

LÝ THUYẾT TÍNH TOÁN

BÀI 6: VĂN PHẠM PHI NGỮ CẢNH

Phạm Xuân Cường
Khoa Công nghệ thông tin
cuongpx@tlu.edu.vn

1. Khái niệm
2. Định nghĩa hình thức
3. Văn phạm nhập nhằng
4. Dạng chuẩn tắc Chomsky

Khái niệm

- Văn phạm phi ngữ cảnh = Context-free Grammar (CFG)
- **CFG**: Là một phương pháp mạnh hơn để mô tả ngôn ngữ
- Ứng dụng:
 - Bộ biên dịch trong các ngôn ngữ lập trình
 - Bộ phân tích trong các trình biên dịch và thông dịch
- Ví dụ:
$$E \rightarrow E + T \mid T$$
$$T \rightarrow T \times F \mid F$$
$$F \rightarrow (E) \mid a$$

Một văn phạm gồm có:

- Tập các **quy tắc thay thế** \equiv các **sản xuất**
- Mỗi quy tắc là một dòng bao gồm 1 ký hiệu và 1 xâu được ngăn cách bởi dấu mũi tên
- Ký hiệu \equiv **biến** \equiv Các ký hiệu in hoa
- Ký hiệu kết thúc \equiv Các ký hiệu in thường, số hoặc ký tự đặc biệt
- **Biến ban đầu** thường xuất hiện bên trái của quy tắc trên cùng

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

$$T \rightarrow T \times F \mid F$$

$$F \rightarrow (E) \mid a$$

- Dẫn xuất

$$\begin{aligned} E &\Rightarrow E + T \Rightarrow T + T \Rightarrow F + T \Rightarrow a + T \\ &\Rightarrow a + F \Rightarrow a + (E) \Rightarrow a + (T) \\ &\Rightarrow a + (T \times F) \Rightarrow a + (F \times F) \\ &\Rightarrow a + (a \times F) = a + (a \times a) \end{aligned}$$

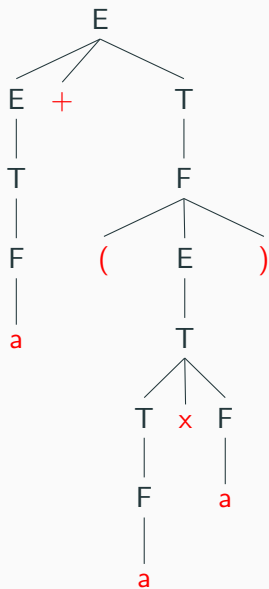
Cũng có thể viết:

$$E \xRightarrow{*} a + (a \times a)$$

$$E \xRightarrow{*} a + (E) \Rightarrow a + (T) \xRightarrow{*} a + (a \times a)$$

- Dẫn xuất trái nhất: Luôn lựa chọn dẫn xuất ở bên trái
 $\dots \Rightarrow F + T \Rightarrow a + T \Rightarrow \dots a + (a \times a)$
- Dẫn xuất phải nhất: Luôn lựa chọn dẫn xuất ở bên phải
 $\dots \Rightarrow F + T \Rightarrow F + F \Rightarrow \dots a + (a \times a)$

Cây dẫn xuất



Định nghĩa hình thức

$$\text{CFG: } \mathbf{G} = (\mathbf{V}, \Sigma, \mathbf{R}, \mathbf{S})$$

Trong đó:

- \mathbf{V} là tập hữu hạn gồm các **biến**
- Σ là tập hữu hạn các ký hiệu kết thúc $\Sigma \cap \mathbf{V}$
- \mathbf{R} tập các quy tắc
- \mathbf{S} biến bắt đầu

Định nghĩa 1

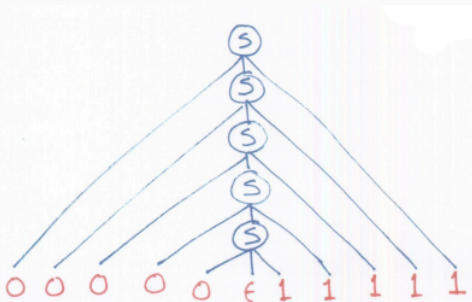
Ngôn ngữ của văn phạm là $\{w | w \in \Sigma^* \text{ và } S \xRightarrow{*} w\}$

Định nghĩa 2

Một **ngôn ngữ phi ngữ cảnh** (CFL) là ngôn ngữ được tạo ra bởi một **văn phạm phi ngữ cảnh** (CFG)

Ví dụ CFL

- $S \rightarrow (S) | SS | \varepsilon$
 $A = \{\varepsilon, (), ()(), (())(), \dots\}$
- Ngôn ngữ $B = \{0^n 1^n \mid n \geq 0\}$
 $S \rightarrow \varepsilon$
 $S \rightarrow 0S1$



Ví dụ

Cho CFG sau:

$E \rightarrow E + E$

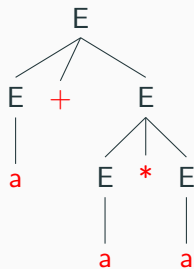
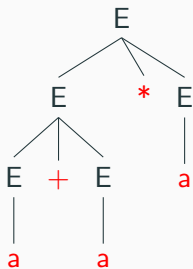
$\rightarrow E \times E$

$\rightarrow (E)$

$\rightarrow a$

$a+a*a$

→ Mỗi cây dẫn xuất đều có duy nhất một cây dẫn xuất trái nhất và duy nhất một cây dẫn xuất phải nhất



Văn phạm nhập nhằng

Chuỗi nhập nhằng:

- Có nhiều hơn 2 cây dẫn xuất \Leftrightarrow Có nhiều cách để tạo ra chuỗi đó

Văn phạm nhập nhằng:

- Một văn phạm là nhập nhằng nếu một vài chuỗi có thể được sinh ra bởi nhiều cách

Văn phạm nhập nhằng:

$$E \rightarrow E + E$$

$$\rightarrow E \times E$$

$$\rightarrow (E)$$

$$\rightarrow a$$

Văn phạm không nhập nhằng:

$$E \rightarrow E + T$$

$$\rightarrow T$$

$$T \rightarrow T \times F$$

$$\rightarrow F$$

$$F \rightarrow (E)$$

$$\rightarrow a$$

Định lý 1

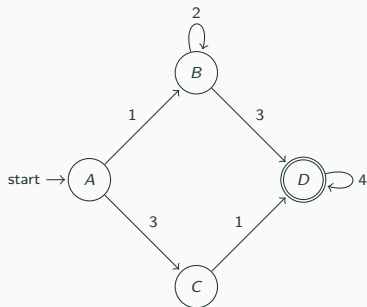
Mọi ngôn ngữ chính quy đều là phi ngữ cảnh

Chứng minh

Ý TƯỞNG: Cho một DFA, xây dựng một văn phạm có thể tạo ra cùng 1 ngôn ngữ với DFA

- Chuyển các trạng thái thành các biến
- Chuyển trạng thái bắt đầu thành biến bắt đầu
- Chuyển các cạnh thành các quy tắc
- Thêm 1 quy tắc ϵ cho mỗi trạng thái kết thúc

Ví dụ



$A \rightarrow 1B$

$A \rightarrow 3C$

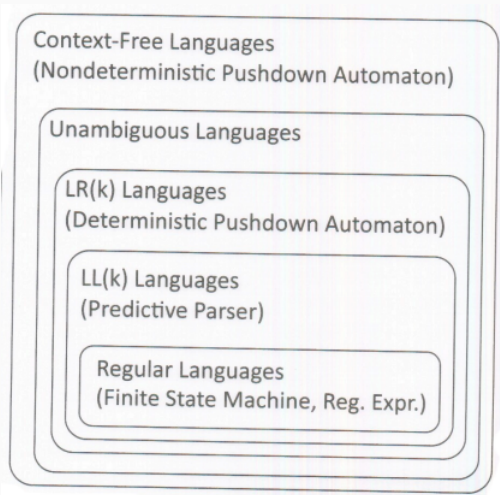
$B \rightarrow 2B$

$B \rightarrow 3D$

$C \rightarrow 1D$

$D \rightarrow 4D$

$D \rightarrow \epsilon$



Dạng chuẩn tắc Chomsky

Dạng chuẩn tắc Chomsky

Định nghĩa

Một văn phạm phi ngữ cảnh ở dạng chuẩn tắc Chomsky nếu tất cả các quy tắc của nó có dạng:

$A \rightarrow BC$

$A \rightarrow a$

Trong đó,

- a là một ký hiệu kết thúc
- A, B, C là các biến bất kỳ, B, C không thể là biến bắt đầu

Ngoài ra ta có thêm quy tắc: $S \rightarrow \epsilon$ với S là biến bắt đầu

Định lý 2

Mọi ngôn ngữ phi ngữ cảnh nào cũng được sinh ra bởi một văn phạm phi ngữ cảnh ở dạng chuẩn tắc Chomsky

Chứng minh định lý 2

Chứng minh định lý 2:

Với mọi CFG ta chuyển chúng về dạng chuẩn tắc Chomsky

- Bước 1: Đảm bảo rằng biến bắt đầu không xuất hiện bên phải bên phải của quy tắc \Leftrightarrow Thêm một biến bắt đầu mới
- Bước 2: Loại bỏ các quy tắc có dạng $A \rightarrow \varepsilon$
- Bước 3: Khử tất các các quy tắc đơn vị $A \rightarrow B$
- Bước 4: Loại bỏ các quy tắc có nhiều hơn 2 biến ở phần bên phải

$$A \rightarrow BCDE$$

$$A \rightarrow Bcde$$

- Bước 5: Đảm bảo rằng chỉ còn tồn tại các quy tắc có dạng sau:

$$A \rightarrow BC$$

$$A \rightarrow a$$

Cho văn phạm sau:

$$S \rightarrow ASA \mid aB$$

$$A \rightarrow B \mid S$$

$$B \rightarrow b \mid \epsilon$$

Hãy chuyển về dạng chuẩn tắc Chomsky

- Bước 1: Thêm biến bắt đầu mới

$$S_0 \rightarrow S$$

$$S \rightarrow ASA \mid aB$$

$$A \rightarrow B \mid S$$

$$B \rightarrow b \mid \varepsilon$$

- Bước 2: Loại bỏ các quy tắc $A \rightarrow \epsilon$

$$S_0 \rightarrow S$$

$$S \rightarrow ASA \mid aB$$

$$A \rightarrow B \mid S$$

$$B \rightarrow b \mid \epsilon$$

$$S_0 \rightarrow S$$

$$S \rightarrow ASA \mid aB \mid a$$

$$A \rightarrow B \mid S \mid \epsilon$$

$$B \rightarrow b$$

$$S_0 \rightarrow S$$

$$S \rightarrow ASA \mid aB \mid a$$

$$A \rightarrow B \mid S \mid \epsilon$$

$$B \rightarrow b$$

$$S_0 \rightarrow S$$

$$S \rightarrow ASA \mid aB \mid a \mid SA \mid AS \mid S$$

$$A \rightarrow B \mid S$$

$$B \rightarrow b$$

- Bước 3: Khử tất các các quy tắc đơn vị $A \rightarrow B$

$S_0 \rightarrow S$

$S \rightarrow ASA \mid aB \mid a \mid SA \mid AS \mid S$

$A \rightarrow B \mid S$

$B \rightarrow b$

Loại bỏ $S \rightarrow S$

$S_0 \rightarrow S$

$S \rightarrow ASA \mid aB \mid a \mid SA \mid AS$

$A \rightarrow B \mid S$

$B \rightarrow b$

Loại bỏ $S_0 \rightarrow S$

$S_0 \rightarrow ASA \mid aB \mid a \mid SA \mid AS$

$S \rightarrow ASA \mid aB \mid a \mid SA \mid AS$

$A \rightarrow B \mid S$

$B \rightarrow b$

- Bước 3: Tiếp

Ta có:

$S_0 \rightarrow ASA \mid aB \mid a \mid SA \mid AS$

$S \rightarrow ASA \mid aB \mid a \mid SA \mid AS$

$A \rightarrow B \mid S$

$B \rightarrow b$

Loại bỏ $A \rightarrow S$

$S_0 \rightarrow ASA \mid aB \mid a \mid SA \mid AS$

$S \rightarrow ASA \mid aB \mid a \mid SA \mid AS$

$A \rightarrow b \mid ASA \mid aB \mid a \mid SA \mid AS$

$B \rightarrow b$

Loại bỏ $A \rightarrow B$

$S_0 \rightarrow S$

$S \rightarrow ASA \mid aB \mid a \mid SA \mid AS$

$A \rightarrow b \mid S$

$B \rightarrow b$

- Bước 4: Loại bỏ các quy tắc có nhiều hơn 2 biến ở phần bên phải. Ví dụ:

Thay $A \rightarrow BCDE$ bằng

$A \rightarrow BA_1$

$A_1 \rightarrow CA_2$

$A_2 \rightarrow DE$

$S_0 \rightarrow A**SA** \mid aB \mid a \mid SA \mid AS$

$S \rightarrow A**SA** \mid aB \mid a \mid SA \mid AS$

$A \rightarrow b|A**SA** \mid aB \mid a \mid SA \mid AS$

$B \rightarrow b$

$S_0 \rightarrow AA_1 \mid aB \mid a \mid SA \mid AS$

$S \rightarrow AA_1 \mid aB \mid a \mid SA \mid AS$

$A \rightarrow b|AA_1 \mid aB \mid a \mid SA \mid AS$

$A_1 \rightarrow SA$

$B \rightarrow b$

- Bước 5: Thay thế $A \rightarrow bC$ bằng $A \rightarrow A_1C$ và $A_1 \rightarrow b$

$S_0 \rightarrow AA_1 \mid aB \mid a \mid SA \mid AS$
 $S \rightarrow AA_1 \mid aB \mid a \mid SA \mid AS$
 $A \rightarrow b \mid AA_1 \mid aB \mid a \mid SA \mid AS$
 $A_1 \rightarrow SA$
 $B \rightarrow b$

Thêm quy tắc $A_2 \rightarrow a$

$S_0 \rightarrow AA_1 \mid A_2B \mid a \mid SA \mid AS$
 $S \rightarrow AA_1 \mid A_2B \mid a \mid SA \mid AS$
 $A \rightarrow b \mid AA_1 \mid A_2B \mid a \mid SA \mid AS$
 $A_1 \rightarrow SA$
 $A_2 \rightarrow a$
 $B \rightarrow b$

→ Đây là dạng chuẩn tắc Chomsky

Questions?