Lecture 6. Regular Languages (Ngôn ngữ chính quy)

Hà Chí Trung, BM KHMT, KCNTT, HVKTQS

hct2009@yahoo.com

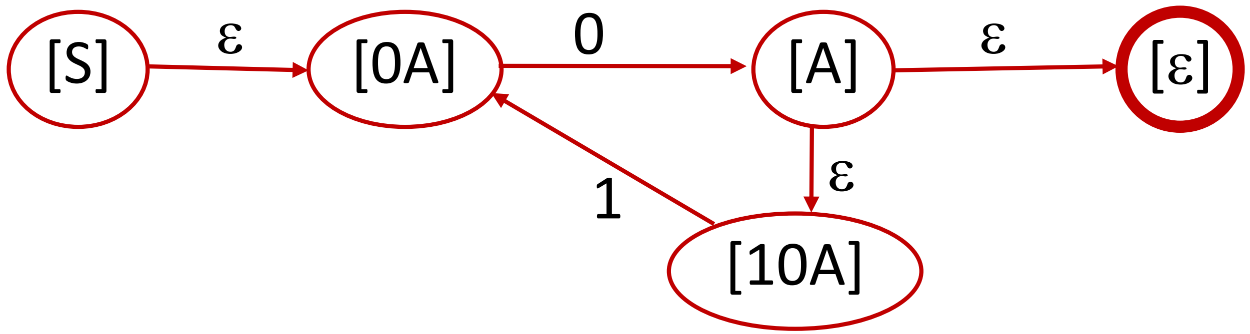
01685-582-102

Sự tương đương giữa RG và FA

- **Định lý:** Tập rỗng, tập tổng quát, tập hữu hạn các chuỗi được xây trên bảng chữ cái hữu hạn là tập hợp chính quy.
- *Chứng minh:*
  - ❖Cách 1: Dựa trên khái niệm biểu thức chính quy đã học trong bài trước.
  - ❖Cách 2: Giải thuật xây dựng **FA** cho tập hữu hạn.
- **Định lý:** Nếu L được sinh ra từ một văn phạm chính quy thì L là tập hợp chính quy.
  - ❖Ý nghĩa: như vậy, một RG có thể được biểu diễn bởi một FA.
  - ❖Chứng minh: Chỉ ra giải thuật xây dựng FA tương đương.

Giải thuật biến đổi từ RG sang FA

- **Giải thuật** xây dựng **FA** cho **văn phạm tuyến tính phải**:
  1) Xây dựng tập Q gồm các trạng thái có dạng [α] với α là S hoặc chuỗi hậu tố của vế phải một luật sinh nào đó trong P.
  2) Nếu A là một biến và $(A \rightarrow \alpha) \in P$: **$\delta$([A], ε) = {[α]}**
  3) Nếu a là một ký hiệu kết thúc: **δ([aα], a) = {[α]}**
  4) Trạng thái bắt đầu **[S]**, trạng thái kết thúc **[ε]**



Giải thuật biến đổi từ RG sang FA

- **Giải thuật** xây dựng **FA** cho **văn phạm tuyến tính trái** được xây dựng dựa trên tính chất sau:
  - ❖ Cho văn phạm G = < Σ, Δ, S, P >, và G' = < Σ, Δ, S, P' >, trong đó nếu:

$$P' = \{ A \rightarrow \alpha \mid A \rightarrow \alpha^R \in P \},$$

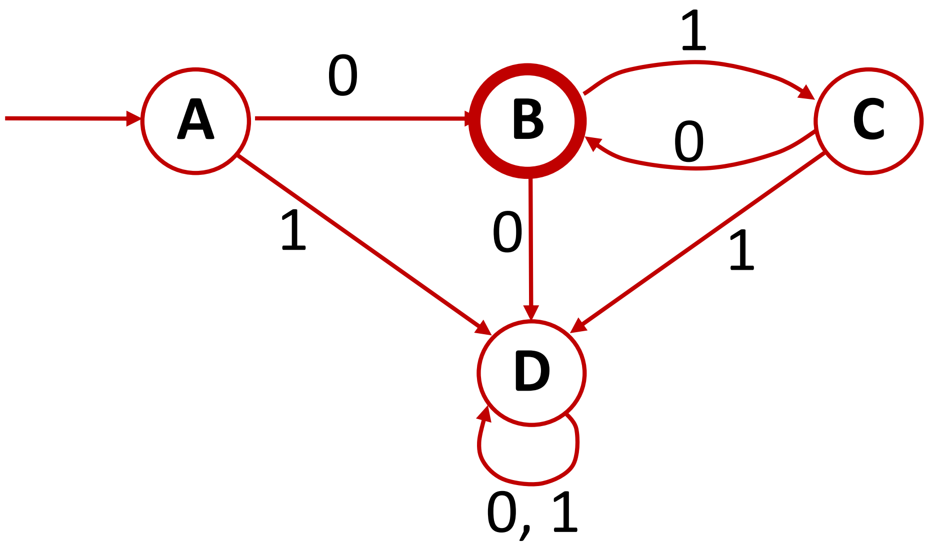
    thì $L(G')^R = L(G)$.
- Như vậy ta có thể xây dựng giải thuật theo 3 bước:
  - ❖ Xác định văn phạm tuyến tính phải G' = < Σ, Δ, S, P' >
  - ❖ Xây dựng NFA cho G';
  - ❖ Đảo ngược chiều các cạnh của NFA này, vị trí kết thúc trở thành vị trí bắt đầu và ngược lại.

Giải thuật biến đổi từ FA sang RG

- **Định lý:** Nếu L là một tập hợp chính quy thì L được sinh ra từ một RG (văn phạm tuyến tính trái hoặc tuyến tính phải) nào đó.
- *Ý nghĩa*: một FA có thể được biểu diễn bởi một RG.
- **Giải thuật xây dựng RG tuyến tính phải cho FA**:
- **Cách 1: xét hàm chuyển trạng thái δ:**
  - Nếu $\boldsymbol{\delta(p, a) = q}$, ta có luật sinh: $\boldsymbol{p \rightarrow aq}$
  - Nếu q là trạng thái kết thúc, ta có luật sinh $\boldsymbol{p \rightarrow a}$
  - Nếu $\boldsymbol{q_0}$ là trạng thái kết thúc, thêm vào: $\boldsymbol{S \rightarrow q_0 \mid \varepsilon}$
- **Cách 2:**
  - Nếu $\boldsymbol{\delta(p, a) = q}$, ta có luật sinh: $\boldsymbol{p \rightarrow aq}$
  - Nếu q là trạng thái kết thúc, ta có luật sinh $\boldsymbol{q \rightarrow \varepsilon}$

Giải thuật biến đổi từ FA sang RG

Ví dụ: xét DFA cho **0(10)*** sau:



$A \rightarrow 0B \mid 1D \mid 0$
$B \rightarrow 0D \mid 1C$
$C \rightarrow 0B \mid 1D \mid 0$
$D \rightarrow 0D \mid 1D$

Do D không có ích nên có thể rút gọn G như sau:

$A \rightarrow 0B \mid 0$
$B \rightarrow 1C$
$C \rightarrow 0B \mid 0$

Giải thuật biến đổi từ FA sang RG

- **Giải thuật xây dựng RG tuyến tính trái cho FA:**
  - Bắt đầu với một NFA cho L^R
  - Đảo ngược chuỗi vế phải cho tất cả mọi luật sinh của văn phạm vừa thu được

Bổ đề bơm cho tập hợp chính quy

- Pumping lemma: nếu L là **RS** thì có tồn tại hằng số **n** sao cho nếu **z** là một từ bất kỳ thuộc L và **|z| ≥ n** thì ta có thể viết z dưới dạng:

  z=uvw với |uv| ≤ n, |v| ≥ 1, và ∀i ≥ 0 ta có **$uv^iw \in L$**

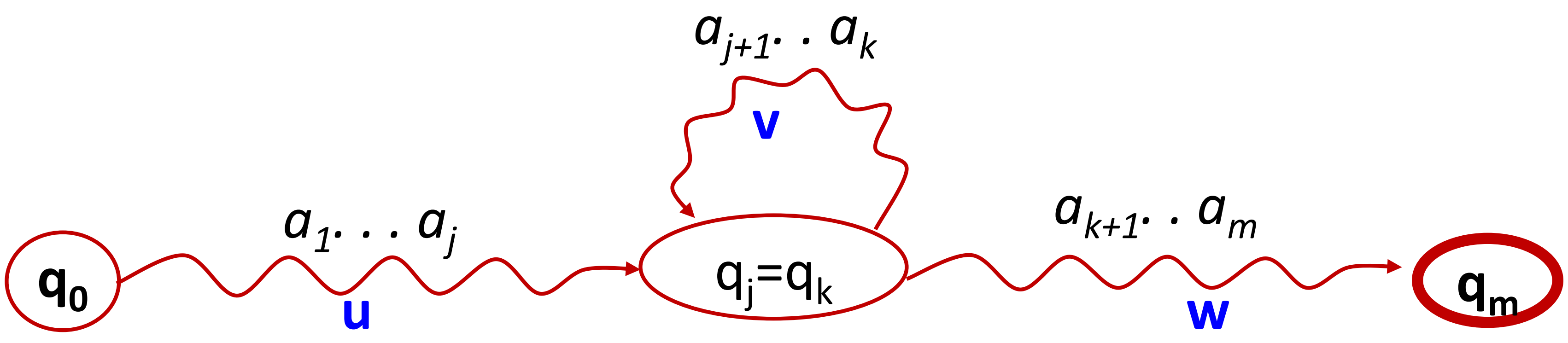
- **Cách khác:** (∀L)(∃n)(∀z)[z thuộc L và | z | ≥ n ta có (∃u, v, w)(z = uvw, | uv | ≤ n, | v | ≥ 1 và (∀i)($uv^iw \in L$))]
- **Chứng minh:** Giả sử L là ngôn ngữ chính quy → tồn tại DFA A=(Q, Σ, δ, q_0, F) có n trạng thái chấp nhận L.
  - Xét chuỗi nhập z = $a_1a_2 \ldots a_m$, m ≥ n. Với mỗi i=1,2,…,m, ta đặt $\delta(q_0, a_1a_2 \ldots a_i) = q_i$, theo nguyên lý Dirichlet, phải có ít nhất 2 trạng thái nào đó trùng nhau.
  - Giả sử đó là hai số nguyên j và k sao cho 0 ≤ j < k ≤ n thỏa mãn $q_j = q_k$.

Bổ đề bơm cho tập hợp chính quy

- $z \in L \rightarrow q_m \in \mathbf{F} \rightarrow a_1 \ldots a_j a_{k+1} \ldots a_m \in L(A) \rightarrow$

  $\mathbf{a_1 \ldots a_j (a_{j+1} \ldots a_k)^i a_{k+1} \ldots a_m \in L(A)}$, với $i \geq 0$

- Vì $j < k$ nên chuỗi $a_{j+1} \ldots a_k$ có độ dài ít nhất bằng 1 và vì $k \leq n$ nên $k-j \leq n$. Chuỗi đó tạo thành một **vòng lặp**:



- Vòng lặp trong hình trên có thể được lặp lại số lần tùy ý, do đó chuỗi $a_1 \ldots a_j (a_{j+1} \ldots a_k)^i a_{k+1} \ldots a_m \in L(M)$, $\forall i \geq 0$.

Bổ đề bơm cho tập hợp chính quy

- **Ứng dụng của bổ đề bơm:** dùng để chứng tỏ một tập hợp không là **RS** theo phương pháp **phản chứng**:
  - Chọn L mà bạn cần chứng tỏ đó không là RL.
  - Chọn hằng số n, hằng số được đề cập đến trong bổ đề bơm.
  - Chọn chuỗi $z \in L$. Chuỗi z phụ thuộc vào hằng số n.
  - Giả thiết phân chuỗi z thành các chuỗi con u, v, w theo ràng buộc $|uv| \leq n$ và $|v| \geq 1$
  - Mâu thuẫn sẽ phát sinh theo bổ đề bơm bằng cách chỉ ra với u, v và w xác định theo giả thiết, có tồn tại một số i mà ở đó $uv^iw \notin L$. Từ đó có thể kết luận rằng L không là ngôn ngữ chính quy. Chọn lựa giá trị cho i có thể phụ thuộc vào n, u, v và w.

Bổ đề bơm cho tập hợp chính quy

- **Ví dụ:** chứng minh tập hợp **L = { 0^{i^2} | i $\in \mathrm{N}$, i ≥ 1}** không làp tập hợp chính quy
- **Chứng minh:**
  - ❖Giả sử L là tập chính quy → tồn tại DFA chấp nhận L. Gọi n là số trạng thái của DFA.
  - ❖Xét chuỗi z = 0^{n^2}. Theo bổ đề bơm: z=uvw với 1≤ |v|≤ n và $uv^iw \in L$
  - ❖Xét i = 2, ta phải có $uv^2w \in L$
  - ❖Mặt khác: n^2 = |z| = |uvw| < |uvvw|≤ $n^2 + n < (n+1)^2$
  - ❖Do n^2 và $(n+1)^2$ là 2 số chính phương liên tiếp nên luv^2wl không thể là một số chính phương, hay uv^2w không thuộc L (trái giả thiết).

Bổ đề bơm cho tập hợp chính quy

- **Ví dụ:** Xét ngôn ngữ $L=\{0^i1^i\}$ giả sử L là ngôn ngữ chính quy. Ta phải có n sao cho $w=xyz$; $|w|\geq n$, $|xy|\leq n$, $y\neq\varepsilon$ và $xy^kz\in L$. Tuy nhiên xét chuỗi 0^n1^n

Tính đóng của tập hợp chính quy

- Một phép toán trên các tập RS được gọi là **đóng** nếu kết quả của phép toán đó là tập hợp cùng loại.
- **Định lý 4:** tập hợp chính quy đóng với các phép toán: hợp, nối kết và bao đóng Kleen (bao đóng sao).
- **Định lý 5:** tập hợp chính quy đóng với phép lấy phần bù.
- **Định lý 6:** tập hợp chính quy đóng với phép giao.
- **Định lý 7:** tập hợp các chuỗi được chấp nhận bởi FA có n trạng thái là: không rỗng nếu và chỉ nếu nó chấp nhận chuỗi có độ dài nhỏ hơn n, hoặc vô hạn nếu và chỉ nếu nó chấp nhận chuỗi có độ dài m với $n \leq m < 2n$.
- **Định lý 8:** Tồn tại giải thuật để xác định hai FA là tương đương (chấp nhận cùng một ngôn ngữ)