

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

PHẠM VIỆT HOÀI

BẢO VỆ CƠ SỞ DỮ LIỆU BẰNG KỸ THUẬT THỦY VĂN DỰA
VÀO GIẢI THUẬT DI TRUYỀN VÀ THUẬT TOÁN MỞ RỘNG
HIỆU

LUẬN VĂN THẠC SĨ KHOA HỌC MÁY TÍNH

Thái Nguyên - 2015

MỤC LỤC

LỜI CẢM ƠN	5
CHƯƠNG 1.TỔNG QUAN	7
1.1.Một số khái niệm cơ bản	7
1.1.1.Khái niệm cơ sở dữ liệu	7
1.1.2.Mô hình cơ sở dữ liệu quan hệ.....	7
1.1.3.Thủy văn số	10
1.1.4.Thủy văn CSDL quan hệ	13
1.2. Các tính chất của thủy văn trên CSDL quan hệ.....	15
1.2.1.Khả năng có thể phát hiện	15
1.2.2.Tính bền vững và dễ vỡ	16
1.2.3.Khả năng cập nhật phần tăng thêm	16
1.2.4.Không dễ cảm nhận được	16
1.3.Một số phép tấn công trên hệ thủy văn CSDL quan hệ	16
1.3.1.Cập nhật thông thường	17
1.3.2.Tấn công có chủ đích.....	17
1.4.Các ứng dụng chủ yếu của thủy văn trên CSDL quan hệ	17
1.4.1.Bảo vệ bản quyền CSDL	18
1.4.2.Xác thực tính toàn vẹn của CSDL.....	18
1.4.3.Dán nhãn	18
1.4.4.Điều khiển truy cập	18
1.4.5.Điều khiển sao chép.....	19
1.5.Thủy văn CSDL không thuận nghịch bằng phương pháp LSB	19
1.5.1.Thuật toán nhúng thủy văn	21
1.5.2.Thuật toán trích dấu thủy văn	22
CHƯƠNG 2. THỦY VĂN CSDL QUAN HỆ BẰNG GIẢI THUẬT DI TRUYỀN VÀ PHƯƠNG PHÁP MỞ RỘNG	24
1.6.Một số phương pháp thủy văn trên CSDL quan hệ	24
1.7.Phương pháp mở rộng hiệu.....	26

1.7.1.Phép biến đổi Haar nguyên trên một cặp giá trị	26
1.7.2.Phương pháp mở rộng hiệu.....	28
1.8.Thủy vân mở rộng hiệu kết hợp với thuật toán di truyền	39
1.8.1.Giải thuật di truyền.....	40
1.8.2.Hàm thích nghi	42
1.8.3.Nhúng thủy vân	44
1.8.4.Trích dấu thủy vân.....	46
1.8.5.Phân tích độ phức tạp tính toán.....	47
CHƯƠNG 3.XÂY DỰNG PHẦN MỀM THỦY VÂN CƠ SỞ DỮ LIỆU	
QUAN HỆ	49
1.9.Tổng quan về .NET	49
1.9.1.Khái niệm .NET	49
1.9.2.Visual Studio.Net	50
1.9.3.Giới thiệu ngôn ngữ lập trình C#	51
1.10.Hệ quản trị cơ sở dữ liệu SQL Service 2008	53
1.10.1.Tổng quan về SQL.....	53
1.10.2.Hệ quản trị cơ sở dữ liệu SQL Server 2008	55
1.11.Mô hình hoạt động của phần mềm.....	55
1.12.Một số giao diện của phần mềm	57
KẾT LUẬN	59

LỜI CẢM ƠN

Để hoàn thành được luận văn này tôi đã nhận được rất nhiều sự động viên, giúp đỡ của nhiều cá nhân và tập thể.

Trước hết, tôi xin bày tỏ lòng biết ơn sâu sắc và sự tôn kính đến

PGS. TS. Phạm Văn Ất đã hướng dẫn tôi thực hiện nghiên cứu luận văn của mình.

Xin cùng bày tỏ lòng biết ơn chân thành tới TS Đỗ Văn Tuấn các thầy cô giáo, người đã đem lại cho tôi những kiến thức bổ trợ, vô cùng có ích trong trong thời gian thực hiện đề tài và trong những năm học vừa qua.

Đồng thời xin gửi lời cảm ơn chân thành tới Ban Giám hiệu, Phòng Đào tạo sau đại học, Đại học công nghệ thông tin & truyền thông đã tạo điều kiện cho tôi trong quá trình học tập.

Cuối cùng tôi xin gửi lời cảm ơn đến gia đình, bạn bè, những người đã luôn bên tôi, động viên và khuyến khích tôi trong quá trình thực hiện đề tài nghiên cứu của mình.

ĐẶT VẤN ĐỀ

Truy cập Internet một cách dễ dàng đã thúc đẩy sự tăng trưởng của nhiều ngành nghề khác nhau và nhiều lĩnh vực khác. Ngày nay, việc chia sẻ thông tin trực tuyến là một hoạt động quan trọng đối với kinh doanh và nghiên cứu, không những vậy, nó còn liên quan đến việc mua bán của cơ sở dữ liệu. Ví dụ, chia sẻ dữ liệu liên quan đến thời tiết, thị trường chứng khoán, tiêu thụ điện năng, y tế, khoa học... vv được thực hiện thường xuyên. Do đó, có một nhu cầu lớn trong việc cung cấp bảo mật cơ sở dữ liệu để ngăn cản việc sao chép và phân phối bất hợp pháp trong môi trường internet. Kỹ thuật thủy văn đang được xem là một trong những giải pháp đạt được kết quả cao trong việc bảo vệ bản quyền và xác thực tính toàn vẹn của dữ liệu số. Vì vậy những năm gần đây xuất hiện một hướng nghiên cứu mới là thủy văn thuận nghịch Trong giấu tin và thủy văn thuận nghịch thường sử dụng các phép biến đổi nguyên khả nghịch như dịch chuyển Histogram, Wavelet nguyên, phép biến đổi tương phản, phép biến đổi theo thuật toán di truyền, phép biến đổi mở rộng hiệu, Trong số đó kỹ thuật giấu tin và thủy văn dựa trên phép biến đổi theo giải thuật di truyền và mở rộng hiệu đạt được hiệu quả cao nên đang được nhiều người quan tâm. Chính vì vậy em chọn đề tài “*Bảo vệ CSDL bằng kỹ thuật thủy văn dựa vào giải thuật di truyền và thuật toán mở rộng hiệu*” làm luận văn tốt nghiệp của mình.

Nội dung luận văn được tổ chức thành 3 chương: Chương 1. Tổng quan, Chương 2. Thủy văn CSDL bằng giải thuật di truyền và phương pháp mở rộng hiệu, Chương 3. Phần mềm thủy văn cơ sở dữ liệu quan hệ.

CHƯƠNG 1. TỔNG QUAN

1.1. Một số khái niệm cơ bản

Phần này trình bày một số khái niệm cơ bản về cơ sở dữ liệu, thủy văn số nói chung và thủy văn cơ sở dữ liệu nói riêng.

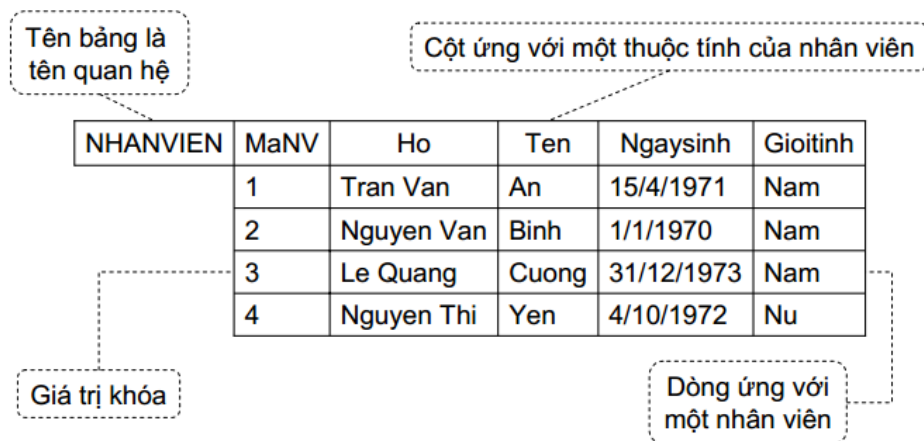
1.1.1. Khái niệm cơ sở dữ liệu

Cơ sở dữ liệu là bộ sưu tập về các loại dữ liệu tác nghiệp bao gồm: chữ viết, văn bản, đồ họa, hình ảnh,... được lưu trữ dưới dạng file dữ liệu trên băng từ, đĩa từ,... Cấu trúc lưu trữ dữ liệu tuân theo các nguyên tắc dựa trên lý thuyết toán học. Cơ sở dữ liệu là tài nguyên thông tin dùng chung cho nhiều người, bất kỳ người dùng nào có quyền sử dụng, khai thác toàn bộ hay một phần đều có thể sử dụng và khai thác nó nhằm thỏa mãn nhu cầu khai thác thông tin

1.1.2. Mô hình cơ sở dữ liệu quan hệ

Mô hình cơ sở dữ liệu quan hệ được tiến sĩ E.F. Codd phát triển vào năm 1970. Loại cơ sở dữ liệu này đã loại bỏ được những vấn đề liên quan đến các thiết kế cơ sở dữ liệu khác. Sử dụng mô hình này có thể giảm thiểu việc dư thừa dữ liệu, tiết kiệm được không gian lưu trữ và có khả năng truy xuất dữ liệu hiệu quả. Hiện nay cơ sở dữ liệu quan hệ đang được sử dụng phổ biến cho các ứng dụng.

Mô hình cơ sở dữ liệu quan hệ lưu trữ dữ liệu trên một hoặc nhiều bảng, mỗi bảng gồm nhiều hàng (row), nhiều cột (column) và được gọi là một quan hệ. Trong thực tế các hàng, các cột được gọi tên theo thuật ngữ là bản ghi (record) và trường (field).

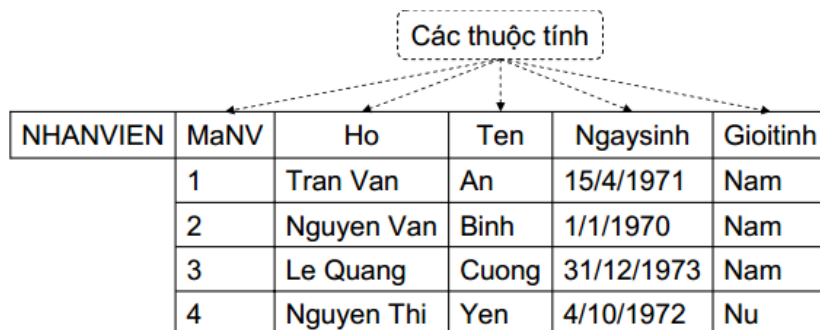


Hình 1.1.Biểu diễn quan hệ

- **Thực thể:** là đối tượng có trong thế giới thực mà ta cần mô tả các đặc trưng của nó.

Ví dụ: Thực thể NHANVIEN. Các đặc trưng của thực thể NHANVIEN gồm: Mã nhân viên (1), Họ (Trần Văn), Tên (An) Ngày sinh (15/4/1971),...

- **Thuộc tính:** Là tên các cột của quan hệ, các dữ liệu thể hiện các đặc trưng của thực thể, mỗi thuộc tính có một kiểu dữ liệu tương ứng.



Hình 1.2.Biểu diễn thuộc tính

- **Miền thuộc tính:** Tập các giá trị có thể có của thuộc tính gọi là miền giá trị của thuộc tính đó gồm có tên, kiểu dữ liệu, khuôn dạng,...

Ví dụ: Ten: kiểu dữ liệu và Text, Ngaysinh, kiểu dữ liệu: date/time, khuôn dạng:dd/mm/yyyy

- Bộ dữ liệu: Là các dòng (bản ghi) của quan hệ không kể dòng tên của các thuộc tính, mỗi bộ chứa giá trị cụ thể của các thuộc tính.

NHANVIEN	MaNV	Ho	Ten	Ngaysinh	Gioitinh
	1	Tran Van	An	15/4/1971	Nam
	2	Nguyen Van	Binh	1/1/1970	Nam
	3	Le Quang	Cuong	31/12/1973	Nam
	4	Nguyen Thi	Yen	4/10/1972	Nu

Bộ dữ liệu

Hình 1.3.Biểu diễn bộ dữ liệu

- Quan hệ: Cho $R = \{A_1, A_2, \dots, A_n\}$ (trong đó R là tên quan hệ, A_1, A_2, \dots, A_n là các thuộc tính) là một tập hữu hạn các thuộc tính khác rỗng. Mỗi thuộc tính $A_i \in R$; $i = 1 \div n$ có một miền giá trị xác định.

Mô hình cơ sở dữ liệu quan hệ thường biểu diễn một quan hệ trên tập các thuộc tính bằng bảng 2 chiều. Các thuộc tính của quan hệ là các cột A_1, A_2, \dots, A_n , các hàng của quan hệ được gọi là các bộ (bản ghi), ký hiệu r và giá trị của mỗi bộ gồm m thành phần. Quan hệ r được biểu diễn bởi bảng sau:

	A_1	A_2	...	A_n
r_1	$r_1. A_1$	$r_1. A_2$...	$r_1. A_n$
r_2	$r_2. A_1$	$r_2. A_2$...	$r_2. A_n$
...
r_m	$r_m. A_1$	$r_m. A_2$...	$r_m. A_n$

Ví dụ:

SINHVIEN(MaSV	Ten	Ngaysinh	Điachi)
HCN01	Hải	12/10/1987	Hà Nội
HCN02	Bình	06/08/1989	Hải Phòng
HCN03	Minh	28/10/1990	Nghệ An
HCN04	Hoa	17/09/1991	Quảng Ninh

Trong đó các thuộc tính là MaSV: Mã sinh viên; Ten: tên; Ngaysinh: Ngày sinh; Diachi: Địa chỉ. Bộ giá trị là (HCN01, Hải, 12/10/1987, Hà Nội).

- Lược đồ quan hệ: Tập tất cả các thuộc tính trong một quan hệ cùng với mối liên hệ giữa chúng được gọi là lược đồ quan hệ.

- Khóa của quan hệ: Trong một lược đồ quan hệ luôn tồn tại một thuộc tính hay một tập các thuộc tính có khả năng biểu diễn duy nhất các thuộc tính còn lại. Nói cách khác giá trị của một hay nhiều thuộc tính có thể xác định duy nhất giá trị các thuộc tính khác. Ví dụ Số chứng minh thư sẽ xác định được tất cả thông tin về người mang chứng minh thư đó. Tập thuộc tính có tính chất như trên được gọi là khoá của lược đồ quan hệ. Thông thường trong một lược đồ quan hệ có thể tồn tại nhiều khoá. Trong số đó sẽ chọn một khoá làm khoá chính sao cho đơn giản và không nhập nhằng thông tin. Giá trị các thành phần của khoá không thể nhận giá trị null hay các giá trị không xác định.

Khóa của quan hệ R xác định trên tập thuộc tính $R = \{A_1, A_2, \dots, A_n\}$ là tập con $K \subset R$ sao cho bất kỳ hai bộ khác nhau $t_1, t_2 \in r$ luôn thỏa $t_1.K \neq t_2.K$.

Khóa là khái niệm quan trọng trong thiết kế cơ sở dữ liệu quan hệ. Khóa thường được áp dụng trong việc tìm kiếm và cập nhật dữ liệu trong các quan hệ cơ sở dữ liệu.

1.1.3. Thủy vân số

Thủy vân số là kỹ thuật nhúng thêm một lượng thông tin (dấu thủy vân) vào dữ liệu đa phương tiện như: văn bản, hình ảnh, âm thanh và cơ sở dữ liệu (dữ liệu đa phương tiện). Việc nhúng dấu thủy vân vào dữ liệu đa phương tiện có thể làm giảm chất lượng dữ liệu nhưng nó là dấu vết để chứng minh quyền sở hữu hoặc dùng để phát hiện ra sự đối trái phép trên dữ liệu trong quá trình trao đổi thông tin.