FCA VÀ TIẾN HÓA ONTOLOGY TỰ ĐỘNG

ThS. Nguyễn Thanh Long[(1)], ThS. Huỳnh Nhứt Phát[(2)]

[(1)]Trường Đại học Công nghiệp Thực phẩm TP.HCM,[(2)]Trường Đại học Nguyễn Tất Thành

Ngày gửi bài: 06/5/2016 Ngày chấp nhận đăng: 11/6/2016

TÓM TẮT

Bài báo này giới thiệu những ý tưởng chính về phân tích khái niệm hình thức một cách cơ bản mà không sử dụng các định nghĩa toán học hình thức. Người đọc có thể hiểu được kỹ thuật quan trọng về việc biểu diễn tri thức đồ họa, cụ thể là sơ đồ khái niệm Lưới. Ngoài ra, bài báo còn trình bày một số ứng dụng về phân tích khái niệm hình thức.

Từ khóa: Phân tích khái niệm hình thức, khái niệm Lưới, các hình thức theo ngữ cảnh

FCA AND ONTOLOGY AUTOMATION EVOLUTION

ABSTRACT

This paper introduces the main idea about a fundamental way of Formal Concepts Analysis without using the formal mathematical definitions. The reader can understand important techniques of graphical knowledge representation, namely conceptual diagram. In addition, the paper also presents some applications of Formal Concepts Analysis.

Key words: Concept lattices, Contextual forms, Formal Concepts Analysis.

1. TỔNG QUAN VỀ FCA (FORMAL CONCEPTS ANALYSIS)

1.1. Giới thiệu

FCA là một phương pháp phân tích dữ liệu được phát triển phổ biến thông qua các miền khác nhau. FCA mô tả mối quan hệ giữa tập các đối tượng và tập các thuộc tính cụ thể. Dữ liệu này thường xuất hiện trong nhiều lĩnh vực hoạt động của con người.

FCA đưa ra hai loại dữ liệu, dữ liệu vào và dữ liệu ra. Trước tiên, ta tìm hiểu là mạng khái niệm. Mạng khái niệm là tập các khái niệm hình thức với dữ liệu vào được phân cấp theo thứ tự của mối quan hệ subconcept-superconcept. Các khái niệm hình thức là các nhóm cụ thể đại diện cho các khái niệm theo quy luật tự nhiên, chẳng hạn như "sinh vật sống trong nước", "xe hơi với các hệ thống điều khiển bánh xe", "một số có thể chia hết cho 3 và 4",... Kế tiếp, dữ liệu ra của FCA là tập các thuộc tính liên quan. Thuộc tính liên quan mô tả sự phụ thuộc cụ thể mang tính hợp lệ về dữ liệu như "mọi số chia hết cho 3 và 4 thì chia hết cho 6", "mọi cán bộ với độ tuổi trên 60 thì phải nghỉ hưu",...

Tính năng phân biệt của FCA là sự tích hợp của ba thành phần về quá trình xử lý khái niệm dữ liệu. Cụ thể là, việc phát hiện và suy luận với các khái niệm về dữ liệu, việc phát hiện và suy luận với các phụ thuộc về dữ liệu, và trực quan hoá dữ liệu. Các khái niệm, các phụ thuộc, và khả năng có thể kết chúng lại thành khối.

Sự tích hợp của các thành phần này làm cho FCA trở thành một công cụ đủ mạnh có thể ứng dụng vào các vấn đề khác. Ví dụ như tổ chức phân cấp của các kết quả tìm kiếm trang web thành các khái niệm dựa trên các chủ đề phổ biến, phân tích dữ liệu biểu hiện gen, phục hồi thông tin, phân tích và hiểu được mã nguồn phần mềm, gỡ lỗi, khai thác dữ liệu, kỹ thuật thiết kế phần mềm, ứng dụng Internet bao gồm phân tích và tổ chức các văn bản, soạn thảo e-mail, phân loại chú thích, và các dự án phân tích dữ liệu.

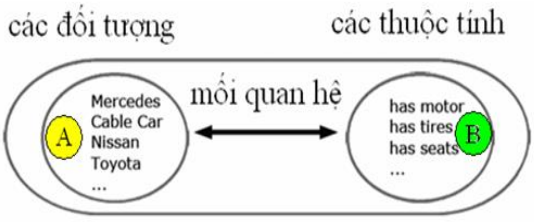
1.2. Khái niệm

Chúng ta hãy xét một ví dụ về khái niệm "xe hơi", điều gì khiến chúng ta gọi một đối tượng là "xe hơi" ?

Mọi đối tượng có các thuộc tính nhất định sau đây sẽ được gọi là "xe hơi":

- ❖ Chiếc xe có các lốp xe
- ❖ Chiếc xe có động cơ
- ❖ Chiếc xe có mục đích vận chuyển
- ❖ Chiếc xe có nhiều chỗ ngồi,…

Việc mô tả khái niệm "xe hơi" này dựa trên tập đối tượng liên quan đến tập thuộc tính:



Hình 1 - Mối quan hệ giữa các đối tượng và thuộc tính

Vậy: Các đối tượng, thuộc tính và mối quan hệ sẽ hình thành một khái niệm.

Do đó, khái niệm được cấu thành bởi hai phần: A là tập các đối tượng và B là tập các thuộc tính và chúng có mối quan hệ nhất định.

Nhận xét:

- Tất cả đối tượng thuộc tập A sẽ có tất cả thuộc tính thuộc tập B
- Tất cả thuộc tính thuộc tập B được chia sẻ cho tất cả đối tượng thuộc tập A
- A được gọi là phần mở rộng của khái niệm, B được gọi là phần nội dung của khái niệm.

1.3. Ngữ cảnh hình thức

Ví dụ, chúng ta xét bảng tham khảo chéo trong FCA sau đây:

I	y_1	y_2	y_3	y_4
x_1	×	×	×	×
x_2	×		×	×
x_3		×	×	×
x_4		×	×	×
x_5	×			

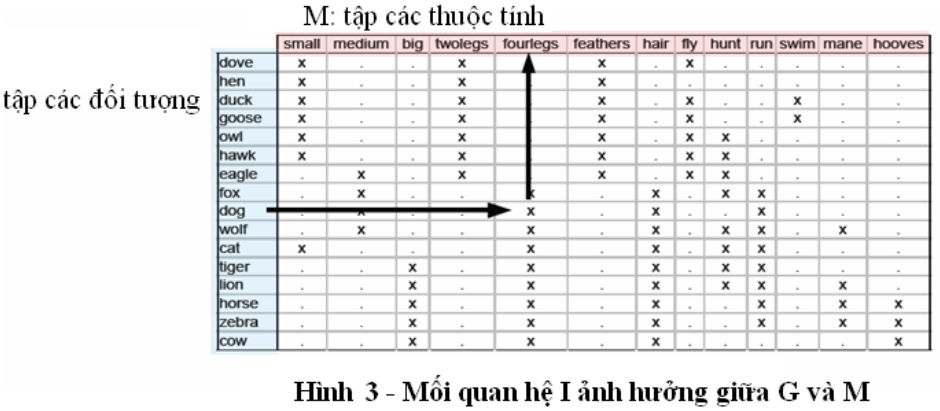
Hình 2 - Bảng tham khảo chéo

Trong bảng này, mô tả mối quan hệ giữa các đối tượng (đại diện bởi các hàng của bảng) và các thuộc tính (đại diện bởi các cột của bảng). Trong bảng chứa giá trị × (được gọi là thuộc tính logic), nó chỉ ra rằng đối tượng tương ứng có thuộc tính tương ứng.

Một cách hình thức, bảng tham khảo chéo đại diện bởi một ngữ cảnh hình thức.

Định nghĩa 1 (ngữ cảnh hình thức): Một ngữ cảnh hình thức là bộ ba ⟨X, Y, I⟩. Trong đó X và Y là các tập khác rỗng và I là một quan hệ hai ngôi giữa X và Y, tức là, $I \subseteq X \times Y$.

Đối với một ngữ cảnh hình thức, các phần tử x thuộc X được gọi là các đối tượng và các phần tử y thuộc Y được gọi là các thuộc tính. Cặp ⟨x, y⟩ ∈ I cho biết đối tượng x có thuộc tính y. Đối với một bảng tham khảo chéo đã cho với n hàng và m cột, tương ứng với ngữ cảnh hình thức ⟨X, Y, I⟩ bao gồm một tập X = {x1,. . . , xn}, một tập Y = {y1,. . . , ym}, và mối quan hệ I được xác định bởi: cặp $\langle x_i, y_j \rangle \in I$, nếu và chỉ nếu thuộc tính logic của bảng tương ứng với hàng i và cột j chứa giá trị ×.



	small	medium	big	twolegs	fourlegs	feathers	hair	fly	hunt	run	swim	mane	hooves
dove	x	.	.	x	.	x	.	x	.	.	.	.	.
hen	x	.	.	x	.	x	.	.	.	.	.	.	.
duck	x	.	.	x	.	x	.	x	.	.	x	.	.
goose	x	.	.	x	.	x	.	x	.	.	x	.	.
owl	x	.	.	x	.	x	.	x	x	.	.	.	.
hawk	x	.	.	x	.	x	.	x	x	.	.	.	.
eagle	.	x	.	x	.	x	.	x	x	.	.	.	.
fox	.	x	.	.	x	.	x	.	x	x	.	.	.
dog	.	.	.	.	x	.	x	.	.	x	.	.	.
wolf	.	x	.	.	x	.	x	.	x	x	.	x	.
cat	x	.	.	.	x	.	x	.	x	x	.	.	.
tiger	.	.	x	.	x	.	x	.	x	x	.	.	.
lion	.	.	x	.	x	.	x	.	x	x	.	x	.
horse	.	.	x	.	x	.	x	.	.	x	.	x	x
zebra	.	.	x	.	x	.	x	.	.	x	.	x	x
cow	.	.	x	.	x	.	x	.	.	.	.	.	x

Hình 3 - Mối quan hệ I ảnh hưởng giữa G và M

1.4. Khái niệm hình thức

❖ Về định nghĩa toán học của các khái niệm hình thức, chúng ta tìm hiểu các toán tử đạo hàm " ' ".

Cho một tập các đối tượng $A \subseteq X$, A' được định nghĩa như sau:

A' = {tất cả thuộc tính trong Y được chia sẻ bởi các đối tượng của A}

Cho một tập các thuộc tính của $B \subseteq Y$, B' được định nghĩa như sau:

B' = {tất cả các đối tượng trong X có tất cả các thuộc tính của B}.

Ví dụ 1 Cho bảng tham khảo chéo trong FCA ở Hình 2:

Chúng ta có:

– $\{x_2\}' = \{y_1, y_3, y_4\}$, $\{x_2, x_3\}' = \{y_3, y_4\}$

– $\{x_1, x_4, x_5\}' = \emptyset$

– $X' = \emptyset$, $\emptyset' = Y$

– $\{y_1\}' = \{x_1, x_2, x_5\}$, $\{y_1, y_2\}' = \{x_1\}$

– $\{y_2, y_3\}' = \{x_1, x_3, x_4\}$, $\{y_2, y_3, y_4\}' = \{x_1, x_3, x_4\}$

– $\emptyset' = X$, $Y' = \{x_1\}$

❖ Khái niệm hình thức là khái niệm cơ bản của FCA. Khái niệm hình thức được định nghĩa như sau:

Định nghĩa 2 (khái niệm hình thức): Một khái niệm hình thức trong ngữ cảnh hình thức ⟨X, Y, I⟩ là một cặp ⟨A, B⟩ với $A \subseteq X$ và $B \subseteq Y$ sao cho A' = B và B' = A.

Cho một khái niệm hình thức ⟨A, B⟩ trong ngữ cảnh hình thức ⟨X, Y, I⟩, trong đó A là phần mở rộng và B là phần nội dung của khái niệm hình thức ⟨A, B⟩.

Khái niệm hình thức được mô tả bằng lời như sau: Cặp ⟨A, B⟩ là một khái niệm hình thức nếu và chỉ nếu A chỉ chứa các đối tượng chia sẻ cho tất cả các thuộc tính từ B và B chỉ chứa các thuộc tính được chia sẻ bởi tất cả các đối tượng từ A.

Ví dụ 2 (khái niệm hình thức). Cho bảng sau:

I	y_1	y_2	y_3	y_4
x_1	×	×	×	×
x_2	×		×	×
x_3		×	×	×
x_4		×	×	×
x_5	×			

Hình 4 - Khối được chọn đại diện cho khái niệm hình thức

Hình chữ nhật được đánh dấu đại diện cho khái niệm hình thức

⟨A1, B1⟩ = ⟨{x1, x2, x3, x4}, {y3, y4}⟩

Bởi vì:

{x1, x2, x3, x4}' = {y3, y4} và {y3, y4}' = {x1, x2, x3, x4}.

Ngoài ra, còn có thêm các khái niệm hình thức khác. Chúng được đại diện bởi các hình chữ nhật được đánh dấu sau đây:

I	y_1	y_2	y_3	y_4
x_1	×	×	×	×
x_2	×		×	×
x_3		×	×	×
x_4		×	×	×
x_5	×			

I	y_1	y_2	y_3	y_4
x_1	×	×	×	×
x_2	×		×	×
x_3		×	×	×
x_4		×	×	×
x_5	×			

I	y_1	y_2	y_3	y_4
x_1	×	×	×	×
x_2	×		×	×
x_3		×	×	×
x_4		×	×	×
x_5	×			

Hình 5 - Các khái niệm hình thức khác đại diện bởi hình chữ nhật được đánh dấu

Tức là:

⟨A2, B2⟩ = ⟨{x1, x3, x4}, {y2, y3, y4}⟩

⟨A3, B3⟩ = ⟨{x1, x2}, {y1, y3, y4}⟩

⟨A4, B4⟩ = ⟨{x1, x2, x5}, {y1}⟩.

Ví dụ minh họa: Cho bảng sau

	nhỏ	trung bình	lớn	hai chân	bốn chân	lông vũ	lông thú	bay	săn bắt	chạy	bơi	bòm	móng guốc
bồ câu	x	.	.	x	.	x	.	x	.	.	.	.	.
gà	x	.	.	x	.	x	.	.	.	.	.	.	.
vịt	x	.	.	x	.	x	.	x	.	.	x	.	.
ngỗng	x	.	.	x	.	x	.	x	.	.	x	.	.
cú	x	.	.	x	.	x	.	x	x	.	.	.	.
diều hâu	x	.	.	x	.	x	.	x	x	.	.	.	.
đại bàng	.	x	.	x	.	x	.	x	x	.	.	.	.
cáo	.	x	.	.	x	.	x	.	x	x	.	.	.
chó	.	x	.	.	x	.	x	.	.	x	.	.	.
sói	.	x	.	.	x	.	x	.	x	x	.	x	.
mèo	x	.	.	.	x	.	x	.	x	x	.	.	.
cọp	.	.	x	.	x	.	x	.	x	x	.	.	.
sư tử	.	.	x	.	x	.	x	.	x	x	.	x	.
ngựa	.	.	x	.	x	.	x	.	.	x	.	x	x
ngựa vằn	.	.	x	.	x	.	x	.	.	x	.	x	x
bò	.	.	x	.	x	.	x	.	.	.	.	.	x

Hình 6 - Bảng tham khảo chéo về một số đặc điểm của một số loài vật

Chọn bất kỳ tập các đối tượng A, ví dụ: A = {vịt}.

Suy ra các thuộc tính A' = {nhỏ, hai chân, lông vũ, bay, bơi}

Suy ra (A')' = { nhỏ, hai chân, lông vũ, bay, bơi}' = {vịt, ngỗng}

(A", A') = ({ vịt, ngỗng}, {nhỏ, hai chân, lông, bay, bơi}) là một khái niệm hình thức.

1.5. Mạng khái niệm

Theo Port-Royal, một khái niệm được xác định bởi một tập các đối tượng và một tập các thuộc tính. Các khái niệm được sắp thứ tự bằng cách sử dụng một mối quan hệ subconcept-superconcept. Mối quan hệ subconcept-superconcept dựa vào quan hệ bao hàm trên các đối tượng và thuộc tính. Một cách hình thức, mối quan hệ subconcept-superconcept được định nghĩa như sau:

Định nghĩa 3 (sắp thứ tự subconcept-superconcept): Cho các khái niệm hình thức ⟨A1, B1⟩ và ⟨A2, B2⟩ của ngữ cảnh hình thức ⟨X, Y, I⟩, đặt ⟨A1, B1⟩ ≤ ⟨A2, B2⟩ khi và chỉ khi A1 ⊆ A2 (B2 ⊆ B1).

Trong đó:

– ≤ đại diện cho việc sắp thứ tự subconcept-superconcept.

– ⟨A1, B1⟩ ≤ ⟨A2, B2⟩ nghĩa là ⟨A1, B1⟩ cụ thể hơn so với ⟨A2, B2⟩ (⟨A2, B2⟩ thì tổng quát hơn so với ⟨A1, B1⟩).

Ví dụ 3. Hãy xét những khái niệm hình thức sau đây từ ví dụ 2:

⟨A1, B1⟩ = ⟨{x1, x2, x3, x4}, {y3, y4}⟩

⟨A2, B2⟩ = ⟨{x1, x3, x4}, {y2, y3, y4}⟩

⟨A3, B3⟩ = ⟨{x1, x2}, {y1, y3, y4}⟩

⟨A4, B4⟩ = ⟨{x1, x2, x5}, {y1}⟩.

Khi đó:

⟨A3, B3⟩ ≤ ⟨A1, B1⟩,

⟨A3, B3⟩ ≤ ⟨A4, B4⟩, ⟨A2, B2⟩ ≤ ⟨A1, B1⟩

⟨A1, B1⟩ || ⟨A4, B4⟩ (không thể so sánh được)

⟨A2, B2⟩ || ⟨A4, B4⟩ (không thể so sánh được).

Tập của tất cả các khái niệm hình thức của một ngữ cảnh hình thức đã cho được gọi là mạng khái niệm, một khái niệm cơ bản của FCA.

Định nghĩa 4 (mạng khái niệm): Ký hiệu B(X, Y, I) là tập của tất cả các khái niệm hình thức của ngữ cảnh hình thức ⟨X, Y, I⟩, tức là:

$B(X, Y, I) = \{\langle A, B\rangle \in 2^X \times 2^Y \mid A' = B, B' = A\}$.

Tập của tất cả các khái niệm hình thức B(X, Y, I) có thể sắp thứ tự subconcept-superconcept được gọi là mạng khái niệm của ngữ cảnh hình thức ⟨X, Y, I⟩.

Như vậy ⟨B(X, Y, I), ≤⟩ là mạng khái niệm.

Ví dụ 4. Hãy xét bảng tham khảo chéo sau đây:

		a	*b*	*c*	*d*	*e*	*f*	*g*	*h*	*i*
con đỉa	1	×	×					×		
cá trắp	2	×	×					×	×	
con ếch	3	×	×	×				×	×	
con chó	4	×		×				×	×	×
cá thu	5	×	×		×		×			
cây lau	6	×	×	×	×		×			
cây đậu	7	×		×	×	×				
cây bắp	8	×		×	×		×			

Hình 7 - Bảng tham khảo cho biết thuộc tính của một số loài vật

a: cần nước để sinh sống, b: cuộc sống trong nước, c: cuộc sống trên đất liền, d: nhu cầu chất diệp lục để sản sinh chất dinh dưỡng, e: hạt giống hai lá mầm, f: hạt giống một lá mầm, g: có thể di chuyển xung quanh, h: có chân tay, i: nuôi nấng con cái của mình.

Ngữ cảnh hình thức tương ứng ⟨X, Y, I⟩ chứa các khái niệm hình thức sau: $C_0 = \langle\{1, 2, 3, 4, 5, 6, 7, 8\}, \{a\}\rangle$, $C_1 = \langle\{1, 2, 3, 4\}, \{a, g\}\rangle$,

$C_2 = \langle\{2, 3, 4\}, \{a, g, h\}\rangle$, $C_3 = \langle\{5, 6, 7, 8\}, \{a, d\}\rangle$,

$C_4 = \langle\{5, 6, 8\}, \{a, d, f\}\rangle$, $C_5 = \langle\{3, 4, 6, 7, 8\}, \{a, c\}\rangle$,

$C_6 = \langle\{3, 4\}, \{a, c, g, h\}\rangle$, $C_7 = \langle\{4\}, \{a, c, g, h, i\}\rangle$,

$C_8 = \langle\{6, 7, 8\}, \{a, c, d\}\rangle$, $C_9 = \langle\{6, 8\}, \{a, c, d, f\}\rangle$,

$C_{10} = \langle\{7\}, \{a, c, d, e\}\rangle$, $C_{11} = \langle\{1, 2, 3, 5, 6\}, \{a, b\}\rangle$,

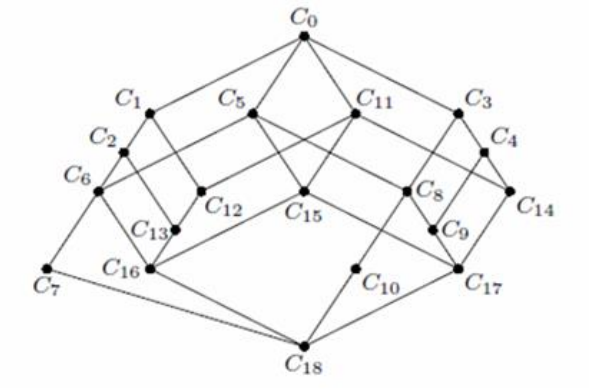
$C_{12} = \langle\{1, 2, 3\}, \{a, b, g\}\rangle$, $C_{13} = \langle\{2, 3\}, \{a, b, g, h\}\rangle$,

$C_{14} = \langle\{5, 6\}, \{a, b, d, f\}\rangle$, $C_{15} = \langle\{3, 6\}, \{a, b, c\}\rangle$,

C_{16} = ⟨{3}, {a, b, c, g, h}⟩, C_{17} = ⟨{6}, {a, b, c, d, f}⟩,

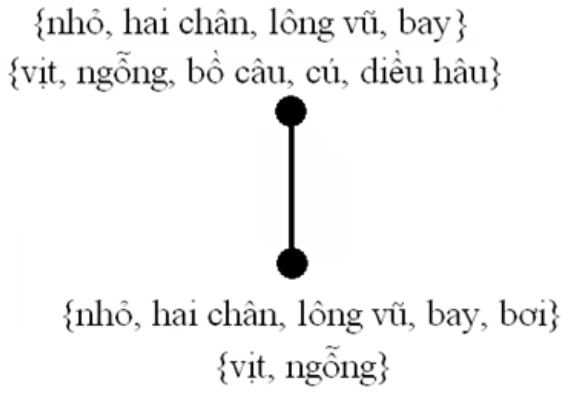
C_{18} = ⟨{}, {a, b, c, d, e, f, g, h, i}⟩.

Mạng khái niệm ⟨B(X, Y, I), ≤⟩ tương ứng được mô tả trong hình sau đây:



Hình 8 - Mạng khái niệm trong ví dụ 4

- Khái niệm hình thức (A", A') = ({vịt, ngỗng}, {nhỏ, hai chân, lông, bay, bơi}) đại diện trong sơ đồ tuyến tính là một nút:



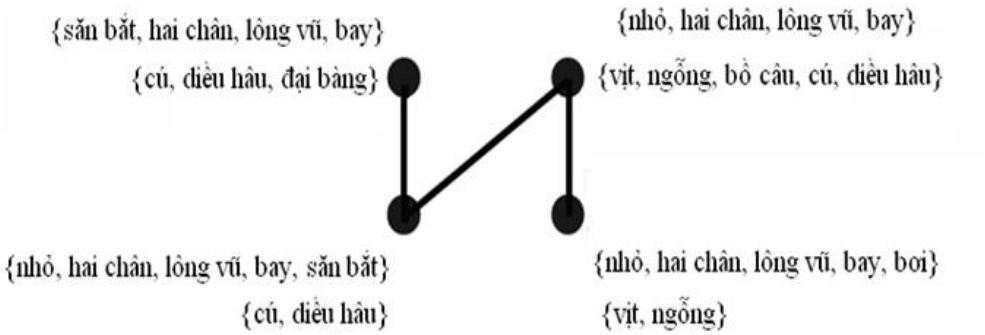
Hình 9 - Mỗi nút là một khái niệm hình thức

- Xét một khái niệm hình thức khác:

(B", B') = ({vịt, ngỗng, bồ câu, cú, diều hâu}, {nhỏ, hai chân, lông, bay}).

- Khái niệm hình thức (A", A') được gọi là subconcept của (B", B') và (B", B') được gọi là superconcept của (A", A').
- (A", A') được vẽ phía dưới của (B", B') và kết nối nhau bởi một đường thẳng.
- Từ đó, ta có thể thêm các khái niệm hình thức khác vào sơ đồ mở rộng:

- ({cú, diều hâu}, {lông, hai chân, nhỏ, bay, đi săn})
- ({cú, diều hâu, đại bàng}, {lông, hai chân, bay, đi săn})
- cộng với các mối quan hệ



Hình 10 - Sơ đồ mở rộng cho các khái niệm hình thức khác

- …

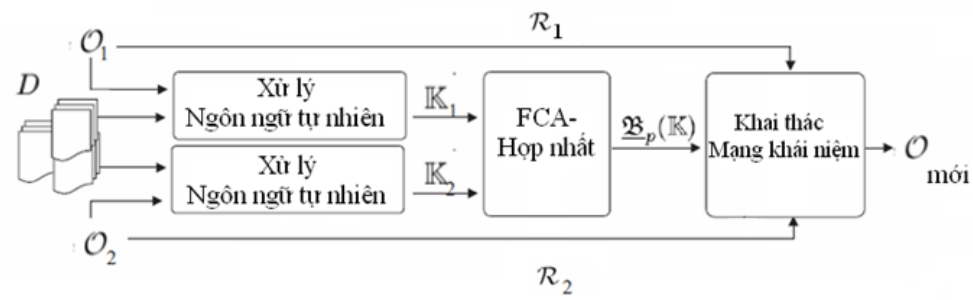
Một số phương pháp có thể suy ra tất cả các khái niệm hình thức: thuật toán của Ganter, thuật toán của Lindig,…

2. ỨNG DỤNG FCA TRONG TỰ ĐỘNG HOÁ TIẾN HOÁ ONTOLOGY

Chúng ta sẽ tìm hiểu một số ứng dụng FCA trong tự động hoá tiến hoá ontology.

***Ứng dụng 1*:** Sự trợ giúp của FCA trong việc giải quyết vấn đề hợp nhất các ontology [4].

Quá trình hợp nhất ontology bằng cách đưa vào hai nguồn ontology (hoặc hơn) và trả về một ontology hợp nhất dựa trên các ontology nguồn đã cho. Kết quả là ontology có thể được sử dụng để biên dịch giữa các ứng dụng dựa trên các ontology nguồn tương ứng của chúng. Các kết quả có chất lượng cao của quá trình hợp nhất sẽ luôn luôn cần con người tham gia để có thể thực hiện sự đánh giá dựa trên kiến thức nền. Như vậy, tất cả các phương pháp tiếp cận hợp nhất nhằm hỗ trợ kỹ sư tri thức.



Hình 11- FCA-phương pháp tiếp cận hợp nhất ontology

Đối với mỗi ontology nguồn, nó trích xuất các thể hiện từ một tập các tài liệu văn bản của miền cụ thể bằng cách áp dụng các kỹ thuật xử lý ngôn ngữ tự nhiên (xem phần bên trái của Hình 11). Bằng cách này, ngữ cảnh được tính toán cho mỗi ontology nguồn. Các đối tượng của nó là các tài liệu, và các thuộc tính của nó là các khái niệm ontology. Một khái niệm ontology sẽ liên quan đến tài liệu nếu và chỉ nếu nó tồn tại trong tài liệu. Các ngữ cảnh được ghép lại với nhau, mạng khái niệm được lược bớt và tính toán với thuật toán Titanic [1] (xem ở giữa Hình 1). Mạng khái niệm thực hiện sự phân cấp, sự phân nhóm các khái niệm của các ontology nguồn. Nó được khai thác và tương tác lẫn nhau chuyển thành ontology được hợp nhất.

Trong phương pháp này, các khái niệm ontology được đồng nhất với các thuộc tính FCA, và không đồng nhất với các khái niệm hình thức. Các thuộc tính sẽ là đầu vào của FCA, trong khi các khái niệm hình thức sẽ là phần hiển thị ở đầu ra của FCA.

***Ứng dụng 2*:** Sự trợ giúp của FCA trong việc khai thác các khái niệm của ontology trong văn bản [4].

Trên thực tế, các khái niệm có thể được tìm thấy trong các văn bản ở các cấp độ khác nhau tùy thuộc minh bạch của các loại văn bản được xem xét, chẳng hạn như một số văn bản chứa các khái niệm rõ ràng dưới hình thức các định nghĩa như "một con hổ là một động vật có vú" hoặc "các động vật có vú như hổ, sư tử hay voi". Một số nhà nghiên cứu tìm hiểu các mô hình để tìm ra sự phân loại hoặc mối quan hệ của các khái niệm trong văn bản. Nhiệm vụ chúng ta phải làm thế nào để tìm ra chúng. Một giải pháp đưa ra là các khái niệm từ các văn bản được phân tích và được sử dụng như thế nào, hơn là tìm kiếm các định nghĩa rõ ràng về chúng.

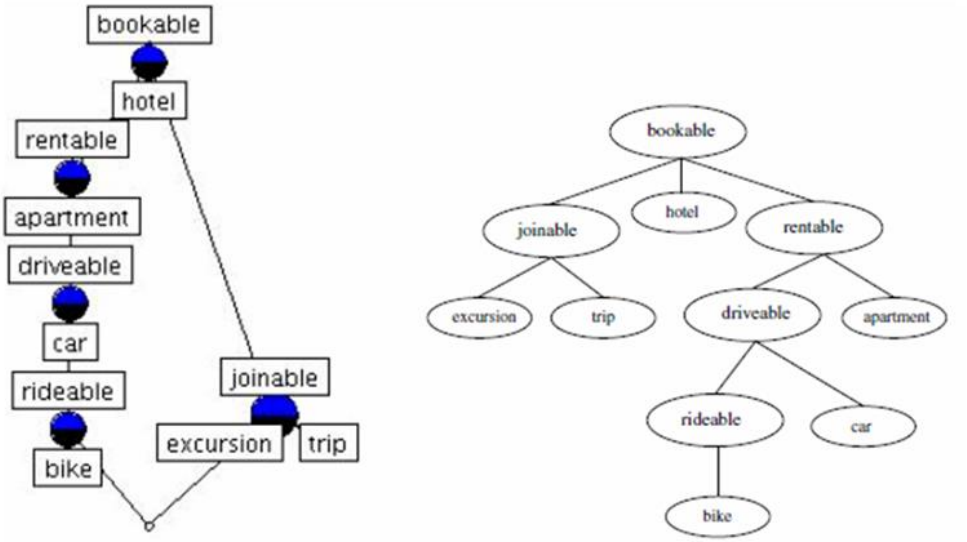
Trong giải pháp này, giả thuyết phân tán giả định rằng các khái niệm giống nhau về mức độ sẽ được dùng chung cho các ngữ cảnh giống nhau.

Giả sử rằng chúng ta quan tâm đến một số khái niệm trong lĩnh vực du lịch trong việc phân tích các văn bản liên quan đến lĩnh vực này. Bằng cách nhìn vào những động từ cũng như các đối tượng trực tiếp của những động từ ấy, chúng ta có thể suy ra một ngữ cảnh hình thức như mô tả trong Hình 12.

	đăng ký	thuê	lái	cưỡi	tham gia
khách sạn	x				
căn hộ	x	x			
xe hơi	x	x	x		
xe đạp	x	x	x	x	
cuộc du ngoạn	x				x
chuyến đi xa	x				x

Hình 12. Quan hệ giữa đối tượng và động từ như là ngữ cảnh hình thức

Chúng ta giả định rằng các mối quan hệ trong Hình 12 có thể là ít hoặc nhiều và tất cả các mối quan hệ không xảy ra trong văn bản được coi là các trường hợp không thực hiện được. Chúng ta có thể nhóm các đối tượng vào các lớp hoặc thậm chí tạo thành một hệ thống phân cấp các khái niệm bằng việc phân tích các ngữ cảnh dùng chung của chúng. Trong hầu hết các kỹ thuật phân nhóm, người ta cố gắng nhóm các khái niệm xuất hiện trong các văn bản thành các lớp có ý nghĩa hoặc phân cấp các khái niệm. Trong giải pháp này, FCA sử dụng cấu trúc các khái niệm trừu tượng. Mạng khái niệm của ngữ cảnh hình thức hiển thị trong Hình 3.12 được mô tả bên trong Hình 13.



Hình 13 - Mạng khái niệm và thứ tự bộ phận đối với ví dụ du lịch

Như vậy, FCA có thể hỗ trợ để chuyển đổi mạng khái niệm thành hệ thống phân cấp các khái niệm của ontology như hiển thị trong hình bên phải cùa Hình 13 bằng cách loại bỏ phần tử đáy của mạng khái niệm.

***Ứng dụng 3*:** Sự trợ giúp của FCA trong hệ thống quản lý email dựa trên ontology.

Hệ thống quản lý email chuẩn lưu trữ các mail trực tiếp từ cấu trúc cây của các kho hồ sơ và các hệ thống quản lý tập tin. Điều này có lợi thế là các cây có cấu trúc đơn giản và có thể giải thích dễ dàng cho người mới tiếp cận sử dụng. Bất lợi là tại thời điểm lưu trữ email người sử dụng phải thấy trước cách thức mà mail sẽ phục hồi lại.

Chúng ta sẽ tìm hiểu khái niệm quản lý Email CEM. Nó sử dụng một ontology đơn giản để lưu trữ các email. Ontology bao gồm một hệ thống phân cấp các khái niệm, cùng với kho từ vựng. Hệ thống phân cấp của ontology có thể là tập bất kỳ được sắp thứ tự bộ phận, đa thừa kế. Các Email có thể được ấn định cho nhiều khái niệm của ontology. Trong thực tế, lợi thế của FCA là hỗ trợ sự quản lý và các tác vụ phục hồi các email dựa trên ontology.

Từ góc độ của FCA trong ngữ cảnh hình thức, các đối tượng FCA là các email, và các thuộc tính FCA là các khái niệm của ontology.

***Ứng dụng 4*:** Sự trợ giúp của FCA trong hệ thống quản lý đối với các kho tri thức được phân tán.

Trên một máy tính cá nhân, có thể tổ chức các nguồn tài nguyên theo nhu cầu người sử dụng. Trong trường hợp các nguồn tài nguyên lưu trữ từ xa, máy tính không thể thực hiện được việc lưu trữ của chúng vì không thuộc thẩm quyền của người sử dụng. Thông qua việc sử dụng siêu văn bản, tài liệu từ xa có thể được liên kết và phục hồi khi cần thiết, vấn đề của việc tìm kiếm và tổ chức tài liệu này từ xa trở nên quan trọng hơn.

Phần mềm tổ chức giám sát CoursewareWatchdog là một phần của dự án PADLR (Personalized Access to Distributed Learning Repositories) được xây dựng dựa trên phương pháp tương đương để hỗ trợ truy cập của người sử dụng đến tài liệu nghiên cứu. Phần mềm này được xây dựng trong tầm điều khiển của Karlsruhe Ontology và Semantic Web

Framework KAON. Nó thiết lập trong phạm vi của hệ thống quản lý ontology, và các kỹ thuật duyệt cho phép duyệt các ontology và cơ sở tri thức.

Các ontology dựa trên hai loại quan hệ: quan hệ phân cấp và không phân cấp. Trong mỗi loại quan hệ, người ta sử dụng một kỹ thuật thích hợp.

Đối với quan hệ phân cấp sẽ thông qua mạng khái niệm của ngữ cảnh hình thức.

Ngược lại, phần mềm tổ chức giám sát CoursewareWatchdog sẽ xem xét các mối quan hệ không phân cấp trong ontology. Những quan hệ này đại diện cho các liên kết giữa các phần tử khác nhau của ontology (ví dụ, các "giảng viên" của một "khoá học" nên được liên kết với nó bằng một mối quan hệ "holdsCourse"). Trình duyệt quan hệ là một kỹ thuật bao gồm việc cung cấp các liên kết cho người sử dụng. Ngoài việc duyệt thông thường cùng các siêu liên kết, các liên kết được phân loại phù hợp với ontology. Nó có thể điều hướng và khai thác ontology theo các mối quan hệ của ontology.

Trong khuôn khổ của FCA, các ontology có thể được xem xét (theo một số ràng buộc) như là đa ngữ cảnh. Người ta hiện đang cố gắng làm thế nào để hình thức hoá mối quan hệ này, khai thác và để tích hợp chúng chặt chẽ hơn. Đặc biệt, người ta muốn tăng cường hơn nữa sự hỗ trợ của FCA trong hệ thống quản lý đối với kho tri thức phân tán dựa trên sự tiến hoá ontology.

***Ứng dụng 5*:** Sự trợ giúp của FCA trong việc suy ra mạng khái niệm và kích thước của nó được trình bày trong logic mô tả đối với sự tiến hoá ontology.

Phương pháp tốt nhất để suy ra mạng khái niệm từ một tập dữ liệu là sự mở rộng khái niệm. Nó cho phép suy ra các thuộc tính đơn trị từ các thuộc tính đa trị, sau đó chúng được đưa vào để tính toán cho mạng khái niệm. Tuy nhiên, việc mở rộng khái niệm vẫn còn yêu cầu dữ liệu được thể hiện trong mối quan hệ (cơ sở dữ liệu) với tên đối tượng là một khoá chính. Sự mở rộng khái niệm như vậy không thể xử lý với nhiều hơn một mối quan hệ. Trong FCA, các mối quan hệ đã được mã hoá trong việc định nghĩa (đa trị) đa ngữ cảnh, cho phép chuyển đổi đa ngữ cảnh thành một cấu trúc có ý nghĩa của mạng khái niệm.

Để minh họa cho phương pháp tiếp cận này, chúng ta xét ví dụ sau: Cho cơ sở dữ liệu bao gồm hai quan hệ hiển thị ở nửa trên của Hình 14. Về mặt FCA nó là đa trị đa ngữ cảnh, về mặt logic mô tả (DL) nó là một A-Box (Assertion Box) [4]. Giả sử chúng ta muốn phân loại những người uống rượu vang. Với tập các định nghĩa hiển thị ở phần dưới của Hình 14, về mặt logic mô tả (DL) nó là một T-Box (Terminological Box), chúng ta có thể xác định các đặc trưng của những người thích uống rượu vang. Những định nghĩa này có thể đem lại hai trường hợp: phạm vi định hướng theo dữ liệu và phạm vi định hướng theo lý thuyết logic.

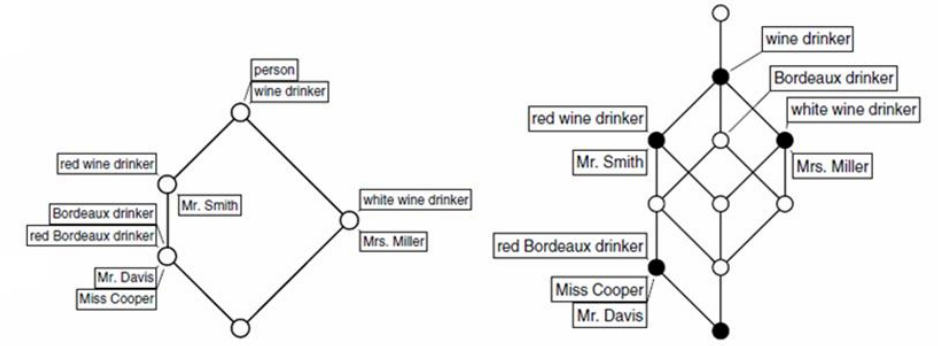
Việc thực hiện phạm vi định hướng theo dữ liệu [5] được hiển thị ở bên trái của Hình 15. Nó được suy ra từ cơ sở dữ liệu bằng cách thu hẹp tập các thể hiện của khái niệm 'Person', và chọn các thuộc tính của tất cả các khái niệm được định nghĩa trong T-Box. Từ sơ đồ, có thể cho thấy trường hợp những người uống rượu Bordeaux thì cũng uống được rượu vang đỏ. Tuy nhiên, sơ đồ không được rõ ràng nếu quan hệ này ảnh hưởng cho tất cả các đối tượng (nếu nó buộc bởi định nghĩa trong T-Box)

drinks	
Person	Wine
Mr. Smith	Casa Solar
Mrs. Miller	Staehle
Miss Cooper	Figeac
Mr. Davis	Figeac
Mr. Davis	Casa Solar

Wines	red wine	white wine	Bordeaux	Price
Figeac	×		×	49,90
Staehle		×		14,90
Casa Solar	×			5,95

wine drinker := person ⊓ ∃ drinks.(red wine ⊔ white wine)
red wine drinker := person ⊓ ∃ drinks.red wine
white wine drinker := person ⊓ ∃ drinks.white wine
Bordeaux drinker := person ⊓ ∃ drinks.Bordeaux
red Bordeaux drinker := person ⊓ ∃ drinks.(red wine ⊓ Bordeaux)

Hình 14 - Đa trị đa ngữ cảnh (bên trên) và định nghĩa về phạm vi logic (bên dưới)



Hình 15 - Kết quả của phạm vi định hướng theo dữ liệu (trái) và phạm vi định hướng theo lý thuyết logic

Phạm vi định hướng theo lý thuyết [4] có sự phân biệt này trong quá trình tính toán. Chúng được xem xét trong tất cả các kết nối có thể xảy ra của các thuộc tính được định nghĩa trong T-Box.

Việc thực hiện phạm vi định hướng theo lý thuyết của ví dụ được hiển thị ở phía bên phải của Hình 15. Phạm vi định hướng theo dữ liệu được nhúng trong nó như là sự kết hợp một phần của mạng. Trong sơ đồ, ta có thể thấy sự quan hệ dựa trên tập thực tế của các thể hiện, nó cũng cho thấy các thuộc tính kết hợp có thể xảy ra.

Chúng ta thấy rằng định nghĩa không bắt buộc người uống rượu Bordeaux phải uống uống rượu vang đỏ (họ cũng có thể chỉ uống Bordeaux trắng). Cũng có thể thấy rằng 'người uống rượu vang' là chung nhất cho các thuộc tính được xác định.

Các phương pháp này có thể được áp dụng cho DL (Description Logics) bất kỳ. Nó cần một thuật toán hợp lý để xác định một thể hiện có thuộc về A-Box, có trong từng mô hình, đến một khái niệm đã cho trong T - Box hay không. Nếu có thể nhiều hơn 5-6 định nghĩa trong T-Box, chúng phải được nhóm lại theo chủ đề thành các tập con nhỏ hơn để trở thành sự mở rộng của kích thước hợp lý; mỗi tập con làm tăng sự mở rộng logic đã cho.

3. KẾT LUẬN

Chúng ta đã tìm hiểu các phương pháp và kỹ thuật của các ứng dụng FCA hỗ trợ tốt cho sự tiến hoá ontology, và đó là bước đầu tiên được thực hiện để thiết lập các liên kết giữa các lý thuyết cơ bản. Các liên kết này phải được tăng cường hơn nữa và được khai thác triệt để cho việc thiết lập một môi trường toàn diện đối với việc xử lý tri thức về khái niệm.

TÀI LIỆU THAM KHẢO

[1]. G. Stumme, R. Taouil, Y.Bastide, N.Pasqier, and L.Lakhal. Computing iceberg concept lattices with Titanic. *J. on Knowledge and Data Engineering*, 42(2):189–222, 2002.

[2]. GanterB.,WilleR.:FormalConceptAnalysis.MathematicalFoundations.Springer,1999.

[3]. GanterB.,StummeG.,WilleR.:FormalConceptAnalysis.FoundationsandApplications. Springer, 2005.

[4]. Philipp Cimiano, Andreas Hotho, Gerd Stumme, and Julien Tane, Conceptual Knowledge Processing with Formal Concept Analysis and Ontologies, Institute for Applied Informatics and Formal Description Methods (AIFB) University of Karlsruhe, D–76128 Karlsruhe, Germany, 2005.

[5]. S. Prediger and G. Stumme. Theory-driven logical scaling. conceptual information systems meet description logics. In E. Franconi et al, editor, Proc. 6th Intl. Workshop Knowledge Representation Meets Databases, Heidelberg. CEURWorkshop Proc, 2004.

[6]. S. Prediger. Logical scaling in formal concept analysis. In D.Lukose, H.Delugach, M.Keeler, L.Searle, and J.F. Sowa, editors, Conceptual structures: Fulfilling Peirce's dream, Heidelberg. Springer, 1997.