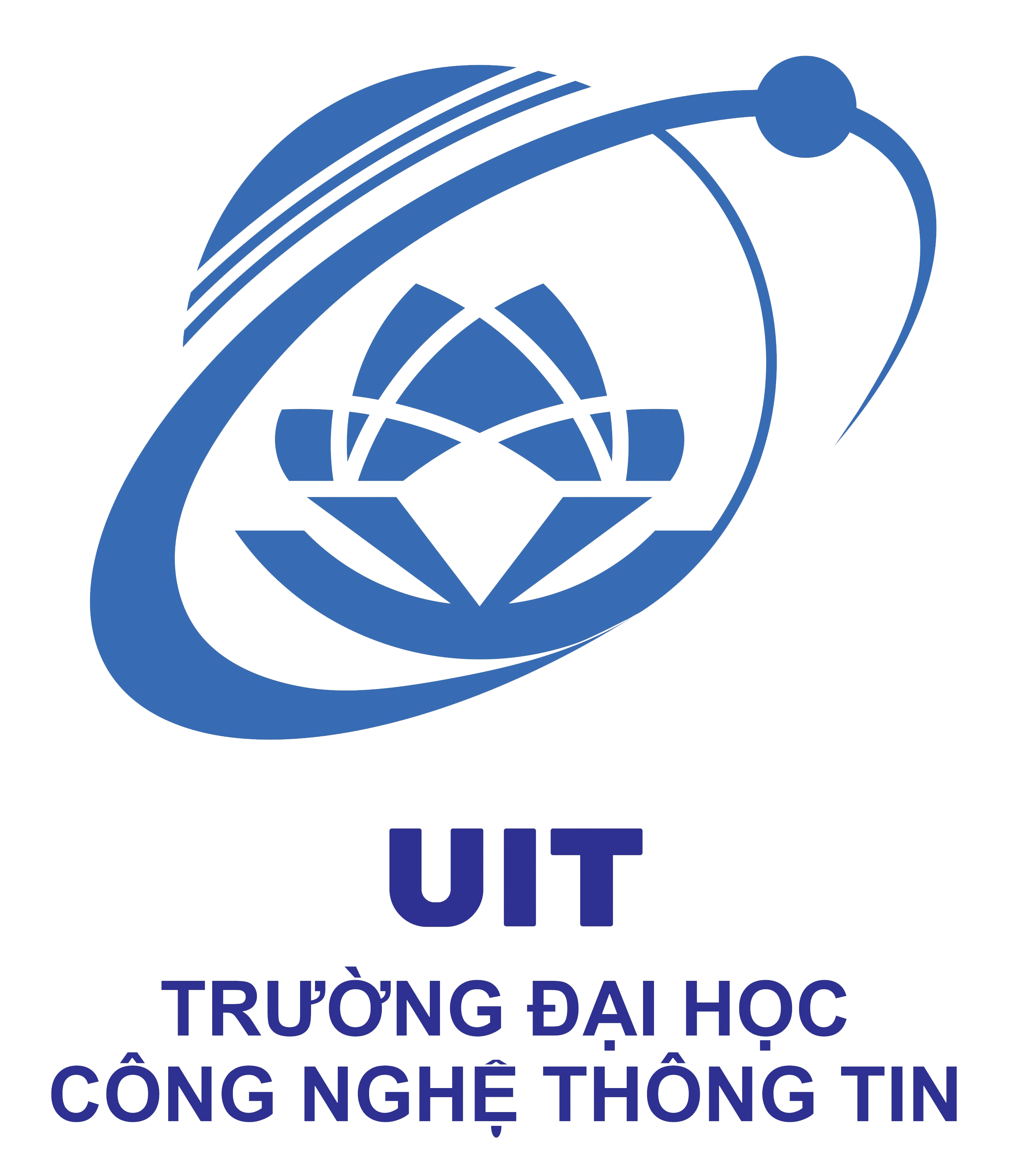
**TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN**

**ĐẠI HỌC QUỐC GIA THÀNH PHỐ HỒ CHÍ MINH**

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ ★★★ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**KHOA: KHOA HỌC MÁY TÍNH**



**Các Hệ Thống Hỏi-Đáp [CS323.H11]**

**BÁO CÁO**

**Natural Language Interface to**

**a DataBase**

**GV Lý Thuyết: Nguyễn Trọng Chỉnh**

**Sinh Viên thực hiện:**

**1. Lê Thành Đạt 13520199**

**2. Huỳnh Văn Tâm 13520735**

**TP.Hồ Chí Minh, tháng 12 năm 2016**

**Lời cảm ơn**

Trước tiên, chúng em xin gửi lời cảm ơn và lòng biết ơn sâu sắc nhất tới ThS. Nguyễn Trọng Chỉnh -người đã chỉ bảo và hướng dẫn cho chúng em trong quá trình thực hiện đồ án cuối kỳ .

Đồ án cuối kỳ “xây dựng một hệ thống hỏi đáp theo hướng chú giải tri thức NLIDB ” của môn Các Hệ Thống Hỏi-Đáp là đồ án giúp cho chúng em hiểu về hệ thống hỏi-đáp mà chúng em đã học. Đồ án cũng tạo điều kiện cho em có thể viết các chương trình cơ bản để truy xuất dữ liệu từ câu truy vấn tự nhiên bằng phương pháp NLIDB , cũng như phát triển thành phần mền liên quan đến các hệ thống hỏi-đáp sau này .

Trong quá trình thực hiện đồ án , chúng em sẽ không tránh khỏi những sai sót .Chúng em rất mong nhận được sự thông cảm và góp ý của quý thầy để chúng em có thể nắm bắt chắc hơn về kiến thức , giúp hoàn thiện và hiểu sâu về các hệ thống hỏi-đáp.  
  
Chúng em xin chân thành cảm ơn thầy !

Mục Lục

[**I . Cơ sở dữ liệu được sử dụng trong hệ thống hỏi-đáp NLIDB :** 4](#_Toc470309543)

[**II. Tập câu hỏi được sử dụng để truy vấn thông tin và truy vấn SQL tương ứng trên dữ liệu đã chọn** 5](#_Toc470309544)

[**III. Xây dựng quy tắc chuyển từ các thành phần trong cây cú pháp của câu hỏi sang ý nghĩa tương ứng theo câu truy vấn SQL**. 8](#_Toc470309545)

[**IV. Bộ phân tích cú pháp cho tập câu hỏi:** 10](#_Toc470309546)

[**V. Thực hiện Demo một số truy vấn trên một hệ quản trị CSDL** 14](#_Toc470309547)

[liệt kê thông tin của người có tên là NGUYỄN HỮU THÀNH 14](#_Toc470309548)

[liệt kê tên , địa chỉ , phương tiện di chuyển của các ứng viên có giới tính là nữ và lương yêu cầu trên 3 14](#_Toc470309549)

[hãy cho biết các loại học vấn của các ứng viên 15](#_Toc470309550)

[liệt kê tất cả thông tin của các ứng viên có vị trí ứng tuyển thuộc cộng tác viên 15](#_Toc470309551)

[**VI. Đánh giá kết quả và nhận xét :** 16](#_Toc470309552)

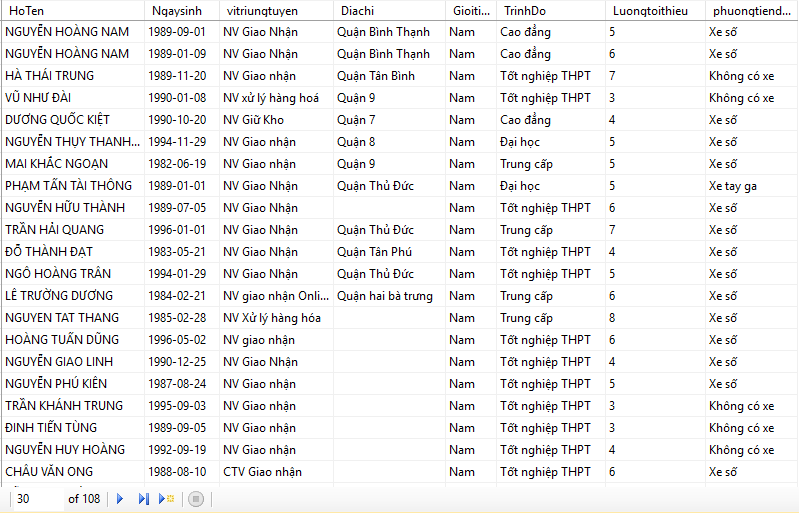
[1. Ngày/tháng/năm hoặc tên riêng thì không thể định nghĩa toàn bộ trong tập luật của Prolog được . 16](#_Toc470309553)

[2. Lỗi sinh ra câu truy vấn Sql lỗi cú pháp 16](#_Toc470309554)

[**VII. Lời Cam Đoan** 17](#_Toc470309555)

# **I . Cơ sở dữ liệu được sử dụng trong hệ thống hỏi-đáp NLIDB :**

Dữ liệu được sử dụng trong hệ thống hỏi-đáp NLIDB là bộ dữ liệu quản lý các ứng viên đang tuyển dụng . Bộ dữ liệu gồm 8 cột gồm họ tên , ngày sinh , vị trí ứng tuyển , địa chỉ , giới tính ,trình độ học vấn , lương yêu cầu tối thiểu , phương tiện di chuyển .



# **II. Tập câu hỏi được sử dụng để truy vấn thông tin và truy vấn SQL tương ứng trên dữ liệu đã chọn**

|  |  |  |
| --- | --- | --- |
| **Stt** | **Câu truy vấn bằng ngôn ngữ tự nhiên** | **Câu truy vấn SQL tương ứng** |
| 1 | liệt kê tên của các ứng viên | Select hoten from UngVien |
| 2 | hãy cho biết thông tin của các ứng viên có năm sinh là 1993 | Select \* from UngVien where year(ngaysinh) = 1993 |
| 3 | hãy cho biết tên của các ứng viên có tháng sinh là 7 và năm sinh là 1987 | Select hoten from UngVien where month(ngaysinh) =7 and year(ngaysinh) = 1987 |
| 4 | liệt kê tên của các ứng viên sinh tháng 7 và năm 1987 | Select hoten from UngVien where month(ngaysinh) =7 and year(ngaysinh) = 1987 |
| 5 | liệt kê tất cả thông tin của ứng viên được sinh ngày 07/12/1987 | Select \*  from UngVien  where ngaysinh='07/12/1987' |
| 6 | hãy cho biết thông tin của các ứng viên có số tuổi là 20 | Select \*  from UngVien  where DATEDIFF(year,Ngaysinh,GETDATE())=20 |
| 7 | liệt kê tên, địa chỉ , lương yêu cầu của các ứng viên muốn ứng tuyển vị trí nhân viên giao nhận | Select HoTen,diachi,luongtoithieu  from UngVien  where vitriungtuyen=N'NV giao nhận' |
| 8 | liệt kê tên của các ứng viên có địa chỉ ở Quận Thủ Đức | Select HoTen  from UngVien  where Diachi=N'Quận Thủ Đức' |
| 9 | liệt kê tên của các ứng viên có giới tính là nữ | Select HoTen  from UngVien  where Gioitinh=N'Nữ' |
| 10 | liệt kê tên, học vấn của các ứng viên có địa chỉ không rõ ràng | Select HoTen,trinhdo  from UngVien  where Diachi='' |
| 11 | cho biết lương yêu cầu của ứng viên có trình độ là Tốt nghiệp THPT | Select luongtoithieu  from UngVien  where TrinhDo=N'Tốt nghiệp THPT' |
| 12 | liệt kê thông tin của các ứng viên có lương yêu cầu là 5 | Select \*  from UngVien  where Luongtoithieu=5 |
| 13 | liệt kê tên của các ứng viên có vị trí ứng tuyển thuộc nhân viên | select hoten  from ungvien  where vitriungtuyen like N'%NV%' |
| 14 | hãy cho biết thông tin của các ứng viên có phương tiện di chuyển là xe số | Select \*  from UngVien  where phuongtiendichuyen =N'xe số' |
| 15 | liệt kê thông tin của các ứng viên có lương yêu cầu lớn hơn hoặc bằng 5 | Select \*  from UngVien  where Luongtoithieu >= 5 |
| 16 | liệt kê tên của các ứng viên có phương tiện di chuyển là xe số hoặc phương tiện di chuyển là xe tay ga | Select Hoten  From UngVien  Where phuongtiendichuyen = N'xe số' OR phuongtiendichuyen = N'xe ga' |
| 17 | liệt kê tên của các ứng viên có lương yêu cầu từ 3 đến 5 | Select HoTen  from UngVien  where Luongtoithieu between 3 and 5 |
| 18 | liệt kê các loại phương tiện di chuyển của các ứng viên | Select DISTINCT phuongtiendichuyen  from UngVien |
| 19 | hãy cho biết các loại vị trí ứng tuyển của các ứng viên | Select DISTINCT vitriungtuyen  from UngVien |
| 20 | hãy cho biết các loại học vấn của các ứng viên | Select DISTINCT TrinhDo  from UngVien |
| 21 | liệt kê tên của các ứng viên muốn ứng tuyển vị trí nhân viên giao nhận và đang ở Quận Tân Bình | Select hoten  from UngVien  where vitriungtuyen= N'NV giao nhận' and Diachi= N'Quận Tân Bình' |
| 22 | liệt kê tên , địa chỉ , phương tiện di chuyển của các ứng viên có giới tính là nữ và lương yêu cầu trên 3 | select Hoten, Diachi,phuongtiendichuyen  from UngVien  where Gioitinh=N'Nữ' and Luongtoithieu>3 |
| 23 | cho biết lương yêu cầu cao nhất của các ứng viên | select max(Luongtoithieu)  from ungvien |
| 24 | liệt kê tất cả thông tin của các ứng viên | select \* from ungvien |
| 25 | hãy cho biết vị trí ứng tuyển của ứng viên có tên là HOÀNG TUẤN DŨNG | select vitriungtuyen  from UngVien  where hoten=N' HOÀNG TUẤN DŨNG ' |
| 26 | hãy cho biết số lượng của các ứng viên có lương yêu cầu là 3 | select count(\*)  from ungvien  where Luongtoithieu=3 |
| 27 | hãy cho biết thông tin của các ứng viên có số tuổi trên 20 | Select \*  from UngVien  where DATEDIFF(year,Ngaysinh,GETDATE())>20 |
| 28 | liệt kê tên của các ứng viên có vị trí ứng tuyển thuộc cộng tác viên | select hoten  from ungvien  where vitriungtuyen like N'%CTV%' |
| 29 | liệt kê tên của các ứng viên có lương yêu cầu từ 3 - 7 | Select HoTen  from UngVien  where Luongtoithieu between 3 and 7 |
| 30 | cho biết lương yêu cầu thấp nhất của các ứng viên | select min(Luongtoithieu)  from ungvien |
| 31 | cho biết thông tin của ứng viên có trình độ là Đại Học | Select luongtoithieu  from UngVien  where TrinhDo=N'Đại Học' |
| 32 | liệt kê tên của các ứng viên có lương yêu cầu từ 3 tới 7 | Select HoTen  from UngVien  where Luongtoithieu between 3 and 7 |

# **III. Xây dựng quy tắc chuyển từ các thành phần trong cây cú pháp của câu hỏi sang ý nghĩa tương ứng theo câu truy vấn SQL.**

1. Liệt kê tên của các ứng viên

Select hoten from UngVien

from

select

Hoten

ungvien

s(SEM) -->vp(SEM).

vp(SEM)--> vb(VB),sbar(SBAR), {append(VB,SBAR,SEM)}.

sbar(SEM)-->np(SEM).

np(SEM)-->nn(NN),pp(PP),{append(NN,PP,SEM)}.

pp(SEM)-->in(IN),np(NP),{append(IN,NP,SEM)}.

np(SEM)-->nn(SEM).

vb("Select ")-->[liệt,kê].

nn("Hoten ")-->[tên].

in("From ")-->[của].

nn("UngVien ")--> [các,ứng,viên].

1. Liệt kê tên của các ứng viên có năm sinh là/nhỏ hơn/lớn hơn X ? (trong đó X là năm sinh) ?

Select hoten from UngVien where year(ngaysinh) = X

s(SEM) -->vp(SEM).

vp(SEM)--> vb(VB),sbar(SBAR), {append(VB,SBAR,SEM)}.

sbar(SEM)-->np(NP),vp(VP),{append(NP,VP,SEM)}.np(SEM)-->nn(NN),pp(PP),{append(NN,PP,SEM)}.

pp(SEM)-->in(IN),np(NP),{append(IN,NP,SEM)}.

np(SEM)-->nn(SEM).

np(SEM)-->cd(SEM).

vb("Select ")-->[liệt,kê].

nn("Hoten ")-->[tên].

in("From ")-->[của].

nn("UngVien ")--> [các,ứng,viên].

vb("where ")-->[có].

nn("year(ngaysinh) " )--> [năm,sinh].

vb("> ")-->[lớn,hơn].

vb("< ")-->[nhỏ,hơn].

vb("= ")-->[là];[bằng].

cd(SEM)-->[X],

{

float(X),with\_output\_to(codes(S1),write(X)),append(S1,"",SEM);

integer(X),with\_output\_to(codes(S1),write(X)),append(S1,"",SEM)

}.

# **IV. Bộ phân tích cú pháp cho tập câu hỏi:**

Sau khi xây dựng phân tích cú pháp của tập câu hỏi ,dưới đây bộ phân tích cú pháp tổng hợp từ 32 câu trên

s(SEM)-->vp(SEM).

vp(VB)-->vb(VB).

vp(SEM)-->vb(VB),np(NP),{append(VB,NP,SEM)}.

vp(SEM)-->vb(VB),np(NP),adjp(ADJP),{append(VB,NP,T1),append(T1,ADJP,SEM)}.

vp(SEM)-->vb(VB),sbar(SBAR),{append(VB,SBAR,SEM)}.

np(SEM)-->nn(NN),pp(PP),{append(NN,PP,SEM)}.

np(SEM)-->nn(SEM).

np(SEM)-->cd(SEM).

np(SEM)-->nn(NN),np(NP),{append(NN,NP,SEM)}.

np(SEM)-->nn(NN),cc(CC),np(NP),{append(NN,CC,T1),append(T1,NP,SEM)}.

np(SEM)-->cd(CD),cc(CC),np(NP),{append(CD,CC,T1),append(T1,NP,SEM)}.

np(SEM)-->nn(NN),np(NP1),cc(CC),np(NP2),{append(NN,NP1,T1),append(T1,CC,T2),append(T2,NP2,SEM)}.

pp(SEM)-->in(IN),np(NP),{append(IN,NP,SEM)}.

pp(SEM)-->in(IN),np(NP),pp(PP),{append(IN,NP,T1),append(T1,PP,SEM)}.

sbar(SEM)-->np(NP),vp(VP),{append(NP,VP,SEM)}.

sbar(SEM)-->np(NP),vp(VP),cc(CC),sbar(SBAR),{append(NP,VP,T1),append(T1,CC,T2),append(T2,SBAR,SEM)}.

adjp(SEM)-->jj(SEM).

vb("Where ")-->[có];[được];[muốn];[sinh].

vb("Select ")-->[liệt,kê];[cho,biết];[hãy,cho,biết].

vb("> ")-->[lớn,hơn];[trên].

vb("< ")-->[nhỏ,hơn];[dưới].

vb("= ")-->[là];[bằng];[ở].

vb(">= ")-->[lớn,hơn,hoặc,bằng].

vb("<= ")-->[bé,hơn,hoặc,bằng].

vb(" like ") -->[thuộc].

nn("Hoten ")-->[tên];[họ,và,tên];[họ,tên].

nn("UngVien ")--> [ứng,viên];[các,ứng,viên];[những, người];[người].

nn("year(ngaysinh) ")--> [năm,sinh].

nn("month(ngaysinh) ")--> [tháng,sinh].

nn("month(ngaysinh) = ")--> [tháng].

nn("year(ngaysinh) = ")--> [năm].

nn(" ngaysinh = ")-->[sinh,ngày];[sinh,vào].

nn(" Gioitinh ") -->[giới,tính].

nn(" Diachi ")-->[địa,chỉ].

nn(" diachi= ") -->[đang,ở].

nn(" ngaysinh ")-->[ngày,sinh].

nn(" vitriungtuyen ")-->[vị,trí,ứng,tuyển].

nn(" count(\*) ")-->[số,lượng].

nn(" vitriungtuyen= ") -->[ứng,tuyển,vị,trí].

nn(" trinhdo ")-->[trình,độ];[học,vấn].

nn(" luongtoithieu ")-->[lương,yêu,cầu];[lương,tối,thiểu];[lương].

nn(" phuongtiendichuyen ") -->[phương,tiện,di,chuyển];[xe].

nn("DATEDIFF(year,Ngaysinh,GETDATE()) ") -->[số,tuổi];[tuổi].

nn("N'NV Giao Nhận'") -->[nhân,viên,giao,nhận].

nn("N'CTV Giao nhận'") -->[cộng,tác,viên,giao,nhận].

nn("N'NV Xử Lý hàng hóa '") -->[nhân,viên,xử ,lý ,hàng ,hóa].

nn("N'NV phân hàng '") -->[nhân,viên,phân, hàng].

nn("N'%CTV%'") -->[cộng,tác,viên].

nn("N'%NV%'") -->[nhân,viên].

nn("N'nữ'")-->[nữ].

nn("N'nam'")-->[nam].

nn("N'xe số'")-->[xe,số].

nn("N'xe ga'")-->[xe,tay,ga].

nn("N'không có xe'")-->[không,có,xe].

nn(" DISTINCT ")-->[các,loại].

nn(" \* ")-->[tất,cả,thông,tin];[tất,cả,các,thông,tin];[thông,tin].

nn(" 'datetime' ") -->[datetime].

nn(" N'specificname' ")-->[specificname].

nn(" max(Luongtoithieu) ") -->[lương,cao,nhất];[lương,yêu,cầu,cao,nhất];[lương,tối,thiểu,cao,nhất].

nn(" min(Luongtoithieu) ") -->[lương,thấp,nhất];[lương,yêu,cầu,thấp,nhất];[lương,tối,thiểu,thấp,nhất].

in(" From ")-->[của].

in(" Between ") -->[từ].

cc(" OR ")-->[hoặc];[hay].

cc(" AND ")-->[và];[đến];[-];[tới].

cc(" , ")-->[,].

jj(" = '' ") -->[không,rõ,ràng].

jj(" not ") -->[không,phải].

cd(SEM)-->[X],

{

float(X),with\_output\_to(codes(S1),write(X)),append(S1,"",SEM);

integer(X),with\_output\_to(codes(S1),write(X)),append(S1,"",SEM)

}.

test(Term, String) :-

with\_output\_to(char(String), write(Term)).

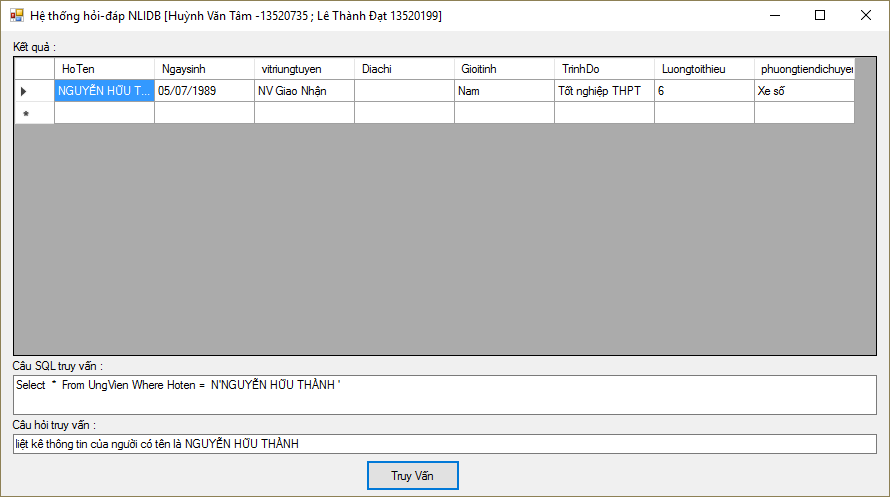
getcd(A,S):-nnp(B,A,[]),with\_output\_to(atom(S),writef(B,"%s")),!.

getquery(A,S):-s(B,A,[]),with\_output\_to(atom(S),writef(B,"%s")),!.

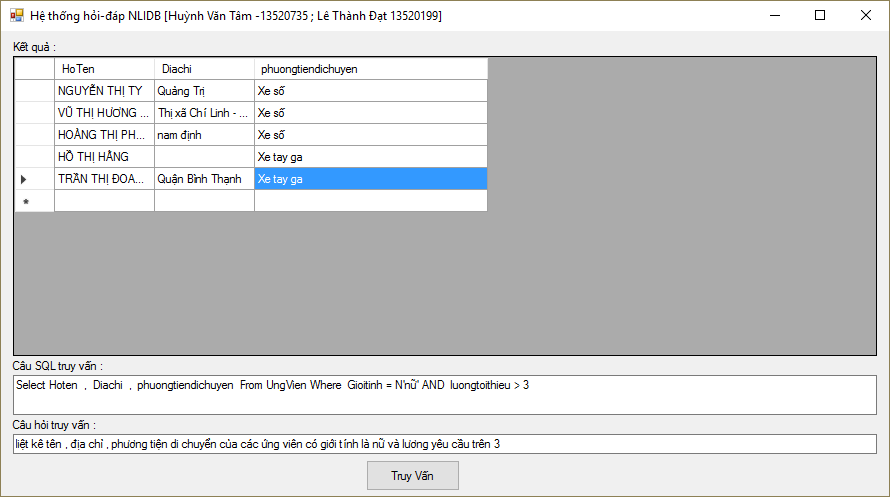
%getquery([cho,biết,tên,của,các,ứng,viên,có,năm,sinh,là,1997],S).

# **V. Thực hiện Demo một số truy vấn trên một hệ quản trị CSDL**

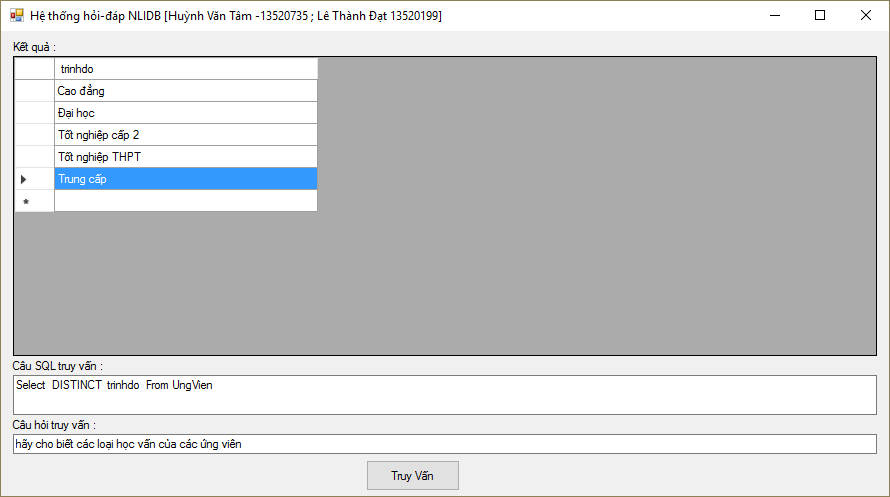
## liệt kê thông tin của người có tên là NGUYỄN HỮU THÀNH



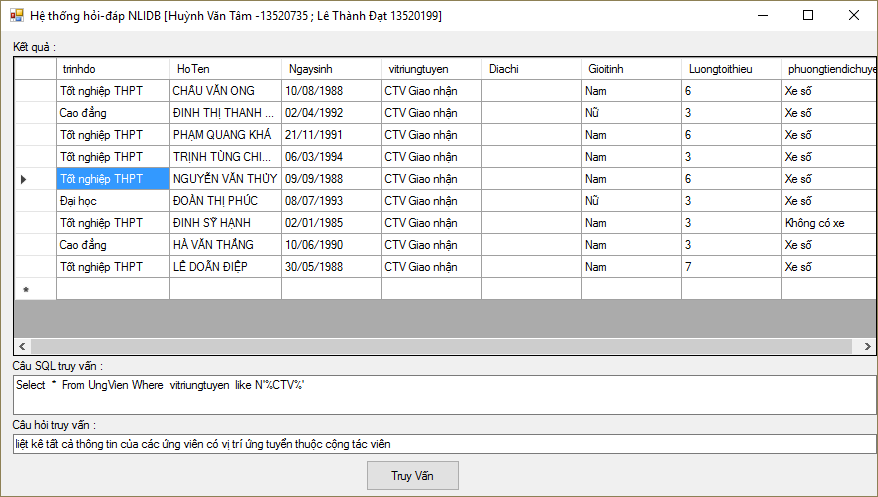
## liệt kê tên , địa chỉ , phương tiện di chuyển của các ứng viên có giới tính là nữ và lương yêu cầu trên 3



## hãy cho biết các loại học vấn của các ứng viên



## liệt kê tất cả thông tin của các ứng viên có vị trí ứng tuyển thuộc cộng tác viên



# **VI. Đánh giá kết quả và nhận xét :**

Trong quá trình thực hiện việc truy vấn , chúng em có một sốvấn đề sau :

1. **Ngày/tháng/năm hoặc tên riêng thì không thể định nghĩa toàn bộ trong tập luật của Prolog được .**

Để giải quyết việc này , trước khi đưa chuỗi sang Prolog xử lý , trong C# thay thế các tên riêng hoặc ngày/tháng/năm bằng một chuỗi tạm thời (datetime và specialName) . Đồng thời lưu trữ các chuỗi này vào các biến .   
Trong prolog thì định nghĩa sẵn 2 chuỗi này thể thực hiện việc truy xuất ra SQL .   
Khi đã có câu truy vấn SQL , khi trở lại C# , ta thay thế chuỗi tạm thời khi bằng chuỗi tên riêng hoặc chuỗi ngày/tháng/năm thực sự .

Vd: liệt kê tên của các ứng viên có ngày sinh là 17/01/1987

B1 : Replace tên riêng hoặc dd/mm/yyyy bằng chuỗi datetime hoặc specialName  
getquery([liệt ,kê ,tên, của, các, ứng ,viên ,có ,ngày,sinh ,là , **datetime**],S).

B2 : khi đã sinh được câu truy vấn SQL như sau :  
 Select Hoten From UngVien Where ngaysinh=’**datetime**’  
 Ta thay datetime bằng chuỗi dd/mm/yyyy ban đầu , ta được câu truy vấn SQL hoàn chỉnh:  
 🡪 **Select Hoten From UngVien Where ngaysinh=’ 17/01/1987’**

1. **Lỗi sinh ra câu truy vấn Sql lỗi cú pháp**

Liệt kê tên các ứng viên --> select hoten Ungvien  
or

Liệt kê tên của các ứng viên có giới tính nữ 🡪 selecte hoten from Ungvien where gioitinh N‘nữ’

Lỗi này hiện tại chúng em vẫn chưa giải quyết được .

# **VII. Lời Cam Đoan**

Nhóm chúng em xin cam đoan về nội dung đồ án , cũng như souce code hoàn toàn do nhóm tự thiết kế và lập trình, hoàn toàn không sao chép nội dung cơ bản từ các đồ án khác , và sản phẩm tạo ra là do bản thân nhóm em tự mình xây dựng nên.