



TRƯỜNG ĐẠI HỌC THỦY LỢI
Khoa Công nghệ thông tin - Bộ môn Khoa học máy tính

LÝ THUYẾT TÍNH TOÁN

Tên giảng viên: Đinh Phú Hùng

Email: hungdp@tlu.edu.vn

Điện thoại: 0912509973

1. Khái niệm
2. Bộ đề Bơm
3. Tổng kết chương 1

Khái niệm

- **Ngôn ngữ chính quy:** Ngôn ngữ được đoán nhận bởi một DFA nào đó

→ **Ngôn ngữ không chính quy** là gì?

Ví dụ: Xét các ngôn ngữ sau trên bộ chữ $\Sigma = \{0,1\}$ là **chính quy** hay **không chính quy**

$$B = \{0^n 1^n \mid n \geq 0\}$$

$$C = \{w \mid w \text{ có số ký hiệu } 0 \text{ bằng số ký hiệu } 1\}$$

$$D = \{w \mid w \text{ có số lần xuất hiện xâu con } 01 \text{ và } 10 \text{ là bằng nhau}\}$$

- **Ngôn ngữ chính quy:** Ngôn ngữ được đoán nhận bởi một DFA nào đó

→ **Ngôn ngữ không chính quy** là gì?

Ví dụ: Xét các ngôn ngữ sau trên bộ chữ $\Sigma = \{0,1\}$

$B = \{0^n 1^n | n \geq 0\} \rightarrow$ **Không chính quy**

$C = \{w | w \text{ có số ký hiệu } 0 \text{ bằng số ký hiệu } 1\} \rightarrow$ **Không chính quy**

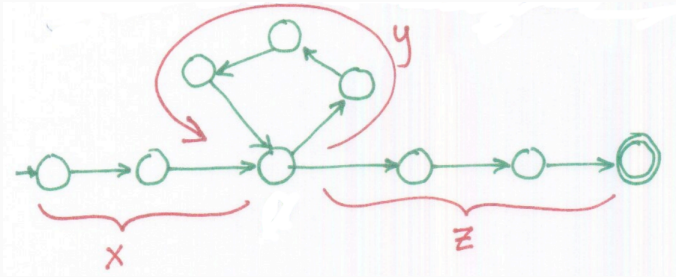
$D = \{w | w \text{ có số lần xuất hiện xâu con } 01 \text{ và } 10 \text{ là bằng nhau}\} \rightarrow$ **Chính quy**

→ Làm sao để chứng minh một ngôn ngữ là không chính quy?

- Hãy tưởng tượng một FSM có thể tạo ra các chuỗi rất dài
Ví dụ: Một DFA có $|Q| = 5$
Làm sao để tạo ra một chuỗi dài \rightarrow Đi theo chu trình
Nếu không theo chu trình thì chuỗi dài nhất được sinh ra là bao nhiêu?
 $\rightarrow |s| \leq 5$
- Tất cả các chuỗi ≥ 5 đều phải đi theo một chu trình nào đó
 - Nếu ta có thể đi theo một chu trình **n lần** thì chuỗi được sinh ra đó sẽ nằm trong ngôn ngữ mà FSM đó đoán nhận
 - Nếu ta bỏ qua chu trình đó thì chuỗi được sinh ra vẫn sẽ nằm trong ngôn ngữ mà FSM đó đoán nhận

Ví dụ

Xét một FSM sau:



→ Tất cả các chuỗi s được sinh ra có dạng $s = xy^iz$ đều thuộc ngôn ngữ A mà máy FSM đoán nhận

Độ dài dẫn xuất

- Nếu A là ngôn ngữ chính quy và s là một xâu đủ dài thuộc A ($|s| \geq p$) thì s có thể được viết như sau: $s = xyz$
- p được gọi là **độ dài dẫn xuất** (pumping length)
- Tất cả các ngôn ngữ chính quy có một thuộc tính đặc biệt
Nếu ngôn ngữ không có thuộc tính này \rightarrow Là ngôn ngữ không chính quy

Bổ đề Bơ

Bổ đề Bơm (Pumping Lemma)

Nếu A là một ngôn ngữ chính quy, thì tồn tại một số p sao cho nếu s là một chuỗi bất kỳ thuộc A có độ dài ít nhất là p , thì s có thể được chia ra làm 3 phần $s=xyz$ thỏa mãn các điều kiện sau:

1. $xy^iz \in A \ \forall i \geq 0$
2. $|y| > 0$
3. $|xy| \leq p$

- Sử dụng bổ đề Bơm để chứng minh một ngôn ngữ A là không chính quy

Ý TƯỞNG: (Chứng minh bằng phản chứng)

- Giả sử A là chính quy
- Nó có một độ dài dẫn xuất p
- Tất cả các xâu trong A có độ dài lớn hơn p ($|s| \geq p$) có thể chia làm 3 đoạn $s = xyz$
- Chọn 1 xâu như vậy trong A
- Chia nó làm 3 đoạn xyz
- Chỉ ra rằng $xy^iz \notin A$ bằng cách
 - Xét tất cả các trường hợp mà s có thể chia thành 3 đoạn
 - Chỉ ra rằng không có trường hợp nào thỏa mãn 3 điều kiện của bổ đề Bơm

→ Mâu thuẫn, do đó kết luận A không phải là chính quy

Ví dụ 1

Cho ngôn ngữ $B = \{0^n 1^n \mid n \geq 0\}$

Hãy chứng minh ngôn ngữ B là không chính quy

Chứng minh:

- Giả sử B là chính quy $\rightarrow B$ có một độ dài dẫn xuất p
- Xâu chúng ta lựa chọn để chỉ ra phản chứng là: $s = 0^p 1^p$
- Xét các trường hợp có thể chia s thành 3 đoạn xyz
 - y nằm trong phần chuỗi 0
 - y nằm trong phần chuỗi 1
 - y nằm trong cả phần chuỗi 0 và chuỗi 1

The diagram shows the string $0^7 1^7$ written in green. Below the string, three red curly braces are used to partition it into three segments labeled x , y , and z from left to right. Segment x contains 4 zeros, segment y contains 3 zeros, and segment z contains 7 ones. This illustrates a case where the pumping length $p=7$ is used to partition the string.

Ví dụ 1

- Xét TH 1: 000011111 $\rightarrow xy^2z = 0000|000011111$
Xâu xy^2z có thuộc B hay không?

Ví dụ 1

- Xét TH 1: 000011111 $\rightarrow xy^2z = 0000|000011111$
Xâu xy^2z có thuộc B hay không? $\rightarrow xy^2z \notin B$
- Tương tự, TH 2: 0000011111 $\rightarrow xy^2z = 000001111|1111 \notin B$
- TH3: 0000011111 $\rightarrow xy^2z = 0000011|0011111 \notin B$
- Ngoài ra theo điều kiện 3:
 - TH1: $|xy| = |0000| = 4 \leq p = 5 \rightarrow \text{True}$
 - TH2: $|xy| = |000001111| = 9 \leq p = 5 \rightarrow \text{False}$
 - TH3: $|xy| = |0000011| = 7 \leq p = 5 \rightarrow \text{False}$
- Có các mâu thuẫn nên giả thiết là sai $\rightarrow B$ là ngôn ngữ không chính quy

- Cho ngôn ngữ $C = \{w \mid w \text{ có số ký hiệu 0 bằng số ký hiệu 1}\}$
 $= \{ww \mid w \in \{0,1\}^*\}$

Hãy chứng minh ngôn ngữ C là không chính quy

- Bài 1.29, 1.46 - 1.49 Sách giáo trình

Tổng kết chương 1

Tổng kết chương 1

- Ngôn ngữ chính quy được đoán nhận bởi ????
- DFA \Leftrightarrow NFA ????
- Biểu thức chính quy biểu diễn ????
- Thế nào là ngôn ngữ không chính quy ????

Questions?