

BỘ GIÁO DỤC VÀ ĐÀO TẠO  
TRƯỜNG ĐẠI HỌC QUẢN LÝ VÀ CÔNG NGHỆ HẢI PHÒNG

---



# ĐỒ ÁN TỐT NGHIỆP

NGÀNH : CÔNG NGHỆ THÔNG TIN

Sinh viên : Nguyễn Tiên Thành  
Giảng viên hướng dẫn : TS. Lê Văn Phùng

HẢI PHÒNG – 2023

**BỘ GIÁO DỤC VÀ ĐÀO TẠO  
TRƯỜNG ĐẠI HỌC QUẢN LÝ VÀ CÔNG NGHỆ HẢI PHÒNG**

-----

**PHẦN TỬ NGOẠI LAI ĐỐI VỚI KHÓA TRONG MÔ  
HÌNH CSDL QUAN HỆ VÀ ỨNG DỤNG TRONG QUẢN  
LÝ KẾT QUẢ TỐT NGHIỆP TẠI TRƯỜNG ĐẠI HỌC  
QUẢN LÝ VÀ CÔNG NGHỆ HẢI PHÒNG**

**ĐỒ ÁN TỐT NGHIỆP ĐẠI HỌC HỆ CHÍNH QUY  
NGÀNH: CÔNG NGHỆ THÔNG TIN**

**Sinh viên : Nguyễn Tiến Thành  
Giảng viên hướng dẫn : TS. Lê Văn Phùng**

**HẢI PHÒNG – 2023**

BỘ GIÁO DỤC VÀ ĐÀO TẠO  
TRƯỜNG ĐẠI HỌC QUẢN LÝ VÀ CÔNG NGHỆ HẢI PHÒNG

---

## NHIỆM VỤ ĐỀ TÀI TỐT NGHIỆP

**Sinh viên:** Nguyễn Tiến Thành

**Mã SV:** 1812111015

**Lớp** : CT2201M

**Ngành** : Công nghệ thông tin

**Tên đề tài:** Phân tử ngoại lai đối với khóa trong mô hình CSDL quan hệ và ứng dụng trong quản lý kết quả tốt nghiệp tại trường Đại học Quản lý và Công nghệ Hải Phòng

## **NHIỆM VỤ ĐỀ TÀI**

### **1. Nội dung và các yêu cầu cần giải quyết trong nhiệm vụ đề tài tốt nghiệp.**

- Tìm hiểu về phụ thuộc hàm và phân tử ngoại lai trong CSDL quan hệ.
- Tìm hiểu phương pháp phát hiện phân tử ngoại lai đối với Khóa trong mô hình CSDL quan hệ.
- Ứng dụng tìm phân tử ngoại lai để quản lý kết quả tốt nghiệp tại trường Đại học Quản Lý và Công nghệ Hải Phòng.

### **2. Các tài liệu cần thiết.**

- Lê Thành Hà (2016), Nghiên cứu phụ thuộc hàm và khóa trong CSDL quan hệ, Luận văn Thạc sỹ, Đại học Sư phạm 2 Hà Nội.
- Lê Văn Phùng, Quách Xuân Trường (2017), Khai phá dữ liệu, Tái bản lần 1, Nhà xuất bản Thông tin và Truyền Thông.
- Lê Văn Phùng (2018), Cơ sở dữ liệu quan hệ và công nghệ phân tích – thiết kế tái bản lần 1, Nhà xuất bản Thông tin và Truyền thông.

### **3. Địa điểm thực tập tốt nghiệp**

Trường ĐH QL&CN Hải Phòng

## MỤC LỤC

<b>LỜI CẢM ƠN.....</b>	<b>Error! Bookmark not defined.</b>
<b>MỞ ĐẦU.....</b>	<b>Error! Bookmark not defined.</b>
<b>CHƯƠNG 1:TỔNG QUAN VỀ MÔ HÌNH CSDL QUAN HỆ</b>	<b>Error! Bookmark not defined.</b>
1.1. Khái quát về mô hình CSDL quan hệ .....	7
1.2. Vai trò của Khóa trong mô hình quan hệ .....	7
1.2.1. Khái niệm Khóa.....	7
1.2.2. Một số thuật toán liên quan đến khóa.....	8
1.2.3. Hệ bằng nhau .....	15
<b>CHƯƠNG 2: PHẦN TỬ NGOẠI LAI ĐỐI VỚI KHÓA TRONG MÔ HÌNH CSDL QUAN HỆ.....</b>	<b>17</b>
2.1. Khái niệm phần tử ngoại lai và vai trò của nó trong mô hình CSDL quan hệ..	17
2.2. Khái niệm phần tử ngoại lai đối với phụ thuộc hàm và khóa.....	20
2.3. Mô hình phát hiện các phần tử ngoại lai .....	21
2.3.1. Định nghĩa.....	21
2.3.2. Phân loại các phần tử ngoại lai trong CSDL quan hệ.....	22
2.4. Định lý nhận biết phần tử ngoại lai theo khóa .....	23
2.5. Thuật toán xác định phần tử ngoại lai theo khóa .....	23
<b>CHƯƠNG 3: ỨNG DỤNG PHẦN TỬ NGOẠI LAI ĐỂ QUẢN LÝ KẾT QUẢ TỐT NGHIỆP TRƯỜNG ĐẠI HỌC QUẢN LÝ VÀ CÔNG NGHỆ</b>	
<b>HẢI PHÒNG .....</b>	<b>25</b>
3.1. Bài toán đặt ra và mục tiêu chương trình .....	25
3.2. Mục tiêu chương trình .....	Error! Bookmark not defined.
3.3. Yêu cầu và các chức năng chính của chương trình thử nghiệm ứng dụng phần tử ngoại lai đối với Khóa trong quản lý kết quả tốt nghiệp tại Trường Đại học QL&CN Hải Phòng	25
3.3.1. Yêu cầu của chương trình thử nghiệm .....	25

3.3.2. Quy trình xử lý .....	26
3.4. Các giao diện chính của chương trình.....	<b>34</b>
3.4.1. Giao diện trang giới thiệu .....	34
3.4.2. Giao diện nhập file dữ liệu .....	34
3.4.3. Giao diện kết quả khi có phần tử ngoại lai.....	35
3.4.4. Giao diện kết quả khi không có phần tử ngoại lai .....	35
<b>KẾT LUẬN</b> .....	<b>Error! Bookmark not defined.</b>
<b>TÀI LIỆU THAM KHẢO</b> .....	<b>Error! Bookmark not defined.</b>

# CHƯƠNG 1:TỔNG QUAN VỀ MÔ HÌNH CSDL QUAN HỆ

## 1.1. Khái quát về mô hình CSDL quan hệ

Mô hình CSDL quan hệ là một loại cơ sở dữ liệu lưu trữ và cung cấp quyền truy cập vào các điểm dữ liệu có liên quan đến nhau. Cơ sở dữ liệu quan hệ dựa trên mô hình quan hệ, một cách trực quan, đơn giản để biểu diễn dữ liệu trong bảng. Trong cơ sở dữ liệu quan hệ, mỗi hàng trong bảng là một bản ghi với một ID duy nhất được gọi là khóa. Các cột của bảng chứa các thuộc tính của dữ liệu và mỗi bản ghi thường có mỗi thuộc tính, giúp dễ dàng thiết lập mối quan hệ giữa các điểm dữ liệu.

## 1.2. Vai trò của Khóa trong mô hình quan hệ

### 1.2.1. Khái niệm Khóa

Khóa chính là hình ảnh của cột mã số hay số thứ tự (vì số thứ tự không thể trùng nhau được).

Khóa đóng một vai trò rất quan trọng vì nhờ có nó người ta mới tìm kiếm được (tìm kiếm bản ghi). Phép toán tìm kiếm bản ghi trong file dữ liệu là ghép toán quan trọng nhất vì chỉ sau khi tìm kiếm xong thì người ta mới tiến hành loại bỏ bản ghi ấy hoặc bổ sung một bản ghi mới vào trước hoặc sau bản ghi mà ta được tìm được.

**VD: ta có bảng quan hệ**

**BANHANG**

<b>MAHANG</b>	<b>TENHANG</b>	<b>SOLUONG (chiếc)</b>
VT0001	Vô tuyến	1000
TL0002	Tủ lạnh	500
RA2012	Radio	2000

Trong bảng trên mã số mặt hàng (MAHANG) là khóa. Mỗi giá trị mã hàng đều xác định duy nhất một loại mặt hàng trong quan hệ BAN\_HANG.

### 1.2.2. Một số thuật toán liên quan đến khóa

**Thuật toán 1: Tính bao đóng của một tập các thuộc tính đối với tập các phụ thuộc hàm trên sơ đồ quan hệ**

**Input :**  $s = \langle R, F \rangle$  là một sơ đồ quan hệ trong đó

$R = (a_1, a_2, \dots, a_n)$  là tập hữu hạn các thuộc tính.

$F$  là tập các phụ thuộc hàm và  $A \subseteq R$

**Output:**  $A^+$  là bao đóng của  $A$  đối với  $F$ .

Nhớ rằng  $A^+ = \{a: A \rightarrow \{a\} \in F^+\}$ .

$A \rightarrow B \in F^+$  nếu và chỉ nếu  $B \subseteq A^+$ .

**Phương pháp:** Lần lượt tính các tập thuộc tính  $A_0, A_1$  như sau:

1).  $A_0 = A$

2).  $A_i = A_{i-1} \cup \{a\}$  nếu  $\exists (C \rightarrow D) \in F, \{a\} \in D$  và  $C \subseteq A_{i-1}$

3). Rõ ràng  $A = A_0 \subseteq A_1 \subseteq \dots \subseteq A_i \subseteq R$  và  $R$  hữu hạn nên tồn tại  $i$  sao cho  $A_i = A_{i+1}$

Khi ấy thuật toán dừng và  $A_i$  chính là  $A^+$

Ví dụ:

Xét sơ đồ quan hệ  $s = \langle R, F \rangle$  trong đó

$$R = \{c, t, h, r, s, g\}$$

$$F = \begin{cases} \{c\} \rightarrow \{t\} \\ \{h, r\} \rightarrow \{c\} \\ \{h, t\} \rightarrow \{r\} \\ \{c, s\} \rightarrow \{g\} \\ \{h, s\} \rightarrow \{r\} \end{cases}$$

Tính  $\{h, r\}^+$  ?

$$A_0 = \{h, r\}$$

$$A_1 = \{h, r, c\} \quad \text{do } \{h, r\} \rightarrow \{c\} \in F$$

$$A_2 = \{h, r, c, t\} \quad \text{do } \{c\} \rightarrow \{t\} \in F$$

$$A_3 = \{h, r, c, t\} = A_2$$

$$\text{Vậy } \{h, r\}^+ = \{h, r, c, t\}$$



## Thuật toán 2: Tính bao đóng cho một tập bất kỳ trên quan hệ r

**Input:**  $r = \{h_1, h_2, \dots, h_m\}$  là một quan hệ trên  $R = \{a_1, a_2, \dots, a_n\}$ ,  $A \subseteq R$

**Output:**  $A_r^+$

Bước 1: Từ r xây dựng một tập  $E_r = \{E_{ij} : 1 \leq i < j \leq m\}$

$$E_{ij} = \{a : a \in R \text{ và } h_i(a) = h_j(a)\}$$

Bước 2: Từ  $E_r$  xây dựng một tập

$$M = \{B \in P(R) : \text{Tồn tại } E_{ij} \in E_r : E_{ij} = B\}$$

ở đây  $P(R)$  là tập các tập con của R

Bước 3:  $A_r^+$  được tính như sau:

$$A_r^+ = \begin{cases} \bigcap B \text{ nếu tồn tại } B \in M : A \subseteq B, \text{ (giao của tất cả các tập} \\ A \subseteq B \text{ trong } M \text{ chứa nó (A))} \\ R \text{ ngược lại.} \end{cases}$$

Có thể thấy rằng  $E_{ij} = \emptyset$  (trong trường hợp hai dòng i và j không có cột nào trùng nhau về giá trị, có nghĩa rằng chúng khác nhau hoàn toàn).

Không bao giờ có  $E_{ij} = R$  (có nghĩa rằng  $E_{ij} =$  toàn bộ không gian), vì nếu xảy ra thì có hai dòng trùng nhau, theo định nghĩa quan hệ thì không cho phép có hai dòng trùng nhau.

### Ví dụ: Xét quan hệ r

A	B	C	D	E
1	1	0	0	0
0	1	0	1	1
0	0	2	0	0
2	0	2	0	1

Tính  $\{B, C\}_r^+$

$E_{12} = \{B, C\}, E_{13} = \{D, E\}, E_{14} = \{D\}$

$E_{23} = \{A\}, E_{24} = \{E\}$

$E_{34} = \{B, C, D\}$

$M = \{\{A\}, \{B, C\}, \{B, C, D\}, \{D\}, \{D, E\}, \{E\}\}$

Vậy  $\{B, C\}_r^+ = \{B, C\} \cap \{B, C, D\} = \{B, C\}$

### **Thuật toán 3: Tìm khóa tối thiểu của một quan hệ**

Khi giải quyết các bài toán thông tin quản lý, người ta thường sử dụng các hệ quản trị cơ sở dữ liệu mà trong đó chứa cơ sở dữ liệu quan hệ. Các phép xử lý đối với bài toán này thường là tìm kiếm bản ghi sau đó thêm bản ghi mới, thay đổi nội dung cơ bản ghi hoặc xóa bản ghi. Trong các thao tác trên, việc tìm kiếm bản ghi là rất quan trọng. Muốn tìm được bản ghi trong file dữ liệu thì chúng ta phải xây dựng khóa của file dữ liệu đó.

Có 2 thuật toán tìm khóa của quan hệ và lược đồ quan hệ. Tìm khóa ở đây chính là tìm khóa tối thiểu.

**Vào:**  $r = \{h_1, \dots, h_m\}$  là một quan hệ trên tập thuộc tính  $R = \{a_1, \dots, a_n\}$

**Ra:**  $K$  là một khóa tối thiểu của  $r$ .

#### **Phương pháp:**

Bước 1: Tính  $E_r = \{A_1, A_2, \dots\}$  trong đó  $E_r$  là các hệ bằng nhau.

Bước 2: Tính  $M_r$  là các hệ bằng nhau cực đại.

Bước 3: Lần lượt tính các thuộc tính  $K_0, K_1, \dots, K_n$  theo qui tắc:

$K_0 = R = \{a_1, \dots, a_n\}$  hoặc  $K_0$  là một khóa đã biết.

$$K_i = \begin{cases} K_{i-1} - \{a_i\} & \text{nếu: không tồn tại } A \in M_r : K_{i-1} - \{a_i\} \subseteq A \\ & \text{(bao } K_i) \\ K_{i-1} & \text{trong trường hợp ngược lại} \end{cases}$$

Bước 4: Đặt  $K = K_n$ . Khi đó  $K$  là khoá tối tiểu.

**Nhận xét:** Nếu ta thay đổi thứ tự các thuộc tính của  $r$  bằng thuật toán này chúng ta có thể tìm được một khoá tối tiểu khác.

**Ví dụ: Cho quan hệ  $r$**

A	B	C	D	E
1	1	0	1	0
1	0	0	1	1
2	1	1	0	2
1	0	1	1	0

Bước 1:  $E_{12} = \{A, C, D\}$        $E_{13} = \{B\}$        $E_{14} = \{A, D, E\}$

$E_{23} = \emptyset$        $E_{24} = \{A, B, D\}$        $E_{34} = \{C\}$

Bước 2:  $M_r = \{ \{A, C, D\}, \{A, D, E\}, \{A, B, D\} \}$

Bước 3: (KH:  $P(M_r)$ )-phần tử của  $M_r$

$K_0 = \{A, B, C, D, E\}$

Xét  $K_1 = K_0 - \{A\} = \{B, C, D, E\} \not\subset P(M_r) \Rightarrow K_1 = \{B, C, D, E\}$ .

Xét  $K_2 = K_1 - \{B\} = \{C, D, E\} \not\subset P(M_r) \Rightarrow K_2 = \{C, D, E\}$ .

Xét  $K_3 = K_2 - \{C\} = \{D, E\} \subset P(M_r) \Rightarrow K_3 = \{C, D, E\}$ .

Xét  $K_4 = K_3 - \{D\} = \{C, E\} \not\subset P(M_r) \Rightarrow K_4 = \{C, E\}$ .

Xét  $K_5 = K_4 - \{E\} = \{C\} \subset P(M_r) \Rightarrow K_5 = \{C, E\}$ .

Vậy  $\{C, E\}$  là một khoá tối tiểu của  $r$ .

Cũng ví dụ trên nhưng ta thay đổi tập thuộc tính theo thứ tự là  $\{A, C, E, B, D\}$ . Ta đi tìm khoá tối tiểu của  $r$

Bước 1 và bước 2 giống như trên

Bước 3:

$$K_0 = \{A, C, E, B, D\}$$

$$\text{Xét } K_1 = K_0 - \{A\} = \{C, E, B, D\} \not\subset P(M_r) \Rightarrow K_1 = \{C, E, B, D\}.$$

$$\text{Xét } K_2 = K_1 - \{C\} = \{E, B, D\} \not\subset P(M_r) \Rightarrow K_2 = \{E, B, D\}.$$

$$\text{Xét } K_3 = K_2 - \{E\} = \{B, D\} \subset P(M_r) \Rightarrow K_3 = K_2 = \{E, B, D\}.$$

$$\text{Xét } K_4 = K_3 - \{B\} = \{E, D\} \subset P(M_r) \Rightarrow K_4 = K_3 = \{E, B, D\}.$$

$$\text{Xét } K_5 = K_4 - \{D\} = \{E, B\} \not\subset P(M_r) \Rightarrow K_5 = \{E, B\}.$$

Vậy khoá tối tiểu của  $r$  là  $\{E, B\}$ .

Như vậy thay đổi thứ tự của các thuộc tính, ta sẽ có những tập khoá tối tiểu khác nhau .

**Ví dụ : Cho quan hệ  $r$**

A	B	C	D	E
2	2	1	4	1
1	2	1	2	2
1	1	2	2	1
3	1	3	1	2

$$\text{Bước 1: } E_{12} = \{B, C\}$$

$$E_{13} = \{E\}$$

$$E_{14} = \emptyset$$

$$E_{23} = \{A, D\}$$

$$E_{24} = \{E\}$$

$$E_{34} = \{B\}$$

$$\text{Bước 2: } M_r = \{\{B, C\}, \{E\}, \{A, D\}\}$$

Bước 3:

\*Các thuộc tính được lấy theo thứ tự  $r_1 = \{A, B, C, D, E\}$

$$K_0 = R = \{A, B, C, D, E\}$$

$$\text{Xét } K_1 = K_0 - \{A\} = \{B, C, D, E\} \not\subset P(M_r) \Rightarrow K_1 = \{B, C, D, E\}.$$

$$\text{Xét } K_2 = K_1 - \{B\} = \{C, D, E\} \not\subset P(M_r) \Rightarrow K_2 = \{C, D, E\}.$$

$$\text{Xét } K_3 = K_2 - \{C\} = \{D, E\} \not\subset P(M_r) \Rightarrow K_3 = \{D, E\}.$$

$$\text{Xét } K_4 = K_3 - \{D\} = \{E\} \subset P(M_r) \Rightarrow K_4 = K_3 = \{D, E\}.$$

$$\text{Xét } K_5 = K_4 - \{E\} = \{D\} \subset P(M_r) \Rightarrow K_5 = K_4 = \{D, E\}.$$

Vậy khoá tối thiểu của  $r$  là  $\{D, E\}$ .

\*Các thuộc tính được lấy theo thứ tự  $r_2 = \{E, D, C, B, A\}$

$$K_0 = R = \{E, D, C, B, A\}$$

$$\text{Xét } K_1 = K_0 - \{E\} = \{D, C, B, A\} \not\subset P(M_r) \Rightarrow K_1 = \{D, C, B, A\}.$$

$$\text{Xét } K_2 = K_1 - \{D\} = \{C, B, A\} \not\subset P(M_r) \Rightarrow K_2 = \{C, B, A\}.$$

$$\text{Xét } K_3 = K_2 - \{C\} = \{B, A\} \not\subset P(M_r) \Rightarrow K_3 = \{B, A\}.$$

$$\text{Xét } K_4 = K_3 - \{B\} = \{A\} \subset P(M_r) \Rightarrow K_4 = K_3 = \{B, A\}.$$

$$\text{Xét } K_5 = K_4 - \{A\} = \{B\} \subset P(M_r) \Rightarrow K_5 = K_4 = \{B, A\}.$$

Vậy khoá tối thiểu khác của quan hệ là  $\{B, A\}$ .

#### **Thuật toán 4: Tìm khóa tối thiểu cho một sơ đồ quan hệ**

**Vào:** sơ đồ quan hệ  $s = \langle R, F \rangle$  trong đó

$F$  là tập các phụ thuộc hàm

$R = \{a_1, \dots, a_n\}$  là tập các thuộc tính

**Ra:**  $K$  là tối thiểu của  $s$

**Phương pháp:** Tính liên tiếp các tập thuộc tính  $K_0, K_1, \dots, K_n$  như sau:

$$K_0 = R = \{a_1, \dots, a_n\}$$

$$K_i = \begin{cases} K_{i-1} & \text{nếu } K_{i-1} - \{a_i\} \rightarrow R \notin F^+ \\ K_{i-1} - \{a_i\} & \text{nếu ngược lại.} \end{cases}$$

$K = K_n$  là khoá tối tiểu.

Ta có thể dùng công thức tương đương:

$$K_i = \begin{cases} K_{i-1} - \{a_i\} & \text{nếu } \{K_{i-1} - a_i\}^+ = R \\ K_{i-1} & \text{nếu ngược lại} \end{cases}$$

### Nhận xét:

- Thay đổi thứ tự các thuộc tính của R bằng thuật toán trên chúng ta có thể tìm được một khoá tối tiểu khác.
- Nếu như đã biết A là một khoá nào đó thì có thể đặt  $K_0 = A$ , ta vẫn tìm ra được khoá tối tiểu và thời gian tìm nhanh hơn.

Ví dụ: Giả sử  $s = \langle F, R \rangle$  là một lược đồ quan hệ trong đó:

$$R = \{a, b, c, d\}$$

$$F = \{\{a, b\} \rightarrow \{d\}, \{c\} \rightarrow \{b\}\}$$

Tìm khoá tối tiểu của sơ đồ quan hệ

Áp dụng thuật toán trên ta có:

$$+ K_0 = R = \{a, b, c, d\}$$

+ Tính  $K_1$

$$\text{Xét } K_1 = K_0 - \{a\} = \{b, c, d\}$$

$$\{b, c, d\}^+ = \{b, c, d\} \neq R$$

$$\text{Vậy } K_1 = \{a, b, c, d\}. \quad (K_1 = K_0)$$

+ Tính  $K_2$

Xét  $K_2 = K_1 - \{b\} = \{a, c, d\}$

$$\{a, c, d\}^+ = \{a, b, c, d\} = R$$

Vậy  $K_2 = \{a, c, d\}$

+Tính  $K_3$

Xét  $K_3 = K_2 - \{c\} = \{a, d\}$

$$\{a, d\}^+ = \{a, d\} \neq R$$

Vậy  $K_3 = \{a, c, d\}$  ( $K_3 = K_2$ )

+Tính  $K_4$

Xét  $K_4 = K_3 - \{d\} = \{a, c\}$

$$\{a, c\}^+ = \{a, b, c, d\} = R$$

Vậy  $K_4 = \{a, c\}$

Vậy khoá tối thiểu là  $\{a, c\}$ .

### 1.2.3. Hệ bằng nhau

#### Định nghĩa

Giả sử  $r = \{h_1, h_2, \dots, h_m\}$  là một quan hệ trên  $R = \{a_1, a_2, \dots, a_n\}$

Đặt  $E_r = \{E_{ij} : 1 \leq i < j \leq |r|\}$  trong đó  $E_{ij} = \{a \in R : h_i(a) = h_j(a)\}$ ,  $|r| = m$

$E_r$  được gọi là hệ bằng nhau của  $r$

Giả sử  $M_r = \{A \in P(R) : \exists E_{ij} = A, \nexists E_{pq} : A \subset E_{pq}\}$

Khi đó  $M_r$  được gọi là hệ bằng nhau cực đại của  $r$ .

#### Nhận xét:

Hệ bằng nhau và hệ bằng nhau cực đại đóng một vai trò quan trọng trong các thuật toán thiết kế cũng như mối quan hệ giữa các lớp quan hệ và lớp các phụ thuộc hàm trong quá trình nghiên cứu cấu trúc logic của lớp các phụ thuộc hàm.

**Ví dụ:** Cho quan hệ r.

	a <sub>1</sub>	a <sub>2</sub>	a <sub>3</sub>	a <sub>4</sub>	a <sub>5</sub>
	1	1	0	1	0
	1	0	0	3	0
	3	1	1	3	1
	5	1	0	3	0

$$E_r = \left\{ \begin{array}{l} E_{12} = \{a_1, a_3, a_5\} \\ E_{13} = \{a_2\} \\ E_{14} = \{a_2, a_3, a_5\} \\ E_{23} = \{a_4\} \\ E_{24} = \{a_3, a_4, a_5\} \\ E_{34} = \{a_2, a_4\} \end{array} \right.$$

$$M_r = \left\{ \begin{array}{l} \{a_1, a_3, a_5\} \\ \{a_2, a_3, a_5\} \\ \{a_3, a_4, a_5\} \\ \{a_2, a_4\} \end{array} \right.$$

Vậy  $M_r$  là hệ bằng nhau cực đại của r.

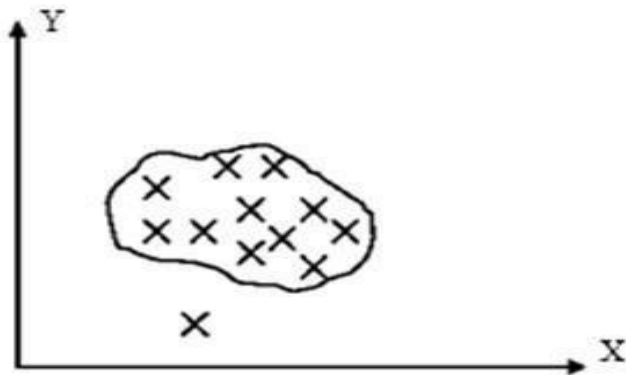


## CHƯƠNG 2: PHẦN TỬ NGOẠI LAI ĐỐI VỚI KHÓA TRONG MÔ HÌNH CSDL QUAN HỆ

2.1. Khái niệm phần tử ngoại lai và vai trò của nó trong mô hình CSDL quan hệ

a) Khái niệm phần tử ngoại lai.

Một cách hình thức người ta có thể định nghĩa phần tử ngoại lai (outliers) của một tập dữ liệu là các phần tử mà theo một cách nhìn nào đó có các đặc tính không giống với tập hợp đa số còn lại của tập dữ liệu. Chẳng hạn trong hình dưới đây cho thấy một phần tử ngoại lai theo vị trí hình học:



*Hình 1 Phần tử ngoại lai trong tập điểm có tọa độ  $(x,y)$  trên mặt phẳng có giá trị tung độ  $y$  nhỏ hơn hẳn các phần tử khác của tập hợp*

Các khái niệm về ngoại lai đầu tiên có nguồn gốc từ lĩnh vực thống kê. Barnett và Lewis định nghĩa: một phần tử ngoại lai là một quan trắc hoặc một tập con các quan trắc mà sự xuất hiện của chúng trái ngược với những quan trắc còn lại. Phần tử ngoại lai cũng có thể được hiểu như một quan trắc mà giá trị của nó khác biệt quá nhiều so với những quan trắc khác gây cho người người ta nghi ngờ rằng nó đã được thực hiện bằng một kỹ thuật khác.

Có nhiều cách định nghĩa và hiểu khác nhau về phần tử ngoại lai. Tuy nhiên chúng có điểm chung là: phần tử ngoại lai của một file dữ liệu là những phần tử của file dữ liệu có sự khác biệt đáng kể đối với những phần tử còn lại. Và khi tiến hành xác định

phần tử ngoại lai, trước hết người ta đưa ra định nghĩa, sau đó sẽ xây dựng phương pháp để xác định.

Cho một sơ đồ quan hệ  $(R,F)$ , với tập thuộc tính  $R = \{A_1, A_2, \dots, A_n\}$  và tập các phụ thuộc hàm  $F$  đúng trên  $R$ . Gọi  $F^+$  là bao đóng của  $F$  (theo Hệ tiên đề Armstrong). Giả sử  $r$  là một bảng dữ liệu có các trường (thuộc tính) và miền giá trị trùng với quan hệ trên sơ đồ quan hệ  $(R,F)$ . Ta gọi  $r$  là bảng dữ liệu trên  $R$ . Bảng dữ liệu này có thể chứa những bộ trùng nhau. Kí hiệu  $T$  là tập các ràng buộc và qui tắc mà các phần tử của  $R$  thỏa mãn.

#### b) Vai trò của nó trong mô hình CSDL quan hệ

Cho một bảng dữ liệu  $r$  trên một tập thuộc tính  $R$ . Kí hiệu  $T$  là tập các qui tắc, ràng buộc (gọi là các luật) mà các phần tử của  $r$  phải tuân theo. Phần tử ngoại lai của  $r$  là những phần tử của bảng dữ liệu này không tuân theo một trong các qui tắc, ràng buộc đó.

Một phần tử của bảng dữ liệu được hiểu là một bộ các giá trị của các thuộc tính.

Các qui tắc, ràng buộc được đề cập bao gồm những ràng buộc về cấu trúc của CSDL (khóa, phụ thuộc hàm, các dạng chuẩn phải tuân theo đối với một quan hệ và các ràng buộc theo ngữ nghĩa phụ thuộc vào yêu cầu, ý nghĩa của ứng dụng mà trong đó CSDL được sử dụng).

Phần tử ngoại lai giữ một vai trò đặc biệt quan trọng trong mô hình CSDL quan hệ, đặc biệt là đối với phụ thuộc hàm của bảng dữ liệu  $r$ .

Cho  $r$  là một bảng dữ liệu trên sơ đồ quan hệ  $(R,F)$ . Giải thiết  $r$  là một quan hệ. Ta gọi một cặp bộ  $t_1, t_2 \in r$  không thỏa mãn điều kiện phụ thuộc hàm của  $F$  là cặp phần tử ngoại lai đối với phụ thuộc hàm của bảng dữ liệu  $r$ .

Người ta biểu diễn một cách hình thức như sau:

Giả sử  $X \rightarrow Y$  là một phụ thuộc hàm thuộc  $F$ . Khi đó cặp  $t_1, t_2 \in r$  là cặp phần tử ngoại lai đối với phụ thuộc hàm  $X \rightarrow Y$  nếu:

$$t_1(X) = t_2(X) \text{ và } t_1(Y) \neq t_2(Y).$$

Khái niệm phần tử ngoại lai đi cùng với mô hình CSDL quan hệ ngày nay đã ngày càng đi sâu vào mọi mặt trong đời sống kinh tế - xã hội. Chúng dùng để:

- Phát hiện xâm nhập (phát hiện các hoạt động nguy hiểm (phá vỡ thâm nhập và các hình thức khác của máy tính lạm dụng) trong một hệ thống máy tính liên quan từ một vấn đề bảo mật. Khác với hệ thống hành vi bình thường, phát hiện xâm nhập là một ứng cử viên hoàn hảo cho việc áp dụng các kỹ thuật phát hiện ngoại lai).

- Phát hiện gian lận (liên quan đến hoạt động tội phạm xảy ra trong các tổ chức thương mại, các tổ chức như ngân hàng, các công ty thẻ tín dụng, cơ quan bảo hiểm, các công ty điện thoại di động, thị trường chứng khoán,... Người sử dụng độc hại có thể là khách hàng thực tế của tổ chức hoặc phải dùng đến hành vi trộm cắp danh tính (giả làm khách hàng). Các hoạt động phát hiện nhằm mục đích phát hiện tiêu thụ trái phép các nguồn tài nguyên được cung cấp bởi tổ chức để ngăn chặn thiệt hại kinh tế).

- Phát hiện bảo hiểm yêu cầu bồi thường gian lận (ví dụ xe hơi gian lận bảo hiểm. Các cá nhân và tổ chức bên yêu sách và các nhà cung cấp thao tác yêu cầu bồi thường hệ thống xử lý cho các tuyên bố trái phép và bất hợp pháp. Các dữ liệu trong lĩnh vực này để phát hiện gian lận đến từ các văn bản trình của các bên tranh chấp).

- Phát hiện gian lận trong y tế công cộng (Dữ liệu có thể có giá trị ngoại lai do một số lý do như tình trạng bệnh nhân bất thường hoặc thiết bị đo đạc lỗi hoặc lỗi ghi âm. Hầu hết các ngoại lai hiện tại kỹ thuật phát hiện trong này nhằm mục đích phát hiện tại miền hồ sơ bất thường (ngoại lai điểm)).

- Phát hiện thiệt hại công nghiệp (đơn vị công nghiệp bị thiệt hại do liên tục sử dụng và hao mòn thông thường, thiệt hại như vậy cần phải được phát hiện sớm để ngăn chặn sự leo thang hơn nữa và gây tổn thất dẫn đến thiệt hại. Các dữ liệu trong phạm vi này thường là cảm biến dữ liệu được ghi bằng các cảm biến khác nhau và

thu thập cho phân tích. Ví dụ như Phát hiện lỗi trong đơn vị cơ khí và thành phần như động cơ, tua-bin, dầu chảy trong đường ống,... Các vết nứt trong dầm, thùng trong khung máy bay, dữ liệu không lường trước được sử dụng cho lỗi phát hiện ở các đơn vị cơ khí,...).

- Phát hiện gian lận trong Xử lý hình ảnh (Phát hiện ngoại lai ở đây nhằm phát hiện những thay đổi trong một hình ảnh theo thời gian (phát hiện chuyển động) hoặc trong các khu vực mà xuất hiện bất thường trên hình ảnh tĩnh. Tên miền này bao gồm các hình ảnh vệ tinh, công nhận chữ số, quang phổ, hình ảnh X quang vú, và giám sát video,... Các yếu tố được gây ra bởi chuyển động hoặc chèn đối tượng hoặc thiết bị lỗi. Các dữ liệu có không gian cũng như đặc điểm thời gian. Mỗi điểm dữ liệu có một vài các thuộc tính liên tục như màu sắc, kết cấu,... Các giá trị ngoại lai thú vị là những điểm hoặc bất thường hoặc khu vực trong ảnh (điểm và sự chênh lệch theo ngữ cảnh)).

- Phát hiện sai sót trong mạng cảm biến (ngoại lai trong dữ liệu thu thập hoặc có thể bao hàm một hoặc nhiều cảm biến bị lỗi (Các ứng dụng cảm biến phát hiện lỗi), hoặc các cảm biến sự kiện phát hiện (ứng dụng phát hiện xâm nhập).

## 2.2. Khái niệm phần tử ngoại lai đối với phụ thuộc hàm và khóa

a) khái niệm phần tử ngoại lai đối với khóa

Định nghĩa:

Cho bảng dữ liệu  $r$  được giả thiết là một quan hệ trên sơ đồ quan hệ  $(R, F)$ ;  $B$  được giả thiết là tập các khóa tối thiểu của  $r$ . Cặp phần tử  $(t_i, t_j)$  với  $t_i, t_j \in r$  ( $i \neq j$ ) là một cặp ngoại lai đối với khóa nếu như với một khóa nào đó  $K \in B$  mà ta có:

$$t_i(K) = t_j(K)$$

Trường hợp nếu  $t_i(K) = t_j(K)$  và  $t_i(R \setminus K) = t_j(R \setminus K)$  tức chúng trùng nhau hoàn toàn về giá trị trên các thuộc tính, có nghĩa là ta có  $t_i(R) = t_j(R)$  thì cặp  $(t_i, t_j)$  được gọi là cặp ngoại lai tầm thường (theo qui ước thì trong một quan hệ không được có 2 bộ trùng nhau hoàn toàn). Còn nếu  $t_i(K) = t_j(K)$  nhưng  $t_i(R \setminus K) \neq t_j(R \setminus K)$  thì ta có cặp ngoại lai theo phụ thuộc hàm  $K \rightarrow R$ .

Trong thực tế việc có những cặp ngoại lai tầm thường là có thể xảy ra (do cập nhật dữ liệu sai, hoặc có sự gian lận cố tình nhập nhiều lần cùng một bản ghi vào một bảng dữ liệu). Việc phát hiện những cặp bộ ngoại lai theo khóa nhằm mục đích giúp chúng ta xử lý các trường hợp này.

Định lý sau làm cơ sở cho việc phát hiện các cặp ngoại lai đối với khóa của bảng dữ liệu

b) Khái niệm phần tử ngoại lai đối với phụ thuộc hàm

Cho  $r$  là một bảng dữ liệu trên sơ đồ quan hệ  $(R, F)$ . Giả thiết  $r$  là một quan hệ. Người ta gọi một cặp bộ  $t_1, t_2 \in r$  không thỏa mãn điều kiện phụ thuộc hàm của  $F$  là cặp phần tử ngoại lai đối với phụ thuộc hàm của bảng dữ liệu  $r$ .

Có thể biểu diễn một cách hình thức như sau:

Giả sử  $X \rightarrow Y$  là một phụ thuộc hàm thuộc  $F$ . Khi đó cặp  $t_1, t_2 \in r$  là cặp phần tử ngoại lai đối với phụ thuộc hàm  $X \rightarrow Y$  nếu  $t_1(X) = t_2(X)$  và  $t_1(Y) \neq t_2(Y)$ .

Dưới đây trình bày thuật toán xác định các cặp ngoại lai đối với phụ thuộc hàm.

Giả sử  $r = \{t_1, t_2, \dots, t_m\}$  là bảng dữ liệu trên cơ sở quan hệ  $(R, F)$ . Tập  $E_r$  được xác định như sau:

$$E_r = \{E_{i,j}: 1 \leq i < j \leq m \text{ và } E_{i,j} = \{a \in R; t_i(a) = t_j(a)\}\}.$$

Gọi  $E_r$  là hệ bằng nhau của  $r$ .

## 2.3. Mô hình phát hiện các phần tử ngoại lai

### 2.3.1. Định nghĩa

Cho một sơ đồ quan hệ  $(R, F)$ , với tập thuộc tính  $R = \{A_1, A_2, \dots, A_n\}$  và tập các phụ thuộc hàm  $F$  đúng trên  $R$ . Gọi  $F^+$  là bao đóng của  $F$  (theo Hệ tiên đề Armstrong). Giả sử  $r$  là một bảng dữ liệu có các trường (thuộc tính) và miền giá trị trùng với quan hệ trên sơ đồ quan hệ  $(R, F)$ . Ta gọi  $r$  là bảng dữ liệu trên  $R$ . Bảng dữ liệu này có thể chứa những bộ trùng nhau. Kí hiệu  $T$  là tập các ràng buộc và qui tắc mà các phần tử của  $R$  thỏa mãn.

### *Định nghĩa:*

Cho một bảng dữ liệu  $r$  trên một tập thuộc tính  $R$ . Kí hiệu  $T$  là tập các qui tắc, ràng buộc (gọi là các luật) mà các phần tử của  $r$  phải tuân theo. Phần tử ngoại lai của  $r$  là những phần tử của bảng dữ liệu này không tuân theo một trong các qui tắc, ràng buộc đó.

Một phần tử của bảng dữ liệu được hiểu là một bộ các giá trị của các thuộc tính. Các qui tắc, ràng buộc được đề cập bao gồm những ràng buộc về cấu của CSDL (khóa, phụ thuộc hàm, các dạng chuẩn phải tuân theo đối với một quan hệ và các ràng buộc theo ngữ nghĩa phụ thuộc vào yêu cầu, ý nghĩa của ứng dụng mà trong đó CSDL được sử dụng).

#### 2.3.2. Phân loại các phần tử ngoại lai trong CSDL quan hệ

Tùy theo các loại ràng buộc đối với các phần tử trong một quan hệ của CSDL quan hệ ta cũng có những loại phần tử ngoại lai đối với từng trường hợp đó (phần tử vi phạm các ràng buộc tương ứng). Tùy theo ngữ cảnh và yêu cầu của bài toán thực tế mà các khái niệm, định nghĩa, phương pháp xác định phần tử ngoại lai sẽ được đưa ra.

Các phần tử ngoại lai trong CSDL quan hệ được đề cập tới trong nhiều tài liệu bao gồm [2]:

1. Phần tử ngoại lai đối với phụ thuộc hàm
2. Phần tử ngoại lai đối với khóa
3. Phần tử ngoại lai đối với phụ thuộc hàm dạng đặc biệt (dạng tỷ lệ, dạng bằng nhau)
4. Phần tử ngoại lai đối với dạng chuẩn
5. Phần tử ngoại lai đối với phụ thuộc hàm xấp xỉ

Trong phạm vi tìm hiểu của khoá luận, em chỉ đề cập tới phần tử ngoại lai đối với phụ thuộc hàm là dạng phần tử ngoại lai cơ bản đối với CSDL quan hệ và có ý nghĩa hỗ trợ trong công việc quản lý giảng dạy của

các trường. Các phần tử ngoại lai trong CSDL quan hệ được đề cập tới trong khoá luận chỉ chủ yếu tập trung vào Khoá.

#### 2.4. Định lý nhận biết phần tử ngoại lai theo khoá

Cho bảng dữ liệu  $r$  được giả thiết là quan hệ trên sơ đồ quan hệ  $(R, F)$ ;  $B$  là tập các khoá tối thiểu của  $r$ ;  $E_r$  là hệ bằng nhau của  $r$ . Khi đó nếu ta có  $E_{i,j} \in E_r$  chứa một khoá  $K$  nào đó (tức là  $K \subseteq E_{i,j}$ ) thì cặp phần tử  $(t_i, t_j)$  với  $t_i, t_j \in r$  (trung ứng với  $E_{i,j}$ ) là một cặp ngoại lai đối với khoá.

##### Chứng minh:

Theo định nghĩa khoá của một sơ đồ quan hệ, nếu  $K$  là một khoá thì ta có  $K \rightarrow R$ .

Giả sử  $K \subseteq E_{i,j}$ , ta có  $t_i(K) = t_j(K)$ . Do  $K \rightarrow R$  nên ta phải có:  $t_i(R) = t_j(R)$ .

Vì vậy:

- Nếu  $t_i(R) = t_j(R)$  ta có một cặp ngoại lai tầm thường (vi phạm điều kiện  $r$  là một quan hệ).

- Nếu  $t_i(K) = t_j(K)$  và  $t_i(R \setminus K) \neq t_j(R \setminus K)$  ta sẽ có một cặp ngoại lai đối với phụ thuộc hàm  $K \rightarrow R$  (theo định nghĩa của khoá). Định lý được chứng minh.

#### 2.5. Thuật toán xác định phần tử ngoại lai theo khoá

Input: Cho sơ đồ quan hệ  $(R, F)$ ; tập các khoá  $B$ ;

Bảng dữ liệu  $r = \{t_1, t_2, \dots, t_m\}$  trên  $R$ .

Output: Tập các cặp ngoại lai theo khoá OUTLI2

Bước 1: Tính hệ bằng nhau  $E_r = \{E_{i,j}: 1 \leq i < j \leq m \text{ và } E_{i,j} = \{a \in R; t_i(a) = t_j(a)\}\}$ .

Bước 2: Với mỗi khoá  $K \in B$  và với mỗi tập  $E_{i,j} \in E_r$  ta kiểm tra điều kiện  $K \subseteq E_{i,j}$ . Nếu đúng lưu cặp  $(t_i, t_j)$  vào OUTLI2.

Tập OUTLI2 là tập các phần tử ngoại lai theo khoá.

Ví dụ 2.4: Cho sơ đồ quan hệ (R,F) với  $R = \{A, B, C, D, E\}$  với các khóa là BE, AE. Xét bảng dữ liệu trên sơ đồ quan hệ (R,F) cho ở bảng sau (được giả thiết là một quan hệ)

A	B	C	D	E
0	1	2	1	3
1	2	2	1	3
0	1	2	1	4
3	1	2	1	1
3	2	1	3	1

Tính  $E_r$ :  $E_{1,2} = CDE$ ;  $E_{1,3} = ABCD$ ;  $E_{1,4} = BCD$ ;  $E_{1,5} = \emptyset$ ;  $E_{2,3} = CD$ ;  $E_{2,4} = CD$ ;  $E_{2,5} = B$ ;  $E_{3,4} = BCD$ ;  $E_{3,5} = \emptyset$ ;  $E_{4,5} = AE$ .

Ta có:  $M_r = \{CDE, ABCD, AE\}$

Ta thấy có cặp  $(t_4, t_5)$  là cặp ngoại lai đối với khóa AE (vì  $E_{4,5}=AE$  tức là nó chứa AE và như vậy AE không thể là khóa được). Kiểm tra lại và nếu sửa lại giá trị  $t_4(A)$  hoặc  $t_5(A) = 4$  thì AE sẽ bảo đảm là khóa của bảng dữ liệu.



## **CHƯƠNG 3: ỨNG DỤNG PHẦN TỬ NGOẠI LAI ĐỂ QUẢN LÝ KẾT QUẢ TỐT NGHIỆP TRƯỜNG ĐẠI HỌC QUẢN LÝ VÀ CÔNG NGHỆ HẢI PHÒNG**

### **3.1. Bài toán đặt ra và mục tiêu chương trình**

Hiện tại trong quá trình quản lý kết quả tốt nghiệp tại Trường Đại học Quản lý và Công nghệ Hải Phòng. Thời gian từ lúc sinh viên hoàn thành tất cả các môn học đến thời điểm xét công nhận tốt nghiệp sẽ mất một khoảng thời gian khoảng 2 tuần để sinh viên, giáo viên chủ nhiệm và cố vấn học tập kiểm tra và rà soát lại tất cả điểm trong thời gian học. Trong khoảng thời gian này do lý do khách quan có thể sẽ có sự sai khác trong kết quả tốt nghiệp của các sinh viên. Do đó để kiểm tra lại các kết quả sai khác này sẽ mất rất nhiều thời gian. Chính vì vậy, cần phải có một mô-đun riêng để kiểm tra việc đánh giá kết quả tốt nghiệp của sinh viên có đúng hay không, bảo đảm khi công bố kết quả cho sinh viên là đúng và tránh sự khiếu nại rắc rối thay vì hiện nay rất bất cập do khối lượng dữ liệu lớn.

### **3.2. Yêu cầu và các chức năng chính của chương trình thử nghiệm ứng dụng phần tử ngoại lai đối với Khóa trong quản lý kết quả tốt nghiệp tại Trường Đại học QL&CN Hải Phòng**

#### **3.2.1. Yêu cầu của chương trình thử nghiệm**

Yêu cầu của chương trình thử nghiệm cần bảo đảm yêu cầu quản lý kết quả tốt nghiệp đã nêu ở trên. Trong chương trình thử nghiệm cần thể hiện rõ ứng dụng nội dung nghiên cứu về phần tử ngoại lai đối với khóa ở chương 2.

+Input: Các bảng dữ liệu danh mục gồm danh sách học phần, danh sách lớp học, danh sách sinh viên (do nhà trường cung cấp) và bảng điểm các môn học phần theo từng ngành của từng lớp, bảng BAOCÁO kết quả đánh giá tốt nghiệp (do giáo viên chủ nhiệm các lớp báo cáo).

+Nội dung xử lý: thực hiện tự động công thức tính điểm trong QUY CHẾ TÍNH ĐIỂM ở mục C. Ghi điểm trung bình chung cho cột TBC\_TL vào bảng mới tính

được: KETQUA (bảng kiểm tra). So sánh với bảng BAOCÁO kết quả học tập (bảng input) để phát hiện các bản ghi sai lệch (phần tử ngoại lai).

+Output: Xác định các bản ghi cụ thể (sinh viên cụ thể) bị sai sót và điều chỉnh.

### 3.3.2. Quy trình xử lý

Minh họa cho quy trình xử lý này được thực hiện trong ngành Đại học CNTT của Trường.

**Bước 1-** Lấy 1 tệp 5 sheet input từ EXCEL

#### **BẢNG 01 - DANH SÁCH HỌC PHẦN**

Mã ngành: .....

Tên ngành: CNTT

STT	Học Phần	Số TC
1	Giao_duc_the_chat_1	1
2	Toán_hoc_1	3
3	Logic_hoc	3
4	Nhap_mon_tin_hoc	3
5	Quan_tri_mang	3
6	Toan_giai_tich_1	3
7	Vat_ly	3

**BẢNG 2-LỚP HỌC**

<b>Mã lớp</b>	<b>Mã ngành</b>	<b>Năm nhập học</b>	<b>SSi số</b>	<b>Tên GVCN</b>

**BẢNG 3-DANH SÁCH SINH VIÊN**

<b>Mã Sinh Viên</b>	<b>Họ đệm</b>	<b>Tên</b>	<b>Ngày sinh</b>	<b>Nơi sinh</b>	<b>Gioi tinh</b>	<b>Lop hoc</b>
12101200188	Trần Thị	Hằng	3/21/1994	Tỉnh Nam Định	Nữ	CT2201
12101200101	Lê Thị	Chinh	8/10/1994	Tỉnh Phú Thọ	Nữ	CT2201
12101200128	Đỗ Thị	Hương	7/1/1994	Tỉnh Thanh Hóa	Nữ	CT2201
12101200050	Nguyễn Thị	Thúy	7/28/1994	Tỉnh Thái Bình	Nữ	CT2201
12101200892	Đỗ Thị	Mai	8/29/1994	TP Hải Phòng	Nữ	CT2201
12101200742	Lê Trung	Dũng	4/19/1994	Tỉnh Thanh Hóa	Nam	CT2201
12101200238	Nguyễn Thị	Hạnh	11/9/1995	Tỉnh Nghệ An	Nữ	CT2201
12101200413	Chu Minh	Đức	10/14/1995	TP Hà Nội	Nam	CT2201

## BẢNG 4-Bảng điểm trong máy

(gọi là bảng báo cáo)

Ngành CNTT

MSV	Giao_duc_t he_chat_1	Toanhoc1	Logic_hoc	Nhap_mo n_tin_hoc	Quan_tri_ hoc	Toan_giai _tich_1	Vat_ly	...
12101200188	2	2	2,5	3	2	2	3	
12101200101	2,5	3	2	1,5	2	2	2	
12101200128	2,5	2,5	2	3	2	2,5	3	
12101200050	1,5	2,5	2,5	3,2	2,5	3,2	3	
12101200892	2	2	2	3	2	3	2	
12101200742	4	2	2	2	2,5	2,5	3	
12101200238	3	1	2	2,5	2,5	2,5	2	
12101200413	3,5	3	3	2,5	3,5	3,2	3,5	

**Bước 2-** Dựa vào công thức tính điểm trong QUYCHẾTÍNHĐIỂM để tính ra Bảng KETQUA:

$$A_p = \frac{\sum_{i=1}^{N_p} a_{ip} * n_{ip}}{\sum_{i=1}^{N_p} n_{ip}}$$

Bảng KETQUA (Output- Kết quả tính trên máy)

Ngành CNTT

STT	Mã SV	STC TL	TBC TL	GDTC	GDQP	CC Ngoại ngữ	Kết quả xét	Số môn điểm F
1	12101200188	18	2,25	Đạt	Đạt	Đạt	Loại trung bình khá	0
2	12101200101	18	2,08	Đạt	Đạt	Đạt	Loại trung bình khá	0
3	12101200128	18	2,5	Đạt	Đạt	Đạt	Loại khá	0
4	12101200050	18	2,82	Đạt	Đạt	Đạt	Loại khá	0
5	12101200892	18	2	Đạt	Đạt	Đạt	Loại trung bình	0
6	12101200742	18	2,33	Đạt	Đạt	Đạt	Loại trung bình khá	0
7	12101200238	18	2,08	Đạt	Đạt	Đạt	Loại trung bình	0
8	12101200413	18	3,12	Đạt	Đạt	Đạt	Loại khá	0

STC TL - số tín chỉ tích lũy

TBC TL - điểm trung bình chung tích lũy

**Bước 3-** thực hiện phép hợp 2 tập KETQUA và BAOCÁO:

Việc so sánh bảng KETQUA với bảng BAOCÁO bằng cách thực hiện phép hợp chúng lại với nhau: KETQUA  $\cup$  BAOCÁO (nối hai tập và bỏ hai dòng hoàn toàn trùng nhau)

S TT	Mã SV	STC TL	TBC TL	GDTC	GDQP	CC Ngoại ngữ	Kết quả xét	Số môn điểm F
11	12101200188	18	3,04	Đạt	Đạt	Đạt	Loại khá	0
22	12101200101	18	2,43	Đạt	Đạt	Đạt	Loại trung bình khá	0
33	12101200128	18	2,29	Đạt	Đạt	Đạt	Loại trung bình khá	0
44	12101200050	18	2,8	Đạt	Đạt	Đạt	Loại khá	0
55	12101200892	18	2,55	Đạt	Đạt	Đạt	Loại khá	0
66	12101200742	18	2,55	Đạt	Đạt	Đạt	Loại khá	0
67	12101200238	18	2,42	Đạt	Đạt	Đạt	Loại trung bình khá	0
68	12101200413	18	2,9	Đạt	Đạt	Đạt	Loại khá	0

BẢNG 5-BAOCAO (là bảng rút gọn của BẢNG 5-Báo cáo Kết quả xét tốt nghiệp)

Kết quả thực hiện phép hợp KETQUA  $\cup$  BAOCAO (nối 2 bảng và bỏ các dòng trùng nhau) sẽ có

Bảng KETQUA  $\cup$  BAOCAO:

STT ( $t_i$ )	Mã SV	STC TL	TBC TL	GDTC	GDQP	CC Ngoại ngữ	Kết quả xét	Số môn điểm F
$Tt_1$	12101200050	18	2,8	Đạt	Đạt	Đạt	Loại khá	0
2	12101200101	18	2,43	Đạt	Đạt	Đạt	Loại trung bình khá	0
2	12101200128	18	2,29	Đạt	Đạt	Đạt	Loại trung bình khá	11
4	12101200188	18	3,04	Đạt	Đạt	Đạt	Loại khá	0
5	12101200238	18	2,42	Đạt	Đạt	Đạt	Loại trung bình khá	1
6	12101200413	18	2,9	Đạt	Đạt	Đạt	Loại khá	0
7	12101200742	18	2,55	Đạt	Đạt	Đạt	Loại khá	0
$Tt_8$	12101200892	18	<b>2,49</b>	Đạt	Đạt	Đạt	Loại trung bình khá	0
$Tt_9$	12101200892	18	<b>2,55</b>	Đạt	Đạt	Đạt	Loại khá	0

Phát hiện có 2 dòng chung khóa là  $t_8$  và  $t_9$

**Bước 4-** Phát hiện cặp phần tử ngoại lai của bảng KETQUA U BAOCÁO bằng cách rà soát có 2 dòng có khóa chính trùng nhau (Mã SV). Trong ví dụ này phát hiện mã sinh viên trùng nhau ở 2 dòng là 12101200892

Nếu có 2 dòng có khóa chính trùng nhau thì đó chính là cặp phần tử ngoại lai

**Bước 5-** Thông báo kết quả không có ngoại lai (tức là báo cáo đúng) hoặc có những cặp phần tử ngoại lai (trong ví dụ này có cặp phần tử ngoại lai là dòng 8 và dòng 9)

**Bước 6-** Kiểm tra từng cặp phần tử ngoại lai (nếu có, chẳng hạn ở đây là dòng 8 và dòng 9) để xác định dòng nào đúng, dòng nào sai tính trung bình điểm sai để tính lại cho người ấy.

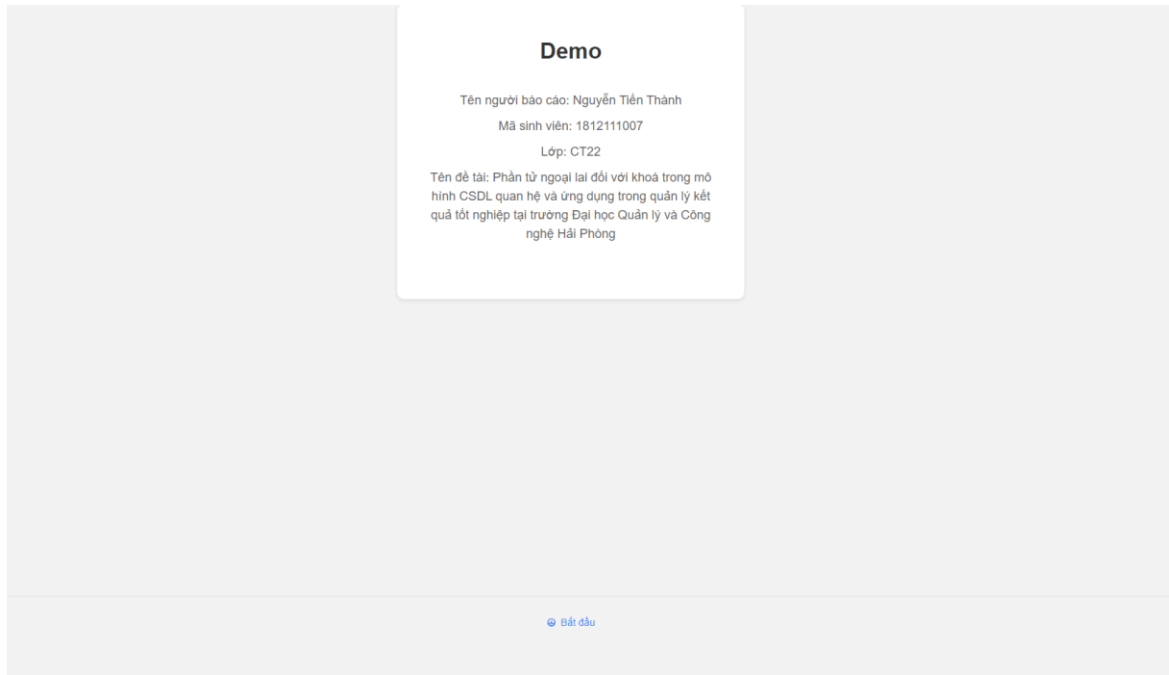


**BẢNG 07- Bảng điểm lưu (sổ điểm-bảng gốc)**

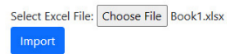
<b>STT</b> <b>((t<sub>i</sub>))</b>	<b>Mã SV</b>	<b>STC TL</b>	<b>TBC TL</b>	<b>GDTC</b>	<b>GDQP</b>	<b>CC</b> <b>Ngoại</b> <b>ngữ</b>	<b>Kết quả xét</b>	<b>Số môn</b> <b>điểm F</b>
T <sub>t1</sub>	12101200050	18	2,8	Đạt	Đạt	Đạt	Loại khá	0
2	12101200101	18	2,43	Đạt	Đạt	Đạt	Loại trung bình khá	0
2	12101200128	18	2,29	Đạt	Đạt	Đạt	Loại trung bình khá	11
4	12101200188	18	3,04	Đạt	Đạt	Đạt	Loại khá	0
5	12101200238	18	2,42	Đạt	Đạt	Đạt	Loại trung bình khá	1
6	12101200413	18	2,9	Đạt	Đạt	Đạt	Loại khá	0
7	12101200742	18	2,55	Đạt	Đạt	Đạt	Loại khá	0
<b>T<sub>t</sub></b>	<b>12101200892</b>	18	<b>2,49</b>	Đạt	Đạt	Đạt	Loại trung bình khá	0
<b>T<sub>t</sub></b>	<b>12101200892</b>	18	<b>2,55</b>	Đạt	Đạt	Đạt	Loại khá	0

### 3.3. Các giao diện chính của chương trình

#### 3.3.1. Giao diện trang giới thiệu



#### 3.4.2. Giao diện nhập file dữ liệu



### 3.4.3. Giao diện kết quả khi có phần tử ngoại lai

[Import File Excel](#)

#### Kết quả

Mã SV	STC	TL	TBC	TL	GDTC	GDQP	CC	Ngoại ngữ	Kết quả	Số môn điểm F
1210120005018	2.82	Đạt	Đạt	Đạt					Loại khá	0
1210120010118	2.08	Đạt	Đạt	Đạt					Loại trung bình	0
1210120012818	2.5	Đạt	Đạt	Đạt					Loại khá	0
1210120018818	2.25	Đạt	Đạt	Đạt					Loại trung bình khá	0
1210120023818	2.08	Đạt	Đạt	Đạt					Loại trung bình	0
1210120041318	3.12	Đạt	Đạt	Đạt					Loại khá	0
1210120074218	2.33	Đạt	Đạt	Đạt					Loại trung bình khá	0
1210120089218	2	Đạt	Đạt	Đạt					Loại trung bình	0

#### Báo cáo

Mã SV	STC	TL	TBC	TL	GDTC	GDQP	CC	Ngoại ngữ	Kết quả	Số môn điểm F
1210120005018	2.82	Đạt	Đạt	Đạt					Loại khá	0
1210120010118	2.08	Đạt	Đạt	Đạt					Loại trung bình	0
1210120012818	1	Đạt	Đạt	Đạt					Loại khá	0
1210120018818	2.25	Đạt	Đạt	Đạt					Loại trung bình khá	0
1210120023818	2.08	Đạt	Đạt	Đạt					Loại trung bình	0
1210120041318	3	Đạt	Đạt	Đạt					Loại khá	0
1210120074218	2.33	Đạt	Đạt	Đạt					Loại trung bình khá	0
1210120089218	2	Đạt	Đạt	Đạt					Loại trung bình	0

#### Dữ liệu không hợp lệ:

Mã SV	STC	TL	TBC	TL	GDTC	GDQP	CC	Ngoại ngữ	Kết quả	Số môn điểm F
12101200128	18	1	Đạt	Đạt	Đạt	Đạt	Đạt		Loại khá	0
12101200413	18	3	Đạt	Đạt	Đạt	Đạt	Đạt		Loại khá	0

### 3.4.4. Giao diện kết quả khi không có phần tử ngoại lai

[Import File Excel](#)

#### Kết quả

Mã SV	STC	TL	TBC	TL	GDTC	GDQP	CC	Ngoại ngữ	Kết quả	Số môn điểm F
1210120005018	2.82	Đạt	Đạt	Đạt					Loại khá	0
1210120010118	2.08	Đạt	Đạt	Đạt					Loại trung bình	0
1210120012818	2.5	Đạt	Đạt	Đạt					Loại khá	0
1210120018818	2.25	Đạt	Đạt	Đạt					Loại trung bình khá	0
1210120023818	2.08	Đạt	Đạt	Đạt					Loại trung bình	0
1210120041318	3.12	Đạt	Đạt	Đạt					Loại khá	0
1210120074218	2.33	Đạt	Đạt	Đạt					Loại trung bình khá	0
1210120089218	2	Đạt	Đạt	Đạt					Loại trung bình	0

#### Báo cáo

Mã SV	STC	TL	TBC	TL	GDTC	GDQP	CC	Ngoại ngữ	Kết quả	Số môn điểm F
1210120005018	2.82	Đạt	Đạt	Đạt					Loại khá	0
1210120010118	2.08	Đạt	Đạt	Đạt					Loại trung bình	0
1210120012818	2.5	Đạt	Đạt	Đạt					Loại khá	0
1210120018818	2.25	Đạt	Đạt	Đạt					Loại trung bình khá	0
1210120023818	2.08	Đạt	Đạt	Đạt					Loại trung bình	0
1210120041318	3.12	Đạt	Đạt	Đạt					Loại khá	0
1210120074218	2.33	Đạt	Đạt	Đạt					Loại trung bình khá	0
1210120089218	2	Đạt	Đạt	Đạt					Loại trung bình	0

#### Không có phần tử ngoại lai

## KẾT LUẬN

Đề tài tốt nghiệp “Phần tử ngoại lai đối với khóa trong mô hình CSDL quan hệ và ứng dụng trong quản lý kết quả tốt nghiệp tại trường Đại học Quản lý và Công nghệ Hải Phòng” đã có ý nghĩa khoa học và tính thực tế nhất định. Việc ứng dụng chủ đề về phần tử ngoại lai trong đời sống kinh tế - xã hội – giáo dục đang ngày được mở rộng và phát triển.

Nội dung đồ án về cơ bản đã đáp ứng đầy đủ các yêu cầu cần giải quyết đề ra trong nhiệm vụ đề tài tốt nghiệp:

- Tổng quan được những vấn đề chung nhất về mô hình CSDL quan hệ và vai trò của Khóa trong mô hình quan hệ dữ liệu.
- Tìm hiểu rõ về phần tử ngoại lai đối với vai trò của nó trong mô hình CSDL quan hệ và mô hình phát hiện phần tử ngoại lai đối với phụ thuộc hàm và khóa.
- Biết ứng dụng lý thuyết phần tử ngoại lai để kiểm tra đúng sai kết luận xếp loại học lực và danh hiệu cho sinh viên trường ĐH QL&CN Hải Phòng.
- Phần thử nghiệm chương trình tuy còn đơn giản nhưng đã chứng tỏ được khả năng nhận thức và khả năng vận dụng kiến thức lý thuyết học được trong nhà trường vào giải quyết những bài toán thực tế.

Kết quả đạt được trong đồ án là sự cố gắng của bản thân em và chất lượng giáo dục của nhà trường. Em xin chân thành cảm ơn các Thầy Cô Trường Đại Học Quản Lý và Công Nghệ Hải Phòng, đặc biệt là Thầy Lê Văn Phùng người đã trực tiếp hướng dẫn em thực hiện đồ án này.