

LỜI CẢM ƠN

Trước tiên tôi xin chân thành cảm ơn *PGS.TS. Đặng Văn Đức, Viện Công nghệ thông tin, Viện Hàn lâm Khoa học và Công nghệ Việt Nam*, người đã định hướng và tận tình hướng dẫn, giúp đỡ tôi trong quá trình thực hiện luận văn tốt nghiệp.

Tôi xin chân thành cảm ơn các thầy giáo, cô giáo trường Đại học Công nghệ thông tin và truyền thông - Đại học Thái Nguyên đã tận tình truyền đạt các kiến thức, quan tâm, động viên trong thời gian tôi học tập và nghiên cứu tại Trường.

Tôi xin chân thành cảm ơn Trường THPT Chuyên Bắc Kạn đơn vị tôi đang công tác đã hết sức tạo điều kiện để tôi có thể hoàn thành nhiệm vụ học tập của mình.

Cho phép tôi gửi lời cảm ơn tới các bạn học cùng CK12I - lớp chuyên ngành Khoa học máy tính, đã giúp đỡ, chia sẻ kinh nghiệm, cung cấp các tài liệu hữu ích trong thời gian tôi học tập, nghiên cứu tại Trường cũng như trong quá trình thực hiện luận văn tốt nghiệp vừa qua.

Vì lượng kiến thức thực tế còn ít nên trong luận văn của em khó tránh khỏi những hạn chế và khiếm khuyết, em rất mong nhận được sự đóng góp ý kiến của các thầy giáo, cô giáo và các bạn để bản thân em có thể hoàn thành tốt hơn kiến thức của mình.

Em xin trân trọng cảm ơn!

Thái Nguyên, tháng 9 năm 2015

Hạ Thị Thảo

LỜI CAM ĐOAN

Tôi xin cam đoan bản luận văn “Các cấu trúc dữ liệu trong hệ thống thông tin địa lý.” là công trình nghiên cứu của tôi dưới sự hướng dẫn khoa học của PGS.TS. Đặng Văn Đức, tham khảo các nguồn tài liệu đã được chỉ rõ trong trích dẫn và danh mục tài liệu tham khảo. Các nội dung công bố và kết quả trình bày trong luận văn này là trung thực và chưa từng được ai công bố trong bất cứ công trình nào.

Thái Nguyên, tháng 09 năm 2015

Hạ Thị Thảo

DANH MỤC CÁC TỪ VIẾT TẮT VÀ THUẬT NGỮ

Ký hiệu/từ viết tắt	Viết đầy đủ	Ý nghĩa
CSDL	Cơ sở dữ liệu	Cơ sở dữ liệu
ESRI	Environmental Systems Research Institute	Viện nghiên cứu Hệ thống môi trường Mỹ
GIS	Geographic Information System	Hệ thống thông tin địa lý
I/O	Input/Output	Nhập/Xuất
XUB	X on Upper Bound	X ở trên biên
YUB	Y on Upper Bound	Y ở trên biên
XLB	X on Lower Bound	X ở dưới biên
YLB	Y on Lower Bound	Y ở dưới biên

DANH MỤC CÁC BẢNG

Bảng 2.1: Các trường hợp của phép chèn vào cây tứ phân điểm.....	33
Bảng 2.2: Mô tả bốn cảnh của nút N trong cây tứ phân MX	35
Bảng 3.1 Các nút lệnh trên thanh công cụ	56

DANH MỤC CÁC HÌNH

Hình 1.1 Hệ thông tin địa lý (ESRI)	3
Hình 1.2: Sự ảnh hưởng của lựa chọn kích thước tế bào	5
Hình 1.3: Trật tự không gian.....	6
Hình 1.4: Số liệu vectơ được biểu thị dưới dạng điểm (Point).....	8
Hình 1.5: Số liệu vectơ được biểu thị dưới dạng Arc	8
Hình 1.6: Số liệu vectơ được biểu thị dưới dạng vùng (Polygon)	9
Hình 1.7: Các nhóm chức năng trong GIS	12
Hình 2.1: Cây k-d tương ứng khi cho các điểm có sẵn.....	20
Hình 2.2: Lưới bản đồ dựng cây	21
Hình 2.3: Trình tự chèn vào cây 2-d	22
Hình 2.4: Phép chèn cây k-d trên bản đồ	23
Hình 2.5: Cách phân hoạch mặt phẳng bởi các điểm xã trên cây tứ phân điểm.....	30
Hình 2.6: Tiến trình chèn vào cây tứ phân điểm.....	31
Hình 2.7: Mô hình một cây tứ phân điểm	32
Hình 2.8: Trình tự chèn vào cây tứ phân MX	36
Hình 2.9: Phép chèn điểm vào cây tứ phân MX	37
Hình 2.10: Sơ đồ R - Tree	39
Hình 2. 11: Bản đồ mẫu mô tả cách nhóm các hình chữ nhật minh họa cây R	40
Hình 2.12: Ví dụ về tập hợp các đoạn thẳng được nhúng với lưới 4 x 4	41
Hình 2.13: Tập hợp các đoạn thẳng và không gian được bao bởi các hình chữ nhật.....	42
Hình 2.14: Trình tự chèn vào cây R.....	44
Hình 2.15: Bản đồ mô tả phép chèn trong cây R	45
Hình 2.16: Mô tả phép tách.....	47
Hình 3.1: Mô hình Use Case của hệ thống	51
Hình 3.2: Giao diện chính của chương trình.....	55
Hình 3.3: Bản đồ sau khi hiển thị cả lớp đường và lớp điểm.	56
Hình 3.4. Bản đồ hiển thị lớp điểm.....	57
Hình 3.5 Truy vấn vùng trên bản đồ lớp điểm.....	57
Hình 3.6: Kết quả của truy vấn trên hình 3.5	58
Hình 3.7: Truy vấn vùng trên bản đồ tổng thể.....	58
Hình 3.8: Kết quả của truy vấn vùng trên bản đồ tổng thể	59

MỤC LỤC

LỜI CẢM ƠN	i
LỜI CAM ĐOAN	ii
DANH MỤC CÁC TỪ VIẾT TẮT VÀ THUẬT NGỮ	iii
DANH MỤC CÁC BẢNG.....	iv
DANH MỤC CÁC HÌNH.....	iv
MỤC LỤC.....	v
MỞ ĐẦU.....	1
CHƯƠNG 1: TỔNG QUAN VỀ HỆ THỐNG THÔNG TIN ĐỊA LÝ (GIS).....	3
1.1 Một số khái niệm cơ bản về GIS.....	3
1.2 Cấu trúc dữ liệu địa lý	3
1.2.1 Dữ liệu không gian.....	4
1.2.2 Dữ liệu phi không gian	9
1.3 Các chức năng của GIS	11
1.4 Tìm kiếm và phân tích dữ liệu không gian	13
1.4.1 Tìm kiếm nội dung trong vùng không gian	13
1.4.2 Tìm kiếm trong khoảng cận kề	14
1.4.3 Tìm kiếm hiện tượng và thao tác bao phủ (overlay).....	14
1.4.4 Nội suy và mô hình hóa bề mặt	15
1.4.5 Phân tích đường đi và đường dẫn	15
1.4.6 Mô hình hóa tương tác không gian	15
1.4.7 Đồ họa và tương tác	16
CHƯƠNG 2: MỘT SỐ CẤU TRÚC DỮ LIỆU SỬ DỤNG TRONG GIS.....	17
2.1 Vấn đề lưu trữ và chỉ mục dữ liệu địa lý.....	17
2.2 Cây k-d (k-d tree)	18
2.2.1 Cấu trúc nút.....	19
2.2.2 Chèn và tìm kiếm trong cây 2-d.....	20
2.2.3 Xóa trong cây 2-d.....	24
2.2.4 Truy vấn khoảng trong cây 2-d.....	25
2.2.5 Cây k-d với $k \geq 2$	28
2.3 Cây tứ phân điểm (Point quadtree)	28
2.3.1 Chèn và tìm kiếm trong cây tứ phân điểm	30
2.3.2 Thao tác xoá trên cây tứ phân điểm	31
2.3.3 Truy vấn khoảng trong cây tứ phân điểm	33

2.4 Cây tứ phân Matrix MX (MX-Quadtree).....	34
2.4.1 Chèn và tìm kiếm trong MX-Quadtree	35
2.4.2 Thao tác xoá trong MX-Quadtree	37
2.4.3 Truy vấn khoảng trong MX-Quadtree	38
2.5 Cây R (R - Tree).....	39
2.5.1 Chèn và tìm kiếm trong R-Tree	42
2.5.2 Xoá trong R-Tree	45
2.5.3 Thuật toán tách nút (Node Splitting)	46
2.5.4 Cây R* (R*- Tree)	47
2.6 So sánh các cây dữ liệu	48
CHƯƠNG 3: XÂY DỰNG CHƯƠNG TRÌNH THỬ NGHIỆM.....	50
3.1 Lựa chọn bài toán thử nghiệm và công nghệ sử dụng	50
3.1.1 Phát biểu bài toán.....	50
3.1.2 Cách giải quyết	50
3.2 Mô tả dữ liệu thử nghiệm.....	52
3.3 Phân tích và thiết kế chương trình thử nghiệm	53
3.3.1 Công cụ xây dựng chương trình	53
3.3.2 Đặc tả chức năng của chương trình.....	53
3.4 Đánh giá kết quả thu được	54
3.4.1 Cài đặt và thử nghiệm	54
3.4.2 Kết quả thử nghiệm.....	54
3.4.3 Nhận xét kết quả thu được	59
3.4.4 Hiệu quả truy vấn khi sử dụng các cấu trúc dữ liệu cây	59
KẾT LUẬN	61
TÀI LIỆU THAM KHẢO.....	62
PHỤ LỤC.....	63

MỞ ĐẦU

Ngày nay, hệ thống thông tin địa lý – Geographic Information System(GIS) đã trở nên phổ biến và là yếu tố không thể thiếu đối với hầu hết các chuyên ngành. GIS đánh giá được hiện trạng của các quá trình, các thực thể tự nhiên, kinh tế-xã hội thông qua các chức năng thu thập, quản lý, truy vấn, phân tích và tích hợp thông tin. Nó được gắn với một nền bản đồ số nhất quán trên cơ sở tọa độ của các dữ liệu đầu vào. Về bản chất GIS là một hệ thống. Các yếu tố cấu thành GIS là phần cứng, phần mềm, các cơ sở dữ liệu, phương pháp và con người. Trong các yếu tố này, cơ sở dữ liệu có vai trò cực kỳ quan trọng trong hệ thống góp phần làm nên sức mạnh của hệ thống.

Tình hình nghiên cứu, kết quả nghiên cứu đã có:

Cấu trúc dữ liệu là một thành phần quan trọng nhất và được coi là lõi của hệ thống thông tin địa lý GIS. Do đó đây là vấn đề được đề cập đến rất nhiều trên thế giới, tuy nhiên cách tiếp xúc tiếp xúc vấn đề có nhiều khía cạnh khác nhau đặc biệt là ở Việt Nam. Để hình thức hoá, trừu tượng hoá dữ liệu GIS người ta cần đến các kỹ thuật như cây k-chiều, cây tứ phân, cây R... Đồng thời còn phải quan tâm đến các phương pháp cài đặt các kỹ thuật này. Các đề tài nghiên cứu đã có chỉ là ứng dụng CSDL GIS để xây dựng ứng dụng chứ chưa có đề tài nghiên cứu thể hiện các cấu trúc dữ liệu GIS cụ thể. Do đó tác giả tập trung nghiên cứu các kỹ thuật trên các cây dữ liệu GIS.

Có thể nói cấu trúc dữ liệu là phần khung và bản chất nhất của các hệ thống GIS, nó là cơ sở của các giải thuật GIS cũng như nói đến khả năng lưu trữ khai thác, phát triển hệ thống dữ liệu. Những nghiên cứu của luận văn không có gì mới so với thế giới, nhưng là mới ở nước ta. Xuất phát từ thực tế đó tôi chọn đề tài “Các cấu trúc dữ liệu trong hệ thống thông tin địa lý.”

Trong khuôn khổ luận văn, tôi trình bày một số vấn đề cơ bản về hệ thống thông tin địa lý (GIS), các kỹ thuật truy vấn không gian trong GIS. Mô tả cấu trúc, các phép toán xây dựng, chèn, xóa, duyệt, truy vấn trên cấu trúc dữ liệu sử dụng

trong GIS. Trong đó, tập trung nghiên cứu và cài đặt thử nghiệm một số cấu trúc dữ liệu không gian.

Bố cục của luận văn bao gồm phần mở đầu, phần kết luận và ba chương nội dung được tổ chức như sau:

Chương 1: Tổng quan về hệ thống thông tin địa lý

Chương này trình bày tổng quan về hệ thống thông tin địa lý, các kỹ thuật truy vấn không gian trong GIS, khả năng của GIS.

Chương 2: Một số cấu trúc dữ liệu sử dụng trong GIS

Chương này mô tả cấu trúc, các phép toán chèn, xoá, duyệt, truy vấn trên các kỹ thuật chỉ mục và tìm kiếm không gian như: cây k-d (k-d Tree), cây tứ phân(Quadtree), cây R (R Tree), R* Tree và so sánh giữa chúng

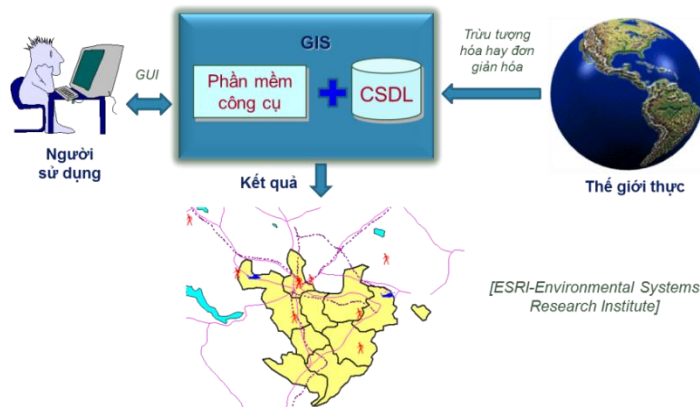
Chương 3: Xây dựng chương trình thử nghiệm

Cài đặt thử nghiệm cây tứ phân điểm. Chương trình được cài đặt từ cơ sở dữ liệu đã có định dạng bằng Shapefile, với ngôn ngữ lập trình C#.NET cùng với thư viện hỗ trợ SharpMap.

CHƯƠNG 1: TỔNG QUAN VỀ HỆ THỐNG THÔNG TIN ĐỊA LÝ (GIS)

1.1 Một số khái niệm cơ bản về GIS

GIS - Geographic Information System hay hệ thống thông tin địa lý được hình thành từ ba khái niệm địa lý, thông tin và hệ thống. Khái niệm “địa lý” liên quan đến các đặc trưng về không gian, vị trí. Các đặc trưng này liên kết trực tiếp đến các đối tượng trong không gian. Chúng có thể là vật lý, văn hoá, kinh tế, tự nhiên... Khái niệm “thông tin” đề cập đến phần dữ liệu được quản lý bởi GIS. Đó là các dữ liệu về thuộc tính và không gian của đối tượng. GIS có tính “hệ thống” tức là hệ thống GIS được xây dựng từ các module. Việc tạo module giúp thuận lợi trong việc quản lý và hợp nhất.



Hình 1.1 Hệ thống tin địa lý (ESRI) [1]

GIS là một hệ thống có ứng dụng rất lớn. Từ năm 1980 đến nay đã có rất nhiều các định nghĩa được đưa ra, tuy nhiên không có định nghĩa nào khái quát đầy đủ về GIS vì phần lớn chúng đều được xây dựng trên khía cạnh ứng dụng cụ thể trong từng lĩnh vực. Có ba định nghĩa được dùng nhiều nhất:

GIS là một hệ thống thông tin được thiết kế để làm việc với các dữ liệu trong một hệ toạ độ quy chiếu. GIS bao gồm một hệ cơ sở dữ liệu và các phương thức để thao tác với dữ liệu đó. GIS là một hệ thống nhằm thu thập, lưu trữ, kiểm tra, tích hợp, thao tác, phân tích và hiển thị dữ liệu được quy chiếu cụ thể vào trái đất.

GIS là một chương trình máy tính hỗ trợ việc thu thập, lưu trữ, phân tích và hiển thị dữ liệu bản đồ.

1.2 Cấu trúc dữ liệu địa lý

Một cơ sở dữ liệu của hệ thống thông tin địa lý có thể chia ra làm 2 loại dữ liệu cơ bản: dữ liệu không gian và phi không gian. Mỗi loại có những đặc điểm riêng và chúng khác nhau về yêu cầu lưu giữ số liệu, hiệu quả, xử lý và hiển thị.

1.2.1 Dữ liệu không gian

Dữ liệu không gian là những mô tả số của hình ảnh bản đồ, chúng bao gồm tọa độ, quy luật và các ký hiệu dùng để xác định một hình ảnh bản đồ cụ thể trên từng bản đồ. Hệ thống thông tin địa lý dùng các số liệu không gian để tạo ra một bản đồ hay hình ảnh bản đồ trên màn hình hoặc trên giấy thông qua thiết bị ngoại vi. Hệ thống tin địa lý sử dụng hai mô hình dữ liệu cơ bản để biểu diễn các đặc trưng không gian: mô hình dữ liệu raster và mô hình dữ liệu vector. Mô hình không gian đặc biệt quan trọng vì cách thức biểu diễn thông tin sẽ ảnh hưởng đến khả năng hiển thị đồ họa của hệ thống.

1.2.1.1 Mô hình dữ liệu raster

Đây là phương pháp biểu diễn các đặc trưng địa lý bằng các điểm ảnh. Được hình thành dựa trên cơ sở quan sát nền thế giới thực.

Mô hình dữ liệu raster hay còn gọi là lưới tế bào hình thành nên cho một số hệ thống tin địa lý. Các hệ thống trên cơ sở raster hiển thị, định vị và lưu trữ dữ liệu đồ họa nhờ sử dụng các ma trận hay lưới tế bào. Độ phân giải dữ liệu raster phụ thuộc vào kích thước của tế bào hay điểm ảnh, chúng có thể khác nhau từ vài đêximet đến vài kilômet. Trong cấu trúc dữ liệu raster, point có thể được biểu diễn bằng một cell. Line được biểu diễn bởi một tập các cell có hướng xác định, độ rộng của line bằng chiều rộng của một cell. Polygon được biểu diễn bởi một dãy các cell nằm kề sát nhau. Tiến trình xây dựng lưới tế bào được mô tả như sau đây:[1]

Giả sử phủ một lưới lên bản đồ, dữ liệu raster được lập bằng cách mã hóa mỗi tế bào bằng một giá trị dựa theo các đặc trưng trên bản đồ, độ chính xác của một đối tượng phụ thuộc vào kích thước hay độ phân giải của các tế bào lưới.