INSTITUTO POLITÉCNICO NACIONAL

CENTRO DE INVESTIGACIÓN EN COMPUTACIÓN

ANÁLISIS ESPACIAL Y SUPERVISADO DE DATOS CRIMINALÍSTICOS UTILIZANDO UN SISTEMA DE INFORMACIÓN GEOGRÁFICA

TESIS QUE PARA OBTENER EL GRADO DE

MAESTRÍA EN CIENCIAS DE LA COMPUTACIÓN

PRESENTA

ING. JOSÉ JAVIER RUIZ GÓMEZ

DIRECTORES DE TESIS

DR. MIGUEL JESÚS TORRES RUIZ
DR. JOSÉ GIOVANNI GUZMÁN LUGO

MÉXICO, D.F., MAYO 2013

SIP-14 bis

INSTITUTO POLITÉCNICO NACIONAL
SECRETARÍA DE INVESTIGACIÓN Y POSGRADO

ACTA DE REVISIÓN DE TESIS

En la Ciudad de México, D.F. siendo las 12:00 horas del día 27 del mes de mayo de 2013 se reunieron los miembros de la Comisión Revisora de la Tesis, designada por el Colegio de Profesores de Estudios de Posgrado e Investigación del:

Centro de Investigación en Computación

para examinar la tesis titulada:

"Análisis espacial y supervisado de datos criminalísticos utilizando un sistema de información geográfica"

Presentada por el alumno:

Ruiz	**Gómez**	**José Javier**
Apellido paterno	Apellido materno	Nombre(s)

Con registro:

A	1	1	0	8	6	5

aspirante de: **MAESTRÍA EN CIENCIAS DE LA COMPUTACIÓN**

Después de intercambiar opiniones los miembros de la Comisión manifestaron ***APROBAR LA TESIS***, en virtud de que satisface los requisitos señalados por las disposiciones reglamentarias vigentes.

LA COMISIÓN REVISORA

Directores de Tesis

Dr. Miguel Jesús Torres Ruiz

Dr. José Giovanni Guzmán Lugo

Dr. Oleksiy Pogrebnyak

Dr. Cornelio Yáñez Márquez

M. en C. José María Ramos Rodríguez

Dr. Marco Antonio Moreno Ibarra

PRESIDENTE DEL COLEGIO DE PROFESORES

Dr. Luis Alfonso Villa Vargas

INSTITUTO POLITECNICO NACIONAL
CENTRO DE INVESTIGACION
EN COMPUTACION
DIRECCION

INSTITUTO POLITÉCNICO NACIONAL

SECRETARÍA DE INVESTIGACIÓN Y POSGRADO

CARTA CESIÓN DE DERECHOS

En la Ciudad de México el día 27 del mes Mayo del año 2013, el que suscribe **José Javier Ruiz Gómez** alumno del Programa de **Maestría en Ciencias de la Computación** con número de registro **A110865** adscrito a **Centro de Investigación en Computación**, manifiesta que es autor intelectual del presente trabajo de Tesis bajo la dirección del **Dr. Miguel Jesús Torres Ruiz** y cede los derechos del trabajo intitulado "*Análisis Espacial y Supervisado de Datos Criminalísticos utilizando un Sistema de Información Geográfica*" al Instituto Politécnico Nacional para su difusión, con fines académicos y de investigación.

Los usuarios de la información no deben reproducir el contenido textual, gráficas o datos del trabajo sin el permiso expreso del autor y/o director del trabajo. Este puede ser obtenido escribiendo a la siguiente dirección **javier.rugo@gmail.com**. Si el permiso se otorga, el usuario deberá dar el agradecimiento correspondiente y citar la fuente del mismo.

José Javier Ruiz Gómez

Nombre y firma

Resumen

El problema de seguridad pública en nuestro país es una constante en aumento, por citar un ejemplo, solamente en la Ciudad de México se presentaron en el año 2012 alrededor de 203,398 averiguaciones previas de delitos del fuero común. Por tanto, el contar con herramientas para analizar la información delictiva es una prioridad, lo cual es una tarea que se encuentra en etapas tempranas de desarrollo y análisis, comparado con otros países.

En el presente trabajo de tesis, se desarrolla una metodología basada en un Sistema de Información Geográfica (SIG) que genera una serie de análisis basados en técnicas espaciales y reconocimiento de patrones, los cuales generan un compendio de información relacionada con patrones de incidencia delictiva en puntos de interés, a tres niveles de análisis: *micro*, *meso* y *macro*. Como caso de estudio, se utiliza el Directorio Estadístico Nacional de Unidades Económicas (DENUE) de la delegación Gustavo A. Madero, con la finalidad de aplicar la metodología.

La metodología está compuesta por dos técnicas de análisis, utilizando métodos de generación de patrones y evaluación. La primera está basada en la densidad delictiva, utilizando la métrica de *City Block* para la etapa de evaluación. La segunda está relacionada con la secuencia de eventos delictivos, desarrollando una aproximación de secuencias y comparando con la *distancia de Levenshtein* estas secuencias temporales.

Posterior a la evaluación de los patrones, se lleva a cabo una interpolación de la información ponderada, usando una capa de datos raster, la cual se genera aplicando un algoritmo de triangulación para obtener diagramas de Voronoi. El resultado es una representación raster que indica las incidencias delictivas sobre una zona geográfica.

Abstract

The public security problem in our country is constantly increasing, to cite one example; only in Mexico City occurred in 2012 around 203,398 inquiries from common crimes. Therefore, for having tools to analyze criminal information is a priority, which is a task that is in early stages of development and analysis, compared with other countries.

In this dissertation, we develop a methodology based on a Geographic Information System (GIS) that generates a set of approaches based on spatial analysis and pattern recognition, which obtain a compendium of information related to crime rate patterns at points of interest, three levels of analysis: *micro*, *meso* and *macro* are proposed. As a case study, we used the National Statistical Directory Economic Units (DENUE) of the delegation Gustavo A. Madero, in order to apply the proposed methodology.

The methodology is composed of two analysis approaches, using methods of pattern generation and evaluation. The first is based on the criminal density, using the *City Block* metric for the evaluation stage. The second relates to the sequence of criminal events, developing an approximation of the sequences and comparing these temporal sequences with *Levenshtein distance*.

After the evaluation of the patterns, a weighted interpolation of the data using a raster data layer is carried out. This layer is obtained applying a triangulation algorithm in order to generate Voronoi diagrams. The result is a raster representation that indicates crime statistics over a geographical area.

Agradecimientos

El presente trabajo representa mi esfuerzo personal y el de personas que creyeron en mi para terminar este trabajo, primeramente quiero agradecer a mi familia, sin su apoyo esto no sería una realidad, agradecer a mis amigos por la confianza brindada y por los consejos de que no me echara para atrás en momentos difíciles, a Dzoara por exhortarme a ser mejor y apoyarme en estas decisiones, a mis asesores que sin sus consejos y sin su apoyo hubiera sido un camino más difícil además por la paciencia y las observaciones que hicieron en el transcurso del presente trabajo.

Gracias a todos

Tabla de contenido

Resumen 2

Abstract 3

Agradecimientos 4

Índice de figuras 8

Indice de tablas 10

Capítulo 1. Introducción 11

1.1 Generalidades 11

1.2 Enfoque de la tesis 17

1.3 Justificación 17

1.4 Planteamiento del problema 19

1.5 Hipótesis 22

1.6 Objetivos 23

1.6.1 General 23

1.6.2 Particulares 23

1.7 Organización de la tesis 23

Capítulo 2. Estado del arte 25

2.1 Desarrollos de sistemas SIG para el análisis criminal 25

2.1.1 Desarrollo de herramientas para el mapeo del crimen usando sistemas de información geográficos en Nueva Zelanda (Andy Gilmour y JillBarclay) 25

2.1.2 Minería de datos aplicada a la detección de patrones delictivos en Argentina (F. Valenga, E. Fernández, H. Merlino, D. Rodríguez, C. Procopio, P. Britosy R. García-Martínez, 2008) 27

2.1.3 Empleo de sistemas de información geográfica en el estudio del miedo al delito (L.Vozmediano , C. San Juan Guillen, 2006) 29

2.1.4 Criminología Ambiental (Environmental Criminology) 32

2.1.4.1 Modelo de análisis *34*
2.1.4.2 Análisis de problemas *34*
2.1.4.3 Análisis de operaciones *35*
2.1.4.4 Perfiles geográficos 35
2.1.5 Sistemas de información geográfica y el control de delitos en el espacio: caso de la comuna de Santiago (C.Contrera, 2002) 36
2.2 Desarrollos de análisis de ontologías con contexto geográfico 42
2.2.1 Sistema de soporte de decisiones basado en una ontología para la investigación de procesos criminales (D. Dzemydiene & E. Kazemikaitiene, 2005) 42
Capítulo 3. Marco teórico 49
3.1 Resumen 49
3.2 Ontologías 49
3.3 Bases de datos relacionales 53
3.3.1 Diseño conceptual de bases de datos modelo relacional 54
3.3.2 Transformación de modelo Entidad-Relación a Modelo Relacional 55
3.3.3 Modelado de bases de datos usando ontologías 57
3.4 Reconocimiento de patrones 58
3.5 La distancia de edición (distancia de Levenshtein) 60
3.5.1 Definiciones 60
3.5.2 Algoritmo de cálculo 62
3.5.3 Aplicaciones 63
3.6 Distancia Manhattan (Métrica City block) 64
3.6.1 Definiciones 64
3.6.2 Algoritmo de cálculo 64
3.6.3 Aplicaciones 65
3.7 Definición de densidad delictiva 65
Capítulo 4. Metodología 67
4.1 Introducción 67
4.2 Arquitectura del sistema 67
4.3 Marco general de trabajo 68

4.4 Diagrama de componentes 70
4.4.1 Etapa de adquisición 70
4.4.2 Etapa de conexión 71
4.4.3 Etapa de análisis 73
4.4.3.1 Proceso de generación de patrones 77
4.4.3.2 Patrones de densidad delictiva 78
4.4.3.3 Patrones de incidencia temporal 80
4.4.3.4 Proceso de análisis de patrones 81
4.4.4 Etapa de validación 83
4.4.4.1 Métrica para el cálculo de densidad 84
4.4.4.2 Métrica para el cálculo de incidencias temporales 85
4.4.5 Etapa de visualización 87
Capítulo 5. Pruebas y resultados 89
5.1 Introducción 89
5.2 Cálculo de densidades 89
5.2.1 Densidad total 91
5.2.2 Comparación de datos de densidades 94
5.2.3 Comparativa de 3 Niveles de Análisis 98
5.3 Patrones temporales 101
5.3.1 Variación de rangos de valores 105
Capítulo 6. Conclusiones y trabajos futuros 108
6.1 Introducción 108
6.2 Aportaciones 109
6.3 Limitaciones 110
6.4 Trabajos Futuros 110
Referencias 112

Índice de figuras

Figura 1.1 Modelo conceptual de GIS propuesto por Tomlinson 12
Figura 1.2 Estudio de proximidad de plantas nucleares en EUA, representando su zona de afectación nuclear (M.Fernandez,2011) 20
Figura 1.3 Ejemplo de centroide en un polígono no convexo 20
Figura 1.4 Ejemplo de interpolación de datos 21
Figura 1.5 Imagen de reconocimiento de patrones a través de una red neuronal 21
Figura 2.1 Distribución de Clusters 29
Figura 2.2 Densidad del Delito en Donostia San Sebastian, lugar seleccionado para prueba 31
Figura 2.3 Comparativo de Informaicón recaudada en cuanto a persepción del delito 32
Figura 2.4 Formato empleado para el reporte de delitos 38
Figura 2.5 Ejemplificaci'on del artículo sobre la técnica de representación de imagenes por medio de coropletas 40
Figura 2.6 Representación de la densidad de delitos en la comuna de Santiago 41
Figura 2.7 Análisis de densidad del delito aplicando coropletas 41
Figura 2.8 Ejemplo de diagrama de clase del objeto de caracter del delito en UML 44
Figura 2.9 Secuencia diagrama del proceso de pruebas de investigación forense 46
Figura 2.10 Vizualización de relaciones ontológicas del sistema 47
Figura 3.1 Ejemplo de definición de relación con ONTOLINGUA 50
Figura 3.2 Ejemplo de declaración de funciones con ONTOLINGUA 50
Figura 3.3 Ejemplo de generación de una clase para una ontología en ONTOLINGUA 51
Figura 3.4 Especificación de los requisitos de una ontología descrito por Barchini, Álvarez, Herrera y Trejo 53
Figura 3.5 Transformación de entidades fuertes 55
Figura 3.6 Transformación de relación de varios a varios 56
Figura 3.7 Transformación de entidades de orden n 56
Figura 3.8 Esquema de relaciones uno a muchos 57
Figura 3.9 Diagrama Entidad-Relación para el dominio de viaje 58
Figura 3.10 Matriz de Similitud de las palabras DIFÍCIL y FÁCIL 63
Figura 3.11 Comparativo entre la distancia Manhattan(coloreada en tono azul) y la distancia Euclidiana(descrita con color verde) 65
Figura 4.1 Arquitectura general de la aplicación 67
Figura 4.2 Arquitectura general de la aplicación por componentes 69
Figura 4.3 Diagrama de componentes de la aplicación 70
Figura 4.4 Modelo de datos propuesto para el sistema 72
Figura 4.5 Caso de uso de generación de patrones 78
Figura 4.6 Densidad de delitos distribuidos espacialmente 79
Figura 4.7 Caso de uso de patrones de incidencia temporal 80
Figura 4.8 Patrones de incidencia temporal 82
Figura 4.9 Diagramas de Voronoi obtenidos 83
Figura 4.10 Representación visual de un mapa de densidad de delitos en el sistema 87
Figura 5.1 Opción para el cálculo de patrones 89

Figura 5.2 Patrones de densidad generados ____ *90*
Figura 5.3 Datos de entrada para calcular el análisis de densidad de delitos ____ *91*
Figura 5.4 Valores obtenidos al aplicar el cálculo con la métrica de City Block ____ *91*
Figura 5.5 Resultado del análisis de densidad delictiva, mostrando detalles de la información con respecto a un foco delictivo representado por un punto ____ *92*
Figura 5.6 Resultado obtenido al analizar el id 21301 ____ *95*
Figura 5.7 Análisis comparativo entre los puntos que describen incidencia delictiva obtenidos por el sistema ____ *98*
Figura 5.8 Comparación de análisis a tres niveles de detalle ____ *100*
Figura 5.9 Datos de entrada para el cálculo de patrones temporales con n=10 ____ *102*
Figura 5.10 Patrones temporales generados ____ *102*
Figura 5.11 Datos de entrada para generar el análisis temporal ____ *103*
Figura 5.12 Valores de la comparación para un análisis temporal a nivel micro (DENUE id = 21301) ____ *103*
Figura 5.13 mapa de similitud temporal a nivel micro para (DENUE id = 21301) ____ *104*
Figura 5.14 Mapa de similitud temporal a nivel meso para (DENUE id = 21301) ____ *105*
Figura 5.15 Mapa de similitud temporal a nivel meso para (DENUE id = 21301), con un patrón de n = 20 ____ *106*
Figura 5.16 Agregación de puntos a los analisis cuando aumenta el nivel de detalle ____ *107*

Índice de tablas

Tabla 3.1 Ejemplo de una tabla en el modelo Entidad Relación.....***¡Error! Marcador no definido.***
Tabla 4.1 Pseudo código del algoritmo para el cálculo de la Triangulación de Voronoi. *74*
Tabla 4.2 Tabla de insidencia de delitos por tipo *79*
Tabla 4.3 Ordenamiento de delitos clasificados por su tipo. *81*
Tabla 4.4 Aplicación del método de City Block. *84*
Tabla 4.5 Código correspondiente a la métrica densidad de delitos. *84*
Tabla 4.6 Código correspondiente a la métrica de incidencias temporales. *85*
Tabla 5.1 Datos relativos a un punto de incidencia delictiva referente al DENUE 21301. ..... *93*
Tabla 5.2 Datos relativos a un punto de incidencia delictiva referente al DENUE 3229. ..... *95*

Capítulo 1. Introducción

En este capítulo se presentan las generalidades de este trabajo de tesis, describiendo los objetivos, la justificación del mismo, así como el planteamiento del problema, lo cual conduce a una hipótesis sobre el mismo. Finalmente, se presenta una sección que describe la organización de esta tesis.

1.1 Generalidades

Desde el comienzo de los tiempos, se han utilizado los mapas para representar información y los eventos que ocurren sobre la superficie terrestre. Es así, como a través de los siglos las técnicas cartográficas se han ido desarrollando, sirviendo actualmente para múltiples objetivos, convirtiendo a los Sistemas de Información Geográfica, en una herramienta esencial para el estudio y seguimiento de procesos espaciales ocurridos a toda escala.

En la era de la visualización científica, los Sistemas de Información Geográfica (GIS de *Geograhic Information Systems*) han incrementado la potencialidad del análisis de información y de la investigación científica, comparado con el empleo de la cartografía tradicional, no solo por la relación existente entre las técnicas cartográficas y la informática para la descripción y análisis mediante un amplio uso de símbolos, lo cual ha permitido el uso de mapas en diferentes áreas como son la astronomía, aplicación de sensores remotos, la geología, la climatología, la criminología, entre otras, lo cual le confiere un nuevo campo de potencialidades [1].

A menudo, se considera que los mapas son únicamente una herramienta de visualización y representación de objetos geográficos. Sin embargo, los mapas tienen un papel de gran alcance en el proceso de investigación, análisis y presentación de resultados relacionados con la información geográfica en diferentes dominios de aplicación.

En esta interpretación restrictiva, un mapa es más que una imagen o parte de una base de datos. [2] y [3] señalan la distinción entre el pensamiento visual y la comunicación visual en el uso de mapas y gráficos, además de que un mapa representa una sólida base para

ser una herramienta fundamental para la investigación científica por la información que en ella se representa.

Un SIG es un producto horizontal por la variedad de usos que tiene en los entornos industriales, esta razón es suficiente para que no exista una definición simplista. Sin embargo, no carece de una idea general que brinde una compresión de ellos en un ámbito más general. Tomlinson propuso un modelo conceptual y holístico que permite describir los componentes principales de un Sistema de Información Geográfica [4]. El modelo general propuesto está provisto de los elementos que se muestran en la Figura 1.1.

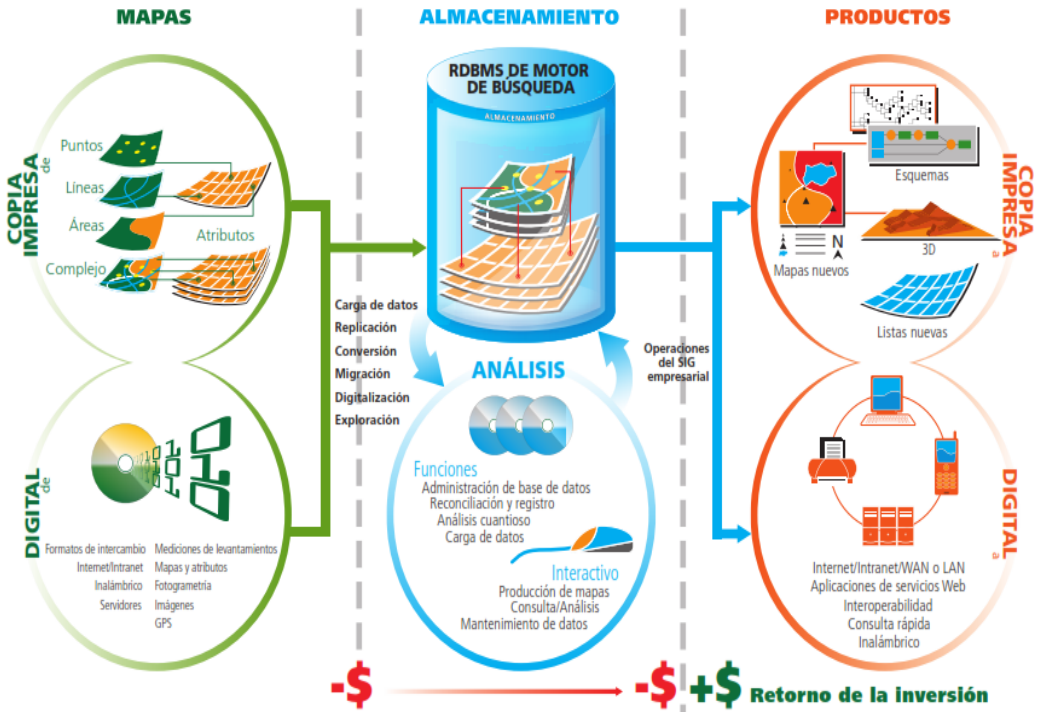


Figura 1.1 Modelo conceptual de GIS propuesto por Tomlinson

Para poder hacer un análisis de este modelo, debe estudiarse parte por parte .La parte de la izquierda representa los datos espaciales que no se han procesado, estos datos tienen la peculiaridad de tener una componente geográfica, representada por las figuras de la izquierda (zona de color verde); dichos datos son representados por tres primitivas básicas (dependiendo de la escala) pueden ser puntos en un espacio, líneas o polígonos. A su vez, dichos datos tienen una descripción que los caracteriza teniendo como base una localización como información descriptiva. Estos datos pueden ser capturados de diferentes formas, por ejemplo, utilizando un sistema GPS o a través de la digitalización de

mapas a lo que representa el segundo componente del modelo de Tomlinson, que son las bases de datos y el almacenamiento digital de los datos (parte central del modelo), en ella se almacenada la información con un formato y depuración de los mismos para ser representada a priori, siendo esta la base de información de un SIG.

La parte derecha del modelo, está representada por la aplicación propiamente, de la cual cabe señalar que está descrita por una herramienta de cómputo y un objetivo particular. Por último, en el modelo se contempla el capital humano, del cual dependen los esfuerzos por interpretar esa información y sacar provecho a las ventajas que ofrece un SIG.

Los sistemas de información geográfica se han utilizado para el estudio de los aspectos espaciales relacionados con el temor al delito en contextos urbanos, para lo cual podemos citar un caso concreto: el estudio del miedo al delito en [5].

A partir de un conjunto de encuestas que fueron aplicadas a una muestra de residentes, se estudió la distribución del delito objetivo en la ciudad, a partir de datos judiciales y la distribución del miedo al delito en tres escenarios urbanos distintos [5], lo que permite describir la forma de vincular aplicaciones de los SIG a un contexto psicológico-social.

En otro caso de estudio, en Argentina se realizó un análisis de los patrones delictivos empleando técnicas de minería de datos y estadística a un conjunto de datos obtenidos de un sistema de geoprocesamiento espacial; por medio de la minería de datos se elaboró un proceso iterativo de búsqueda de información no trivial en grandes volúmenes de datos con el propósito de generar información lo más similar posible a la que podría generar un experto humano: por medio de patrones, identificando asociaciones, cambios, así como anomalías y estructuras significativas [6]. Con el análisis anterior se generó un valor agregado y nuevo conocimiento, partiendo de un conocimiento previo el cual fue sometido a técnicas de análisis de datos y estadística aplicada para el desarrollo de dichos estudios.

El estudio vinculado del crimen, tratando de estudiarla como una actividad que puede relacionarse a un contexto geográfico delimitado, ha permitido el trazado de mapas criminales, los cuales son una herramienta de análisis para tratar de determinar los patrones conductuales y tendencias. Este estudio se basa en la determinación de patrones

que permiten realizar una serie de estudios analíticos del problema, el conductual y el de los recursos disponibles sobre el mismo sistema y el mapa criminal [7].

A lo largo de una década y media, los SIG han buscado la representación y el mapeo de los eventos geográficos, los cuales tienen características muy específicas como son el lugar de incidencia así como un tipo de evento que los caracteriza. En el caso de las actividades humanas se les pueden categorizar de manera geográfica, siendo las actividades criminales un aspecto puramente antropológico del ser humano, dichas actividades pueden ser categorizadas en este tipo de aplicaciones [8].

En un sentido estricto, cualquier sistema que permita la representación y el análisis de la información geográfica es un SIG, siempre y cuando se sigan las características básicas y representativas de estas aplicaciones. Para algunos autores un SIG es un software cuyo propósito es la manipulación de los datos geográficos. Sin embargo, las aplicaciones comerciales son cerradas, costosas y propietarias, lo cual no permite el desarrollo de manera integral de aplicaciones de propósito específico.

Considerando los ejemplos que fueron descritos anteriormente, estos estudios de carácter específico pueden aplicarse a diferentes ramas o ciencias, a las cuales se les pueda representar con base en eventos o fenómenos con una relación espacio-temporal, este es el caso de la criminología.

La criminología es una ciencia de carácter multidisciplinaria que basa sus fundamentos en conocimientos propios de la sociología, la psicología y la antropología social, tomando para ello el marco conceptual que delimita el derecho. La criminología estudia las causas del crimen y preconiza los remedios del comportamiento antisocial del hombre. Las áreas de investigación criminológicas incluyen la incidencia y las formas del crimen, así como sus causas y consecuencias. También reúnen las reacciones sociales y las regulaciones gubernamentales respecto al crimen. Debido a su carácter multidisciplinario, la integración y el desarrollo de herramientas que permitan obtener información para su investigación y estudio, proporciona la apertura para el desarrollo de sistemas que puedan procesar información referente a este fenómeno.

Dentro de la criminología existe un subgénero (la criminología ambiental) que se dedica particularmente a estudiar los sucesos criminales como resultado de la relación entre ofensores y objetivos delictivos potenciales que se dan en puntos específicos del espacio y del tiempo [9]. Dicha rama es la que se puede emplear dentro de nuestro campo de estudio, por su evidente relación con los SIG.

Aunque formalmente, este fenómeno no representa completamente una distribución continua, varios eventos criminalísticos pueden ocurrir en diversos puntos aislados en un espacio geográfico particular. Por lo tanto, es posible estimar los valores entre los puntos conocidos para poder ser capaces de realizar un mapa criminal de la zona como lo realiza Fritz en su estudio descrito en [7]. Hoy en día, y hablando concretamente de nuestra sociedad, las actividades criminales siempre han sido una actividad prioritaria a considerar en diferentes zonas del país, haciéndolo de esta manera un fenómeno problemático a lo largo no solo de México sino de otros países traspasando nuestros límites territoriales.

La relevancia de este fenómeno puede determinarse mediante un perfil geográfico, siendo ésta una técnica de análisis de la actividad espacial o geográfica de los delincuentes que se aplica en la investigación criminal. Con esta técnica se está logrando dotar a investigadores policiales y criminólogos de un perfil del fenómeno que incluye movimientos geográficos, siendo el objetivo fundamental poder inferir o describir el comportamiento de eventos delictivos para así prevenir el desarrollo del mismo y determinar las zonas de impacto mediante un SIG [10].

En muchos casos prácticos, esta técnica a sido utilizada con éxito en las investigaciones de delitos seriales, en los que la policía no cuenta con pistas ni indicios forenses que puedan guiar una investigación. En estos casos, el disponer de un perfil geográfico puede reducir el número de zonas de vigilancia para un tipo de delito, mejorando así el despliegue de operativos a los sectores geográficos con mayor incidencia delictiva, auxiliando de esta manera en la toma de decisiones. A este tipo de zonas afectadas se les conoce comúnmente como "*hotspots*", definidos como los lugares donde existe una mayor incidencia de un evento delictivo en [11]. Estos estudios han ayudado significativamente en las particularidades de problemáticas relacionadas con un contexto geográfico, respondiendo generalmente a la siguiente pregunta: ¿Qué cosa está próxima a otra?

Estos análisis han generado estudios sobre las diferentes zonas de influencia que yacen sobre un lugar a un determinado fenómeno, particularmente en el caso de estudio de la criminología.

De igual forma, partiendo de estos lugares "*hotspots*" resulta más sencillo el desarrollo de técnicas que permitan deducir patrones, ya que de esta manera se define la estructura básica de lo que será una entidad nueva, la cual se encuentre relacionada con los eventos particulares, permitiendo el desarrollo de una capa temática espacial para la creación y adquisición de información más enriquecida. Esta información plasmada en mapas temáticos aporta los siguientes datos:

- Integra diferentes tipos de información de una manera coherente y relacionada.
- Provee una visualización intuitiva de eventos y casos relacionados, describiendo eventos y patrones que a investigadores forenses y criminalistas les puede aportar pistas.
- Genera la base para el análisis espacial y las relaciones de eventos.
- Genera una perspectiva del alcance del problema y puede permitir desplegar recursos de manera más eficiente para su desarrollo.

Asimismo, bajo la definición de estos *hotspots* y de las entidades del crimen como tal, existe la posibilidad de generar una red neuronal, con la finalidad de conocer las relaciones que puedan existir entre entidades, cuyo valor aparente no estén ligadas pero que con base en un análisis más detallado, se puedan encontrar relaciones entre dos eventos; es decir, crímenes con relación espacial, lo que ha permitido, por ejemplo, el caracterizar los centros de venta de droga próximos a la ubicación de teléfonos públicos [12].

También es para considerar el uso de las superposiciones, de las cuales se obtendrán nueva información juntando las capas generadas por los análisis de proximidad, recordando que fundamentalmente los SIG trabajan bajo capas, de las cuales desprendemos información relacionada a puntos, líneas y áreas, de esta manera se generará información nueva al combinar datos de dos capas.

En consecuencia, el presente trabajo está orientado en el desarrollo de un Sistema de Información Geográfica enfocado al análisis delictivo e interpretación de los resultados referentes a este fenómeno social, utilizando técnicas de procesamiento semántico de datos espaciales, así como de reconocimiento de patrones.

1.2 Enfoque de la tesis

Esta tesis se encuentra orientada a ofrecer un análisis espacial, utilizando técnicas de análisis espacial y reconocimiento de patrones para determinar el comportamiento de la criminalidad en un caso de estudio particular, enfocándose en algún área geográfica específica, implementando una aplicación SIG de escritorio.

Con esta tesis se intenta implementar una serie de características que puedan ser usadas en un sistema de información geográfica. La finalidad de esta aplicación consiste en proyectar actividades criminales a través de los históricos; de esta manera, se pretende realizar un estudio comparativo de zonas con base en sus incidencias. Con lo anterior, se propone una capa temática con los resultados que este tipo de proyección arroje para que puedan servir en un análisis detallado por parte de los investigadores relacionados con el campo.

La tesis promueve el desarrollo de una base de conocimiento, en donde se defina de manera explícita una conceptualización del dominio, la cual se utilizará en trabajos futuros para el diseño de una base de datos geográfica y para algunos procesos de análisis semántico.

1.3 Justificación

El índice delictivo ha cobrado un gran marco de estudio, los cartógrafos, los sociólogos, los criminólogos, los ingenieros y los investigadores han visto la necesidad de contar con herramientas y técnicas que permitan un mejor entendimiento de este fenómeno para la toma de decisiones. Por lo tanto, esta tesis busca contribuir en el desarrollo de ese entendimiento, mediante el desarrollo de una aplicación que permita realizar un análisis de los datos geográficos con una temática delictiva para el análisis de inteligencia delictiva.

Los agentes de la ley se enfrentan día a día a retos relacionados con mejorar el desempeño de sus actividades y para ello es necesario contar con información que les permita poder desplegar sus recursos de manera más eficiente, ya que no basta con saber la ubicación de las incidencias reportadas, porque es necesario generar estrategias para prevenir nuevos incidentes.

Las necesidades que generalmente se presentan para los investigadores y forenses son las siguientes:

1. Encontrar patrones de incidencia de los cuales los agentes e investigadores se puedan basar para proyecciones, estudios, desarrollos y técnicas que los lleven al mejor entendimiento del fenómeno y su relación con sus lugares determinados.
2. Contar con herramientas de predicción y proyección del delito para poder desarrollar e implementar operativos en búsqueda de combatir el fenómeno delictivo de manera más eficiente.
3. Buscar la predicción de patrones y relaciones entre dos actividades criminales, para encontrar de esta manera secuencias de incidentes delictivos que tengan relación durante un periodo de tiempo y espacio específicos.

Actualmente en el país, no se han desarrollado herramientas que permitan el estudio de este tipo de eventos. Además, las ramas de investigación para posibles proyectos futuros, quedan abiertas para el desarrollo de perfiles más complejos, donde se pueden involucrar estudios criminológicos, sociales, psicológicos, cartográficos, antropológicos, estadísticos, matemáticos, por nombrar algunos.

Otro planteamiento a considerar en el contexto de la incidencia delictiva, es el contar con información cartográfica en formato digital, la cual pueda ser utilizada para realizar los análisis espaciales y estadísticos correspondientes sobre las bases de datos que se tienen; con la finalidad de aprovechar la gran cantidad de información con que se cuenta y hacer un análisis de inteligencia policiaca que sea de utilidad en el control, atención y prevención del delito.

1.4 Planteamiento del problema

Hoy en día, las agencias encargadas para el control y el combate a la delincuencia, han enfrentado tanto un incremento en dichas actividades como un aumento en la complejidad del desarrollo de los mismos, es común encontrar que los crímenes responden a patrones establecidos de bandas o acciones recurrentes.

Supongamos el siguiente caso, existe una organización criminal que opera en determinada zona, se puede asumir que su objetivo es el robo a negocios de una zona específica y su manera de operar es la siguiente: efectúan el robo a mano a armada en un negocio de una zona y posterior a esto aprovechan la confusión ayudándose de las vías rápidas obteniendo un vehículo también mediante el robo.

En este caso particular, tenemos dos tipos de delitos, con dos reportes distintos y dos averiguaciones que estudiando los delitos de forma aislada se pueden interpretar como crímenes diferentes, sin embargo, al analizar la situación en conjunto claramente apreciamos una secuencia lógica en la que se involucran incidentes en una zona y tiempo específicos que están relacionados.

El problema es el siguiente ¿Cómo sería posible encontrar este tipo de relaciones mediante técnicas de análisis para poder predecir patrones y zonas de alta incidencia delictiva?

Esta pregunta se responde con el uso de técnicas detalladas de análisis, planteado primeramente un análisis de proximidad (un ejemplo de este tipo de análisis se puede ver en la Figura 1.2). Retomando nuestro casos hipotético, tenemos el reporte de un robo en determinada zona a un establecimiento, en este caso también tenemos el reporte de un robo a vehículo en una zona próxima en circunstancias de tiempo que varían poco.

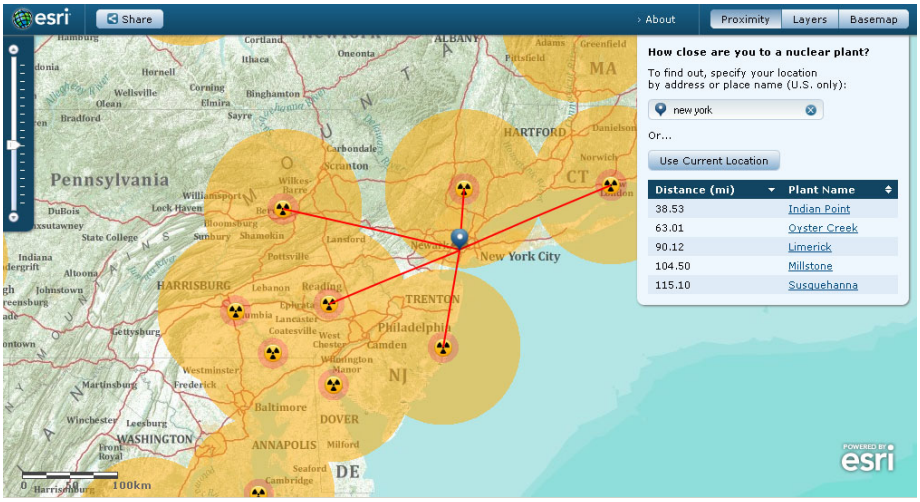


Figura 1.2 Estudio de proximidad de plantas nucleares en EUA, representando su zona de afectación nuclear (M.Fernandez,2011)

De igual manera en el análisis propuesto, se pretende generar una capa con la zona de afectación característica de un delito, previamente identificando el centroide ponderado de donde se alberga la mayor cantidad de eventos delictivos determinados para un delito X, ver Figura 1.3.

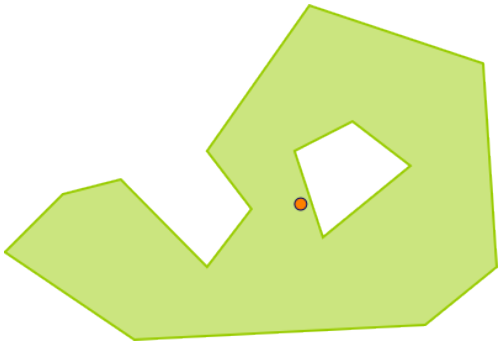


Figura 1.3 Ejemplo de centroide en un polígono no convexo

Una vez que se tiene identificado el centroide y estas "zonas calientes" (los *hotspots*), es necesario buscar relaciones en la información utilizando espacios de periodo temporales, por lo que el análisis de la proximidad del fenómeno no es suficiente.

Esto generara una serie de datos nuevos, a los cuales será necesario aplicar una técnica de interpolación (ver Figura 1.4) para categorizarlos en el espacio físico, con esto también se puede proyectar el fenómeno.

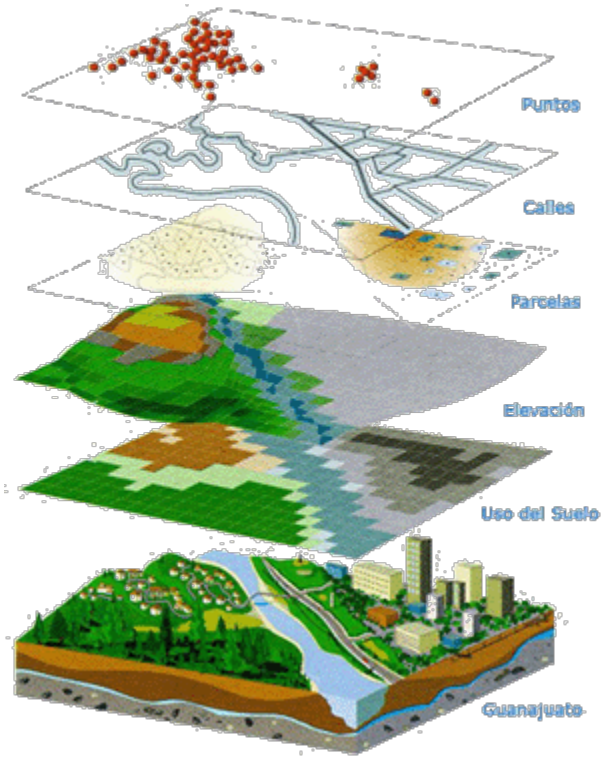


Figura 1.4 Ejemplo de interpolación de datos

Ahora bien, en un inicio se planteó la relación existente entre eventos delictivos aparentemente aislados, por lo cual se propone el uso de técnicas de reconocimiento de patrones sintáctico (ver Figura 1.5), con el propósito de poder descubrir tendencias ,así como también llevar un patrón estadístico más completo.

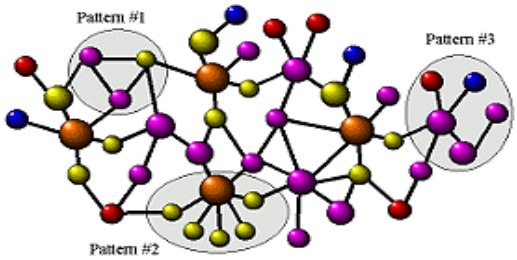


Figura 1.5 Imagen de reconocimiento de patrones a través de una red neuronal

Por lo anteriormente descrito, será necesario el desarrollo de un sistema donde se tengan los datos de los reportes de incidencias delictivas que puedan facilitar su consulta y despliegue de la información de una manera intuitiva y natural, pero que al mismo tiempo brinde el soporte con herramientas que estén facultadas con los análisis mencionados. Esto ayudará a la toma de decisiones referentes a las acciones a realizar respecto a este

tipo de actividades, siendo éstas un punto fuerte, crítico y vital de las operaciones realizadas por las agencias que vinculen estadística criminal así como las gubernamentales que están encargadas en el combate de este fenómeno.

Para realizar estos análisis será necesarios hacer uso de un SIG, ya que permite contar con una serie de herramientas que ayudan al desarrollo de las técnicas analíticas antes mencionadas, considerando para ello una herramienta que sea de código abierto para facilitar también el desarrollo y disminuir los costos.

1.5 Hipótesis

La presente tesis tiene el objetivo de proveer mecanismos de análisis que puedan brindar respuesta a las siguientes interrogantes sobre el campo de la actividad criminal:

- ¿En qué zonas es más propenso un tipo de delito sobre otro?
- ¿Qué relaciones existen entre los tipos de delito?
- ¿Qué evolución tendrá la actividad criminal para un tiempo determinado?
- ¿Dónde se encuentran las áreas de incidencia delictiva y qué rangos presentan?

Por lo tanto, la hipótesis de investigación se enfoca en el diseño de una base de datos geográfica, a partir de una representación conceptual (ontología de aplicación); en donde esta estructura permita el diseño de la base de datos geográfica, así como también sirva de entrada para el análisis semántico, complementado con el análisis espacial para determinar la incidencia delictiva en alguna área geográfica específica. Con lo anterior, se pueden generar mapas sobre índices delictivos, correlación de delitos, estrategias de combate, comportamiento de diversos fenómenos, patrones de búsqueda en un ámbito espacio-temporal, tendencias y proyecciones delictivas en áreas geográficas, entre otros resultados.

Para poder resolverlas preguntas descritas anteriormente, suponemos que con el desarrollo de un SIG se pueden encontrar las relaciones existentes sobre los tipos de delito en una zona, dichas relaciones pueden ser encontradas y proyectadas con un conjunto de técnicas bien definidas, como es el caso de los análisis de proximidad, con lo cual se determinan las zonas que tienen más índice delictivo sobre un hecho,

determinando el factor de distancia entre los eventos y proyectando esa zona como un sector con un índice crítico correspondiente a una escala establecida.

1.6 Objetivos

En esta sección se definen los objetivos tanto general y particulares que se pretenden con el desarrollo de la presente tesis.

1.6.1 General

Diseñar e implementar un Sistema de Información Geográfica (SIG) para mejorar las tareas relacionadas con el análisis delictivo empleando técnicas de análisis espacial y reconocimiento de patrones.

1.6.2 Particulares

- Diseñar una aplicación SIG de escritorio para mostrar los resultados de un análisis delictivo.
- Diseñar una base de datos geográfica, basada en una representación conceptual dentro del dominio de aplicación de la incidencia delictiva.
- Generar vectores de patrones de densidad delictiva y cadenas de secuencia de densidad temporal.
- Evaluación de patrones para generar un mapa comparativo de los puntos referentes a la actividad criminal.

1.7 Organización de la tesis

Esta tesis está compuesta por seis capítulos considerando que al inicio de cada uno se presenta una introducción sobre la información que en él se encuentra, asimismo con las conclusiones y un resumen general tratado en cada uno de ellos. A continuación se describen los aspectos más importantes que se abordan en cada uno de los capítulos que integran el presente trabajo de tesis:

Capítulo 1. Este capítulo presenta una introducción general a la tesis, proporcionando también los objetivos del trabajo, el enfoque, la justificación que señala la relevancia del porque se está desarrollando este trabajo, la hipótesis de investigación, así como los alcances y limitaciones del trabajo.

Capítulo 2. En este capítulo se detalla el estado del arte y el enfoque que se le ha dado a este trabajo. El capítulo concluye con una crítica sobre el estado del arte y sobre lo propuesto en el desarrollo de un SIG Web orientado al análisis criminal.

Capítulo 3. En este capítulo se presenta el marco teórico de la tesis, del cual parten todos los elementos primordiales de los cuales se basa el desarrollo de la misma.

Capítulo 4. Este capítulo describe la metodología que ha sido empleada en el desarrollo de la tesis y como está situada en el desarrollo de un sistema que permita relacionar la criminología con los aspectos más naturales de un SIG.

Capítulo 5. En este capítulo se presentan y discuten los resultados obtenidos al aplicar la metodología en el caso de estudio en particular, describiendo de esta manera las validaciones que se siguieron para la adquisición de dichos resultados.

Capítulo 6. El capítulo está destinado para las conclusiones del trabajo, haciendo énfasis en los alcances logrados, así como en las limitaciones y en las posibilidades de trabajos futuros.

Capítulo 2. Estado del arte

En este capítulo se describirán los principales trabajos relacionados con el desarrollo de herramientas y análisis criminal, ya que estos trabajos son un punto de partida para el desarrollo de este trabajo de tesis; se describen cada uno de los trabajos junto con sus aportes, para de esta manera tratar de determinar la posibilidad de usarlo en la presente propuesta. Se mencionarán los resultados de técnicas y análisis que se han empleado en estudios de criminalística, así como propuestas enfocadas al desarrollo de nuevas técnicas.

2.1 Desarrollos de sistemas SIG para el análisis criminal

En esta sección se estudian desarrollos previos que han tenido lugar en otras partes del mundo relacionados con la creación y el análisis de herramientas SIG, cuyo principal objetivo ha sido el control de las actividades delictivas.

2.1.1 Desarrollo de herramientas para el mapeo del crimen usando sistemas de información geográficos en Nueva Zelanda (Andy Gilmour y Jill Barclay)

A mediados de la década de los 90s, la policía de Nueva Zelanda desarrolló un nuevo sistema de registro criminal [13]. Debido a una visita que hicieron a norte América, el Centro de Información y Comunicación Tecnológica (ICT) de la policía de Nueva Zelanda, analizó la factibilidad de contar con la implementación propia de un SIG para uso local. Dentro de algunas de sus funciones, se podían hacer consultas de los sistemas policiacos sobre patrones de actividad criminal en determinadas zonas geográficas.

Las problemáticas que se presentaron al desarrollar dicha aplicación (citando en un orden de mayor a menor medida) estuvieron enfocadas en dos grandes factores; en el primero de ellos, se desconocía el funcionamiento y los paquetes de una aplicación SIG, en consecuencia, parte del grupo de analistas involucrado, no visualizaba el potencial del desarrollo de la herramienta, cuestionando entre otras cosas tanto la rentabilidad como la usabilidad de la aplicación.

El sistema fue desarrollado con una tecnología cliente-servidor, inicialmente la aplicación fue instalada en 19 sitios, posteriormente pasó a un total de 33 lugares. El sistema que había sido implementado tenía un mantenimiento costoso y difícil, es por ello que se consideró contar con una alternativa web, ya que este tipo de aplicaciones reducen en gran medida tanto el costo como la complejidad asociados al mantenimiento.

Por lo anterior, se decidió cambiar la estructura de la aplicación pero se optó por utilizar una arquitectura basada en el cliente, consiguiendo un ahorro en lo que concierne al mantenimiento e incrementando el acceso a la aplicación por parte de los usuarios finales.

Con el crecimiento del sistema, se le asignó el nombre de MAPS *(Map-Based Analytical Police System)*. Actualmente, el sistema sigue funcionado como MAPS versión 4.0 el cual realiza una vinculación de diferentes fuentes de información para su operación, de las cuales destacan las siguientes:

El *Communications and ResourceDeployment* (CARD), el cual es un centro de información del que se pueden extraer todas las llamadas de emergencia que operan en tres centros CARD, transfiriendo información de consulta para el MAPS, el cual puede usar la información que recibe para mapearla en pantalla, con 15 minutos de retraso, después de que la información fue capturada en alguno de los centros.

La otra fuente de información con la que cuenta es el *National Intelligence Application* (NIA) el cual administra los registros e información de operación de la policía de Nueva Zelanda, la cual contiene entre otras cosas, información de inteligencia y de investigaciones.

El MAPS es un sistema que tiene ya más de 10 años en funcionamiento, con lo cual es un sistema que ha madurado en muchos aspectos, uno de los que podemos mencionar, es el de las limitantes que se tuvieron en un inicio en lo que respecta al entendimiento de las tecnologías que se usan en la implementación de un sistema SIG.

Algo a tomar en consideración, es la cohesión que debe existir en los datos, si un sistema de esta naturaleza es alimentado con información de diferentes bases de datos, se debe tener cuidado en que exista una relación de conceptos, así como verificar que las

entidades o las relaciones de una base de datos preserven la semántica con lo existente en otras fuentes de información.

2.1.2 Minería de datos aplicada a la detección de patrones delictivos en Argentina (F. Valenga, E. Fernández, H. Merlino, D. Rodríguez, C. Procopio, P. Britosy R. García-Martínez, 2008)

Esta investigación [6] hace énfasis en el análisis de los registros criminales, considerados por los autores como parte fundamental para la prevención del delito, así como para el diseño efectivo de políticas y planes de prevención. En Argentina, este tipo de análisis se ha realizado históricamente mediante herramientas estadísticas, descriptivas o deductivas, considerando fundamentalmente variables y relaciones primarias en los registros criminales.

Su investigación profundiza sobre los resultados obtenidos con la estadística descriptiva, buscando la interrelación de las variables más complejas que conllevan al contexto real, lo cual requiere un tratamiento estadístico más detallado, que permitió experimentar con técnicas más profundas de análisis de variables.

En general, el tamaño de las bases de datos está basado en aspectos como la capacidad y eficiencia de almacenamiento y no en su posterior uso o análisis. Por esta razón, en muchos casos, los registros almacenados son demasiado grandes o complejos para poder analizarse y superan el alcance de la estadística .

La Minería de Datos es un proceso iterativo de búsqueda de información no trivial en grandes volúmenes de datos que busca generar información similar a la que podría generar un experto humano: patrones, asociaciones, cambios, anomalías y estructuras significativas como lo mencionan en [6] su investigación, lo cual fue indispensable para el desarrollo de la investigación considerando las limitantes de la estadística clásica.

Sumando a ello, en el desarrollo de su análisis y refiriéndose al aspecto criminal, la gran cantidad de variables y vertientes que ocurren en el entorno real de un fenómeno criminal, fue la base para la justificación de uso de la minería de datos; con el propósito de realizar un análisis estadístico más profundo del fenómeno en su entorno.

Otros aspectos fundamentales que sirvieron de motivación para este trabajo fueron dos aplicaciones existentes ya desarrolladas para ese fin, COPLINK, el cual es un sistema desarrollado en la Universidad de Arizona en el año de 1997, el cual, posee dos características indiscutibles, integración de información con diferentes bases de datos de corporaciones policiales con base en su módulo CopLink Conect y el despliegue de búsqueda de patrones en los datos compuestos a través de su segunda componente, CopLink Detect.

El otro sistema es OVER que fue creado en el año 2000 y está especializado en casos de robo a domicilios particulares. Se enfoca en la identificación de recursos y la elaboración de estrategias de prevención, realizando un análisis de los recursos trata de encontrar aquellos que son críticos y con base en ellos hacer más eficientes los procesos de investigación y combate.

De igual manera, OVER analiza la información y clasifica aquella que será relevante para la investigación, por medio de la parte forense así como los hechos registrados en las declaraciones, de esta manera, se abarca uno de sus objetivos primordiales, el desarrollo de una análisis espacio-temporal para confirmar las suposiciones de las tendencias y patrones dictaminados.

Haciendo uso del algoritmo *k*-means simple se tratan de generar clases (clusters) de la información obtenida, después de lo cual se emplea una inducción para dar explicación al comportamiento de forma más descriptiva, esto se obtiene utilizando el algoritmo J48.De esta manera se genera un árbol clasificador para categorizar en dichos clusters el tipo de delito y sus características, dando como resultado un valor agregado significativo, esto se puede apreciar en la Figura 2.1.

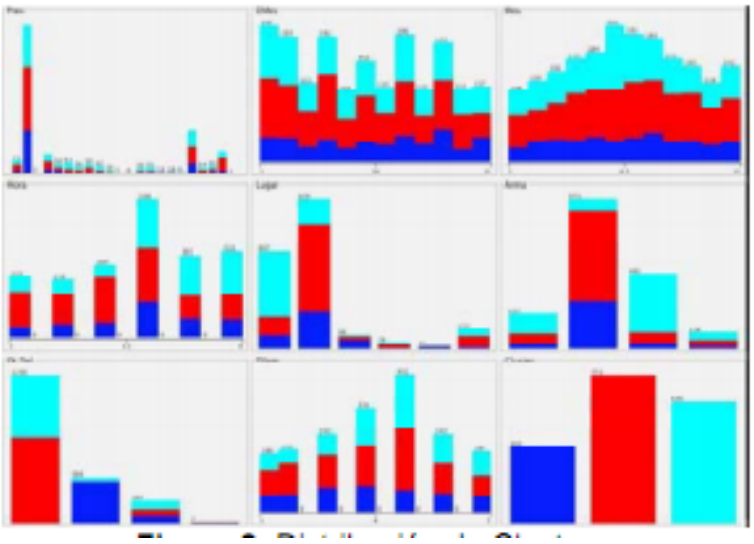


Figura 2.1. Distribución de los clusters.

Hay aspectos importantes a considerar, uno de ellos es el hecho de que se debe contar con volúmenes adecuados de información para su tratamiento con este tipo de técnicas, siendo necesario tener o vincular diferentes fuentes de información. Sin embargo, el resultado analítico que se genera con este tipo de estudio, es de fácil entendimiento y manipulación por un sector del público ajeno al área de los sistemas de información.

2.1.3 Empleo de sistemas de información geográfica en el estudio del miedo al delito (L.Vozmediano , C. San Juan Guillen, 2006)

La investigación de este artículo [5] se centra en el estudio de las causales del miedo al delito, actividad que lleva tiempo de ser analizada (desde la época de los años 70) por la criminología y otras disciplinas, como son la sociología, la geografía, el urbanismo ,la arquitectura y también desde la psicología, particularmente la psicología social y ambiental. Tanto los factores que originan el miedo al delito como las consecuencias del propio miedo, implican a todas las áreas mencionadas.

El conjunto de variables está integrado por los siguientes elementos: el sexo, la edad, las capacidades de afrontamiento y control; en definitiva, factores de vulnerabilidad relativa frente al delito .El sexo y la edad son las variables socio demográficas que se asocian comúnmente con el miedo al delito. Si bien, la evidencia empírica al respecto es, en

ocasiones, contradictoria, muchos estudios afirman que tanto las mujeres como los adultos mayores experimentan un mayor temor ante el delito, lo cual es indicado en las investigaciones [14], [15] y [16], pero otros investigadores encontraron lo opuesto, o mantienen que esta relación es débil o no existe, acorde con lo presentado en [17]. En cuanto a las capacidades de afrontamiento y control, una mayor percepción de control ante la situación podría reducir el miedo o la inseguridad experimentada.

Las variables psicosociales se relacionan con las dinámicas y cohesión vecinales. Usando variables entre las que se pueden mencionar: la satisfacción residencial y el apego al barrio y el sentido de comunidad, que son variables que mostraban un escaso poder predictivo frente a posicionamientos más razonados, tales como la confianza en la acción policial [18].

Realizado lo anterior, se distinguieron niveles de clases de variables relevantes en relación al miedo al delito, y al delito mismo, para analizar la *estructura* del espacio urbano. Las características estructurales de un lugar son un factor que inhibe las relaciones sociales, de esa manera, lo hace más susceptible para la ocurrencia de delitos, y en consecuencia, tienen mayor efecto en provocar un incremento en el miedo al delito; por lo tanto se trataría de un "espacio crimípeto" definido en [19]. Estos espacios, debido a sus características espaciales, son susceptibles para ser nichos delictivos, los cuales también se ven afectados por factores como el tiempo y el análisis de la degradación de espacios, donde destaca la perspectiva de los "actos incívicos" [20]; así como la teoría de "ventanas rotas" [21], la cual menciona, entre otros puntos interesantes, que un espacio con signos de degradación indicaría un control social debilitado, produciendo más deterioro, incremento del temor al delito, evasión de un espacio y finalmente, un aumento del delito. Desde esta perspectiva, se puede tener una correlación entre el desorden y la probabilidad del delito. Estas variables relacionadas con la *estructura*, así como las involucradas con la *degradación,* son mencionadas como parte del estudio de la influencia de los actos criminales.

Los resultados obtenidos en este estudio se pueden consultar en la Figura 2.2.

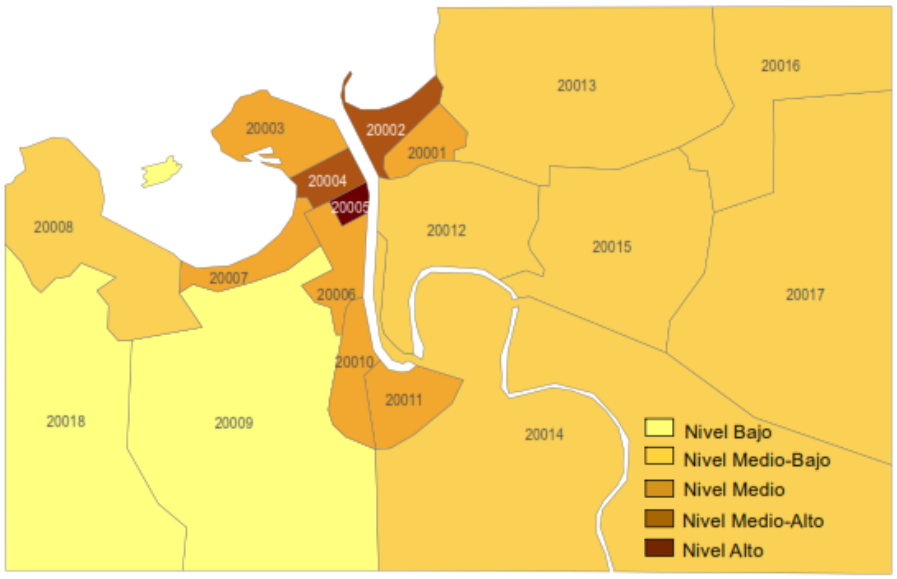


Figura 2.2. Densidad del delito en Donostia San Sebastian, lugar seleccionado para las pruebas.

El primer resultado del proceso descrito ,es un mapa de densidad de delito registrado judicialmente para cada código postal de la ciudad. En el SIG, cada una de las parcelas que representa un código postal, tiene su valor asociado en una variable denotada como "número de delitos"; este número, dividido entre el valor que representa la extensión de la parcela, fue crucial para proporcionar el índice de densidad de delito (representado en el mapa), que en su conjunto resultó fundamental para poder realizar comparaciones entre diferentes parcelas, a pesar de la disparidad de área de cada una de ellas.

En el segundo mapa (consultar Figura 2.3) se hizo una clasificación de las zonas referidas en tres niveles socio-económico, y se efectuó una superposición de los resultados que se generaron previamente, con el objetivo de realizar una comparación de la distribución de estas variables con un delito registrado judicialmente; de lo cual se concluyó que la inseguridad podría estar desempeñando un papel relevante en la génesis y comportamiento del miedo al delito. Este miedo al delito puede estar afectado por los factores estructurales del espacio, las configuraciones urbanísticas, por ser una localidad propensa a la inmigración, entre otros puntos. Asimismo, un factor involucrado es la percepción del estatus socio-cultural del individuo, ya que cuando la diferencia de clases es rigurosamente elevada, el parámetro de calidad de vida también lo es y el sentimiento de inseguridad se incrementa.

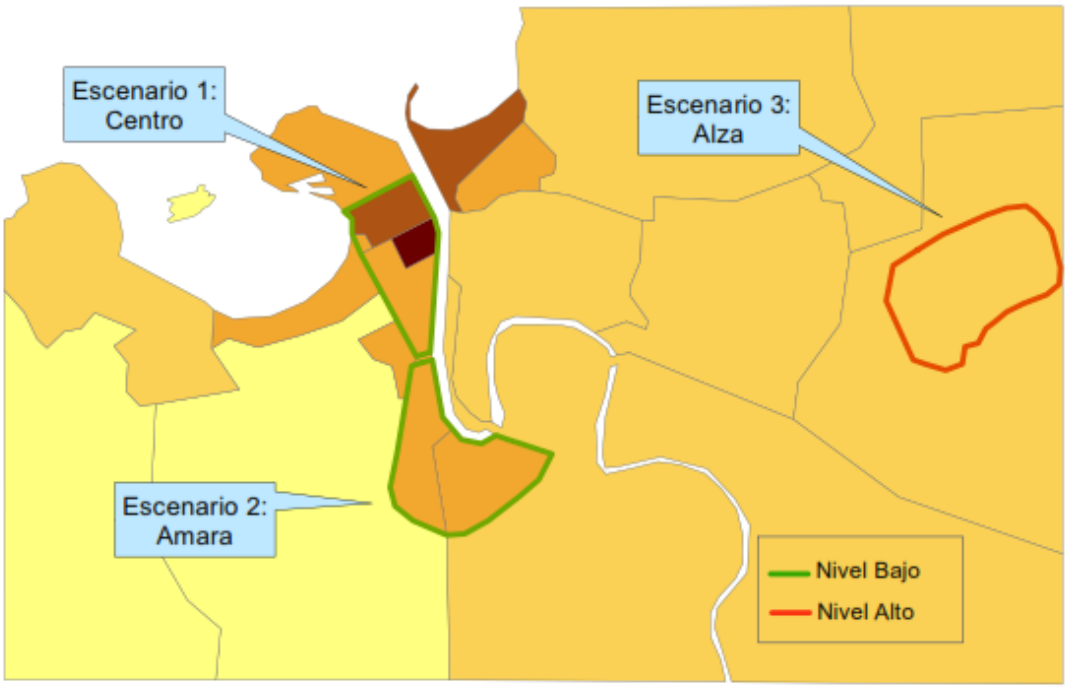


Figura 2.3. Comparativo de la información recaudada en cuanto a la percepción del delito.

Se puede observar que la percepción del delito no siempre corresponde a la incidencia real, ya que es un factor completamente psicológico que corresponde a variables de niveles de satisfacción y en general de la misma percepción del entorno, siendo factor a considerar, la comparativa entre los acontecimientos reales y la medición de la percepción.

2.1.4 Criminología Ambiental (Environmental Criminology)

La Criminología Ambiental [22] tuvo su origen intelectual desde inicios del año 1800, en donde se investigaron las primeras estadísticas sobre la delincuencia y las convicciones francesas asignadas por delitos violentos y la propiedad a nivel de departamento. Otros estudiosos ingleses analizaron los elementos geográficos de la delincuencia durante los años 1850 y 1860. En Estados Unidos, en el año de 1925, Clifford Shaw & McKay Henry, de la Escuela de Sociología de Chicago, estudiaron la ecología social de Chicago, basándose en las residencias de los delincuentes conocidos y organizar estas áreas en zonas concéntricas, generando así un modelo urbano.

La Criminología Ambiental fue creciendo como una especialidad dentro de la criminología, el estudio de la delincuencia, los delincuentes, la conducta delictiva, junto con las medidas correctivas correspondientes, la cual introduce el concepto de la geografía o la dimensión espacial del delito. Se centra en el papel del espacio y del lugar en relación con el tiempo, la ubicación y el carácter de los delitos, diferenciándose de la criminología en general en el aspecto de que el gran grueso de la ciencia se centra en tres dimensiones de otros tipos de delitos -el aspecto legal, el delincuente, y la víctima u objetivo- mientras que la Criminología Ambiental, se enfoca exclusivamente en el contexto en el que actúa la delincuencia, como lo define Fritz en uno de sus artículos. Los criminólogos ambientales se centran en el estudio del caso criminal formulando preguntas acerca de dónde y cuándo se cometen delitos [22]. Esto va encaminado al desarrollo de un mapa de actividad criminal que pueda apoyar en el entendimiento de las causales del fenómeno y el cómo contrarrestar dicha problemática.

Por otra parte, la Criminología Ambiental se base en tres puntos de vista teóricos (teoría de la elección racional, la teoría de la actividad de rutina, y la teoría del patrón de crimen). La perspectiva de la elección racional sugiere una explicación de cómo los delincuentes seleccionan sus objetivos al cometer sus crímenes [23]. El enfoque de la actividad de rutina, se encarga de explicar los hechos delictivos, en virtud de la confluencia de un delincuente motivado y un objetivo deseable o la oportunidad que se cruzan en el mismo lugar y hora. Por último, para la identificación de patrones delictivos, consideran la interacción de los delincuentes con su entorno físico y social, lo cual determina su elección de objetivos, donde es de gran apoyo la sociología de la función, que describe los hábitos y cotidianidades del infractor.

Las opciones para los posibles delincuentes se definen con base en sus movimientos, sus localidades de residencia, los lugares donde regularmente se desenvuelven, así como la forma en que se construyeron sus entornos sociales y físicos [22]. Actualmente, los criminólogos ambientales se concentran en contar con herramientas para la toma de decisiones, para poder definir un perfil geográfico, determinar la organización espacial de la incidencia delictiva, realizar análisis de puntos críticos y definir acciones para prevenir el delito, a través de Diseño Ambiental (CPTED) que es fundamental para el análisis y previsión [24]. En el año de 1986, en la ciudad de Minneapolis se analizó un total de 323,000 llamadas a la policía local y se detectó que un pequeño número de "*hot spots*"

concentraban la mayor parte de la incidencia de delitos de la ciudad [24], además se determinó que el 3% de los lugares concentraban el 50% de las llamadas a la policía. Lo anterior permite resaltar la importancia que tiene el análisis de los "*hotspots*" para estudiar su comportamiento y de esta manera tener una mejor identificación de las zonas con alta incidencia delictiva, para poder determinar la recurrencia de crímenes en estas zonas y poder trazar lo que comúnmente se conoce como “mapa del crimen”.

El mapa del crimen se basa en el análisis de la delincuencia, para la identificación de tendencias de la delincuencia, de los patrones criminales y las estrategias operativas, con el propósito de elaborar informes administrativos que son fundamentales para la toma de decisiones. En [25] se ofrece una descripción detallada del análisis de la delincuencia como una profesión y como una aplicación metodológica de la investigación. Este análisis permite apoyar en las actividades que se describen a continuación.

2.1.4.1 Modelo de análisis

El mapeo del crimen se utiliza para identificar los elementos espaciales y temporales de los patrones de la delincuencia y su serie de tiempo. Ambos (el patrón geográfico y la secuencia cronológica) deben ser evaluados sobre una base regular para desarrollar un perfil. En una tentativa de prevenir el delito futuro a través de la intervención, el perfil se utiliza para detener a los criminales mediante la consulta de bases de datos de delincuentes y proporcionar pistas en una determinada investigación.

2.1.4.2 Análisis de problemas

El mapeo del crimen se ha convertido en un instrumento para llevar a cabo la resolución de problemas conocidos en los círculos policiales como SARA, o la exploración, el análisis, la respuesta y la evaluación de los problemas de una comunidad. Permite a los funcionarios realizar pruebas previas y posteriores de las iniciativas de la policía y la comunidad para poder medir el desplazamiento de la delincuencia.

2.1.4.3 Análisis de operaciones

El mapeo del crimen también se utiliza para asignar recursos, realizar un análisis comparativo de la carga de trabajo, el diseño más eficiente de programas y alinear los límites de atención de cada sector de la policía, llamada redistribución de distritos. De este modo, los agentes están en mejores condiciones de responder a las llamadas de una manera más oportuna y eficaz.

2.1.4.4 Perfiles geográficos

Son desarrollados usando los principios de la criminología ambiental, rutina de actividades, la elección racional de técnicas y patrones delictivos específicos para poder identificar delincuentes en serie. Se centra en puntos críticos como son: el contacto con la víctima, el asalto local, y serie de pruebas del delito.

Acorde a las conclusiones del autor, esto ha fundamentado que durante las dos últimas décadas los gobiernos estatales y locales, en los Estados Unidos, han utilizado los mapas de la delincuencia para desarrollar planes estratégicos. Mediante el uso de una amplia variedad de información geográficamente referenciada, se trata de identificar jurisdicciones en situación de riesgo, coadyuvar en la planificación urbana definir proyectos de renovación, así como compartir una amplia variedad de datos a través de las instituciones sociales(las escuelas, el departamento de policía, áreas de control de libertad condicional, servicios sociales , tribunales), donde no existía un intercambio de datos hasta antes de la llegada de los sistemas de información geográfica. Los departamentos de libertad condicional y de libertad bajo palabra combinando un SIG y los Sistemas de Posicionamiento Global, pueden llevar un control y monitoreo constante de los individuos que se encuentran bajo un arresto domiciliario por medio de dispositivos de vigilancia electrónica.

2.1.5 Sistemas de información geográfica y el control de delitos en el espacio: caso de la comuna de Santiago (C. Contrera, 2002)

La finalidad de esta investigación [26] fue la utilizar un modelo de análisis espacial para la elaboración de mapas delictivos, por medio de la generación de mapas coropléticos de densidad de delitos a partir de la ocurrencia eventos por hectárea, basados en la información primaria de eventos delictivos identificados. Esta investigación tomó como base el trabajo descrito en [27] el cual menciona que para la representación de un mapa de áreas en primer lugar es necesario tener en cuenta la naturaleza de la variable. Si ésta consiste en datos absolutos, las soluciones más idóneas se basan en emplear símbolos puntuales escalados o realizando mapas de densidades de puntos. Si por el contrario, la variable es un índice o tasa, resulta más adecuado el uso de tonos o tramas seriadas, las cuales indican la posición del área en una escala de densidad, esta es la razón fundamental por la que para este caso de estudio se emplearon este tipos de mapas.

El control y procesamiento de los datos constituyen una ventaja para los SIG, ya que estos sistemas pueden basarse en diferentes fuentes de datos para efectuar su análisis. En el caso del control del delito para la comuna de Santiago, fue necesario crear plantillas para el acceso a los datos que estaban en un formato específico de DBase. De la información obtenida se aplicaron los procesos de especialización correspondientes. Otras características que se pueden citar son las siguientes:

- El **formato magnético** de la información permite un almacenamiento físico compacto, reduciendo costes en su mantenimiento y acelerando el acceso de los datos.
- El **mantenimiento** de la información espacial estructurada.
- La **facilidad de actualización** de la información.
- La **accesibilidad a la información**, facilita su revisión y actualización, contribuyendo a que la información corresponda más fielmente a la realidad geográfica que representa.
- El **análisis de la información**, donde el formato digital facilita el tratamiento de la información geo-referenciada mediante la aplicación de operadores matemáticos o lógicos que, combinados, facilitan la implementación de modelos de procesamiento cartográfico.

Los Sistemas de Información modernos son una opción viable y más que comprobada que la utilización de métodos de análisis manuales, aunque lo importante [28], en el caso específico de los sistemas geográficos se destacan las siguientes características:

- ***La integración y combinación de la información*****,** donde obteniendo los datos requeridos, independiente de la procedencia de los archivos y en un formato acorde a los requerimientos necesarios, se hace posible la producción de mapas e informes.

- ***La precisión y resolución*****,** donde correlacionando las distintas fuentes de información y mejorando la capacidad de edición y corrección de la información, se identifican errores o equivocaciones de una manera sistémica.

- ***La reducción del trabajo manual*****,** se ahorra tiempo debido a las posibilidades de compilar la información y producir mapas en forma automatizada gracias a aplicaciones directas. Se elimina casi por completo las actividades manuales, y se destina el personal a otras actividades de planificación.

Con base en el estudio realizado en los datos de prueba, se realizó el seguimiento de la información obtenía de los años 93 al 99 con un formato de información como se ilustra en la Figura 2.4:

I - 1993 MUESTRA 4206 DELITOS

MESES ENE-FEB-JUN-JUL-AGOST-SEPT-OCT-NIV-DIC

TIPO DELITO	%		LUGAR ROBO	%
HOMICIDIOS	0.1%		VÍA PÚBLICA	57.26%
HURTO	18.7%		LOCAL COMERCIAL	16.12%
ROBO	81.2%		CASA/HABITACIÓN	10.57%
VIOLACIÓN	0.1%		FABRICA/TALLER/OFICINA	6.31%
			OTROS	9.74%
TOTAL	100.0%		TOTAL	100.00%

FUENTE : CARABINEROS DE CHILE

Figura 2.4. Formato empleado para el reporte de delitos.

Al analizar los datos, se observó que la tendencia de la región se situaba en un tipo de delito en particular. En este caso, los robos y sus diferentes variantes ocupaban la mayor concentración de los delitos. De esta manera, al identificar esta tendencia, se pudo llevar más a fondo un análisis al estudiar las características de dichos eventos, es decir, relacionar su naturaleza a una situación en particular, para lo cual se refinó el análisis y se concluyó que dichos eventos tenían una mayor ocurrencia en la vía pública.

Con base en estos resultados se tomó la determinación de aumentar el patrullaje. La tendencia en un periodo de 6 meses fue (como se esperaba)que se comenzó a notar una disminución en los reportes de actividad delictiva en la zona. Lo anterior permite ejemplificar la importancia de contar con un mapa del crimen, ya que ayuda a la visualización de la problemática en general y de esta manera apoyar en la toma de decisiones.

Esta etapa es una aproximación a un desarrollo más complejo, en el cual los autores mencionan que es importante señalar unos aspectos de relevancia a considerar en el proceso del análisis urbano, los cuales involucran aspectos sociales, perceptuales/emocionales y físicos. Al definirse estos aspectos se pudo acotar la problemática y el alcance, planteándose varios aspectos a considerar sobre las posibles causas de las tendencias del comportamiento social de la incidencia delictiva.

En el primer ámbito de estudio, en el caso social, se involucra el concepto de la definición de barrio, citando a Henry Lefevre*" corresponde a un espacio social, cuya existencia es más bien simbólica que funcional o estructural. Refleja una forma de organización del espacio y de la ciudad"*. Se considera que el barrio tiene una semi existencia, donde se constituyen relaciones interpersonales más o menos durables y profundas. Es el mayor de los grupos sociales pequeños y el más pequeño de los grandes grupos. En este aspecto, los autores mencionan lo que es el sentido de pertenencia y la visualización [6] demostraron en sus estudios de la percepción del delito, del cual parte este sentido de pertenencia y percepción en general de las problemáticas, así como con las costumbres del barrio como una entidad social.

El segundo aspecto o ámbito perceptual, va relacionado a encontrar una medida para calificar la habitabilidad de la zona con base en parámetros que sean considerados objetivo [28] que involucra la capacidad de llegar a otras personas, actividades, recursos, servicios, información o lugares, incluyendo la cantidad y diversidad de los elementos. En nuestro análisis, la proximidad es un elemento fundamental ya que de él derivan una serie de fenómenos sociales, los cuales involucran una complejidad diferente pero que puede ser considerada para investigaciones. Este aspecto es tan amplio que puede fácilmente abarcar un estudio particular.

El tercer aspecto que se considera como parte fundamental de estudio, es el ámbito físico o de control. Referido más al deterioro de la calidad debida urbana, como un elemento que contribuye al debilitamiento de los mecanismos de control social ejercido por los vecinos sobre su propio territorio geográfico, entendiendo como la calidad de vida una serie de variables que se implican en un entorno [29], en cuanto al grado de uso y el acceso a los espacios, actividades, su creación, reparación, modificación y dirección, los cuales son controlados por aquellos que los usan, trabajan o residen en los mismos. Lo describen como un nivel de control individual del habitante de un lugar al interactuar con su entorno.

Una vez que se delimitó el tema, se procedió a buscar la implementación de técnicas de despliegue y análisis de la información obtenida. En este caso se usaron *coropletas* (ver Figuras 2.5, 2.6 y 2.7), donde fue necesario (para la representación de mapas de áreas) el considerar la naturaleza de la variable y generar los datos absolutos para su

representación escalada con la finalidad de aplicar mapas de densidades como lo visto en [28].

En este caso de estudio se siguió el siguiente procedimiento para la preparación de los datos:

- La información se adjudica al área que le corresponda.
- Los datos se agrupan en clases, correspondiendo a cada una de ellas una trama específica.
- Las áreas son tramadas según las áreas en que se incluyan.
- Opcionalmente, se borran las fronteras de la clase en la que se incluyan.
- Se digitan las intervenciones en las unidades respectivas
- Se someten los datos a un riguroso test de estandarización para descartar errores de digitación.
- Se geo-codifica la base de datos, elaborando coberturas digitales de puntos con atributos asociados.
- Se establece relación topológica de los puntos con el espacio (densidad de puntos por hectárea) a través de un procedimiento de conversión puntual a vectorial.
- Se emite el plano conservando simbologías para la posterior comparación de áreas.

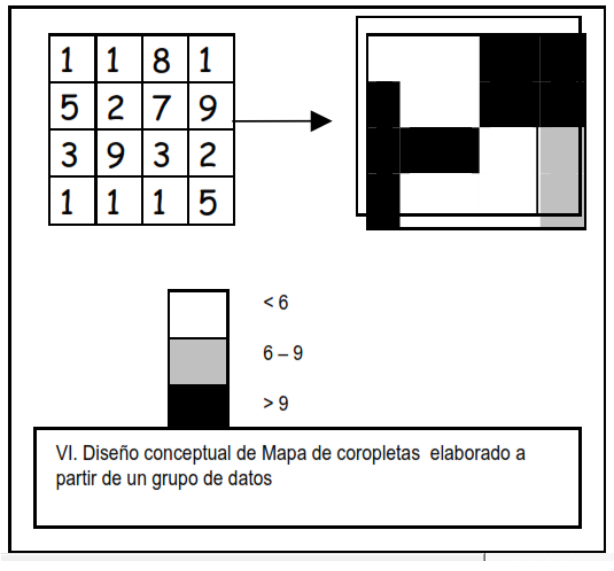


Figura 2.5. Ejemplificación del artículo sobre la técnica de representación de imágenes por medio de coropletas

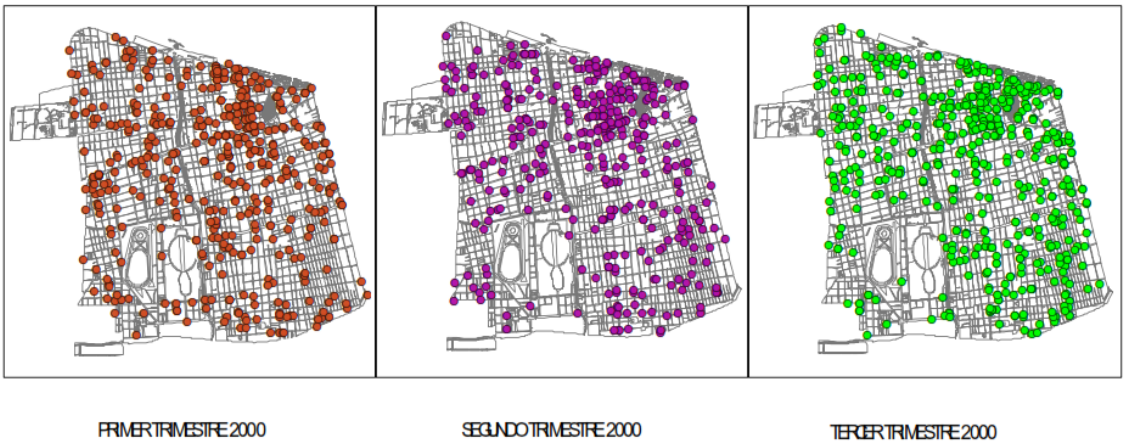


Figura 2.6. Representación la densidad de delitos en la comuna de Santiago.

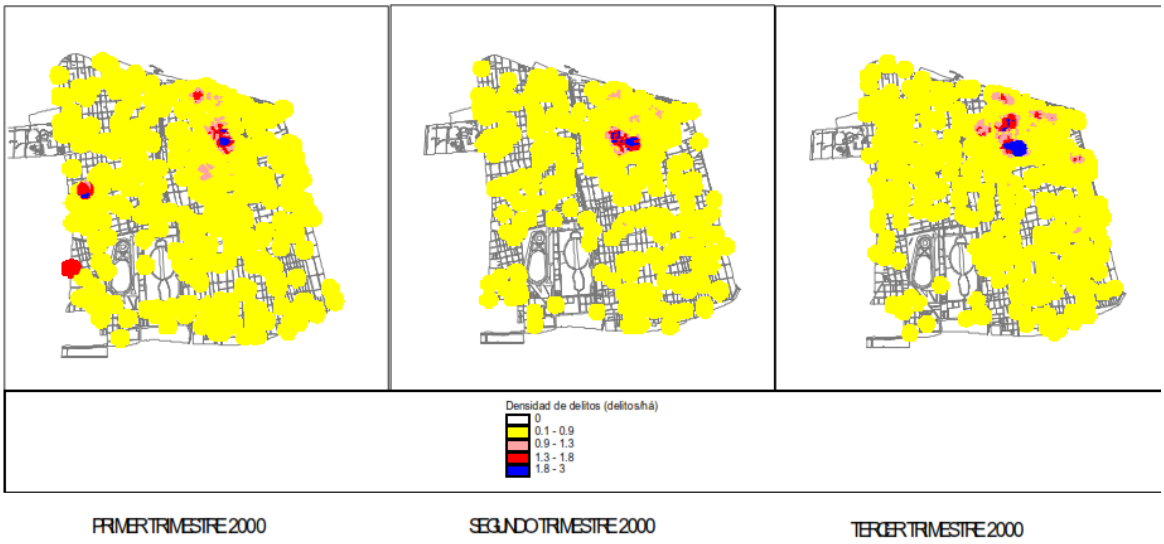


Figura 2.7. Análisis de densidad del delito aplicando coropletas.

Como conclusión, este análisis representa de primera mano un desarrollo de visualización e identificación de zonas vulnerables a la incidencia del delito. Con la técnica desarrollada, se forma un "mapa del delito" con base en los ámbitos descritos anteriormente lo cual permite definir una tendencia para el desarrollo de nuevas aplicaciones, sobre todo en el aspecto de análisis de patrones. En el trabajo se concluye que un posible patrón o hipótesis del comportamiento delictivo corresponde a regiones con pocas relaciones interpersonales y poca urbanización, la cual está relacionada en su mayoría con el sentido de pertenencia de barrio y el grado de control del entorno. Aunque se consideraron los

aspectos físicos como un factor para los delitos, en el estudio se determinó que fueron poco representativos.

2.2 Desarrollos de análisis de ontologías con contexto geográfico

En esta sección se presentan estudios relacionados a involucrar técnicas de procesamiento de información basadas en el contexto, empleando ontologías en un ámbito geográfico como técnicas para la recuperación de información, normalización de datos, descripción de clases o soporte a la toma de decisiones con base en las características de las entidades descritas.

2.2.1 Sistema de soporte de decisiones basado en una ontología para la investigación de procesos criminales (D. Dzemydiene & E. Kazemikaitiene, 2005)

Los delitos complejos y dinámicos requieren nuevas formas de extracción de información y representación del conocimiento [30]. El análisis de la delincuencia basándose en una ontología, busca la obtención y representación de un modelo estructural de la información sobre los delitos con lo cual se determinan las reglas principales, la forma de adquirir la información en este caso en particular referente a la investigación forense y el crimen para obtener sus causas primarias.

El desarrollo propuesto en el artículo corresponde a un sistema de soporte de decisiones (DSS) para poder actuar "amablemente" en términos de fiabilidad al reproducir el comportamiento cognitivo de los investigadores de delitos, el cual exige una profunda reflexión, conceptos, principios, y los escenarios relacionados directamente con el crimen [30].

Se pueden clasificar los problemas que surgen en la investigación de un crimen en dos grandes grupos: los problemas de una naturaleza puramente criminal y aquellos asociados con el derecho penal, procesal penal, la organización y la criminología. El primer grupo de problemas se puede resolver mediante el empleo de la información del análisis del crimen, mientras que para los problemas del otro grupo se necesita más información sobre el derecho penal, el procesal penal y similares, lo cual sería significativo para la

investigación del crimen. En donde es muy importante considerar la forma de representar el conocimiento sobre investigación criminal y los métodos para construir los sistemas de toma de decisiones. Para ello, se describió el sentido de la aplicación, señalando los problemas de toma de decisiones con base en la creación de la ontología de la investigación del delito.

Se mencionan algunos esfuerzos similares, que actualmente se desarrollan en áreas tan diversas como la psicología, la medicina, el análisis de la delincuencia y la informática, que se centran en el análisis y la gestión de imágenes con cierto apoyo en las técnicas de ingeniería del lenguaje. La fuerza de la inteligencia vinculada (sistema FLINTS9) combina términos forenses y físicos probatorios (hits) con el fin de mostrar los vínculos entre los delincuentes y las pruebas recopiladas en un delito, para producir un perfil de los delincuentes y de los delitos cometidos. Otros sistemas como IMPRESS, Pedernales, Locard y las imágenes de flujo de trabajo, se emplean para la construcción del perfil de un habitual criminal, llevando así un seguimiento de las pruebas que tienen que ver con el movimiento de los objetos expuestos relacionados con el delito (de la escena del crimen a través de la corte). Este seguimiento se realiza de nuevo a través de una descripción de la clase).

El desarrollo de una ontología de aplicación ayudó a crear el marco conceptual y de esta manera se pudo definir la recolección, acumulación, almacenamiento, tratamiento y transmisión de toda la información importante en una investigación, para establecer las condiciones óptimas para tomar decisiones en la investigación de una serie de crímenes, desarrollando un conjunto de los posibles *modus operandi* de un crimen en particular. Este conjunto de posibles alternativas de *modus operandi* crimen puede definirse como:

$$AC = \{a \in A \mid \exists\, s \in S, (\forall\, c \in C, (S \Rightarrow c)) \wedge (S \Rightarrow a)\}, \qquad (2.1)$$

Donde *AC* es el conjunto de todas las alternativas posibles o *modus operandi* de la delincuencia (por ejemplo, el suicidio, accidente, o el asesinato), *S* es el conjunto de todos los escenarios del crimen consistentes, y *C* es el conjunto de todas las piezas recogidas de pruebas del delito.

La metodología de la aplicación del sistema de soporte de decisión se basa en la ontología creada (ver Figura 2.8) constituyendo un vocabulario específico utilizado para describir una realidad determinada, además de un conjunto de supuestos explícitos sobre el significado del vocabulario.

El lenguaje *L* con el vocabulario *V* son vistos como una función de asignación de elementos (pruebas) a los delitos *D* y elementos de *R* símbolos determinantes de *V*. El papel de la ontología se puede considerar como un conjunto de axiomas lógicos diseñados para explicar el significado del vocabulario [32], esto se describe mediante la Ecuación 2.2.

$$I:V \rightarrow D \cup R \qquad (2.2)$$

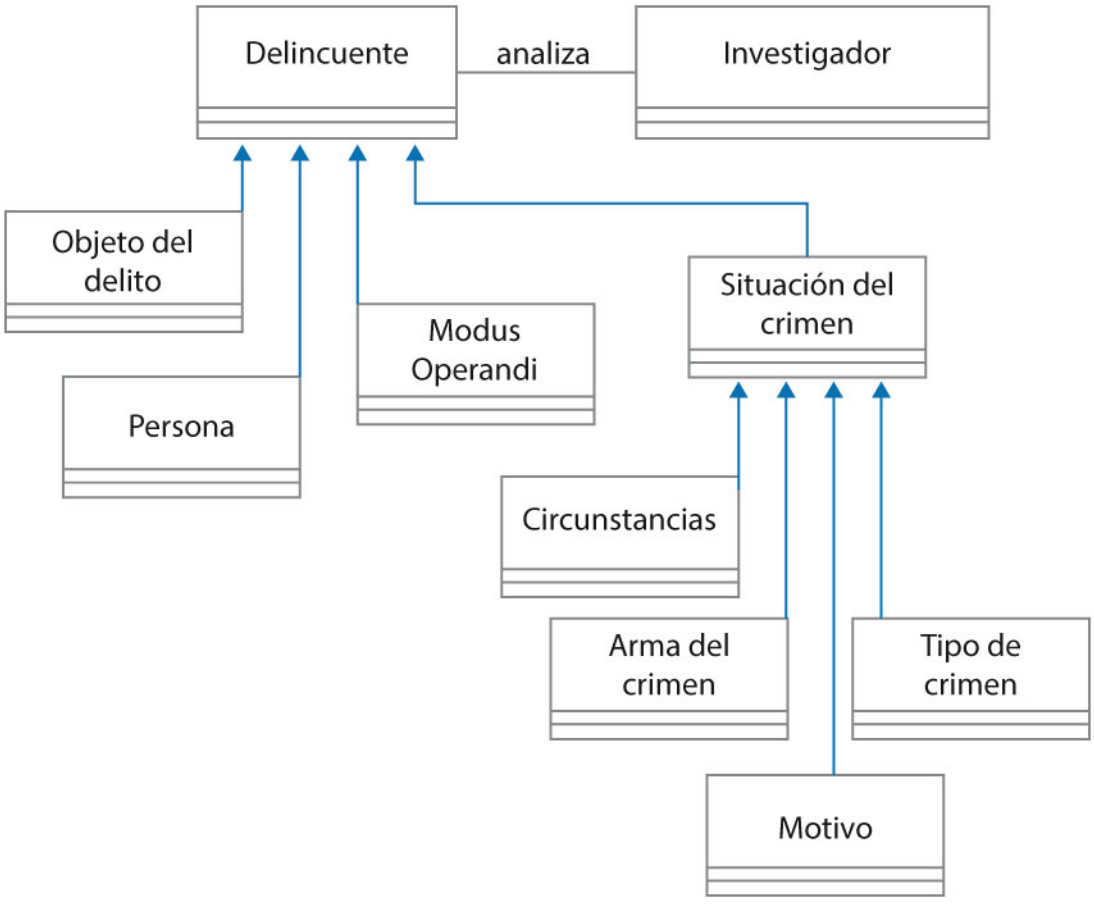


Figura 2.8. Ejemplo del diagrama de clase de objeto del carácter del delito en UML propuesto por Dzemydiene & EgleKazemikaitiene

Debido al propósito y la naturaleza de las ontologías los autores clasifican las ontologías de acuerdo con su nivel de generalidad:

Ontología de nivel superior: Describe conceptos muy generales, tales como espacio, tiempo, materia, objetos, etc., que son independientes de un problema en particular de dominio.

Ontología de dominio y la ontología de trabajo: Describen el vocabulario relacionado con el área específica de trabajo, generando así una especialización más detallada de la ontología de nivel superior.

Ontología de aplicación: Los conceptos se describen en un dominio particular. Se trata con conceptos que a menudo corresponden a las funciones desempeñadas por las entidades de dominio en la ejecución de una de terminada actividad.

De esta manera se definen tres cosas particularmente: conceptos vocabulario y actividades, todo esto para situar en contexto estas relaciones específicas, considerando también una secuencia de la manera en que se opera para la obtención y la manipulación de las pruebas (ver Figura 2.9), ya que estos términos entran en las relaciones que tuvieron que ser definidas en la ontología.

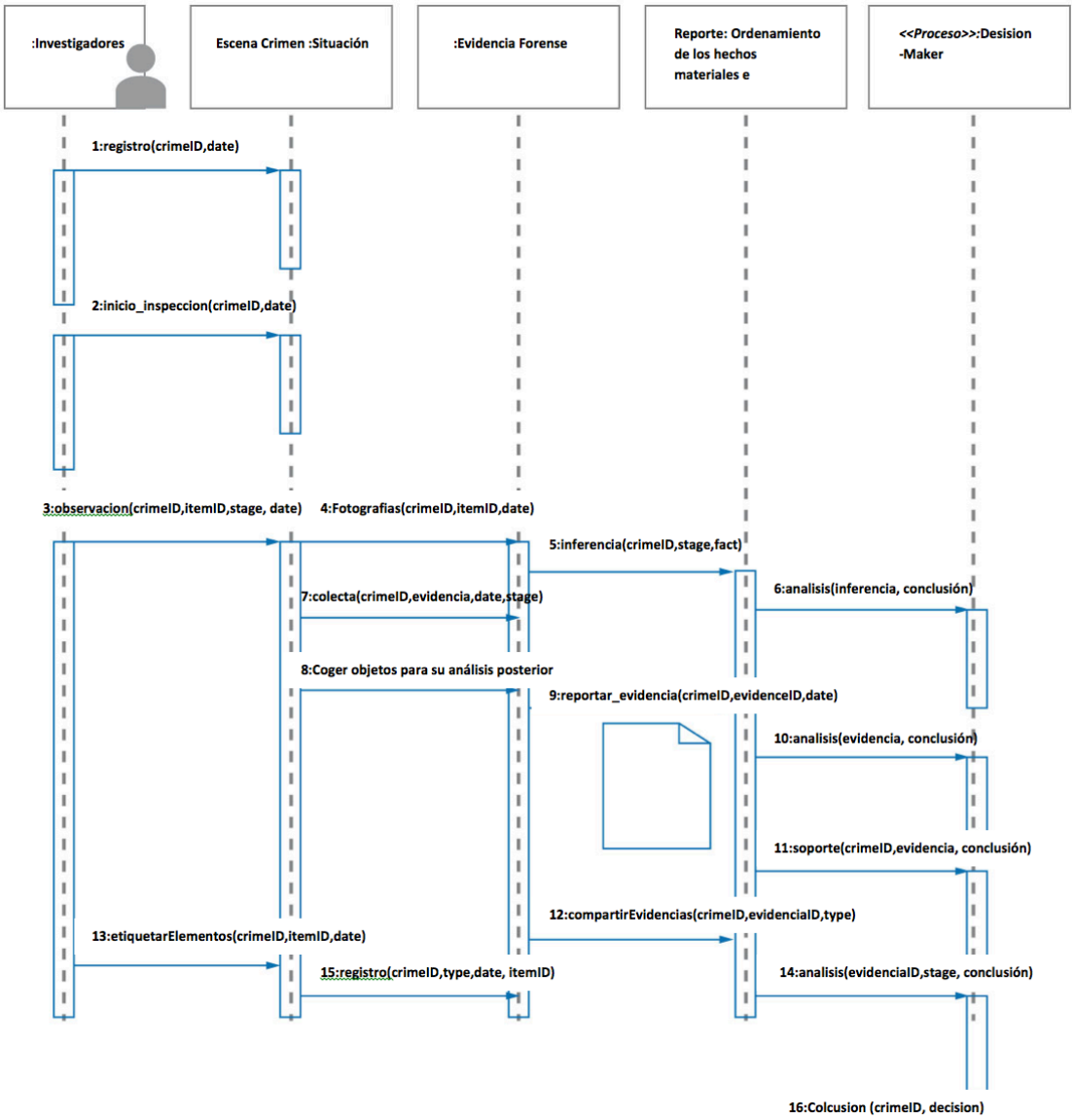


Figura 2.9. Secuencia diagrama del proceso de pruebas de investigación forense propuesto por Dzemydiene & EgleKazemikaitiene.

En la figura anterior se ejemplifica el proceso de obtención y manipulación de pruebas. Además de la peculiaridad de la línea de investigación, se sigue un protocolo establecido para la obtención de las prueba. En un caso en particular se ve el ejemplo de las definiciones del proceso ,en este caso, el análisis fue realizado a crímenes en contra de la vida(ver Figura 2.10) la cual sale de la definición de un procedimiento de investigación, donde se siguen los siguiente pasos:

- Se inspeccionan los alrededores del lugar buscando información importante para la obtención de pruebas circunstanciales. Este es conocido como el proceso de observación.
- Se examina el cuerpo de la víctima para poder determinar la causa de la muerte y otras particularidades que corresponden al conocimiento de los médicos forenses.
- El investigador examina el cuerpo y la ropa con el fin de identificar a la persona, el sexo, las huellas dactilares, sangre, saliva, esperma u otros rastros en el cuerpo.

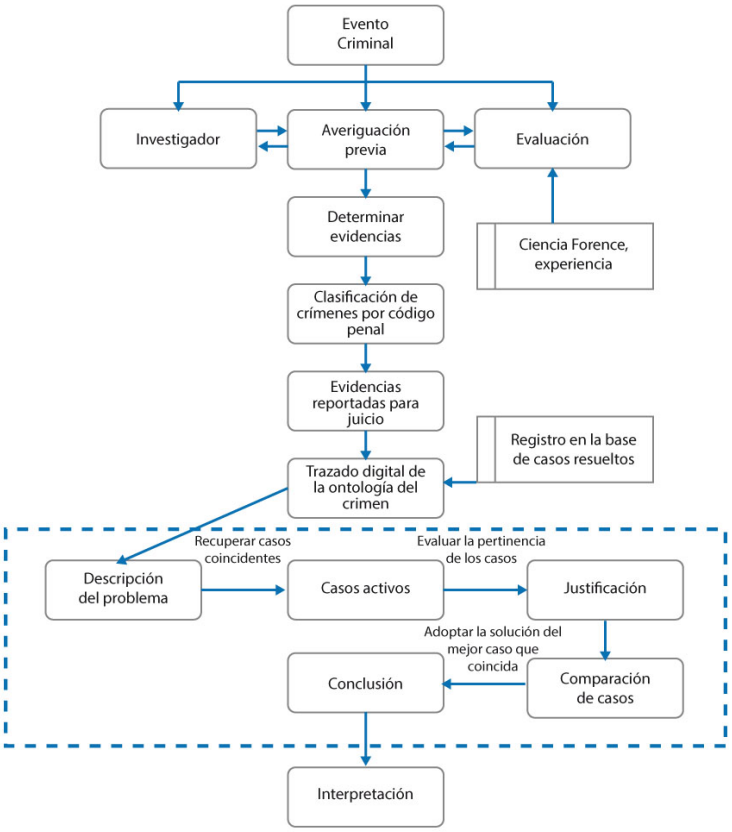


Figura 2.10. Visualización de las relaciones ontológicas del sistema según Dzemydiene y Kazemikaitiene.

De este proceso preliminar se obtiene la primera hipótesis sobre las causas y los hechos ocurridos.

La principal ventaja del enfoque que aquí se presenta es la posibilidad de integrar la ontología en el algoritmo de generación del espacio de escenario. El enfoque para el desarrollo del sistema de apoyo en la toma de decisiones utilizó técnicas de razonamiento

basadas en modelos y casos de uso de las investigaciones criminales. El sistema de apoyo permite la combinación de eventos delictivos y los estados que producen un determinado conjunto de pruebas de la base de conocimiento, la ontología fue útil en la generación de fragmentos de escenarios de situaciones inconsistentes. Una adecuada investigación de los delitos depende de la calidad del sistema de información utilizado como asesoramiento.

Capítulo 3. Marco teórico

3.1 Resumen

En este capítulo se plasman los fundamentos teóricos de las técnicas y algoritmos que fueron tomados como base para formular la metodología del presente trabajo de tesis. Se describen las características que presentan las herramientas y los fundamentos de las técnicas o estudios para presentar en el siguiente capítulo la metodología desarrollada. Se inicia con un estudio de conceptos que sirven como base para la generación de un modelo de datos.

3.2 Ontologías

Gruber en [30] define una ontología como "una especificación explícita de una conceptualización", ya que debe quedar claro que un "explícito" es un objeto concreto, el símbolo de nivel de objeto. Según Neches en [31], una ontología define los términos y relaciones básicas para definir extensiones de vocabulario, involucrando para su definición, la identificación de términos y relaciones básicas entre términos, las reglas para combinarlos y proporcionar definiciones de términos y relaciones. Por último, una ontología incluye no solo los términos que son definidos explícitamente en la ontología, sino además los términos que pueden ser inferidos utilizando reglas.

Considerando lo descrito en [32], [33] y [34] se puede obtener una definición más amplia de una ontología, la cual se define como una especificación formal de una conceptualización compartida, donde la *conceptualización* se refiere a un modelo abstracto de algún fenómeno en el mundo, teniendo identificados los conceptos relevantes de los fenómenos, donde las restricciones y reglas están explícitamente definidas reflejando el conocimiento consensuado definido sobre una rama del conocimiento.

En otras palabras, según lo explicado en [35], independientemente del ámbito en que se desarrollen, es necesario aclarar que la base para una ontología es la conceptualización,

junto con un vocabulario para referirse a las entidades de un dominio particular; es decir, las ontologías para representar el conocimiento precisan los siguientes [31]:

- **Conceptos:** Son las ideas básicas que se intentan formalizar. Los conceptos pueden ser clases de objetos, métodos, planes, estrategias, procesos de razonamiento, etc.

- **Relaciones:** Representan la interacción y enlace entre los conceptos, formando la taxonomía del dominio. Las relaciones básicas son: *sub-clase-de*, *parte-de*, *conectada-a*. Un ejemplo de definición de relación se puede observar en la Figura 3.1.

```
(define-relation flightNumber (?flight ?flightNumber)
   "It identifies the flight company and the number"
:def
  (and (Flight ?flight)
       (String ?flightNumber)))
```

Figura 3.1 Ejemplo de definición de relación en ONTOLINGUA.

- **Funciones:** Son un tipo concreto de relación donde se identifica un elemento mediante el cálculo de una función que considera varios elementos de la ontología, ver Figura 3.2

- **Instancias:** Se utilizan para representar objetos determinados de un concepto, son casos concretos de las clases.

```
(define-function Pays (?room ?discount) :-> ?finalPrice
   "Price of the room after applying the discount"
:def (and (Room ?room) (Number ?discount)
          (Number ?finalPrice)
          (Price ?room ?price))
:lambda-body
          (- ?price (/ (* ?price ?discount) 100)))
```

Figura 3.2 Ejemplo de declaración de funciones en ONTOLINGUA.

- **Axiomas:** Son teoremas que se declaran sobre relaciones que deben cumplir los elementos de la ontología. Los axiomas deben proveerse para definir la semántica o el significado de los términos.

En la Figura 3.3 se ejemplifica la definición de una clase.

```
(define-class Research-Topic (?X)
   ""
:def
 (and (name ?X String) (description ?X String) (approaches ?X String)
       (research-groups ?X Research-Group) (researchers ?X Researcher)
       (related-topics ?X Research-Topic) (sub-topics ?X Research-Topic)
       (events ?X Event) (journals ?X Journal) (projects ?X Project)
       (application-areas ?X String) (products ?X Product)
       (bibliographies ?X URL) (mailing-lists ?X EMail) (webpages ?X URL)
       (international-funding-agencies ?X Funding-Agency)
       (national-funding-agencies ?X Funding-Agency)
       (author-of-ontology ?X Researcher)
       (date-of-last-modification ?X Date))
:axiom-def
 (and (Template-Facet-Value Cardinality name Research-Topic 1)
       (Template-Facet-Value Minimum-Cardinality description Research-Topic 1)
       (Template-Facet-Value Minimum-Cardinality researchers Research-Topic 1)
       (Template-Facet-Value Minimum-Cardinality related-topics Research-Topic 1)
       (Template-Facet-Value Minimum-Cardinality bibliographies Research-Topic 1)
       (Template-Facet-Value Minimum-Cardinality author-of-ontology Research-Topic 1)
       (Template-Facet-Value Cardinality date-of-last-modification Research-Topic 1)))
```

Figura 3.3 Ejemplo de generación de una clase para una ontología en ONTOLINGUA

Un indicador de la complejidad de una ontología es el conjunto de relaciones conceptuales. Fox y Gruninger sostienen que una ontología se define como un vocabulario más una especificación del significado de dicho vocabulario [35]. Esta visión permite distinguir ontologías basadas en el grado de formalidad en la especificación del significado. Las ontologías informales usan un lenguaje natural, las ontologías semi formales proporcionan axiomatizaciones débiles tales como las taxonomías, mientras que las ontologías formales definen la semántica del vocabulario por una axiomatización completa y efectiva. Debido a la formalidad y la manera de representar un campo de dominio, las ontologías funcionan de una manera excelente para la determinación de modelos de datos.

Se puede observar que cualquier sistema de información tiene su propia ontología implícita, ya que se le atribuye un significado a los símbolos usados según una visión particular del mundo. Sin embargo, de manera explícita, una ontología puede tener diferentes roles en dicho sistema.

Teniendo en cuenta los beneficios que ofrecen las ontologías en los sistemas de información, se pueden tener dos perspectivas respecto al rol que desempeñan:

- Como un soporte para el análisis conceptual de métodos y técnicas para los sistemas

- Como un soporte para el diseño, desarrollo y uso de los sistemas de información. En esta perspectiva se analizan dos dimensiones: la visión de los desarrolladores, concerniente a la manera en que una ontología ayuda o se usa para desarrollar el sistema, así como la visión del usuario, la cual es relativa a la manera en que una ontología facilita la tarea del mismo al interactuar con él.

Adicionalmente, las ontologías también pueden ser clasificadas según la formalidad en la que son representadas en :

- *Altamente informales.* Si son expresadas en lenguaje natural.

- *Semi-informales.* Si son expresadas en una forma restringida y estructurada del lenguaje natural

- *Semi-formales.* Si son expresadas en un lenguaje artificial y definido formalmente (Ontolingua, OWL, como ejemplos).

- *Rigorosamente formales.* Si proporcionan términos definidos meticulosamente con semántica formal, teoremas y completitud.

Es factible usar una ontología que modele el dominio de aplicación y proporcione un vocabulario para la especificación de requisitos (ER) del sistema a diseñar. El rol que la ontología juega en la especificación varía con el grado de formalidad y automatización de la metodología de diseño.

En una aproximación informal, las ontologías facilitan el proceso de identificación de los requisitos de un sistema y el entendimiento de las relaciones entre los componentes del

sistema, ver Figura 3.4. Esto es importante cuando se desarrollan, aportando como ya se describió anteriormente, una especificación declarativa de los sistemas de información.

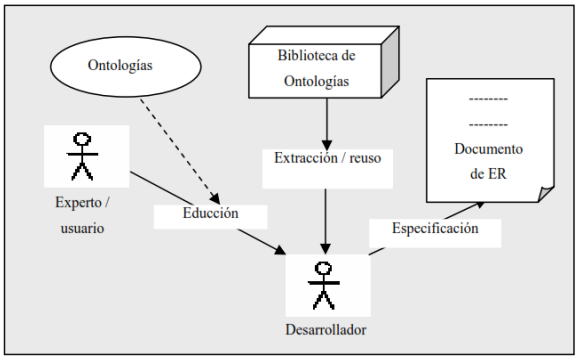


Figura 3.4 Especificación de los requisitos de una ontología descrito por Barchini, Álvarez, Herrera y Trejo.

3.3 Bases de datos relacionales

Son un tipo de bases de datos en las cuales, los datos se muestran en forma de tablas y relaciones. Este es el modelo más usado en el desarrollo de sistemas. A finales de los años 60, Edgar Frank Codd definió las bases del modelo relacional. Trabajaba para la empresa IBM, que tardó un poco en implementar este tipo de bases de datos. Pocos años después el modelo se empezó a implementar cada vez más, hasta ser el modelo de bases de datos más popular. Los objetivos del modelo propuesto por Codd [36] son los siguientes:

- *Independencia física.* La forma de almacenar los datos no debe influir en su manipulación lógica.
- *Independencia lógica.* Las aplicaciones que utilizan la base de datos no deben ser modificadas como consecuencia de que se modifiquen elementos de la base de datos.
- *Flexibilidad.* La base de datos ofrece fácilmente distintas vistas en función de los usuarios y aplicaciones.
- *Uniformidad.* Las estructuras lógicas siempre tienen una única forma conceptual (las tablas).
- *Sencillez.*

Las bases de datos relacionales se basan en el uso de tablas (también son denotadas con el nombre de relaciones). Las tablas se representan gráficamente como una estructura rectangular formada por filas y columnas. Cada columna almacena información sobre una propiedad determinada de la tabla (a esta propiedad se conoce como atributo), Cada fila o tupla posee una ocurrencia o ejemplar de la instancia o relación representada por la tabla (ver Tabla 3.1).

Tabla 3.1 Ejemplo de una tabla en el modelo Entidad Relación

NOMBRE

atributo 1	**atributo 2**	**atributo 3**	**....**	**atributo n**	
valor 1,1	valor 1,2	valor 1,3	....	valor 1,n	← **tupla 1**
valor 2,1	valor 2,2	valor 2,3	....	valor 2,n	← **tupla 2**
.....	**.....**	.....	**....**	.....	....
valor m,1	valor m,2	valor m,3	....	valor m,n	← **tupla m**

3.3.1 Diseño conceptual de bases de datos modelo relacional

Restricciones

Se trata de ciertas condiciones de obligado cumplimiento por los datos de la base de datos. Existen restricciones de varios tipos, las cuales se describen brevemente a continuación:

Inherentes. Son aquellas que no son determinadas por los usuarios, sino que son definidas por el hecho de que la base de datos es relacional. Por ejemplo:

- No puede haber dos tuplas iguales.
- El orden de las tuplas no importa.
- El orden de los atributos no importa.
- Cada atributo sólo puede tomar un valor en el dominio en el que está inscrito.
- **Semánticas.** El modelo relacional permite a los usuarios incorporar restricciones personales a los datos. Las principales son:
    - *Clave primaria.* Hace que los atributos marcados como clave primaria no puedan repetir valores.

- *Unicidad.* Impide que los valores de los atributos marcados de esa forma, puedan repetirse.
- *Obligatoriedad.* Prohíbe que el atributo marcado de esta forma no tenga ningún valor.
- *Integridad referencial.* Prohíbe colocar valores en una clave externa que no estén reflejados en la tabla donde ese atributo es clave primaria.
- *Regla de validación.* Condición que debe de cumplir un dato concreto para que sea actualizado.

3.3.2 Transformación de modelo Entidad-Relación a Modelo Relacional

Transformaciones de entidades fuertes. En principio las entidades fuertes del modelo Entidad Relación son transformados al modelo relacional siguiendo estas instrucciones:

- Entidades. Las entidades pasan a ser tablas.
- Atributos. Los atributos pasan a ser columnas.
- Identificadores principales. Pasan a ser claves primarias
- Identificadores candidatos. Pasan a ser claves candidatas.

Esto hace que la transformación se realice de la forma que se describe en la Figura 3.5.

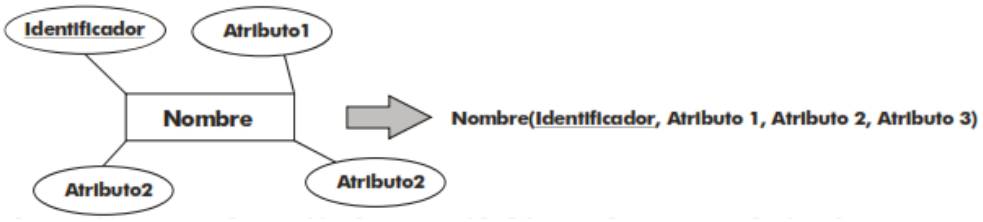


Figura 3.5 Transformación de entidades fuertes.

Transformación de relaciones. La idea inicial es transformar a cada relación en una tabla en el modelo relacional, pero se debe hacer una distinción según el tipo de relación:

- *Relaciones varios a varios* .En las relaciones varios a varios, la relación se transforma en una tabla cuyos atributos son: los atributos de la relación y las

claves de las entidades relacionadas (que pasarán a ser claves externas). La clave de la tabla la forman todas las claves externas (ver Figura 3.6):

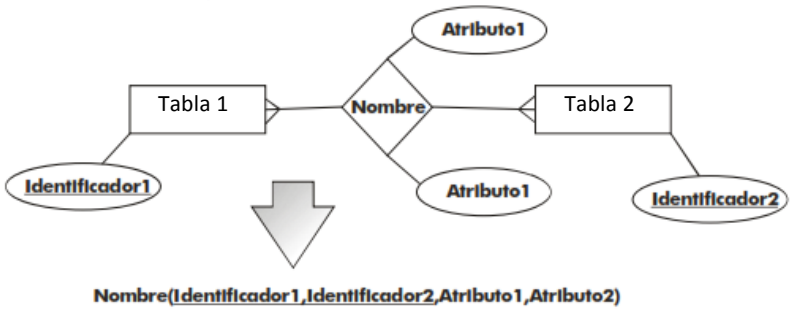


Figura 3.6 Transformación de relación de varios a varios.

- *Relaciones de orden n*. Las relaciones ternarias, cuaternarias y n-arias que unen más de dos relaciones se transforman en una tabla que contiene los atributos de la relación más los identificadores de las entidades relacionadas. La clave la forman todas las claves externas, como se ilustra en la Figura 3.7:

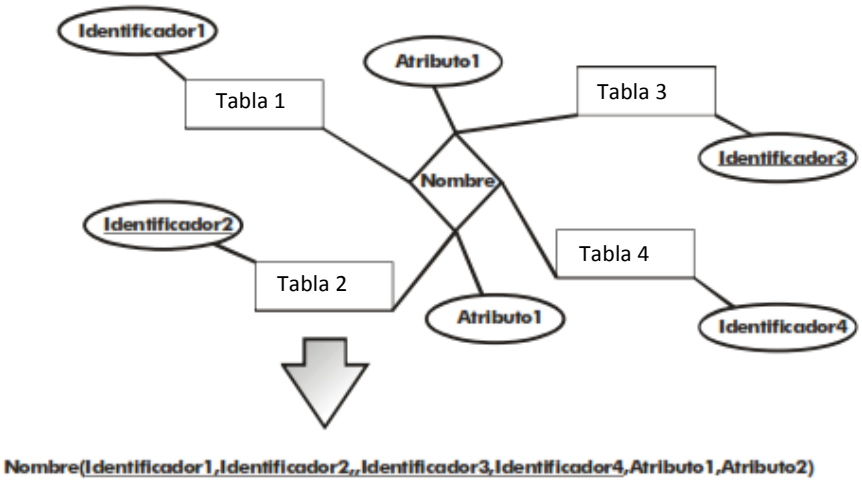


Figura 3.7 Transformación de entidades de orden n.

- Las relaciones binarias de tipo uno a muchos no requieren ser transformadas en una tabla en el modelo relacional. En su lugar, la tabla del lado *varios* (tabla relacionada) incluye como clave externa el identificador de la entidad del lado *uno* (tabla principal):

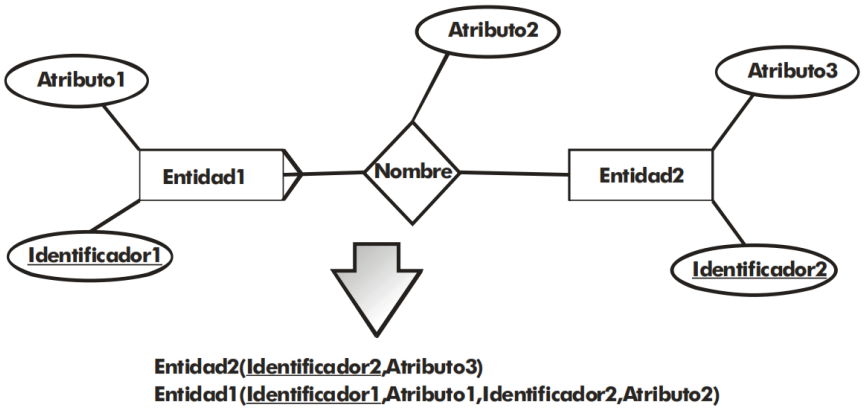


Figura 3.8 Esquema de relaciones uno a mucho

3.3.3 Modelado de bases de datos usando ontologías

Para modelar el dominio de viaje usando un diagrama Entidad-Relación (ER), con la extensión común de relaciones de generalización entre entidades, se pueden modelar las clases a través de entidades ER. Las clases pueden estar organizadas en taxonomías de clases con la relación de generalización entre entidades de ER .Así mismo, los atributos se pueden describir a través de los valores de las columnas de la tabla.

Para poder representar relaciones Ad hoc entre las clases de la ontología, se puede hacer uso de relaciones entre las entidades. Estas relaciones pueden tener cualquier aridad. En la siguiente figura (ver Figura 3.9) se representan las relaciones binarias *arrivalPlace* y *DeparturePlace* y la relación ternaria *connects*, que permiten modelar un viaje *(Travel)*.

En la notación ER se pueden modelar cardinalidades en las relaciones. Por ejemplo, se define que *Travel* tiene al menos un *arrivalPlace* y ésta solo una en otro caso con la notación ER es posible representar atributos de relación, por ejemplo: se puede definir *connectionLength* como un atributo de la relación *connects*.

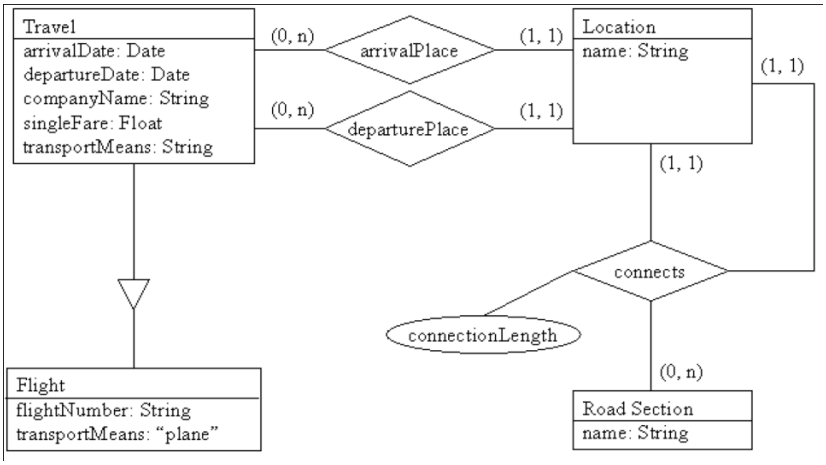


Figura 3.9 Diagrama Entidad-Relación para el dominio de viaje

El lenguaje para la representación de las bases de datos relacionales *Strutcture Query Lenguague* (SQL) permite crear instancias a través de la sentencia insert. Sin embargo, SQL no es apropiado para representar restricciones formales en una forma declarativa, puesto que las reglas de consistencia deben especificarse usando triggers o procedimientos almacenados.

Las vistas de SQL son útiles para representar inferencias a partir de las tablas de la BD. Se ha empleado esta tecnología para construir ontologías, aunque esta tecnología no siempre es apropiada, el modelo resultante resulta ligero y fácil de utilizar en aplicaciones que requieran una manipulación de grandes volúmenes de datos.

3.4 Reconocimiento de patrones

El reconocimiento de patrones es la ciencia que se encarga de la descripción y clasificación (reconocimiento) de objetos, personas, señales, representaciones, etc. Esta ciencia trabaja con base en un conjunto previamente establecido de todos los posibles objetos individuales a reconocer. El reconocimiento de patrones se clasifican generalmente de acuerdo con el tipo de procedimiento de aprendizaje utiliza para generar el valor de salida teniendo la siguiente clasificación:

- **Aprendizaje supervisado**: El aprendizaje supervisado supone que se ha proporcionado un conjunto de datos de entrenamiento, el cual consiste de un

conjunto de instancias que han sido debidamente etiquetados de forma manual con una salida correcta.

- **Aprendizaje no supervisado:** El aprendizaje no supervisado asume datos de entrenamiento que no ha sido manualmente etiquetados. Se trata de encontrar patrones inherentes en estos datos, los cuales se pueden emplean para determinar un valor de salida correcto para las instancias de datos nuevos.

- **Semi-aprendizaje supervisado**: Una combinación de los dos citados anteriormente y que recientemente se ha explorado. Utiliza una combinación de datos etiquetados y no etiquetados (típicamente un pequeño conjunto de datos etiquetados combinado con una gran cantidad de datos no etiquetados).

Hay que tener en cuenta que a veces se utilizan diferentes términos para describir los procedimientos correspondientes del aprendizaje supervisado y del no supervisado. Por ejemplo, el equivalente de clasificación no supervisada es normalmente conocido como agrupamiento. Por otra parte, en algunos campos la terminología es diferente: por ejemplo, en ecología de comunidades, el término "clasificación" se utiliza para referirse a lo que comúnmente se conoce como *"clustering"*.

La pieza de datos de entrada para la que se genera un valor de salida se denomina formalmente una instancia. La instancia se describe por un vector de características, que en su conjunto constituye una descripción de todas las características conocidas de la instancia (estos vectores de características puede ser vistos como la definición de una serie de puntos en un espacio multidimensional adecuado, donde los métodos para la manipulación de vectores en los espacios vectoriales pueden ser aplicados a ellos, tales como el cálculo del producto punto o el ángulo que se forma entre dos vectores). Típicamente, las características son de tipo:

- **Categórico.** También conocido como nominal, donde una característica puede tomar uno de los valores de un conjunto cuyos elementos son desordenados, por ejemplo el tipo de género que puede tomar el valor de "masculino" o "femenino", o el tipo de sangre que puede tomar alguno de los siguientes valores: "A", "B", "AB" u "O").

- **Ordinal.** Que consiste en un conjunto de elementos ordenados, por ejemplo, "grande", "mediano" o "pequeño", con valores enteros (un recuento del número de veces que aparece una palabra en un correo electrónico) o de valor real (una medición de la presión arterial).

Con frecuencia, los datos categóricos y ordinales se agrupan juntos, del mismo modo para los datos de valores enteros y valores reales. Adicionalmente, muchos algoritmos funcionan solo en términos de los datos categóricos y requieren que un valor real o un valor entero de datos sea discretizado en grupos (por ejemplo, menos de 5, entre 5 y 10, o mayor que 10).

3.5 La distancia de edición (distancia de Levenshtein)

La distancia de edición basada en la métrica de Levenshtein [37], tiene como propósito inicial medir semejanzas entre dos secuencias de símbolos, ha sido fuertemente utilizada en el área de la informática en problemas de procesamiento de lenguaje natural y también en el campo de la bio-informática, permitiendo la comparación de secuencias protéicas y genómicas.

3.5.1 Definiciones

A continuación se describirán una serie de definiciones importantes, relacionadas con la distancia de Levenshtein.

Definición de similitud de cadena

Es necesario definir el concepto del grado de semejanza que pueden tener dos cadenas. El concepto de similitud es en si, la cantidad de operaciones de edición o de cambio que son necesarias para transformar una secuencia en otra. Estas operaciones de edición o cambio no son más que las de insertar y borrar los símbolos necesarios para que una cadena quede igual a la otra. Para dejar claro el concepto de similitud se requieren las siguiente definiciones [37]:

Definición 1. Sean $x = \langle x_1, x_2, ..., x_n \rangle$ y $y = \langle y_1, y_2, ..., y_m \rangle$ dos secuencias finitas de símbolos en algún alfabeto finito C, se dice que y es una sub secuencia de x, denotado por $y \subset x$, si existe un conjunto de índices $\{i_1, i_2, ..., i_m\}$ en x, donde $1 \le i_k \le n$, y $1 \le k \le m$, tales que $i_1 < i_2 < .... < i_m$ y que $y = \langle x_{i_1}, x_{i_2}, ..., x_{i_m} \rangle$.

Definición 2. Sean las secuencias finitas y, x_a, x_b se dice que y es una subsecuencia común para las secuencias x_a y x_b, denotado por $y \subset (x_a, x_b)$, si y solo si $y \subset x_a$ y $y \subset x_b$.

Definición 3. La similitud entre dos secuencias $x, y \in C^*$, denotada por $S(x,y)$ está dada por $S(x,y) = \max\{|z| \subset (x,y)\}$; con $z \in C^*$.

Donde $|z|$ indica la longitud de la secuencia z, es decir, la cantidad de símbolos que contiene.

- Cuando $S(x,y) = 0$ esto implica que x no tiene símbolos comunes con y, lo cual significa que no existe una cadena similar contemplando lo siguiente: $0 \le S(x,y) \le \min\{|x|, |y|\}$.

Definición de la métrica Levenshtein

La distancia de Levenshtein se describe ampliamente en [25], en donde se cita que esta distancia se define entre dos cadenas $x, y \in C^*$, con $n = |x|, m = |y|$, está definida por medio de la expresión:

$$D(x,y) = n + m - 2S(x,y) \qquad (3.1)$$

Donde $S(x,y)$ es la similitud entre las secuencias x e y. Los límites de esta distancia se logran, por un lado cuando la similitud entre las secuencias comparadas es nula, y en el otro extremo, cuando la similitud entre las secuencias comparadas es máxima.

Cuando la similitud es nula (secuencias sin símbolos comunes), la distancia $n+m$. Cuando la similitud es máxima (se comparan dos secuencias iguales), la distancia es

igual a 0. Con base en lo anterior, la distancia de Levenshtein está acotada por los siguientes límites (ver Ecuación 3.2).

$$0 \leq D(x,y) \leq n+m \qquad (3.2)$$

3.5.2 Algoritmo de cálculo

En esta sección se describirá brevemente la forma de realizar el cálculo de la distancia de Levenshtein, para lo cual se considerará un ejemplo, que consistirá en analizar la matriz de cambios de comparación de las palabras FÁCIL y DIFÍCIL. El algoritmo trabaja de la siguiente manera [37]:

Sean n la longitud de la cadena$_1$ (cadena origen) y m la longitud de la cadena$_2$ (cadena destino), el algoritmo usa una matriz de tamaño $(n+1)\times(m+1)$,se va a representar esta matriz de modo que su origen (0,0) se encuentra en la esquina superior izquierda, en consecuencia las filas crecen de arriba hacia abajo. Vamos a ver los pasos que da el algoritmo para computar dicha distancia:

Primer paso:

1. Se asigna a n el valor de la longitud de la cadena concurrente.
2. Se asigna a m el valor de la longitud de la cadena "vidas".
3. Se crea una matriz de n filas y m columnas. Es decir, se colocará en vertical la palabra de partida, y en horizontal la palabra objetivo.
4. Se inicializa la primera fila de 0 a m y la primera columna de 0 a n.
5. *Segundo paso:* Para poblar la matriz (ver Figura 3.10) se debe tener en cuenta que el elemento de coordenadas (i,j) será el mínimo entre:
6. La celda inmediatamente superior más 1: $matriz[i-1][j]+1$.
7. La celda inmediatamente a la izquierda más 1: $matriz[i][j-1]+1$.
8. La celda diagonal superior izquierda ,en este caso hay dos alternativas:
    a. Si $cadena_1[i-1] = cadena_2[j-1]$, entonces sería el valor de la diagonal tal cual: $matriz[i-1][j-1]$.

b. Si $cadena_1[i-1]! = cadena_2[j-1]$, entonces sería el valor de la diagonal más 1, es decir: $matriz[i-1][j-1]+1$.

		F	A	C	I	L
	0	1	2	3	4	5
D	1	1	2	3	4	5
I	2	2	2	3	3	4
F	3	2	3	3	4	4
I	4	3	3	4	3	4
C	5	4	4	3	4	4
I	6	5	5	4	3	4
L	7	6	6	5	4	3

Figura 3.10 Matriz de similitud de las palabras DIFÍCIL y FÁCIL

3.5.3 Aplicaciones

Como se mencionó anteriormente, la distancia de edición ha incursionado de ciertas áreas del conocimiento para funciones como las siguientes áreas fundamentalmente:

- Sistemas para la revisión de faltas ortográficas automatizada en textos.
- Sistemas de reconocimiento de voz
- Sistemas para el análisis de ADN y cadenas proteicas
- Sistemas para la detección de plagios.
- Taxonomías.

Siendo más específicos su labor en el hábito en los clasificadores es fundamental, para la biología, han sido utilizados como parte de la implementación de árboles ultra métricos.

Otra aplicación práctica donde ha sido de utilidad, es en la implementación de algoritmos de *k-means* para el cálculo de centroides proteícos en las secuencias de ADN, esta distancia de cambio, también se utiliza en la comparativa y cálculo de secuencias representativas.

3.6 Distancia Manhattan (Métrica City block)

Definida por Herman Minkowski en el siglo XIX, esta distancia descrita ampliamente en [38], [40] permite calcular la ***geometría Taxicab***, siendo una forma en la que la función de distancia usual o métrica de geometría euclidiana se sustituye por un nuevo indicador en el que la distancia entre dos puntos es la suma de las diferencias absolutas de su coordenadas. La métrica *Taxicab* también se conoce como distancia rectilínea, distancia de la ciudad de bloques, distancia Manhattan, o la longitud de Manhattan, con las correspondientes variaciones en el nombre de la geometría.

3.6.1 Definiciones

La distancia d_1 entre dos vectores p,q en un espacio vectorial fijo real de n dimensiones, con sistema de coordenadas cartesianas, es la suma de las longitudes de las proyecciones del segmento de línea entre los puntos en los ejes de coordenadas. Formalmente en [26], se define:

$$d_1(p,q) = \|p-q\|_1 = \sum_{i=1}^{n} |p_i - q_i| \tag{3.3}$$

Donde los vectores p,q están expresados como:

$$p = (p_1, p_2, ..., p_n) \text{ y}$$
$$q = (q_1, q_2, ..., q_n) \tag{3.4}$$

Considerando lo anterior, en un plano bidimensional, la distancia de taxi entre (p_1,p_2) y (q_1,q_2) es $|p_1 - q_1| + |p_2 - q_2|$.

3.6.2 Algoritmo de cálculo

A continuación se describen los pasos que se deben seguir para calcular esta distancia [38]:

1. Se ingresan dos vectores con el mismo número de k componentes.
2. Se recorren los dos vectores simultáneamente, en su *i*-ésima componente.

3. Se resta la *i*-ésima componente del vector 2 a la *i*-ésima componente del vector 1, obteniendo como resultado final el valor absoluto de la diferencia.
4. Se acarrea esta diferencia para sumarle a la diferencia de las demás componentes hasta que $i = k$.

3.6.3 Aplicaciones

Originalmente, se empleó la distancia Manhattan para resolver el problema relacionado con el cálculo de distancias en espacios metropolitanos, en el que cotidianamente existen obstáculos, a diferencia de la distancia Euclidiana, para obtener resultados más prácticos ya que considera los movimientos en una retícula (el nombre Manhattan viene de la isla y la distribución de sus calles en forma de retícula), ver Figura 3.11. Por otra parte, en el trazado de rutas de camiones y de distribución es comúnmente recurrida, debido a que contempla este tipo de situaciones espaciales. En análisis de los problemas clásicos de la computación sobre las estrategias de juego en el ajedrez, es usada para los cálculos de los movimientos de las piezas del juego.

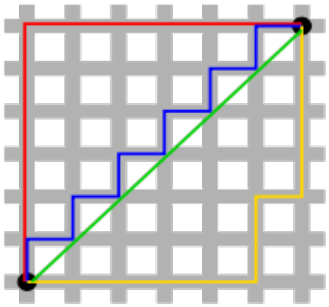


Figura 3.11 Comparativo entre la distancia Manhattan(coloreada en tono azul) y la distancia Euclidiana(descrita con color verde).

3.7 Definición de densidad delictiva

La densidad es un término utilizado en diferentes áreas de estudio, en el contexto social se utiliza la definición de densidad de población, también denominada *población relativa*, la cual se refiere al número promedio de habitantes de un área urbana o rural con relación a una unidad de superficie dada [39]. Su fórmula se describe de la siguiente forma:

$$Densidad = \frac{Delitos}{Área} \qquad (3.5)$$

Dada la anterior definición de densidad, podemos definir a la densidad de delitos como la una medida que representa la frecuencia de eventos delictivos sobre un área proyectada en un círculo relacionado con un punto de interés; lo cual se define como sigue:

$$Densidad = \frac{Evento_Delictivo}{Área} \tag{3.6}$$

La densidad delictiva, corresponde a los eventos producidos sobre un área determinada, en este caso con una temática delictiva [40].

Capítulo 4. Metodología

4.1 Introducción

En este capítulo se describen los elementos y tecnologías utilizadas para el desarrollo del prototipo de la tesis, primeramente se aborda la arquitectura del sistema, se describen a grandes rasgos los componentes y las herramientas utilizadas, ubicando su interacción entre ellas. Posteriormente, se aborda el proceso general y el desarrollo de los modelos empleados para el análisis.

4.2 Arquitectura del sistema

El sistema ha sido implementado en Java y la componente web de la aplicación ha sido construida utilizando la API de GeoTools. Asimismo, el enfoque de este marco de trabajo radica en el uso de software libre, por tal motivo, se utiliza como motor de base de datos a PostgreSQL, particularmente la extensión para trabajar con datos geo espaciales PostGIS. En la Figura 4.1 se presenta la arquitectura general de la aplicación.

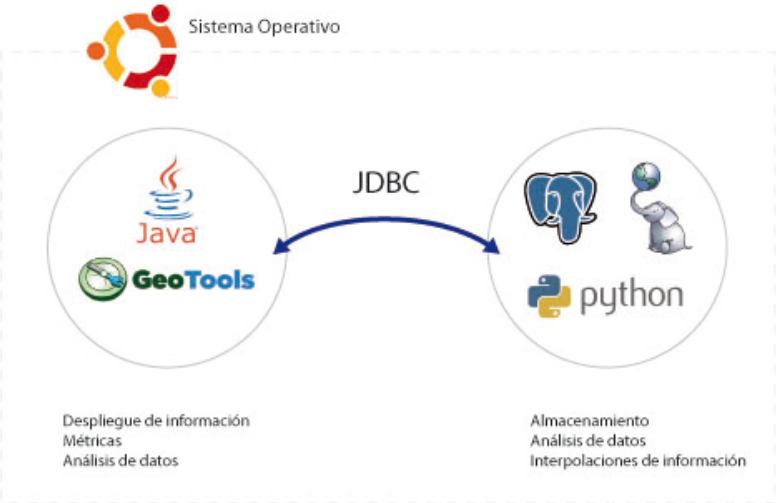


Figura 4.1 Arquitectura general de la aplicación

Por otra parte, se puede observar que la arquitectura está compuesta de de dos grandes elementos, que a su vez están definidos por varias componentes o*plug-ins* de los cuales dependen.

La aplicación por motivos de compatibilidad de versiones, fue desarrollada en Linux Ubuntu 12.10, para el desarrollo de los módulos de despliegue, análisis e implementación de métricas éstas fueron desarrolladas en Java. Asimismo, el despliegue de la información geográfica fue utilizado el marco de trabajo de GeoTools.

Con respecto a la parte de almacenamiento, se cuenta con una base de datos de Postgres 9.0 con el *pluging* de datos geográficos PostGIS 1.5.1, los cuales son el núcleo de almacenamiento de la aplicación. Por motivos de eficiencia, algunos componentes de desarrollo y análisis son implementados directamente en la parte de la base de datos como procedimientos, para dichas funciones fue necesario apoyarse en el lenguaje script PL/Python.

Por último, se tiene una conexión a través de un JDBC que hace referencia a Postgres para Java, con la finalidad de interactuar con estos dos elementos libremente.

4.3 Marco general de trabajo

El marco general de trabajo para esta tesis se compone de las etapas de adquisición de datos, conexión, análisis y métricas. Estas etapas se encargan de realizar tareas específicas, las cuales van desde la forma de obtención de los datos, la conexión a la base de datos para extraer la información, análisis de la información delictiva con base en métodos de interpolación y clasificación, así como métricas para medir el grado de confiabilidad, mediante técnicas basadas en reconocimiento de patrones.

En la Figura 4.2 se muestra el diagrama a bloques de la metodología para el análisis delictivo.

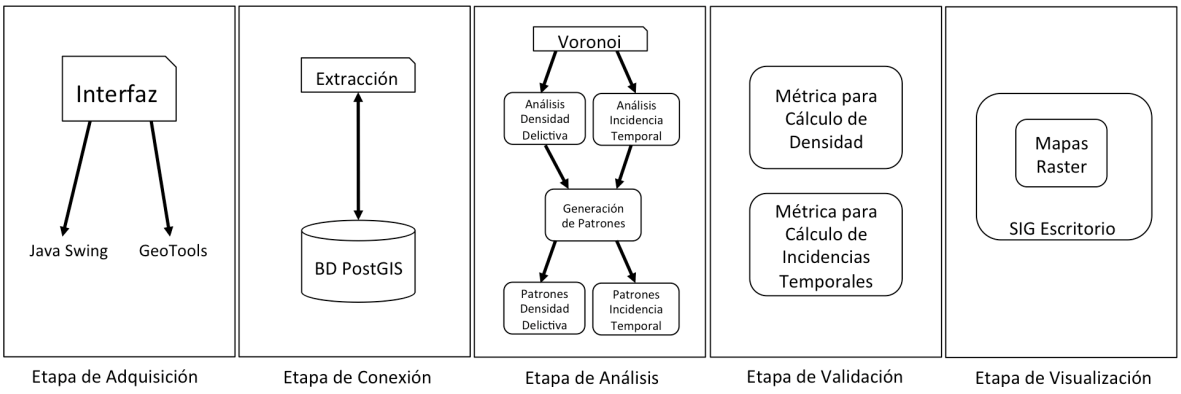


Figura 4.2 Arquitectura general de la aplicación por componentes.

Como se observa en la figura anterior, el proceso inicia con la generación de una interfaz gráfica de usuario, para determinar el número de clases a generar, con respecto a los patrones delictivos que se deseen clasificar.

Posteriormente, es necesario realizar una extracción de estos datos para su análisis, los cuales se encuentran almacenados en una base de datos implementada en PostGIS; una vez que se tienen los datos delictivos extraídos, se aplica una triangulación basada en el método de diagramas de Voronoi para contar con una clasificación previa de las áreas con incidencia delictiva similares. Para ello, se aplican dos métodos de análisis: el primero enfocado en la densidad delictiva, el cual permite identificar la frecuencia de delitos por unidad de área y el segundo, de incidencia temporal, el cual se enfoca en conocer el patrón temporal sobre la incidencia delictiva.

Por otra parte, en la etapa de validación, se lleva a cabo un proceso de verificación de los patrones generados; es decir, se analiza la similitud de características mediante dos enfoques: el primero es una métrica para el cálculo de la densidad y el segundo es una métrica para el cálculo de las incidencias temporales.

Finalmente los resultados de estos análisis se muestran en una aplicación SIG de escritorio. Cabe señalar que los resultados obtenidos se muestran en mapas con formato raster.

En la sección 4.4 se describe el diagrama de componentes que ha sido definido para esta aplicación.

4.4 Diagrama de componentes

En la Figura 4.3 se muestra el diagrama de componentes de la aplicación, el cual está compuesto por el componente “interfaz”, el componente “análisis”, el componente “conexión” y el componente “métricas”.

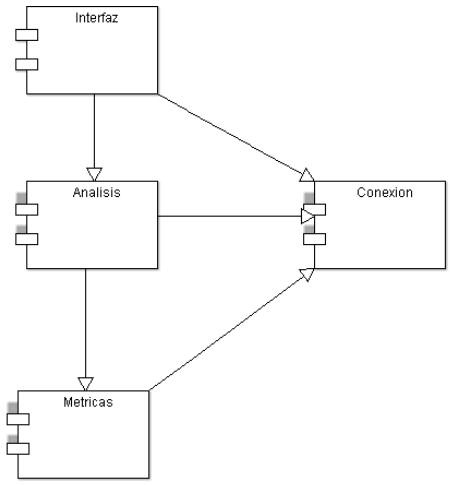


Figura 4.3 Diagrama de componentes de la aplicación.

4.4.1 Etapa de adquisición

En esta etapa se define el bloque de componentes relacionado con la interfaz del sistema. Por tanto, en este componente se tienen todos los mecanismos para la interacción y el despliegue de la información en el prototipo, la cual le da sentido a la información almacenada en el despliegue de la aplicación. La interfaz está generada con dos componentes básicos:

- **Java Swing** (utilizando el *plugin* de Eclipse WindowsBuilder) es un framework para desarrollar interfaces gráficas con Java y con independencia de la plataforma. Utiliza un simple modelo de programación por hilos, y posee las siguientes características:

    - Independencia de plataforma.

- Extensibilidad. Es una arquitectura altamente particionada los usuarios pueden proveer sus propias implementaciones modificadas para sobrescribir las implementaciones por defecto. Se pueden extender clases existentes proporcionando alternativas de implementación para elementos esenciales.

- **GeoTools.** Es una biblioteca para sistemas de información geográfica que permite el despliegue de datos conforme a los estándares de la compañía ESRI. Está escrito en Java y se basa de componentes de Swing para su implementación.

En el diagrama de componentes se puede observar que éste tiene interacción directa con otros dos componentes: el de *análisis* con el que se encarga de mandar llamar las acciones correspondientes según las acciones definidas en la interfaz gráfica de la aplicación, dependiendo de los parámetros de entrada y de los componentes seleccionados es la acción correspondiente a seguir. Con los componentes de conexión se comunica directamente para el despliegue de información de la base de datos en tablas de información y mapas.

4.4.2 Etapa de conexión

La etapa de conexión contempla la estructura de la base de datos, en la cual se realizan las operaciones de análisis, generación y recuperación de los patrones. El modelo inicial de datos propuesto se muestra en la Figura 4.4.

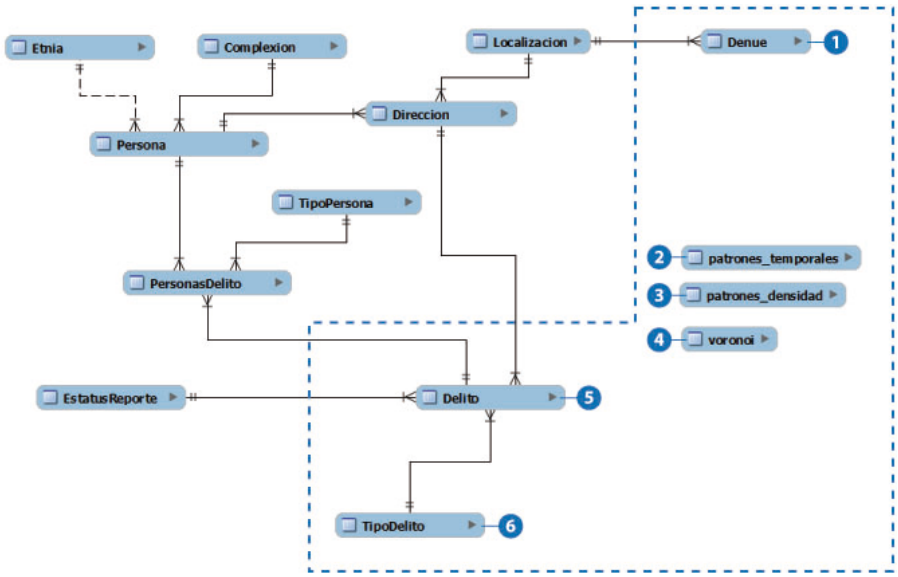


Figura 4.1 Modelo de datos propuesto para el sistema.

En el modelo inicial se contemplaba generar un análisis con más variables, pero debido a la falta de información en los rubros no señalados, se trabajó con la información contenida y generada en las tablas que están marcadas en los recuadros, por lo que la definición de datos se da conforme a la información de las siguientes tablas:

DENUE (Directorio Estadístico Nacional de Unidades Económicas). El DENUE proporciona información actualizada de todas las actividades económicas que tienen lugar en el territorio nacional. Sobresalen las que se desarrollan en la industria manufacturera, el comercio y los servicios privados, sectores en los que se encuentra el mayor número de unidades económicas y de personal ocupado. Las actividades de transporte y construcción se captaron por empresa, por lo que sólo se encuentra en el directorio, la unidad económica matriz o informante de cada razón social.

Para el caso de la presente tesis, la capa de información del DENUE corresponde a los puntos de interés en nuestro estudio, a partir de estos puntos los análisis se efectúan para calcular la *densidad delictiva* y la *densidad temporal*, la cual más adelante será detallada.

Delito. Corresponde a la información puntal del delito a analizar, en la cual está vinculado el tipo de delito, la localización y la fecha de ocurrencia. Los datos contemplados pero sin información corresponden a los testigos y los acusados, los cuales corresponden a un tipo de persona.

Esta información corresponde a la capa de datos con la que los puntos del DENUE se cruzan para la generación de sus patrones delictivos (a partir del DENUE).

Un aspecto a considerar es que para el presente trabajo y en el caso de estudio se utilizan datos sintéticos debido a la naturaleza sensible de contar con datos reales.

Tipo_Delito. En esta tabla se presenta una relación del tipo de delito con un identificador, este identificador es el que se utiliza para formar los patrones y la tabla de Tipo_Delito es más un catálogo de tipos para las clases de delito.

Patrones_temporales. Tabla de soporte que genera la información de patrones de carácter temporal, la información de esta tabla es generada al momento de efectuar los procesos.

Patrones_densidad. Tabla que contiene información de patrones de densidad de delitos por área, de la misma forma que la tabla temporal, ésta genera la información específica con el proceso.

Voronoi. Otra tabla de soporte, la cual ayuda a la interpolación de datos a través de la información obtenida, para generar una triangulación de Voronoi. Para calcular esta tabla se corre un proceso generado en pl/python que calcula la triangulación con base en la información *geométrica* obtenida de una tabla en PostGIS.

4.4.3 Etapa de análisis

En esta etapa se lleva a cabo el análisis de los delitos, con la finalidad de establecer grupos homogéneos que puedan ser clasificados, con base en características similares. Para ello, se ha implementado el método de interpolación de Voronoi para obtener las áreas de incidencia común, a partir de una tabla de puntos que representan los *delitos* de una zona en particular. Primeramente se ejecuta un procedimiento, basado en el algoritmo de *Fortune* [40], con la finalidad de calcular la triangulación de Voronoi.

Cabe señalar que este algoritmo fue seleccionado al ser de propagación lineal y presentar un tiempo de ejecución de *n log(n),* siendo una cantidad de puntos considerable para generar la interpolación, podría llegar a ser exhaustiva con otros algoritmos que resultaban menos eficaces. Además existe muy buena documentación de las implementaciones sobre este algoritmo. En contraparte, el algoritmo de *k-means* está más orientado para la agrupación (clustering) de clases.

En la Tabla 4.1 se describe el pseudo código del algoritmo propuesto para esta tarea.

Tabla 4.1 Pseudo código del algoritmo para el cálculo de la Triangulación de Voronoi.

dejar que $*(z)$sea la transformación $*(z) = (z_x, z_y + d(z))$, donde $d(z)$ es la distancia euclídea entre zy el sitio más cercano
vamos a Tser la "línea de playa"
dejar que $R_p$sea la región cubierta por el sitio p.
dejar que C_{pq}sea el rayo límite entre sitios py q.
vamos a $p_1, p_2, \cdots, p_m$ser los lugares con mínima $y$coordenada, ordenada por $x$la coordenada
$Q \leftarrow S - p_1, p_2, ..., p_m$
crear rayos iniciales límites verticales **aunque no** IsEmpty ( ) **no** ← DeleteMin ( ) **caso** de un sitio en : $C^0_{p1,p2}, C^0_{p2,p3}, ...C^0_{pm-1,pm}$
$T \leftarrow *(R_{p1}), C^0_{p1,p2}, *(R_{p2}), C^0_{p2,p3}, ..., *(R_{p_{m-1}}), C^0_{p_{m-1},p_m}, *(R_{p_m})$
$Q\, p\, Q\, p\, p*(V)$
encontrar la ocurrencia de una región $*(R_q)$en Tque contiene p,
entre corchetes por C_{rq}la izquierda y C_{qs}en la derecha
crear nuevos rayos de frontera C^-_{pq}y C^+_{pq}con bases p
reemplazan $*(R_q)$con $*(R_q), C^-_{pq}, *(R_p), C^+_{pq}, *(R_q)$en T
borrar de Qcualquier intersección C_{rq}y C_{qs}
se insertan en Qcualquier intersección C_{rq}y C^-_{pq}
se insertan en Qcualquier intersección C^+_{pq}y es un vértice de Voronoi : $C_{qs}\, p*(V)$
vamos a $p$ser la intersección de C_{qr}la izquierda y C_{rs}la derecha

dejar que C_{uq}sea el vecino izquierdo C_{qr}y
dejar que C_{sv}sea el vecino de la derecha de la C_{rs}de T
crear un nuevo rayo límite C^0_{qs}si $q_y = s_y$,
o crear C^+_{qs}si $p$es correcto en el mayor de qy s,
crear contrario C^-_{qs}
sustituir $C_{qr}, *(R_r), C_{rs}$con nueva creación C_{qs}en T
borrar de Qcualquier intersección C_{uq}y C_{qr}
eliminar en Qcualquier intersección C_{rs}y C_{sv}
insertar en Qcualquier intersección C_{uq}y C_{qs}
insertar en Qcualquier intersección C_{qs}y C_{sv}
registro $p$como la cumbre de C_{qr}y C_{rs}y la base de la C_{qs}
salida de los segmentos de contorno C_{qr}y **endcase ENDWHILE**
salida restantes rayos límites enC_{rs}

T

Posteriormente se establece una conexión a la base de datos de PostGIS para ejecutar una llamada que permita realizar la inserción de valores a la tabla diseñada para tal fin. En este caso la llamada es la siguiente:

```
Insert into voronoi(select * from voronoi('denue', 'geom') as (id integer,
                                the_geom geometry));
```

En este caso, tanto la generación de los patrones como el cálculo de Voronoi son procesos que pueden tardar un poco; por esta razón se optó por mantener los valores en las tablas de apoyo para no re calcular en cada ejecución del programa dichos valores.

El método de Voronoi se aplica primero para tener la referencia de un polígono de incidencia de cada uno de estos puntos, estos polígonos de afectación son los que serán visualizados para representar el mapa delictivo de la zona y a su vez son reutilizados en ambos cálculos.

De encontrarse nuevos puntos de estudio, sería necesario recalcular esta capa de información para poder vincular la afectación de la zona con respecto a los puntos ya contemplados.

En esta etapa, se lleva a cabo el análisis de los patrones delictivos, con la finalidad de obtener representaciones cartográficas que permitan analizar y visualizar la distribución espacial de los delitos.
Para ello, se proponen dos métodos orientados a generar este tipo de representaciones: el análisis de densidad delictiva y el análisis de incidencia temporal. Ambos métodos han sido desarrollados en Java y cuentan con un correspondiente análisis espacial que se realiza en PostGIS.

- *Análisis de densidad delictiva.* Es el análisis de la frecuencia de eventos delictivos en un radio determinado.

- *Análisis de incidencia temporal.* El análisis temporal, es la frecuencia de eventos delictivos con respecto a un periodo de tiempo.

Ambos análisis pueden ser desarrollados a tres niveles: *micro*, *meso* y *macro*, la única diferencia es el nivel de detalle seleccionado del universo en el análisis de datos.

El análisis *micro*, representa un nivel de detalle mayor, ya que se contempla un área de estudio y análisis con un radio menor de los trescientos metros, el cual sirve para localizar eventos delictivos aledaños.

El análisis *meso*, representa un nivel de detalle medio, el cual se calcula con un radio de afectación de delitos no mayor a los setecientos metros.

El análisis *macro*, es el detalle general del análisis, contempla mayor distancia calculándose a un radio de mil quinientos metros.

El *radio de afectación* es la distancia de la cual se basa el análisis para tomar datos para la generación de patrones, por lo que se podría decir que el radio de afectación corresponde en realidad a la medida real en metros del nivel de detalle deseado.

Los niveles de análisis corresponden como se mencionó a un "nivel de detalle" este nivel se puede entender como una analogía al observar una pintura a la distancia, entre mayor distancia se tenga, vemos la pintura en su totalidad; a medida que nos aproximamos comenzamos a encontrar detalles de las pinceladas que crearon el paisaje que vemos pintado, de esta manera entre mayor nivel de detalle tenemos un panorama general de nuestro mapa delictivo se presenta y entre menor distancia vemos los detalles son más específicos.

El análisis de *densidad delictiva* y de *incidencia temporal*, presentan una secuencia de eventos similares, por lo que pueden ser analizados con los mismos casos de uso, las diferencias radican en las secuencias que siguen y los métodos que invocan para generar patrones y las métricas necesarias para medir dichos análisis.

El proceso de generación de resultados sigue dos procesos, uno para generar los patrones con base en la información almacenada y otro proceso para poder realizar un análisis compartido de dichos patrones.

4.4.3.1 Proceso de generación de patrones

Para el proceso de generación de patrones se realizaron las siguientes tareas:

1. El usuario selecciona el tipo de análisis y el nivel de detalle, los cuales se ingresan como parámetros.

2. Se envía la petición y se generan los patrones, con base en estos parámetros se decide el nivel de detalle a considerar y el tipo de patrones a generar. Los datos a los cuales se les genera el patrón corresponden a la tabla del DENUE. Estos valores son ponderados y se relacionan con la tabla de DELITOS. En este caso, cada punto de DENUE se compara con los puntos de los reportes de delitos, tomando en consideración los parámetros.

3. Los patrones son cargados a la base de datos, correspondiendo a una tabla de patrones para cada tipo de análisis.

4. Finalmente, se envía una respuesta al usuario de que la acción es generada.

Este proceso es importante ya que los resultados de los análisis están vinculados a los análisis desarrollados en este proceso. Por tanto, éste proporciona la ventaja de que si se ingresan nuevos datos a la base de datos éstos se pueden comparar y visualizar.

El proceso de generación de patrones en la aplicación por motivos del caso de estudio se ha dejado a que haga análisis de todos los puntos en el universo, pero los métodos desarrollados contemplan el desarrollo del patrón por punto, con lo que si existen nuevos puntos a considerar estos pueden generarse el patrón sin problemas considerando el universo existente (ver Figura 4.5).

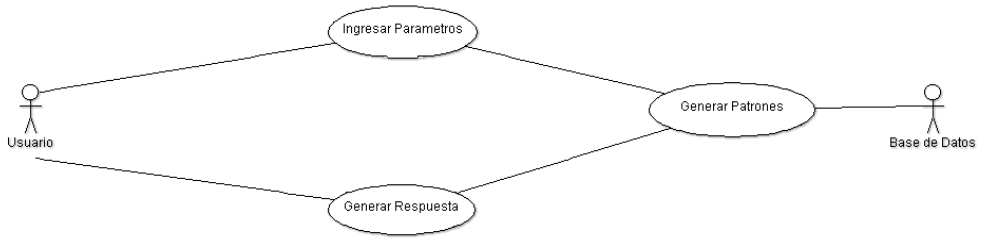


Figura 4.5 Caso de uso de generación de patrones

4.4.3.2 Patrones de densidad delictiva

Los patrones de densidad están formados por todos los puntos de la tabla de DELITOS, los cuales se encuentran distribuidos en un radio específico, según el nivel de detalle.

Los patrones se van formando con la sumatoria de incidencias del mismo tipo de delito, siempre y cuando estén dentro de la circunferencia formada por el radio de distancia correspondiente a cada nivel de detalle, lo cual se denota a través de la ecuación de la sumatoria de la métrica de *city block* (ver Ecuación 3.3).

En la Figura 4.6 se muestra una representación espacial en donde se puede apreciar que se define un radio que permite generar la densidad de los delitos de con su distribución espacial. En esta figura se puede observar el esquema general de cómo se calculan estos patrones; se tiene un punto de interés (color blanco), del cual dependiendo del nivel de detalle se genera una circunferencia que envuelve a los puntos (delitos), los datos

contenidos dentro de esta circunferencia son considerados en el análisis, ya que están dentro del radio de afectación considerado en el análisis.

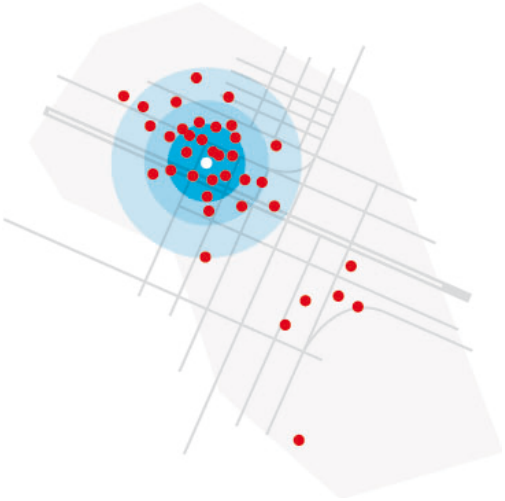


Figura 4.6 Densidad de delitos distribuidos espacialmente

Por ejemplo, para este caso en la Tabla 4.2 se tiene un punto, entonces de acuerdo con la distribución espacial y el radio definido se tiene la siguiente cantidad de delitos:

Tabla 4.2 Tabla de incidencia de delitos por tipo

Tipo de Delito	Cantidad de Delitos
1	3
2	1
3	1
4	3
5	1
6	0

De acuerdo con la información presentada en la Tabla 4.2, se genera un vector que contiene la relación anterior de la siguiente manera:

$$V_d = \{3,2,1,1,3,1,0\}$$

En donde cada *R* del vector corresponde a una clase de delito.

4.4.3.3 Patrones de incidencia temporal

Los patrones de incidencia temporal, al igual que la densidad se proyectan en una circunferencia con base en el valor del nivel de análisis seleccionado. Las diferencias radican en dos aspectos fundamentales, el primero y el más importante, se ordenan los puntos con respecto a la fecha de incidencia, ordenando la primer incidencia en tiempo con respecto a la última que hacen referencian a los puntos que se encuentren dentro del radio del punto de estudio.

La segunda se refiere al límite de entrada, ya que se puede limitar a una cantidad máxima de puntos a considerar para el análisis. Por omisión el valor dado es de 10 puntos en el ordenamiento, de esta manera se ordenan los puntos y se van ingresando a un vector temporal donde los puntos son almacenados para su posterior clasificación. Cabe señalar que estos puntos representan los delitos distribuidos espacialmente en un área determinada. Un ejemplo de lo anterior se muestra en la Figura 4.7:

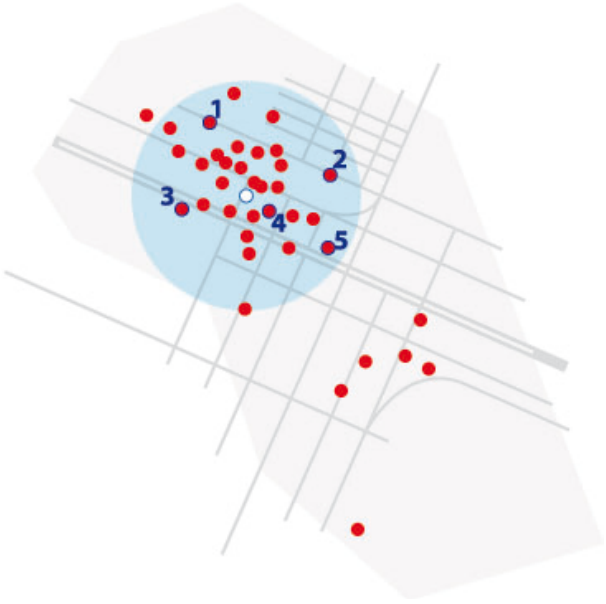


Figura 4.7 Patrones de incidencia temporal.

En la figura anterior se observa el punto de interés marcado con color azul y se aprecia una numeración del 1 al 5, suponiendo que dejamos el límite del patrón en 5 puntos estos números corresponderían al ordenamiento temporal de los delitos, es decir, el primero en ocurrir y así sucesivamente.

Posteriormente, es necesario desplegar la clase a la que pertenece este delito, tal como se muestra en la Tabla 4.3.

Tabla 4.3 Ordenamiento de delitos clasificados por su tipo.

Orden	Tipo Delito
1	3
2	3
3	1
4	2
5	1

Por tanto, de esta manera se genera una cadena correspondiente a este ordenamiento, representada por: Cadena = “33121”. En consecuencia, esta cadena es lo que se considera como el patrón temporal armado.

4.4.3.4 Proceso de análisis de patrones

Para realizar el análisis de patrones, es necesario que las tablas de soporte tengan almacenados los datos, ya que el proceso se basa directamente en las consultas y resultados de la información contenida en estas tablas. En la Figura 4.8 se muestra el proceso de análisis de patrones.

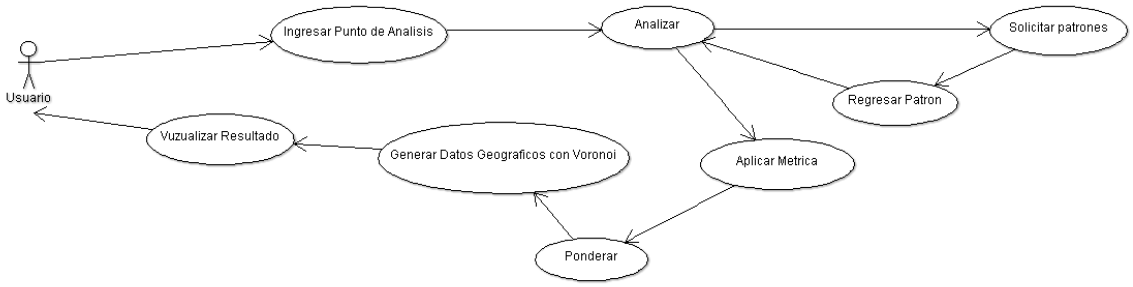


Figura 4.8 Caso de uso de patrones de incidencia temporal

El proceso descrito en el caso de uso anterior, consiste fundamentalmente de las siguientes tareas:

1. Se solicita el punto de interés del DENUE con el que se va a comparar el resto del universo, el objetivo es trazar un mapa del delito conforme a la información del punto de interés para saber qué zonas tienen un patrón similar.

2. Se inicia el análisis recuperando la información de los patrones correspondientes (temporales o de densidad según el caso elegido).

3. Se recuperan los patrones y se inicia la ponderación de la información con base en las métricas, y se genera un resultado ponderado con respecto a la similitud de los puntos en cuestión con respecto al punto de interés.

4. Se interpolan estos datos, utilizando el método de triangulación de Voronoi para obtener una capa de datos geográficos que represente el mapa de la situación del delito.

5. Los diagramas de Voronoi obtenidos son desplegados con base en la información ponderada que representa estas áreas. Esta información genera una serie de secciones con colores diferentes según el caso de su similitud.

6. Este proceso es similar en los dos análisis, ya que lo que varía es el patrón y la métrica empleada para cuantificarlos. En la Figura 4.9 se muestra un ejemplo de los patrones clasificados de acuerdo con el método de interpolación de Voronoi.

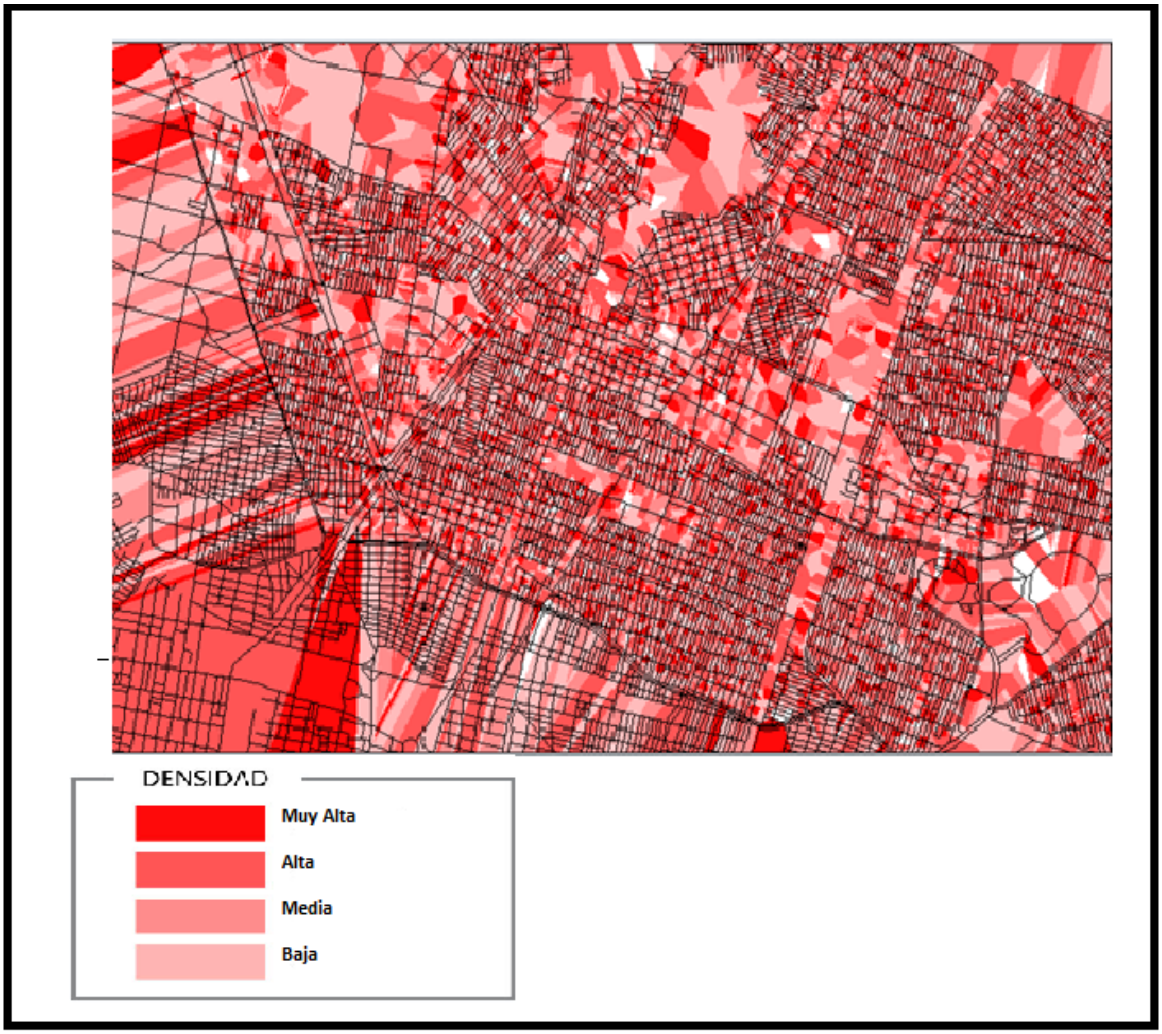


Figura 4.9 Diagramas de Voronoi obtenidos.

4.4.4 Etapa de validación

En esta etapa se lleva a cabo un proceso de validación, el cual se basa en un conjunto de métricas que corresponden también al análisis delictivo. Con estas métricas es posible visualizar de mejor manera los cambios en la ponderación y de ser necesario tener y emplear las métricas en otro tipo reporte.

4.4.4.1 Métrica para el cálculo de densidad

Para el cálculo y la ponderación de las densidades delictivas se optó por usar la métrica de *City Block* definida en la Ecuación 3.3, de la cual se toman los valores de cada componente del vector y se restan contra la misma componente de otro vector, tomando el valor absoluto de la operación; se efectúa una suma de los valores escalares de las componentes de cada *R* del vector, obteniendo así un valor escalar que corresponde a la similitud de vectores, siendo el cero el valor de igualdad entre ambos vectores.

Asimismo, lo anterior se ejemplifica en la Tabla 4.4.

Tabla 4.4 Aplicación del método de City Block.

Vector	R_1	R_2	R_3	R_4	
A	5	3	5	7	
B	9	3	2	6	
Resultado	4	0	3	1	8

El número **8** corresponde a la sumatoria de las *R* resultantes de restar *A* y *B* componente a componente.

La implementación de esta métrica se define como *densidad de delitos*. En la Tabla 4.5 se muestra el código correspondiente a la implementación.

Tabla 4.5 Código correspondiente a la métrica densidad de delitos.

```
public int diferenciaDensidadDelito(int[] patron_a, int [] patron_b){
              int metrica=0;
              for(int i=0; i<patron_a.length ; i++){
                    metrica= metrica + (patron_a[i]-patron_b[i]);
              }
              if(metrica<0)
                    return metrica*(-1);
       return metrica;
}
```

4.4.4.2 Métrica para el cálculo de incidencias temporales

La métrica seleccionada para ponderar los valores de similitud entre los patrones temporales, es la métrica basada en la distancia de Levenshtein (Ecuacion 3.1), ya que es efectiva y muy utilizada para la verificación de patrones con ordenamientos jerárquicos.

De esta manera, se compara el patrón generado en la cadena temporal que previamente se formó.

La implementación de la métrica se describe en la Tabla 4.6, en donde se muestra el algoritmo basado en la distancia de Levenshtein (descrito en la sección 3.5.2).

Tabla 4.6 Código correspondiente a la métrica de incidencias temporales.

```
package Metricas;
public class Levenshtein {
             private String texto;
             private String patron;
             private String TokenMin;
             public Levenshtein(String texto, String patron)
                 {
             this.texto = texto;
             this.patron = patron;
             this.TokenMin = "";
                 }
             public int getDistLev(String s, String t)
                 {
             int n = s.length();
             int m = t.length();
             int[][] d = new int[n + 1][ m + 1];
             if (n == 0)
                     {
             return m;
                     }
             if (m == 0)
                     {
             return n;
                     }
             for (int i = 0; i <= n; d[i][0] = i++)
             { }
             for (int j = 0; j <= m; d[0][j] = j++)
             { }
```

```
            for (int i = 1; i <= n; i++)
                    {
            for (int j = 1; j <= m; j++)
            {
            int costo;
            if(t.charAt(j-1) == s.charAt(i-1))
            costo = 0;
            else
            costo = 1;
            d[i][j]=Math.min(Math.min(d[i-1]
                            [j]+1,d[i][j-1]+1), d[i-1][j-1] +
costo);
                        }
                    }
            return d[n][m];
                }
            public int getDistanciaMinima()
                {
            String[] tokens = texto.split(" ");
                    String cadmin = "";
            int distminima = 99999;

            for(int i=0; i<tokens.length; i++)
            {
                      //obtenemos la distancia por cada palabra del
texto total
            int dist=this.getDistLev(tokens[i],patron);
            if(dist<distminima)
                        {
            distminima = dist;
            cadmin = tokens[i];
                        }
                    }
            this.TokenMin = cadmin;
            return distminima;
                }
            public String getTokenMin()
            {
                    return this.TokenMin;
                }
            }
```

4.4.5 Etapa de visualización

En esta etapa se lleva a cabo la visualización de los datos. Particularmente la información que se despliega es en formato raster, ya que corresponde directamente a los métodos de interpolación y de clasificación utilizados.

En este caso, los resultados obtenidos se despliegan en un sistema de información geográfica de escritorio, presentando mapas raster que reflejan la densidad de delitos, así como de incidencia temporal.

Los resultados son presentados de manera coloreada, la cual dependiendo del análisis tendrá una posible interpretación. Para los patrones de densidad, los colores entre más roja sea la zona más alta la densidad de delitos es representada. En el caso de los patrones temporales representan zonas con un comportamiento criminal similar en cuanto a tipos de delitos y estos eventos están abiertos a un rango de tiempo, entre más clara la zona es, tendrá un comportamiento similar, entre más oscura este comportamiento es diferente. En la Figura 4.10 se muestra un ejemplo de lo descrito anteriormente.

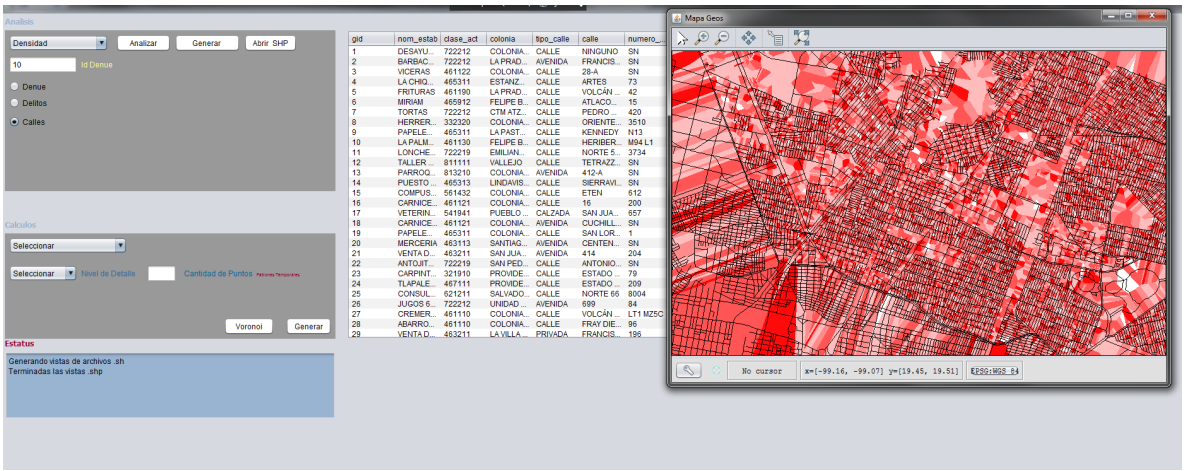


Figura 4.10 Representación visual de un mapa de densidad de delitos en el sistema.

El proceso de visualización involucra dos cosas principalmente, la primera con el botón de "generar" se realiza un proceso de consulta de la base de datos del postgres y se hace una transformación con la función de pgsql2shp del postgis, esta función convierte una tabla de atributos espaciales a un archivo .shp. Los archivos .shp son un estándar para la

visualización de datos en los sistemas de información geográficos y son estos mismos los que la herramienta de visualización del geotools puede abrir e interpretar.

La representación se hace por capas, estas capas dependen del orden de inserción en nivel en el que estarán, se deja por debajo la capa nueva de visualización de raster de delitos y por encima se agregan las capas de delitos, calles y DENUE.

Capítulo 5. Pruebas y resultados

5.1 Introducción

En este capítulo se presentan los resultados obtenidos, aplicando la metodología propuesta mediante la implementación del prototipo. Asimismo, se interpretan estos resultados a través de un caso de estudio para cada uno de los métodos propuestos en este trabajo de tesis.

Primeramente, se seleccionará cada uno de los tipos de análisis. De esta manera se procederá paso a paso para mostrar los resultados. Los datos en cuestión de análisis son 48,886 registros del DENUE, de los cuales se obtienen sus respectivos patrones, cruzando contra 152 puntos delictivos esparcidos por toda la delegación de la Gustavo A. Madero en la Ciudad de México.

Los 152 registros de delitos están divididos en 6 clases de interés, las cuales sirven para definir los patrones, siendo estos datos un supuesto conjunto representativo, no acorde en su totalidad a la situación actual de la Delegación.

5.2 Cálculo de densidades

Primeramente, se calculan los patrones correspondientes a este nivel de detalle, para esto, dentro del prototipo es necesario seleccionar la opción de “Cálculos”, donde se elige la opción deseada como se muestra en la Figura 5.1.

Figura 5.1 Opción para el cálculo de patrones.

Posteriormente, es necesario seleccionar los patrones de densidad, así como el nivel de detalle que en este caso será *micro*, después de esto, al dar click en la opción de "Generar", se procede a realizar el cálculo de los patrones, ya que se definió un radio de interés por omisión (al elegir el parámetro *micro*). Por tanto los patrones obtenidos se muestran en la Figura 5.2.

	clase integer	idDenue integer	patrones character(100)
1	812410	4333	3,10,9,5,0,4
2	811493	48317	2,3,5,5,5,3
3	434313	48318	14,19,11,15,18,16
4	461160	48319	11,12,7,8,10,5
5	461110	48320	10,12,8,8,11,5
6	811499	48321	12,21,11,16,16,16
7	463113	48322	14,20,9,13,14,16
8	722219	48323	14,21,12,13,16,16
9	464112	48324	1,1,1,1,2,3
10	465211	48325	3,3,6,4,5,2
11	811430	48326	12,17,9,11,14,14
12	812110	48327	10,19,10,15,12,17
13	463211	48328	11,17,9,11,15,15
14	931410	48329	3,7,6,6,4,3
15	811115	48330	1,4,4,2,0,2
16	465311	48331	3,9,7,3,5,5
17	332710	48332	1,2,3,2,2,3
18	311830	48333	12,20,11,16,16,16
19	464111	48334	14,20,9,13,14,16
20	461122	48335	14,20,12,16,18,17
21	461170	48336	12,20,10,17,16,17
22	722212	48337	6,14,7,10,7,12
23	561432	48338	10,10,7,7,10,2
24	461110	48339	5,5,5,3,5,4
25	316211	48340	0,1,1,1,1,2
26	465311	48341	11,16,8,10,11,7
27	812110	48342	11,16,9,11,11,9
28	531311	48343	8,20,10,14,11,15
29	461110	48344	12,19,9,13,15,15

Figura 5.2 Patrones de densidad generados.

Hay dos posibilidades a seguir después de obtener los patrones para la parte del análisis; la primera es que se puede realizar un análisis general, en la cual todos los patrones se comparan con el patrón ideal de 0, y la segunda es seleccionar un punto de interés para generar un mapa de similitud de comportamiento.

5.2.1 Densidad total

Para continuar con los datos es necesario seleccionar o introducir un punto de interés (id Denue) y ejecutar el análisis de densidad, presionando el botón de analizar (ver Figura 5.3).

Figura 5.3 Datos de entrada para calcular el análisis de densidad de delitos.

El botón de "Analizar" invoca el método de comparación de patrones basados en la métrica de *City Block* , después de finalizar con un valor discreto, el cual representa el cálculo de la densidad, tal como se muestra en la Figura 5.4.

	idDenue integer	clase integer	nombre character(15	geom geometry	discretizador integer
1	37997	467111	TLAPALERIA	01010000001	45
2	38008	461190	EXPENDIO DE	01010000005	10
3	38010	811499	COMPOSTURAS	0101000000D	40
4	38020	722212	COCINA EL B	0101000000D	2
5	38022	722219	CARNICERIA	0101000000E	2
6	38024	465912	EL GRAN DET	01010000002	8
7	38036	461110	CASA FELIPE	01010000005	18
8	38044	461110	MISCELANEA	01010000007	13
9	38053	541330	DESPACHO DE	0101000000B	14
10	38055	463218	VENTA DE SO	01010000006	15
11	38057	624412	RES DE ESTA	0101000000F	13
12	38067	622112	HOSPITAL GE	01010000000	17
13	38080	467111	REFACCIONES	01010000002	18
14	38091	722212	COCINA ECON	01010000002	2
15	38093	461110	MISCELANEA	01010000004	9
16	38100	811121	TALLER DE R	0101000000C	5
17	38113	465914	VENTA DE AR	01010000002	7
18	38115	532291	ALQUILER DE	01010000000	19
19	38123	463212	ROPA PARA B	0101000000C	18
20	38128	461110	MISCELANEA	01010000000	8
21	38130	325510	FERRO MEXIC	01010000001	12
22	38150	811111	TALLER MECA	01010000006	42
23	38151	713943	GIMNASIO	0101000000D	25
24	38153	468212	AUTOPARTES	0101000000C	28

Figura 5.4 Valores obtenidos al aplicar el cálculo con la métrica de City Block.

Posteriormente, presionando el botón “Generar” se obtiene la capa de datos raster que representa a los diagramas de Voronoi. Esta opción genera un archivo que corresponde a las zonas con mayor o menor índice de densidad de delitos. Para presentar esta información, los valores con color rojo más intenso reflejan una densidad delictiva mayor y los valores con rojo de menor intensidad reflejan las áreas con menor densidad. Asimismo, los shapefiles cargados sirven de entrada para el visualizador *swing* de GeoTools, con la finalidad de abrir el visualizador, con la opción de "Abrir SHP".

Por tanto, es posible seleccionar las capas de información complementarias correspondientes al DENUE, delitos y calles para su visualización y análisis. En esta prueba utilizamos las capas de DENUE y calles para observar la información obtenida.

En la Figura 5.5 se muestra la interfaz del sistema, presentando los resultados del mapa de densidad delictiva, el cual está compuesto por el diagrama de Voronoi y los datos vectoriales. Asimismo, se presenta en detalle información relativa a un delito en particular, el cual es representado por un punto.

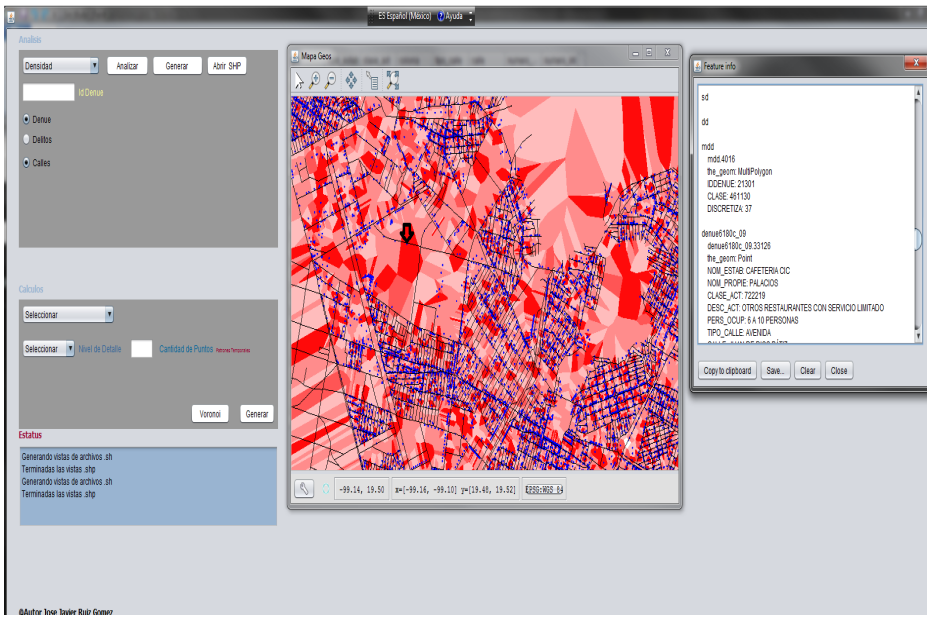


Figura 5.5 Resultado del análisis de densidad delictiva, mostrando detalles de la información con respecto a un foco delictivo representado por un punto

Como podemos observar, en la zona se ha seleccionado un punto de fuerte incidencia de delitos y podemos obtener su detalle, tal como se muestra a continuación en la Tabla 5.1.

Tabla 5.1 Datos relativos a un punto de incidencia delictiva referente al DENUE 21301.

```
Pos x=-99.1414 y=19.5043
msd

sd
dd

mdd
mdd.4016
the_geom: MultiPolygon
IDDENUE: 21301
    CLASE: 461130
    DISCRETIZA: 37

denue6180c_09
denue6180c_09.33126
the_geom: Point
NOM_ESTAB: CAFETERIA CIC
    NOM_PROPIE: PALACIOS
    CLASE_ACT: 722219
    DESC_ACT:    OTROS    RESTAURANTES    CON
SERVICIO LIMITADO
    PERS_OCUP: 6 A 10 PERSONAS
    TIPO_CALLE: AVENIDA
    CALLE: JUAN DE DIOS BÁTIZ
    NUMERO_EXT: SN
    EDIFICIO:
    NUMERO_INT:
    COLONIA: COLONIA ZACATENCO
    CENTRO_COM:
    NUM_LOCAL:
    COD_POSTAL: 7300
    CVE_EDO: 09
    ENTIDAD: DISTRITO FEDERAL
    CVE_MUN: 005
    MUNICIPIO: GUSTAVO A. MADERO
    CVE_LOC: 0001
    LOCALIDAD: GUSTAVO A. MADERO
    AGEB: 0718
    MANZANA: 003
    TELEFONO: 5519934264
    CORREOELEC:
    WWW:
    TIPO_ESTAB: FIJO
    ALTA: JULIO 2010
ESTATUS: 0
    LATITUD: 19.504496503
    LONGITUD: -99.14132856

calles
```

```
calles.1346
the_geom: MultiLineString
LENGTH: 1467.96028
    ID: 128536
    CLAVE: e14a29
    FC: 3165
    TIPO: Tercer orden
    FNODE: 2565
    LPOLY: 2
    RPOLY: 2
    TNODE: 2792
```

La información anteriormente descrita muestra información de todas las capas contenidas, Denue, valor discreto y calles de la cual solo nos interesa la parte marcada de los datos utilizaremos para comparar zonas.

5.2.2 Comparación de datos de densidades

Con la finalidad de comparar los datos de las densidades obtenidas, se hace uso de los mismos patrones, para obtener información significativa. Por tanto, se realiza un análisis del punto del punto DENUE **21301**, y se coloca este valor en el punto de **id** y se vuelve a ejecutar la opción de "Analizar"; con lo cual se genera nuevamente una capa de datos raster que muestra la información referente a la comparación de las densidades. El resultado obtenido se muestra en la Figura 5.6.

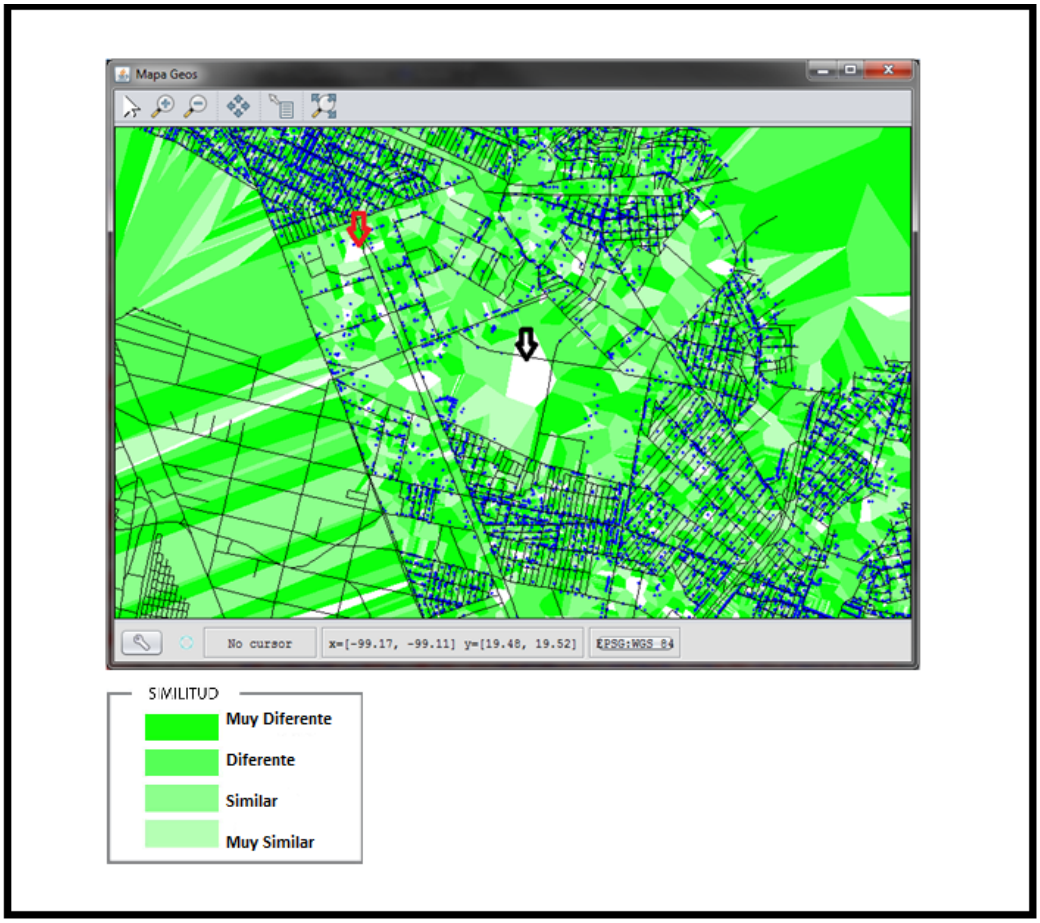


Figura 5.6 Resultado obtenido al analizar el id 21301.

A partir de la figura anterior, se puede observar que la zona de interés ahora está marcada de color blanco y señalada con una flecha en color negro. Entonces se puede apreciar otro punto en la zona de color blanco, señalada en la Figura 5.6 con una flecha en color rojo. Este punto es otro punto de interés; ya que ahora la interpretación que se le da al análisis es de similitud de características; lo cual indica que entre más claro sea el color, mayor similitud presentan las zonas conforme a su actividad delictiva y su densidad de delitos (en este caso las tonalidades inician con un verde más intenso hacia uno más suave).

En la Tabla 5.2 se presenta la información referente al punto señalado con la flecha en color rojo.

Tabla 5.2 Datos relativos a un punto de incidencia delictiva referente al DENUE 3229.

Pos x=-99.1552 y=19.5137
msd
sd

```
dd

mdd

mdd.4016
the_geom: MultiPolygon
IDDENUE: 3229
    CLASE: 326198
    DISCRETIZA: 0
denue6180c_09
denue6180c_09.3329
the_geom: Point
    NOM_ESTAB: BULK MOLDING COMPOUNDS MEXICO
NOM_PROPIE: BULK MOLDING COMPOUNDS MEXICO SA DE CV
    CLASE_ACT: 326198
    DESC_ACT: FABRICACION DE OTROS PRODUCTOS DE PLASTICO CON REFORZAMIENTO
    PERS_OCUP: 11 A 30 PERSONAS
    TIPO_CALLE: EJE VIAL
    CALLE: CENTRAL (LÁZARO CÁRDENAS)
    NUMERO_EXT: 802
    EDIFICIO:
    NUMERO_INT:
    COLONIA: NUEVA INDUSTRIAL VALLEJO
    CENTRO_COM:
    NUM_LOCAL:
    COD_POSTAL: 7720
    CVE_EDO: 09
    ENTIDAD: DISTRITO FEDERAL
    CVE_MUN: 005
    MUNICIPIO: GUSTAVO A. MADERO
    CVE_LOC: 0001
    LOCALIDAD: GUSTAVO A. MADERO
    AGEB: 0489
    MANZANA: 002
    TELEFONO: 51190203
    CORREOELEC:
    WWW:
    TIPO_ESTAB: FIJO
    ALTA: JULIO 2010
ESTATUS: 0
    LATITUD: 19.513936376
    LONGITUD: -99.15494835

denue6180c_09.39271
the_geom: Point
NOM_ESTAB: KOBREX SA DE CV
    NOM_PROPIE: KOBREX SA DE CV
    CLASE_ACT: 434225
    DESC_ACT: COMERCIO AL POR MAYOR DE EQUIPO Y MATERIAL ELECTRICO
    PERS_OCUP: 11 A 30 PERSONAS
    TIPO_CALLE:
    CALLE: CALZADA VALLEJO
    NUMERO_EXT: 1361
    EDIFICIO:
```

```
    NUMERO_INT: B
    COLONIA: COLONIA NUEVA INDUSTRIAL VALLEJO
    CENTRO_COM:
    NUM_LOCAL: 1361
    COD_POSTAL: 7700
    CVE_EDO: 09
    ENTIDAD: DISTRITO FEDERAL
    CVE_MUN: 005
    MUNICIPIO: GUSTAVO A. MADERO
    CVE_LOC: 0001
    LOCALIDAD: GUSTAVO A. MADERO
    AGEB: 0489
    MANZANA: 002
    TELEFONO: 5550398800
    CORREOELEC: jvallejo@kobrex.com
    WWW: WWW.KOBREX.COM
    TIPO_ESTAB:
    ALTA: JULIO 2010
    ESTATUS: 1
    LATITUD: 19.513653310
    LONGITUD: -99.15522944

calle
```

En este caso, se aprecia que este valor es 0, ya que éste denota la *similitud máxima*, lo cual se puede interpretar que entre más próximo es el valor a 0, mayor semejanza presentan las áreas definidas por los diagramas de Voronoi.

En la Figura 5.7 se aprecia la comparativa entre los resultados obtenidos para el punto id=21301 (derecha) y el punto id=3229 (izquierda). En ambos casos los resultados son muy similares; sin embargo, dependiendo del ranking de valores los puntos de similitud varían.

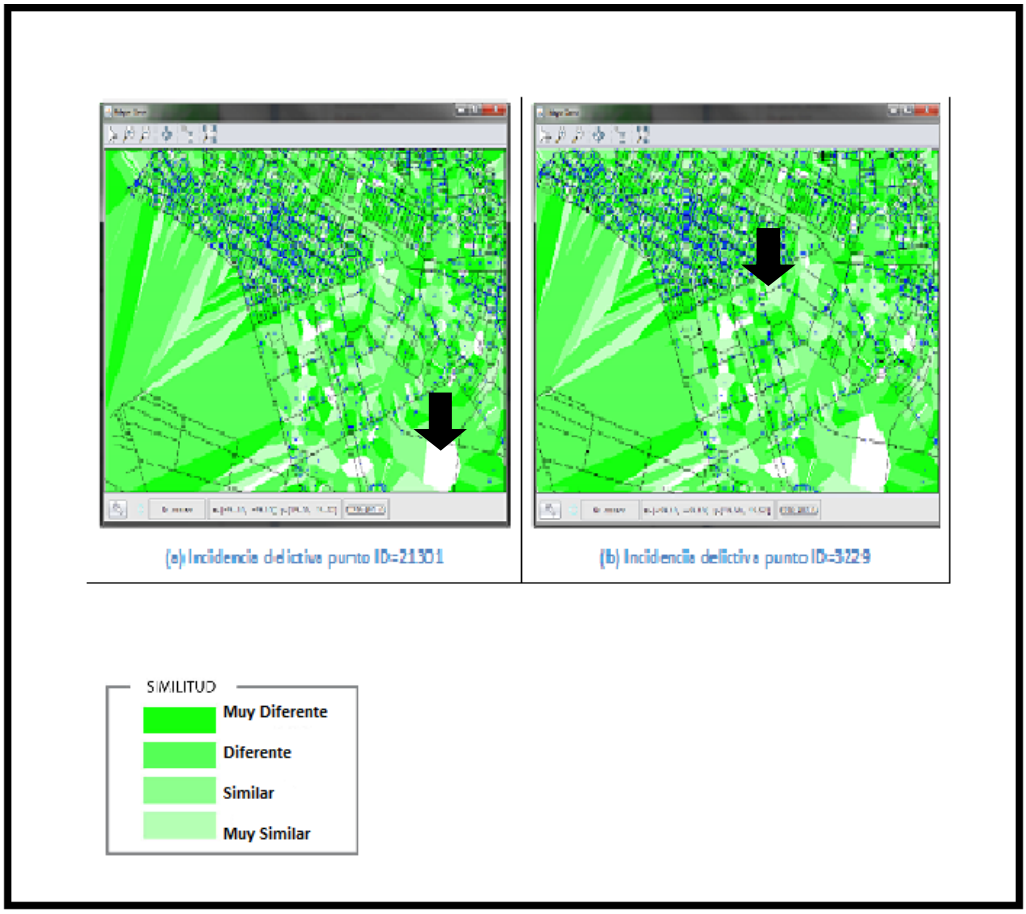
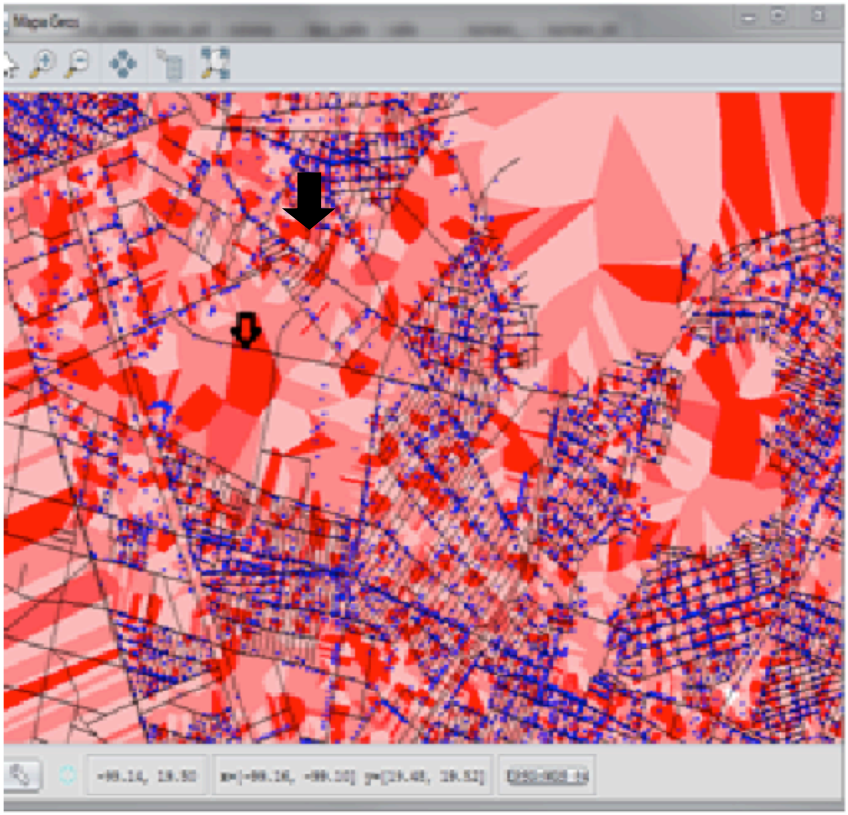


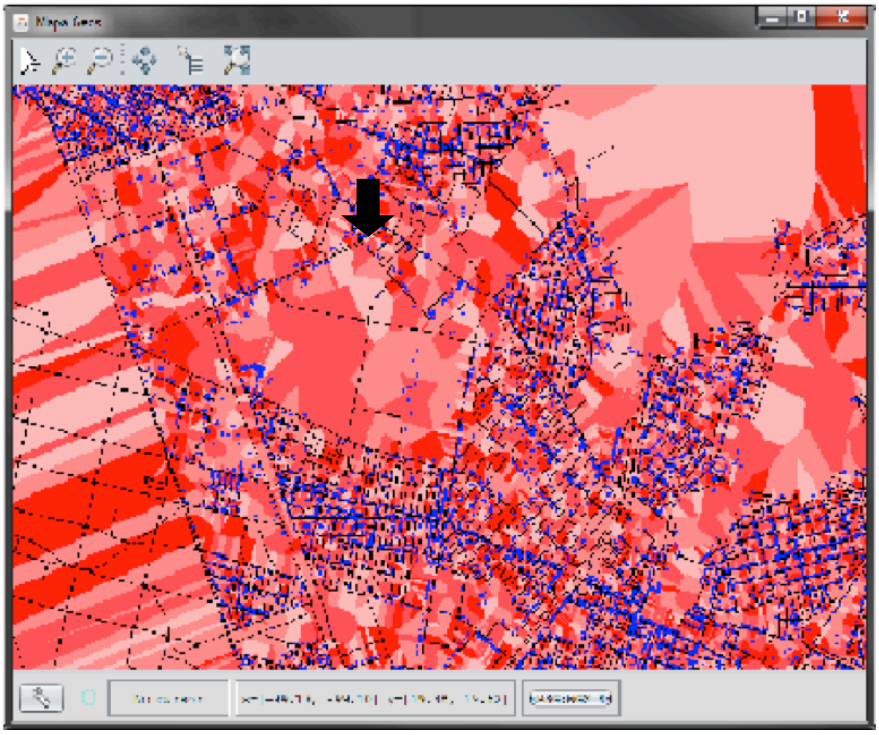
Figura 5.7 Análisis comparativo entre los puntos que describen incidencia delictiva obtenidos por el sistema.

5.2.3 Comparativa de 3 Niveles de Análisis

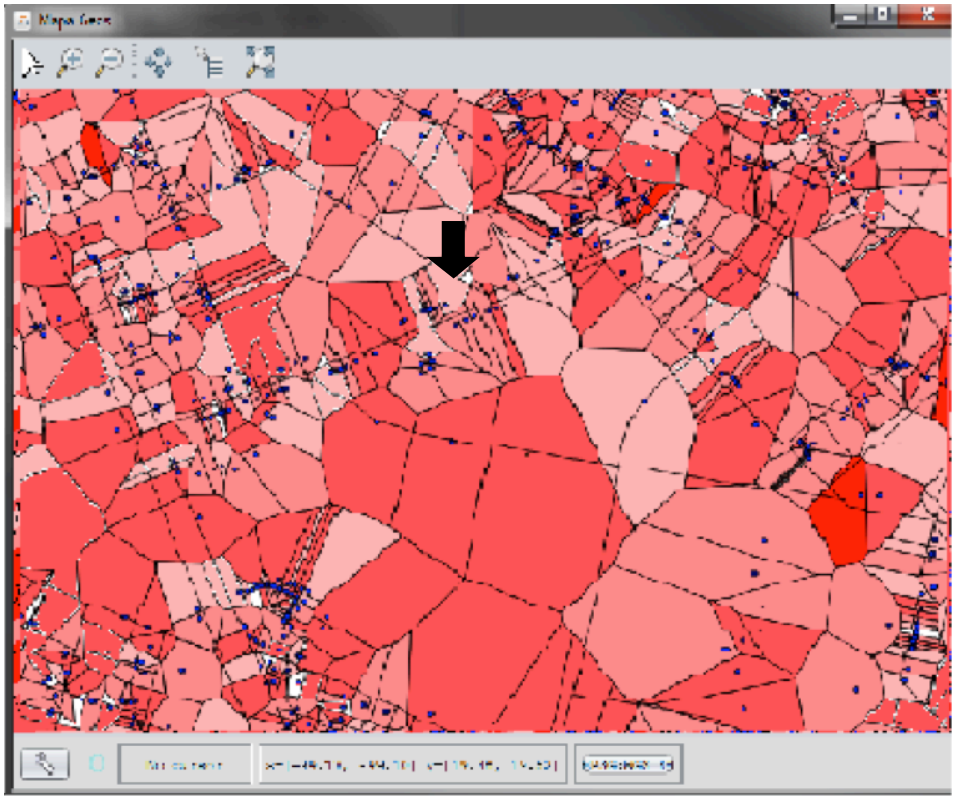
En la Figura 5.8 se aprecian los 3 niveles de detalle que son posibles realizar, en este caso veremos los resultados a nivel micro, *meso* y *macro;* a diferencia delo anterior en donde se presentaron los resultados referentes a la densidad total.



(a) Nivel *micro*



(b) Nivel *meso*



(c) Nivel *macro*

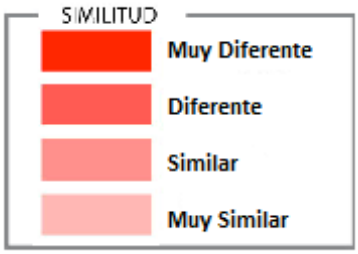


Figura 5.8 Comparación de análisis a tres niveles de detalle.

De acuerdo con la Figura 5.8, se puede apreciar que en (a) se tiene el resultado del análisis a un nivel *micro*, en (b) a un nivel *meso* y en (c) a un nivel *macro*. Estos niveles van de mayor a menor detalle y podemos concluir que los valores de las regiones tienden a "estabilizarse". Asimismo, zonas que ya habían sido evaluadas en cierta categoría pueden cambiar su valor.

Este cambio se debe a que al agregar más zonas de cálculo, los puntos que estén próximos a las fronteras de otros puntos, adquirirán un valor mayor por tener más zonas de análisis, Otra causa es que el rango de valores que se toma, se obtiene a partir del valor máximo y mínimo de los puntos de interés, si todos los puntos aumentan su valor también, los

rangos podrán aumentar o disminuir en cantidades, generando un decremento o incremento de los rangos dependiendo el caso de estudio.

De esta variación de rangos de distancia (nivel de detalle), podemos observar que esta "estabilización de valores" no es más que el cambio de la media estadística en todo nuestro universo de estudio, al cambiar la media, los valores generalmente cambian todo el panorama del mapa, ya que consecuentemente de aumentar el rango, aumentamos la cantidad de puntos que se analizan por patrón, esto genera este aumento (o disminución) del valor promedio de la densidad.

5.3 Patrones temporales

Los patrones temporales consideran la fecha de ocurrencia en su ordenamiento, este tipo de análisis a diferencia del obtenido por densidades, forzosamente necesita un punto de interés para trazar un mapa del punto de interés contra el resto de puntos que forman el universo de discurso.

La razón de esto es en la métrica de Leveinstein, definida en la (Ecuación 3.1) y el patrón que se forma, ya que si comparamos con la cadena vacía por ejemplo, los resultados arrojados por el resto de los puntos daría una cantidad igual, ya que mide los cambios necesarios para que una cadena sea convertida en otra cadena, y debido a que el tamaño de la cadena es fijo, este análisis no es descriptivo a menos que se compare el universo con un punto de interés.

Para poder calcular un patrón temporal tenemos que indicar la cantidad de puntos máximos que deben ser considerados en la cadena. Para ello, utilizamos un campo donde se proporciona la cantidad de puntos máxima (ver Figura 5.9).

Figura 5.9 Datos de entrada para el cálculo de patrones temporales con *n*=10

De acuerdo con la Figura 5.9, se elige utilizar una *n* máxima de 10, al presionar el botón "Generar", la aplicación desarrolla un buffer al nivel de detalle seleccionado (en este caso nivel *micro*), obteniendo los puntos de interés y ordenándolos cronológicamente de manera ascendente para obtener las ocurrencias. Por tanto, para construir la cadena se coloca el tipo de delito al que ese evento delictivo corresponde. En la Figura 5.10 se muestra la tabla generada por el sistema, mostrando la carga de la cadena, ya con el ordenamiento de los patrones.

	clase integer	idDenue integer	patrones character var
1	461170	810	1564436162
2	461110	811	1461142145
3	722212	812	6443623126
4	811121	813	1564436622
5	811111	814	1654461622
6	463310	815	1564436223
7	463310	816	1544612234
8	465311	817	1656464361
9	722212	818	6443623126
10	811430	819	6433126441
11	461110	820	5644362312
12	722212	821	1465464616
13	931210	822	1564436622
14	561432	823	1544612234
15	811121	824	1564436223
16	461110	825	1461142145
17	722212	826	1656443616
18	465311	827	1462453464
19	812210	828	6433126441
20	621211	829	1564436162
21	811114	830	1564436223
22	611132	831	1461142145
23	463211	832	1465634643
24	722219	833	1656443616
25	465311	834	1461142145
26	311830	835	1461245634
27	722212	836	6443623126
28	722110	837	6331644263
29	465311	838	6443623126

Figura 5.10 Patrones temporales generados.

Continuando con el ejemplo anterior, utilizamos el mismo valor de inicio del punto del DENUE con el id = 21301 y comenzamos con la ponderación basada en la métrica de Levenshtein (ver Figura 5.11).

Figura 5.11 Datos de entrada para generar el análisis temporal.

Al generar el análisis, las ponderaciones se van colocando en una tabla intermedia, justo como en el caso de los patrones de densidad, los resultados se presentan en la Figura 5.12.

	idDenue integer	clase integer	nombre character(15	geom geometry	discretizador integer
1	6299	722212	COCINA LA T	01010000005	7
2	6302	461110	ABARROTES J	01010000007	8
3	6304	811112	SERVICIO EL	0101000000E	8
4	6306	311830	TORTILLERIA	01010000006	8
5	6312	463310	ZAPATERIA	0101000000A	6
6	6313	811420	REPARADORA	01010000007	0
7	6314	461110	ABARROTES L	01010000004	0
8	6322	461130	VERDULERIA	01010000003	8
9	6328	722212	COMIDA RAPI	0101000000F	0
10	6336	465311	PAPELERIA L	01010000001	8
11	6345	461110	TIENDA	0101000000B	8
12	6350	811313	INGENIERIA	01010000004	8
13	6353	464111	FARMACIA SA	01010000009	8
14	6360	311830	TORTILLERIA	0101000000A	7
15	6361	311812	PANIFICADOR	01010000009	8
16	6364	434311	RECUPERADOR	0101000000C	9
17	6372	561431	PAPELERIA C	0101000000F	8
18	6374	621211	CONSULTORIO	0101000000D	8
19	6381	468211	MOFLES MADE	01010000007	0
20	6383	461110	TIENDA	01010000007	0
21	6392	722212	JUGOS	01010000002	8
22	6393	468420	CAMBIO DE A	0101000000F	6
23	6398	461170	PALETAS DE	01010000004	4
24	6404	466212	TELEFONIA C	01010000001	7
25	6405	461110	VENTA DE DU	01010000000	0
26	6411	811211	EL CROW	0101000000A	0
27	6418	811491	CERRAJERIA	01010000002	6
28	6420	434225	DISTRIBUCIO	0101000000F	6
29	6433	532493	INTERNATION	01010000008	8
30	6443	611112	JARDIN DE N	01010000002	8
31	6447	461130	FRUTAS	01010000004	8
32	6451	811119	CLUTCH Y FR	01010000004	8

Figura 5.12 Valores de la comparación para un análisis temporal a nivel micro (DENUE id = 21301).

De acuerdo con los valores mostrados en la Figura 5.12, se puede apreciar que los valores son muy diferentes a los generados por la otra medición; esto tiene mucho sentido, ya que el valor máximo diferencial entre dos patrones es 10, debido al tamaño máximo de la cadena y lo que mide la métrica que son los cambios necesarios para que una cadena sea igual a otra.

El tamaño mínimo es de 0, considerando en este caso que los parámetros son iguales. Tomando en consideración lo anterior, en la Figura 5.13 se obtienen los siguientes resultados:

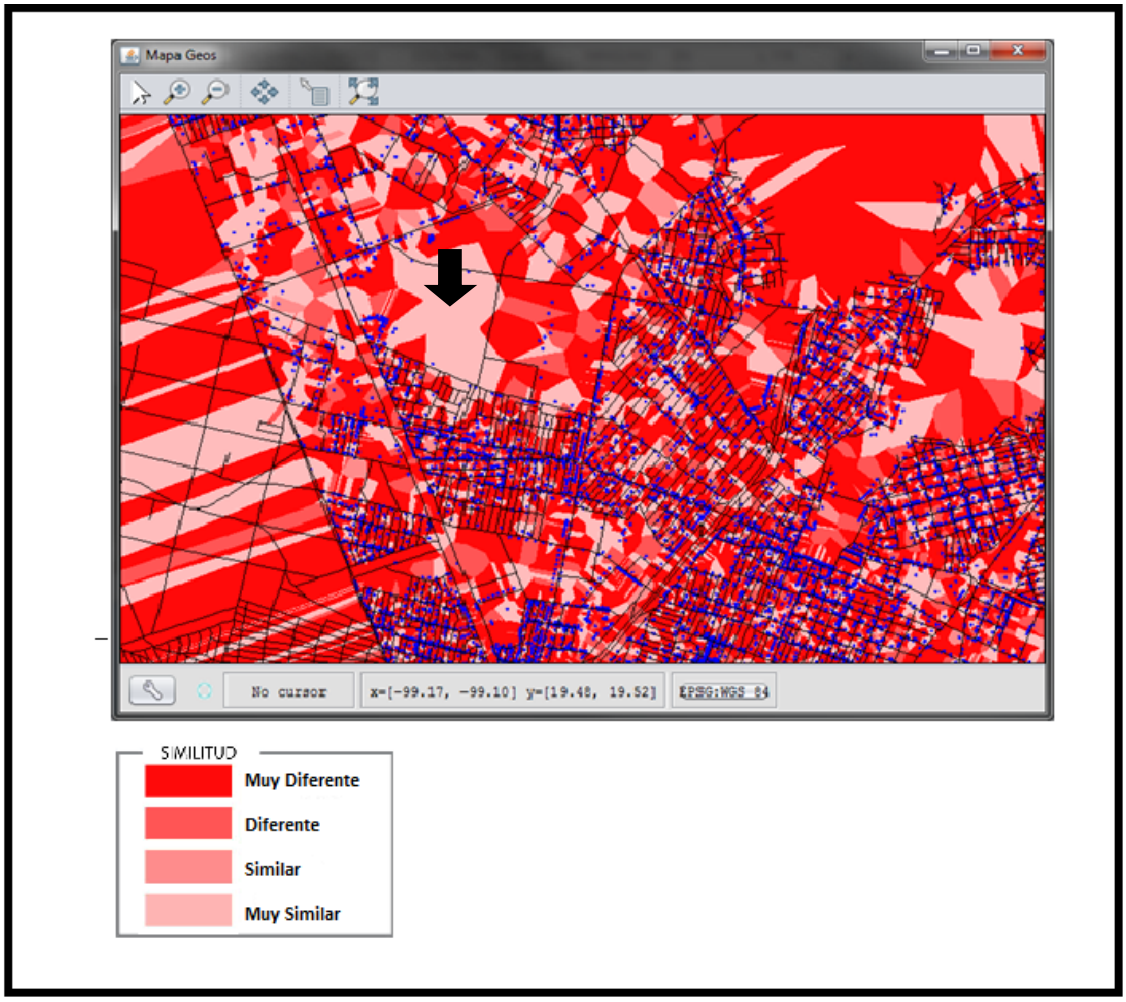


Figura 5.13 mapa de similitud temporal a nivel micro para (DENUE id = 21301).

El análisis se interpreta de la siguiente manera, los datos que presentan mayor similitud se muestran de manera más clara, los que tienen mayor diferencia presentan un tono más oscuro en la degradación.

Asimismo, lo primero que se nota en la información que está siendo generada, es que se tienen zonas más definidas, esto se debe a que el rango es menor y esto ocasiona que la información entre zonas sea más óptima.

Lo segundo corresponde a la cantidad de puntos similares; este patrón además de depender de la distancia entre puntos, depende de su temporalidad al estar acomodados por lo que eventos que igual tengan una distancia un poco mayor, por fecha de ocurrencia quedan adelante de la cadena, los eventos ocurridos fuera del rango se descartan, por lo que zonas que comparten dichos eventos, seguramente tendrán en sus cabeceras de cadena, la información que ocurre primero hasta la última.

5.3.1 Variación de rangos de valores

La variación de rangos se presenta cuando la tendencia que se presenta ocurre que a mayor rango para la obtención de los datos, se eliminan categorías, tal como se muestra en la Figura 5.14.

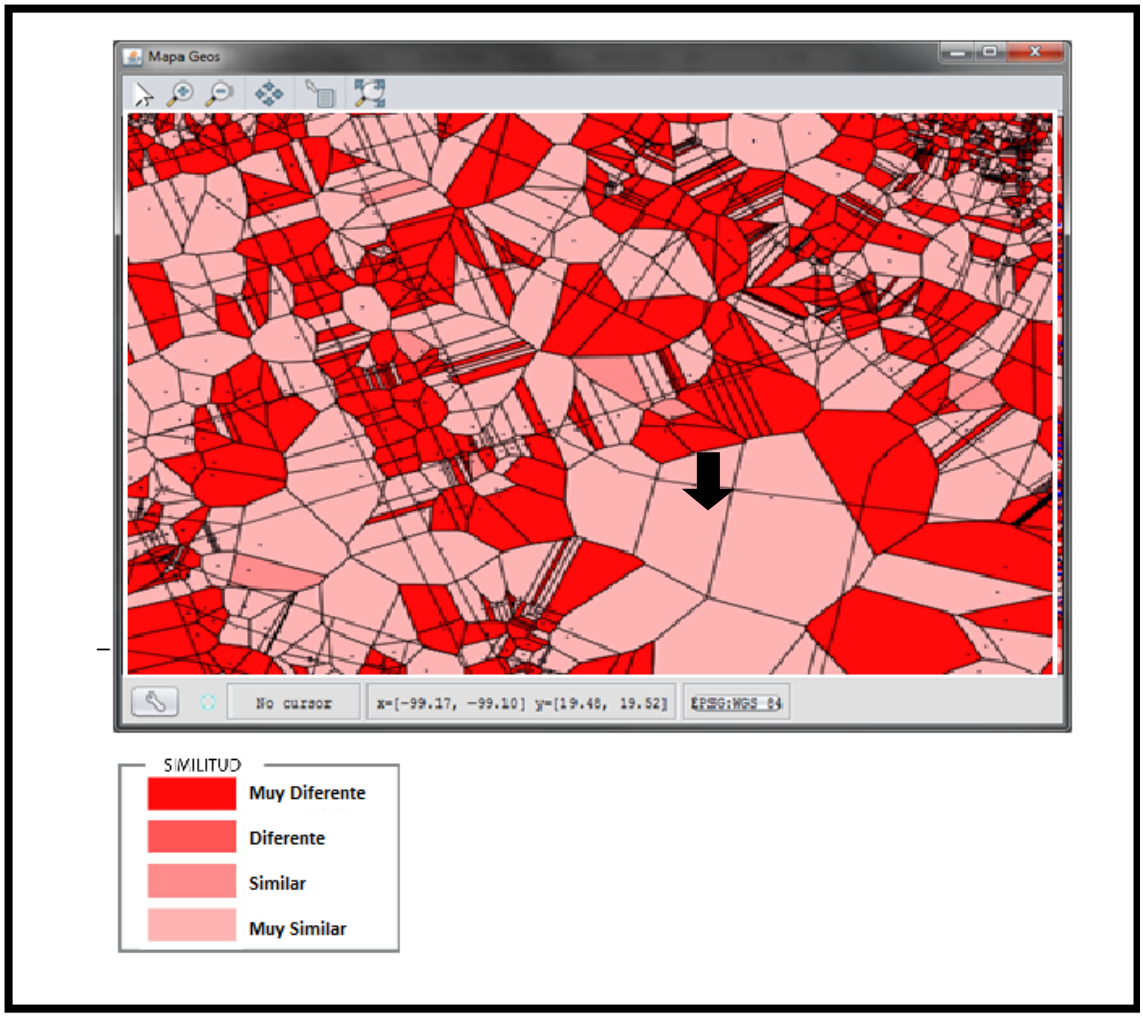


Figura 5.14 Mapa de similitud temporal a nivel meso para (DENUE id = 21301).

De acuerdo con la Figura 5.14, podemos observar que los rangos que predominan son o muy parecidos al punto de interés o muy diferentes. Esto se debe a que al tomar el mayor rango predominan los valores que tienen una ocurrencia temporal similar.

Otro factor a considerar en este fenómeno, es la cantidad total de puntos a ser tomados en cuenta en el arreglo, entre mayor sea éste, los rangos tienden a aumentar. En la Figura 5.15 se presenta un aumento al valor inicial de *n*, el cual va de 10 a 20, lo cual se aprecia en la figura.

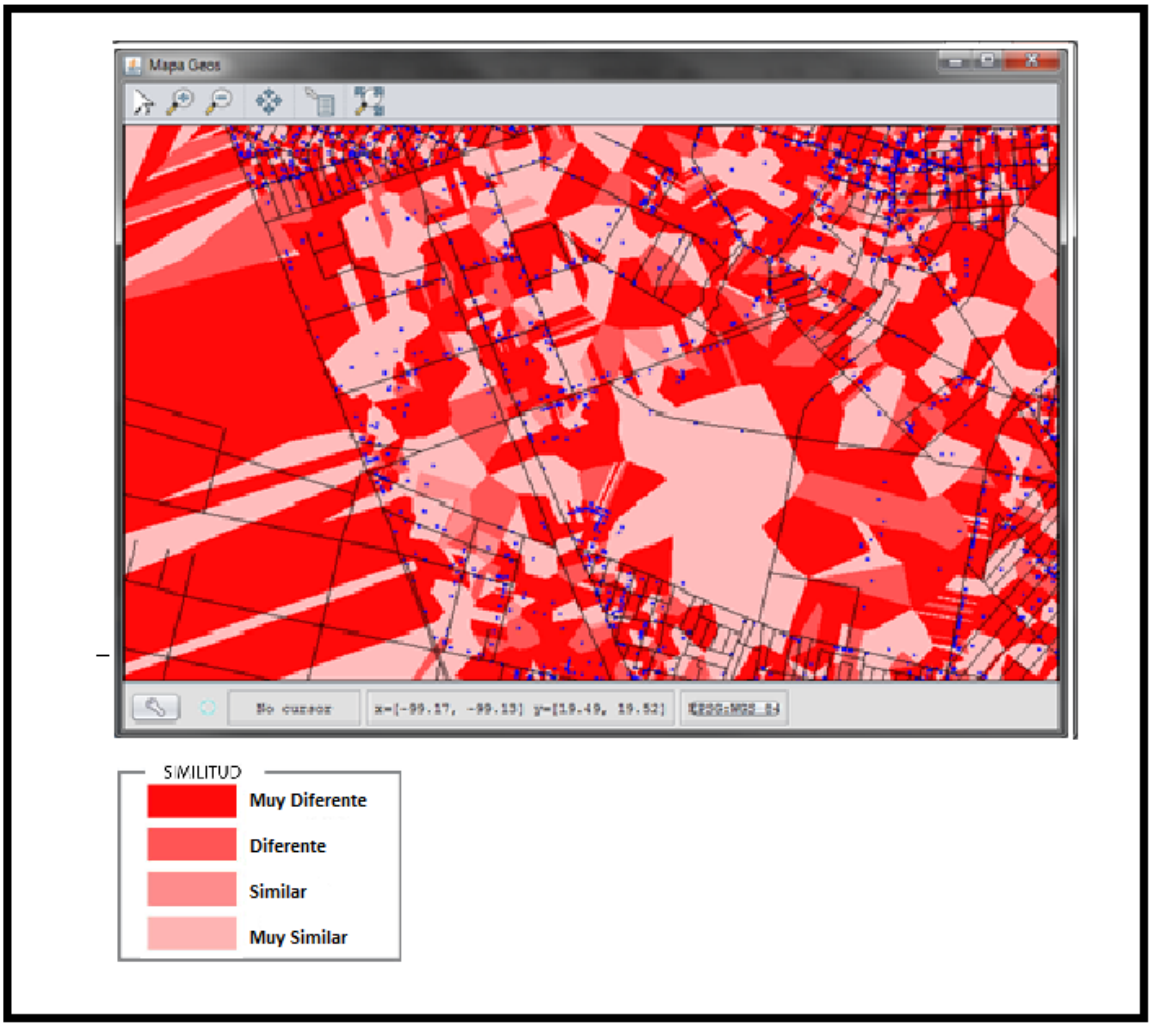


Figura 5.15 Mapa de similitud temporal a nivel meso para (DENUE id = 21301), con un patrón de n = 20.

Finalmente, se puede observar que el rango se amplía y se tienen nuevamente zonas parecidas al análisis anterior, a un nivel menor y esto se debe al ordenamiento de los datos temporales. Lo anterior, se muestra en la Figura 5.16.

Figura 5.16 Agregación de puntos a los análisis cuando aumenta el nivel de detalle

Al ampliar el rango como vemos, ampliamos el área de análisis y estos valores se le suman a la densidad que ya tenía el punto, esto provoca que se aumente su valor y que los rangos entre regiones se reduzcan.

Capítulo 6. Conclusiones y trabajos futuros

6.1 Introducción

Con base en el trabajo desarrollado en la presente tesis, existen algunas conclusiones importantes que es necesario destacar. Estas conclusiones se enfocan directamente al contexto delictivo estudiado en este trabajo.

En este sentido, el fenómeno delictivo es complejo y requiere de un profundo análisis por parte de especialistas en este ámbito, bajo el contexto social, ya que este fenómeno es continuo y presenta comportamientos muy diversificados.

Asimismo, este fenómeno varía conforme al tiempo y lugar, ya que se presentan relaciones entre las diferentes entidades involucradas en este contexto. De igual forma, podemos concluir que mientras más elevado sea el rango o área de estudio, el análisis que se puede obtener puede ser más real.

Esto se debe que en comparativa, las zonas pequeñas, presentan índices delictivos mayores, con respecto al análisis en zonas más grandes. Sin embargo, al realizar el análisis a un nivel *meso*, los rangos de las zonas cambian y los valores tienden a estabilizarse.

De igual forma, es necesario resaltar que los estudios a los tres niveles de detalle propuestos: *micro*, *meso* y *macro*; corresponden a diversos universos de interés y con base en el enfoque propuesto en la metodología, se pueden apreciar diversas tendencias con respecto a los índices delictivos.

Por otra parte, es importante señalar que los objetivos trazados para esta tesis fueron cumplidos, ya que se ha diseñado e implementado un Sistema de Información Geográfica con la finalidad de brindar un soporte en las tareas relacionadas con el análisis delictivo, empleando técnicas de análisis espacial y reconocimiento de patrones.

Asimismo se diseñó una aplicación SIG de escritorio para mostrar los resultados de un análisis delictivo, implementando una base de datos geográfica, basada en una representación conceptual dentro del dominio de aplicación de la incidencia delictiva.

Bajo este contexto, se han generado vectores de patrones de densidad delictiva y cadenas de secuencia de densidad temporal, con el objetivo de evaluar los patrones y generar un mapa comparativo de los puntos referentes a la actividad criminal.

6.2 Aportaciones

Como principales aportaciones podemos citar las siguientes:

- Diseño de una metodología para el análisis delictivo en tres niveles: *micro*, *meso* y *macro*.

- Implementación de un sistema de información geográfica de escritorio para la visualización de mapas raster delictivos.

- Implementación de métricas para la comparación de resultados: métrica para el cálculo de la densidad delictiva y métrica para el cálculo de incidencias temporales.

- Diseño e implementación de métodos para la clasificación y reconocimiento de patrones, así como de análisis de características con base en el método de diagramas de Voronoi y la distancia de Levenshtein.

- Diseño e integración de diversas herramientas para el almacenamiento y procesamiento de los datos: PostGIS y GeoTools.

- Generación de información isométrica para la representación de dos capas de información espacial y su interpolación.

6.3 Limitaciones

Las principales limitaciones presentadas en este trabajo de tesis son las siguientes:

- El conjunto de técnicas propuesto, únicamente fueron probadas para calcular y cruzar dos capas de datos representadas como puntos.

- Es necesario generar archivos de apoyo para la visualización, lo cual implica que el tiempo de ejecución puede incrementarse sustancialmente con respecto al tamaño del conjunto de datos.

- El acceso a la información representa una limitación debido a su naturaleza, por considerarse como datos sensibles al contexto.

- El sistema no puede ser compartido y funciona básicamente como un sistema de escritorio, con lo cual no pueden ser visualizados los resultados en otras plataformas.
- El caso de estudio está limitado a una delegación, lo cual en el punto de las fronteras el análisis no puede considerarse del todo asertivo.
-

6.4 Trabajos Futuros

De acuerdo con la experiencia adquirida en el desarrollo de esta tesis, se proponen los siguientes trabajos futuros:

- Incluir el conjunto completo de datos, para ello sería necesario generar estos datos y adecuarlos a la base de datos para la aplicación del análisis.

- Optimizar los algoritmos de generación de capas de datos para que el sistema pueda trabajar más rápido.

- Trabajar en la información obtenida de zonas de interés y vincular con zonas de orígenes de criminales y víctimas para observar el factor de migración de estas actividades.

- Diseñar un perfil de tipo de criminal por la zona donde opera o por su lugar de origen, esto corresponde a la información vincula da en el modelo que no fue utilizada.

- Migrar el sistema a una plataforma web para que pueda ser utilizada de manera remota y sea más sencilla el ingreso de datos.

- Ampliar la zona de estudio para analizar los comportamientos similares entre delegaciones para buscar un comportamiento más amplio e implementar métodos para la tendencia de análisis.

- Utilizar otros campos de conocimiento para aplicar el estudio de densidades a otras disciplinas para ver el comportamiento de los fenómenos que involucren densidades de eventos y secuencias de ocurrencia de los mismos.

Referencias

1. M.F. Goodchild, B. Buttenfield & J. Wood. (1994). *"Introduction to Visualizing Data Validity"*, In: H.M. Hearnshaw & D.J. Unwin (Eds.), Visualization in Geographical Information Systems, pp. 141-149.

2. A.M. MacEachren. (1994). *"Some Truth with Maps: A Primer on Design & Symbolization"*, Washington D. C., Association of American Geographers, pp. 234-257.

3. D. DiBiase. (1990). *"Visualization in the Earth Sciences"*, Earth and Mineral Sciences Bulletin, Vol. 59, No. 2, pp. 13-18.

4. R. Tomlinson. (2007). *"Thinking about GIS: Geographic Information System Planning for Managers"*, ESRI Press, 3rd Edition, Redmond, CA, USA, pp. 34-47.

5. L. Vozmediano & S.J. Guillen. (2006). *"Empleo de Sistemas de Información Geográfica en el Estudio del Miedo al Delito"*, Revista Española de Investigación Criminológica (REIC), No. 4, pp. 118-132.

6. F. Valenga, E. Fernández & H. Merlino, (2008). *"A Portlet-Based Service-Oriented Architecture for a Second Generation Portal"*, In: Proceedings of the VII Jornadas Iberoamericanas de Ingeniería del Software e Ingeniería del Conocimiento", Guayaquil Ecuador, pp. 121-128.

7. M.S. Fritz & D.P. MacKinnon. (2007). *"Required sample size to detect the mediated effect*", Journal of Psychological Science, No. 18, pp. 233-239.

8. M.F. Goodchild (2009), *"Geographic Information Systems and Science: Today and Tomorrow"*, The Earth and Planetary Science, Proceedings of the 6th International Conference on Mining Science and Technology, Vol. 1, pp. 1037–1043.

9. P.J. Brantingham & P.L. Cohmen. (1981), *"Environmental Criminology"*, Waveland Press, 1st Edition, New York, USA, pp. 137-146.

10. V. Garrido & P. López. (2006). *"El rastro del asesino: El perfil psicológico de los criminales en la investigación policial"*, Universidad de Barcelona, pp. 45-47.

11. J.E. Eck, S. Chainey, J.G. Cameron, M. Leitner & R.E. Wilson. (2005). *"Mapping Crime: Understanding Hotspots"*, U.S. Department of Justice, National Institute of Justice Washington, D.C, USA, pp. 1-27.

12. P.N. Grabosky, R.G. Smith & G. Dempsey. (2001). *"Electronic theft: unlawful acquisition in cyberspace"*, Cambridge University Press, Boston, MA, USA, pp. 101-114.

13. A. Glimour & J. Barclay. (2008). *"Developing Geographical Information Systems and Crime Mapping Tools in New Zealand"*, Crime Mapping Case Studies. Practice and Research. West Sussex, pp. 204-210.

14. M. Warr. (1984). *"Fear of victimization: Why are women and the elderly more afraid",* Journal of Social Science Quaterly, No. 65, pp. 681-702.

15. W. Skogan & M. Maxfield. (1981). *"Coping With Crime"*, Beverly Hills, CA, USA, Sage Publications, pp. 45-49.

16. D.J. Evans & M. Fletcher. (2000). *"Fear of crime: testing alternative hypotheses"*, Journal of Applied Geography, No. 20, pp. 395-411.

17. R. LaGrange & L. Ferraro. (1992). *"Perceived Risk and Fear of Crime: Role of Social and Physical Incivilities"*, Journal of Research in Crime and Delinquency, Vol. 29, No. 3, pp. 311–334.

18. B. Fernández-Ramírez & J.A. Corraliza. (1998). *"Generalidad y especificidad en la explicación del miedo al delito"*, Apuntes de Psicología, Vol. 16, No. 1-2, pp. 173-186.

19. C. San Juan. (2000). *"Theories of design and designs of theory in the environmental interventions"*, Bulletin of People Environment Studies, Vol. 15, pp. 15-16.

20. J. Wilson & G. Kelling. (1982). *"The police and neighborhood safety: Broken windows"*, The Atlantic Monthly", No. 249, pp. 29-38.

21. D. Lewis & M. Maxfield. (1980). *"Fear in the Neighbourhoods: An Investigation of the Impact of Crime"*, Journal of Research in Crime and Delinquency, No. 17, pp. 160-89.

22. P. Brantingham & L. Brantingham. (1991). *"Environmental Criminology. Prospect Heights"*, Waveland Press, Illinois, USA, pp. 23-28.

23. J. Eck & D. Weisburd. (1995). *"Crime Places in Crime Theory"*, First Chapter in Crime and Place, Monsey, Criminal Justice Press, 55-58.

24. L.W. Sherman. (1995). *"Hotspots of Crime and Criminal Careers of Places"*, Technical Report, University of Maryland, College Park, Washington, D.C., USA, pp. 35-52.

25. J.P. Gottlieb, M.G. MacAvoy & C.J. Bruce. (1994). *"Neural responses related to smooth-pursuit eye movements and their correspondence with electrically elicited smooth eye movements in the primate frontal eye field"*, Journal of Neurophysiol, Vol. 72, No. 4, pp. 1634–1653.

26. G. Orueta. (1995). *"Prácticas de Análisis Espacial"*, Reporte Técnico, Universidad de Barcelona, Oikos-Tau, pp. 384-385.

27. W. Kerfoot & M. Lynch. (1987). *"Branchiopod Communities: Associations with Planktivorous Fish in Space and Time"*, Predation, University Press New England, Hanover, NH, pp. 367-378.

28. O. Newman. (1972). *"Defensible space. Crime prevention through urban design"*, London, MacMillan Press, pp. 89-93.

29. D. Dzemydiene & E. Kazemikaitiene. (2005). *"Development of the Component-Based Knowledge Management System for Assistance in Crime Investigation Processes"*,

30. T. Gruber & R. Toward. (1991). *"Principles for the Design of Ontologies Used for Knowledge Sharing",* In: N. Guarino & R. Poli (Eds.), Proceedings of the Conference on Formal Ontology in Conceptual Analysis and Knowledge Representation, Trento, Italy, IOS Press, pp.119-131.

31. R. Neches, R. Fikes, T. Finin, T. Gruber, R. Patil, T. Senator & W.R. Swartout. (1991). *"Enabling Technology for Knowledge Sharing"*, AI Magazine, Vol. 12, No. 3, pp. 1-36.

32. W.N. Borst. (1997). *"Construction of Engineering Ontologies for Knowledge Sharing and Reuse",* Universiteit of Twente, Enschede, pp. 88-92.

33. R. Studer. (1998). *"Knowledge Engineering Principles and Methods"*. Journal of Data & Knowledge Engineering, Vol. 25, No. 1, pp. 161-197.

34. G. Barchini, M. Álvarez, S. Herrera & M. Trejo. (2007). *"El rol de las ontologías en los SI"*, Revista de la Facultad de Ciencias Exactas y Tecnologías, Universidad Nacional de Santiago del Estero, No. 14, pp. 1-5.

35. K. Fox & M. Gruninger. (1998). *"Ontologies to Support Process Integration in Enterprise Engineering"*, Journal of Computational and Mathematical Organization Theory, Vol. 6, No. 4, pp. 381-394.

36. D.M. Tsou & P.C. Fisher. (1982). *"Decomposition of a relation scheme into Boyce-Codd Normal Form"*, Journal of AC SIGACT News, Vol. 14, No. 3, pp. 23-29.

37. L. Yujian & L. Bo. (2007). *"A Normalized Levenshtein Distance Metric"*, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 29, No. 6, pp. 1091-1095.

38. D. Cassidy & P.A. Quinn. (2007). *"How the Irish invented Slang: The Secret Language of the Crossroads"*, Counter Punch Press, pp. 90-92.

39. A. Coelho, C.A. Bramblett, L.B. Quick & S.S. Bramblett. (1976). *"Resource availability and population density in primates: a socio-bioenergetics analysis of the energy budgets of Guatemalan howler and spider monkeys"*, Journal of Primates, Springer-Verlag, Vol. 17, No.1, pp. 63-80.

40. M. De Berg, O. Cheong & M. Van Kreveld. (2008). *"Computational Geometry: Algorithms and Applications"*, Springer-Verlag, pp. 145-201.

41. Entidades federativas de México por Densidad de Población. (2013). Internet: http://es.wikipedia.org/wiki/Anexo:Entidades_federativas_de_M%C3%A9xico_por_densidad_de_poblaci%C3%B3n