

# ĐỊNH GIÁ TRUY VẤN ĐỐI VỚI CHƯƠNG TRÌNH DATALOG MÔ TẢ

Trương Công Tuấn<sup>1</sup>

<sup>1</sup> Trường Đại học Khoa học, Đại học Huế

tctuan\_it\_dept@yahoo.com

**TÓM TẮT:** Hiện nay lĩnh vực nghiên cứu của cơ sở dữ liệu (CSDL) suy diễn và logic mô tả đã được mở rộng theo nhiều hướng khác nhau, trong đó việc kết hợp logic mô tả với CSDL suy diễn là một hướng nghiên cứu đang được nhiều người quan tâm. Chương trình Datalog là một lớp chương trình quan trọng của CSDL suy diễn. Bài báo này tập trung vào cú pháp, ngữ nghĩa của chương trình Datalog mô tả ALC - sự kết hợp của logic mô tả ALC và chương trình Datalog và đề xuất phương pháp định giá truy vấn đối với lớp chương trình này.

**Từ khóa:** Cơ sở dữ liệu suy diễn, chương trình Datalog, logic mô tả, hợp giải SLD.

## I. MỞ ĐẦU

Trong những thập niên vừa qua, chương trình Datalog và logic mô tả là hai lĩnh vực nghiên cứu đã được nhiều người quan tâm và đã có nhiều ứng dụng hữu ích trong các lĩnh vực khác nhau.

Logic mô tả là họ các ngôn ngữ hình thức, được ứng dụng rất hiệu quả để mô tả tri thức ontology và đóng vai trò quan trọng để xây dựng Web ngữ nghĩa, với mục đích tăng khả năng liên kết giữa các tài nguyên và khả năng hiểu được thông tin trong máy tính [3]. Logic mô tả rất thích hợp cho việc biểu diễn tri thức có cấu trúc theo các lớp và các mối quan hệ giữa các lớp nhưng không thích hợp cho việc biểu diễn các truy vấn phức tạp.

Bên cạnh các nghiên cứu về logic mô tả, việc nghiên cứu cơ sở dữ liệu (CSDL) suy diễn cũng đã đạt được nhiều thành tựu quan trọng cả về lý thuyết và ứng dụng thực tiễn. CSDL suy diễn là sự mở rộng của CSDL quan hệ, có khả năng xử lý suy diễn cũng như xử lý thông tin không đầy đủ và có những ứng dụng hữu ích trong các lĩnh vực như: Trí tuệ nhân tạo, hệ hỗ trợ quyết định, phân tích tài chính,... Trong CSDL suy diễn, Datalog là một lớp chương trình quan trọng, Datalog là ngôn ngữ khai báo, ở đó các chương trình không được tạo ra từ các câu lệnh cũng như từ các hàm mà được tạo ra chủ yếu dựa trên tập các vị từ, thích hợp cho việc biểu diễn tri thức bằng các quy tắc suy diễn. Tuy nhiên, chương trình Datalog lại sử dụng mô hình dữ liệu EDB tương tự như mô hình dữ liệu quan hệ, mô hình này có nhiều hạn chế trong việc biểu diễn tri thức.

Định giá truy vấn là một trong các vấn đề quan trọng đối với chương trình Datalog. Có 2 phương pháp cơ bản để định giá truy vấn đối với chương trình Datalog: phương pháp định giá theo kiểu dưới lên và phương pháp định giá theo kiểu trên xuống. Phương pháp định giá theo kiểu dưới lên có ưu điểm là việc tìm lời giải câu truy vấn thường kết thúc. Tuy nhiên, nó không xem xét câu truy vấn trong suốt quá trình định giá, từ đó dẫn đến việc tính toán nhiều sự kiện không liên quan đến câu truy vấn. Phương pháp định giá theo kiểu trên xuống có điểm khởi đầu của việc tính toán là từ đích truy vấn và chúng sẽ không tính các sự kiện không liên quan đến câu truy vấn. Phép hợp giải SLD là một phương pháp định giá truy vấn theo kiểu trên xuống rất hiệu quả đối với chương trình Datalog, đã được dùng làm nền tảng cho hầu hết các hệ thống lập trình logic như SWI Prolog, DLV,...

Hiện nay, một hướng nghiên cứu nhằm tích hợp giữa logic mô tả với chương trình Datalog đã và đang được nhiều người quan tâm [4], [5], [6], [7], [8], [9], [10]. Ưu điểm của việc kết hợp này là đã tận dụng được những thành quả nghiên cứu trong cả hai lĩnh vực nêu trên nhằm hỗ trợ xử lý lập luận suy diễn trong Web ngữ nghĩa cũng như nhằm đưa ra một công cụ hiệu quả trong việc biểu diễn tri thức. Trong bài báo này, tôi tập trung vào phương pháp định giá truy vấn đối với chương trình Datalog mô tả ALC – là sự mở rộng của phép hợp giải SLD trên lớp chương trình Datalog được mở rộng với logic mô tả ALC. Phần còn lại của bài báo được tổ chức như sau: Trong phần II, III tôi nhắc lại về logic mô tả ALC và chương trình Datalog, phần IV trình bày về cú pháp và ngữ nghĩa chương trình Datalog mô tả ALC và phần V tập trung vào phương pháp định giá truy vấn đối với chương trình Datalog mô tả ALC. Cuối cùng, kết luận được nêu trong phần VI.

## II. LOGIC MÔ TẢ ALC

### A. Giới thiệu logic mô tả

Logic mô tả là một họ các ngôn ngữ hình thức rất thích hợp cho việc biểu diễn và suy luận tri thức trong một miền quan tâm cụ thể. Trong logic mô tả, miền quan tâm được mô tả thông qua các thuật ngữ về cá thể, khái niệm, vai trò và các tạo từ. Mỗi cá thể đại diện cho một đối tượng, mỗi khái niệm đại diện cho một tập các đối tượng có chung tính chất và mỗi vai trò đại diện cho một quan hệ hai ngôi giữa các đối tượng hoặc giữa các đối tượng và các giá trị dữ liệu. Các khái niệm phức được xây dựng từ các khái niệm, tên vai trò và tên cá thể bằng cách kết hợp với các tạo từ. Một hệ thống logic mô tả cho phép mô tả các khái niệm có liên quan với nhau và các tri thức tiềm ẩn. Các tri thức tiềm ẩn này có thể được suy luận từ những tri thức đã được biểu diễn thông qua các dịch vụ suy luận hoặc các bộ suy luận.

Logic mô tả được xây dựng dựa vào ba thành phần cơ bản gồm tập các cá thể, tập các khái niệm nguyên tố (có thể hiểu như các lớp) và tập các vai trò nguyên tố. Các logic mô tả khác nhau được đặc trưng bởi tập các *tạo từ khái niệm* và *tạo từ vai trò* mà nó được phép sử dụng để xây dựng nên các khái niệm phức, vai trò phức từ khái niệm nguyên tố và vai trò nguyên tố.

Trong phần tiếp theo sẽ giới thiệu ngôn ngữ logic mô tả ALC [3].

## B. Logic mô tả ALC

### 1. Cú pháp

Logic mô tả ALC là ngôn ngữ logic mô tả có các quy tắc cú pháp đơn giản, ALC cho phép dùng các tạo từ: phép phủ định  $\neg$ , phép giao  $\sqcap$ , phép hợp  $\sqcup$ , lượng từ hạn chế tồn tại  $\exists$ , lượng từ hạn chế với mọi  $\forall$ .

*Định nghĩa 2.1* (Khái niệm) Cho  $\Sigma_C$  là tập các khái niệm nguyên tố,  $\Sigma_R$  là tập các vai trò nguyên tố,  $\Sigma_C \cap \Sigma_R = \emptyset$ . Các *khái niệm* của logic mô tả ALC được định nghĩa đệ quy như sau:

- (i) Nếu  $A \in \Sigma_C$  thì  $A$  là một khái niệm của ALC,
- (ii) Nếu  $C, D$  là các khái niệm và  $R \in \Sigma_R$  là một vai trò thì  $\top, \perp, \neg C, C \sqcap D, C \sqcup D, \forall R.C, \exists R.C$  là các khái niệm của ALC. Các ký hiệu có ý nghĩa như trong bảng sau đây:

**Bảng 1.** Cú pháp của ngôn ngữ logic mô tả ALC

Ký hiệu	Ý nghĩa
$\top$	Khái niệm đỉnh
$\perp$	Khái niệm đáy
$\neg A$	Phủ định khái niệm nguyên tố
$C \sqcap D$	Giao hai khái niệm
$C \sqcup D$	Hợp hai khái niệm
$\forall R.C$	Hạn chế phổ dụng của khái niệm $C$ bởi vai trò $R$
$\exists R.C$	Hạn chế tồn tại của khái niệm $C$ bởi vai trò $R$

Khái niệm đỉnh ( $\top$ ) đại diện cho tất cả đối tượng, khái niệm đáy ( $\perp$ ) không đại diện cho bất kỳ đối tượng nào.

*Ví dụ 2.1* Giả sử ta có các khái niệm nguyên tố *Human* để chỉ các đối tượng là con người, *Female* để chỉ những người có giới tính nữ, *hasChild* là một vai trò nguyên tố biểu thị rằng một đối tượng này có con là đối tượng kia. Lúc đó ta có thể định nghĩa các khái niệm phức sau đây:

$Human \sqcap Female$  để biểu diễn người phụ nữ,

$Human \sqcap \neg Female$  để biểu diễn người không phải là phụ nữ,

$Human \sqcap \exists hasChild.\top$  để biểu diễn người có con,

$Human \sqcap \forall hasChild.\perp$  để biểu diễn người không con,

$Human \sqcap \forall hasChild.Female$  để biểu diễn người có tất cả các con đều là con gái.

### 2. Ngữ nghĩa

Ngữ nghĩa của logic mô tả ALC được cho theo các *thể hiện*, ta có định nghĩa sau:

*Định nghĩa 2.2* (Thể hiện) Một *thể hiện* của logic mô tả ALC, ký hiệu  $\mathcal{I}$ , là một cặp  $(\Delta^{\mathcal{I}}, \cdot^{\mathcal{I}})$ , trong đó  $\Delta^{\mathcal{I}}$  là một tập khác rỗng các cá thể, được gọi là *miền* của  $\mathcal{I}$  và  $\cdot^{\mathcal{I}}$  là một hàm thể hiện, hàm  $\cdot^{\mathcal{I}}$  ánh xạ mỗi cá thể với một phần tử của  $\Delta^{\mathcal{I}}$ , mỗi khái niệm nguyên tố  $A$  với một tập con  $A^{\mathcal{I}}$  của  $\Delta^{\mathcal{I}}$ , mỗi vai trò nguyên tố  $R$  với một quan hệ hai ngôi  $R^{\mathcal{I}} \subseteq \Delta^{\mathcal{I}} \times \Delta^{\mathcal{I}}$ . Thể hiện  $\mathcal{I}$  được mở rộng với các khái niệm phức trong ngôn ngữ logic mô tả ALC như sau:

$$\top^{\mathcal{I}} = \Delta^{\mathcal{I}}$$

$$\perp^{\mathcal{I}} = \emptyset$$

$$(\neg C)^{\mathcal{I}} = \Delta^{\mathcal{I}} \setminus C^{\mathcal{I}}$$

$$(C \sqcap D)^{\mathcal{I}} = C^{\mathcal{I}} \cap D^{\mathcal{I}}$$

$$(C \sqcup D)^{\mathcal{I}} = C^{\mathcal{I}} \cup D^{\mathcal{I}}$$

$$(\forall R.C)^{\mathcal{I}} = \{a \in \Delta^{\mathcal{I}} \mid \forall b \in \Delta^{\mathcal{I}}, \text{ nếu } (a, b) \in R^{\mathcal{I}} \text{ thì } b \in C^{\mathcal{I}}\}$$

$$(\exists R.C)^{\mathcal{I}} = \{a \in \Delta^{\mathcal{I}} \mid \exists b \in \Delta^{\mathcal{I}}, (a, b) \in R^{\mathcal{I}} \text{ và } b \in C^{\mathcal{I}}\}$$

**Ví dụ 2.2 Giả sử ta có:**

Tập các cá thể  $\Sigma_I = \{Trung, Huong, Dung\}$

Tập các khái niệm nguyên tố:  $\Sigma_C = \{Human, Doctor, Teacher, Male\}$ , trong đó:

- Human*            để chỉ các đối tượng là con người
- Doctor*           để chỉ các đối tượng là bác sĩ
- Teacher*          để chỉ các đối tượng là giáo viên
- Male*             để chỉ các đối tượng có giới tính nam.

Tập các vai trò nguyên tố  $\Sigma_R = \{childOf, marriedTo\}$ , trong đó:

- childOf*           để chỉ đối tượng này là con của đối tượng kia
- marriedTo*       để chỉ đối tượng này kết hôn với đối tượng kia

Ta định nghĩa một thể hiện  $\mathcal{I} = (\Delta^{\mathcal{I}}, \cdot^{\mathcal{I}})$ , trong đó:

$$\Delta^{\mathcal{I}} = \{A, B, C, E, F, G\}$$

$\cdot^{\mathcal{I}}$  ánh xạ mỗi cá thể thuộc  $\Sigma_I$  với một phần tử của  $\Delta^{\mathcal{I}}$  như sau:

$$Trung^{\mathcal{I}} = A, Huong^{\mathcal{I}} = B, Dung^{\mathcal{I}} = C$$

$\cdot^{\mathcal{I}}$  ánh xạ mỗi khái niệm nguyên tố thuộc  $\Sigma_C$  với một tập con của  $\Delta^{\mathcal{I}}$  như sau:

- $Human^{\mathcal{I}} = \{A, B, C, E, F, G\}$
- $Doctor^{\mathcal{I}} = \{A, B, C\}$
- $Teacher^{\mathcal{I}} = \{E, F, G\}$
- $Male^{\mathcal{I}} = \{A, C\}$

$\cdot^{\mathcal{I}}$  ánh xạ mỗi vai trò nguyên tố thuộc  $\Sigma_R$  với một tập con của  $\Delta^{\mathcal{I}} \times \Delta^{\mathcal{I}}$  như sau:

- $childOf^{\mathcal{I}} = \{(A, C), (B, C)\}$
- $marriedTo^{\mathcal{I}} = \{(A, E), (E, A)\}$

**Định nghĩa 2.3**

(i) Khái niệm  $C$  gọi là được bao hàm trong khái niệm  $D$ , ký hiệu  $C \sqsubseteq D$  nếu  $C^{\mathcal{I}} \subseteq D^{\mathcal{I}}$  với mọi thể hiện  $\mathcal{I}$ , và  $C$  tương đương  $D$ , ký hiệu  $C \equiv D$ , nếu  $C \sqsubseteq D$  và  $D \sqsubseteq C$ .

(ii) Một khẳng định dạng  $A \equiv C$  được gọi là một định nghĩa khái niệm, trong đó  $A$  là tên của một khái niệm nguyên tố và  $C$  là một khái niệm.

(iii) Một mệnh đề bao hàm là một mệnh đề có dạng  $C \sqsubseteq D$ , trong đó  $C, D$  là hai khái niệm.

**Định nghĩa 2.4**

(i) Một tập hữu hạn  $\mathcal{T}$  các mệnh đề bao hàm được gọi là TBox.

(ii) Một thể hiện  $\mathcal{I}$  là mô hình của mệnh đề bao hàm  $C \sqsubseteq D$  nếu  $C^{\mathcal{I}} \subseteq D^{\mathcal{I}}$  và  $\mathcal{I}$  là mô hình của TBox  $\mathcal{T}$  nếu nó là thỏa tất cả các khẳng định bao hàm trong  $\mathcal{T}$ .

**Định nghĩa 2.5**

(i) Một khẳng định cá thể là có dạng  $a:C$  và một khẳng định vai trò có dạng  $aRb$ , trong đó  $C$  là một khái niệm và  $R$  là một vai trò,  $a, b$  là các tên cá thể. Một khẳng định thành viên là một khẳng định cá thể hoặc một khẳng định vai trò.

(ii) Một *ABox*  $\mathcal{A}$  là một tập hữu hạn các khẳng định thành viên, nghĩa là các khẳng định có dạng  $a:C$  hoặc  $aRb$ .

(iii) Một thể hiện  $\mathcal{I}$  thỏa mãn khẳng định  $a:C$  nếu  $a^{\mathcal{I}} \in C^{\mathcal{I}}$ , và thỏa mãn  $aRb$  nếu  $(a^{\mathcal{I}}, b^{\mathcal{I}}) \in R^{\mathcal{I}}$ .  $\mathcal{I}$  được gọi là mô hình của *ABox*  $\mathcal{A}$  nếu  $\mathcal{I}$  thỏa tất cả các khẳng định trong  $\mathcal{A}$ .

*Định nghĩa 2.6* (Cơ sở tri thức) Một cơ sở tri thức  $L$  của logic mô tả ALC là một cặp TBox  $\mathcal{T}$  và ABox  $\mathcal{A}$ . Một thể hiện  $\mathcal{I}$  là mô hình của  $L$  nếu  $\mathcal{I}$  là mô hình của cả  $\mathcal{T}$  và  $\mathcal{A}$ .

### III. CHƯƠNG TRÌNH DATALOG

Phần này chỉ trình bày một số khái niệm cơ sở của chương trình Datalog. Chi tiết đầy đủ về logic bậc nhất và chương trình Datalog có thể xem trong [1], [2].

#### A. Cú pháp

*Định nghĩa 3.1*

(i) Hằng hoặc biến là *hạng thức*.

(ii) *Nguyên tố* là biểu thức có dạng  $p(t_1, \dots, t_n)$ , trong đó  $p$  là vị từ  $n$  ngôi,  $t_1, \dots, t_n$  là các hạng thức.

*Định nghĩa 3.2* Một chương trình Datalog bao gồm một tập hữu hạn các mệnh đề xác định có dạng:

$$p \leftarrow q_1, \dots, q_n \quad (n \geq 0) \quad (1)$$

trong đó  $p, q_i$  là các nguyên tố có đối là hằng hoặc biến,  $p$  được gọi là đầu và  $q_1, \dots, q_n$  được gọi là thân của mệnh đề. Dấu phẩy (,) trong mệnh đề (1) thay thế cho phép hội ( $\wedge$ ).

Khi  $n = 0$ , mệnh đề (1) trở thành  $p \leftarrow$  và được gọi là mệnh đề đơn vị, ký hiệu " $\leftarrow$ " có thể bỏ qua. Ngữ nghĩa của mệnh đề đơn vị  $p \leftarrow$  là với mọi phép thay thế các biến trong  $p$  bởi các hằng thì sẽ làm cho  $p$  đúng. Mệnh đề đơn vị có đối là hằng còn được gọi là *sự kiện*. Một tập các sự kiện của chương trình Datalog được gọi là cơ sở dữ liệu EDB.

#### B. Ngữ nghĩa

*Định nghĩa 3.2* (Vũ trụ Herbrand/Cơ sở Herbrand/Thể hiện Herbrand) Cho  $P$  là chương trình Datalog. *Vũ trụ Herbrand* của  $P$ , ký hiệu  $U_P$  là tập tất cả các hằng của  $P$ . *Cơ sở Herbrand* của  $P$ , ký hiệu  $B_P$  là tập tất cả các nguyên tố nền của  $P$ . *Thể hiện Herbrand* của  $P$  là một tập con của cơ sở Herbrand  $B_P$  của  $P$ .

*Định nghĩa 3.3* *Thể hiện Herbrand* của  $P$  là một tập con của cơ sở Herbrand  $B_P$  của  $P$ . Một thể hiện Herbrand  $I$  được gọi là *mô hình Herbrand* (hoặc đơn giản là mô hình) của  $P$  nếu mọi mệnh đề của  $P$  là đúng theo thể hiện  $I$ .

*Định nghĩa 3.4* Ngữ nghĩa của chương trình Datalog  $P$  là mô hình Herbrand nhỏ nhất của  $P$ .

Mô hình Herbrand nhỏ nhất của  $P$  có thể tính được nhờ việc lặp toán tử  $T_P$ ,  $T_P$  được định nghĩa như sau:

*Định nghĩa 3.5* (Toán tử hệ quả trực tiếp) Cho  $P$  là chương trình Datalog. Hiện hành nền của một mệnh đề  $r$  trong  $P$  là mệnh đề nhận được từ  $r$  bằng cách thay các biến trong  $r$  bởi các hằng trong  $P$ . Gọi  $ground(P)$  là tập các hiện hành nền của các mệnh đề trong  $P$ . Ký hiệu  $2^{B_P}$  là tập các tập con của  $B_P$ . Toán tử hệ quả trực tiếp của  $P$  là ánh xạ  $T_P: 2^{B_P} \rightarrow 2^{B_P}$  được định nghĩa như sau: với mỗi  $I \in 2^{B_P}$ ,

$$T_P(I) = \{A \in B_P \mid \exists A \leftarrow A_1, \dots, A_n \in ground(P) \text{ sao cho } \{A_1, \dots, A_n\} \subseteq I\}$$

*Định lý 3.1* [1] Cho  $P$  là chương trình Datalog. Lúc đó toán tử  $T_P$  đơn điệu và điểm bất động nhỏ nhất của  $T_P$  là mô hình Herbrand nhỏ nhất  $M_P$  của  $P$ .

*Mệnh đề 3.1* [1] Cho  $P$  là chương trình Datalog. Mô hình Herbrand nhỏ nhất  $M_P$  của  $P$  là giới hạn của dãy  $T_P \uparrow n$ ,  $n \in \mathbb{N}$ , trong đó  $T_P \uparrow 0 = \emptyset$ ,  $T_P \uparrow (i+1) = T_P(T_P \uparrow i)$ .

Ví dụ 3.1 Xem chương trình Datalog  $P$  sau:

$$\begin{aligned} path(X, Y) &\leftarrow edge(X, Y) \\ path(X, Z) &\leftarrow edge(X, Y), path(Y, Z) \end{aligned}$$

giả sử cơ sở dữ liệu EDB là tập  $\{edge(1,2), edge(2,3), edge(3,4), edge(4,5)\}$ .

Dùng toán tử  $T_P$ , mô hình Herbrand nhỏ nhất của  $P$  tính được như sau:

$$M_P = \{edge(1,2), edge(2,3), edge(3,4), edge(4,5), path(1,2), path(2,3), path(3,4), path(4,5), path(1,3), path(2,4), path(3,5), path(1,4), path(2,5), path(1,5)\}.$$

#### IV. CHƯƠNG TRÌNH DATALOG MÔ TẢ ALC

Trong phần này sẽ trình bày cú pháp và ngữ nghĩa của chương trình Datalog mô tả ALC  $KB = (L, P)$ . Một cách không hình thức,  $KB$  bao gồm một cơ sở tri thức ALC  $L$  và một chương trình Datalog mở rộng  $P$ . Đối với chương trình Datalog mở rộng  $P$ , mỗi khái niệm, vai trò của logic mô tả ALC lần lượt có thể xem là vị từ 1-ngôi và vị từ 2-ngôi trong  $P$ , được gọi là vị từ logic mô tả. Các vị từ chỉ xuất hiện trong chương trình Datalog mở rộng và không xuất hiện trong  $L$  được gọi là vị từ Datalog. Trong chương trình Datalog mở rộng  $P$ , các vị từ có thể là vị từ logic mô tả hoặc là vị từ Datalog. Nguyên tố logic mô tả là nguyên tố mà vị từ của nó là vị từ logic mô tả.

##### A. Cú pháp

*Định nghĩa 4.1* (Mệnh đề Datalog mở rộng) *Mệnh đề Datalog mở rộng* là công thức có dạng:

$$p \leftarrow q_1, \dots, q_n \quad (n \geq 0) \quad (1)$$

trong đó  $p$  là nguyên tố,  $q_i$  là các nguyên tố hoặc nguyên tố logic mô tả có đối là hằng hoặc biến.

*Ví dụ 4.1* Xét một cơ sở tri thức có các khái niệm *Sinhvien*, *Hocphan*, *Chude* lần lượt chỉ các đối tượng sinh viên, học phần, chủ đề của học phần. Xem mệnh đề Datalog mở rộng sau đây:

$$dahoc(X,Z) \leftarrow thidau(X,Y), monhoc(Y,Z), X : Sinhvien, Y : Hocphan, Z : Chude$$

trong đó nguyên tố *thidau(X,Y)* có nghĩa sinh viên  $X$  thi đậu học phần  $Y$  và *monhoc(Y,Z)* có nghĩa  $Z$  là chủ đề của học phần  $Y$ . Các nguyên tố logic mô tả  $X : Sinhvien, Y : Hocphan, Z : Chude$  trong mệnh đề này là các khẳng định của logic mô tả. Mệnh đề Datalog mở rộng này định nghĩa vị từ *dahoc*, nguyên tố *dahoc(X,Z)* có nghĩa sinh viên  $X$  đã học xong chủ đề  $Z$  nếu sinh viên  $X$  thi đậu kỳ thi của học phần  $Y$  và  $Z$  là chủ đề của học phần  $Y$ .

*Định nghĩa 4.2* (Chương trình Datalog mô tả ALC) Một *chương trình Datalog mô tả ALC*  $KB = (L, P)$  bao gồm một cơ sở tri thức ALC  $L$  và một tập hữu hạn  $P$  các mệnh đề Datalog mở rộng.

##### B. Ngữ nghĩa

Các khái niệm về vũ trụ/thể hiện/mô hình Herbrand của chương trình Datalog mô tả ALC  $KB = (L, P)$  được mở rộng tương tự như trong trường hợp của chương trình Datalog. Một thể hiện  $I$  của  $KB$  được gọi là mô hình của  $KB$  nếu nó thỏa mãn các mệnh đề của  $P$  và cơ sở tri thức  $L$ . Ngữ nghĩa của  $KB$  được mở rộng từ ngữ nghĩa của chương trình Datalog.

*Định nghĩa 4.3* Ngữ nghĩa của chương trình Datalog mô tả ALC  $KB = (L, P)$  là mô hình Herbrand nhỏ nhất  $M_{KB}$  của  $KB$ .

Tương tự như chương trình Datalog,  $M_{KB}$  có thể tính được bằng toán tử  $T_{KB}$  được định nghĩa như sau:

*Định nghĩa 4.4* (Toán tử hệ quả trực tiếp đối với  $KB$ ) Cho  $KB = (L, P)$  là chương trình Datalog mô tả ALC. Toán tử hệ quả trực tiếp đối với  $KB$  là ánh xạ  $T_{KB}: 2^{B_{KB}} \rightarrow 2^{B_{KB}}$  được định nghĩa như sau: Với mỗi  $I \in 2^{B_{KB}}$ ,

$$T_{KB}(I) = \{A \in B_P \mid \exists A \leftarrow A_1, \dots, A_n \in ground(P) \text{ sao cho } A_i \in I \text{ nếu } A_i \text{ là nguyên tố và } A_i \text{ thỏa } L \text{ nếu } A_i \text{ là nguyên tố logic mô tả, } i \in \{1, \dots, n\}\}$$

*Định lý 4.5* Cho  $KB = (L, P)$  là chương trình Datalog mô tả ALC. Lúc đó toán tử  $T_{KB}$  đơn điệu và điểm bất động nhỏ nhất của  $T_{KB}$  là mô hình Herbrand nhỏ nhất  $M_{KB}$  của  $KB$ .

*Chứng minh:*  $T_{KB}$  là mở rộng trực tiếp của  $T_P$ , vì vậy tính đúng đắn của định lý được suy ra trực tiếp từ định lý 3.1.

*Ví dụ 4.2* Xét chương trình Datalog mô tả ALC  $KB = (L, P)$ , trong đó:

Cơ sở tri thức ALC  $L$  bao gồm các khái niệm:  $GV =$  Giảng viên,  $GVQL =$  Giảng viên làm công tác quản lý,  $CBGD =$  Cán bộ giảng dạy,  $DAY =$  Giảng dạy,  $SV =$  Sinh viên,  $CHUDE =$  Chủ đề,  $HP =$  Học phần,  $HPNC =$  Học phần nâng cao,  $HPCB =$  Học phần cơ bản. Giả sử:

- 1) Giảng viên là cán bộ giảng dạy.
- 2) Giảng viên làm công tác quản lý là Giảng viên và không dạy bất kỳ học phần nào.
- 3) Tập các học phần được phân thành các học phần cơ bản và các học phần nâng cao.
- 4) Hung là Giảng viên và Hung dạy học phần *tmt*, Mai là Giảng viên và dạy các học phần nâng cao, Hoa là sinh viên, *tmt* là học phần nâng cao, *cstt* và *ltlg* là các chủ đề.

Lúc đó cơ sở tri thức ALC  $L$  được mô tả như sau:

$$GV \sqsubseteq CBGD$$

$$GVQL = GV \sqcap \neg \forall DAY.HP$$

$$HPNC \sqcup HPCB = HP$$

$$HPNC \sqcap HPCB \sqsubseteq \perp$$

$$Hung:GV, Hung DAY ttnt,$$

$$Mai:GV \sqcap \forall DAY. HPNC,$$

$$Hoa: SV, ttnt: HPNC,$$

$$cstt: CHUDE, ltlg: CHUDE$$

Chương trình Datalog mở rộng  $P$  gồm các mệnh đề sau:

$$dahoc(X, Z) \leftarrow thidau(X, Y), monhoc(Y, Z), X: SV, Y: HP(Y), Z: CHUDE$$

$$huongdan(X, Y) \leftarrow dahoc(X, Z), chuyengia(Y, Z), X: SV, Z: CHUDE, Y: CBGD \sqcap \exists DAY. HPNC$$

$$huongdan(X, Y) \leftarrow X: SV, Y: GVQL$$

$$thidau(Hoa, ttnt), monhoc(ttnt, cstt), monhoc(ttnt, ltlg),$$

$$chuyengia(Hung, cstt), chuyengia(Mai, ltlg)$$

Các nguyên tố trong  $P$  có ý nghĩa như sau:

$huongdan(X, Y)$ : Sinh viên  $X$  có thể được Giảng viên  $Y$  hướng dẫn,

$dahoc(X, Y)$ : Sinh viên  $X$  đã học chủ đề  $Y$  trong chương trình học,

$chuyengia(X, Y)$ : Giảng viên  $X$  là một chuyên gia về chủ đề  $Y$ ,

$thidau(X, Y)$ : Sinh viên  $X$  đã thi đậu kỳ thi của học phần  $Y$ ,

$monhoc(X, Y)$ : Chủ đề  $Y$  được nghiên cứu trong học phần  $X$ .

Ngữ nghĩa của chương trình Datalog mô tả ALC  $KB$  được tìm bởi toán tử  $T_{KB}$  là:

$$M_{KB} = \{ thidau(Hoa, ttnt), monhoc(ttnt, cstt), monhoc(ttnt, ltlg), chuyengia(Hung, cstt), chuyengia(Mai, ltlg), \\ dahoc(Hoa, cstt), huongdan(Hoa, Hung) \}$$

## V. ĐỊNH GIÁ TRUY VẤN ĐỐI VỚI CHƯƠNG TRÌNH DATALOG MÔ TẢ ALC

Trong phần này sẽ trình bày phương pháp định giá truy vấn đối với chương trình Datalog mô tả ALC, phương pháp này là sự mở rộng của phép hợp giải SLD để định giá truy vấn đối với chương trình Datalog [2].

*Định nghĩa 5.1* (Đích)

1. Một *đích* (hoặc *truy vấn*)  $G$  đối với chương trình Datalog mô tả  $KB = (L, P)$  là công thức có dạng:

$$false \leftarrow q_1, \dots, q_m \quad (1)$$

trong đó  $n \geq 0$ , mỗi  $q_i$  là một nguyên tố hoặc một ràng buộc khẳng định cá thể của logic mô tả. (1) có thể viết đơn giản:

$$\leftarrow q_1, \dots, q_m$$

2. Một *câu trả lời* đối với  $KB \cup G$  là một phép thế  $\theta$  đối với các biến của  $G$ . Ta nói  $\theta$  là *câu trả lời đúng* đối với  $KB$  nếu  $G\theta$  là hệ quả logic của  $KB$ .

*Định nghĩa 5.2* (Hợp giải) Cho  $G$  là đích có dạng:

$$\leftarrow q_1, \dots, q_i, \dots, q_m$$

và  $C$  là mệnh đề Datalog mở rộng có dạng:

$$p \leftarrow b_1, \dots, b_n$$

Lúc đó việc *hợp giải* đích  $G$  với mệnh đề  $C$  sẽ nhận được đích  $G'$  được dẫn xuất từ  $G$  và  $C$  bằng cách dùng phép thế  $\theta$  nếu các điều kiện sau đây thỏa mãn:

(i)  $q_i \in \{q_1, \dots, q_m\}$  là một nguyên tố, gọi là *nguyên tố chọn* trong đích  $G$ ,

(ii)  $\theta$  là phép thế tổng quát nhất sao cho  $p\theta = q_i\theta$ ,

(iii)  $G'$  là đích  $\leftarrow (q_1, \dots, q_{i-1}, b_1, \dots, b_n, q_{i+1}, \dots, q_m)\theta$ .

*Định nghĩa 5.3* (Dẫn xuất) Cho  $(L, P)$  là chương trình Datalog mô tả ALC và  $G_0$  là một đích. Một *dẫn xuất* đối với  $(L, P) \cup G$  bao gồm:

1. Một dãy các đích (có thể hữu hạn hoặc vô hạn):  $G_0, \dots, G_n, \dots$
2. Một dãy các mệnh đề Datalog mở rộng của  $P$ :  $C_0, \dots, C_n, \dots$
3. Một dãy các phép thế  $\theta_1, \dots, \theta_n, \dots$  sao cho với mỗi  $G_{i+1}$  là hợp giải của  $G_i$  và  $C_i$  bằng cách dùng  $\theta_{i+1}$ .

Một dẫn xuất có thể kết thúc với đích cuối cùng có dạng  $c_1, \dots, c_k$  trong đó  $c_i$  là ràng buộc khẳng định cá thể của logic mô tả, được gọi là *mệnh đề rỗng có ràng buộc* hoặc đơn giản là mệnh đề rỗng.

*Định nghĩa 5.4* (Dẫn xuất thành công, dẫn xuất thất bại)

(i) Một *dẫn xuất thành công* khi đích kết thúc bởi mệnh đề rỗng.

(ii) Một *dẫn xuất thất bại* khi đích kết thúc không phải là mệnh đề rỗng, nguyên tố chọn trong đích này không thể hợp nhất với đầu của mọi mệnh đề trong chương trình.

*Định nghĩa 5.5* (Phép thế tính câu trả lời) Cho  $(L, P)$  là chương trình Datalog mô tả ALC và  $G$  là một đích. Một *phép thế để tính câu trả lời* của dẫn xuất thành công đối với  $(L, P) \cup G$  là một phép thế  $\theta$  nhận được bằng cách thu hẹp phép hợp  $\theta_1, \theta_2, \dots, \theta_n$  chỉ trên các biến của  $G$ , trong đó  $\theta_1, \theta_2, \dots, \theta_n$  là dãy các phép thế được dùng trong dẫn xuất thành công đối với  $(L, P) \cup G$ .

*Ví dụ 5.1* Xét chương trình Datalog mô tả  $KB = (L, P)$  trong ví dụ 4.2 và đích  $G_0: \leftarrow \text{huongdan}(\text{Hoa}, \text{Hung})$  với ý nghĩa là “Hoa không được giảng viên Hung hướng dẫn”.

Với đích  $G_0$ , tìm được mệnh đề trong  $P$  có ký hiệu vị từ đầu là *huongdan*:

$\text{huongdan}(X, Y) \leftarrow \text{dahoc}(X, Z), \text{chuyengia}(Y, Z), X: \text{SV}, Z: \text{CHUDE}, Y: \text{CBGD} \sqcap \exists \text{DAY.HPNC}$

Hợp nhất  $\text{huongdan}(\text{Hoa}, \text{Hung})$  và  $\text{huongdan}(X, Y)$  nhận được phép thế  $\theta_1 = \{X/\text{Hoa}, Y/\text{Hung}\}$ . Bước dẫn xuất đầu tiên sinh ra đích  $G_1$ :

$G_1: \leftarrow \text{dahoc}(\text{Hoa}, Z), \text{chuyengia}(\text{Hung}, Z), \text{Hoa}: \text{SV}, Z: \text{CHUDE}, \text{Hung}: \text{CBGD} \sqcap \exists \text{DAY.HPNC}$

Với đích  $G_1$ , nguyên tố chọn  $\text{dahoc}(\text{Hoa}, Z)$ , lúc này ta tìm được mệnh đề trong  $P$  là:

$\text{dahoc}(X, Z) \leftarrow \text{thidau}(X, Y), \text{monhoc}(Y, Z), X: \text{SV}, Y: \text{HP}, Z: \text{CHUDE}$ . Trong bước hợp giải thứ hai, nhận được  $\theta_2 = \{X/\text{Hoa}\}$ . Bước dẫn xuất thứ hai sinh ra đích  $G_2$ :

$G_2: \leftarrow \text{thidau}(\text{Hoa}, Y), \text{monhoc}(Y, Z), \text{chuyengia}(\text{Hung}, Z), \text{Hoa}: \text{SV}, Y: \text{HP}, Z: \text{CHUDE}, \text{Hung}: \text{CBGD} \sqcap \exists \text{DAY.HPNC}$

Với đích  $G_2$ , nguyên tố chọn  $\text{thidau}(\text{Hoa}, Y)$ , ta tìm được sự kiện:  $\text{thidau}(\text{Hoa}, \text{tnt})$ . Phép thế nhận được ở bước này là  $\theta_3 = \{Y/\text{tnt}\}$ . Bước dẫn xuất thứ ba sinh ra đích  $G_3$ :

$G_3: \leftarrow \text{monhoc}(\text{tnt}, Z), \text{chuyengia}(\text{Hung}, Z), \text{Hoa}: \text{SV}, \text{tnt}: \text{HP}, Z: \text{CHUDE}, \text{Hung}: \text{CBGD} \sqcap \exists \text{DAY.HPNC}$

Với đích  $G_3$ , nguyên tố chọn  $\text{monhoc}(\text{tnt}, Z)$ , tìm được sự kiện  $\text{monhoc}(\text{tnt}, \text{cstt})$ . Phép thế nhận được ở bước này là  $\theta_4 = \{Z/\text{cstt}\}$ . Bước dẫn xuất thứ tư sinh ra đích  $G_4$ :

$G_4: \leftarrow \text{chuyengia}(\text{Hung}, \text{cstt}), \text{Hoa}: \text{SV}, \text{tnt}: \text{HP}, \text{cstt}: \text{CD}, \text{Hung}: \text{CBGD} \sqcap \exists \text{DAY.HPNC}$

Với đích  $G_4$ , nguyên tố chọn  $\text{chuyengia}(\text{Hung}, \text{cstt})$  ta tìm được mệnh đề đơn vị  $\text{chuyengia}(\text{Hung}, \text{cstt})$  để hợp giải, vì vậy ta nhận được mệnh đề rỗng:

$\leftarrow \text{Hoa}: \text{SV}, \text{tnt}: \text{HP}, \text{cstt}: \text{CHUDE}, \text{Hung}: \text{CBGD} \sqcap \exists \text{DAY.HPNC}$

Như vậy dẫn xuất đối với  $(L, P) \cup G_0$  là thành công và nhận được câu trả lời đúng đối với đích

$G_0: \leftarrow \text{huongdan}(\text{Hoa}, \text{Hung})$ .

nghĩa là sinh viên Hoa được hướng dẫn bởi giảng viên Hung.

*Định lý 5.6* Phép hợp giải để định giá đích  $G$  đối với chương trình Datalog mô tả ALC  $KB = (L, P)$  là đúng đắn, nghĩa là mọi phép thế để tính câu trả lời đối với  $KB \cup G$  là câu trả lời đúng.

*Chứng minh.* Gọi  $G$  là đích  $\leftarrow q_1, \dots, q_m$  và  $\theta_1, \dots, \theta_n$  là dãy các phép thế được dùng trong hợp giải của  $KB \cup \{G\}$ . Ta chứng minh  $(q_1, \dots, q_m)\theta_1 \dots \theta_n$  là hệ quả logic của  $KB$  bằng cách dùng phép quy nạp theo chiều dài của phép hợp giải.

Với  $n = 1$ , đích  $G$  có dạng  $\leftarrow q_1$ , chương trình  $P$  có mệnh đề đơn vị có dạng  $A \leftarrow$  và  $q_1 \theta_1 \equiv A \theta_1$ . Vì  $A \theta_1 \leftarrow$  là một thể hiện của mệnh đề đơn vị của  $P$ , điều này suy ra  $q_1 \theta_1$  là hệ quả logic của  $KB$ .

Tiếp theo, ta giả sử kết luận của định lý là đúng đối với phép thế để tính câu trả lời của hợp giải  $KB \cup \{G\}$  có chiều dài là  $n-1$ . Lúc này dẫn xuất thành công nhận được là:  $G_0 \xrightarrow{C_0} G_1 \dots G_{n-1} \xrightarrow{C_{n-1}} \text{mệnh đề rỗng}$ .

trong đó  $G_0$  là đích đầu tiên  $\leftarrow q_1, \dots, q_m$ . Giả sử  $q_j$  là nguyên tố chọn trong bước dẫn xuất đầu tiên của đích  $G_0$  và  $C_0$  là mệnh đề trong  $P: B_0 \leftarrow B_1, \dots, B_k$  ( $k \geq 0$ ). Lúc đó  $q_j \theta_1 = B_0 \theta_1$  và đích  $G_1$  có dạng:

$$\leftarrow (q_1, \dots, q_{j-1}, B_1, \dots, B_k \wedge q_{j+1}, \dots, q_m) \theta_1 \quad (1)$$

Theo giả thiết quy nạp thì công thức:  $(q_1, \dots, q_{j-1}, B_1, \dots, B_k, q_{j+1}, \dots, q_m) \theta_1 \dots \theta_n$  là hệ quả logic của  $KB$ . Do đó:

$$(B_1, \dots, B_k) \theta_1 \dots \theta_n \quad (2)$$

cũng là hệ quả logic của  $KB$ . Từ (1) ta cũng có:  $(q_1, \dots, q_{j-1}, q_{j+1}, \dots, q_m) \theta_1 \dots \theta_n$  (3)

là hệ quả logic của  $KB$ . Lúc đó từ (2) và từ:  $(B_0 \leftarrow B_1, \dots, B_k) \theta_1 \dots \theta_n$  là hệ quả logic của chương trình nên:

$$B_0 \theta_1 \dots \theta_n \quad (4)$$

là hệ quả logic của  $KB$ . Vậy từ (3) và (4) ta có:  $(q_1, \dots, q_{j-1}, B_0, q_{j+1}, \dots, q_m) \theta_1 \dots \theta_n$  là hệ quả logic của  $KB$ . Do  $\theta_1$  là phép thế của  $B_0$  và  $q_j$  nên  $B_0$  có thể thay bởi  $q_j$  trong (4). Vậy  $(q_1, \dots, q_m) \theta_1 \dots \theta_n$  là hệ quả logic của  $KB$ .

## VI. KẾT LUẬN

Bài báo đã tập trung trình bày cú pháp và ngữ nghĩa của chương trình Datalog mô tả ALC – một sự tích hợp của logic mô tả ALC và chương trình Datalog. Ngữ nghĩa của chương trình Datalog mô tả ALC là sự mở rộng tự nhiên của ngữ nghĩa chương trình Datalog. Ngoài ra, bài báo cũng đưa ra một cơ chế suy diễn để định giá truy vấn đối với chương trình Datalog mô tả ALC. Phương pháp này là sự mở rộng tự nhiên của phép hợp giải SLD để định giá truy vấn đối với chương trình Datalog. Đối với trường hợp chương trình Datalog mô tả ALC có chứa phủ định trong các mệnh đề, ngữ nghĩa và phương pháp định giá truy vấn sẽ được tiếp tục được xem xét. Trong khuôn khổ của bài báo tôi chưa trình bày đến vấn đề này.

## TÀI LIỆU THAM KHẢO

- [1] Abiteboul S., Hull R., Vianu V., Foundation of Databases, Addison Wesley Publishing, MA, 1995.
- [2] Apt K. R., Logic Programming, Elsevier Science Publishers, 1990.
- [3] Baader F., D. Calvanese, D. McGuinness, D. Nardi, and P. F. Patel-Schneider, The Description Logic Handbook: Theory, Implementation, and Applications, Cambridge University Press, 2003.
- [4] Cuenca Grau, B., Motik, B., Stoilos, G., Horrocks, I.: “Computing Datalog rewritings beyond Horn ontologies”. In: IJCAI. pp. 832-838, 2013.
- [5] Eiter, T., Ianni, G., Lukasiewicz, T., Schindlauer, R., Tompits, H., “Combining answer set programming with description logics for the semantic web”. Artificial Intelligence 172(12-13), 1495-1539, 2008.
- [6] Heymans S., T. Eiter, G. Xiao, “Tractable reasoning with DL-programs over Datalog-rewritable description logics”, in: H. Coelho, R. Studer, M. Wooldridge (Eds.), ECAI 2010, Proc. 19th European Conference on Artificial Intelligence, vol. 215 of Frontiers in Artificial Intelligence and Applications, IOS Press, 35–40, 2010.
- [7] Kaminski, M., Nenov, Y., Cuenca Grau, B.: Datalog rewritability of disjunctive Datalog programs and its applications to ontology reasoning. in: C. E. Brodley, P. Stone (Eds.), Proc. 28th AAAI Conference on Artificial Intelligence, AAAI, 1077–1083, 2014.
- [8] Mei, J., Lin, Z., Boley, H., Li, J., & Bhavsar, V. C., The DatalogDL: Combination of deduction rules and description logics, Computational Intelligence, 23(3), pp 356-372, 2007.
- [9] Shen, Y. D., “Well-supported semantics for description logic programs”. In Proceedings of the 22nd International Joint Conference on Artificial Intelligence (IJCAI-11), 2011.
- [10] Thomas Lukasiewicz, “A novel combination of answer set programming with description logics for the semantic web”. IEEE Transactions on Knowledge and Data Engineering 22(11), 1577-1592, 2010.
- [11] Xiao G., Heymans S., and Eiter T., “DReW: a reasoner for Datalog-rewritable description logics and dl-programs”. In Informal Proc. 1st Int’l Workshop on Business Models, Business Rules and Ontologies, 2010.

## QUERY EVALUATION FOR DESCRIPTION DATALOG PROGRAM

Truong Cong Tuan

**ABSTRACT:** Currently, the field of deductive database research and description logic has been extended in many different directions, where the combination of description logic with the deductive database has been of great interest. The Datalog program is an important class of the deductive database. This paper focuses on the syntax and semantics of the description Datalog program ALC - a combination of the ALC description logic and the Datalog program, and proposes a query evaluation method for this program class.