

## **CAPITULO 1.**

### **INTRODUCCIÓN**

El uso de tecnologías para la búsqueda de conocimiento de información como la minería de datos geoespaciales, nos permite la localización de datos relevantes y útiles para la toma de decisiones en diferentes contextos tales como estudios en una zona urbana o descubrimiento de posibles rutas de escape en caso de cualquier posible desastre. Por estas razones el poder interpretar los resultados obtenidos de dichas tecnologías de manera clara y amigable es de gran prioridad en la toma de decisiones.

#### **1.1 Descripción del Problema**

Este proyecto de tesis tiene como propósito aplicativo proporcionar una interfaz de visualización, interacción e interpretación de los resultados obtenidos de la aplicación de minería de datos geoespaciales realizada por el Mtro. Manuel Pech Palacio para su proyecto de tesis doctoral. El problema radica al poder interpretar tales resultados para mostrarlos de manera visual ahorrándole tiempo útil al usuario final.

Actualmente existe un sistema que visualiza los patrones encontrados por la minería de datos en el contexto de la tesis doctoral de Manuel Pech Palacio del cual se partirá para el desarrollo de GeoMVisio, sin embargo no permite la interacción con la visualización de los resultados presentados para Tesis Doctoral de Manuel Pech Palacio reduciendo la posibilidad de interpretación al usuario.

### **1.3 Solución del Problema**

Para poder cumplir con el propósito de esta tesis el procedimiento a seguir será el siguiente. Primero se tendrá que comprender a fondo la función del sistema actual para entender la manera en que se visualizan los resultados de la minería de datos geoespaciales y saber que se deberá mejorar. Segundo se investigarán tecnologías existentes que ayuden a GeoMVisio en la interacción con las imágenes visualizadas y cotejar las funcionalidades que estas herramientas aporten para decidir si su integración vale la pena. Finalmente se procederá a crear el producto final de esta tesis, para esto se realizarán las mejoras que se recomiendan para el desarrollo final.

### **1.4 Minería de Datos**

La minería de datos se define como la extracción de conocimiento a partir de grandes cantidades de datos. Otra manera de llamarlo es el paso esencial para el proceso de descubrimiento de conocimiento de grandes bases de datos [Han, 2001]. Esta extracción de datos se lleva a cabo a partir de 7 pasos principales:

1. Limpieza de Datos (Eliminación de datos).
2. Integración de Datos (Combinación de datos).
3. Selección de Datos (Obtención de datos relevantes de una base de datos).
4. Transformación de Datos. (Los datos son representados correctamente para la minería de datos).

5. Minería de Datos (Procesado de datos a través de algoritmos para extraer patrones de datos).
6. Evaluación de Datos (Interpretación de los datos verdaderamente importantes).
7. Presentación de Conocimiento (Presentación al usuario del conocimiento extraído).

La minería de datos geoespaciales, como su nombre lo indica se encarga de extraer información a partir de bases de datos geoespaciales, que son aquellos que representan un espacio que ocupa un lugar geográficamente hablando (por ejemplo un río, un edificio, una isla). Los resultados que arrojan son patrones geoespaciales interesantes contenidos en una base de datos geográfica para un uso adecuado de la persona que realizó la minería de datos, según su campo de trabajo.

Estos datos son recopilados por diferentes tipos de aplicaciones como los Sistemas de Información Geográfica. Este campo es de gran ayuda en diferentes áreas de primordial interés como son urbanismo, planeación civil de rescate o evacuación en caso de un siniestro, planificación de edificación y construcción, economía, entre otros.

La minería de datos cuenta con diferentes sistemas para realizar la tarea de búsqueda de conocimiento. Un ejemplo es el sistema Subdue.

### **1.3.1 Visualización en el contexto de Minería de Datos**

La interpretación de los resultados de minería de datos es de suma importancia. Para ello debe de existir una manera de poder visualizar los patrones descubiertos. Debido a esto un sistema de minería de datos debe ser capaz de desplegar los patrones descubiertos de diferentes maneras según se acople a la información que fue representada, por ejemplo en forma de grafos, diagramas, entre otros.

El poder visualizar los patrones descubiertos de diferentes maneras ayuda al usuario para comprenderlos e identificar patrones de interés que representen un conocimiento útil o para poder guiar una búsqueda en un futuro [Han, 2001].

### **1.5 Sistema de minado de datos basado en grafos Subdue.**

Subdue es una herramienta basada en grafos desarrollada en lenguaje de programación C++ desarrollado en la universidad de Texas en Arlington [Subdue, 2005]. Este es un sistema de descubrimiento de conocimiento que encuentra patrones estructurados y relacionados.

Subdue representa la información por medio de conjuntos de grafos dirigidos o no dirigidos etiquetados cuyos componentes son representados por vértices etiquetados relacionados por medio de arcos también etiquetados.

Subdue emplea el principio de longitud de descripción mínima (*MDL* por las siglas en inglés de *minimum description length*) para guiar la búsqueda hacia las subestructuras más apropiadas. Aunque cuenta con otros dos métodos

para evaluar las subestructuras candidatas. El primer es “*Set Cover*” y el segundo es “*Size*”.

Subdue ha sido aplicado exitosamente en diferentes áreas entre las que se encuentran bioinformática, lucha contra el terrorismo, análisis de redes sociales, aviación y geología.

Cabe mencionar que esta tecnología no interviene directamente con el propósito de esta tesis, ya que la única interacción que se tiene con Subdue tiene que ver con los archivos de salida que arroja. Más adelante se explicará con más detalle la parte de Subdue que concierne a esta tesis.

### **1.6 Objetivo General**

Crear una interfaz que permita visualizar, discriminar y manipular las subestructuras obtenidas a partir de los archivos de resultados que proporciona el sistema Subdue. Así el usuario experto final podrá ver los patrones e interactuar con estos para poder decidir si es de importancia según sus necesidades y conocimientos.

### **1.7 Objetivos Específicos**

- Analizar y comprender el funcionamiento del sistema Subdue.
- La lectura de las subestructuras descubiertas por Subdue.
- Relación de subestructuras descubiertas por Subdue.

- Visualización de la subestructura descubierta en modo grafo.
- Darle al grafo la funcionalidad de un hipergrafo; es decir, la navegación a través de grafos.
- Implementar niveles de visualización.
- Impresión de la subestructura encontrada ya sea por medio de impresora y/o archivo JPEG.

## **1.8 Limitaciones**

Las limitaciones que se presentan básicamente conciernen a que no se interactúa directamente con el sistema que genera los archivos de entrada para el software. El resultado final de GeoMVisio depende de que los archivos que recibe sean correctos. Así también se presentan aquellas limitaciones que se traigan al integrar alguna tecnología de soporte.

## **1.9 Hardware y Software**

### **1.9.1 Hardware**

- PC
- Ambiente Windows XP Profesional

### **1.9.2 Software**

- Java JDK 1.5.0\_03
- Paquete de clases Grappa

- Aplicación Dot (Graphviz)

### **1.10 Estructura de documentos de tesis.**

La estructura de esta tesis esta organizada de la siguiente manera:

- I.** El capítulo 1 Introducción. Este capítulo introduce al tema de tesis, así como el proyecto al cual pertenece, nos menciona las herramientas y tecnologías base de la tesis.
- II.** Capítulo 2 Conceptos Generales. El segundo capítulo menciona y explica los conceptos generales y las herramientas que se usan para poder entender la noción general del documento y el software de esta tesis
- III.** Capítulo 3 Arquitectura. En este capítulo se observara la arquitectura creada para esta tesis y como se integran los diferentes sistemas y tecnologías relacionados con GeoMVisio para cumplir el objetivo.
- IV.** Capítulo 4 Análisis y Diseño. Este capítulo explica el análisis del problema para buscar las soluciones a los diferentes objetivos de la tesis para poder realizar el diseño de la misma. Se muestra el diseño del mismo y el modelo de clases que se creo para realizar la función de la tesis.

- V.** Capítulo 5 Implementación. Este capítulo muestra la manera en que se integraron las herramientas de soporte en el software de esta tesis exponiéndolo por medio del modelado del sistema. Se muestran de igual manera la implementación del trabajo final.
- VI.** Capítulo 6 Conclusiones y Trabajo a Futuro. Finalmente este capítulo presenta las conclusiones del trabajo de tesis. Aborda también lo trabajos a futuro para mejorar esta tesis.