

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

LƯƠNG VĂN NHẬT

**NGHIÊN CỨU CƠ SỞ DỮ LIỆU ĐỐI TƯỢNG CHUYỂN ĐỘNG
VÀ ỨNG DỤNG**

LUẬN VĂN THẠC SĨ KHOA HỌC MÁY TÍNH

NGƯỜI HƯỚNG DẪN
TS. NGUYỄN NHƯ SƠN

THÁI NGUYÊN 2015

MỤC LỤC

LỜI CAM ĐOAN	iv
LỜI CẢM ƠN.....	v
DANH MỤC TỪ VIẾT TẮT	vi
MỞ ĐẦU	1
Chương 1. TỔNG QUAN VỀ CSDL ĐỐI TƯỢNG CHUYỂN ĐỘNG VÀ CÁC CÔNG CỤ PHÁT TRIỂN	3
1.1. Cơ sở dữ liệu.	3
1.1.1. Định nghĩa cơ sở dữ liệu	3
1.1.2. Ưu điểm của việc sử dụng CSDL	3
1.1.3. Tổ chức CSDL	4
1.1.4. Hệ quản trị cơ sở dữ liệu.	5
1.2. Cơ sở dữ liệu đối tượng chuyển động.....	6
1.2.1. Điểm chuyển động	7
1.2.2. Vùng chuyển động	9
1.2.3. Đặc tính của các đối tượng chuyển động.....	10
1.2.4. Mô hình dữ liệu khái quát cho đối tượng chuyển động.....	11
1.2.5. Công nghệ khai phá dữ liệu	11
1.3. Cơ sở dữ liệu không gian	12
1.3.1. Mô phỏng khái niệm không gian	13
1.3.2. Mở rộng mô hình dữ liệu và ngôn ngữ truy vấn.....	15
1.4. Cơ sở dữ liệu thời gian.....	18
1.4.1. Quản lý thời gian trong cơ sở dữ liệu dạng chuẩn	18
1.4.2. Miền thời gian	19
1.5. Dịch vụ công nghệ điện toán đám mây Google App Engine (GAE)...	19
1.5.1. Tổng quan về điện toán đám mây	19
1.5.2. Google App Engine (GAE).....	20
1.6. Hệ thống định vị toàn cầu GPS - Global Positioning System	27
1.6.1. Phần không gian:.....	28
1.6.2. Phần kiểm soát:	29
1.6.3. Phần sử dụng:	29
1.7. Kết luận	29

Chương 2. HỆ THỐNG LƯU TRỮ CSDL ĐỐI TƯỢNG CHUYỂN ĐỘNG	30
2.1. Tổng quan về hệ thống lưu trữ CSDL đối tượng chuyển động	30
2.2. Hệ thống lưu trữ CSDL đối tượng chuyển động Secondo.....	30
2.2.1. Giới thiệu.....	30
2.2.2. Kiến trúc	31
2.2.3. Quản lý dữ liệu đối tượng chuyển động:	36
2.2.4. Truy vấn dữ liệu	39
2.2.5. Thuật toán tối ưu hóa: Tìm đường đi ngắn nhất của một đồ thị thứ tự vị từ (predicate order graph - POG).....	41
2.3. Kết luận	46
Chương 3. THIẾT KẾ VÀ THỬ NGHIỆM TRÊN HỆ THỐNG LƯU TRỮ DỮ LIỆU CHUYỂN ĐỘNG	47
3.1. Giới thiệu bài toán.....	47
3.1.1. Cấu trúc chương trình	48
3.1.2. Phân tích hệ thống.....	49
3.1.3. Mô hình hoạt động	50
3.2. Cài đặt và cấu hình Secondo	50
3.2.1. Cài đặt Secondo trên Ubuntu 14.04	50
3.2.2. Cấu hình Secondo lần đầu trong Hệ điều hành Ubuntu 14.04.....	51
3.2.3. Giao diện hệ thống Secondo	52
3.3. Thử nghiệm	53
3.3.1. Thao tác trên Secondo	53
3.3.2. Thao tác trên thiết bị được thử nghiệm.....	54
3.4. Đánh giá	59
3.5. Kết luận	60
KẾT LUẬN VÀ HƯỚNG PHÁT TRIỂN.....	61
TÀI LIỆU THAM KHẢO	63
PHỤ LỤC:	64

LỜI CAM ĐOAN

Tôi xin cam đoan toàn bộ nội dung quyển luận văn này là do tôi tự sưu tầm, tra cứu và xây dựng đáp ứng nội dung yêu cầu của đề tài.

Nội dung bản luận văn chưa từng được công bố hay xuất bản dưới bất kỳ hình thức nào và cũng không được sao chép từ bất kỳ một công trình nghiên cứu nào.

Phần mã nguồn của chương trình do tôi thiết kế và xây dựng, trong đó có sử dụng một số thư viện chuẩn và các thuật toán được các tác giả xuất bản công khai và miễn phí trên mạng Internet.

Nếu sai tôi xin hoàn toàn chịu trách nhiệm.

Thái Nguyên, tháng 08 năm 2015

Tác giả

Lương Văn Nhất

LỜI CẢM ƠN

Sau một thời gian tìm hiểu và nghiên cứu, luận văn “*Nghiên cứu cơ sở dữ liệu đối tượng chuyển động và ứng dụng*” đã hoàn thành. Ngoài sự cố gắng của bản thân, tôi đã nhận được nhiều sự khích lệ từ phía nhà trường, thầy cô, gia đình và bạn bè.

Lời đầu tiên em xin bày tỏ lòng biết ơn sâu sắc đến thầy giáo - TS. Nguyễn Như Sơn – Viện Công nghệ thông tin thuộc Viện Hàn lâm Khoa học và Công nghệ Việt Nam là giáo viên hướng dẫn tôi đã tận tình giúp đỡ trong suốt thời gian làm luận văn.

Em xin chân thành cảm ơn các thầy, cô đã luôn nhiệt tình giúp đỡ, truyền đạt những kiến thức và kinh nghiệm quý báu trong suốt quá trình học tập.

Xin chân thành cảm ơn các anh, các chị và các bạn học viên học cùng lớp Cao học K11I đã luôn động viên, giúp đỡ và chia sẻ với tôi những kinh nghiệm học tập trong suốt khoá học.

Tôi xin chân thành cảm ơn Ban giám hiệu trường THCS Ninh Nhất, các thầy cô giáo trong trường đã tạo điều kiện tốt nhất để tôi có thể hoàn thành khoá học này.

Mặc dù đã có nhiều cố gắng, song do điều kiện nghiên cứu và trình độ, luận văn không tránh khỏi những khiếm khuyết. Em chân thành mong nhận được sự đóng góp ý kiến của các thầy, cô giáo và bạn bè đồng nghiệp gần xa.

Một lần nữa em xin cảm ơn!

Người thực hiện luận văn

Lương Văn Nhất

DANH MỤC TỪ VIẾT TẮT

- API	Application Programming Interface
- CSDL	Cơ sở dữ liệu
- DBMS	Database Management System
- GAE	Google App Engine
- GPS	Global Position System
- JVM	Java Virtual Machine
- KTG	Không gian-thời gian (Spatio-temporal)
- LBS	Location-based service
- MOD	Moving Objects Databases
- GUI	Graphic User Interface

DANH MỤC HÌNH

Hình 1.1. Kiểu dữ liệu không gian (point, line, region).....	6
Hình 1.2: Điểm thay đổi rời rạc (trái) - Điểm thay đổi liên tục (phải).	8
Hình 1.3: Trình diễn rời rạc của một điểm chuyển động.	9
Hình 1.4. Vùng thay đổi rời rạc (trái) - Vùng thay đổi liên tục (phải).	9
Hình 1.5. Công nghệ khai phá dữ liệu từ MOD.....	12
Hình 1.6. Điểm, đường và vùng	14
Hình 1.7. Phân vùng.....	15
Hình 1.8. Mạng.....	15
Hình 1.9. Kiến trúc hoạt động của GAE.	22
Hình 1.10: Phân bố các vệ tinh GPS	28
Hình 2.1. Kiến trúc hệ thống Secondo	31
Hình 2.2. Kiến trúc thô của nhân trong hệ thống Secondo	32
Hình 2.3 Mô hình quản lý dữ liệu chuyển động.....	36
Hình 2.4. Đồ thị POG với ba vị trí từ p , q , r	41
Hình 2.5. Hình xây dựng đệ quy của đồ thị POG	43
Hình 2.6. Hai truy vấn đồ thị	44
Hình 3.1. Mô hình bài toán tìm kiếm	47
Hình 3.2. Mô hình hoạt động của ứng dụng	50
Hình 3.3. Một số lệnh khởi tạo hệ thống Secondo	51
Hình 3.4. Giao diện hệ thống Secondo	53
Hình 3.5. Thực thi lệnh ifconfig trên Ubuntu.....	54
Hình 3.6. Giao diện chính.....	55
Hình 3.7. Giao diện theo dõi hành trình của thiết bị.....	56
Hình 3.8. Giao diện truy vấn thông tin địa điểm	57
Hình 3.9. Kết quả truy vấn.....	58
Hình 3.10. Kết quả tìm đường đi.....	59

MỞ ĐẦU

1. Lý do chọn đề tài:

Gần đây, các nhà nghiên cứu bắt đầu khám phá lĩnh vực về chuyển động của các đối tượng như nhiều phương tiện giao thông, điện thoại di động, ... Một số công việc (ví dụ, Geolife trong dự án Microsoft) sẽ cố gắng để phát hiện phương tiện giao thông ngoài trời, người sử dụng điện thoại di động, ... từ dữ liệu GPS để nắm bắt vị trí và đường đi của chúng. Vấn đề này thuộc xu hướng xây dựng các hệ thống CSDL qui mô cực lớn liên quan tới các thông tin thời gian và không gian đang được triển khai mạnh mẽ trên thế giới. Ví dụ cụ thể là các CSDL của đối tượng chuyển động (Moving Object Database - MOD), một dạng của CSDL không gian-thời gian (spatio-temporal - KTG).

Đề tài: “*Nghiên cứu cơ sở dữ liệu đối tượng chuyển động và ứng dụng*” tập trung tìm hiểu những khái niệm cơ bản về công nghệ CSDL đối tượng chuyển động dựa theo mã nguồn mở Secondo và ứng dụng xây dựng chương trình tìm kiếm địa điểm dựa trên việc xác định vị trí và đường đi của người sử dụng điện thoại di động.

2. Đối tượng và phạm vi nghiên cứu của đề tài:

- CSDL đối tượng chuyển động [6];
- Công nghệ CSDL MOD dựa theo mã nguồn mở Secondo [11];
- Công nghệ điện toán đám mây của Google [9];
- Dữ liệu GPS của thiết bị Smartphone.

3. Hướng nghiên cứu của đề tài:

- Nghiên cứu cơ sở khoa học, công nghệ CSDL đối tượng chuyển động dựa theo mã nguồn mở Secondo

- Lưu trữ dữ liệu GPS của thiết bị Smartphone lên máy chủ CSDL Secondo
- Cài đặt ứng dụng demo “*Xác định vị trí và đường đi của thiết bị Smartphone*”

4. Phương pháp nghiên cứu:

- Nghiên cứu cơ sở khoa học, công nghệ về CSDL đối tượng chuyển động dựa theo mã nguồn mở Secondo;
- Tìm hiểu nguồn thông tin từ các bài báo, Internet, ...
- Phương pháp thực nghiệm.

Chương 1. TỔNG QUAN VỀ CSDL ĐỐI TƯỢNG CHUYỂN ĐỘNG VÀ CÁC CÔNG CỤ PHÁT TRIỂN

1.1. Cơ sở dữ liệu.

1.1.1. Định nghĩa cơ sở dữ liệu

Cơ sở dữ liệu là tập hợp các dữ liệu có liên quan với nhau, được lưu trữ trên máy tính, có nhiều người sử dụng và được tổ chức theo một mô hình. Dữ liệu là những sự kiện có thể ghi lại được và có ý nghĩa. Ví dụ như:

- Họ và tên: Lương Văn Nhất.
- Ngày sinh: 22/8/1985.
- Địa chỉ: Xã Quỳnh Lưu, huyện Nho Quan, tỉnh Ninh Bình.
- Điện thoại: 0946864885.
- Tên cơ quan: Trường THCS Ninh Nhất.

Tất cả các CSDL đều có thể được biểu diễn bởi hệ thống các thực thể, thuộc tính và mối quan hệ giữa các thực thể.

1.1.2. Ưu điểm của việc sử dụng CSDL

Từ khái niệm trên, ta thấy rõ ưu điểm nổi bật của CSDL là:

- Giảm sự trùng lặp thông tin xuống mức thấp nhất và do đó đảm bảo được tính nhất quán và toàn vẹn dữ liệu.
- Đảm bảo sự độc lập giữa dữ liệu và chương trình ứng dụng: Cho phép thay đổi cấu trúc, dữ liệu trong cơ sở dữ liệu mà không cần thay đổi chương trình ứng dụng.
- Trừu tượng hoá dữ liệu (Data Abstraction): Mô hình dữ liệu được sử dụng để làm ẩn lưu trữ vật lý chi tiết của dữ liệu, chỉ biểu diễn cho người sử dụng mức khái niệm của cơ sở dữ liệu.
- Nhiều khung nhìn (multi-view) cho các đối tượng người dùng khác nhau: Đảm bảo dữ liệu có thể được truy xuất theo nhiều cách khác