BỘ GIÁO DỤC VÀ ĐÀO TẠO
ĐẠI HỌC ĐÀ NẴNG

VÕ VĂN LUẬN

NGHIÊN CỨU SPATIAL OLAP XÂY DỰNG CỔNG THÔNG TIN BẢN ĐỒ KINH TẾ XÃ HỘI THÀNH PHỐ ĐÀ NẴNG

Chuyên ngành: KHOA HỌC MÁY TÍNH
Mã số: 60.48.01

TÓM TẮT LUẬN VĂN THẠC SĨ KỸ THUẬT

Đà Nẵng - Năm 2011

Công trình được hoàn thành tại
ĐẠI HỌC ĐÀ NẴNG

Người hướng dẫn khoa học: **PGS. TS. Lê Văn Sơn**

Phản biện 1: **PGS. TS. Tăng Tấn Chiến**

Phản biện 2: **PGS. TS. Đoàn Văn Ban**

Luận văn được bảo vệ trước Hội đồng chấm Luận văn tốt nghiệp thạc sĩ kỹ thuật họp tại Đại học Đà Nẵng vào ngày 18 tháng 6 năm 2011.

* *Có thể tìm hiểu luận văn tại:*
- Trung tâm Thông tin - Học liệu, Đại học Đà Nẵng.
- Trung tâm Học liệu, Đại học Đà Nẵng.4

MỞ ĐẦU

1. Lý do chọn đề tài

Hiện nay, Công nghệ thông tin được ứng dụng rộng rãi trong nhiều lĩnh vực của đời sống xã hội. Dữ liệu được thu thập và lưu trữ trong quá trình ứng dụng công nghệ thông tin ngày càng được tích luỹ nhiều lên.

Trong xu thế phát triển hiện nay, nhu cầu truy vấn, xử lý thông tin ngày càng đa dạng, nhiều câu hỏi mang tính chất định tính yêu cầu xử lý thông tin nhanh trên một khối lượng dữ liệu rất lớn đã có sẵn từ nhiều nguồn khác nhau. Từ đó, một khuynh hướng kỹ thuật mới phát triển thay thế các phương pháp quản trị và khai thác cơ sở dữ liệu truyền thống, đó là xử lý phân tích trực tuyến (OLAP) dựa vào kho dữ liệu.

Với ưu điểm về tốc độ truy cập thông tin nhanh, chính xác, kho dữ liệu đang được áp dụng một cách rộng rãi trong lĩnh vực quản lý tài chính, ngân sách, kinh doanh… và được ứng dụng xử lý phân tích trực tuyến phân tích nhằm mục đích xác định xu hướng, mô hình của hành vi, tìm kiếm mối quan hệ và phụ thuộc giữa các dữ liệu khác nhau. Các kết quả của những phân tích này là cơ sở cho việc ra các quyết định khác nhau. Kho dữ liệu và ứng dụng OLAP là những thành phần cốt lõi của các hệ thống hỗ trợ ra quyết định [4].

Việc ứng dụng kho dữ liệu và khai phá tri thức vào dữ liệu không gian sẽ mang đến kết quả trực quan, dễ hiểu và dễ tiếp cận, đặc biệt là khi công nghệ không gian được kết hợp với công nghệ phi không gian như việc kết hợp một hệ thống thông tin địa lý (GIS) với một kho dữ liệu. Với sự phát triển của khái niệm kho dữ liệu không gian, các công cụ thông thường được sử dụng để khai thác kho dữ liệu phi không gian không đủ để phân tích đầy đủ các thành phần không

gian địa lý của các dữ liệu không gian, thành phần địa lý này rất quan trọng trong việc mô tả kho dữ liệu không gian. Một giải pháp mới sau đó đã được phát triển, bao gồm kết hợp thế mạnh của GIS, với những thế mạnh của các công cụ OLAP thành một loại ứng dụng mới, ứng dụng Spatial OLAP hay SOLAP.

Việc nghiên cứu và triển khai ứng dụng Spatial OLAP đã và đang được triển khai trong nhiều lĩnh vực khác nhau ở các nước trên thế giới. Tại Việt Nam, vấn đề này tương đối còn mới mẻ, tuy nhiên, cũng đang được nghiên cứu và dần đưa vào triển khai trong thực tế.

Mặc khác, thành phố Đà Nẵng đang đẩy nhanh ứng dụng công nghệ thông tin trong công việc quản lý và điều hành của các cấp lãnh đạo trong đó có việc ứng dụng công nghệ GIS cho việc quản lý. Bên cạnh đó, việc cung cấp các số liệu kinh tế xã hội của thành phố và phân tích các số liệu theo nhu cầu quản lý gặp rất nhiều khó khăn. Như vậy, việc nghiên cứu công cụ phân tích trực tuyến các số liệu kinh tế xã hội của thành phố kết hợp với việc hiển thị số liệu trực quan trên bản đồ là nhu cầu cấp thiết.

Xuất phát từ những lý do trên, tôi chọn đề tài “Nghiên cứu Spatial OLAP xây dựng cổng thông tin bản đồ kinh tế xã hội TP Đà Nẵng” với mục tiêu tập trung vào việc nghiên cứu lý thuyết spatial OLAP để làm cơ sở, nền tảng cho việc xây dựng cổng thông tin bản đồ kinh tế xã hội phục vụ nhu cầu phân tích số liệu một cách trực quan nhằm đưa ra quyết định nhanh chóng, chính xác.

2. Mục đích nghiên cứu

Kết quả nghiên cứu nhằm mục đích nắm vững lý thuyết về kho dữ liệu, GIS, OLAP và SOLAP để xây dựng cổng thông tin bản đồ kinh tế xã hội (KTXH) TP Đà Nẵng trên cơ sở ứng dụng các công cụ phân tích dữ liệu trực trực quan, sinh động các số liệu tổng hợp KTXH

của TP Đà Nẵng nhằm phục vụ công tác chỉ đạo điều hành và nghiên cứu thống kê. Để hoàn thành mục đích, ý tưởng đề ra cần nghiên cứu các nội dung như sau:

- Nghiên cứu cơ sở lý thuyết: về kho dữ liệu, xử lý phân tích trực tuyến, hệ thống thông tin địa lý, dữ liệu không gian, kho dữ liệu không gian, spatial OLAP;

- Nghiên cứu chi tiết và xây dựng kho dữ liệu KTXH từ các số liệu của niên giám thống kê;

- Nghiên cứu công cụ hiển thị bản đồ và công cụ lập trình về Web-GIS;

- Xây dựng cổng thông tin bản đồ KTXH.

3. Đối tượng và phạm vi nghiên cứu

- **Đối tượng nghiên cứu**

Dữ liệu: bao gồm các nguồn dữ liệu thông tin tổng hợp KTXH của TP Đà Nẵng; các bản đồ địa giới hành chính của thành phố.

Công nghệ: bao gồm các công nghệ Web-GIS phục vụ hiển thị trực quan dữ liệu đầu ra, kho dữ liệu, OLAP, SOLAP phục vụ phân tích dữ liệu trên cơ sở tích hợp nhiều nguồn dữ liệu khác nhau, kỹ thuật xây dựng cổng thông tin điện tử cung cấp giao diện Web.

Các lý thuyết về hệ thống thông tin địa lý được sử dụng như nền tảng lý luận trong quá trình xây dựng hệ thống.

- **Phạm vi nghiên cứu**

Nghiên cứu số liệu các chỉ tiêu tổng hợp KTXH từ các nguồn niên giám thống kê của TP Đà Nẵng: Niên giám thống kê của thành phố, quận, huyện;

Nghiên cứu các bản đồ địa giới hành chính của TP Đà Nẵng;

Công nghệ Web-GIS, .NET Framework, SQL Server 2005, cổng thông tin mã nguồn mở DotnetNuke.

4. Phương pháp nghiên cứu

- **Phương pháp phân tích**

Thu thập và phân tích dữ liệu thông tin tổng hợp KTXH từ các niên giám thống kê đã được công bố.

Thu thập dữ liệu không gian: bản đồ địa giới hành chính của TP Đà Nẵng.

- **Phương pháp tiếp cận hệ thống**

Nghiên cứu tài liệu, ngôn ngữ và các công nghệ liên quan, phục vụ cho việc tìm hiểu và xây dựng: kho dữ liệu, công cụ xử lý phân tích dữ liệu trực tuyến, cổng thông tin điện tử, công nghệ xử lý bản đồ và ứng dụng spatial OLAP.

- **Phương pháp thực nghiệm**

Thực hiện việc cài đặt, thử nghiệm chọn lựa bộ công cụ, kết hợp việc xây dựng bản đồ và phân tích dữ liệu.

Kiểm tra, thử nghiệm sản phẩm đề tài và đưa ra nhận xét, đánh giá kết quả đạt được.

5. Ý nghĩa khoa học và thực tiễn của đề tài

- **Ý nghĩa về mặt khoa học**

Ứng dụng công nghệ thông tin vào việc nghiên cứu và phân tích dữ liệu KTXH TP Đà Nẵng;

Nghiên cứu và giới thiệu khả năng kết hợp giữa việc phân tích dữ liệu phi không gian với dữ liệu không gian trên bản đồ;

Giới thiệu khả năng kết hợp các công nghệ OLAP, GIS và cổng thông tin để xây dựng cổng thông tin bản đồ;

- **Ý nghĩa về thực tiễn**

Cung cấp kho dữ liệu các chỉ tiêu tổng hợp KTXH của TP Đà Nẵng phục vụ công tác nghiên cứu và phân tích số liệu KTXH;

Cung cấp giao diện đơn giản, dễ sử dụng, có tính linh hoạt cao nhằm hỗ trợ việc phân tích dữ liệu, ra quyết định một cách thuận tiện với kết quả hiển thị trực quan trên bản đồ và biểu đồ.

Sản phẩm phục vụ đắc lực, kịp thời và có độ chính xác cao cho các cán bộ lãnh đạo, cán bộ quản lý, các cán bộ làm công tác chuyên môn trong lĩnh vực quy hoạch, lập kế hoạch phát triển KTXH;

6. Cấu trúc luận văn

Nội dung chính của luận văn này được chia thành ba chương với nội dung như sau:

Chương 1 – Cơ sở lý thuyết: Nội dung chính là tìm hiểu, nghiên cứu lý thuyết liên quan đến vấn đề nghiên cứu của đề tài, gồm kho dữ liệu, xử lý phân tích trực tuyến (OLAP), hệ thống thông tin địa lý, cổng thông tin và spatial OLAP.

Chương 2 – Nghiên cứu dữ liệu KTXH TP Đà Nẵng và Xây dựng kho dữ liệu: Tiến hành thu thập, phân tích dữ liệu KTXH và xây dựng cấu trúc của kho dữ liệu phục vụ xây dựng ứng dụng.

Chương 3 – Xây dựng cổng thông tin bản đồ KTXH TP Đà Nẵng: Phân tích chức năng và xây dựng cổng thông tin đồng thời tiến hành chạy thử nghiệm, đánh giá kết quả.

CHƯƠNG 1
CƠ SỞ LÝ THUYẾT

1.1. KHO DỮ LIỆU

1.1.1. Khái niệm

Theo định nghĩa do W. H. Inman đưa ra: Kho dữ liệu (Data Warehouse) được hiểu là một "tập hợp dữ liệu tương đối ổn định, gắn thời gian, được tích hợp theo chủ đề nhằm hỗ trợ quá trình ra quyết định trong công tác quản lý" [6]. Kho dữ liệu có những đặc điểm sau:

Theo chủ đề (Subject-oriented): Không phải tất cả các dữ liệu đều được tập hợp vào kho dữ liệu, chỉ xử lý những dữ liệu có ích theo các chủ đề.

Tích hợp (Intergated): Đặc điểm nổi bật thứ hai của kho dữ liệu là tích hợp. Đây là đặc điểm quan trọng nhất [6].

Tính ổn định (Nonvolatile): Dữ liệu trong kho dữ liệu là một bản sao lưu trữ vật lý của dữ liệu được chuyển đổi từ môi trường tác nghiệp và được lưu trữ lâu dài.

Gắn thời gian (Time-variant): Dữ liệu trong kho được tạo ra và gắn liền với một giá trị thời gian nhất định.

1.1.2. So sánh cơ sở dữ liệu và kho dữ liệu

1.1.3. Kiến trúc kho dữ liệu

1.1.4. Nguồn dữ liệu

1.1.5. Kho dữ liệu chủ đề (Data Mart)

1.1.6. Bảng sự kiện (Fact Table)

Bảng sự kiện là bảng chứa các dữ liệu có tính chất đo lường (measurement), gọi là độ đo (measure). Độ đo mô tả một trong những khía cạnh định lượng của các sự kiện để phân tích**.** Hay nói khác, độ đo trong kho dữ liệu được dùng để minh họa cho một trường dữ liệu chứa một giá trị đo lường được.

Bảng sự kiện chứa các dữ liệu là đối tượng phục vụ cho việc phân tích, đánh giá về một lĩnh vực, một chủ đề.

1.1.7. Bảng chiều (Dimension Table)

Bảng chiều chứa các thông tin miêu tả nghiệp vụ. Trong không gian của mô hình đa chiều, các thông tin này được gọi là thuộc tính của chiều. Bảng chiều mô tả và bổ sung thông tin cho bảng sự kiện.

Bảng chiều đóng một vai trò quan trọng trong kho dữ liệu. Do bảng chiều chứa các thông tin miêu tả nghiệp vụ, nên đây là nguồn được dùng để sàng lọc dữ liệu, gộp nhóm, và tạo các nhãn trình bày.

1.1.8. Lược đồ dữ liệu

1.1.8.1. Lược đồ hình sao

1.1.8.2. Lược đồ bông tuyết

1.1.8.3. Lược đồ ngân hà

1.1.9. Quá trình ETL (Extract, Transform, Load)

1.2. XỬ LÝ PHÂN TÍCH TRỰC TUYẾN (OLAP)

1.2.1. Khái niệm

Xử lý phân tích trực tuyến (Online Analytical Processing – OLAP) là một kỹ thuật sử dụng các thể hiện dữ liệu đa chiều, gọi là khối (cube), nhằm cung cấp khả năng truy xuất nhanh đến dữ liệu của kho dữ liệu để phân tích và tạo lập các báo cáo đa chiều phục vụ việc ra quyết định.

1.2.2. Mô hình dữ liệu đa chiều

1.2.3. So sánh OLAP với OLTP

1.2.4. Các mô hình lưu trữ hỗ trợ OLAP

1.2.4.1. Mô hình Multidimensional OLAP (MOLAP)

1.2.4.2. Mô hình Relational OLAP (ROLAP)

1.2.4.3. Mô hình Hybird OLAP (HOLAP)

1.2.4.4. So sánh các mô hình lưu trữ OLAP

1.3. HỆ THỐNG THÔNG TIN ĐỊA LÝ (GIS)

1.3.1. Khái niệm

Hệ thống thông tin địa lý (Geographic Information System – GIS) là một công nghệ dựa vào máy tính để xây dựng bản đồ, phân tích và xử lý các đối tượng tồn tại và các sự kiện xảy ra trên trái đất.

1.3.2. Bản đồ

1.3.2.1. Các tính chất của bản đồ

1.3.2.2. Các yếu tố nội dung của bản đồ địa lý

1.3.3. Dữ liệu GIS

Có 2 dạng dữ liệu được sử dụng trong kỹ thuật GIS là dữ liệu không gian và phi không gian.

Dữ liệu không gian: là những mô tả hình ảnh bản đồ được số hoá theo một khuôn dạng nhất định mà máy tính hiểu được.

Dữ liệu phi không gian (dữ liệu thuộc tính): được trình bày dưới dạng các ký tự hoặc số, hoặc ký hiệu để diễn tả đặc tính, đặc điểm và các hiện tượng xảy ra tại các vị trí địa lý xác định

1.3.3.1. Mô hình dữ liệu không gian

1.3.3.2. Mô hình dữ liệu phi không gian

1.3.3.3. Mối quan hệ giữa dữ liệu không gian và phi không gian

1.3.4. Web-GIS

1.3.4.1. Khái niệm

1.3.4.2. Hoạt động của Web-GIS

1.3.4.3. Tiềm năng của Web-GIS

1.3.4.4. Các phương thức phát triển Web-GIS

1.4. SPATIAL OLAP (SOLAP)

1.4.1. Khái niệm

SOLAP có thể được định nghĩa là “một nền tảng trực quan được xây dựng đặc biệt nhằm hỗ trợ nhanh chóng, dễ dàng phân tích

và khai thác dữ liệu không gian, thời gian với cách tiếp cận đa chiều bao gồm mức độ kết hợp các thành phần hiển thị có sẵn trong bản đồ cũng như hiển thị ở dạng bảng và biểu đồ" [5].

Giao diện của một công cụ SOLAP gồm có hai phần chính: vùng hiển thị và bảng điều hướng. Vùng hiển thị cho phép người dùng xem các thông tin mong muốn trong các hình thức của một trong nhiều bản đồ, thống kê sơ đồ hoặc bảng với khả năng thay đổi thứ tự hiển thị để xem các thông tin trong cách có ý nghĩa nhất và với khả năng đồng bộ hóa tất cả các hiển thị. Tất cả các hiển thị là động với ý nghĩa người dùng có thể sử dụng các thao tác khác nhau để "định hướng" bên trong dữ liệu thông qua các thành phần hiển thị (ví dụ như một đa giác trên bản đồ, một quán bar ở một biểu đồ thanh, một tế bào trong một bảng).

1.4.2. Chiều không gian (Spatial Dimension)

Chiều không gian là chiều có một hoặc nhiều thuộc tính chứa kiểu dữ liệu có thể là chữ số hoặc hình học. Chiều không gian có 3 loại được sử dụng: chiều không gian phi hình học (non-geometric), chiều không gian hình học (geometric) và chiều không gian hỗn hợp (mixed hoặc geometric-to-non-geometric).

Chiều không gian phi hình học: là chiều chỉ chứa các dữ liệu phi hình học [4]. Dữ liệu chiều không gian chỉ là dữ liệu trên danh nghĩa như là phi hình học hoặc đại diện cho bản đồ có liên quan đến các thành phần chiều [5]. Hiện tại, chiều không gian phi hình học đã được sử dụng trong các công cụ OLAP truyền thống. Khi sử dụng với công cụ SOLAP, chiều không gian phi hình học sẽ thực hiện giống như các chiều mô tả khác và dữ liệu hình học cho phép biểu diễn các thành phần của chiều trên bản đồ sẽ không được sử dụng

Chiều không gian hình học: Tất cả các chiều từ mức thấp nhất đến mức cao hơn đều là dữ liệu hình học [4]. Chiều không gian hình học tối đa hoá tiềm năng khám phá mối quan hệ không gian và mối tương quan đó không ăn khớp với các ranh giới xác định trước.

Chiều không gian hỗn hợp: Các chiều không gian hỗn hợp bao gồm các hình hình học cho một tập hợp con của các mức chi tiết. Thành phần của chiều hình học và hỗn hợp có thể được hiển thị trên bản đồ bằng cách sử dụng các biến hình ảnh có liên quan đến các giá trị của độ đo khác nhau trong khối dữ liệu đang được phân tích [5].

1.4.3. Độ đo không gian (Spatial Measure)

Độ đo không gian gồm có 2 kiểu: kiểu thứ nhất là hình học và kiểu thứ hai là số. Độ đo không gian kiểu hình học là tập hợp các hình học biểu diễn các đối tượng không gian tương ứng với một kết hợp đặc biệt các thành phần chiều từ một đến nhiều chiều không gian hình học [5]. Độ đo không gian kiểu hình học bao gồm một tập hợp các toạ độ, yêu cầu một hoạt động hình học như là kết hợp các không gian, một không gian hợp nhất hoặc một không gian giao nhau, được tính toán. Một ví dụ của độ đo không gian hình học là một tập hợp các đa giác giao nhau kết hợp từ các đa giác thành phần của hai chiều không gian hình học như địa giới hành chính và con suối.

Độ đo không gian kiểu số là kết quả của việc tính toán độ dài không gian hoặc tính toán topo. Ví dụ của độ đo không gian kiểu số là "bề mặt" và "khoảng cách".

1.5. CỔNG THÔNG TIN (PORTAL)

1.5.1. Khái niệm

1.5.2. Phân loại cổng thông tin

1.5.3. Các tính năng của cổng thông tin

CHƯƠNG 2
NGHIÊN CỨU DỮ LIỆU KINH TẾ – XÃ HỘI TP ĐÀ NẴNG VÀ XÂY DỰNG KHO DỮ LIỆU

2.1. NGHIÊN CỨU DỮ LIỆU KTXH TP ĐÀ NẴNG

2.1.1. Hệ thống thông tin tổng hợp KTXH TP Đà Nẵng

2.1.1.1. Xác định nguồn dữ liệu thu thập

Hàng năm, Cục Thống kê thành phố Đà Nẵng biên soạn và xuất bản niên giám thống kê TP Đà Nẵng. Niên giám thống kê này cung cấp các số liệu phục vụ nhu cầu nghiên cứu về kinh tế – xã hội (KTXH) trên địa bàn TP Đà Nẵng.

Ngoài ra, nhằm phục vụ công tác nghiên cứu thông tin về KTXH trên địa bàn của từng quận, huyện phục vụ công tác chỉ đạo điều hành của lãnh đạo, phòng Thống kê quận, huyện tiến hành biên soạn và phát hành niên giám thống kê quận, huyện của mình.

Nghiên cứu dữ liệu KTXH chủ yếu tập trung nghiên cứu các số liệu được công bố trong các niên giám thống kê nhằm đảm bảo tính thống nhất và tin cậy của số liệu.

2.1.1.2. Nhóm số liệu trong niên giám thống kê quận, huyện

2.1.1.3. Nhóm số liệu trong niên giám thống kê thành phố

2.1.2. Sử dụng số liệu KTXH

2.1.2.1. Yêu cầu

Số liệu KTXH sử dụng từ các niên giám thống kê phải đảm bảo khả năng phục vụ cho việc phân tích đa chiều, trong đó có chiều không gian.

2.1.2.2. Phân tích, thu thập dữ liệu

Sau khi tiến hành tìm hiểu, phân tích chi tiết, cụ thể tất cả số liệu từ các niên giám thống kê và căn cứ vào tiêu chí thu thập dữ liệu,

các số liệu sẽ được xem xét, thu thập theo chủ đề. Đây là cơ sở để xây dựng các kho dữ liệu chủ đề và từ đó hình thành kho dữ liệu.

Dữ liệu dân số: Dữ liệu dân số trong niên giám thống kê là dữ liệu dân số trung bình. Dân số trung bình là dân số được tính bình quân cho một kỳ nghiên cứu nhất định, thường là một năm.

Dữ liệu dân số được cung cấp đầy đủ trong niên giám thống kê thành phố, quận, huyện. Niên giám thống kê thành phố cung cấp số liệu chi tiết về dân số đến cấp quận, huyện. Niên giám thống kê quận, huyện cung cấp số liệu dân số chi tiết đến cấp xã, phường. Do đó, nguồn dữ liệu phục vụ xây dựng kho dữ liệu là niên giám thống kê quận, huyện. Bảng 2.1 minh hoạ dữ liệu dân số trong cuốn Niên giám thống kê thành phố năm 2007 và Bảng 2.2 minh hoạ dữ liệu dân số trong cuốn Niên giám thống kê quận Cẩm Lệ năm 2007.

Bảng 2.1 Dân số trung bình năm 2007

(*Nguồn: Niên giám thống kê thành phố Đà Nẵng năm 2007* [2])

Đơn vị tính: người

	Tổng số	**Nam**	**Nữ**	**Thành thị**	**Nông thôn**
Tổng số	**806.744**	**393.335**	**413.409**	**699.834**	**106.910**
Hải Châu	195.105	93.885	101.220	195.105	-
Thanh Khê	167.287	82.442	84.845	167.287	-
Sơn Trà	119.968	58.602	61.366	119.968	-
Ngũ Hành Sơn	54.066	26.594	27.472	54.066	-
Liên Chiểu	95.088	47.060	48.028	95.088	-
Cẩm Lệ	68.320	33.153	35.167	68.320	-
Hoà Vang	106.910	51.599	55.311	-	106.910

Bảng 2.2 Dân số, diện tích và mật độ dân số năm 2007

(*Nguồn: Niên giám thống kê quận Cẩm Lệ năm 2007 [3]*)

	Dân số trung bình (người)		Diên tích (km2)	Mật độ (người/km2)
	Tổng số	Nữ		
Tổng số	**68.320**	**35.167**	**33,8**	**2.055**
Hoà An	13.476	6.744	3,2	4.257
Hoà Phát	10.263	5.248	6,6	1.566
Khuê Trung	15.383	7.898	3,0	5.106
Hoà Xuân	11.160	5.790	10,1	1.101
Hoà Thọ Đông	9.985	5.257	2,7	3.746
Hoà Thọ Tây	8.053	4.230	8,2	908

Dữ liệu tăng trưởng dân số: Có 2 yếu tố của sự tăng trưởng dân số sẽ được đề cập đến: tăng tự nhiên (sinh, chết) và tăng cơ học (chuyển đến, chuyển đi). Tuy nhiên, số liệu được thu thập là số liệu về tỷ lệ tăng tự nhiên.

Dữ liệu giá trị sản xuất các ngành: Bao gồm giá trị sản xuất công nghiệp, nông nghiệp và thương mại, dịch vụ.

Dữ liệu giá trị sản xuất công nghiệp: Giá trị sản xuất công nghiệp được tính theo giá thực tế của người sản xuất và giá so sánh.

Dữ liệu giá trị sản xuất nông nghiệp: Giá trị sản xuất nông nghiệp được tính theo giá thực tế và giá so sánh.

Dữ liệu y tế: Dữ liệu y tế cung cấp những số liệu tổng hợp về cơ sở y tế, số giường bệnh, số cán bộ y tế.

Dữ liệu giáo dục – đào tạo: là số liệu về số trường học, số lớp học, số học sinh và số giáo viên.

Dữ liệu tổng sản phẩm trong nước (GDP): là số liệu tổng hợp phản ánh kết quả cuối cùng của các hoạt động sản xuất, kinh doanh của nền kinh tế trong thời kỳ nhất định, thường là một năm.

Dữ liệu tổng mức bán hàng hóa và dịch vụ tiêu dùng xã hội: là số liệu tổng hợp phản ánh toàn bộ giá trị hàng hoá và doanh thu dịch tiêu dùng đã bán trực tiếp cho người tiêu dùng của các đơn vị có cơ sở kinh doanh thương nghiệp, dịch vụ trong một khoảng thời gian nhất đinh, thường là một năm.

Dữ liệu hoạt động du lịch trên địa bàn

Dữ liệu hoạt động xuất nhập khẩu

Dữ liệu biến động chỉ số giá cả hàng hóa: Chỉ số giá tiêu dùng là số liệu tương đối phản ánh xu hướng và mức độ biến động giá tiêu dùng cho sinh hoạt của đời sống của cá nhân và gia đình.

2.1.3. Đánh giá số liệu

Qua phân tích dữ liệu từ niên giám thống kê thành phố và niên giám thống kê các quận huyện, dữ liệu thu thập được, tuỳ vào lĩnh vực, chủ đề, có thể có dữ liệu phân cấp theo đơn vị hành chính cấp quận, huyện (có thể cấp xã, phường) hoặc dữ liệu ở cấp thành phố.

Đối với cấp số liệu chỉ có ở cấp thành phố, số liệu được thu thập từ niên giám thống kê thành phố. Đối với số liệu có đến cấp quận huyện, số liệu có thể được thu thập từ các niên giám thống kê thành phố, quận, huyện. Tuỳ vào từng chỉ tiêu, có thể sử dụng số liệu từ niên giám thống kê thành phố hoặc quận, huyện. Đối với số liệu có đến cấp xã phường, số liệu được thu thập từ các niên giám thống kê quận, huyện.

Thông qua việc phân tích và đánh giá số liệu, các số liệu thu thập được sẽ có sự thống nhất giữa các quận, huyện nhằm tạo thuận lợi cho công tác so sánh, phân tích sau này.

2.2. XÂY DỰNG KHO DỮ LIỆU

Căn cứ vào dữ liệu đã được phân chia theo các chủ đề, tiến hành xây dựng kho dữ liệu chủ đề; trên cơ sở đó, tích hợp các kho dữ liệu chủ đề để hình thành kho dữ liệu. Lược đồ sử dụng để xây dựng kho dữ liệu chủ đề là lược đồ hình sao.

2.2.1. Bảng chiều

Tuỳ vào lĩnh vực, chủ đề khác nhau, các kho dữ liệu chủ đề sẽ có các chiều cần phân tích là giống nhau hoặc khác nhau. Chiều được sử dụng hầu hết trong các kho dữ liệu chủ đề là chiều thời gian và địa phương. Với dữ liệu như Bảng 2.3, phân tích dữ liệu GDP được thực hiện với các chiều: thời gian, thành phần kinh tế và ngành kinh tế.

Bảng 2.3 Tổng sản phẩm quốc nội theo giá cố định 1994

(*Nguồn: Niên giám thống kê thành phố Đà Nẵng năm 2007 [2]*)

Đơn vị tính: triệu đồng.

	2004	**2005**	**2006**	**2007**
Tổng	**5.460.211**	**6.214.308**	**6.776.118**	**7.545.443**
Thành phần kinh tế				
Kinh tế nhà nước	3.073.381	3.476.894	3.437.074	3.788.286
Kinh tế ngoài nhà nước	1.834.588	2172.642	2.706.873	3.019.991
Kinh tế có vốn đầu tư nước ngoài	430.139	486.795	491.069	553.453
Thuế nhập khẩu	122.133	99.997	141.102	183.713
Ngành kinh tế				
Nông, lâm nghiệp, thuỷ sản	339.001	373.461	333.559	346.806
Công nghiệp, xây dựng	2.750.026	3.207.434	3.428.366	3.543.741
Dịch vụ	2.371.174	2.633.413	3.194.193	3.654.896

Cùng với các chủ đề được hình thành trong quá trình thực hiện phân tích, thu thập dữ liệu từ các niên giám thống kê, các bảng chiều cũng được hình thành trong kho dữ liệu.

2.2.1.1. Bảng chiều thời gian

2.2.1.2. Bảng chiều địa phương

2.2.1.3. Bảng chiều thành phần kinh tế

2.2.1.4. Bảng chiều ngành kinh tế

2.2.1.5. Bảng chiều loại cơ sở y tế

2.2.1.6. Bảng chiều cấp học

2.2.1.7. Bảng chiều loại trường

2.2.1.8. Bảng chiều loại khách

2.2.1.9. Bảng chiều giới tính

2.2.2. Kho dữ liệu chủ đề dân số

2.2.2.1. Đặc tả bảng sự kiện dân số

2.2.2.2. Lược đồ hình sao kho dữ liệu chủ đề dân số

2.2.3. Kho dữ liệu chủ đề tăng trưởng dân số

2.2.3.1. Đặc tả bảng sự kiện tăng trưởng dân số

2.2.3.2. Lược đồ hình sao kho dữ liệu chủ đề tăng trưởng dân số

2.2.4. Kho dữ liệu chủ đề giá trị sản xuất các ngành

2.2.4.1. Đặc tả bảng sự kiện giá trị sản xuất các ngành

2.2.4.2. Lược đồ hình sao kho dữ liệu chủ đề giá trị sản xuất các ngành

2.2.5. Kho dữ liệu chủ đề giá trị sản xuất công nghiệp

2.2.5.1. Đặc tả bảng sự kiện giá trị sản xuất công nghiệp

2.2.5.2. Lược đồ hình sao kho dữ liệu chủ đề giá trị sản xuất công nghiệp

2.2.6. Kho dữ liệu chủ đề giá trị sản xuất nông, lâm, thuỷ sản

2.2.6.1. Đặc tả bảng sự kiện giá trị sản xuất nông, lâm, thuỷ sản

2.2.6.2. Lược đồ hình sao kho dữ liệu chủ đề giá trị sản xuất nông, lâm, thuỷ sản

2.2.7. Kho dữ liệu chủ đề tổng sản phẩm trong nước (GDP)

2.2.7.1. Đặc tả bảng sự kiện tổng sản phẩm trong nước (GDP)

2.2.7.2. Lược đồ hình sao kho dữ liệu chủ đề GDP

2.2.8. Kho dữ liệu chủ đề tổng mức bán hàng hóa và dịch vụ xã hội

2.2.8.1. Đặc tả bảng sự kiện tổng mức bán hàng hóa và dịch vụ xã hội

2.2.8.2. Lược đồ hình sao kho dữ liệu chủ đề tổng mức bán hàng hóa và dịch vụ xã hội

2.2.9. Kho dữ liệu chủ đề kim ngạch xuất nhập khẩu

2.2.9.1. Đặc tả bảng sự kiện kim ngạch xuất nhập khẩu

2.2.9.2. Lược đồ hình sao kho dữ liệu chủ đề kim ngạch xuất nhập khẩu

2.2.10. Kho dữ liệu chủ đề y tế

2.2.10.1. Đặc tả bảng sự kiện y tế

2.2.10.2. Lược đồ hình sao kho dữ liệu chủ đề y tế

2.2.11. Kho dữ liệu chủ đề giáo dục

2.2.11.1. Đặc tả bảng sự kiện giáo dục

2.2.11.2. Lược đồ hình sao kho dữ liệu chủ đề giáo dục

2.2.12. Kho dữ liệu chủ đề du lịch

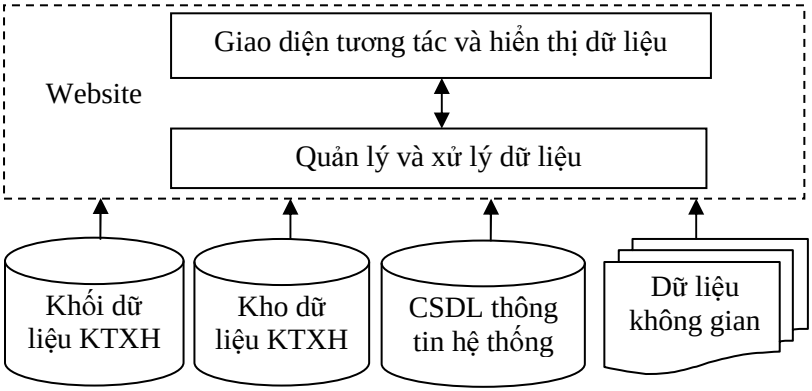
2.2.12.1. Đặc tả bảng sự kiện du lịch

2.2.12.2. Lược đồ hình sao kho dữ liệu chủ đề du lịch

CHƯƠNG 3
XÂY DỰNG CỔNG THÔNG TIN BẢN ĐỒ KINH TẾ XÃ HỘI THÀNH PHỐ ĐÀ NẴNG

3.1. MÔ HÌNH HỆ THỐNG

Cổng thông tin bản đồ KTXH thành phố Đà Nẵng cung cấp một giao diện thống nhất trên nền web phục vụ nhu cầu phân tích dữ liệu KTXH và hiển thị dữ liệu trực quan dưới dạng bảng, biểu đồ và bản đồ.



Hình 3.1 Mô hình hệ thống cổng thông tin bản đồ KTXH

Mô hình hệ thống cổng thông tin bản đồ KTXH bao gồm các khối như sau:

- Kho dữ liệu KTXH: Lưu trữ dữ liệu KTXH.

- Khối dữ liệu KTXH: Lưu trữ các khối dữ liệu đa chiều phụ vụ phân tích với nguồn dữ liệu là kho dữ liệu KTXH.

- Dữ liệu không gian: Các tập tin không gian lưu trữ ở dạng shapfile về địa giới hành chính các địa phương.

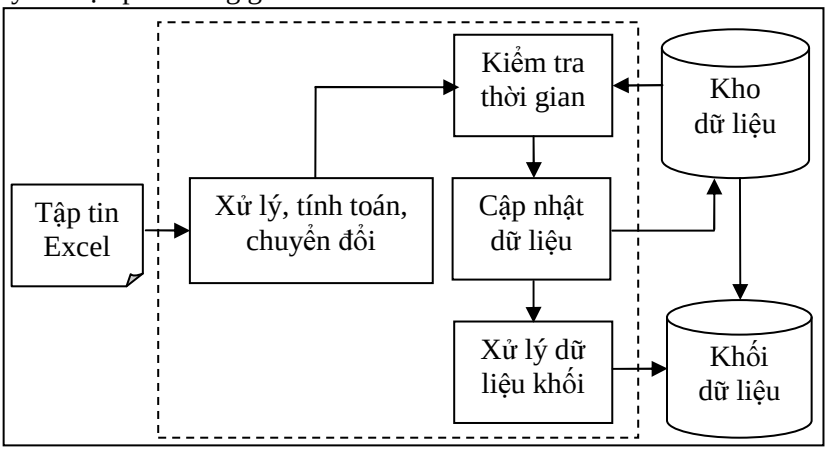
- CSDL thông tin hệ thống: Lưu trữ các thông tin để quản lý hệ thống như người dùng, phân quyền người dùng,...

- Thành phần website: Quản lý và xử lý dữ liệu, giao diện tương tác và hiển thị dữ liệu.

3.2. XÂY DỰNG CÁC CHỨC NĂNG

3.2.1. Thành phần chức năng tích hợp dữ liệu

Thành phần xử lý bản đồ bao gồm hai thành phần chức năng năng chính, đó thành phần xử lý dữ liệu không gian và thành phần xử lý dữ liệu phi không gian.

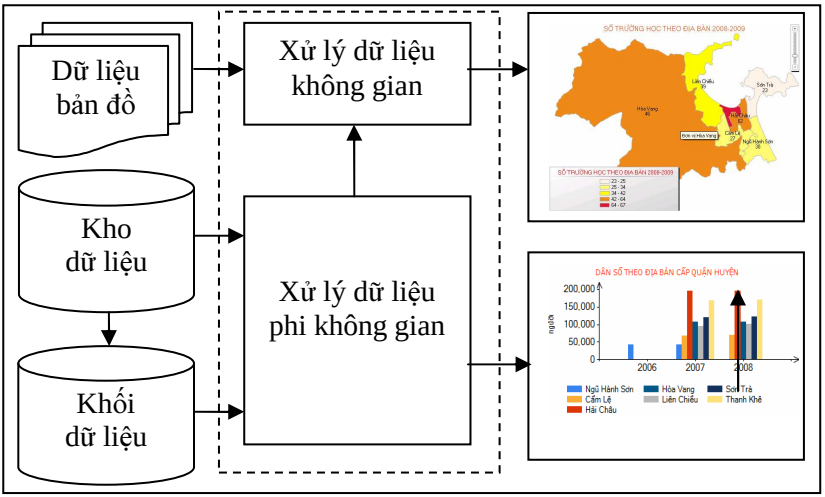


Hình 3.2 Mô hình hoạt động thành phần tích hợp dữ liệu

3.2.2. Thành phần xử lý bản đồ

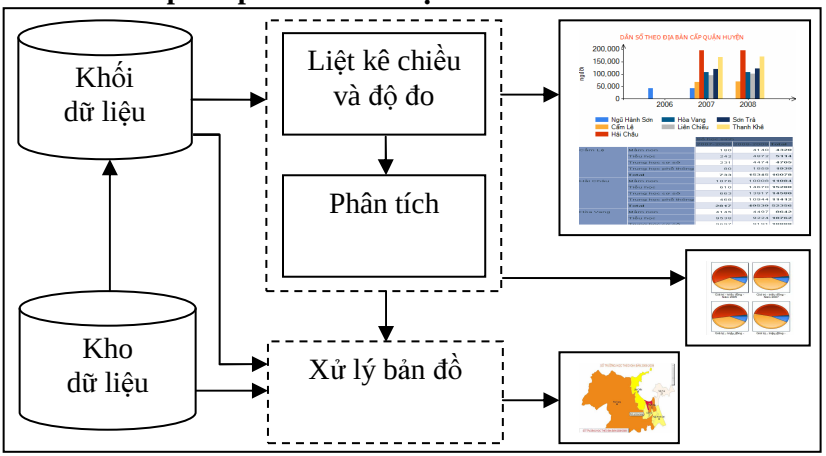
Thành phần chức năng xử lý dữ liệu không gian: thực hiện đọc dữ liệu không gian, hiển thị trên môi trường web và cho phép thực hiện các chức năng phóng đại (zoom), di chuyển (move).

Thành phần chức năng xử lý dữ liệu phi không gian: đảm bảo nạp dữ liệu phi không gian để thực hiện kết hợp với thành phần chức năng xử lý dữ liệu không gian để tạo ra một giao diện trực quan hướng dữ liệu trên bản đồ. Thành phần này cũng cung cấp chức năng xem số liệu dưới dạng các biểu đồ, bảng biểu số liệu nhằm chi tiết hoá dữ liệu để so sánh, đánh giá những số liệu hiển thị trên bản đồ.



Hình 3.3 Mô hình hoạt động thành phần xử lý bản đồ

3.2.3. Thành phần phân tích dữ liệu



Hình 3.4 Mô hình hoạt động thành phần phân tích số liệu

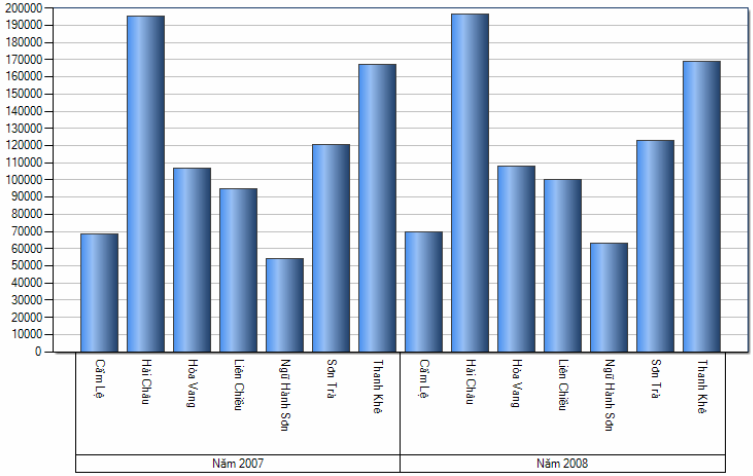
3.3. CÀI ĐẶT HỆ THỐNG

3.4. THỬ NGHIỆM HỆ THỐNG

3.4.1. Phân tích dữ liệu dân số

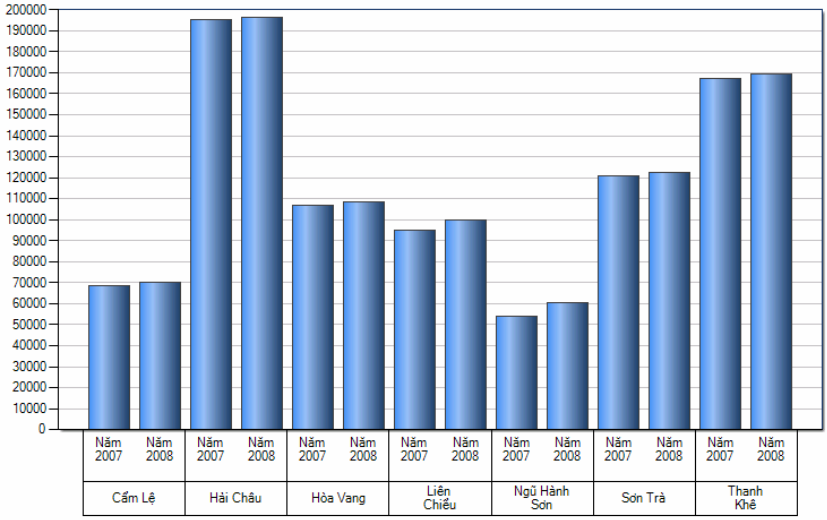
Phân tích dữ liệu về dân số dân số thể hiện trên bảng biểu, biểu đồ và bản đồ. Dữ liệu hiển thị dạng biểu đồ cung cấp khung nhìn trực

quan về dân số giữa các địa phương, ngoài ra còn so sánh sự tăng dân số qua các năm. Hình 3.5 minh hoạ phân tích dữ liệu dân số nhằm nhấn mạnh việc so sánh dân số giữa các quận, huyện trong các năm khác nhau. Hình 3.6 phân tích dữ liệu dân số nhằm nhấn mạnh việc so sánh dân số giữa các năm trong cùng một quận, huyện.

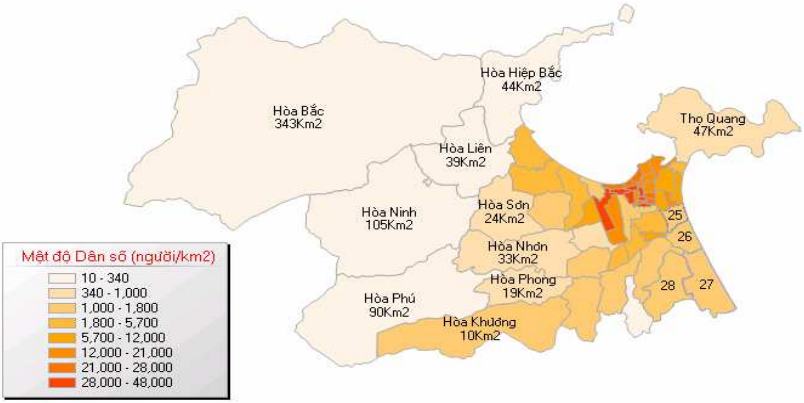


Hình 3.5 So sánh dân số giữa các quận, huyện cùng một năm

Bên cạnh dữ liệu dân số được hiển thị trên biểu đồ, dữ liệu dân số được biểu diễu trên bản đồ nhằm cung cấp cái nhìn trực quan đến sự phân bố dân cư trong thành phố. Hiển thị dữ liệu dân số trên bản đồ theo 10 khoảng dữ liệu, có nghĩa là những đơn vị cấp phường có cùng màu sắc thì dữ liệu có thể khác nhau nhưng thuộc cùng 1 màu. Hình 3.7 minh hoạ dữ liệu về sự phân bố dân cư hiển thị trên bản đồ. Hình 3.7 cho thấy dân cư đươc phân bố tập trung ở nội thành, thuộc quận Hải Châu, vùng màu đỏ trên bản đồ.



Hình 3.6 So sánh dân số giữa các năm của các quận, huyện



Hình 3.7 Phân bố dân cư thể hiện trên bản đồ

3.4.2. Phân tích dữ liệu GDP

3.4.3. Phân tích Giá trị sản xuất công nghiệp (SXCN)

3.5. ĐÁNH GIÁ KẾT QUẢ

Cổng thông tin bản đồ KTXH TP Đà Nẵng đã cung cấp một cái nhìn trực quan về số liệu tổng hợp KTXH TP Đà Nẵng; giúp so sánh dữ liệu để xem xét, phân tích, đánh giá về một chủ đề thuận lợi hơn.

Các chức năng được thử nghiệm dễ dàng, dễ sử dụng, hoạt động ổn định, hiển thị kết quả nhanh chóng và rất thuận lợi.

Với dữ liệu về KTXH được thu thập, tích hợp vào kho dữ liệu và xây dựng khối dữ liệu theo từng chủ đề, hệ thống đã cung cấp được những chức năng cơ bản nhằm tập trung vào việc nghiên cứu, phân tích số liệu nhằm xác định chiến lược phát triển KTXH và công tác chỉ đạo, điều hành liên quan đến lĩnh vực quy hoạch, lập kế hoạch phát triển KTXH TP Đà Nẵng.

KẾT LUẬN VÀ HƯỚNG PHÁT TRIỂN

1. Kết luận

Đề tài "Nghiên cứu Spatial OLAP xây dựng cổng thông tin bản đồ KTXH TP Đà Nẵng" là sự kết hợp giữa việc nghiên cứu các lý thuyết liên quan đến nhiều lĩnh vực và các dữ liệu thực tế về kinh tế xã hội để xây dựng cổng thông tin bản đồ kinh tế xã hội.

Về lý thuyết, luận văn đã đề cập một cách hệ thống các kiến thức cơ bản về các lĩnh vực: kho dữ liệu, xử lý phân tích dữ liệu trực tuyến, hệ thống thông tin địa lý, spatial OLAP và giới thiệu khả năng kết hợp các công nghệ OLAP, GIS và cổng thông tin để xây dựng cổng thông tin bản đồ KTXH.

Về thực tiễn ứng dụng, luận văn đã xây dựng kho dữ liệu các chỉ tiêu tổng hợp KTXH của TP Đà Nẵng và ứng dụng công nghệ OLAP trong phân tích dữ liệu. Đồng thời, xây dựng cổng thông tin bản đồ KTXH để phân tích dữ liệu từ kho dữ liệu, hiển thị dữ liệu trực quan trên bản đồ và biểu đồ với giao diện đơn giản, dễ sử dụng, có tính linh hoạt cao nhằm hỗ trợ ra quyết định một cách thuận tiện.

2. Hướng phát triển

Thu thập dữ liệu với khối lượng dữ liệu lớn hơn, đa dạng hơn như đa dạng về nguồn dữ liệu, dữ liệu thu thập nhiều năm, thời gian thu thập đến quý, tháng…

Phát triển thêm thành phần chức năng thu thập và tích hợp dữ liệu để có thể thu thập dữ liệu từ nhiều nguồn dữ liệu khác.

Cho phép xử lý phân tích trực tuyến một các đầy đủ với dữ liệu không gian để cho phép thể hiện kết quả phân tích độ đo không gian, phân tích nhiều chiều và phân cấp chiều trên bản đồ.