Trí Tuệ Nhân Tạo

(Artificial Intelligence)

Lê Thanh Hương

Viện Công nghệ thông tin và Truyền thông
Trường Đại Học Bách Khoa Hà Nội

Nội dung môn học

Chương 1. Tổng quan

Chương 2. Tác tử thông minh

Chương 3. Giải quyết vấn đề

Chương 4. Tri thức và suy diễn

- Giới thiệu về logic
- Logic định đề
- Logic vị từ

Chương 5. Học máy

Giới hạn của Logic định đề

- Hãy xét ví dụ sau đây:
  - Tuấn là một sinh viên của HUST
  - Mọi sinh viên của HUST đều học môn Đại số
  - Vì Tuấn là một sinh viên của HUST, nên Tuấn học môn Đại số
- Trong logic định đề:
  - Định đề *p*: "Tuấn là một sinh viên của HUST"
  - Định đề *q*: "Mọi sinh viên của HUST đều học môn Đại số"
  - Định đề *r*: "Tuấn học môn Đại số"
  - Nhưng: (trong logic định đề) *r* không thể suy ra được từ *p* và *q*!

Logic vị từ (FOL): Ví dụ

- Ví dụ nêu trên có thể được biểu diễn trong logic vị từ bởi các biểu thức (logic vị từ) sau
  - *HUST_Student(Tuan)*: "Tuấn là một sinh viên của HUST"
  - *∀x:HUST_Student(x) → Studies_Algebra(x)*: "Mọi sinh viên của HUST đều học môn Đại số"
  - *Studies_Algebra(Tuan)*: "Tuấn học môn Đại số"
- Trong logic vị từ, chúng ta có thể chứng minh được:

{HUST_Student(Tuan), ∀x:HUST_Student(x) → Studies_Algebra(x)} ⊢ Studies_Algebra(Tuan)

- Với ví dụ trên, trong logic vị từ:
  - Các ký hiệu *Tuan*, *x* được gọi là các **phần tử** (*Tuan* là hằng, *x* là biến)
  - Các ký hiệu *HUST_Student* và *Studies_Algebra* là các **vị từ**
  - Ký hiệu *∀* là **lượng từ với mọi**
  - Các phần tử, các vị từ và các lượng từ cho phép biểu diễn các biểu thức

FOL: Ngôn ngữ (1)

- 4 kiểu **ký hiệu (symbols)**
  - **Hằng (Constants)**: Các tên của các đối tượng trong một lĩnh vực bài toán cụ thể (ví dụ: *Tuan*)
  - **Biến (Variables)**: Các ký hiệu mà giá trị thay đổi đối với các đối tượng khác nhau (ví dụ: *x*)
  - **Ký hiệu hàm (Function symbols)**: Các ký hiệu biểu diễn ánh xạ (quan hệ hàm) từ các đối tương của miền (domain) này sang các đối tượng của miền khác (ví dụ: *plus*)
  - **Các vị từ (Predicates)**: Các quan hệ mà giá trị logic là đúng hoặc sai (ví dụ: *HUST_Student* and *Studies_Algebra*)
- Mỗi ký hiệu hàm hoặc vị từ đều có một tập các tham số
  - Ví dụ: *HUST_Student* và *Studies_Algebra* là các vị từ có 1 tham số
  - Ví dụ: *plus* là một ký hiệu hàm có 2 tham số

FOL: Ngôn ngữ (2)

- Một **phần tử (term)** được định nghĩa (truy hồi) như sau
  - Một hằng số là một phần tử
  - Một biến là một phần tử
  - Nếu $t_1, t_2, \ldots, t_n$ là các thành phần và f là một ký hiệu hàm có n tham số, thì $f(t_1, t_2, \ldots, t_n)$ là một phần tử
  - Không còn gì khác là một phần tử
- Các ví dụ của phần tử (term)
  - *Tuan*
  - *2*
  - *friend(Tuan)*
  - *friend(x)*
  - *plus(x,2)*

FOL: Language (3)

- **Các nguyên tử (Atoms)**
  - Nếu $t_1,t_2,\ldots,t_n$ là các thành phần (terms) và p là một vị từ có n tham số, thì $p(t_1,t_2,\ldots,t_n)$ là một nguyên tử (atom)
  - Ví dụ: `HUST_Studies(Tuan)`, `HUST_Studies(x)`, `Studies_Algebra(Tuan)`, `Studies(x)`
- **Các biểu thức (Formulas)** được định nghĩa như sau
  - Một nguyên tử (atom) là một biểu thức
  - Nếu ϕ và ψ là các biểu thức, thì $\neg\phi$ và $\phi\wedge\psi$ là các biểu thức
  - Nếu ϕ là một biểu thức và x là một biến, thì $\forall x{:}\phi(x)$ là một biểu thức
  - Không còn gì khác là một biểu thức
- Lưu ý: $\exists x:\phi(x)$ được định nghĩa bằng $\neg\forall x:\neg\phi(x)$

FOL: Ngữ nghĩa (1)

- Một **phép diễn giải (interpretation)** của một biểu thức ϕ được biểu diễn bằng cặp $<\mathcal{D},\mathcal{I}>$
- **Miền giá trị (Domain)** $\mathcal{D}$ là một tập khác rỗng
- **Hàm diễn giải (Interpretation function)** $\mathcal{I}$ là một phép gán giá trị đối với mỗi hằng, ký hiệu hàm, và ký hiệu vị từ – sao cho:
  - Đối với hằng `c`: $\mathcal{I}(\mathtt{c}) \in \mathcal{D}$
  - Đối với ký hiệu hàm (có *n* tham số) `f`: $\mathcal{I}(\mathtt{f})$: $\mathcal{D}^{n} \rightarrow \mathcal{D}$
  - Đối với ký hiệu vị từ (có *n* tham số) `P`: $\mathcal{I}(\mathtt{P})$: $\mathcal{D}^{n} \rightarrow \{\mathtt{true}, \mathtt{false}\}$

FOL: Ngữ nghĩa (2)

- **Diễn giải đối với một biểu thức logic vị từ**. Giả sử ϕ, ψ và λ là các biểu thức vị từ
  - Nếu ϕ là $\neg\psi$, thì $\mathcal{I}(\phi)$=`sai` nếu $\mathcal{I}(\psi)$=`đúng`, và $\mathcal{I}(\phi)$=`đúng` nếu $\mathcal{I}(\psi)$=`sai`
  - Nếu ϕ là $(\psi\wedge\lambda)$, thì $\mathcal{I}(\phi)$=`sai` nếu $\mathcal{I}(\psi)$ hoặc $\mathcal{I}(\lambda)$ là `sai`, và $\mathcal{I}(\phi)$=`true` nếu cả $\mathcal{I}(\psi)$ và $\mathcal{I}(\lambda)$ là `đúng`
  - Giả sử $\forall x:\phi(x)$ là một biểu thức, thì $\mathcal{I}(\forall x:\phi(x))$=`đúng` nếu $\mathcal{I}(\phi)(d)$=`đúng` với mọi giá trị $d \in \mathcal{D}$

FOL: Ngữ nghĩa (3)

- Một biểu thức ϕ là **thỏa mãn được (satisfiable)** nếu và chỉ nếu tồn tại một phép diễn giải $<\mathcal{D}, \mathcal{I}>$ sao cho $\mathcal{I}(\phi)$ – Chúng ta ký hiệu là: $\models_{\mathcal{I}} \boldsymbol{\phi}$
- Nếu $\models_{\mathcal{I}} \phi$, thì chúng ta nói rằng $\mathcal{I}$ là một **mô hình (model)** của ϕ. Nói cách khác, $\mathcal{I}$ **thỏa mãn (satisfies)** ϕ
- Một biểu thức là **không thể thỏa mãn được (unsatisfiable)** nếu và chỉ nếu không tồn tại bất kỳ phép diễn giải nào
- Một biểu thức ϕ là **đúng (valid)** nếu và chỉ nếu mọi phép diễn giải $\mathcal{I}$ đều thỏa mãn ϕ – Chúng ta ký hiệu là: $\models \boldsymbol{\phi}$

Lượng tử logic Với mọi

- Cú pháp của lượng tử logic **Với mọi** (universal quantifier): $\forall$<*Biến*$_1$,…,*Biến*$_n$>: <*Mệnh đề*>
- Ví dụ: Tất cả (mọi) sinh viên đang ngồi học trong lớp K4 đều chăm chỉ

$\forall$x: *Ngoi_trong_lop*(x,K4) $\Rightarrow$ *Cham_chi*(x)

- Mệnh đề ($\forall$x: *P*) là đúng trong một mô hình *m*, khi và chỉ khi *P* đúng với *x* là **mỗi** (mọi) đối tượng trong mô hình đó
- Tức là, mệnh đề ($\forall$x: *P*) tương đương với sự **kết hợp (và)** của tất cả các trường hợp của *P*

Ngoi_trong_lop(Hue,K4) $\Rightarrow$ *Cham_chi*(Hue)
$\land$ *Ngoi_trong_lop*(Cuong,K4) $\Rightarrow$ *Cham_chi*(Cuong)
$\land$ *Ngoi_trong_lop*(Tuan,K4) $\Rightarrow$ *Cham_chi*(Tuan)
$\land$ …

Lượng tử logic Tồn tại

- Cú pháp của lượng tử logic Tồn tại (existential quantifier): $\exists$<*Biến*$_1$,…,*Biến*$_n$>: <*Mệnh đề*>
- Ví dụ: Tồn tại (có) sinh viên đang ngồi học trong lớp K4, và là sinh viên chăm chỉ:

$\exists$x: *Ngoi_trong_lop*(x,K4) $\wedge$ *Cham_chi*(x)

- Mệnh đề ($\exists$*x: P*) là đúng trong một mô hình *m*, khi và chỉ khi *P* là đúng với *x* là **một** đối tượng trong mô hình đó
- Tức là, mệnh đề ($\exists$*x: P*) tương đương với phép **tuyển (hoặc)** của các trường hợp của *P*

Ngoi_trong_lop(Hue,K4) $\wedge$ *Cham_chi*(Hue)
$\vee$ *Ngoi_trong_lop*(Cuong,K4) $\wedge$ *Cham_chi*(Cuong)
$\vee$ *Ngoi_trong_lop*(Tuan,K4) $\wedge$ Cham_chi(Tuan)
$\vee$ …

Các đặc điểm của các lượng từ logic

- Tính hoán vị:
  - $(\forall x\ \forall y)$ là tương đương với $(\forall y\ \forall x)$
  - $(\exists x\ \exists y)$ là tương đương với $(\exists y\ \exists x)$
- Tuy nhiên, $(\exists x\ \forall y)$ **không** tương đương với $(\forall y\ \exists x)$
  - $\exists x\ \forall y$: Yeu(x,y) - "Trên thế giới này, tồn tại (có) một người mà người đó yêu quý tất cả mọi người khác"
  - $\forall y\ \exists x$: Yeu(x,y) - "Trên thế giới này, mọi người đều được ít nhất một người khác yêu thích"
- Mỗi lượng từ logic ($\exists$ hoặc $\forall$) đều có thể được biểu diễn bằng lượng từ kia
  - $(\forall x$: Thich(x,Kem)) là tương đương với $(\neg\exists x$: $\neg$Thich(x,Kem))
  - $(\exists x$: Thich(x,BongDa)) là tương đương với $(\neg\forall x$: $\neg$Thich(x,BongDa))

Sử dụng logic vị từ

Biểu diễn các phát biểu trong ngôn ngữ tự nhiên

- "*x* là anh/chị/em của *y*" tương đương với "*x* và *y* là anh em ruột"

  $\forall$x,y: *Anh_chi_em(x,y)* $\Leftrightarrow$ *Anh_em_ruot(x,y)*

- "Mẹ của *c* là *m*" tương đương với "*m* là phụ nữ và *m* là bậc cha mẹ của *c*"

  $\forall$m,c: *Me*(c) = m $\Leftrightarrow$ *(Phu_nu*(m) $\wedge$ *Cha_me*(m,c))

- Quan hệ "anh em ruột" có tính chất đối xứng

  $\forall$x,y: *Anh_em_ruot*(x,y) $\Leftrightarrow$ *Anh_em_ruot*(y,x)

Bài tập

Chuyển đổi các phát biểu sau sang logic vị từ:

1. Tất cả các sinh viên đều chăm học
2. Có một số sinh viên
3. Một số sinh viên chăm học
4. Mỗi sinh viên đều thích một sinh viên nào đó
5. Mỗi sinh viên đều thích một sinh viên khác
6. Có một sinh viên được tất cả sinh viên khác thích

Bài tập

Chuyển đổi các phát biểu sau sang logic vị từ:

1. Tất cả các sinh viên đều chăm học
2. Có một số sinh viên
3. Một số sinh viên chăm học
4. Mỗi sinh viên đều thích một sinh viên nào đó
5. Mỗi sinh viên đều thích một sinh viên khác
6. Có một sinh viên được tất cả sinh viên khác thích

Các phép biến đổi tương đương

1. Loại bỏ dấu suy ra

$\alpha \leftrightarrow \beta \Rightarrow (\alpha \rightarrow \beta) \wedge (\beta \rightarrow \alpha)$

$\alpha \rightarrow \beta \Rightarrow \neg\alpha \vee \beta$

2. Chuyển phủ định vào trong ngoặc

$\neg(\alpha \vee \beta) \Rightarrow \neg\alpha \wedge \neg\beta$

$\neg(\alpha \wedge \beta) \Rightarrow \neg\alpha \vee \neg\beta$

$\neg\neg\alpha \Rightarrow \alpha$

$\neg\forall x, \alpha \Rightarrow \exists x, \neg\alpha$

$\neg\exists x, \alpha \Rightarrow \forall x, \neg\alpha$

3. Đặt tên các biến khác nhau

$\forall x,\ \exists y, (\neg P(x) \vee \exists x, Q(x,y)) \Rightarrow \forall x_1, \exists x_2, (\neg P(x_1) \vee \exists x_3, Q(x_3, y_2))$

Ví dụ

a. John owns a dog
$\exists x.\ D(x) \wedge O(J,x)$
$D(Fido) \wedge O(J, Fido)$

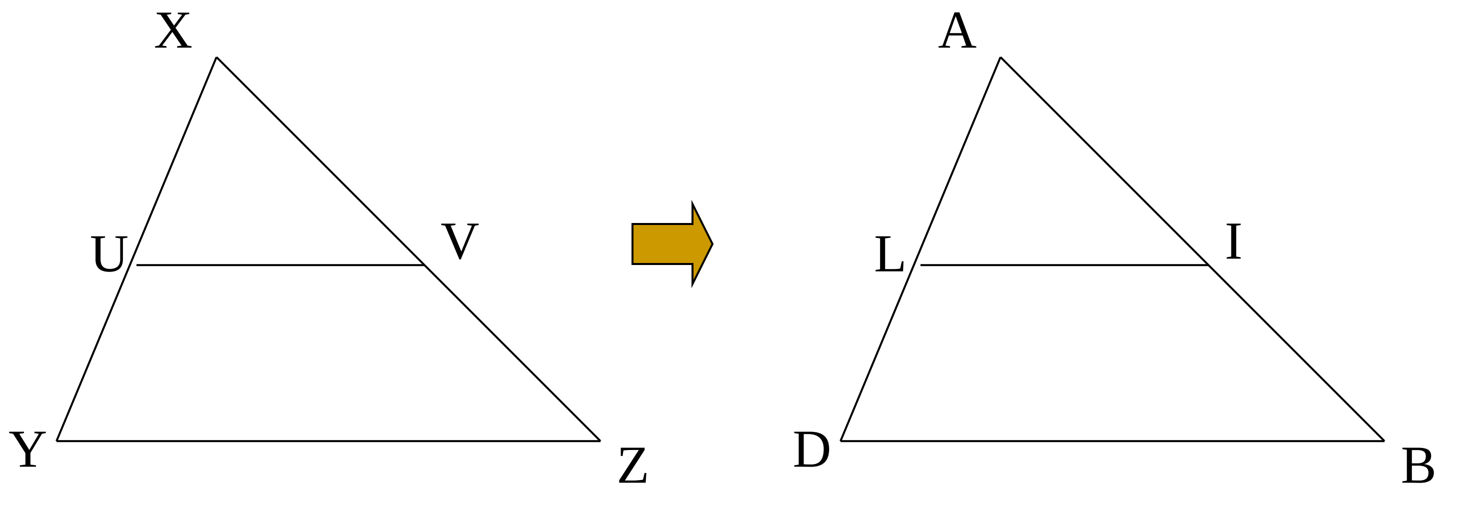
b. Anyone who owns a dog is a lover-of-animals
$\forall x.\ (\exists y.\ D(y) \wedge O(x,y)) \rightarrow L(x)$
$\forall x.\ (\neg\exists y.\ (D(y) \wedge O(x,y)) \vee L(x)$
$\forall x.\ \forall y.\ \neg(D(y) \wedge O(x,y)) \vee L(x)$
$\forall x.\ \forall y.\ \neg D(y) \vee \neg O(x,y) \vee L(x)$
$\neg D(y) \vee \neg O(x,y) \vee L(x)$

c. Lovers-of-animals do not kill animals
$\forall x.\ L(x) \rightarrow (\forall y.\ A(y) \rightarrow \neg K(x,y))$
$\forall x.\ \neg L(x) \vee (\forall y.\ A(y) \rightarrow \neg K(x,y))$
$\forall x.\ \neg L(x) \vee (\forall y.\ \neg A(y) \vee \neg K(x,y))$
$\neg L(x) \vee \neg A(y) \vee \neg K(x,y)$

Phép gán trị

VD: Định lý đường trung bình:

r_1: trđ(U,XY) $\wedge$ trđ(V,XZ) $\Rightarrow$ ss(UV,YZ)



Phép gán trị θ ={A/X,B/Z,D/Y,L/U,I/V}:

- $r_1\theta$: trđ(L,AD) $\wedge$ trđ(I,AB) $\Rightarrow$ ss(LI,DB)

Hợp giải Robinson cho logic vị từ

1. Viết mỗi GT_i, $\neg$KL trên 1 dòng
2. Đưa GT_i, $\neg$KL về dạng chuẩn CNF

$\forall x_1 \forall x_2 \ldots \forall x_n\ [p_1(\ldots) \vee \ldots \vee p_n(\ldots)] \wedge [q_1(\ldots) \vee \ldots \vee q_m(\ldots)] \quad (*)$

3. Tách mỗi dòng (*) thành các dòng con:

$\cancel{\forall x_1 \forall x_2 \ldots \forall x_n}\ [p_1(\ldots) \vee \ldots \vee p_n(\ldots)]$

$\cancel{\forall x_1 \forall x_2 \ldots \forall x_n}\ [q_1(\ldots) \vee \ldots \vee q_m(\ldots)]$

tất cả đều với $\forall$

4. Hợp giải:

u) $\neg p(x_1, x_2, \ldots, x_n) \vee q(\ldots)$

v) $p(y_1, y_2, \ldots, y_n) \vee r(\ldots)$

$\Rightarrow$ w) $q(\ldots) \vee r(\ldots)$ với phép gán trị

$$\theta = \left\{ \frac{z_1}{x_1}, \frac{z_1}{y_1}, \ldots, \frac{z_n}{x_n}, \frac{z_n}{y_n} \right\}$$

5. Vô lý xảy ra khi

i) $\neg p(x_1, x_2, \ldots, x_n)$

ii) $p(y_1, y_2, \ldots, y_n)$

với phép gán trị $\theta = \left\{ \frac{z_1}{x_1}, \frac{z_1}{y_1}, \ldots, \frac{z_n}{x_n}, \frac{z_n}{y_n} \right\}$

Ví dụ về bước 4

- Sử dụng phép gán trị nào để hợp giải

  P(a,x,b), và

  $\neg$P(y,z,z)

  Phép gán trị $\theta = \left\{ \frac{a}{y}, \frac{b}{z}, \frac{b}{x} \right\}$

  - P(a,b,b)
  - $\neg$P(a,b,b)

Ví dụ về bước 4 (tiếp)

- Sử dụng phép gán trị nào để hợp giải

  $P(a,x,x,b)$, và

  $\neg P(y,y,z,b)$

Ví dụ về bước 4 (tiếp)

- Cho các sự kiện p(a,b), p(c,d), q(d,c,c) đúng
- Cho luật

  $p(x,y) \wedge q(y,x,x) \Rightarrow r(x,y)$
- Sử dụng các phép gán trị với luật trên, hãy đưa ra các sự kiện mới đúng.

- Gợi ý:
  - Thử với $p(x,y) \equiv p(a,b)$ hoặc $p(x,y) \equiv p(c,d)$

Ví dụ về hợp giải

$$\forall x \quad P(x) \rightarrow Q(x)$$

$$\forall x \quad \neg P(x) \rightarrow R(x)$$

$$\forall x \quad Q(x) \rightarrow S(x)$$

$$\forall x \quad R(x) \rightarrow S(x)$$

Chuyển về dạng chuẩn

$$1. \neg P(x) \vee Q(x)$$

$$2. P(x) \vee R(x)$$

$$3. \neg Q(x) \vee S(x)$$

$$4. \neg R(x) \vee S(x)$$

Hợp giải 1 và 3

$$5. \neg P(x) \vee S(x)$$

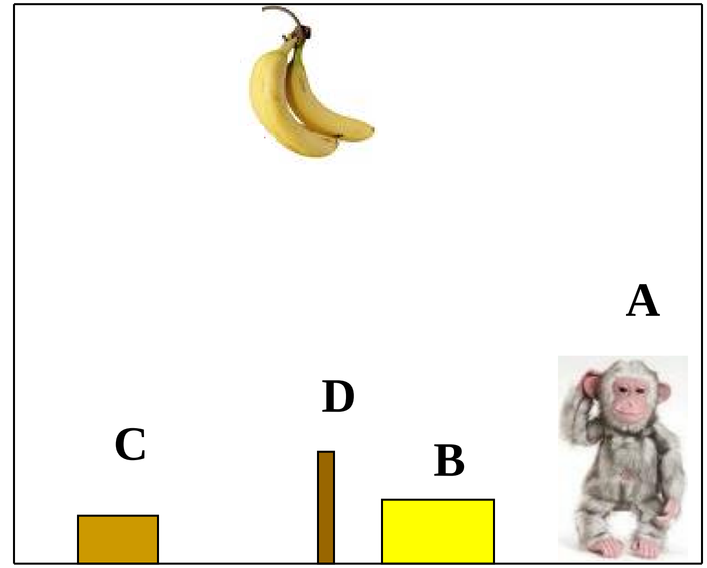
Hợp giải 2 và 5

$$6. R(x) \vee S(x)$$

Hợp giải 4 và 6

$$7. S(x)$$

Bài toán khỉ - chuối



1. tại(C,1)
2. tại(B,3)
3. tại(A,4)
4. tại(D,2)
5. tại(A,x) $\Rightarrow$ tại (A,y)
6. tại(A,x) $\wedge$ tại(O,x) $\Rightarrow$ tại(A,y) $\wedge$ tại(O,y)
7. tại(A,x) $\wedge$ tại(O,x) $\Rightarrow$ trên(A,O)
8. tại(A,x) $\wedge$ tại(O1,x) $\wedge$ tại(O2,x) $\Rightarrow$ trên(O1,O2)

KL: tại(B,2) $\wedge$ trên(C,B) $\wedge$ trên(A,C) $\wedge$ trên(D,A)

$\neg$KL: $\neg$ tại(B,2) $\vee$ $\neg$ trên(C,B) $\vee$ $\neg$ trên(A,C) $\vee$ $\neg$ trên(D,A)

Bài tập 1

- Cho tập các phát biểu:
  - John owns a dog
  - Anyone who owns a dog is a lover of animals
  - Lovers of animals do not kill animals

- Chứng minh:
  - John does not kill animals.

Bài tập 2

- Nếu xem một ai đó lừa dối người khác là kẻ bịp bợm và
- Bất kỳ ai đồng tình với kẻ bịp bợm cũng là kẻ bịp bợm.
- Trong tập thể có một người nhút nhát đồng tình với kẻ lừa dối
- thì chắc chắn có 1 tên bịp bợm tính tình nhút nhát.

Bài tập 3

1. Fred là con chó giống Collie.
2. Sam là chủ của nó.
3. Hôm nay là thứ bảy.
4. Thứ bảy trời lạnh.
5. Fred là con chó được huấn luyện.
6. Chó spaniel và (chó collie được huấn luyện) là chó tốt.
7. Nếu một con chó tốt và có ông chủ thì nó sẽ đi cùng ông chủ.
8. Nếu thứ bảy và ấm thì Sam ở công viên.
9. Nếu thứ bảy và không ấm thì Sam ở viện bảo tàng.

- Hỏi fred ở đâu? $\exists X$ loc(fred,X)

1. Fred là con chó giống Collie.
2. Sam là chủ của nó.
3. Hôm nay là thứ bảy.
4. Thứ bảy trời lạnh.
5. Fred là con chó được huấn luyện.
6. Chó spaniel và (chó collie được huấn luyện) là chó tốt.
7. Nếu một con chó tốt và có ông chủ thì nó sẽ đi cùng ông chủ.
8. Nếu thứ bảy và ấm thì Sam ở công viên.
9. Nếu thứ bảy và không ấm thì Sam ở viện bảo tàng.

- Hỏi fred ở đâu? $\exists X$ loc(fred,X)

1. collie(Fred).
2. owner(Sam, Fred).
3. day(sat).
4. cold(sat).
5. trained(Fred).
6. spaniel(X) $\vee$ (collie(X) $\wedge$ trained(X)) $\rightarrow$ gooddog(X).
7. gooddog(X) $\wedge$ owner(Y,X) $\wedge$ loc(Y,Z) $\rightarrow$ loc(X,Z).
8. day(sat) $\wedge$ $\neg$cold(sat) $\rightarrow$ loc(Sam, park).
9. day(sat) $\wedge$ cold(sat) $\rightarrow$ loc(Sam,museum).