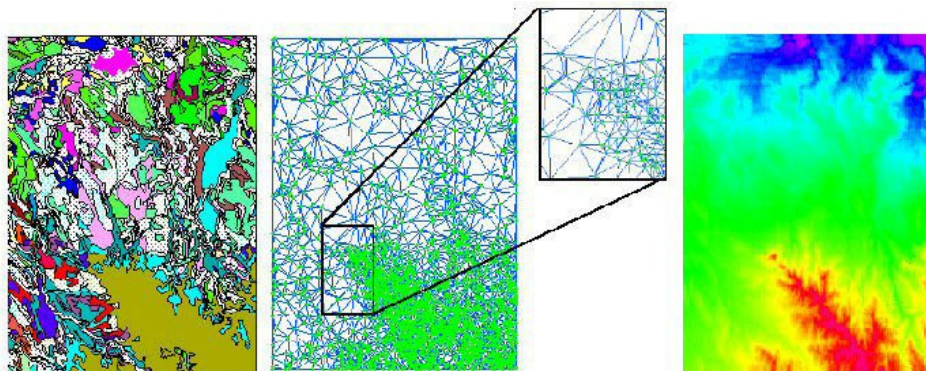


Bộ Giáo dục và Đào tạo
Trường Đại học Nông nghiệp 1 Hà nội

GIÁO TRÌNH HỆ THỐNG THÔNG TIN ĐỊA LÝ



TRẦN THỊ BĂNG TÂM

Hà nội 2006

BỘ GIÁO DỤC VÀ ĐÀO TẠO
TRƯỜNG ĐẠI HỌC NÔNG NGHIỆP HÀ NỘI

Trần Thị Băng Tâm

Giáo trình
Hệ thống Thông tin Địa lý

Hà nội, tháng 2 năm 2006

Lời nói đầu

Trong những năm gần đây, hệ thống thông tin địa lý (Geographical Information System gọi tắt là GIS), đã trở thành một ngành khoa học rất được quan tâm trong nhiều lĩnh vực nghiên cứu và ứng dụng khoa học kỹ thuật trên toàn cầu. GIS đã thực sự trở thành một yếu tố quan trọng trong hệ thống công nghệ thông tin và không thể thiếu được trong mọi chương trình đào tạo đại học và trên đại học.

Các môn học liên quan đến Hệ thống thông tin địa lý đã được đưa vào giảng dạy ở Trường Đại học nông nghiệp 1 từ nhiều năm qua nhưng bài giảng chủ yếu vẫn dựa vào các tài liệu bằng tiếng Anh. Thực tế đòi hỏi cần có một giáo trình GIS tiếng Việt để phục vụ công tác dạy và học, nhất là cho đối tượng sinh viên ở các trường đại học nông nghiệp được thuận tiện hơn. Giáo trình này là một cố gắng của các cán bộ giảng dạy môn học GIS thuộc bộ môn Địa chính để cung cấp tài liệu tiếng Việt về GIS. Giáo trình này nhằm giới thiệu những kiến thức cơ bản về GIS và chủ yếu cho đối tượng sinh viên và cán bộ của các trường đại học nông nghiệp.

Nội dung của giáo trình này được trình bày trong bảy chương và hai phụ lục.

Chương một giới thiệu quá trình hình thành Hệ thống thông tin địa lý, các nội dung chính, các thành phần của GIS, và các chức năng hoạt động chính của một hệ phần mềm GIS thông thường, các ứng dụng của GIS trong các ngành kinh tế khoa học kỹ thuật tự nhiên và xã hội đồng thời cũng trình bày mối quan hệ giữa các thành phần của GIS để bạn đọc có thể hiểu rõ định nghĩa về GIS.

Chương hai nói về hệ thống tham chiếu không gian trong GIS, bắt đầu bằng việc cung cấp cho người đọc các kiến thức chung về bản đồ, sau đó trình bày các hệ tọa độ phổ biến hiện nay trên thế giới và cuối cùng sử dụng những kiến thức về bản đồ và hệ tọa độ để giải thích về các hệ tham chiếu không gian đặc trưng cho HTTTĐL.

Cấu trúc dữ liệu hệ thống thông tin địa lý được trình bày trong Chương ba bởi vì hiểu biết về cấu trúc lưu trữ dữ liệu của một HTTTĐL bao giờ cũng rất hữu ích kể cả cho những người sử dụng HTTTĐL. Cấu trúc dữ liệu ảnh hưởng đến cả khối lượng bộ nhớ cần thiết và hiệu suất quá trình phân tích dữ liệu. Chương này trước hết trình bày các khái niệm về lưu trữ dữ liệu thông tin dưới hình thức cơ sở dữ liệu, sau đó giới thiệu các dạng lưu trữ đặc thù cho các HTTTĐL cùng với các ưu nhược điểm của từng dạng

Chương bốn giới thiệu các công đoạn nhập dữ liệu vào hệ thống GIS, và chỉnh lý dữ liệu cả về sai số và quy đổi hệ thống tọa độ và tham chiếu. Các dạng biểu thị kết quả cũng được trình bày ở cuối chương.

Chương năm chuyên về mô hình số độ cao (Digital Elevation Model, DEM). Sự biểu diễn bằng số độ cao địa hình một khái niệm liên quan chặt chẽ đến GIS và được ứng dụng rộng rãi trong để giải quyết nhiều vấn đề thực tiễn không chỉ trong giới hạn biểu diễn địa hình. Chương này trình bày khái niệm DEM, phương pháp thành lập DEM và các ứng dụng của nó.

Chương sáu là về các phương pháp phân tích dữ liệu trong hệ thống thông tin địa lý. Sự khác nhau giữa GIS và các phần mềm đồ họa là khả năng biến đổi các dữ liệu không gian gốc thành các câu trả lời cho các mục đích của người sử dụng, khả năng phân tích các dữ liệu không gian và phi không gian, khả năng tái hiện chuỗi từ cơ sở dữ liệu bằng các chức năng bất biến và đa biến của các phương pháp thống kê sử dụng các phương

pháp nội và ngoại suy. Chương này sẽ trình bày khả năng giải quyết các vấn đề, và những yêu cầu cơ bản của cơ sở dữ liệu trong GIS. Cuối cùng, khả năng phân tích dữ liệu không gian và thuộc tính của GIS sẽ được giải thích, cơ bản bằng cách kết nối hai loại dữ liệu trong GIS, dữ liệu không gian và dữ liệu thuộc tính.

Chương bảy trình bày các vấn đề liên quan đến chất lượng dữ liệu, sai số và biến động dữ liệu trong tự nhiên. Vấn đề làm sao bảo đảm hạn chế được sai số và độ tin cậy trong các quá trình thu thập, xử lý, cập nhập và phân tích dữ liệu là cực kỳ quan trọng trong bất cứ ngành khoa học kỹ thuật nào, kể cả trong GIS. Chương này trình bày các vấn đề về biến động sai số dữ liệu, nguồn gốc và nguyên do các sai số, vấn đề truyền tải sai số trong GIS và hậu quả của chúng.

Ngoài bảy chương chính, chúng tôi thấy cần thiết phải thêm hai phần phụ lục để người đọc có thể tham khảo (a) về tình hình phát triển các phần mềm GIS, cả phần mềm trên thị trường và phần mềm miễn phí nguồn mở; và (b) về xu hướng phát triển của GIS trên cơ sở tham khảo những tạp chí chuyên ngành mới xuất bản gần đây. Phụ lục này giới thiệu xu hướng phát triển mới của GIS khi kết hợp với Internet, GPS và các thiết bị Mobile kết nối không dây (wireless) với mong muốn làm cho người đọc cảm nhận được một phần những gì đang diễn ra trên thế giới trong ngành GIS.

Tác giả xin cảm ơn TS. Nguyễn Duy Bình (Viện Khí tượng Thủy văn và Môi trường TP Hồ Chí Minh) đã nhiệt tình hợp tác và giúp đỡ về chuyên môn trong quá trình viết giáo trình. Xin cảm ơn các đồng nghiệp trong và ngoài trường Đại học nông nghiệp 1 đã đọc bản thảo và góp ý để hoàn thành cuốn sách này

Mặc dù đã cố gắng rất nhiều nhưng do trình độ có hạn nên chắc chắn giáo trình này không thể tránh được thiếu sót. Rất mong nhận được mọi ý kiến đóng góp để lần xuất bản sau giáo trình sẽ có chất lượng cao hơn, góp phần tốt hơn cho nhiệm vụ giảng dạy và học tập môn GIS ở các trường đại học nông nghiệp.

Hà nội, tháng 2 năm 2006

Tác giả

MỤC LỤC

Chương	Trang
Lời nói đầu	ii
Danh mục Bảng.....	vii
Danh mục Hình vẽ.....	viii
 Chương 1 GIỚI THIỆU VỀ HỆ THỐNG THÔNG TIN ĐỊA LÝ (GIS) VÀ KHẢ NĂNG ỨNG DỤNG GIS TRONG CÁC NGÀNH KINH TẾ, KỸ THUẬT.....	 1
1.1 Vai trò của Hệ thống thông tin địa lý	2
1.2 Lịch sử phát triển hệ thống thông tin địa lý	4
1.3 Những nội dung chính một hệ GIS đề cập tới	5
1.4 Yêu cầu đối với một hệ GIS.....	5
1.5 Định nghĩa về GIS	6
1.6 Các thành phần chính của Hệ thống thông tin địa lý.....	8
1.7 Một số ứng dụng GIS trong các ngành khoa học kinh tế và kỹ thuật	15
Câu hỏi ôn tập	20
 Chương 2 MỘT SỐ KHÁI NIỆM VỀ HỆ THỐNG THAM CHIẾU KHÔNG GIAN	 21
2.1 Khái niệm về bản đồ	21
2.2 Các đặc tính của bản đồ	24
2.3 Hệ thống tọa độ (Coordinate system)	27
2.4 Phép chiếu bản đồ (Map projection).....	29
2.5 Hệ qui chiếu VN-2000.....	33
Câu hỏi ôn tập:	34
 Chương 3 CẤU TRÚC DỮ LIỆU GIS	 36
3.1 Các khái niệm cơ sở.....	36
3.2 Dữ liệu không gian	38

3.2.1 Cấu trúc dữ liệu Raster	38
3.2.2. Cấu trúc dữ liệu véc tơ	42
3.3 Dữ liệu thuộc tính.....	51
Câu hỏi ôn tập	52
 Chương 4 NHẬP VÀ CHỈNH LÝ DỮ LIỆU TRONG HỆ THỐNG THÔNG TIN ĐỊA LÝ	 54
4.1 Nhập dữ liệu.....	54
4.2 Biên tập và Chỉnh sửa dữ liệu không gian	59
4.3 Nhập dữ liệu thuộc tính phi không gian	66
4.4 Kết nối dữ liệu không gian và dữ liệu thuộc tính	66
4.5 Hiển thị đầu ra của dữ liệu địa lý	67
Câu hỏi ôn tập	67
 Chương 5 MÔ HÌNH SỐ ĐỘ CAO (DEM).....	 68
5.1 Giới thiệu chung.....	68
5.2 Phương pháp biểu thị DEM	69
5.3 Phương pháp xây dựng DEM	71
5.4 Các sản phẩm ứng dụng DEM.....	73
Câu hỏi ôn tập	78
 Chương 6 PHÂN TÍCH DỮ LIỆU TRONG GIS.....	 80
6.1 Mục đích và khả năng giải quyết vấn đề của các hệ thống GIS	80
6.2 Chức năng quản lý cơ sở dữ liệu.....	82
6.3 Các chức năng phân tích dữ liệu không gian.....	83
Kết luận	91
Câu hỏi ôn tập	91
 Chương 7 CHẤT LƯỢNG DỮ LIỆU, SAI SỐ VÀ BIẾN ĐỘNG DỮ LIỆU.....	 94
7.1 Giới thiệu chung.....	94
7.2 Sai số có nguồn gốc rõ ràng.....	94
7.3 Sai số trong quá trình thu thập dữ liệu	97
7.4 Sai số trong quá trình phân tích dữ liệu.....	99
 Trường Đại học Nông nghiệp Hà Nội – Giáo trình Hệ thống thông tin Địa lý	 v

7.5 Kết luận	103
Câu hỏi ôn tập	103
Tài liệu tham khảo	104
Phụ lục A GIỚI THIỆU MỘT SỐ PHẦN MỀM GIS	106
A.1 Hệ phần mềm trên thị trường.....	106
A.2 Hệ phần mềm nguồn mở	108
Phụ lục B XU HƯỚNG PHÁT TRIỂN CỦA GIS	110
Bảng Từ viết tắt.....	116
Danh mục từ vựng	118
Bảng tra thuật ngữ	125

Danh mục Bảng

Bảng	Trang
Bảng 2.1 Một số ellipsoids được sử dụng chính thức trên thế giới	Error! Bookmark not defi
Bảng 3.1 Nén theo hàng cột	41
Bảng 3.2 So sánh mô hình dữ liệu dạng vector và raster.....	50
Bảng 5.1 Phương pháp biểu thị mặt cong địa hình.....	69
Bảng 5.2 Sản phẩm ứng dụng DEM trong GIS.....	73
Bảng 7.1 Phân loại sai số trong GIS theo nguồn gốc	95
Bảng A.1 Các phần mềm GIS phổ biến trên thị trường.....	107
Bảng A.2 Các dự án phần mềm nguồn mở GIS, GPS và xử lý ảnh.....	108
Bảng A.3 Các website về phần mềm nguồn mở GIS và về hội nghị Web...	109

Danh mục Hình vẽ

Hình	Trang
Hình 1.1 GIS là sự thể hiện của thế giới thực bao gồm nhiều đặc tính địa lý được thể hiện theo các lớp dữ liệu đại diện	2
Hình 1.2 Một bản đồ có thể là tập hợp của nhiều lớp thông tin chuyên đề khác nhau	3
Hình 1.3 Chiết xuất thông tin từ nhiều lớp dữ liệu	3
Hình 1.4 Các nội dung hoạt động của một GIS	5
Hình 1.5 Sơ đồ khái niệm về một hệ thống TTDL	6
Hình 1.6 Phân loại các Hệ thống thông tin.....	7
Hình 1.7 Phân biệt hệ thống thông tin địa lý và hệ thống thông tin đất.....	8
Hình 1.8 Các thành phần phần cứng của GIS.....	9
Hình 1.9 Các chức năng thành phần chính của một hệ quản trị dữ liệu địa lý	9
Hình 1.10 Sơ đồ cấu trúc của chức năng nhập dữ liệu.....	10
Hình 1.11 Cấu trúc của một cơ sở dữ liệu địa lý	11
Hình 1.12 Chức năng chuyển đổi dữ liệu	11
Hình 1.13 Chức năng trình bày kết quả.....	11
Hình 1.14 Nguồn dữ liệu vào của một hệ GIS	13
Hình 1.15 Quan hệ giữa các thành phần của GIS	14
Hình 1.16 Sơ đồ phối hợp giữa công tác quản lý và công nghệ GIS.....	15
Hình 2.1 Bản đồ là một hệ thống truyền tải các thông tin không gian	22
Hình 2.2 Sơ đồ hoạt động xử lý và tạo lập bản đồ số trên mạng.....	24
Hình 2.3 Khái niệm các đặc tính địa lý trên bản đồ.....	25
Hình 2.4 Ví dụ so sánh giữa tỉ lệ bản đồ khác nhau	26
Hình 2.5 Hệ tọa độ địa lý.....	Error! Bookmark not defined.
Hình 2.6 Đường trục trái đất.....	<u>29</u>
Hình 2.7 Phép chiếu lập thể cực (azimuthal).....	31
Hình 2.8 Phép chiếu hình nón (Lambert)	31
Hình 2.9 Phép chiếu hình trụ đứng (Mercator).....	31
Hình 2.10 Phép chiếu hình trụ và lập thể ngang (Transverse Mercator)	32
Hình 2.11 Phép chiếu UTM.....	33

Hình 3.1 Cấu trúc dữ liệu RASTER	39
Hình 3.2 Khái niệm về các lớp dữ liệu và tổ chức dữ liệu raster	40
Hình 3.3 Nén chain-code.....	41
Hình 3.4 Nén block-code	42
Hình 3.5 Nén cây tứ phân (quadtrees).....	43
Hình 3.6 Các dữ liệu vector dạng điểm đường vùng trong tọa độ phẳng x, y.....	43
Hình 3.7 Các cấu trúc vector dạng đường.....	45
Hình 3.8 Cấu trúc vector dạng vùng.....	46
Hình 3.9 Vùng với cấu trúc topo riêng biệt.....	47
Hình 3.10 Vùng với cấu trúc topo mạng đầy đủ	47
Hình 3.11 Tính chất liên tục của Topology.....	48
Hình 3.12 Tính chất tạo vùng của Topology.....	49
Hình 3.13 Tính chất kề cận của Topology	49
Hình 3.14 Mô hình quan hệ giữa dữ liệu không gian và dữ liệu thuộc tính	52
Hình 4.1 Sơ đồ nhập dữ liệu cho một HTTT địa lý.....	55
Hình 4.2 Raster hoá dữ liệu.....	57
Hình 4.3 Nhận dạng tự động đường trên bản đồ quét	59
Hình 4.4 Tìm kiếm điểm trong đa giác	61
Hình 4.5 Loại bỏ điểm dư thừa.....	62
Hình 4.6 Chuyển đổi hệ tịnh tiến hệ tọa độ.....	64
Hình 4.7 Chuyển đổi góc quay hệ tọa độ.....	64
Hình 4.8 Phát hiện và sửa lỗi dữ liệu không gian dạng vector.....	65
Hình 5.1 Ví dụ mạng TIN	70
Hình 5.2 Mạng TIN với sự thay đổi kích thước lưới đặc trưng	71
Hình 5.3 Ứng dụng TIN để biểu thị sự biến động độ cao địa hình	71
Hình 5.4 Sơ đồ Voronoi	72
Hình 5.5 Biểu đồ khối biểu thị sự biến động của độ cao địa hình	74
Hình 5.6 Bản đồ đường đồng mức với khoảng cách độ cao 5 m.....	75
Hình 5.7 Bản đồ địa hình ở Hình 5.6 nhưng sử dụng mô hình DEM	75
Hình 5.8 Vector hướng dốc	76
Hình 5.9 Ứng dụng DEM để biểu thị địa hình	77
Hình 5.10 Hình ảnh tô bóng của bản đồ địa hình vùng núi đá.....	78

Hình 6.1 Sơ đồ phân tích dữ liệu trong hệ thống GIS.....	80
Hình 6.2 Thay đổi hệ tọa độ và phép chiếu	84
Hình 6.3 Truy cập phân loại và đo đạc.....	86
Hình 6.4 Thao tác đào và đắp	87
Hình 6.5 Sơ đồ Venn	88
Hình 6.6 Chồng ghép bản đồ	88
Hình 6.7 Tạo vùng đệm và chức năng dùng lưới.....	89
Hình 6.8 Phép tính đại số trên bản đồ.	90
Hình 6.9 Chồng xếp đại số hai lớp dữ liệu không gian.....	90
Hình B.1 Sơ đồ làng Catalhöyük (năm 6500 trước CN).....	110
Hình B.2 Màn hình mở đầu của Google Earth	113

Chương Một

GIỚI THIỆU VỀ HỆ THỐNG THÔNG TIN ĐỊA LÝ (GIS) VÀ KHẢ NĂNG ỨNG DỤNG GIS TRONG CÁC NGÀNH KINH TẾ, KỸ THUẬT

Chương này giới thiệu quá trình hình thành Hệ thống thông tin địa lý (HTTTĐL hay GIS), các nội dung chính, các thành phần của GIS, và các chức năng hoạt động chính của một hệ phần mềm GIS thông thường, đồng thời cũng trình bày mối quan hệ giữa các thành phần của GIS để bạn đọc có thể hiểu rõ định nghĩa về GIS. Các ứng dụng của GIS trong các ngành kinh tế khoa học kỹ thuật tự nhiên và xã hội được trình bày trong chương này cho thấy GIS có thể là một công cụ hữu hiệu trong công tác quản lý các nguồn tài nguyên thiên nhiên và nâng cao nhận thức của con người về bảo vệ môi trường.

Trong những năm gần đây, hệ thống thông tin địa lý (Geographical Information System gọi tắt là GIS), đã trở thành một ngành khoa học rất được quan tâm trong nhiều lĩnh vực nghiên cứu và ứng dụng trên toàn cầu. GIS đã trở thành một yếu tố quan trọng trong hệ thống công nghệ thông tin và như vậy đã trở nên không thể thiếu trong quá trình hỗ trợ ra mọi quyết định quan trọng của loài người.

Những người ra quyết định ở mọi tầng lớp (chính trị, tổ chức tư nhân, kinh tế kỹ thuật và công nghiệp) đã nhận thức được rõ ràng về tầm quan trọng của mối quan hệ giữa số liệu về vị trí địa lý và các thông tin khác, và họ đang sử dụng GIS trong mọi quá trình ra quyết định của mình.

Thế giới của chúng ta đang không ngừng biến chuyển, theo đà những quá trình tự nhiên và bị ảnh hưởng sâu sắc do những hoạt động của con người. Trái đất ngày càng có mật độ dân số cao hơn và các đô thị ngày càng lớn hơn. Hiện nay công nghệ khoa học kỹ thuật có ảnh hưởng lớn đối với cuộc sống mạnh mẽ hơn bất cứ thời đại nào trong lịch sử, và quá trình chuyên môn hóa cũng diễn ra với tốc độ ngày càng mạnh mẽ. Giao lưu giữa những người trên thế giới đã và đang trở nên thường xuyên hơn và mỗi cá nhân cũng có điều kiện đón nhận đủ các loại thông tin. Nói một cách khác đó là quá trình toàn cầu hóa và theo nhiều nhà nghiên cứu thì các nguy cơ đối với cuộc sống của loài người cũng trở nên khó có thể xác định được hơn. Hoạt động của con người đã ảnh hưởng đáng kể đến môi trường trái đất, đến đa dạng sinh học của thế giới, và đến tài nguyên thiên nhiên có thể là nguồn sống của các thế hệ tương lai. Các bằng chứng về những nguy cơ kể trên chúng ta có thể nhìn thấy ở khắp mọi nơi trên thế giới. Các thành phố đang trở nên to lớn hơn, tiêu thụ ngày càng nhiều tài nguyên hơn và các công trình đang tràn ra các vùng trước đây còn là nguyên dạng của tự nhiên.

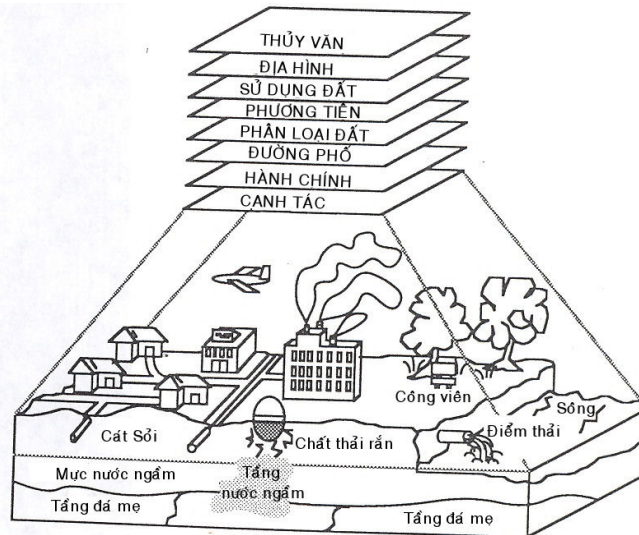
Các nghiên cứu về khí tượng đã chứng minh trái đất chúng ta đang nóng dần lên và sẽ là nguyên nhân của nhiều hiện tượng tự nhiên không thể lường trước được. Các dòng đại dương và khí hậu có thể sẽ thay đổi. Sinh vật sống trong các đại dương không những phải chịu tác động của nhiệt độ không khí tăng lên mà còn bị các nguồn ô nhiễm của con người làm cho điêu đứng. Mặc khác, mực nước biển có thể sẽ dâng cao, không những làm biến động môi trường vùng ven biển mà có thể làm biến mất các cộng đồng dân cư các vùng đó. Đây chỉ là những ví dụ rõ ràng về tác động của con người hiện nay nhưng cũng đủ để ta thấy nguy cơ phía trước có thể to lớn đến chừng nào. Loài người bắt buộc phải tự mình rút ra những bài học và phải phải học thật nhanh để quản lý phát triển bền vững, bảo đảm cho sự tồn tại lâu dài của chính con người.

Hệ thống thông tin địa lý (GIS) đã trở thành một khung công cụ tổng hợp rất hiệu quả nhằm quản lý các hoạt động cả của con người và thiên nhiên bởi vì nó giúp ta tổng hợp và phân tích mọi dữ liệu phức tạp, và đưa ra kết quả để tất cả mọi người, từ các nhà khoa học, nghiên cứu, nhà vạch định kế hoạch và công chúng, đều có thể cảm nhận được. Chúng ta đang bước vào một thời kỳ ứng dụng của công nghệ GIS có thể tổ chức và tổng hợp tốt hơn các dữ liệu khoa học, đánh giá và định hướng các biến động theo thời gian và không gian, xây dựng các mô hình, phương pháp số học và hệ thống nhằm tìm ra các mối quan hệ, các biện pháp quản lý trong việc sử dụng và bảo vệ các nguồn tài nguyên thiên nhiên gắn với sự phát triển kinh tế - xã hội trên toàn cầu.

Với sự phát triển của công nghệ GIS, các nhà làm chính sách, các nhà khoa học tự nhiên và xã hội, các tổ chức kinh tế doanh nghiệp tư nhân và Nhà nước, đều có thể sử dụng GIS như một công cụ kỹ thuật để giải quyết các vấn đề cấp thiết của thời đại như tình trạng triệt phá rừng, sự xuống cấp của môi trường, vấn đề đô thị hoá, dự báo về những biến động của khí hậu. GIS cung cấp cho con người những công cụ mạnh nhất để có thể xây dựng, tổ chức, xử lý và quản lý các dữ liệu cung cấp các thông tin trợ giúp cho các chuyên gia về GIS và các nhà quản lý trong việc ra các quyết định đúng đắn, các giải pháp hữu hiệu cho các vấn đề trên..

1.1 Vai trò của Hệ thống thông tin địa lý

Trong vài thập kỷ gần đây, địa lý học đã có nhiều phát kiến về vận dụng những kỹ thuật thông tin, trong đó bao gồm những phương pháp mới về mô hình toán học và mô hình thống kê đồng thời cũng giới thiệu những nguồn thông tin mới như dữ liệu viễn thám.



Hình 1.1 GIS là sự thể hiện của thế giới thực bao gồm nhiều đặc tính địa lý được thể hiện theo các lớp dữ liệu đại diện

Trong khung cảnh của những phát kiến này, hệ thống thông tin địa lý (HTTĐL) đã đóng một vai trò quan trọng như là một kỹ thuật tổ hợp. HTTĐL được coi là sự liên kết của một số các kỹ thuật khác nhau vào thành một tổng thể hài hòa. HTTĐL là kỹ thuật mạnh mẽ cho phép nhà địa lý tổ hợp dữ liệu của họ và áp dụng những phương pháp hỗ trợ phân tích địa lý truyền thống như phân tích chồng ghép bản đồ và mô hình hoá, trên máy tính. Với HTTĐL nhiều điều có thể thực hiện được như là vẽ bản đồ, xây dựng mô

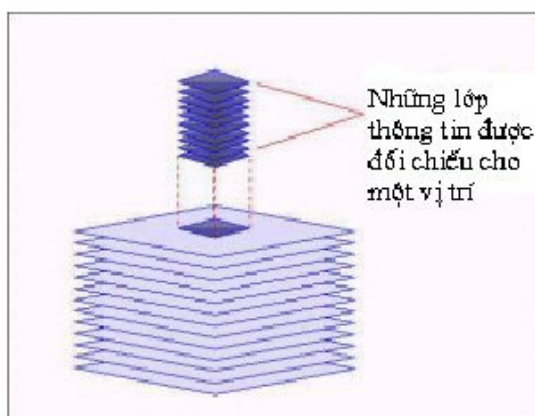
hình, hỏi đáp và phân tích một lượng lớn dữ liệu. Tất cả đều được lưu giữ trong một cơ sở dữ liệu. Sự phát triển của HTTĐL dựa trên nền tảng của nhiều ngành khoa học khác nhau như: địa lý, bản đồ học, đo vẽ ảnh, viễn thám, khảo sát, trắc địa, xây dựng công trình, thống kê, khoa học máy tính, dân tộc học và nhiều nhánh khác của khoa học trái đất và khoa học tự nhiên. HTTĐL có thể được xem xét như một cơ sở dữ liệu số trong đó dữ liệu thông tin được liên kết trong một hệ tọa độ không gian. HTTĐL cho phép:

1. Nhập dữ liệu (các bản đồ, ảnh hàng không, vệ tinh, và các nguồn khác);
2. Lưu trữ dữ liệu, truy nhập và hỏi đáp;
3. Chuyển đổi dữ liệu, phân tích và mô hình hoá;
4. Báo cáo dữ liệu (các bản đồ, báo cáo và sơ đồ).

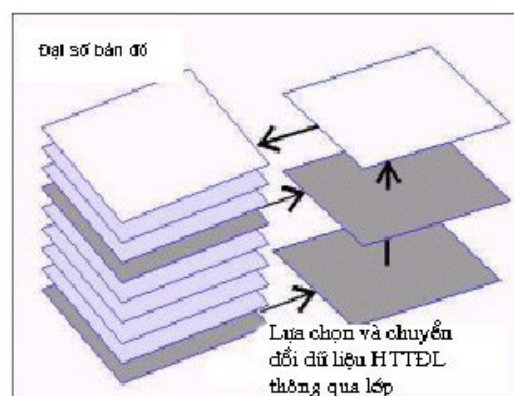
Đối tượng chính của các ứng dụng HTTĐL là sự chồng ghép của những lớp thông tin khác nhau thông qua việc sử dụng nhiều nguồn dữ liệu đa dạng được xây dựng trên một bản đồ cơ sở địa hình. Điều quan trọng là tất cả các lớp thông tin phải được đăng ký với cùng một hệ quy chiếu không gian phổ dụng. Với chức năng đo đạc và tổ hợp những thành phần khác nhau, HTTĐL cho ta nhìn thấy và quản lý tất cả những gì đang diễn ra trên bề mặt Trái đất (Hình 1.1).

HTTĐL cho phép tìm kiếm những loại hình không gian, xử lý và cho những mối quan hệ giữa những lớp chuyên đề khác nhau. Một bản đồ có thể là lớp chồng xếp của nhiều bản đồ chuyên đề cùng được đăng ký vào một hệ tọa độ chung (Hình 1.2) HTTĐL cũng cho phép tách biệt một hay nhiều lớp từ các lớp thông tin nguồn. Thông tin của những lớp cụ thể có thể được tập hợp và chuyển sang lớp mới để tiếp tục phân tích về sau. Đây là quá trình thực hiện các phép tính đại số trên bản đồ (hình 1.3).

Cùng với khoa học bản đồ, kỹ thuật HTTĐL đã tăng cường hiệu quả và sức mạnh phân tích của việc thành lập bản đồ truyền thống. Một hệ HTTĐL có thể được sử dụng để sản xuất những hình ảnh không phải chỉ là những bản đồ mà còn là những bản vẽ kỹ thuật và sản phẩm hình họa khác. Những hình ảnh này cho phép những nhà quan sát cảm nhận các đối tượng của họ theo cách mà chưa bao giờ nhìn thấy trong các tài liệu trước đây. Những hình ảnh này rất có ích trong việc dẫn dắt tới các khái niệm kỹ thuật của HTTĐL dành cho những người không chuyên nghiên cứu khoa học.



Hình 1.2 Một bản đồ có thể là tập hợp của nhiều lớp thông tin chuyên đề khác nhau



Hình 1.3 Chiết xuất thông tin từ nhiều lớp dữ liệu

1.2 Lịch sử phát triển hệ thống thông tin địa lý

1. Giai đoạn 1: Từ 1960s - 1970s

GIS được sử dụng mang tính đơn lẻ, cá nhân, hệ thống thiếu tính linh hoạt. Có thể kể ra một số tác giả và hệ GIS đầu tiên trong giai đoạn này như R. Tomlinson & Canada Geographic Information System (CGIS), H. Fisher & SYMAP mapping package

2. Giai đoạn 2: từ giữa 1970s đến đầu những năm 1980s

Chủ yếu là sự truyền bá về GIS, ít phát kiến mới, tập trung chủ yếu ở các cơ quan nghiên cứu Nhà nước

3. Giai đoạn 3 : Từ 1980 -1990

Do sự phát triển của kinh tế thị trường, các phần mềm GIS nổi tiếng như ArcInfo in 1982 by ESRI (Environmental Systems Research Institute) Mapinfo ra đời sự phát triển của GIS được chấp nhận

4. *Giai đoạn 4: Từ cuối 1980s đến nay* đánh dấu sự tiến bộ vượt bậc của công nghệ sản xuất máy tính điện tử. Cấu hình máy vi tính ngày càng mạnh và giá thành của cả phần cứng và phần mềm đều hạ. Sự tiến bộ vượt trội của bản đồ vẽ trên máy tính so với bản đồ giấy (nhanh hơn, đẹp hơn, chứa nhiều thông tin hơn, dễ cập nhật, lưu trữ tiện lợi, giá trị sử dụng cao, sai số kỹ thuật và ngẫu nhiên thấp

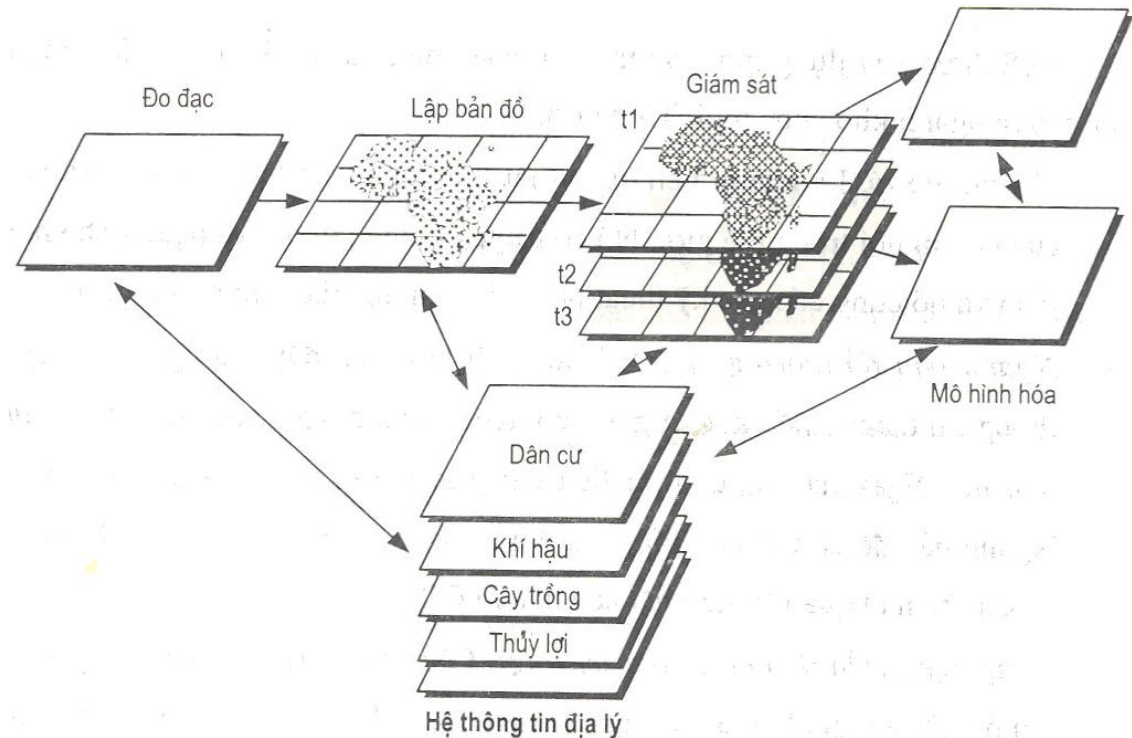
5. *Xu hướng hiện nay của GIS* (xem phụ lục B): Trong tương lai GIS sẽ được phát triển một cách toàn diện hơn về các nội dung sau:

- Thay đổi cộng đồng sử dụng GIS từ cơ quan chính phủ sang thành phần kinh doanh (mang tính thị trường)
- Các phần cứng và phần mềm GIS xuất hiện ngày càng nhiều với giá thành ngày càng hạ, dung lượng lớn, cấu hình cao, tính độc lập của hệ thống cao hơn
- Môi quan hệ dữ liệu được tăng cường về mặt bản quyền dữ liệu, Sở hữu và phân phối dữ liệu (Data copyrights, ownerships, and privacy)
- Chất lượng số liệu được chú ý (Census data, demographic data)
- Ứng dụng các mô hình không gian ngày càng nhiều
- Sự tích hợp của các sách lược phát triển và các kỹ thuật ngoại vi
- Cách thể hiện dữ liệu (Data visualization) và Web GIS
- Môi trường máy tính (Computer environment)
 - Windows vs. UNIX
 - PCs vs. Workstation

Tương lai không xa, khi mà giá thành và tốc độ đường truyền internet, kể cả mạng không dây, được cải tiến vượt bậc thì khả năng ứng dụng GIS trên cơ sở web sẽ có thể lại bùng nổ. Các công nghệ Mobile như WiFi, và 2.5 và 3G cellular, sẽ đem lại khả năng mới cho GIS.

1.3 Những nội dung chính một hệ GIS đề cập tới

GIS cho biết một vật thể A ở đâu? (tên vùng, mã vùng, điểm quan trắc theo hệ tọa độ đã biết).



Hình 1.4 Các nội dung hoạt động của một GIS

GIS nêu và giải quyết những điều kiện thoả mãn với vị trí của một vật thể. Vật thể A có quan hệ thế nào với vật thể B, có bao nhiêu kiểu A trong một khoảng cách hay một diện tích, một phạm vi không gian đã cho trước; Tính toán khoảng cách ngắn nhất giữa 2 vật thể, chi phí có hiệu quả nhất hoặc phát hiện những cản trở từ điểm X đến điểm Y nào đó; Mô tả sự kết hợp giữa các vật thể và các thuộc tính của chúng.

GIS sử dụng cơ sở dữ liệu số để lập mô hình mô tả những biến đổi của thế giới thực trong phạm vi một vùng hay nhiều vùng theo một kịch bản trong một quá trình hoàn cảnh cụ thể với sự biến đổi của thời gian.

GIS đưa ra những khả năng lựa chọn dựa trên các kết quả đã được phân tích chính xác giúp cho những nhà làm chính sách những nhà đầu tư có thể ra được các quyết định có hiệu quả trong lĩnh vực hoạt động của mình.

GIS có khả năng đề xuất các mô hình hoạt động một khi các vấn đề trên đã được thực hiện.

Những nội dung kết quả mà HTTTĐL hiện đại đưa ra không có gì khác với những kết quả của những phương pháp phân tích truyền thống, tuy nhiên nó đã chứng tỏ là một công cụ hữu hiệu giúp cho các nhà khoa học chuyên môn giải quyết được các vấn đề một cách chính xác và nhanh chóng.

1.4 Yêu cầu đối với một hệ GIS

Một hệ GIS hoạt động có hiệu quả phải thực hiện được các yêu cầu sau:

- Xử lý dễ dàng nhanh chóng với số lượng dữ liệu địa lý lớn;
- Có khả năng tách chọn chi tiết dữ liệu theo miền, vùng hoặc theo chuyên đề;
- Có khả năng tìm kiếm theo các tính chất đặc trưng đặc biệt của một hay nhiều đối tượng địa lý;
- Có khả năng phân tích các dữ liệu không gian, liên kết các dữ liệu không gian và thuộc tính;
- Có khả năng cập nhật dữ liệu nhanh chóng với chi phí thấp;
- Có khả năng mô hình hoá dữ liệu và đề ra các phương án chọn lựa;
- Có khả năng trao đổi dữ liệu với các hệ thống thông tin khác, có khả năng để phát triển tiếp với các chức năng tiện ích khác;
- Có khả năng kết xuất dữ liệu ra với các hình thức khác nhau.

1.5 Định nghĩa về GIS

Tùy theo các chức năng, nhiệm vụ của mỗi hệ thống tin địa lý mà có nhiều định nghĩa khác nhau, tuy nhiên các khái niệm về GIS đều dựa trên 3 yếu tố quan trọng là: dữ liệu đầu vào, hệ thống vi tính số kỹ thuật cao và khả năng phân tích số liệu không gian

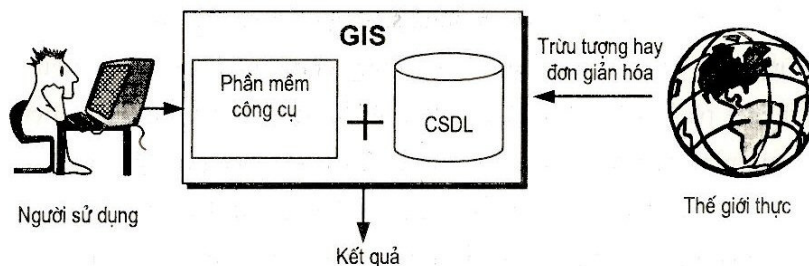
Định nghĩa theo chức năng: GIS là một hệ thống bao gồm 4 hệ con: dữ liệu vào, quản trị dữ liệu (quản lý và phân tích dữ liệu) và dữ liệu ra.

GIS là tập hợp của các thuật toán: Trong một hệ thống tin địa lý có thể sử dụng các phương pháp tính đại số, hình học từ đơn giản đến phức tạp. Các phép đo đạc có thể thực hiện với các phép tính đơn giản như tính diện tích, đo chiều dài, tính sai số trung bình đến các phép toán giải tích cao cấp trong quá trình xử lý phân tích số liệu.

Định nghĩa theo mô hình cấu trúc dữ liệu: gồm các cấu trúc dữ liệu được sử dụng trong các hệ thống khác nhau (cấu trúc dạng Raster và vector).

Về mặt công nghệ: GIS là công nghệ thông tin để lưu trữ, phân tích và trình bày các thông tin không gian và phi không gian. Công nghệ GIS có thể coi là một tập hợp hoàn chỉnh các phương pháp và các phương tiện nhằm sử dụng và lưu trữ các đối tượng.

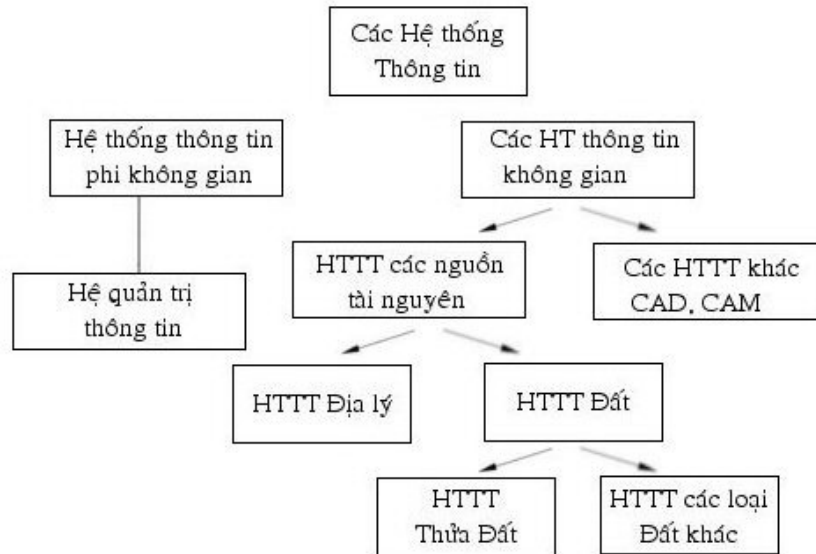
GIS là hệ thống trợ giúp ra quyết định: GIS có thể coi là một hệ thống trợ giúp việc ra quyết định, có thể tích hợp các số liệu không gian trong một cơ chế thống nhất. GIS được sử dụng để cung cấp thông tin nhanh hơn và hiệu quả hơn cho các nhà hoạch định chính sách. Các cơ quan chính phủ dùng GIS trong quản lý các nguồn tài nguyên thiên nhiên, trong các hoạt động quy hoạch, mô hình hoá và quan trắc.



Hình 1.5 Sơ đồ khái niệm về một hệ thống TTĐL

Định nghĩa tổng quát theo BURROUGHTS, 1986 "GIS như là một tập hợp các công cụ cho việc thu nhập, lưu trữ, thể hiện và chuyển đổi các dữ liệu mang tính chất không gian từ thế giới thực để giải quyết các bài toán ứng dụng phục vụ các mục đích cụ thể" (Hình 1.5).

GIS được coi là một công cụ để tổng hợp các dữ liệu không gian đã thu nhận theo các tỉ lệ, khoảng thời gian khác nhau và trong các khuôn dạng dữ liệu khác nhau.

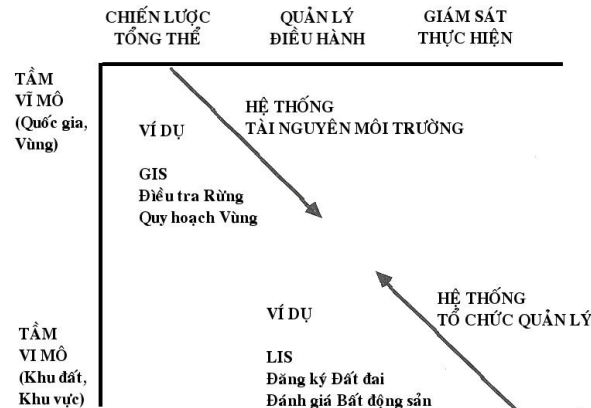


Hình 1.6 Phân loại các Hệ thống thông tin

Trong hình 1.6, hệ thống thông tin không gian CAD phần lớn chỉ chú ý đến việc trình bày và hiển thị các dữ liệu hình họa ít quan tâm đến tính chất thuộc tính của các dữ liệu. Các tính chất thuộc tính có thể hoặc không cần thiết phải có trong quá trình phân tích dữ liệu. Một CAD có thể là một thành phần quan trọng của GIS nhưng bản thân một phần mềm đồ họa không thể giải quyết được tất cả những yêu cầu và kết quả trông đợi như đối với một hệ thống GIS. Vì vậy một phần mềm đồ họa chỉ có thể là một nền tảng tốt để phát triển GIS. Hệ thống TTĐL có nhiều điểm chung với CAD, có thể dùng để thiết kế, vẽ nhiều vật thể, đối tượng kỹ thuật trong một phạm vi rộng. Cả GIS và CAD đều cần thiết khả năng thể hiện các thực thể liên quan trong một hệ thống quan trắc, thể hiện các thuộc tính phi không gian và các mối quan hệ không gian. Khối lượng lớn và sự khác biệt của các dữ liệu đầu vào và các nét đặc trưng riêng biệt của các phép phân tích dữ liệu của hệ GIS là sự khác biệt với hệ CAD. Điều này có thể làm cho một phần mềm của CAD trở thành không thích hợp với một GIS và ngược lại.

Một GIS không chỉ đơn thuần là một hệ thống máy tính chuyên làm bản đồ mặc dù nó có thể tạo ra bản đồ với các tỉ lệ khác nhau, trong các hệ qui chiếu khác nhau và với hệ thống ký hiệu khác nhau. GIS là công cụ để phân tích dữ liệu. Ưu điểm lớn của GIS là giúp chúng ta xác định được các mối quan hệ không gian giữa các đối tượng địa lý. GIS không coi lưu trữ một bản đồ hoàn chỉnh hoặc lưu trữ ảnh hay khung nhìn đặc biệt là chính mà chủ yếu GIS lưu trữ các số liệu giúp chúng ta có thể tạo ra các khung nhìn và vẽ chúng theo mục đích mong muốn. Bởi vậy cần phải phân biệt GIS là một hệ thống thông tin không gian đặc biệt (Hình 1.7). Hệ thống thông tin đất (LIS) thực chất là một

GIS hoạt động ở tầm vi mô nghiên cứu các lĩnh vực quản lý và sử dụng các thông tin tài nguyên đất.



Hình 1.7 Phân biệt hệ thống thông tin địa lý và hệ thống thông tin đất

1.6 Các thành phần chính của Hệ thống thông tin địa lý

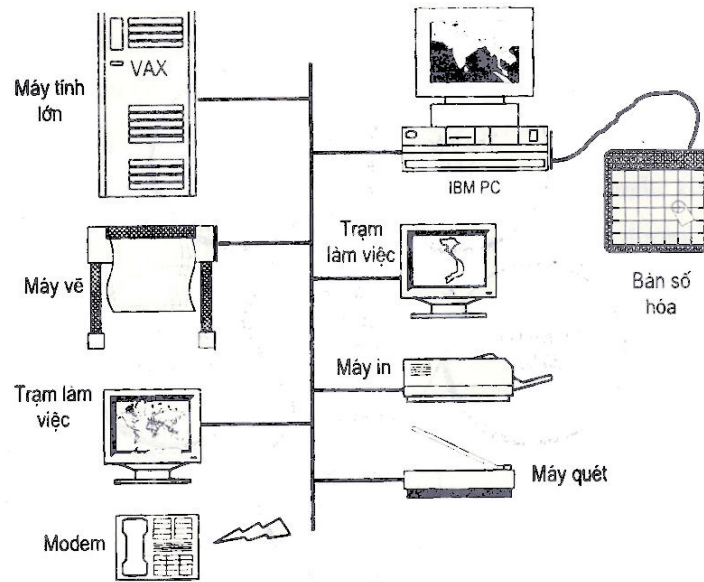
GIS đòi hỏi sự cung cấp một tập hợp các công cụ và phương pháp để người sử dụng có thể tổ chức thao tác và biểu diễn dữ liệu địa lý cho lĩnh vực áp dụng riêng của mình. Kỹ thuật là phần cứng và phần mềm của công nghệ. Công nghệ bao gồm cơ sở khoa học và kỹ thuật để thể hiện các nguồn dữ liệu khác nhau. Vì vậy các thành phần cơ bản để tạo nên một hệ GIS bao gồm các thiết bị tin học (phần cứng của hệ thống), chương trình quản trị dữ liệu (phần mềm của hệ thống), nguồn nhân lực sử dụng hệ thống, nguồn dữ liệu và phương pháp sử dụng hệ thống.

1.6.1 Phần cứng (hardware)

Các thành phần cứng thông thường của một Hệ thống thông tin địa lý thông thường được thể hiện trong Hình 1.8.

Phần cứng của một hệ GIS gồm máy vi tính, cấu hình và mạng công việc của máy tính, các thiết bị ngoại vi nhập xuất dữ liệu và lưu trữ dữ liệu. Các máy tính có thể làm việc độc lập hoặc có thể được đặt vào một mạng liên kết. Các thiết bị nhập dữ liệu như bàn số hóa hoặc máy quét dùng để chọn lọc các đặc tính địa lý từ một bản đồ hay ảnh nguồn vào hệ thống máy tính dưới dạng dữ liệu số vectơ hay ma trận dạng lưới.

Bộ phận điều khiển trung tâm (CPU) được nối với bộ phận lưu trữ (diskdrive) làm nhiệm vụ lưu trữ dữ liệu và chương trình máy tính. Các thiết bị ngoại vi khác như máy in, máy vẽ (plotter) thường được dùng để trình bày, hiển thị và in các dữ liệu kết quả đã được xử lý. Các ổ đĩa DVD, CD, modem được sử dụng đồng thời trong việc lưu trữ các dữ liệu đầu vào và ra của hệ thống hay đóng vai trò chuyển dữ liệu giữa các hệ thống thông tin với nhau. Người sử dụng có thể thể hiện dữ liệu như bản đồ trên màn hình từ máy tính và các thiết bị ngoại vi như máy quét, máy in.

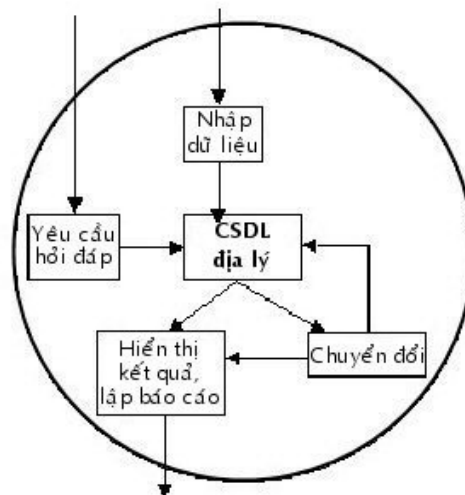


Hình 1.8 Các thành phần phần cứng của GIS

1.6.2 Phần mềm (software)

Một hệ GIS thường có chương trình máy tính có khả năng lưu trữ và quản trị các dữ liệu địa lý gồm năm phụ hệ kỹ thuật (subsystem) chủ yếu sau:

- Nhập dữ liệu;
- Lưu trữ dữ liệu và quản lý dữ liệu;
- Chuyển đổi dữ liệu;
- Hiển thị dữ liệu và báo cáo kết quả.
- Giao diện với người dùng.

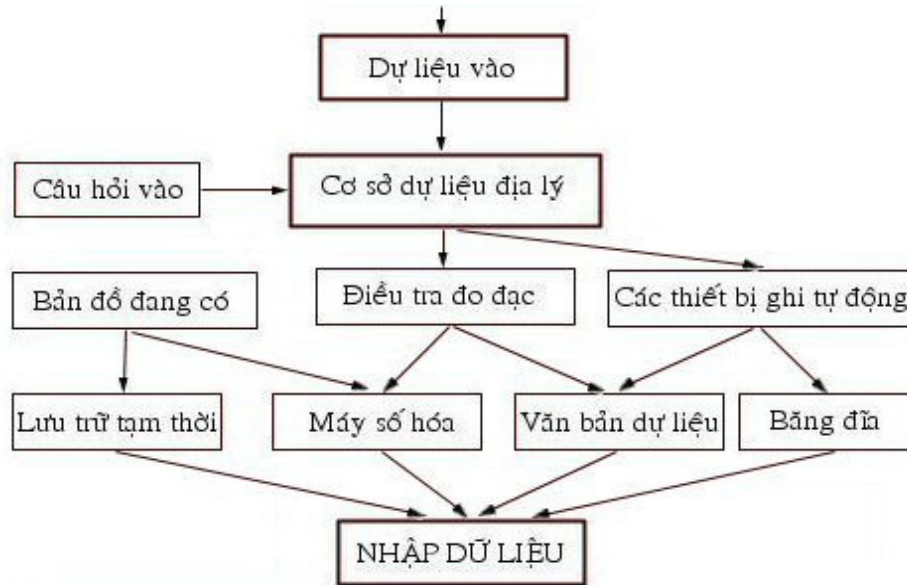


Hình 1.9 Các chức năng thành phần chính của một hệ quản trị dữ liệu địa lý

❖ Chức năng nhập dữ liệu

Thành phần nhập dữ liệu bao gồm tất cả các tác vụ liên quan đến chuyển đổi dữ liệu thu thập được ở khuôn dạng bản đồ có sẵn, số liệu thực địa, các bộ phận thu cảm ứng (bao gồm ảnh hàng không, ảnh vũ trụ và các cách thu thập dữ liệu gián tiếp khác) thành dạng số. Có rất nhiều công cụ máy tính khác nhau để thực hiện công việc này: Bàn phím, bàn số hoá, tệp text, máy quét ảnh, băng từ, đĩa cứng.

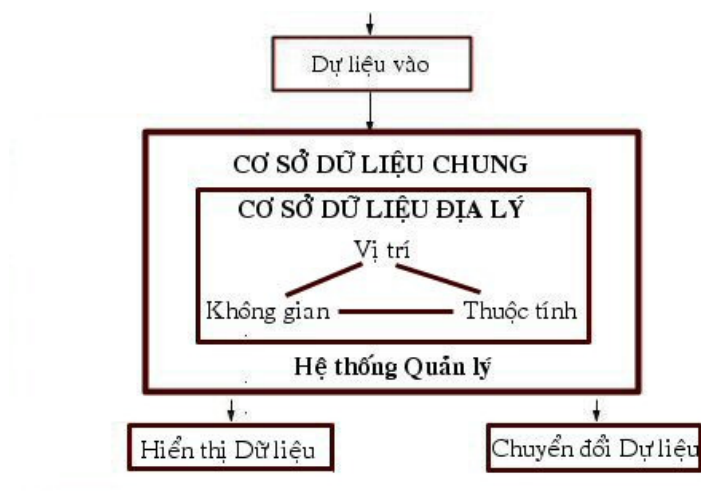
Chức năng nhập dữ liệu có các thành phần thể hiện theo sơ đồ Hình 1.10. Nhập dữ liệu và chỉnh sửa dữ liệu là các thao tác rất cần thiết để xây dựng một cơ sở dữ liệu địa lý (sẽ được trình bày kỹ ở chương 4).



Hình 1.10 Sơ đồ cấu trúc của chức năng nhập dữ liệu

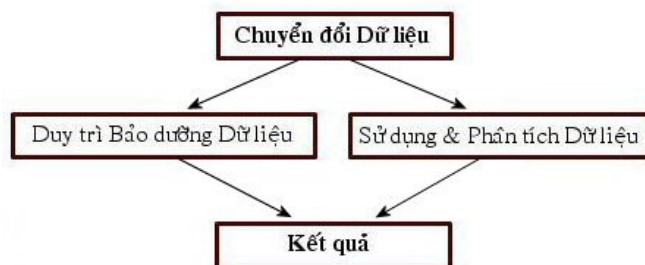
Lưu trữ dữ liệu và quản lý dữ liệu

Là chức năng hoạt động quan trọng nhất của một phần mềm GIS (Hình 1.11). Dữ liệu với các tính chất như vị trí, các liên kết (quan hệ không gian) và các thuộc tính của các nguyên tố địa lý như điểm, đường, vùng đại diện cho các thực thể trên bề mặt quả đất được các chương trình máy tính tổ chức cơ sở dữ liệu theo Hệ thống quản trị dữ liệu (DBMS). Cấu trúc và phương pháp tổ chức dữ liệu sẽ được trình bày kỹ hơn ở chương 3.



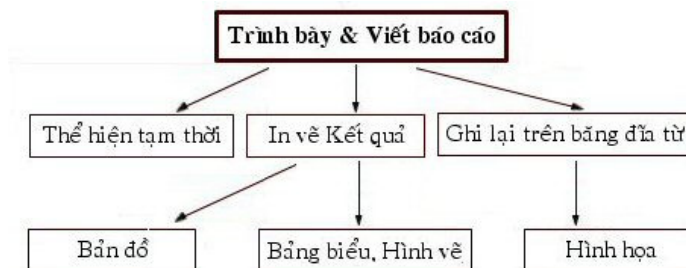
Hình 1.11 Cấu trúc của một cơ sở dữ liệu địa lý

Chức năng thể hiện sự trao đổi, chuyển đổi dữ liệu của phần mềm một HTTTDL rất phong phú giúp cho người sử dụng có thể duy trì bảo dưỡng và cập nhật các dữ liệu đã có sẵn hoặc có thể nhập xuất dữ liệu với một hay nhiều phần mềm tin học khác đồng thời giúp cho khả năng phân tích các dữ liệu được thực hiện linh hoạt và chính xác hơn.



Hình 1.12 Chức năng chuyển đổi dữ liệu

Chức năng trình bày kết quả được coi là sự hiển thị dữ liệu và các kết quả phân tích đến người sử dụng. thông tin sau các chức năng nhập, chỉnh lý và phân tích các dạng dữ liệu. Hình 1.13. cho thấy kết quả có thể trình bày với nhiều hình thức như thể hiện tạm thời trên màn hình máy tính, in vẽ kết quả dưới dạng bảng, biểu, bản đồ hoặc ghi lại trên băng, đĩa từ, phim, v.v.



Hình 1.13 Chức năng trình bày kết quả

Các sản phẩm phần mềm

Trên thị trường thế giới hiện nay có các phần mềm GIS có khả năng đáp ứng với rất nhiều ứng dụng khác nhau. Một số phần mềm có thể tải được thậm chí miễn phí từ internet. Người lựa chọn phần mềm cần hiểu rõ được khả năng đáp ứng của phần mềm mà mình lựa chọn trên cơ sở bốn chức năng cơ bản: nhập dữ liệu, chỉnh lý dữ liệu, phân tích dữ liệu và biểu thị kết quả. Các công ty cung cấp phần mềm GIS chính trên thế giới hiện nay là ESRI, Intergraph, Landmark Graphic và Mapinfo. Các phần mềm có thể được sử dụng trên các hệ điều hành WINDOWS 9X, WINDOWS NT, WINDOW XP, Unix, và Machintosh. Phụ lục A giới thiệu chi tiết đặc điểm của các phần mềm GIS phổ biến hiện nay trên thế giới.

Các sản phẩm phần mềm cho các bản đồ số: Đối tượng của các phần mềm này là số hoá, vẽ và thành lập bản đồ, dùng để quản lý các tư liệu bản đồ số, sửa chữa, cập nhật các thông tin trên bản đồ, xuất bản bản đồ (MicroStation, AutoCAD).

Các sản phẩm quản trị bản đồ: Các sản phẩm này cũng có các chức năng cập nhật thông tin bản đồ, ngoài ra thêm chức năng quản trị các thông tin bản đồ và thông tin thuộc tính của bản đồ. Chúng có khả năng liên kết dữ liệu bản đồ với các dữ liệu phi không gian. Các chức năng chủ yếu là thiết lập bản đồ thống kê theo các thuộc tính, hiển thị và in ấn (MapInfo, Arc/View).

Các sản phẩm phần mềm quản trị và phân tích không gian: Các sản phẩm này là mức phát triển cao hơn, ngoài các chức năng quản trị, cập nhật thông tin, kết nối dữ liệu thuộc tính, các phần mềm này còn có thêm các chức năng phân tích dữ liệu không gian. Các phép toán không gian cơ bản hiện nay đã trở thành mặc nhiên trong hệ GIS. Với các chức năng phân tích này người ta đã hoàn thiện không gian dữ liệu hình học trong CSDL GIS (Arc/Info, ArcGIS, ILWIS, MGE, Geomedia Professional).

1.6.3 Dữ liệu GIS

Có thể coi thành phần quan trọng nhất trong một hệ GIS là dữ liệu. Các dữ liệu không gian và dữ liệu thuộc tính. Hệ GIS sẽ kết hợp dữ liệu không gian với các nguồn dữ liệu khác, thậm chí có thể sử dụng DBMS để tổ chức lưu trữ và quản lý dữ liệu.

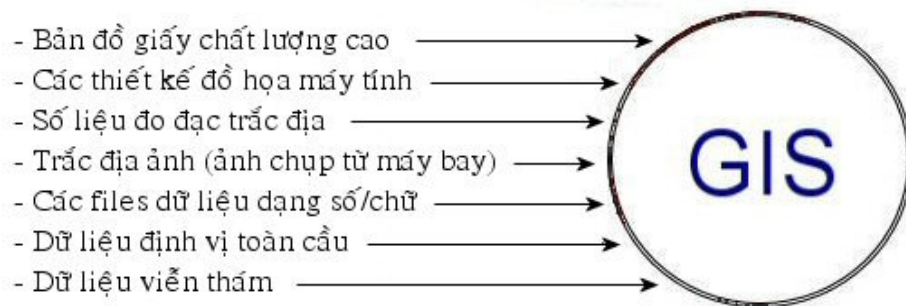
Các tác vụ liên quan đến dữ liệu là:

Thu thập dữ liệu :

Sự hình thành từ những năm 1960 và phát triển tốc độ của máy vi tính vào công tác xây dựng bản đồ và xử lý các thông tin không gian đã dẫn đến sự ra đời của các hệ thống thông tin địa lý (GIS) để thu nhận và xử lý các dữ liệu không gian từ nhiều nguồn dữ liệu khác nhau (Hình 1.14).

Nguồn dữ liệu vào của HTTTĐL rất đa dạng chủ yếu ở hai loại dữ liệu chính là dữ liệu không gian và dữ liệu thuộc tính. Các dữ liệu không gian mà GIS có thể thu nhận bao gồm các bản đồ giấy có giá trị sử dụng cao. Đây là những bản đồ được vẽ và in từ máy vi tính chứa các thông tin đáp ứng được các yêu cầu về chuyên môn, cơ sở pháp lý và thời gian phản ánh thông tin. Các số liệu đo đạc bằng các phương pháp trắc địa mặt đất như đo góc và khoảng cách bằng máy kinh vĩ điện tử, toàn đạc điện tử có thể được cung cấp trực tiếp hay gián tiếp cho hệ thống để thành lập nên các bản đồ cơ sở với tỉ lệ lớn. Nguồn ảnh chụp từ máy bay (phép quang trắc) bằng các máy chụp ảnh đặc biệt có thể nhập trực tiếp vào các phần mềm GIS bằng các công cụ quang trắc hiện đại. Độ chính xác của phương pháp phụ thuộc vào tỷ lệ ảnh và tầm bay của máy bay. Một ảnh chụp có tỉ lệ 1:10000 thì độ chính xác của dữ liệu véc tơ trên ảnh là 3 – 20 cm.

Hệ thống định vị toàn cầu (GPS) cung cấp dữ liệu trắc địa vệ tinh rất hiệu quả trong việc thành lập hệ thống mốc đo đạc ở các cấp toàn cầu, quốc gia và cơ sở. GPS là hệ thống đo đạc bằng vệ tinh bao gồm một chùm 24 vệ tinh quay quanh Trái đất ở độ cao hơn 2000 km. Các vệ tinh GPS cùng phát sóng tín hiệu chính xác đến máy thu GPS với các mã thông tin P (Precision code) và CA (Charge Acquisition code). Có 2 loại máy thu là GPS quân sự và GPS dân sự. Mã P phức tạp hơn mã CA nên chỉ có máy thu GPS quân sự mới nhận biết được với độ chính xác khoảng 1m. Máy thu GPS dân sự nhận biết được mã CA với độ chính xác từ 15m – 100m. Máy thu GPS hiển thị vị trí bằng chữ và số của lưới tọa độ chọn trước. Các lưới tọa độ thường được sử dụng là UTM và kinh vĩ độ. Khi sử dụng GPS ta có dữ liệu véc tơ 2 D hay 3 D. Các dữ liệu dựa trên Hệ tọa độ tâm Trái đất WGS-84. Các số liệu GPS được GIS sử dụng để kiểm soát tọa độ xác định vị trí chính xác của các đối tượng địa lý trong hệ thống.



Hình 1.14 Nguồn dữ liệu vào của một hệ GIS

Ngày nay dữ liệu viễn thám (remote sensing, RS) đã trở thành nguồn cung cấp thông tin đầu vào hiện đại, kịp thời và chính xác cho GIS. Có thể nói công nghệ viễn thám (RS) đã giúp cho GIS ngày càng hoàn thiện hơn. Dữ liệu viễn thám là các dữ liệu thu thập được từ xa bằng các thiết bị cảm biến trên máy bay hay vệ tinh. Các dữ liệu này đặc trưng cho các vùng rộng lớn được đo đạc và phân tích, không bị ảnh hưởng bởi các vùng lân cận. Ảnh viễn thám cho chúng ta có thể quan sát và cập nhật các thông tin về Trái đất thường xuyên và hiệu quả nhất. Một số ảnh vệ tinh đang được sử dụng rộng rãi trên thế giới hiện nay là ảnh LANDSAT (USA) được phóng lên quỹ đạo năm 1972, ảnh Spot (Pháp) được phóng năm 1986, ảnh ASTER, Quick Bird của Nhật bản phóng năm 1995. Viễn thám cho chúng ta bức tranh rõ nét nhưng rời rạc về các hiện tượng trên Trái đất nhưng GIS lại có khả năng phân tích chi tiết và liên kết các bức tranh đó trở thành hệ thống. Với công nghệ giải đoán ảnh bằng mắt hay xử lý ảnh số bằng máy tính qua kỹ thuật phân lớp (nhận dạng mẫu) có thể nhận biết được các đối tượng trên ảnh. Những kỹ thuật của viễn thám đi đôi với khả năng phân tích các bản đồ, sự tạo ra những mô hình thống kê, vẽ các bản đồ, HTTĐL tạo ra bức tranh tổng thể chặt chẽ, không đơn điệu rời rạc về các đối tượng nghiên cứu. Đồng thời quyết định của HTTĐL là công cụ mạnh mẽ vượt qua những vấn đề địa lý và môi trường. Ngày nay công nghệ 3S (GPS, RS và GIS) đã giúp kho tàng kiến thức khoa học hiểu biết về hệ thống Trái đất ngày càng phong phú và giá trị hơn.

GIS cũng thu nhận các số liệu thuộc tính là các thông tin đặc trưng đi kèm với các dữ liệu không gian của các đối tượng trên bản đồ hay trên ảnh. Đây là các dữ liệu ở dạng văn bản hoặc các số liệu thống kê thu được trong công tác điều tra dã ngoại, hoặc là các số liệu phân tích trong phòng thí nghiệm v.v. được lưu trữ dưới dạng các tệp tin dạng chữ hoặc dạng số có thể nhập trực tiếp hoặc gián tiếp vào hệ thống GIS.

Các dữ liệu được thu nhận vào một hệ GIS có thể từ nhiều nguồn khác nhau như từ hệ thống quan trắc môi trường, hoặc từ các cơ sở dữ liệu đã có sẵn. Các dữ liệu không gian có chứa các thuộc tính như các dữ liệu thô, số liệu thí nghiệm hoặc các chỉ số phải được chuyển đổi sang dạng dữ liệu vào của GIS. Các dữ liệu không gian ban đầu cũng ở dạng như bản đồ giấy, các file dữ liệu dạng ảnh máy bay hay ảnh viễn thám, v.v. Ngoài ra các dữ liệu địa lý khác như tỉ lệ bản đồ (ảnh), phép chiếu, vị trí địa lý, năm thu thập dữ liệu và khả năng thực hiện cũng phải được xác định và chuẩn hóa để đưa vào hệ thống.

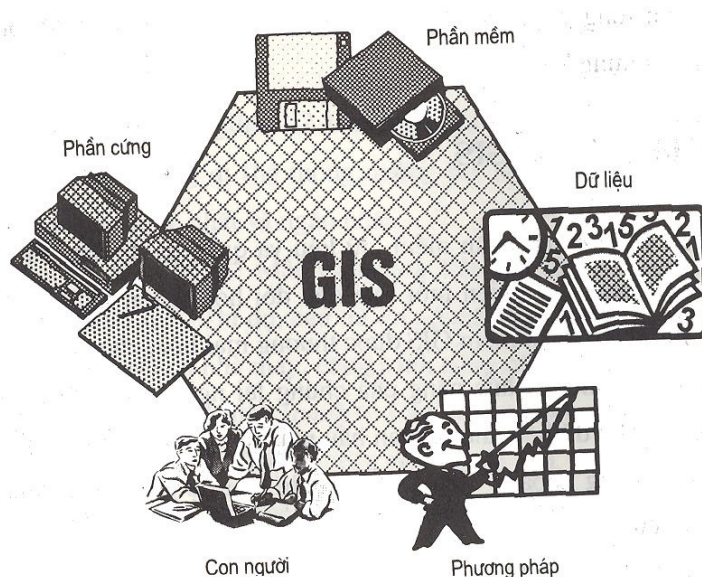
Nhập dữ liệu:

Đây là giai đoạn thực hiện các thao tác như trút dữ liệu số hóa, quét, hoặc nhập dữ liệu từ bàn phím. Quá trình lựa chọn các dữ liệu để nhập vào hệ thống gắn liền với sự trợ giúp của các thiết bị ngoại vi như bàn số hóa, máy quét đòi hỏi thời gian và nhân công kỹ thuật. Việc nhập dữ liệu từ cơ sở dữ liệu khác cũng đòi hỏi phải định dạng lại (reformat) khuôn dạng dữ liệu. Những dữ liệu địa lý thường là dữ liệu khó định dạng nhất. Sự chuyển đổi khuôn dạng dữ liệu là yêu cầu để có thể làm việc được với các dữ liệu GIS và phải được đặt ra trong qui trình nhập dữ liệu đối với cán bộ kỹ thuật.

Các thao tác chỉnh lý và thể hiện dữ liệu cho phép thông báo các dữ liệu không gian với bất kỳ một hay tổ hợp các thuộc tính đi kèm dưới dạng bản đồ và bảng thuộc tính.

Phân tích dữ liệu GIS đòi hỏi chương trình máy tính thực hiện các thao tác như tính toán diện tích, khoảng cách, phép đệm, chồng xếp không gian, thống kê không gian hoặc các hoạt động logic, v.v. Khả năng và sự linh hoạt của một hệ GIS có thể chứng minh qua quá trình phân tích dữ liệu. Thiết kế, xây dựng chính xác một cơ sở dữ liệu sẽ đáp ứng được các yêu cầu của người sử dụng và tránh được những giả thuyết sai lệch.

Mô hình hóa các dữ liệu không gian cũng là một chức năng rất quan trọng. Một hệ GIS có thể chứa mô hình phương pháp số hoặc có thể là một môi trường máy tính độc lập. Khi sử dụng một hệ thống riêng biệt để mô hình hóa các kết quả đưa vào hệ thống GIS cần phải được xem xét, chỉnh sửa và được lưu giữ như một phần của cơ sở dữ liệu cho các mục đích phân tích khác.

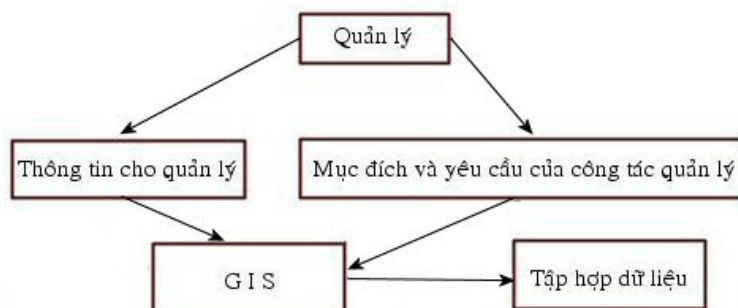


Hình 1.15 Quan hệ giữa các thành phần của GIS

1.6.4 Người sử dụng

Các yếu tố về kỹ thuật (phần cứng, phần mềm và cơ sở dữ liệu) của một hệ thống thông tin địa lý sẽ không có hiệu quả nếu như thiếu kỹ năng sử dụng của con người, không được vận dụng vào một hoàn cảnh thích hợp và thông nhất với chính sách phát triển của Nhà nước.

Nguồn nhân lực để vận hành một hệ GIS bao gồm các cán bộ vận hành, cán bộ kỹ thuật chuyên môn và các nhà quản lý. Cán bộ vận hành là những đồ họa viên thiết kế trình bày bản đồ theo các tiêu chuẩn và qui phạm về hệ thống ký hiệu và bản đồ, các cán bộ thu nhận dữ liệu có nhiệm vụ biến đổi dữ liệu bản đồ giấy sang dạng số cho một hệ GIS. Cán bộ kỹ thuật bao gồm các nhà phân tích thông tin (information analyst) giải quyết các vấn đề đặc biệt và đáp ứng các yêu cầu thông tin cho người sử dụng. Cán bộ quản lý hệ thống có trách nhiệm bảo trì hoạt động của hệ thống (phần cứng, phần mềm). Lập trình viên là những người sử dụng những ngôn ngữ máy tính để đưa các vấn đề chuyên môn đã được chuẩn bị bởi phân tích viên trở thành các chương trình. Người quản lý dữ liệu trợ giúp cho các phân tích viên, lập trình viên và người sử dụng trực tiếp tổ chức các đặc tính địa lý thành các lớp (layer), xác minh nguồn dữ liệu, gán mã cho các dữ liệu phi không gian và giải trình thông tin về nội dung của cơ sở dữ liệu.



Hình 1.16 Sơ đồ phối hợp giữa công tác quản lý và công nghệ GIS

Nhà quản lý có nhiệm vụ tổ chức và điều hành thường xuyên sự thực hiện một dự án GIS và quản lý việc sản xuất các kết quả thông tin (output data) theo các yêu cầu của tổ chức, sản phẩm thông tin đầu ra phải đáp ứng được các yêu cầu chuyên môn và phải được xã hội chấp nhận.

1.6.5 Phương pháp

GIS là một loại hệ thống thông tin đặc biệt nên tùy từng mục đích và hoàn cảnh ứng dụng cụ thể mà lựa chọn và thiết kế hệ thống cho phù hợp. Muốn một hệ GIS hoạt động có hiệu quả đòi hỏi phải có sự phối hợp tốt giữa các nhà quản lý, khoa học chuyên môn và các kỹ sư thiết kế xây dựng hệ thống. Xây dựng một hệ GIS đơn giản hay hiện đại là tùy thuộc vào hoàn cảnh cụ thể mục đích và yêu cầu cung cấp thông tin cho các lĩnh vực chuyên môn. Một dự án GIS chỉ thành công khi nó được quản lý tốt và người sử dụng hệ thống phải có kỹ năng tốt, nghĩa là phải có sự phối hợp tốt giữa công tác quản lý và công nghệ GIS.

1.7 Một số ứng dụng GIS trong các ngành khoa học kinh tế và kỹ thuật

Công nghệ GIS đã có vai trò rất lớn trong xây dựng bản đồ và phân tích không gian trong việc thu nhận dữ liệu, phân tích dữ liệu và thể hiện dữ liệu như xây dựng bản đồ địa hình, bản đồ địa chính, trong các thiết kế đồ họa, thiết kế xây dựng, địa chất, trong

nghiên cứu toán học của các biến không gian, trong khoa học đất, điều tra đo đạc, trắc địa ảnh qui hoạch và phát triển nông thôn, xây dựng mạng lưới giao thông công cộng, trong phân tích và xử lý ảnh viễn thám. Sau đây là một số ứng dụng của GIS trong một số lĩnh vực khoa học và kỹ thuật quan trọng.

HTTĐL và cảnh quan

Trong khung cảnh của nghiên cứu cảnh quan, HTTĐL được sử dụng trong đánh giá tác động môi trường, sự tương thích môi trường, lập mô hình nước ngầm, quản lý tài nguyên, phát triển chính sách, đưa ra quyết định và thông qua công chúng.

Vai trò của HTTĐL trong nghiên cứu cảnh quan sinh thái nhằm cung cấp một cấu trúc dữ liệu cho lưu trữ và quản lý hiệu quả hệ thống sinh thái trên những vùng có phạm vi không gian lớn như:

- Cho phép tập hợp và phân chia dữ liệu không gian ở các tỉ lệ;
- Xác định vùng nghiên cứu hoặc vùng nhạy cảm môi trường;
- Hỗ trợ phân tích thống kê không gian của sự phân bố sinh thái;
- Nâng cao khả năng chiết xuất thông tin viễn thám;

Cung cấp dữ liệu đầu vào và các tham số cho mô hình hoá hệ thống sinh thái (Stow, 1993), HTTĐL năng động cung cấp cơ sở cho mô hình hoá mô phỏng của những quá trình cảnh quan động phụ thuộc thời gian) (Tài liệu dự án "Capacity building for sustainable development", Viện Địa lý Việt nam).

HTTĐL trong khoa học đất

Sự sử dụng HTTĐL trong khoa học đất không phải là vấn đề mới mà đã được nhiều tài liệu giới thiệu từ những năm 70. Tuy nhiên trong quá khứ những sự quan sát đã được thực hiện tại những điểm hiện trường và thông tin về đất chỉ dựa trên dữ liệu phẫu diện đất (mô tả, phân tích trong phòng thí nghiệm). Ngày nay, với những kỹ thuật tổ hợp của thành lập bản đồ, nội suy viễn thám và mô hình hoá, HTTĐL trở nên rất hiệu quả cho những người điều tra về đất và những nhà quy hoạch môi trường. Bản đồ đất thế giới tỉ lệ 1: 5.000000 được thành lập (FAO 1989) dựa trên cơ sở dữ liệu đất và địa hình (SOTER) với đầy đủ các dữ liệu không gian và tính chất lý hóa học của đất với đặc điểm địa hình của tất cả các loại đất trên thế giới. Trước đây những nhà khoa học đất quan tâm đến các vấn đề nghiên cứu gắn liền với thành lập bản đồ đất, ngày nay đã cố gắng miêu tả sự biến đổi không gian của đất. Với kỹ thuật HTTĐL, thông tin có thể được cung cấp dưới những dạng khác nhau cho những người sử dụng đất từ người quy hoạch đô thị ở các nước phương Tây tới chuyên gia tài nguyên đất ở các nước đang phát triển. HTTĐL ngày nay được thiết lập để khảo sát thực tế về đất toàn cầu, được sử dụng chủ yếu cho sản xuất bản đồ, lập các bản đồ thích nghi để đáp ứng yêu cầu người sử dụng thông tin về mục đích đặc biệt và cho mô hình hoá quá trình môi trường.

Các cơ quan khảo sát thổ nhưỡng quốc gia như Dịch vụ bảo tồn đất USDA sử dụng Hệ thống thông tin đất đai như một công cụ tổng quan cho thành lập bản đồ và cung cấp thông tin thổ nhưỡng cho người sử dụng. Sau đây là một số ứng dụng về bản đồ phân loại đất với các thuộc tính liên quan:

- Xây dựng bản đồ thể hiện cơ cấu của đất, xem xét các biến đổi đặc tính mỗi khoanh đất; Đổi tên, phân loại những vùng đất khác nhau; Chồng ghép logic các lớp thông tin đất với các lớp thông tin khí hậu, thảm che phủ đất, địa mạo, đá mẹ; Tính toán giá trị thuộc tính từ những lớp thông tin khác nhau hoặc ước tính

khối lượng đất đất hàng năm mất đi do xói mòn kết hợp sử dụng phương trình tổng hợp xói mòn đất (USLE) phổ dụng;

- Tạo một cơ sở dữ liệu đất;
- Sử dụng dữ liệu mẫu diện: Ước tính, mô tả định lượng về tính chất đất bao gồm các trị trung bình, giá trị phương sai của tất cả những tính chất định lượng của đất;
- Sản xuất những bản đồ thuộc tính đơn bằng nội suy dữ liệu mẫu diện đất trong nghiên cứu ô nhiễm đất; Tạo mô hình ước tính sản lượng mùa vụ hoặc mô hình định lượng khác từ mẫu diện đất và dữ liệu khác;
- Phân tích nhu cầu sử dụng phân bón cấp vùng và quốc gia.

Thử nghiệm phần mềm HTTĐL trên những tiêu bản đất và tính toán nhu cầu phân bón của đất. Hỏi đáp tìm kiếm được dựa trên cơ sở dữ liệu địa lý (bản đồ) và dữ liệu phi địa lý (nhóm đất lớn nhất, góc dốc, mức độ xói mòn, tưới tiêu, nhiễm mặn kiềm, tính sỏi, đá, phân loại tiềm năng sử dụng đất, sử dụng đất, dự đoán tình trạng dinh dưỡng NPK và pH) cho mỗi vùng.

Xử lý dữ liệu hỏi đáp trên cơ sở dữ liệu đất

Bảng thuộc tính được tạo ra trong Microsoft Excell và hỏi đáp dựa trên tính chất của đất. Kết thúc của mỗi hỏi đáp, một lớp được thiết lập cho một tính chất và tạo ra một bản đồ của mỗi lớp. Các bản đồ này có thể là:

- Bản đồ hiện trạng N, P và K;
- Bản đồ xác định những nhóm đất chính với những tổ hợp tính chất khác nhau của đất;
- Bản đồ xói mòn đất.

Phân tích ước tính sản lượng cây trồng

Vì sự biến đổi của địa hình, dự báo những sản lượng khác nhau sẽ có trên những vị trí khác nhau. Tổ hợp độ dốc và độ sâu của đất được vận dụng để ước tính sự biến đổi sản lượng mong muốn trên các vùng với địa hình và đất khác nhau.

Ở Mỹ cục bảo vệ môi trường (US EPA), cơ quan đo đạc, địa chất (USGS), cơ quan bảo vệ đất (USSCS), cơ quan nghiên cứu vũ trụ và hàng không quốc gia (NASA), cục quản lý bầu trời và đại dương (NOAA) đã sử dụng công nghệ GIS ở mức tầm cỡ hàng đầu thế giới trong một số lĩnh vực sau:

Qui hoạch sử dụng đất

Trợ giúp các cơ quan quản lý Nhà nước và địa phương trong qui hoạch vùng và qui hoạch sử dụng đất các cấp.

Giảm bớt các ảnh hưởng bất lợi do sự phát triển (ví dụ qui hoạch phát triển đô thị) đối với các vùng cảnh quan tự nhiên.

HTTĐL và sự biến động toàn cầu

Các nhà khoa học và nghiên cứu môi trường ứng dụng GIS trong nghiên cứu các biến đổi môi trường toàn cầu dưới các tác động ảnh hưởng của các yếu tố tự nhiên như khí hậu, thời tiết và ảnh hưởng tác động của con người về mặt kinh tế văn hoá xã hội nhằm tìm ra các giải pháp thích hợp bảo vệ trái đất khỏi các nguy cơ hiểm họa của thiên nhiên.

Sự biến động toàn cầu (môi trường) là kết quả của những quá trình có thể được đặt vào tình cảnh quá khứ và hiện tại mà hiệu ứng của nó rất có thể ảnh hưởng tới tương lai. Những quá trình này rất khác biệt từ vùng nọ sang vùng kia vì bản chất tự nhiên và con người đang làm ảnh hưởng tới sự thay đổi môi trường. Sự phức tạp của cảnh quan và của tác động tương hỗ giữa các nhân tố ảnh hưởng tạo cho nó sự cần thiết để tiếp cận các nguyên lý nghiên cứu khác nhau về thay đổi môi trường.

Hiện nay, cộng đồng khoa học đã nhận ra những hậu quả tác động của hoạt động con người, kỹ thuật HTTĐL đang là một công cụ cần thiết trong nỗ lực tìm hiểu quá trình thay đổi toàn cầu. Những bản đồ khác nhau và những nguồn thông tin địa phương có thể được tổ hợp trong những mô hình mô phỏng sự tương tác của những hệ thống tự nhiên phức tạp. Cơ sở dữ liệu HTTĐL phục vụ nghiên cứu biến đổi toàn cầu chắc chắn sẽ mang tính toàn cầu về bản chất, đa không gian, đa thời gian và đa nguyên lý. Những cơ sở dữ liệu này sẽ là rất lớn. Những nghiên cứu thay đổi toàn cầu đòi hỏi tiếp cận nguyên lý quan sát trái đất phải là một hệ thống đầy đủ, cần thiết phải nâng cao phương pháp vận dụng các cơ sở dữ liệu (ví dụ phân tích các loại hình biến đổi, khuynh hướng biến đổi và sự liên kết giữa các loại hình biến đổi đó).

Chương trình khoa học hệ thống trái đất NASA

Mục tiêu của chương trình toàn cầu này là thu được những hiểu biết khoa học về toàn bộ hệ thống trái đất mức độ toàn cầu bằng những mô tả những phần trong các hợp phần của nó và những tương tác đã và tiếp tục tiến hoá như thế nào theo thời gian (Ủy ban Khoa học Hệ thống Trái đất, 1988; Clark và cộng sự, 1991).

Những ứng dụng của GIS trong quân sự cũng có nhiều thành tích nổi trội trong thời gian gần đây như xác định những vị trí thích hợp cho các tuyến phòng thủ quân sự, các vùng chiến lược để triển khai các hoạt động quân sự phù hợp với các loại vũ khí hiện đại.

Vai trò của HTTĐL trong việc điều tra dân số đã tăng thực sự trong những năm gần đây. Không chỉ trong việc điều tra dân số mà HTTĐL còn có tầm quan trọng rất cao trong kế hoạch hoá thống kê và phân tích thống kê dữ liệu dân số trước và sau mỗi giai đoạn. Một trong những hệ HTTĐL thực hiện vấn đề này là hệ thống US TIGER. Những nhiệm vụ của hệ thống TIGER là:

- cung cấp dữ liệu địa lý, thống kê cập nhật và gắn kết dữ liệu trong không gian của toàn bộ các bang ở Hoa kỳ;
- cung cấp hệ thống thành lập bản đồ tự động;
- cung cấp hệ thống dữ liệu bảng biểu thông qua những đơn vị hành chính tiêu chuẩn hoặc những tập hợp thống kê;
- cung cấp hệ thống thành lập và xuất bản bản đồ.

Những đòi hỏi này có thể được đạt được bằng cách thiết kế một hệ thống cơ sở dữ liệu không gian. Hệ thống cơ sở dữ liệu không gian TIGER bao gồm bốn dạng file (tệp) chính:

- những tệp dữ liệu không gian phân chia theo quốc gia;
- những tệp phân dữ liệu không gian chia theo địa hạt;
- liệt kê dữ liệu địa lý;
- những tệp dữ liệu làm việc tạm thời.

Sử dụng đất và sự thay đổi bề mặt (Land Use and Land Cover Change)

Chương trình này mang tính hợp tác toàn cầu. Các nhà khoa học thế giới trong các lĩnh vực khoa học kỹ thuật ứng dụng các hệ thống công nghệ cao như GIS, RS (viễn thám), và GPS kết hợp với các hệ thống nghiên cứu xã hội con người, v.v. nhằm tìm ra các nguyên nhân tác động đến những biến đổi của bề mặt trái đất, nghiên cứu các mối liên quan, ảnh hưởng và kết quả giữa các vấn đề sử dụng đất và các diễn biến thay đổi của các lớp phủ bề mặt đất (Land Cover).

HTTĐL cho phát triển nông thôn

Hiện nay, dân số thế giới đã đạt khoảng hơn 6 tỉ người. Người ta dự đoán rằng con số này sẽ tăng gấp đôi vào năm 2050. Sự tăng dân số chủ yếu ở những nước đang phát triển và điều này đi cùng với hiện tượng di cư dân số từ nông thôn ra thành thị. Sự tăng trưởng cơ học này sẽ gây ra những vấn đề nghiêm trọng trong công nghiệp, môi trường và sức khỏe con người. Không còn nghi ngờ gì nữa quy hoạch nông thôn sẽ đóng một vai trò quan trọng. Để tìm cách giải quyết những vấn đề này, những dữ liệu tin cậy gần đây về số lượng dân cư, tính chất và sự phân bố dân cư phải luôn sẵn có. Những nguồn thông tin về dân tộc học (thống kê và những yêu cầu thông tin) có chi phí lớn và không đầy đủ trong đa số các trường hợp ở những nước đang phát triển. Trong khung cảnh này, kỹ thuật viễn thám được tổ hợp với HTTĐL để điều tra và xử lý dữ liệu cho các nghiên cứu dân tộc học.

Ở Việt nam công nghệ GIS mới được chú ý đến trong vòng 15 năm trở lại đây. Tuy nhiên do những hạn chế về công nghệ, kỹ thuật và quản lý GIS mới chỉ được tập trung ở một số cơ quan nghiên cứu trung ương của một số ngành như nông lâm nghiệp, địa chất, khí tượng thủy văn, giao thông vận tải, quốc phòng. Phần lớn mới chỉ dừng lại ở mức xây dựng cơ sở dữ liệu cho các dự án nghiên cứu. Một số phần mềm lớn về GIS như ARCINFO, MAPINFO, MAPPING OFFICE đã được sử dụng ở nhiều nơi để xây dựng lại các bản đồ địa hình, địa chính, hiện trạng trên phạm vi toàn quốc. Sự kết hợp giữa công nghệ viễn thám và GIS đã bắt đầu được ứng dụng trong một số nghiên cứu về nông lâm nghiệp như trong công tác điều tra qui hoạch rừng (viện điều tra qui hoạch rừng), công tác điều tra và đánh giá và qui hoạch đất nông nghiệp của viện qui hoạch, thiết kế nông nghiệp, v.v. đã mở ra các triển vọng đáng chú ý trong việc ứng dụng GIS.

Ở các thành phố lớn GIS được ứng dụng trong qui hoạch kiến trúc đô thị, biểu diễn phối cảnh thành phố, điều hành mạng lưới giao thông, cứu hỏa cứu thương, quản lý bất động sản, cấp thoát nước đô thị, v.v.

Xây dựng và thành lập Bản đồ chuyên đề

Ở Việt nam hiện nay GIS có thể coi là một công cụ hữu hiệu giúp cho các cơ quan quản lý đất đai như đăng ký thống kê, qui hoạch, nông nghiệp, thủy lợi, v.v. xây dựng các bản đồ chuyên đề:

- Bản đồ địa hình;
- Bản đồ địa chính;
- Bản đồ đất nông nghiệp, lâm nghiệp;
- Bản đồ thủy lợi;
- Bản đồ hiện trạng sử dụng đất, bản đồ qui hoạch sử dụng đất;
- Bản đồ các vùng sinh thái.

Câu hỏi ôn tập

1. Hệ thống thông tin địa lý là gì?
2. Lịch sử phát triển của hệ thống thông tin địa lý? Xu hướng hiện nay của GIS?
3. Phân tích các nội dung cơ bản của một HTTĐL
4. Các yêu cầu đối với một HTTĐL hoạt động tốt là gì?
5. Phân tích vai trò của các thành phần trong một HTTĐL. Đối với các nước đang và chậm phát triển thành phần nào có ý nghĩa quyết định trong vấn đề xây dựng và phát triển một HTTĐL.
6. Nêu sự khác nhau cơ bản giữa các phần mềm tin học GIS như ARCINFO, MAPINFO, ILLWIS, ARCVIEW và các phần mềm đồ họa khác như AutoCAD, MicroStation, PhotoShop, và CoralDraw.
7. Nêu các ứng dụng của GIS trong các ngành khoa học kỹ thuật và kinh tế trên thế giới.
8. Hãy nêu tóm tắt nội dung một dự án GIS đã hoặc đang thực hiện ở Việt nam trong lĩnh vực quản lý tài nguyên mà bạn đã tìm hiểu?

Chương Hai

MỘT SỐ KHÁI NIỆM VỀ HỆ THỐNG THAM CHIỀU KHÔNG GIAN

Trước khi giới thiệu về các hệ tham chiếu không gian được sử dụng trong Hệ thống Thông tin Địa lý (HTTTĐL), ta cần hiểu rõ các khái niệm cơ bản về bản đồ và các hệ tọa độ. Chương này cung cấp cho người đọc các kiến thức chung về bản đồ, sau đó trình bày các hệ tọa độ phổ biến hiện nay trên thế giới và cuối cùng sử dụng những kiến thức về bản đồ và hệ tọa độ để giải thích về các hệ tham chiếu không gian đặc trưng cho HTTTĐL.

2.1 Khái niệm về bản đồ

Quá trình đồ họa được coi như là sự kết giao giữa khoa học và nghệ thuật. Các yếu tố căn bản của đồ họa kết hợp giữa lý luận và thực tiễn được thể hiện chung cho tất cả các loại bản đồ.

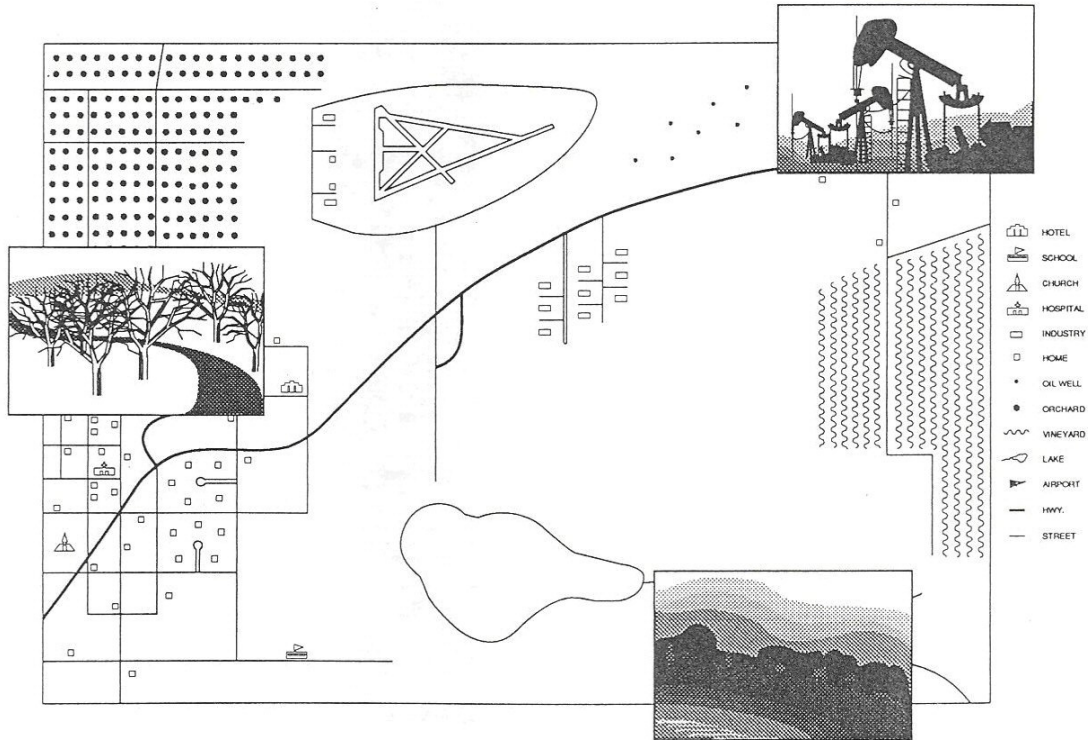
Bản đồ theo Robinson (1984) được định nghĩa như sau

Bản đồ là sự trình bày thu nhỏ điển hình về diện tích. Như vậy mỗi một bản đồ phải thể hiện được mối quan hệ đã xác định giữa diện tích thực tế và diện tích trên bản đồ. Mối quan hệ này quan trọng trước hết. Tỷ lệ bản đồ sẽ giới hạn cả về hình thức và cách trình bày các thông tin trên bản đồ.

Bản đồ là sự chuyển đổi tọa độ của một bề mặt không bằng phẳng trên quả đất về một lưới chiếu phẳng. Việc chọn phép chiếu sẽ ảnh hưởng đến việc một bản đồ sẽ được sử dụng ra sao?

Bản đồ là sự khái quát hoá thế giới hiện thực: Bản đồ là sự đại diện đồ họa cho một vùng, vì vậy nó thể hiện dữ liệu mà người vẽ đã lựa chọn cho mục đích sử dụng. Các thông tin đã được trình bày trên bản đồ đã được phân loại và đơn giản hoá để tăng khả năng làm việc của người sử dụng với bản đồ chứa các ký hiệu đại diện cho các thành phần của thế giới thực. Một số ký hiệu của số ít bản đồ được chấp nhận với ý nghĩa phổ biến nhưng cũng có một vài loại bản đồ có các qui định đặc biệt về ký hiệu trình bày dữ liệu với sự thay đổi của các ký hiệu như đường, điểm, màu sắc, dạng chữ hoặc các patterns (nét kẻ gạch).

Bản đồ bao gồm cả dữ liệu thông tin quan trọng đầu vào của một Hệ thống Thông tin Địa lý (HTTTĐL) cũng như là sự thể hiện các kết quả phân tích đầu ra của hệ thống đó. Bản đồ cũng bao gồm 2 yếu tố cơ bản của thực thể: Giới hạn thể hiện vị trí của thực thể trong một không gian 2 chiều và các thuộc tính tại giới hạn thể hiện các số đo về số lượng và chất lượng của thực thể tại vị trí đó. Từ các tính chất cơ bản này một sự thay đổi về các tính chất quan hệ không gian và các tính chất hình học cũng có thể xác định được như khoảng cách, hướng, sự liên tục và độ chính xác. Vì vậy bản đồ là một công cụ rất hữu hiệu trong việc truyền tải các mối quan hệ không gian.



Hình 2.1 Bản đồ là một hệ thống truyền tải các thông tin không gian

Bản đồ địa hình có thể coi là loại thông tin nền (basemap) cho mọi bản đồ chuyên môn khác. Các bản đồ này thường được gọi là bản đồ chuyên đề vì nó chỉ chứa thông tin về một chuyên đề nào đó. Để thành lập được bản đồ chuyên đề người ta chỉ việc chồng xếp lớp thông tin chuyên đề lên lớp thông tin nền (địa hình).

Thuật ngữ "Bản đồ chuyên đề" được sử dụng rất rộng rãi và được sử dụng không chỉ cho bản đồ về những mục tiêu chung chung như "đất", "nước" mà còn có được dùng biểu diễn rất chi tiết các tính chất, thuộc tính của đối tượng nghiên cứu. Ví dụ độ phì nhiêu của đất, độ nhiễm mặn, độ chua, các đặc tính lý học của đất v.v. Các chuyên đề có thể là về định tính mà cũng có thể về định lượng.

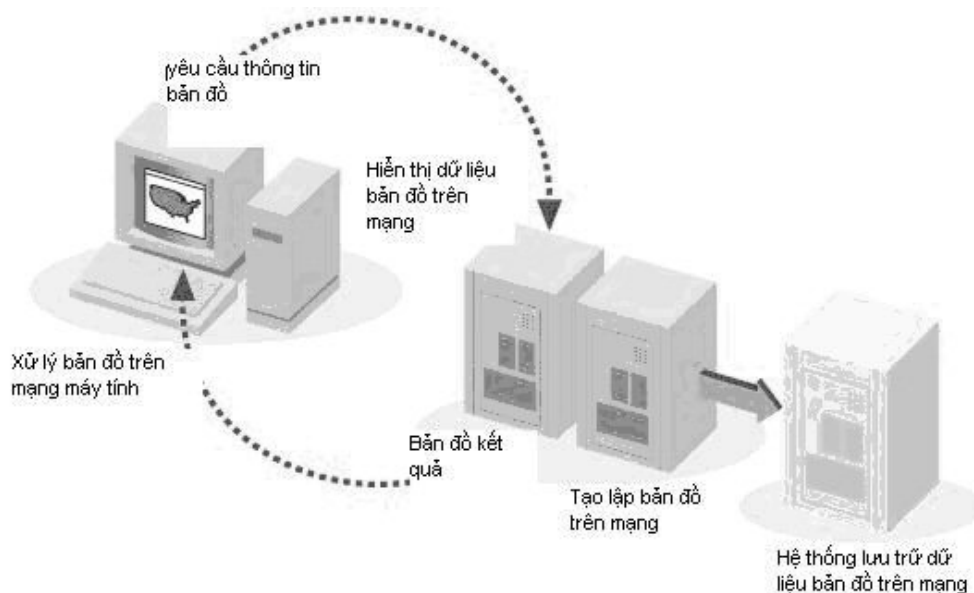
Cả hai kiểu trên đều có thể được biểu diễn dưới dạng các bản đồ phân bố theo giá trị (choropleth maps) và các giá trị đặc trưng cho mỗi vùng phân chia. Dữ liệu định lượng cũng có thể được mô tả trên bản đồ bằng cách mô hình hoá toán học thành một mặt cong liên tục. Theo mặt cong này người ta có thể biểu diễn các đường đồng mức giống như đường đồng mức trên bản đồ địa hình.

Mọi bản đồ có một điểm chung: đó là dữ liệu bản đồ được vẽ trên giấy hoặc phim. Thông tin được biểu diễn dưới dạng điểm, đường và vùng. Các thực thể địa lý được thể hiện trên bản đồ bằng nhiều kỹ xảo, chẳng hạn các ký hiệu hoặc các màu sắc hoặc text, mà ý nghĩa giải thích của chúng được để trong phần chú giải; ở đó người ta ghi lại thông tin một cách chi tiết hơn.

Bản đồ giấy vẽ tay cùng với chú giải tạo thành một cơ sở dữ liệu cực kỳ quan trọng cho công tác thu thập dữ liệu, mã hoá các thông tin. Bản đồ giấy có một số đặc trưng sau:

- Khối lượng dữ liệu rất lớn, được phân loại một cách rõ ràng, hiểu được và thể hiện được trên giấy, do đó có nhiều thông tin thứ yếu có thể bị bỏ qua.
- Bản đồ được vẽ chính xác và thể hiện các đối tượng phức tạp đều phải rất rõ ràng. Một bản đồ có thể được chia thành các mảnh bản đồ có tỉ lệ lớn hơn.
- Khi dữ liệu đã được đưa lên bản đồ, nó không dễ dàng xoá được hoặc bổ sung tổng hợp thêm dữ liệu khác.
- Bản đồ giấy là các dữ liệu tĩnh. Nó rất khó khăn áp dụng phép phân tích không gian một cách định lượng.

Về cơ bản, bản đồ vẽ tay là bản đồ thể hiện trạng thái nhận thức trong một thời điểm. Con người đã biết vẽ và thể hiện màu sắc trên bản đồ giấy từ rất lâu tuy nhiên thể hiện các kết quả phân tích trên bản đồ giấy là một công việc rất mệt mỏi. Bản đồ số với những ưu điểm về tính năng lưu trữ và phân tích số liệu với một mức chi phí phù hợp đã tạo ra các khả năng mới cho người sử dụng. Chỉ cần kích chuột máy tính và với bộ mã cho các đối tượng địa lý máy tính có thể phân tích, vẽ và thể hiện các số liệu bản đồ thành các lớp chuyên đề khác nhau. Tất nhiên để tạo được các bản đồ số đòi hỏi người vẽ phải được đạt các yêu cầu nhất định về chuyên môn tin học. Các số liệu bản đồ phải hoàn toàn chính xác. Kết quả thể hiện trên bản đồ phải có tiêu chí và nội dung rõ ràng. Vẽ bản đồ số là khâu được thực hiện sau khi đã thu thập và phân tích số liệu GIS. Với sự trợ giúp của Internet và các công cụ đi kèm ta có thể tạo và xuất bản bản đồ trên mạng trực tuyến một cách dễ dàng hơn cho các tình huống khác nhau. Một website của một thành phố có thể sử dụng bản đồ để hướng dẫn người dân và khách du lịch tra cứu địa điểm của các nhà hàng, khách sạn, bảo tàng và các dịch vụ cần thiết. Các nhà kinh doanh có thể dùng bản đồ để đặt cửa hàng và văn phòng giao dịch. Các chính quyền địa phương có thể sử dụng bản đồ để trình bày các số liệu trung bình về thu nhập kinh tế của các vùng họ quản lý. Phương pháp lập bản đồ trên mạng (web mapping) sẽ giúp họ tạo ra các bản đồ thậm chí cả các bản đồ trung gian với các công cụ thông thường như Map Server, Open EV và Post GIS (xem chi tiết ở Phụ lục A và B). Các công cụ này sẽ giới thiệu cách tìm kiếm, thu thập, hiểu, sử dụng và chia sẻ các dữ liệu bản đồ trên mạng. Lập bản đồ số là một lĩnh vực đang phát triển song song với việc thu thập và phân tích dữ liệu GIS. Bản đồ số có thể phối hợp các dữ liệu địa lý, dữ liệu GPS và các công cụ quản lý dữ liệu vào một nguồn thông tin bản đồ để người sử dụng không mất nhiều thời gian và công sức để tìm kiếm thông tin cần thiết.



Hình 2.2 Sơ đồ hoạt động xử lý và tạo lập bản đồ số trên mạng

Hiện nay người ta có nhiều nguồn dữ liệu hiện đại hơn như ảnh hàng không, ảnh viễn thám. Những dữ liệu này cho phép nhìn thấy các sự thay đổi dưới mặt đất theo thời gian như sự xói mòn đất, sự thay đổi của thảm thực vật, rừng, v.v. Thế nhưng ảnh viễn thám hay ảnh hàng không không phải là bản đồ. Chúng ở khuôn dạng số và lưu trữ trong các đĩa từ. Các dữ liệu số thể hiện đối tượng thực tế không ở dạng điểm, đường hay vùng mà chúng ở dưới dạng ma trận điểm, mỗi điểm ảnh được gọi là pixel (picture element). Đây là một nguồn tài liệu hết sức phong phú, khách quan và kịp thời cho nhiều mục đích khác nhau. Những người lập bản đồ địa hình, địa lý, địa mạo đều có thể sử dụng dữ liệu này để tách ra các thông tin cần thiết cho mình. Quá trình phân tích thông tin trên ảnh được tiến hành theo nhiều phương pháp rất khác nhau và được trợ giúp bởi các công cụ toán học, vật lý học và tin học.

2.2 Các đặc tính của bản đồ

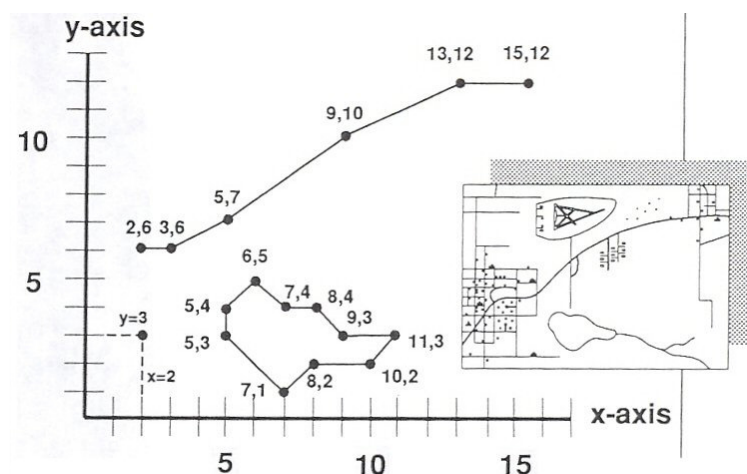
Các đặc tính của bản đồ bao gồm các đặc tính biểu thị và đặc tính đặc thù. Đặc tính biểu thị là các điểm, đường, vùng, ký hiệu và chú thích. Đặc tính đặc thù bao gồm tỷ lệ bản đồ, độ phân giải, độ chính xác và phạm vi bao phủ của bản đồ.

2.2.1 Đặc tính biểu thị

Đặc tính điểm là sự thể hiện sự định vị độc nhất đặc trưng cho một vật thể trên bản đồ có đường biên hay hình dáng quá nhỏ không thể thể hiện thành một đường hay vùng mà phải dùng một ký hiệu (symbol) đặc biệt hay nhãn (label) để thể hiện vị trí điểm. Trên hình 2.3 giá trị tọa độ $x = 2$, $y = 3$ thể hiện đặc tính địa lý điểm

Đặc tính đường là một chuỗi tọa độ liên tiếp nối liền nhau để thể hiện cho các vật thể có hình dạng tuyến tính nhưng quá hẹp để có thể thể hiện thành vùng hoặc để thể hiện một đặc tính địa lý không có chiều rộng như các đường đồng mức.

Đặc tính vùng thể hiện sự khép kín đường biên của một vùng tương đối đồng nhất trên bản đồ (Hình 2.3).



Hình 2.3 Khái niệm các đặc tính địa lý trên bản đồ

Các đặc tính khác có tính diễn tả là ký hiệu và ghi chú. Bản đồ giấy biểu thị các đặc tính diễn tả bằng các ký hiệu đồ họa. Ví dụ các đường giao thông các cấp khác nhau được biểu thị bằng các ký hiệu đường khác nhau hoặc/và với các màu sắc khác nhau; các hồ nước được tô đều bằng màu xanh lục; các khu rừng có màu xanh lá cây; v.v.

2.2.2 Đặc tính đặc thù

Tỉ lệ (Map scale)

Tỷ lệ là sự phóng to hay thu nhỏ theo yêu cầu một vùng của trái đất trên một bản đồ để thể hiện tỉ lệ khoảng cách trên một trang bản đồ hay một ảnh đến chiều dài thực tế trên bề mặt trái đất. Giá trị của tỉ lệ được thể hiện bằng các tham số đại lượng cho biết phép đo trên bản đồ và thực địa có cùng một đơn vị đo. Ví dụ:

1 cm trên bản đồ = 10000 cm trên thực địa Bản đồ có tỉ lệ 1: 10000

1 inch trên bản đồ = 2000 feet trên thực địa

1 inch trên bản đồ = 24000 inches trên thực địa Bản đồ có tỉ lệ 1: 24000.

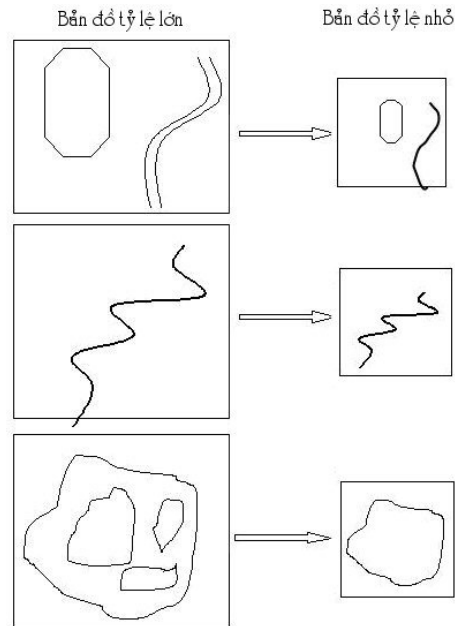
Cần chú ý là tỉ lệ luôn đề cập về khoảng cách theo chiều ngang tuyến tính chứ không phải là phép đo về diện tích hay độ cao

Tỉ lệ của một bản đồ phụ thuộc vào lượng thông tin và độ lớn của vùng sẽ được thể hiện trên bản đồ. Bản đồ có tỉ lệ lớn sẽ trình bày các đặc tính địa lý một cách chi tiết hơn nhưng chỉ thể hiện được vùng nhỏ hơn vì số thu nhỏ của bản đồ lớn hơn (ví dụ tỉ lệ 1:10000). Bản đồ có tỉ lệ nhỏ (1:250000) có thể trình bày được một vùng rộng lớn nhưng mức độ thể hiện chi tiết sẽ nhỏ hơn vì hệ số thu nhỏ nhỏ hơn.

Độ phân giải (Map resolution)

Là sự thể hiện sự chính xác, rõ ràng về vị trí và hình dạng của các đặc tính bản đồ đã được xác định theo một tỉ lệ bản đồ đã cho trước. Đối với một bản đồ tỉ lệ lớn độ phân giải của các đặc tính đạt gần với các đặc tính thực tế hơn vì mức độ thu nhỏ của bản đồ từ thực địa nhỏ hơn. Khi tỉ lệ bản đồ giảm dần thì độ phân giải cũng sẽ thấp đi vì khi đó các đặc tính sẽ được làm trơn, đơn giản dần thậm chí có thể sẽ mất đi trên bản đồ.

Khi tỉ lệ bản đồ giảm các đặc tính có chiều dài và hẹp như sông suối sẽ được đơn giản thành dạng đường (line) và các vùng nhỏ sẽ được thể hiện thành dạng điểm (point). Kích cỡ và các đại lượng tối thiểu đôi khi được gọi là đơn vị bản đồ nhỏ nhất. Giá trị này phải được coi như qui tắc khi biên soạn bản đồ. Ví dụ như các đặc tính sông suối phải được thể hiện thành đường (line) nếu như chiều rộng của chúng nhỏ hơn 1/8 inch và thành đặc tính điểm khi có chiều dài ngắn hơn 1/4 inch trên bản đồ.



Hình 2.4 Ví dụ so sánh giữa tỉ lệ bản đồ khác nhau

Theo TOBLER (1987) định nghĩa độ phân giải không gian cho các dữ liệu địa lý là sự thể hiện địa phận hình học bị phân chia bởi số vùng quan sát được chuẩn hoá theo các chiều không gian. Địa phận trên bản đồ hay ảnh là diện tích được phủ trùm theo các vùng quan sát. Ví dụ diện tích của nước Mỹ là 6 triệu km² và có 50 bang do đó nguyên tố phân giải trung bình của một bản đồ thể hiện một bang có thể là:

$$\begin{aligned} \text{nguyên tố phân giải trung bình} &= \sqrt{\frac{\text{diện tích}}{\text{số vùng quan sát}}} \\ &\approx \sqrt{\frac{6 \times 10^6 \text{ km}^2}{50}} \approx 346 \text{ km} \end{aligned}$$

Như vậy chúng ta có thể tính giá trị đại diện cho độ phân giải không gian cho một tập dữ liệu. Nếu ta tăng số vùng quan sát kích cỡ của nguyên tố phân giải trung bình sẽ giảm ví dụ như khi ta tăng số vùng quan sát của bản đồ nước Mỹ lên 3141 (3141 tỉnh) thì giá trị trên sẽ là:

$$= \sqrt{\frac{6 \times 10^6 \text{ km}^2}{3141}} = 43 \text{ km}$$

Như vậy khi chúng ta sẽ có nhiều thông tin thì yếu tố phân giải trung bình sẽ giảm lúc đó ta có một tập dữ liệu có độ phân giải cao hơn và ngược lại độ phân giải thấp sẽ có ít lượng quan sát hơn và yếu tố phân giải trung bình sẽ lớn hơn.

Độ chính xác của bản đồ (accuracy)

Ngoài độ phân giải bản đồ cũng còn cần có độ chính xác về đặt đúng vị trí các điểm và đường trên trang bản đồ. Các sai số có thể ảnh hưởng đến chất lượng bản đồ và các sản phẩm khác của hệ thống thông tin địa lý. Sai số sẽ được xem xét trong Chương 7 và nói chung có các nguồn gốc như sau:

- Các nguồn dữ liệu có thể quá cũ không còn giá trị sử dụng
- Tỷ lệ bản đồ có thể gây cản trở đến kiểu, chất lượng và số lượng dữ liệu
- Số lượng thông tin không đầy đủ để xác định các vật thể không gian theo yêu cầu
- Các vấn đề về thời gian, tài chính, và nhân lực không cho phép sản xuất ra một sản phẩm theo đúng các tính chất đã được yêu cầu
- Sự thay đổi tự nhiên của các phép đo gốc ban đầu như sự thiếu chính xác của các số liệu gốc do các vấn đề như số liệu thực địa không chuẩn, sai số của các thiết bị. Các số liệu thuộc tính thu từ nhiều nguồn khác nhau, v.v.
- Sai số trong quá trình xử lý: gồm các sai số kỹ thuật sinh ra trong quá trình tính toán hoặc các sai số logic làm cho ta thể hiện số liệu không đúng Tất cả các sai số trên tập trung cả trên một bản đồ riêng biệt hoặc trong một tầng dữ liệu của GIS. Khi xử lý nhiều lớp dữ liệu cùng một lúc các lớp dữ liệu riêng biệt có thể không tương thích về tỷ lệ, độ chính xác và làm ảnh hưởng việc tạo ra sản phẩm bản đồ có chất lượng

Vùng phủ bản đồ (Map extent)

Phần trải vùng phủ của một bản đồ là vùng trên bề mặt quả đất đã được trình bày trên bản đồ. Nó giới hạn phần được phủ trùm của vùng đó. Phạm vi phủ thường được thể hiện bằng một hình chữ nhật chứa toàn bộ các đặc tính của bản đồ. Kích cỡ của khung thường phụ thuộc vào tỷ lệ của bản đồ. Một bản đồ được vẽ với tỷ lệ nhỏ hơn sẽ đại diện cho một vùng lớn hơn do các kích thước của các đặc tính bản đồ đã được giảm với mức độ lớn hơn. Ví dụ một bản đồ tỷ lệ 1:50000 có độ trải lớn hơn và phủ trùm một vùng rộng hơn so với bản đồ tỷ lệ 1: 25000.

2.3 Hệ thống tọa độ (Coordinate system)

Để xây dựng được một hệ thống tọa độ đo toàn cầu, người ta đã coi bề mặt trái đất như là bề mặt của một hình ellipsoid tròn xoay. Hệ ellipsoid chung cho thế giới là WGS-84. Có rất nhiều hình ellipsoid tròn xoay có thể xấp xỉ được bề mặt trái đất. Ví dụ ellipsoid Clark (1880). Hiện có khoảng 15 ellipsoid trên thế giới được dùng để xấp xỉ bề mặt trái đất. Ở nước ta trước đây thường dùng ellipsoid Krasovsky cho các bản đồ trên phạm vi miền bắc và Clarke cho các bản đồ địa hình làm cho miền Nam.

Hệ trục tọa độ được gọi là lưới bản đồ vì chúng được biểu diễn bởi tập các đường thẳng ngang và đường thẳng đứng cách đều nhau và cắt nhau.

Khi đã xác định được một ellipsoid, người ta tiến hành lập một hệ thống các mốc trắc địa trên bề mặt trái đất. Hệ thống mốc trắc địa này dùng làm cơ sở để tính toán tọa độ bất kỳ điểm nào khác nằm trên bề mặt trái đất.

Để lấy mốc tính cho chiều cao, người ta lấy mực nước biển trung bình làm mốc 0. Mọi cao độ sau này đều tính so với mực nước biển trung bình. Ví dụ ở hệ tọa độ VN2000 điểm mốc độ cao nước ta đặt tại Hải phòng và điểm gốc tọa độ quốc gia đặt tại

Viện nghiên cứu địa chính Hà nội. Mỗi quốc gia đều có một lưới tọa độ chia theo các cấp 0, I, II, III và IV. Mật độ lưới cấp 0 có khoảng cách giữa các điểm là 100 km, cấp I là 25-30 km cấp II (15 - 20km) và III (cấp cơ sở) là 2-5 km. Lưới địa chính cơ sở là cố định phù trùm cho cả nước không phụ thuộc vào phương pháp đo vẽ bản đồ địa chính. Sau lưới địa chính cơ sở là lưới địa chính 1 và lưới địa chính 2 được tính toán xây dựng dựa vào tỉ lệ và phương pháp đo vẽ bản đồ địa chính. Ví dụ đo vẽ bản đồ địa chính tỉ lệ 1/500 đòi hỏi mật độ điểm địa chính 1 và 2 phải dày hơn so với đo vẽ bản đồ địa chính tỉ lệ 1/1000.

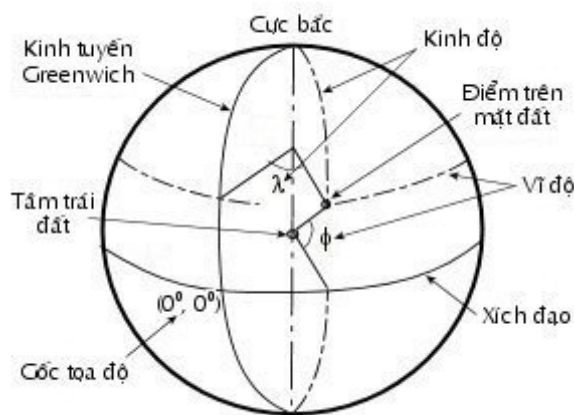
Vị trí của một đối tượng bất kỳ trên bề mặt trái đất được xác định theo 2 kiểu hệ tọa độ khác nhau: tọa độ địa lý và tọa độ vuông góc trong một lưới chiếu phẳng

Hệ tọa độ địa lý kinh độ, vĩ độ: là hệ tọa độ hình cầu bao gồm các kinh độ và vĩ độ xác nhận vị trí của các điểm trên trái đất. Chúng là các góc được đo từ trung tâm Trái đất đến một điểm trên bề mặt. Vĩ độ được đo về phía Bắc hoặc Nam còn kinh độ được đo phía Đông và Tây bán cầu. Mỗi một vùng địa lý được xác định bởi độ lệch góc Bắc hoặc Nam từ đường xích đạo (độ lệch vĩ độ) hoặc độ lệch góc kinh độ từ đường kinh tuyến gốc. Các đường phân tuyến của các kinh độ gọi là kinh tuyến kéo dài từ cực Bắc xuống Nam. Đường kinh tuyến 0 độ (đi qua vùng Greenwich của nước Anh) gọi là kinh tuyến gốc. Các phân tuyến của vĩ độ khoanh quả đất thành các vòng tròn song song gọi là các đường vĩ tuyến. Đường vĩ tuyến đi qua vĩ độ 0 gọi là Xích đạo. Kinh độ và vĩ độ được đo bằng độ, phút và giây. Vĩ độ tại điểm Bắc cực là 90° và tại điểm Nam cực là -90° . Kinh độ có giá trị dương đến 180° từ kinh độ gốc về phía Đông và có giá trị âm đến -180° từ kinh độ gốc về phía Tây.

Hệ tọa độ kinh độ và vĩ độ không thuận tiện cho việc xác định khoảng cách, diện tích của các đối tượng

Hệ tọa độ phẳng UTM (Universal Transverse Mercator)

Là hệ tọa độ phẳng quốc tế được xây dựng theo phép chiếu UTM. Do các phép chiếu riêng qui định cho mỗi vùng nên hệ UTM có độ chính xác cao (xem phần phép chiếu UTM).



Hình 2.5 Hệ tọa độ địa lý

2.4 Phép chiếu bản đồ (Map projection)

2.4.1. Ellip tròn xoay

Người ta giả thiết nếu Trái đất hình cầu của chúng ta không bị lồi lõm bởi địa hình bề mặt thì có thể giống như một hình ellip tròn xoay. Mô hình ellip tròn xoay (ellipsoid) cần thiết cho tính toán các khoảng cách lớn. Thiết bị GPS sử dụng mô hình trái đất tròn xoay để tính toán tọa độ các vị trí trên trái đất trên cơ sở các tín hiệu thu được từ ít nhất bốn vệ tinh. Một ellipsoid có bán kính xích đạo a (bán kính lớn) và bán kính cực b (bán kính nhỏ). Hình ảnh từ vệ tinh cho thấy hình dạng Trái đất có khác với ellipsoid tròn xoay tuy nhiên ảnh hưởng không đáng kể đến bản đồ tỉ lệ lớn nên ellipsoid vẫn sử dụng nhiều trong ứng dụng tính toán. Ellipsoid tròn xoay được sử dụng để xác định các mốc đo đạc. Rất nhiều các ellipsoid được nhiều quốc gia sử dụng tuy nhiên chưa có một ellipsoid nào được thống nhất cho toàn cầu (Bảng 2.1). Ở Việt nam trong thời kỳ Pháp ellipsoid Clark 1880 được sử dụng cho toàn quốc. Trước 1975 Miền Nam sử dụng ellipsoid Everest 1930 cho hệ tham chiếu và miền Bắc nước ta sử dụng ellipsoid Kravosky từ những năm 1960.

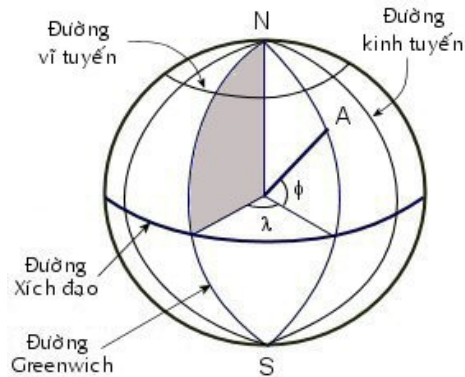
Độ dẹt các cực Trái đất được xác định bởi tỉ lệ $f = (a-b)/a$, trong đó a là bán kính đường tròn lớn, b là bán kính đường tròn nhỏ. Độ dẹt được biểu diễn là $1/f$. Số e là độ lệch tâm được tính như sau:

$$e^2 = 2f - f^2 \text{ hay } f = 1 - (1 - e^2)^{1/2}$$

Tuy độ dẹt f là rất nhỏ vẫn phải chú ý trong tính toán đối với bản đồ có tỉ lệ lớn hơn hay bằng 1/100 000. Với những bản đồ tỉ lệ nhỏ như bản đồ châu lục, bản đồ khu vực hay bản đồ thế giới thì có thể bỏ qua.

Bảng 2.1 Một số ellipsoids được sử dụng chính thức trên thế giới

Năm	Tên ellipsoid	Chiều dài trục		Độ dẹt (f)
		Trục dài (a)	Trục ngang (b)	
1980	GRS80	6378137	6356752.3	1/298.257
1972	WGS72	6378135	6356750.5	1/298.26
1940	Krasovsky	6378245	6356863.0	1/298.3
1924	International	6378388	6356911.9	1/297
1880	Clack 1880	6378249.1	6356514.9	1/293.46
1886	Clack 1866	6378206.4	6356583.8	1/294.98
1841	Bessel	6377397.2	6356079.0	1/299.15



Hình 2.6 Đường trục trái đất

Hình dạng Trái đất là một hình cầu không bằng phẳng bị biến dạng bởi nhiều yếu tố như lực hút, lực ly tâm, sự đa dạng của các thành phần trái đất các ảnh hưởng của các quá trình kiến tạo địa chất, địa lý và tác động của con người. Một hình cầu rất bất tiện cho các công việc lưu trữ thể hiện và phân tích dữ liệu và đòi hỏi chi phí cao. Hệ tọa độ đề các và hệ tọa độ địa lý chỉ có thể thể hiện trong không gian 3 chiều, bản đồ giấy chỉ thể hiện được hai chiều nên phải xây dựng hệ tọa độ khác phù hợp với hệ tọa độ trên giấy. Do đó cần có một phép ánh xạ từ bề mặt trái đất lên mặt phẳng hai chiều. Phép ánh xạ này còn gọi là phép chiếu.

2.4.2 phép chiếu

Phép chiếu bản đồ dùng để thể hiện toàn bộ hay một phần của bề mặt trái đất hình cầu trên một mặt phẳng gọi là bản đồ 2 chiều. Tuy nhiên thể hiện một bề mặt 3 chiều về mặt phẳng 2 chiều sẽ liên quan đến sự lệch đối với một số đại lượng như hình dạng, khoảng cách, diện tích và hướng. Có rất nhiều phép chiếu khác nhau được sử dụng. Sự khác biệt giữa các phép chiếu là bình thường vì không có một phép chiếu riêng biệt nào có thể đáp ứng một cách hoàn thiện (không chính xác một cách tuyệt đối) cho tất cả người sử dụng. Mỗi một vùng địa lý trên Trái đất đều có thể sử dụng một hệ thống phép chiếu riêng biệt tuy nhiên phép chiếu nào cũng cần phải đảm bảo được các tính chất sau: đúng hướng, đúng vị trí tương đối giữa các điểm, đúng hình dạng, diện tích và đúng khoảng cách. Các mặt chiếu cơ bản sử dụng là mặt phẳng, hình trụ và hình nón. Điểm hay đường tiếp xúc giữa mặt chiếu với ellíp tròn xoay gọi là tiếp điểm hay giao tuyến. Độ biến dạng tại tiếp điểm hay giao tuyến này bằng 0

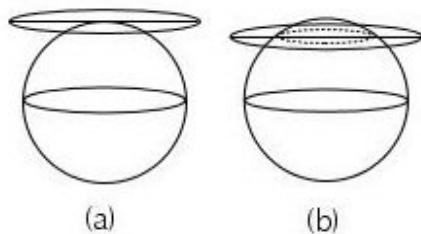
Do mặt ellipsoid và mặt phẳng không đồng phôi nên phép chiếu của mặt ellipsoid lên mặt phẳng là không bảo toàn được tất cả các tính chất hình học của mặt ellipsoid. Thường thì người ta sử dụng hai loại phép chiếu: Phép chiếu bảo toàn góc và phép chiếu bảo toàn độ dài.

Phân loại các phép chiếu căn cứ vào hình dạng, vị trí của mặt chiếu người ta phân loại như sau:

Theo hình dạng

Phép chiếu lập thể cực (Azimuthal).

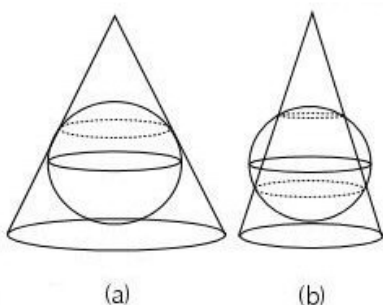
Phép chiếu azimuthal còn gọi là phép chiếu lập thể cực (H.2.7). Đây là phép chiếu đơn giản nhất. Người ta chiếu bề mặt của trái đất lên một cho mặt phẳng tiếp xúc với một điểm trên bề mặt trái đất. Các đối tượng trên bản đồ được tạo ra bằng chuyển đổi vị trí của chúng từ địa cầu sang mặt phẳng. Trong phép chiếu mặt phẳng chuẩn, kinh tuyến và các đường thẳng vĩ tuyến là các đường tròn đồng tâm hướng về cực và chúng chỉ có một điểm duy nhất không bị biến dạng là tiếp điểm.



Hình 2.7 Phép chiếu lập thể cực (azimuthal)

Phép chiếu này có thể chọn ở 2 vùng cực của hình cầu. Phép chiếu này có độ biến dạng nhỏ tại tiếp điểm, giữ nguyên được hình dáng vật thể đồng dạng với hình chiếu của nó trên mặt phẳng.

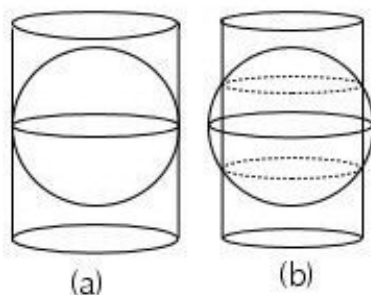
Phép chiếu hình nón (Conical)



Hình 2.8 Phép chiếu hình nón (Lambert)

Trong phép chiếu hình nón các đường kinh độ và vĩ độ được chiếu lên hình nón tiếp tuyến với địa cầu dọc theo hình tròn nhỏ nào đó (hình 2.8). Thông thường người ta chọn trục của hình nón trùng với trục quay của Trái đất. Trong trường hợp này vòng tròn trục giao giữa hình nón và Trái đất là vĩ tuyến. Vĩ tuyến được chiếu thành các cung của các đường tròn đồng tâm; kinh tuyến được chiếu thành các tia mà góc ở giữa chúng là đều nhau. Tỷ lệ trong phép chiếu này bị thay đổi nhanh về các hướng. Vì vậy nó chỉ phù hợp với các vùng vĩ độ trung bình, đặc biệt những vùng có hướng Đông - Tây. Nước Mỹ thường sử dụng phép chiếu này. Phép chiếu này còn được sử dụng trong hệ thống chiếu Lambert, hệ thống chiếu Bonnor vật thể chiếu gần kinh tuyến trung ương và nằm giữa khu vực chiếu không bị biến dạng về diện tích, càng xa kinh tuyến giữa càng bị biến dạng.

Phép chiếu hình trụ (cylindrical)



Hình 2.9 Phép chiếu hình trụ đứng (Mercator)

Là phép chiếu hình trụ thể hiện sự cắt hoặc giao của một hình trụ với hình cầu (H.2.9). Hình trụ tiếp xúc sẽ cắt bề mặt hình cầu và tạo ra một đường tròn lớn như trường hợp phép chiếu hình nón và cũng thể hiện tính tương đương về khoảng cách. Trong trường hợp tổng quát thì giao tuyến của hình trụ và trái đất là đường xích đạo. Tiếp tuyến này được gọi là đường chuẩn và

nó không bị biến dạng trên mặt phẳng chiếu. Độ biến dạng tăng nhanh về 2 cực. Các đường kinh tuyến cách đều nhau còn vĩ tuyến thì song song nhưng không cách đều nhau nữa.

Phân loại phép chiếu theo vị trí

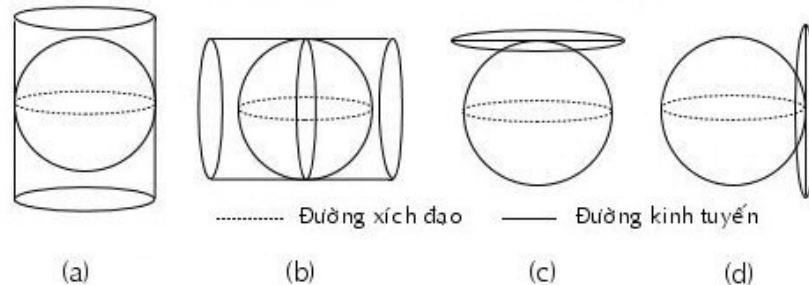
❖ *Phép chiếu chuẩn (normal)*

Giao tuyến mặt chiếu hình trụ với trái đất tại xích đạo; Trục của hình nón trùng với trục Trái đất. Giao điểm của mặt phẳng chiếu với Trái đất tại cực

Phép chiếu ngang: giao tuyến của mặt chiếu hình trụ với Trái đất tại các kinh tuyến; Trục của hình nón vuông góc với trục Trái đất; Giao điểm của mặt phẳng chiếu với Trái đất tại điểm trên đường xích đạo

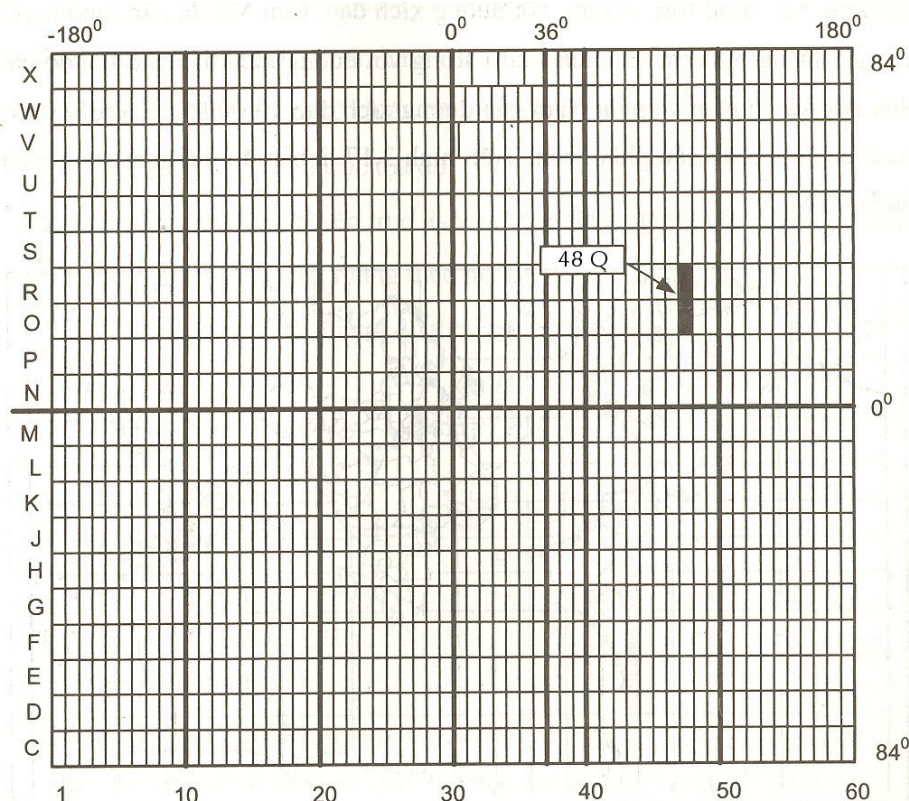
❖ *Phép chiếu xiên: (oblique)*

Giao tuyến của mặt phẳng chiếu hình trụ tại các điểm khác hai trường hợp trên, trục của hình nón tạo thành góc với trục Trái đất



Hình 2.10 Phép chiếu hình trụ và lập thể ngang (Transverse Mercator)

Hệ thống phép chiếu UTM: Phép chiếu Transverse Mercator thường được xây dựng cho bản đồ có độ chính xác cao. Phép chiếu này hầu như không làm biến dạng các vùng gần hai cực địa cầu. Nếu hình trụ quay 90 độ từ phương thẳng đứng về nằm ngang khi đó coi xích đạo theo hướng Bắc Nam. Phép chiếu hình trụ ngang Mercator thể hiện chính xác về khoảng cách và hình dạng bề mặt quả đất nên được sử dụng như một phương pháp chiếu chuẩn quốc tế (lưới chiếu UTM -Universal Tranverse Mercator).



Hình 2.11 Phép chiếu UTM

Phép chiếu toàn cầu UTM chia bề mặt trái đất thành 60 múi theo vĩ tuyến mỗi múi 6 độ (°) và đánh số từ 1 đến 60 kể từ kinh độ 180 Tây. Mỗi múi kéo dài từ vĩ độ 84 Nam đến 80 vĩ độ Bắc. Phép chiếu UTM còn chia bề mặt Trái đất thành các dải 8° từ đường xích đạo được gán với các ký tự từ C đến X, không sử dụng các ký tự I và O). Mỗi múi có kinh tuyến trục ví dụ múi 48° có kinh tuyến trục là 105° Đông. Khi chọn phép chiếu trước hết phải xác định thuộc tính quan tâm: vị trí, hình dạng hay kích thước. Có thể chọn phép chiếu phù hợp cho các vùng sau:

- chọn phép chiếu hình trụ cho các vùng nhiệt đới;
- chọn phép chiếu hình nón cho các vùng ôn đới ;
- chọn phép chiếu phẳng cho các vùng cực Bắc và cực Nam.

Trong thực tế, các vùng được chọn các phép chiếu trên có độ biến dạng nhỏ nhất.

Phần đất liền và các đảo ven biển Việt nam nằm trọn trong hai múi chiếu 48 và 49. Múi thứ nhất từ kinh tuyến 102° đến 108°, kinh tuyến trung ương là 105°, múi thứ 2 từ 108° đến 114°, kinh tuyến trung ương là 111°, sử dụng múi chiếu quốc tế là 6° . Các bản đồ tỉ lệ từ 1/5000 và lớn hơn sử dụng múi chiếu 3° với kinh tuyến trung ương 102°, 105°, 108° và 111°.

2.5 Hệ qui chiếu VN-2000

Trước khi có hệ qui chiếu VN-2000, do điều kiện lịch sử ở Việt nam đã từng sử dụng ba hệ qui chiếu khác nhau:

(a) Ellipsoid qui chiếu Clark từ thời Pháp sử dụng cho cả Đông dương, điểm gốc tại Hà nội, sử dụng lưới chiếu tọa độ phẳng Bonne (giả nón đứng, đồng diện tích);

(b) Hệ qui chiếu Ellipsoid Everest của Mỹ cho Nam Á, lưới chiếu tọa độ phẳng UTM (trụ ngang cắt Ellipsoid, đồng góc); và

(c) Hệ qui chiếu Ellipsoid Kravsovsky của Liên xô, lưới chiếu tọa độ Gausse-Kruger trụ ngang tiếp xúc ellipsoid, đồng góc) gọi là hệ qui chiếu Hà nội 72.

Việc lựa chọn một hệ qui chiếu quốc gia mới là một yêu cầu cần thiết vì cả 3 hệ qui chiếu trên đều chưa thật phù hợp với địa phận lãnh thổ nước ta.

Hệ qui chiếu VN-2000 có các thông số cụ thể như sau:

Là hệ qui chiếu WGS-84 với kích thước:

- Bán trục lớn $a = 6378137,0$ m

- Độ dẹt $f = 1: 298,257223563$

Giá trị tọa độ điểm gốc được định vị phù hợp với lãnh thổ Việt nam; Điểm gốc tọa độ quốc gia đặt tại Viện Nghiên cứu địa chính, đường Hoàng Quốc Việt, Hà nội.

Hệ tọa độ phẳng; Lưới chiếu nón đồng góc cho bản đồ 1/1.000.000; Lưới chiếu trụ ngang đồng góc UTM với múi chiếu 6° cho bản đồ địa hình cơ bản, bản đồ nền, bản đồ hành chính quốc gia ở các tỉ lệ từ 1/500.000 đến 1/25.000 có hệ số điều chỉnh biến động chiều dài $k_0 = 0,9996$; Lưới chiếu trụ ngang đồng góc UTM với múi chiếu 3° cho bản đồ cơ bản, bản đồ nền, bản đồ hành chính quốc gia từ 1/10.000 tới 1/2.000, ($k_0 = 0,9999$)

Chia múi và phân mảnh hệ thống bản đồ địa hình cơ bản theo hệ thống UTM quốc tế và danh pháp bản đồ như hệ thống hiện hành có chú thích thêm danh pháp theo hệ quốc tế

Câu hỏi ôn tập:

1. Tại sao nói bản đồ là hệ thống truyền tải các thông tin không gian?
2. Phân biệt bản đồ cơ sở và bản đồ chuyên đề, bản đồ số và bản đồ giấy?
3. Định nghĩa tỉ lệ bản đồ và đặc trưng của tỉ lệ bản đồ. Cho biết tỉ lệ của một bản đồ biết rằng 1 inch trên bản đồ tương ứng với 2.500 feet trên thực
4. Trình bày mối quan hệ giữa tỷ lệ (Scale), độ phân giải (Resolution) và độ phủ trùm (map extent) của bản đồ. Cho ví dụ.
5. Định nghĩa, phân loại các phép chiếu bản đồ cơ bản? Phép chiếu nào thông dụng cho toàn cầu hiện nay? Ở Việt nam hiện nay sử dụng phép chiếu nào?
6. Các hệ thống tọa độ thường được sử dụng trên thế giới? Tại sao hệ thống tọa độ phẳng UTM lại được sử dụng trong GIS?

Chương Ba

CẤU TRÚC DỮ LIỆU GIS

Các hệ thống thông tin có thể được tổ chức lưu trữ theo nhiều cách khác nhau. Vấn đề chọn lựa cấu trúc lưu trữ dữ liệu không gian là một trong những quyết định quan trọng của thời sơ khai thiết kế hệ thống thông tin địa lý (HTTTĐL). Hiểu biết về cấu trúc lưu trữ dữ liệu của một HTTTĐL bao giờ cũng rất hữu ích kể cả cho những người sử dụng HTTTĐL vì cấu trúc dữ liệu ảnh hưởng đến cả khối lượng bộ nhớ cần thiết và hiệu suất quá trình phân tích dữ liệu. Ngay cả khi tiến hành công tác thu thập dữ liệu cũng cần phải quyết định dạng lưu trữ dữ liệu. Trong nghiên cứu dữ liệu môi trường ta vẫn thường phải biến chuyển dữ liệu sang các dạng lưu trữ khác nhau. Do vậy hiểu biết về cấu trúc dữ liệu rất quan trọng, nhất là về ảnh hưởng của việc chuyển hóa dạng lưu trữ ảnh hưởng thế nào đến chất lượng dữ liệu. Chương này trước hết trình bày các khái niệm về lưu trữ dữ liệu thông tin dưới hình thức cơ sở dữ liệu, sau đó giới thiệu các dạng lưu trữ đặc thù cho các HTTTĐL cùng với các ưu nhược điểm của từng dạng.

3.1 Các khái niệm cơ sở

3.1.1 Lưu trữ dữ liệu địa lý trong máy tính

Đặc điểm dữ liệu trong các HTTTĐL khác biệt với dữ liệu ở các hệ thống thông tin khác (ngân hàng, thư viện, quản lý khách hàng hàng không, y tế, v.v.) ở chỗ chúng bao gồm cả thông tin về vị trí không gian, thậm chí cả các mối liên hệ topo không gian và các thông tin diễn tả (thuộc tính) các vật thể trong hệ thống dữ liệu.

Dữ liệu địa lý biểu thị các vật thể trên bề mặt trái đất bao giờ cũng được thể hiện trên mặt giấy phẳng hai chiều bằng ba thực thể chính: điểm (points), đường (lines) và vùng (polygons) được mã hóa vào trong hệ thống CSDL địa lý. Các thực thể này biểu thị tất cả các vật thể địa lý trên mặt đất để mô hình hóa thế giới thực tế.

Các điểm như là vị trí các giếng khoan dầu mỏ hay giếng khoan nước ngầm, đường như là trung tâm của các đường giao thông, của sông ngòi, là hai thực thể cơ bản của quan niệm đơn giản hóa sự biểu thị các vật thể trong thế giới thực này. Khi muốn thể hiện các vùng được bao quanh bởi đường biên nào đó (như địa giới đơn vị hành chính, hay các hồ nước) thì ta chú ý thể hiện các đường biên khép kín bao quanh vùng đó mà trong ngôn ngữ dữ liệu không gian người ta gọi là vùng (polygon).

Người ta sử dụng một hệ thống tọa độ nhất định để liên hệ dữ liệu địa lý với vị trí thực tế của vật thể mà dữ liệu đại diện trên bề mặt trái đất. Hệ thống tọa độ có thể là hệ cục bộ cho các vùng nghiên cứu nhỏ hoặc là hệ thống tọa độ quốc gia hay quốc tế như UTM (Universal Transverse Mercator).

Ngoài các dữ liệu không gian như diễn tả ở trên, trong các HTTTĐL còn tồn tại loại dữ liệu khác không kém phần quan trọng là dữ liệu thuộc tính phi không gian. Loại dữ liệu này thường chiếm một phần bộ nhớ không nhỏ trong các HTTTĐL. Ví dụ đối với một giếng khoan thì ngoài thông tin về vị trí địa lý của nó còn có một loạt các dữ liệu thuộc tính như độ sâu, ngày khoan giếng, lưu lượng bơm, chủ nhân, người vận hành, v.v. Nhiều HTTTĐL được thiết kế bao gồm những công cụ đặc biệt cho việc quản lý những dữ liệu thuộc tính. Vấn đề lưu trữ và quản lý các loại dữ liệu cũng như các quan hệ giữa chúng trong HTTTĐL bắt buộc phải thiết kế và xây dựng hệ thống cơ sở dữ

liệu nhằm đạt được mục đích sử dụng hiệu quả nhất cả về khối lượng tích trữ và tốc độ xử lý.

3.1.2 Tổ chức cơ sở dữ liệu trong máy tính

Trước khi tìm hiểu về phương cách dữ liệu được tích trữ và quản lý như thế nào trong các HTTTĐL, ta nên biết được những kiến thức cơ bản về tích trữ và quản lý dữ liệu chung hiện nay trên thế giới. Mặc dù người sử dụng HTTTĐL không bắt buộc phải thông thạo mọi chi tiết về việc máy tính lưu trữ dữ liệu nhưng kiến thức về các phương pháp cấu trúc dữ liệu sẽ giúp họ hiểu được sự vận hành của hệ thống và về các ưu nhược điểm của nó. Hiện nay trên thị trường sách có rất nhiều tài liệu về cấu trúc, thiết kế, xây dựng và ứng dụng các hệ thống quản lý CSDL (database management systems, DBMS) nên người đọc có thể dễ dàng bổ sung kiến thức sâu rộng hơn về vấn đề này.

Công nghệ tin học bắt đầu có những biến chuyển vượt bậc trong thập kỷ những năm 60 và 70 đã dẫn đến những bước phát triển mới về cơ sở dữ liệu như CSDL (lưu trữ thông tin) và hệ thống quản lý CSDL hay là DBMS (hệ thống tin học để quản lý CSDL). Hiện tại các phần mềm DBMS phức tạp đang được ứng dụng rộng rãi để quản lý các CSDL to lớn như là hệ thống thống kê dân số của cả quốc gia hay dữ liệu thống kê toàn cầu.

Một CSDL là một tập hợp thông tin về loại vật thể nào đó và các quan hệ giữa chúng. Ví dụ CSDL về tên người và địa chỉ. Tên người có thể phân loại theo tên khách hàng, bạn bè hay thành viên gia đình. Mục đích của việc thu thập dữ liệu và quản lý dữ liệu trong khuôn khổ bất kỳ hệ thống CSDL nào là tốc độ truy cập dữ liệu và cập nhập dữ liệu nhanh chóng và thuận tiện. DBMS là chương trình máy tính cung cấp các công cụ về nhập dữ liệu, lưu trữ dữ liệu, truy vấn dữ liệu và xuất dữ liệu theo yêu cầu một cách có hiệu suất cao. Trên thị trường phần mềm hiện nay có rất nhiều DBMS (Microsoft Access, MS SQL Server, Oracle, Apache, v.v.). Mỗi DBMS được thành lập chuyên dụng cho một số hệ thống máy tính, hệ điều hành hay loại CSDL nhất định và người sử dụng cần xác định loại DBMS nào là thích hợp nhất cho điều kiện và mục đích sử dụng của mình.

Tập tin và vấn đề truy cập tập tin

Các tập tin (files) trong CSDL được sắp xếp theo một trình tự nhất định nhằm tạo điều kiện lưu trữ và cập nhập thuận tiện. Ta có thể phân làm ba kiểu sắp xếp tích trữ như sau:

Danh mục tập tin đơn giản: Các tập tin được lưu trữ theo thứ tự thời điểm mà tập tin đó được nhập vào CSDL. Bất kỳ tập tin mới nào đều được đặt ở vị trí cuối cùng của danh sách. Loại hình lưu trữ như vậy có ưu điểm là nhập dữ liệu mới dễ dàng nhưng sẽ rất khó khăn khi cần truy cập một tập tin nhất định nào đó.

Danh sách sắp xếp thứ tự: Lưu trữ loại này tương tự như sắp xếp các từ vựng trong một quyển từ điển theo thứ tự vần ABC. Việc thêm tập tin mới cần phải có động tác tạo khoảng trống trước khi nhập nó vào CSDL. Tuy vậy việc truy cập tập tin sẽ nhanh hơn vì ta không phải tìm kiếm từ đoạn đầu của CSDL.

Tập tin có gắn chỉ số (index): Phương pháp này khắc phục được các nhược điểm của cả hai hình thái tích trữ danh sách đơn giản và danh sách thứ tự giảm dần do việc mỗi một tập tin được gắn thêm một chỉ số thuộc tính. Tuy vậy cũng vẫn tồn tại nhược điểm: Không thích hợp cho CSDL với sự cập nhập liên tục cả dữ liệu mới và xóa dữ liệu cũ vì trong trường hợp đó thì phải cập nhập thêm cả các chỉ số.

Cấu trúc cơ sở dữ liệu

Một CSDL bao gồm một tập hợp các tệp tin. Người ta phân biệt các CSDL thành ba loại chính: cấu trúc dạng cây (hierarchical), dạng lưới (network) và dạng quan hệ (relational database).

Cấu trúc dạng cây (hierarchical) Cấu trúc này dựa trên quan hệ giữa các tệp tin, chủ yếu là loại quan hệ cha-con hay một-đến-nhiều, tương tự như quan hệ gia đình. Loại quan hệ này rất quen thuộc đối với các nhà sinh vật học khi họ phân loại động thực vật. Cấu trúc dạng cây này giả thiết có thể truy cập bất cứ thành phần nào của CSDL với một chìa khóa diễn tả toàn bộ cấu trúc CSDL. Giả thiết nữa là tồn tại một sự tương ứng giữa chìa khóa và các thuộc tính của thành phần đó.

Các ưu nhược điểm của cấu trúc hình cây này là (a) cấu trúc rõ ràng, dễ hiểu, dễ dàng cập nhập và mở rộng; (b) truy cập nhanh chóng nếu thông qua các chìa khóa nhưng sẽ khó khăn khi sử dụng các thuộc tính; và (c) cấu trúc CSDL cứng nhắc và không thể thay đổi; (d) cần có một tệp tin phụ để lưu trữ các chỉ số và có thể có một lượng lớn số lần nhắc lại cùng một thuộc tính dẫn đến sự tăng không cần thiết bộ nhớ lưu trữ.

Cấu trúc dạng lưới (network) Loại cấu trúc này cho phép kết nối giữa các tệp tin trong CSDL, đặc biệt đối với dữ liệu đồ họa mà trong đó việc thể hiện kết nối giữa các vật thể nằm kề nhau có tầm quan trọng thiết yếu. Ví dụ nếu ta sử dụng cấu trúc dạng cây cho dữ liệu đồ họa thì sẽ rất khó khăn thể hiện liên kết giữa các vật thể có vị trí kề cạnh nhau vì các đường kết nối chỉ hạn chế dịch chuyển lên xuống theo trình tự cha-con.

Cấu trúc dạng quan hệ (relational database) Trong CSDL dạng quan hệ, các dữ liệu được lưu trữ thành từng mẫu tin (records hay tuples, bao gồm một danh sách được sắp xếp tuần tự trong một bảng hai chiều (còn có tên là quan hệ). Mỗi bảng dữ liệu thường được lưu trữ như là một tệp tin riêng biệt trong đó các quan hệ được thiết kế sao cho đạt được sự tìm kiếm dữ liệu hiệu quả nhất.

CSDL loại quan hệ có các ưu điểm: (a) rất linh hoạt và có thể thỏa mãn mọi yêu cầu về tìm kiếm dữ liệu với công thức logic cùng như công thức đại số; và (b) nhiều dạng dữ liệu có thể dễ dàng được truy vấn, kết hợp hay so sánh cũng như được cập nhập thêm hay xóa đi.

Tuy vậy nó cũng có một số nhược điểm (a) Việc truy vấn xảy ra theo trình tự các tệp tin (bảng) trên cơ sở các quan hệ nên có thể cần thời gian đáng kể để máy tính thực hiện các động tác này; (b) Công việc thiết kế CSDL đòi hỏi kỹ năng chuyên môn cao để có thể đạt được cấu trúc hợp lý cho khả năng truy vấn sau này và như vậy có thể giá thành thiết kế và xây dựng sẽ tăng cao hơn so với việc sử dụng các loại cấu trúc khác.

Cấu trúc dữ liệu phổ biến hiện nay trên thế giới đều theo dạng CSDL quan hệ, kể các CSDL cho các HTTTĐL. Sau đây ta sẽ xét lần lượt cấu trúc hai loại dữ liệu trong các HTTTĐL là dữ liệu không gian và dữ liệu thuộc tính, bao gồm cả cách thức lưu trữ và kết nối giữa chúng.

3.2 Dữ liệu không gian

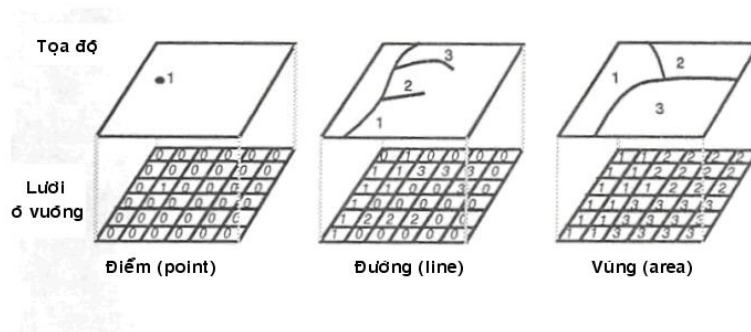
Đây là dạng dữ liệu cơ bản của một hệ thống thông tin địa lý. Dạng dữ liệu này bao gồm các thông tin có tính đồ họa chỉ rõ hình dạng, phạm vi không gian, vị trí địa lý của một thực thể trong thế giới thực được khái quát hóa thành các đặc tính địa lý như điểm, đường hay vùng trên bản đồ (ảnh) dưới dạng raster hoặc vector.

3.2.1 Cấu trúc dữ liệu Raster

a. Định nghĩa

Cấu trúc raster là một trong những cấu trúc dữ liệu đơn giản nhất trong GIS. Nó còn được có tên là “tổ chức theo ô vuông của dữ liệu không gian” (cellular organization of spatial data). Với cấu trúc dữ liệu dạng raster, giá trị của thông số nghiên cứu như độ cao địa hình (m), loại đất, sinh khối thực vật (gr/m^2), v.v. phải được xác định cho từng ô lưới (cell) và ta sẽ có một ma trận không gian. Trong cấu trúc dữ liệu dạng raster (Hình 3.1), các vật thể trên bản đồ có thể được mô tả bằng các yếu tố (map feature) sau:

- Điểm (point): là một ô vuông (single grid cell).
- Đường (line): là tập hợp các ô lưới vuông nối tiếp nhau và sắp xếp theo một hướng nhất định.
- Vùng (area): là một tập hợp khép kín các ô vuông lưới có vị trí liên kề nhau.



Hình 3.1 Cấu trúc dữ liệu RASTER

Hình 3.1 thể hiện trực quan cấu trúc dữ liệu raster của các thực thể điểm, đường và vùng.

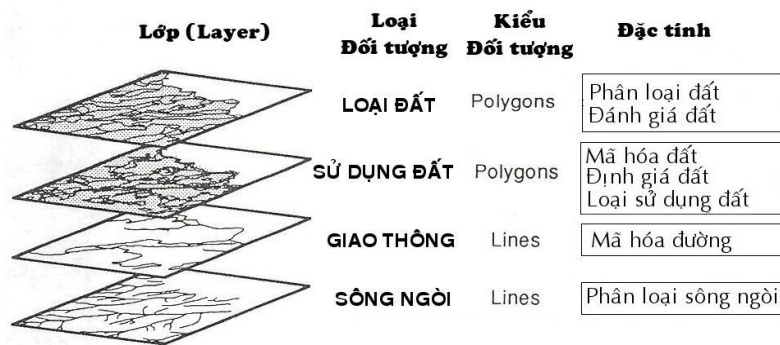
Cấu trúc dữ liệu raster đơn giản nhất là một ma trận ô lưới (ma trận hàng và cột). Mỗi một ô vuông (gridcell) được đặt theo một hàng và cột và mang một số đại diện cho kiểu hay giá trị của thuộc tính của đặc tính địa lý (code có trên bản đồ). Cấu trúc dữ liệu trên cũng cho thấy trên mặt phẳng 2 chiều các dữ liệu địa lý không được thể hiện một cách liên tục mà từng phần nên ảnh hưởng đáng kể đến tính toán độ dài hay diện tích cho các đặc tính bản đồ khi kích thước của các gridcell đại diện cho các đặc tính càng lớn.

Tầng (lớp) dữ liệu: vì mỗi cell trong một bảng 2 chiều chỉ mang một số nên các thuộc tính địa lý khác nhau phải được thể hiện theo các tập hợp ma trận riêng biệt gọi là lớp chồng xếp. Thế giới thực là sự thể hiện tổng hợp của một tập hợp chồng xếp các lớp dữ liệu mang các đặc tính khác nhau (địa hình, loại đất, đường, sông suối, nhà cửa, dân cư, v.v.) hình vẽ. Phép chồng xếp rất quan trọng trong công tác xử lý ảnh số, ảnh viễn thám và là cơ sở của hầu hết các quá trình xử lý ảnh raster

b. Tổ chức dữ liệu raster trong GIS

Mục đích tổ chức dữ liệu raster trong GIS là để tối ưu hoá việc nhập dữ liệu, giảm bớt sự lưu trữ và các yêu cầu về xử lý dữ liệu. Các dữ liệu raster được tổ chức như sau (Hình 3.3):

Khi đã xác định được các đối tượng và thuộc tính của chúng, chúng ta có thể tổ chức các đặc tính địa lý thành các tầng (lớp) dữ liệu như trong Hình 3.2. Có nhiều yếu tố ảnh hưởng đến tổ chức lớp trong cơ sở dữ liệu địa lý và chúng khác nhau trong mỗi kiểu ứng dụng. Cách thông thường nhất là tổ chức lớp theo kiểu của các đặc tính (điểm, đường, vùng) và nhóm theo các chuyên đề. Các lớp được tổ chức sao cho các điểm, đường và vùng được lưu trữ trong các lớp riêng biệt. Ví dụ như các nhà ở mang đặc tính điểm có thể được lưu riêng thành một lớp trong khi đường giao thông đại diện cho đặc tính đường thì được lưu trong một lớp khác. Các đặc tính có thể tổ chức theo chuyên đề mà nó đại diện. Ví dụ dòng chảy và đường đi lại tuy cùng mang đặc tính đường nhưng có thể tổ chức thành 2 lớp khác nhau. Các thuộc tính đi kèm đặc tính địa lý thủy hệ (sông suối) có thể là tên gọi, phân loại (sông hoặc kênh rạch) và vận tốc dòng chảy trong khi thuộc tính của đường giao thông là tên gọi, kết cấu bề mặt và độ rộng v.v. Do sự khác nhau rất xa về thuộc tính nên chúng phải được lưu trữ vào các lớp khác nhau nhưng phải cùng chung một vùng quan trắc địa lý (cùng một hệ thống tham chiếu không gian).



Hình 3.2 Khái niệm về các lớp dữ liệu và tổ chức dữ liệu raster

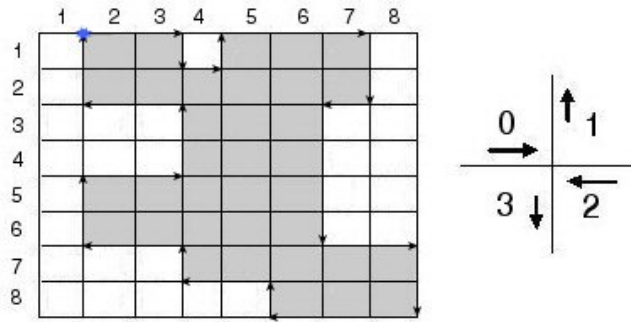
c. Các phương pháp nén lưu trữ dữ liệu raster

Khi mỗi một ô vuông được gán một giá trị thuộc tính thì ta sẽ phải lưu trữ (n hàng $\times$ m cột $\times$ 3 giá trị tọa độ x, y và giá trị thuộc tính $= n \times m \times 3$) số liệu cho mỗi lớp dữ liệu raster. Đây là trường hợp khi ta muốn số hóa bản đồ địa hình thành một ma trận độ cao. Nếu trong nhiều tập hợp thành phần gồm nhiều ô vuông có chứa cùng một giá trị thuộc tính thì ta có thể áp dụng các thuật toán lưu trữ nén để giảm đáng kể bộ nhớ của máy tính. Các phương pháp nén này như ta sẽ thấy ở phần tiếp theo, chỉ có thể được sử dụng khi ta không có quan hệ loại một-đa phương của thông số thuộc tính với các tọa độ ngang của các ô vuông. Có bốn phương pháp nén cơ bản dữ liệu dạng raster: chain codes, run-length codes, block-codes và quadrees.

Phương pháp nén theo đường biên của vùng (chain codes)

Các đường biên của một vùng (vùng gồm các ô tô đậm trong Hình 3.3) có thể được thể hiện bằng một dãy các vec tơ đơn vị liên tiếp nhau theo hướng 4 phương. Các hướng này được qui ước bằng các số (hướng Đông = 0, Bắc = 1, Tây = 2, Nam = 3). Ví dụ ta muốn lưu trữ dữ liệu vùng được tô đậm trong Hình 3.3 bằng cách biểu thị đường biên ngoài của vùng với mã chain-code, ta bắt đầu từ đỉnh phía trên, bên phải của cell hàng 1, cột 1 và đặt mã lần lượt theo chiều kim đồng hồ ta sẽ có dãy giá trị sau thể hiện biên của vùng tô đậm:

$0^2, 3, 0, 1, 0^3, 3^2, 2, 3^4, 0^2, 3^2, 2^3, 1, 2^2, 1, 2^2, 1^2, 0^2, 3^2, 2^2, 1^2$



Hình 3.3 Nén chain-code

Chain codes là một phương thức lưu trữ nén dữ liệu raster hiệu quả và nó cho phép tiến hành dễ dàng các phép tính chu vi và diện tích, nhận biết vùng lõm lõm hoặc vị trí thay đổi hướng đột ngột. Tuy vậy nó gây khó khăn cho phép phân tích chồng xếp như kết hợp và ghép bản đồ nếu không biến đổi dữ liệu trở về dạng không nén. Nhược điểm nữa là vấn đề dư thừa dữ liệu vì tất cả các đường biên phải lưu trữ hai lần.

Phương pháp nén theo hàng cột (run-length codes)

Phương pháp này cho phép các điểm trên mỗi đơn vị bản đồ được lưu trữ theo hàng từ trái qua phải bắt đầu bằng ô (cell) đầu tiên đến ô cuối của mỗi vùng con. Trong Hình 3.3 ví dụ trên ta có thể viết như ở Bảng 3.1

Bảng 3.1 Nén theo hàng cột

Hàng	Cột
hàng 1:	2,3 5,7
hàng 2:	2,7
hàng 3:	4,6
hàng 4:	4,6
hàng 5:	2,6
hàng 6:	2,6
hàng 7:	4,8
hàng 8:	6,8

Ta thấy vùng nghiên cứu với 64 ô vuông (cell) đã được lưu trữ chỉ bằng 18 chữ số (9 cặp số) và như vậy đã giảm xuống đáng kể bộ nhớ cần thiết để lưu trữ dữ liệu. Phương pháp này phù hợp cho việc lưu trữ dữ liệu trong các máy tính có dung lượng bộ nhớ nhỏ. Tuy nhiên việc nén quá nhiều dữ liệu có thể sẽ gây trở ngại cho quá trình xử lý và phân tích dữ liệu.

Phương pháp nén theo khối (Block codes)

Quan niệm về lưu trữ dữ liệu kiểu run-length codes có thể được tổng quát hóa ra mặt hai chiều khi sử dụng các khối ô vuông với kích thước khác nhau để trải lên toàn bộ vùng bản đồ cần thể hiện. Hình 3.4 trình bày cách ứng dụng phương pháp nén block codes cho dữ liệu vùng tô đậm ở Hình 3.3. Cấu trúc dữ liệu sau khi nén bao gồm chỉ ba con số: góc của khối ô vuông (có thể là tâm điểm hoặc góc trái phía dưới) và bán kính của khối ô vuông. Trong ví dụ ở Hình 3.4 thì vùng nghiên cứu được mã hóa dạng nén block code $2^n \times 2^n$ bằng 7 khối 1 đơn vị ô vuông và 7 khối 4 đơn vị ô vuông. Do mỗi góc khối ô vuông cần 2 số liệu để lưu trữ nên cả vùng cần $3 \times 14 = 42$ con số.

Rõ ràng là nếu bản đồ bao gồm càng lập được nhiều khối ô có kích thước càng lớn thì phương pháp nén block codes càng hiệu quả. Phương pháp này còn có ưu thế khi thực hiện kết hợp và giao cắt các bản đồ và khi thực hiện công việc kiểm tra sự co giãn về hình dạng của bản đồ.

	1	2	3	4	5	6	7	8
1								
2								
3								
4								
5								
6								
7								
8								

Hình 3.4 Nén block-code

Phương pháp nén cây tứ phân (quadtree block)

Phương pháp nén thứ tư này dựa trên cơ sở chia liên tục của dạng ma trận $2^n \times 2^n$ thành các thành phần dạng cây tứ phân. Cả vùng bản đồ được chia thành bốn phần liên tục trong khi thỏa mãn điều kiện giữ các ô vuông con phải nằm trọn trong vùng nghiên cứu. Giới hạn thấp nhất của phép chia là một pixel. Hình 3.5 cho ví dụ sự chia liên tục của một vùng thành những khối 1/4. Cấu trúc block này có thể trình bày dưới dạng cây tứ phân gọi là quadtree. Toàn bộ mảng gồm $2^n \times 2^n$ điểm là nút gốc của cây tứ phân và chiều cao lớn nhất của cây là n tầng. Mỗi nút có bốn nhánh cây là NW, NE, SW và SE. Nút lá tương ứng với quadtree mà nó không cần phải tiếp tục chia tiếp để bao trùm vùng nghiên cứu. Như vậy mỗi nút sẽ lưu trữ trong máy tính với độ lớn 2 bit và để lưu trữ vùng bản đồ như trong Hình 3.5 ta cần 2 bit / quadtree $\times (4+16+24+28)$ quadtree = 144 bit.

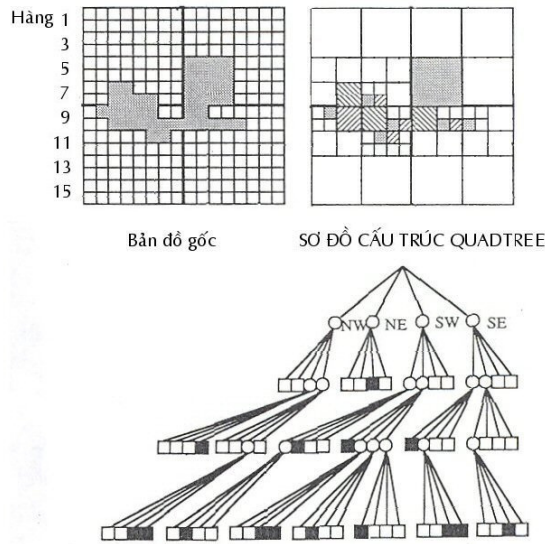
Phương pháp nén hình cây tứ phân có các ưu điểm so với các phương pháp lưu trữ nén khác: (a) dễ tính toán diện tích chu vi của các vùng có hình dạng chuẩn; và (b) có thể giảm bớt sự lưu trữ với các độ phân giải khác nhau do đặc điểm nó có thể thay đổi độ phân giải và giảm lưu trữ ở các khu vực có giá trị thông số thuộc tính đồng đều.

Tuy vậy những vấn đề nảy sinh đối với phương pháp này là: (a) các vùng có cùng kích thước và hình dạng có thể được biểu thị bằng những quadtree khác nhau nên sẽ gây khó khăn cho việc phân tích nhận hình dạng; và (b) rất khó khăn nếu ta muốn chia nhỏ một vùng hay tạo lỗ trống không một vùng của bản đồ.

Tóm lại khi bản đồ bao gồm $n \times n$ ô vuông (cell) và mỗi ô chứa một giá trị thông số thuộc tính khác nhau thì ý định giảm lượng bộ nhớ lưu trữ sẽ rất khó mà thực hiện được. Hạn chế của cấu trúc dữ liệu loại raster chủ yếu do nó tạo nên khối lượng lưu trữ lớn. Tuy vậy khi tồn tại những vùng đồng nhất về giá trị thông số thuộc tính thì ta có thể sử dụng một trong bốn phương pháp lưu trữ nén đã trình bày ở trên.

Phương pháp nén run-length code có thể là hiệu quả nhất trong việc hiển thị và lưu trữ các vùng khi các kích cỡ các pixel lớn so với kích thước của bản đồ. Khi độ phân giải tăng cao và số lượng các pixel trong một vùng tăng thì có thể phương pháp nén block code hay quadtree sẽ trở nên phương án cần áp dụng. Block code còn có thêm ưu

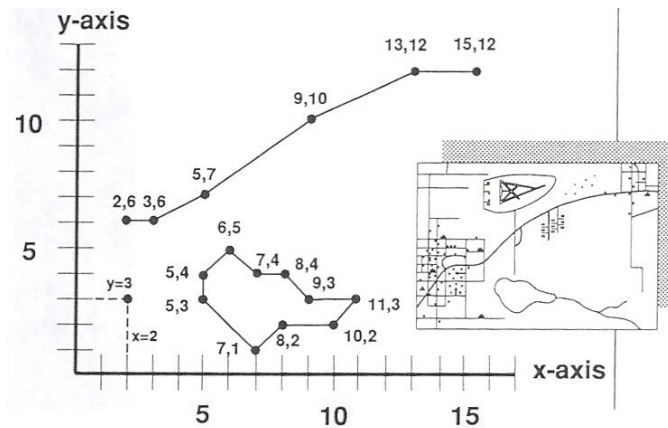
điểm ở chỗ nó có khả năng biến đổi độ phân giải tùy theo mức độ chi tiết của các vùng thành phần.



Hình 3.5 Nén cây tứ phân (quadtree)

3.2.2 Cấu trúc dữ liệu dạng Vector

Theo quan niệm toán học thì vector bao gồm một điểm (với cặp tọa độ x, y trong không gian phẳng hay x, y, z trong không gian ba chiều), một khoảng cách và một chỉ số hướng. Đối với dữ liệu GIS dạng vector, ta coi như thực thể có thể xuất phát từ bất cứ vị trí nào chứ không bị giới hạn bởi độ phân giải (kích thước của ô vuông nhỏ nhất) như trong quan niệm tổ chức dữ liệu dạng raster.



Hình 3.6 Các dữ liệu vector dạng điểm đường vùng trong tọa độ phẳng x, y

Ví dụ để lưu trữ dữ liệu biểu thị một vật thể hình tròn thì trong cấu trúc raster ta tạo mã code cho tập hợp những ô vuông sơ cấp nằm trong phạm vi hình tròn trong khi đối với cấu trúc dạng vector thì ta chỉ cần lưu trữ tọa độ điểm tâm vòng tròn và giá trị bán kính của hình tròn.

a. Định nghĩa

Cấu trúc dữ liệu vector (Hình 3.6) là một cố gắng nhằm thể hiện chính xác các đối tượng trong thế giới thực lên bản đồ số bằng giá trị liên tục của các cặp tọa độ và xác định chính xác mối quan hệ không gian của các đối tượng. Do khoảng không gian được coi là liên tục chứ không lấy giá trị trung bình nên tất cả các vị trí không gian, các khoảng cách và các kích thước đều có thể được hiển thị chính xác tỷ lệ theo giá trị thực của chúng.

Tuy vậy, thực sự thì không thể tồn tại sự thể hiện chính xác tuyệt đối mà phụ thuộc vào mức độ thể hiện của máy tính. Thực ra thì bất cứ thiết bị vector nào cũng bị giới hạn bởi một khoảng cách cơ bản mặc dù khoảng cách này có giá trị vô cùng nhỏ so với độ phân giải của hầu hết thiết bị raster. Bên cạnh giả thiết về sự chính xác thể hiện tọa độ thì cấu trúc vector sử dụng các công thức quan hệ chính xác nên có thể lưu trữ dữ liệu phức tạp hiệu quả về mặt bộ nhớ.

b. Các thực thể vector

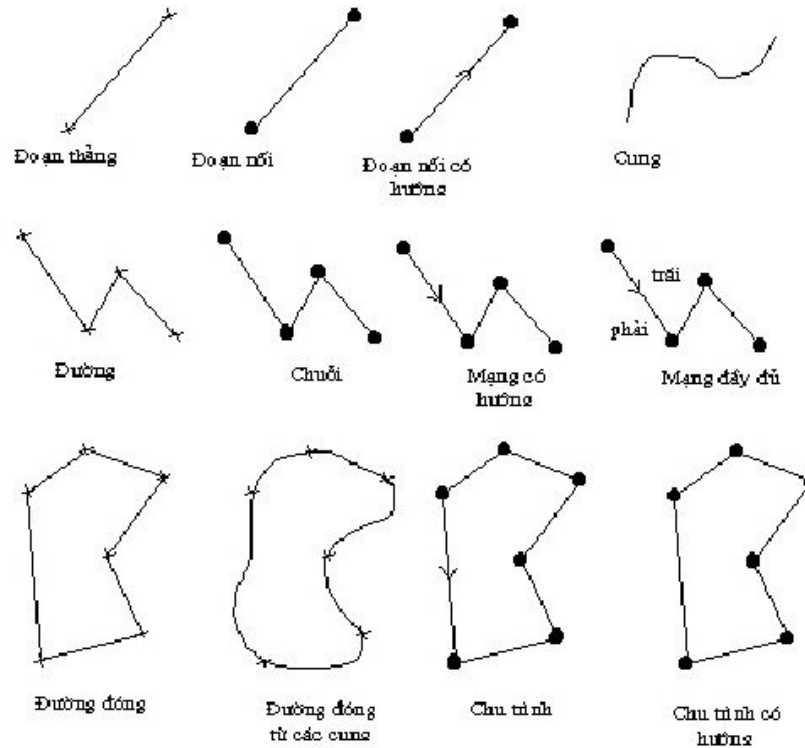
Cấu trúc vector coi vật thể tự nhiên là tập hợp các thực thể không gian cơ sở (điểm, đường và vùng) và tổ hợp giữa các thực thể này. Các thực thể sơ cấp này được thành lập trên cơ sở các cặp tọa độ của các điểm trong một hệ tọa độ nhất định.

Thực thể điểm (point entity) là các thực thể địa lý được xác định bởi một cặp tọa độ (x, y) duy nhất. Ngoài ra thì các dữ liệu mô tả điểm đó như ký hiệu, tên gọi, v.v. cũng được lưu trữ cùng với cặp tọa độ.

Thực thể đường (line entity, Hình 3.7) được định nghĩa như là tập hợp các thực thể địa lý được xác định bằng những đoạn thẳng có ít nhất hai hay nhiều cặp tọa độ. Một đường đơn giản nhất được tạo thành khi lưu trữ một điểm đầu và một điểm cuối (hai cặp tọa độ X, Y) và có thể kèm theo dữ liệu ký hiệu thể hiện nó trên bản đồ. Một cung (arc), một kênh (chain) hay một chuỗi (string) là tập hợp của n cặp tọa độ biểu thị một đường cong phức tạp và liên tục. Người ta có thể giảm đáng kể bộ nhớ lưu trữ nếu có thể biểu diễn các đường cong bằng phương trình toán học vì khi đó chỉ cần lưu trữ các thông số của phương trình thay vì lưu trữ giá trị của tất cả các cặp tọa độ.

Khái niệm mạng lưới đường (line network): Tổ hợp các đoạn thẳng đơn giản không chứa các thông tin liên quan đến mạng lưới và sẽ gây khó khăn trong các phân tích đường tiêu nước, hệ thống giao thông, v.v. Để có thể tạo được một cấu trúc mạng, trong đó máy tính dễ dàng xác định thứ tự liên quan giữa các đường, người ta phải tạo các quan hệ ràng buộc, chỉ ra sự liên kết giữa các đoạn thẳng. Line Network là sự tổ hợp của các đoạn thẳng gắn với sự thể hiện liên tục quan hệ giữa chúng qua các điểm nút (node). Các node cho biết hướng và xác định góc mà mỗi đoạn cong gắn vào nút, tạo thành mối quan hệ không gian cho toàn bộ mạng lưới (topology). Hình 3.7 là minh họa cho các loại đường khác nhau được thể hiện dạng vector trong GIS.

Thực thể vùng (polygon entity, region or area): Vùng là một đối tượng hình học 2 chiều. Vùng có thể là một đa giác đơn giản hay hợp của nhiều đa giác đơn giản. Do một vùng được cấu tạo từ các đa giác nên cấu trúc dữ liệu của đa giác phải ghi lại được sự hiển diện của các thành phần này và các phần tử cấu tạo nên đa giác.



Hình 3.7 Các cấu trúc vector dạng đường

c. Phân loại cấu trúc dạng Vector

Mục đích của cấu trúc dữ liệu vector dạng vùng là nhằm biểu thị được các tính chất topo không gian của các thực thể vùng (như hình dạng, quan hệ kề cận, và phân cấp) để sau đó các thực thể sơ cấp cấu thành này có thể thể hiện được các vùng lên màn hình và dễ dàng được xử lý và phân tích như các bản đồ chuyên đề. Trước khi diễn tả và phân tích các loại cấu trúc dữ liệu vector ta cần liệt kê các yêu cầu của hệ thống GIS đối với mạng lưới (network polygon):

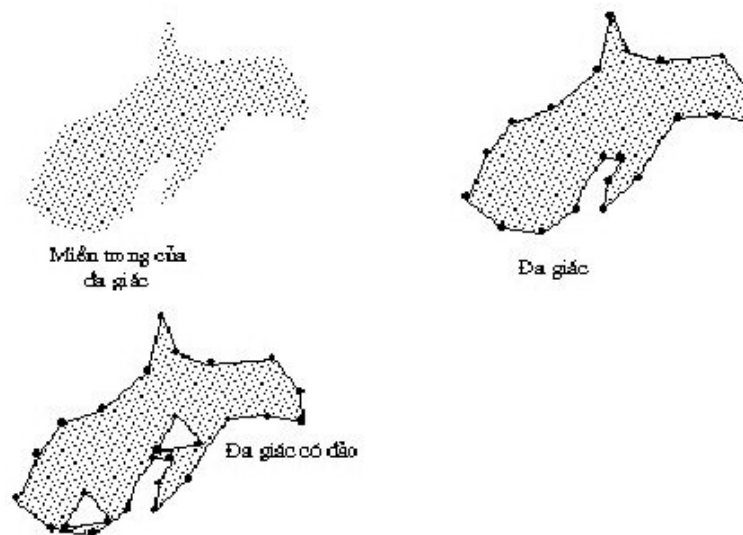
- Mỗi một polygon sơ cấp trên bản đồ sẽ có một giá trị diện tích, chu vi và hình dạng duy nhất. Trong cấu trúc vector ta không có thực thể sơ cấp cấu thành như kiểu ô vuông cell của dữ liệu cấu trúc dạng raster;
- Các dữ liệu polygon phải kèm theo dữ liệu về các thực thể giáp ranh tương tự như mạng lưới đường (line network);
- Các polygon khác nhau trong một bản đồ chuyên dụng không nhất thiết phải được thể hiện đồng mức. Các vùng kích thước nhỏ có thể nằm trọn trong các vùng lớn.

Các phương pháp để thể hiện vùng sử dụng cấu trúc vector có thể được phân loại như: (a) Thể hiện bằng các đa giác thông thường; (b) Thể hiện bằng điểm tương ứng duy nhất cho từng đa giác, chứa các cặp tọa độ của các đỉnh; (c) Thể hiện thành một hệ thống với các quan hệ topo rõ ràng; và (d) Cấu trúc mạng topo.

Polygon đơn giản (thông thường): Phương pháp đơn giản nhất để thể hiện một đa giác là mở rộng cách lưu trữ đường cong như là một tập hợp các cặp tọa độ XY trên đường biên của đa giác (Hình 3.8). Tên hay ký hiệu của vùng mà đa giác đại diện được

lưu trữ dưới dạng các ký tự văn bản (text) để người sử dụng biết thông tin riêng của vùng. Mặc dù đây là cách biểu thị đơn giản nhưng nó còn tồn tại nhiều nhược điểm.

- Các đường biên của các vùng kề nhau phải được số hoá và lưu trữ hai lần. Điều này rất dễ tạo ra sai số khi ghép biên và sinh ra các vùng nhỏ mới không có ý nghĩa giữa các đường biên chung;
- Không có thông tin về các quan hệ láng giềng;
- Không thể hiện được sự hiện diện của các đa giác con nằm trọn bên trong một đa giác lớn khác;
- Sẽ rất khó khăn khi muốn kiểm tra được các lỗi trên các đường biên như đường biên không khép kín, đa giác có cạnh cắt chéo vào phía trong (weird polygons).



Hình 3.8 Cấu trúc vector dạng vùng

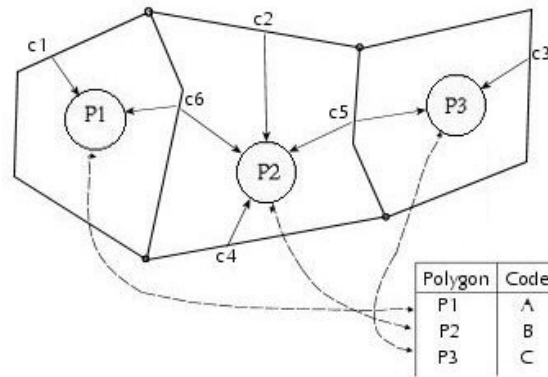
Polygon với các điểm tọa độ lưu trữ theo thứ tự: trong loại lưu trữ này, các cặp tọa độ điểm được đánh số liên tiếp theo một hướng nhất định và cũng được gắn với một chỉ số cho biết điểm đó thuộc về polygon nào.

Loại lưu trữ này tránh được sự trùng lặp lưu trữ các đường biên của đa giác nhưng các vấn đề về truy xét lân cận và nhận biết bao trùm vẫn không giải quyết được.

Polygon với cấu trúc topo riêng: Loại lưu trữ dữ liệu không gian này có một bảng dữ liệu riêng biệt để chứa các thông tin về thành phần các đường trong từng polygon (Hình 3.9).

DIME (Dual Independent Map Encoding) của Cục Điều tra Dân số Hoa kỳ (US Bureau of Census) là một trong những cấu trúc loại polygon có cấu trúc topo riêng biệt. Thực thể sơ cấp là một đoạn thẳng xác định bởi hai điểm ở hai đầu; đường cong thì được xác định bởi một tập hợp các đoạn thẳng. Mỗi đoạn thẳng được gắn hai chỉ dẫn ở hai đầu và mã thể hiện các polygon ở hai phía của nó.

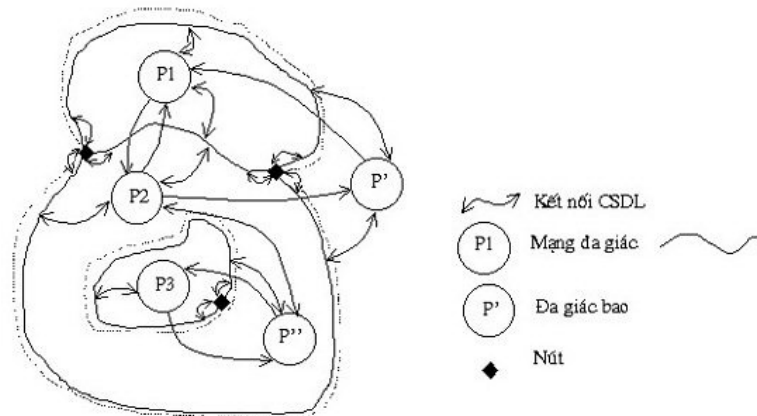
Loại cấu trúc dữ liệu vector này vẫn chưa tạo điều kiện tốt cho việc phân tích dữ liệu (chồng xếp hay tìm kiếm) và vẫn còn phải lưu trữ nhiều dữ liệu thừa.



Hình 3.9 Vùng với cấu trúc topo riêng biệt

Polygon với cấu trúc mạng topo hoàn chỉnh là loại cấu trúc vector được sử dụng trong hầu hết các phần mềm vector GIS hiện nay. Nó được trình bày trực quan như trong sơ đồ Hình 3.10, và sẽ được tiếp tục được bàn luận đến trong phần tiếp theo về tô pô cấu trúc dữ liệu vector.

Cấu trúc polygon với mạng topo hoàn chỉnh này cho phép giải quyết vấn đề các polygon con nằm trọn trong polygon lớn (đảo polygon), vấn đề xử lý các sai số đường biên và vùng cận kề, v.v.



Hình 3.10 Vùng với cấu trúc topo mạng đầy đủ

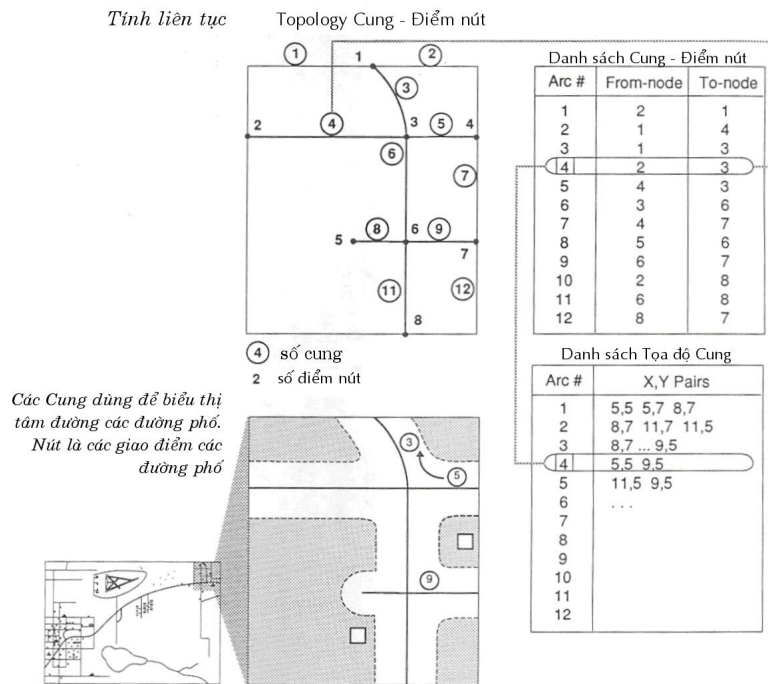
d. Quan hệ trong cấu trúc mạng vector (vector topology)

Trong bản đồ số, mối quan hệ không gian giữa các đối tượng trên bản đồ gọi là topology. Đối với một bản đồ, topology cho biết sự liên tục giữa các đối tượng, chỉ ra các polygon kề nhau và có thể xác định được một đối tượng như một vùng là một tập hợp của một loạt đối tượng khác (đường). Sau đây là mô tả các tính chất của topology chủ yếu dựa trên tài liệu của ESRI (1991).

Topology có 3 tính chất quan trọng: (a) tính liên tục (connectivity): các cung, đường nối với nhau tại các node; (b) tính tạo vùng (area definition): các cung hoặc đường (line) nối lần lượt với nhau thành vòng khép kín, bao quanh một vùng, tạo thành một polygon;

và (c) tính tiếp giáp (contiguity): các cung (arc) đều có hướng và được gán chỉ số biểu thị cho các polygon ở bên phải và bên trái nó.

❖ *Tính liên tục (connectivity)*



Hình 3.11 Tính chất liên tục của Topology

Theo định nghĩa, vị trí tọa độ các điểm (các cặp x, y) dọc theo một cung (arc hoặc vertices) xác định hình dạng của cung (arc) đó. Điểm cuối của cung (arc) gọi là điểm nút (node).

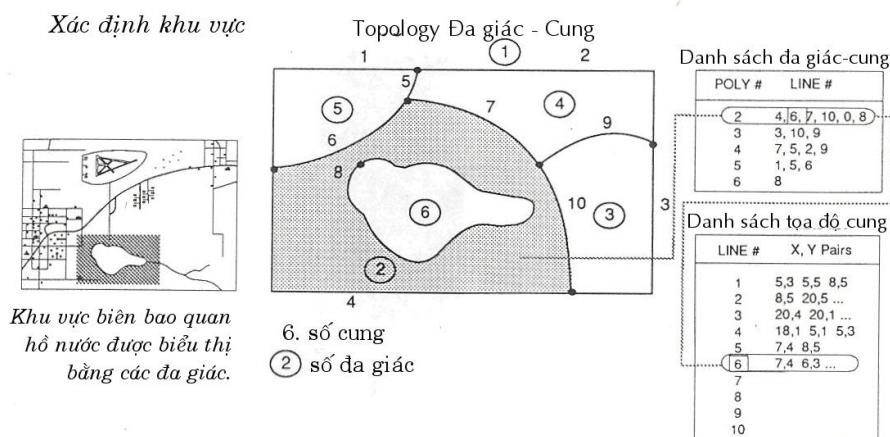
Mỗi cung (arc) bao giờ cũng có 2 nút: nút đầu và nút cuối (from-node và to-node). Các arc chỉ có thể nối được với nhau tại các điểm cuối của chúng (điểm nút). Để biết được cung nào nối với cung nào (những arc nào nối với nhau) thì chương trình máy tính chỉ việc rà soát dữ liệu ở tất cả các nút. Theo ví dụ trong Hình 3.11 thì các cung 3, 4, 5 và 6 gặp nhau tại nút số 3. Như vậy máy tính biết được có thể dịch chuyển dọc theo arc số 5 để đến arc số 3 vì chúng có chung một điểm node 3 nhưng không thể rẽ trực tiếp từ arc 5 đến arc 9 vì chúng không có chung điểm node nào cả.

❖ *Tính tạo vùng*

Polygon là một chuỗi các cặp tọa độ x, y nối liên tiếp với nhau để khép kín thành một vùng. Tuy vậy trong tổ chức dữ liệu vector được nhiều phần mềm GIS áp dụng, thì vùng (polygon) được lưu trữ dưới dạng tập hợp các cung (arc) bao quanh nó. Máy tính lưu trữ danh sách các vùng cùng với các cung tạo ra mỗi vùng, gọi là Polygon-Arc list, bên cạnh danh sách tọa độ các cung (Arc coordinate list) như thường lệ. Trong ví dụ Hình 3.12 thì polygon số 2 được xác định bởi các arc số 4, 6, 7, 10, và 8 tạo thành. Vì cung 8 tạo nên đa giác 6 là đa giác đảo nằm trong polygon số 2 nên trong danh sách đa giác-cung phải mã hóa thêm chữ số 0 trước số 8).

Mặc dù tên chỉ số của một cung có thể xuất hiện nhiều lần trong Danh sách Đa giác - Cung (trong Hình 3.12 thì tên cung số 6 có cả trong hàng cho vùng 2 và vùng 5), nhưng dữ liệu về cung đó chỉ được lưu trữ một lần duy nhất ở Danh sách Tọa độ Cung.

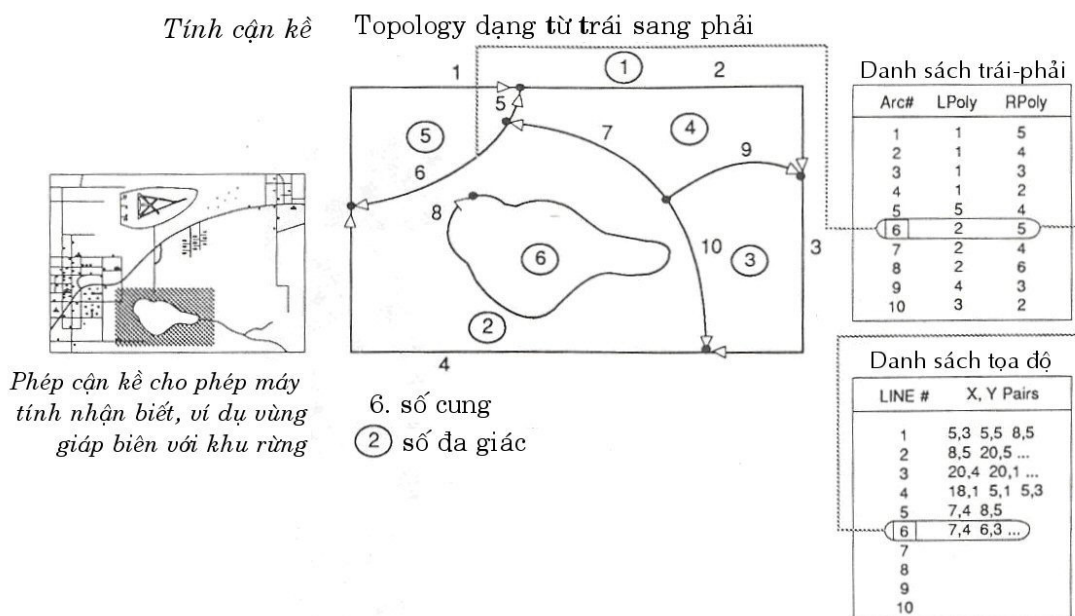
Việc tổ chức lưu trữ 2 danh sách bảng riêng biệt (đa giác – cung và tọa độ cung) như vậy đã tránh được sự lưu trữ lặp tọa độ các cung tạo nên đa giác và đồng thời bảo đảm được việc các đường biên chung của 2 đa giác kề nhau cũng trùng khớp lên nhau.



Hình 3.12 Tính chất tạo vùng của Topology

❖ Tính kề cận (contiguity)

Vì mỗi một cung (arc) đều được gán cho chỉ số hướng (node đầu và node cuối) nên có thể lưu trữ danh sách liệt kê các polygon nào nằm ở bên phải hoặc bên trái của từng cung (arc), gọi là Danh sách trái-phải (Left-right list). Như vậy ta có thể nhận ra các polygon nào kề cận nhau nếu chúng có chung ít nhất một cung. Ví dụ trong Hình 3.13 ta thấy polygon số 2 ở bên trái và polygon số 5 ở bên phải của arc số 6. Qua đó máy tính nhận ra ngay polygons số 2 và số 5 là liền kề nhau. Phần mềm Arc/Info sử dụng polygon chỉ số 1 để lưu trữ vùng ở bên ngoài đường biên của bản đồ.



Hình 3.13 Tính chất kề cận của Topology

3.2.3 So sánh mô hình dữ liệu Raster và Vector

Cấu trúc lưu trữ dữ liệu không gian trong một phần mềm GIS có thể bao gồm thông tin tôpô khi cần thiết. Thông tin về tôpô không những chỉ diễn tả vị trí của vật thể trong không gian mà còn chỉ ra những quan hệ không gian của vật thể với các vật thể cận kề nó. Những thông tin tôpô đặc biệt quan trọng đối với rất nhiều loại phân tích dữ liệu không gian như tự động nhận biết lỗi nhập dữ liệu, tạo cửa sổ biểu thị vật thể và giao diện trên màn hình, ứng dụng tính toán mạng lưới, xác định nếu một điểm có nằm trong một vùng, tạo vùng đệm, chồng xếp bản đồ, và các xử lý cắt bản đồ khác. Tuy vậy nếu ứng dụng phần mềm không đòi hỏi các thông tin chi tiết về quan hệ không gian thì có thể việc lưu trữ các thông tin tôpô chỉ làm phức tạp hóa không cần thiết cấu trúc của cơ sở dữ liệu. Điều này chỉ làm cho các động tác quản lý như cập nhật, mở rộng CSDL trở nên kém hiệu quả. Ví dụ chỉ cần cấu trúc dữ liệu vector không bao gồm thông tin tôpô, cho công việc thể hiện bản đồ lên màn hình và in ấn bản đồ.

Vấn đề so sánh ưu nhược điểm của hai loại cấu trúc dữ liệu không gian phổ biến raster và vector đã được nghiên cứu và trình bày trong rất nhiều sách giáo khoa về GIS. Mặc dù bản chất lưu trữ hai loại dữ liệu này hoàn toàn khác nhau nhưng việc so sánh sẽ giải thích được phần nào cho các khả năng ứng dụng của chúng. Các chỉ số so sánh cơ bản dựa trên khối lượng bộ nhớ để lưu trữ (hiệu suất lưu trữ dữ liệu), tốc độ truy cập, độ ổn định, hiệu suất xử lý và phân tích dữ liệu, độ chính xác khi hiển thị dữ liệu, v.v.

Mô hình raster có ưu điểm là có cấu trúc dữ liệu đơn giản, dễ dàng chồng xếp các dữ liệu, cho nhiều khả năng biến động không gian ở mức cao, có hiệu quả trong việc tăng dày và thể hiện các dạng ảnh số. Nhược điểm của mô hình là lưu trữ và nén dữ liệu gặp khó khăn, khó thể hiện các mối quan hệ vị trí không gian, chất lượng ảnh không cao. Muốn khắc phục điều này thường phải tăng độ phân giải, giảm kích thước của ô vuông cơ sở, pixel (tăng số lượng pixel) nhưng điều đó lại tăng khối lượng dữ liệu lưu trữ và tất nhiên ảnh hưởng lớn đến hiệu suất việc quản lý dữ liệu lưu trữ và công tác phân tích sau này.

Bảng 3.2 So sánh mô hình dữ liệu dạng vector và raster

	Raster	Vector
1. Nhập dữ liệu	Nhanh	Chậm
2. Khối lượng dữ liệu	Lớn	Nhỏ
3. Chất lượng đồ họa	Trung bình	Tốt
4. Cấu trúc dữ liệu	Đơn giản	Phức tạp
5. Độ chính xác hình học	Thấp	Cao
6. Khả năng phân tích vùng	Tốt	Kém
7. Khả năng phối hợp các lớp dữ liệu	Tốt	Kém
8. Khả năng tạo lập bản đồ	Đơn giản	Phức tạp

Mô hình vector có ưu điểm hơn hẳn raster trong việc lưu trữ dữ liệu. Có hiệu quả cao khi phân tích dữ liệu với mô hình topo ví dụ như trong phân tích mô hình mạng (network analysis). Hình ảnh bản đồ có chất lượng cao khi in ấn và thể hiện trên màn

hình. Tuy vậy mô hình vectơ có cấu trúc dữ liệu phức tạp gây trở ngại trong các phân tích dữ liệu như chồng xếp và xử lý cắt bản đồ. Bảng 3.2 tóm tắt các ưu nhược điểm khi sử dụng hai loại mô hình cấu trúc dữ liệu này.

3.3 Dữ liệu thuộc tính

Dữ liệu thuộc tính (DLTT) là các thông tin đi kèm với các dữ liệu không gian chỉ ra các tính chất đặc trưng cho mỗi đối tượng điểm, đường và vùng trên bản đồ. Phần lớn dữ liệu thuộc tính được lưu trữ trong các tệp tin riêng biệt và cấu trúc lưu trữ của chúng hiện nay được xây dựng dựa trên quan niệm về hệ thống quản lý cơ sở dữ liệu (database management systems, DBMS) như đã đề cập ở phần đầu của chương này

3.3.1 Các nhóm thông tin thuộc tính

Trong ngành môi trường và tài nguyên, người ta phân các thông tin thuộc tính thành các nhóm như sau:

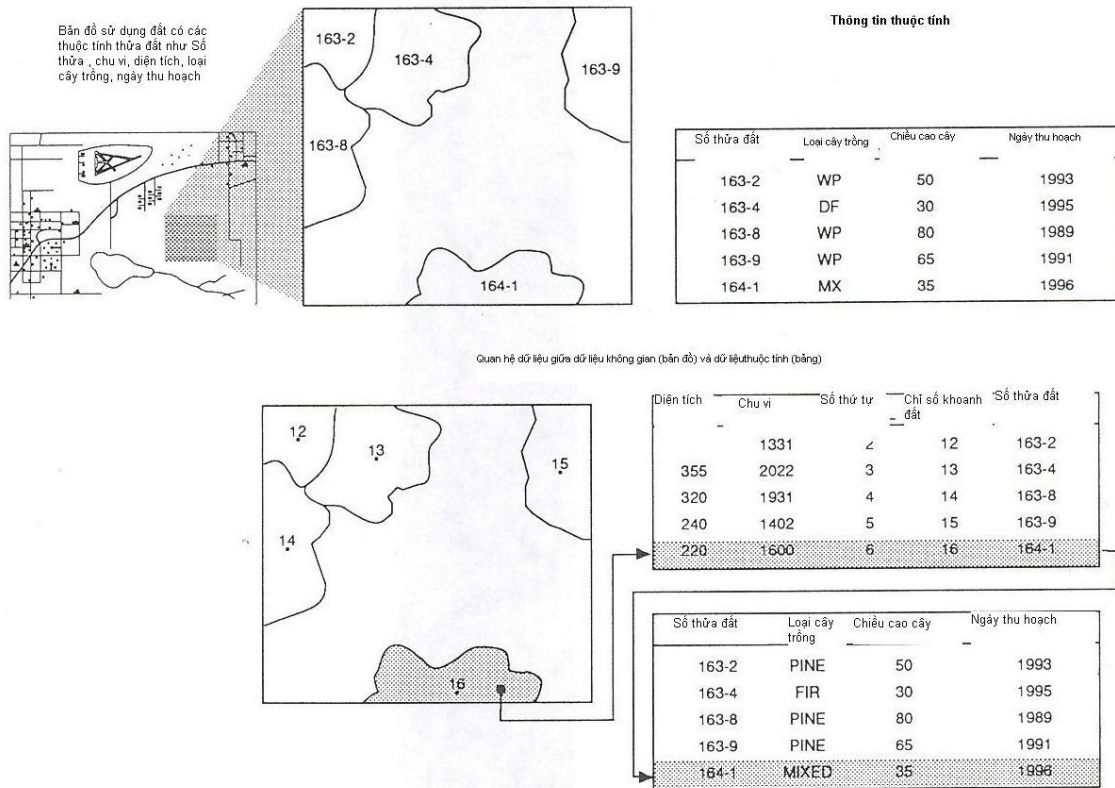
- *thông tin cơ sở hạ tầng*: gồm các thông tin hỗ trợ cho các lớp dữ liệu không gian như thông tin về các loại đường giao thông, hệ thống thủy lợi, thủy văn, về mạng lưới điện, mạng lưới cấp thoát nước, về các công trình cơ sở hạ tầng, v.v.
- *thông tin kinh tế văn hoá xã hội*: các thông tin về mạng lưới hành chính, dân cư, phát triển kinh tế, đầu tư giáo dục y tế, an ninh quốc phòng, v.v.
- *thông tin môi trường*: các thông tin về ô nhiễm môi trường, về hạn hán ngập lụt, sâu bệnh, về các thảm họa, rủi ro do thiên nhiên gây ra.
- *thông tin đất đai, địa chính*: các thông tin về tài nguyên đất, hệ thống thừa đất, đăng ký đất đai, hệ thống thuế đất, định giá đất và tài sản, hệ thống pháp luật đất đai, v.v.

Các thông tin thuộc tính thường được lưu trữ dưới dạng các tệp tin dữ liệu của các hệ quản trị dữ liệu như DBASE, FOXPRO, EXELS, ACCES, ORACLE. Thông thường các phần mềm GIS như MAPINFO, ARCINFO, ARCVIEW, MGE, ILLWIS thường có thêm phần chức năng quản trị cơ sở dữ liệu thuộc tính dưới dạng các tệp *.DAT, *.DBF.

3.3.2 Cấu trúc quan hệ giữa dữ liệu không gian và thuộc tính

Ví dụ cấu trúc quan hệ giữa hai loại dữ liệu quan trọng này trong GIS về tình hình sử dụng đất được minh họa trong Hình 3.14. Các mối quan hệ này thể hiện sự kết nối giữa 2 tệp dữ liệu có cùng một hay nhiều thuộc tính chung theo các kiểu quan hệ 1-1 (một tệp đến một tệp khác), 1-N (một tệp đến N tệp khác) hay N-M (nhiều tệp đến nhiều tệp khác). Ví dụ quan hệ giữa thửa đất và mã thửa đất là 1-1, giữa tờ bản đồ và các thửa đất là 1-N, giữa các thửa đất và các mục đích sử dụng đất là N-M.

Để thể hiện các quan hệ đó người ta sử dụng các trường khoá chung trong các tệp tin quan hệ. Các trường chung này có tên là khoá quan hệ (key field), ví dụ trường số thửa đất trong các bảng ở Hình 3.14 là trường khóa quan hệ.



Hình 3.14 Mô hình quan hệ giữa dữ liệu không gian và dữ liệu thuộc tính

Câu hỏi ôn tập

1. Hãy trình bày về mô hình Raster.
2. Hãy trình bày về mô hình Vector.
3. Thế nào là lớp dữ liệu? Vai trò của lớp dữ liệu trong quản lý và phân tích dữ liệu.
4. Các phương pháp nén dữ liệu raster Chain -code, run - length code, block - code, quade tree. Hãy nêu các ưu nhược điểm của mỗi phương pháp?
5. Pixel là gì? Vai trò của pixel trong mô hình dữ liệu raster? Cho ví dụ.
6. Hãy trình bày về mô hình Vector.
7. Trình bày ưu và nhược điểm của hai mô hình Vector và Raster.
8. Trình bày mô hình quan hệ không gian trong cấu trúc Vector (Vector topology).
9. Định nghĩa dữ liệu thuộc tính? Nêu các dạng dữ liệu thuộc tính cơ bản của GIS?
10. Mô hình quan hệ dữ liệu dạng giữa dữ liệu không gian và thuộc tính trong GIS?
11. Mô hình phân cấp dữ liệu không gian và thuộc tính trong GIS?

Chương Bốn

NHẬP VÀ CHỈNH LÝ DỮ LIỆU TRONG HỆ THỐNG THÔNG TIN ĐỊA LÝ

Quá trình xuất nhập và chỉnh lý dữ liệu có tầm quan trọng thiết yếu, quyết định cho thành công của bất cứ dự án GIS nào. Chương này giới thiệu các công đoạn nhập dữ liệu vào hệ thống GIS, và chỉnh lý dữ liệu cả về sai số và quy đổi hệ thống tọa độ và tham chiếu. Các dạng biểu thị kết quả cũng được trình bày ở cuối chương.

4.1 Nhập dữ liệu

Nhập dữ liệu là hoạt động gán mã cho các dữ liệu và ghi nhận chúng vào cơ sở dữ liệu (CSDL). Có 2 loại dữ liệu được nhập là dữ liệu không gian chỉ ra vị trí địa lý của dữ liệu và dữ liệu thuộc tính mô tả các dữ liệu không gian.

Dữ liệu được đưa vào CSDL thông qua các chức năng đầu vào của HTTTĐL. Nhập dữ liệu là một quá trình mã hoá lưu trữ và tổ chức dữ liệu vào CSDL. Đây là một quá trình rất quan trọng của hệ thống ảnh hưởng đến độ chính xác và tính logic của dữ liệu trong CSDL, do đó các thiết bị nhập số liệu cũng như các phần mềm nhập số liệu đều phải đảm bảo độ chính xác và thông thường có chi phí tương đối cao

Trong HTTTĐL, các dữ liệu không gian và phi không gian được lưu trữ và được tổ chức quản lý theo các tiêu chuẩn chặt chẽ của một CSDL. Người ta có thể phân tách thành ba công đoạn chính trong quá trình nhập dữ liệu mới vào HTTTĐL:

- Nhập dữ liệu không gian;
- Nhập dữ liệu thuộc tính; và
- Kết nối hai loại dữ liệu trên.

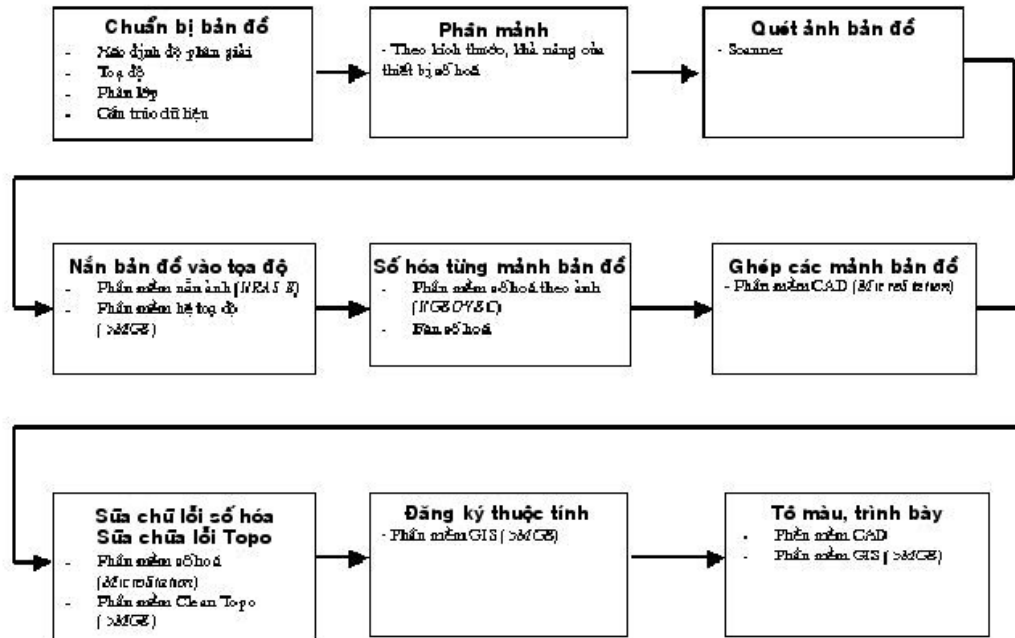
4.1.1 Nhập dữ liệu không gian

Dữ liệu đầu vào thường là các bản đồ giấy, các bản vẽ tay, các ảnh hàng không, ảnh chụp vệ tinh, điểm lấy mẫu và các dữ liệu khác đo gián tiếp bằng các thiết bị cảm ứng.

Phương pháp nhập dữ liệu phụ thuộc vào cấu trúc dữ liệu và phần mềm quản trị CSDL đang sử dụng. Hình 4.1 mô tả qui trình nhập dữ liệu không gian sử dụng hệ thống phần mềm MGE của Intergraph.

các phương pháp nhập dữ liệu không gian thông dụng là:

- Nhập số liệu bằng tay,
- Số hoá bằng bàn số hoá (digitizer) hoặc
- Sử dụng máy quét (scanner).



Hình 4.1 Sơ đồ nhập dữ liệu cho một HTTT địa lý

Nhập dữ liệu dạng vector

Hiện nay các dữ liệu bản đồ lưu trữ trong CSDL GIS thường được để dạng cấu trúc vector. Các dữ liệu được số hoá từ các bản đồ giấy cũng hay được số hoá dưới cấu trúc vector

Nhập dữ liệu vector bằng tay: Dữ liệu này là điểm, đường hay vùng được nhập từ bàn phím, theo các tọa độ chính xác hoặc nhập từ tệp dữ liệu ASCII.

Số hoá bằng bàn số hoá: Phương pháp này thường dùng để số hoá các bản đồ nguồn có tỉ lệ lớn. Số hoá bằng bàn số là quá trình đưa dữ liệu trên bản đồ giấy vào máy tính nhờ sự hỗ trợ của bàn số hoá. Bàn số hoá này cho ta biết tọa độ (X, Y) của một điểm bất kỳ trên bàn số hoá. Tọa độ này sẽ được chuyển đổi thành tọa độ tương ứng trên thực tế khi biết tọa độ các điểm khống chế.

Cấu tạo của bàn số hoá có hai thành phần 1) Bàn từ tính và 2) Con chuột cảm ứng.

Để số hoá được dữ liệu bản đồ giấy, người ta đặt bản đồ lên bàn số hoá, khai báo các mốc tọa độ khống chế và kích chuột vào các đối tượng trên bản đồ để nhận được tọa độ.

Bàn số hoá sẽ cho phép ta nhập ba kiểu dữ liệu chính:

- Điểm,
- Đường,
- Chữ,
- Các đối tượng vùng được xây dựng trên cơ sở cấu trúc topo mạng đa giác.

Do vậy việc xây dựng cấu trúc topo cho dữ liệu số hoá là rất quan trọng. Nó bao gồm từ quá trình cắt đường, sửa chữa lỗi đến quá trình tạo topo cho dữ liệu (xem thêm phần quan hệ topo trong cấu trúc dữ liệu dạng véc tơ, chương 3).

Các đặc tính địa lý trên bản đồ có thể được số hóa theo 2 kiểu sau:

- point mode: mỗi cặp tọa độ được ghi lại sau mỗi lần nhấn chuột;
- stream mode: các cặp tọa độ được ghi lại liên tục dọc theo các cung đường mà không cần nhấn chuột.

Trong một khoảng cách nhất định nào đó một điểm được ghi lại bất cứ lúc nào khi con chuột chuyển động theo hướng x hoặc y, hoặc là trong một khoảng thời gian số cặp tọa độ đã định trước được tạo ra theo từng giây. Kiểu số hóa này thường được sử dụng để số hóa các đường cong như là sông suối hoặc đường bình độ.

Các giá trị tọa độ được ghi lại bởi kỹ thuật viên số hóa sẽ phụ thuộc vào đặc điểm của các bản số hóa. Để biến đổi các giá trị này vào hệ thống tham chiếu bản đồ, các điểm giao trong lưới chiếu (lưới ô vuông) hay các góc của bản đồ cũng phải được số hóa. Sau đó sẽ so sánh giá trị tọa độ của bản số hóa và giá trị tọa độ thực của bản đồ gốc. Sử dụng phép chuyển đổi toán học, bất cứ tọa độ của một điểm bất kỳ nào được số hóa cũng sẽ được chuyển đổi vào hệ thống tham chiếu. Dạng đơn giản nhất của phép chuyển đổi sẽ là:

$$X = ax + by + c$$

$$Y = dx + ey + f$$

Trong đó X, Y là tọa độ thực của điểm trên bản đồ; x, y là tọa độ của bản số hóa a, b, c, d, e, f là những hệ số mà giá trị của chúng được tính dựa trên các cặp điểm tọa độ gốc. Giả sử có 3 cặp tọa độ gốc tương ứng với 3 điểm đã biết do đó sẽ có 6 phương trình cho 6 giá trị chưa biết từ a đến f. Nếu như có nhiều hơn 3 cặp giá trị tọa độ của 3 điểm có thể sử dụng cả phương pháp bình phương nhỏ nhất để tính các giá trị X, Y từ các giá trị x, y của các điểm tương ứng được số hóa.

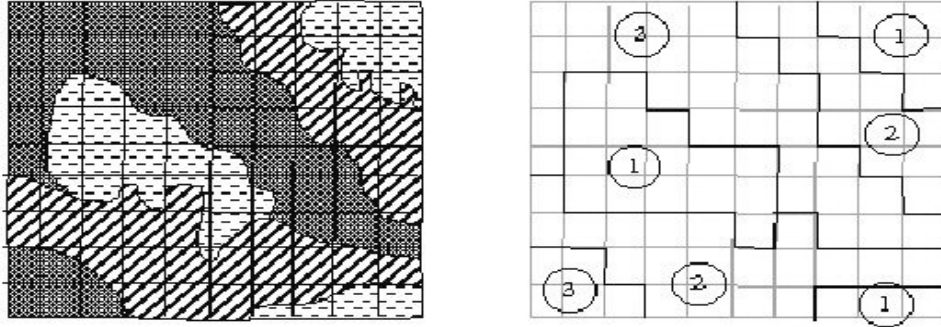
Trước khi số hóa đòi hỏi phải phân loại các đặc tính địa lý cần số hóa. Điều này có thể thực hiện bằng cách gán cho chúng một trị số mang ý nghĩa riêng như Đường = 1, Nhà ở = 2, v.v. và nhập thủ công từ bàn phím hoặc sử dụng con chuột với các phím số hoặc sử dụng các hộp trình đơn (menu box). Với hộp này máy tính sẽ biết phải ghi lại các đối tượng địa lý nào tiếp theo tương tự như kiểu mã đối tượng trước đó.

Độ chính xác của các bản số hóa có trong khoảng 0.01 mm. Tuy vậy trong thực tế khó đạt được như vậy vì còn phụ thuộc vào kỹ năng của kỹ thuật viên số hóa.

Số hoá đòi hỏi người kỹ thuật viên phải có các kỹ năng về tổ chức dữ liệu, các kỹ thuật số hoá. Ngoài ra đây là một quá trình tiêu tốn nhiều thời gian và rất dễ tạo ra nhiều sai sót về kỹ thuật như bỏ sót dữ liệu hoặc dư thừa dữ liệu làm ảnh hưởng đến các quá trình kỹ thuật tiếp theo như chỉnh sửa, phân tích và xử lý dữ liệu.

Nhập dữ liệu raster

Nhập dữ liệu raster bằng tay: Đối với hệ thống này, mọi điểm, đường, vùng đều được biến thành các cell. Phương pháp thông dụng nhất được diễn ra như sau: Đầu tiên chọn kích cỡ lưới ô, sau đó chồng lên bản đồ. Quá trình phân loại các ô vuông sẽ được thực hiện bằng tay. Giá trị tại từng ô nhận được từ bản đồ sẽ được ghi lại vào máy tính. Giả sử một lưới raster là 100×100 . Như vậy có 10000 giá trị được nhập sẽ là một công việc hết sức mệt mỏi. Hình 4.2 mô tả quá trình chuyển dữ liệu bản đồ giấy thành dữ liệu raster.



Hình 4.2 Raster hoá dữ liệu

Nhập dữ liệu bằng phương pháp quét

Hiện nay các dữ liệu bản đồ lưu trữ trong CSDL thường được để dưới cấu trúc vector. Các dữ liệu được số hoá từ các bản đồ giấy cũng hay được số hoá dưới cấu trúc vector. Ưu điểm của cách này là có độ chính xác cao, cập nhật dễ dàng và chiếm ít bộ nhớ lưu trữ. Số hoá trực tiếp từ bản đồ thành raster khó sửa chữa và phụ thuộc rất nhiều vào độ phân giải của dụng cụ. Do đó nếu người sử dụng làm việc với HTTTĐL sử dụng cấu trúc dữ liệu raster thì các dữ liệu vector phải được chuyển thành dữ liệu raster. Ở đây các tham số quan trọng là kích thước cell trong lưới ô kết quả. Nếu chọn kích cỡ cell lớn thì tiết kiệm bộ nhớ lưu trữ dữ liệu nhưng dữ liệu sẽ kém chính xác; nếu cần độ chính xác dữ liệu cao thì phải đặt kích cỡ cell nhỏ và kết quả sẽ có dung lượng lớn.

Chuyển đổi từ vector sang raster. là tìm tập pixel trong không gian raster trùng khớp với vị trí của điểm đường hay đa giác trong biểu diễn vector. Tổng quát tiến trình biến đổi là tiến trình xấp xỉ với vùng không gian cho trước, mô hình raster chỉ có khả năng xác định các vị trí có tọa độ nguyên.

Các thiết bị quét ảnh (scanner) được chế tạo mô phỏng lại hoàn toàn quá trình tạo dữ liệu raster bằng tay. Về thực chất, các thiết bị này chuyển đổi các tín hiệu ánh sáng phản xạ từ bản đồ thành giá trị pixel và ghi chúng lại theo một khuôn dạng nhất định. Chất lượng dữ liệu raster quét phụ thuộc vào thiết bị quét, về độ phân giải theo mỗi chiều của mặt phẳng

Quét ảnh là phương pháp nhanh đòi hỏi chi phí cao về thiết bị, có 2 loại máy quét:

- loại scan bản đồ dạng raster thường đi liền với software để biến đổi dữ liệu raster thành vector sau khi quét
- loại scan bản đồ dạng đường liên tiếp ví dụ như scan đường đồng mức, các bản đồ địa chất.

Quá trình quét (scanning) dựa trên nguyên tắc đơn giản là một điểm trên phần nào của bản đồ có thể có một hay 2 màu khác nhau (đen, trắng). Máy quét đưa vào một chùm ánh sáng laze có năng lượng thấp và một buồng cảm ứng (camera) có độ phân giải cao. Buồng camera được gắn với một dụng cụ cảm ứng không gian. máy quét có 2 loại: loại phẳng (flat bed) và loại dạng hình ống có thể xoay được. Độ phân giải được điều chỉnh theo chế độ trong các khoảng khác nhau. Các máy hiện đại có thể đạt tới độ phân giải 25- 50 micrômét. Do vậy số lượng pixel sẽ rất lớn và chiếm một dung lượng lưu trữ đáng kể. Ví dụ một bản đồ đen trắng có kích cỡ ô vuông là 50 cm với kích cỡ pixel là 50 micro met (0.05mm) có thể tạo ra 100 triệu pixel. Nếu cỡ pixel là 25 micromet thì sẽ có 400 triệu pixel. Nếu độ phân giải thấp kích cỡ pixel lớn thì tốc độ

scan nhanh nhưng quét được ít dữ liệu hoặc bỏ sót dữ liệu. Nếu độ phân giải cao máy sẽ quét chậm khi đó khối lượng dữ liệu cần lưu trữ sẽ rất lớn. Chất lượng dữ liệu raster quét phụ thuộc vào thiết bị quét, về độ phân giải theo mỗi chiều của mặt phẳng quét và độ phân giải theo tín hiệu phản xạ từ bản đồ. Thông thường trên thị trường Việt nam hiện có các máy quét với độ phân giải từ 600 – 2400 dpi (dot per inch), nghĩa là có thể phân biệt được từ 0.05 – 0.01 mm trong khi đó mắt thường chỉ phân biệt được 0.1 mm. Có 2 loại scanner chính: Loại scanner đen trắng và loại scanner màu. Loại đen trắng chỉ nhận được mức năng lượng phản xạ lại từ bản đồ, loại kia thì có bộ phận phân tích thành 3 dải màu cơ bản và ghi chúng thành các cấp độ khác nhau. Thường thì mỗi dải màu được phân thành 256 – 1024 cấp độ.

Thiết bị quét ảnh là thiết bị rất chính xác, do đó các scanner khổ lớn có công nghệ quét và hiệu chỉnh sai số hoàn toàn khác so với các scanner khổ nhỏ.

Các scanner thường được cài đặt chương trình phần mềm làm tăng cường chất lượng của ảnh quét trong quá trình phân tích sau khi quét.

Mục đích của biến đổi raster sang vector là làm phát sinh các đường biên bản đồ từ ảnh quét. Tiếp theo là gán chỉ danh cho các đường phát sinh đó. mức cao hơn là hệ thống có khả năng nhận dạng chữ viết và các đối tượng có hình dạng khác nhau trên bản đồ. Trong trường hợp ảnh vệ tinh thì phải phân lớp ảnh thành các hiện tượng đã biết. Có rất nhiều thuật toán để chuyển đổi dữ liệu raster sang vector. Người ta quan tâm đến 3 kiểu chuyển đổi sau.

- Nhận dạng vùng,
- Nhận dạng đường,
- Nhận dạng các ký tự.

Nhận dạng vùng thường được ứng dụng để chuyển đổi kết quả xử lý số liệu trong HTTTĐL raster sang vector với mục đích lưu trữ, tra cứu và in ấn. Các bản đồ vector chỉ đầu ra của một hệ thống. Trong hầu hết các HTTTĐL raster đều có chức năng này.

Nhận dạng đường thường được ứng dụng để nhập số liệu từ bản đồ quét thành dữ liệu vector. Quá trình nhận dạng này bao giờ cũng bắt đầu từ ảnh quét. Có 2 kiểu nhận dạng đường: Nhận dạng tự động, và Nhận dạng bán tự động.

Nhận dạng tự động là quá trình nhận dạng ảnh quét trong đó không có sự tham gia chỉnh sửa của con người. Kiểu nhận dạng này thường dùng để nhận dạng các bản đồ quét có chất lượng tương đối tốt. Quy trình nhận dạng theo trình tự sau (Hình 4.3):

- Tăng cường chất lượng ảnh quét (xoá các pixel thừa, làm trơn ảnh, v.v.),
- Lọc xương ảnh để nhận dạng đường,
- Chuyển đổi xương ảnh thành vector.

Chuyển đổi từ raster sang vector. Mục đích của biến đổi raster sang vector là làm phát sinh các đường biên bản đồ từ ảnh quét. Tiếp theo là gán chỉ danh cho các đường phát sinh đó. Mức cao hơn là hệ thống có khả năng nhận dạng chữ viết và các đối tượng có hình dạng khác nhau trên bản đồ. Trong trường hợp ảnh vệ tinh thì phải phân lớp ảnh thành các hiện tượng đã biết.

Nhận dạng ký hiệu bao hàm cả nhận dạng ký tự, chủ yếu dùng để nhận dạng các bảng số, văn bản. Nhận dạng ký hiệu ít khi được dùng để nhận dạng các ký hiệu hay ký tự trên bản đồ.

Nhận dạng tự động có 2 nhược điểm chính:

- Yêu cầu chất lượng ảnh quét cao, đòi hỏi quá trình sơ bộ xử lý công phu,
- Không cho phép hiệu chỉnh thông số trong quá trình xử lý nhận dạng, do đó vẫn phải kiểm tra, sửa chữa trên bản kết quả.

Hiện nay trên thế giới có nhiều phần mềm nhận dạng đường trên bản đồ quét và một số scanner đã cài đặt chương trình nhận dạng bản đồ.

Số hoá trực tiếp trên ảnh quét

Để khắc phục các nhược điểm trên, người ta thường dùng phương pháp nhận dạng bán tự động. Quá trình số hoá trên ảnh quét được trợ giúp bởi một số công cụ phần mềm nhận dạng (như công cụ IDRAS C, GEOVEC của Hệ phần mềm Mapping office). Người dùng hiển thị ảnh bản đồ quét lên trên màn hình và kích chuột vào đường trên bản đồ ảnh, phần mềm sẽ tự động số hoá dọc theo đường đó đến khi nào cắt phải đường khác thì dừng lại. Lúc đó người dùng lại kích chuột để chỉ hướng cho chương trình tiếp tục nhận dạng.

Hiện nay ở Việt nam đã có nhập công nghệ số hoá theo phương pháp này. Đây là một công nghệ chính xác và quan trọng nhất là nó kiểm soát được lỗi khi số hoá.

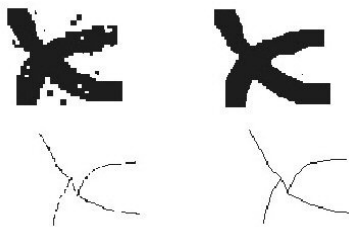
Các dạng số liệu nhập bằng các thiết bị khác:

Các ảnh hàng không và ảnh vũ trụ được thu thập từ các thiết bị cảm ứng từ xa, đặt trên máy bay hay vệ tinh. Các dữ liệu raster này có thể được nhập trực tiếp vào CSDL bằng các chức năng nắn đọc ảnh của HTTTĐL. Đặc điểm quan trọng cần chú ý là tính chất số liệu phụ thuộc vào thông số kỹ thuật của bộ phận cảm ứng.

Dữ liệu nội suy từ các dữ liệu khác bằng phương pháp toán học .Trong một số trường hợp các dữ liệu được nhập vào hệ thống nhờ vào kết quả phân tích nội suy toán học. Ví dụ như các số liệu đo đạc độ cao người ta có thể căn cứ vào giá trị độ cao của các điểm chuẩn để nội suy ra các điểm lảng giềng để có đầy đủ số liệu xây dựng bản đồ đường đồng mức (địa hình)

Meta data

Trong trường hợp số liệu được nhập vào HTTTĐL từ nguồn số liệu có sẵn từ một hệ thống thông tin khác (các số liệu chi tiết hoặc tổng hợp từ các mạng máy tính chuyên ngành khác) được gọi là meta data



Hình 4.3 Nhận dạng tự động đường trên bản đồ quét

4.2 Chỉnh sửa dữ liệu không gian

Khi số liệu đã được nhập vào hệ thống chúng cần được xử lý theo trình tự như sau:

4.2.1 Rà soát lỗi trong quá trình số hóa

Bước đầu tiên trong xử lý số liệu là rà soát các lỗi đối tượng điểm, đường và vùng sinh ra trong quá trình nhập dữ liệu (số hóa ảnh, bản đồ). Các lỗi về điểm thường gặp trong quá trình số hóa là sự sai lệch về vị trí (tọa độ) của các điểm dữ liệu so với tọa độ trong hệ tọa độ qui chiếu đã chọn. Trong mô hình dữ liệu dạng vector sự phát hiện ra sai lệch về vị trí của các điểm giao (node) của các cung, đường thẳng có ý nghĩa quan trọng trong việc chỉnh lý các dữ liệu không gian

Các lỗi các đối tượng dạng đường cần phải chỉnh sửa trong quá trình chỉnh lý số liệu là các đường thừa do khi nhập dữ liệu đối tượng được số hóa nhiều lần hoặc sinh ra do các điểm cuối của đường hoặc điểm giao của 2 hay nhiều đường chưa đúng vị trí thực. Theo ESRI (1994) trong trường hợp này có thể phân biệt 2 loại đường cần phải chỉnh sửa là đường cắt chưa tới (Under shooting arc) khi các đường chưa giao nhau và đường cắt quá (Over shooting arc) khi điểm nút giao của các đường lệch đi so với vị trí thực.

Các lỗi về vùng thông thường là các lỗi về thừa vùng hay thiếu vùng sinh ra trong quá trình nhập dữ liệu. Thừa vùng khi các đường biên của vùng được số hóa nhiều lần sinh ra các đa giác vụn (silver polygon) hoặc khi điểm cuối của các đường cắt quá giao nhau cũng có thể tạo thành các đa giác thừa lạ (weird polygon).

Vì các đối tượng đường, điểm được số hoá là các dữ liệu cơ bản của hệ thống nên chúng phải có các cấu trúc dữ liệu phù hợp. Các tọa độ và các thuộc tính của các đối tượng điểm đường vùng cần được thể hiện trên máy tính sao cho tất cả các đặc tính và các điểm riêng biệt đều có thể được xoá, thay đổi các mã(code) đặc tính hoặc có thể thay đổi vị trí. Trong quá trình chỉnh lý các lỗi khi số hoá bản đồ cần phải loại bỏ hoặc cập nhật lại dữ liệu bản đồ, hay các số liệu thuộc tính. Trong quá trình chỉnh lý các lỗi trên cần áp dụng một số thuật toán có liên quan như: Thiết lập topology để loại bỏ bớt các điểm và đường dư thừa, tạo lập các bảng nút (node topology), bảng cung (arc topology) và bảng vùng (polygon topology).

Khi số hóa đa giác mỗi đa giác chỉ cần 1 điểm để đặt nhãn. Hai loại lỗi liên quan đến điểm đặt nhãn trong đa giác có thể xảy ra là: thiếu hay thừa nhãn trong một đa giác. Lỗi về nhãn thường do khi số hóa ta đặt nhiều điểm tham chiếu cho đa giác

Các lỗi sinh ra khi số hóa vùng bao gồm:

- Xâu đoạn thẳng lơ lửng (dangling) do vẽ các đường cắt quá hay cắt chưa tới;
- Không gán nhãn cho các đa giác;
- Thừa nhãn cho một đa giác.

Điều quan trọng trong mạng các xâu đoạn thẳng để hình thành đa giác là các xâu phải kết thúc chính xác tại các điểm nút nơi các xâu gặp nhau. Hiệu chỉnh các xâu lơ lửng bằng cách chuyển dịch đầu cuối các xâu vào đúng vị trí của điểm nút. Chức năng này trong GIS gọi là chụp (snap), có khả năng chụp 2 hay nhiều điểm mà cách nhau một khoảng cách cho trước.

4.2.2 Các thao tác cơ bản với đường và điểm

Các đường thẳng cắt nhau

Nhiều ứng dụng đòi hỏi tìm giao điểm của 2 hay nhiều đoạn thẳng như khi xác định điểm giao nhau khi xây dựng topo của các đường thẳng được số hoá rồi rạc hay khi tìm

giao điểm của đường thẳng với đường biên bản đồ. Thông thường cần phải xác định điểm giao khi các trường hợp sau:

- Các đoạn thẳng kết thúc bằng các nút, chúng có thể cắt nhau,
- Những đường có độ dài không xác định có thể cắt hoặc không cắt nhau.

Nhiều công thức toán học có thể sử dụng để biểu diễn các đoạn thẳng như phương trình tuyến tính tổng quát biểu diễn đường thẳng trong không gian 2 chiều:

$$ux + vy + w = 0$$

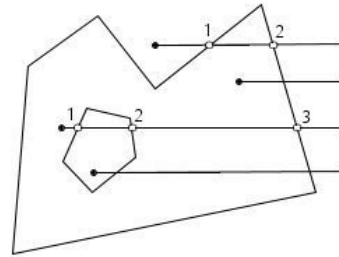
hay có thể viết dưới dạng khác: $y = a + bx$ trong đó b là hệ số góc và a là điểm chặn.

Trong GIS để biểu diễn đoạn thẳng phải có:

- phương trình đường thẳng,
- tọa độ các đầu nút.

Tìm kiếm điểm trong đa giác

Thủ tục nhằm xác định một điểm độc lập hay điểm nút của một đường thẳng có nằm trong một đa giác hay không dựa trên định lý nửa đường thẳng của Jordan. Giả sử ta có một điểm x, y và một đa giác (Hình 4.4). Để biết điểm nằm trong hay ngoài đa giác hãy phát sinh nửa đường thẳng bắt đầu từ điểm đã biết kéo dài đến tọa độ tột cùng của đa giác. Sau đó tính tổng số điểm giao của nửa đường thẳng với đa giác. Nếu tổng điểm là lẻ thì điểm nằm trong đa giác, nếu chẵn thì nằm ngoài đa giác. Để dễ tính số tổng điểm hãy vẽ nửa đường thẳng song song với trục X. Thủ tục trên đúng cho cả đa giác đảo và đa giác lõm. Tuy nhiên không áp dụng khi điểm nằm trên cạnh hay đỉnh của đa giác. Nếu đa giác có nhiều cạnh thì phải sử dụng chữ nhật bao. Nếu điểm nằm ngoài chữ nhật bao thì cũng nằm ngoài đa giác. Sau đó định lý nửa đường thẳng được áp dụng chỉ khi nào điểm nằm trong chữ nhật bao.



Hình 4.4 Tìm kiếm điểm trong đa giác

Xác định trọng tâm

Nhiều khi cần phải xác định 1 điểm để đại diện cho đa giác hay đường thẳng như chọn vị trí hiển thị nhãn bản đồ hay đo khoảng cách đến đối tượng khác, ta phải xác định 1 vị trí trong đa giác gọi là trọng tâm, xấp xỉ ở điểm giữa đa giác. Có thể chọn vài cách sau:

- Chọn bằng trực giác. Đa số trường hợp người ta sử dụng phương pháp đơn giản này, khi gán nhãn cho các vùng người dùng có thể chọn vị trí thích hợp trong đa giác để tạo nhãn cho vùng.
- Điểm được xác định từ các đỉnh của đa giác, lấy giá trị trung bình từ toàn bộ tọa độ của các đỉnh

$$X_c = \frac{\sum_{i=1}^N X_i}{N}$$

$$Y_c = \frac{\sum_{i=1}^N Y_i}{N}$$

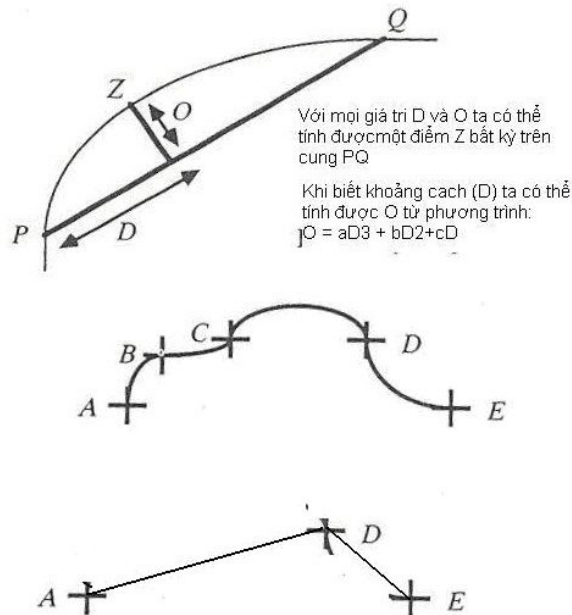
Trong trường hợp nhiều đỉnh tập trung về một phía đa giác thì trọng tâm sẽ dịch về phía chúng, trong trường hợp này sử dụng trọng lực thay cho trọng tâm các đỉnh hay tính từ trọng tâm của đường tròn nội ngoại tiếp hay chữ nhật ngoại tiếp. Nếu đa giác quá lõm thì phải dịch trọng tâm của đa giác vào phía trong.

- Được tính toán như tâm của hình chữ nhật, đường tròn nội hay ngoại tiếp;

Loại bỏ điểm dư thừa

Tất cả các phương pháp số hóa bản đồ đều làm phát sinh nhiều điểm và đoạn thẳng không cần thiết (lỗi thừa hay thiếu điểm hoặc đoạn thẳng). Số liệu mà máy tính nhận từ bản số hóa là các tọa độ điểm của một lớp dữ liệu (data layer).

$$E = x_1y_1, x_2y_2...x_ny_n$$



Hình 4.5 Loại bỏ điểm dư thừa

Các tọa độ này được phát sinh do người sử dụng nhấn bàn phím một cách ngẫu nhiên. Vì vậy 2 điểm liên tiếp được phát sinh có thể trùng nhau, có thể cùng nằm trên một đường thẳng và có thể gần nhau đến mức không cần thiết. Sau khi đã tạo bảng quan hệ topology tức là đã tạo bảng nút và bảng cung có thể tiến hành loại bỏ bớt các điểm dư thừa (Hình 4.5).

Coi tọa độ thứ nhất của cung là điểm chốt và tọa độ cuối là điểm di động. Hai điểm này sẽ tạo thành một đoạn thẳng AE (x1 y1, x5 y5). tính khoảng cách từ các điểm nằm giữa điểm chốt và điểm di động ta có các giá trị d1, d2, d3. Nếu tất cả các khoảng cách

này đều nhỏ hơn một giá trị D thì AE thuộc về lớp dữ liệu đang xét và các điểm nằm giữa điểm chốt và điểm di động bị loại bỏ. Điểm cố định mới là điểm di động hiện hành. Điểm có khoảng cách $d > D$ và ở xa điểm chốt nhất trở thành điểm di động mới (trên hình vẽ 4.6) điểm x_4, y_4 không thỏa mãn điều kiện trên được lựa chọn và trở thành điểm di động mới. Công việc lại tiếp tục với đoạn thẳng nối từ điểm chốt tới điểm di động mới. Như vậy tất cả các điểm đã đóng vai trò điểm cố định sẽ được lưu giữ lại cho lớp dữ liệu. Công thức tính khoảng cách từ các điểm x_i, y_i tới đoạn thẳng $Ax + By + C = 0$ là:

$$d = S / (A^2 + B^2)^{1/2}$$

với $S = A x_i + B y_i + C$

Chuyển đổi tọa độ và hệ qui chiếu

Nhiều hệ GIS không lưu trữ tọa độ theo khuôn mẫu độ, phút giây mà lưu trữ hệ tọa độ thập phân. Do vậy khi hiển thị và khi nhập tọa độ đặc trưng (tọa độ kiểm soát) phải chuyển đổi giữa chúng. Công thức biến đổi độ phút, giây sang khuôn mẫu độ thập phân như sau:

$$\text{Tọa độ thập phân} = \text{độ} + \frac{\text{Phút} + \frac{\text{Giây}}{60}}{60}$$

Đổi ngược lại từ độ thập phân sang độ, phút, giây giữ phân nguyên của số thập phân là độ thí dụ đổi 42,758333 thì sẽ là 42 độ, số phút bằng phần kết quả nguyên của phép nhân $0.758333 \times 60 = 45,5119980$, tính tương tự cho số giây ta được xấp xỉ 30 giây. Kết quả là 42 độ 45 phút và 30 giây.

Việc sử dụng các công thức và thủ tục khác nhau để biến đổi qua lại các tọa độ địa lý và tọa độ lưới phụ thuộc vào việc xác định rõ ràng các phép chiếu đã sử dụng. Trong thực tế có rất nhiều bản đồ thiếu các thông tin về phép chiếu nên thường phải dự đoán phép chiếu đã sử dụng để tìm ra phép chiếu phù hợp nhất hoặc phải biến đổi hệ tọa độ chưa rõ về hệ tọa độ rõ ràng với giả thiết nhận dạng được chính xác các điểm điều khiển; các vị trí có tọa độ chính xác trong cả 2 hệ tọa độ. Các phần mềm GIS hiện nay thường có chức năng chuyển đổi tọa độ và hệ qui chiếu về lưới chiếu phẳng UTM và Hệ qui chiếu WGS – 84. Tất cả các dữ liệu về hệ qui chiếu, hệ tọa độ - độ cao nhà nước được xác định trên cơ sở lưới đo đạc vệ tinh GPS và được lưu giữ trong dạng: tên điểm, cấp hạng, tọa độ X, tọa độ Y, độ cao H, các yếu tố của ellip sai số. Các trị đo của lưới tọa độ, độ cao được lưu trữ ở dạng trị đo nối giữa các điểm, sai số đo.

Biến đổi hệ trục tọa độ là chuyển đổi tượng trên hệ trục tọa độ cố định và cố định đối tượng, chuyển dịch hệ trục. Ví dụ trên Hình 4.6 cho thấy x, y là vị trí đối tượng trước khi biến đổi và u, v là vị trí sau khi biến đổi. Có 2 nhóm biến đổi chính là biến đổi affine và biến đổi đường cong. Biến đổi affine giữ cho các đường song song vẫn song song theo 4 kiểu cơ bản sau:

a) Chuyển dịch gốc tọa độ không xoay

$$u = x - a$$

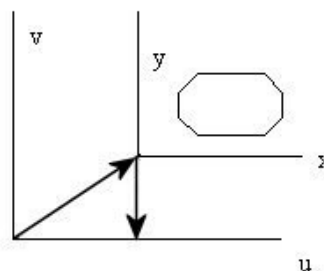
$$v = y - b$$

b) Co giãn đối tượng: cố định gốc và trục, thay đổi tỉ lệ (kích thước, đối tượng)

$$u = cx$$

$$v = dy$$

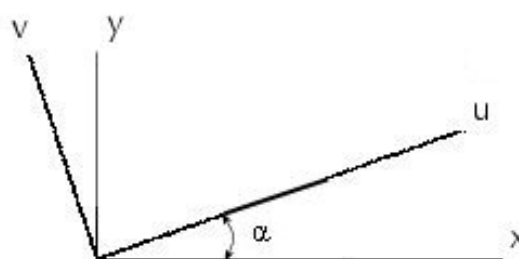
c) Xoay: gốc cố định, Trục xoay quanh gốc một góc α



Hình 4.6 Chuyển đổi hệ tịnh tiến hệ tọa độ

$$u = x \cos(\alpha) + y \sin(\alpha)$$

$$v = x \sin(\alpha) + y \cos(\alpha)$$



Hình 4.7 Chuyển đổi góc quay hệ tọa độ

d) Phản chiếu: đảo chiều trục đối tượng là ảnh đảo:

$$u = x$$

$$v = c - y$$

e) Biến đổi affine phức hợp: tổ hợp của các kiểu biến đổi trên:

$$u = a + bx + cy$$

$$v = d + ex + fy$$

Biến đổi cong: Khi phương trình biến đổi affine có thể được mở rộng thành số mũ, đường song song sẽ không còn song song và có thể trở thành đường cong hay bề mặt cong. Một giải pháp biểu diễn mối quan hệ các ảnh cong với một bản đồ phẳng chính xác là dựa trên phép chiếu biến đổi bản đồ.

Không có một phương pháp chính xác nào để hiệu chỉnh bản đồ vì dữ liệu có được không tuân theo một phép chiếu chuẩn hoặc theo một hệ tọa độ chuẩn như ảnh có thể xiên, thang đo không đồng đều. Để khắc phục người ta sử dụng phương pháp hiệu chỉnh điểm điều khiển mặt đất thực hiện so sánh vị trí của đối tượng trong dữ liệu với vị trí trên bản đồ tin cậy. Những vị trí này gọi là điểm khống chế mặt đất (ground control point) là điểm có thể nhận ra bằng các tọa độ chính xác của nó trong một hệ tọa độ đã xác định. Việc lựa chọn vị trí và số lượng điểm khống chế mặt đất là rất quan trọng. Các điểm khống chế thực được phân bố đều trên toàn bộ ảnh, đặc biệt là ở các đường viền, cạnh và không chỉ tập trung trên một vùng nhỏ. Để tránh sai sót khi thu nhận bản đồ như mô tả trên người ta áp dụng các phương pháp nắn chỉnh hình học để chuyển từ tọa độ ban đầu gọi là tọa độ sai về tọa độ chính xác hơn (tọa độ đúng).

Biên vẽ tổng hợp bản đồ từ tỉ lệ lớn hơn hoặc ghép mảnh hay các tờ bản đồ, thành lập các bản đồ chuyên đề theo những chủ đề nhất định. Bản đồ chuyên đề dùng để hiển thị các thuộc tính của các đối tượng bản đồ dưới dạng đồ họa. Thông thường căn cứ vào các thuộc tính của các đối tượng bản đồ chuyên đề biểu diễn dữ liệu bằng cách làm bóng, mã hoá màu, phủ các điểm với mật độ điểm khác nhau hay hiển thị các biểu tượng với các kích cỡ khác nhau trên các vùng của bản đồ

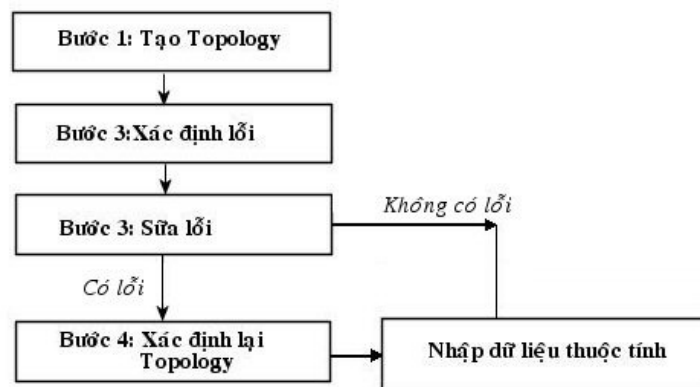
Thiết lập quan hệ không gian (topology)

Các hình thể trên bề mặt Trái đất được ánh xạ lên một mặt phẳng không gian 2 chiều như các điểm, đường và vùng. Hệ thống tọa độ x,y được sử dụng để lưu trữ các thông tin trên bản đồ. Mỗi điểm được lưu trữ như một cặp tọa độ, đường được lưu trữ như hàng loạt tọa độ có thứ tự, vùng được lưu trữ như một loạt các đoạn thẳng bao quanh vùng. Đó là các khái niệm để định nghĩa chính xác các mối quan hệ không gian. Các kiểu quan hệ khác nhau được thể hiện bằng một danh sách các đặc điểm, ví dụ như một vùng được định nghĩa bởi các cung tạo thành đường biên bao quanh nó. (xem thêm chương 3, phần cấu trúc quan hệ dữ liệu dạng vectơ).

Việc tạo lập và lưu trữ các mối quan hệ topo sẽ rất hiệu quả trong việc xây dựng Hệ thống thông tin địa lý. Nó cho phép xử lý dữ liệu nhanh hơn trên một tập dữ liệu lớn hơn. Khái niệm topo giúp thuận tiện cho các chức năng phân tích, ví dụ như mô hình luồng chảy qua các đường liên thông trong một mạng, ghép các vùng lân cận có cùng một tính chất đặc trưng, xác định các đặc tính kề cận và chồng ghép các đặc tính địa lý.

Xây dựng topology là quá trình thiết lập mã hóa các quan hệ giữa các điểm, các cung/đường và các vùng. Trong quá trình thiết lập các bảng nút, bảng cung và bảng đường được kiến tạo. Các bước chính để tạo mối quan hệ topo bao gồm:

- sắp xếp dữ liệu trong tệp lưu tọa độ bản đồ sao cho tọa độ theo trục Y tăng dần.
- Loại bỏ bớt các điểm và đường dư thừa.
- Tạo các bảng nút, bảng cung và bảng vùng.



Hình 4.8 Phát hiện và sửa lỗi dữ liệu không gian dạng vectơ

Hình 4.8 mô tả quá trình chỉnh sửa lỗi dữ liệu không gian là quá trình hoạt động luân phiên giữa xây dựng topology và sửa lỗi với công cụ ARCEDIT của phần mềm GIS – ARC/INFO. Dữ liệu không gian chỉ được hoàn chỉnh sau khi tất cả các thao tác

như soát lỗi, thiết lập topology, sửa lỗi với công cụ và hiệu chỉnh hệ tọa độ đã được thực hiện.

4.3 Nhập dữ liệu thuộc tính phi không gian

Dữ liệu thuộc tính phi không gian là những tính chất, đặc điểm riêng mà thực thể không gian cần đến để thể hiện trong HTTTĐL. Chúng không phải là các dữ liệu không gian. Ví dụ một con đường cần được số hoá như một tập các pixel nối với nhau trong cấu trúc dữ liệu raster hoặc là một thực thể dạng đường trong cấu trúc vector. Đường trong HTTTĐL còn được thể hiện với một màu nào đó hoặc ký hiệu hoặc một vài con số đi kèm theo. Các con số đi kèm này có thể là kiểu của đường, dạng bề mặt đường, phương pháp xây dựng, ngày xây dựng, v.v. Tất cả các số liệu này đều được gán chung cho một thực thể, do đó sẽ rất hiệu quả nếu chúng ta ghi và quản lý chúng riêng. Các dữ liệu này có chung một mã khoá với thực thể mà nó gắn với. Khi cần, lần theo mã khoá đó, người ta sẽ nhanh chóng khôi phục toàn bộ số liệu về thực thể.

Dữ liệu thuộc tính phi không gian thông thường được các Hệ quản trị cơ sở dữ liệu (HQTCSDL) quản lý. Hiện nay đa phần các HTTTĐL chuyên nghiệp đều dựa vào một HQTCSDL quan hệ để quản lý số liệu thuộc tính phi không gian của mình. Các HTTTĐL nhỏ hơn thì quản lý số liệu dưới dạng ASCII hay sử dụng các khuôn dạng EXCEL, DBASE thành các bảng riêng biệt.

Đối với các HQTCSDL quan hệ, người dùng sẽ nhập số liệu tuân thủ các qui tắc của một HQTCSDL quan hệ. Quá trình nhập số liệu diễn ra như sau:

Thiết lập CSDL mới nếu chưa có CSDL. Nếu đã có CSDL, khởi động CSDL. Đây là các bảng thuộc tính (attribute Table) thường được đi kèm khi nhập các dữ liệu không gian. Ví dụ khi số hóa bản đồ đất thì các bảng thuộc tính dạng *.dbf kèm theo có chứa các thông tin về chỉ số thửa đất, chu vi, diện tích và một số các thông tin hỗ trợ khác như các đặc tính lý hóa học đất, loại đất. Các thông tin này phải được nhập vào cơ sở dữ liệu bằng các mở các bảng tương ứng để nhập số liệu hoặc tạo các bảng thuộc tính quan hệ hoàn toàn mới. Cần chú ý rằng CSDL thuộc tính chỉ được tạo ra sau khi đã hoàn thiện xong CSDL không gian.

4.4 Kết nối dữ liệu không gian và dữ liệu thuộc tính

Việc kết nối giữa các đối tượng trên bản đồ và các bản ghi chứa các thông tin mô tả tương ứng trên cơ sở dữ liệu được thực hiện đơn giản trên cơ sở mối quan hệ 1-1

Điều kiện để thực hiện được thao tác kết nối hai dạng dữ liệu không gian và phi không gian là các dữ liệu không gian phải được nhập vào hệ thống và phải được tiền xử lý. Điều này có nghĩa là các đặc tính địa lý (điểm, đường và vùng) phải được đặt đúng vị trí thực và được bổ sung với đầy đủ các thông tin cần thiết với các mẫu tin tương ứng trong bảng thuộc tính. Việc kết nối giữa các đối tượng trên bản đồ và các bản ghi chứa các thông tin mô tả tương ứng trên cơ sở dữ liệu được thực hiện đơn giản trên cơ sở mối quan hệ 1-1 giữa các đối tượng địa lý trên bản đồ và các mẫu tin tương ứng mô tả thuộc tính của nó. Số duy nhất để tạo mối liên kết là mã (code) của đối tượng, gọi là chỉ số liên kết thuộc tính.

Thông thường người ta sử dụng một chương trình phần mềm để thực hiện công việc kết nối này. Đối với các thực thể không gian, người dùng sẽ phải nhập các mã khoá một cách trực tiếp cho từng thực thể. Đối với dữ liệu thuộc tính phi không gian thì người dùng nhập mã khoá vào CSDL quan hệ thông qua HQTCSDL. Người dùng sẽ dùng chương trình phần mềm kết nối để khai báo cho HTTTĐL.

4.5 Hiện thị đầu ra của dữ liệu địa lý

Biểu diễn dữ liệu đầu ra là một chức năng quan trọng của bất kỳ một HTTTĐL nào. Chức năng này không phải chỉ cho phép người dùng xem kết quả tính toán mà còn được sử dụng nhiều ngay trong quá trình nhập, sửa chữa dữ liệu. Đây là một cách hội thoại tốt nhất giữa người dùng với hệ thống, trợ giúp cho người dùng có được những kết luận chính xác hơn trong các bước xử lý số liệu tiếp theo.

Dữ liệu của HTTTĐL có thể được trình bày ở đầu ra theo nhiều cách khác nhau:

1. Hiện thị trên các thiết bị màn hình. Hiện thị trên màn hình: Thông tin được biểu diễn trên màn hình bằng các màu sắc theo các tiêu chuẩn hiện thị của bản đồ. Các công cụ quen thuộc sẽ giúp người dùng phóng to, thu nhỏ, nhìn toàn cảnh, xoay các góc khác nhau để hiện thị.

2. In ra giấy hoặc phim bằng các thiết bị in, thiết bị vẽ,

3. Chuyển đổi sang khuôn dạng khác hoặc đưa ra thành các tệp tin.

Câu hỏi ôn tập

1. Trình bày mục đích của hoạt động nhập dữ liệu không gian? Cho biết phương pháp nhập dữ liệu không gian thủ công?

2. Nguyên lý hoạt động của phương pháp nhập dữ liệu bằng bàn số hóa?

3. Các công việc chuẩn bị dữ liệu trước khi nhập dữ liệu vector từ bản đồ giấy ảnh hưởng thế nào đến các hoạt động về sau với dữ liệu?

4. Ưu và nhược điểm của phương pháp nhập dữ liệu dạng vector?

5. Xác định độ phân giải của máy quét dữ liệu Raster có tầm quan trọng thế nào đối với chất lượng công việc nhập dữ liệu?

6. Mô tả quá trình nhập dữ liệu raster bằng phương pháp truyền thống và phương pháp quét dữ liệu? Vì sao phương pháp quét dữ liệu kết hợp giữa người dùng và phần mềm trợ giúp lại phổ biến ở Việt nam hiện nay?

7. Trình bày nguyên nhân các lỗi thường gặp trong quá trình nhập dữ liệu?

8. Chỉnh sửa dữ liệu không gian bao gồm các công việc gì? Ý nghĩa của các công việc đó trong việc hoàn thiện dữ liệu không gian

9. Ý nghĩa của việc xây dựng quan hệ topo trong chỉnh lý dữ liệu không gian

10. Cách tổ chức và nhập dữ liệu thuộc tính trong GIS?

11. Nêu điều kiện để liên kết dữ liệu không gian và dữ liệu thuộc tính trong GIS?

12. Trình bày một số đặc điểm của các công cụ xuất dữ liệu cho một hệ GIS?

Chương Năm

MÔ HÌNH SỐ ĐỘ CAO (DEM)

Mô hình số độ cao (Digital Elevation Model, DEM) là sự biểu diễn bằng số độ cao địa hình một khái niệm liên quan chặt chẽ đến GIS và được ứng dụng rộng rãi trong để giải quyết nhiều vấn đề thực tiễn không chỉ trong giới hạn biểu diễn địa hình. Chương này trình bày khái niệm DEM, phương pháp thành lập DEM và các ứng dụng của nó.

5.1 Giới thiệu chung

Khác với các đơn vị sử dụng đất, phân loại đất và địa chất, độ cao địa hình có xu hướng biến đổi liên tục nên không thể biểu diễn chúng bằng bản đồ chuyên dụng thông thường. Tất nhiên trong địa hình tồn tại những vùng thay đổi độ cao đột ngột như vực sâu, khe núi nhưng phần lớn chỉ là sự thay đổi cục bộ.

Thông thường thì sự thay đổi độ cao địa hình được thể hiện bằng một loạt đường đồng mức mà các điểm trên một đường đồng mức có cùng một giá trị độ cao. Các đường này là các đường cong khép kín mà trong GIS người ta gọi là các polygons. Bằng phương pháp này thì yếu tố địa hình cũng được thể hiện và lưu trữ trong GIS như trong các bản đồ số chuyên dùng khác. Tuy vậy phương pháp biểu thị đó chưa phải là tối ưu khi sử dụng phương pháp số để phân tích và để mô hình hóa. Người ta cần một phương pháp tốt hơn để hiển thị và phân tích loại dữ liệu thay đổi liên tục (tương tự như số đo độ cao địa hình) và phương pháp đó là mô hình số độ cao.

Bất kỳ sự biểu thị bằng số sự thay đổi liên tục của độ cao trong không gian đều được gọi là mô hình số độ cao (Digital Elevation Model, DEM). Nó có thể là độ cao tuyệt đối của các điểm trên bề mặt quả đất, độ cao của các tầng đất, hoặc của mực nước ngầm. DEM còn nhiều tên gọi khác trong tiếng Anh như là Digital Terrain Model (DTM), Digital Terrain Data (TDD) và Digital Terrain Elevation Data (DTED). Ngoài ứng dụng để biểu thị địa hình, DEM còn có thể được ứng dụng để thể hiện sự thay đổi liên tục trong khoảng không hai chiều của bất kỳ thông số môi trường khác nào.

Sự cần thiết của DEM Mô hình số độ cao có vô vàn ứng dụng trong thực tiễn, đặc biệt phổ biến là những ứng dụng sau:

1. Lưu trữ dữ liệu bản đồ số địa hình trong các cơ sở dữ liệu (CSDL) quốc gia.
2. Giải quyết tính toán đào đắp đất trong thiết kế đường và các dự án kỹ thuật công chánh khác.
3. Biểu thị ba chiều trực quan điều kiện địa hình có mục đích quân sự (thiết kế hệ thống đạn đạo, huấn luyện phi công) và cho mục đích thiết kế và quy hoạch cảnh quan (kiến trúc cảnh quan).
4. Phân tích tầm quan sát xuyên địa hình (tương tự dùng cho mục đích quân sự và thiết kế cảnh quan).
5. Thiết kế xác định vị trí cho đường giao thông và cho đập nước.
6. Phân tích thống kê và so sánh các loại địa hình.
7. Tính toán và thành lập bản đồ độ dốc, bản đồ hướng dốc, bản đồ hình dạng mái dốc để từ đó thành lập ảnh địa hình trực quan có hình bóng (ứng dụng

- trong nghiên cứu tầng địa chất hay dự báo khả năng xói mòn đất và dòng chảy mặt).
8. Sử dụng làm bản đồ nền hay bản đồ tích hợp với các bản đồ chuyên dụng như bản đồ loại đất, loại sử dụng đất hay thảm thực vật.
 9. Sử dụng như là dữ liệu vào cho các mô hình mô phỏng cảnh quan và các quá trình tự nhiên liên quan đến cảnh quan môi trường.
 10. Khi thay thế độ cao bằng một trong các thông số thuộc tính khác thì DEM có thể biểu thị trực quan dạng mặt cong cho vấn đề quãng thời gian hành trình, giá thành, dân số, mức độ ô nhiễm, mực nước ngầm, v.v.

5.2 Phương pháp biểu thị DEM

Sự biến đổi giá trị độ cao địa hình trên một vùng đất có thể được mô hình hóa theo nhiều cách. DEM có thể được biểu thị và lưu trữ dưới dạng hàm số toán học ba chiều (phương trình mặt phẳng) hay dưới dạng các điểm hoặc các đường hình ảnh như liệt kê ở Bảng 5.1.

Bảng 5.1 Phương pháp biểu thị mặt cong địa hình

<i>A. Phương pháp toán học</i>	
Toàn vùng	
	Dãy Fourier Đa thức bậc bốn bội
Chi tiết	
	Chia vùng đồng đều Chia vùng không đồng đều
<i>B. Phương pháp vật thể bản đồ</i>	
Đường đồng mức (đường bình độ ngang)	
Đường mặt cắt dọc	
Điểm (ma trận độ cao) hay mạng lưới đều (regular rectangular grid, GRID)	
Vector: Mạng không đồng đều tam giác (triangular irregular network, TIN)	

5.2.1 Phương pháp toán học

Phương pháp toán học để biểu thị mặt cong địa hình chủ yếu dựa vào các hàm số toán ba chiều và có khả năng mô phỏng với độ nhẵn rất cao các mặt địa hình phức tạp. Phương pháp cục bộ chia vùng mô phỏng ra thành các miếng bé hình vuông hoặc hình dạng tùy ý có diện tích tương tự nhau và độ cao của từng miếng sẽ được ước lượng dựa trên độ cao các điểm đã quan trắc trong miếng đó. Với mục đích bảo đảm sự liên tục của độ dốc qua đường biên giữa các miếng con thì người ta sử dụng các hàm số đối trọng (weighting functions). Các hàm số xấp xỉ rời rạc (piecewise approximation) rất ít khi được sử dụng trong việc thành lập bản đồ số nhưng lại rất phổ biến trong hệ thống máy tính hỗ trợ thiết kế (CAD, computer added design).

5.2.2 Phương pháp vật thể bản đồ

Phương pháp sử dụng vật thể đường đầu tiên truyền thống trong bản đồ học để biểu diễn bề mặt địa hình là sử dụng đường bình độ hay còn gọi là đường đồng mức. Mọi điểm nằm trên cùng một đường đồng mức sẽ có cùng một giá trị độ cao.

Phương pháp sử dụng mặt cắt dọc để biểu diễn độ cao có ứng dụng thuận tiện để phân tích độ dốc vùng nghiên cứu. Tuy nhiên như đã đề cập ở trên hai phương pháp sử dụng đường trên không thuận tiện cho mục đích phân tích dữ liệu trong GIS. Vì vậy phương pháp chung nhất trong hệ GIS là sử dụng mô hình lưới đều GRID (Regular Rectangular Grid) hay lưới tam giác không đều TIN (Triangular Irregular Network).

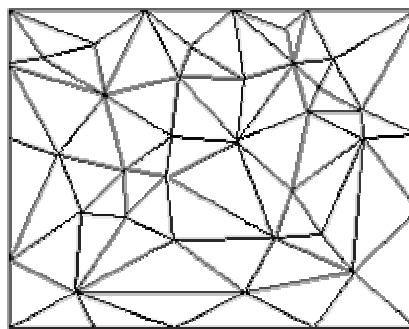
Mô hình lưới đồng đều hay còn gọi là ma trận độ cao được thành lập từ việc phân tích lập thể ảnh hàng không hoặc có thể thông qua việc nội suy từ lưới dữ liệu quan trắc độ cao. Do máy tính có khả năng xử lý ma trận dễ dàng nên dữ liệu loại mô hình GRID này rất phổ biến, được sử dụng cho các hệ GIS dạng raster. Trong mô hình raster GRID này vùng địa hình được chia thành các ô (cell) trên cơ sở hàng và cột. Mỗi một ô chứa độ cao của điểm trung tâm của ô. Ma trận độ cao được sử dụng để thành lập đường đồng mức, tính toán độ dốc, hướng dốc và xác định đường biên các lưu vực sông.

Tuy vậy phương pháp lưới đồng đều này có các nhược điểm sau:

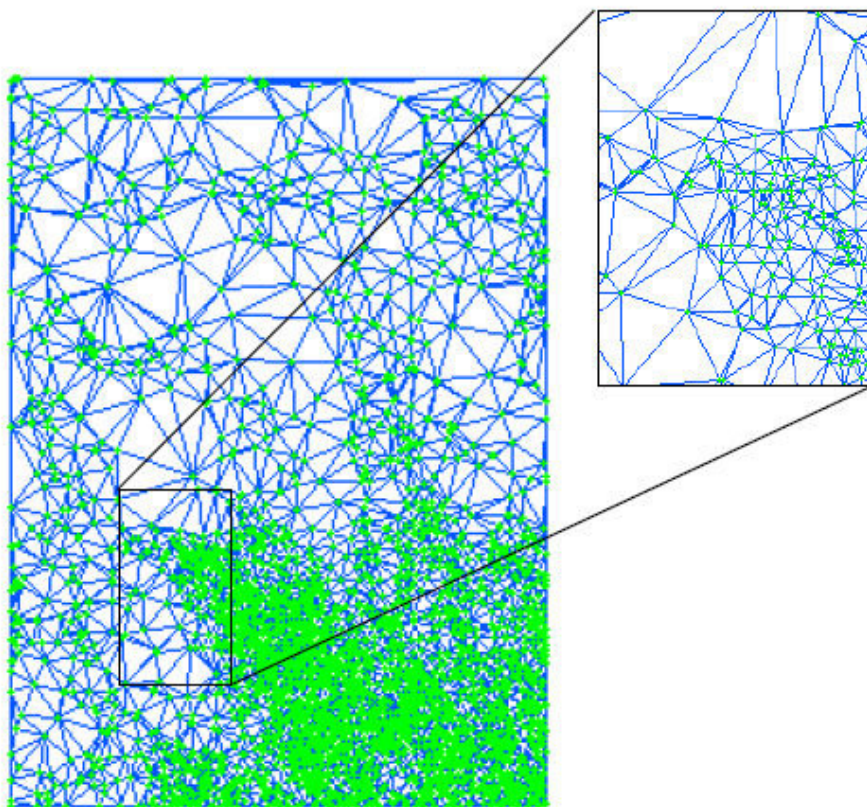
- Tồn tại số lượng dữ liệu không cần thiết tại các vùng có địa hình đồng nhất;
- Không có khả năng thích ứng để biểu thị các vùng có địa hình phức tạp trừ lúc thay đổi toàn bộ kích thước ma trận.

Như vậy lưới đồng đều không có khả năng biểu thị các vùng địa hình thay đổi đột ngột như các khe vực, hố lồi lõm và sông ngòi. Hạn chế này có thể gây sự nhầm lẫn trong khi đánh giá kết quả phân tích địa hình.

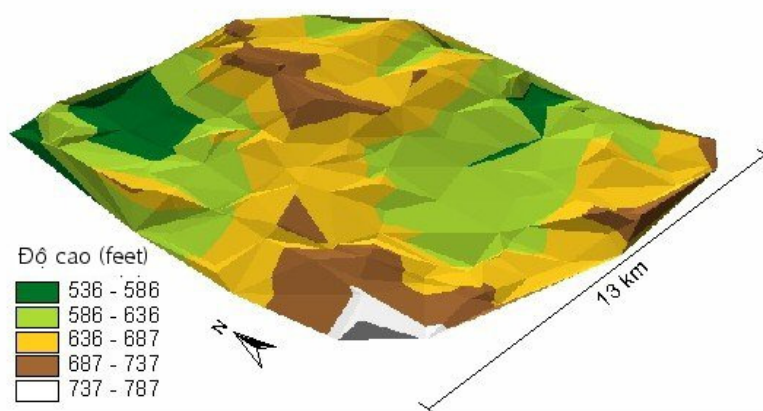
TIN được coi là phương pháp thuận tiện và kinh tế hơn. Mô hình TIN là thể hiện vector của cấu trúc địa hình, bao gồm các dây tam giác không đều không phủ lên nhau và bao trùm toàn bộ bề mặt địa hình, mỗi một tam giác xác định một mặt phẳng. TIN, theo khái niệm hình học là tập các đỉnh nối với nhau thành các tam giác. Mỗi một tam giác được giới hạn bởi 3 điểm đặc trưng về giá trị X,Y và Z (độ cao). Các tam giác này hình thành một bề mặt 3 phía, có độ dốc và hướng dốc. TIN có khả năng biểu diễn bề mặt liên tục từ tập điểm dữ liệu rời rạc và được coi như tập hợp các tam giác có các thuộc tính về độ dốc, diện tích và hướng. Hình 5.1 thể hiện cấu mô hình TIN và Hình 5.2 trình bày mô hình TIN trong thực tế khi thường phải thể hiện sự thay đổi kích thước lưới theo yêu cầu biến đổi của dữ liệu. Hình 5.3 là ví dụ về áp dụng TIN và kỹ thuật tô bóng để thể hiện địa hình độ cao một khu vực.



Hình 5.1 Ví dụ mạng TIN



Hình 5.2 Mạng TIN với sự thay đổi kích thước lưới đặc trưng



Hình 5.3 Ứng dụng TIN để biểu thị sự biến động độ cao địa hình

5.3 Phương pháp xây dựng DEM

Phương pháp chụp ảnh lập thể

Phương pháp này dùng một dụng cụ chụp ảnh chuyên dùng để chụp một số lượng lớn điểm mẫu với các giá trị X, Y, Z từ các ảnh lập thể hay viễn thám; sau đó các điểm được nội suy thành các ô vuông đồng nhất (grid). Phương pháp này tốn thời gian và đòi hỏi kỹ thuật chụp ảnh cao và số điểm kiểm soát phải nhiều nên ít khi được áp dụng.

Nội suy từ các đường đồng mức

Đây là phương pháp tiêu chuẩn để xây dựng DEM trong môi trường GIS. Đối với một khu vực, một số thông tin về địa hình có sẵn, việc xây dựng một DEM từ các đường đồng mức phải qua một số bước sau:

* Bước 1: Số hóa các đường đồng mức, có thể thực hiện qua một trong hai cách sau.

Số hóa tự động quét ảnh (scanning): chuyển các thông tin từ ảnh chụp hay bản đồ sang dạng tệp tin raster. Để có kết quả tốt, bản đồ đường đồng mức không nên kèm các thông tin khác. Sau đó bản đồ được chuyển sang dạng vector bằng các phần mềm chuyên dụng nhưng mỗi đường đồng mức phải được gán mã bằng tay. Nếu ảnh nguồn không rõ ràng thì phương pháp này tốn công hơn việc số hóa bằng bàn số hóa (digitizing).

Số hóa bằng thủ công: Dùng bàn số hóa để số hóa các đường đồng mức vẫn được coi là phương pháp tiêu chuẩn để xây dựng một DEM. Mỗi đường đồng mức được số hóa riêng lẻ và được gán mã thể hiện độ cao tương ứng.

*Bước 2: Raster hóa các đường đồng mức: được thực hiện bởi các chức năng rasterizing của các phần mềm chuyên dụng. Vấn đề quan trọng ở đây là việc chọn kích thước của các pixel mà các đường đồng mức chạy qua được tự động gán giá trị bằng độ cao của chính đường đồng mức đó.

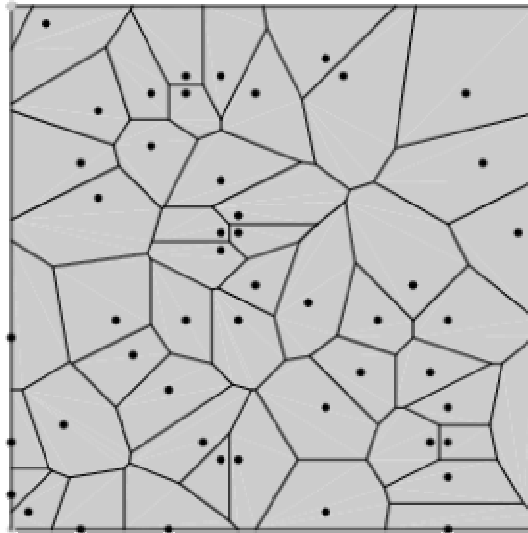
*Bước 3: Nội suy các đường đồng mức đã được raster hóa: Từ các đường bình độ chuẩn được raster hóa có thể nội suy ra các đường đồng mức khác, do vậy mỗi pixel trong bản đồ sẽ nhận giá trị cho điểm trung tâm của pixel.

*Bước 4. Xây dựng mô hình TIN (Hình 5.4), thường được thực hiện với sơ đồ Voronoi.

Sơ đồ Voronoi:

Giả sử trong một mạng điện thoại của thành phố, mỗi máy điện thoại sẽ được nối với một cột điện thoại gần nhất do vậy ta phải chia thành phố thành nhiều vùng, mỗi vùng có duy nhất một cột và khoảng cách từ mỗi vị trí trong vùng đến cột trong vùng đó là ngắn nhất. Kết quả của phân hoạch này là sơ đồ Voronoi (Hình 5.4).

Sơ đồ Voronoi có thể được tóm tắt như sau. Gọi $P = \{ p_1, p_2, \dots, p_n \}$ là tập hợp n điểm nằm trong mặt phẳng hai chiều. Ta chia (phân hoạch) mặt phẳng thành n đa giác sao cho bất kỳ điểm vị trí nào nằm trong một đa giác i đều có khoảng cách đến điểm i ngắn hơn khoảng cách từ nó đến các điểm vị trí p_k khác. Sơ đồ các đa giác này gọi là sơ đồ Voronoi $V(p_i)$ và được biểu diễn bằng ngôn ngữ toán học như sau:



Hình 5.4 Sơ đồ Voronoi

$$V(p_i) = \{x : |p_i - x| \leq |p_j - x|, \forall j \neq i\}$$

Sơ đồ Voronoi có rất nhiều ứng dụng trong hình học giải tích, hình học đồ họa và GIS:

- * Xác định vùng lân cận gần nhất (Nearest neighbor search) – Khi phải xác định vùng lân cận gần nhất của một điểm (vị trí) cho trước trong tổng số N điểm thì vùng đó chính là đa giác bao quanh điểm đó trong sơ đồ Voronoi.
- * Xác định vị trí phục vụ hợp lý (Facility location) – Ví dụ mạng lưới cửa hàng siêu thị muốn lập một cửa hàng mới và điều đầu tiên là xác định vị trí mới thích hợp. Vị trí mới này phải thỏa mãn yêu cầu ít ảnh hưởng nhất đến lượng khách hàng của các siêu thị đang vận hành hay nói cách khác là càng xa các siêu thị hiện có càng tốt. Người ta có thể sử dụng sơ đồ Voronoi bằng cách so sánh và phân tích tất cả các cạnh thẳng trong sơ đồ của vị trí các siêu thị hiện có.
- * Hình tròn rỗng lớn nhất (Largest empty circle) – Ví dụ ta cần tìm một vùng đất lớn chưa phát triển (dân cư và dịch vụ công cộng) để xây một nhà máy mới. Điều kiện là mảnh đất đó phải càng cách ly được tối đa các điểm dân cư hay công cộng. Đây là bài toán tương tự như trường hợp xác định vị trí hợp lý.
- * Quy hoạch đường (Path planning) – Khi các điểm vị trí trong sơ đồ là các trở ngại bất thuận lợi cho giao thông mà đường đi cần tránh xa thì các cạnh của đa giác trong sơ đồ Voronoi chính là các đoạn đường bảo đảm tránh được xa nhất các trở ngại.

Trong GIS, sơ đồ Voronoi được áp dụng để hình thành các chức năng biến đổi đối tượng raster sang vector nhờ kỹ thuật xây dựng mô hình TIN.

5.4 Các sản phẩm ứng dụng DEM

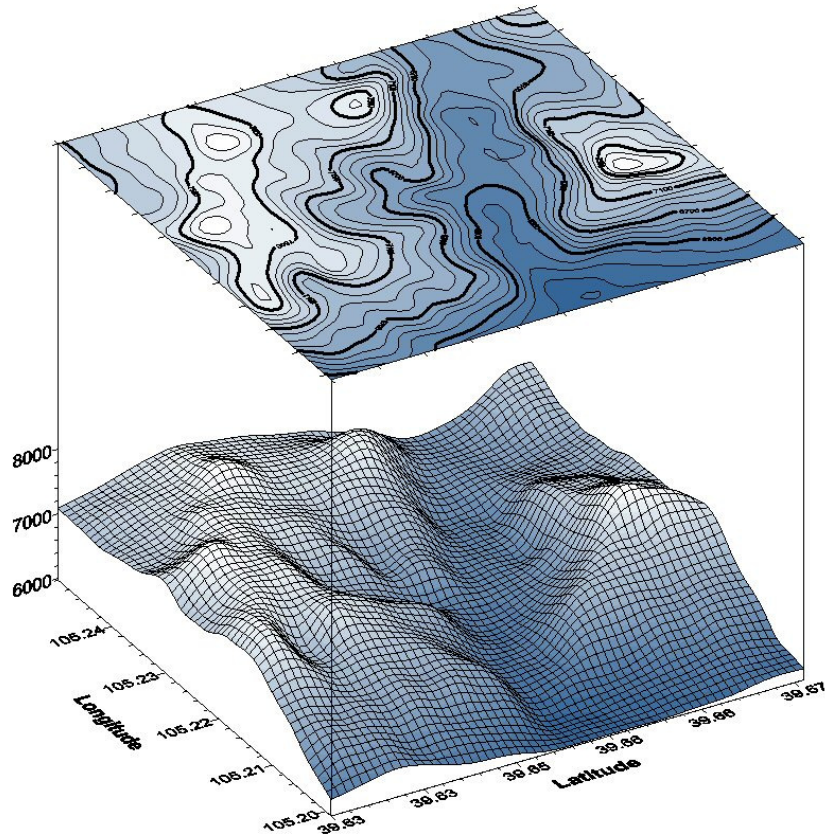
Kết quả các ứng dụng của DEM trong GIS có thể được tóm tắt như trong Bảng 5.2.

Bảng 5.2 Sản phẩm ứng dụng DEM trong GIS

- | |
|---|
| <ul style="list-style-type: none"> a) Biểu đồ khối, lát cắt dọc và ngang b) Tính toán thể tích các khối c) Bản đồ đường đồng mức d) Đường quan sát nhìn thấy e) Bản đồ độ dốc, độ lồi, độ lõm và hướng dốc f) Bản đồ địa hình tô bóng mặt khuất g) Xác định đường biên của lưu vực sông ngòi và vùng tiêu nước |
|---|

Biểu đồ khối

Biểu đồ khối là một trong các kết quả phổ biến của DEM. Nó cho phép xem xét trực quan ba chiều sự thay đổi trong không gian hai chiều của giá trị một thông số ta quan tâm. Thông số này không nhất thiết phải là độ cao địa hình. Hiện nay có rất nhiều phần mềm có khả năng tạo ra loại biểu đồ khối này từ tập hợp dữ liệu X, Y và Z. Ví dụ kết quả loại biểu đồ khối được trình bày trong Hình 5.5.



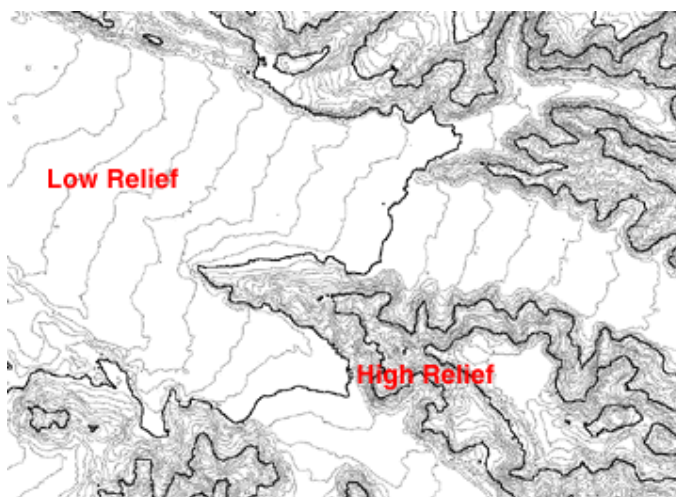
Hình 5.5 Biểu đồ khối biểu thị sự biến động của độ cao địa hình

Tính toán thể tích các khối

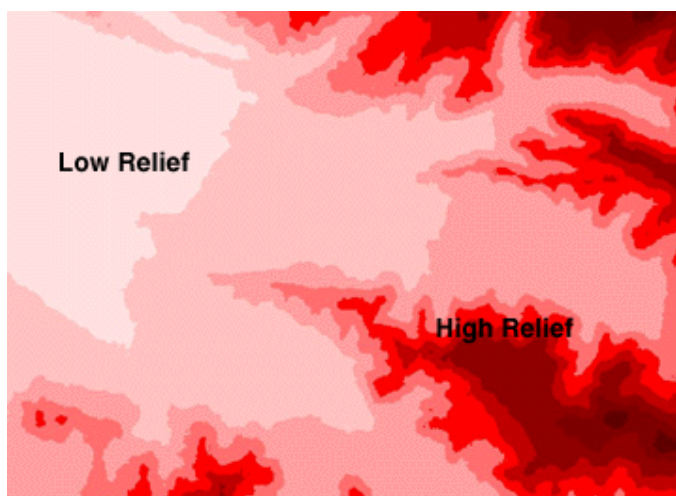
Trong thiết kế các công trình dân dụng, ví dụ tính toán san phẳng đất trong nông nghiệp và xây dựng, hay thiết kế đường giao thông, việc tính toán khối lượng đào đắp đất luôn là vấn đề rất quan trọng và cần thiết. Phương pháp phổ biến là xây dựng một mô hình DEM cho vùng đất nghiên cứu sau khi hoàn thành công việc đo đạc địa hình. Sau đó một mô hình DEM khác được thành lập để thể hiện tình trạng địa hình sau khi đã thực hiện công trình (sau khi đã san phẳng nếu là thiết kế san phẳng đất). Khối lượng đào đắp sẽ được tính toán dựa trên sự khác biệt giữa hai mô hình DEM.

Bản đồ đường đồng mức

Các đường đồng mức có thể dễ dàng thành lập bằng cách phân loại giá trị độ cao của các ô (cell) theo một thang độ độ cao nhất định và sau đó thể hiện các loại độ cao đó bằng các đường phân giới hay thể hiện bằng màu sắc. Bản đồ đường đồng mức thường được tạo ra bằng cách cắt giao tiếp mô hình DEM địa hình với các mặt phẳng ngang theo phân loại độ cao địa hình. Hình 5.6 cho ví dụ về bản đồ đường đồng mức địa hình và Hình 5.7 cũng là bản đồ địa hình nhưng thể hiện bằng mô hình DEM.



Hình 5.6 Bản đồ đường đồng mức với khoảng cách độ cao 5 m



Hình 5.7 Bản đồ địa hình ở Hình 5.6 nhưng sử dụng mô hình DEM

Đường quan sát nhìn thấy

Khả năng quan sát đóng một vai trò quan trọng trong các hoạt động quân sự, thông tin liên lạc sử dụng microwave và các nghiên cứu cảnh quan du lịch. Việc xác định tầm quan sát trên bản đồ giấy rất khó khăn do số lượng các lát cắt dọc cần xem xét rất lớn.

Tầm quan sát được xác định trên bản đồ số có mạng TIN bằng phương pháp truy vấn đường đi (tracking procedure). Phương pháp này là một biến đổi của thuật toán đường ẩn (hidden line algorithm).

Bản đồ độ dốc, độ lồi, độ lõm và hướng dốc

Trước khi mô hình độ cao DEM xuất hiện thì người ta sử dụng rất nhiều kỹ thuật để đánh giá độ dốc và độ lồi lõm của địa hình. Với DEM thì các công việc này trở nên nhanh chóng và thuận tiện, không cần nhiều công sức như trước đây.

Sau khi dữ liệu độ cao địa hình đã được chỉnh lý và thể hiện bằng mô hình TIN thì ta có thể sử dụng nhiều công cụ của phần mềm TIN để tính toán độ dốc, hướng dốc và độ lồi lõm của vùng nghiên cứu. Sau đây là những công thức tính toán chủ yếu.

Xác định hướng dốc

Góc của véc tơ tổng hợp giữa các vector D_X và D_Y sẽ cho chúng ta hướng dốc (Hình 5.7). Điểm cần lưu ý là phải tính toán giá trị + hay – trong bản đồ D_X và D_Y

$$\text{Hướng dốc } \% = 57,29579 \times \arctan\left(\frac{D_x}{D_y}\right)$$

$$\text{Góc dốc } \beta = 57,29579 \times \arctan\left(\frac{\text{dodoc } \%}{100}\right)$$

trong đó số 57,29579 là hệ số chuyển đổi từ gradient sang độ. Giá trị có thể tính từ -90^0 đến $+90^0$.

Tính toán độ dốc

Độ dài của vector (góc dốc) đã được xác định theo chiều X và Y có thể tính theo định lý Pythagore.

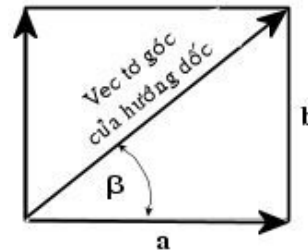
$$\text{Độ dốc } \% = \sqrt{\frac{(D_x^2 + D_y^2)}{k_c^2}} \times 100$$

trong đó là D_X gradient theo trục X; D_Y là gradient theo trục Y; và k_c là kích thước của pixel.

Để xác định sự sai khác về độ dốc theo từng mét, người ta chia độ dốc tính theo công thức trên cho kích thước của pixel.

Tính độ dốc theo độ:

$$\text{Góc dốc (độ)} = 57,29579 \times \arctan\sqrt{\frac{(D_x^2 + D_y^2)}{k_c^2}}$$



Hình 5.8 Vector hướng dốc

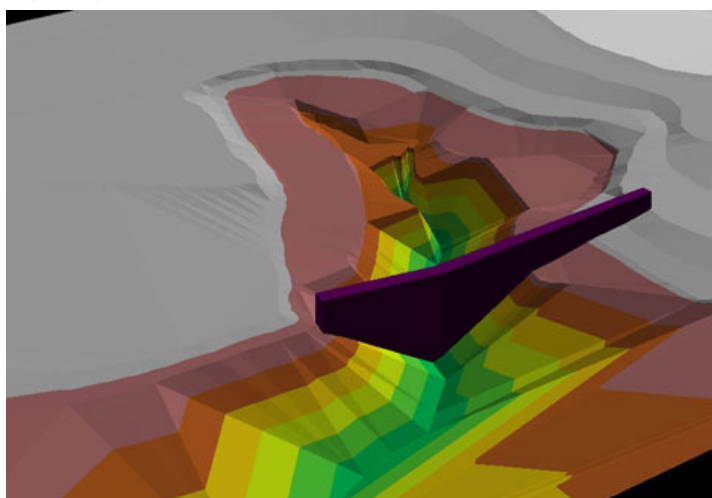
Bản đồ địa hình tô bóng mặt khuất

Các nhà vẽ bản đồ đã tạo ra nhiều phương pháp để cải tiến việc thể hiện trực quan của bản đồ, nhất là dạng địa hình của vùng đồi núi. Một trong những kỹ thuật thành công nhất là cách tô bóng địa hình được tạo ra chủ yếu bởi trường phái nghệ nhân bản đồ người Thụy sỹ và Áo. Kỹ thuật này có nguồn gốc ở thể loại nghệ thuật hội họa thời phục hưng, với việc đánh bóng và thể hiện ánh sáng để thể hiện hình ba chiều. Phương pháp thủ công này dùng bút vẽ để đánh bóng mặc dù thường tạo ra hình ảnh rất ấn tượng nhưng có giá thành rất cao và chủ yếu phụ thuộc vào kỹ năng và trí tưởng tượng của nghệ sỹ. Người ta cũng cho rằng các nghệ nhân bản đồ này phải là người sinh sống ở vùng núi.

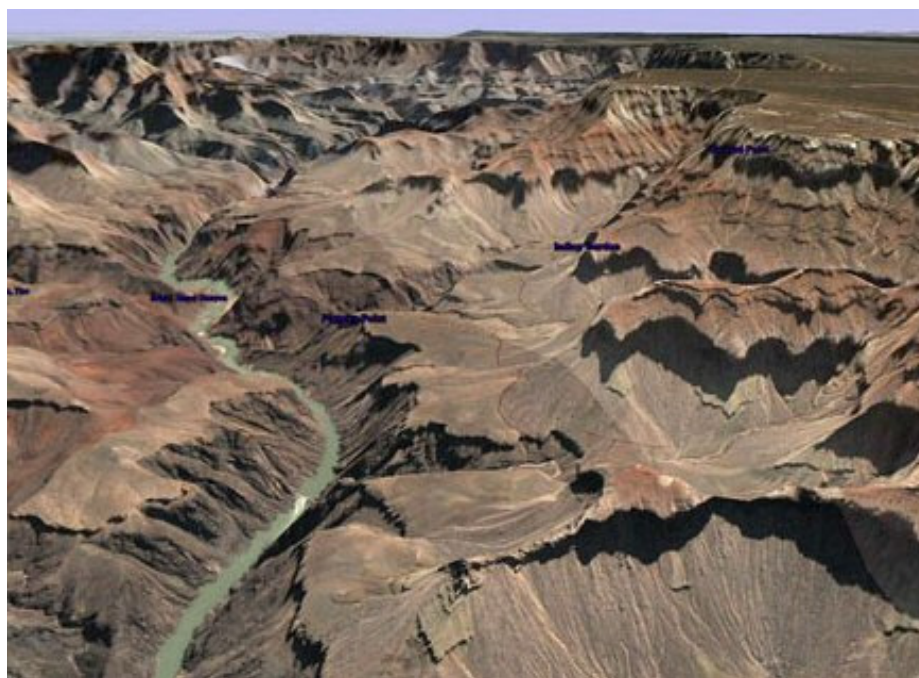
Từ khi xuất hiện bản đồ số, rất nhiều chuyên gia bản đồ đã nghĩ đến khả năng tạo bóng cho bản đồ địa hình một cách tự động, chính xác và dễ lặp lại. Nguyên tắc phương pháp tạo bóng địa hình là hình dung hình ảnh địa hình được chiếu sáng từ một vị trí nhất định. Kết quả tương tự một ảnh hàng không do việc sử dụng các gam màu sáng tối khác nhau. Tuy vậy bản đồ số địa hình có nhiều điểm khác biệt so với ảnh hàng không. Trước

hết bản đồ tô bóng này không biểu thị đúng hình ảnh mặt đất mà chỉ là của bề mặt đã số hóa của mặt đất. Thứ hai là nguồn ánh sáng tưởng tượng trong bản đồ tô bóng thường được chọn ở góc bằng hoặc lớn hơn 45° phía trên đường chân trời phía tây-bắc. Góc chiếu sáng này có tính chất nhân tạo mà rất khó có khả năng xảy ra trong thực tế thiên văn. Điểm khác biệt thứ ba nằm ở bản chất của mô hình độ cao DEM bởi mô hình này đã là sự đơn giản hóa bởi số lượng hạn chế số điểm dữ liệu và không thể thể hiện hết tất cả các chi tiết của địa hình thực tế.

Hình 5.9 và 5.10 cho ta thấy hai ứng dụng phổ biến của mô hình DEM. Hình 5.9 là một ví dụ về sử dụng kỹ thuật tô bóng cho bản đồ địa hình vùng hồ chứa có đập nước và Hình 5.8 là một ví dụ cho bản đồ địa hình vùng núi đá Garand Canyon, Hoa kỳ. Bạn đọc có thể tìm hiểu chi tiết về kỹ thuật tạo bóng cho bản đồ số địa hình với mô hình TIN trong Burrough (1987).



Hình 5.9 Ứng dụng DEM để biểu thị địa hình



Hình 5.10 Hình ảnh tô bóng của bản đồ địa hình vùng núi đá

Xác định đường dòng và đường biên của lưu vực sông ngòi và vùng tiêu nước

Khi các đường sông ngòi và tuyến tiêu nước không được số hóa bằng bàn số hoá, thì với việc sử dụng mô hình TIN, ta có một công cụ tự động xác định các đường dòng này. Phương pháp thủ công cổ truyền trên cơ sở xem xét tỉ mỉ các hình ảnh viễn thám hoặc bản đồ địa hình thường rất tốn công sức và rất dễ tạo sai số, nhất là đối với ảnh vệ tinh có chứa các mảng bị nhiễu.

Xác định đường dòng – Để vạch được đường dòng trên mô hình TIN, người ta phải xác định được vị trí các điểm lồi và lõm của địa hình. Phương pháp đơn giản nhất là xét lần lượt từng mảng bốn ô (cells) và đánh dấu ô có độ cao cao nhất và ô độ cao nhỏ nhất. Khi xem xét hết tất cả các cặp bốn ô của bản đồ thì nối liền các ô có đánh dấu để từ đó xác định đường dòng chảy hay đường tiêu nước. Những phương pháp khác cũng tương tự nhưng tăng số lượng ô xem xét đồng thời lên thành 3 x 3 hay cao hơn nữa.

Xác định đường biên của lưu vực sông ngòi – Lưu vực một con sông hay nhánh sông được hiểu là diện tích tập trung nước mưa của sông hay của nhánh sông đó. Trình tự logic xác định các đường biên lưu vực bắt đầu bằng việc tính độ dốc và hướng dốc cho từng ô (cell). Sau đó chương trình tìm kiếm các vùng thượng lưu của từng điểm trên dòng chảy sông. Thuật toán cũng tương tự như việc xác định đường dòng, bắt đầu từ điểm cuối hạ lưu và xem xét từng cặp 3 x 3 ô. Đối tượng xem xét ở đây là hướng dốc của từng ô chứ không chú ý đến độ dốc như trong trường hợp xác định đường dòng chảy. Chi tiết các thuật toán này được trình bày chi tiết hơn Burrough (1987).

Câu hỏi ôn tập

1. Nêu khái niệm về mô hình số độ cao(DEM) ?
2. Các ứng dụng của DEM vào thực tiễn?
3. Phân biệt giữa mô hình số độ cao GRID và TIN?
4. Khái niệm về sơ đồ Voronoi? Quan hệ giữa Voronoi và TIN?
5. Trình bày một số sản phẩm của DEM trong GIS?
6. DEM được ứng dụng như thế nào trong công tác xây dựng bản đồ địa hình?

Chương Sáu

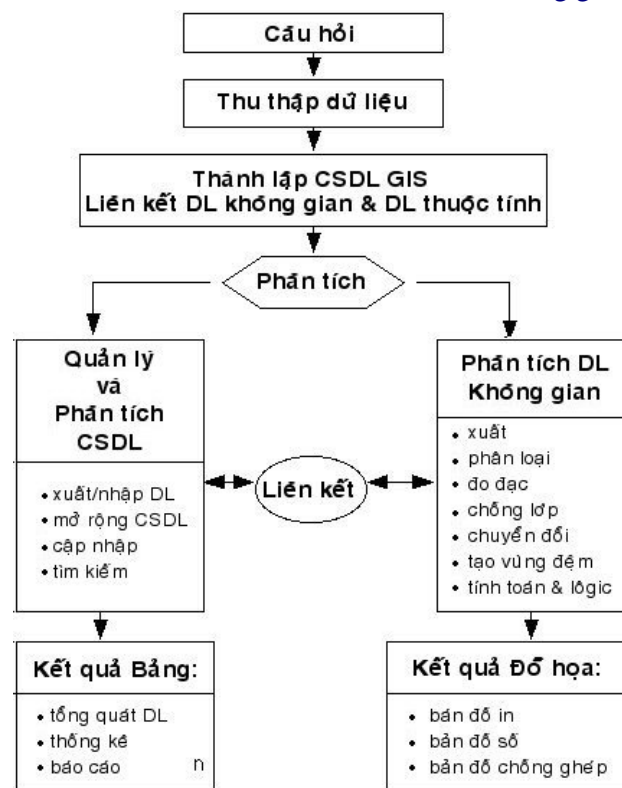
PHÂN TÍCH DỮ LIỆU TRONG GIS

Sự khác nhau giữa GIS và các phần mềm đồ họa là khả năng biến đổi các dữ liệu không gian gốc thành các câu trả lời cho các mục đích của người sử dụng, khả năng phân tích các dữ liệu không gian và phi không gian, khả năng tái hiện chuỗi từ cơ sở dữ liệu bằng các chức năng bất biến và đa biến của các phương pháp thống kê sử dụng các phương pháp nội và ngoại suy. Chương này sẽ trình bày khả năng giải quyết các vấn đề, và những yêu cầu cơ bản của cơ sở dữ liệu trong GIS. Cuối cùng, khả năng phân tích dữ liệu không gian và thuộc tính của GIS sẽ được giải thích, cơ bản bằng cách kết nối hai loại dữ liệu trong GIS, dữ liệu không gian và dữ liệu thuộc tính.

6.1 Mục đích và khả năng giải quyết vấn đề của các hệ thống GIS

Phân tích dữ liệu trong GIS bao gồm khâu chồng xếp bản đồ (map overlay) và phân tích các mối quan hệ không gian có trong cơ sở dữ liệu, là sự thể hiện chức năng phân tích địa lý và quản lý cơ sở dữ liệu.

Sơ đồ trong hình 6.1 diễn tả cấu trúc quá trình phân tích GIS để đạt được mục đích sử dụng. Trước hết ta cần xác định mục đích phân tích dựa trên yêu cầu của người sử dụng và các câu hỏi mà hệ thống GIS có thể trả lời. Dữ liệu sau khi được thu thập có thể được số hóa đối với dữ liệu không gian hay thành lập bảng số liệu đối với dữ liệu thuộc tính, kể cả công đoạn kết nối dữ liệu thuộc tính với tọa độ không gian.



Hình 6.1 Sơ đồ phân tích dữ liệu trong hệ thống GIS

Tất cả dữ liệu sau đó được nhập vào hệ thống kể cả các mối quan hệ để trở thành một hệ thống liên hoàn cả hai loại dữ liệu. Quá trình xử lý phân tích bao hàm cả dữ liệu không gian và dữ liệu thuộc tính sử dụng các liên kết GIS giữa chúng. Kết quả có thể phân thành hai loại, kết quả dạng bảng số (bảng tổng kết tóm lược dữ liệu, bảng kết quả thống kê và các báo cáo) và kết quả dạng đồ họa (bản đồ dạng giấy, bản đồ số dạng tệp tin và bản đồ kết quả của các phép chồng ghép).

Trước hết ta cần hiểu rõ các vấn đề mà hệ thống GIS có thể giải quyết. Mặc dù đã được đề cập đến trong chương một nhưng các khả năng giải quyết vấn đề của GIS vẫn sẽ được nhắc lại ở đây để làm bạn đọc dễ dàng theo dõi các chức năng phân tích ở phần tiếp sau của chương này. Các ứng dụng GIS được hình thành từ những câu hỏi thực tế đã được đặt ra đối với dữ liệu liên quan đến vị trí không gian. Những câu hỏi phổ biến là:

- Có những gì ở một vị trí địa lý nhất định (phân tích vị trí)?
- Khi nào thì một số điều kiện về vị trí địa lý được thỏa mãn?
- Các vật thể và quá trình đã thay đổi như thế nào theo thời gian?
- Các dữ liệu không gian có dạng như thế nào?
- Điều gì sẽ xảy ra nếu ta thay đổi một số điều kiện?

Phân tích vị trí

Dữ liệu trên bản đồ giấy có thể chỉ ra vị trí địa lý của vật thể nhưng không thể giải thích được tạo sao lại như vậy. Ví dụ một ảnh hàng không có thể cho ta biết ngô phát triển rất tốt ở một số khu vực trên cánh đồng nhưng ta không thể tìm được thông tin tại sao lại như vậy. Phép phân tích GIS có thể giúp ta tạo ra sự liên kết giữa tình trạng phát triển của ngô, loại đất, và chế độ nước bằng cách xem xét đồng thời bản đồ số về cây trồng, về phân bố loại đất và về độ ẩm đất. Hơn thế, GIS có khả năng phân tích những quan hệ phức hợp này một cách nhanh chóng và lưu trữ những dữ liệu quan tâm và kết quả phân tích.

Thỏa mãn điều kiện không gian

Người sử dụng rất hay muốn biết sàng lọc để tìm ra những dữ liệu địa lý thỏa mãn một số điều kiện nhất định. Ví dụ ta dự định bổ sung giếng khoan khai thác nước ngầm cho một điểm dân cư và muốn xác định vị trí của các giếng đó. Các giếng này cần phải nằm không xa quá 15 km so với điểm dân cư và phải ở gần sát đường ống cung cấp nước chính đã có sẵn. Như vậy ta cần biểu thị các yêu cầu đó thành ngôn ngữ địa lý và sau đó kết hợp với các điều kiện cục bộ để tìm ra vị trí các giếng khoan tương lai.

Phân tích biến đổi theo thời gian

Quan niệm của chúng ta về thế giới tự nhiên cũng như chính thế giới tự nhiên luôn thay đổi theo thời gian. Quá trình đô thị hóa làm phân bố sử dụng đất biến động liên tục suốt trong mấy thập kỷ gần đây. Những hiểu biết về thế giới tự nhiên của chúng ta nếu không cập nhật theo thời gian thì sẽ trở thành vô nghĩa. Ví dụ một nhà nghiên cứu quan tâm đến quan hệ giữa phân bố các loại sử dụng đất và luật phân vùng quy hoạch. Nếu anh ta có được lưu trữ của các dữ liệu này trong một khoảng thời gian dài thì hệ thống GIS có thể tìm ra quan hệ đó.

Phân bố không gian

Người sử dụng GIS thường muốn tìm hiểu về xu thế và dạng phân bố của dữ liệu trong không gian. Nói cách khác là liệu hai hay nhiều vật thể hay thông số thay đổi có

tương tự với nhau không trong không gian? Ví dụ nếu ta thấy số lượng tai nạn giao thông thống kê trên đường cao tốc với tốc độ xe 65 dặm (mile)/ giờ có quan hệ gì với số tai nạn trên đường có xe chạy với vận tốc 55 mile/ giờ.

Đánh giá hậu quả các phương án

Việc xây dựng các phương án (hay kịch bản) là kết quả của câu hỏi loại “Điều gì sẽ xảy ra nếu ...?”. Ví dụ tai họa gì sẽ xảy ra với các vùng ven biển nếu nhiệt độ toàn cầu tăng lên, băng ở bắc và nam cực tan ra, và mực nước biển dâng cao? Ở đây, người sử dụng phải kết hợp với các mô hình mô phỏng để dự báo hậu quả và thành lập bản đồ tác động của nhiệt độ tăng lên các vùng ven biển.

Ở ví dụ trên hậu quả phụ thuộc vào mực nước biển sẽ dâng lên cao bao nhiêu. Mực nước biển lại phụ thuộc vào lượng băng sẽ tan và sự biến đổi của chế độ mưa và bốc hơi nước mà hai hiện tượng sau cũng bị tác động bởi biến động của nhiệt độ. Nếu các phân tích dựa trên các giả thiết khác nhau về biến đổi của hiện tượng tan băng, chế độ mưa và bốc hơi nước như là hàm số của sự thay đổi nhiệt độ toàn cầu, thì người ta có thể dự báo các mức tác động khác nhau lên môi trường các vùng ven biển.

Tuy vậy, trước khi sử dụng các công cụ phân tích để trả lời các câu hỏi do dự án sử dụng GIS đặt ra, ta cần xử lý dữ liệu không gian và dữ liệu thuộc tính để chúng tương thích với hệ thống GIS của dự án. Các chức năng xử lý dữ liệu này thường là một bộ phận quan trọng của bất kỳ hệ thống GIS nào.

6.2 Chức năng quản lý cơ sở dữ liệu

Các khả năng xử lý dữ liệu (data manipulation) không chỉ bao gồm các chức năng thông thường của một hệ thống quản lý dữ liệu (DBMS) nhằm quản lý và tìm kiếm dữ liệu có trong cơ sở dữ liệu (cho cả dữ liệu thuộc tính và dữ liệu không gian), mà còn cả khả năng nhập và chuyển đổi dữ liệu từ những dạng chuẩn khác nhau hay từ những nguồn khác nhau, khả năng chỉnh lý dữ liệu và xóa bỏ dữ liệu không còn có tác dụng. Bình thường thì dữ liệu thu thập được là từ các nguồn khác nhau và chúng không bao giờ tương thích ngay lập tức về dạng lưu trữ để có thể nhập ngay vào hệ thống GIS.

Một hệ thống GIS cần phải cho phép người quản trị thực hiện được các hoạt động: (1) nhập dữ liệu từ các nguồn bên ngoài GIS, (2) dễ dàng cập nhật và sửa đổi dữ liệu, và (3) truy cập để trả lời các câu hỏi tìm kiếm (query) dữ liệu. Phần mềm chuyên dụng về hệ thống quản lý CSDL (DBMS) thường là một thành phần quan trọng của bất cứ hệ thống GIS nào và sẽ cung cấp các công cụ để thực hiện các công việc đã đề cập ở trên. Người sử dụng cũng có thể dùng phần mềm về quản lý CSDL bên ngoài hệ thống GIS. Sau đây là cụ thể các động tác với CSDL khi sử dụng phần mềm quản lý CSDL thông dụng.

6.2.1 Nhập và mở rộng CSDL

Dữ liệu nhập vào hệ thống GIS, nhất là dữ liệu thuộc tính, thường ở dạng tệp tin tiêu chuẩn ASCII (American Standard Code for Information Interchange). ASCII là tiêu chuẩn dạng dữ liệu mà hầu hết các hệ thống máy tính đều có thể đọc được. Một hệ GIS có thể có khả năng nhập được các tệp tin dạng khác.

Phương pháp nhập dữ liệu khác là mở rộng CSDL bằng cách sử dụng bàn phím. Bất kỳ CSDL nào cũng có thể tạo ra một lượng dữ liệu có hạn bằng phương pháp này (khoảng vài trang giấy số liệu).

Hệ quản lý CSDL cũng cho phép kiểm tra sai số tự động khi dữ liệu mới được nhập vào hay dữ liệu cũ được chỉnh sửa. Tuy nhiên không phải tất cả mọi sai số đều có thể

được phát hiện tự động nên người sử dụng rất cần thiết phải theo dõi sát quá trình nhập và chỉnh sửa dữ liệu.

Hệ thống GIS cũng phải có công cụ để xuất dữ liệu sang các dạng khác nhau. Trong quá trình đó dữ liệu được ghi lại dưới dạng lưu trữ chung nhất (ví dụ ASCII) vào tệp tin để sau đó có thể được nhập vào các hệ máy tính hay hệ phần mềm khác.

6.2.2 Cập nhập dữ liệu thuộc tính

Công đoạn phổ biến nữa là cập nhập và chỉnh sửa dữ liệu thuộc tính khi cần thiết. Bởi vì không ai có thể hình dung được trong tương lai sẽ cần dữ liệu gì nên hệ thống GIS phải được trang bị công cụ thân thiện nhằm chỉnh sửa và cập nhập dữ liệu. Dữ liệu thuộc tính là loại dữ liệu rất hay thay đổi giá trị nên tính thời sự của cơ sở dữ liệu phụ thuộc rất nhiều vào khả năng cập nhập và chỉnh sửa của nó.

6.2.3 Tìm kiếm trên CSDL

Sử dụng cơ sở dữ liệu để trả lời những câu hỏi liên quan đến dữ liệu thường được thực hiện thông qua quá trình phân tích cơ sở dữ liệu. Quá trình tìm kiếm này cho ra kết quả ở dạng bảng số. Những kết quả bảng này có thể là một phần của các báo cáo tóm tắt hoặc chúng được nhập vào các phần mềm khác (như Microsoft Excel) để tiếp tục được xử lý.

Để thực hiện việc tìm kiếm (query), ta cần thành lập các biểu thức logic biểu thị các điều kiện. Những điều kiện này biểu thị những vật thể địa lý nào cần phân tích và chúng sẽ được phân tích ra sao. Kết quả sẽ là một phần của cơ sở dữ liệu. Một số biểu thức logic có thể rất đơn giản với chỉ một điều kiện nhưng nhiều biểu thức rất phức tạp, thường thì bao gồm rất nhiều điều kiện.

Ví dụ ta có một CSDL gồm vị trí, kích thước và các đặc điểm khác (như tên chủ nhân) của các hồ nước trong một quốc gia và người sử dụng đặt câu hỏi đơn giản “Trong tỉnh X có những hồ nước nào?”. Để trả lời câu hỏi đó, hệ thống quản lý CSDL sẽ tạo nên một bộ dữ liệu con và biểu thị nó dưới dạng bảng (hàng và cột). Bộ dữ liệu con này chỉ cần thỏa mãn một điều kiện duy nhất: các hồ ở trong địa giới tỉnh X. Phức tạp hơn, nếu câu hỏi của người sử dụng là “Có những hồ nước nào trong tỉnh X thuộc sở hữu cá nhân và có diện tích mặt hồ lớn hơn 50 ha?” thì đây là tìm kiếm với nhiều điều kiện. Bộ dữ liệu kết quả gồm các hồ phải thỏa mãn các điều kiện về địa giới, về chủ nhân và về diện tích mặt hồ.

Có những câu hỏi bắt buộc hệ thống CSDL phải sử dụng cả thuật toán logic và phân tích toán học để giải quyết như tính toán mật độ dân số, tính toán tổng kết thống kê, và chuyển đổi hệ thống đo lường.

6.3 Các chức năng phân tích dữ liệu không gian

Các phần mềm hệ thống CSDL và phần mềm GIS đều được trang bị công cụ để tiến hành phân tích CSDL nhưng phần mềm GIS khác biệt ở khả năng phân tích bản đồ (dữ liệu không gian). Ta sẽ dễ hình dung hơn khi tưởng tượng phân tích dữ liệu không gian như việc sắp đặt các lớp bản đồ GIS. Các lớp này tương tự như những bản đồ giấy có nền trong suốt mà ta có thể chồng xếp lên nhau. Thường thì một lớp bản đồ GIS chỉ chứa một loại dữ liệu (gọi là theme). Khi chưa có máy tính thì người ta phải thực hiện việc phân tích thủ công với các bản đồ có nền trong suốt.

Nếu giải quyết một vấn đề dữ liệu có tính không gian, người sử dụng phải xác định hiện tượng cần nghiên cứu và trình tự tiến hành phân tích dữ liệu. Hệ thống GIS cung

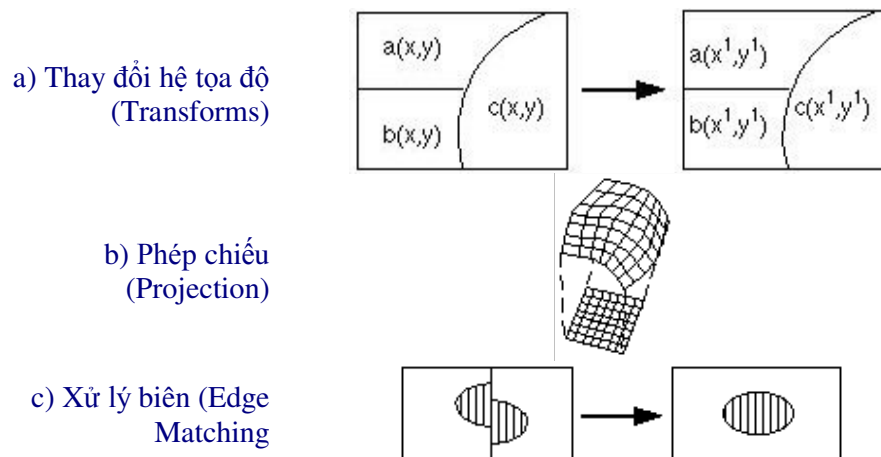
cấp một tập hợp các công cụ "tools" hay còn gọi là các chương trình máy tính cho phép người sử dụng thực hiện một loạt các động tác xử lý đối với bản đồ số và các dữ liệu thuộc tính. Các động tác phân tích này bao gồm các tìm kiếm không gian, các xử lý dữ liệu và phân tích không gian như sẽ được trình bày tiếp theo.

Các chức năng phân tích và xử lý dữ liệu thuộc tính trong GIS sử dụng các biểu thức logic cũng tương tự như các công cụ đối với cơ sở dữ liệu thông thường như đã diễn tả ở phần đầu chương. Khi phân tích dữ liệu không gian (bản đồ), trừ chức năng chồng ghép thì các hoạt động khác có thể được thực hiện trên một lớp dữ liệu duy nhất hay trên hai hay nhiều lớp dữ liệu.

Các công cụ phân tích dữ liệu không gian trong GIS có thể được phân thành bảy loại: (1) thay đổi hệ tọa độ và phép chiếu, (2) Phép nội suy (3) truy xét, phân loại và đo đạc, (4) chồng bản đồ theo logic và trực quan, (5) tạo vùng đệm và nối mạng lưới, (6) phép tính toán đại số, và (7) biểu thị kết quả.

Bạn đọc cần chú ý là tên gọi các chức năng này có thể khác nhau ở các phần mềm GIS khác nhau nhưng thực chất của chúng đều tương tự như nhau. Các chức năng hay công cụ này có thể phải kết hợp theo một tuần tự nhất định nào đó thì mới hoàn thành được một mục đích đã được đặt ra. Trong chương này thì các chức năng được thể hiện qua Hình 6.2 đến Hình 6.6.

6.3.1 Thay đổi hệ tọa độ và phép chiếu



Hình 6.2 Thay đổi hệ tọa độ và phép chiếu

Sơ đồ ở Hình 6.2 diễn tả chức năng thay đổi hệ tọa độ, phép chiếu và xử lý biên. Chức năng thay đổi hệ tọa độ, phép chiếu cho phép người sử dụng GIS loại trừ những bất hợp lý giữa các lớp dữ liệu. Dữ liệu bản đồ có thể được biến đổi từ hệ tọa độ và phép chiếu này sang hệ tọa độ và phép chiếu khác.

Những động tác này thuộc loại xử lý chuẩn bị để loại trừ những sai số tọa độ, phép chiếu không đồng bộ khi dữ liệu được thu thập, nhất là sử dụng các nguồn khác nhau. Hoạt động phổ biến là đăng ký bản đồ với mục đích đưa các lớp bản đồ khác nhau lên cùng một hệ tọa độ hay hệ tham chiếu để chúng cùng khớp nhau trên diện tích thực tế. Dữ liệu khi thu thập có thể chứa đựng các đại lượng đo lường khác nhau, phổ biến nhất là có các hệ tọa độ khác nhau (như hệ kinh vĩ tuyến, hệ UTM, Universal Transverse Mercator). Một hệ thống GIS phải có khả năng xử lý những khác biệt loại này như diễn tả ở Hình 6.2a.

Hiện nay trên thế giới người ta sử dụng nhiều hệ phép chiếu khác nhau nhằm thể hiện bề mặt ba chiều của trái đất thành mặt phẳng hai chiều. Thông thường người sử dụng GIS không thể đặt yêu cầu về hệ tham chiếu khi thu thập dữ liệu. Nhiều bản đồ có thể sử dụng hệ chiếu Mercator hình trụ, bản đồ khác sử dụng hệ Albers hình nón, và nhiều bản đồ sử dụng hệ chiếu phẳng. Chức năng thay đổi phép chiếu (Hình 6.2b) cũng rất phổ biến trong các phần mềm GIS.

Một vấn đề cũng rất phổ biến là khi ghép các mảnh bản đồ. Phần lớn các vật thể chạy qua đường biên của hai mảnh tiếp giáp nhau phần lớn cần phải chỉnh lý về vị trí và nhiều khi cả về hình dạng (Hình 6.2c). Người ta thường phân bố sai số khép này đồng đều về hai phía của đường giáp biên. Ví dụ ta ghép nhiều mảnh bản đồ có chứa đường giao thông ta sẽ phải thực hiện động tác chỉnh lý các đoạn đường đi qua đường giáp biên để làm cho các đường giao thông trở nên liên tục.

6.3.2 Phép nội suy

Trong trường hợp nguồn dữ liệu khuyết thiếu một số điểm, đường hay vùng thì cần phải thực hiện hình thức nội suy (giải đoán) để tạo thêm dữ liệu. Có nghĩa là dữ liệu của một hay nhiều điểm trong không gian xung quanh được sử dụng để tìm ra các giá trị mới cho các điểm khuyết thiếu mà trước đó không thể đo đạc hay quan trắc được. Có nhiều phương pháp nội suy, ví dụ;

- Nội suy giá trị gần nhất: giá trị của dữ liệu điểm cần tìm thông qua giá trị điểm chuẩn (được đo trực tiếp) gần nhất;
- Nội suy tuyến tính qua đường thẳng nối 2 điểm;
- Nội suy dùng đa thức (spline) dựa trên 3 hay nhiều điểm cho trước;
- Nội suy stochastic trên cơ sở tạo số ngẫu nhiên và các tham số fractal với một số điểm; và
- Nội suy trên cơ sở mô hình.

Phương pháp phổ biến nhất là nội suy điểm dựa trên giá trị trung bình của các điểm láng giềng. Giả sử chọn n điểm láng giềng hay trong vòng tròn có bán kính cho trước, giá trị e của điểm cần nội suy sẽ là:

$$e = \frac{\sum_{i=1}^n (S_i / d_i)}{\sum_{i=1}^n (1 / d_i)}$$

trong đó S_i là giá trị điểm mẫu cách điểm nội suy một khoảng d_i .

$d_i = 0$ nếu vị trí điểm nội suy trùng với điểm mẫu. Biểu thức tổng quát để tính giá trị e sẽ là:

$$e = \sum_{i=1}^n (W_i / S_i)$$

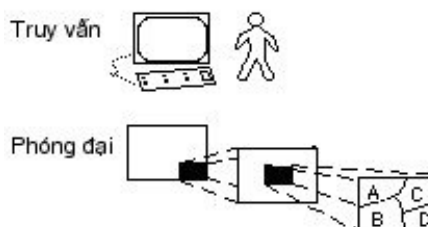
trong đó W_i là trọng lực áp dụng cho từng điểm mẫu, thông thường nó tỉ lệ nghịch với khoảng cách từ điểm mẫu S_i tới điểm đánh giá. Tổng của W_i luôn có giá trị bằng 1.

Trong thực tế nội suy hay được áp dụng cho mô hình hóa bề mặt khi cần phải giải đoán các giá trị mới cho bề mặt 2 chiều dựa trên các điểm độ cao lân cận.

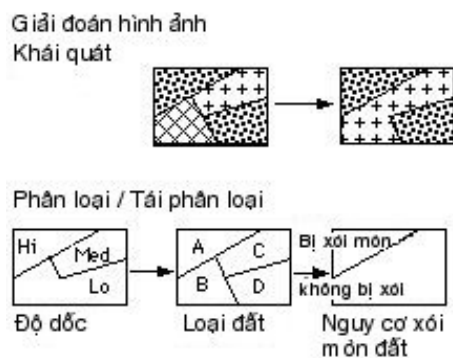
6.3.3 Truy vấn dữ liệu, phân loại và phép đo dữ liệu không gian

Truy vấn dữ liệu, phân loại và đo đạc là ba chức năng riêng biệt nhưng rất hay được sử dụng kết hợp với nhau. Ví dụ một người sử dụng GIS quan tâm đến vấn đề thu nhập dân cư mong có được bản đồ riêng của từng khu vực trong đó chia rõ các vùng nhỏ với tổng thu nhập các hộ gia đình tương đương nhau hay ở trong một giới hạn cho trước.

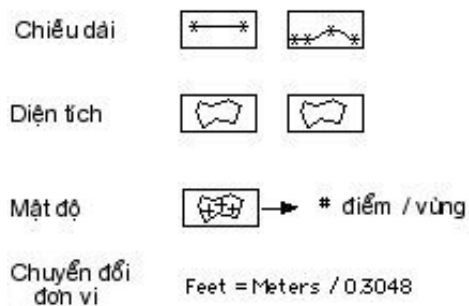
a) Truy vấn không gian (Spatial Retrieval)



b) Phân loại (Deliniation and Classification)



c) Đo đạc (Measurement)



Hình 6.3 Truy cập phân loại và đo đạc

Ở phần các chức năng của CSDL ta thấy ví dụ tìm các hồ nước theo tiêu chuẩn chủ sở hữu và diện tích mặt hồ và kết quả được biểu thị theo dạng bảng số. Trong hệ GIS, dữ liệu còn có thể truy cập theo tiêu chuẩn vị trí của chúng và theo các quan hệ không gian với các vật thể địa lý khác. Quá trình này được gọi là truy vấn không gian (Hình 6.3a). Ví dụ, các hồ nước có diện tích từ 0 đến 1 ha có thể được tô một màu, từ 1 đến 5 ha tô màu khác, v.v., và kết quả ta có một bản đồ màu biểu thị các hồ nước ứng với độ lớn diện tích khác nhau. Phương cách biểu thị theo không gian như vậy rõ ràng có hiệu quả hơn nhiều so với cách biểu thị bằng bảng biểu nếu chỉ sử dụng CSDL.

Khi bản đồ được biểu thị trên màn hình máy tính thì ta có thể phóng đại một vùng nào đó để xem xét chi tiết (Hình 6.3b). Động tác này có thể được thực hiện với con chuột hay với việc đặt điều kiện chỉ rõ tọa độ giới hạn vùng xem xét. Kết quả các vật

thể (điểm, đường và đa giác) nằm trong vùng xem xét có thể được đặt vào một lớp bản đồ mới để tăng tốc độ xử lý phân tích tiếp theo.

Phân loại không gian đòi hỏi phải có nhiều lớp dữ liệu và những dữ liệu thuộc tính không gian này sẽ được kết hợp, chồng ghép theo một trình tự đến khi đạt được kết quả mong muốn (Hình 6.3b). Ví dụ, việc xác định các vùng dân cư với mức sống nghèo khổ khác nhau có thể được xác định trên cơ sở dữ liệu về tỷ lệ thất nghiệp, mức trợ cấp của chính phủ, và dữ liệu thống kê về thu nhập. Việc kết hợp những dữ liệu này sẽ cho biết xu thế phân bố tổng hợp và có thể vạch ra những đường đồng đẳng về vấn đề mà ta quan tâm. Những dạng phân bố kết quả có thể được tiếp tục phân tích nếu thấy cần thiết (ví dụ, xác định những vùng ưu tiên thiết lập văn phòng cho các dịch vụ xã hội).

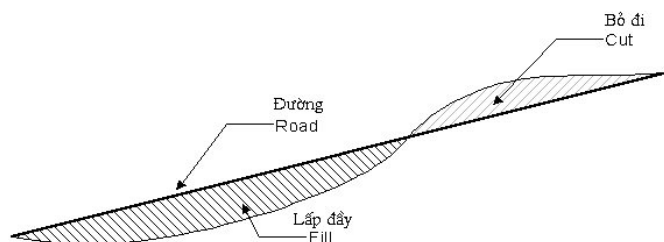
Công việc xác định chiều dài của đường giao thông, sông ngòi, diện tích của khu dân cư, mật độ dân số, v.v có thể thực hiện được với chức năng đo đạc của hệ thống GIS (Hình 6.3c). Một hệ thống GIS cho phép thực hiện việc đo đạc đơn giản và phức hợp. Khả năng đo đạc tự động giúp cho hiệu quả các công việc loại này tăng lên rất nhiều.

Với chức năng đo đạc, hệ thống GIS có cung cấp khả năng chuyển đổi đơn vị đo, ví dụ chiều dài có thể chuyển đổi từ đơn vị Anh sang hệ mét và ngược lại. Chiều cao và chiều dài sườn dốc thường là kết quả khi phân tích các quan hệ ba chiều trong mô hình độ cao. Khoảng cách mia, góc đo so với phương bắc và các số đo trắc địa có thể được biến đổi thành vị trí, khoảng cách và góc đo địa hình để thành lập bản đồ địa hình.

Người sử dụng GIS nhiều khi muốn biết số lần xảy ra của một hiện tượng. Ví dụ trong việc nghiên cứu ô nhiễm nước ngầm họ muốn xác định vị trí của các giếng khoan có thể bị ô nhiễm và có thể muốn tìm các khu vực có khả năng bị ô nhiễm. Trong trường hợp này ta phải có cả bản đồ chứa đựng dữ liệu về mức độ mẫu nước bị ô nhiễm trong các giếng (theo không gian) và số lần lấy các mẫu nước đó (tần suất theo thời gian).

Trong các nghiên cứu khác, người ta mong muốn tìm ra sự phân loại một thông số theo không gian dựa trên dữ liệu không gian và dữ liệu thuộc tính. Khi đó dữ liệu trong vùng nghiên cứu phải được xếp loại, kết hợp với một thang phân loại để xác định các đường đồng đẳng chia diện tích thành các vùng có giá trị thông số đồng nhất. Việc phân tích quan điểm chính trị của cộng đồng là một ví dụ loại này, trong đó dữ liệu không gian (vị trí điểm dân cư và dân số) có thể kết hợp với dữ liệu phi không gian (quan điểm chính trị) để lập bản đồ quan điểm chính trị và tính toán diện tích tương ứng.

Chức năng đo đạc khác nữa liên quan đến việc xác định kích thước theo không gian ba chiều. Thao tác 'đào và đắp' được biết tới trong ngành xây dựng thuộc loại thao tác này. Ví dụ việc tính toán tổng khối lượng đất đắp đi (đào) hoặc lấp đầy vào chỗ trống (đắp) để xây dựng đường giao thông (Hình 6.4).

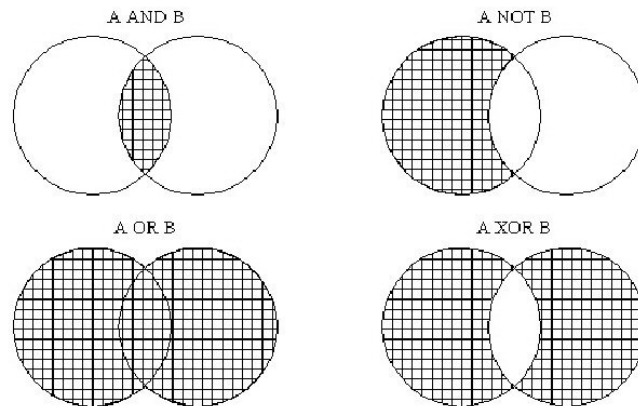


Hình 6.4 Thao tác đào và đắp

Trong ba chức năng truy vấn, phân loại và đo đạc của hệ thống GIS như thể hiện tóm tắt ở sơ đồ Hình 6.3 thì chức năng truy vấn hay còn gọi là “hỏi đáp tìm kiếm” rất hay được sử dụng thường xuyên. Ta nên chú ý là sự hỏi đáp trên dữ liệu chuyên đề có thể được thực hiện trên một hoặc nhiều tính chất của các đối tượng.

Quá trình đặt câu hỏi là quá trình lựa chọn thông tin từ tập hợp dữ liệu dựa trên những điều kiện được định rõ. Trường hợp lựa chọn một tính chất (điều kiện đơn) phép tính đại số được sử dụng để cấu thành sự lựa chọn bao gồm tập hợp phép tính đại số bằng, lớn hơn, nhỏ hơn và tổ hợp của 3 phép đó ($=$, $>$, $<$, $<>$, $\geq$, $\leq$). Ví dụ lựa chọn tất cả những thửa đất chuyên dùng có diện tích lớn hơn hoặc bằng 1000 m² để có thể sử dụng vào việc xây dựng một trạm y tế.

Khi nhiều điều kiện đơn được tổ hợp lại để tạo thành những điều kiện phức tạp (lựa chọn nhiều hơn 1 tính chất của đối tượng) thì người ta cần sử dụng toán tử Boolean (AND, OR, XOR, NOT) để thiết lập sự lựa chọn phức tạp. Kiểu đặt vấn đề này có thể được minh họa trên sơ đồ Venn (xem hình 6.5). Những vùng đánh bóng diễn biến kết quả lựa chọn (Bernhardsen 1999).

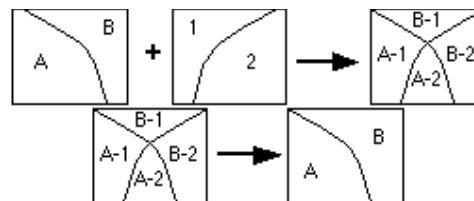


Hình 6.5 Sơ đồ Venn

6.3.4 Chức năng chồng ghép và phép logic

a) Chồng bản đồ
(Overlay)

b) Ghép bản đồ
(Dissolve)



Hình 6.6 Chồng ghép bản đồ

Chức năng chồng các lớp bản đồ cho phép người sử dụng đặt các lớp dữ liệu lên nhau trên cơ sở các quan hệ không gian. Ghép bản đồ tạo ra các loại dữ liệu tổng quát hơn trong bản đồ (Hình 6.6).

Phép chồng ghép và khả năng đánh giá các quan hệ không gian có thể là chức năng được biết đến nhiều nhất của các hệ thống GIS. Quan hệ giữa các lớp dữ liệu có thể được truy vấn thông qua các biểu thức toán học (lô gic) hoặc/và bằng trực quan hóa hình.

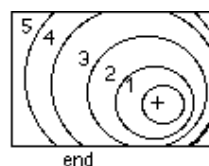
Phép chồng ghép bản đồ sử dụng các biểu thức logic hoặc các hàm không gian và tích trữ kết quả trong CSDL GIS như là các lớp dữ liệu mới (Hình 6.6a). Do các lớp đã được biến đổi đồng nhất về tọa độ và hệ tham chiếu nên chúng có thể được chồng khớp lên nhau về mặt kích thước không gian. Ví dụ các đường sông ngòi (lines) và vị trí đã quan sát thấy động vật quý hiếm như hươu nai (points) có thể được tích hợp với dữ liệu về diện tích rừng (polygons) để tạo nên một CSDL mới diễn tả quan hệ giữa ba yếu tố đó thành bản đồ khu vực sống của động vật.

Chức năng ghép bản đồ cho phép người sử dụng biến một lớp dữ liệu phức tạp thành một lớp dữ liệu mới trong đó các đường phân cách hai vùng có cùng một giá trị thông số bị loại trừ (dissolve) như diễn tả trong hình 6.6b). Kết quả là một lớp có giá trị dữ liệu tổng quát hơn. Chức năng này thực chất là chức năng ngược của phép chồng xếp bản đồ. Ví dụ việc ghép dữ liệu không gian của một bản đồ sở hữu về các loại chủ nhân của các vùng đất bao gồm chủ nhân là chính quyền địa phương, chính quyền trung ương, các xí nghiệp công nghiệp tư nhân, các vùng đất sở hữu tư nhân, sẽ dẫn đến việc thành lập quan hệ không gian giữa sở hữu công cộng và sở hữu tư nhân.

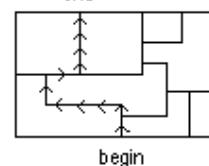
Phép logic là việc sử dụng các lệnh logic để tạo ra các lớp dữ liệu mới và chọn ra các đặc tính địa lý mới từ bảng thuộc tính. Sự lựa chọn logic dựa trên các biểu thức luận lý hoặc các giá trị thuộc tính. Các giá trị thuộc tính được lựa chọn sẽ được chọn, sát nhập hoặc loại bỏ để tạo ra lớp dữ liệu mới

6.3.5 Tạo vùng đệm và nối mạng

a) Tạo vùng đệm
(Buffering)



b) Nối mạng
(Networks)



Hình 6.7 Tạo vùng đệm và chức năng dùng lưới

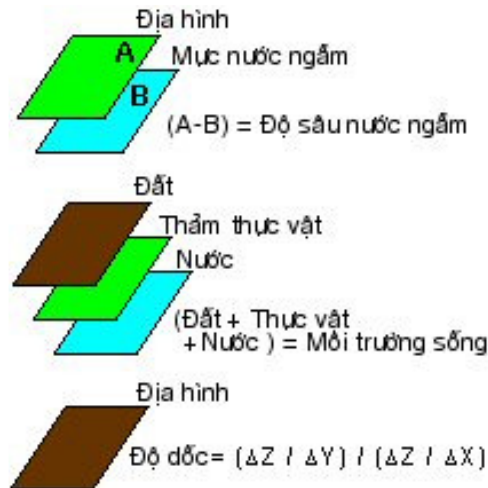
Chức năng tạo vùng đệm nhằm xác định vùng xung quanh, sát với một vật thể ta quan tâm trên bản đồ và chức năng nối mạng là để xác định đường chuyển động trong một vùng. Ví dụ về tạo vùng đệm là tạo vùng xung quanh một vật thể mà khoảng cách mọi điểm trong vùng đó không cách xa vật thể một khoảng cách nhất định (diện tích cách đường quốc lộ 10 km). Chức năng nối mạng xem xét chuyển động của một vật thể theo hướng của một vật thể khác trên bản đồ (ví dụ hướng chuyển động chính của hoạt động giao thông trên một đường cao tốc).

Hai loại phân tích dữ liệu không gian này xem xét vùng phụ cận của một vật thể hay vấn đề nối các vật thể điểm trong bản đồ (Hình 6.7). Ví dụ ta có hai lớp dữ liệu không gian, một biểu thị một con đường và dữ liệu kia về khu rừng gồm cả đường kính trung bình các cây lấy gỗ. Ta giả thiết người sử dụng muốn khả năng khai thác gỗ (cây gỗ với đường kính lớn hơn một giá trị nhất định) trong vòng 150 mét khoảng cách từ con đường. Phép tạo vùng đệm của GIS có thể được áp dụng để tạo ra vùng rừng ở phía hai bên con đường. Sau đó các dữ liệu thuộc tính về cây gỗ có thể được kết hợp phân tích để tạo ra một lớp dữ liệu mới.

Chức năng nối mạng thường được sử dụng để đánh giá các phương án tối ưu đường chuyển động và phân bổ tài nguyên (Hình 6.7b). Cụ thể là xác định đường đi hợp lý giữa hai điểm hoặc xác định vùng ảnh hưởng của một dịch vụ công cộng. Ví dụ, vùng phục vụ của dịch vụ chữa cháy, đường phân phối bưu phẩm, đường đi dự định của dịch vụ cấp cứu. Trong tất cả những ví dụ này thì hệ thống GIS sẽ phân tích yếu tố khoảng cách, vận tốc cho phép trên đường giao thông, và các thông số giao thông khác (như mật độ giao thông, mức độ kiểm soát giao thông) để xác định các đường đi hợp lý khác nhau kết nối giữa hai điểm (vị trí không gian).

6.3.6 Phép tính đại số

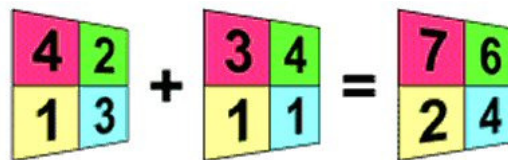
Phép tính đại số
(Map Algebra Utilities)



Hình 6.8 Phép tính đại số trên bản đồ.

Chức năng sử dụng phép tính đại số cho phép GIS xác định quan hệ toán học giữa các lớp dữ liệu. Toàn bộ các bản đồ có thể được gộp với nhau (phép cộng), cắt lẫn nhau (phép trừ), nhân và chia các dữ liệu thuộc tính theo những điều kiện hay quy tắc do người sử dụng đặt ra. Ví dụ một lớp dữ liệu mới từ dữ liệu về độ cao địa hình và dữ liệu về mực nước ngầm để thành lập lớp dữ liệu về các khoảng cách từ mặt đất đến mực nước ngầm (Hình 6.8).

Map algebra (Đại số bản đồ) có thể được sử dụng để phân tích không gian các dữ liệu dạng raster. Các biểu thức của algebra cũng tương tự như các hoạt động chồng xếp nhưng các thành phần phép tính đại số được sử dụng trong việc kết hợp hoặc so sánh các lớp dữ liệu (Hình 6.9).



Hình 6.9 Chồng xếp đại số hai lớp dữ liệu không gian

Map algebra cho biết cách so sánh các lớp dữ liệu thông qua chức năng thuật toán. Trong sơ đồ này giá trị của các cell trùng khớp trong hai lớp dữ liệu được tính bằng tổng của chúng trong lớp dữ liệu kết quả biểu thức xây dựng là: $[\text{grid1}] + [\text{grid2}]$

Trong ESRI ArcView, biểu thức đại số bản đồ (Map algebra expressions) được xây dựng trong chức năng Map Calculator để phân tích một hay nhiều lớp dữ liệu. Kết quả là tạo ra một lớp dữ liệu mới sau khi áp dụng biểu thức đại số (được thành lập từ sử dụng bàn phím hoặc nhấn chuột vào lớp dữ liệu toán tử) và các yêu cầu.

6.3.7 Công cụ biểu thị kết quả

Công cụ cuối cùng mà chương này đề cập đến là khả năng biểu thị kết quả của hệ thống GIS. Phần biểu thị kết quả đã được trình bày ở chương một về các thành phần của một hệ thống GIS nên không cần thiết phải nhắc lại chi tiết ở chương này.

Kết quả quá trình phân tích có thể được trình bày ở dạng in ấn ra giấy, lưu trữ dạng tệp tin, hoặc có thể được trình bày trên màn hình máy tính. Khả năng của phần mềm GIS và của thiết bị đầu ra như máy in, máy vẽ sẽ quyết định cho chất lượng của sự biểu thị kết quả phân tích. Trong phần lớn các trường hợp, hệ thống GIS cho phép bản đồ số, bảng tổng kết, và các tệp tin cơ bản có thể nhập được vào các hệ thống GIS khác.

Các kết quả đồ họa in ấn, mà phần lớn là các bản đồ, yêu cầu hệ thống GIS phải bao gồm cả một số lượng ký hiệu đồ họa và cho phép sử dụng các dạng lưu trữ đồ họa khác nhau. Các ký hiệu này chủ yếu dùng để biểu thị đồ họa các loại đường, đa giác hoặc điểm nhằm thể hiện các vật thể hay hiện tượng địa lý cũng như dùng để chú thích kết quả phân tích.

Kết luận

Bằng khả năng quản lý và phân tích tổng hợp cả bản đồ và dữ liệu thuộc tính, Geographic Information Systems đã ngày càng được ứng dụng rộng rãi trong hầu khắp các lĩnh vực hoạt động của con người. Một ứng dụng GIS bắt đầu bằng việc đặt ra mục đích hợp với khả năng của hệ thống GIS, rồi tiếp đến là bước thu thập dữ liệu bao gồm cả dữ liệu không gian và dữ liệu thuộc tính. Mục đích ứng dụng bao giờ cũng là động lực việc thu thập dữ liệu cũng như quá trình xây dựng mô hình. Ở giai đoạn phân tích tiếp theo, cả dữ liệu không gian (bản đồ) và dữ liệu thuộc tính được chuyển hóa và xử lý thông qua các công cụ phân tích mà phần mềm GIS cung cấp.

Chương này giới thiệu tổng quan quá trình phân tích riêng biệt cho CSDL và bản đồ nhưng thực ra hai loại dữ liệu này có quan hệ mật thiết với nhau trong hệ thống GIS. Thực sự thì sức mạnh của các ứng dụng GIS nằm ở đặc điểm có thể truy cập và phân tích đồng thời cả dữ liệu không gian và dữ liệu thuộc tính. Các chức năng phân tích được giới thiệu ở đây theo nghĩa rộng được coi là rất phổ biến cho mọi hệ thống GIS nhưng không phải là tất cả những gì các hệ thống GIS hiện nay có thể thực hiện được. Chuyên ngành GIS đã và đang liên tục phát triển và cùng với sự tốc độ nhanh chóng của tin học hiện đại, chắc chắn GIS sẽ cung cấp cho người sử dụng nhiều công cụ xử lý và phân tích ngày càng mạnh hơn.

Sự phổ cập rộng rãi của GIS cũng có nguồn gốc từ ảnh hưởng của nó tới những nhà quản lý qua các công cụ hỗ trợ quyết định và công cụ về hệ thống chuyên gia. Đây là hai loại công cụ có tác động sâu rộng nhưng cần được cung cấp dữ liệu thật nhanh chóng và đáng tin cậy. Sự phát triển của GIS như vậy sẽ phụ thuộc vào sự phát triển của hệ thống cơ sở dữ liệu, vào các phương pháp bảo đảm thu thập dữ liệu chính xác và lưu trữ dữ liệu một cách bền vững. Ở chương tiếp theo sẽ đề cập đến vấn đề này.

Câu hỏi ôn tập

1. Hãy diễn tả cấu trúc quá trình phân tích GIS để đạt được mục đích sử dụng nào đó?

2. Hãy nêu các vấn đề mà một hệ GIS có thể giải quyết?
3. Bạn hiểu thế nào về chức năng quản trị cơ sở dữ liệu của GIS?
4. Khái niệm về phân tích không gian trong GIS?
5. Trình bày các phương pháp nội suy thường gặp trong GIS?
6. Cho các ví dụ về chức năng truy vấn, phân loại và đo đạc của GIS?
7. Bản chất của phép chồng xếp bản đồ là gì? Các dữ liệu thuộc tính được xử lý như thế nào trong quá trình chồng xếp ?
8. Các nội dung của Sơ đồ Venn là gì?
9. Ý nghĩa của phép tạo vùng đệm trong thực tiễn như thế nào? Cho một vài ví dụ minh họa?
10. Tại sao gọi là phép đại số bản đồ

Chương Bảy

CHẤT LƯỢNG DỮ LIỆU, SAI SỐ VÀ BIẾN ĐỘNG DỮ LIỆU

Trong tổng chi phí cho một Hệ thống Thông tin Địa lý (HTTTĐL) thì công việc thu thập dữ liệu, chỉnh lý, nhập dữ liệu và phân tích dữ liệu chiếm một phần rất lớn. Do vậy vấn đề làm sao bảo đảm hạn chế được sai số và độ tin cậy trong các quá trình thu thập, xử lý, cập nhập và phân tích dữ liệu trở nên cực kỳ quan trọng. Chương này trình bày các vấn đề về biến động sai số dữ liệu, nguồn gốc và nguyên do các sai số, vấn đề truyền tải sai số trong HTTTĐL và hậu quả của chúng.

7.1 Giới thiệu chung

Hệ thống Thông tin Địa lý (HTTTĐL) được công nhận là một công cụ đầy hiệu quả nhưng đồng thời cũng cần chi phí lớn để thành lập và vận hành nó. Kinh phí không những cần thiết để lắp đặt phần cứng, cài đặt phần mềm và đào tạo nhân lực, mà còn nhằm vào việc thu thập dữ liệu, chỉnh lý, nhập dữ liệu và phân tích dữ liệu. Theo Burrough (1987) thì chi phí cho thành phần dữ liệu thường chiếm quá 50% tổng số kinh phí cho một HTTTĐL. Quá trình thu thập dữ liệu qua đo đạc và quan trắc thực địa đòi hỏi khối lượng lớn về sức người và của. Trong ba thập kỷ qua, với sự phát triển của công nghệ viễn thám, ảnh hành không, hệ thống định vị toàn cầu (GPS) cùng với kỹ thuật quan trắc tự động thì khối lượng dữ liệu thu được ngày càng trở nên to lớn. Tuy vậy, trước khi dữ liệu không gian và thuộc tính được sử dụng trong các quá trình phân tích thì bao giờ người ta cũng giả thuyết sai số trong các dữ liệu đó hầu như không đáng kể. Thêm vào đó, với kỹ thuật đồ bản máy tính hiện nay, các sản phẩm có chất lượng in ấn cao rất dễ tạo nên cảm giác tin tưởng vào các kết quả phân tích cũng như chất lượng của các dữ liệu đầu vào. Nhiều nghiên cứu về thổ nhưỡng đã cho ta biết là những đường cong biên trên bản đồ phân hạng đất dù được thành lập cẩn thận đến mức nào đi nữa thì cũng không thể thể hiện được một thực tế hiển nhiên về tính chất nhiều khi là ngẫu nhiên và fuzzy của sự phân bố các loại đất, nhất là cấu trúc vật lý của đất.

Do vậy vấn đề làm sao bảo đảm hạn chế được sai số và độ tin cậy trong các quá trình thu thập, xử lý, cập nhập và phân tích dữ liệu trở nên cực kỳ quan trọng. Bảng 7.1 chỉ ra các yếu tố là nguyên nhân tạo ra sai số trong dữ liệu GIS. Khái niệm sai số ở đây bao gồm cả lượng biến động ngẫu nhiên trong tự nhiên. Nhóm I bao hàm các sai số dễ nhận biết nhất. Nhóm II gồm các sai số không rõ ràng và chỉ được phát hiện ra nếu xem xét dữ liệu cẩn thận hoặc trong quá trình chỉnh lý dữ liệu. Nhóm III có thể là nhóm quan trọng nhất vì nó bao gồm các sai số tạo ra trong quá trình phân tích dữ liệu. Các sai số loại này chỉ có thể được phát hiện khi người sử dụng không những có hiểu biết về dữ liệu mà còn phải thông thạo về cấu trúc dữ liệu và các thuật toán phân tích dữ liệu.

Sau đây chúng ta xem xét cụ thể từng loại nguồn gốc gây ra sai số trong GIS.

7.2 Sai số có nguồn gốc rõ ràng

7.2.1 Dữ liệu lạc hậu

Các tập dữ liệu dành cho một dự án không bao giờ được thu thập trong cùng một thời điểm trừ trường hợp dự án loại nhỏ. Các cơ sở nghiên cứu thực hiện dự án bắt buộc phải sử dụng các dữ liệu có sẵn đã công bố dưới dạng các bản đồ và các báo cáo trong

đó có thể bao gồm cả các ảnh viễn thám được thu thập mới nhất. Điều hiển nhiên là mức độ tin cậy của các dữ liệu môi trường và xã hội sẽ giảm dần theo thời gian trừ trường hợp dữ liệu địa chất. Mặc dù sự thay đổi các tầng địa chất vẫn diễn ra với tốc độ không đáng kể so với tốc độ biến đổi của các hiện tượng tự nhiên khác như tầng đất trồng trọt, chế độ thủy văn, thảm thực vật và phân bố chế độ sử dụng đất, nhưng dữ liệu địa chất vẫn có thể trở nên không có giá trị nếu chúng được thu thập theo những tiêu chuẩn và thiết bị đã lạc hậu. Rất nhiều dữ liệu cũ, thu thập theo dạng viết tay đã phải bị hủy bỏ vì sẽ quá tốn kém để xử lý chúng.

Bảng 7.1 Phân loại sai số trong GIS theo nguồn gốc

I. Nguồn gốc rõ ràng
1. Dữ liệu lạc hậu
2. Vùng phủ của dữ liệu – hoàn toàn và không hoàn toàn
3. Tỷ lệ bản đồ
4. Mật độ quan trắc/lấy mẫu
5. Mức độ liên quan
6. Dạng lưu trữ
7. Mức độ cập nhập
8. Chi phí
II. Nguồn gốc do quá trình thu thập dữ liệu
9. Độ chính xác địa lý
10. Độ chính xác của dữ liệu thuộc tính
11. Biến động dữ liệu trong khi cập nhập và thiếu khách quan
III. Nguồn gốc trong quá trình phân tích dữ liệu
12. Sai số máy tính
13. Quá trình phân tích không gian do lỗi logic và do chồng ghép
14. Vấn đề phân loại và tổng quát
Phương pháp, định nghĩa phân loại và tính nội suy

7.2.2 Vùng phủ của dữ liệu

Điều mong muốn của những người sử dụng dữ liệu là toàn bộ diện tích vùng nghiên cứu sẽ được quan trắc hay thu thập dữ liệu với mật độ đồng đều. Tuy vậy trong thực tế người ta vẫn phải làm việc với các mật độ dữ liệu khác nhau ở những vùng khác nhau. Ngay cả ở các nước phát triển, vẫn còn những vùng phải sử dụng bản đồ đất tỷ lệ 1:25 000 đến 1:50 000. Điều cần lưu ý hơn là trong 30-40 năm gần đây, các quan niệm và định nghĩa về loại đất, phương pháp thể hiện lên bản đồ và cả cách thức điều tra đất đã thay đổi. Những vấn đề lịch sử này dẫn đến vấn đề xử lý đường biên khi thành lập bản đồ đất và nhiều khi bắt buộc phải có các điều tra đất thực địa mới.

Trong trường hợp độ bao phủ của dữ liệu thu thập được lên vùng nghiên cứu không hoàn chỉnh thì cần phải xác định phương cách để đạt được độ đồng đều về dữ liệu cho toàn bộ vùng nghiên cứu. Hai phương án phổ biến là (a) thu thập dữ liệu bổ sung và (b) tái tạo dữ liệu bằng phương pháp nội suy hay tương tự cho vùng thiếu khuyết dữ liệu.

7.2.3 Tỷ lệ bản đồ

Bản đồ giấy thường được sản xuất cho từng loại chuyên ngành riêng biệt nhưng với công nghệ máy tính, nhất là cơ sở dữ liệu thì bản đồ số có thể chứa đựng cả chi tiết các quan trắc thực địa. Các cơ sở cung cấp bản đồ luôn có sẵn mức độ chi tiết của thông tin cho từng loại tỷ lệ. Vấn đề ở đây là phải căn cứ trên yêu cầu về mục đích nghiên cứu mà chọn tỷ lệ bản đồ thích hợp. Tỷ lệ bản đồ quá lớn sẽ gây khó khăn cho các công đoạn phân tích dữ liệu không gian và thuộc tính.

7.2.4 Mật độ quan trắc dữ liệu

Đã có rất nhiều nghiên cứu về độ dày phân bố các điểm quan trắc hay điểm lấy mẫu tùy thuộc vào yêu cầu của từng loại bản đồ với các tỷ lệ khác nhau. Tuy vậy các cơ sở cung cấp bản đồ rất ít khi cung cấp các thông tin về các dữ liệu đã thu thập để xây dựng các bản đồ đó. Hiện nay thái độ như vậy đang có chiều hướng thay đổi. Ví dụ Viện Điều tra Đất Hà lan cung cấp cả thông tin về các công ty điều tra đất đã thực hiện công tác điều tra dữ liệu thành lập bản đồ. Các thông tin này còn bao gồm cả vị trí và phân loại các điểm lấy mẫu. Trung tâm điều tra đất của Anh quốc cũng cung cấp cho khách hàng mua bản đồ những thông tin tương tự.

Khi có đầy đủ các thông tin về địa điểm, thời gian lấy mẫu và các thông tin liên quan thì người sử dụng bản đồ có thể sử dụng những công cụ thống kê như lý thuyết thống kê lấy mẫu, kỹ thuật xử lý thống kê địa vật lý, v.v. để xem xét và đánh giá với mức độ tin cậy hợp lý cho các kết quả phân tích dữ liệu GIS.

Tóm lại mật độ hay phân bố các điểm quan trắc hay lấy mẫu không phải là điều kiện đủ để đánh giá chất lượng dữ liệu. Người ta còn cần biết mật độ quan trắc/lấy mẫu của dữ liệu có thích hợp cho mục đích nghiên cứu GIS của mình hay không.

7.2.5 Ý nghĩa của dữ liệu

Không phải tất cả dữ liệu không gian địa lý đều trực tiếp đại diện cho vấn đề quan tâm trong các nghiên cứu phân tích dữ liệu GIS. Lý do là có nhiều dữ liệu không thể thu thập được hoặc đòi hỏi chi phí quá cao để thu thập được. Ví dụ các bộ phận cảm ứng trong viễn thám thu nhận tín hiệu điện từ để ước tính các loại hình sử dụng đất, lượng chất hữu cơ của thảm thực vật, độ ẩm đất, độ phì đất, mức độ xói mòn đất. Các quan hệ giữa các đại lượng điện từ và các đại lượng biểu thị đã được nghiên cứu từ trước nên thông tin thu thập được qua các đại lượng điện từ có thể trở thành nguồn dữ liệu có giá trị.

Công việc hiệu chỉnh bộ phận cảm ứng đã trở thành thành phần rất quan trọng của ngành viễn thám. Thường thì một nhóm pixels có màu sắc đồng nhất trên bức ảnh viễn thám sẽ được xác định như là “training set”. Nhóm pixels này sẽ được so sánh với các quan trắc thực địa tiến hành trên mặt đất để biết được nó đại diện cho loại cây trồng, loại đất hay bất kỳ vật thể nào ở mặt đất mà ta cần nghiên cứu. Những nhóm pixels khác trên ảnh cũng được ước đoán theo nguyên tắc kết hợp với khảo sát mặt đất với các “training set” khác.

Bản đồ của một dự án nhất định thường được thành lập sử dụng phương pháp tương tự với các bản đồ cùng loại. Kỹ thuật toán thống kê có vai trò quan trọng trong suốt quá trình giải đoán như vậy.

7.2.6 Dạng dữ liệu

Có ba vấn đề liên quan đến dạng dữ liệu khi chúng được lưu trữ trong máy tính. (a) Vấn đề dữ liệu được lưu chuyển như thế nào giữa các hệ thống máy tính khác nhau. Đây có thể là môi trường lưu chuyển như đĩa mềm, băng từ, lưu chuyển qua mạng, v.v. hoặc có thể là mật độ lưu trữ dữ liệu (bits per inch) hay thể loại ký tự được sử dụng (ASCII, EBCDIC hay binary). (b) Vấn đề sắp xếp dữ liệu hay còn được gọi là cấu trúc dữ liệu, raster hay vector. Nếu là raster thì kích thước của pixel là bao nhiêu? Liệu hệ thống máy tính của tổ chức sử dụng dữ liệu có cần phải chuyển đổi dạng dữ liệu hay không. (c) Vấn đề cuối cùng liên quan đến chính bản thân dữ liệu như tỷ lệ, hệ tham chiếu đã được sử dụng, và thể loại của dữ liệu.

Để dữ liệu có thể được lưu chuyển dễ dàng thì dữ liệu cần được lưu trữ dưới dạng phổ thông nhất, cho phép càng nhiều hệ thống máy tính sử dụng được càng tốt. Dạng lưu trữ dữ liệu không nhất thiết phải là dạng nén chiếm ít bộ nhớ. Mỗi quốc gia đều đã tiêu chuẩn hóa dạng lưu trữ của dữ liệu nhưng thật ngạc nhiên là trên thế giới vẫn chưa đạt được một tiêu chuẩn thống nhất. Điều đó cũng chứng tỏ tốc độ phát triển nhanh chóng của các ứng dụng phân tích dữ liệu số không gian địa lý.

7.2.7 Mức độ cho phép sử dụng

Không phải mọi dữ liệu không gian địa lý đều có thể cung cấp cho mọi người sử dụng vì theo quy định quốc gia thì một số dữ liệu đã được gắn với một mức độ bảo mật nhất định. Bên cạnh yếu tố quân sự đối với bản đồ địa hình (liên quan đến mô hình số địa hình) trong dữ liệu GIS thì vấn đề thủ tục hành chính có thể gây trở ngại đáng kể cho vấn đề lưu thông dữ liệu. Tương tự, giá thành và dạng lưu trữ cũng có thể làm cho dữ liệu không thể chuyển giao. Các cơ quan chính phủ hoặc quốc tế có thể cho ta biết rõ những yêu cầu và đặc tính lưu chuyển của dữ liệu không gian.

7.2.8 Giá thành

Mọi công việc thu thập, xây dựng dữ liệu đầu vào, chuyển đổi hay biến đổi hệ tham chiếu dữ liệu đều cần có kinh phí. Người quản lý bất kỳ dự án nào cũng phải ước tính các chi phí này, nhất là khi so sánh việc sử dụng dữ liệu thu thập từ các nguồn cung cấp và việc khởi động công tác điều tra mới. Công việc số hóa bản đồ giấy sử dụng digitizer có thể trở nên rất tốn kém nếu dùng bản đồ và số liệu viết tay. Tiến hành số hóa với scanner có thể giảm bớt chi phí khi sử dụng bản đồ đường đồng mức hoặc ảnh. Nhiều công việc liên quan đến dữ liệu có thể giao cho các công ty chuyên trách theo phương thức hợp đồng sẽ tiết kiệm kinh phí so với việc sử dụng nhân viên không chuyên trách.

7.3 Sai số trong quá trình thu thập dữ liệu

7.3.1 Sai lệch vị trí

Tầm quan trọng của độ chính xác dữ liệu địa lý phụ thuộc chủ yếu vào dạng tích trữ của dữ liệu. Dữ liệu trắc địa đối với các vật thể cụ thể như đường sá, nhà cửa, đường biên các thửa ruộng và sử dụng đất, thường được thu thập với độ chính xác cao về vị trí. Các thiết bị điện tử trắc địa hiện đại hiện nay cho phép đạt được độ chính xác đo đạc đến từng decimét. Ngược lại, các đường biên không rõ ràng như đường phân loại đất và thảm thực vật thì phụ thuộc vào sự đánh giá của người lập bản đồ khi xem xét dữ liệu thực địa. Đối với bản đồ phân loại thực vật thì ranh giới giữa hai loại thực vật rất không rõ ràng và bao giờ cũng có vùng chuyển tiếp mà trong khu vực đó hai loại thực vật sẽ song song tồn tại có thể như là kết quả của sự chuyển tiếp về khí hậu, loại đất, chế độ

nước và địa hình. Sự biến đổi về chế độ thủy văn và loại độ dốc thực tế cũng không bao giờ diễn ra đột ngột như các đường đồng đẳng biểu thị trong các bản đồ.

Sai lệch vị trí địa lý có thể là kết quả của sự đánh giá không hợp lý khi quan trắc thực địa, sự co giãn của giấy bản đồ, hay sai số trong quá trình vectơ hóa sau khi số hóa bản đồ bằng scan. Những sai số cục bộ về vectơ hóa có thể được chỉnh sửa bằng việc sử dụng giao diện hỏi đáp và sai lệch chung do dẫn nở giấy thì chỉnh sửa bằng kỹ thuật biến đổi “tấm mỏng cao su” (rubber sheeting methods).

7.3.2 Sai lệch dữ liệu thuộc tính

Độ chính xác của dữ liệu thuộc tính phụ thuộc vào mức độ tin cậy của dữ liệu khi chúng được gán vào các thực thể điểm, đường và vùng của bản đồ. Người ta phân biệt độ chính xác về chất lượng và độ chính xác về số lượng. Độ chính xác về chất lượng liên quan đến bản chất của dữ liệu thuộc tính, ví dụ một thửa ruộng hoặc một đơn vị sử dụng đất đang trồng lúa thì bị chú thích là trồng khoai tây. Độ chính xác về số lượng liên quan đến tính khách quan trong ước tính thuộc tính, ví dụ một máy đo pH không được hiệu chỉnh tốt có thể gây sai số cao hơn một đơn vị pH cho tất cả các số đo.

7.3.3 Biến động dữ liệu

Sự biến động dữ liệu (hay là sai số) có thể xảy ra với dữ liệu không gian theo nhiều cách khác nhau. Sai số trong quá trình cập nhập dữ liệu có thể được chỉnh sửa dễ dàng nếu chúng sẽ dẫn đến các kết quả phân tích bất bình thường nhưng nếu các kết quả phân tích không đột biến thì sẽ rất khó nhận biết được.

Sai số đo đạc. Dữ liệu thu thập có thể có chất lượng kém do quá trình quan trắc đo đạc thực địa diễn ra thiếu tính tin cậy, không chính xác và sử dụng thiết bị không đạt chuẩn mực. Người đọc cần phân biệt rõ khái niệm đúng (accuracy) và chính xác (precision). Số đo đúng là số đo rất gần với số thực tế, hay là mức độ gần sát với số thực. Theo quan điểm thống kê thì độ chính xác biểu thị sự biến động của dữ liệu đo (thường là độ biến động chuẩn, standard deviation) xung quanh số trung bình. Trong ngành máy tính thì độ chính xác được biểu thị bằng số chữ số sau dấu thập phân.

Sai số thực địa. Người đi quan trắc thực địa có tầm quan trọng đặc biệt đối với chất lượng dữ liệu mà họ thu thập. Các tiêu chuẩn, phương pháp và trình tự quan trắc cần được thống nhất và có tài liệu chỉ dẫn rõ ràng sẽ giúp cho người đi thực địa nâng cao chất lượng công việc. Tuy vậy người sử dụng dữ liệu quan trắc môi trường nên hiểu rõ là các tập dữ liệu quan trắc bao giờ cũng có độ chính xác khác nhau nếu chúng được thu thập bởi những tổ chức khác nhau và trong các thời điểm khác nhau.

Ở các tổ chức chuyên về đo đạc thực địa lớn thì người ta thường lưu trữ đánh giá chất lượng quan trắc của từng cá nhân thực hiện công việc. Các dữ liệu phụ này có thể nên được công bố cùng với dữ liệu quan trắc mặc dù có sự phản đối của phần lớn người quan trắc thực địa. Cách thức hiệu quả hơn để tăng độ chính xác quan trắc là chấn chỉnh và tiêu chuẩn hóa tất cả các vấn đề liên quan đến công việc thu thập dữ liệu thực địa như tiêu chuẩn hóa kỹ thuật đo đạc, tiêu chuẩn hóa biểu mẫu, và thành lập các tiểu ban liên quan giữa các nhà quản lý và người đi thực địa để công việc thu thập dữ liệu có thể được thực hiện theo những tiêu chuẩn chất lượng cao nhất.

Sai số trong phòng thí nghiệm. Người ta thường cho rằng chất lượng dữ liệu thu thập được trong phòng thí nghiệm bao giờ cũng tốt hơn so với dữ liệu thu thập ngoài thực địa. Mặc dù các dữ liệu phân tích của cùng một mẫu tuân thủ theo cùng một phương pháp trong một phòng thí nghiệm thường cho ra kết quả giống nhau nhưng sự

giống nhau về kết quả đó không bao giờ lặp lại khi thực hiện ở các phòng thí nghiệm khác nhau. Một nghiên cứu so sánh kết quả phân tích trong phòng thí nghiệm của Trung tâm thông tin quốc tế về đất (the International Soil Reference and Information Center, Wageningen, 1986) cho biết biến động về kết quả phân tích của cùng một mẫu đất có thể là hơn $\pm 11\%$ đối với thành phần sét, $\pm 20\%$ đối với mức độ trao đổi ion ($\pm 25\%$ cho thành phần sét), $\pm 10\%$ đối với mức độ bão hòa và ± 0.2 đơn vị đối với số đo pH. Đúng là tác động của việc tiêu chuẩn hóa các phòng thí nghiệm là vô cùng to lớn.

Biến động không gian và chất lượng bản đồ. Rất nhiều bản đồ chuyên ngành, đặc biệt là các bản đồ về tài nguyên thiên nhiên như thảm thực vật và phân loại đất, đã không xem xét đến yếu tố biến động cục bộ của số đo. Vấn đề này đang trở nên đề tài được nghiên cứu rất nhiều trong thời gian gần đây, đặc biệt trong lĩnh vực điều tra phân loại đất, vật lý đất và nước ngầm. Các nghiên cứu chú trọng đến mức độ biến động của các đặc tính và các phương pháp đánh giá phân loại và nội suy.

Trước đây, các bản đồ phân loại đất coi mức độ biến động 15% cho phép để đánh giá mức đồng nhất của một thông số đặc tính đất. Độ biến động này được định nghĩa như là các số liệu quan trắc không đúng với số liệu chuẩn gắn liền với chủ giải các mức độ của thông số. Quan niệm về biến động về thông số Z ở vị trí x trong một đơn vị loại đất được thể hiện bằng phương trình sau

$$Z(x) = \mu + \alpha_j + \varepsilon$$

trong đó μ là giá trị trung bình thực tế của Z trên phạm vi toàn bộ bản đồ, α_j là giá trị sai lệch giữa μ và giá trị trung bình trong đơn vị đất và ε là đại lượng biến động có phân bố chuẩn.

Trong thực tế thì phương pháp thể hiện biến động dữ liệu trên trong phạm vi một đơn vị phân loại đất không thể hiện đầy đủ sự biến động các đặc tính của đất. Nguyên do được trình bày như sau:

- Đối với cùng một thông số thì mức độ biến động sẽ khác nhau trong phạm vi các đơn vị phân loại khác nhau;
- Mức độ biến động của đại lượng ε thường có quan hệ hàm mũ với kích thước không gian của đơn vị bản đồ (diện tích);
- Sự biến động giá trị của các thông số trong một đơn vị bản đồ không tuân theo quy luật thống kê ngẫu nhiên (phân bố chuẩn) mà phụ thuộc vào phương thức lấy mẫu và phương pháp nội suy khi thành lập bản đồ, nhất là bản đồ ở diện rộng.

7.4 Sai số trong quá trình phân tích dữ liệu

Sai số trong quá trình phân tích dữ liệu GIS là hậu quả của sai số do máy tính (ví dụ sai số làm tròn), do các giả thiết đồng nhất trong quá trình phân tích không gian cũng như do quá trình phân loại.

7.4.1 Sai số máy tính

Sai số do máy tính là loại sai số rất khó nhận biết và chủ yếu do khả năng tích trữ con số trong bộ nhớ của máy tính chỉ theo độ chính xác của bộ xử lý.

Sai số làm tròn số là một trong sai số phổ biến của bộ xử lý máy tính và là hậu quả của mức độ chính xác cho phép tối hạn của bộ xử lý trung tâm. Ví dụ mỗi biến số của chương trình xử lý được lưu trữ chỉ với một hoặc hai byte thay vì một số byte vô giới

hạn. Trong các máy tính thế hệ cũ, với mức độ tích trữ và xử lý dữ liệu ở dạng 8 bit và 16 bit thì sai số làm tròn khi cần thực hiện nhiều phép tính liên tiếp sẽ rất đáng kể.

Gruenberg (1984) công bố một ví dụ minh họa sai số khi bình phương liên tiếp 27 lần số 1,0000001 (tương tự phép tính với số mũ 134217728) và kết quả là máy tính tạo ra một sai số đến 1200 % so với kết quả chính xác.

Trong các phép xử lý số liệu không gian, nhất là với dạng raster, dữ liệu đều ở dạng số nguyên nên bộ xử lý của máy tính luôn thực hiện phép làm tròn số sau khi tính toán chu vi và diện tích. Sai số làm tròn sẽ trở nên đáng kể nếu chương trình xử lý không có phần khai báo tăng độ chính xác xử lý (single and double precision).

7.4.2 Sai số do phân tích địa lý

Phần lớn các phân tích dữ liệu không gian đều được thực hiện với giả thiết (a) dữ liệu là đồng đều, (b) phương pháp số hóa đã được thực hiện không sai sót (infallible), (c) công việc chồng ghép bản đồ chỉ là việc giao cắt các đường biên và kết nối mạng đường cong, (d) các đường biên được coi là không có độ dày, (e) mọi thuật toán có tính xác định và không có yếu tố biến đổi ngẫu nhiên, và (f) phân loại các thông số tự nhiên (thuộc tính trong bản đồ) theo lớp đã được thực hiện hoàn chỉnh. Các vấn đề này cần được nhìn nhận kỹ càng về các dữ liệu cơ sở, nhất là về các yếu tố và kỹ thuật đã sử dụng để thu thập và thành lập bản đồ.

7.4.3 Sai số do phân loại và tổng quát hóa

Người ta nhận thấy một lượng lớn các vấn đề dữ liệu khi đưa vào sử dụng trong các hệ thống GIS có nguyên nhân từ các phương pháp phân loại, tổng quát hóa và nội suy đã được sử dụng cho các dữ liệu trước khi chúng được nhập vào hệ thống GIS. Burrough (1997) trình bày khá chi tiết về vấn đề này.

7.4.4 Quá trình phân tích xử lý dữ liệu không gian

Các phép phân tích dữ liệu trong GIS thường liên quan đến chồng ghép hai hay nhiều lớp mạng không gian. Mạng không gian có thể bao gồm mạng các ô đồng đều hoặc các đa giác không đồng dạng. Phép chồng dữ liệu có thể vì mục đích chuyển đổi dữ liệu, như đặt dữ liệu dạng vector của một mạng các hình đa giác lên một mạng vuông với một độ phân giải nhất định để thu được dữ liệu dạng raster. Mục đích khác có thể là để kết hợp dữ liệu bằng cách chồng lên nhau hai lớp dữ liệu raster hoặc hai lớp dữ liệu dạng đa giác. Mục đích của việc chồng cắt bản đồ có thể chỉ đơn giản là về mặt hình học hoặc là xử lý các thuộc tính của các ô (pixels) hoặc các đa giác (polygons).

Sau đây ta sẽ xét đến các sai số có thể sinh ra như là hậu quả của (a) biểu thị mạng các đa giác (polygons) bằng mạng ô vuông (raster), (b) phép biến đổi số học hoặc logic hai lớp dữ liệu, và (c) chồng cắt hai lớp mạng đa giác. Cuối cùng của chương này sẽ trình bày vấn đề sử dụng đường biên để biểu thị sự biến đổi các thông số tự nhiên.

Quá trình raster hóa một bản đồ vector

Sai số có thể xuất hiện khi chuyển đổi một lớp dữ liệu vector thành lớp raster. Đầu tiên và là sai số dễ nhận biết khi mỗi ô pixel chỉ có thể gán được một giá trị thuộc tính như là giá trị trung bình của thuộc tính đó trong vùng ô vuông. Các ảnh LANDSAT cũ có độ phân giải $80\text{ m} \times 80\text{ m}$ và giá trị của thông số trong mỗi ô được ước tính trên cơ sở lượng phản xạ trung bình trên toàn bộ diện tích 6400 m^2 . Nếu một phần của ô vuông có sự hiện diện của vật thể có độ phản xạ ánh sáng cao như đường sá hay bãi cát thì rất

có thể cả ô vuông sẽ được miêu tả như đường hoặc bãi biển. Loại sai số như vậy rất phổ biến khi ô pixel có kích thước đáng kể.

Sai số cũng có thể trở nên đáng kể khi một bản đồ vector được biến đổi thành dạng raster do các đường biên của đa giác không bao giờ trùng với đường ranh giới các ô vuông. Rất nhiều nghiên cứu đã đưa ra các công thức ước tính tổng sai số do phép biến đổi dữ liệu không gian này.

Quá trình số hóa bản đồ

Nguồn gốc sai số trong quá trình số hóa chủ yếu từ sai số của bản đồ gốc và sai số do hình thức biểu thị qua số hóa.

Bên cạnh sai số do sự co giãn của bản đồ giấy thì sai số trong quá trình số hóa là do các đường biên trên bản đồ giấy bao giờ cũng có một độ dày nhất định. Một đường biên dày 1 mm trên bản đồ tỷ lệ 1:1250 sẽ chiếm diện tích có chiều rộng 1,25 m thực tế. Một bản đồ đất có kích thước 400 m × 600 m có thể bao gồm 24000 mm tổng chiều dài các đường biên thửa ruộng và chiếm một diện tích 10% diện tích tổng số của bản đồ. Nguyên tắc là lấy đường ở giữa đường biên trên bản đồ giấy làm đường biên khi số hóa nhưng thực tế không thể thỏa mãn được nguyên tắc này. Khi số hóa bằng bản số hóa thì việc lựa chọn đường biên do bàn tay con người thực hiện còn khi sử dụng kỹ thuật scan thì việc số sẽ liên quan đến các thuật toán của chương trình scan. Sai số tiếp theo là do việc các đường cong chỉ được biểu thị bằng các đoạn thẳng ngắn (vertices) khi số hóa.

Quá trình chồng ghép hai hay nhiều lớp mạng đa giác

Khi chồng ghép bản đồ đã số hóa thì xuất hiện một số lớn các đa giác nhỏ, nhất là khi các đường biên phức tạp có độ biến động không gian lớn. Với kỹ thuật máy tính hiện nay thì các đa giác con này có thể được loại trừ theo nguyên tắc lựa chọn ngẫu nhiên theo hai phía của đường biên hoặc theo nguyên tắc phức hợp bình phương bé nhất dựa trên độ chính xác do người thực hiện phép chồng ghép lựa chọn trước. Dù sao thì các sai số cũng xuất hiện vì khi loại bỏ các đa giác nhỏ thì đường biên thực tế đã không được biểu thị chính xác tuyệt đối.

Quá trình tính toán thông số thuộc tính

Vấn đề truyền tải sai số khi thực hiện các phép tính trên các dữ liệu thuộc tính nhiều khi trở thành rất đáng kể, nhất là khi kết hợp tính toán trên nhiều lớp bản đồ khác nhau. Sự truyền tải hay là tích lũy sai số thường xuất hiện khi thực hiện các phép tính số học hoặc phép tính logic.

Để minh họa ta xét sự tích lũy này qua các phép tính số học cộng, trừ, nhân chia và lũy thừa. Giả sử giá trị x của một thông số thuộc tính trên một bản đồ X có biến động ngẫu nhiên là ε . Biểu diễn toán học là $x \pm \varepsilon x$. Ví dụ như là lượng nước trong đất mà cây trồng có thể sử dụng và ta muốn kết hợp bản đồ lượng nước X này với bản đồ hiệu suất tưới nước Y trong đó hiệu suất sử dụng nước tưới cũng có độ biến động tự nhiên $y \pm \varepsilon y$. Nếu ta giả thiết hai thông số này hoàn toàn độc lập với nhau và mỗi lượng biến động εx và εy có giá trị 20% thì sai số của tổng lượng nước cây trồng có thể sử dụng $u = (x + y)$ sẽ vào khoảng 28%. Đối với các phép tính đồ bản thì các sai số loại này có thể tích lũy đến mức đáng lo ngại, nhất là đối với các phép tính nhân, chia và hàm mũ hay logarit.

Ví dụ sau đây về mô hình ước tính lượng đất xói mòn sẽ làm rõ hơn mức độ sai số được tính lũy qua quá trình tính toán dữ liệu thuộc tính khi chồng ghép bản đồ. Ta nghiên cứu ước lượng lượng đất có khả năng bị bào mòn do mưa trên vùng đất trồng ngô trong khoảng thời gian 40 năm. Lượng đất xói mòn được tính theo một biểu thức phổ biến gọi là công thức tổng hợp xói mòn đất (Universal Soil Loss Equation, USLE).

$$A = R \times K \times L \times S \times C \times P$$

Trong đó A là lượng đất bị bào mòn hàng năm (tấn ha^{-1}); R là sức xói mòn do mưa; K là sức xói mòn của đất; L là chỉ số chiều dài sườn dốc; S là độ dốc sườn dốc (%); C là chỉ số canh tác; và P là chỉ số liên quan đến các biện pháp chống xói mòn đất đã sử dụng trên vùng đất tính toán.

Nguồn dữ liệu cho nghiên cứu này bao gồm các bản đồ chuyên đề về phân bố mưa và về giá trị trung bình của các thông số về đất và điều kiện khí tượng. Các thông số R , L và S được ước tính từ các công thức tương quan.

Thông số R. Giá trị của R trong USLE được tính từ công thức do FAO

$$R = 0,11 abc + 66$$

Trong đó a là lượng mưa trung bình hàng năm (cm); b là lượng mưa (cm) lớn nhất hàng ngày xảy ra trong vòng 2 năm; c là tổng lượng mưa lớn nhất của một trận mưa (cm) xảy ra trong vòng 2 năm. Các nghiên cứu cho vùng đã chỉ ra giá trị của các thông số này như sau.

$$a = 172,5 \pm 20 \text{ cm}; b = 5,41 \pm 1,1 \text{ cm}; c = 2,25 \pm 0,5 \text{ cm}$$

Ta có thể nhận thấy giá trị các biến động chuẩn khá đáng kể do không có nhiều dữ liệu đo đạc trong vùng.

Như vậy ta tính được giá trị của thông số $R = 297 \text{ cm}$ với độ biến động chuẩn $\pm 72 \text{ cm}$ trên năm.

Thông số K. Theo kết quả các nghiên cứu về xói mòn đất thì giá trị của K bằng khoảng 0.1 đối với loại đất laterit và do độ biến động rất lớn của các loại đất nên giá thiết độ biến động chuẩn của K khoảng 50% ($\pm 50\%$) là hợp lý.

Thông số L. Thông số L được ước tính theo công thức Wischmeier và Smith (1978)

$$L = (l/22,1)^w$$

Trong đó l là chiều dài sườn dốc trung bình trong khu vực nghiên cứu (mét). Nếu coi chiều dài sườn dốc tiêu biểu trong vùng nghiên cứu có giá trị bằng 100m $\pm 20\text{m}$ thì ta có giá trị của L là 2,13m ± 0.045 .

Thông số S. Thông số độ dốc được tính theo công thức tương quan parabol

$$S = 0,0065 s^2 + 0,0454 s + 0,065$$

Với s là độ dốc tiêu biểu đo bằng % (Wischmeier và Smith, 1978). Đối với độ dốc bằng 10 ± 2 % (trên cơ sở phân tích bản đồ độ dốc) ta có giá trị của $S = 1,169 \pm 0,122$.

Thông số C. Ước tính giá trị chỉ số cây trồng C lớn nhất là 0,63 nhưng vì cây trồng không che phủ mặt đất suốt năm nên giá trị C của cây ngô theo Róe (1975) giao động giữa 0,4 và 0,9 trong khi độ biến độ chuẩn là ± 15 .

Thông số P. Chỉ số biện pháp chống xói mòn đã được áp dụng P được ước tính có giá trị bằng 0,5 $\pm 0,1$.

Khi đã có đầy đủ giá trị các thông số thì nếu áp dụng USLE ta thu được lượng đất hàng năm bị xói mòn trong khu vực nghiên cứu là 23 $\pm 14,8 \text{ tấn ha}^{-1} \text{ năm}^{-1}$. Điều đó có nghĩa là hàng năm bề mặt vùng đất canh tác bị sụt đi 0,23 $\pm 15 \text{ cm năm}^{-1}$ hay 9 $\pm 6 \text{ cm}$ trong 40 năm. Như vậy mức độ biến động của xói mòn đất trong vùng nghiên cứu trong điều kiện khí hậu, đất và địa hình được ước tính nằm trong khoảng từ 3 đến 15 cm.

Ta có thể thấy đây là một dao động quá lớn. Dao động này là kết quả của các phép nhân liên tiếp trong mô hình tính toán. Điều này chứng tỏ là nếu người sử dụng khi tính toán kết hợp các dữ liệu thuộc tính trong GIS với các giá trị dữ liệu thô thì có khả năng kết quả tính toán mô hình cuối cùng sẽ không còn có giá trị dự báo.

7.5 Kết luận

Chương này giới thiệu tổng quan các nguồn sai số và độ biến động của giá trị các thông số tự nhiên mà chúng có thể làm giảm độ tin cậy của hệ thống thông tin địa lý (GIS). Nhiều loại sai số có thể được nhận biết và loại trừ trước khi cập nhập chúng vào hệ thống GIS hoặc có thể xác định được trong quá trình nhập dữ liệu. Loại sai số khác sinh ra trong quá trình xử lý và phân tích dữ liệu như sai số trong quá trình số hóa hay phân tích chồng ghép cắt dữ liệu không gian, thì sẽ khó phát hiện hơn.

Sự biến động tự nhiên theo không gian của giá trị các thông số tự nhiên có lẽ là loại sai số có tầm quan trọng mà phần lớn người sử dụng GIS khó cảm nhận được. Các sai số này được giấu kín qua những đường cong đẹp đẽ và những vùng tô màu hoàn hảo. Người sử dụng ít khi tiếp cận được với những thông tin liên quan đến quá trình thu thập số liệu để thành lập bản đồ. Khi dữ liệu thuộc tính có độ biến thiên đáng kể thì các mô hình toán học càng phức tạp bao nhiêu thì sẽ càng tạo nên tích lũy sai số biến động to lớn bấy nhiêu và có thể làm cho kết quả mà mô hình đạt được trở nên vô nghĩa.

Sau đây là những điểm cần đặc biệt chú ý để giảm bớt được tác động của các sai số và các biến động không gian trong tự nhiên.

- Xác định kỹ thuật phân loại tốt nhất và thích hợp nhất;
- Cải tiến kỹ thuật nội suy để tránh việc sử dụng các giá trị trung bình hay các vị trí lấy mẫu đặc trưng đối với các vùng nghiên cứu có diện tích lớn;
- Giảm thiểu những sai số trong quá trình thu thập dữ liệu hiện trường, nhất là những sai số phổ biến.

Câu hỏi ôn tập

1. Ý nghĩa của việc đánh giá sai số trong GIS? Các nguyên nhân dẫn đến sai số?
2. Hãy phân tích cụ thể về các nguyên nhân biến động dữ liệu dẫn đến sai số?
3. Có những sai số nào thường gặp trong quá trình phân tích dữ liệu không gian?
4. Hãy trình bày cụ thể về mỗi loại sai số trên?

Tài liệu tham khảo

- Aronoff, S.** 1989. *Geographic Information Systems: A management perspective*. WDL Publications, Ottawa, Canada. 294 p.
- Bernhardsen, T.** *Geographic Information Systems : An Introduction*. John Wiley & Sons, Inc., Indianapolis, U.S. 2002.
- Berry, J. K.** 1987. *Fundamental operations in computer-assisted map analysis*. Int. J. of Geographic Info. Systems 1:119-136.
- Berry, J. K.** 1989. *Beyond mapping: Spatial data analysis*. pp. 10-16. In The GIS Sourcebook. GIS World, Inc., Ft. Collins, CO. 180 p.
- Burrough, P.A.** 1986. *Principle of Geographical Information Systems for Land Resources Assessment*. Oxford University Press. U.K. 186 pages.
- Burrough, Peter A. và Rachael McDonnell, P. A.** *Principles of Geographical Information Systems (Spatial Information Systems)*; 2nd edition. Oxford Press, 1998.
- Đặng Văn Đức.** *Hệ thống Thông tin Địa lý*. Nhà XB Khoa học và Kỹ thuật, Hà nội, 2001.
- Dangermond, J.** 1986. *The software toolbox approach to meeting the user's need for GIS analysis*. pp. 66-75. In Geographic Information Systems Workshop Proceedings, Atlanta, GA, April 1-4, 1986. Am. Soc. for Photogrammetry and Remote Sensing, Falls Church, VA. 426 p.
- Darius J. Bartlett and Jennifer L. Smith.** *GIS for coastal zone management*. CRC Press LLC, Boca Raton, Florida, 2002.
- de Berg, M., M. van Kreveld, M. Overmans, and O. Schwarzkopf.** 2000. "Voronoi Diagrams: The Post Office Problem." Ch. 7 in Computational Geometry: Algorithms and Applications, 2nd rev. ed. Springer-Verlag, Berlin. pp. 147-163.
- DeMers, Michael N.** *Fundamentals of Geographic Information Systems*. John Wiley & Sons, Inc., Indianapolis, U.S. 2002.
- ESRI (Environmental Systems Research Institute),** *Understanding GIS: The Arc/Info Method*. ESRI Inc., New York, U.S., 2001.
- FAO,** *Land & water Resource Information Systems*, Roma, Italy 1999
- G.D.Lodwick, M.Feuchtwanger,** *Land information systems*, Curtin University of Technology, School of Land Survey and Land information, AUSTRALIA, 1989
- Lee Y. and Walsh, P.A.** 1984. *Map projections for use with the geographical information systems*. Oxford University Press. U.K. 288 pages.
- Meijerink, N.F.** 1994. *Coordinate systems and georeferencing*. In "Introduction to the use of GIS for practical hydrology". American Society of Civil Engineers (ASCE). U.S.A. 467 pages.
- Mitchell, Tyler.** *Web Mapping Illustrated*, O'Reilly Media, Inc., Sebastopol, CA, U.S., 2005.
- Nicolas Chrisman,** *Exploring Geographic Informatin Systems*, University of Washington, John Wiley & Sons, Inc., U. S, 2000

- Nguyễn Đình Dương, Eddy Nierynck, Phạm Ngọc Hồ và Luc Hens**, *Dự án xây dựng năng lực phát triển bền vững: Ứng dụng kỹ thuật Viễn thám và Hệ thống thông tin địa lý cho qui hoạch môi trường*, Hà nội 1999
- Nguyễn Trọng Bình, Trần thị Băng Tâm**, *Giáo trình hệ thống thông tin địa lý cho chuyên ngành quản lý đất đai*, Trường Đại học nông nghiệp 1, Hà nội ,1996
- Pat Hohl,Editor**, *GIS data conversion: Strategies, Techniques, and Management*, OnWord Press, USA, 1998.
- Paul A Longley. Michael F Googchild, David J Maguire and David W**, *Geographic Information Systems and Science*, John Wiley & Son Inc. England, 2001
- Peter F. Dale, John D. McLaughlin**, *Land Information Management*, Clarendon Press.Oxford, 1989
- Rhind, D.** 1989. *Why GIS?* Arc News, Environmental Systems Research Institute, Redlands, CA. 11(3):28-29.
- Rich, David W.** *Relational management and display of site environmental data*. Lewis Publishers, Boca Raton, Florida, U.S. 2002.
- Star, Jeffrey và Estes, John.** *Geographic Information Systems: An Introduction*. Prentice Hall, New Jersey, U.S., 1990.
- Trần thị Băng Tâm.** *Ứng dụng GIS xây dựng cơ sở dữ liệu tài nguyên đất đai phục vụ nông nghiệp phát triển bền vững*. Báo cáo đề tài NCKH cấp Bộ 2000-2003, Đại học nông nghiệp 1, 2004.
- Tổng cục Địa chính.** *Xây dựng Hệ qui chiếu và Hệ tọa độ quốc gia Việt nam*, 1999
- Tung Fung, Poh Chin Lai,Hui Lin and Anthony Gar-On Yeh**, *GIS in Asia*, selected Paper of the Asia GIS/LIS and Spatial Analysis Conference, GISASIAPACIFIC, 1997
- US Environmental Protection Agency.** *Geographic Information System(GIS) Guidelines Document*. Washington, DC, 1992
- Viện nghiên cứu địa chính, Bộ Tài nguyên và Môi trường.** *Hướng dẫn áp dụng Hệ tọa độ quốc gia mới và xây dựng qui trình kỹ thuật chuyển đổi tư liệu bản đồ từ Hệ VN72 sang Hệ VN 2000*. Hà nội năm 2001

Phụ lục A

GIỚI THIỆU MỘT SỐ PHẦN MỀM GIS

Sự thành công của một dự án HTTTĐL phụ thuộc vào nhiều yếu tố về lựa chọn thiết kế cơ sở dữ liệu, lựa chọn phần cứng và phần mềm, đào tạo và sử dụng nhân lực, thu thập và xử lý dữ liệu, v.v., trong đó đặc biệt quan trọng là quá trình xét chọn phần mềm GIS. Trên thị trường hiện nay có rất nhiều phần mềm GIS với đầy đủ những chức năng GIS. Bên cạnh đó, quan niệm về phần mềm mở (open softwares) về GIS đang được áp dụng ngày một rộng rãi, kể cả công nghệ liên kết GIS với Web. Phụ lục này không có ý định đề cập đến vấn đề chọn lựa phần mềm mà chỉ giới thiệu mang tính liệt kê một số phần mềm phổ biến nhất hiện nay về GIS.

A.1 Hệ phần mềm trên thị trường

Hệ phần mềm ESRI

ESRI (the Environmental Systems Research Institute) cung cấp các ứng dụng thiết kế cơ sở dữ liệu, tự động hóa cơ sở dữ liệu (CSDL), cài đặt phần mềm và dịch vụ hỗ trợ kỹ thuật cho phần mềm GIS. Các phần mềm phổ biến nhất hiện nay của ESRI là Arc/Info và ArcView.

Arc/Info là phần mềm GIS chuyên dụng chạy trên nền UNIX và Windows NT. Nó cho phép người sử dụng tạo ra và quản lý CSDL không gian to lớn với số lượng lớn người cập nhập đồng thời, thực hiện những phân tích dữ liệu không gian có các dạng khác nhau và cho phép sản xuất bản đồ số với chất lượng cao. Arc/Info là hệ GIS nhằm tích trữ, chỉnh lý, phân tích quản lý và thể hiện dữ liệu cấu trúc không gian dạng vectơ. Kể từ phiên bản 7.1.2, Arc/Info có thêm một môi trường phụ trợ lập trình như Visual Basic.

ArcView hiện nay được coi là phần mềm GIS được sử dụng rộng rãi nhất trên máy tính cá nhân và cũng có thể là phần mềm bản đồ phổ biến nhất trên thế giới. ArcView có giao diện hình họa thân thiện và cho phép thể hiện trực quan dữ liệu, tìm kiếm, phân tích dữ liệu, đồng thời bao gồm cả chức năng chỉnh sửa dữ liệu không gian. Với ArcView người sử dụng có thể thành lập bản đồ số và bổ sung bản đồ từ các dữ liệu có sẵn một cách dễ dàng. Các module phụ trợ cho ArcView như Spatial Analysis, Network Analysis, v.v. cho phép người sử dụng giải quyết rất nhiều vấn đề khác nhau liên quan đến dữ liệu địa lý.

Các phần mềm khác của ESRI bao gồm Database Integrator, ArcCAD, ArcStorm, ArcTools, ArcPress, ArcGis và MapObjects. MapObjects cung cấp nhiều công cụ được thiết kế dành cho người lập trình muốn bổ sung các chức năng GIS cho các phần mềm tin học của riêng mình.

Intergraph

Công ty Intergraph chuyên sản xuất phần mềm đồ họa là một trong những tên tuổi nổi bật trong thị trường thế giới về phần mềm GIS. GeoMedia là sản phẩm GIS đầu tiên của Intergraph sử dụng công nghệ Jupiter. Sản phẩm GIS hàng đầu của công ty này là MGE (Modular GIS Environments) được tung ra thị trường lần đầu tiên vào năm 1988. MGE sử dụng MicroStation như là công cụ xử lý đồ họa và là phiên bản cho máy tính cá nhân PC. Sau đó MGE được liên kết với bộ Intergraph Office: Civil Office,

Municipal Office, Mapping Office và GIS Office. Hệ Office này nhằm thỏa mãn các nhu cầu khác nhau của người sử dụng. GIS Office là phiên bản chính chạy trên nền UNIX và Mapping Office chạy trên nền DOS/Windows là phiên bản dành cho các máy tính để bàn.

Phiên bản 8.0 là phiên bản mới nhất của MGE với MicroStation V7 nhưng không có kế hoạch phát triển tiếp tục MGE vì sự hợp tác của Intergraph với MicroStation không còn thuận lợi như trước. Intergraph chuyển sang tập trung phát triển GeoMedia với phiên bản mới nhất là GeoMedia 06.00.34.20. Phiên bản này của GeoMedia có đầy đủ các chức năng phân tích GIS như truy vấn dữ liệu không gian và thuộc tính, tạo vùng đệm, chồng xếp bản đồ và thành lập bản đồ chuyên dụng cũng như các công cụ phục vụ mô phỏng và kết nối với Web.

MapInfo

MapInfo cung cấp phần mềm GIS có đặc thù khác biệt so với phần mềm của các công ty khác. MapInfo tích trữ dữ liệu cấu trúc không gian không hoàn toàn ở dạng vector. Nó biểu thị dữ liệu cả ở các dạng vector và raster khác nhau, trong khi liên kết các hình bản đồ, đồ thị và bảng dữ liệu. Phần mềm bao gồm cả công cụ cho phép người sử dụng chỉnh sửa các thành phần cơ sở của bản đồ với khả năng chọn lựa đồng thời nhiều vật thể để xóa, sao chép và chồng ghép.

MapInfo ProServer là phiên bản chạy trên máy chủ (server) cung cấp khả năng tìm kiếm và xử lý dữ liệu trên mạng máy tính. Nó cũng cho phép sử dụng phần mềm trên cơ sở giao tiếp qua Internet Web browser.

Bảng A.1 Các phần mềm GIS phổ biến trên thị trường

	Địa chỉ	Phần mềm
1. ESRI	Environmental Systems Research Institute (ESRI) 380 New York Street Redlands, CA 92373, USA www.esri.com	ArcGIS ARC/INFO, ArcView
2. Intergraph	Intergraph Corporation P.O. Box 240000 170 Graphics Drive Madison, AL 35758, USA www.intergraph.com	Modular GIS Environment (MGE) GeoMedia
3. MapInfo	MapInfo One Global View Troy, New York 12180, USA www.mapinfo.com	MapInfo Professional

Phiên bản mới nhất năm 2005 là MapInfo Professional v8.0. Các thành phần khác của hệ thống phần mềm MapInfo bao gồm: MapInfo Professional® là sản phẩm hàng

đầu về GIS; MapInfo MapX® là công cụ để những người lập trình thành lập các công cụ GIS vào phần mềm của họ; MapInfo® MapXtreme là phần mềm server về GIS có thể chạy trên nền Java và Windows; MapInfo® MapXtend® là công cụ phát triển các ứng dụng về bản đồ cho các thiết bị cầm tay nối mạng không dây; và MapXtreme® 2004 được dành cho những người lập trình GIS trên nền Microsoft dot Net.

A.2 Hệ phần mềm nguồn mở

Đã từ nhiều năm nay, khái niệm phần mềm nguồn mở (FOSS – Free and Open Source Software) đã bắt đầu có ảnh hưởng sâu rộng trong công nghệ thông tin và lĩnh vực GIS, viễn thám và GPS cũng không vượt ra ngoài phạm vi đó. Khi nói đến phần mềm nào đó có “nguồn mở” thì có nghĩa là mã nguồn của phần mềm đó được truy cập, chuyển tải, sửa đổi và sử dụng miễn phí và tự do. Hầu hết người sử dụng không quan tâm đến mã nguồn mà chỉ sử dụng chương trình phần mềm đã được dịch ra ngôn ngữ máy tính, nhưng đối với những người lập trình muốn tiếp tục phát triển phần mềm thì rất quan trọng. Đơn giản là vì quá trình dịch mã nguồn phần mềm từ ngôn ngữ lập trình (ví dụ C++, Visual C, C#, Visual Basic, v.v.) sang ngôn ngữ máy tính là quá trình một chiều, có nghĩa là khi mã nguồn đã được dịch ra ngôn ngữ máy, ở dạng executable, thì không có cách gì chuyển ngược lại thành mã nguồn.

Bảng A.2 Các dự án phần mềm nguồn mở GIS, GPS và xử lý ảnh

	Địa chỉ	Phần mềm
1.	GRASS	GIS
2.	QGIS, OpenEv, jGRASS	Desktop GIS
3.	GDAL	Data translation
4.	PostgreSQL/PostGIS	Spatially enabled BDMS
5.	PROJ4	Coordinate transform
6.	OSSIM, OpenDragon	Image processing
7.	MapServer, GeoServer, Deegree	Web mapping
8.	GRASSLinks	Web GIS
9.	PrimaGIS	Spatially enabled content management system
10.	Manuscript, CartoWeb, Maplab, Chameleon	Web mapping application development environment
11.	GPSdrive	GPS navigation
12.	GPSTalk	GPS format conversion
13.	Gpsd (DGPS)	GPS TCP Daemon Service
14.	GPSTk	GPS application development toolkit
15.	Geospatial	Clearinghouse Isite Z 39.50

Những ai quan tâm đến công nghệ thông tin đều biết là phần mềm nguồn mở đã và đang tác động mạnh mẽ đến thị trường phần mềm. FireFox có lúc đã làm Microsoft Internet Explorer điêu đứng. Hệ điều hành Linux đang ngày càng hoàn thiện và đã có thể đáp ứng được yêu cầu của những người sử dụng máy tính bình thường, và tất nhiên đang bào mòn thị trường của Microsoft Windows. Hệ điều hành server Apache và ngôn ngữ lập trình web PHP đã được hơn 50% websites trên thế giới sử dụng.

Trong lĩnh vực GIS thì hiện nay phần mềm GRASS là một trong nhiều chương trình nguồn mở đang ngày càng được sử dụng. Bảng A.2 liệt kê một số dự án phát triển các phần mềm liên quan đến GIS, GPS và xử lý ảnh viễn thám, trong đó dự án MapServer sẽ được trình bày chi tiết trong Phụ lục B. Bạn đọc có thể tham khảo các website ở Bảng A.3 về các phần mềm miễn phí này cũng như về các vấn đề liên quan.

Bảng A.3 Các website về phần mềm nguồn mở GIS và về hội nghị Web

Địa chỉ	Nội dung
www.freegis.org	Phần mềm GRASS
www.mapserverfoundation.org	Phần mềm Mapserver
www.ing.unitn.it/~grass/conferences/GRASS_2002/	Bàn luận về phần mềm nguồn mở (FOSS) GRASS 2002
www.omsug.ca/osgis2004/	Hội nghị Web về FOSS GIS năm 2004
www.gisws.media.osaka-cu.ac.jp/grass04/	Danh mục các FOSS
www.mapserver.gis.umn.edu/community/conferences/mum3	Hội nghị Web về FOSS 2005
www.foss4g2006.org	Hội nghị Web về FOSS 2006 đang diễn ra

Phụ lục B

XU HƯỚNG PHÁT TRIỂN CỦA GIS

Suốt ba thập kỷ qua GIS đã được ứng dụng ngày càng rộng rãi và với sự xuất hiện của Internet, Viễn thám và Hệ thống định vị toàn cầu (Global Positioning Systems, GPS) GIS đang có một triển vọng phát triển lên tầm cao mới. Phụ lục này sẽ nhìn nhận GIS từ góc độ bản đồ và giới thiệu xu hướng phát triển mới của GIS khi kết hợp với Internet, GPS và các thiết bị Mobile kết nối không dây (wireless) với mong muốn làm cho người đọc cảm nhận được một phần những gì đang diễn ra trên thế giới trong ngành GIS.

Đã từ hàng ngàn năm nay, bản đồ đóng một vai trò rất quan trọng trong mọi hoạt động của loài người. Năm 1961, một trong những bản đồ cổ nhất đã được phát hiện khi khai quật khảo cổ ở Thổ Nhĩ Kỳ. Khu vực Catalhöyük bao gồm một tụ điểm dân cư tồn tại vào khoảng năm 6500 trước Công nguyên (CN). Bản đồ được tìm thấy vẽ trên tấm gốm thể hiện mặt nằm ngang của một tụ điểm dân cư và bao gồm cả một miệng núi lửa đang hoạt động. Mặc dù không được vẽ theo một tỷ lệ nào đó nhưng rõ ràng nó biểu thị được phương sắp xếp của các công trình. Hình B1. là một phần in lại của bản đồ cổ đại đó.



Hình B.1 Sơ đồ làng Catalhöyük (năm 6500 trước CN),

Sơ đồ có nền là hình một núi lửa đang hoạt động (Nguồn: © John Brock, Measure & Map magazine, 2001)

Bản đồ đó được ước tính tồn tại 6500 năm trước CN cũng có nghĩa là loài người đã bắt đầu vẽ bản đồ trước khi họ bắt đầu học viết. Tác dụng của bản đồ thể hiện ở chỗ nó cung cấp cho người sử dụng một lượng thông tin lớn một cách trực quan rõ ràng và cô đọng.

Từ những bản đồ đầu tiên của loài người nó đã chỉ rõ vị trí không gian của các vật thể. Tuy vậy bản đồ giấy không thể bao gồm nhiều dữ liệu phi không gian và không thể thể hiện đầy đủ thực tế của một vùng. Thế giới thực tế bao gồm nhiều vật thể có quan hệ địa lý với nhau. Một bản đồ là một mô hình đơn giản hóa của thế giới thực tại, và chủ yếu nó chỉ thể hiện vật thể nằm trên mặt phẳng của mặt đất.

Thời đại hoàng kim của bản đồ giấy và cũng là của những nhà trắc địa ở Châu Âu, là thời kỳ bùng nổ của những thám hiểm xuyên lục địa. Bản đồ là dụng cụ không thể thiếu khi người ta muốn đi đến những vùng đất mới. Những công cụ kèm theo là com pa, bàn tính và đồng hồ. Bản đồ số, hệ thống thông tin địa lý (GISs), và các thiết bị xác định vị trí là biểu hiện của thời đại hoàng kim mới của bản đồ và các nhà trắc địa. Những công cụ tạo đà phát triển của thời đại này là máy tính, Internet, và Hệ thống định vị toàn cầu (GPS).

Trong thế giới hiện đại, việc truyền đạt thông tin rõ ràng và ấn tượng là một yếu tố quyết định trong mọi lĩnh vực hoạt động, dù là thương mại, khoa học kỹ thuật hay chính trị. Và hầu như mọi thông tin dù là với dữ liệu dân số, bầu cử, hay môi trường thì đều ít nhiều mang tính phân bố không gian. Tương tự như vai trò của đồ thị trong toán học, bản đồ có thể biểu thị vật thể rõ ràng và dễ hiểu hơn nhiều so với bất kỳ một dạng bảng số liệu nào bởi vì người quan sát bản đồ có thể dễ dàng liên hệ thông tin với vị trí địa lý một cách trực giác.

Tuy vậy bản đồ giấy vẫn còn rất nhiều nhược điểm như sử dụng không thuận tiện và rất khó và tốn kém để cập nhật mới. Bản đồ giấy khi đến tay người sử dụng thì bao giờ cũng đã lạc hậu so với thực tế. Nói cách khác thì bản đồ giấy chỉ lưu trữ được dữ liệu trong quá khứ và không thể có tính thời sự.

Bản đồ số cho phép những khả năng mới và một trong những yếu tố quan trọng là nó có thể thay đổi dễ dàng và nhanh chóng (tính động) nên nó có thể biểu thị một phần thế giới thực cả về mặt thời gian. Sự xuất hiện và phổ cập nhanh chóng của Internet càng làm cho khả năng thể hiện thực của bản đồ số nâng cao. Tuy vậy đặc tính động thái này của bản đồ số làm cho công việc thu thập và quản trị dữ liệu trở nên tốn kém vượt bậc.

Tuy vậy, như đã đề cập đến trong Chương một, GIS đã trở thành một tổ hợp công cụ đặc biệt hiệu quả để quản lý mọi hoạt động của con người và mọi hiện tượng của tự nhiên bởi vì nó khả năng tổng hợp phân tích đồng thời những dạng dữ liệu phức tạp và trình bày kết quả cả ở dạng trực quan cho cả những người làm công tác nghiên cứu khoa học kỹ thuật và khoa học xã hội, những nhà hoạch định kế hoạch và cả công chúng bình thường. Hệ thống GIS kết hợp mọi dữ liệu từ những đo lường vật lý, phân tích hóa học và mô hình máy tính trong một hệ thống đồng nhất nên có thể bảo đảm một sự phát triển bền vững. Tóm lại GIS giúp chúng ta hiểu rõ hơn các quá trình tự nhiên đã và đang xảy ra, cả về tầm vĩ mô và tầm vi mô. GIS góp phần làm cho các hoạt động của loài người ngày càng trở nên có hiệu suất cao hơn, có giá thành hạ hơn, và làm cho các quá trình ra quyết định của con người hợp lý hơn. GIS trình bày các kết quả ở dạng trực quan đồ họa nên nó góp phần tăng cường trao đổi thông tin, sự hiểu biết lẫn nhau và tinh thần hợp tác giữa các cộng đồng loài người.

Đặc biệt trong thời gian vài năm gần đây, sự phát triển nhanh chóng của các thiết bị mobile đã tạo nên động lực mới cho GIS. Tuy vậy các hệ thống GIS hiện tại vẫn chưa có khả năng đáp ứng hiệu quả những yêu cầu mới này. Ta xét một hệ thống dẫn đường GPS trong ô tô có thể chỉ ra vị trí hiện tại của ô tô và hướng dẫn người lái ô tô đến địa điểm theo yêu cầu nhưng do dữ liệu có sẵn trong hệ thống vẫn ở dạng tĩnh, rất khó khăn để có thể được cập nhật liên tục, nên hệ thống dẫn đường hiện nay chưa thể coi là có tính thực sự thời sự (real-time information). Những dịch vụ loại này có thể là dẫn đường để tránh ùn tắc, tìm trạm xăng có giá dịch vụ rẻ nhất.

Tương lai không xa, khi mà giá thành và tốc độ đường truyền internet, kể cả mạng không dây, được cải tiến vượt bậc thì khả năng ứng dụng GIS trên cơ sở web sẽ có thể lại bùng nổ. Các công nghệ Mobile như WiFi, và 2.5 và 3G cellular, sẽ đem lại khả năng mới cho GIS.

Một trong các khả năng đó là mở rộng đưa các chức năng của hệ thống GIS từ những máy tính để bàn hay mạng cục bộ LAN lên mạng Internet. Các chức năng của sự kết nối hệ thống GIS và mạng Web bao gồm việc phân phối dữ liệu không gian (bản đồ), hiển thị bản đồ tĩnh và bản đồ động, hỏi đáp và phân tích GIS trên Web.

Theo các nhiều nhà chuyên môn về GIS trên thế giới thì chúng ta đang chứng kiến một hiện tượng mới về GIS: phát triển GIS trên mạng Web và sẽ trở thành một dạng phân bố GIS toàn cầu.

Thuật ngữ chuyên môn mới của hiện tượng đó người ta gọi là GeoWeb; một mạng lưới rộng rãi toàn cầu về phân phối, chia sẻ dữ liệu, hợp tác phát triển và vận hành đồng thời GIS theo thời gian thực (real time) trên toàn thế giới. Các hệ thống GIS riêng rẽ trên thế giới sẽ ngày càng kết nối với nhau ở mức độ chia sẻ về dữ liệu và khả năng vận hành đến mức ta có thể gọi nó là hệ thống toàn cầu của những hệ thống. Tất nhiên sẽ xuất hiện vô vàn vấn đề cần phải giải quyết: vấn đề tiêu chuẩn hóa mô hình số liệu, vấn đề tạo cổng web (web portals), và vấn đề liên quan đến tinh thần cởi mở của con người, vấn đề giữa dữ liệu tư nhân và dữ liệu nhà nước, v.v.

Chúng ta đang bước vào kỷ nguyên của xã hội thông tin với ngày càng nhận được và phải xử lý khối lượng thông tin càng lớn hơn. Ví dụ chỉ trong vòng vài năm tới ta sẽ thấy số lượng vệ tinh nhân tạo được phóng lên quỹ đạo trái đất, tăng lên hàng trăm lần, cũng như khối lượng dữ liệu ngày càng to lớn từ GPS, từ các hệ thống tự động thu thập dữ liệu (dataloggers) và số lượng website và web portals. Yêu cầu xử lý dữ liệu có tính không gian sẽ tăng lên không ngừng và chần chừ GeoWeb hay khái niệm tương tự sẽ phát triển nhanh chóng như bất kỳ hiện tượng nào có động lực yêu cầu của số đông nhân loại.

Khả năng vận hành đồng thời (interoperability) là yếu tố cơ bản của GeoWeb. Và để khả năng này trở thành hiện thực trên toàn cầu thì cần có một chương trình chuyển đổi các dạng lưu trữ dữ liệu, hệ quy chiếu tọa độ, tỷ lệ và sơ đồ vào trong những CSDL thống nhất, dễ tương thích được với mọi hệ thống. Công nghệ chuyển đổi này chính là cấu trúc thống nhất của mạng ảo (cyber structure) và phương pháp tính mạng (grid computing). Các hệ máy tính bao gồm nhiều bộ xử lý chạy đồng thời (multicore processor) cho ta thấy sức mạnh tính toán đang tăng nhanh, các băng tải internet cũng không ngừng được tăng rộng thêm, các thiết bị lưu trữ ngày càng có khả năng lưu trữ lớn. Bên cạnh đó, các tiêu chuẩn và công nghệ về Web như Extensible Markup Language (XML), và Simple Object Extensible Protocol (SOAP), sẽ ngày càng trở nên quan trọng đối với Web.

Người ta cho rằng ngày càng sẽ có nhiều thiết bị di động (mobile) truy cập GeoWeb, và sẽ cần thiết làm cho hệ GeoWeb trở nên real-time hơn. Các phần mềm GIS sẽ ngày càng có tính dịch vụ nhiều hơn. Những công nghệ GIS hiện đại có thể sẽ thiên về các chức năng dịch vụ trên Web về dữ liệu không gian qua việc nâng cao khả năng cho các server về việc nhân dữ liệu và cập nhập dữ liệu qua đường Web.

Để ví dụ về xu hướng phát triển này ta nghiên cứu ứng dụng mới nhất Google Earth, được coi là sự kiện GIS và Viễn thám của năm 2005, và MapServer, một phần mềm nguồn mở đang có chiều hướng phát triển mạnh, nhằm tạo ra các ứng dụng GIS trên nền Web.

Google Earth

Google Earth (www.earth.google.com) là sản phẩm bản đồ từ việc xử lý ảnh vệ tinh của NASA có kết hợp với mô hình địa hình và những ngôi nhà lớn theo dạng đồ họa ba chiều.

Do có khả năng tìm kiếm nên người sử dụng có thể “bay” đến mọi địa điểm trên thế giới để nhìn thấy chi tiết đến mức các đường phố. Dịch vụ này cho phép người sử dụng truy cập miễn phí và dễ dàng kho dữ liệu ảnh viễn thám đồ sộ của NASA mà người bình thường cho đến nay không thể biết được. Khả năng thể hiện hình ảnh ba chiều sự thay đổi mọi địa hình trên trái đất và chi tiết đến mức độ đường phố mọi địa điểm trên thế giới, kể cả các địa điểm nhạy cảm về quân sự và chính trị, trên bất cứ máy tính của một người bình thường có kết nối Internet nào thực sự đã gây bất ngờ và sự khâm phục của những người làm công tác GIS và RS. Tất nhiên đồng thời cũng tạo nên nỗi lo âu trong chính phủ nhiều nước trên thế giới (Ấn độ, Australia, Thái lan, Hà lan, Hàn quốc và Sri Lanka). Dù sao thì Google Earth đã và đang tồn tại hiển nhiên và sẽ trở thành một ví dụ sống động về khả năng mạnh mẽ của GIS khi nó được phát triển lên mạng Internet.



Hình B.2 Màn hình mở đầu của Google Earth

MapServer

Công cụ để kết nối GIS với web cũng đã được các công ty phát triển phần mềm GIS bổ sung vào các sản phẩm của họ. ESRI, Intergraph, MapInfo đều đã có những module để biểu thị và phân tích bản đồ trên nền web. Cộng đồng những người phát triển phần mềm nguồn mở (đã được trình bày trong Phụ lục A) cũng tạo ra nhiều phần mềm tương tự nhưng đặc biệt là họ cho phép cộng đồng GIS trên thế giới sử dụng miễn phí, đồng thời với việc cho phép chuyển tải tự do mã nguồn, phần mềm có khả năng tập hợp được trí tuệ lập trình của hàng ngàn chuyên viên lập trình trên thế giới.

Một trong các phần mềm nguồn mở thể hiện bản đồ và truy vấn GIS trên Web được công nhận đang có triển vọng là MapServer (<http://mapserver.gis.umn.edu>). Đây là công cụ hiển thị bản đồ trên môi trường web tương tự như là CGI (Common Gateway Interface) script hay là một ứng dụng riêng biệt qua giao diện API (Application Program Interface) với việc sử dụng công cụ lập trình. Nói cách khác thì MapServer là môi trường phát triển nguồn mở để xây dựng các ứng dụng GIS có khả năng kết nối với Web. Trường Đại học Minnesota thành lập và phát triển MapServer có trợ giúp của NASA và Cục Quản lý Tài nguyên Thiên nhiên Minnesota (US Department of Natural Resources). Hiện nay MapServer được cài đặt và sử dụng để phát triển các ứng dụng bản đồ trên nền Web ở khoảng 20 tổ chức trên thế giới.

Sau đây là một số ứng dụng của được phát triển thông qua việc sử dụng MapServer API. (hoặc qua CGI như đã trình bày ở trên) Khi kết hợp với phần mềm nguồn mở về CSDL MySQL và ngôn ngữ lập trình cũng nguồn mở như là PHP, những ứng dụng này có thể dễ dàng được trang bị thêm nhiều công cụ hữu hiệu mà chỉ riêng với MapServer thì rất khó thực hiện. Những ứng dụng giới thiệu ở đây không có gì là mới mẻ nhưng chúng chứng tỏ những tác dụng mà MapServer có thể thực hiện được.

Công cụ trợ giúp phân phối bất động sản (Real Estate Sales Tool) — Khi liên kết tọa độ kinh vĩ độ (lat/long coordinates) cho mỗi khu đất hay nhà cửa vào phần mềm hiển thị hàng hóa (multiple listing service) thì người ta có thể tạo nên danh mục catalog

có tính không gian đồ họa (ví dụ việc nhấn chuột để tìm kiếm, việc các cửa sổ thông tin hiện lên khi con trỏ đi qua các điểm trọng tâm).

Theo dõi tiến trình có tính thời gian thực (Real-Time Track and Trace) — Khi dùng thiết bị GPS để xác định vị trí tức thời (thời gian thực) và chuyển giao dữ liệu vị trí đến thiết bị di động có công nghệ cellular 2.5 và 3G, MapServer có thể thành lập ứng dụng chỉ đường tức thời. Cơ sở dữ liệu nguồn mở MySQL là một CSDL phục vụ tốt cho việc lưu trữ loại dữ liệu địa lý này.

Cố vấn chỉ đường tránh tắc nghẽn giao thông tức thời (Real-Time Traffic Advisories and Congestion Avoidance) — Với việc thu thập dữ liệu mật độ các luồng giao thông thông qua cập nhập tọa độ GPS, địa chỉ đường phố và các điểm nút giao thông— MapServer có thể tính toán mức độ giao thông thực tế theo thời gian thực, thể hiện các thông tin đó lên web (để các thiết bị di động có thể cập nhập), và đưa ra chỉ dẫn về các luồng đường hợp lý.

Bảng Từ viết tắt

AGIS – Automatic Geographic Information Systems

AID – Agency for International Development

API – Application Program Interface

ASP – Active Server Pages

B/W – Black and White

BIL – Band Interleaved by Line

BIP – Band Interleaved by Pixel

BPI – Bits per inch

BPS – Bits per second

BSQ – Band Sequential

CAD – Computer Added Design

CAM – Computer Added Manufacturer

CBD – Central Business District

CCRS – Canada Center for Remote Sensing

CCT – Computer Compatible Tape

CGI – Common Gateway Interface

CIR – Color Infrared Film

CPU – Central Processing Unit

CRT – Cathode Ray Tube

CZCS – Coastal Zone Color Scanner

D/A – Digital to Analog Converter

DBMS – Database Management Systems

DEM – Digital Elevation Model

DGPS – Differential Global Positioning Systems

DIME – Dual Independent Map Encoding

DLG – Digital Line Graph

DMA – Defense Mapping Agency

DN – Digital Number

EPA – Environmental Protection Agency

ESA – European Space agency

ESRI – Environmental Systems Research Institute

ESRI – Environmental Systems Research Institute

FL – Focal Length

FOSS	– Free and Open Source Software
FOV	– Field of View
GBIS	– Geo-Based Information System
GIS	– Geographical Information Systems
GLONASS	– Global Navigation Satellite System
GPS	– Global Positioning Systems
HDDT	– High-Density Digital Tape
HOM	– Hotine Oblique Mercator
HTML	– HyperText Markup Language
HTTP	– HyperText Transfer Protocol
IFOV	– Instantaneous Field of View
IGIS	– Integrated Geographic Information Systems
IRG	– Inter-record Gap
ISP	– Internet Service Provider
LANDSAT	– Land Remote Sensing Satellite
LIS	– Land Information Systems
MSS	– National Aeronautics and Space Administration
NASA	– National Aeronautics and Space Administration
NOAA	– National Oceanic and Atmospheric Administration
NSF	– National Science Foundation
OGC	– Open Geographic Information System Consortium
OS	– Operating System
PI	– Photo Interpretation or Photo Interpreter
QA	– Quality Assurance
SAR	– Synthetic Aperture Radar
SCA	– Suitability/Capability Analysis
SNR or S/N	– Signal to Noise Ratio
SPOT	– Systeme Probatoire d’Observation de la Terre
TM	– Thematic Mapper
UNESCO	– United Educational, Scientific and Cultural Organization
USGS	– United States Geological Survey
UTM	– Universal Transverse Mercator
UV	– Ultraviolet
VICAR	– Video Information Communication and Retrieval
ZTS	– Zoom Transfer Scope

Danh mục từ vựng

Analog – Sự biểu thị liên tục của các đại lượng vật lý (điện thế vôn, dòng điện ampe, độ sáng) trái ngược với sự biểu thị rời rạc hay còn gọi là “Biểu thị số (digital)”.

Ancillary Data – Dữ liệu bổ sung.

ASCII (American Standard Code for Information Exchange) – Bộ mã tiêu chuẩn Hoa kỳ cho việc trao đổi thông tin – Bộ mã cho các ký tự chữ cái phổ biến nhất để biểu thị dữ liệu. Mỗi một chữ cái hay số được biểu thị bằng một số từ 0 đến 127. Công ty IBM mở rộng mã ASCII lên thành 255 để biểu thị các ký tự đồ họa và các ngôn ngữ khác tiếng Anh.

Bàn số hóa (Digitizer) – Thiết bị để chuyển hóa các thông tin vật thể bản đồ vào máy tính. Chúng bao gồm một con chuột hay bút trỏ cảm ứng (stylus or cursor), một bàn điện tử (a tablet), và các chip điện tử.

Băng từ (Magnetic tape) – Băng bằng nhựa plastic hẹp và mỏng, phủ một lớp vật liệu từ, được dùng để lưu trữ các tín hiệu dưới dạng từ.

BASIC (Beginner’s All-Purpose Symbolic Instruction Code) – Ngôn ngữ lập trình để phát triển các chương trình theo cách của đối thoại; là ngôn ngữ dễ sử dụng trong việc lập trình cho máy tính. BASICA là một dạng cải tiến của BASIC cho máy vi tính IBM và GW-BASIC thường được sử dụng cho máy vi tính nhãn hiệu khác IBM. Sau đó Microsoft phát triển thành Visual BASIC và Visual BASIC dot NET.

Batch file – Tập tin có chứa một tập hợp các dòng lệnh để hệ điều hành máy tính thực hiện tự động. Các tập tin này phải có đuôi “.bat” để hệ điều hành có thể nhận biết.

Baud – Đơn vị đo tốc độ truyền tải tín hiệu dữ liệu.

BIOS (Basic Input-Output System) – hoạt động như là một thành phần trung gian giữa phần cứng của máy tính và các phần mềm, đồng thời cung cấp các dịch vụ như là đọc và ghi vào bộ nhớ cũng như các đĩa cứng của máy tính, v.v. Chương trình này được các thiết bị đặc biệt cài đặt vào các chip ROM. Thường thì các chương trình BIOS này không bị thay đổi trong suốt quá trình sử dụng một máy tính.

Bít (Binary digit) – Đơn vị thông tin bé nhất của máy tính; bit được viết tắt từ BInary digiT (ký số nhị phân).

Bit map – Hình ảnh được biểu thị bằng ma trận các ô vuông sơ đẳng (pixels).

Bộ dữ liệu (Data set) – Một tập hợp những giá trị thu thập được khi phân tích một mẫu nghiên cứu; một nhóm những dữ liệu thu thập được dưới những điều kiện tương tự và như vậy có thể được phân tích cùng một lúc.

Bộ lưu trữ bằng trống từ (Magnetic-drum storage) – Bộ lưu trữ bao gồm một trống hình trụ quay được phủ bằng một lớp có từ tính. Dữ liệu dưới dạng nhị phân được lưu trữ bằng các chấm từ li ti sắp xếp theo rãnh ghi xung quanh trống. Đầu từ ghi và đọc được kết hợp với mỗi rãnh ghi để cho có thể chọn được các rãnh ghi mong muốn. Khi trống quay, dữ liệu trên các rãnh được đọc hoặc ghi một cách tuần tự.

Bộ lưu trữ đệm (Buffer storage): 1- Thiết bị có thể tạm thời lưu trữ thông tin trong một quá trình truyền thông tin. 2- Bộ lưu trữ thứ cấp được dùng riêng cho truyền dữ liệu giữa bộ lưu trữ bên trong và bộ lưu trữ bên ngoài. 3- Bộ lưu trữ được dùng để tạo điều kiện cho truyền dữ liệu giữa các đôi thiết bị dự trữ có tốc độ đầu vào và đầu ra không

phù hợp. 3- Thiết bị đầu vào trong đó thông tin được tập hợp từ bộ lưu trữ bên ngoài hoặc thứ cấp và trữ lại sẵn sàng để chuyển đến bộ lưu trữ bên trong.

Bộ nhớ (Memory) – Một tổ chức của đơn vị lưu trữ, chủ yếu dùng để truy cập thông tin và dữ liệu. Các loại bộ nhớ bao gồm trống, đĩa, lõi hoặc các bộ nhớ bằng vi mạch. Những bộ lưu trữ truy cập cực nhanh và tại đó thực hiện các lệnh và dữ liệu được sử dụng được gọi là bộ nhớ chính.

Bộ nhớ Cache (Cache memory) – Một vùng đặc biệt của bộ nhớ có tốc độ cao lưu giữ các thông tin để máy tính có thể truy xuất nhanh.

Bộ nhớ chính (Main memory) – Thông thường là bộ lưu trữ nhanh nhất của một máy tính và là nơi thực hiện các lệnh.

Bộ nhớ động (Dynamic memory) – Lưu trữ dữ liệu trên một thiết bị hoặc bằng cách khác cho phép dữ liệu có thể dịch chuyển hoặc thay đổi theo thời gian.

Bus – Là đường dẫn thông tin giữa tấm motherboard và các tấm adapter như là tấm video display hoặc tấm kiểm soát các phần cứng phụ kiện (drive controller cards).

Byte – Nhóm các số nhị phân kế liên nhau (thường là 8), tạo thành đơn vị dữ liệu cơ sở của máy tính. Do được lưu trữ tương đương một ký tự nên byte cũng là đơn vị cơ sở để đo sức chứa của máy tính.

C.G. – Đồ họa máy tính (C.G. Computer Graphics). Vẽ đồ họa bằng máy tính.

CAD/CAM (Computer Added Design/Manufacture) – Thiết kế có máy tính hỗ trợ/Chế tạo có máy tính hỗ trợ. Sự hoạt động của các máy tính trong thiết kế và sản xuất sản phẩm.

Cập nhật (Update) – Sửa đổi một tập tin chủ bằng các thông tin hiện có dựa theo các thủ tục đã được xác định.

Cặp tọa độ (Coordinates) – Một cặp gồm hai giá trị nhằm xác định vị trí của một điểm, ví dụ như một trạm đo trên bản đồ. Các cặp tọa độ bao giờ cũng dựa trên một hệ tọa độ nào đó. Hệ tọa độ chota biết tỷ lệ và đơn vị đo của các giá trị tọa độ, ví dụ hệ tọa độ hai chiều Cartesian (XY) hoặc hệ hình cầu (kinh vĩ tuyến).

Chất lượng dữ liệu (Data quality) – Toàn bộ thành phần và những đặc điểm của dữ liệu có khả năng thỏa mãn các mục đích đã được xác định; ước tính tổng các mức độ thỏa mãn các yêu cầu đối với dữ liệu.

Chế bản văn phòng (DTP, Desktop Publishing) – Sử dụng máy tính, máy in và phần mềm chuyên dụng để thiết kế, sắp xếp trang và in một tài liệu.

Chỉ số chất lượng của dữ liệu (Data quality indicators) – Các chỉ số thống kê và các diễn tả chất lượng được sử dụng để đánh giá mức độ đạt yêu cầu sử dụng của dữ liệu. Các chỉ số chính là độ khách quan, độ chính xác, độ tương thích, mức độ đầy đủ và mức độ đại diện.

Chương trình đồ họa (Graphics program) – Phần mềm để tạo và xử lý đồ họa như là các chương trình đồ họa, vẽ đồ thị, biểu đồ và CAD.

Chương trình nguồn (Source program) – Chương trình máy tính được viết bằng ngôn ngữ lập trình trước khi được dịch sang ngôn ngữ máy.

Chuyển đổi hệ tọa độ (Coordinate conversion) – Chuyển đổi tọa độ của các vật thể từ hệ tọa độ này sang hệ tọa độ khác, ví dụ từ hệ kinh tuyến – vĩ tuyến sang hệ UTM (Universal Transverse Mercator).

Cơ sở dữ liệu quan hệ (Relational Database) – Tập hợp các dữ liệu trong đó các mục và trường đều có liên quan với nhau.

CPU, Đơn vị xử lý trung tâm (Control Processing Unit) – Một nhóm các mạch thực hiện các chức năng cơ bản của máy tính bao gồm 3 phần: đơn vị điều khiển, đơn vị số học và logic, và đơn vị xuất/nhập.

Dải sóng (Band) – Một tập hợp nhất định các loại sóng điện từ có các bước sóng khác nhau.

Device driver – Chương trình máy tính cài đặt để thiết lập các thông số giao tiếp giữa máy tính và một phần mềm hay một thiết bị phụ cận như ổ đĩa mềm, máy in, máy vẽ, bàn số hóa hay máy quét.

Độ chính xác (Accuracy) – Là mức độ phù hợp giữa các giá trị quan trắc được và các giá trị chuẩn mực định trước theo tiêu chuẩn nào đó. Các chênh lệch giá trị có thể có nguồn gốc từ sai số ngẫu nhiên (precision) hoặc từ sai số hệ thống (bias) xảy ra trong suốt quá trình lấy mẫu và phân tích.

Độ nhạy cảm (Sensitivity) – Mức độ biến động bé nhất mà thiết bị đo đạc có thể phân biệt được.

Đối số (Argument): 1– Thừa số tham chiếu đã biết trước, cần thiết để tìm mục muốn tìm (hàm số) trong bảng. 2– Một biến số mà giá trị của nó quyết định giá trị của một hàm số. Đối số của một hàm được liệt kê trong ngoặc đơn sau tên của hàm số, khi mà hàm số đó được sử dụng. Việc tính toán theo định nghĩa của hàm số sẽ bị ảnh hưởng bởi các biến số được dùng làm đối số.

DOS – Hệ điều hành (Disk Operating System) – Phần mềm hệ điều hành, nó điều khiển đĩa và truy cập các tập tin.

Dữ liệu số (Digital data) – Thông tin được thể hiện qua một bộ mã bao gồm một chuỗi các phân tử rời rạc.

Đường cong hiệu chỉnh (Calibration curve) – Đồ thị đường cong quan hệ giữa số đo thực tế và số liệu do bộ phận cảm ứng cung cấp.

Đường đồng mức (Contour liné) – Là đường cong trên bản đồ địa hình mà mọi điểm trên một đường có cùng một giá trị độ cao. Các đường đồng mức biểu thị cho dạng của một mặt cong.

FORTRAN (FORmula TRANslator) – Ngôn ngữ lập trình thông dụng nhất được xây dựng cho các chương trình ứng dụng khoa học và kỹ thuật.

Giao diện (Interface) – Ranh giới chung giữa các hệ thống xử lý dữ liệu tự động hoặc các phần của một hệ thống. Trong truyền thông và hệ thống cơ sở dữ liệu, nó có thể bao gồm mã, khuôn thức, tốc độ hoặc các thay đổi nếu cần thiết.

Giao diện với người sử dụng (User interface) – Thành phần của một phần mềm chuyên về truyền tải thông tin giữa người sử dụng và máy tính.

Giao thức (Protocol) – Một tập các nguyên tắc, mã và tín hiệu chuẩn mực để trao đổi thông tin giữa các hệ thống. Trong vận hành các thiết bị phòng thí nghiệm là quy định các bước tiến hành lấy mẫu và phân tích mẫu mà người thực hiện phải triệt để tuân thủ theo. Giao thức còn có thể gọi là định ước hay định chuẩn.

Gigabyte – Số đo bằng khoảng một tỷ (10^6) byte, viết tắt là GB.

GIS (Geographic Information Systems) – Hệ thống thông tin địa lý được sử dụng để nhập, quản lý, xử lý và phân tích dữ liệu không gian cùng với các thông tin thuộc tính trên bản đồ.

GML – Geography Markup Language. Tiêu chuẩn về ngôn ngữ truyền thông tin dữ liệu địa lý trên Web.

Hệ điều hành (Operating system) – Phần mềm cơ bản điều khiển hoạt động của phần cứng, và quản lý các tập dữ liệu mà không cần con người tác động.

Hệ số tương quan (Correlation coefficient) – Một số có giá trị giữa -1 và +1 để biểu thị mức độ tương quan tuyến tính giữa hai biến số. Hệ số này có giá trị càng gần với -1 hoặc +1 thì hai biến số đó càng có quan hệ tuyến tính hơn (mức tương quan cao hơn). Giá trị gần với zero có nghĩa là hầu như không có tương quan gì giữa hai biến số.

Hệ thống chủ động (Active System) – Trong viễn thám, một hệ thống chủ động sẽ có trang bị nguồn phát điện từ trường hay sóng bức xạ và hệ thống dò cũng sẽ thu nhận các tín hiệu do các nguồn đó bị các vận thể quan trắc phản hồi hay phát tán.

Hệ thống Thông tin Địa lý Tự động (Automatic GIS) – Một HTTĐL hoàn toàn dựa trên kỹ thuật máy tính.

Hệ thống tọa độ Cartesian (Cartesian coordinate system) – Hệ tọa độ trong mặt phẳng, trong đó hai trục X và Y vuông góc với nhau và có cùng một tỷ lệ.

HTML (HyperText Markup Language) – Ngôn ngữ để truyền tải dữ liệu trên Internet đang được hầu hết các trang Web sử dụng hiện nay.

Kênh tuyến (Bus) – Kết nối truyền thông bao gồm một bộ các đường dẫn hoặc dây nối các phần khác nhau của hệ thống phần cứng máy tính, và qua đó dữ liệu được truyền và nhận bằng các mạch khác nhau trong hệ thống.

Khai thác dữ liệu (Data mining) – Truy cập và phân tích dữ liệu thuộc về một CSDL khác để xem xét xu thế và các đặc tính của dữ liệu.

Ký tự (Character) – Một ký tự của tập hợp các ký hiệu cơ bản chẳng hạn các ký tự của bàn phím máy đánh chữ. Các ký hiệu thường bao gồm các số thập phân từ 0 đến 9, các chữ cái từ A đến Z, các dấu câu, các ký hiệu động tác và các ký hiệu đơn khác mà máy tính có thể đọc, ghi và lưu trữ được.

LAN (Local Area Network), Mạng cục bộ – Mạng máy tính trong đó các thiết bị đầu cuối và các thiết bị khác ở gần nhau được nối với nhau bằng cáp dẫn.

Lệnh (Command) – Một xung điện tử, tín hiệu hoặc một tập hợp các tín hiệu để bắt đầu, kết thúc hoặc tiếp tục các thao tác.

Lỗi (Bug): 1- Tất cả các lỗi về điện, điện tử hay cơ học làm ảnh hưởng đến hoạt động của máy tính. Cũng có thể là những sai sót trong chương trình; 2- Lỗi trong thiết kế một chương trình hoặc một máy tính.

Lớp bản đồ (Layer) – Trong GIS, đây là dữ liệu cho một thông số duy nhất của một vật thể nhất định ví dụ như mạng giao thông hay mạng đường tiêu nước. Tương tự như khái niệm Coverage và Theme.

Lưới kê ô (Grid) – Một lưới ô là biểu thị toán học của một mặt cong được thành lập trên cơ sở số liệu gán cho các điểm. Có nhiều thuật toán để tạo mặt cong từ một tập hợp điểm duy nhất và kết quả có thể là các mặt cong khác nhau.

Lưu đồ (Flowchart) – Một biểu thị đồ họa về các thao tác liên tiếp của máy, bằng các đường các ký hiệu để mô tả các hoạt động như tính toán, thay thế, so sánh, nhảy, sao chép, đọc ghi, v.v.

Lưu trữ (Storage) – Từ chung chỉ các loại thiết bị có khả năng cất giữ thông tin.

Mã hóa (Code) – Một hệ thống các ký hiệu tiêu biểu cho dữ liệu hoặc lệnh trong máy tính hoặc máy lập bảng.

Mã Macro (Macro code) – Một hệ mã tập hợp một nhóm các lệnh máy tính vào trong một từ mã đơn.

Mảng (Array) – Một chuỗi các vật thể được sắp xếp theo một trình tự nhất định.

Mạng máy tính (Computer Network) – Cấu hình kết nối mật thiết và tinh xảo các máy tính cách rời nhau.

Mẫu tin, Bảng ghi (Record): 1- Một tập hợp của một hay nhiều trường liên tiếp trong chủ thể liên quan, như bảng lương người làm công. 2- Một nhóm các sự kiện liên quan hoặc trường của các thông tin được xem như một đơn vị,

Máy chủ (Host computer) – Máy tính điều khiển chính trong một hệ thống gồm nhiều máy tính phân bố ở các vị trí khác nhau hoặc nhiều người sử dụng.

Máy tính nhân bản (Clone) – Các máy tính có cấu trúc tương thích với máy tính cá nhân IBM. Ví dụ những loại phổ biến này là Compaq, Gateway, và Dell; những loại máy tính nhân bản ít tên tuổi hơn gọi là máy tính không tên (no-name clones).

Máy vẽ (Plotter) – Plotters là thiết bị sử dụng bút vẽ, phun mực, hoặc lực hút điện từ để sản xuất các bản vẽ trên giấy, thường là có màu sắc. Kích thước bản vẽ trên giấy có các khổ thay đổi từ 8.5" x 11" đến 36" x 42" hoặc lớn hơn.

Metadata – Thông tin về dữ liệu (Data about data). Ví dụ các diễn tả về nội dung của một loại dữ liệu là metadata của dữ liệu đó.

MS-DOS (Microsoft Disk Operating System) – Hệ điều hành tiêu chuẩn một người sử dụng, do công ty Microsoft soạn thảo cho máy tính 80 x 86. Đây là hệ điều hành máy tính phổ biến nhất trong những năm 80.

Ngẫu nhiên (Random) – Không có kế hoạch, mục đích, hoặc hình dáng trước.

Ngôn ngữ C – Một loại ngôn ngữ lập trình bậc cao được xây dựng chủ yếu để lập trình các chương trình hệ thống cấu trúc. Trên cơ sở ngôn ngữ C, người ta đã và đang phát triển ngôn ngữ lập trình bậc cao C++ (C plus-plus) và C# (C sharp) tạo thuận lợi cho lập trình dạng đối tượng và mạng internet.

Ngôn ngữ lập trình bậc cao (High-level programming language) – Ngôn ngữ lập trình máy tính dễ học và cho phép người sử dụng viết chương trình bằng cách dùng các từ và lệnh, dễ hiểu và tương tự như các từ tiếng Anh; chương trình phải được dịch ra ngôn ngữ máy thì mới sử dụng được với máy tính..

Ngôn ngữ máy (Machine language) – Một ngôn ngữ được dùng để viết các câu lệnh dưới dạng mà bộ xử lý trung tâm của máy tính có thể nhận biết trực tiếp và thực hiện.

Nhị phân (Binary) – Một hệ thống số, dựa trên cơ sở 2, khi viết chỉ dùng số 0 và số 1, khác với hệ thống số thông dụng sử dụng cơ sở 10 (số thập phân).

Ô vuông sơ đẳng (Pixel) – Một pixel, viết tắt từ chữ PICture ELe ment, là một chấm trong một hình ảnh, nó cũng là đơn vị biểu thị độ phân giải của màn hình.

Phần bù (Complement) – Phần bù của một số với N được xác định bằng cách lấy (N-1) trừ đi tổng chữ số của số đã cho, sau đó cộng thêm một đơn vị vào chữ số có vị trí quan trọng nhất và thực hiện đủ các phép nhớ. Thí dụ phần bù với 2 của số nhị phân 11010 là 00110; phần bù với 10 của số thập phân 456 là 544.

Phần cứng (Hardware) – Các bộ phận cơ, từ tính, điện và điện tử, hoặc các cấu thành vật lý khác của máy tính.

Phần mềm nguồn mở (open source software) – Phần mềm nguồn mở là phần mềm miễn phí, bao gồm cả mã nguồn, từ các nguồn khác nhau kể cả từ các chính phủ và các tổ chức và cá nhân. Các tiêu chuẩn và quy định về phần mềm nguồn mở hiện nay đã được chuẩn hóa và có thể tìm hiểu qua địa chỉ <http://www.opensource.org>.

Phép chiếu (projection) Lambert – Phương pháp biểu thị tọa độ mặt hình cầu (kinh vĩ độ, latitude-longitude) lên trên một mặt phẳng sử dụng một mặt hình trụ song song với trục nối hai cực của trái đất.

Phép tham chiếu (Projecton) – Trong ngành bản đồ phép chiếu là phương pháp biểu thị mặt đất cong ba chiều lên bản đồ mặt phẳng hai chiều. Có nhiều loại phép chiếu khác nhau dựa trên phương cách làm thế nào để áp một mặt phẳng lên mặt cong của quả cầu trái đất. Các loại phép chiếu phổ biến là Mercator, Albers Equal Area, và Lambert Conformal Conic.

Planimeter – Thiết bị để tính diện tích một vùng trên giấy.

Polyline – Một nhóm các đoạn thẳng nối tiếp nhau được biểu thị và xử lý như một thực thể duy nhất.

RAM (Random access memory), Bộ nhớ truy cập ngẫu nhiên – Bộ nhớ cho phép truy cập đến bất kỳ vị trí nào và theo các thứ tự bất kỳ, thông thường dưới dạng các vi mạch. SSố đo của RAM được tính bằng (MB) đối với hệ thống nhỏ và (GB) với hệ thống lớn. Lượng RAM càng lớn thì chương trình càng to và dữ liệu cần nhiều có thể được nạp vào bộ nhớ của máy tính. Xem ROM để biết thêm chi tiết.

Raster – Dữ liệu được biểu thị như một mảng các điểm. Thiết bị hiển thị Raster bao gồm cả máy in và màn hình. Dữ liệu raster khác hẳn với dữ liệu dạng vector, trong đó sử dụng các đoạn thẳng để biểu thị các vật thể hình ảnh.

ROM (Read-Only Memory) – ROM khác với RAM ở điểm phải cần đến thiết bị đặc biệt mới có thể ghi thông tin lên chip ROM được, và khi nó đã được lắp đặt trong máy tính thì máy tính chỉ có thể đọc dữ liệu từ ROM mà không thể ghi bất cứ thông tin nào vào chúng. ROM được sử dụng cho các chương trình cứng, không bao giờ thay đổi như BIOS trên máy tính dạng IBM. Xem RAM để biết thêm chi tiết.

Sai số hệ thống (Systematic error) – Một sai lệch so với số liệu đúng xảy ra liên tục trong quá trình thu thập hay phân tích dữ liệu, chủ yếu do yếu tố con người và do phương pháp không chuẩn mực.

Sai số ngẫu nhiên (Random error) – Độ sai lệch của số liệu quan trắc so với số liệu đúng mà chúng có giá trị không theo một dạng cố định và hoán toàn có tính ngẫu nhiên. Người ta nhận thấy các sai số này thường có phân bố xác suất chuẩn (normal).

Scanner, Máy quét – Thiết bị thuộc loại quang học cho phép chuyển tải liên tục một hình ảnh, bản vẽ, bản đồ và văn bản vào lưu trữ trong máy tính dưới dạng hình ảnh.

SQL – Structured Query Language. Ngôn ngữ lập trình rất được phổ biến để truy cập dữ liệu trong các cơ sở dữ liệu.

Tệp (file): Một tập hợp của các tài liệu có liên quan với nhau và được xem là một đơn vị.

Thang độ đậm nhạt (Gray-scale) – Là thang độ các mức màu đậm nhạt mà một thiết bị hiển thị (màn hình, máy in) có thể tạo ra.

Thiết bị hiển thị (Display) 1– Thiết bị trình bày thông tin hoặc hình ảnh; 2– Sự trình bày cho xem dữ liệu trên màn hình.

Thiết bị ngoại vi (Peripheral) – là bất cứ thiết bị nào không nhất thiết phải lắp đặt để một máy tính có thể hoạt động được, hoặc là thiết bị lắp đặt ngoài vỏ máy tính. Màn hình, máy in, con chuột, máy vẽ, bàn số hóa, máy quét là những ví dụ về thiết bị phụ cận.

Thời gian truy cập (Access Time or Cycle Time) – Lượng thời gian cần thiết cho bộ nhớ hoặc thiết bị nhận hoặc chuyển dữ liệu, tính từ lúc yêu cầu truy tìm thông tin được đưa ra.

Thông tin phản hồi (Feedback) – Sử dụng một phần của toàn bộ đầu ra của một máy, một quá trình hoặc một hệ thống vào đầu vào của giai đoạn khác, như trường hợp áp dụng cho mục đích tự hiệu chỉnh. Những hệ thống hoặc những chương trình thông tin phản hồi như vậy sử dụng quá trình so sánh liên tục giữa đầu vào và đầu ra để tạo lập những sự hiệu chỉnh cần thiết. Hệ thống thông tin phản hồi được xem là một hệ thống tự hiệu chỉnh nếu nó là một vòng lặp khép kín.

Thư mục (Directory) – Bảng mục lục tên và thông tin về các tệp tin trong một thiết bị lưu trữ.

Thuật toán (Algorithm): 1– Thủ tục xác định để tính từng bước nhằm đạt được kết quả đã cho; thường là một thủ tục đã được đơn giản hóa để giải một bài toán phức hợp; cũng là một biểu diễn đầy đủ của một số hữu hạn các phép toán. 2– Một quá trình hay một tập hợp xác định của các quy tắc dẫn đến và đảm bảo kết quả mong muốn từ số liệu đầu vào đã cho. Một chuỗi các công thức và/hoặc các bước đại số/lô gic để tính toán hoặc xác định một bài toán đã cho hay các nguyên tắc xử lý.

Thuộc tính (Attribute) – Thuộc tính của một thực thể. Một thực thể có thể có nhiều thuộc tính.

Tọa độ tương đối (Relative coordinates) – Tọa độ được xác định trên cơ sở gốc tọa độ cục bộ chứ không sử dụng hệ tọa độ toàn cầu.

Trường dữ liệu (Field) – Tập hợp của một hay nhiều ký tự được coi như toàn thể; tập hợp của một hay nhiều cột trong một bảng dữ liệu được dùng một cách thống nhất để ghi các thông tin tương tự.

Tương quan (Correlation) – Mức độ liên hệ giữa hai biến số, có quan hệ với hệ số tương quan (correlation coefficient).

UNIX– Một hệ điều hành chuyên dùng cho các máy tính mini.

Virtual disk – See RAM disk.

WYSIWYG (What You See Is What You Get), Làm gì thấy vậy – Thuật ngữ sử dụng trong phần mềm văn bản, chế phẩm in ấn và cả trong phần mềm tạp trang web, có nghĩa là màn hình sẽ hiển thị tài liệu đúng dạng như bản in/trang web cuối cùng.

Bảng tra thuật ngữ

3

3G cellular, 121

A

Apache, 39, 117

API, 122, 123, 125

Arc/Info, 15, 52, 112, 114

ArcGIS, 15, 115

ArcView, 98, 114, 115

ASCII, 59, 71, 89, 104, 127

B

bản đồ

bản đồ giấy, 5, 37, 72, 82, 88, 104, 108, 119, 120

bản đồ học, 3, 76

bản đồ số, 26, 37, 74, 76, 82, 83, 87, 88, 90, 98, 103, 114, 120

lớp bản đồ, 4, 90, 91, 94, 96, 109

lớp dữ liệu, 30, 42, 54, 96, 98

tỷ lệ bản đồ, 27, 103

bản đồ, 3, 4, 5, 6, 11, 12, 14, 15, 16, 17, 18, 19, 20, 21, 22, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 37, 41, 44, 47, 48, 51, 52, 54, 55, 58, 59, 60, 61, 62, 63, 64, 65, 66, 68, 69, 72, 76, 78, 82

khoảng cách, 7, 30, 33, 66

biểu thị kết quả, 14, 58, 91, 98

binary, 104, 127

bộ phận cảm ứng, 103, 129

byte, 107, 128, 130

C

C#, 116, 131

C++, 116, 131

CAD, 9, 76, 125, 128

CAM, 125, 128

CGI, 122, 123, 125

chất lượng bản đồ, 106

chi phí, 8, 58, 101, 103, 104

chiến lược phát triển, 7

chỉnh sửa dữ liệu, 89, 114

chu vi, 71

chức năng dịch vụ, 122

D

đa giác, 47, 49, 52, 59, 60, 64, 65, 66

đặc tính địa lý, 11, 17, 27, 28, 96

dân tộc học, 3, 21

dạng lưu trữ, 38, 89, 99, 104, 121

đặt câu hỏi, 95

địa hình, 3, 17, 19, 21, 22, 25, 26, 30, 32, 42, 76, 78

địa lý, 2, 3, 6, 7, 11, 12, 14, 16, 19, 20, 21, 22, 25, 26, 29, 32, 33, 41, 42, 47, 58, 68, 72, 86

địa lý học, 2

điểm, 64, 69

điểm, 7, 10, 18, 25, 26, 27, 29, 30, 32, 33, 34, 35, 44, 47, 49, 51, 53, 54, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 71, 76, 78, 79, 92, 93

điểm quan trắc, 103

điểm trung tâm, 76, 78

diện tích, 7, 16, 27, 28, 29, 33, 35, 41, 44, 45, 48, 77, 95

độ biến động, 105, 106, 107, 109, 110, 129

đo đạc thực địa, 105, 106

đô thị hóa, 88

đối tượng, 4, 6, 9, 10, 25, 26, 32, 33, 34, 47, 51, 60, 64, 65, 66, 68, 69, 79, 95

dữ liệu

cập nhập dữ liệu, 39, 40, 53, 88, 89, 101, 102, 105, 110, 114, 120, 122, 123

dữ liệu, 2, 3, 4, 5, 6, 7, 8, 9, 11, 12, 14, 15, 16, 17, 18, 19, 20, 21, 24, 25, 26, 27, 29, 30, 33, 41, 42, 43, 44, 46, 48, 53, 54, 55, 56, 58, 59, 60, 61, 62, 63, 64, 65, 69, 70, 71, 72, 77, 86, 90, 92, 95, 96, 98

bản đồ, 4, 15, 16, 25, 26, 59, 60, 61, 65

chuyển đổi dữ liệu, 12, 14, 62
 cơ sở dữ liệu, 3, 14, 16, 17, 19, 20, 21, 25, 55, 86
 dạng số, 5, 12, 16, 63
 lưu thông dữ liệu, 104
 lưu trữ dữ liệu, 11, 43, 44, 54, 61
 lưu trữ thông tin, 4
 nhân dữ liệu, 122
 nhập dữ liệu, 11, 12, 13, 16, 42, 58, 64
 phân tích, 3, 6, 7, 8, 9, 10, 14, 15, 16, 17, 18, 20, 24, 25, 26, 33, 54, 61, 62, 76, 86, 90, 98
 quản lý dữ liệu, 17, 26
 thu nhận, 5, 9, 15, 16, 17, 69
 thu thập dữ liệu, 102
 truy nhập, 3
 dữ liệu dữ liệu, 61, 62, 63, 64, 65, 69, 70, 71, 72
 dữ liệu hình học, 15
 dữ liệu phi không gian, 15, 17
 dữ liệu quan trắc, 76, 106
 đường biên, 27, 43, 44, 49, 52, 59, 62, 64, 65

E

ESRI, 14, 51, 64, 98, 112, 114, 115, 122, 125
 Extensible Markup Language, 121

F

FireFox, 117
 FOSS, 116, 118, 126

G

GDAL, 116
 GeoMedia, 115, 116
 GeoServer, 117
 Geospatial, 117
 GeoWeb, 121
 giá thành, 7
 GPS, 5, 6, 21, 26, 30, 68, 101, 116, 117, 119, 120, 121, 123, 126

H

hệ tham chiếu, 24, 91, 92, 96, 104
 hệ thống dẫn đường, 120

hệ thống quản lý CSDL, 13, 15, 39, 54, 89, 90, 125

hệ thống thông tin địa lý, 1, 4, 16, 22, 30
 GIS, 1, 2, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 19, 20, 21, 22, 23, 26, 30, 37, 38, 41, 42, 46, 47, 48, 50, 52, 53, 55, 56, 57, 58, 59, 65, 68, 70, 72, 74, 76, 78, 79, 80, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 101, 102, 103, 104, 107, 108, 110, 111, 112, 113, 114, 115, 116, 117, 118, 119, 120, 121, 122, 126, 130, 131
 HTTĐL, 3, 6, 17, 18, 19, 20, 21, 22, 24, 130

Hệ thống thông tin địa lý
 GIS, 65, 68

hệ thống tọa độ
 hệ tọa độ, 3, 7, 69
 hệ tọa độ không gian, 3

hiệu chỉnh, 70, 103, 105, 129, 133

I

Intergraph, 14, 58, 115, 116, 122
 Internet, 115, 117, 119, 120, 121, 122, 126, 130

K

khoa học máy tính, 3
 khoa học trái đất, 3
 khoa học tự nhiên, 2, 3
 khối lượng đào đắp, 95
 không gian ba chiều, 95

L

lạc hậu, 102, 120
 LAN, 121, 130
 Landmark Information System, 10, 126
 Linux, 117
 lô gic, 96

M

mạng cục bộ, 121
 mạng không dây, 116, 120
 Map Calculator, 98

MapInfo, 15, 115, 116, 122
 mật độ quan trắc, 103
 máy tính để bàn, 115, 121
 máy vi tính, 5, 11
 Mecator, 38
 MGE, 15, 55, 58, 115, 116
 Microsoft, 19, 39, 90, 116, 117, 127, 131
 MicroStation, 15, 22, 115
 mô hình, 2, 3, 6, 7, 8, 9, 16, 18, 19, 20, 25, 30, 53, 54, 56, 59, 64, 74, 76, 78, 80, 92, 93
 mô hình hoá, 3, 8, 9, 18, 25
 mô hình độ cao, 82, 83, 94
 DEM, 22, 74, 75, 78, 80, 81, 82, 83, 84, 125
 mobile, 120, 121
 Modular GIS Environments, 115
 môi trường, 2, 4, 6, 7, 15, 16, 18, 19, 20, 21, 55
 điều kiện, 7, 95
 MySQL, 123

N

người sử dụng, 5, 11, 16, 18, 25, 27, 33, 61, 86
 nội suy điểm, 92

O

OSSIM, 117

P

phân loại, 37, 39, 40, 42, 49, 60, 61, 74, 81, 91, 93, 94, 95, 102, 103, 105, 106, 107, 108, 111
 phần mềm
 nguồn mở, 116, 117, 118, 122, 123, 132
 phần mềm, 10, 11, 14, 15, 16, 17, 19, 21, 22, 55, 58, 62, 63, 68, 71, 78, 86
 đồ họa, 10, 86
 lập thể, 5, 34, 78
 vật thể, 7, 10, 27, 30, 34, 35
 phân tích dữ liệu
 chồng ghép bản đồ, 96, 107, 109
 chồng xếp bản đồ, 86

chu vi, 44, 45, 48, 71, 107
 chức năng đo đạc, 94
 chuyển đổi đơn vị, 94
 hỏi đáp, 3, 19, 95
 hỏi đáp tìm kiếm, 95
 overlay, 86
 phép tính đại số, 98
 tìm kiếm, 3, 8, 26, 89, 90

phát triển kinh tế, 7
 phép chiếu, 37, 91, 92, 132
 phép nội suy, 18, 19, 64, 78, 92, 93
 phòng thí nghiệm, 18
 PHP, 117, 123
 phương án, 7, 88, 97, 103
 phương pháp nội suy, 92
 phương pháp thống kê, 86
 polygons, 38, 49, 52, 74, 96, 108
 PostgreSQL, 117
 PrimaGIS, 117

Q

QGIS, 116
 quan hệ không gian, 10, 24, 47, 51, 56, 69, 70, 86
 qui hoạch, 17, 19, 21, 22

R

raster, 41, 42, 43, 44, 45, 46, 47, 48, 53, 54, 56, 59, 61, 62, 63, 71, 72, 76, 78, 79, 98, 104, 107, 108, 115, 132
 Raster, 61

S

sai số, 9, 50, 58, 68, 85, 89, 91, 92, 101, 102, 105, 107, 108, 109, 110, 111, 129, 133
 làm tròn, 107
 nguồn gốc, 110
 sản xuất, 4, 7, 17, 18, 22, 30
 Simple Object Extensible Protocol, 121
 sơ đồ, 3, 13, 79, 95, 98
 số hóa, 43, 60, 61, 64, 67, 71, 72, 83, 85, 86, 104, 105, 107, 108, 109, 110, 127, 129, 133

SOAP, 121

sử dụng đất, 55, 74, 75, 88, 102, 103, 105

T

tệp tin, 6, 39, 40, 54, 55, 78, 87, 89, 98, 127, 133

thí nghiệm, 6, 106, 130

thiết bị, 6, 11, 47, 98, 102, 105, 116, 119, 120, 121, 123, 127, 128, 129, 130, 131, 132, 133

thiết bị di động, 121, 123

thời gian thực, 120, 121, 123

thửa đất chuyên dùng, 95

tìm kiếm, 39, 115

tổ hợp, 69

topology, 47, 51, 56, 65, 67, 69, 70

trắc địa, 3, 32

truy cập, 39, 40, 49, 53, 89, 91, 94, 99, 116, 121, 122, 128, 129, 132, 133

tự động, 53, 83, 85, 89, 94, 101, 114, 121, 127, 129

U

UNIX, 114, 115, 133

USLE, 19, 109, 110

UTM, 6, 33, 35, 36, 37, 38, 68, 91, 126, 129

V

vấn đề, 2, 6, 7, 16, 18, 20, 21, 22, 30, 95

vận hành đồng thời, 121

vector, 9, 11, 41, 46, 47, 48, 49, 50, 51, 52, 53, 54, 59, 61, 62, 63, 64, 70, 71, 72, 76, 78, 80, 82, 83, 104, 108, 114, 115, 132

vị trí địa lý, 1, 38, 41, 87, 88, 105, 120

viễn thám, 2, 3, 6, 16, 17, 18, 21, 26, 42, 78

 đo vẽ ảnh, 3

 khảo sát, 3, 18

 vệ tinh, 3, 30, 59, 62, 63

Visual Basic, 114, 116

Visual C, 116

W

web portals, 121

WiFi, 121

Windows, 114, 115, 116, 117

Windows NT, 114

X

xử lý biên, 91

xử lý dữ liệu, 5, 70, 89, 90, 107, 108, 114, 115, 121, 129

 chuyển đổi hệ tọa độ, 91

 tạo vùng đệm, 53, 91, 97, 115

 thống kê, 2, 3, 5, 6, 15, 16, 18, 20, 21, 22

 tổ hợp, 3, 16, 18, 19, 20, 21, 47, 69, 95

 xóa bỏ dữ liệu, 89

 xử lý biên, 91

