

Bài 7

Phân tích cú pháp tiền định

ONE LOVE. ONE FUTURE.

Phân tích tiền định

- Tư tưởng chính của giải thuật phân tích cú pháp trên xuống **quay lui**
 - Bắt đầu từ gốc, phát triển xuống các nút cấp dưới
 - Chọn một sản xuất và thử xem có phù hợp với xâu vào không
 - Quay lui nếu lựa chọn dẫn đến ký hiệu được sinh bởi văn phạm không phù hợp ký hiệu đang xét
- Có thể tránh được **quay lui**?
 - Cho sản xuất $A \rightarrow \alpha \mid \beta$ bộ phân tích cú pháp cần chọn giữa α và β
- Làm thế nào?
 - Cho ký hiệu không kết thúc A và ký hiệu xem trước t, sản xuất nào của A chắc chắn sinh ra một xâu bắt đầu bởi t?

Phân tích tiền định

- Nếu có hai sản xuất: $A \rightarrow \alpha \mid \beta$, ta mong muốn có một phương pháp rõ ràng để chọn đúng sản xuất cần thiết
- Định nghĩa:
 - Với α là một xâu chứa ký hiệu kết thúc và không kết thúc, $x \in \text{FIRST}(\alpha)$ nếu từ α có thể suy dẫn ra xy (x chứa 0 hoặc 1 ký hiệu)
 - Nếu $\text{FIRST}(\alpha)$ và $\text{FIRST}(\beta)$ không chứa ký hiệu chung ta biết phải chọn $A \rightarrow \alpha$ hay $A \rightarrow \beta$ khi đã xem trước một ký hiệu

Phân tích tiền định

- Tính FIRST(X):

- Nếu X là ký hiệu kết thúc $\text{FIRST}(X)=\{X\}$
- Nếu $X \rightarrow \varepsilon$ là một sản xuất thì thêm ε vào $\text{FIRST}(X)$
- Nếu X là ký hiệu không kết thúc và $X \rightarrow Y_1Y_2\dots Y_n$ là một sản xuất ,
 - Thêm $\text{FIRST}(Y_1)$ vào $\text{FIRST}(X)$
 - Thêm $\text{FIRST}(Y_{i+1})$ vào $\text{FIRST}(X)$ nếu $\text{FIRST}(Y_1), \dots, \text{FIRST}(Y_j)$ chứa ε
- Tính $\text{FIRST}(\alpha)$ tương tự bước thứ ba trong tính $\text{FIRST}(X)$

Phân tích tiền định

- Nếu ta có sản xuất để chọn là $A \rightarrow \alpha$ với $\alpha = \varepsilon$ hoặc $\alpha \Rightarrow^* \varepsilon$? Ký hiệu nào sẽ là ký hiệu đầu tiên được sinh bởi một dạng câu chứa A?
- Có thể mở rộng A nếu ta biết rằng **tồn tại một dạng câu mà ký hiệu đang xét xuất hiện sau A**, nghĩa là ký hiệu đang xét thuộc $\text{FOLLOW}(A)$
- Định nghĩa:
 - Với A là ký hiệu không kết thúc, $x \in \text{FOLLOW}(A)$ nếu và chỉ nếu S có thể suy dẫn ra $\alpha A x \beta$, $|x| = 1$ hoặc $x = \varepsilon$ (khi ấy β cũng là ε)

Tính FOLLOW

- FOLLOW(S) chứa ϵ (EOF)
- Với các sản xuất dạng $A \rightarrow \alpha B \beta$, mọi ký hiệu trong FIRST(β) trừ ϵ tham gia vào FOLLOW(B)
- Với các sản xuất dạng $A \rightarrow \alpha B$ hoặc $A \rightarrow \alpha B \beta$ trong đó FIRST(β) chứa ϵ , FOLLOW(B) chứa mọi ký hiệu của FOLLOW(A) và ϵ (hoặc \$)

Phân tích tiền định

- Với các khái niệm
 - FIRST
 - FOLLOW
- Ta có thể xây dựng bộ phân tích cú pháp mà **không đòi hỏi quay lui**
- Chỉ có thể xây dựng bộ phân tích cú pháp như vậy cho những văn phạm đặc biệt
- Loại văn phạm như vậy bao gồm văn phạm một số ngôn ngữ lập trình đơn giản, chẳng hạn KPL, PL/0, PÁSCAL-S

Bảng phân tích tiền định

- Dùng cho bộ sinh phân tích cú pháp
- Đầu vào của giải thuật: văn phạm G và xâu w
- Căn cứ
 - Ký hiệu đang xét
 - Ký hiệu đang ở đỉnh stack
- Quyết định
 - Thay thế ký hiệu không kết thúc
 - Chuyển con trỏ sang ký hiệu tiếp
 - Chấp nhận xâu
 - Thông báo lỗi

Vào: Văn phạm phi ngũ cảnh LL(1) G
Xâu w

Các thành phần cơ bản

- Stack
- Bảng phân tích
- Bảng vào
- Chương trình phân tích

Mô tả các thành phần

- Băng vào chứa xâu cần phân tích, kết thúc bằng \$ (EOF)
- Stack giống như stack D2 của bộ phân tích cú pháp top down quay lui, # ở đáy của stack. Ban đầu S ở đỉnh stack, trên ký hiệu #.
- Bảng phân tích $M[A,a]$ với A là một ký hiệu của văn phạm, a là ký hiệu kết thúc hoặc \$.

Hoạt động của bộ phân tích cú pháp

- Nếu stack còn lại # (đáy), đầu đọc chỉ \$ (EOF), dừng và đoán nhận xâu.
- If X=a (ký hiệu kết thúc đang xét trên xâu vào) và không là \$, xóa X trên đỉnh stack , chuyển đầu đọc sang ô kế tiếp.
- Nếu X là ký hiệu không kết thúc, bộ PTCP tra bảng phân tích cú pháp M, tìm ô M[X,a], thay thế ký hiệu đỉnh stack (X) bằng vẽ phải sản xuất trong ô (nếu có). Nếu là ô rỗng -> ERROR, gọi thủ tục thông báo lỗi.

Bảng phân tích LL(1)

- Dùng cho bộ sinh phân tích cú pháp
- Căn cứ
 - Ký hiệu đang xét
 - Ký hiệu đang ở đỉnh stack
- Quyết định
 - Thay thế ký hiệu không kết thúc
 - Chuyển con trỏ sang ký hiệu tiếp
 - Chấp nhận xâu

Giải thuật xây dựng bảng phân tích

1. Với mỗi sản xuất $A \rightarrow \alpha$ của văn phạm G, thực hiện các bước 2 và 3.
2. Với mỗi ký hiệu kết thúc $a \in \text{FIRST}(\alpha)$, thêm $A \rightarrow \alpha$ vào $M[A,a]$.
3. If ε thuộc $\text{FIRST}(\alpha)$, thêm $A \rightarrow \alpha$ vào $M[A,b]$ với mỗi b thuộc $\text{FOLLOW}(A)$. If ε thuộc $\text{FIRST}(\alpha)$, và $\$$ thuộc $\text{FOLLOW}(A)$, thì thêm $A \rightarrow \alpha$ vào $M[A,\$]$
4. Các ô $M(a,a)$ với a là ký hiệu kết thúc, thêm hành động “đẩy”
5. $M[\#, \$] = \text{"nhận"}$
6. Các ô còn lại đánh dấu là “lỗi”.

Ví dụ

- Văn phạm:

```
E → TE'  
E' → +TE' | ε  
T → FT'  
T' → *FT' | ε  
F → (E) | id
```

$$\text{FIRST}(+TE') = \{+\}$$

$$\text{FOLLOW}(E') = \{\$,)\}$$

$$\text{FIRST}(*FT') = \{*\}$$

$$\text{FOLLOW}(T') = \{+, \$,)\}$$

$$\text{FIRST}((E)) = \{()\}$$

$$\text{FIRST}(id) = \{id\}$$

Văn phạm này LL(1)
có thể xây dựng bộ phân tích tiền định

Bảng phân tích

	$FIRST(+TE') = \{+\}$	$FIRST(E) = \{[, id\}$	$FOLLOW(E') = \{\$,)\}$
E		$E \rightarrow \epsilon$	
E'	$E' \rightarrow +TE'$		$E' \rightarrow \epsilon$
T		$T \rightarrow FT'$	$T \rightarrow FT'$
T'	$T' \rightarrow \epsilon$		$T' \rightarrow \epsilon$
F		$F \rightarrow (E)$	$F \rightarrow id$
+	Đây	Đây	Đây
*		Đây	Đây
(Đây
)			Nhận
id			
#			

Phân tích xâu vào id*id sử dụng bảng phân tích và stack

Bước	Stack	Xâu vào	Hành động kế tiếp
1	#E	id*id\$	E → TE'
2	#E'T	id*id\$	T → FT'
3	#E'T'F	id*id\$	F → id
4	#E'T'id	id*id\$	đẩy id
5	#E'T'	*id\$	T' → *FT'
6	#E'T'F*	*id\$	đẩy *
7	#E'T'F	id\$	F → id
8	#E'T'id	id\$	đẩy id
9	#E'T'	\$	T' → ε
10	#E'	\$	E' → ε
11	#	\$	nhận