

## Chương 2:

## BỘ PHÂN TÍCH TỪ NGUỒN

### ① Nhiệm vụ.

- Mục tiêu: Tìm chuỗi dài nhất các ký tự đầu vào
- Nội dung:
  - Duyệt từng ký tự của VB nguồn.
  - XD từ vựng
  - Nhận dạng từ tổ → giải pháp tiếp theo

→ Từ vựng: Đơn vị nhỏ nhất trong NNLT.

→ Từ tổ (token): Chỉ các từ vựng có ngữ nghĩa ngữ pháp.

VĐ: "Hôm nay", "trời", "đẹp" ⇒ Từ vựng.

↓                      ↓                      ↓

trạng từ              DT                      TT ⇒ Từ tổ.

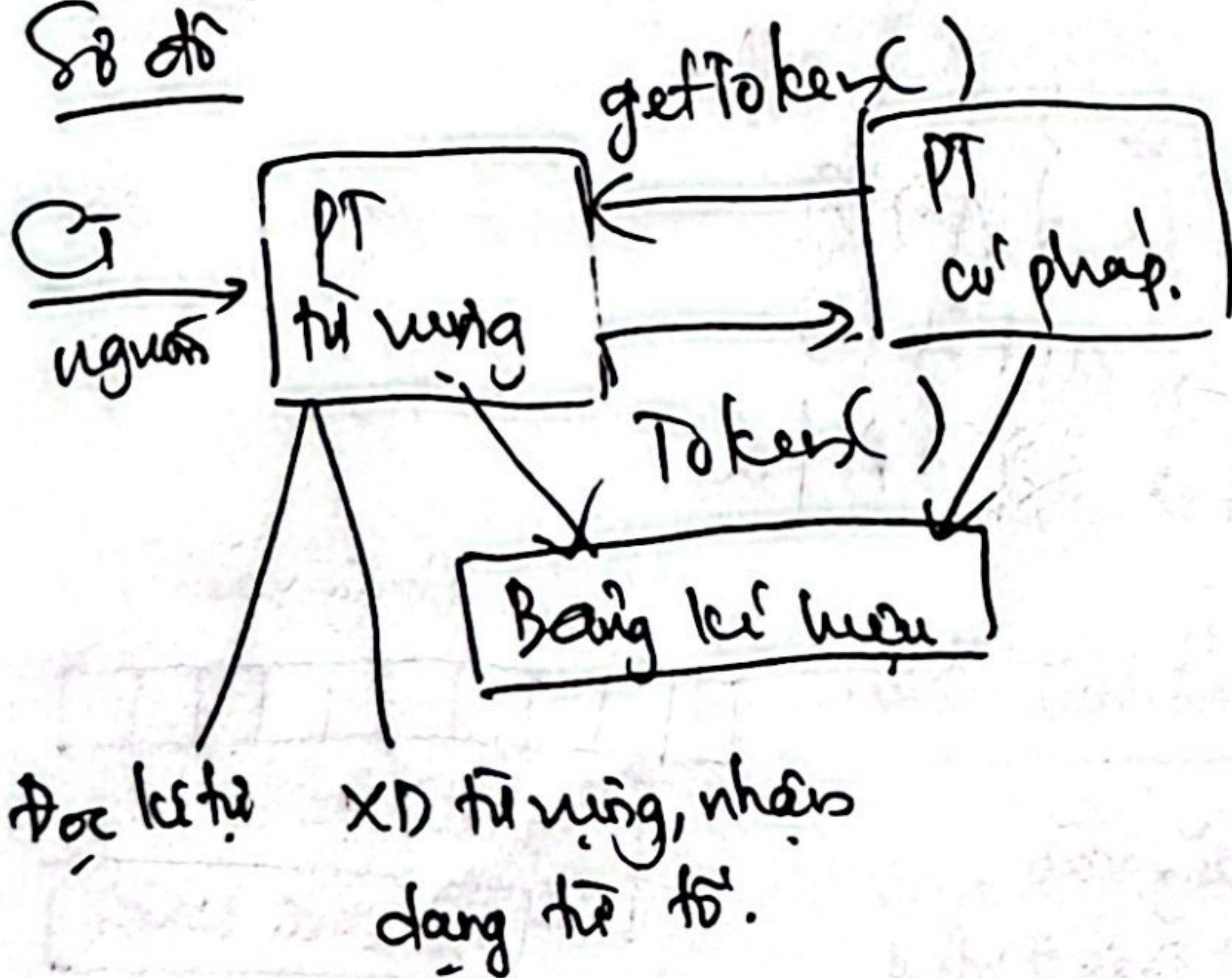
⇒ Bộ phân tích từ vựng nhận dạng từ tổ.

↓ Biết ∈ tính tương ứng

Từ tổ  $\xrightarrow{\text{tất}}$  Bộ phân tích cú pháp.

∈ tính  $\xrightarrow{\text{sd}}$  bộ sinh mã.

Sơ đồ



⊛ Mẫu (Pattern) → Luật mô tả từ tổ

↓ mô tả bởi bthức chính quy

### ② Biểu thức chính quy

- Ngôn ngữ: Set cơ string (câu / xâu).
- Câu: Dãy hữu hạn các từ / ký hiệu
- Từ: Tập từ 1 hoặc chữ hữu hạn.
- Ngôn ngữ: Có thể là vô hạn / hữu hạn.
- mô tả bởi bthức chính quy

### Biểu thức chính quy.

$\Sigma$  - bảng chữ của 1 ngôn ngữ.

- $\phi$ : biểu thức chính quy bđ ngôn ngữ  $\phi$
- $\epsilon$ : bđ ngôn ngữ  $\{\epsilon\}$
- $a \in \Sigma$ ,  $a$  là bđ ngôn ngữ  $\{a\}$

→  $r \rightarrow R \Rightarrow r+s$  là bthức chính quy biến  $R$  US

$s \rightarrow S \Rightarrow r \cdot s$   $r^*$   $R \cdot S$   $R^*$

Chú ý:  $(r+s) \rightarrow (r|s)$   $(r \cdot s) \rightarrow rs$ .

$(a|\epsilon) \leftrightarrow (a|)$

Định thứ tự ưu tiên.

- $(*)$  ưu tiên hơn  $(.)$
- $(.)$  ưu tiên hơn  $(+)$

VĐ:  $(0 + (1(0^*))) = (0 + (10^*)) = 0 + 10^*$

VĐ:  $\Sigma = \{a, b\}$  là 1 bảng chữ.

$e_1 = a^* + b^* \Rightarrow L(e_1) = \{\epsilon, a, aa, aaa \dots b, bb \dots\}$

$e_2 = a^* b^* \Rightarrow L(e_2) = \{\epsilon, a, b, ab, aa, bb, aaa \dots\}$

$e_3 = a(a+b)^* \Rightarrow L(e_3) = \{\epsilon, a, aa, ab, aaa, aab, aba, abb \dots\}$

Chú ý: Xâu  $\leftarrow$  tập hợp các từ hợp bởi ký hiệu  $a, b$ .

$a$  chỉ các  $b$  chỉ số.  $\Rightarrow L(e_3)$  - bthức chứa tên

Tính của bthức chính quy.

$$r+s = s+r$$

$$r+r = r$$

$$r+s+t = (r+s)+t = r+(s+t).$$

$$(r \cdot s) \cdot t = r \cdot (s \cdot t) = r \cdot s \cdot t.$$

$$r \cdot \epsilon = \epsilon r = r \quad r + \phi = \phi + r = r.$$

$$r(s+t) = rs + rt.$$

$$r + r^* = r^* \quad (r + \epsilon)^* = r^* \quad (r^*)^* = r^*$$

$$r \cdot r^* = r^* r = r^+.$$

$$(r+s)^* = (r^* s^*)^*$$







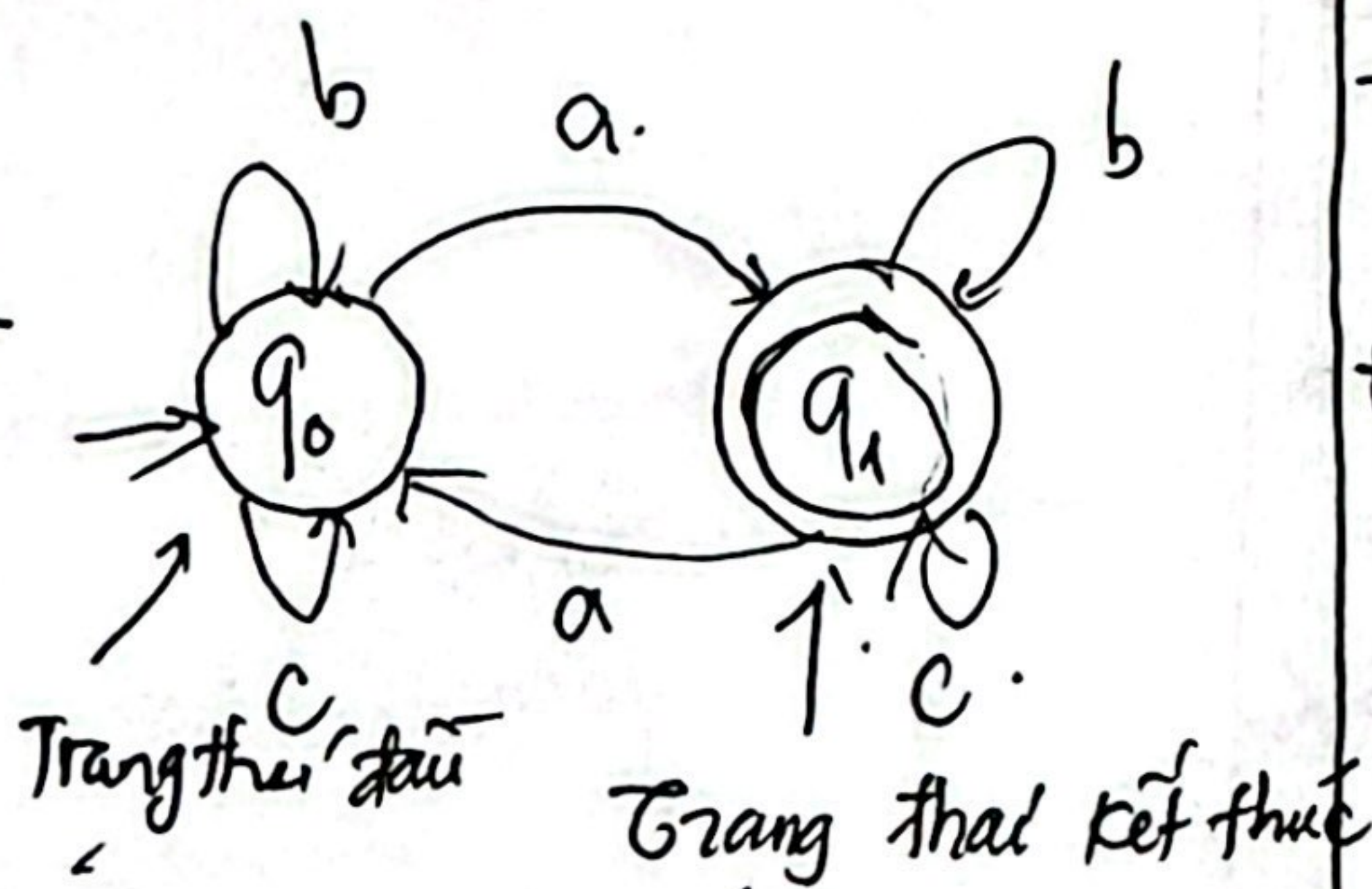
Đ đoán nhận x là  $w \in \Sigma^*$ .

- Đặt <sup>đơn vị</sup> ~~nhỏ~~ <sup>nhỏ</sup> hàng vào.  $\begin{cases} \text{Mỗi kí hiệu 1 ô.} \\ \# \rightarrow \text{kí hiệu} \\ \rightarrow \text{đặt ở cuối ô.} \end{cases}$

- Ô tô mất khỏi đầu q.
- Đầu dọc trở vào ký hiệu trái nhất w.
- Ô tô mất ở trạng thái k → xác định đầu
- ↳ Ngc lại, k.

10  $\Sigma = \{a, b, c\}$   $Q = \{q_0, q_1\}$   $q_0 = 0.$   
 $F = \{q_1\}$

$g$	$a$	$b$	$c$
$q_0$	$q_1$	$q_0$	$q_0$
$q_1$	$q_0$	$q_1$	$q_1$



$$\textcircled{p} \xrightarrow{a} \textcircled{q} \rightarrow \delta(p, q) = q$$

Note: Xâu trên hai chữ  $\{a, b, c\}$  có lẽ kì lạ

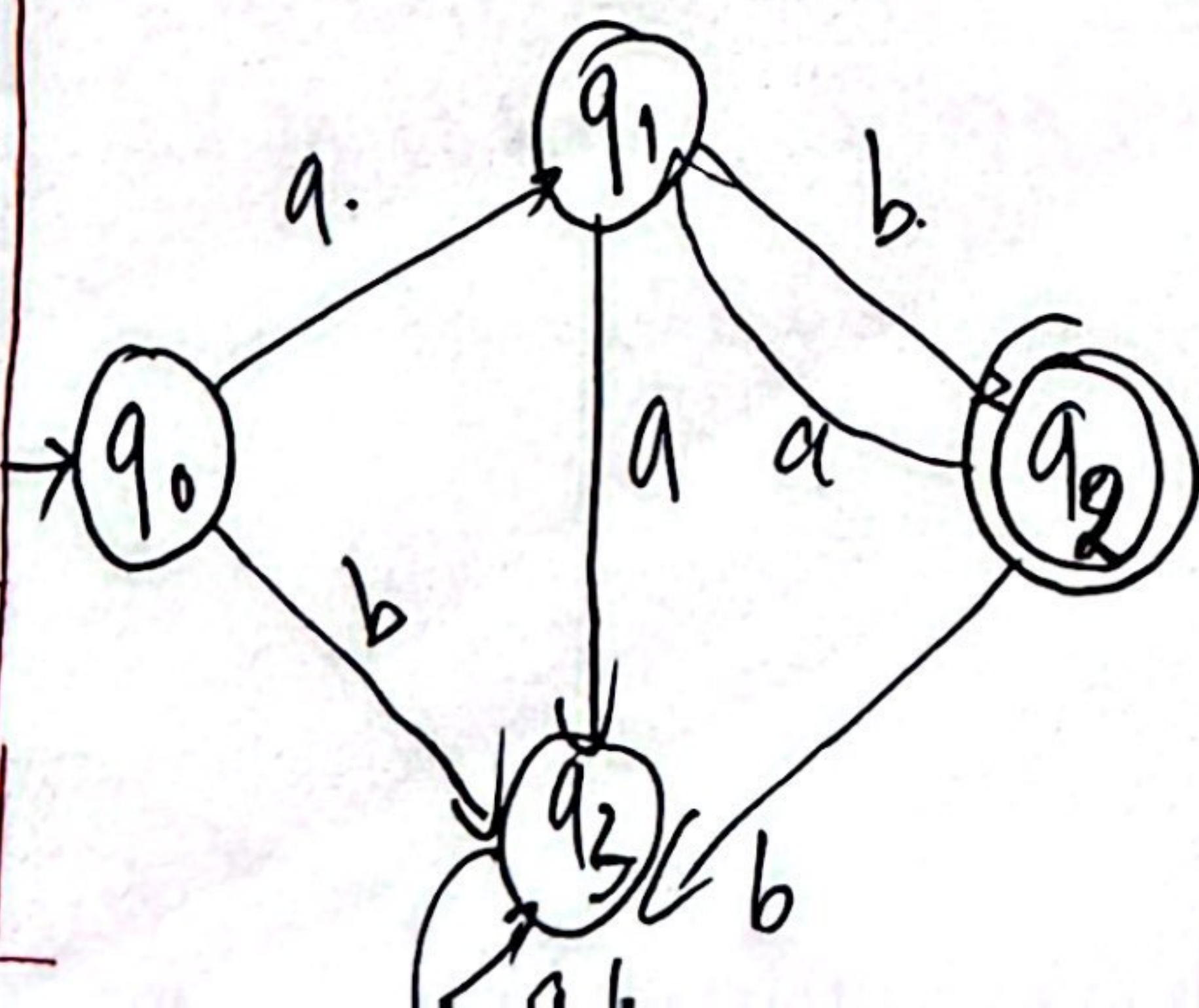
⊗ ĐTS mất hạn do chủ

OTD (DFA)  $M = (\Sigma, Q, \delta, q_0, F)$

- $\Sigma$ : bộ chữ.
- $Q$ : Tập hữu hạn các trạng thái của bộ đtđ.
- $Q \cap \Sigma = \emptyset$
- $\delta: Q \times \Sigma \rightarrow Q$ : Hàm dịch chuyển.
- $q_0 \in Q$ : T<sup>z</sup> ban đầu
- $F \subseteq Q$ : Tập các trạng thái cuối.

W.  $M = (\Sigma, Q, \delta, q_0, F)$   $\Sigma = \{a, b\}$   
 $\Sigma = \{a\}$   $Q = (q_i, i = 0, 3); q_0 = q_0; F = \{q_2\}$

$\delta$	$a$	$b$
$q_0$	$q_1$	$q_3$
$q_1$	$q_3$	$q_2$
$q_2$	$q_1$	$q_3$
$q_3$	$q_3$	$q_3$



$\Rightarrow k'$  bắt đầu tại  $b \rightarrow Rn vào t^2$  chế

$$L(N) = \{ab\}^* \{ab\}^{n \geq 0} = ab(ab)^n = (ab)^+$$

⊗ Hình trạng của FA

- Hình trạng của 1FA là  
1 xâu đang gx

④  $q \in \mathbb{Q}$ :  $r^2$  htar

①  $x \in \Sigma^*$  → phân chia xét của x vào

- Chuyển đổi hình trạng:  $\begin{cases} a \in \Sigma \\ b \in \Sigma^* \end{cases}$   
 Đầu vào là  $a$ ,  
 +)  $x = ay$ .

\*)  $\exists x (q, x) = b \Rightarrow q^x = q \cdot a_y \Rightarrow p_y.$

$qx \rightarrow py$ : bđm hình trạng

Vd:  $q_0 \underline{a} b a a b \Rightarrow q_1 \underline{b} a a b \Rightarrow q_2 a a b$

$$\Rightarrow q_1 ab \Rightarrow q_3 b \Rightarrow q_3 \Rightarrow 100$$

Hình trạng đầu: 90w 93°F

$q_0 \in W \Rightarrow q_{n+1} \in F$ : Xâu đưa đầu nh

Algor nghĩ đơn giản nhận bởi DFA  $\rightarrow L(M)$

$$LCM) = \{w \mid w \in \Sigma^* \text{ v.a. } q_0 w \Rightarrow^* p \in F\}$$

⑧ Đến một hạn hạn không đơn giản

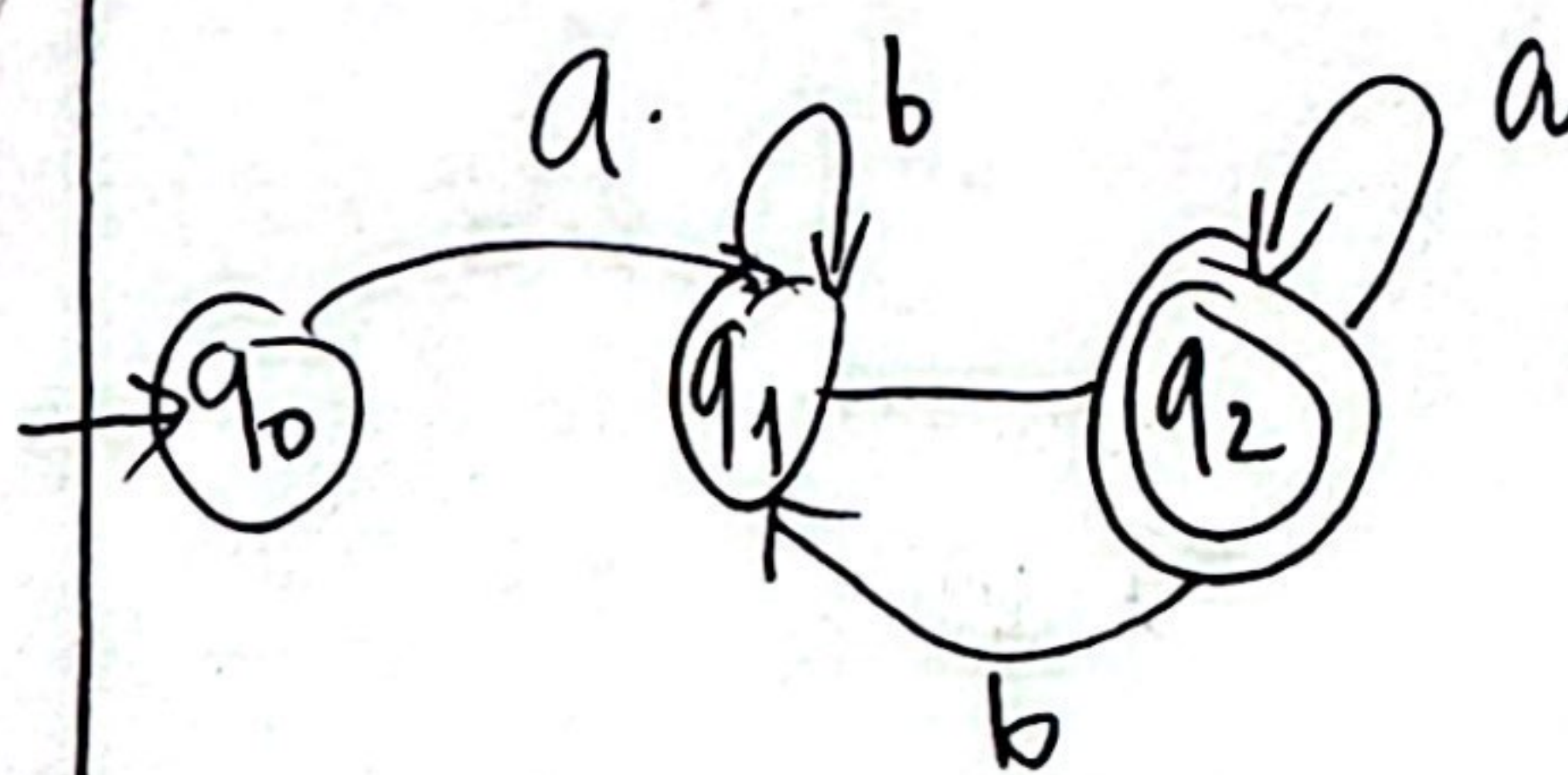
OHK:  $M = (\Sigma, Q, \delta, q_0, F)$ .

$\Sigma, Q, q_0, F$  giống OHA.  $\boxed{S = Q \times (\Sigma^* \varepsilon) \rightarrow 2^Q}$

$\angle^Q$ : tập hợp phần tử của  $Q$ .

- Ngõn ngĩa:  $L(M) = \{w \mid w \in Z^* \text{ và } \exists q_0 w \rightarrow^* q \text{ ep}\}$

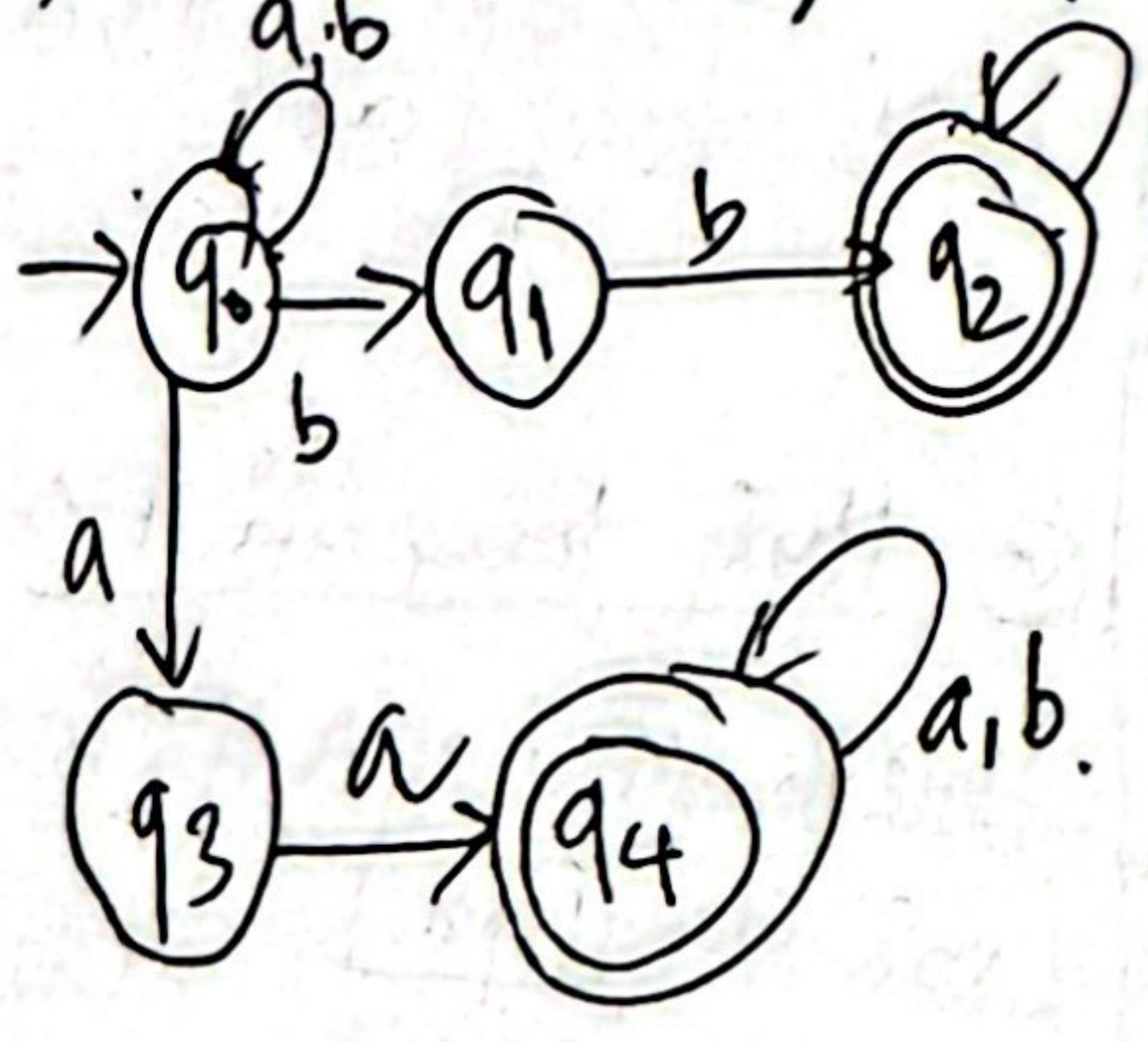
w. Vgl.:  $\underline{a}(a+b)^2 \underline{a}$  bei  $\delta$  format  $h^2$ .





10.  $M = (\{a, b\}, q_0 \rightarrow q_4, \delta, q_0, \{q_2, q_4\})$

$\delta$	a	b
$q_0$	$q_0, q_3$	$q_0, q_1$
$q_1$	$\emptyset$	$q_2$
$q_2$	$q_2$	$q_2$
$q_3$	$q_4$	$\emptyset$
$q_4$	$q_4$	$q_4$



$q_0 abaab \Rightarrow q_0 baab \Rightarrow q_0 aab \Rightarrow q_0 ab$   
 $\Rightarrow q_0 b \Rightarrow q_1$  (m)

$q_0 abaab \Rightarrow q_0 baab \Rightarrow q_0 aab \Rightarrow q_3 ab$   
 $\Rightarrow q_4 b \Rightarrow q_4 \Rightarrow$  Xấu đi đến  
 nên.

$\Rightarrow$  Đoạn nhận xấu có 2 kí tự a hoặc b liên tiếp.

(\*) DFA và NFA.

$\hookrightarrow L(DFA) = L(NFA)$ .

$\hookrightarrow$  DFA và NFA là tương đương

$\downarrow$  DFA  $\supset$  NFA tự do

(\*) Thuật toán đoạn nhận xấu của OHP.

$\rightarrow$  Phân tích bằng

$\leftarrow$  Phg pháp diễn giải

4. Phân tích từ vựng của ngôn ngữ PL10