

Capítulo 4

Tecnologías en los Sistemas de Información Geográfica

4. Tecnologías en los Sistemas de Información Geográfica

La incursión de la tecnología en diversos campos de la ciencia logró una notable evolución en cuanto a los métodos de procesamiento y almacenamiento de datos, permitiendo de una manera más rápida y sencilla el realizar análisis complejos y resolver problemas. Un ejemplo de este gran cambio es el manejo de información cartográfica digital, que de una representación de tipo espacial del medio natural (vegetación, suelos y ríos) y humano (carreteras, uso del suelo) aumentó su campo de acción hacia la creación de mapas de gran complejidad y el desarrollo de modelos digitales sobre el terreno.

Debido a estos antecedentes los Sistemas de Información Geográfica (SIG) han comenzado a cobrar gran importancia como herramienta en la toma de decisiones; no solo en la geografía, sino también en aquellas ciencias que están enfocadas principalmente en la planificación territorial, la resolución de problemas ambientales y socioeconómicos [López, Posada y Moreno, 2000] y en los últimos años hacia la mercadotecnia. Su aparición en la industria a finales del siglo XX se debe principalmente al bajo costo de los programas y el alto nivel de procesamiento de datos. En la actualidad empresas mexicanas como PEMEX y CEMEX ocupan esta herramienta para guiar sus excavaciones y actividades de campo, lo que les ha permitido ahorrar millones de pesos [Morales, 2001]. En el comercio, estos sistemas han ayudado a planear la factibilidad de localización de nuevas sucursales según factores como demografía, áreas de influencia, búsquedas de rutas y recorridos mínimos, por ejemplo.

4.1 Sistemas de información Geográfica (SIG)

Según la *National Center for Geographic Information and Analysis* (N.C.G.I.A.) podemos definir a un SIG como: "... un sistema de hardware, software y procedimientos diseñados para facilitar la obtención, gestión, manipulación, análisis, modelado y salida de datos espacialmente referenciados, para resolver problemas complejos de planificación y gestión" [NCGIA, 2002]. Otra definición es la que da la *Association for Geographic Information* (AGI) y el Departamento de Geografía de la Universidad de Edimburgo los cuales lo definen como: "...un sistema de ordenadores para obtener, almacenar, integrar, manipular, analizar y representar datos relativos a la superficie terrestre" [AGI, 2002]. Al leer estas 2 definiciones podemos concluir que ambas lo ven como un sistema que los ayuda a manipular, analizar y representar información.

4.1.1 Inicios

Los inicios sobre estos sistemas se remonta hasta principios de la década de los 60's, cuando bajo una estructura *ráster* y *vectorial*¹ que combinaba la cartografía con los datos necesarios para la gestión forestal de Canadá, Roger Tomlinson con apoyo de IBM desarrolló el *Canadian Geographic Information System* (C.G.I.S.). Poco tiempo después Ian McHarg escribe *Design with Nature*, este libro plantea el concepto de S.C.A (Análisis de Capacidad /Susceptibilidad) el cual se convertiría con el tiempo en la metodología SIG. Este es un método

¹ Estos conceptos serán definidos más adelante en la Sección 4.1.3.

manual que aplica superposiciones transparentes de matrices binarias, sin embargo tenía diversos problemas como la imposibilidad de ponderar variables (por ser binario), gran determinismo y el aumento de la dificultad de uso cada vez que se trataba de combinar una mayor cantidad de documentos.

Es en los 70's cuando universidades norteamericanas se interesan por el desarrollo de SIG's ocupando el concepto S.C.A de McHarg. El resultado fue un conjunto de sistemas matriciales o *raster* que se caracterizan por la sencillez, el bajo costo, el no poder manejar atributos y ser aplicables a espacios muy limitados. Los más notables son *SYMAP* y *GRID* de la Universidad de Harvard y el *Map Analysis Package (MAP)* de Yale. Otro sistema que salió por esas mismas fechas fue *DIME* el cual fue el primer sistema en manejar una topología completa. A finales de esta década el Laboratorio de Harvard desarrolla *ODYSSEY*, el cual es un SIG *vectorial* con superposición de polígonos mediante geometría coordinada.

Debido al gran estudio realizado por estos laboratorios, los años 80's se convirtieron en la época de mayor auge y desarrollo de los SIG's, principalmente vectoriales. Es aquí cuando se vuelven un producto industrial, ejemplo de ello es *ARC/INFO* del Instituto de Investigaciones de Sistemas Ambientales (ESRI). En la actualidad los SIG's se distinguen por ser sistemas que integran cada vez más los conceptos *ráster* y *vectoriales*, el aumento de comunicación entre sistemas e interfaces de usuario y el uso de herramientas basadas en la metodología orientada a objetos.

El siguiente paso en el desarrollo de los SIG es su integración en los llamados sistemas de soporte de decisiones, en los sistemas de sobremesa (divulgación de la cartografía y de la Información Geográfica), los sistemas y servidores de información geográfica en red y distribuidos (Internet) y los llamados SIG móviles el cual aplicamos en esta tesis (aplicación de los SIG en el ámbito de dispositivos móviles).

4.1.2 Funcionalidad

Existen al menos cinco argumentos básicos que fundamentan la utilización de un SIG (Domínguez, 2000), los cuales son:

- Permite gestionar un gran volumen de información a diferentes escalas y proyecciones.
- Realiza comparaciones entre escalas y perspectivas emulando una cierta capacidad de representación de diferentes lugares al mismo tiempo, esto es algo que llamamos *análisis vicariante*.
- Admite multiplicidad de aplicaciones y desarrollos; poniendo a nuestra disposición herramientas informáticas estandarizadas.
- Diferencia entre cambios cualitativos y cuantitativos; aportándonos una gran capacidad de cálculo.
- Integra espacialmente datos tabulares y geográficos junto a cálculos sobre variables (topología).

Debido a esto es posible afirmar que los SIG son una herramienta cada vez más utilizada e imprescindible para todas aquellas personas que utilizan información geográfica. Las dos respuestas fundamentales que un SIG contesta por medio de los mapas (según Cebrián, 1994) son:

- ¿Cuáles son las características de las localizaciones incluidas en un área dada?
- ¿Cuál es la distribución espacial de un cierto tipo de objeto?

4.1.3 Cartografía Digital

Tradicionalmente podemos definir a los mapas como archivos que almacenan objetos geométricos tales como polígonos, líneas y puntos. Sin embargo hay otro tipo de mapas que mediante el uso de matrices tratan al territorio como un todo continuo en un sistema cartesiano. En el primer caso hablamos de un **SIG vectorial** (Fig. 4.1) y en el segundo de un **SIG matricial o ráster** (Fig. 4.2).

Un sistema vectorial significa que representa el territorio a través de vectores, los cuales se localizan en el espacio por medio de coordenadas que especifican un origen y un destino. Las características del objeto representado van asociadas al mismo en forma de atributos, ejemplo de estos son: clase, tamaño, color, tipo de línea y grosor de la línea.

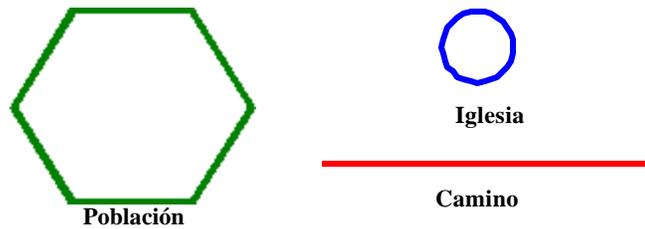


Figura 4.1 Representación Vectorial

Los SIG matricial o ráster también llamados mapa de bits, son aquellos que realizan sus cálculos a través de una estructura matricial en la que cada celda o píxel tiene un valor y una localización determinadas, logrando con ello que todo el territorio descrito tenga un valor. Podemos asignar valores binarios cuando queramos representar la existencia o ausencia de un objeto o podemos ocupar valores decimales para representar variables continuas. Esta modalidad de mapas se destina para imágenes vía satélite y fotos aéreas.

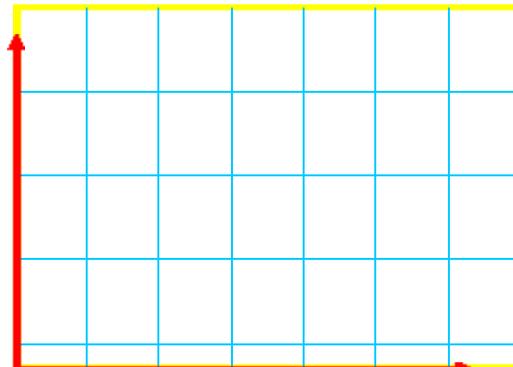


Figura 4.2 Representación Matricial

En ambos sistemas la diferencia fundamental se basa en la estructura de los datos espaciales, las relaciones topológicas, el volumen físico de la información y en los métodos de análisis.

4.1.4 Procesamiento de la información

Un SIG tiene una dependencia total de las fuentes de información de las que se alimenta, es decir, dependemos de la escala, el ajuste y la veracidad de los datos. Según Domínguez (2000) será imposible superar esta barrera a menos que nosotros seamos nuestra propia fuente de información.

Los datos que maneja un SIG podemos dividirlos en 2 grupos: Tablas y Gráficos. Las tablas están formadas por todos aquellos datos de estructura tabular que pueden ser relacionados con el territorio o superficie a analizar. El segundo grupo incluye en sí la cartografía y los productos derivados de la captación de imágenes desde un medio exterior por ejemplo: fotografía aérea e imágenes de satélite. Estos datos se relacionan entre sí en el sistema mediante lo que denominamos *apuntadores*, lo cual sería un dato común al mapa y a la base de datos.

Debido a lo anterior podemos afirmar que la información supone el 80% del buen funcionamiento de un SIG [Domínguez, 2000]. Una gran parte de los esfuerzos se centra en las tareas de armonización, corrección y puesta en marcha de los datos. Por ello, se hace imprescindible se tenga en cuenta el análisis de la información a utilizar, así como los métodos de carga al sistema y el control sobre los datos. En la actualidad hay muchos organismos y empresas que se dedican a proporcionar esta información, ejemplo de ello es el INEGI aquí en México y el EUROSTAT en Europa.

4.1.5 Análisis Espacial usando un SIG

El mayor provecho de estos sistemas se obtiene cuando es posible generar información a partir de la relación que se genera entre los datos geoespaciales y los existentes referente a los temas que resulten de interés. Las principales herramientas que se usan en este tipo de análisis son:

- **Superposición.** Esta se considera la herramienta básica del análisis espacial. Permite por medio de métodos matriciales o vectoriales el acumular capas de información referidas a una misma área, el resultado es la obtención de análisis más complejos y detallados. Es una fuente importante de cartografía analítica y sintética. Según Domínguez (2000), esta herramienta debe responder preguntas tales como: ¿Qué es común?, ¿Qué es diferente? y los relativos a “incluido en, pertenece a”, por ejemplo.
- **Área de influencia.** Son aquellas que a partir de una entidad espacial y de acuerdo a un conjunto de variables definen una nueva entidad en el espacio. Por lo general estas nuevas entidades son del estilo de corredores (*buffers*), círculos o coronas y figuras irregulares. Otra posibilidad es la *segmentación dinámica*, la cual permite dividir una línea en relación a los diferentes valores que posea en cada segmento

pudiendo de este modo dar diferentes anchuras a un mismo corredor en base al valor de la variable de cada segmento.

- **Análisis de vecindad.** Estos son usuales en los sistemas *ráster*. Permiten, mediante la aplicación de algoritmos, conocer cómo se relaciona un objeto geográfico con su entorno y viceversa.
- **Análisis de redes.** Esta herramienta se apoya de la topología. Puede construir redes de cualquier tipo (hidrográficas, carreteras, transportes, eléctricas...) siempre que mantengan su característica de sistema (dirección, conexión, etc.).
- **Modelos Digitales de Elevaciones.** Herramienta clásica. Estos tienen diferentes nombres en función de la técnica utilizada para el levantamiento o de la variable a representar. Las técnicas utilizadas varían desde la utilización de modelos de triangulación (comunes en geodesia) a la realización de matrices cuadrangulares aportando un valor de z a cada celda. Una de las ventajas de estos modelos es la posibilidad de obtener perfiles o cortes.
- **Modelización y prospectiva.** En esta herramienta el SIG puede ser utilizado como una herramienta de simulación y de prospección a partir de la modelización del territorio.

4.1.6 Campos de aplicación y herramientas de software

El uso y aplicación de los SIG ha ido en aumento partir de la década de los 80's ya que es cuando los sistemas se vuelven un *producto industrial*. Hay que diferenciar en este punto que las aplicaciones son diferentes a los campos de aplicación. La primera se refiere a dar una solución a un problema en concreto por ejemplo sistemas como ArcView 3.2 (Spatial Analyst) e Idrisi 32, y la segunda se enfoca a las áreas donde su influencia resolvería problemas. Comenzaremos con los campos de aplicación:

- **Administración Pública.** Este campo podemos decir que ha sido el gran motor del desarrollo e implantación de los SIG. La mayor parte de los organismos vinculados de una u otra forma con la organización territorial, el medio ambiente, la gestión catastral, etc., han incorporado esta tecnología. Aunque en muchos lados su desempeño no ha sido el óptimo, con el mejoramiento de la información estos se volverán sistemas muy robustos.
- **Carácter socioeconómico.** Se incluyen aplicaciones del tipo de localización de servicios y negocios, análisis financieros y de mercado, gestión del patrimonio y geomarketing.
- **Medio ambiente.** Este es el campo más tradicional y frecuente de desarrollo de aplicaciones. Se opina que el estudio del Medio Ambiente

encaja mejor en la lógica de análisis de los sistemas ráster, debido en gran medida a que los primeros sistemas de este sector se desarrollaron bajo este formato y a que los estudios medio ambientales suelen utilizar variables continuas que se representan mejor en esos sistemas.

- **Utilidades.** Este suele incluir aquellos apartados referidos básicamente a redes de conducción de energía (agua, electricidad, etc.).
- **Otros campos de aplicación.** Otro campos de los que podemos hablar son de la educación, seguridad, controles de navegación, análisis electorales e investigación, por mencionar algunos.

Como pudimos ver los SIG son una tecnología que poco a poco comienza a tener participación en ámbitos que hace algunos años nunca se hubiera podido pensar. Esto se debe principalmente a que estos sistemas poco a poco se han ido adaptando a las necesidades de los usuarios y al constante cambio en los dispositivos, ejemplo de ello es el otro campo de aplicación: Internet.

- **Internet.** Según Harder (98) Internet no cambia la naturaleza básica de los SIG, la pone "on line". Los servidores de mapas tiene su primer desarrollo en 1993 con el Xerox PARC Map Viewer, el cual era un servidor HTML que ya no está en funcionamiento. Los avances más recientes van enfocados hacia las estructuras cliente - servidor, SIG

basados en la tecnología orientada a objetos y SIG distribuidos. En la actualidad un SIG que funcione en la red mundial puede utilizarse para buscar rutas, localizar servicios, conectarse a bases de datos de organismos públicos como el INEGI, realizar análisis de diversas índoles, predicción del tiempo y procesar datos procedentes de la teledetección, por citar algunos ejemplos. A continuación se mostrará cuales son actualmente las formas de funcionar de un SIG de acuerdo a una clasificación hecha por Domínguez (2000):

- Los más simples son las aplicaciones estáticas, es decir, el servidor web simplemente pone a disposición del usuario una imagen en formato GIF o JPG. Estos mapas sólo muestran localizaciones. Por ejemplo los mapas que brinda el sitio de Terra.com.mx [Terra, 2002], en la Fig. 4.3.
- Una variación de estas aplicaciones serían aquellas que se actualizan cada determinado tiempo. Por ejemplo la página del CENAPRED [CENAPRED, 2002]. En esta pagina la imagen del volcán se actualiza cada minuto (véase Fig.4.4).

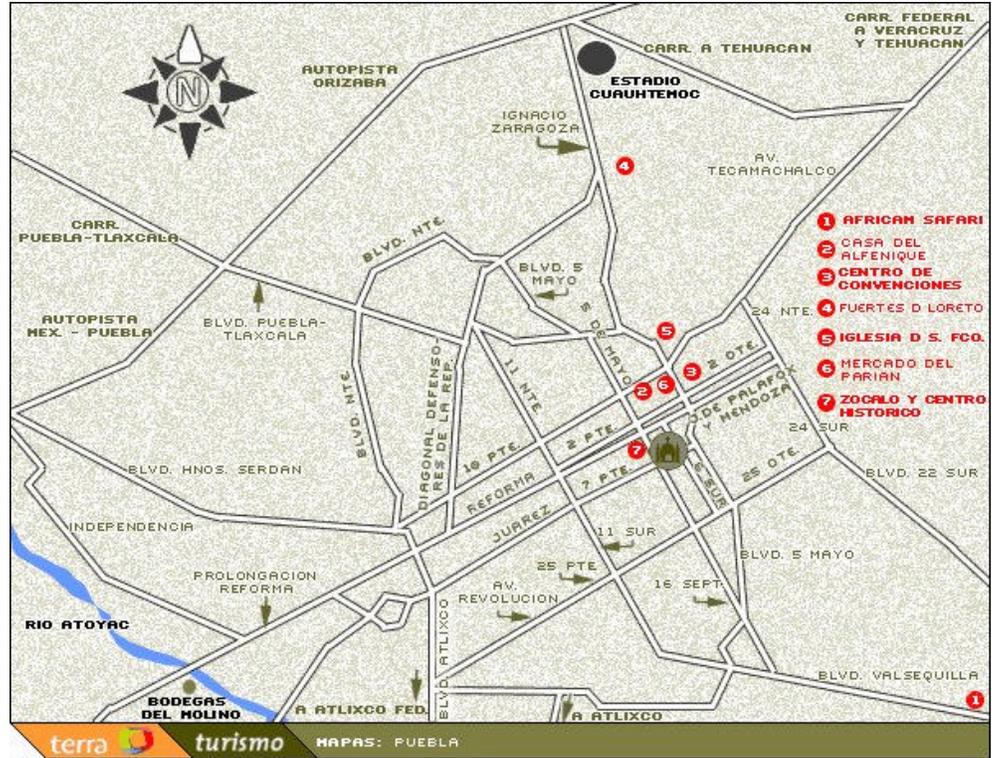


Fig. 4.3 Mapa de la Ciudad de Puebla

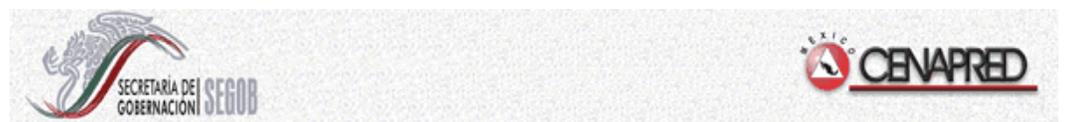


Fig. 4.4 Secretaría de Gobernación - CENAPRED

- Subiendo de complejidad tenemos aquellas aplicaciones en las cuales el usuario puede generar su propio mapa, es decir, tenemos un SIG funcionando que acepta peticiones del usuario y da como respuesta un mapa. Este es el servicio que da el Instituto de Estadística de Andalucía [IEA, 2002] que a partir de mapas y estadísticas el visitante puede generar mapas de provincias de esta comunidad autónoma en base a una variable dada (Fig. 4.5).

Instituto de Estadística
de Andalucía **IEA**

Figura 4.5 Instituto de Estadística de Andalucía

- El siguiente paso son aquellos sitios en los cuales el resultado es un mapa producto de un análisis espacial como pueden ser búsquedas geográficas, búsquedas condicionadas, etc. Ejemplo de este es la aplicación del sitio de Visa: “Atm locator” [VISA, 2002], en su sitio uno puede encontrar todos los cajeros automáticos que existen en la zona en la que el usuario se encuentra (Fig. 4.6).



Fig. 4.6 Atm locator

- Otra aplicación es en la que el resultado son mapas producto de un procesamiento de datos geográficos, en este caso el servidor procesa o transforma los datos almacenados como respuesta a la petición del usuario.
- Los servicios de datos públicos. En este caso un organismo pone a disposición del usuario sus datos geográficos para ser descargados en su computadora. Ejemplo de este servicio es el que brinda el Servicio Geológico de los Estados Unidos [USGS,2002], del cual se pueden descargar las líneas de

costa de todo el mundo a diferentes escalas y en diferentes formatos (Fig. 4.7).



Fig. 4.7 Servicio Geológico de los Estados Unidos

- Servicio de venta de datos. El objetivo es el mismo que el anterior solo que aquí la empresa u organismo cobran por facilitar la información.

Como pudimos observar el fenómeno Internet está marcando la pauta en los desarrollos SIG. Sin embargo de manera paralela, los SIG móviles que engloba la telefonía móvil y la tecnología WAP (*Wireless Access Protocol*) cada día van teniendo avances más notables; seguidos muy de cerca por la introducción de la tecnología XML (*eXtensible Markup Language*) en los nuevos sistemas.

4.2 El uso del estándar XML

Desde la introducción del concepto XML a finales de la década de los 90's, han surgido una gran cantidad de ramas, estándares y reglas que han hecho pensar a la comunidad informática en general si es necesario hacer tantas reformas a la estructura actual. Mucho se habla sobre las virtudes y defectos de

esta nueva tecnología, pero a decir verdad pocas son las personas que realmente entienden este nuevo concepto.

Inicialmente diremos que XML no es un lenguaje de marcado o etiquetado como lo diría su traducción al español, sino un paquete de herramientas que nos permite crear, modelar y usar lenguajes de marcado. Según T.Ray (2001) debido a esta característica tan peculiar, dentro de poco tiempo tendremos varios lenguajes de marcado que cubrirán cualquier tipo de documento o aplicación.

4.2.1 Que es XML

T. Ray (2001) define XML o “eXtensible Markup Language” de acuerdo a 3 niveles:

- a) Nivel 1: Un protocolo que nos permite tener acceso al contenido y manejo de información.
- b) Nivel 2: Una familia de tecnologías que pueden hacer todo, desde formateo de documentos hasta filtración de datos.
- c) Nivel 3: Una filosofía para el manejo de información que busca la máxima utilidad y flexibilidad de los datos por medio del refinamiento más puro o de la forma más estructurada.

La mayoría de los usuarios sólo se limitan a ocupar el primer nivel de esta tecnología. Tomando como referencia ese punto, podemos definir a XML como un lenguaje de marcado para documentos que contienen información estructurada

[Walsh, 98]. El concepto “información estructurada” se define como la unión de: contenido (es decir, palabras e imágenes) y alguna indicación de cual es el rol que juega ese contenido (por ejemplo: pie de pagina). Una característica fundamental de XML es que es independiente de la representación que se le de, es decir, sólo almacena los datos el tratado y la presentación puede ser tan variable como se desee.

El grupo responsable de desarrollar XML es el: *XML Working Group*, el cual fue promovido por el *World Wide Web Consortium (W3C)* en el año de 1996 y fue presidido por Jon Bosak de *SUN Microsystems*. La versión 1.0 de XML es una recomendación por el W3C desde el 10 de febrero de 1998. Esta especificación es un subconjunto del estándar SGML (Standard Generalized Markup Language) que data de 1986 [Bosak, 2000] definido por ISO 8879 y que fué desarrollado inicialmente por IBM en la década de los 70's.

4.2.2 Estructura de XML

Inicialmente diremos que XML puede parecer similar a HTML solo que existe una diferencia: un documento XML contiene datos que se auto-definen exclusivamente y un documento HTML contiene datos mal definidos mezclados con elementos de formato [Reino, 2000].

Cuando se piensa en un documento, uno puede imaginarse una serie de palabras, organizadas en párrafos, secciones y capítulos. En otras palabras

registros leídos por humanos (libros, artículos, ensayos, etc); pero un documento XML es algo más general. Se considera como la unidad básica de información XML conformado por elementos y otros *tags* de una manera jerárquica; el documento puede contener texto como en un artículo, pero no es necesario que lo tenga [T. Ray, 2001].

Una forma de entender cual es la estructura de un documento XML , es observando un ejemplo:

```
<?xml version="1.0" encoding="UTF-7" standalone="yes"?>
<!DOCTYPE MENSAJE SYSTEM "mensajeEmail.dtd">
<mensajeEmail>
  <remitente>
    <nombre>Miguel Méndez R.</nombre>
    <email>migmenr@hotmail.com</email>
  </remitente>
  <destinatario>
    <nombre>Dr. David Sol</nombre>
    <email>sol@mail.udlap.mx</email>
  </destinatario>
  <fecha>28 de Octubre de 2002</fecha>
  <asunto>Proyecto de Tesis</asunto>
  <texto>
    <parrafo>
      Este es un ejemplo de cómo es la estructura de
      un XML. Es muy
      sencillo. Y lo usaremos para el Proyecto de
      Tesis.
    </parrafo>
  </texto>
</mensajeEmail>
```

Todo documento XML puede ser representado en forma de árbol. Cada parte del documento ya sea elemento, atributo, texto o un comentario tiene una

representación como un nodo [T. Ray, 2001]. Ya que esto facilita de gran manera su manipulación.

En la figura 4.8 se tiene un diagrama de árbol que representa el mensajeEmail, las cajas con recuadro rojo representan los seis elementos. El elemento “mensajeEmail” con recuadro color marrón se le denomina *elemento raíz*, debido a que encapsula a todos los demás elementos. Las cajas con recuadro azul o *hojas* representan el contenido del documento.

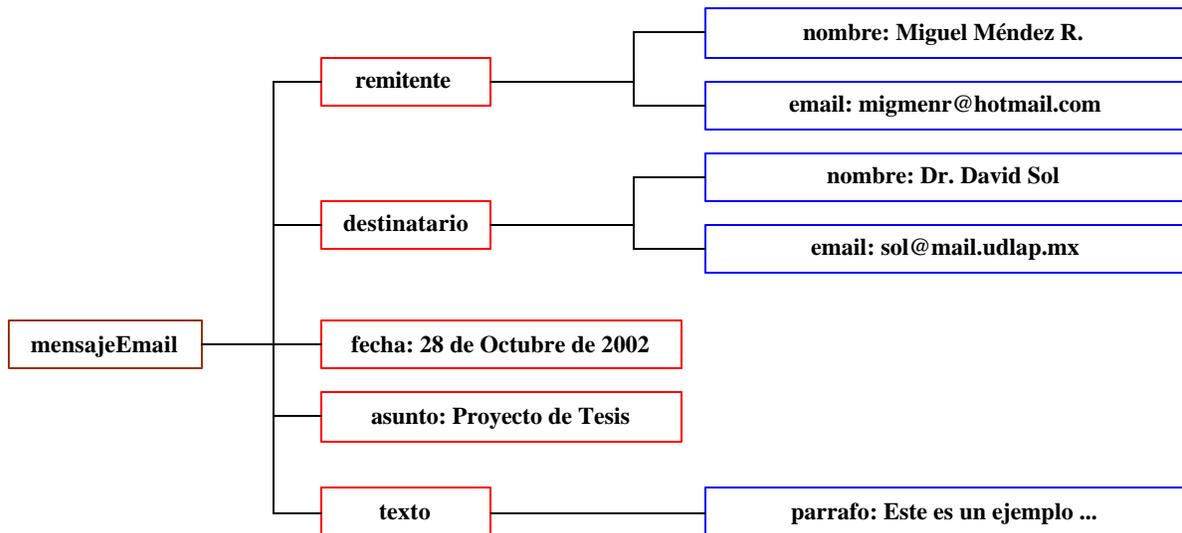


Fig. 4.8 Diagrama de árbol para el ejemplo mensajeEmail

A pesar de que XML fue diseñado para que los humanos lo pudieran interpretar, no se intenta que sea un documento final, es decir, no se espera a que se abra el documento en un editor de texto y se encuentre formateado; se necesitan de otros recursos para pulir la presentación como lo son las hojas de estilo (*stylesheets*) [Ray, 2001].

4.2.2.1 Documentos XML bien formados

Según Barbero (1999), los documentos bien formados son aquellos que cumplen las especificaciones del lenguaje respecto a las reglas sintácticas, sin estar sujetos a unos elementos fijados en un DTD² y seguir una estructura jerárquica muy estricta.

Como hemos podido observar, la sintaxis de XML y HTML es muy parecida entre sí, pero hay ciertas diferencias [Reino, 2000] que hacen resaltar a un documento XML:

- a) **Estructura jerárquica de los elementos:** Esta estructura se debe seguir en cuanto a las etiquetas que delimitan sus elementos. Una etiqueta debe estar correctamente “incluida” en otra y los elementos con contenido deben estar correctamente “cerrados”.
- b) **Etiquetas vacías:** Se debe seguir un formato específico para estos casos `<elemento-sin-contenido/>`
- c) **Un solo elemento raíz:** Forzosamente debe existir solo un elemento inicial que englobe a todos los demás.

² Definición de tipo de Documento (DTD): Es una definición de los elementos que puede haber en el elemento XML, y su relación entre ellos, sus atributos, posibles valores, etc. Es decir, es la definición de la gramática.

- d) **Valores de atributos:** El valor de un atributo forzosamente tendrá que estar encerrado en comillas simples (') o dobles (").
- e) **Tipo de letra, espacios en blanco (espacios, tabuladores y saltos de línea):** XML es sensible al tipo de letra utilizado, es decir, diferencia entre mayúsculas y minúsculas, En cuanto a los espacios en blanco éstos son ignorados por los procesadores XML en general y sólo son utilizados para dar mayor legibilidad al código.
- f) **Asignación de nombres:** Según la especificación XML 1.0

Un nombre [empieza] con una letra o uno o más signos de puntuación, y [continúa] con letras, dígitos, guiones, rayas, dos puntos o puntos, denominados de forma global como caracteres de nombre. Los nombres que empiezan con la cadena "xml", se reservan para la estandarización de esta o de futuras versiones de esta especificación.

- g) **Marcado y datos:** Las construcciones como etiquetas, referencias de entidad y declaraciones se denominan "marcas". Estas son la parte del documento que el procesador XML espera entender. El resto del documento que se encuentra entre las marcas, son los datos que resultan entendibles por las personas.

Ahora, para tener una idea más clara sobre un documento XML debemos conocer su estructura, una vez entendiéndola se sigue la misma idea para los demás lenguajes de marcado:

a) **Prólogo:** Aunque es opcional, los documentos XML pueden empezar con una línea que describa:

- 1) La **versión** de XML. Por ejemplo: 1.0
- 2) **Tipo de codificación** que tiene: US-ASCII (7 bits) o UTF-8 (código Unicode). El más común para nosotros por el uso con lenguajes europeos y por los caracteres del español usamos UTF-7 o ISO-8859-1.
- 3) Además se puede incluir una **declaración de documento** autónomo, es decir, que controla que componentes de la DTD son necesarios para completar el procesamiento del documento.
- 4) **Declaración de tipo de documento:** Define que tipo de documento estamos creando para ser procesado correctamente, es decir, valida y define los datos que contiene nuestro documento XML. Adicional a esto se pone un identificador público (PUBLIC) para poder hacer referencia a dicho DTD o mediante un Identificador Universal de Recursos (URI) precedido por la palabra SYSTEM (Fig. 4.9).

```
<?xml version="1.0" encoding="UTF-7" standalone="yes"?>  
<!DOCTYPE MENSAJE SYSTEM "mensajeEmail.dtd">
```

Fig. 4.9 Declaración de Tipo de Documento

- b) **Elementos:** Estos componentes pueden tener contenido o bien ser elementos vacíos. Cada uno de ellos tiene una sintaxis diferentes:

Sintaxis del elemento con contenido (Fig. 4.10):

```
<nombre atributoN="valorN"> contenido </nombre>
```

Figura 4.10 Sintaxis de elemento con contenido

Sintaxis de un elemento vacío (Fig. 4.11):

```
<nombre atributoN="valorN"/>
```

Figura 4.11 Sintaxis de elemento vacío.

- c) **Atributos:** En ocasiones es necesario transmitir una mayor cantidad de información sobre un elemento, de la que el nombre o el contenido pueden proveer. El uso de atributos nos permite describir información que afecta la conducta del elemento de una manera clara y efectiva. Por ejemplo (Fig. 4.12) podemos dar al elemento una etiqueta única o podemos describir un propiedad del elemento [Ray, 2001]. Los valores siempre deben ir entre comilla sencilla (') o doble (").

Ejemplo de un atributo:

```
Distancia = '15'  
<cita texto = "'Hola, como estas?', esta fría la noche">
```

Fig. 4.12 Ejemplo de declaración de atributos.

d) **Entidades Predefinidas:** Están compuestas de un nombre y un valor.

Funcionan como una macro, cuando la aplicación procesa un documento genera una tabla, en la cual graba los nombres de las entidades existentes y su valor. Cuando encuentra el nombre de una entidad en el contenido del documento, lo intercambia por el valor correspondiente (Espinosa, 2002). Se declaran una sola vez y pueden ser usadas cuantas veces sea necesario. Las entidades se declaran en el prologo, específicamente en la declaración del tipo de documento (Fig. 4.13).

```
<!ENTITY nombre "valor">
```

Figura 4.13 Declaración de Entidades.

La manera de llamar una entidad es la siguiente: **&nombre;**

Este es un ejemplo de como funcionan (Fig. 4.14):

<pre><?xml versión=1.0?> <!DOCTYPE mensajeEmail> <mensajeEmail> <remitente>Miguel Méndez</remitente> <destinatario>Héctor Montiel</destinatario> <parrafo> Este es un ejemplo de cómo funciona una identidad en un archivo XML. Por cierto la clase será dada en el salón B312 del Edif. 31 en la UDLA-P. Por cualquier duda aquí esta mi email <email>prueba@hotmail.com</email> -- Saludos PD. Recuerda que es el salon B312 del Edif. 31. </parrafo> </mensajeEmail></pre>	<pre><?xml version=1.0?> <!DOCTYPE mensajeEmail [<!ENTITY email "<email>prueba@hotmail.com</email>"> <!ENTITY salon "salón B312 del Edif. 31"> <!ENTITY uni "UDLA-P"> <!ENTITY firma "-- Saludos">]> <mensajeEmail> < remitente > Miguel Méndez </ remitente > < destinatario > Héctor Montiel </ destinatario > < parrafo > Este es un ejemplo de cómo funciona una identidad en un archivo XML. Por cierto la clase será dada en el &salon; en la &uni;. Por cualquier duda aquí esta mi email &email; &firma; PD. Recuerda que es el &salon; </parrafo > </mensajeEmail></pre>
<i>mensajeEmail sin uso de entidades</i>	<i>mensajeEmail con el uso de entidades</i>

Fig. 4.14 Ejemplo de cómo funcionan las Entidades.

Tal vez pueda parecer que el uso de la entidades hace más complicado el uso de un archivo XML, sin embargo si analizamos el ejemplo con las entidades veremos que el texto se reduce considerablemente y nos permite el reuso de determinadas partes. Dentro de estas mismas entidades tenemos las que se denominan entidades de carácter, es decir que representan solamente un carácter. En esta categoría tenemos 2 tipos:

1) **Entidades de carácter predefinidos:** Este grupo representa a los caracteres que no pueden ser utilizados en el contenido de los elementos. La especificación de XML, nos provee de las siguientes entidades predefinidas, con el fin de poder expresar los caracteres de manera segura (Tabla 4.1).

Nombre	Valor
Amp	&
Apos	'
Gt	>
Lt	<
Quot	"

Tabla 4.1 Entidades de carácter predefinidos

2) **Entidades de carácter numerado:** XML al soportar unicode, tiene acceso a miles de símbolos, letras e ideogramas [Ray, 2001]. Para tener acceso a estos elementos se utilizan las entidades de carácter numerado. El nombre de estas entidades es de la forma: #n, en donde n es la posición que ocupa el carácter en unicode. Por

ejemplo si se requiere utilizar el símbolo “ç”, se usa la siguiente entidad `ç` o su equivalente en hexadecimal `ç`;

- e) **Bloques CDATA:** Significa *character data*. Se utiliza cuando se necesita escribir caracteres prohibidos como `&`, `<`, `>`, etc. En este bloque se le dice al documento que no interprete el contenido, sino que solo lo trate como texto. La sintaxis es la siguiente (Fig. 4.15):

```
<![CDATA [ texto ]]>
```

Fig. 4.15 Ejemplo de declaración de un Bloque CData

El único texto que no puede ir dentro del contenido del bloque CDATA es la cadena `']]>` que representa el fin del bloque.

- f) **Comentarios:** Estos tienen el mismo formato que en HTML. También son ignorados en el momento del procesado de información. Su sintaxis es la siguiente (Fig. 4.16):

```
<!--mensaje -->
```

Fig. 4.16 Declaración de un Comentario

Como pudimos ver esto es la estructura de un archivo XML, no se entrará más en detalles debido a que no es el tema en esta tesis.

4.2.3 Aplicaciones y usos de XML

Debido a su gran flexibilidad el estándar XML se puede usar para una infinidad de trabajos y aportar muchas ventajas en muchos campos prácticos. Según Álvarez (2002) algunas aplicaciones en específico serían:

- **Comunicación de datos:** Si la información se transfiere en este formato, cualquier aplicación podría escribir un documento de texto plano con los datos que estaba manejando en formato XML y otra aplicación recibir esta información y trabajar con ella.
- **Migración de datos:** Si tenemos que mover los datos de una base de datos a otra sería muy sencillo siempre y cuando las dos trabajasen en formato XML.
- **Aplicaciones web:** Hasta ahora cada navegador interpreta la información a su manera por lo que en la mayoría de las ocasiones se tienen que hacer funciones para un navegador y la misma pero con formato diferente para otro. Con XML tenemos una sola aplicación que maneja los datos y para cada navegador podremos tener una hoja de estilo o similar para aplicarle el estilo adecuado. Si mañana nuestra aplicación debe correr en WAP solo tenemos que crear una nueva hoja de estilo o similar.
- **Nuevos lenguajes de marcado (*Markup Languages*):** Esta podría ser llamada como la aplicación principal de XML. La facilidad de poder crear nuevos lenguajes para el procesamiento de información; ejemplo claro de ellos

es GML (*Geography Markup Language*), el cual es utilizado en esta tesis y será descrito a continuación.

4.3 GML

GML o "*Geography Markup Language*", es un lenguaje basado en XML que nos ayuda a representar, intercambiar y almacenar información de tipo geográfica. Este lenguaje está basado en un estándar desarrollado por el *OpenGIS Consortium* (OGC). En su fase inicial empresas como *Oracle Corporation*, *Galdos Systems Inc.*, *MapInfo*, *CubeWerx* y *Compusult Ltd.* realizaron algunas pruebas a través de una serie de demostraciones pertenecientes al *Consortium's Web Mapping Test Bed* (WMT) llevado a cabo en septiembre de 1999.

De acuerdo a lo anterior uno podría pensar que información geográfica y la representación gráfica de la información resulta algo similar, sin embargo no es así. Al hablar de información geográfica tratamos de capturar información acerca de la geometría y las propiedades de los objetos que se encuentran en el mundo. La manera en la que simbolizamos esa información en un mapa, es decir, el color o el grosor de línea que usamos se relaciona más con la visualización [Lake, 2000].

De la misma manera que cualquier aplicación de XML, GML representa información geográfica por medio de texto. Permite visualizar archivos de una manera simple y clara, además de que facilita la actualización de la información de

manera muy rápida. Uno de los primeros trabajos en esta campo fue realizado por la Provincia de Columbia Británica la cual desarrollo el formato denominado SAIF. En la actualidad esta provincia cuenta con un acervo de mas de 7000 archivos en este formato [Lake, 2000].

4.3.1 Estructura

Como hemos leído, GML está basado en un modelo abstracto de geografía desarrollado por el OGC. Este describe el mundo en términos de entidades geográficas llamadas “*features*”.

4.3.1.1 Entidades geográficas (Features)

De acuerdo con la OGC un *feature* es una abstracción de un fenómeno real que esta asociado con a una ubicación relativa con la tierra. Esencialmente no es más que una lista de propiedades y geometrías. Las propiedades están compuestas por un nombre, tipo y valor; y las geometrías por grupos de figuras básicas tales como: punto, línea, curva, superficies y polígonos.

Así mismo, GML provee la manera de representar las coordenadas de un elemento geométrico, donde este elemento se denomina <coordinates> (Fig. 4.17). Los valores dentro de este elemento se manejan por parejas (x,y) separados por coma y cada tupla se separa por un espacio en blanco o *return*.

```
<coordinates >  
570945.0990594306,2101418.49307865  
570916.2888732497,2101460.885209745  
570916.700447338,2101494.222710897  
</coordinates >
```

Fig. 4.17 Declaración de coordenadas.

Una vez entendido este esquema podemos avanzar a explicar como representar las formas geométricas vistas anteriormente según se usan en esta tesis.

a) Point (Punto): Este representa un punto con coordenadas (x,y) en el plano cartesiano (Fig. 4.18).

```
<Point >  
  <coordinates >  
    73582.2551207771,106414.46793016  
  </coordinates >  
</Point >
```

Fig. 4.18 Declaración de Punto

b) LineString (línea): Representa una secuencia de líneas rectas, un camino cerrado indica que la ultima coordenada (x,y) es igual a la primera (Fig. 4.19).

```
<LineString >  
  <coordinates >510320.152,2122945.  
    937 510320.146,2122953.343  
    510305.263,2122954.818  
  </coordinates >  
</LineString >
```

Figura 4.19 Declaración de una línea.

c) LinearRing: Representa una secuencia cerrada de líneas rectas, la ultima coordenada (x,y) debe de ser igual a la primera (Fig. 4.20).

```
<LinearRing>  
  <coordinates>  
    570945.0990594306,2101418.4930786  
    570916.2888732497,2101460.88520975707  
    29.0226630731,2101863.4046681  
    573465.5787761798,2103072.60933955710  
    02.7194317926,2101492.5764145  
    570945.0990594306,2101418.4930786  
  </coordinates>  
</LinearRing>
```

Figura 4.20 Declaración de una secuencia de líneas cerradas.

d) Polygon (Polígono): Representa un polígono (Fig. 4.21).

```
<Polygon>  
  <outerBoundaryIs>  
    <LinearRing>  
      <coordinates>  
        571906.6429502425,2108956.0757805724  
        571934.9288294474,2108906.575491964  
        571871.2856012364,2108779.289035542  
        571779.3564938207,2108574.216411307  
        571256.067728531,2109465.2216062597  
        571517.7121111759,2109260.1489820243  
        571906.6429502425,2108956.0757805724  
      </coordinates>  
    </LinearRing>  
  </outerBoundaryIs>  
</Polygon>
```

Figura 4.21 Declaración de un Polígono

Además de las entidades básicas, existen otras que surgen a raíz de unir varias unidades básicas:

e) Multipoint (Multipunto): Este lo podemos definir como un conjunto de puntos (Fig. 4.22).

```
<MultiPoint >
<pointMember>
<Point >
<coordinates >
571256.067728531,2109465.2216062597
</coordinates >
</Point >
</pointMember>
<pointMember>
<Point >
<coordinates >
571871.2856012364,2108779.289035542
</coordinates >
</Point >
</pointMember>
</MultiPoint >
```

Figura 4.22 Declaración de un Multipunto.

- f) **MultiLineString (Multilínea)**: Podemos definirlo como un conjunto de LineStrings (Fig. 4.23).

```
<MultiLineString >
<lineStringMember >
<LineString >
[...]
</LineString >
</lineStringMember >
<lineStringMember >
<LineString >
[...]
</LineString >
</lineStringMember >
</MultiLineString >
```

Figura 4.23 Declaración de una Multilínea

- g) **MultiPolygon (MultiPolígono)**: Es un conjunto de polígonos (Fig. 4.24).

```
<MultiPolygon >
<polygonMember >
<Polygon >
[...]
</Polygon >
</polygonMember >
<polygonMember >
<Polygon >
[...]
</Polygon >
</polygonMember >
</MultiPolygon >
```

Figura 4.24 Declaración de un Multipolígono

Hasta ahora éstas son las formas con las cuales podemos trabajar en un archivo GML. Nosotros podemos jugar con ellas de la forma que deseemos y además darles el significado que más se apegue a nuestras necesidades.

4.3.1.2 Colección de Entidades Geográficas (Feature Collections)

GML introduce el elemento Feature Collections como base del documento GML es decir el elemento raíz. Éste consta de un elemento llamado Bounding Box y una colección de features [Cuthbert, 2000]. La función principal de este elemento raíz es demarcar el punto de valores (x,y) menores y el punto de valores (x,y) mayores de las geometrías que engloba.

4.3.2 Tecnologías de las que depende GML y sus beneficios

GML depende de 3 tecnologías en específico para tener un correcto funcionamiento [Espinosa, 2002] :

- a) **XML**: Un documento GML tiene que ser validado de la misma manera que cualquier documento XML.
- b) **SVG, VML (Microsoft's Vector Markup)**: Estas son aplicaciones XML para poder representar gráficos vectoriales. Estas aplicaciones están orientadas a la apariencia, es decir, se preocupan por las formas, color, grueso de líneas, transparencia, etc.
- c) **XSLT**: Para poder generar un mapa a partir de un documento GML, es necesario su transformación a un formato de visualización como lo es

SVG o VML. Esto requiere asociar a cada entidad geográfica con un estilo grafico (color, textura, etc.) por medio de XSLT.

En cuanto a los beneficios que ofrece esta tecnología podemos decir que nos da una mayor facilidad en el manejo y edición de mapas, ya que nos permite tener un mayor control sobre ellos. Otro punto a favor es el poder tener mayor cantidad de información de graficación permitiendo la manipulación de consultas más complejas.

4.3.3 Aplicaciones

En la actualidad este lenguaje se encuentra en la versión 2.0 y está en espera de que se libere la versión 3.0 a finales de este año.

GML fue adoptado recientemente por el Servicio Geográfico del Reino Unido como medio para entregar "todos" sus datos geográficos para toda Gran Bretaña. También Canadá ha adoptado recientemente GML, convirtiéndolo en la clave para la estandarización de su desarrollo en Infraestructuras Geo-Espaciales (CGDI).

En la actualidad, en el laboratorio del CENTIA dedicado al diseño de Sistemas de Información Geográfica en la UDLA-P este se ha convertido en el método estándar de guardado y procesamiento de información. Ejemplo de ello es esta tesis, el cual es usado dada su sencillez y tamaño como medio para enviar

información de una PDA a una computadora de escritorio y viceversa de una manera rápida; la principal ventaja que ofrece este tipo de formato es el poder manejar el contenido de estos archivos como si fuera texto plano.

El unir la funcionalidad de un SIG, la tecnología *wireless* y el estándar de OpenGis en un solo proyecto parecería complicado, sin embargo, veremos una solución propuesta para los investigadores de campo que trabajan en la periferia del volcán uniendo estos tres conceptos.