

ĐẠI HỌC THÁI NGUYÊN
TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN VÀ TRUYỀN THÔNG

TRẦN HỮU CƯỜNG

**TỔNG HỢP VÀ PHÂN TÍCH DỮ LIỆU TỪ CÁC ĐỊA
PHƯƠNG TRÊN NỀN WEB CÓ SỰ TRỢ GIÚP
CỦA CÔNG NGHỆ GIS**

Chuyên ngành: Khoa học máy tính

Thái Nguyên - 2014

LỜI CẢM ƠN

Sau một thời gian nghiên cứu, cùng với sự giúp đỡ tận tình của thầy giáo PGS.TS. Nguyễn Văn Vy đến nay em đã hoàn thành luận văn về đề tài “Tổng hợp và phân tích dữ liệu từ các địa phương trên nền web có sự trợ giúp của công nghệ GIS”

Em xin trân trọng cảm ơn thầy PGS.TS. Nguyễn Văn Vy người đã trực tiếp hướng dẫn, giúp đỡ em trong quá trình làm luận văn.

Em xin cảm ơn quý thầy cô trong trường Đại học Công Nghệ Thông Tin và Truyền Thông - Đại Học Thái Nguyên đặc biệt là các quý thầy cô trong Viện Hàn lâm Khoa học và Công nghệ Việt Nam đã chỉ dạy cho em những kiến thức hết sức quý báu trong thời gian em học tập.

Vì lý do thời gian và vì lượng kiến thức thực tế còn hạn chế nên trong luận văn chắc chắn không tránh khỏi những khiếm khuyết, mong thầy cô giáo cùng các bạn đóng góp ý kiến để tôi có thể củng cố kiến thức và tiếp tục hoàn thiện tài liệu tốt hơn.

Em xin chân thành cảm ơn!

Trần Hữu Cường

MỤC LỤC

LỜI CẢM ƠN.....	I
MỤC LỤC.....	II
TÓM TẮT.....	IV
MỘT SỐ THUẬT NGỮ VÀ TỪ VIẾT TẮT.....	V
MỞ ĐẦU	8
CHƯƠNG 1 KHÁI QUÁT VỀ CÔNG NGHỆ GIS VÀ BÀI TOÁN TỔNG HỢP PHÂN TÍCH SỐ LIỆU.....	4
1.1. Công nghệ GIS	4
1.1.1. Định nghĩa GIS	4
1.1.2. Thành phần dữ liệu GIS.....	4
1.1.2.1. Dữ liệu không gian	5
1.1.2.2. Dữ liệu thuộc tính	6
1.1.3. Các chức năng chính của GIS.....	7
1.1.4. Cơ sở dữ liệu phân tán dạng GIS.....	8
1.1.5. Hệ quản trị CSDL lựa chọn	11
1.2. WebGIS.....	11
1.2.1. Kiến trúc WebGIS và các bước xử lý.....	11
1.2.2. Các chuẩn trao đổi WebGIS hiện nay.....	18
1.2.3. Web và Webservice	20
1.2.4. Đặc điểm của WebGIS.....	21
1.2.5. Những công cụ được sử dụng cho WebGIS	23
1.3 Hoạt động nghiệp vụ theo dõi dữ liệu	25
1.3.1. Tổ chức thu thập và tổng hợp dữ liệu	25
1.3.2. Khai báo dữ liệu.....	26
1.3.3. Điều tra định kỳ hay đột xuất.....	26
1.3.4. Sơ đồ mô tả quá trình thu thập dữ liệu báo cáo	27
CHƯƠNG 2: MỘT SỐ KỸ THUẬT PHÂN TÍCH DỮ LIỆU VÀ DỰ BÁO.....	28
2.1. Một số kỹ thuật phân tích dữ liệu	28
2.1.1. Kỹ thuật toán định lượng	28
2.1.2. Kỹ thuật phân tích dựa vào bản đồ và GIS.....	31
2.1.3. Kỹ thuật viễn thám phân tích chỉ số thực vật (NDVI).....	29
2.1.4. Kỹ thuật tổng hợp.....	31
2.2. Một số phương pháp dự báo	36
2.2.1. Hồi quy tuyến tính đa biến.....	36
2.2.2. Phương pháp dự báo dịch dựa vào giá trị trung bình cộng 2 độ lệch chuẩn (Mean + 2SD).....	37
2.2.3. Dự báo dựa trên phân bố Poisson.....	38
2.3 Phân tích và dự báo	42
2.3.1. Hoạt động nghiệp vụ tại các cấp.....	42
2.3.2. Các chức năng nghiệp vụ.....	44
2.3.3. Mô hình vật lý của hệ thống	45
2.3.4. Mô hình logic của hệ thống	46
2.3.5. Thiết kế các phân hệ của hệ thống	48

2.3.5.1. Sơ đồ các phân hệ.....	46
2.3.5.2. Thiết kế mô hình phân hệ các tuyến	49
2.3.5.3. Thiết kế phân hệ tuyến xã.....	48
2.3.5.4. Phân hệ tuyến huyện	51
2.3.5.5. Phân hệ tuyến tỉnh.....	51
CHƯƠNG 3 CHƯƠNG TRÌNH THỬ NGHIỆM.....	51
3.1. Phân tích yêu cầu và lựa chọn môi trường.....	51
3.1.1. Phân tích dữ liệu dịch bệnh.....	51
3.1.2. Thiết kế dữ liệu logic	56
3.1.3. Thiết kế dữ liệu vật lý	60
3.1.4. Ứng dụng	66
3.1.5. Phân hệ khai thác trên nền web	69
3.2. Lập trình và cài đặt hệ thống.....	72
3.2.1. Lựa chọn môi trường	72
3.2.2. Lập trình và cài đặt hệ thống	73
3.2.3. Vận hành và thử nghiệm hệ thống	73
3.2.4. Đánh giá.....	73
3.3. Một số kết quả.....	71
3.3.1. Những kết quả của luận văn.....	74
3.3.2. Những hạn chế và hướng phát triển.....	77
TÀI LIỆU THAM KHẢO	78
PHỤ LỤC: CÁC MẪU BÁO CÁO VÀ TỔNG HỢP	79

TÓM TẮT

GIS ra đời và được phát triển mạnh trong những năm gần đây. Cùng với sự bùng nổ về công nghệ của Internet, GIS cũng đã phát triển công nghệ cho phép chia sẻ thông tin thông qua mạng toàn cầu bằng cách kết hợp GIS và Web hay còn gọi là WebGIS. Bên cạnh đó, xu hướng chia sẻ dữ liệu, phát triển phần mềm trên công nghệ mã nguồn mở cũng đang được quan tâm ở các nước đang phát triển vì nhiều lợi ích mà nó mang lại. Vì thế việc nghiên cứu ứng dụng WebGIS trên cơ sở mã nguồn mở sẽ mang lại khả năng chia sẻ thông tin địa lý rộng rãi cho các ngành.

Nội dung chính của đề tài là tìm hiểu về hệ thống thông tin địa lý GIS và ứng dụng của chúng. Nghiên cứu về WebGIS, khả năng xây dựng ứng dụng WebGIS trên cơ sở mã nguồn mở, trên cơ sở đó ứng dụng xây dựng WebGIS phục vụ công tác *“Tổng hợp và phân tích dữ liệu từ các địa phương trên nền web có sự trợ giúp của công nghệ GIS”*.

Đề tài đi sâu về các kỹ thuật phân tích vì vậy sẽ không đề cập nhiều đến vấn đề tổng hợp số liệu

Kết quả của đề tài đã trình bày các nghiên cứu lý thuyết về WebGIS: phân loại, tìm hiểu các chiến lược phát triển WebGIS, tìm hiểu phần mềm xây dựng WebGIS trên cơ sở mã nguồn mở là MapServer và đã xây dựng ứng dụng WebGIS giám sát tình hình và cảnh báo xu hướng lây truyền của bệnh sốt rét trên phạm vi cả nước.

MỘT SỐ THUẬT NGỮ VÀ TỪ VIẾT TẮT

Viết tắt	Tên đầy đủ
GIS	Geographic Information System - Hệ thống tin địa lý
ESRI	Environmental System Research Institute- Viện nghiên cứu môi trường
DBMS	Data Base Manager System – Hệ quản trị cơ sở dữ liệu.
GUI	Graphical User Giao diện - Giao diện đồ hoạ người sử dụng.
CSDL	Cơ sở dữ liệu.
WWW	World Wide Web - mạng toàn cầu.
HTML	HyperText Markup Language - Ngôn ngữ siêu văn bản.
HTTP	HyperText Transfer Protocol - Giao thức truyền siêu văn bản.
KST	Ký sinh trùng
SR	Sốt rét
BNSR	Bệnh nhân sốt rét
SRLH	Vùng sốt rét lưu hành
PCSR	Phòng chống sốt rét
YTDP	Y tế dự phòng
XN	Xét nghiệm
TT	Tuyên truyền

DANH MỤC HÌNH VẼ

Hình 1.1: Mô hình dữ liệu không gian	5
Hình 1.2: Dữ liệu Raster	6
Hình 1.3: Chức năng chính của GIS	8
Hình 1.4 : Geodatabase trong ArcGIS	9
Hình 1.5: Kiến trúc webgis.....	12
Hình 1.6: Các dạng yêu cầu từ phía client.....	15
Hình 1.7: Cấu hình Server-side	17
Hình 1.8: Cấu hình Client - Side	17
Hình 1.9: Khả năng xử lý Gis.....	18
Hình 1.10: Client hỏi thông tin.....	19
Hình 1.11: Chuỗi URL	19
Hình 1.12 : Kiến trúc Web một máy chủ	20
Hình 1.13: Mô hình WebService.....	21
Hình 1.14: Mô hình vận hành của giao thức SOAP.....	23
Hình 1.15: Sơ đồ nghiệp vụ thu thập dữ liệu báo cáo.....	26
Hình 2.1: Mô hình hệ thống vật lý	44
Hình 2.2: Mô hình logic của hệ thống.....	45
Hình 2.3: Sơ đồ các thành phần của hệ thống	47
Hình 2.4: Mô hình tổng thể các tuyến.....	48
Hình 2.5: Mô hình xử lý của phân hệ tuyến xã	49
Hình 2.6 : Mô hình xử lý.....	51
Hình 3.1: Mô hình khái niệm của dữ liệu ER	54
Hình 3.2: Sơ đồ liên kết mô hình dữ liệu của hệ quan hệ thống	57
Hình 3.3: Kiến trúc hệ thống quản lý và dự báo dịch hại trực tuyến.....	64

DANH MỤC CÁC BẢNG

Bảng 1.1: DBMS.....	9
Bảng 2.1: Quy trình báo cáo tuyến xã.....	41
Bảng 2.2: Các chức năng của hệ thống theo dõi, giám sát dịch bệnh sốt rét.....	45
Bảng 2.3: Cập nhật bản đồ dịch bệnh	51
Bảng 2.4: Phân hệ tổng hợp, phân tích dự báo	52
Bảng 3.1: Các thực thể và các thuộc tính	53
Bảng 3.2: Mỗi quan hệ giữa các thực thể.....	53
Bảng 3.3: Lớp dữ liệu không gian.....	60
Bảng 3.4: thông tin tỉnh.....	58
Bảng 3.5: Các thôn.....	58
Bảng 3.6: Danh sách các xã/phường.....	58
Bảng 3.7: Danh sách các huyện (quận).....	59
Bảng 3.8: Bệnh nhân SR.....	59
Bảng 3.9: Danh mục thuốc và vật tư.....	59
Bảng 3.10: Danh mục loại bệnh sốt rét.....	60
Bảng 3.11: Kiểu xét nghiệm thực hiện.....	60
Bảng 3.12: Mục đích sử dụng thuốc.....	60
Bảng 3.13 Hình thức tuyên truyền.....	61
Bảng 3.14: Đơn vị y tế thuộc xã.....	61
Bảng 3.15: Điều trị bệnh nhân.....	61
Bảng 3.16: Cấp thuốc cho các mục tiêu sử dụng.....	62
Bảng 3.17: Tình trạng sử dụng vật tư và thuốc của đơn vị y tế.....	62
Bảng 3.18: Thống kê việc xét nghiệm	62
Bảng 3.19 Thống kê số người chết.....	63
Bảng 3.20 công tác tuyên truyền.....	63
Bảng 3.21. Đánh giá.....	68

MỞ ĐẦU

1. Đặt vấn đề

Việc theo dõi dịch bệnh (như dịch sốt rét, sốt xuất huyết dịch nở mồm long móng ở gia súc dịch chân tay miệng và một số bệnh lạ mới được phát hiện...) đã phát sinh ở các địa phương toàn quốc để có thông báo kịp thời cho các địa phương phòng tránh và có giải pháp thích hợp khi tình hình có nguy cơ bùng phát là công việc rất quan trọng và được tiến hành thường xuyên ở bộ Y tế. Trong nhiều năm qua, công việc này đã được tiến hành có kết quả nhưng vẫn chưa đảm bảo về mặt thời gian vì cách làm việc chủ yếu vẫn là thủ công: Các báo cáo và văn bản được gửi về cơ quan quản lý cấp bộ rồi tổng hợp, ra thông báo và các quyết định cho địa phương. Quá trình này thường diễn ra chập chạp, chưa đáp ứng được yêu cầu quản lý như mong muốn: nhiều đợt dịch như nở mồm long móng, sốt rét, chân tay miệng, sốt xuất huyết. Các chỉ đạo không theo kịp tình hình diễn biến của dịch bệnh, và có những hậu quả, thiệt hại đáng tiếc [].

Trong tình thế như trên, với sự tiến bộ nhanh chóng của công nghệ thông tin, luận văn cố gắng áp dụng các tiến bộ công nghệ mới cho hoạt động này. Vì thế đề tài: "*Tổng hợp và phân tích dữ liệu từ các địa phương trên nền web có sự trợ giúp của công nghệ GIS*" được đề xuất nhằm hoàn thiện một bước công tác theo dõi và chỉ đạo kịp thời các dịch bệnh hàng năm, hạn chế đến mức thấp nhất có thể tác hại của chúng.

Các dữ liệu về dịch bệnh do các địa phương thu thập, chúng thường gắn với phạm vi quản lý của địa phương, với dân cư và các điều kiện tự nhiên trên không gian địa bàn quản lý. Việc theo dõi các dịch bệnh đòi hỏi phải tổng hợp từ dưới lên, và cần được phân tích, đánh giá tình hình, tiến hành các dự báo để thông báo lại cho các địa phương biết và có những giải pháp đề phòng, ngăn chặn dịch bệnh kịp thời, đúng nơi, đúng đối tượng.

Công nghệ GIS đã phát triển [1], việc tổ chức thu thập dữ liệu cũng như chuyển tải và biểu diễn thông tin trên nền web có thể sử dụng công nghệ này [6,7,8].

GIS cung cấp cho người sử dụng khả năng thay đổi, đồng thời cũng xác định những gì có thể xảy ra khi có sự thay đổi dữ liệu đó.

Nói cách khác, GIS cung cấp cho người sử dụng những mô hình khác nhau để dữ liệu bản đồ là thành phần chính trong cơ sở dữ liệu của GIS. Các bản đồ gắn chặt với thế giới thực và luôn được bổ sung những thông tin mới theo sự diễn biến của thực tế.

Nền tảng của thông tin hình học trong GIS là bản đồ đã được số hoá ở dạng nào đó để có thể thực hiện từ những phép tính đơn giản nhất: *đo đạc diện tích, chu vi, chiều dài, vị trí ...* đến những phép tính phức tạp nhất: *mở rộng diện tích, xác định giao của nhiều vùng diện tích ...* là những bài toán khá phổ biến trong quản lý và nghiên cứu khoa học.

Từ các cơ sở dữ liệu (CSDL) thông thường, GIS làm trực quan hóa, thuận tiện và cùng một lúc cung cấp cho người sử dụng nhiều thông tin khác nhau được tổng hợp lại.

Ví dụ, Sở y tế của một tỉnh cần thông tin về mật độ ký sinh trùng của tỉnh để lựa chọn một địa danh làm điểm nghiên cứu tại một huyện nào đó, trên màn hình là bản đồ của huyện với mật độ ký sinh trùng từng khu vực được thể hiện bằng các màu khác nhau, sở y tế chỉ việc chọn một khu vực có mật độ ký sinh trùng cao và bấm “chuột” vào điểm đó, trên màn hình sẽ hiện lên các thông số về: *số dân, mật độ ký sinh trùng, khí tượng thủy văn, ...* của khu vực cần tìm

Các phần mềm GIS áp dụng tối đa công nghệ GIS để có thể tạo ra hệ thống tự động lập bản đồ và phương tiện xử lý dữ liệu thông minh, nhờ hệ chuyên gia,..

Trước những diễn biến ngày càng phức tạp về tình hình dịch bệnh SR trên cả nước, việc có một công cụ quản lý các hoạt động trong lĩnh vực phòng chống SR trên địa bàn để có thể đáp ứng nhu cầu thông tin nhanh chóng, chính xác, trợ giúp sự chỉ đạo kịp thời cho các Trung tâm y tế dự phòng, Trung tâm phòng chống SR là việc làm rất cần thiết. Việc quản lý các đối tượng này sẽ rất trực quan và hiệu quả nếu được xây dựng trên nền công nghệ thông tin địa lý (GIS).

2. Mục tiêu nghiên cứu của đề tài:

- Nghiên cứu các công nghệ liên quan đến phát triển hệ thống GIS (Geographical Information System) trên nền Web.
- Phát triển ứng dụng GIS trợ giúp việc cảnh báo dịch bệnh tại địa phương từ cấp cơ đến cấp tỉnh/thành phố một cách hiệu quả.

3. Nhiệm vụ nghiên cứu

- Nghiên cứu mô hình dữ liệu không gian được tổ chức trên WebGIS
- Mô hình hệ thống GIS véctor: Kiến trúc kết nối GIS và Web
- Khảo sát thu thập dữ liệu bản đồ, dữ liệu một loại bệnh SR từ cơ sở đến cấp tỉnh/thành phố và đề xuất mô hình dữ liệu lưu trữ trong hệ thống GIS .
- Xây dựng thử nghiệm chương trình demo Hệ thống thông tin tích hợp Web-GIS trên cơ sở mã nguồn mở, với khả năng cảnh báo dịch bệnh SR.