

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

ĐẶNG BÌNH NINH

**TRUY VẤN THÔNG KÊ CƠ SỞ DỮ LIỆU MỘT CÁCH MỀM DẪO DỰA
TRÊN LOGIC MỜ**

GVHD: PGS.TS HỒ CẨM HÀ

LUẬN VĂN THẠC SĨ KHOA HỌC MÁY TÍNH

U

THÁI NGUYÊN 2015

LỜI CAM ĐOAN

Tôi xin cam đoan luận văn này của tự bản thân tôi tìm hiểu, nghiên cứu.

Các tài liệu tham khảo được trích dẫn và chú thích đầy đủ. Nếu không đúng tôi xin hoàn toàn chịu trách nhiệm.

Tác giả luận văn

Đặng Bình Ninh

LỜI CẢM ƠN

Trong thời gian nghiên cứu và thực hiện luận văn này, tôi đã nhận được sự giúp đỡ nhiệt tình của cơ quan, gia đình và các cá nhân. Tôi xin bày tỏ lời cảm ơn sâu sắc nhất tới tất cả các tập thể, cá nhân đã tạo điều kiện giúp đỡ tôi trong suốt quá trình thực hiện nghiên cứu luận văn này.

Trước hết tôi xin trân trọng cảm ơn Ban Giám Hiệu, các thầy giáo, cô giáo phòng Sau đại học trường Đại học Công Nghệ Thông Tin & Truyền Thông, cùng các thầy cô giáo, những người đã trang bị kiến thức cho tôi trong suốt quá trình học tập.

Đặc biệt, tôi xin bày tỏ sự kính trọng và lòng biết ơn chân thành nhất đến cô giáo - PGS. TS. Hồ Cẩm Hà, người cô đã tận tình hướng dẫn khoa học và giúp đỡ tôi trong suốt quá trình học tập, nghiên cứu và hoàn thành luận văn.

Tôi chân thành gửi lời cảm ơn đến các cán bộ, giảng viên khoa Giáo dục Mầm Non, trường Đại học Hồng Đức, Thanh Hóa, nơi tôi đang công tác đã tạo mọi điều kiện trong công việc để tôi có thể hoàn thành luận văn này.

Xin chân thành cảm ơn đến gia đình, người thân, bạn bè đã giúp đỡ và động viên tôi trong suốt thời gian học tập cũng như trong thời gian thực hiện luận văn.

Do thời gian nghiên cứu có hạn, luận văn của tôi không tránh khỏi thiếu sót, rất mong nhận được sự đóng góp của các thầy cô giáo cùng toàn thể bạn đọc.

Xin chân trọng cảm ơn!

Thanh Hóa, ngày 08 tháng 10 năm 2015

TÁC GIẢ LUẬN VĂN

Đặng Bình Ninh

MỤC LỤC

LỜI CAM ĐOAN.....	i
LỜI CẢM ƠN	iii
MỤC LỤC.....	iv
DANH MỤC CÁC KÝ HIỆU, CÁC CHỮ VIẾT TẮT.....	vi
DANH MỤC CÁC BẢNG VÀ HÌNH VẼ	vii
MỞ ĐẦU.....	1
CHƯƠNG I: TỔNG QUAN VỀ CƠ SỞ DỮ LIỆU	4
1.1 Cơ sở dữ liệu và các truy vấn	4
1.1.1 Khái niệm cơ sở dữ liệu.....	4
1.1.2 Mô hình CSDL quan hệ.....	6
1.1.2.1 Khóa của quan hệ [1].....	7
1.1.2.2 Phụ thuộc hàm	8
1.1.2.3 Đại số quan hệ	8
1.1.3 Các truy vấn trên mô hình quan hệ.....	11
1.1.3.1 SQL	11
1.1.3.2 MySQL	13
1.1.3.3 SQL server.	14
1.2 Ứng dụng lý thuyết tập mờ trong truy vấn CSDL.....	15
1.2.1 Tập mờ và thông tin không chắc chắn	15
1.2.1.1 Tập mờ [5]	15
1.2.1.2 Các kiểu hàm thuộc.....	16
1.2.2 Logic mờ.....	18
1.2.3 Lượng từ mờ	19
1.2.4 Ứng dụng logic mờ trong truy vấn CSDL	21
1.2.4.1 Cơ sở dữ liệu mờ.....	21
1.2.4.2 Cơ sở dữ liệu quan hệ	23
CHƯƠNG 2 MỘT PHƯƠNG PHÁP TRUY VẤN THỐNG KÊ MỀM DẪO ..	25

2.1 Quan niệm truy vấn mềm dẻo trên logic mờ	25
2.1.1 Khái niệm truy vấn mềm dẻo	25
2.1.2 Một số ngôn ngữ truy vấn CSDL mềm dẻo [7].....	26
2.2 Lượng từ hóa với ngôn ngữ tự nhiên	29
2.2.1 Lượng từ ngôn ngữ.....	29
2.2.2 Tóm tắt ngôn ngữ sử dụng logic mờ với lượng hóa ngôn ngữ tự nhiên.....	30
2.2.2.1 Giới thiệu mô hình toán tắt SAITETIQ [7]	30
2.2.2.2 Tóm tắt ngôn ngữ theo phương pháp tiếp cận của Yager.	33
2.3 Protoforms – Một phương pháp xây dựng thống kê mềm dẻo.....	38
2.4 Một truy vấn mờ thêm vào như một phương tiện để thực hiện thống kê bằng ngôn ngữ tự nhiên	41
2.5 Mô tả quá trình truy vấn	43
2.5.1. Biểu thức của truy vấn.....	43
2.5.2 Đánh giá truy vấn.....	45
2.5.3 Thuật toán lựa chọn	45
2.5.4 Phân loại	47
2.5.5 Từ tóm tắt đến các bộ trong một câu trả lời.	49
CHƯƠNG 3 CÀI ĐẶT THỬ NGHIỆM.....	50
3.1 Mô tả mô hình và giải pháp sử dụng trong truy vấn mềm dẻo.....	50
3.1.1 Mô hình truy vấn mềm dẻo	50
3.1.2 Giải pháp được sử dụng trong truy vấn mềm dẻo	51
3.2 Nội dung thử nghiệm	52
3.2.1 Xác định miền tham chiếu và xác định điều kiện mờ.....	53
3.2.2 Xử lý truy vấn.	54
3.2.3 Lọc dữ liệu	54
3.3 Thuật toán	54
3.4 Kết quả thực hiện.....	59
KẾT LUẬN	65
TÀI LIỆU THAM KHẢO.....	66
Tiếng việt.....	66

DANH MỤC CÁC KÝ HIỆU, CÁC CHỮ VIẾT TẮT

Từ viết tắt	Nghĩa tiếng anh	Nghĩa tiếng việt
DBMS	Database Managemet System	Hệ quản trị cơ sở dữ liệu
FQUERY	Fuzzy Query	Công cụ truy vấn mờ
LL	Lower Limit	Giới hạn dưới
UL	Upper Limit	Giới hạn trên
CSDL		Cơ sở dữ liệu
CNTT		Công nghệ thông tin

DANH MỤC CÁC BẢNG VÀ HÌNH VẼ

<i>Bảng 2.1: Năm loại tóm tắt cơ bản.</i>	39
<i>Bảng 2.2 Bảng quan hệ</i>	48
<i>Hình 1.1. Hàm thuộc hình thang</i>	17
<i>Hình 1.2 Hàm thuộc hình tam giác</i>	18
<i>Hình 1.3 Tập mờ Q và Q^*</i>	20
<i>Hình 2.1 Biến ngôn ngữ cho bảng VẬT LIỆU</i>	31
<i>Hình 2.2 Một phần của bảng VẬT LIỆU</i>	32
<i>Hình 2.3 Một phần của bản tóm tắt hệ thống phân cấp cho VẬT LIỆU</i>	32
<i>Hình 2.4 Ví dụ về hàm thành viên của một giá trị mờ</i>	Error! Bookmark not defined.
<i>Hình 2.5 Thành phần của một truy vấn mờ [8]</i>	42
<i>Hình 2.6 đoạn chương trình mô phỏng [7]</i>	47
<i>Hình 3.1 Mô hình truy vấn mềm dẻo</i>	51
<i>Hình 3.2 Hình minh họa việc nhập khái niệm mờ</i>	54
<i>Hình 3.3 Sơ đồ khối thuật toán thiết lập thuộc tính mờ</i>	56
<i>Hình 3.4 Sơ đồ khối thuật toán cập nhật các giá trị mờ</i>	57
<i>Hình 3.5 Sơ đồ khối thuật toán thực hiện câu lệnh truy vấn</i>	58
<i>Hình 3.6 Màn hình menu chính</i>	59
<i>Hình 3.7 Xác định miền tham chiếu</i>	60
<i>Hình 3.8 Nhập khái niệm mờ cho thuộc tính trẻ</i>	60
<i>Hình 3.9 Nhập khái niệm mờ cho thuộc tính cao</i>	61
<i>Hình 3.10 Chọn điều kiện để thực hiện thống kê CSDL</i>	62
<i>Hình 3.11 Thực hiện truy vấn và hiển thị kết quả</i>	63

MỞ ĐẦU

1. Lý do chọn đề tài

Mô hình cơ sở dữ liệu quan hệ do Codd E.F đề xuất từ năm 1970 đã đạt được những kết quả hoàn chỉnh về lý thuyết và ứng dụng, cho đến nay các mô hình cơ sở dữ liệu quan hệ vẫn được dùng rất phổ biến.

Với lượng dữ liệu được lưu trữ trong cơ sở dữ liệu, để có thể tìm kiếm được những thông tin hữu ích và cần thiết chúng ta cần phải sử dụng đến hệ thống tìm kiếm tự động dựa trên mô hình cơ sở dữ liệu quan hệ với các câu lệnh truy vấn cho phép người dùng truy xuất dữ liệu trong cơ sở dữ liệu. Nếu như không có truy vấn thì dữ liệu lưu trữ đó không thể sử dụng được, giống như việc chúng ta cất giữ một món đồ nhưng không thể tìm thấy hoặc không dùng đến thì việc lưu trữ đó trở nên vô nghĩa.

Tuy nhiên mô hình đó sẽ gặp một số hạn chế trong trường hợp tìm kiếm những thông tin mơ hồ, không đầy đủ, không chắc chắn (gọi chung là dữ liệu mờ), loại dữ liệu này được con người sử dụng thường xuyên trong thực tế. Ví dụ như thay vì muốn có “Danh sách nhân viên có độ tuổi từ 25 đến 30 trong công ty” thì chúng ta lại cần biết “Danh sách những nhân viên trẻ có công trình khoa học được công bố trên tạp chí uy tín thế giới”, hoặc là, thay vì kiểm tra xem “Có bao nhiêu nhân viên được khen thưởng” thì chúng ta cần biết “Có phải hầu hết những nhân viên có tiềm năng đều được đào tạo ở nước ngoài hay không”...

Đã có rất nhiều công trình nghiên cứu đưa ra nhiều cách tiếp cận với các câu hỏi liên quan đến thông tin không đầy đủ, không chính xác trên các cơ sở dữ liệu kinh điển (thường được gọi là hỏi mềm dẻo cơ sở dữ liệu) nhằm giải quyết nhu cầu truy vấn trong những tình huống nêu trên, đặc biệt là những câu hỏi mang tính thống kê để nhận định tình hình của người làm quản lý. Các cách xử lý loại truy vấn như vậy được đề xuất chủ yếu dựa trên cơ sở tiếp cận của lý thuyết tập mờ, hầu hết cách tác giả đều xây dựng ngôn ngữ truy vấn với mong muốn thao tác mềm dẻo, chính xác với dữ liệu mờ. Và người đi tiên phong trong

lĩnh vực này là Lotfi A.Zadeh, ông đã đề xuất khái niệm mờ từ những khái niệm mơ hồ, không rõ ràng, không chắc chắn và hình thức toán học nó bằng tập mờ (*Fuzzy set*), trên cơ sở đó lý thuyết tập mờ được hình thành. Dựa trên lý thuyết tập mờ của Lotfi A.Zadeh các nhà khoa học đã tiếp cận và phát triển theo nhiều hướng khác nhau. Chúng ta có thể tìm thấy các kết quả này qua các công trình của D. Dubois, H. Prade, C.S. George Lee, H.J. Zimmermann, T.J. Ross, R. Fuller, J.J. Buckley, R. Kruse, D. Nauck, N.K. Kasabov, W. Pedrycz,... Và rất nhiều bài báo đã được đăng trên các tạp chí uy tín quốc tế như: “*Querying a summary of database*” của W. A. Voglozin, G. Raschia, L. Ughetto, N. Mouaddib, và một bài báo khác là “*Linguistic database summaries and their protoforms: towards natural language based knowledge discovery tools*” của Janusz Kacprzyk, Sławomir Zadrozny.

Trên những cơ sở về lý thuyết và thực tiễn nói trên, với ý nghĩa khoa học và khả năng ứng dụng của chủ đề đã đề cập trên đây, em chọn đề tài “***Truy vấn thống kê cơ sở dữ liệu một cách mềm dẻo dựa trên logic mờ***” làm đề tài cho luận văn tốt nghiệp của mình.

2. Đối tượng, phạm vi nghiên cứu

*** Đối tượng nghiên cứu:**

- Lý thuyết logic mờ và các mô hình ứng dụng của logic mờ.
- Truy vấn kết tập và CSDL quan hệ.

*** Phạm vi nghiên cứu:**

- Về mặt lý thuyết: Luận văn tập trung nghiên cứu các kiến thức có liên quan đến logic mờ đặc biệt là những truy vấn mềm dẻo (mang tính ngôn ngữ) - truy vấn mờ.
- Về mặt ứng dụng: Cài đặt thử nghiệm một module xử lý truy vấn mềm dẻo khi khai thác một CSDL được quản trị bởi MS. Access .

3. Hướng nghiên cứu của đề tài.

- Tìm hiểu các cách tiếp cận, các quan điểm về truy vấn mềm dẻo CSDL đã có.