando la ontología se definieron qué clases pueden tener instancias. Aprovechando esto, dichas instancias se representan explícitamente en una base de datos, lo cual significa que la representación explícita representa de forma conceptual a las propiedades y relaciones de los datos geográficos.
- La representación conceptual permite definir una convergencia entre las representaciones, las descripciones y las interpretaciones de los datos geográficos. Esto se debe a que la representación conceptual se basa en los conceptos de la ontología. La ontología representa un medio para representar, describir e interpretar los datos en este caso la interpretación se refiere a la verificación de la consistencia.
- Las tuplas almacenadas en la representación conceptual permiten que cualquier persona pueda entender la semántica de la relación sin ver representación gráfica alguna.
- Adicionalmente, la representación conceptual permite manipular la información geográfica sin considerar aspectos que normalmente se han utilizado en enfoques tradicionales como pares de coordenadas, unidades de medida, ángulos, escala, sistema de referencia, geoide, elipsoide, etc. Además, no importa la dimensión del objeto geográfico, su representación conceptual siempre será igual. Así que, esto difiere de los GIS tradicionales que se basan en primitivas geométricas de representación.
- La consistencia se verifica con base en mediciones de la confusión. La confusión es una medida asimétrica que depende del contexto, la cual tiene como propósito presentar mediante un valor numérico la similitud que existe entre dos conceptos organizados jerárquicamente. Aunque la ontología no posee una estructura jerárquica, puede emplearse la confusión en los fragmentos de la ontología que posean dicha estructura.

- La confusión ha probado ser muy útil para este caso de estudio, ya que nos permite definir la distancia semántica entre dos conceptos. Esto facilita la definición de diversas clases de verificaciones de los datos generalizados.
- Utilizando esta metodología, la verificación de los datos generalizados puede implantarse en cualquier manejador de base de datos, ya sea espacial o no.
- Se desarrollaron medidas para evaluar la similitud a nivel local y global. Con esto, se pudo satisfacer los requisitos de calidad en los datos geográficos para diversos grupos de usuarios.

## 6.2 Contribuciones

Las contribuciones de la presente tesis son:

- Es uno de los primeros trabajos enfocados hacia la realización de análisis espacial a nivel conceptual mediante la identificación de inconsistencias ocasionadas por la generalización, evitando el uso de los aspectos tradicionales de la cartografía.
- Se realizó un análisis exhaustivo de las propiedades que poseen las curvas de nivel y los ríos de una red hidrológica en diferentes escalas (1:50,000 y 1:250,000). Obtenido como principal resultado la caracterización de las relaciones existentes entre ríos y curvas de nivel.
- Se realizó una conceptualización del dominio topográfico con el propósito de generar una ontología que describe las relaciones entre los ríos y las curvas de nivel.
- Se definió una medición para la sinuosidad con base en los componentes del modelo vectorial.
- Se desarrolló la medición de una relación topológica refinada con métricas para cuantificar la relación *>Pasa- por<* entre los ríos y las curvas de nivel.

- Se desarrolló una representación conceptual basada en una conceptualización del dominio topográfico.
- Se desarrolló una metodología enfocada hacia la identificación de inconsistencias en datos generalizados basada en una representación conceptual, evitando el uso de aspectos tales como coordenadas, sistemas de referencia, proyecciones cartográficas, entre otros.
- Se desarrollaron medidas locales y globales para cuantificar la similitud semántica de los datos fuente y los datos generalizados.
- Con base en las medidas locales y globales, se identificaron los requerimientos que deben cumplir los datos generalizados para diversos grupos de usuarios.
- Una conjunto de programas basados en un Sistema de Información Geográfica, que ejecutan la metodología propuesta. En particular el método desarrollado fue implantado en ARC/Info, pero bien podría implantarse en otro GIS.

Se han publicado los siguientes artículos derivados de la presente investigación:

- Moreno-Ibarra, M.: Semantic Similarity Applied to Generalization of Geospatial Data, Second International Conference on Geospatial Semantics (GeoS 2007). Lecture Notes in Computer Science, Springer-Verlag, Berlin Heidelberg New York (2007) (Aceptado para publicación).

En este artículo se presenta de forma detallada la metodología desarrollada en la presente tesis y el caso de estudio analizado en esta tesis. Cabe señalar que este artículo fue arbitrado por tres miembros de un comité internacional de reconocido prestigio en las áreas relacionadas con esta tesis. Además, GeoS 2007 es el congreso más grande que existe en el mundo en lo que se refiere a los aspectos conceptuales y semánticos en GIS.

- Martínez, M., Moreno, M., Torres, M. Levachkine, S.: Representing the semantic content of Topological Relations into Spatial Databases, Proceedings of the Third International Workshop Information on Fusion and Geographic Information Systems (IF&GIS 2007). Lecture Notes in Geoinformation and Cartography, Springer-Verlag, Berlin Heidelberg New York (2007) 223- 233.

Este artículo describe el procedimiento para representar conceptualmente un conjunto de seis relaciones topológicas, mediante un procedimiento de análisis topológico en dos niveles: intrínseco y extrínseco.

- Verastegui, K., Martínez, M. Moreno, M., Levachkine, S., Torres, M.: (2006) Incorporating Semantics into GIS Applications, Lecture Notes in Artificial Intelligence, Vol. 4251. Springer-Verlag, Berlin Heidelberg New York (2006) 698-705.

En este artículo se presenta de forma general la metodología expuesta en esta tesis. Se presenta la generación de la representación conceptual para dos casos de estudio. El primero se basa en las relaciones topológicas definidas por el INEGI. El segundo presenta los fundamentos para generar la representación conceptual de las propiedades geométricas de los datos geospaciales. Cada uno de los casos de estudio presentados se corresponden con los resultados obtenidos en dos trabajos de tesis de maestría.

- Moreno, M., Levachkine, S.: Buscando las Características Cualitativas y Cuantitativas para la Modificación Eficiente de Escala en Capas Vectoriales. Revista Computación y Sistemas, Vol. 9 No 2 (2006) 104-118.

En este artículo se presenta de forma detallada el sistema de generalización expuesto en este trabajo de tesis. Además, se presenta el análisis de las propiedades de los objetos geográficos relacionados con esta tesis.

### 6.3 Trabajo futuro

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

- Uno de los principales obstáculos que presenta el desarrollar ontologías es que no existen conceptualizaciones muy detalladas sobre datos geográficos, lo que dificulta su desarrollo. Considero que se debe realizar un trabajo en conjunto con el INEGI, con el propósito de explotar las ventajas que poseen los sistemas de información geoespaciales que utilizan ontologías.
- Definir la equivalencia entre conceptos que pertenecen a dos ontologías distintas usando las restricciones. Por ejemplo, suponga una ontología que tiene una clase *nada\_sinuoso* con la restricción Sinuosidad = 0 y por otro lado una ontología que tiene una clase *recto* con la restricción Sinuosidad = 0. Si bien *nada\_sinuoso* y *recto* son diferentes, mediante las restricciones podemos deducir que son equivalentes.
- Aplicar la metodología utilizando propiedades y relaciones adicionales para verificar la generalización de otras capas de datos geográficos.
- Desarrollar nuevas relaciones topológicas refinadas con métricas para poder abarcar otros casos de estudio.
- Aplicar la metodología propuesta a otros casos de estudio distintos a la generalización automática. Por ejemplo, medir la similitud entre dos modelos digitales de elevación con base en una representación conceptual de conceptos geomorfológicos [Moreno et al., 2005]. En este caso se puede utilizar una jerarquía que incluya la rugosidad y las formas del terreno. La definición de similitud se realiza midiendo la confusión sobre la jerarquía. Para implantar un método robusto se deben realizar algunos algoritmos adicionales, ya que existen modelos digitales de elevación con distintas resoluciones. Adicionalmente, se podrían generar medidas de similitud para identificar si un DEM en particular cumple con las características necesarias para ser utilizado en una determinada aplicación. También, podría usarse para comparar modelos digitales de

elevación de distintas épocas, permitiendo la evaluación de los cambios que ha sufrido alguna región en particular. De igual forma, podría aplicarse para comparar DEMs elaborados por distintas dependencias como son el INEGI y el CIC-IPN.

- Aprovechar que la representación conceptual es compacta y que la búsqueda espacial se reduce a la consulta en una tabla, para implantar aplicaciones geográficas en dispositivos móviles, los cuales disponen de poco espacio de almacenamiento y tienen una capacidad de procesamiento limitada.
- Aplicar este método para verificar la consistencia semántica de los datos. Por ejemplo, prohibir que una localidad esté sobre el mar.
- Agregar más relaciones y poder contestar consultas sobre una o más relaciones espaciales (*al norte de, más cercana, junto a, a corta distancia por automóvil, rodeado de, más lejano que*, entre otras).
- Utilizar otros casos de estudio para la representación conceptual y mostrar su utilidad en GIS orientados al análisis espacial. Actualmente, se están desarrollando mecanismos de análisis basados en la representación conceptual que resuelvan consultas a respuestas más concretas como, ¿Cuál es el mejor lugar para ubicar una pizzería? y ¿Cuál es el mejor lugar para comprar una casa de acuerdo a mi perfil?
- Debido a que la representación conceptual es muy compacta puede usarse para aplicaciones en dispositivos móviles como PDAs<sup>83</sup>, enfocándose principalmente hacia operaciones de análisis espacial que con los métodos tradicionales requieren de operaciones complejas y grandes volúmenes de datos.
- Desarrollar una medida para evaluar la similitud de estructuras de datos que no posean estructura jerárquica.

---

<sup>83</sup> Personal Digital Assistant, (Ayudante personal digital)

# REFERENCIAS

[AGENT, 1999] Project ESPRIT, AGENT,;: (1999) Selection of Basic Measures, Technical report, AGENT Consortium, URL: <http://agent.ign.fr/>.

[AGENT, 2000] Project ESPRIT, AGENT: (2000) Map Generalisation by Multi-Agent Technology, AGENT Consortium, URL: <http://agent.ign.fr/>.

[Armstrong, 1991] Armstrong, M.P: (1991) 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.

[Bader, 1997] Bader, M.: (1997) Methoden zur Erkennung und Lösung von metrischen Konflikten in der Polygongeneralisierung, Master’s thesis. Department of Geography, University of Zurich.

[Bader et al., 1999] Bader, M., Barrault, M., Regnaud, N., Mustiere, S., Duchene, C., Ruas, A., Fritsch, E., Lecordix, F. and Barillot X.: (1999) AGENT Workpackage D2 - Selection of Basic Algorithms, Technical report, AGENT Consortium.

[Bader y Barrault, 2000] Bader, M. and Barrault, M.: (2000) Improving Snakes for Linear Feature Displacement in Cartographic Generalization, In Proceedings 5th International Conference on GeoComputation, Manchester, U.K.

[Bader, 2001] Bader, M.: (2001) Energy Minimizing Methods for Feature Displacement in Map Generalization, Ph.D. thesis. University of Zurich, Department of Geography  
URL: <http://www.geo.unizh.ch/gis/mbader/>,

[Barkowsky, 1997] Barkowsky, T.: (1997), Cognitive requirements on making and interpreting maps. In S. Hirtle, A. Frank (Eds.), Spatial information theory: A theoretical basis for GIS. Proceedings COSIT 97. Springer-Verlag, Berlin, Heidelberg, 1329, 347–361.

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

[Beard, 1991] Beard, K.: (1988) Multiple representations from a detailed database: a scheme for automated generalization, Ph.D. thesis. University of Wisconsin-Madison.

[Bigus y Bigus, 2001] Bigus, J. and Bigus, J.: (2001) Intelligente Agenten mit Java programmieren. eCommerce und Informationsrecherche automatisieren. Addison-Wesley, Munich, Germany.

- [Brassel y Weibel, 1988] Brassel, K. and Weibel, R.: (1988) A Review and Conceptual Framework of Automated Map Generalization. *International Journal of Geographic Information Systems*, 2(3): 229- 244.
- [Burghardt, 2000] Burghardt, D.: (2000) Automatisierung der kartographischen Verdrängung mittels Energieminimierung, Ph.D. thesis. Technical University of Dresden, Department of Geodesy.
- [Buttenfield, 1991] Buttenfield, B.P.: (1991) Map Generalization. Making Rules for Knowledge Representation. In McMaster R.B. (ed). Longman, London, UK.
- [Campbell, 1993] Campbell, J: (1993): Map use and analysis; 2nd edition. Wm.C.Brown Publishers, USA, 428P.
- [Carnap, 1956] Carnap, R.: (1956) Meaning and Necessity, 2nd edition. Chicago: University of Chicago Press, USA.
- [Chavez y Maes, 1996] Chavez, A. and Maes, P.: (1996) Kasbah, An Agent Marketplace for Buying and Selling Goods. Proceedings 1st International Conference on the Practical Application of Intelligent Agents and Multi-Agent Technology, London, U.K., 75–90.
- [Chrisman, 1983] Chrisman, N. R.: (1983) Epsilon Filtering: A technique For Automated Scale Changing. 43<sup>rd</sup> annual ACSM Meeting, Washington D.C., 322-331
- [Clancey, 1993] Clancey, W. J.: (1993) The Knowledge Level Reinterpreted: Modelling Socio-Technical Systems. *International Journal of Intelligent Systems*, 8, 33-49.
- [Cocchiarella, 1991] Cocchiarella, N. B.: (1991) Formal Ontology. In H. Burkhardt and B. Smith (eds.). *Handbook of Metaphysics and Ontology*. Philosophia Verlag, Munich: 640-647.
- [Cook y Jones, 1989] Cook, A. and Jones, C.:(1989) Rule-based cartographic name placement with prolog. Proceedings AutoCarto 9, Baltimore, USA, 231–240.
- [Cromley, 1991] Cromley, R. G.: (1991) Hierarchical Methods to Line Simplification. *Cartography and Geographic Information Systems*, 18(2), 125-131.
- [Deveau, 1985] Deveau, T.J.: (1985) 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.
- [Douglas y Peucker, 1973] Douglas, D. and Peucker T,: (1973) 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.



[Duchêne et al., 2002] Duchêne, C., Regnauld, N., Ruas, A.: (2002) *Le modèle AGENT*, dans : Ruas A. (sous la direction de). Généralisation et représentation multiple. Hermès Science Publications, Paris, France, 369-383.

[Eco, 1976] Eco, U: (1976) A Theory of semiotics. Bloomington, Ind.: Indiana University Press, USA.

[Edwardes y Mackaness, 2000] Edwardes, A. and Mackaness, W. (2000) 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),: Taylor&Francis, London, UK, 161–173.

[Egenhofer, 1989] Egenhofer, M.: (1989) A Formal Definition of Binary Topological Relationships, Third International Conference on Foundations of Data Organization and Algorithms (FODO), Paris, France, Lecture Notes in Computer Science, Springer-Verlag, Berlin, Germany, 367, 457-472.

[Egenhofer, 1990] Egenhofer, M.: (1990) Interaction with geographical information systems via Spatial queries. Journal of Visual Languages and Computing, 1(4), 389-413.

[Egenhofer y Franzosa, 1991] Egenhofer, M., Franzosa, R.: (1991) Point-set topological relations. Internationan Journal of Spatial Information Systems, 5, 161-174.

[Egenhofer et al., 1994] Egenhofer, M., Mark, D. and Herring, J.: (1994) 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, USA.

[Egenhofer y Mark, 1995] Egenhofer, M. and D. Mark.: (1995) Naive geography. In Frank and W. Kuhn (eds). Spatial Information Theory: a theoretical basis for GIS. Lecture Notes in Computer Science, Springer-Verlag, Berlin, Germany, 988, 1-16.

[Enderton, 1972] Enderton, H. B.: (1972) A Mathematical Introduction to Logic. San Diego, CA: Academic Press, USA.

[ESRI, 1999] Environmental Systems Research Institute (ESRI): (1999), Understanding GIS: The ARC/INFO Method, ESRI Press, USA.

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

[Fonseca, 2001] Fonseca, F.: (2001) Ontology-Driven Geographic Information Systems, PhD Thesis. The University of Maine, Department of Spatial Information Science and Engineering, USA.

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

- [Galanda, 2003] Galanda, M.: (2003) Automated Polygon Generalization in a Multi Agent System, PhD Thesis. University of Zurich, Department of Geography, Switzerland.
- [Genesereth y Nilsson, 1987] Genesereth, M. R. and Nilsson, N. J.: (1987) Logical Foundation of Artificial Intelligence. Morgan Kaufmann, Los Altos, California.
- [Gruber, 1993] Gruber, T.: (1993) Toward principles for the design of ontologies used for knowledge sharing. *International Journal of Human-Computer Studies*, Special issue: The Role of Formal Ontology in the Information Technology, 43(5-6), 907-928.
- [Guarino, 1995] Guarino, N.: (1995) Formal ontology, conceptual analysis and knowledge representation. *International Journal of Human-Computer Studies*, Special issue: The Role of Formal Ontology in the Information Technology, 43(5-6), 625-640.
- [Guarino, 1998] Guarino, N.: (1998) Formal Ontology and Information Systems. In: N. Guarino(Ed.). *Formal Ontology in Information Systems*. IOS Press, Amsterdam, Netherlands. 3-15.
- [Harrie, 1999] Harrie, L.: (1999) The Constraint Method for Solving Spatial Conflicts in Cartographic Generalization. *Cartography and Geographic Information Science*, 26(1), 55-69.
- [Herbert y Joao, 1992] Herbert, G. and Joao, E.: (1991) Automating map design and generalisation: A review of systems and prospects for future progress in the 1990's. SERRL Working Report 27", Technical report, Department of Geography, Birkbeck College.
- [Head, 1991] Head, C. G.: (1991) Mapping as language or semiotic system: Review and comment. In D. M. Mark & A. U. Frank (Eds.). *Cognitive and linguistic aspects of geographic space*. Dordrecht, Boston, London: Kluwer Academic Publishers, 237-262.
- [Hiperdictionary] Online dictionary, URL: <http://www.hyperdictionary.com>., accesado en: Diciembre de 2005.
- [Hoffman y Singh, 1997] Hoffman, D.D., and Singh, M. (1997) Saliency of visual parts. *Cognition*, 63, 29-78.
- [Højholt, 2000] Højholt, P.: (2000) Solving Space Conflicts in Map Generalization: Using a Finite Element Method. *Cartography and Geographic Information Science*, 27(1), 65-74.
- [ICA, 1973] International Cartographic Association: (1973) *Multilingual Dictionary of Technical Terms in Cartography*, Steiner, Wiesbaden.
- [Imhof, 1972] Imhof, E.: (1972) *Thematische Kartographie*, Berlin, New York: Walter de Gruyter.

[INEGI, 1993] Instituto Nacional de Estadística, Geografía e Informática (1993), Base de datos geográficos. Modelo de datos Vectorial. Instituto Nacional de Estadística, Geografía e Informática (INEGI), México.

[INEGI, 1995] Instituto Nacional de Estadística, Geografía e Informática: (1995), Base de datos geográficos. Diccionario de datos topográficos escala 1:250,000, Instituto Nacional de Estadística, Geografía e Informática(INEGI), México.

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

[INEGI, 1997] Instituto Nacional de Estadística, Geografía e Informática, Base de datos geográficos. Diccionario de datos topográficos escala 1:1,000,000. Instituto Nacional de Estadística, Geografía e Informática (INEGI), México.

[ISO, 2005] International Organization for Standardization: (2005) Information Technology —Environmental Data Coding Specification (EDCS), International Standard ISO/IEC 18025:2005(E),  
URL: [http://standards.iso.org/ittf/PubliclyAvailableStandards/ISO\\_IEC\\_18025\\_Ed1.html](http://standards.iso.org/ittf/PubliclyAvailableStandards/ISO_IEC_18025_Ed1.html)

[Jaakkola, 1998] Jaakkola, O. (1998) Multi-Scale Categorical Databases with Automatic Generalization Transformations Based on Map Algebra. *Cartography and Geographic Information Systems*, 25(4), 195–207.

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

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

[Kilpeläinen, 1995] Kilpeläinen, T., and Sarjakoski, T.: (1995) Incremental generalization for multiple representations of geographical objects. *GIS and Generalization: Methodology and Practice*. In: J.C. Muller, J.P. Lagrange, R. Weibel. Bristol(eds). Taylor & Francis, 209-218.

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

[Kufoniyi, 1995] Kufoniyi, O.: (1995) Spatial Coincidence Modelling, Automated Databases Updating and Data Consistency in Vector GIS, Enschede: International Institute for Aerospace Survey and Earth (ITC).

[Kulik et al., 2005] Kulik, L., Duckham, M., and Egenhofer, M.: (2005) Ontology-driven map generalization. *Journal of Visual Languages and Computing*, 16 (3): 245-267.

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

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

[Le Men, 1996] Le Men H. : (1996) Generalisation cartographique pour l'occupation du sol: application au passage V1 'a V2 pour la BD Carto R\_”, Revue internationale de géomatique, 6(2-3), 227–248.

[Levachkine y Guzman-Arenas, 2004] Levachkine S., y Guzman-Arenas, A.: (2004). Hierarchies measuring qualitative variables. Lecture Notes in Computer Science, Springer-Verlag, 2945, 262-274.

[Levachkine y Guzman-Arenas, 2007] Levachkine, S., Guzmán-Arenas, A.: (2007) Hierarchy as a new data type for qualitative variables. Expert Systems with Applications: An International Journal, 32(3), 899-910.

[Levesque, 1984] Levesque, H. J.: (1984). Foundations of a functional approach to knowledge representation. Artificial Intelligence, 23, 155-212.

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

[Loibl y Totzer, 2002] Loibl, W. and Totzer, T.: (2002) Growth and Densification Processes in Suburban Landscapes - a Spatial Agent – Simulation. Proceedings 5th AGILE Conference on Geographic Information Science, Mallorca, Spain, 23–30.

[Mark et al., 1989] Mark D.M. and Csillag F.: (1989) The Nature of Boundaries on ‘Area Class’ Maps”, Cartographica, 26(1), 65–77.

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

[McMaster, 1986] McMaster, R.B.: (1986) A statistical analysis of mathematical measures of linear simplification. The American Geographer, 13 (2), 103-116.

[McMaster y Shea, 1992] McMaster, R., Shea, K. S.: (1992), Generalization in Digital Cartography, Association of American Geographers, U.S.A.

[Molenaar, 1989] Molenaar, M.: (1989) Single valued vector maps – a concept in GIS, Geo-Informationssysteme, 2, 18-26.

[Molenaar *et al.*, 1994] Molenaar, M., Kufoniya, O. and Bouloucos, T.: (1994) Modelling topological relationships in vector maps, *Advances in GIS Research*, pages 112-126, 1994.

[Molenaar, 1998] Molenaar, M.: (1998) *An Introduction to the Theory of Spatial Object Modelling for GIS*, Department of Geo-Informatics, International Institute for Aerospace Survey and Earth Science, Enschede, Netherlands.

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

[Moreno, 2001] Moreno, M., (2001) *La Generalización Automática de la Información Geográfica Multiescala*, Tesis de Maestría. Centro de Investigación en Computación, México.

[Moreno *et al.*, 2002] Moreno, M., Torres M., Levachkine, S.: (2002) Automatic Modification In Scale of Hydrological Networks, Twenty Second Annual ESRI International User Conference, ESRI, San Diego, California, USA.  
e-Proceedings. URL: <http://gis.esri.com/library/userconf/proc02/pap0173/p0173.htm>

[Moreno *et al.*, 2005] Moreno, M., Levachkine, S., Torres, M., Quintero, R., Guzmán, G.: (2005) Automatic Geomorphometric Analysis for Digital Elevation Models. *Lecture Notes in Computer Science*, 3684, 374 – 381.

[Müller *et al.*, 1995] Müller, J.C., Weibel, R., Lagrange, J.P. and Salgé F.: (1995) “Generalization: State of the Art and Issues”, in Müller J.C., Lagrange J.P. and Weibel R. (eds.), *GIS and Generalization. Methodology and Practice*. GISDATA 1, Taylor&Francis, 3–17.

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

[Nedas *et al.*, 2007] Nedas, K., Egenhofer, M., Wilmsen D.: (2007) Metric Details of Topological Line-Line Relations. *International Journal of Geographical Information Science*, 21(1), 21–48.

[Newell, 1982] Newell, A.: (1982) The knowledge level. *Artificial Intelligence*, 18(1), 87-127.

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

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

- [Noy y McGuinness, 2001] Noy, N. F., McGuinness D. L.: (2001) *Ontology Development 101: A Guide to Creating Your First Ontology*. 2001. D.L. SMI technical report SMI-2001-0880. URL:  
[http://protege.stanford.edu/publications/ontology\\_development/ontology101.html](http://protege.stanford.edu/publications/ontology_development/ontology101.html).
- [Nutter, 1987] Nutter, J. T.: (1987) *Epistemology*. In: S. Shapiro (ed.). *Encyclopedia of Artificial Intelligence*. John Wiley, USA.
- [OGC, 1999] Open Geospatial Consortium, Inc. (1999) *Abstract Specification Overview*, <http://www.opengeospatial.org/>.
- [OGC, 2005] Open Geospatial Consortium, Inc. (2005) *OpenGIS Reference Model*, <http://www.opengeospatial.org/>.
- [O'Hare y Jennings, 1996] O'Hare, G.M.P. and Jennings, N.R. (eds) (1996) *Foundations of Distributed Artificial Intelligence*. John Wiley & Sons, Inc, New York USA.
- [Papadias y Theodoris, 1997] Papadias, D. and Theodoris, Y.: (1997) *Spatial relations, minimum bounding rectangles, and spatial data structures*. *International Journal of Geographical Information Science*, 11(2), 111-138.
- [Plazanet et al., 1998] Plazanet, C., Bigolin, M. and Ruas, A.: (1998) *Experiments with Learning Techniques for Spatial Model Enrichment and Line Generalization*. *GeoInformatica*, 2(4), 315–333.
- [Poli, 1996] Poli, R.: (1996) *Ontology and Knowledge Organization*. In *Proceedings of 4th Conference of the International Society of Knowledge Organization (ISKO 96)*. Washington, USA.
- [Ratajski, 1967] Ratajski, L.: (1967) *Phénomènes des points de généralization*. *International Yearbook of Cartography*, 7, 143-151.
- [Regnauld, 2001] Regnauld, N.: (2001) *Contextual Building Typification in Automated Map Generalization*. *Algorithmica*, 30(2), 312-333.
- [Reichenbacher, 1995] Reichenbacher, T.: (1995) *Eine Methode für den Wissenserwerb in der kartographischen Generalisierung durch Interaktionsaufzeichnung und induktives Lernen*, Master's thesis. University of Zurich, Department of Geography.
- [Rosgen, 1994] Rosgen, D. L.: (1994) *A Classification of Natural Rivers*. *Catena*, Elsevier 22(3), 169-199.
- [Ruas, 1995] Ruas, A.: (1995) *Multiple paradigms for automated map generalization: Geometry, topology, hierarchical partitioning and local triangulation*. *Proceedings AutoCarto 12*, Charlotte, USA, 69–78.

[Ruas y Plazanet, 1996] Ruas, A. and Plazanet, C.: (1996) Strategies for Automated Generalization. Proceedings 7<sup>th</sup> International Symposium on Spatial Data Handling (Advances in GIS Research II), Delft: Taylor&Francis, 6.1–6.17.

[Ruas, 1999] Ruas, A. : (1999) Modèle de généralisation de données géographiques `a base de constraints et d'autonomie, Ph.D. thesis. Université de Marne-la-Vallée.

[Ruas, 2000] Ruas, A.: (2000) The Roles of Meso Objects for Generalisation, Proceedings 9th Symposium on Spatial Data Handling, Beijing, China, 3b50–3b63.

[Salitchev, 1972] Salitchev, K.A. (1972) Cartografía (Traducción del Ruso), Editorial del pueblo, Cuba.

[Shea y McMaster, 1989] Shea, K.S., McMaster, R.B.: (1989) Cartographic Generalization in a Digital Environment: When and How to Generalize, Proceedings, Ninth International Symposium on Computer-Assisted 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.

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

[Smith, 1995] Smith, B.: (1995) Formal Ontology, Commonsense and Cognitive Science. International Journal of Human-Computer Studies, Special issue on Formal Ontology, Conceptual Analysis and Knowledge Representation, 43, 641-667.

[Spiess, 1990] Spiess, E.: (1990) Generalisierung in thematischen Karten, In: Swiss Society of Cartography (ed.), Kartographisches Generalisieren, Baden: SSC, Kartografische Publikationsreihe, 10, 63–70.

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

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

[Tversky, 1981] Tversky, B.: (1981) Distortions in memory for maps. Cognitive Psychology, 13, 407-433.

[Ubeda y Egenhofer, 1997] Ubeda, T. and Egenhofer, M.: (1997) Topological Errors Correcting in GIS. Advances in Spatial Databases, Fifth International Symposium on Large Spatial Databases, SSD '97. Lecture Notes in Computer Science, Springer-Verlag, 1262, 283-297.

[Verastegui, 2007] Verastegui, K.: (2007) Conceptualización de Aspectos Geométricos en el Contexto de Datos Geoespaciales, Tesis de Maestría. Centro de Investigación en Computación, México.

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

[Merrian-Vebster, 2007] Dictionary, URL: <http://www.merriam-webster.com>, accesado en: Enero 2007.

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

[Weibel, 1995] Weibel, R.: (1995) 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.

[Weibel, 1996] Weibel, R.: (1996) A Typology of Constraints of Line Simplification. Proceedings 7th International Symposium on Spatial Data Handling (Advances in GIS Research II), Taylor&Francis, Delft, The Netherlands, 9A.1–9A.14.

[Weibel y Dutton, 1998] Weibel, R. and Dutton, G.: (1998) Constraint-based Automated Map Generalization. Proceedings 8th International Symposium on Spatial Data Handling, 214–224.

[Weiss, 1999] Weiss, G. (ed.) (1999) Multiagent Systems. A modern Approach to Distributed Artificial Intelligence, Cambridge, Massachusetts: The MIT Press, USA.

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

[Wikipedia, 2007] Enciclopedia libre, URL: <http://es.wikipedia.org/>, accesado en: junio 2007.

[Worboys, 1996] Worboys M.F.: (1996). Metrics and topologies for geographic spatial. Advances in GIS research II : proc. of 7th International Symposium on Spatial Data Handling, Kraak and Molenaar (eds), Taylor and Francis, 365-375.

[Wooldridge, 1999] Wooldridge, M.: (1999) Intelligent Agents. In: Weiss G. (ed.), Multiagent Systems. A modern Approach to Distributed Artificial Intelligence, MIT Press, USA, 27–78.

[WordNet, 2007] Semantic lexicon, URL: <http://wordnet.princeton.edu/>, accesado en: mayo de 2007.

[Yaolin, 2002] Yaolin, L.: (2002) Artificial Neural Network Method for Evaluating Basic Land Price. Journal of Geo-Information Science, 4(4), 1-8.

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



# ANEXO I CÓDIGO FUENTE

```
/* alturas.aml
/* Este programa asigna altitudes a los nodos de una cobertura de rios
/*A----- Arguments -----
/*

&severity &error &routine bailout
&ECHO &OFF
&args cover topo tolalt

    &call chkargs
    &call existe1
    &call existe2
    &type -----
    &type   Iniciando asignación de alturas a los nodos
    &type -----
    &call asignacion

/* &call ordenar
&MESSAGES &POPUP
&type ----- ASIGNACION DE ALTURAS TERMINADA -----
----
&messages &off
&messages &on

&return

&routine chkargs
/*-----
&s cover = [UPCASE %cover%]
/*&s = buildaat.FALSE.
&if [NULL %cover%] &then &do
    &call usage
    &call bailout
&end
&else &do
    &if not [EXISTS %cover% -COVER] &then &do
        &type esta cobertura no existe---%cover% -----
        &call bailout
    &end
&end
&return

&routine asignacion
/*****
    &type
    &type --- INICIANDO ASIGNACION ---
    &type
    additem %cover%-A.Pat %cover%-A.Pat NODE# 4 5 B;
    &type
    &type --- Creando la tabla nal(node arc list) ---
    &type
    &r buildnal %cover%
    &type
```

```

&type --- Nal(node arc list) Terminada ---
&type

&type
&type --- Ordenando Indices ---
&type
&r info1 %cover%-A
&type
&type --- Indices Ordenados ---
&type

/* une la tabla que contiene la lista de las
/* ALTURAS PARA CADA NODO CON LA LISTA DE NODOS
&type
&type --- uniendo nal con alturas ---
&type
joinitem %cover%.nal %cover%-A.pat %cover%.nal NODE# NODE#
&type
&type --- unido nal con alturas ---
&type

/* ELIMINA LOS ATRIBUTOS QUE NO SE REQUIEREN

&sv cov [translate [entryname %cover%]]

&type
&type --- eliminando atributos no requeridos ---
&type

DROPIITEM %cover%.nal %cover%.nal AREA;
DROPIITEM %cover%.nal %cover%.nal PERIMETER;
DROPIITEM %cover%.nal %cover%.nal %cov%-A#;
DROPIITEM %cover%.nal %cover%.nal %cov%-A-ID;

/*SE CREAN 2 TABLAS Y SE
/* AÑADE UN NUEVO ATRIBUTO A C/U

&type
&type --- añadiendo atributos de altura inicial y final ---
&type

additem %cover%.nal %cover%-I.DAT FNODE# 4 5 B
additem %cover%.nal %cover%-F.DAT TNODE# 4 5 B

/* SE ASIGNA EL VALOR DE LOS NODOS ALOS ATRIBUTOS AÑADIDOS
&r info2 %cover%-I %cover%-F

/* SE ELIMINAN LOS ATRIBUTOS NO EMPLEADOS
&TYPE %cover%-F.DAT %cover%-I.DAT
DROPIITEM %cover%-F.DAT %cover%-F.DAT ARC#
DROPIITEM %cover%-F.DAT %cover%-F.DAT NODE#
DROPIITEM %cover%-I.DAT %cover%-I.DAT ARC#
DROPIITEM %cover%-I.DAT %cover%-I.DAT NODE#

&sv exis = [exists %cover%-AN -cover]
&if not %exis% &then
&do
&type
&copy %cover% %cover%-AN
&end

```

```

&else
  &do
    &if %exis% &then
      &type
      &type ---- La cobertura ya existia se eliminará
      &type
      kill %cover%-AN all
      copy %cover% %cover%-AN
    &end
/* SE UNEN LAS TABLAS QUE CONTIENEN LAS ALTURAS DE LOS NODOS
/* CON LA TABLA DE LOS ARCOS
&type --- AÑADIENDO ALTURAS A NODOS FINALES
&sv cov [translate [entryname %cover%]]
JOINITEM %cover%-AN.AAT %cover%-I.DAT %cover%-AN.AAT FNODE# %cov%-AN-ID
&type --- AÑADIENDO ALTURAS A NODOS INICIALES
JOINITEM %cover%-AN.AAT %cover%-F.DAT %cover%-AN.AAT TNODE# %cov%-AN-ID

&return

&routine ordenar

  AE
  &STATION 9999
  EDIT %COVER%-AN
  DE ARCS ARROW
  DRAW
  EF ARCS
  SELECT ALL
  SELECT FOR I-ALTURA < F-ALTURA
  DRAWSELECT
  FLIP
  &sv exis = [exists %cover%-AC -cover]

&if not %exis% &then
  &do
    SAVE %COVER%-AC
  &end
&else
  &do
    &if %exis% &then
      &type
      &type ---- La cobertura ya existia se eliminará
      &type
      kill %cover%-AC YES
      SAVE %COVER%-AC

    &end

  Q
&return

/*-----
&routine existel
/*-----
&sv exis = [exists %cover%-B -cover]
&if not %exis% &then
  &do
    &call buff
  &end
&else
  &do
    &if %exis% &then
      &type

```

```

    &type ---- La cobertura ya existia se eliminará
    &type
    kill %cover%-B all
    &call buff
    &end
&return

/*-----
&routin existe2
/*-----
&sv exis = [exists %cover%-A -cover]
&if not %exis% &then
    &do
        &call alt
    &end
&else
    &do
        &if %exis% &then
            &type
            &type ---- La cobertura ya existia se eliminará
            &type
            kill %cover%-A all
            &call alt
        &end
    &end
&return

/* -----
&routin alt
/* -----
clip %topo% %cover%-B %cover%-A poly 0.00000000001
&sv cov [translate [entryname %topo%]]
DROPIITEM %cover%-a.pAT %cover%-a.pAT %cov%#
DROPIITEM %cover%-a.pAT %cover%-a.pAT %cov%-id
&return

/* -----
&routin buff
/* -----
buffer %cover% %cover%-B # # 0.0000005 0.00000001 node
&return

/*-----
&routin usage
/*-----
&type Usage &r alturas <covertura> -- Sintaxis Correcta
&return /*

/*-----
&routin bailout
/*-----
&severity &error &ignore
/*&call exit
&return &error Error ...

/* correccion.aml
/* Este programa corrige la dirección del flujo en una red de ríos
/* cover - El nombre de la cobertura que se le asignaran altitudes

&severity &error &routin bailout

```

```

&ECHO &OFF
&args cover grid
&sv cov [translate [entryname %cover%]]
&sv co = [before %cov% -]
&sv .c [translate [entryname %cover%]]
&sv .g [translate [entryname %grid%]]
&call ordenar
q
n

&MESSAGES &POPUP
&type ----- CORRECCION DE ALTURAS TERMINADA -----
&messages &off
&messages &on
&return

&routine draw
    DE ARCS ARROW
    ap vistaDEM
    DRAW
&return

&routine ordenar

    AE
    &STATION 9999
    EDIT %COVER%
    mape %cover%
    &call draw
    EF ARCS
    SELECT ALL
    &sv total = [show number select]
    &type El numero total de arcos es %total%
    &sv index = 1
    &type ----- Iniciando correccion automática -----
    SELECT FOR I-ALTURA < F-ALTURA
    DRAWSELECT
&MESSAGES &POPUP
&type ----- CORRECCION AUTOMATICA -----
&messages &off
&messages &on
    FLIP
    Draw
ef arcs
&MESSAGES &POPUP
&messages &off
&messages &on
&do &until %index% eq %total%
    &type arco %index% de %total%
    select %cov%# = %index%
    &sv ini = [show ARC %index% ITEM I-ALTURA]
    &sv fin = [show ARC %index% ITEM F-ALTURA]
    &TYPE ALTURA INICIAL %INI%, ALTURA FINAL %FIN%
    &IF %INI% = %FIN% &THEN
        &do
            drawselect
            &type indice %index%
            &sv num-ver = [show ARC %index% NPNTS]
            &sv nodo = [show ARC %index% Vertex %num-ver% ]
            &sv nodo1 = [before %nodo% ,]
            &sv nodo2 = [after %nodo% ,]
            &type %nodo% y X%nodo1% y Y%nodo2%
            &sv nodo = [show ARC %index% Vertex 1]

```

```

&sv nodo3 = [before %nodo% ,]
&sv nodo4 = [after %nodo% ,]
&sv d1 = [before %nodo1% .]
&sv d2 = [before %nodo2% .]
&sv d3 = [before %nodo3% .]
&sv d4 = [before %nodo4% .]
&if %d1% > %d3% &then
  &do
    &sv pa = %d3%
    &sv d3 = %d1%
    &sv d1 = %pa%
  &end
&if %d2% > %d4% &then
  &do
    &sv pa = %d4%
    &sv d4 = %d2%
    &sv d2 = %pa%
  &end
&sv ventana = [show window Corregir-arco-%index% exists]
&IF %ventana% = .FALSE. &then
  &do
    &type nodo %d1% - %d2% - %d3% - %d4%      f
    &type windows create Corregir-arco-%index% %d1% %d2% %d3% %d4%
    windows create Corregir-arco-%index% %d1% %d2% %d3% %d4%
  &end
  &else
    &do
      &type nodo %d1% - %d2% - %d3% - %d4%
      &type windows extent Corregir-arco-%index% %d1% %d2% %d3% %d4%
      windows extent Corregir-arco-%index% %d1% %d2% %d3% %d4%
    &end
&MESSAGES &POPUP
&s selecc [getchoice 'Invertir arco' Continuar Cancelar]
&messages &off
&messages &on
&IF %selecc% = 'Invertir arco' &then
  &do
    flip
    windows destroy Corregir-arco-%index%
    &sv index = %index% + 1
  &end
  &else
    &IF %selecc% = 'Continuar' &then
      &do
        &type ----- Arco Siguiente -----
        &sv index = %index% + 1
        draw
      &end
    &else
      &IF %selecc% = 'Cancelar' &then
        &do
          &sv index = %total%
          &type ----- Cancelar -----
        &end
      &end
    &end
  &else
    &do
      &TYPE ARCO %index% --- OK
      &sv index = %index% + 1
    &END
    &type -----
  &end
&end

```

```

&MESSAGES &POPUP
&s selecc [getchoice Guardar Cancelar]
&messages &off
&messages &on
&type %cov%
&IF %selecc% = 'Guardar' &then
  &do
    &sv exis = [exists %co%-AC -cover]
    &if not %exis% &then
      SAVE %CO%-AC
    &else
      &do
        &if %exis% &then
          &type ---- La cobertura ya existia se eliminará
          kill %co%-AC YES
          &type ---- Guardando Correccion de alturas -----
          SAVE %CO%-AC
        &end
      &end
    &else
      &IF %selecc% = 'Cancelar' &then
        &type
          &type ---- CANCELO CORRECCION -----
&return

/* funcion.aml
/* Coloca el atributo funcion y define su utilidad dentro del sistema
/*
/*A----- Arguments -----
/*
/* cover - El nombre de la cobertura que se le asignaran altitudes

&severity &error &routine bailout
&ECHO &OFF
&args cov

  &sv cove [translate [entryname %cov%]]
  &sv cover = [before %cove% -]
  &type %cover%
/* llama a las funciones
&call chkargs
&call existe

  &MESSAGES &POPUP
  &type ----- Funcion Hecho
  &messages &off
  &messages &on

&return

/* -----
&routine existe
/* -----
&sv exis = [exists %cover%-f -cover]
&if not %exis% &then
  &do
    copy %cover%-s %cover%-f
    ADDITEM %cover%-f.aat %cover%-f.aat funcion 2 3 C
    &call nodosrep
  &end

```

```

&else
  &do
    &if %exis% &then
      &type
      &type ---- La cobertura ya existia se eliminará
      &type
      kill %cover%-f all
      copy %cover%-s %cover%-f
      /* build %cover%-f
      &type %cover%-f
      ADDITEM %cover%-f.aat %cover%-f.aat funcion 2 3 C
      &call nodosrep
    &end
  &return

/* -----
&routine nodosrep
/* -----
  &s infofile [translate [entryname %cover%-f]].AAT
  /*&r maxid %cover%-f arc FNODE#
  &r max-min-id %cover%-f arcs FNODE#
  &sv .maxid = %.max%
  &sv i
  &sv ii
  ae
  &station 9999
  edit %cover%-f
  de arcs
  ef arcs
  &messages &off &info

  select all
  &type hecho p
  calculate funcion = 'P'
  nselect
  save
  &type %.maxid%

  &do i = 1 &to %.maxid% &by 1
    SELECT TNODE# = %i%
    &if [show number select] = 0 &then
      &do
        SELECT FNODE# = %i%
        &type El arco %i% es de Entrada
        calculate funcion = 'I'
      &end
    SELECT FNODE# = %i%
    &if [show number select] = 0 &then
      &do
        SELECT TNODE# = %i%
        &type El arco %i% es de Salida
        calculate funcion = 'O'
      &end

  &end
  save
  &messages &on
  draw
  q

&return

/*-----

```



```
&routine chkargs
/*-----
  &s cover = [UPCASE %cover%]
  &if [NULL %cover%] &then &do
    &call usage
    &call bailout
  &end
  &else &do
    &if not [EXISTS %cover% -COVER] &then &do
      &type Esta Cobertura no existe ---%cover% -----
      &call bailout
    &end
  &end
&end
&return

/*-----
&routine usage
/*-----
&type Usage &r funcion <cobertura> -- Esta es la sintaxis correcta
&return /*

/*-----
&routine bailout
/*-----
&severity &error &ignore
/*&call exit
&return &error Error ...

/*subsis1.aml
/* Asigna, en la cobertura de los rios, el subsistema al que pertenecen los arcos
/*A----- Arguments -----
/* cover - El nombre de la cobertura
/*

&severity &error &routine bailout
&ECHO &OFF
&args cov tama

/* llama a las funciones

&sv cove [translate [entryname %cov%]]
&sv cover = [before %cove% -]
&call chkargs
&call existe3
&MESSAGES &POPUP
&type ----- Subsistemas Hecho -----
&messages &off
&messages &on

&return
/*-----
&routine existe3
/*-----
&sv exis = [exists %cover%-bb -cover]

&if not %exis% &then
  &do
    &type
    &type ---- Haciendo buffer a %cover%-bb
    &type
    buffer %cover% %cover%-bb # # %tama%
```

```

        &call sys
    &end
&else
    &do
        &if %exis% &then
            &type
            &type ---- La cobertura ya existia se eliminará
            &type
            kill %cover%-bb all
            &type
            &type ---- Haciendo buffer a %cover%-bb
            &type
            buffer %cover% %cover%-bb # # %tama%
            &call sys
        &end
    &return

/*-----
&routine chkargs
/*-----
    &s cover = [UPCASE %cover%]
    &if [NULL %cover%] &then &do
        &call usage
        &call bailout
        &end
    &else &do
        &if not [EXISTS %cover% -COVER] &then &do
            &type Esta Cobertura no existe ---%cover% -----
            &call bailout
        &end
    &end
&return

&routine sys

    additem %cover%-bb.pat %cover%-bb.pat subsis 4 5 I
    &r info4 %cover%-BB
    &sv exis = [exists %cover%-s -cover]
    &if not %exis% &then
        &do
            &type no hay pez
        &end
    &else
        &do
            &if %exis% &then
                kill %cover%-s all
            &end

    intersect %cover%-ac %cover%-bb %cover%-S line
    DROPITEM %cover%-s.aat %cover%-s.aat %cover%-BB#
    DROPITEM %cover%-s.aat %cover%-s.aat AREA
    DROPITEM %cover%-s.aat %cover%-s.aat PERIMETER
    DROPITEM %cover%-s.aat %cover%-s.aat %cover%-BB-ID
    DROPITEM %cover%-s.aat %cover%-s.aat INSIDE
&return

/*-----
&routine usage
/*-----
&type Usage &r subsis <covertura> -- Esta es la sintaxis correcta
&return /*

```

```

/*-----
&routine bailout
/*-----
&severity &error &ignore
/*&call exit
&return &error Error ...

/* sig.aml
/* Calcula el arco siguiente en una capa de ríos
/*A----- Arguments -----
/* cover - El nombre de la cobertura que se le asignaran altitudes

&args cover

/* &r maxid %cover% arc %cover%#
&r max-min-id %cover% arcs %cover%#
&sv .maxid = %.max%

&type el maximo de arcos es %.maxid%
&messages &off &info
tables
&sv i
SELECT %cover%.aat
&type colocando arco siguiente
&do i = 1 &to %.maxid% &by 1
  Reselect FNODE# = 0
  nselect
  Reselect %cover%# = %i%
  &if [show number select] = 0 &then
    &do
      &sv x = i
      &type el arco %i% es de entrada
      nselect
    &end
  &else
    &do
      &sv sig = [show record %i% FNODE#]
      nselect
      Reselect TNODE# = %sig%
      &if [show number select] > 0 &then
        &do
          &sv aux %sig%
          &do
            &if [show number select] <> 0 &then
              Calculate sig = %i%
              &type el arco siguiente de %i% es %sig%
            &end
          &end
        &end
      &end
    &end
  &end
  q
&messages &on
&type -----
&type      Fin de Siguiete
&type -----

&return

/* ant.aml
/* Calcula el arco anterior en una capa de ríos

```

```

/*A----- Arguments -----
/* cover - El nombre de la cobertura que se le asignaran altitudes

&args cover

&r max-min-id %cover% arcs %cover%#
&sv .maxid = %.max%
  &type el maximo de arcos es %.maxid%
  &messages &off &info
  tables
  &sv i

SELECT %cover%.aat
&type colocando arco anterior
&do i = 1 &to %.maxid% &by 1
  Reselect FNODE# = 0
  nselect
  Reselect %cover%# = %i%
  &if [show number select] = 0 &then
    &do
      &sv s i
      /* &type el arco %i% es de salida
      nselect
      &end
    &else
      &do
        &sv ant = [show record %i% TNODE#]
        nselect
        Reselect FNODE# = %ant%
        &if [show number select] > 0 &then
          &do
            &sv aux %ant%
            &do
              &if [show number select] > 0 &then
                Calculate ant = %i%
                &type el arco anterior de %i% es %ant%
              &end
            &end
          &end
        &end
      &end
    &end
  &end
  q
&messages &on
&type -----
&type      Fin de Anterior
&type -----

&return

```

```

/* CLAJER.aml
/* Clasifica capas de ríos utilizando CLAJER
/*A----- Arguments -----
/* cover - El nombre de la cobertura que se le asignaran altitudes
&severity &error &routine bailout
&ECHO &OFF
&args cov
&sv cove [translate [entryname %cov%]]
&sv cover = [before %cove% -]
&sv .IDX = 0
&call chkargs

```

```

&call existe
&r idx %cover%-o1

&return

&routine cla1
  &type ----- Clasificando Rios orden 1 -----
  &r sig %cover%-c
  &r ant %cover%-c

  &r max-min-id %cover%-c arcs subsis
  &sv .maxid = %.max%
  &sv maxsub = %.maxid%
  &r max-min-id %cover%-c arcs %cover%-c#
  &sv .maxid = %.max%

TABLES
&sv i
&sv ii
&sv longsub
&sv longmax
&sv pas 0
&sv sg
  /* Comienza a ordenar
&messages &oN
SELECT %cover%-c.aat
&do ii = 1 &to %maxsub% &by 1 /* 1
  &sv longmax 0
  &sv mayor 0
  &do i = 1 &to %.maxid% &by 1 /* 2
    &type arco %i% del sistema %ii%
    RESELECT FNODE# = 0
    NSELECT
    /* Selecciona el arco de entrada al sistema
    RESELECT %cover%-c# = %i% and FUNCION = 'I' and subsis = %ii%
    &if [show number select] = 1 &then
      &do /* 3
        &sv aux = %i%
        &sv longactual = [show record %aux% length]
        &sv longsub = %longactual%
        &sv pas 0
        &do &while %pas% = 0 /* 4
          &sv fun = 0
          &sv sg = [show record %aux% sig]
          &type el arco %aux% del sub %ii%
          &if %sg% <> 0 &then
            &do /* 5
              &sv fun [show record %sg% funcion]
              &sv aux = %sg%
              RESelect %cover%-c# = %aux%;
              &sv longactual = [show record %aux% length]
              &sv longsub = %longactual% + %longsub%
              &type la longitud es %longsub%
            &end /* 5
          &if %fun% = 'P' &then
            &sv pas = 0
          &else
            &if %fun% = 'O' &then
              &do /* 6
                &sv pas = 999
                &if %longsub% > %longmax% &then
                  &do /* 7
                    &sv longmax %longsub%

```

```

                                &sv mayor %i%
                                &end /* 7
                                &end /* 6
                                &end /* 4
                                &end /* 3
                                &type 5
/*                                nselect
                                &end /*2

/* Asignar valor al atributo de orden y a la longitud de la subruta

&type la longitud maxima es %longmax%
/*&type el arco donde inicia es %mayor%
/*&messages &off &info
&messages &on

reselect FNODE# = 0
nselect
RESelect %cover%-c# = %mayor%
&type -----
&type      Añadiendo ORDEN a rios
&type -----
&if [show number select] = 1 &then
  &do
    Calculate orden = 1
    Calculate lsub = %longmax%
    Calculate ID-RI = %.IDX%
  &end
&sv aux %mayor%
&sv pas 0
&sv fun = 0
&do &while %pas% = 0 /* 4
  &if %aux% <> 0 &then
    &do                                /* 5
      &sv sg [show record %aux% sig]
      &if %sg% <> 0 &then
        &do
          &sv fun [show record %sg% funcion]
          &sv aux = %sg%
        &end
      &end
    &nselect
  RESelect %cover%-c# = %aux%;
  &if [show number select] = 1 &then
    &do
      Calculate orden = 1
      Calculate lsub = %longmax%
      Calculate ID-RI = %.IDX%
      /*&type Añadiendo orden al subsistema %ii%
    &end                                /* 5
    &sv %sg% = %aux%
    &if [show number select] = 0 &then
      &do
        &sv pas = 999
      &end
    &if %fun% = 'P' &then
      &do
        &sv pas = 0
      &end
    &else
      &if %fun% = 'O' &then
      &do
        &sv pas = 999
      &end
    &end
  &end
&end

```

```

        &end
    &end /* 4
    *
    &sv .IDX = %.IDX% + 1
&end /* 1
q
ae
edit %cover%-c
ef arcs
de arcs
drawselect
Select orden = 1
put %cover%-1;
delete
save
Select all;
&if [show number select] <> 0 &then
    &do
        calculate sig = 0;
        calculate ant = 0;
    &end
save
q
    &r sig %cover%-c
&r ant %cover%-c
ae
edit %cover%-c
ef arcs
de arcs
select sig = 0 and ant = 0;
&if [show number select] <> 0 &then
    &do
        calculate orden = 2;
        Calculate lsub = length
        &type copiando a %cover%-01
        Put %cover%-01
        delete;
    &end
save
ef arcs
Select all;
&if [show number select] <> 0 &then
    &do
        calculate sig = 0;
        calculate ant = 0;
        save
    &end
&messages &on
q

&SV ORDEN = 2

    &do orden = 2 &to 4 &by 1

        &type                INICIA CLASIFICACION  DE ORDEN %ORDEN%
        &r sig %cover%-c
        &r ant %cover%-c
        &call SUBSIS
        &r max-min-id %cover%-c arcs sub
        &sv .maxid = %.max%
        &sv maxsub = %.maxid%
        &r max-min-id %cover%-c arcs %cover%-c#
        &sv .maxid = %.max%

```

```

TABLES
&sv i = 0
&sv ii = 0
&sv longsub = 0
&sv longmax = 0
&sv pas 0
&sv sg = 0
select %cover%-c.aat
&do ii = 1 &to %maxsub% &by 1
&sv longmax 0
&sv mayor 0
&do i = 1 &to %.maxid% &by 1
  RESELECT FNODE# = 0
  NSELECT
  reselect %cover%-c# = %i% and ant = 0 and SUB = %ii%
  &if [show number select] = 1 &then
    &do
      &sv aux = %i%
      &sv longactual = [show record %aux% length]
      &sv longsub = %longactual%
      &sv pas 0
      &do &while %pas% = 0
        &sv fun = 0
        &sv sg = [show record %aux% sig]
        &if %sg% <> 0 &then
          &do
            &sv fun [show record %sg% funcion]
            &sv aux = %sg%
            RESelect %cover%-c# = %aux%;
            &sv longactual = [show record %aux% length]
            &sv longsub = %longactual% + %longsub%
            &type la longitud es %longsub%
          &end
        &if %sg% <> 0 &then
          &sv pas = 0
        &else
          &if %sg% = 0 &then
            &do
              &sv pas = 999
              &if %longsub% > %longmax% &then
                &do
                  &sv longmax %longsub%
                  &sv mayor %i%
                &end
              &end
            &end /* 7
          &end /*4
        &end/*3
      &end /* 2
    &messages &on
    &sv pas 0
    reselect FNODE# = 0
    nselect
    ReSelect %cover%-c# = %mayor%
    &if [show number select] = 1 &then
      &do
        Calculate orden = %orden%
        Calculate lsub = %longmax%
        Calculate ID-RI = %.IDX%
      &end
    &sv aux %mayor%
    &sv pas 0
    &if %mayor% = 0 &then
      &sv pas = 999

```



```

&do &while %pas% = 0
  &sv sg = [show record %aux% sig]
  &sv aux %sg%
  reselect FNODE# = 0
  nselect
  Reselect %cover%-c# = %aux%
  &if %sg% = 0 &then
    &sv pas = 999
  &if [show number select] = 1 &then
    &do
      Calculate orden = %orden%
      Calculate lsub = %longmax%
      Calculate ID-RI = %.IDX%
    &end
  &if [show number select] = 0 &then
    &sv pas = 999
  &if %sg% <> 0 &then
    &do
      &sv pas = 0
    &end
  &else
    &if %sg% = 0 &then
      &do
        &sv pas = 999
      &end
    &end /* 12
  &sv .IDX = %.IDX% + 1
&end
ARCEDIT
edit %cover%-c
ef arcs
de arcs
drawselect
Select orden = %orden%
put %cover%-1;Y;
delete
save
Select all;
&if [show number select] <> 0 &then
  &do
    calculate sig = 0;
    calculate ant = 0;
  &end
save
q

&r sig %cover%-c
&r ant %cover%-c
ae
edit %cover%-c
ef arcs
de arcs
select sig = 0 and ant = 0;
&if [show number select] <> 0 &then
  &do
    calculate orden = %orden% + 1;
    Calculate lsub = length
    Put %cover%-o1;Y;
    select sig = 0 and ant = 0;
    delete;
    save
  &end

```

```

Select all;

&if [show number select] <> 0 &then
    &do
        calculate sig = 0;
        calculate ant = 0;
        save
    &end
&messages &on
q

DROPITEM %cover%-c.aat %cover%-c.aat sub
&end

&return

/* -----
&routine subsis
/* -----
    &sv exis = [exists %cover%-buff -cover]
    &if not %exis% &then
        &do
            buffer %cover%-c %cover%-buff # # 1
        &end
    &else
        &if %exis% &then
            &do
                &type
                &type ---- La cobertura ya existia se eliminará
                &type
                kill %cover%-buff all
                buffer %cover%-c %cover%-buff # # 1
            &end

additem %cover%-buff.pat %cover%-buff.pat sub 4 5 I
&r info3 %cover%-BUFF
&sv exis = [exists %cover%-c1 -cover]
&if not %exis% &then
    &do
        &type no hay pez
    &end
&else
    &if %exis% &then
        kill %cover%-c1 all

intersect %cover%-c %cover%-buff %cover%-c1 line
DROPITEM %cover%-c1.aat %cover%-c1.aat %cover%-BUFF#
DROPITEM %cover%-c1.aat %cover%-c1.aat AREA
DROPITEM %cover%-c1.aat %cover%-c1.aat PERIMETER
DROPITEM %cover%-c1.aat %cover%-c1.aat %cover%-BUFF-ID
DROPITEM %cover%-c1.aat %cover%-c1.aat INSIDE
DROPITEM %cover%-c1.aat %cover%-c1.aat %cover%-C-ID
DROPITEM %cover%-c1.aat %cover%-c1.aat %cover%-C#
DROPITEM %cover%-c1.aat %cover%-c1.aat %cover%-AC-ID
DROPITEM %cover%-c1.aat %cover%-c1.aat %cover%-AC#
Kill %cover%-c all
rename %cover%-c1 %cover%-c
DROPITEM %cover%-c1.aat %cover%-c1.aat %cover%-C1-ID
&return

/* -----
&routine existe

```

```
/* -----
&sv exis = [exists %cover%-c -cover]
&sv exis1 = [exists %cover%-1 -cover]
&sv exis22 = [exists %cover%-o1 -cover]

&if %exis22% &then
  &do
    kill %cover%-o1 all
  &end
&if not %exis% &then
  &do
    copy %cover%-f %cover%-c
    &call anade
    &call cla1
  &end
&else
  &do
    &if %exis% &then
      &type
      &type ---- La cobertura ya existia se eliminará
      &type
      kill %cover%-c all
      &if %exis1% &then
        &do
          &type Eliminando cobertura clasificada
          Kill %cover%-1 all
        &end
      copy %cover%-f %cover%-c
      &call anade
      &call cla1
    &end
  &end
&return

/* -----
&routine chkargs
/*-----
&s cover = [UPCASE %cover%]
&if [NULL %cover%] &then &do
  &call usage
  &call bailout
&end
&else &do
  &if not [EXISTS %cover%-F -COVER] &then &do
    &type Esta Cobertura no existe ---%cover% -----
    &call bailout
  &end
&end
&return

/*-----
&routine usage
/*-----
&type Usage &r clajer <covertura> -- Esta es la sintaxis correcta
&return /*

/*-----
&routine bailout
```

```

/*-----
  &severity &error &ignore
  /*&call exit
  &return &error Un Error

/* -----
&routine anterior
/* -----
  /*coloca el valor del arco anterior
  &sv i
  /* &r maxid %cover%-c arc %cover%-c#

&r max-min-id %cover% arcs %cover%-c#
&sv .maxid = %.max%

&type el maximo de arcos es %.maxid%
&messages &off &info
&do i = 1 &to %.maxid% &by 1
  Select %cover%-c# = %i%
  &if [show number select] = 0 &then
    &do
      /*&type
      &type el arco %i% es de salida
      /*&type
    &end
  &else
    &do
      &sv ant = [show arc %i% TNODE#]
      Select FNODE# = %ant%
      &if [show number select] > 0 &then
        &do
          /*
          drawselect
          &sv aux %ant%
          &if [show number select] <> 0 &then
            Calculate ant = %i%
            /*&type
            &type El arco anterior de %i% es %ant%
            /*&type
          &end
        &end
      &end
    &end
  &messages &on
&return

/* -----
&routine anade
/* -----
  additem %cover%-c.aat %cover%-c.aat orden 4 5 I
  additem %cover%-c.aat %cover%-c.aat subr 4 5 I
  additem %cover%-c.aat %cover%-c.aat lsub 8 18 F
  additem %cover%-c.aat %cover%-c.aat sig 4 5 I
  additem %cover%-c.aat %cover%-c.aat ant 4 5 I
  additem %cover%-c.aat %cover%-c.aat ID-RI 4 5 I
&return

/* cambio.aml
/* generaliza una red hidrologica
/*A----- Arguments -----
/*

```

```
/* cover - El nombre de la cobertura que se le asignaran altitudes

&severity &error &routine bailout
&ECHO &OFF
&args cov

/* llama a las funciones

/*****
&sv cove [translate [entryname %cov%]]
&sv cover = [before %cove% -]
&type %cover%
/*****

&call chkargs

    &call existe
    &call gen

        &MESSAGES &POPUP
        &type ----- Hecho
        &messages &off
        &messages &on

&return

/*-----
&routine gen
/*-----
    ae
    edit %cover%-100
    editfeature arcs
    drawenvironment arcs
    ef arcs
    select all
    unsplit none
    save
    /*backcoverage r3-o
    /*backenvironment arcs
    select all
    generalize 350 bendsimplify
    grain .5
    splinemethod default
    spline
    save
    q

    clean %cover%-100 # # # line

&return

&routine existe

&sv exis = [exists %cover%-100 -cover]
&if not %exis% &then
    &do
        copy %cover%-0 %cover%-100
    &end
&else
```

```

&do
  &if %exis% &then
    &type
    &type ---- La cobertura ya existia se eliminará
    &type
    kill %cover%-100 all
    copy %cover%-0 %cover%-100
  &end
  &call selec
&return

&routine selec

/*Calculando máximos
&r max-min-id2 %cover%-100 arc lsub 1
&sv maxo1 = %.max%
&r max-min-id2 %cover%-100 arc lsub 2
&sv maxo2 = %.max%
&r max-min-id2 %cover%-100 arc lsub 3
&sv maxo3 = %.max%
&r max-min-id2 %cover%-100 arc lsub 4
&sv maxo4 = %.max%

&type Comienza la seleccion de los arcos de primero y segundo orden
&station 9999
ae
  edit %cover%-100
  editfeature arcs
  drawenvironment arcs
  ef arcs

  &sv criterio1 = %maxo1% / 5
  &sv criterio2 = %maxo2% / 3
  &sv criterio3 = %maxo3% / 3
  &sv criterio4 = %maxo3% / 3

  /*****
select lsub > 2500
  nselect
delete
save
save
q

&sv exis = [exists %cover%-1-bis -cover]
&if not %exis% &then
  &do
    &type -- no hay
  &end
&else
&do
  &if %exis% &then
    &type
    &type ---- La cobertura ya existia se eliminará
    &type
    kill %cover%-1-bis all
  &end
  copy %cover%-100 %cover%-1-bis

&return

```

```

&routine chkargs
/*-----
&s cove = [UPCASE %cover%]
&if [NULL %cove%-o] &then &do
  &call usage
  &call bailout
&end
&else &do
  &if not [EXISTS %cove%-o -COVER] &then &do
    &type Esta Cobertura no existe ---%cover%-o -----
    &call bailout
  &end
&end
&return

/*-----
&routine usage
/*-----
&type Usage &r cambio <cobertura> -- Esta es la sintaxis correcta
&return /*

/*-----
&routine bailout
/*-----
&severity &error &ignore
/*&call exit
&return &error Un Error

/* Predesc.aml
/* Genera las descripciones semánticas de ríos y curvas de nivel/*
/*
/*A----- Arguments -----
/*
/* r - capa de rios
/* c - capa de curvas
/*
&args r c

&sv ri [translate [entryname %r%]]
&sv rio = [before %ri% -]
&sv cu [translate [entryname %c%]]
&sv curva = [before %cu% -]

&type ri %ri%
&sv esc = [after %ri% -]
&sv num1 = [before %ri% -]
&sv num = [after %num1% R]
&sv exis = [exists %ri%-N -cover]
&sv exis1 = [exists %cu%-N -cover]

&type esc %esc% num %num% cu %cu% ri %ri%
&type Kill %ri%-n all

&if %exis% &then
  Kill %ri%-n all
&if %exis1% &then
  Kill %cu%-N all

/* Asigna identificadores a las capas %ri%-n %cu%-n %num%
&r preparal %ri%-n %cu%-n %num% %esc%
&sv exis = [exists rc%num%-esc%-N -cover]
&if %exis% &then

```

```

Kill rc%num%-esc%-N all
/*&type append rc%num%-esc%-N line features

/* Une las capas de rios y curvas
append rc%num%-esc%-N line features
%ri%-N
%cu%-N
;
Y
Y
/***** UNSPLIT

CLEAN rc%num%-esc%-N # # 0.000000001 LINE

ae
edit rc%num%-esc%-N
ef arcs
select all
unsplit name
save
q
/* parte por la mitad los arcos para definir relaciones
&r partirmitad rc%num%-esc%-N

RENODE rc%num%-esc%-N
CLEAN rc%num%-esc%-N # # 0.000000001 LINE

&R BUILDNAL.AML rc%num%-esc%-N

&r ordenanal rc%num%-esc%-N
joinitem rc%num%-esc%-N.aat rc%num%-esc%-N.nal rc%num%-esc%-N.aat rc%num%-
esc%-N#

&sv exis = [exists rc%num%-esc%-2 -cover]
&if %exis% &then
    Kill rc%num%-esc%-2 all

&sv exis = [exists BUFF-RC%num%-esc% -cover]
&if %exis% &then
    Kill BUFF-RC%num%-esc% all

&sv exis = [exists CLIP-RC%num%-esc% -cover]
&if %exis% &then
    Kill CLIP-RC%num%-esc% all

AE
EDIT rc%num%-esc%-N
EF ARCS
SELECT INTER = 'I'
    NSELECT
    DELETE
SELECT ALL

&SV max [show number select]
&do i = 1 &to %max% &by 1
    SELECT NUMREL = %i%
    RESELECT TYPE = 'R'
    &SV selec [show number select]
    &if %selec% > 2 &then
        &DO
            SELECT NUMREL = %i%
            DELETE
        &END

```



```

&END
  SAVE rc%num%-&esc%-2
Q
RENODE rc%num%-&esc%-2
BUFFER rc%num%-&esc%-2 BUFF-RC%num%-&esc% # # 75 0.01 NODE
CLIP rc%num%-&esc%-2 BUFF-RC%num%-&esc% CLIP-RC%num%-&esc% LINE
CLEAN CLIP-RC%num%-&esc% # # 0.0000000001 LINE
RENODE CLIP-RC%num%-&esc%
&r elimina CLIP-RC%num%-&esc%
RENODE CLIP-RC%num%-&esc%
&r desc-rc1 CLIP-RC%num%-&esc%
&r concept CLIP-RC%num%-&esc%
&return

/* Prepara.aml
/* prepara las capas de rios y cuevas de nivel para generar descripciones/*A-----
----- Arguments -----
/*
/* cover-r - capa de rios
/* cover-c - capa de curvas
/* num - num consecutivo para identificar las capas
/* esc - fuente-1, generalizado-100

&args cover-r cover-c num esc
&type %cover-r% %num%
&type %cover-c% %num%
&type -----

&r name-rio %cover-r% %num% %esc%
/*
&r name-curva2 %cover-c% %num% %esc%
&r name-curva2 %cover-c% %num% %esc%

&sv rio = [before %cover-r% -]
&sv curva = [before %cover-c% -]

dropitem %cover-c%.aat %cover-c%.aat ID
dropitem %cover-c%.aat %cover-c%.aat CLAVE
dropitem %cover-c%.aat %cover-c%.aat FC
dropitem %cover-c%.aat %cover-c%.aat TIPO
dropitem %cover-c%.aat %cover-c%.aat ELEVATION
dropitem %cover-c%.aat %cover-c%.aat ID-C
dropitem %cover-c%.aat %cover-c%.aat SINUOUS
dropitem %cover-r%.aat %cover-r%.aat %rio%-AC#
dropitem %cover-r%.aat %cover-r%.aat %rio%-AC-ID
dropitem %cover-r%.aat %cover-r%.aat F_ALTURA
dropitem %cover-r%.aat %cover-r%.aat I_ALTURA
dropitem %cover-r%.aat %cover-r%.aat ID
dropitem %cover-r%.aat %cover-r%.aat CLAVE
dropitem %cover-r%.aat %cover-r%.aat FC
dropitem %cover-r%.aat %cover-r%.aat CONDICION
dropitem %cover-r%.aat %cover-r%.aat CONDICI
dropitem %cover-r%.aat %cover-r%.aat SUBSIS
dropitem %cover-r%.aat %cover-r%.aat FUNCION
dropitem %cover-r%.aat %cover-r%.aat SUBR
dropitem %cover-r%.aat %cover-r%.aat LSUB
dropitem %cover-r%.aat %cover-r%.aat SIG
dropitem %cover-r%.aat %cover-r%.aat ANT
dropitem %cover-r%.aat %cover-r%.aat ID-RI
dropitem %cover-r%.aat %cover-r%.aat ORDEN

/* name-rio.aml

```

```

/* asigna identificadores a las capas de ríos
/*A----- Arguments -----
/*
/* cov - capa de rios
/* num - num consecutivo para identificar las capas
/* esc - fuente-1, generalizado-100
&args cov num esc
&sv cove [translate [entryname %cov%]]
&sv cover = [before %cove% -]
&type cover %cover% cove %cove%
copy %cover%-%esc% %cover%-%esc%-N
additem %cover%-%esc%-N.aat %cover%-%esc%-N.aat name 16 17 C
additem %cover%-%esc%-N.aat %cover%-%esc%-N.aat TYPE 2 3 C
&sv rio IO_
    ae
    edit %cover%-%esc%-N
    ef arcs
    select all
    &sv maxarcos [show number select]
    &do ii = 1 &to %maxarcos% &by 1
        select %cover%-%esc%-N# = %ii%
        &sv numero = [show arc %ii% ITEM ID-RI]
        &sv cadena %rio%%numero%
        &type R%cadena%
        CALCULATE type = 'R'
        CALCULATE NAME = 'R'%cadena%
        /*&type %1%
    &end
save
q
&type rio
&return

/* name-curva.aml
/* asigna identificadores a las capas de curvas de nivel
/*A----- Arguments -----
/*
/* cov - capa de curvas de nivel
/* num - num consecutivo para identificar las capas
/* esc - fuente-1, generalizado-100

&args cov num esc

/*****
&sv cove [translate [entryname %cov%]]
&sv cover = [before %cove% -]
&type %cover%
/*****

&sv exi = [exists %cover%-%esc%-q -cover]
&if %exi% &then
    kill %cover%-%esc%-q all

&sv exi = [exists %cover%-%esc%-n -cover]
&if %exi% &then
    kill %cover%-%esc%-n all

dropitem %cover%-%esc%.aat %cover%-%esc%.aat ID-C
copy %cover%-%esc% %cover%-%esc%-q
&call ids

```

```

additem %cover%-esc%-N.aat %cover%-esc%-N.aat name 16 17 C
additem %cover%-esc%-N.aat %cover%-esc%-N.aat TYPE 2 3 C
&sv rio urva_
  ae
  edit %cover%-esc%-N
  ef arcs
  select all
  &sv maxarcos [show number select]
  &do ii = 1 &to %maxarcos% &by 1
  select %cover%-esc%-N# = %ii%
  &sv numero = [show arc %ii% ITEM ID-C]
  &sv cadena %rio%numero%
  &type C%cadena%
  CALCULATE TYPE = 'C'
  CALCULATE NAME = 'C'%cadena%
  &end
  save
  q
  DROPITEM %cover%-esc%-n.aat %cover%-esc%-n.aat ID-c
&return

/* -----
&routine ids
/* -----
  &sv exis = [exists %cover%-esc%-v -cover]
  &if not %exis% &then
    &do
      buffer %cover%-esc%-q %cover%-esc%-v # # 3
    &end
  &else
    &if %exis% &then
      &do
        kill %cover%-esc%-v all
        buffer %cover%-esc%-q %cover%-esc%-v # # 3
      &end

additem %cover%-esc%-v.pat %cover%-esc%-v.pat ID-C 4 5 I
&r info3-3 %cover%-esc%-V

  &sv exis = [exists %cover%-esc%-n -cover]
  &if not %exis% &then
    &do
      &type no hay pez
    &end
  &else
    &if %exis% &then
      kill %cover%-esc%-n all
  intersect %cover%-esc%-q %cover%-esc%-v %cover%-esc%-n line
  DROPITEM %cover%-esc%-n.aat %cover%-esc%-n.aat %cover%-esc%-v#
  DROPITEM %cover%-esc%-n.aat %cover%-esc%-n.aat AREA
  DROPITEM %cover%-esc%-n.aat %cover%-esc%-n.aat PERIMETER
  DROPITEM %cover%-esc%-n.aat %cover%-esc%-n.aat %cover%-esc%-v-id
  DROPITEM %cover%-esc%-n.aat %cover%-esc%-n.aat %cover%-esc%-q-ID
  DROPITEM %cover%-esc%-n.aat %cover%-esc%-n.aat %cover%-esc%-q#
  DROPITEM %cover%-esc%-n.aat %cover%-esc%-n.aat INSIDE
&return

/* partirmital-curva.aml
/* fragmenta los arcos para analizar la relación entre ríos y curvas de nivel

```

```

/*A----- Arguments -----
/*
/* cover - capa de curvas de nivel y ríos
&args cover

tolerance %cover% edit 2
tolerance %cover% nodesnap 2
tolerance %cover% weed 2
tolerance %cover% grain .5
tolerance %cover% snap 2
tolerance %cover% fuzzy 2

AE
EDIT %cover%
EF ARCS
weedtolerance 0.000001
COORDINATE KEYBOARD
de arcs nodes
SELECT ALL
&SV max [show number select]
&type Seleccionados %max%
&do i = 1 &to %max% &by 1 /* 1
    &sv num-ver = [show ARC %i% NPNTS]
    &type el numero de vertices del arco %i% es %num-ver%
    &type -----
    &if %num-ver% = 2 &then
        &do
            &sv v1 = [show ARC %i% Vertex 1]
            &sv v2 = [show ARC %i% Vertex 2]
            &sv v1X = [before %v1% ,]
            &sv v1Y = [after %v1% ,]
            &sv v2X = [before %v2% ,]
            &sv v2Y = [after %v2% ,]
            &sv med = %v1X% + %v2X%
            &sv medX = %med% / 2
            &sv med = %v1Y% + %v2Y%
            &sv medY = %med% / 2
            &type medX %medX% medY %medY%
            select %cover%# = %i%
            split
            %medX%, %medY%;
        &end
    &else
        &do
            &sv mitad [round [calc %num-ver% / 2]]
            &sv nodo = [show ARC %i% Vertex %mitad% ]
            select %cover%# = %i%
            drawselect
            split
            %nodo%;
        &end
    &end
&end
COORDINATE cursor
save
q
&return

/*N buildnal.aml
/*P----- Purpose -----
/*
/* crea una Node Arc List (NAL)
/*

```

```
/*A----- Arguments -----
/* cov - capa a la que se le construirá una NAL

&severity &error &routine bailout
&args c
&sv cover [translate [entryname %c%]]
/* Check arguments
&call chkargs

/* Turn off the messages.
&s old$messages = [SHOW &messages]
&messages &off &all

/* Check if %cover% has an AAT.
&call chkaat

/* Build the NAL
&call nal

&call exit
&return

/*-----
&routine chkargs
/*-----
&s cover = [UPCASE %cover%]
&s builtaat = .FALSE. /* Set to .TRUE. in routine chkaat if no AAT
&if [NULL %cover%] &then &do
    &call usage
    &call bailout
&end
&else &do
    &if not [EXISTS %cover% -COVER] &then &do
        &type Coverage %cover% does not exist.
        &call bailout
    &end
&end
&return

/*-----
&routine chkaat
/*-----
&if not [EXISTS %cover% -LINE] &then &do
    &s builtaat = .TRUE.
    BUILD %cover% LINE
&end
&return

/*-----
&routine nal
/*-----
&type -----
&type Creando Node Arc List (NAL) para %cover%...
&type -----
&if [EXISTS %cover%.nal -INFO] &then &do
    &s del$stat = [DELETE %cover%.nal -INFO] &then &do
&end
&data arc info > /dev/null
ARC
DEFINE %cover%.NAL
NODE#,4,5,B
ARC#,4,5,B
[UNQUOTE '']
```

```

REDEFINE
1,FNODE#,4,5,B
1,TNODE#,4,5,B
[UNQUOTE '']
SELECT %cover%.AAT RO
REL %cover%.NAL 1 BY FNODE# APPEND
CALC $lARC# = %cover%#
REL %cover%.NAL 1 BY TNODE# APPEND
CALC $lARC# = %cover%#
SELECT %cover%.NAL
SORT NODE#,ARC#
Q STOP
&end
&return

/*-----
&routine usage
/*-----
&type Usage &r buildnal <cover>
&return /*

/*-----
&routine exit
/*-----
&if %builtaat% &then &do
    DROPFEATURES %cover% LINE ATTRIBUTES
&end
&do item &list fnode# tnode#
    &if [EXISTS %cover%.nal -INFO] &then &do
        &if [ITEMINFO %cover%.nal -INFO %item% -EXISTS] &then &do
            DROPITEM %cover%.nal %cover%.nal %item%
        &end
    &end
&end
&if [VARIABLE old$messages] &then &do
    &if [SHOW &messages] ne %old$messages% &then &do
        &messages %old$messages%
    &end
&end
&return

/*-----
&routine bailout
/*-----
&severity &error &ignore
&call exit
&return &error Error en buildnal.aml...

/*N ordenanal.aml
/*P----- Purpose -----
/*
/* ordena una Node Arc List (NAL)
/*
/*A----- Arguments -----
/* cover - capa que posee la NAL que se quiere ordenar

&args cover

ADDITEM %cover%.NAL %cover%.NAL INTER 2 3 C
ADDITEM %cover%.NAL %cover%.NAL NUMREL 4 5 b

```

```

&r max-min-id %cover% arcs FNODE#
&sv maxFN = %.max%
&r max-min-id %cover% arcs TNODE#
&sv maxTN = %.max%

&if %maxFN% > %maxTN% &then
  &sv MAX = %maxFN%
&else
  &sv MAX = %maxTN%

TABLES
&sv i 1
&sv ii 1
SELECT %cover%.NAL
&do i = 1 &to %MAX% &by 1
  &TYPE NODO %i%
  RESELECT NODE# = 0
  nselect
  RESELECT NODE# = %i%
  /*&sv 1 [show number select]
  /*&type seleccionados ----- %1%
  &if [show number select] = 4 &then
  &do
    CALCULATE INTER = 'I'
    CALCULATE NUMREL = %ii%
    &sv ii = %ii% + 1
  &end
  &ELSE
    &do
      CALCULATE INTER = 'N'
      CALCULATE NUMREL = 0
    &end
  &END
  RESELECT NODE# = 0
  nselect
  RESELECT INTER = 'I'
  NSELECT
  purge
  Y;
  ALTER ARC#,%cover%#,,,,
  Q
&return

/*N descrc.aml
/*P----- Purpose -----
/*
/* realiza la medición de la relación pasa por
/*
/*A----- Arguments -----
/* cover - capa a medir

&args cover
dropitem %cover%.aat %cover%.aat sin
dropitem %cover%.aat %cover%.aat CLASIF
additem %cover%.aat %cover%.aat sin 8 9 F 4
additem %cover%.aat %cover%.aat CLASIF 4 5 I
/***** OBTENER NUMERO DE RELACIONES
&r max-min-id %cover% arcs NUMREL
&s RELACION_MAX = %.max%
/*****FIN DE OBTENER NUMERO DE RELACIONES
ae

```

```

EDIT %cover%
EF ARCS
&s rel = 1
/***** ANALISIS DE RELACIONES
&do rel = 1 &to %RELACION_MAX% &by 1
/***** OBTENER OBJETOS QUE COMPONEN LA RELACION
  Select NUMREL = %rel%
  Reselect type = 'R'
  &SV selec [show number select]
  &if %selec% = 2 &then
    &do
      /*****
      &type -----
      &type Analizando Relación ( Select NUMREL = %rel% )
      &type &sv obj-R1 [SHOW SELECT 1]
      &type &sv obj-R2 [SHOW SELECT 2]
      &type -----
      &sv obj-R1 [SHOW SELECT 1]
      &sv obj-R2 [SHOW SELECT 2]
    NSELECT
    Select NUMREL = %rel%
    Reselect type = 'C'
    &sv Area_max = 0
    &sv Vertice_max = 0
    &sv obj-C1 [SHOW SELECT 1]
    &sv obj-C2 [SHOW SELECT 2]
      &TYPE &sv tnode-r1 %obj-R1%
      &TYPE %obj-R2%
      &sv tnode-r1 = [show ARC %obj-R1% TNODE#]
      &sv fnode-r2 = [show ARC %obj-R2% FNODE#]
/***** ORDENA LOS OBJETOS RIO (OBJ-R1 ENTRA, OBJ-R2 SALE) *****/
  &type rio-1 %obj-R1% rio-2 %obj-R2% ---
  &if %tnode-r1% <> %fnode-r2% &then
    &do
      &sv paso %obj-r1%
      &sv obj-r1 = %obj-r2%
      &sv obj-r2 = %paso%
    &end
    &sv tnode-r1 = [show ARC %obj-R1% TNODE#]
    &sv fnode-r2 = [show ARC %obj-R2% FNODE#]
    &if %tnode-r1% = %fnode-r2% &then
      &do
        &type ya se armo
      &end
    &type rio-1 %obj-R1% rio-2 %obj-R2% -----
/***** COORDENADAS DE VERTICES EN R1(FNODE#) Y R2(TNODE#)
  &sv Fnode-R1 = [show ARC %obj-r1% Vertex 1 ]
  &sv numnodos = [show ARC %obj-r2% NPNTS]
  &sv Tnode-R2 = [show ARC %obj-r2% Vertex %numnodos%]

/*****FIN DE ORDENAR LOS OBJETOS RIO (OBJ-R1 ENTRA, OBJ-R2 SALE)
/***** ORDENA LOS OBJETOS CURVA (OBJ-C1 PRIMERO, OBJ-C2 SEGUNDO)
  &sv tnode-c1 = [show ARC %obj-C1% TNODE#]
  &sv fnode-c2 = [show ARC %obj-C2% FNODE#]
  &if %tnode-c1% <> %fnode-c2% &then
    &do
      &sv paso %obj-C1%
      &sv obj-c1 = %obj-c2%
      &sv obj-c2 = %paso%
    &end
    &sv tnode-c1 = [show ARC %obj-C1% TNODE#]
    &sv fnode-c2 = [show ARC %obj-C2% FNODE#]
    &if %tnode-c1% = %fnode-c2% &then

```



```

        &do
            &type ya se armo
        &end
/***** FIN DE ORDENAR LOS OBJETOS CURVA (OBJ-C1 PRIMERO, OBJ-C2 SEGUNDO)
/***** NUMERO DE VERICES EN C1 Y C2
        &sv ver-c1 = [show ARC %obj-C1% NPNTS]
        &sv ver-c2 = [show ARC %obj-C2% NPNTS]
        &sv A = 1

/**** OBTENER COORDENADAS PARA LAS CURVAS ****
/**** COORDENADAS DE FNODE EN C1
        &sv nodo = [show ARC %obj-c1% Vertex 1]
        &sv nodo-ini = %nodo%
/**** COORDENADAS DE LOS VERTICES DE C1
        &SV ver-c1 = %ver-c1% - 1
        &do i = 2 &to %ver-c1% &by 1
            &sv nodo = [show ARC %obj-c1% Vertex %i% ]
            &sv nodo%A% = %nodo%
            &TYPE NODO %A% [value nodo%A%] -- REL %rel%
            &sv A = %A% + 1
        &end
/**** COORDENADAS DEL NODO DE INTERSECCION
        &sv nodo-inter = [show ARC %obj-c2% Vertex 1 ]
        &TYPE INTERSECCION EN %nodo-inter%
        &sv nodo-fin = [show ARC %obj-c2% Vertex %ver-c2% ]
        &SV ver-c2 = %ver-c2% - 1
/**** COORDENADAS DE LOS VERTICES DE C2
        &if %ver-c2% = 1 &then
            &do
                &type SOLO TIENE NODOS
                &sv nodo = [show ARC %obj-c2% Vertex 1]
                &sv nodo%A% = %nodo%
                &sv A = %A% + 1
            &end
        &else
            &do
                &type TIENE NODOS Y VERTICES
                &do i = 1 &to %ver-c2% &by 1
                    &sv nodo = [show ARC %obj-c2% Vertex %i% ]
                    &sv nodo%A% = %nodo%
                    &TYPE NODO %A% [value nodo%A%] -- REL %rel%
                    &sv A = %A% + 1
                &end
            &end

        &TYPE LAS CURVAS TIENEN %A% VERTICES

        &do i = 1 &to %A% &by 1
            &TYPE NODO %i% [value nodo%i%] -- REL %rel%
        &end

/**** FIN DE OBTENER COORDENADAS PARA LAS CURVAS ****
/**** DETECTANDO SI LA CURVA ES RECTA &sv
i = 1
        &sv Vertex = [value nodo%i%]
        &sv longcurva = [invdistance %nodo-ini% %vertex%]
/****
        &type Analizando Relación ( Select NUMREL = %rel% )
/****
        Select NUMREL = %rel%
        &TYPE -
        &TYPE NODO INI %nodo-ini%
        &sv maximun = %A% - 1

```

```

&Type      &sv maximun = %A% - 1
           &do i = 2 &to %maximun% &by 1
           &if %i% = 1 &then
             &do
               &sv aux = 1
               &sv Vertex = [value nodo%aux%]
               &type NODO %aux% [value nodo%aux%]
             &end
           &else
             &do
               &sv aux = %i% - 1
               &sv Vertex = [value nodo%aux%]
             &end
           &sv Vertex1 = [value nodo%i%]
           &type NODO %i% [value nodo%i%] i %i%
           &type --
           &sv long = [invdistance %vertex% %vertex1%]
           &sv longcurva = %long% + %longcurva%
         &end
&TYPE NODO fin %nodo-fin%
&TYPE -
&type %longcurva%
&sv Vertex = [value nodo%maximun%]
&sv long = [invdistance %vertex% %nodo-fin%]
&sv longcurva-R = %long% + %longcurva%
&SV longcurva = [round %longcurva-R%]
&type %longcurva-R%
&sv recta-R [invdistance %nodo-ini% %nodo-fin%]
&SV recta = [round %recta-R%]
&type %recta-R%
&sv sin = %longcurva% / %recta%
&type SIN %SIN% REL %REL%Es recta!!!!!!!
&if %sin% = 1 or %sin% <= 1.01 &then
  &do
    &type SIN %SIN% REL %REL%Es recta!!!!!!!
    CALCULATE CLASIF = 0
    CALCULATE SIN = %sin%
  /*    &MESSAGES &POPUP
  /*    &type ----- RECTA-----
  /*    &messages &off
  /*    &messages &on

  &end
/*****FIN DE DETECTANDO SI LA CURVA ES RECTA
&ELSE
  &if %sin% > 1.01 &then
  &DO
/*****
  &type A %A%
  Select NUMREL = %rel%
  CALCULATE SIN = %sin%
  &do I = 1 &to %MAXIMUN% &by 1

/***** CALCULANDO MAXIMA CONCAVIDAD
  &sv lado1 = [invdistance %nodo-ini% %nodo-fin%]
  &sv otrovertice = [value nodo%i%]
  &type este es el vertice -%otrovertice%-
  &sv lado2 = [invdistance %otrovertice% %nodo-ini%]
  &sv lado3 = [invdistance %otrovertice% %nodo-fin%]
  &sv p1 = %lado1% + %lado2% + %lado3%
  &type LADOS %lado1% + %lado2% + %lado3% PAL NODO %i%
  &SV p = %p1% / 2
  &sv t1 = %P% - %lado1%

```

```

        &sv t2 = %P% - %lado2%
        &sv t3 = %P% - %lado3%
        &sv areal = %t1% * %t2% * %t3% * %p%
        &sv area = [sqrt %areal%]
        &type
        &type AREA ----- %area% EN EL NODO %i%
        &type
        &if %area% > %Area_max% &then
            &do
                &sv Area_max = %area%
                &sv Vertice_max = %otrovertice%
            &end
        &end

/*****FIN DE CALCULANDO MAXIMA CONCAVIDAD          &END
        &END

&END

Save
&r calc_peri.aml %cover%

q
&return

/*N concept.aml
/*P----- Purpose -----
/*
/* etapa se sintesis
/*
/*A----- Arguments -----
/* cover - capa

&args c

&sv cov [translate [entryname %c%]]
&sv cover = [after %cov% -]
&type %cover%
&sv exis = [exists CP-%cover% -cover]
&if %exis% &then Kill CP-%cover% all
copy %cov% CP-%cover%
additem CP-%cover%.aat CP-%cover%.aat CPTO 36 37 C
additem CP-%cover%.aat CP-%cover%.aat numCPTO 4 5 I
ae
    edit CP-%cover%
    EF ARCS
    SELECT CLASIF <= 100
    RESELECT CLASIF >= 99
        &if [show number select] <> 0 &then
            &DO
                CALCULATE numCPTO = 7
                &SV CONCEPTO = PASA POR MAXIMA CONVEXIDAD
                CALCULATE CPTO = '%CONCEPTO%
            &END
        SELECT CLASIF <= 98
        RESELECT CLASIF >= 85
            &if [show number select] <> 0 &then
                &DO
                    CALCULATE numCPTO = 6
                    &SV CONCEPTO = PASA CASI POR MAXIMA CONVEXIDAD
                    CALCULATE CPTO = '%CONCEPTO%

```

```

&END
SELECT CLASIF <= 84
RESELECT CLASIF >= 35
  &if [show number select] <> 0 &then
    &DO
      CALCULATE numCPTO = 5
      &SV CONCEPTO = PASA CASI POR CONVEXIDAD
      CALCULATE CPTO = ''%CONCEPTO%
    &END
SELECT CLASIF <= 34
RESELECT CLASIF >= -34
  &if [show number select] <> 0 &then
    &DO
      CALCULATE numCPTO = 4
      &SV CONCEPTO = PASA POR RECTA
      CALCULATE CPTO = ''%CONCEPTO%
    &END
SELECT CLASIF <= -35
RESELECT CLASIF >= -85
  &if [show number select] <> 0 &then
    &DO
      CALCULATE numCPTO = 3
      &SV CONCEPTO = PASA CASI POR CONCAVIDAD
      CALCULATE CPTO = ''%CONCEPTO%
    &END
SELECT CLASIF <= -86
RESELECT CLASIF >= -98
  &if [show number select] <> 0 &then
    &DO
      CALCULATE numCPTO = 2
      &SV CONCEPTO = PASA CASI POR MAXIMA CONCAVIDAD
      CALCULATE CPTO = ''%CONCEPTO%
    &END
SELECT CLASIF <= -99
RESELECT CLASIF >= -100
  &if [show number select] <> 0 &then
    &DO
      CALCULATE numCPTO = 1
      &SV CONCEPTO = PASA POR MAXIMA CONCAVIDAD
      CALCULATE CPTO = ''%CONCEPTO%
    &END
&END

SAVE
Q

DROPitem CP-%cover%.aat CP-%cover%.aat CLASIF
DROPitem CP-%cover%.aat CP-%cover%.aat SIN
&r hacerrel CP-%cover%

&return

/*N hacerrelacion.aml
/*
/*P----- Argumentos-----
/* cover - capa

&ARGS cover

&r max-min-id %cover% arcs NUMREL
&sv MAXREL = %.max%

dropitem %cover%.aat %cover%.aat NAME2

```

```

additem %cover%.aat %cover%.aat NAME2 16 17 C

&sv exis = [exists r-%cover% -COVER]
&if %exis% &then KILL r-%cover% ALL

ae
&sv i 1
&sv ii 1
edit %cover%
ef arcs
  &do i = 1 &to %maxrel% &by 1
    SELECT NUMREL = %i%

    &SV sel [show number select]
    &if %sel% = 0 &then
      &type zero
    &else
      &do
        RESELECT TYPE = 'R'
        &sv obj-R1 [SHOW SELECT 1]
        &type rio 1 %obj-R1%
        nselect
          SELECT NUMREL = %i%
        RESELECT TYPE = 'C'
        &sv obj-C1 [SHOW SELECT 1]
        &type curva 1 %obj-C1%
        &sv NAME_C = [show ARC %obj-C1% ITEM NAME]
        &type %NAME_C%
        select %cover%# = %obj-R1%
        CALCULATE NAME2 = ''%NAME_C%
      &end
    &END

    SELECT NAME2 = ''
  delete
  save r-%cover%
  q

dropitem r-%cover%.AAT r-%cover%.AAT TYPE NODE#

/* confusion.aml
/*
/* calcula la confusion entre 2 representaciones conceptuales
/*
/*A----- Arguments -----
/* cover - capa

&args cov1 cov2 new

&Sv fuente SC-RC6-1
&Sv gener SC-RC6-100

/*****
&sv col [translate [entryname %cov1%]]
&sv col00 [translate [entryname %cov2%]]

/*****

&sv exi = [exists %new% -cover]
&if %exi% &then
  kill %new% all

```

```

&sv exi = [exists %fuente% -cover]
&if %exi% &then
    kill %fuente% all
copy %col% %fuente%

&sv exi = [exists %gener% -cover]
&if %exi% &then
    kill %gener% all
copy %col00% %gener%

additem %fuente%.aat %fuente%.aat llave 35 36 C
additem %gener%.aat %gener%.aat llave 35 36 C

&call fuente-a
&call gener-a

&R info-7 %GENER%.AAT

dropITEM %fuente%.aat cucha12.dat FNODE# length name name2 tnode# lpoly# rpoly#
%fuente%# %fuente%-id inter numrel

JOINITEM %gener%.aat cucha12.dat %gener%.aat llave

&return

/*****
&routine fuente-a
/*****
ae
    edit %fuente%
    ef arcs
    select all
    &sv maxarcos [show number select]
    &do ii = 1 &to %maxarcos% &by 1
        select %fuente%# = %ii%
        &sv rio = [show arc %ii% ITEM Name]
            &sv curva = [show arc %ii% ITEM Name2]
        &sv cadena %rio%%curva%
        &type CALCULATE llave = %cadena%
        CALCULATE llave = '%cadena%
    &end
save
q
&return

/*****
&routine gener-a
/*****
ae
    edit %gener%
    ef arcs
    select all
    &sv maxarcos [show number select]
    &do ii = 1 &to %maxarcos% &by 1
        select %gener%# = %ii%
        &sv rio = [show arc %ii% ITEM Name]
            &sv curva = [show arc %ii% ITEM Name2]
        &sv cadena %rio%%curva%
        &type CALCULATE llave = %cadena%
        CALCULATE llave = '%cadena%
    &end

```

```
save
q
&return

/* COMPARA.AML
/*
/* compara dos representaciones conceptuales
/*
/*A----- Arguments -----
/* cover - capa

&ARGS C1 C100

&sv cov1 [translate [entryname %c1%]]
&sv cov100 [translate [entryname %c100%]]

&sv cover = [after %cov1% -]
&type cover %cover%
&sv cove = [after %cover% -]
&type cove %cove%
&sv capa = [before %cove% -]
&type capa %capa%

&sv exis = [exists desc-%capa% -cover]
&if %exis% &then Kill desc-%capa% all
COPY %COV1% desc-%capa%

&SV AUX
TABLES
SELECT DESC-%CAPA%.AAT
ALTER NAME,NAME-R-1,,,,
ALTER NAME2,NAME-C-1,,,,
ALTER CPTO,CPTO1,,,,
ALTER NUMCPTO,NUMCPTO1,,,,
Q
TABLES
SELECT %COV100%.AAT
ALTER NAME,NAME-R-100,,,,
ALTER NAME2,NAME-C-100,,,,
ALTER CPTO,CPTO100,,,,
ALTER NUMCPTO,NUMCPTO100,,,,
Q

additem DESC-%CAPA%.AAT DESC-%CAPA%.AAT RES-NUM1 4 5 I
/*RES-texto 5 6 C
additem DESC-%CAPA%.AAT DESC-%CAPA%.AAT RES-NUM2 4 5 I
/*RES-num 4 5 i

JOINITEM DESC-%CAPA%.AAT %COV100%.AAT DESC-%CAPA%.AAT NUMREL

AE
  EDIT DESC-%CAPA%
  EF ARCS
  SELECT ALL
  &SV CONT 1
  &sv max [show number select]
  &do i = 1 &to %max% &by 1
    SELECT NUMREL = %i%
    &type relacion %i%
    &sv obj-Rel [SHOW SELECT 1]
```

```

&sv N_riol = [show ARC %obj-Rel% ITEM NAME-R-1]
&sv N_curval = [show ARC %obj-Rel% ITEM NAME-C-1]
&TYPE RIO 1 %N_riol% CURVA 1 %N_curval%
&type --
select all
&type --
&sv ri = [quote %N_riol%]
&sv cu = [quote %N_curval%]
SELECT NAME-R-100 eq %ri%
RESELECT NAME-C-100 EQ %CU%
&sv cont2 [show number select]

&type
&type SELECT NAME-R-100 eq %ri% and NAME-C-100 EQ %CU%
&type
/***** Esta relacion no existe en la otra escala
*****
&if [show number select] = 0 &then
&do
&type no existe
&end
/***** Analizar relación que se conserva,
&if [show number select] = 1 &then
&do
&TYPE -----
&type relación que se conserva
&TYPE -----
SELECT NUMREL = %i%
CALCULATE RES-num1 = 1
&SV CONT %CONT% + 1
&TYPE %CONT%
&end
/***** Analizar relación que se conserva, seguramente es un error
&if [show number select] > 1 &then
&do
&TYPE *****
&type ESTA RELACION ESTA DUPLICADA EN r-100
&TYPE *****
SELECT NUMREL = %i%
CALCULATE RES-num1 = %cont2%
&SV CONT %CONT% + 1
&TYPE %CONT%
&END
&END
/*&TYPE CONT %CONT2%
&SV CONT2 = 1
&do i = 1 &to 16 &by 1
SELECT NUMREL = %i%
&type relacion %i%
&sv obj-Rel [SHOW SELECT 1]
&sv N_riol = [show ARC %obj-Rel% ITEM NAME-R-100]
&sv N_curval = [show ARC %obj-Rel% ITEM NAME-C-100]
&TYPE RIO 1 %N_riol% CURVA 1 %N_curval%
&type --
select all
&type --
&sv ri = [quote %N_riol%]
&sv cu = [quote %N_curval%]
SELECT NAME-R-1 eq %ri%
RESELECT NAME-C-1 EQ %CU%

&type
&type SELECT NAME-R-11 eq %ri% and NAME-C-1 EQ %CU%
&type

```



```
/****** Analizar relación que se conserva
  &if [show number select] = 1 &then
    &do
      &TYPE -----
      &type parece que no hay
      &TYPE -----
    &end
  &if [show number select] > 1 &then
    &do
      /*SELECT NUMREL = %i%
      CALCULATE RES-num2 = 1
      &SV CONT2 %CONT2% + 1
      &TYPE %CONT2%
    &end

    &if [show number select] = 0 &then
      &do
        SELECT NAME-R-100 eq %ri%
        RESELECT NAME-C-100 EQ %CU%
        CALCULATE RES-num2 = 2
        &SV CONT2 %CONT2% + 1
      &END
    &END
  &TYPE CONT2 %CONT2%

&RETURN
```