

Una categorización de las principales funciones de establecimiento y explotación de sistemas de información geográfica

Juan Antonio CEBRIÁN DE MIGUEL

Introducción

En los últimos cinco años, y especialmente en la literatura anglosajona, se ha generalizado definitivamente el uso del término sistema de información geográfica (o *GIS* simplemente) como denominación de base de datos computerizada que contiene información espacial. Conceptos semejantes, tales como base de datos georreferenciados o base de datos espaciales, han caído prácticamente en desuso.

No es este el momento de tratar por extenso el desarrollo histórico de este nuevo modo de manipulación y archivo de la información espacial, pero sí que es necesario señalar que los sistemas de información geográfica cuentan ya en su haber con una tradición de más de veinte años. En el año 1962, en Canadá, se diseña el primer sistema de información geográfica, destinado en este caso al mantenimiento de un inventario de recursos naturales a escala nacional (Tomlinson, 1984).

En las últimas dos décadas el desarrollo en este campo ha sido considerable, interesando, en primer lugar, a los organismos administrativos, pero también a la investigación universitaria y, más recientemente, a empresas comerciales de software y/o hardware. El principal factor de modificación en este proceso lo constituye, sin duda, la continua innovación tecnológica en el campo de la informática (cuatro generaciones de ordenadores han desfilado ante nosotros en el breve intervalo de treinta y cinco años y ya se habla con insistencia de la quinta generación (Gardarin, 1984). En momentos sucesivos los sistemas de información geográfica, al igual que los demás sistemas de proceso de datos, han tenido que adaptarse a las nuevas circunstancias tecnológicas, descubriéndose en cada etapa nuevas posibilidades de manipulación y análisis de la información.

No obstante, el desarrollo del cuerpo de conocimientos en esta área no

constituye un proceso estable ni transparente. Este hecho es debido fundamentalmente a dos razones. La primera de ellas la constituye la existencia de importantes intereses comerciales que impiden la publicación de las soluciones a los problemas planteados. En segundo término, la carencia de un marco teórico general que contribuya a definir objetivos bien formulados, descubrir los problemas ya resueltos y delimitar los temas que requieren una solución. En este artículo se presenta una categorización de algunas de las funciones propias de los sistemas de información geográfica con vistas a definir un esquema que contribuya a la progresiva delimitación de ese marco teórico al que hemos aludido hace un momento.

Funciones de un sistema de información geográfica

De acuerdo con la categorización establecida recientemente en un estudio sobre una de las funciones más importantes en todo sistema de información geográfica (Guevara, 1983), dos unidades fundamentales pueden distinguirse siempre en un sistema de esta índole: el componente operativo o funcional, y la base de datos. El componente funcional es un conjunto de procedimientos u operadores que actúan sobre la información contenida en la base de datos. En este artículo, para evitar ambigüedades, utilizaremos siempre en lo sucesivo el término función para designar a cada uno de los elementos del componente operativo de un sistema de información geográfica.

Más concretamente, una función de un sistema de información geográfica puede ser definida como un procedimiento algorítmico abstracto, o conjunto de éstos, que permite seleccionar, procesar y actualizar el contenido de la base de datos. Cada una de las funciones de un sistema de información geográfica debe ser formulada en los términos más abstractos posibles, independientemente de su implementación concreta en un determinado sistema, de tal manera que pueda ser evaluada, criticada y mejorada desde un punto de vista estrictamente analítico.

Algunos autores postulan la existencia de un tercer componente en un sistema de información geográfica, las estructuras de datos, que constituyen el elemento puente entre el conjunto de funciones y la base de datos. Las estructuras de datos pueden, en algunas ocasiones, reflejar la organización de la base de datos. En otras ocasiones las estructuras de datos contienen información extraída de la base de datos, pero organizada de modo distinto, con vistas a satisfacer los requisitos de la función que ha sido invocada. Normalmente, las estructuras de datos son estados de la memoria interna del ordenador y tienen carácter temporal. La base de datos, en cambio, reside en memoria externa como conjunto integrado de información, permanente entre dos actualizaciones consecutivas.

La clasificación y descripción de funciones que presentamos a conti-

nuación, como ya hemos indicado previamente, tiene un carácter genérico, las específicas configuraciones de diversos sistemas de información geográfica, en función de su orientación concreta (producción de mapas topográficos, catastro, cartografía temática y estadística, manipulación de la información asociada a una red de vías de comunicación, tratamiento de imágenes digitales, etc.) no se consideran. Para llegar a esta relación categórica de funciones, hemos tomado como punto de partida varias publicaciones recientes que, aunque no con un carácter global, han abordado el tema (Dangermond, 1983; Tomlinson y Boyle, 1981; Tomlinson, 1984).

Desde un punto de vista general podemos decir que todo sistema de información geográfica debe ofrecer una solución a cada uno de los tres problemas fundamentales en el campo de la documentación: entrada de información, archivo/recuperación de información y salida de información. Estas tres tareas, las más familiares y las unidas a tener realmente en cuenta por el usuario del sistema, dependen en su concreta realización de la estructura de la base de datos. No obstante, al nivel lógico en que nos movemos, podemos hacer abstracción de esta dependencia al considerar las funciones que resuelven los tres problemas mencionados anteriormente. Esta es la única vía posible de establecer un marco general coherente.

Entrada de información

La entrada de datos en un sistema de información geográfica está condicionada por dos factores fundamentales: la fuente de información espacial y el formato digital de la base de datos (fig. 1).

La información espacial, previa su introducción en el sistema, puede encontrarse ya en formato digital (por ejemplo, una imagen registrada por un «scanner» multiespectral) o estar contenida en algún tipo de documento analógico (por ejemplo, una fotografía aérea o un mapa impreso). En el primer caso las funciones de entrada de información no son otra cosa que procedimientos de transformación de coordenadas y de cambio de formato, posiblemente muy sofisticados, para adaptar los datos originales al esquema de representación de la base de datos.

Si la información espacial ha de ser extraída de algún documento analógico, la entrada de datos al sistema resulta mucho más laboriosa, requiriéndose toda una serie de manipulaciones, cada una de las cuales será el resultado de la función correspondiente. Tres grandes grupos de tareas (Marble, Lauzon y McGranaghan, 1984) pueden distinguirse: preparación de los documentos, digitalización y corrección de errores.

La preparación de los documentos, previa su digitización, es una tarea imprescindible si se pretende un rendimiento eficaz en el módulo de entrada de datos al sistema. Básicamente consiste en la anotación y simpli-

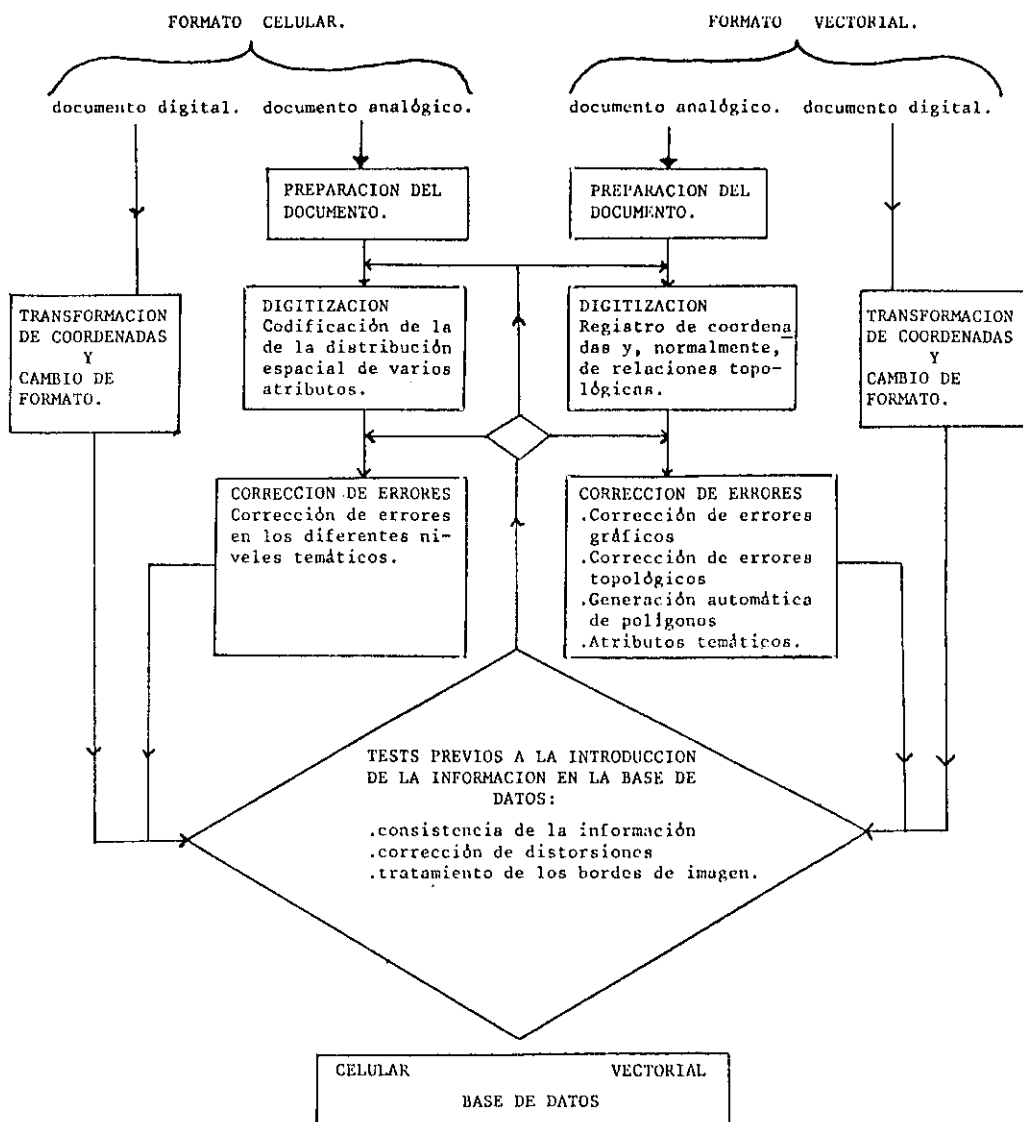


FIG. 1. *Entrada de datos en un sistema de información geográfica.*

ficación gráfica de los documentos a digitizar. No nos detenemos más en este aspecto porque, a pesar de su interés, se trata de una tarea que se realiza siempre sin la asistencia del ordenador.

Las tareas de digitización y corrección dependen en su realización concreta del formato digital de la base de datos.

En un sistema de información geográfica celular el único tipo de entidad espacial es el constituido por elementos superficiales delimitados por una malla ortogonal. Habitualmente, el espaciado de la malla es idéntico en ambas dimensiones y, consecuentemente, los elementos celulares son cuadrados en el plano de proyección del mapa. La localización espacial de cada uno de estos elementos no necesita ser registrada explícitamente ya que puede derivarse implícitamente de la resolución de la malla, de las coordenadas de un punto en la malla (normalmente el vértice inferior izquierdo) y de la posición del elemento en cuestión en una matriz bidimensional. Muy diferentes atributos pueden caracterizar cada célula: unidad administrativa, vegetación, altitud media, tipo de roquedo, etc.

La digitización consiste en la superposición de una malla transparente sobre cada mapa o imagen y la consiguiente codificación del valor del atributo en cada célula. Si el atributo tiene diversos valores en el interior de cada unidad superficial elemental, es necesario tomar una decisión acerca de la información que se ha de registrar. Si se decide registrar un único valor por cada célula, éste puede ser bien el valor dominante o un valor promediado. Si se decide registrar más de un valor es necesario codificar de alguna manera (lo más lógico es recurrir a un orden significativo) la importancia de la ocurrencia de cada valor en el interior de cada célula.

La función de entrada de datos, en este caso, es un diálogo operador-sistema, mediante un terminal alfanumérico, en dos fases. En la primera el operador define los parámetros generales de la entrada de datos, esto es, las características de la malla y la naturaleza del atributo a registrar. En la segunda fase se introducen los valores del atributo sobre cada célula en una secuencia predeterminada.

La corrección de errores en un sistema de esta índole se realiza reproduciendo gráficamente la información registrada en un periférico raster (una impresora de líneas, por ejemplo) e identificando visualmente los errores. Una vez detectadas las anomalías, los valores correctos se introducen en el fichero correspondiente sustituyendo a los inadecuados. Es recomendable realizar esta operación repetidas veces con cada fichero, para garantizar la calidad adecuada en la información registrada.

En un sistema de información geográfica de tipo vectorial las operaciones de digitización y corrección de errores, y consecuentemente las funciones que las realizan, son bastante más complejas. Aunque existen diversos modos de realizar la digitización vectorial de un documento gráfico, el más común en la actualidad es la digitización por arcos (o seg-

mentos). Este modo de digitización vectorial tiene en cuenta ciertas propiedades topológicas de los grafos bidimensionales y fue propuesto por primera vez por el *Bureau of the Census* de los Estados Unidos [DIME, *Dual Independent Map Encoding* (U. S. Department of Commerce, Bureau of the Census, 1970)]. Puntos (nodos), líneas (arcos) y áreas (polígonos) son las unidades elementales de cualquier representación bidimensional, pero nodos y polígonos están claramente relacionados con arcos. Cada arco está limitado por dos nodos; por otra parte, cada arco es siempre frontera entre dos polígonos. Consecuentemente, registrando únicamente la configuración de los arcos y las relaciones topológicas de cada uno de éstos, con el par de nodos que lo limitan y con los dos polígonos de los que es frontera, se obtiene una descripción completa de todos los elementos de un mapa.

Las funciones de asistencia a este tipo de digitización pueden presentar diversos grados de flexibilidad. En un extremo se encuentran las que requieren del operador la introducción de todas las relaciones topológicas de cada segmento en el momento de su digitización («desde nodo», «hasta nodo», «polígono derecha», «polígono izquierda»). En el otro extremo, el operador puede realizar una digitización continua de todos los elementos lineales del mapa (sin codificar el comienzo o final de cada arco), seguida de una digitización de un centroide (punto interior a) por cada polígono identificable. En el último caso, un módulo perteneciente a la función de entrada de datos establece la partición en segmentos y reconoce las relaciones topológicas de cada uno de ellos. A pesar de su atractivo, una entrada de datos tan flexible requiere un tiempo de ordenador considerable. Por este motivo, algunos autores piensan que no es un procedimiento operativo y aconsejan un compromiso entre flexibilidad y tiempo de ordenador.

La digitización de un documento gráfico para su introducción en un sistema de información geográfica de tipo vectorial puede realizarse utilizando diversos sensores (digitizadores de mesa, «scanners», etc.) y puede incluir las más sofisticadas transformaciones, pero el resultado es siempre un conjunto de arcos, o segmentos, descritos por una lista de pares de coordenadas.

La corrección de errores de una digitización vectorial ha de realizarse siempre que sea posible en un terminal con pantalla de rayos catódicos mediante un diálogo interactivo tanto gráfico como alfanumérico. Las funciones de asistencia a esta tarea han de ofrecer un repertorio completo de procedimientos destinados a la corrección de errores gráficos, así como un módulo de examen y corrección de errores en la información referente a las relaciones topológicas de los segmentos.

Si algunos de los polígonos definidos mediante arcos van a desempeñar un papel en el proceso posterior de la información en el sistema, la descripción de su forma se registra de manera que sea directamente accesible. Normalmente este tratamiento puede llevarse a cabo al mismo

tiempo que se comprueba la corrección topológica de la información espacial.

Finalmente, un último test es necesario en el proceso de corrección de errores. Se trata de comprobar que el identificador numérico que se ha asignado durante la digitización a cada entidad espacial en el mapa es correcto y único. La coherencia de esta información es primordial, ya que en bases de datos vectoriales la información espacial es almacenada normalmente en un fichero distinto del que contiene los diversos atributos temáticos que la afectan. En estas condiciones, la única posibilidad de relacionar adecuadamente ambas informaciones la constituye la presencia de un mismo conjunto de identificadores numéricos en cada fichero.

Sólo después de todos los tratamientos previos a los que hemos hecho mención se procede a la codificación de los atributos temáticos de cada una de las entidades espaciales (geométricamente definidas por puntos, líneas o áreas). El resultado es uno o varios ficheros cada uno de cuyos registros contiene la información correspondiente a una entidad espacial. Cada registro comprende tantos campos como atributos se han registrado, siendo el primero de ellos el identificador numérico que relaciona unívocamente cada entidad espacial con su descripción geométrica.

Antes de incluir el resultado de una sesión de digitización y corrección de errores en la base de datos, sea cual sea el formato digital de ésta, todavía se requieren unos test complementarios. Su objetivo es asegurar la coherencia de la información, no ya al nivel de la sesión de trabajo (asegurada por todos los test previos de corrección de errores), sino en el ámbito más amplio de la base de datos. La información geométrica y temática debe ser coherente. El sistema de referencia de coordenadas, bien es único a lo largo de toda la base de datos o, en caso de existir varios sistemas de referencia, en cada bloque de información espacial ha de estar registrado explícitamente el sistema de referencia utilizado. El formato de registro de coordenadas ha de ser homogéneo. Si se comprueba la existencia de distorsiones en el documento que ha servido de base para la digitización, éstas se pueden remover mediante filtros digitales, sin necesidad de redigitizar la información, siempre que no sean muy acusadas. La información temática debe tener un contenido y una organización análoga a lo largo de toda la base de datos.

En sistemas de tipo vectorial es necesario tener en cuenta un aspecto más (en sistemas de tipo celular, de menor resolución, no tiene sentido considerarlo), la continuidad de trazos lineales y características a través de los bordes de los documentos analógicos utilizados en la digitización (mapas o fotos aéreas). La forma más adecuada de resolver este problema es definir una función que se activa cada vez que un nuevo resultado de digitización se va a incluir en la base de datos y que investiga los casos de discontinuidad de líneas en los bordes y los presenta en una pantalla de rayos catódicos para que el operador decida cuándo se trata de

un error de digitización, que la función entonces corrige, o cuando, en cambio, se trata de dos líneas sin relación entre sí.

La entrada de datos al sistema debe tener en cuenta las operaciones de actualización de la base de datos. Para ello es necesaria la presencia de las funciones que permitan la modificación tanto de la información temática como de la espacial en la forma más sencilla posible. Si la información temática incluida en el sistema es controlada, como es habitual hoy día, por un módulo estándar de control de bases de datos alfanuméricos, los problemas de actualización han sido resueltos por los diseñadores de este módulo genérico. Las funciones de actualización de la información gráfica han de ofrecer la posibilidad de visualizar la descripción vigente en la base de datos en una pantalla de rayos catódicos. Así mismo deben contemplar la capacidad de suprimir cualquier elemento gráfico utilizando el cursor de la pantalla para identificarlo. Finalmente, deben tener en cuenta la superposición, con el mayor grado de precisión posible, de los nuevos elementos, que serán digitizados en pantalla e insertados convenientemente en la base de datos, cuando se registre la información presente en la pantalla después de efectuadas las correcciones pertinentes.

Archivo/recuperación de la información

Archivo y recuperación de la información son, realmente, aspectos distintos de una misma realidad: vías de comunicación exterior de la base de datos. En el primer caso, un bloque de información externa recorre un camino en la base de datos hasta encontrar el lugar apropiado donde residir. En el segundo, el mismo camino, normalmente, o algún otro complementario que conduzca a la misma dirección, ha de recorrerse para localizar y recuperar la información previamente introducida en la base de datos.

En la práctica, la información es introducida en el sistema con una estructura que, si bien condiciona en parte el proceso de recuperación, permite un acceso flexible. Consecuentemente, las vías de archivo y recuperación de datos se asemejan solamente en sus rasgos generales.

En el resto de este apartado sólo se van a considerar las funciones de recuperación de la información. Ello es debido a que tratar con cierto detalle de las funciones de archivo implica hacer referencias concretas a la estructura de la base de datos, que es algo que se debe evitar porque nos alejaría claramente del objetivo de este artículo, centrado en el componente funcional del sistema. De cualquier manera, al tratar de las funciones de recuperación quedarán de manifiesto una serie de requisitos que las funciones de archivo han de cumplir, ya que han de ser tales que permitan una realización optimizada de los modos de recuperación que se definan.

Para introducir un principio de claridad en el complejo ámbito de la recuperación de información en un sistema de información geográfica, vamos a distinguir en este proceso dos diferentes fases: extracción de datos y filtrado de los mismos.

Extracción de datos es el proceso de selección de un subconjunto de la documentación contenida en la base de datos, cuyos elementos son capaces de satisfacer por sí mismos o en combinación entre sí una demanda concreta de información.

Mediante sucesivas generalizaciones (McKeown, 1984), podemos llegar a la definición de cuatro modos fundamentales de extracción de datos (fig. 2).

Extracción mediante especificación geométrica: Dado un dominio espacial, definido por las coordenadas de un vértice (punto) o de varios (línea o área), extraer todas las entidades espaciales contenidas total o parcialmente en el dominio, ésto es, su posición geodética, su nombre propio y todos (o algunos) sus atributos.

Extracción mediante condición geométrica: Dado un dominio espacial y una condición geométrica (por ejemplo, un dominio puntual y la condición: en un radio de 4 metros), extraer todas las entidades espaciales que satisfacen la condición.

Extracción mediante especificación simbólica: Dado un nombre propio, extraer la entidad espacial (o entidades) señalada, ésto es, su posición geodética y todos (o algunos) sus atributos temáticos.

Extracción mediante condición simbólica o lógica: Dada una cadena de caracteres alfanuméricos (condición simbólica) o una expresión booleana (condición lógica), extraer todas las entidades espaciales cuyo nombre propio contiene la cadena alfanumérica señalada o cuyos atributos satisfacen positivamente la condición lógica apuntada.

Todos estos modos de extracción podrán efectuarse con mayor eficacia si el sistema ofrece una terminal interactiva en la que en cada momento se dispone de una representación espacial, en una pantalla de rayos catódicos, de la zona en la que se está trabajando y de un diálogo mediante teclado alfanumérico y cursor gráfico.

Filtrado de datos es un concepto genérico con el que nos referimos a todas las manipulaciones llevadas a cabo sobre la información extraída de la base de datos antes de su presentación final.

Es imposible resumir en pocas líneas el innumerable conjunto de posibles transformaciones. La relación que presentamos a continuación sólo pretende ser un marco general que permita la inclusión de cualquiera de aquéllas en alguno de los apartados genéricos de ésta.

Reclasificación de atributos. No es infrecuente que las clases que define el usuario difieran de las presentes en la base de datos. Cualquier sistema debe prever esta posibilidad y ofrecer una solución satisfactoria.

Transformación de coordenadas. Cualquier sistema debe ofrecer la posibilidad de efectuar con sencillez y transparencia las transformaciones

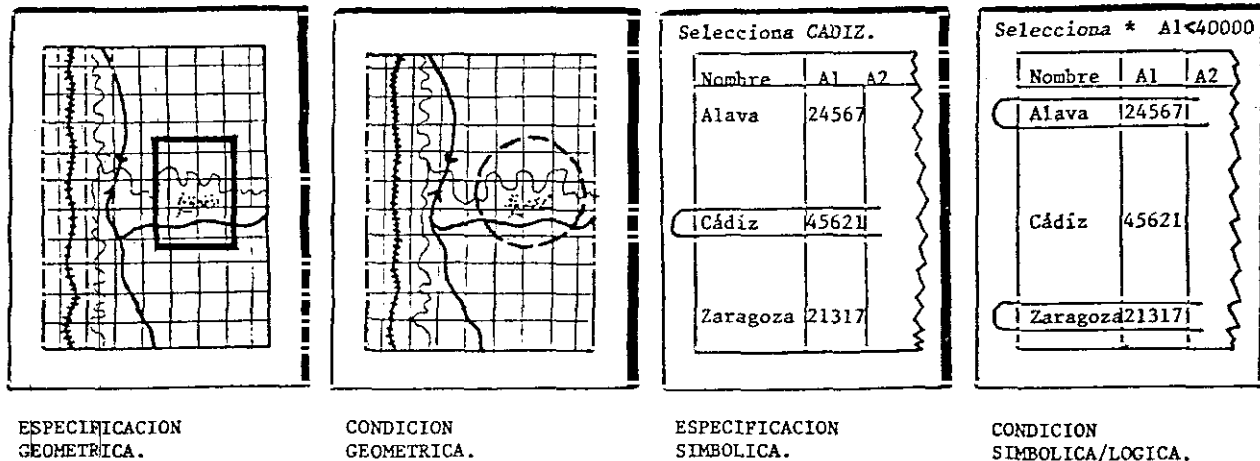


FIG. 2. *Diferentes vías de acceso a la información contenida en la base de datos.*

más comunes: modificación de escala, giros, translaciones y cambios de proyección.

Generalización cartográfica. Automatizar el proceso de generalización cartográfica no es hoy más que un objetivo a largo plazo. No obstante, algunos problemas concretos, los más sencillos (fusión de polígonos con las mismas características o reducción del número de puntos que describen una línea, por ejemplo), han sido ya resueltos y su utilidad demostrada en muchas aplicaciones.

Abstracción cartográfica. Entendemos por tal la representación de la distribución espacial de una característica por algún tipo de función continua que se ajusta satisfactoriamente a los valores de la característica en puntos muestrales. El resultado es siempre menos complejo y, por tanto, más fácilmente comprensible.

Análisis espacial. Esta categoría incluye una amplísima gama de manipulaciones, desde las más simples, como es el caso de las medidas (número de puntos en un polígono, medición de longitudes, áreas o volúmenes), hasta las más complicadas: análisis de redes, intersección de polígonos, análisis de modelos digitales del terreno, aplicación de modelos socioeconómicos, entre otras muchas. Está fuera de lugar el pretender que todo sistema de información geográfica haya de tener incorporadas estas funciones, no así, en cambio, el proponer que esté estructurado de manera que, ante la demanda de un número significativo de usuarios, resulte viable sin mucho esfuerzo la implementación de una nueva capacidad de análisis.

Salida de información

Vamos a considerar este aspecto sólo en sus rasgos generales. Cómo se resuelve técnicamente cada una de las funciones a que vamos hacer alusión se describe en algunas publicaciones recientes (Monmonier, 1982; Cebrián y G. Ferrández, 1984), que el lector interesado puede consultar.

La salida de información de un sistema de información geográfica puede ser textual (alfanumérica) o gráfica. Ambos tipos de información pueden ser presentados en formato digital o en formato analógico. Normalmente el producto analógico es el que se presenta como respuesta al usuario que ha interrogado al sistema. El producto digital contiene información destinada a otro sistema de información geográfica o, en general, a cualquier sistema informatizado.

La información textual analógica es normalmente un conjunto de tablas que presentan documentación proveniente de la base de datos o resultados de análisis realizados sobre ella.

La información gráfica de carácter analógico puede consistir en mapas, gráficos o diagramas.

Ambos tipos de información pueden aparecer impresos en papel o re-

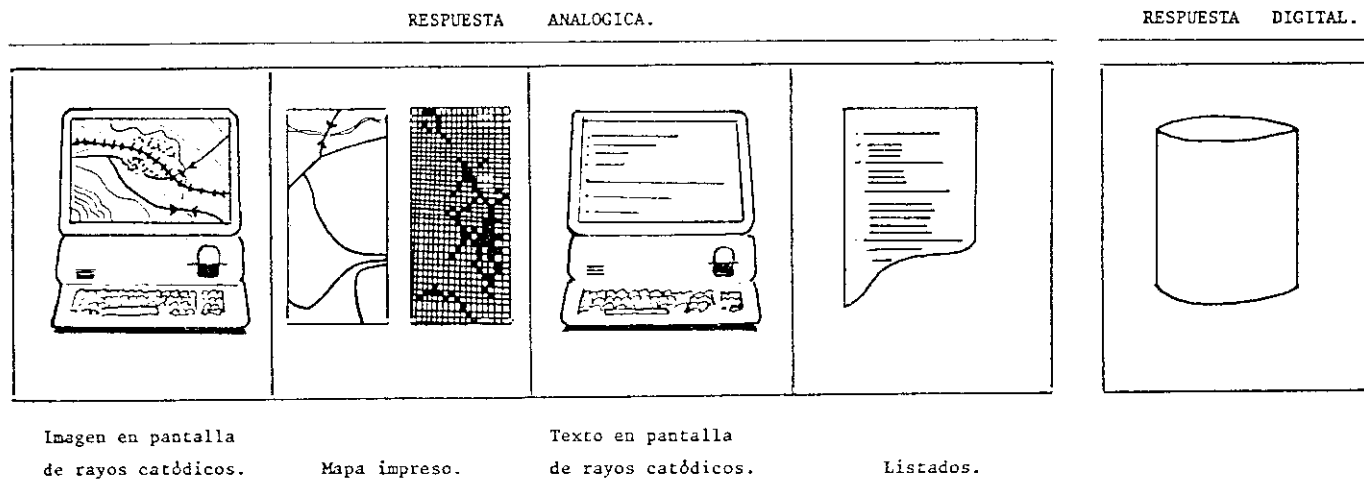


FIG. 3. Diferentes modos de respuesta de un sistema de información geográfica.

producidos temporalmente en una pantalla de rayos catódicos (fig. 3). Es muy recomendable que la salida gráfica de todo sistema pueda ser generada tanto en un periférico raster como en un periférico vectorial.

Buffalo, 23 de abril de 1985

BIBLIOGRAFÍA

- CEBRIÁN, J. A. y, FERRÁNDEZ, M. G. (1984). *Cartografía Temática y Representación Gráfica Mediante Ordenador*. IGN, Madrid, 375 pp.
- DANGERMOND, J. (1983). «A Classification of Software Components Commonly Used in Geographic Information Systems». En Peuquet, D. J. y O'Callaghan, J., eds. *Design and Implementation of Computer-Based Geographic Information Systems*. Amherst, NY: IGU Commission on Geographical Data Sensing and Processing.
- GARDARIN, G. (1984). «Towards the fifth generation of data management systems». En Gardarin, G. y Gelenbe, E., eds. *New Applications of Data Bases*. London, Academic Press, pp. 3-15.
- GUEVARA, J. A. (1983). *A Framework for the Analysis of Geographic Information System Procedures: The Polygon Overlay Problem. Computational Complexity and Polyline Intersection*. Ph. D. thesis, SUNY at Buffalo.
- HERRERO, R.; BOSQUE, J., y CEBRIÁN, J. A. (1980). «Nuevas tendencias en la investigación geográfica: el sistema de información del Instituto Geográfico Nacional (Signa)». *Estudios Geográficos*, n. 161, pp. 447-465.
- MARBLE, D. F.; LAUZON, J. P., y MCGRANAGHAN, M. (1984). «Development of a Conceptual Model of the Manual Digitizing Process». En Marble, D. F., H. W. Calkins and Peuquet, D. J., eds. *Basic Reading in Geographic Information Systems*, SPAD Ltd., Williamsville, pp. 3.19-3.33.
- MCKEOWN, D. M. (1984). «Digital Cartography and Photo Interpretation from a Database Viewpoint». En Gardarin, G. y E. Gelenbe, eds. *New Applications of Data Bases*. London, Academic Press, pp. 19-42.
- MONMONIER, M. S. (1982). *Computer-Assisted Cartography. Principles & Prospects*, Prentice-Hall, 214 pp.
- U.S. Department of Commerce, Bureau of the Census, 1970. *The DIME Geocoding System*, Report n. 4 of the Census Use Study.
- TOMLINSON, R. F.; CALKINS, H. W., y MARBLE, D. F. (1976). *Computer Handling of Geographical Data*. Paris: The UNESCO Press.
- TOMLINSON, R. F., y BOYLE, A. R. (1981). «The State of Development of Systems for Handling Natural Resources Inventory Data». *Cartographica*, vol. 18, n.º 4, pp. 65-95.
- TOMLINSON, R. F. (1984). «Geographic Information Systems- A New Frontier». Keynote Address to the *International Symposium on Spatial Data Handling*, August 20-24, 1984 Zurich, Switzerland.