CHÍ KHOA HỌC № 4 - 1990

ức Giáo

RIEVAL SYSTEMS VỚI CÁC ÒN NGỮ LÀ CÁC RM VÀ FORMULA MỞ RỘNG

rong [1], người ta đã xétvấn đề tối thiểu hóa cấu trúc các Retrival systems $S = [X, Y, \sigma_T]$, ở $\in TREE$, $\sigma_T : X \longrightarrow 2^Y$ là hàm kết quả của cây T. hư vậy lớp $\{S = [X, Y, \sigma_T]/T \in TREE\}$ lớp con của lớp các Retrival systems mà SALTON đã định nghĩa trong [1]. Trong kếtquả chúng tôi muốn nghiên cứu một số tính chất của các Retrival systems tổng quát trên cơ sở ng các khái niệm về "Term" và "Formula" của W. Lipski và M. Marek trong [2].

1. KHÁI NIỆM VỀ TERM VÀ FORMULA MỞ RỘNG

iả sử X là tập không rỗng của các phần tử nào đó.

$$C_X := \{x \in X, f, \omega, F, W, \sim, +, \cdot, \longrightarrow, \longleftrightarrow,], \vee, \wedge, \Longrightarrow, \Longleftrightarrow, \forall, \exists, =\}$$

gọi là Alphabet của X. Các Term và Formula được định nghĩa trên bảng C_X.

ịnh nghĩa 1 (Term)

. Mỗi $x \in X$ gọi là một term.

. Ký hiệu f, ω là các term.

. Nếu t, t_1, t_2 là các term thì $\sim t, t_1 + t_2, t_1 \cdot t_2$, $t_1 \longrightarrow t_2$ và $t_1 \longleftrightarrow t_2$ cũng là các term.

ập các term định nghĩa như trên ta ký hiệu qua TERM.

ịnh nghĩa 2 (Formula)

. Các ký hiệu F, W là các formula.

. Nếu $t_1, t_2 \in TERM$ thì $(t_1 = t_2)$ là formula.

. Nếu H, H_1, H_2 là các formula thì dãy các ký hiệu $(\rceil H)$, $(H_1 \wedge H_2)$, $(H_1 \vee H_2)$, $(H_1 \Longrightarrow H_2)$, $\Longrightarrow H_2)$, $\forall x H$, $\exists x H$ cũng là các formula.

ập các formula định nghĩa như trên ký hiệu qua FORM. Với tập $X \neq \phi$ như trên, ta lấy thêm $\neq \phi$ sao cho $X \cap Y \equiv \phi$. Với cặp $[X, Y]$ ta gọi bộ ba $S = [X, Y, \sigma]$, trong đó $\sigma : X \longrightarrow 2^Y$, rievasystem. Tập các Retrievasystem định nghĩa như trên là $S[X, Y]$. Hay

$$S[X, Y] = \{S = [X, Y, \sigma]/\sigma : X \longrightarrow 2^Y\}.$$

iả sử $S = [X, Y, \sigma] \in S[X, Y]$ và $t \in TERM$ ta định nghĩa giá trị $val(t, S)$ như sau

ịnh nghĩa 3

. $val(f, S) = \phi$ (tập rỗng).

. $val(\omega, S) = Y$.

. $val(x, S) = \sigma(x) \quad \forall x \in X$.

4. $val(\sim t, S) = Y \setminus val(t, S)$.

5. $val(t_1 + t_2, S) = val(t_1, S) \cup val(t_2, S)$.

6. $val(t_1 \cdot t_2, S) = val(t_1, S) \cap val(t_2, S)$.

7. $val(t_1 \longrightarrow t_2, S) = val(\sim t_1, S) \cup val(t_2, S)$.

8. $val(t_1 \longleftrightarrow t_2, S) = (val(\sim t_1, S) \cup val(t_2, S)) \cap (val(\sim t_2, S) \cup val(t_1, S))$.

Giả sử $S = [X, Y, \sigma] \in S[X, Y]$ và $H \in FORM$ ta định nghĩa giá trị $VAL(H, S)$ như

Định nghĩa 4

1. $VAL(F, S) = 0$ (sai)

2. $VAL(W, S) = 1$ (đúng)

3. $VAL(t_1 = t_2, S) = \begin{cases} 1, & \text{nếu } val(t_1, S) = val(t_2, S) \\ 0, & \text{còn lại} \end{cases}$

4. $VAL(\rceil H, S) = \begin{cases} 1, & \text{nếu } VAL(H, S) = 0 \\ 0, & \text{nếu } VAL(H, S) = 1 \end{cases}$

5. $VAL(H_1 \wedge H_2, S) = \begin{cases} 1, & \text{nếu } VAL(H_1, S) = VAL(H_2, S) = 1 \\ 0, & \text{nếu ngược lại} \end{cases}$

6. $VAL(H_1 \vee H_2, S) = \begin{cases} 1, & \text{nếu } VAL(H_1, S) = 1 \text{ hoặc } VAL(H_2, S) = 1 \\ 0, & \text{nếu } VAL(H_1, S) = VAL(H_2, S) = 0 \end{cases}$

7. $VAL(H_1 \Longrightarrow H_2, S) = \begin{cases} 1, & \text{nếu } VAL(H_1, S) = 0 \text{ hoặc } VAL(H_2, S) = 1 \\ 0, & \text{ngược lại} \end{cases}$

8. $VAL(H_1 \Longleftrightarrow H_2, S) = \begin{cases} 1, & \text{nếu } VAL(H_1, S) = VAL(H_2, S) \\ 0, & \text{ngược lại} \end{cases}$

9. $VAL(\forall x H, S) = Om(\{VAL(H, S')/S' \in S[X, Y] \text{ và } val(x', S') = val(x', S)\ \forall x' \in$ ... $x\})$

10. $VAL(\exists x H, S) = eX(\{VAL(H, S')/S' \in S[X, Y] \text{ và } val(x', S') = val(x', S)\ \forall x' \in$ ... $x\})$

Ở đây Om và eX được định nghĩa qua bảng dưới:

	Om	eX
$\{1\}$	1	1
$\{0\}$	0	0
$\{0, 1\}$	0	1

Bây giờ ta đưa vào khái niệm tương đương giữa các term cũng như sự tương đươ[ng] các formula như sau

Định nghĩa 5

1. Hai Term $t_1, t_2 \in TERM$ là tương đương với nhau khi và chỉ khi $val(t_1, S) = va$[...] với mọi $S = S[X, Y]$ (ký hiệu $t_1 \underset{\sim}{T} t_2$).

2. Hai formula $H_1, H_2 \in FORM$ là tương đương với nhau (ký hiệu $H_1 \underset{\sim}{F} H_2$) khi và c[hỉ khi]

$$VAL(H_1, S) = VAL(H_2, S)\ \forall S \in S[X, Y].$$

rằng quan hệ $\underset{\sim}{T}$ ($\underset{\sim}{F}$) cho ta một phân hoạch tương đương $TERM/\underset{\sim}{T}$ ($FORM/\underset{\sim}{F}$) trên ($FORM$).

h lý 1

$TERM/\underset{\sim}{T}$ (cũng như tập $FORM/\underset{\sim}{F}$) là đại số bun.

ứng minh: Kiểm tra lại các tính chất của đại số bun.

ng định lý trên đối với $TERM/\underset{\sim}{T}$ phần tử 0 là "f" còn phần tử đơn vị là "ω". Đối với $/\underset{\sim}{F}$, "F" là phần tử 0 còn "W" là phần tử đơn vị.

2. TẬP MÔ TẢ ĐƯỢC TRONG RETRIEVALSYSTEM VÀ RETRIEVALSYSTEM-CHON

h nghĩa 6

sử $S = [X, Y, \sigma] \in S[X, Y]$ và $A \subseteq Y$. Tập A gọi là mô tả được trong S khi và chỉ khicó $\in TERM$ sao cho $val(t, S) = A$.

gọi $A^S = \{A/A \subseteq Y, \exists t \in TERM$ mà $val(t, S) = A\}$ là tập tất cả các tập mô tả được

h lý 2

mỗi $S \in S[X, Y]$ thì

A^S_Y là tập không rỗng.

A^S_Y đóng đối với các phép toán lấy phần bù, giao và hợp các tập hợp.

$A^S_Y/\underset{\sim}{T}$ là một đại số bun, với phần tử 0 là ϕ, phần tử đơn vị là Y.

ứng minh: Dựa vào định nghĩa.

h nghĩa 7

sử $S = [X, Y, \sigma] \in S[X, Y]$

ọi là hệ chọn khi và chỉ khi với mỗi $y \in Y$ có tồn tại $t \in TERM$ sao cho $val(t, S) = \{y\}$.

h lý 3

$[X, Y, \sigma]$ là một hệ chọn khi và chỉ khi với mỗi tập con bất kỳ trong Y đều là tập mô tả ng S.

ng minh: Dựa vào định nghĩa hệ chọn và term.

h lý 4

sử $t_1, t_2 \in TERM$ và $S = [X, Y, \sigma] \in S[X, Y]$ là hệ chọn. $val(t_1, S) \subseteq val(t_2, S)$ khi và ồn tại $t \in TERM$ sao cho $val(t_1 + t, S) = val(t_2, S)$.

ng minh: Dễ dàng suy từ định nghĩa.

3. KHÁI NIỆM VỀ RETRIEVALSYSTEMS BỘ PHẬN

nghĩa 8

$[x', Y', \sigma']$ được gọi là Retrievalsystem bộ phận của $S = [X, Y, \sigma]$ (ký hiệu $s' \subset S$) nếu u các điều kiện sau đây thỏa mãn

$' \subseteq X$

$' \subseteq Y$

3. Với mỗi $x \in X'$ ta có $\delta'(x) = \delta(x) \cap Y'$

Định nghĩa 9

Giả sử $S = [X, Y, \sigma]$ và $S' = [x', Y', \sigma']$ là các Retrievalsystems.

1. $S \overset{X}{\subset} S'$ khi và chỉ khi $S \subset S'$ và $X = X'$

2. $S \underset{Y}{\subset} s'$ khi và chỉ khi $S \subset S'$ và $Y = Y'$

Dễ ràng suy ra kết quả sau

Định lý 5

1. Nếu $S \subset S'$ thì với mọi $t \in TERM$ ta luôn có

$$val(t, S) = val(t, S') \cap Y.$$

2. Nếu $S \underset{Y}{\subset} s'$ thì với mỗi $H \in FORM$ ta luôn có

$$VAL(H, S) = VAL(H, S').$$

Định lý 6

Giả sử $S = [X, Y, \sigma]$ và $S' = [x', Y', \sigma']$ là hai Retrievalsystems. Nếu $S \subset S$thì sẽ tồn t. Retrievalsystems $S_1 = [X_1, Y_1, \sigma_1]$ và $S_2 = [X_2, Y_2, \sigma_2]$ sao cho các điều kiện sau được thỏa

1. $S \underset{Y}{\subset} S_1 \overset{X}{\subset} S'$

2. $S \overset{X}{\subset} S_2 \underset{Y}{\subset} S'$

Định lý 7

Giả sử $S = [X, Y, \sigma] \in S[X, Y]$.

Nếu $B = \{B_1, B_2, \ldots, B_m\} \in 2^Y$ thì tồn tại một Retrievalsystem $S' = [X', Y', \sigma']$ thỏa hai điều kiện sau

1. $S \underset{Y}{\subset} S'$

2. $B \subseteq A_Y^{S'}$

và rõ ràng với S cho trước, ta xây dựng được

$s' = [x', Y', \sigma']$ sao cho $S \underset{Y}{\subset} S'$, ở đây $X' = TERM/\underset{\sim}{T}$, $Y' = Y$, và $\sigma' = \sigma_S$.

4. KHÁI NIỆM VỀ TỔNG CỦA CÁC RETRIEVALSYSTEM

Giả sử $S_i = [X_i, Y_i, \sigma_i]$ là Retrievalsystems trên $[X_i, Y_i]$ $(i = 1, 2, \ldots, n)$. Ta giả thiết $X_i \cap$ ϕ $(i \neq j)$.

Ta gọi hệ $\bigoplus_{i=1}^{n} S_i = [X, Y, \sigma]$ là tổng của các hệ $S_i (i = 1, 2, \ldots, n)$, trên bộ $[X, Y]$ nếu:

$$X = \bigcup_{i=1}^{n} X_i; \qquad Y = \bigcup_{i=1}^{n} Y_i \text{ và } \forall x \in X \text{ ta có :}$$

$$\sigma(x) = \begin{cases} \sigma_1(x), & \text{nếu } x \in X_1 \\ \vdots & \\ \sigma_n(x), & \text{nếu } x \in X_n \end{cases}$$

õ ràng suy ra kết quả sau

Định lý 8

. *Với mỗi* $j \in \{1, 2, \ldots, n\}$ *ta luôn có* $S_j \subset \bigoplus_{i=1}^{n} S_i$

. *Nếu* $Y_1 = Y_2 = \cdots = Y_n = Y$ *thì*

$$A_Y^{\bigoplus_{i=1}^{n} S_i} = \bigcup_{i=1}^{n} A_Y^{S_i}$$

. *Nếu* $S_i (i = 1, 2, \ldots, n)$ *là hệ chọn thì* $\bigoplus_{i=1}^{n} S_i$ *cũng là hệ chọn.*

5. KHÁI NIỆM VỀ TÍCH MÔ TẢ ĐƯỢC VÀ TÍNH ĐỘC LẬP CỦA NÓ TRONG RETRIEVALSYSTEM

iả sử $S = [X, Y, \sigma] \in \mathcal{S}[X, Y]$. Ta đặt

$$B^S = \{T/t \in TERM,\ \exists A \subseteq Y \text{ sao cho } val(t, S) = A\}.$$

Định nghĩa 10

iả sử $t \in TERM$. t được gọi là tích mô tả được trong $S = [X, Y, \sigma]$ nếu t có dạng: $\cdot t_2^{\varepsilon_2} \ldots t_m^{\varepsilon_m}$, trong đó $t_i \in B^S$, $\varepsilon_i \in \{0, 1\}$, $t_i^{\varepsilon_i} = t_i$ nếu $\varepsilon_i = 1$ còn $t_i^{\varepsilon_i} = \sim t_i$ nếu $\varepsilon_i = 0$ và $= m$.

Định nghĩa 11

iả sử $t = t_1^{\varepsilon_1} \cdot t_2^{\varepsilon_2} \ldots T_m^{\varepsilon_m}$ và $t' = t_1^{\varepsilon_1} \cdot t_2^{\varepsilon_2} \ldots T_m^{\varepsilon_m}$ là hai tích mô tả được trong $S = [X, Y, \sigma]$. i t là phân biệt được với t' (ký hiệu $t \neq t'$) nếu có tồn tại chỉ số $i_0 \in \{1, 2, \ldots, m\}$ sao cho $\varepsilon_{i'_0}$.

Định nghĩa 12

iả sử $S = [X, Y, \sigma] \in \mathcal{S}[X, Y]$ và $t, t' \in TERM$. Ta nói t và t' là độc lập với nhau trong S à chỉ nếu $val(t, S) \cap val(t', S) = \phi$.

Định lý 9

ới mỗi $S = [X, Y, \sigma] \in \mathcal{S}[X, Y]$ *ta có*

. $\bigcup_{t \in B^S} val(t, S) = Y$

. *Nếu t và t' là hai tích mô tả được trong S và $t_1 \neq t_2$ thì t và t' là độc lập với nhau trong*

hứng minh: 1. Suy từ tập B^S là đại số bun. 2. Suy từ định nghĩa.

hư trên chúng ta đã biết: với mỗi $S = [X, Y, \sigma]$ cho ta xác định một tập

$$B^S = \{t/t \in TERM,\ \exists A \subseteq Y \text{ sao cho } val(t, S) = A\},$$

$1 \leq |A| \leq Y$. Trường hợp đặc biệt khi $|A| = 1$ thì ta đặt

$$B_0^S = \{t/t \in TERM,\ \exists y \subseteq Y \text{ sao cho } val(t, S) = \{y\}\}$$

i là tập nhân của S. Rõ ràng $B_0^S \subseteq B^S$ với mọi $S \in \mathcal{S}X, Y$.

Về tập nhân dễ ràng chứng minh kết quả sau đây

Định lý 10

Với mỗi $S = [X, Y, \sigma] \in \mathcal{S}[X, Y]$ ta có

1. Tập B_0^S là duy nhất.
2. Nếu S là hệ chọn thì $t \in B^S$ khi và chỉ khi $t = \sum_{j \in I} t_j$, ở đây $|I| \leq |B_0^S|$ và $t_j \in B_0^S$.

TÀI LIỆU THAM KHẢO

1. Salton G., Automatic information organization and Retrieval. New York 1969 (Mc Gr... Book Company).
2. Lipski W. and Marek W., On information storage and retrieval systems. CCPAS rep... 200, Warsaw (1975).
3. Thiele H., On a graph-theoretic realisation of retrieval systems. F.N.S.E.T., Cachan, ... 4-8, Juillet 1977.
4. Đỗ Đức Giáo, Tạp chí Khoa học, ĐHTH Hà Nội, N. 4, 1986.
5. Đỗ Đức Giáo, Tạp chí Khoa học, ĐHTH Hà Nội, N. 4, 1987.

Do Duc Giao

RETRIEVAL SYSTEMS WITH THE LANGUAGES
IN PUT ARE TERM AND FORMULAS ENLAGE

In the paper [4] we have developed a graph theoretic realization of retrival systems defined by [1].

In this paper we present a mathematical foundations of retrieval systems - based on some idea... in [2].

Bộ môn Tin học - ĐHTH Hà Nội