

# Xây dựng bảng phân tích cú pháp đoán trước không đệ quy

Hoàng Văn Tuấn

Email: [tuanhoang.97dx@gmail.com](mailto:tuanhoang.97dx@gmail.com)

# Phân tích cú pháp

- Mục tiêu:
  - Cho một chuỗi chứa các ký hiệu kết thúc.
  - Xác định xem chuỗi này có thuộc văn phạm của ngôn ngữ đang xét hay không. Nếu có thì hãy xây dựng cây phân tích cú pháp tương ứng.
- Có 2 phương pháp phân tích cú pháp:
  - Phân tích theo chiều từ Trên – Xuống (Top – Down)
  - Phân tích theo chiều từ Dưới – Lên (Bottom – Up)

# Phân tích theo chiều từ Trên – Xuống (Top – Down)

- Cho 1 văn phạm phi ngữ cảnh G và 1 câu cần phân tích w.
- Xây dựng cây phân tích cú pháp cho câu đầu vào với:
  - Ký hiệu bắt đầu của văn phạm làm gốc.
  - Dùng các luật sinh thích hợp để khai triển cây dần xuống cho tới khi sinh ra được câu cần phân tích ở các nút lá.
- Có thể xem quá trình này như cách tìm dẫn xuất trái nhất cho câu.
- Phân tích cú pháp theo chiều từ trên xuống dừng khi và chỉ khi G không có tính đệ quy trái.

## **Xây dựng Parsing table M cho phân tích cú pháp đoán trước không đệ quy**

- Bước 1: Loại bỏ tính đệ quy trái
- Bước 2: Tính tập First
- Bước 3: Tính tập Follow
- Bước 4: Xây dựng Parsing table M

# 1. Loại bỏ tính đệ quy

- Một văn phạm được gọi là đệ quy trái nếu tồn tại một dẫn xuất có dạng  $A \rightarrow^+ A\alpha$
- Các phương pháp phân tích top – down không thể xử lý văn phạm đệ quy trái, do đó cần phải biến đổi văn phạm để loại bỏ tính đệ quy trái.
- Đệ quy trái có 2 loại:
  - Đệ quy trái trực tiếp: Có dạng  $A \rightarrow A\alpha$
  - Đệ quy trái gián tiếp: Gây ra do dẫn xuất của 2 hoặc nhiều bước.

# 1. Loại bỏ tính đệ quy

## Loại bỏ đệ quy trái trực tiếp:

- Nhóm các luật sinh:

$$A \rightarrow A\alpha_1 \mid A\alpha_2 \mid \dots \mid A\alpha_m \mid \beta_1 \mid \beta_2 \mid \dots \mid \beta_n$$

- Thay thế các luật sinh trên bằng các luật sinh sau:

$$A \rightarrow \beta_1 A' \mid \beta_2 A' \mid \dots \mid \beta_n A'$$

$$A' \rightarrow \alpha_1 A' \mid \alpha_2 A' \mid \dots \mid \alpha_m A' \mid \epsilon$$

- Ví dụ:

Thay thế luật sinh  $A \rightarrow A + T \mid T$  bởi

$$A \rightarrow TA'$$

$$A' \rightarrow +TA' \mid \epsilon$$

# 1. Loại bỏ tính đệ quy

Loại bỏ đệ quy trái trực tiếp:

- Ví dụ:

$$E \rightarrow E + T \mid T$$

$$T \rightarrow T * F \mid F$$

$$F \rightarrow (E) \mid \text{id}$$

Khử đệ quy trái cho văn phạm trên:

$$E \rightarrow TE'$$

$$E' \rightarrow +TE' \mid \varepsilon$$

$$T \rightarrow FT'$$

$$T' \rightarrow *FT' \mid \varepsilon$$

$$F \rightarrow (E) \mid \text{id}$$

## 2. Tính First

- Định nghĩa:
  - Nếu  $\alpha$  là một xâu thì  $\text{First}(\alpha)$  là tập hợp các ký hiệu kết thúc mà nó bắt đầu một chuỗi dẫn xuất từ  $\alpha$
  - Nếu  $\alpha \rightarrow^* \epsilon$  thì  $\epsilon$  thuộc  $\text{First}(\alpha)$
- Cách tính First:
  1. Nếu  $X$  là kí hiệu kết thúc thì  $\text{First}(X)$  là  $\{X\}$ .
  2. Nếu  $X \rightarrow \epsilon$  là một luật sinh thì thêm  $\epsilon$  vào  $\text{First}(X)$ .
  3. Nếu  $X \rightarrow Y_1 Y_2 Y_3 \dots Y_k$  là một luật sinh thì:
    - Thêm  $\text{First}(Y_1) - \{\epsilon\}$  vào  $\text{First}(X)$ . Dừng nếu  $\epsilon \notin \text{First}(Y_1)$
    - Thêm  $\text{First}(Y_2) - \{\epsilon\}$  vào  $\text{First}(X)$ . Dừng nếu  $\epsilon \notin \text{First}(Y_2)$
    - .....
    - Thêm  $\text{First}(Y_k) - \{\epsilon\}$  vào  $\text{First}(X)$ . Dừng nếu  $\epsilon \notin \text{First}(Y_k)$
    - Thêm  $\{\epsilon\}$  vào  $\text{First}(X)$



## 2. Tính First

Ví dụ về cách tính First:

- Cho văn phạm:

$$E \rightarrow TE'$$

$$E' \rightarrow +TE' \mid \varepsilon$$

$$T \rightarrow FT'$$

$$T' \rightarrow *FT' \mid \varepsilon$$

$$F \rightarrow (E) \mid \text{id}$$

- Tính First:

$$\text{First}( ( ) ) = \{ ( ) \}$$

$$\text{First}( ) ) = \{ ) \}$$

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

$$\text{First}( + ) = \{ + \}$$

$$\text{First}( * ) = \{ * \}$$

$$\text{First}( T ) = \{ (, \text{id} \}$$

$$\text{First}( E ) = \{ (, \text{id} \}$$

$$\text{First}( E' ) = \{ +, \varepsilon \}$$

$$\text{First}( T' ) = \{ *, \varepsilon \}$$

$$\text{First}( F ) = \{ (, \text{id} \}$$

- Nếu  $X$  là kí hiệu kết thúc thì  $\text{First}(X)$  là  $\{X\}$ .
- Nếu  $X \rightarrow \varepsilon$  là một luật sinh thì thêm  $\varepsilon$  vào  $\text{First}(X)$ .
- Nếu  $X \rightarrow Y_1 Y_2 Y_3 \dots Y_k$  là một luật sinh thì:
  - Thêm  $\text{First}(Y_1) - \{\varepsilon\}$  vào  $\text{First}(X)$ . Dừng nếu  $\varepsilon \notin \text{First}(Y_1)$
  - Thêm  $\text{First}(Y_2) - \{\varepsilon\}$  vào  $\text{First}(X)$ . Dừng nếu  $\varepsilon \notin \text{First}(Y_2)$
  - .....
  - Thêm  $\text{First}(Y_k) - \{\varepsilon\}$  vào  $\text{First}(X)$ . Dừng nếu  $\varepsilon \notin \text{First}(Y_k)$
  - Thêm  $\{\varepsilon\}$  vào  $\text{First}(X)$

# 3. Tính Follow

- Định nghĩa:
  - Nếu  $A$  là một ký hiệu chưa kết thúc thì  $\text{Follow}(A)$  là tập các ký hiệu kết thúc mà nó xuất hiện ngay bên phải  $A$  trong một dạng câu ( $S \rightarrow^* \alpha A a \beta$ )
  - Nếu  $A$  là ký hiệu phải nhất trong một dạng câu ( $S \rightarrow^* \alpha A$ ) thì  $\$$  thuộc  $\text{Follow}(A)$

# 3. Tính Follow

- Cách tính Follow:
  - Tính First() của tất cả các ký hiệu của văn phạm.
  - Áp dụng các quy tắc sau cho đến khi không thể thêm được gì vào mọi tập Follow được nữa:
    - 1) Đặt \$ vào Follow(S), với S là ký hiệu bắt đầu của văn phạm.
    - 2) Nếu  $A \rightarrow \alpha B \beta$  thì thêm  $\text{First}(\beta) \setminus \varepsilon$  vào Follow(B)
    - 3) Nếu 
$$\left[ \begin{array}{l} A \rightarrow \alpha B \\ \left\{ \begin{array}{l} A \rightarrow \alpha B \beta \\ \beta \rightarrow * \varepsilon \end{array} \right. \end{array} \right.$$
 thì thêm Follow(A) vào Follow(B)

# 3. Tính Follow

- Ví dụ:

❖ Văn phạm

$$E \rightarrow TE'$$

$$T \rightarrow FT'$$

$$F \rightarrow (E) \mid \text{id}$$

$$E' \rightarrow +TE' \mid \varepsilon$$

$$T' \rightarrow *FT' \mid \varepsilon$$

❖ Tính Follow

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

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

$$\text{Follow}(F) = \{*, +, \$, )\}$$

# Phân tích cú pháp đoán trước

- ❖ Để dùng được kỹ thuật phân tích cú pháp đoán trước, văn phạm cần có các tính chất sau: với mỗi cặp luật sinh  $A \rightarrow \alpha \mid \beta$
1. Không có ký hiệu kết thúc  $a$  nào mà cả  $\alpha$  và  $\beta$  đều dẫn xuất ra các chuỗi bắt đầu bằng  $a$ .
  2. Chỉ  $\alpha$  hoặc chỉ  $\beta$  dẫn xuất ra  $\epsilon$ .  
( $\rightarrow \text{First}(\alpha) \cap \text{First}(\beta) = \Phi$ )
  3. Nếu  $\beta \Rightarrow^* \epsilon$  thì  $\alpha$  không dẫn xuất được chuỗi nào bắt đầu bằng một ký hiệu kết thúc thuộc tập  $\text{Follow}(A)$   
( $\rightarrow$  Nếu  $\epsilon$  thuộc  $\text{First}(\beta)$  thì  $\text{First}(\alpha) \cap \text{Follow}(A) = \Phi$ ,  
tương tự nếu  $\epsilon$  thuộc  $\text{First}(\alpha)$  )
- $\rightarrow$  Văn phạm có 3 tính chất trên gọi là văn phạm LL(1)

# Văn phạm LL(1)

- **L (Left to right parse)**: mô tả hành động quét chuỗi nhập từ trái sang phải.
- **L (Leftmost derivation)**: biểu thị việc sinh ra một dẫn xuất trái nhất cho chuỗi nhập.
- **1 (1 symbol lookahead)**: tại mỗi bước đầu đọc chỉ đọc trước được 1 token để thực hiện các quyết định phân tích cú pháp.

## **Xây dựng Parsing table M cho phân tích cú pháp đoán trước không đệ quy**

- Bước 1: Loại bỏ tính đệ quy trái
- Bước 2: Tính tập First
- Bước 3: Tính tập Follow
- Bước 4: Xây dựng Parsing table M

## 4. Xây dựng Parsing table M

- ❖ Với mỗi luật sinh  $A \rightarrow \alpha$  của văn phạm, thực hiện các bước:
  1. Với mỗi ký hiệu kết thúc  $a \in \text{FIRST}(\alpha)$ , thêm  $A \rightarrow \alpha$  vào  $M[A, a]$
  2. Nếu  $\epsilon \in \text{FIRST}(\alpha)$ :
    - Nếu  $\epsilon \in \text{FIRST}(\alpha)$  thì đưa luật sinh  $A \rightarrow \alpha$  vào  $M[A, b]$  với mỗi ký hiệu kết thúc  $b \in \text{FOLLOW}(A)$
    - Nếu  $\epsilon \in \text{FIRST}(\alpha)$  và  $\$ \in \text{FOLLOW}(A)$  thì đưa luật sinh  $A \rightarrow \alpha$  vào  $M[A, \$]$
- ❖ Ô còn trống trong bảng tương ứng với lỗi (error).



# 5. Bài tập

- Cho văn phạm:

$$E \rightarrow TE'$$

$$E' \rightarrow +TE' \mid \varepsilon$$

$$T \rightarrow FT'$$

$$T' \rightarrow *FT' \mid \varepsilon$$

$$F \rightarrow (E) \mid \text{id}$$

Hãy xây dựng bảng phân tích cú pháp đoán trước không đệ quy.

Non-terminal	Input symbol					
	id	+	*	(	)	\$
E	$E \rightarrow TE'$			$E \rightarrow TE'$		
E'		$E' \rightarrow +TE'$			$E' \rightarrow \varepsilon$	$E' \rightarrow \varepsilon$
T	$T \rightarrow FT'$			$T \rightarrow FT'$		
T'		$T' \rightarrow \varepsilon$	$T' \rightarrow *FT'$		$T' \rightarrow \varepsilon$	$T' \rightarrow \varepsilon$
F	$F \rightarrow \text{id}$			$F \rightarrow (E)$		

Thank you for  
watching