

Estructura de las bases de datos geográficas.

Introducción.

Vamos a describir aquí la estructura de las bases de datos geográficas en general, presentando un modelo que es una abstracción de los casos más habituales y útiles. Por supuesto, cada sistema particular tendrá sus particularidades pudiéndose apartar más o menos de lo expuesto aquí.

Para abordar el tema presentaremos primero una idea del contenido y propósito de una base de datos geográfica y compararemos las estructuras de diferentes sistemas.

Aquí trataremos únicamente las estructuras de datos de los SIG bidimensionales, que son los más ampliamente difundidos, conocidos y empleados. Estos sistemas gestionan elementos espaciales de cero, una o dos dimensiones sobre la superficie de la tierra, más —en algunos casos— la elevación del elemento. Estos sistemas incluyen habitualmente la posibilidad de gestionar y visualizar modelos tridimensionales de la superficie (continua) terrestre. Las habilidades tridimensionales de estos sistemas son, sin embargo, insuficientes para el manejo de objetos tridimensionales complejos o difusos como corrientes de agua, frentes climáticos, estructuras geológicas, etc.

Los sistemas de CAD actuales que sí tienen características tridimensionales completas son también insuficientes para la gestión de estos geo-elementos complejos: pueden ser capaces de almacenarlos y representarlos pero no de realizar ningún análisis que los involucre. Actualmente existen sistemas —los llamados SIG tridimensionales— cuyo objetivo es tratar este tipo de elementos; gestionar los datos de las geo-ciencias de forma integral. Estos sistemas se encuentran en el extremo de la actual tecnología informática: requieren de los computadores más avanzados, son muy costosos y carecen de la madurez que ha alcanzado el desarrollo de los SIG tradicionales; la experiencia en su uso es todavía insuficiente, por lo que la funcionalidad idónea que debieran presentar todavía no se conoce. Así mismo las operaciones de análisis espacial tridimensional todavía no son suficientemente robustas y eficientes en tiempo como para poder hacer un uso cómodo y fiable de ellas.

Datos Geográficos.

En primer lugar cabe distinguir dos tipos fundamentales de información geográfica: las **entidades geográficas** y las **variables espaciales**. Si bien muchos autores engloban ambos tipos de datos como entidades geográficas, aquí las distinguiremos para clarificar el desarrollo. Las entidades geográficas son elementos que constan de una representación espacial—el fenómeno observable—además de una serie de atributos o valores asociados. Las variables espaciales son funciones que adoptan un valor en cada punto del terreno, como por ejemplo la elevación, la pendiente, tipo de suelo o una imagen de satélite. (En este último caso el valor en cada punto sería la intensidad de radiación reflejada en una banda de energía.) Las entidades geográficas se pueden clasificar por su representación espacial en puntuales, lineales y extensas.

Algunas variables espaciales en lugar de tener una variabilidad continua se agrupan en áreas de valor constante, como por ejemplo el tipo de suelo. Éstas variables pueden ser descritas por tanto por la colección de tales áreas, que como atributo tienen el valor de la variable sobre ellas. Éstas áreas se pueden considerar entidades geográficas, así que en algunos casos hay ambigüedad en cuanto a considerar un dato como entidad o variable. En todo caso no es ningún problema, pues por la naturaleza de esos datos no tiene mayor repercusión el considerarlos de uno u otro tipo.

Veamos algunos ejemplos de entidad geográfica:

Entidad	Expresión espacial	Atributos
Hospital	punto de su localización	nombre, plazas, etc.
Válvula en red de distribución	localización	características técnicas
Carretera	Línea de su eje	Nombre, Tipo de pavimento
Río	Líneas de sus márgenes	Nombre, Caudal medio
Línea de conducción	Línea	Identificación, Datos técnicos
Término Municipal	Area que comprende	Nombre, Datos demográficos
Espacio protegido	Area que comprende	Nombre,...
Lago	Area que comprende	Nombre
Manzana urbana	Area que comprende	Nombre, Datos demográficos

Funcionalidad requerida.

Para comprender la estructuración de la información que requiere el SIG vamos a establecer algunos ejemplos de operaciones que un SIG debe realizar sobre los datos. Éstas son algunas tareas típicas en un SIG:

Selección de datos o áreas de interés.

- Comencemos por algo sencillo: seleccionar aquellos municipios con menos de mil habitantes. En esta tarea interviene un tipo de entidades geográficas: los municipios. La tarea se puede realizar operando únicamente con los atributos de éstas entidades.
- Algo más complejo: ¿Que municipios son atravesados por la carretera *X*? Ahora se necesitan dos tipos de entidades geográficas: municipios y carreteras y la tarea involucra atributos (nombre de la carretera) y representación espacial.
- Otra selección en la misma línea sería determinar los municipios con más de dos institutos y población menor de *X*. Aquí se emplean las entidades geográficas municipios e institutos. El trabajo se puede realizar utilizando únicamente los atributos en el caso de que el nombre del municipio sea un atributo de los institutos; en otro caso se requerirá operar con los elementos espaciales para determinar que institutos se hallan dentro de cada municipio.
- Supongamos que se intenta determinar los propietarios afectados por una obra (carretera, edificación) para su expropiación, notificación, etc. O bien las conducciones (agua, gas, elect.) afectadas por la obra. Queremos

preguntar al sistema por las parcelas y/o conducciones que son atravesadas por la obra o se encuentran a una distancia determinada. Aquí se opera con entidades geográficas como carreteras, edificaciones, parcelas, conducciones, y se necesita un elemento espacial auxiliar: un *buffer* o área que dista menos de un valor dado de una determinada entidad geográfica.

- Para determinar el emplazamiento de una carretera deseamos determinar el impacto de diferentes proyectos. Por ejemplo determinar áreas a menos de 30 metros de la vía con cierto tipo de vegetación, uso, etc. O para situar una nueva sucursal de una entidad bancaria deseamos evaluar distintos emplazamientos en función de la renta de las zonas adyacentes, población, situación de otras entidades financieras. Éstas operaciones son similares a la anterior en cuanto que hacen uso de un buffer. Pero ahora la complejidad aumenta al necesitarse cruzar el buffer con varios tipos de elementos.
- Imaginemos que deseamos determinar las zonas propicias para el cultivo de viñedos: Queremos que el sistema muestre las zonas de tipo de suelo *X*, uso del suelo agrícola, pendiente no superior al 5% y orientada hacia el sol. Además posiblemente nos interesen sólo aquellas de determinados propietarios o municipios. Para ello se necesitan entidades geográficas (parcelas, municipios, uso del suelo) y algunas variables espaciales: tipo del suelo, pendiente, orientación. La complejidad de ésta operación radica en que se deben cruzar todos estos elementos para determinar las áreas que nos interesan.

Búsqueda de correlaciones.

- Para un estudio de mercado deseamos correlacionar la respuesta a un mailing directo con algunas variables demográficas y económicas disponibles, para discretizar el espacio en distritos a usar por el departamento de ventas y distribución. Se requiere la capacidad de aplicar operaciones estadísticas a variables espaciales.
- También resultan los SIG útiles en prospección, predicción, estudios naturales. Por ejemplo se pueden encontrar correlaciones entre atributos espaciales y la presencia de un determinado fenómeno. Así se determinó por ejemplo en un trabajo de la "*Royal Society for the Proteccion of Birds*" Se intentaba determinar la distribución de cierta especie de ave, trabajo que se estimó en 25 años. En la investigación preliminar se encontró una correlación entre una banda energética del Landsat y la localización del ave que resultó muy acertada, eliminando el tiempo previsto de estudio.

Otras Operaciones

- Para evaluar el potencial de polución de aguas freáticas podemos calcular un índice de riesgo, combinando los datos de profundidad a la capa freática, medio acuífero, tipo de suelo, pendiente del terreno, conductividad hidráulica del acuífero, etc. Se necesita una fórmula que modelice el fenómeno en función de las variables mencionadas y la capacidad de operar sobre una serie de variables espaciales para crear otra nueva.

- Para realizar el reparto diario a una serie de establecimientos, deseamos conocer la ruta óptima (minimizando tiempo, distancia o cualquier otro tipo de coste), contando con la red de vías de la ciudad, o bien de carreteras de la comarca. Aquí se necesitan funciones que, utilizando complejos algoritmos, operen sobre la red compuesta por un tipo de entidades geográficas: vías o carreteras. Son fundamentales las relaciones de conectividad entre los elementos.
- Para disponer un plan de emergencia deseamos saber la zona inundable alrededor de un río para un caudal determinado de éste..., o bien el área de influencia de un escape tóxico... Ésta es una tarea muy complicada que requiere un programa que implemente un modelo del fenómeno a simular. Se requiere, pues, un modelo físico–matemático eficientemente computable del sistema y su implementación en un SIG.

Bases de Datos Relacionales.

Veamos en primer lugar que forma tienen las bases de datos tradicionales, es decir, las que contienen únicamente atributos alfanuméricos.

Los SIG gestionan sus propias bases de datos, pero casi todos proveen algún tipo de comunicación o enlace con las bases de datos existentes para unir atributos alfanuméricos a los datos geográficos. El modelo de base de datos de más éxito en general, y que es el utilizado por casi todos los SIG es el modelo relacional, y es el que describiremos aquí. Se debe notar sin embargo el gran auge que las bases de datos orientadas a objetos están teniendo en los sistemas SIG. Sin duda son más adecuadas éstas a las necesidades del SIG pero en la actualidad no cuentan con la madurez suficiente. En cualquier caso las bases de datos orientadas a objetos no son incompatibles con las relacionales, y se pueden considerar una extensión de éstas.

Una base de datos relacional está compuesta por tablas o relaciones. Una tabla tiene una serie de atributos o columnas y está formada por tuplas o filas que tienen un valor para cada uno de los atributos. Veamos un ejemplo en que representamos mediante tablas los atributos de algunas entidades geográficas:

Tabla **Municipio**:

Nombre	Población
Zaragoza	586.000
Tudela	26.058
Pamplona	252.815
Utebo	7.580
Gallur	3.066
Perdiguera	480
Farlete	496

Tabla **Hospital**:

Nombre	Municipio	Plazas
Miguel Servet	Zaragoza	<i>nnnn</i>
Hospital Provincial	Zaragoza	<i>nnnn</i>
Reina Sofía	Tudela	<i>nnnn</i>
San Juan de Dios	Pamplona	<i>nnnn</i>
Virgen del Camino	Pamplona	<i>nnnn</i>

Una entidad de datos (como una entidad geográfica sin representación espacial) viene dada aquí por un conjunto de atributos que constituye una clave primaria, una

identificación única de la entidad. En el ejemplo, municipios y hospitales son entidades de datos cuya clave primaria es su nombre. Una tabla que contenga la clave primaria de una entidad y otras columnas, como las del ejemplo, proporciona atributos a las entidades. En el ejemplo se da el caso de que cada fila de cada tabla es la descripción completa de una entidad pero esto no es necesariamente siempre así.

Los componentes de la base de datos son pues muy simples, y las relaciones complejas se modelan a través de una serie de operaciones entre tablas. Éstas operaciones pueden estar disponibles como un álgebra relacional o a través de un lenguaje del cálculo relacional. Para el usuario lo más habitual es ésto último, y en concreto, el lenguaje más popular de este tipo es SQL—Structured Query Language. Así por ejemplo el siguiente predicado de SQL obtiene los nombres y número de plazas de los hospitales implantados en municipios de más de 100.000 habitantes:

```
Select Hospital.Nombre, Hospital.Plazas
From Hospital, Municipio
Where Hospital.Municipio=Municipio.Nombre
and Municipio.Población>100.000
```

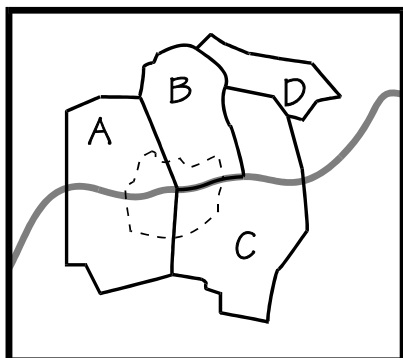
Esta arquitectura es muy flexible y potente. Sin embargo está limitada a la manipulación de valores numéricos y alfabéticos. No podemos por ejemplo determinar la longitud de carreteras que atraviesan un determinado municipio, o las empresas situadas a menos de 2Km de un río, etc. Un lenguaje como el SQL es una buena base para un lenguaje de acceso en SIG, pero necesita operadores y funciones que actúen sobre la parte espacial de los datos, para determinar por ejemplo si una carretera atraviesa un municipio o cual es su longitud.

Bases de Datos de CAD —Elementos Geométricos.

En un CAD cada archivo o diseño es una pequeña base de datos que consta fundamentalmente de elementos gráficos. Éstos se pueden clasificar en puntuales, unidimensionales, y bidimensionales. Fundamentalmente puntos, líneas o polígonos además de etiquetas textuales. (No consideramos aquí los elementos tridimensionales disponibles en los sistemas actuales) A éste tipo de elementos que se encuentran en CAD los llamaremos geométricos, y también se denominan habitualmente **vectoriales**, por estar definidos por una serie de vectores, como son la localización de los vértices de una línea quebrada. La información vectorial tiene la característica de conservar sus propiedades a cualquier escala de representación. Algunos sistemas de CAD permiten también incorporar en la base de datos gráfica imágenes digitales o “mapas de bits”, que son mosaicos de pequeños elementos de imagen rectangulares. La estructura de estas imágenes se denomina *raster* (“rejilla”). En general son estructuras raster las que discretizan el espacio en una retícula regular y almacenan de alguna manera el contenido de cada rectángulo. La distinción entre estructuras vectoriales y *raster* es muy habitual en relación al tratamiento de datos espaciales o gráficos. Por ejemplo la imagen de un monitor o la salida de impresoras matriciales, de chorro de tinta o láser son elementos *raster*, por estar compuestos de pequeños elementos o píxeles, mientras que la forma de trabajo de un trazador de plumillas es vectorial.

Las bases de datos CAD están por lo general particionadas horizontal y verticalmete. Horizontalmente en el sentido de que cada archivo de CAD comprende un diseño o dibujo; una (pequeña) área rectangular u **hoja**. Verticalmente los elementos gráficos estan compartimentadas en **capas** o niveles: cada elemento está en una única capa.

Éstos elementos gráficos nos servirían para la descripción espacial de entidades geográficas pero no de variables espaciales. La división en capas del CAD puede servir para diferenciar los distintos tipos de entidades (munipios, ríos, espacios naturales en la ilustración). La separación en capas nos obligaría a duplicar una línea que, como en el ejemplo, sea a la vez límite de un municipio y curso de un río. La división en hojas nos obliga a delimitar claramente el área de estudio e incluir en el mismo archivo CAD todos los datos de interés.



Un mapa como éste es representado en CAD como un conjunto de líneas y textos.

Aunque no es fundamental en CAD, casi todos los sistemas avanzados permiten que éstas entidades gráficas tengan algún atributo alfanumérico asociado. En particular pueden tener un identificador que sirva de enlace con tablas de datos.

Sin embargo las bases de datos CAD no cuentan con posibilidades de acceso análogas a las de las bases de datos relacionales. Habitualmente no permiten realizar una interrogación con criterios espaciales para determinar que objetos gráficos la cumplen, y mucho menos incluir condiciones sobre los datos alfanuméricos que puedan estar asociados a

los gráficos. Las necesidades de un SIG se parecen conceptualmente mucho más a las de una base de datos que a las de un sistema CAD.

Elementos Geométricos con Atributos.

La idea presentada antes de unir atributos alfanuméricos a elementos gráficos del tipo de los usados en CAD puede ser suficiente para diseñar una sencilla base de datos geográfica si se establece un sistema de archivo que gestione las particularidades de las coordenadas geográficas y no se estructura la información en hojas y capas, sino en conjuntos de datos superponibles.

En una estructura de este tipo cada entidad geográfica estará representada por un elemento geométrico y los atributos alfanuméricos asociados.

Los distintos tipos de entidades geográficas estarían representados por conjuntos de datos con atributos y geometría homogéneos. Así por ejemplo el conjunto de los términos municipales (un tipo de entidad geográfica) tendrían una serie de atributos comunes (nombre del municipio, población, etc) y una representación geométrica homogénea (polígonos que encierran el término).

Tablas de datos con objetos espaciales.

Ésta sencilla estructura de datos (atributos con geometría) puede servir de implementación para un modelo de base de datos geográfica que nace de manera natural de ella y que resulta muy útil por su sencillez y flexibilidad.

El modelo que vamos a describir es una extensión natural de una base de datos relacional. La información estará estructurada en (la veremos como) tablas al igual que en el modelo relacional. Las tablas corresponden a los conjuntos de datos homogéneos descritos antes. A los tipos de atributos habituales —números y textos— que aparecen en las tablas se añade aquí uno nuevo: el objeto espacial, de forma que cada tabla pueda tener como máximo una columna de este tipo, que representa el elemento geométrico asociado a los atributos, es decir la parte espacial de la entidad geográfica. De ésta manera una tabla puede contener elementos geográficos completos, incluyendo atributos y geometría. Por ejemplo:

Tabla **Municipio**:











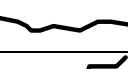

Nombre	Población	Objeto
Zaragoza	586.000	
Tudela	26.058	
Pamplona	252.815	
Utebo	7.580	
Gallur	3.066	
Perdiguera	480	
Farlete	496	

Tabla **Carretera**:

Nombre	Objeto
N232	
A2	
C-137	
C-124	
A68	

Nótese que esta estructura sigue siendo insuficiente para la gestión de variables espaciales con variación continua como elevación, imagen de satélite, etc. En todo caso se podrían añadir tablas de tipo especial cuya representación espacial fuera una estructura raster, como puede ser un imagen de teledetección.

Para ser útil esta estructura se debe contar con un medio de acceso que combine los atributos y la geometría, teniendo en cuenta que los objetos tienen como referencia común un espacio geométrico, y por tanto existen relaciones implícitas entre los objetos espaciales: una carretera atraviesa un municipio, un municipio contiene un hospital, una carretera enlaza con otra, etc. La forma habitual y más versátil de proporcionar esta funcionalidad es mediante un lenguaje de acceso que proporcione operadores y funciones espaciales. Habitualmente este lenguaje es una extensión del SQL visto antes.

Dado que se ha incorporado el objeto espacial como un nuevo tipo fundamental de datos bastaría extender el conjunto de funciones y operadores disponibles en el lenguaje para operar sobre datos numéricos y textuales con otros operadores y funciones que operen sobre objetos espaciales. Las funciones pueden extraer atributos como la longitud de un objeto lineal o el área de uno extenso, o pueden construir un objeto que sea la unión o intersección de otros dos, etc. Se pueden definir operadores lógicos que determinen si dos objetos se intersectan, uno está incluido en otro, etc.

Así por ejemplo para seleccionar el nombre y longitud de las carreteras que atraviesan el municipio de Zaragoza:

```

Select Carretera.Nombre, len(Carretera.Objeto)
From Carretera
Where Carretera.objeto intersects
        (Select Objeto From Municipio
         Where Nombre="Zaragoza")

```

Aquí “len” es una función que opera sobre un objeto, calculando su longitud; “intersects” es un operador lógico que determina si un objeto interseca o corta a otro.



Representación de un mapa con municipios, un río y un espacio natural en el modelo de tablas con objetos.

Con esta estructura, el usuario que visualiza y analiza los datos trabaja sobre elementos de la máxima sencillez. La abstracción de manipular la expresión espacial de los datos como objetos gráficos—un tipo más de atributo—permite un modelo sumamente simple y útil. Independientemente de la estructura subyacente de los datos la visión de éstos como tablas es la forma ideal para un sistema de análisis y visualización como son los llamados *Desktop Mapping*.

Sin embargo la estructura que hemos descrito para dar soporte a estas “tablas con objetos” limita las posibilidades del sistema y en algunos casos afecta a su eficiencia. Para que las operaciones espaciales, esto es, la determinación de las relaciones geométricas intrínsecas de los objetos sean eficientes, las tablas tienen que estar dotadas de un sofisticado mecanismo de indexación espacial.

Limitaciones.

De los requerimientos funcionales que hicimos a un SIG quedan algunos sin resolver, o cuando menos su eficiencia temporal es muy limitada o se requiere una considerable programación. Algunas limitaciones se deben a la estructura descrita y otras al modelo de datos.

El modelo visto no soporta de forma natural variables espaciales, es decir, datos como la altimetría o imágenes de satélite. Así no resultan realizables, o bien requieren comandos u operaciones especiales, —no integradas en el modelo— las tareas que

involucran el cruce de variables espaciales —para determinar áreas de interés, o para crear nuevos datos combinando variables—.

La estructura de datos no mantiene de forma explícita las relaciones entre los objetos —un objeto es vecino de otro, está contenido, lo cruza, etc— hay tareas que se dificultan o necesitan de estructuras internas auxiliares que permitan acceder eficientemente a algunas de estas relaciones. Las tareas que resultan difíciles en este sentido son las que necesitan acceder a un gran número de relaciones entre objetos (relaciones de todos los objetos con todos los demás) ya que si sólo se requieren relaciones de un objeto (incluso con todos los demás) el uso de índices espaciales puede resolver eficientemente el problema. Ejemplos de tales tareas son operaciones complejas sobre redes lineales, como la búsqueda de caminos mínimos, determinar que llaves aislan una sección, simular los cambios de presión al cerrar una válvula, generación de viajes en una red viaria, etc. También resulta difícil el cruce de distintas clases de objetos, o algunos modelos físicos o simulaciones (clima, hidrología, contaminación). Estas limitaciones se pueden solventar mediante estructuras de datos espaciales más complejas, manteniendo no obstante el modelo de trabajo de tablas con objetos espaciales; no es por tanto una limitación inherente al modelo descrito.

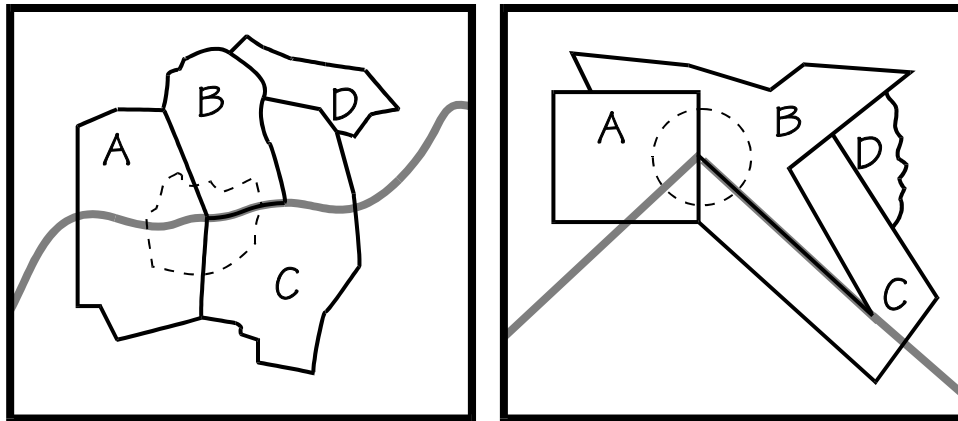
Otra limitación se desprende del hecho de que esta estructura carga toda la gestión de los elementos espaciales al usuario. Esto no supone ningún inconveniente para los elementos puntuales, pero sí para los lineales y extensos. Si el usuario desea modificar la delimitación de un término municipal debe manipular el o los polígonos que lo definen. Pero además debe ocuparse de modificar todos los elementos colindantes con él que lo requieran, como los municipios vecinos, límites comarcales, provinciales, autonómicos, nacionales, pero posiblemente no el curso de un río que coincidía con el antiguo término. Esta limitación, aunque no es estrictamente intrínseca al modelo de tablas con objetos, no resulta fácil de obviar sin complicar el modelo; por lo común la solución involucra no solo estructuras de datos más sofisticadas, sino además que el usuario acceda a ellas y las manipule interactivamente.

Estructura de las Bases de Datos Geográficas con Topología.

Un SIG requiere una estructura de datos que permita realizar eficientemente todas las operaciones detalladas más arriba así como la gestión completa de la base de datos geográfica, combinando de forma homogénea conjuntos de entidades geográficas y variables espaciales.

Geometría y Topología.

Para poder responder eficientemente a las cuestiones que involucren la relación topológica entre elementos (qué elementos colindan con uno dado, cuales lo cruzan, ...) así como para optimizar el almacenamiento de la geometría (límites municipales coincidentes...) es necesario que la base de datos contenga de forma explícita estas relaciones. Así pues aparecen dos estratos en la parte espacial de la base de datos: el dominio geométrico que contiene la descripción de los objetos espaciales, y el dominio topológico que contiene las relaciones entre éstos.

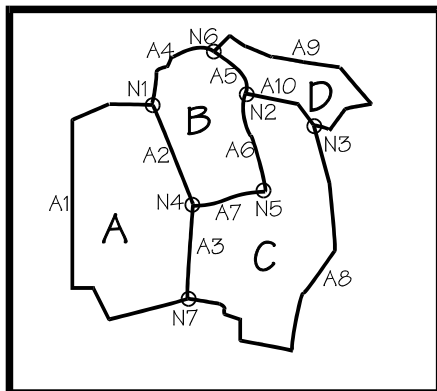


Los dos mapas constan de una serie de polígonos de línea continua (municipios), otro de línea a trazos (espacio protegido) y una línea mas ancha (un río). Los elementos que los componen son arcos (líneas geométricas) que hemos dotado de simbología para su representación gráfica. Cada mapa es geoméricamente diferente, pero ambos son topológicamente equivalentes; la topología implícita en cada una de las geometrías es la misma. (por ejemplo, el area D colinda con B y C, el río atraviesa A y C y delimita parte de B, etc)

Supongamos que queremos responder a una cuestión del tipo de la sexta planteada anteriormente; necesitamos la delimitación de los términos municipales, la de un espacio natural y los cursos de agua, para saber qué municipios se ven afectados por el espacio, por los cursos, etc. En primer lugar tendremos una serie de líneas que componen estas delimitaciones. Estas son entidades geométricas, pero no geográficas. La única información asociada a la geometría es el tipo de elemento representado: límites administrativos, de espacio natural, un curso de agua,... Estos tipos de elementos geométricos se denominan **clases** (de elementos). La mejor forma de clasificación de la geometría es que cada elemento pueda pertenecer a una o más clases; así, en el ejemplo anterior se observa una línea que es a la vez división municipal y curso de agua. Sin embargo algunos sistemas SIG compartimentan la geometría en clases exclusivas (capas, coberturas), obligando, en casos como el anterior, a duplicar algunos elementos. Esta situación plantea dificultades a la hora de modificar la geometría ya que el operador debe tener en cuenta las coincidencias de elementos y reproducir los cambios en todos.

Para responder al tipo de cuestiones planteadas antes es necesario que el sistema componga la estructura topológica de esta geometría, y es en el dominio topológico donde las entidades geográficas tienen su expresión espacial. En nuestro caso la conexión de las líneas que son límites administrativos —entidades geométricas— delimita una serie de áreas —entidades topológicas— que son la expresión espacial de los términos municipales—entidades geográficas.

Como ejemplo, aquí tenemos la forma en que se estructuran las líneas que delimitan unos polígonos o áreas:



Topología de los límites municipales.

Arco	Nodo Inicio	Nodo Fin
A1	N7	N1
A2	N1	N4
A3	N4	N7
A4	N1	N6
A5	N6	N2
A6	N2	N5
A7	N5	N4
A8	N3	N7
A9	N6	N3
A10	N2	N3

Arco	Pol. Dcha.	Pol. Izda.
A1	A	—
A2	A	B
A3	A	C
A4	B	—
A5	B	D
A6	B	C
A7	B	C
A8	—	C
A9	D	—
A10	C	D

Pol.	Arcos
A	A1, A2, A3
B	A4, A5, A6, A7, -A2
C	-A7, -A6, A10, A8, -A3
D	A9, -A10, -A5

Una geometría vectorial como la del ejemplo está compuesta por líneas y puntos. Como paso previo a la estructuración topológica el sistema dividirá las líneas en **arcos** asegurando que los arcos que se van a estructurar conjuntamente solo interseccionan entre sí en los extremos, llamados **nodos**. Un tipo especial de elemento puntual existente en la geometría son los puntos de etiquetado, mostrados en el ejemplo: son entidades gráficas con texto que sirven para proveer de identificadores a los elementos topológicos que el sistema compondrá (polígonos en nuestro caso.) Estos identificadores serán el nexo de unión de los elementos espaciales con los atributos alfanuméricos. De no existir los puntos de etiquetado, los identificadores de la topología deben ser introducidos interactivamente por un operador, o bien ser “inventados” por el sistema.

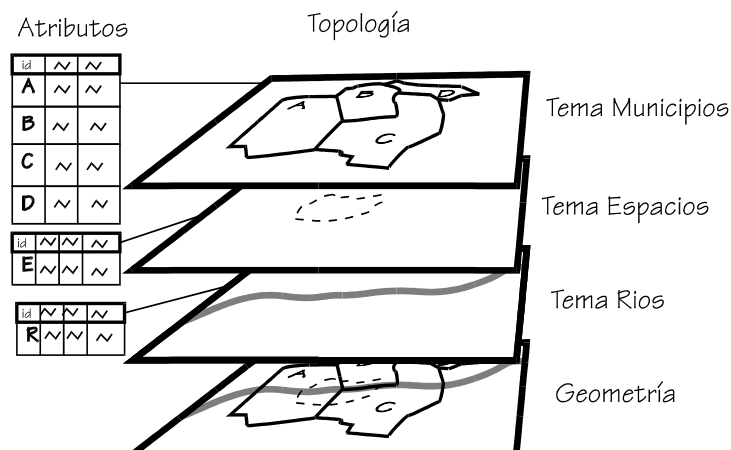
Los elementos topológicos que pueden construirse sobre la base de una geometría de líneas y puntos como la vista pueden ser de tres tipos:

- Áreas extensas (polígonos o regiones): Están delimitadas por arcos que componen su frontera. Éstos arcos establecen la conectividad entre las áreas: cada arco separa dos regiones colindantes.
- Líneas (redes lineales): Compuestas por arcos conexos; conectadas por nodos, que pueden unir un número arbitrario de arcos y por tanto de líneas.
- Entidades puntuales: Su topología es trivial (inexistente). La composición topológica únicamente les asigna un identificador.

El Tema.

Hemos distinguido dos estructuras espaciales. Una es un conjunto de elementos geométricos y la otra es el conjunto de elementos que sus relaciones topológicas delimitan. Cada uno de estos elementos tendrán un identificador único asociado que permitirá asociarle atributos contenidos en tablas de datos como las vistas antes. Esta conjunción de elementos espaciales y atributos es lo que se llama **tema**. Un tema es pues una colección de entidades geográficas homogéneas. La representación espacial

puede ser, como hemos visto, mediante elementos estructurados topológicamente sobre una geometría vectorial; en tal caso el tema se llama **tema vectorial**.



Cada tema vectorial se basa en parte de la geometría (los elementos pertenecientes a una o más clases). Así la misma geometría puede ser compartida por varios temas; en el ejemplo se aprecia que un arco de la geometría pertenece a la clase hidrografía, constituyendo parte de un río pero también pertenece a los límites administrativos, delimitando dos municipios.

Usualmente los temas vectoriales se pueden clasificar en poligonales (o de áreas), lineales (o en red) y puntuales, según el tipo de los elementos topológicos que lo componen.

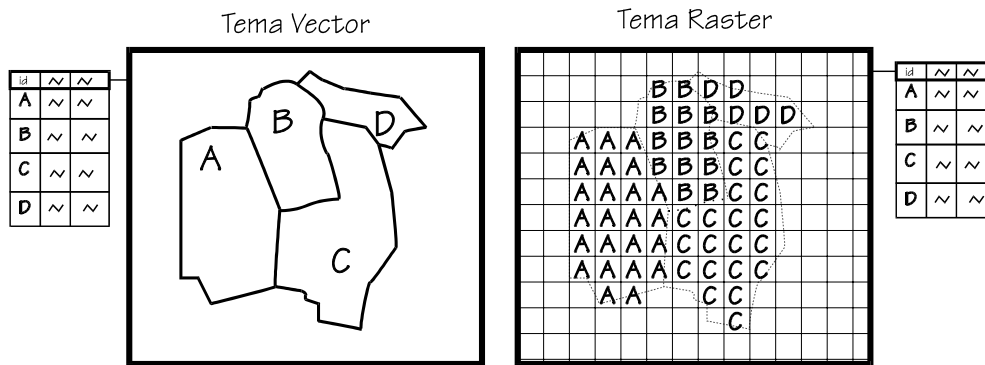
Un tema cuyas entidades cubran totalmente una parte del espacio sin solapamiento, como tipos de suelo, puede considerarse una variable espacial. El valor de la variable en un punto es el de los atributos de la entidad que contiene al punto.

Además de la estructura vectorial con topología, caben otras posibilidades en cuanto a la estructura de los temas, dictadas por la necesidad de gestionar variables espaciales del tipo de altitud, pendiente, imagen de satélite, etc.

Temas “Raster”.

Los datos provenientes de teledetección y fotogrametría son de tipo *raster*: imágenes constituidas por pequeños elementos regulares; pequeños rectángulos dispuestos en filas y columnas. En cada uno de éstos elementos se dispone de un valor (intensidad de la imagen), de forma que constituyen auténticas variables espaciales. La forma natural de almacenar esta información es como una matriz que contenga los valores de cada celdilla. Una tal matriz representando valores de una variable discretizada en elementos regulares es una estructura raster genérica tal como se emplea en un SIG.

Una estructura raster se puede usar para contener cualquier variable espacial y se usa de hecho como implementación alternativa a la vectorial para los temas. Un tema de este tipo se llama **tema raster**. El espacio es discretizado en pequeños rectángulos o cuadrados, de forma que el tamaño que tienen estos elementos es fundamental y determina la resolución del tema.



Un tema raster se puede utilizar no sólo para representar una variable espacial, sino cualquier tipo de entidad geográfica, pero normalmente con la limitación de que las entidades no se solapan. En cada celdilla se almacenaría el identificador de la entidad presente sobre ella. Esta no es la forma natural para almacenar datos como los términos municipales, pero es útil para determinados cálculos debido a la sencillez de proceso que supone el combinar o cruzar varios temas de tipo raster como veremos más adelante.

Históricamente se han diferenciado SIG's vectoriales y raster. Actualmente los principales sistemas SIG combinan ambos tipos de estructura.

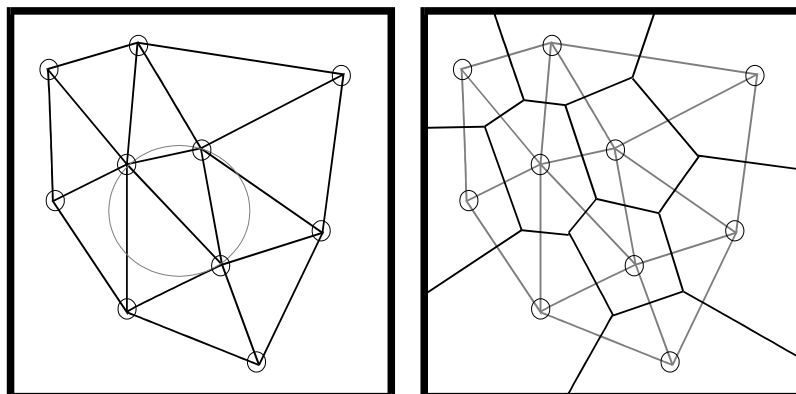
Temas "TIN".

Otra estructura que se utiliza para implementar variables espaciales es el **TIN**—*Triangulated Irregular Network* (Red de Triangulación Irregular), diseñada en 1978 por Peucker y sus colaboradores. Es una estructura vectorial topológica similar a la de los temas vectoriales, con la excepción de que las áreas siempre son triángulos y no se necesitan gestionar islas o agujeros. Se usa fundamentalmente cuando la variable espacial se debe interpolar a partir de una colección de muestras puntuales, como por ejemplo altitudes. El TIN mantiene todos los puntos muestrales, y puede incorporar otros rasgos importantes como líneas de ruptura en el caso de la altimetría. Se adapta eficientemente a la densidad de detalle de cada área, a diferencia de las estructuras raster que contienen una densidad de detalle uniforme en todo el área. Algunas operaciones se pueden realizar eficientemente sobre un TIN. Por ejemplo operaciones usuales sobre la altimetría como cálculo de pendientes, visualización tridimensional, cálculo de perfiles. En cambio su complejidad debido a que contiene explícitamente las relaciones topológicas dificulta su gestión, y algunas operaciones disponibles en raster no son posibles actualmente. Partiendo de un TIN se puede interpolar eficientemente una matriz de valores (estructura raster) del tamaño deseado. Así, además de servir como implementación de una variable espacial, el TIN sirve de estructura intermedia para interpolar un tema raster a partir de muestras puntuales.

Cuando las muestras corresponden a una variable discreta (como el tipo de suelo) la estructura que se crea no es una triangulación basada en los puntos sino su dual: una colección de polígonos que cubren el espacio de forma que cada polígono contenga uno de los puntos muestrales. El valor del atributo dentro de cada polígono se considera constante. Una partición de este tipo del espacio se denomina teselación. La más útil es la teselación de Dirichlet (también diagrama de Voronoi o polígonos de Thiessen). La propiedad fundamental de esta teselación es que el polígono correspondiente a un punto muestral contiene el área que está más cercana a ese punto central

que a cualquier otro. Así al utilizar esta estructura como variable espacial asignamos a cada punto el valor del punto muestral más cercano.

La triangulación dual de la teselación descrita es la triangulación de Delaunay, triangulación basada en los puntos que tiene la propiedad de que el círculo circunscrito de cualquiera de sus triángulos no contiene ningún otro punto muestral. Esta triangulación es usada habitualmente para los TIN.



A la izquierda tenemos la triangulación de Delaunay para 10 puntos y el circun-círculo de uno de los triángulos. A la derecha vemos los polígonos de Thiessen para los mismos puntos.

Si se prevee trabajar con muestras de variables discretas como el tipo de suelo, será conveniente que el SIG disponga de herramientas para crear un tema vectorial de áreas a partir de un conjunto de puntos muestrales, que sea la teselación de Dirichlet de éstos.

Temas MDT

Cuando un tema representa la altimetría del terreno (una variable espacial) se suele denominar MDT (modelo digital del terreno) o MDE (Modelo digital de elevaciones). La implementación puede ser raster o TIN. Un MDT puede ser creado si se disponen de datos de altimetría como pueden ser las curvas de nivel o puntos acotados. Para generar el MDT se necesitan herramientas que generen un tema raster o bien un tema TIN a partir de los datos de altimetría. También es frecuente que los SIG cuenten con otras herramientas especializadas para tratar los MDT. Por ejemplo utilidades para preparar vistas perspectivas del terreno, en las cuales un tema que contenga la altimetría determina la forma del terreno mientras que otros que contengan planimetría o una imagen de satélite se pueden proyectar—*drapping*—sobre él. Otras utilidades pueden generar un tema de pendientes a partir de un MDT, o calcular cortes del terreno—perfiles—o bien calcular el área visible desde un punto—cuencas visuales—, cuencas hidrográficas, etc.

Temas, Temas y más Temas.

De esta forma el tema es un concepto unificador entre los distintos tipos de datos geográficos por una parte y entre las diferentes estructuras de datos que les dan soporte por otra.

El tema, conjunto de entidades geográficas homogéneas, es el equivalente en una base de datos geográfica a las tablas de las bases de datos tradicionales. Contiene todo los elementos para que se pueda interrogar eficientemente; cada elemento tiene una serie de atributos como en una tabla, tiene una relación topológica con los otros elementos del mismo u otros temas que compartan la misma geometría y tiene una expresión geométrica que permite cuantificar su área, perímetro longitud, posición, etc.

Análisis de superposición.

La operación mas importante que realiza un SIG es el análisis de superposición. En éste dos o más temas (e.g. tipo suelo, pendiente) se combinan para crear otro intersección de ambos. (Tradicionalmente este proceso se ha hecho superponiendo hojas transparentes). El nuevo tema contendrá las intersecciones de los elementos de los temas combinados, combinando también los datos asociados. Es un proceso que involucra las geometrías y las topologías. En la topología que se deriva cada elemento tendrá asociados los atributos de todos los elementos que lo han generado (tipo suelo X , pendiente Y). Esta topología derivada puede ser interrogada para responder a preguntas como ¿Qué áreas tienen tipo X y pendiente $<Y$? Las geometrias combinadas se pueden usar para cuantificación de elementos, como la superficie de pendiente $<Y$ % dentro de cierto municipio, o la longitud de un río que atraviesa un municipio.

La superposición es un trabajo computacionalmente muy costoso cuando los temas están en formato vectorial, e involucra habitualmente la interacción de un operador. En cambio se realizan muy rápidamente y de forma automática si los temas son *raster*, pero el resultado estará afectado de un error debido a la discretización.

Creación de nuevos temas.

A veces, combinando temas existentes se intenta crear una información completamente nueva; Hemos mencionado ya la utilidad de herramientas que permitan calcular un tema de pendientes a partir de uno de elevaciones. En general un conjunto de herramientas que permitan operar flexiblemente sobre los temas para generar otros son muy útiles. A estos procesos se les suele denominar **modelado cartográfico**.

Es el caso de un modelo basado en parámetros existentes del terreno (pendiente, tipo, vegetación, elevación, etc.) que nos permita evaluar, por ejemplo un índice de erosionabilidad, riesgo de incendio, de contaminación etc. Las herramientas necesarias para hacer este tipo de análisis deben permitir que en una superposición como la vista antes los atributos se combinen según una determinada fórmula. Por ejemplo asociados a los temas a combinar podemos tener determinadas valoraciones numéricas y estas se pueden combinar con distintos pesos para el cálculo del valor deseado. Normalmente no se cuenta con un modelo perfecto para la combinación de los temas sino que bien se debe ajustar a una determinada región o coyuntura, o bien se debe crear completamente. Así pues es un trabajo de refinamiento interactivo, en el que se harán análisis *raster* —poco precisos— para ajustar el modelo, y luego se calculará con precisión en forma vectorial.

Visualización.

Para poder representar gráficamente los elementos espaciales de un tema vectorial, sea sobre un monitor gráfico o sobre papel (lo cual es una operación básica de

análisis), se les ha de asignar una determinada simbología. Por simbología entendemos aquí los atributos gráficos como el tipo (forma) de las líneas, su grosor, el color, las tramas de relleno, la forma de los símbolos, etc. Así en los ejemplos anteriores se han representado municipios, espacios y ríos con tres tipos de línea diferentes. La simbología no forma parte de los elementos espaciales, si bien cada elemento geométrico puede tener asociada una simbología por defecto. No consideraremos la simbología como parte de la base de datos geográfica. La simbología se elige en el momento de realizar una representación gráfica (mapa) y puede hacerse genéricamente en función de la geometría o del tema, de cuatro maneras fundamentales:

- Simbología por defecto de la geometría: Cada elemento se representa con una tipología propia por omisión, almacenada junto con la geometría. Este tipo de representación no tiene especial interés y no se encuentra en todos los sistemas.
- Simbología por clase: Es una forma de representación en función de la geometría, representando homogéneamente todos los elementos de una misma clase. Suele ser la forma básica de ver la geometría; Se utiliza una tabla de simbología, que puede ser particular para un usuario o para un trabajo, que asocia las clases que se desean representar con la simbología que las distingue.
- Simbología por tema: Los elementos de un conjunto de temas seleccionados se presentan con grafismo uniforme para cada tema. Esta forma de representación es análoga a la anterior y se puede asimilar a la anterior ya que es corriente que un tema utilice elementos geométricos de una sólo clase, por lo que un sistema puede facilitar sólo uno de los dos métodos.
- Simbología en función de los atributos: Se establece para uno o más temas la simbología en función de los atributos de cada elemento. Ésta constituye una representación temática, o **mapa temático** y admite múltiples variantes. Los valores de los atributos que van a determinar el grafismo pueden ser individuales, continuos o agregarse en intervalos de valores. Éstos valores pueden controlar aspectos como el color (o atributos individuales de éste), grosor y tipo de las líneas, forma y tamaño de los símbolos, o incluso pueden parametrizar la generación de símbolos especiales que sean pequeños gráficos locales a cada elemento.

Materialización de las Bases de Datos Geográficas.

En la sección anterior hemos generalizado una estructura de base de datos geográfica. Ahora discutiremos las formas en que esa estructura se puede materializar en los sistemas concretos existentes. Distinguiremos dos apartados en la materialización: implementación y modelo de interacción. Por implementación entenderemos la forma en que un SIG determinado hace efectiva la base de datos sobre el soporte de un sistema operativo. Independientemente de esta implementación el SIG presentará de una manera determinada los datos al usuario, formando un modelo o paradigma con el que interactuar. Ya hemos discutido un modelo tal al hablar de la forma de ver los datos geográficos en tablas relacionales con objetos espaciales como un atributo más.

Implementación.

Las estructuras para una base de datos geográfica descritas se pueden implementar de infinitas maneras; cabe distinguir dos estrategias fundamentales, partiendo del supuesto de que para la gestión de los atributos alfanuméricos se utilizará un sistema convencional de bases de datos, siendo este habitualmente alguno de los ya disponibles en el mercado.

Sistemas Integrados: llamaremos así a aquellos en que el almacenamiento de los datos espaciales se implementa sobre una base de datos convencional al igual que los atributos. Hasta ahora los sistemas integrados han utilizado bases de datos relacionales, que de forma conceptual parecen adecuadas para representar la estructura topológica y las coordenadas de la geometría. En la práctica esto plantea serios problemas de eficacia, especialmente en los accesos a los datos por criterios espaciales. Tiene la ventaja de mostrar automáticamente una visión homogénea de toda la base de datos geográfica. Debido a las dificultades pocos sistemas han elegido hasta ahora esta estrategia, como *System 9* de *Prime*. Actualmente los sistemas integrados están cobrando mayor relevancia, con el advenimiento de las bases de datos orientadas a objetos, que reúnen las condiciones necesarias para almacenar y gestionar eficientemente tanto los datos alfanuméricos como los espaciales. El sistema *SmallWorld*, de *SmallWorld Systems Ltd.* utiliza esta técnica, estando todos los datos soportados por una base de datos orientada a objetos que además presenta otras innovaciones muy importantes en el sistema de control de la concurrencia.

Sistemas Híbridos: para optimizar el rendimiento del acceso y análisis espacial, estos sistemas implementan las estructuras espaciales usando directamente el sistema de archivos del sistema operativo. Estos archivos de formato propio al sistema suelen incorporar sofisticados mecanismos de acceso espacial. La mayor parte de los sistemas de SIG se han diseñado según este esquema.

Paradigmas de Interacción.

Independientemente de la implementación que se haga de la base de datos geográfica, el sistema presenta una visión de ésta al usuario, que constituye un modelo en el que el usuario puede acceder y manipular los datos. Vamos a repasar los principales modelos de interacción disponibles en los SIG, dejando de lado modelos poco consistentes particulares a un sistema determinado.

El concepto de tema que hemos definido es análogo al de tablas con objetos espaciales descrito más arriba. Así un SIG puede presentar al usuario una visión de los temas como tablas con objetos espaciales, posibilitando una explotación sencilla de los temas mediante un lenguaje como el descrito. Un sistema que presenta esta visión es *MapInfo*, de *MapInfo Corporation*, que no tiene estructura topológica explícita —se encuadra en los llamados *Desktop Mapping*— pero sí cuenta con una estructura espacial avanzada para realizar búsquedas espaciales eficientes, economizar espacio en disco y realizar algunas operaciones de análisis geográfico.

Algunos sistemas presentan una visión completa de los temas, incluyendo la estructura topológica y geométrica de éstos, como colección de tablas relacionales. Así se permite en control total de la base de datos dentro de un modelo homogéneo y bien conocido. El coste de este control total es una elevada complejidad de uso. Este modelo es implementado por ejemplo por *EPS Prime Meridian*, de *Essential Planning Systems Ltd*, un sistema con topología y temas de todo tipo que introduce la

noción de servidor de datos geográficos: actúa como un servidor de base datos geográficos a través del lenguaje S²QL. Nótese que los sistemas de implementación integrada, presentan automáticamente este modelo de uso.

Es más usual sin embargo que el usuario no disponga de una visión homogénea en tablas de la base de datos geográfica completa. Normalmente se puede acceder como tablas relacionales a los atributos alfanuméricos de las entidades y a tablas de enlace con los elementos espaciales, que contienen los principales atributos geométricos de éstos. Para acceder a la parte espacial con detalle y operar sobre ella se necesitan procedimientos especiales, ajenos al sistema relacional. Tal es la forma de uso de *ArcInfo*, de *Environmental Systems Research Institute*. Éste último sistema, estructura la información espacial en **coberturas** que compartimentan la geometría y contienen la topología de ésta. Las coberturas de ArcInfo se pueden identificar con temas (con su geometría propia) aunque en rigor una cobertura puede contener un tema poligonal o uno puntual mas otro lineal.

Resumen.

Los datos geograficos son entidades geográficas (un río, un término municipal, un hospital, etc) y variables espaciales (elevación, pendiente, etc)

Las bases de datos relacionales presentan un modelo simple y muy potente de uso de los datos alfanuméricos; todos los datos se estructuran en tablas o relaciones de atributos de tipo textual y numérico.

Los sistemas CAD se limitan a gestionar objetos vectoriales.

Nótese que la simbología (grafismo) que se pueda aplicar a los objetos vectoriales no forma parte de una base de datos geográfica.

El enlace entre objetos vectoriales y atributos en tablas de datos puede resultar una forma sencilla de base de datos geográfica para aplicaciones limitadas.

Para los usos más genéricos de las bases de datos geográficas esta estructura no basta. Cuando se requiere la gestión integral de una base de datos geográfica, con capacidad de creación y combinación de los datos, y posibilidad de operaciones complicadas es necesaria la estructuración topológica de los datos geométricos.

El modelo de datos de un SIG, basado en temas o coberturas cubre todas éstas necesidades. Los temas tienen atributos como las tablas y tienen también una expresión espacial.

En cuanto a su estructura los temas pueden ser:

- Temas vectoriales: contienen objetos como áreas (términos municipales, tipo de suelo), líneas (red de conducción, carreteras) estructurados topológicamente.
- Temas raster: contienen la información discretizada en una cuadrícula. Están caracterizados por el tamaño de los elementos. No tienen estructura topológica pero se pueden combinar y modelar muy eficientemente.
- Temas TIN: es un tipo de tema vectorial que contiene una triangulación sobre un conjunto de puntos o muestras. Útil para interpolar una variable continua a partir de los puntos.

En cuanto a su contenido, un tema puede contener entidades geográficas o variables espaciales indistintamente.

Un caso particular de tema es el MDT: representa la elevación; una variable espacial. Su estructura suele ser raster o TIN.

La estructura en temas de una base de datos geográfica se materializa en una implementación y un modelo de utilización.

La implementación puede ser integrada, cuando un sistema de base de datos tradicional se emplea para almacenar todos los datos, o híbrida si los datos espaciales se almacenan en archivos propios.

En cuanto al modelo de utilización, el más sencillo para el usuario es el modelo relacional extendido con objetos espaciales como un atributo mas, interpretando los temas como tablas.

Otro modelo, más complejo pero homogéneo y coherente, es el que presenta toda la información como tablas de datos alfanuméricas.

Por último, el modelo mas habitual pero menos útil sería un modelo híbrido en el que la gestión de los datos espaciales es independiente de los atributos mediante un sistema de comandos propio.

Bibliografía.

Bosque Sendra, J., 1992, Sistemas de Información Geográfica: Rialp, S.A., Madrid.

Burrough, P.A., 1986, Principles of Geographical Information Systems for Land Resources Assesment: Clarendon, Oxford.

Calvo Melero, M., 1993, Sistemas de información geográfica digitales. Sistemas geomáticos: IVAP, Vitoria.

Healeym, R.G., 1991, Database Management Systems: Longman Group, Essex.

Sacks–Davies, R., McDonell, K.J., Ooi, B.C., 1987: GEOSQL—a query language for geographical information systems. Royal Melbourne Institue of Technology, Melbourne.