BỘ GIÁO DỤC VÀ ĐÀO TẠO
TRƯỜNG ĐẠI HỌC NÔNG NGHIỆP HÀ NỘI

PGS.TS. NGUYỄN VĂN ĐỊNH

GIÁO TRÌNH

AUTOMAT VÀ NGÔN NGỮ HÌNH THỨC

Chương 3
OTOMAT HỮU HẠN VÀ NGÔN NGỮ CHÍNH QUY

Trong chương này, chúng ta sẽ nghiên cứu một mô hình "máy trừu tượng" để đoán nhận ngôn ngữ, đó là các otomat hữu hạn. Chương này gồm các nội dung sau:

3.1. Otomat hữu hạn đơn định

- Ngôn ngữ được đoán nhận bởi otomat đơn định

3.2. Otomat hữu hạn không đơn định

- Ngôn ngữ đoán nhận bởi otomat không đơn định
- Đơn định hóa các otomat
- Sự tương đương giữa các otomat đơn định và không đơn định

3.3. Ngôn ngữ chính quy và biểu thức chính quy

- Sự liên hệ giữa otomat hữu hạn và ngôn ngữ chính quy
- Sự liên hệ giữa otomat hữu hạn và biểu thức chính quy
- Điều kiện cần của ngôn ngữ chính quy

3.1. OTOMAT HỮU HẠN ĐƠN ĐỊNH

Một otomat hữu hạn (*Finite Automata-FA)* là một mô hình tính toán thực sự hữu hạn với sự mô tả quá trình tính toán bởi các *input (đầu vào)* và các *output (đầu ra).* Tại mỗi thời điểm, hệ thống có thể được xác định bởi một trong số các cấu hình nội tại, gọi là các *trạng thái* (*state status*). Mỗi trạng thái của hệ thống thể hiện tóm tắt các thông tin liên quan đến những input đã chuyển qua hệ thống, và xác định các trạng thái kế tiếp trên dãy input tiếp theo.

Mọi cái liên quan đến một otomat hữu hạn đều có kích thước hữu hạn, cố định, và không thể mở rộng trong suốt quá trình tính toán. Các loại otomat khác được nghiên cứu sau này có ít nhất một bộ nhớ vô hạn về tiềm năng. Sự phân biệt giữa các loại otomat khác nhau chủ yếu dựa trên việc thông tin có thể được lưu trữ trong bộ nhớ như thế nào.

Một otomat hữu hạn làm việc theo thời gian rời rạc, như vậy, ta có thể nói về thời điểm "kế tiếp" khi đặc tả hoạt động của một otomat hữu hạn.

Tuy nhiên, nói chung, thông tin ra (output) sản sinh bởi một otomat hữu hạn phụ thuộc vào cả thông tin vào hiện tại lẫn các thông tin vào trước đó. Như vậy otomat có khả năng (với một phạm vi nào đó) ghi nhớ các thông tin vào trong quá khứ của nó. Một cách chi tiết hơn, điều đó có nghĩa như sau: Mỗi otomat hữu hạn có một số hữu hạn trạng thái được lưu ở bộ nhớ trong, tại mỗi thời điểm i, nó ở một trong các trạng thái đó, chẳng hạn q_i. Trạng thái q_{i+1} ở thời điểm sau được xác định bởi q_i và thông tin vào a_i cho ở thời điểm i. Thông tin ra ở thời điểm i được xác định bởi trạng thái q_i (hay bởi cả a_i và q_i). Vì vậy, các otomat hữu hạn còn được gọi là các "*máy trạng thái*".

Trong khoa học máy tính, ta có thể tìm thấy nhiều ví dụ về hệ thống trạng thái hữu hạn, và lý thuyết về các otomat hữu hạn là một công cụ thiết kế hữu ích cho các hệ thống này. Chẳng hạn, trường hợp đơn giản nhất là thiết bị không có bộ nhớ mà ở mỗi thời điểm, thông tin ra chỉ phụ thuộc vào thông tin vào lúc đó. Các thiết bị như vậy là mô hình của các mạch tổ hợp. Hay phức tạp hơn, mỗi trạng thái của một hệ thống chuyển mạch với n cổng vào sẽ là một trong 2^n phép gán của 0 và 1 với những giá trị của input, vì vậy, mỗi bộ chuyển mạch có thể được xem là một hệ thống trạng thái hữu hạn.

Các máy tính hiện đại ngày nay cũng là các máy trạng thái hữu hạn, trạng thái hiện tại của bộ xử lý trung tâm, bộ nhớ trong và các thiết bị lưu trữ liên quan ở mỗi thời điểm bất kỳ là một số rất lớn nhưng hữu hạn của số các trạng thái. Và ngay cả bộ não của con người cũng có thể được xem như một hệ thống trạng thái hữu hạn, vì số các nơ ron thần kinh là số có giới hạn, số này không vượt quá 2^{35}.

Các otomat hữu hạn được chia làm hai loại: otomat hũu hạn đơn định (*Deterministic Finite Automata*-DFA) và otomat hữu hạn không đơn đinh (*Non deterministic Finite Automata*-NFA), và đều có khả năng đoán nhận các ngôn ngữ chính quy, là lớp ngôn ngữ có khả năng ứng dụng rất lớn trong việc xây dựng các ngôn ngữ lập trình. Otomat đơn định có khả năng đoán nhận ngôn ngữ dễ dàng hơn so với otomat không đơn định, nhưng otomat không đơn định lại có ưu điểm là thường có cấu trúc đơn giản hơn. Điều quan trọng là lớp ngôn ngữ đoán nhận bởi hai loại otomat này là tương đương.

Dưới đây chúng ta sẽ lần lượt nghiên cứu cả hai loại otomat này.

3.1.1. Định nghĩa otomat hữu hạn đơn định

Định nghĩa 3.1

Một *otomat hữu hạn đơn định* (*Deterministic Finite Automata-DFA*) là một bộ năm:

$$\boldsymbol{A = < Q, \Sigma, \delta, q_0, F >}$$

trong đó:

- Q là một tập hữu hạn khác rỗng, được gọi là *tập các trạng thái (set of states)*;
- Σ là một bảng chữ cái, được gọi là bảng chữ vào *(input alphabet);*
- δ: $D \rightarrow Q$, là một ánh xạ từ $D \subseteq Q \times \Sigma$ vào Q, được gọi là *hàm chuyển trạng thái (transition function)*;
- $q_0 \in Q$, được gọi là *trạng thái khởi đầu (starting state);*
- $F \subseteq Q$ được gọi là tập *trạng thái kết thúc (final states).*

Nếu $D = Q \times \Sigma$, tức là hàm hai biến $\delta(q, a)$ xác định với mọi cặp giá trị (q, a), với $q \in Q$, $a \in \Sigma$ thì ta nói A là otomat đầy đủ, trái lại, nếu có những cặp giá trị *(q, a)* mà tại đó hàm chuyển $\delta(q, a)$ không xác định, thì otomat A gọi là không đầy đủ. Sau này ta sẽ thấy rằng mọi otomat hữu hạn đều đưa về được otomat hữu hạn đầy đủ tương đương.

Hoạt động của otomat hữu hạn đơn định $A = <Q, \Sigma, \delta, q_0, F>$ với một xâu vào $\omega = a_1a_2\dots a_n$ có thể được mô tả như sau:

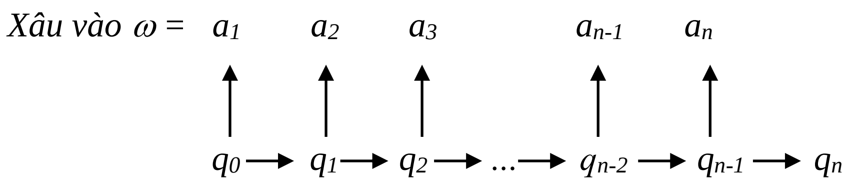
Khi bắt đầu làm việc, otomat ở trạng thái khởi đầu q_0 và đầu đọc đang nhìn vào ô có ký hiệu a_1. Tiếp theo, dưới tác động của ký hiệu vào a_1, otomat chuyển từ trạng thái q_0 về trạng thái mới q_1 xác định bởi hàm chuyển, tức là $\delta(q_0, a_1) = q_1 \in Q$, và đầu đọc chuyển sang phải một ô, tức là nhìn vào ô có ký hiệu a_2. Sau đó otomat A có thể lại tiếp tục chuyển từ trạng thái q_1 nhờ hàm chuyển δ về trạng thái mới $q_2 = \delta(q_1, a_2)$. Quá trình đó sẽ tiếp tục cho tới khi gặp một trong các tình huống sau:

a). Otomat A đọc hết xâu vào ω và chuyển sang trạng thái q_n, tức là A ở trạng thái q_{n-1}, đọc ký hiệu cuối cùng a_n và có $\delta(q_{n-1},a_n) = q_n$, khi đó nếu $q_n \in F$, ta nói rằng A đoán nhận xâu ω, trái lại nếu $q_n \notin F$ thì ta nói otomat A không đoán nhận xâu ω.

b). Hoặc khi otomat A ở trạng thái q_i nào đó, đọc đến ký hiệu a_j của xâu vào, ($j \leq n$) và hàm chuyển $\delta(q_i , a_j)$ không xác định, khi đó otomat A dừng, ta cũng nói A không đoán nhận xâu ω.

Chú ý rằng nếu A là otomat đầy đủ thì tình huống a). luôn xảy ra, tức là xâu ω luôn được đọc hết, còn tình huống b). chỉ có thể xảy ra khi A là otomat không đầy đủ.

Hình 3.1 dưới đây mô tả quá trình đoán nhận xâu *ω* với tình huống otomat A đọc hết xâu vào ω và chuyển về trạng thái q_n. Khi đó nếu $q_n \in F$, ta nói rằng A đoán nhận xâu ω, trái lại nếu $q_n \notin F$ thì ta nói otomat A không đoán nhận xâu ω.



Hình 3.1. Mô tả quá trình đoán nhận xâu ω của otomat A

3.1.2. Biểu diễn otomat hữu hạn đơn định

Hàm chuyển trạng thái là một bộ phận quan trọng của một otomat hữu hạn đơn định. Cho một otomat thực chất là cho hàm chuyển trạng thái của nó, có thể cho dưới dạng bảng chuyển hoặc cho dưới dạng đồ thị chuyển. Ứng với mỗi cách cho hàm chuyển, ta có một cách biểu diễn otomat.

Biểu diễn otomat bằng bảng chuyển trạng thái

Cho ôtômát $A = <Q, \Sigma, \delta, q_0, F>$, với $Q = \{q_0, q_1, q_2,\dots, q_m\}$ là tập trạng thái, và bảng chữ cái $\Sigma = \{a_1, a_2,\dots, a_n\}$, khi đó hàm chuyển có thể cho bởi một bảng gồm m+1 hàng và n cột; trong đó ô (i, j) nằm trên dòng i cột j của bảng được ghi giá trị hàm chuyển $\delta(q_i,a_j)$, và ô (i, j) là ô trống (hoặc ghi kí hiệu Ø) nếu $\delta(q_i,a_j)$ không xác định.

Cho bảng chuyển trạng thái, và chỉ rõ tập trạng thái kết thúc F, quy ước q_0 là trạng thái khởi đầu, ta sẽ hoàn toàn xác định được otomat A.

Bảng chuyển trạng thái của otomat A sẽ có dạng như sau:

Trạng thái	Ký hiệu vào			
	a_1	a_2	…	a_n
q_0	$\delta(q_0, a_1)$	$\delta(q_0, a_2)$	…	$\delta(q_0, a_n)$
q_1	$\delta(q_1, a_1)$	$\delta(q_1, a_2)$	…	$\delta(q_1, a_n)$
q_2	$\delta(q_2, a_1)$	$\delta(q_2, a_2)$	…	$\delta(q_2, a_n)$
…	…	…	…	…
q_m	$\delta(q_m, a_1)$	$\delta(q_m, a_2)$	…	$\delta(q_m, a_n)$

Hình 3.2. Bảng chuyển trạng thái của otomat A

Biểu diễn otomat bằng đồ thị chuyển trạng thái

Cho otomat A = <Q, Σ, δ, q_0, F>. Hàm chuyển δ có thể cho bằng một đa đồ thị có hướng, có khuyên G, được gọi là *đồ thị chuyển trạng thái (hay đồ thị chuyển)* của otomat A. Tập đỉnh của G được gán nhãn bởi các phần tử thuộc Q, còn các cung được gán nhãn bởi các phần tử thuộc Σ, tức là nếu $a \in \Sigma$ và từ trạng thái q chuyển sang trạng thái p theo công thức δ(q, a) = p thì sẽ có một *cung từ đỉnh q tới đỉnh p được gán nhãn a.*

Đỉnh vào của đồ thị chuyển là đỉnh ứng với trạng thái ban đầu q_0. Các đỉnh sẽ được khoanh bởi các vòng tròn, tại đỉnh xuất phát q_0 có mũi tên đi vào, riêng đỉnh với trạng thái kết thúc được phân biệt bởi vòng tròn đậm, hoặc hình vuông…

Việc cho đồ thị chuyển với các quy ước như trên là hoàn toàn xác định được otomat A.

Thí dụ 3.1: Cho hai otomat hữu hạn đơn định:

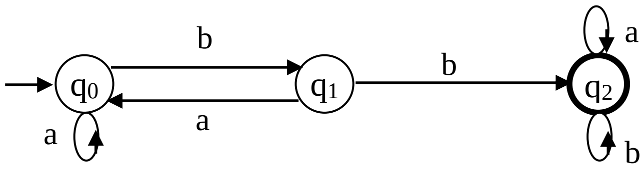
- A_1 = <{q_0, q_1, q_2}, {a, b}, δ, q_0, {q_2}>,

Với $\delta(q_0, a) = q_0$, $\delta(q_0, b) = q_1$, $\delta(q_1, a) = q_0$, $\delta(q_1, b) = q_2$, $\delta(q_2, a) = q_2$, $\delta(q_2, b) = q_2$.

Ta có bảng chuyển trạng thái và đồ thị chuyển trạng thái của otomat A_1 được cho trong hình 3.3 và 3.4:

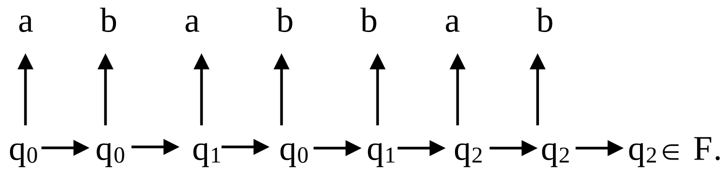
Trạng thái	Ký hiệu vào a	b
q_0	q_0	q_1
q_1	q_0	q_2
q_2	q_2	q_2

Hình 3.3. Bảng chuyển trạng thái của A_1



Hình 3.4. Đồ thị chuyển trạng thái của A_1

Dãy trạng thái của otomat A_1 trong quá trình đoán nhận xâu vàoα = ababbab là:



Hình 3.5. Quá trình đoán nhận xâu α = ababbab của A_1

Như vậy, xâu α được đoán nhận bởi otomat A_1.

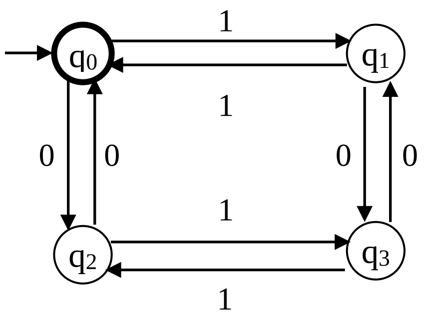
- $A_2 = <\{q_0, q_1, q_2, q_3\}, \{0, 1\}, \delta, q_0, \{q_0\}>$,

trong đó $\delta(q_0, 0) = q_2$, $\delta(q_0, 1) = q_1$, $\delta(q_1, 0) = q_3$, $\delta(q_1, 1) = q_0$, $\delta(q_2, 0) = q_0$, $\delta(q_2, 1) = q_3$, $\delta(q_3, 0) = q_1$, $\delta(q_3, 1) = q_2$.

Ta có bảng chuyển trạng thái và đồ thị chuyển trạng thái của otomat A_2 được cho trong hình 3.6 và 3.7:

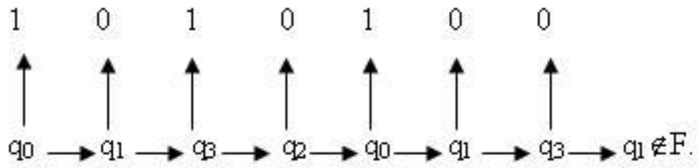
Trạng thái	Ký hiệu vào 0	1
q_0	q_2	q_1
q_1	q_3	q_0
q_2	q_0	q_3
q_3	q_1	q_2

Hình 3.6. Bảng chuyển trạng thái của otomat A_2



Hình 3.7. Đồ thị chuyển trạng thái của otomat A_2

Dãy trạng thái của otomat A_2 trong quá trình đoán nhận xâu vào β = 1010100 được mô tả trong hình 3.8 dưới đây:



Hình 3.8. Quá trình đoán nhận xâu vào β = 1010100

Như vậy, otomat A_2 không chấp nhận xâu β.

Ta có thể mô tả quá trình đoán nhận xâu vào của otomat hữu hạn đơn định đầy đủ A bằng thuật toán mô phỏng sau:

Input :

- Một xâu ω, kết thúc bởi ký hiệu kết thúc file là eof.
- Một otomat hữu hạn đơn định đầy đủ A với trạng thái khởi đầu q_0 và tập trạng thái kết thúc là F.

Output:

- Trả lời "Đúng" nếu A đoán nhận xâu ω.
- Trả lời "Sai" nếu A không đoán nhận xâu ω.

Thuật toán:

```
Begin
        S := q0;
        C := ký hiệu đầu tiên của ω;
        While C < > eof do
                begin
                        S:= δ(S, C);
                        C:= ký hiệu tiếp theo;
                end;
        if S in F return (True)
        else return (False);
End.
```

3.1.3. Ngôn ngữ được đoán nhận bởi otomat đơn định

Để mô tả hình thức quá trình đoán nhận một từ (xâu vào), người ta đưa vào ánh xạ mở rộng δ' từ $Q\times\Sigma^*$ vào Q như trong định nghĩa sau:

Định nghĩa 3.2

Cho otomat hữu hạn đơn định $A = <Q, \Sigma, \delta, q_0, F>$. Mở rộng δ' của δ là một ánh xạ từ $Q\times\Sigma^*$ vào Q được xác định như sau:

1. $\delta'(q, \varepsilon) = q,\ \forall q\in Q,$

2. $\delta'(q, \varphi a) = \delta(\delta'(q, \varphi), a),\ \forall a\in\Sigma,\ \forall q\in Q,\ \forall\omega = \varphi a\in\Sigma^*$ *sao cho* $\delta'(q, \varphi)$ *được xác định.*

Chú ý rằng, ánh xạ δ chỉ khác ánh xạ δ' khi ký hiệu vào là ε, hoặc là một xâu kí hiệu vào ω, do điều kiện 2., trên $Q\times\Sigma$, ta có thể đồng nhất δ' với δ. Nếu không cần phân biệt, từ đây về sau ta viết δ thay cho δ', và được hiểu là ánh xạ δ trên miền $Q\times\Sigma$, hiểu là ánh xạ δ' trên miền $Q\times\Sigma^*$.

Định nghĩa 3.3

Cho otomat hữu hạn đơn định $A = <Q, \Sigma, \delta, q_0, F>$, và một xâu $\omega\in\Sigma^*$. Ta nói ω được đoán nhận bởi A nếu $\delta(q_0, \omega)\in F$. Tập hợp tất cả các từ được đoán nhận bởi otomat A gọi là ngôn ngữ được đoán nhận bởi otomat A, và ký hiệu là T(A), vậy:

$$T(A) = \{\omega\in\Sigma^* \mid \delta(q_0, \omega)\in F\}$$

Từ định nghĩa trên, ta thấy rằng trong đồ thị chuyển của A, xâu $\omega\in\Sigma^*$ được đoán nhận bởi A khi và chỉ khi ω là xâu của các nhãn ứng với một đường đi từ đỉnh q_0 đến một trong các đỉnh kết thúc. Cụ thể, nếu $\omega = a_1a_2\ldots a_k$ thì đường đi là $(q_0, q_1,\ldots, q_k)$ với cung (q_{i-1}, q_i) có nhãn a_i (với $1 \le i \le k$) và $q_k\in F$.

Như vậy, T(A) là tập hợp tất cả xâu ghi trên các đường đi từ q_0 đến một đỉnh kết thúc nào đó của otomat A.

Bổ đề 3.1

Cho otomat hữu hạn đơn định $A = <Q, \Sigma, \delta, q_0, F>$. Khi đó $\forall\omega_1, \omega_2\in\Sigma^*$, $\forall q\in Q$ sao cho $\delta(q, \omega_1\omega_2)$ xác định, ta có:

$$\delta(q, \omega_1\omega_2) = \delta(\delta(q, \omega_1), \omega_2) \quad (1)$$

Chứng minh: Ta chứng minh đẳng thức trên bằng quy nạp theo độ dài của ω_2.

- Khi $|\omega_2| = 1$ hay $\omega_2 = a$, $a\in\Sigma$, ta có $\delta(q, \omega_1 a) = \delta(\delta(q, \omega_1),a)$. Đẳng thức (1) đúng.

. Giả sử đẳng thức (1) đúng với mọi ω_2 có độ dài $|\omega_2| \leq n$. Ta cần chứng minh nó cũng đúng với ω_2 có độ dài $|\omega_2| = n + 1$. Đặt $\omega_2 = \omega'_2a$, với $\omega'_2 \in \Sigma^*$, $|\omega'_2| = n$, $a \in \Sigma$. Ta có $\delta(q, \omega_1\omega_2) = \delta(q, \omega_1\omega'_2a) = \delta(\delta(q, \omega_1\omega'_2), a) = \delta(\delta(\delta(q, \omega_1), \omega'_2), a) = \delta(\delta(q, \omega_1), \omega'_2a) = \delta(\delta(q, \omega_1), \omega_2)$.

Do đó đẳng thức (1) đúng với ω_2 có độ dài n + 1.

Bổ đề được chứng minh.

Chú ý:

1. Với otomat hữu hạn đơn định $A = <Q, \Sigma, \delta, q_0, F>$ bất kỳ, ta luôn có thể xây dựng một otomat hữu hạn đơn định đầy đủ A' tương đương với A.

Thật vậy, lấy $S \notin Q$ (do đó $S \notin F$), đặt $Q' = Q \cup \{S\}$ và $\delta': Q' \times \Sigma \rightarrow Q'$ xác định bởi: $\forall q \in Q$, $\forall a \in \Sigma$, $\delta'(q, a) = \delta(q, a)$ nếu $\delta(q, a)$ được xác định, $\delta'(q, a) = S$ nếu $\delta(q, a)$ không được xác định và $\delta'(S, a) = S$. Khi đó A' là otomat hữu hạn đơn định đầy đủ mà T(A') = T(A).

Ta thường chọn $S = \varnothing$, và không cần bổ sung S vào Q.

2. Trên đồ thị chuyển một otomat đơn định và đầy đủ được thể hiện: với mỗi đỉnh (gán nhãn là một trạng thái) đều có số cung đi ra khỏi đỉnh đó đúng bằng số ký hiệu của bảng chữ cái, mỗi cung ra khác nhau được gán nhãn bằng một chữ cái khác nhau.

Chẳng hạn, các otomat A_1 và A_2 trong thí dụ 3.1 là các otomat đơn định và đầy đủ.

3.2. OTOMAT HỮU HẠN KHÔNG ĐƠN ĐỊNH

3.2.1. Định nghĩa otomat hữu hạn không đơn định

Định nghĩa 3.4

Otomat hữu hạn không đơn định (Nondeterministic Finite Automata-NFA) là một bộ năm:

$$A = <Q, \Sigma, \delta, q_0, F>$$

trong đó Q, Σ, q_0, F như trong định nghĩa của DFA, riêng hàm chuyển trạng thái là một ánh xạ không đơn định $\delta: Q \times \Sigma \rightarrow 2^Q$, ở đây 2^Q là ký hiệu tập tất cả các tập con của Q.

Ở đây ánh xạ δ là một hàm đa trị, cho kết quả là một tập trạng thái, tức là $\forall q \in Q$, $\forall a \in \Sigma$ thì $\delta(q, a) = \{p_1, p_2,\ldots p_k\} \subseteq Q$ (hàm không đơn định), vì vậy otomat A trong định nghĩa trên đây được gọi là không đơn định.

Trong trường hợp δ(q, a) xác định $\forall q \in Q$, $\forall a \in \Sigma$, ta nói otomat A là đầy đủ.

Nếu $\delta(q, a) = \{p_1, p_2,\ldots, p_k\}$ thì ta nói rằng otomat A ở trạng thái q gặp ký hiệu a thì có thể chuyển đến một trong các trạng thái $p_1, p_2,\ldots, p_k$. Nếu $\delta(q, a) = \{p\}$ thì ở trạng thái q gặp ký hiệu a, otomat A chỉ chuyển đến một trạng thái duy nhất p, trường hợp này cũng giống như trong otomat hữu hạn đơn định. Như vậy, ta thấy rằng một otomat hữu hạn đơn định cũng có thể coi là một otomat hữu hạn không đơn định đặc biệt, với kết quả của hàm chuyển $\delta(q, a)$ là một tập chỉ gồm một trạng thái.

Nếu $\delta(q, a)$ không xác định (ta thường viết $\delta(q, a) = \varnothing$) thì ở trạng thái q gặp ký hiệu a, otomat A không thể chuyển đến trạng thái nào, otomat sẽ dừng, (tương tự như với DFA).

Hoạt động của otomat hữu hạn không đơn định $A = <Q, \Sigma, \delta, q_0, F>$ khi cho xâu vào $\omega = a_1a_2\ldots a_n$ có thể được mô tả như sau:

Khi bắt đầu làm việc, otomat ở trạng thái khởi đầu q_0 và đầu đọc đang nhìn vào ô có ký hiệu a_1. Từ trạng thái q_0, dưới tác động của ký hiệu vào a_1, $\delta(q_0, a_1) = \{p_1,\ldots, p_k\}$, otomat xác định trạng thái tiếp theo có thể là một trong các trạng thái $p_1,\ldots, p_k$ và đầu đọc chuyển sang phải một ô, tức là nhìn vào ô có ký hiệu a_2. Tiếp tục với mỗi p_i $(1\leq i \leq k)$ và ký hiệu tiếp theo là a_2, các trạng thái tiếp theo có thể đến được là $\delta(p_1, a_2)\cup\ldots\cup\delta(p_k, a_2)$. Quá trình đó sẽ tiếp tục cho tới khi gặp một trong các tình huống sau:

a). Otomat A đọc hết xâu vào ω và chuyển sang tập trạng thái Q', khi đó có hai khả năng:

(i) Nếu $Q'\cap F \neq \varnothing$, ta nói rằng otomat A đoán nhận ω.

(ii) Nếu $Q'\cap F = \varnothing$, ta nói rằng A không đoán nhận ω.

b). Hoặc ở bước i-1, có $\delta(q_{i-1}, a_{j-1}) = \{p_1,\ldots, p_k\}= Q'$, mà $\forall p_i \in Q'$, hàm chuyển $\delta(q_i , a_j)$ không xác định, khi đó otomat A dừng, ta cũng nói A không đoán nhận xâu ω.

Chú ý rằng nếu A là otomat đầy đủ thì tình huống a). luôn xảy ra, tức là xâu ω luôn được đọc hết, còn tình huống b). chỉ có thể xảy ra khi A là otomat không đầy đủ.

Một otomat hữu hạn không đơn định có thể biểu diễn dưới dạng bảng chuyển hoặc đồ thị chuyển như trong trường hợp otomat hữu hạn đơn định. Nếu $\delta(q, a) = \{p_1, p_2,\ldots, p_k\}$ thì trong đồ thị chuyển có k cung từ q sang $p_1,.., p_k$ được ghi cùng một nhãn a.

Thí dụ 3.2: Cho otomat hữu hạn không đơn định:

$A = <\{q_0, q_1, q_2, q_3, q_4\}, \{0, 1\}, \delta, q_0, \{q_2, q_4\}>$,

Với $\delta(q_0,0) = \{q_0,q_3\}$, $\delta(q_0, 1) = \{q_0,q_1\}$, $\delta(q_1, 0) = \varnothing$, $\delta(q_1, 1) = \{q_2\}$, $\delta(q_2, 0) = \{q_2\}$,

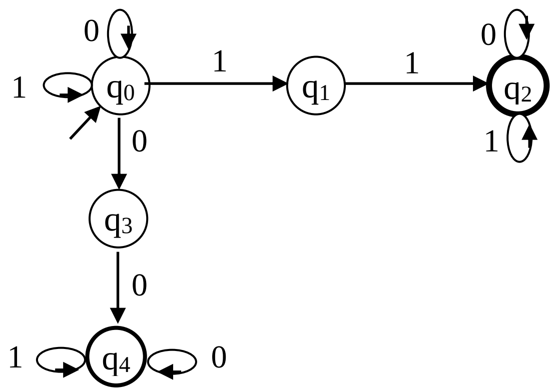
$\delta(q_2, 1) = \{q_2\}$, $\delta(q_3, 0) = \{q_4\}$, $\delta(q_3,1) = \varnothing$, $\delta(q_4, 0) = \{q_4\}$, $\delta(q_4, 1) = \{q_4\}$.

Bảng chuyển trạng thái và đồ thị chuyển trạng thái của otomat A cho trong hình 3.9 và 3.10:

Trạng thái	Ký hiệu vào	
	a_1	a_2
q_0	$\{q_0, q_3\}$	$\{q_0, q_1\}$
q_1	$\varnothing$	$\{q_2\}$
q_2	$\{q_2\}$	$\{q_2\}$
q_3	$\{q_4\}$	$\varnothing$
q_4	$\{q_4\}$	$\{q_4\}$

Hình 3.9. Bảng chuyển của otomat không đơn định A

Otomat A trên đây là không đơn định và không đầy đủ. Dựa vào bản chuyển trạng thái, ta có thể xây dựng đồ thị chuyển trạng thái của A như sau:



Hình 3.10. Đồ thị chuyển của otomat không đơn định A

Chú ý: Trên đồ thị chuyển một otomat không đơn định được thể hiện: tồn tại ít nhất một đỉnh (gán nhãn là một trạng thái) mà có ít nhất hai cung ra khác nhau được gán nhãn bằng cùng một chữ cái.

Chẳng hạn, trong đồ thị chuyển của otomat A trong hình 3.10, tại đỉnh q_0, có hai cung ra cùng ghi kí hiệu ‘0’, nên A là otomat không đơn định. Hơn nữa, A cũng là otomat không đầy đủ.

Từ đồ thị chuyển, có thể thấy rằng otomat A trên đây cũng không phải là otomat đầy đủ, vì có những đỉnh (trạng thái), từ đó có số cung đi ra ít hơn số ký hiệu trong bảng chữ cái (tại đỉnh gán nhãn q_1 và q_3).

3.2.2 Ngôn ngữ được đoán nhận bởi otomat hữu hạn không đơn định

Định nghĩa 3.5

Cho otomat hữu hạn không đơn định A = <Q, Σ, δ, q₀, F>. Mở rộng của δ là ánh xạ δ' từ tập $Q \times \Sigma^*$ *vào* 2^Q *được xác định như sau:*

1) $\delta'(q, \varepsilon) = \{q\}, \forall q \in Q$,

2) $\delta'(q, \varphi a) = \bigcup_{p \in \delta(q,a)} \delta'(p, \varphi)$, $\forall q \in Q, \forall a \in \Sigma, \forall \omega = \varphi a \in \Sigma^*$ *sao cho* $\delta'(q, \varphi)$ *được xác định.*

Cũng như trường hợp otomat hữu hạn đơn định, ta có thể sử dụng ký hiệu δ thay cho δ', ký hiệu này được hiểu là ánh xạ δ trên miền $Q \times \Sigma$, và được hiểu là ánh xạ δ' trên miền $Q \times \Sigma^*$.

Thật vậy, trên miền $Q \times \Sigma$, ta có $\delta'(q, a) = \delta'(q, \varepsilon a) = \bigcup_{p \in \delta(q,a)} \delta'(p, \varepsilon) = \bigcup_{p \in \delta(q,a)} p = \delta(q, a)$, $\forall q \in Q, \forall a \in \Sigma$, tức là trên $Q \times \Sigma$ thì ánh xạ δ' chính là ánh xạ δ.

Định nghĩa 3.6

Cho otomat hữu hạn không đơn định A = <Q, Σ, δ, q₀, F>, và một xâu $\omega \in \Sigma^*$. *Ta nói ω được đoán nhận bởi A nếu* $\delta(q_0, \omega) \cap F \neq \emptyset$. *Tập tất cả các từ được đoán nhận bởi otomat A gọi là ngôn ngữ được đoán nhận bởi otomat A, và ký hiệu là T(A).*

$$T(A) = \{\omega \in \Sigma^* \mid \delta(q_0, \omega) \cap F \neq \emptyset\}.$$

Thí dụ 3.3: Cho otomat hữu hạn không đơn định:

$A = <\{q_0, q_1, q_2\}, \{a, b\}, \delta, q_0, \{q_2\}>$,

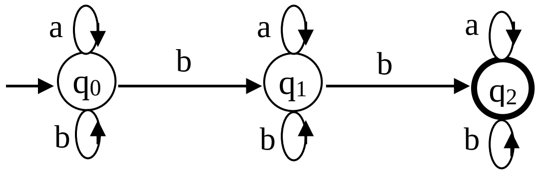
trong đó $\delta(q_0, a) = \{q_0\}$, $\delta(q_0, b) = \{q_0, q_1\}$, $\delta(q_1, a) = \{q_1\}$, $\delta(q_1, b) = \{q_1, q_2\}$,

$\delta(q_2, a) = \{q_2\}$, $\delta(q_2, b) = \{q_2\}$.

Bảng chuyển và đồ thị chuyển của otomat A được cho trong hình 3.11 và 3.12:

Trạng thái	Ký hiệu vào	
	a	b
q_0	$\{q_0\}$	$\{q_0, q_1\}$
q_1	$\{q_1\}$	$\{q_1, q_2\}$
q_2	$\{q_2\}$	$\{q_2\}$

Hình 3.11. Bảng chuyển của otomat A trong thí dụ 3.3



Hình 3.12. Đồ thị chuyển của otomat A trong thí dụ 3.3

Có thể kiểm tra được rằng từ $\omega = a^n b^n \in T(A)$, tuy nhiên otomat A không đoán nhận ngôn ngữ $L = \{ a^n b^n \mid \forall n \geq 1\}$. Ngôn ngữ được đoán nhận bởi otomat A là:

$T(A) = \{\omega_1 b \omega_2 b \omega_3 \mid \omega_1, \omega_2, \omega_3 \in \{a, b\}^*\}$.

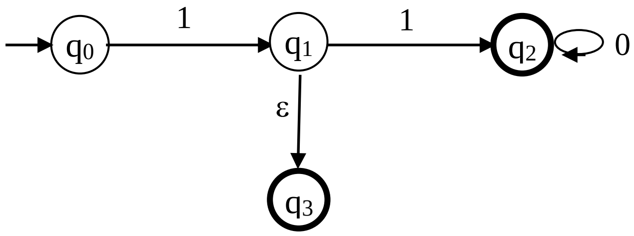
3.2.3 Otomat hữu hạn với các bước chuyển rỗng

Xét otomat hữu hạn không đơn định $A = <Q, \Sigma, \delta, q_0, F>$ với các thành phần như trong định nghĩa 3.4. Nếu hàm chuyển trạng thái được xác định trên miền $Q \times (\Sigma \cup \{\varepsilon\})$, tức là biến thứ hai của hàm chuyển δ chấp nhận giá trị ε, và ta có $\delta(q, \varepsilon) = \{p_1, p_2, \ldots p_k\}$. Nói các khác, khi đang ở trạng thái q, otomat A có thể chuyển sang một trong các trạng thái $p_1, p_2, \ldots p_k$ mà không cần đọc một ký hiệu nào. Một bước chuyển như vậy được gọi là một *bước chuyển" nhắm mắt"* hay *bước chuyển rỗng* của otomat A, hay một *ε- chuyển (ε-move*). Một bước chuyển rỗng từ trạng thái q sang trạng thái p được biểu diễn trên đồ thị chuyển là một cung đi từ q đến p, trên đó có ghi nhãn là ε.

Như vậy, các otomat có ε-chuyển sẽ có hàm chuyển trạng thái là ánh xạ $\delta: Q \times (\Sigma \cup \{\varepsilon\}) \rightarrow 2^Q$.

Dễ dàng thấy rằng, mọi otomat có ε-chuyển đều là các otomat không đơn định, chẳng hạn xét thí dụ sau:

Thí dụ 3.4: Cho otomat A với đồ thị chuyển.



Hình 3.13. Đồ thị chuyển của otomat A trong thí dụ 3.4

Ta có A là một otomat hữu hạn với ε-chuyển $\delta(q_1, \varepsilon) = q_3$. Ta sẽ chứng tỏ rằng A là một otomat không đơn định. Thật vậy:

Từ đồ thị chuyển, ta có: $\delta(q_0, 1) = q_1$, (i)

nhưng ta cũng có: $\delta(q_0, 1) = \delta(q_0, 1\varepsilon) = \delta(\delta(q_0, 1),\varepsilon) = \delta(q_1, \varepsilon) = q_3$. (ii)

Từ (i) và (ii), ta có: $\delta(q_0, 1) = \{q_1, q_3\}$, vậy otomat A là không đơn định.

Dễ thấy rằng, ngôn ngữ được đoán nhận bởi otomat A là: $T(A) = \{1, 110^n \mid n \geq 0\}$.

3.2.4. Các otomat tương đương

Đối với các otomat hữu hạn, người ta quan tâm đến các ngôn ngữ mà nó đoán nhận, bất kể chúng là các otomat loại nào, đơn định hay không đơn định. Nếu hai otomat hữu hạn cùng đoán nhận một ngôn ngữ thì ta nói chúng tương đương, Ta có định nghĩa sau:

Định nghĩa 3.7

Hai otomat hữu hạn $A = <Q, \Sigma, \delta, q_0, F>$ và $A' = <Q', \Sigma', \delta', q'_0, F'>$ được gọi là tương đương nếu $T(A) = T(A')$.

Cho trước một otomat A, việc xác định các otomat tương đương thường nhằm làm đơn giản otomat ban đầu nhờ việc loại bỏ các trạng thái không tham gia vào việc đoán nhận ngôn ngữ, các trạng thái như vậy gọi là các "*trạng thái chết*" (*dead states*). Một trạng thái *p* sẽ là trạng thái chết nếu không có một đường đi nào từ q_0 đến một trạng thái kết thúc mà đi qua *p*.

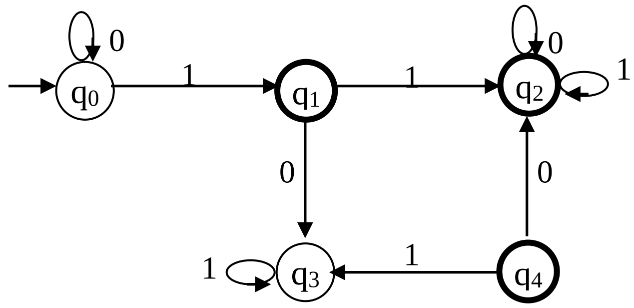
Thí dụ 3.5: Cho otomat hữu hạn: $A = <\{q_0, q_1, q_2, q_3, q_4\},\{0, 1\}, \delta, q_0, \{q_1, q_2, q_4\}>$,

với: $\delta(q_0,0) = q_0$, $\delta(q_0,1) = q_1$, $\delta(q_1,0) = q_3$, $\delta(q_1,1) = q_2$,

$\delta(q_2,0) = q_2$, $\delta(q_2,1) = q_2$, $\delta(q_3,1) = q_3$,

$\delta(q_4,0) = q_2$, $\delta(q_4,1) = q_3$.

Đồ thị chuyển của A là:



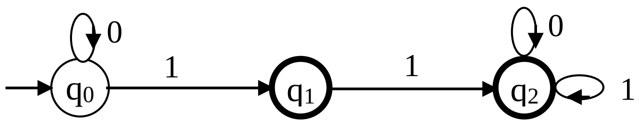
Hình 3.14. Đồ thị chuyển của otomat A trong thí dụ 3.5

Trước hết, ta nhận thấy rằng không có đường đi từ đỉnh q_0 đến đỉnh kết thúc q_4, tức là sẽ không có từ nào được đoán nhận bởi A với đỉnh kết thúc q_4. Ngoài ra, cũng không có một đường đi nào từ q_0 đến đỉnh một đỉnh kết thúc mà đi qua q_3. Như vậy, các trạng thái q_3 và q_4 sẽ không tham gia vào việc đoán nhận các từ, chúng là các trạng thái chết. Trong đồ thị chuyển, ta có thể bỏ đi đỉnh q_3 và q_4 mà không ảnh hưởng đến tập từ được đoán nhận bởi otomat A. Với việc bỏ đi các trạng thái chết q_3 và q_4, ta nhận được otomat A' tương đương với otomat A.

A' = $<\{q_0, q_1, q_2\}, \{0, 1\}, \delta, q_0, \{q_1, q_2\}>$,

trong đó $\delta(q_0,0) = q_0$, $\delta(q_0,1) = q_1$, $\delta(q_1,1) = q_2$, $\delta(q_2,0) = q_2$, $\delta(q_2,1) = q_2$.

Đồ thị chuyển của A' được cho trong hình 3.15 dưới đây:



Hình 3.15. Đồ thị chuyển của otomat A' trong thí dụ 3.5

Các đường đi từ q_0 đến đỉnh kết thúc q_1 ứng với các xâu 0^n1, $n \geq 0$. Các đường đi từ q_0 đến đỉnh kết thúc q_2 ứng với các xâu $0^n11\omega$, $n \geq 0$, $\omega \in \{0, 1\}^*$. Vậy ngôn ngữ được đoán nhận bởi các otomat trên là:

$$T(A) = T(A') = \{0^n1, 0^n11\omega \mid n \geq 0, \omega \in \{0, 1\}^*\}.$$

3.2.5. Vấn đề đơn định hóa otomat

Khi nghiên cứu các otomat và các ngôn ngữ hình thức, chúng ta thường gặp hai bài toán cơ bản:

Bài toán phân tích: Cho một otomat hữu hạn A, hãy xác định ngôn ngữ được đoán nhận bởi otomat này, tức là xác đinh tập T(A). Nói chung, việc xác định ngôn ngữ được đoán nhận bởi otomat đơn định là dễ dàng hơn nhiều so với một otomat không đơn định.

Bài toán tổng hợp: Cho một ngôn ngữ chính quy L, hãy xây dựng một otomat hữu hạn đoán nhận L, tức là tìm otomat A để L = T(A). Bài toán này nói chung khó hơn bài toán thứ nhất, với bài toán này, người ta thường xây dựng một otomat không đơn định để đoán nhận ngôn ngữ L, vì việc này đơn giản hơn xây dựng một otomat đơn định cũng đoán nhận L, sau đó, nếu cần thiết, tìm một otomat đơn định tương đương với otomat vừa tìm được.

Như vậy, với cả hai bài toán trên, vấn đề tìm một otomat đơn định tương đương với một otomat không đơn định cho trước là rất cần thiết, vì nó sẽ giúp cho việc giải quyết cả hai bài toán trên trở nên đơn giản hơn. Vấn đề này được gọi là *đơn định hóa otomat*, được phát biểu như sau:

Cho A = <Q, Σ, δ, q_0, F> là một otomat không đơn định, hãy xây dựng otomat đơn định và đầy đủ M tương đương với otomat A, theo nghĩa cùng đoán nhận một ngôn ngữ, tức là T(M) = T(A).

Việc xây dựng otomat M được thực hiện theo các thuật toán sau đây, được gọi là thuật toán đơn định hóa otomat.

Thuật toán đơn định hóa 1

Input: Otomat hữu hạn không đơn định A = <Q, Σ, δ, q_0, F>

Output: Otomat hữu hạn đơn định M = <Q', Σ, δ', s_0, F'> tương đương với otomat A.

Phương pháp:

Bước 1:

Xây dựng hàm hai biến T: $2^Q \times \Sigma \to 2^Q$ thỏa mãn các điều kiện:

1. $\forall$ q$\in$ Q, $\forall$ a$\in$ Σ thì T(q, a) = δ(q, a),

2. $\forall$ B $\subseteq$ Q mà B là tập trạng thái kết quả của hàm chuyển δ hoặc hàm T, mà |B|> 1 thì:

$$T(B, a) = \bigcup_{p \in B} T(p, a)$$

Bước 2:

Xác định tập trạng thái mới Q' = {s_0, s_1,…, s_k | $k \le 2^{|Q|}-1$}:

1/. Đặt $s_0 = \{q_0\}$, $s_1 = \{q_1\}$,… $s_i = \{q_i\}$, $\forall$ $q_0, q_1, \ldots, q_i \in Q$,

2/. Đặt $s_{i+1} = B_1$, $s_{i+2} = B_2$,… với B_i là tập trạng thái kết quả của hàm chuyển δ, hoặc hàm T, | B_i | > 1.

3/. Nếu otomat A là không đầy đủ, đặt $s_k = \varnothing$ và thêm vào hàm chuyển δ' các giá trị $\delta'(s_k, a) = s_k$ $\forall$ a$\in$ Σ để otomat M trở thành otomat đầy đủ.

4/. Trạng thái khởi đầu của otomat M là s_0.

5/. Tập trạng thái kết thúc của otomat M là F' = {s$\in$ Q' | s $\cap$ F $\neq \varnothing$ }.

Bước 3 :

Xác định hàm chuyển δ’: $Q' \times \Sigma \rightarrow Q'$ của otomat M:

$\forall s \in Q'$, $\forall a \in \Sigma$ thì $\delta'(s, a) = T(s, a)$

Việc chứng minh T(A) = T(M) là khá dễ dàng, dành cho sinh viên như là bài tập.

Thí dụ 3.6: Cho otomat $A = <\{p_0, p_1, p_2\}, \{a, b, c\}, \delta, p_0, \{p_1, p_2\}>$ với hàm chuyển δ cho bởi bảng sau:

δ	a	b	c
p_0	$\{p_1\}$	$\{p_1, p_2\}$	$\{p_2\}$
p_1	$\{p_2\}$		$\{p_0, p_2\}$
p_2	$\{p_1\}$	$\{p_1\}$	$\{p_2\}$

Hình 3.16. Bảng chuyển của otomat A trong thí dụ 3.6

Hãy xây dựng otomat $M = <Q', \{a, b, c\}, \delta', s_0, F'>$ đơn định và đầy đủ, tương đương với otomat A.

1/. Xây dựng hàm T: $2^Q \times \Sigma \rightarrow 2^Q$

$T(p_0, a) = \{p_1\}$, $T(p_0, b) = \{p_1, p_2\}$, $T(p_0, c) = \{p_2\}$,

$T(p_1, a) = \{p_2\}$, $T(p_1, b) = \varnothing$, $T(p_1, c) = \{p_0, p_2\}$,

$T(p_2, a) = \{p_1\}$, $T(p_2, b) = \{p_1\}$, $T(p_2, c) = \{p_2\}$,

$T(\{p_1, p_2\}, a) = T(p_1,a) \cup T(p_2,a) = \{p_2\} \cup \{p_1\} = \{p_1, p_2\}$, $T(\{p_1, p_2\}, b) = \varnothing \cup \{p_1\} = \{p_1\}$, $T(\{p_1, p_2\}, c) = \{p_0, p_2\} \cup \{p_2\} = \{p_0, p_2\}$,

$T(\{p_0, p_2\}, a) = \{p_1\}$, $T(\{p_0, p_2\}, b) = \{p_1, p_2\}$, $T(\{p_0, p_2\}, c) = \{p_2\}$

2/. Đặt $s_0 = \{p_0\}$, $s_1 = \{p_1\}$, $s_2 = \{p_2\}$, $s_3 = \{p_1, p_2\}$, $s_4 = \{p_0, p_2\}$, $s_5 = \varnothing$ ta có tập trạng thái mới là $Q' = \{s_0, s_1, s_2, s_3, s_4, s_5\}$.

3/. Hàm chuyển mới δ’: $Q' \times \Sigma \rightarrow Q'$ được xác định bởi bảng sau:

δ’	a	b	c
s_0	s_1	s_3	s_2
s_1	s_2	s_5	s_4
s_2	s_1	s_1	s_2
s_3	s_3	s_1	s_4
s_4	s_1	s_3	s_2
s_5	s_5	s_5	s_5

Hình 3.17. Bảng chuyển của otomat đơn định M trong thí dụ 3.6

Trạng thái khởi đầu của M là s_0, tập trạng thái kết thúc mới là: F' = $\{s_1, s_2, s_3, s_4\}$.

Rõ ràng otomat M = <$\{s_0, s_1, s_2, s_3, s_4, s_5\}$, {a, b, c}, δ', s_0, $\{s_1, s_2, s_3, s_4\}$> với hàm chuyển δ' cho bởi bảng trên là otomat hữu hạn đơn định và đầy đủ. Có thể thấy rằng otomat M là tương đương với otomat A.

Thuật toán đơn định hóa 2

Input: Otomat hữu hạn không đơn định A = <Q, Σ, δ, q_0, F>

Output: Otomat hữu hạn đơn định M = <Q', Σ, δ', s_0, F'> tương đương với otomat A.

Phương pháp:

Bước 1:

Đặt trạng thái khởi đầu của otomat đơn định M là trạng thái khởi đầu q_0 của otomat không đơn định A. Đặt Q' = $\{[q_0]\}$, và tiếp tục xác định các trạng thái khác của M trong bước 2.

Bước 2:

Với mỗi ký hiệu vào $a \in \Sigma$ và với mỗi trạng thái $[q_i, q_j, \dots q_k]$ trong Q' của M, xác định giá trị hàm chuyển trạng thái δ' như sau:

1. $\delta'([q_i, q_j, \dots q_k], a) = \delta(q_i, a) \cup \delta(q_j, a) \cup \dots \cup \delta(q_k, a) = [q_l, q_m, \dots q_n]$

2. Bổ sung trạng thái mới $[q_l, q_m, \dots q_n]$ vào Q' nhờ phép gán $Q' := Q' \cup \{[q_l, q_m, \dots q_n]\}$

(có thể ký hiệu mỗi trạng thái $[q_i, q_j, \dots q_k]$ của Q' bởi $B_1, B_2, \dots$).

3. Lặp lại bước 2 để xác định các bước chuyển trạng thái của hàm chuyển và bổ sung các trạng thái mới cho M, cho đến khi không có trạng thái mới nào được bổ sung.

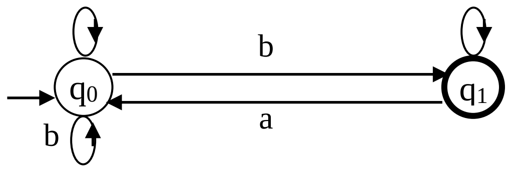
Bước 3:

Trạng thái $B = [q_i, q_j, \dots q_k] \in Q'$ gọi là trạng thái kết thúc nếu ít nhất một trong các trạng thái $q_i, q_j, \dots q_k$ là trạng thái kết thúc của A, tức là $B \in F'$ nếu và chỉ nếu $\{q_i, q_j, \dots q_k\} \cap F \neq \emptyset$. Nếu otomat A chấp nhận ε thì trạng thái khởi đầu $[q_0]$ đồng thời được đặt là trạng thái kết thúc của M'.

Như vậy thuật toán đã xây dựng được otomat đơn định M', dễ dàng thấy rằng T(M') = T(A).

Thí dụ 3.7: Cho otomat không đơn định: A = <$\{q_0, q_1\}$, {a, b}, δ, q_0, $\{q_1\}$>, trong đó $\delta(q_0, a) = \{q_0\}$, $\delta(q_0, b) = \{q_0, q_1\}$, $\delta(q_1, a) = \{q_0, q_1\}$, $\delta(q_1, b) = \varnothing$.

Đồ thị chuyển của A là:



Hình 3.18. Đồ thị chuyển của otomat A trong thí dụ 3.7

Do A là otomat không đơn đinh nên việc xác định tập từ T(A) là khá khó khăn.

Trước hết ta sẽ xây dựng otomat đơn định M = <Q', {a, b}, δ', t_0, F'> tương đương với A.

Theo thuật toán đơn định hóa 2, ta thực hiện các bước sau:

Bước 1: Đặt trạng thái khởi đầu của M là $[q_0]$, và Q' = $\{[q_0]\}$

Bước 2: Tìm các trạng thái mới của M, bằng cách xác đinh các bước chuyển trạng thái với mỗi trạng thái đã có của Q'.

- Với $[q_0] \in Q'$:

  - với ký hiệu vào a: $\delta'([q_0], a) = \delta(q_0, a) = [q_0]$,

  - với ký hiệu vào b: $\delta'([q_0], b) = \delta(q_0, b) = [q_0, q_1]$,

  - bổ sung trạng thái mới: $Q' := Q' \cup \{[q_0, q_1]\} = \{[q_0], [q_0, q_1]\}$

- Tiếp tục xét với $[q_0, q_1] \in Q'$

  - với ký hiệu vào a: $\delta'([q_0, q_1], a) = \delta(q_0, a) \cup \delta(q_1, a) = [q_0, q_1]$,

  - với ký hiệu vào b: $\delta'([q_0, q_1], b) = \delta(q_0, b) \cup \delta(q_1, b) = [q_0, q_1]$,

  Do không có trạng thái mới bổ sung vào Q', kết thúc bươc 2.

Bước 3 : do $[q_0, q_1]$ có chứa q_1 là trạng thái kết thúc của A, nên $[q_0, q_1]$ là trạng thái kết thúc của otomat M

Như vậy, otomat M được xác định bởi :

- $Q' = \{ [q_0], [q_0, q_1]\}$, với ký hiệu $t_0 = [q_0]$ và $t_1 = [q_0, q_1]$, ta có $Q' = \{t_0, t_1\}$, .
- Bảng chữ cái $\Sigma = \{a, b\}$,
- Trạng thái khởi đầu : t_0
- Tập trạng thái kết thúc $F' = \{t_1\}$,

. Hàm chuyển trạng thái δ' được xác định

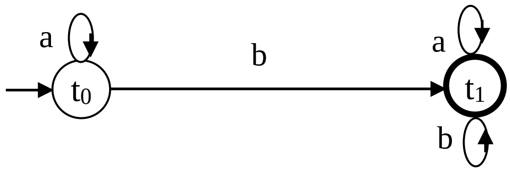
$\delta'(t_0, a) = t_0$, $\delta'(t_0, b) = t_1$, $\delta'(t_1, a) = t_1$, $\delta'(t_1, b) = t_1$.

Ta có bảng chuyển của M:

δ'	a	b
t_0	t_0	t_1
t_1	t_1	t_1

Hình 3.19. Bảng chuyển của otomat đơn định M trong thí dụ 3.7

Rõ ràng otomat M là đơn định và có đồ thị chuyển như sau:



Hình 3.20. Đồ thị chuyển của otomat M trong thí dụ 3.7

Việc xác định ngôn ngữ được đoán nhận bởi otomat M là khá dễ dàng:

$T(A) = T(M) = \{a^n b\omega \mid n \geq 0, \omega \in \{a, b\}^*\}$.

Nhận xét: Với thuật toán 1, việc xây dựng otomat đơn định M là khá dễ dàng, các bước đều được chuẩn hóa, thuận lợi cho việc lập trình. Tuy nhiên thuật toán này có nhược điểm là số trạng thái của M thường là khá lớn, cho nên sau khi tìm được M, cần loại bỏ các trạng thái không tham gia quá trình đoán nhận ngôn ngữ (*dead states*), để làm cho otomat đơn giản hơn.

Thuật toán 2 thường cho kết quả ngắn gọn hơn so với thuật toán 1, chỉ những trạng thái thực sự tham gia vào quá trình đoán nhận ngôn ngữ mới được bổ sung vào tập trạng thái mới Q'. Tuy nhiên, nhược điểm của thuật toán này là việc thực hiện cần phải linh hoạt, các quy tắc không được chuẩn hóa, cho nên khó lập trình. Otomat M có thể không phải là đầy đủ.

Thuật toán 2 phù hợp với việc đơn định hóa các otomat có số trạng thái không lớn, có thể tính trực tiếp, không cần lập trình.

3.2.6. Sự tương đương giữa otomat đơn định và otomat không đơn định

Các định lý dưới đây sẽ cho ta thấy sự tương đương giữa lớp otomat hữu hạn đơn định và lớp otomat hữu hạn không đơn định.

Định lý 3.1

Nếu ngôn ngữ L được đoán nhận bởi một otomat hữu hạn không đơn định thì tồn tại một otomat hữu hạn đơn định đoán nhận L.

Việc chứng minh định lý này được suy từ các thuật toán đơn định hóa otomat.

Định lý 3.2

Lớp ngôn ngữ được sinh bởi otomat hữu hạn đơn định là trùng với lớp ngôn ngữ được sinh bởi otomat hữu hạn không đơn định.

Chứng minh: Ta gọi L_N là lớp ngôn ngữ sinh bởi các otomat hữu hạn không đơn định, L_D là lớp ngôn ngữ sinh bởi các otomat hữu hạn đơn định, ta cần chứng minh $L_N = L_D$. Ta sẽ chứng minh hai bao hàm thức:

- $L_N \subseteq L_D$. Giả sử L là một ngôn ngữ tùy ý thuộc lớp L_N, tức là tồn tại một otomat không đơn định A đoán nhận L, tức là ta có T(A) = L. Theo định lý 2.1, tồn tại một otomat đơn định M tương đương với otomat A, cũng đoán nhận L, tức là L = T(M), vậy L thuộc lớp L_D, do đó ta có $L_N \subseteq L_D$.

- $L_D \subseteq L_N$. Giả sử L là một ngôn ngữ tùy ý thuộc lớp L_D, tức là tồn tại một otomat đơn định M đoán nhận L, ta có T(M) = L. Tuy nhiên, ta luôn luôn có thể xem hàm chuyển đơn định $\delta(q, a) = p \in Q$ trong otomat đơn định như là một trường hợp đơn giản của hàm chuyển không đơn định $\delta(q, a) = \{p\} \in 2^Q$ trong otomat không đơn định. Như vậy, một otomat đơn định có thể được xem là một trường hợp đặc biệt của otomat không đơn định Và vì thế, ngôn ngữ L nói trên có thể xem là được đoán nhận bởi otomat không đơn định. Do đó $L_D \subseteq L_N$.

Từ đó ta có $L_D = L_N$.

Định lý được chứng minh.

Như vậy, từ đây về sau, nếu có một otomat hữu hạn A sao cho L = T(A), thì ta nói ngôn ngữ L được đoán nhận bởi otomat hữu hạn, mà không phân biệt L được đoán nhận bởi otomat đơn định hay không đơn định.

3.3. OTOMAT HỮU HẠN VÀ NGÔN NGỮ CHÍNH QUY

3.3.1. Ngôn ngữ chính quy và biểu thức chính quy

Trong chương trước, theo cách phân loại của Chomsky, thì ngôn ngữ chính quy là ngôn ngữ được sinh bởi các văn phạm chính quy. Trong phần này, ta sẽ định nghĩa các ngôn ngữ chính quy như là các tập hợp từ đặc biệt, được gọi là các tập chính quy. Ta cũng sẽ chỉ ra rằng định nghĩa ngôn ngữ chính quy ở đây là tương đương với định nghĩa ngôn ngữ chính quy thông qua văn phạm chính quy. Đồng thời với các ngôn ngữ chính quy, chúng ta đưa ra các khái niệm về biểu thức chính quy, là công cụ để biểu diễn các ngôn ngữ chính quy.

Định nghĩa 3.8

Cho bảng chữ cái Σ, khi đó ngôn ngữ chính quy (regular languages) được định nghĩa đệ quy như sau:

1. Ngôn ngữ rỗng, $L= \emptyset$ là một ngôn ngữ chính quy

2. Ngôn ngữ chỉ gồm một từ rỗng, $L= \{\varepsilon\}$ là một ngôn ngữ chính quy

3. Ngôn ngữ chỉ gồm một từ là một ký hiệu, $L = \{a\}$, với $a \in \Sigma$, là ngôn ngữ chính quy trên bảng chữ cái Σ.

4. Nếu L_1 và L_2 là hai ngôn ngữ chính quy trên bảng chữ cái Σ thì:

a. $L = L_1 \cup L_2$ là một ngôn ngữ chính quy trên bảng chữ cái Σ,

b. $L = L_1.L_2$ là một ngôn ngữ chính quy trên bảng chữ cái Σ,

c. $L = L_1^*$ (hay L_2^*) là các ngôn ngữ chính quy trên bảng chữ cái Σ.

5. Các ngôn ngữ nhận được bằng cách áp dụng bất kỳ quy tắc nào từ 1-4 trên đây là các ngôn ngữ chính quy.

Có thể chứng minh được rằng ngôn ngữ chính quy theo định nghĩa trên là tương đương với ngôn ngữ chính quy được sinh bởi các văn phạm chính quy.

Thật vậy, có thể xây dựng các văn phạm chính quy sinh ra các ngôn ngữ $\emptyset$ và ngôn ngữ $\{a\}$ (xem thí dụ 2.18). Hơn nữa, trong chương 2 cũng đã chỉ ra rằng lớp các ngôn ngữ chính quy là đóng đối với các phép toán hợp, nhân ghép và lặp trên các ngôn ngữ. Như vậy lớp ngôn ngữ chính quy theo định nghĩa trên đây là trùng với lớp ngôn ngữ chính quy đã được định nghĩa theo văn phạm chính quy.

Người ta quy ước tập chỉ gồm một từ rỗng $\{\varepsilon\}$là một ngôn ngữ chính quy. Ngôn ngữ chính quy có chứa từ rỗng ε, được gọi là các ngôn ngữ chính quy suy rộng. Trong hầu hết các trường hợp, khi không cần phân biệt, để đơn giản, ta dùng khái niệm "ngôn ngữ chính quy" chung cho cả "ngôn ngữ chính quy" và "ngôn ngữ chính quy suy rộng".

Như vậy, từ định nghĩa 3.8, ta có định lý sau:

Định lý 3.3

Mọi ngôn ngữ chính quy trên bảng chữ cái Σ đều nhận được từ các ngôn ngữ hữu hạn bằng cách áp dụng một số hữu hạn lần các phép toán hợp, nhân ghép và phép lặp.

Để diễn đạt các ngôn ngữ chính quy, ta đưa vào khái niệm biểu thức chính quy, được định nghĩa như sau:

Định nghĩa 3.9

Cho bảng chữ cái Σ, khi đó biểu thức chính quy (regular expressions) được định nghĩa đệ quy như sau:

1. $r = \emptyset$ là một biểu thức chính quy biểu diễn ngôn ngữ $L = \emptyset$ (ngôn ngữ rỗng)

2. $r = \varepsilon$ là một biểu thức chính quy biểu diễn ngôn ngữ $L = \{\varepsilon\}$

3. $r = a$ là một biểu thức chính quy biểu diễn ngôn ngữ $L = \{a\}$, với $a \in \Sigma$,.

4. Nếu r_1 và r_2 là hai biểu thức chính quy biểu diễn các ngôn ngữ chính quy L_1 và L_2 trên bảng chữ cái Σ thì:

a. $r = r_1 + r_2$ là biểu thức chính quy biểu diễn ngôn ngữ $L_1 \cup L_2$ trên bảng chữ cái Σ

b. $r = r_1.r_2$ là biểu thức chính quy biểu diễn ngôn ngữ $L_1.L_2$ trên bảng chữ cái Σ

c. $r = r_1^*$ (hay r_2^*) là biểu thức chính quy biểu diễn ngôn ngữ L_1^* (hay L_2^*) trên bảng chữ cái Σ

5. Các biểu thức nhận được bằng cách áp dụng bất kỳ quy tắc nào từ 1-4 trên đây là các biểu thức chính quy.

▪ Nếu r là biểu thức chính quy biểu điễn ngôn ngữ L thì ta viết $L = L(r)$.

Từ định nghĩa ngôn ngữ chính quy và biểu thức chính quy, dễ dàng suy ra định lý sau về các ngôn ngữ chính quy:

Định lý 3.4

Một ngôn ngữ trên bảng chữ cái Σ là chính quy khi và chỉ khi nó được biểu diễn bằng một biểu thức chính quy.

Chú ý:

1. Biểu thức chính quy có chứa ε và phép toán lặp (*) gọi là biểu thức chính quy suy rộng. Trong hầu hết các trường hợp, khi không cần phân biệt, ta dùng khái niệm "biểu thức chính quy" chung cho cả "biểu thức chính quy" và "biểu thức chính quy suy rộng"

2. Trong các biểu thức chính quy ta có thể bỏ qua các dấu ngoặc và quy ước thứ tự ưu tiên của các phép toán là phép lặp, phép nhân ghép và cuối cùng là các phép hợp.

3. Hai biểu thức chính quy được xem là bằng nhau nếu chúng biểu diễn cùng một ngôn ngữ chính quy, tức là, nếu r và s là các biểu thức chính quy, thì: r = s khi và chỉ khi L(r) = L(s)

Nếu r, s, t là các biểu thức chính quy thì ta có các kết quả sau:

- r + s = s + r,
- (r + s) + t = r + (s + t),
- r + r = r,
- (rs)t = r(st),
- r(s + t) = rs + rt, (s + t)r = sr + tr,
- $\varnothing^* = \varepsilon$,
- $(r^*)^* = r^*$.

Có thể chứng minh các kết quả trên bằng cách chỉ ra rằng hai biểu thức chính quy ở hai vế của mỗi đẳng thức đều biểu diễn cùng một ngôn ngữ chính quy. Xin dành việc chứng minh này cho sinh viên như là bài tập.

Thí dụ 3.8: Bảng dưới đây trình bày một số biểu thức chính quy thường gặp và mô tả ngôn ngữ chính quy tương ứng :

Biểu thức chính quy	**Ngôn ngữ chính quy tương ứng**
r = (a+b)*	Tập các xâu gồm các ký hiệu 'a'và 'b' có độ dài tùy ý, kể cả xâu rỗng.
r = (a+b)*abb	Tập các xâu gồm các ký hiệu 'a'và 'b' và được kết thúc bởi xâu con abb
r = ab(a+b)*	Tập các xâu gồm các ký hiệu 'a'và 'b' và được bắt đầu bởi xâu con ab.

r =(a+b)*aa(a+b)*	Tập các xâu gồm các ký hiệu ‘a’và ‘b’ và chứa xâu con aa.
r = a*b*c*	Tập các xâu gồm một số bất kỳ ký hiệu ‘a’, tiếp theo là một xâu gồm một số bất kỳ ký hiệu ‘b’, rồi đến một xâu gồm một số bất kỳ ký hiệu ‘c’. Có thể là xâu rỗng.
r = $a^+b^+c^+$	Tập các xâu gồm ít nhất một ký hiệu ‘a’, tiếp theo là một xâu gồm ít nhất một ký hiệu ‘b’, rồi đến một xâu gồm ít nhất một ký hiệu ‘c’.
r = aa*bb*cc*	Tập các xâu gồm ít nhất một ký hiệu ‘a’, tiếp theo là một xâu gồm ít nhất một ký hiệu ‘b’, rồi đến một xâu gồm ít nhất một ký hiệu ‘c’.
r = (a+b)* (a + bb)	Tập các xâu gồm các ký hiệu ‘a’và ‘b’ và được kết thúc bởi ‘a’ hoặc bb.
r = (aa)*(bb)*b	Tập các xâu chứa một số chẵn ký hiệu ‘a’ (có thể không có ký hiệu ‘a’nào) rồi đến một số lẻ ký hiệu ‘b’.
r = (0+1)*000	Tập các xâu chứa các ký hiệu ‘0’ và ‘1’ và kết thúc bởi ‘000’.
r = (11)*	Tập các xâu gồm một số chẵn ký hiệu ‘1’.

Hình 3.21. Một số biểu thức chính quy

Thí dụ 3.9: Xác định ngôn ngữ chính quy được biểu diễn bởi biểu thức chính quy $r = (01^*+02)1$.

Ta có:

$r = (01^*+02)1 = 01^*1+021$,

vậy ngôn ngữ chính quy biểu diễn bởi r là:

$L(r) = L(01^*1+021) = L(01^*1) \cup L(021) = \{01^n, 021 \mid n \geq 1\}$

3.3.2. Sự liên hệ giữa otomat hữu hạn và ngôn ngữ chính quy

Trong chương trước ta đã thấy rằng với mọi ngôn ngữ chính quy đều tồn tại một văn phạm chính quy sinh ra nó, và ngược lại ngôn ngữ sinh bởi văn phạm chính quy là ngôn ngữ chính quy.

Trong phần này, ta sẽ thấy có một sự liên hệ tương tự như vậy giữa otomat hữu hạn và ngôn ngữ chính quy.

Định lý 3.5

Nếu L là một ngôn ngữ chính quy thì tồn tại một otomat hữu hạn không đơn định A đoán nhận L, tức là T(A) = L.

Chứng minh: Giả sử L là ngôn ngữ chính quy sinh bởi văn pham chính quy G = $<\Sigma, \Delta, S, P>$, tức là L = L(G). Ta xây dựng otomat hữu hạn không đơn định A = $< Q, \Sigma, \delta, q_0, F>$, theo quy tắc:

- $Q = \Delta \cup \{E\}$, với E là ký hiệu mới, $E \notin \Sigma \cup \Delta$,
- $q_0 = S$;
- $F = \begin{cases} \{E\} \text{ - nếu quy tắc } S \to \varepsilon \notin P \\ \{E, S\} \text{ - nếu có quy tắc } S \to \varepsilon \in P, \end{cases}$
- Xác định hàm chuyển δ từ văn phạm G:

  - $\forall A \in \Delta, \forall a \in \Sigma$, đặt $\delta(A, a) = \{B \in \Delta \mid \text{nếu có quy tắc } A \to aB\} \cup \{E \mid \text{nếu có quy tắc } A \to a\}$,

  - $\forall a \in \Sigma$ đặt $\delta(E, a) = \emptyset$.

Otomat A xây dựng như trên là một otomat hữu hạn không đơn định, ta sẽ chứng minh L = T(A):

1. Lấy $\omega \in L$, ta chứng minh rằng $\omega \in T(A)$:

- nếu $\omega = \varepsilon$: trong văn phạm G có quy tắc $S \to \varepsilon \in P$, do đó $S \in F$. Trong trường hợp này $\delta(S, \varepsilon) = \{S\} \subseteq F$ nên $\varepsilon \in T(A)$.

- nếu $\omega = a_1a_2\ldots a_n \neq \varepsilon$: Do G là văn phạm chính quy, nên tồn tại dãy quy tắc $S \to a_1A_1$, $A_1 \to a_2A_2, \ldots, A_{n-1} \to a_n$ trong P sao cho $S \vdash a_1A_1 \vdash a_1a_2A_2 \vdash \ldots \vdash a_1a_2\ldots a_{n-1}A_{n-1} \vdash a_1\ldots a_{n-1}a_n$

- Từ định nghĩa của δ, ta có $A_1 \in \delta(S, a_1)$, $A_2 \in \delta(A_1, a_2), \ldots, A_{n-1} \in \delta(A_{n-2}, a_{n-1})$, $E \in \delta(A_{n-1}, a_n)$. Như vậy, $E \in \delta(S, a_1a_2\ldots a_n)$ hay $\delta(q_0, \omega) \cap F \neq \emptyset$, do đó $\omega \in T(A)$. Vậy $L \subseteq T(A)$.

2. Lấy $\omega \in T(A)$, ta chứng minh rằng $\omega \in L(G)$:

- nếu $\omega = \varepsilon$: $\delta(S, \varepsilon) \cap F \neq \emptyset$ hay $S \in F$, vậy có quy tắc $S \to \varepsilon \in P$, do đó $\varepsilon \in L(G)$

• nếu $\omega = a_1a_2\ldots a_n \neq \varepsilon$: $\delta(S, \omega)\cap F \neq \varnothing$ với $\omega \neq \varepsilon$ hay $E \in \delta(S, \omega)$, do đó tồn tại các trạng thái $A_1, A_2,\ldots, A_{n-1}\in\Delta$ sao cho $A_1\in \delta(S, a_1)$, $A_2\in \delta(A_1, a_2),\ldots, A_{n-1}\in \delta(A_{n-2}, a_{n-1})$, $E\in\delta(A_{n-1}, a_n)$. Từ đó ta có $S\rightarrow a_1A_1$, $A_1\rightarrow a_2A_2,\ldots, A_{n-1}\rightarrow a_n\in P$ hay trong G có một suy dẫn là $S \vdash a_1A_1 \vdash a_1a_2A_2 \vdash\ldots\vdash a_1a_2\ldots a_{n-1}A_{n-1} \vdash a_1\ldots a_{n-1}a_n = \omega$. Vì vậy $\omega\in L$. Hay $T(A) \subseteq L$

Vậy ta đã chứng minh L = T(A), tức là tồn tại một otomat hữu hạn không đơn định đoán nhận L.

Như vậy, nếu gọi $\mathcal{L}_N$ là lớp các ngôn ngữ được đoán nhận bởi otomat hữu hạn không đơn định và $\mathcal{L}_R$ là lớp các ngôn ngữ chính quy thì từ định lý trên ta có $\mathcal{L}_R\subset \mathcal{L}_N$,

Định lý trên cho phép ta xây dựng một otomat hữu hạn không đơn định A tương đương với văn phạm chính quy G cho trước, sao cho T(A) = L(G).

Thí dụ 3.10: Cho ngôn ngữ $L = \{\omega ab^nab \mid n\geq 0, \omega\in \{a, b\}^*\}$. Hãy xác định otomat hữu hạn đoán nhận ngôn ngữ L.

Ta có văn phạm $G = <\{a, b\}, \{S, A, B\}, S, \{S\rightarrow aS, S\rightarrow bS, S\rightarrow aA, A\rightarrow bA, A\rightarrow aB, B\rightarrow b\}>$ là văn phạm chính quy, và dễ thấy rằng L = L(G).

Từ văn phạm chính quy G, ta sẽ xây dựng được otomat không đơn định A để T(A) = L(G) = L. Theo quy tắc trong định lý trên, ta có:

$$A = <\{S, A, B, E\}, \{a, b\}, \delta, S, \{E\}>,$$

với hàm chuyển δ được xác định như sau:

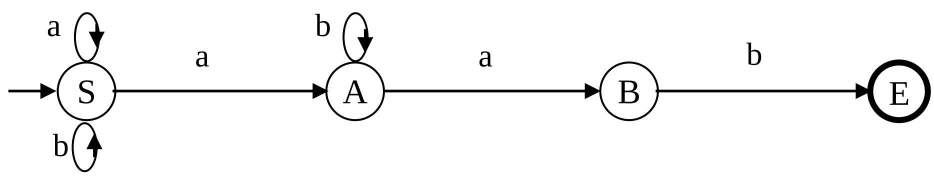
$\delta(S, a) = \{S, A\}$; $\delta(S, b) = \{S\}$,

$\delta(A, a) = \{B\}$; $\delta(A, b) = \{A\}$,

$\delta(B, a) = \varnothing$; $\delta(B, b) = \{E\}$,

$\delta(E, a) = \varnothing$; $\delta(E, b) = \varnothing$.

Đồ thị chuyển của A được cho trong hình 3.22:



Hình 3.22. Đồ thị chuyển của otomat A trong thí dụ 3.10

Theo định lý trên, otomat A đoán nhận ngôn ngữ chính quy L, thật vậy ta có:

$$T(A) = \{\omega ab^nab \mid n \geq 0, \omega\in\{a, b\}^*\} = L$$

Thí dụ 3.11:. Cho văn phạm chính quy G = <{0, 1}, {S, S_1 }, S, P>, với tập quy tắc :

$$P: \left| \begin{array}{l} S \rightarrow 0S, \\ S \rightarrow 0S_1, \\ S_1 \rightarrow 1S_1, \\ S_1 \rightarrow 1. \end{array} \right.$$

Hãy xây dựng otomat hữu hạn A sao cho T(A) = L(G), xác định ngôn ngữ được đoán nhận bởi văn phạm nói trên.

Theo quy tắc trong định lý 3.5, ta có A = < Q, Σ, δ, q_0, F>, với các thành phần như sau:

- Q = {S, S_1, E},
- Σ = {0, 1},
- q_0 = S,
- F = {E} (do trong G không có quy tắc S→ ε)
- Xác định hàm chuyển δ:

  δ(S, 0) = {S, S_1}; δ(S, 1) = ∅,

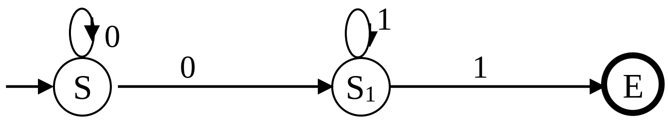
  δ(S_1, 0) = ∅; δ(S_1, 1) = {S_1, E},

  δ(E, 0) = ∅; δ(E, 1) = ∅.

Vậy otomat hữu hạn cùng đoán nhận ngôn ngữ sinh bởi văn phạm G là bộ năm:

A = <{S, S_1, E}, {0, 1}, δ, S, {E}>

Đồ thị chuyển trạng thái của otomat A là:



Hình 3.23. Đồ thị chuyển của otomat A trong thí dụ 3.11

Từ đồ thị chuyển của A, dễ thấy rằng T(A) = {0^n1^m | với n, m ≥ 1}, vậy ngôn ngữ sinh bởi văn phạm G là L(G) = T(A) = {0^n1^m | với n, m ≥ 1}.

Định lý 3.6

Nếu L là ngôn ngữ được đoán nhận bởi một otomat hữu hạn đơn định thì L là một ngôn ngữ chính quy.

Chứng minh: Giả sử L = T(M), với M = <Q, Σ, δ, q_0, F> là một otomat hữu hạn đơn định. Từ otomat M, ta xây dựng văn phạm G = <Σ, Δ, S, P>, theo quy tắc sau:

- Δ = Q,
- S = q_0,
- Tập quy tắc P = {q→ ap | nếu có δ(q, a) = p} ∪ {q→ a | nếu có δ(q, a) = p∈ F}.

Khi đó G chỉ có hai loại quy tắc q→ ap, q→ a, với p, q∈ Δ, a∈ Σ, vậy G là văn phạm chính quy.

1/. Ta chứng minh L(G) = L, với giả thiết ε∉ L.

- Lấy ω = $a_1a_2\ldots a_n$∈ L(G), ω ≠ ε, do ω∈ L(G) nên trong văn phạm G tồn tại suy dẫn $q_0 \models \omega$ hay $q_0 \vdash a_1q_1 \vdash a_1a_2q_2 \vdash\ldots \vdash a_1a_2\ldots a_{n-1}q_{n-1} \vdash a_1\ldots a_{n-1}a_n = \omega$.

Do đó ta có các quy tắc trong P: $q_0\to a_1q_1$, $q_1\to a_2q_2,\ldots, q_{n-1}\to a_{n-1}q_{n-1}$, $q_{n-1}\to a_n$, từ đó suy ra các giá trị của hàm chuyển trong otomat M:

$$\delta(q_0, a_1) = p_1,\ \delta(q_1, a_2) = p_2,\ldots, \delta(q_{n-2}, a_{n-1}) = q_{n-1},\ \delta(q_{n-1}, a_n) = q_n\in F$$

tức là $\delta(q_0, \omega) = q_n\in F$ hay ω∈ T(M) = L

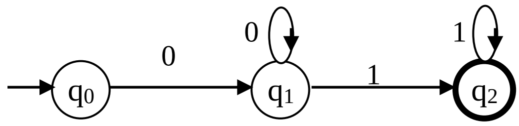
- Lấy ω = $a_1a_2\ldots a_n$∈ L, ω ≠ ε, do ω được đoán nhận bởi M, nên theo định nghĩa ta có $\delta(q_0, \omega) = q_n\in F$. Khi đó tồn tại dãy trạng thái $q_1, q_2,\ldots, q_n$ sao cho $\delta(q_0, a_1) = p_1,$ $\delta(q_1, a_2) = q_2,\ldots, \delta(q_{n-2}, a_{n-1}) = q_{n-1}$, $\delta(q_{n-1}, a_n) = q_n\in F$. Theo cách xây dựng văn phạm G, trong P sẽ có các quy tắc $q_0\to a_1q_1$, $q_1\to a_2q_2,\ldots, q_{n-1}\to a_{n-1}q_{n-1}$, $q_{n-1}\to a_n$, như vậy ta có các suy dẫn trong G: $q_0 \vdash a_1q_1 \vdash a_1a_2q_2 \vdash\ldots \vdash a_1a_2\ldots a_{n-1}q_{n-1} \vdash a_1\ldots a_{n-1}a_n = \omega$, hay $q_0 \models \omega$, vậy ω∈ L(G).

2/. Trong trường hợp ε∈ L, ta xây dựng G' tương đương với G trong đó ký hiệu xuất phát không xuất hiện trong bất kỳ vế phải của quy tắc nào, đồng thời thêm vào G' quy tắc $q_0\to \varepsilon$ để nhận được văn phạm chính quy G' sao cho L(G') = L(G) ∪ {ε}. Vậy ta luôn có L(G) = L. Vậy định lý được chứng minh.

Như vậy, nếu gọi L_D là lớp các ngôn ngữ được đoán nhận bởi otomat hữu hạn đơn định và L_R là lớp các ngôn ngữ chính quy thì từ định lý trên ta có $L_D \subset L_R$,

Định lý trên cho phép ta xây dựng một văn phạm chính quy G tương đương với một otomat hữu hạn đơn định A cho trước, sao cho L(G) = T(A). (Obtain RG from DFA)

Thí dụ 3.12: Cho otomat hữu hạn đơn định A có đồ thị chuyển như hình dưới đây:



Hình 3.24. Đồ thị chuyển của otomat A trong thí dụ 3.12

Hãy xây dựng văn phạm chính quy cùng sinh ngôn ngữ được đoán nhận bởi otomat A.

Từ đồ thị chuyển, ta có thể xác định được otomat A = <{q_0, q_1, q_2}, {0, 1}, δ, q_0, {q_2}>, với hàm chuyển xác định như sau:

$\delta(q_0, 0) = q_1$;

$\delta(q_1, 0) = q_1$; $\delta(q_1, 1) = q_2$,

$\delta(q_2, 1) = q_2$.

Theo thuật toán trong định lý 3.6, ta xây dựng được văn phạm chính quy:

G = <{0, 1}, {q_0, q_1, q_2}, q_0 , P >, với tập quy tắc sinh như sau:

$$P: \left| \begin{array}{l} q_0 \to 0q_1, \\ q_1 \to 0q_1 \mid 1q_2 \mid 1, \\ q_2 \to 1q_2 \mid 1. \end{array} \right.$$

Dễ dàng kiểm tra được rằng L(G) = T(A) = $\{0^n1^m \mid$ với $m, n \geq 1\}$

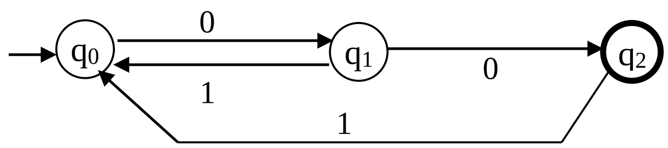
Thí dụ 3.13: Cho otomat hữu hạn đơn định A = <{q_0, q_1, q_2}, {0, 1}, δ, q_0, {q_2}>, trong đó:

$\delta(q_0, 0) = q_1$;

$\delta(q_1, 0) = q_2$; $\delta(q_1, 1) = q_0$,

$\delta(q_2, 1) = q_0$.

Hãy xây dựng văn phạm chính quy cùng sinh ngôn ngữ được đoán nhận bởi otomat A. Ta có đồ thị chuyển của otomat A là:



Hình 3.25. Đồ thị chuyển của otomat A trong thí dụ 3.13

Dễ thấy rằng L = T(A) = $\{(01)^n00, \{\omega 00 \mid \omega \in \{01, 001\}^*\}$.

Ta sẽ xây dựng văn phạm G = <Σ , Δ, S, P> theo quy tắc trong định lý trên

- Δ = {q_0, q_1, q_2},
- S = q_0,
- Tập quy tắc P = {$q_0 \to 0q_1$, $q_1 \to 0q_2$, $q_1 \to 0$, $q_1 \to 1q_0$, $q_2 \to 1q_0$ }.

Như vậy G là văn phạm chính quy, và dễ thấy rằng L(G) = L, vậy L là ngôn ngữ chính quy.

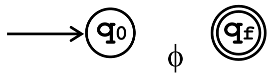
3.3.3. Sự liên hệ giữa otomat hữu hạn và biểu thức chính quy

Định lý 3.7

Cho r là một biểu thức chính quy, khi đó sẽ tồn tại một otomat hữu hạn không đơn định M đoán nhận ngôn ngữ chính quy được xác định bởi r, tức là T(M) = L(r).

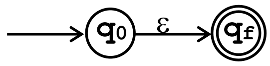
Chứng minh. Dựa trên định nghĩa biểu thức chính quy, ta sẽ xây dựng các thuật toán để xác định otomat không đơn định M từ các biểu thức chính quy tương ứng như sau:

1. r = ∅. Otomat M cần xây dựng có 2 trạng thái $\mathbf{q_0}$ và $\mathbf{q_f}$ với đồ thị chuyển như sau:



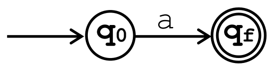
***Hình 3.26. Đồ thị chuyển của otomat tương ứng với biểu thức chính quy* r = ∅**

2. r = ε. Otomat M cần xây dựng có 2 trạng thái $\mathbf{q_0}$ và $\mathbf{q_f}$ với đồ thị chuyển như sau:



***Hình 3.27. Đồ thị chuyển của otomat tương ứng với biểu thức chính quy* r = ε**

3. r = *a*. Otomat M cần xây dựng có 2 trạng thái $\mathbf{q_0}$ và $\mathbf{q_f}$ với đồ thị chuyển như sau:



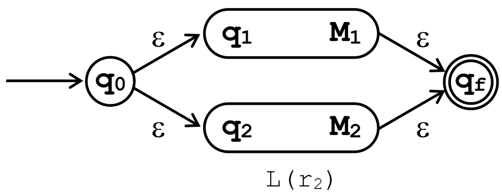
***Hình 3.28. Đồ thị chuyển của otomat tương ứng với biểu thức chính quy* r = a**

4. Giả sử đã xây dựng được các otomat M, M_1 và M_2 tương ứng với các biểu thức chính quy r, r_1 và r_2, ta sẽ xây dựng các otomat tương ứng với các biểu thức chính quy $r_1 + r_2$; $r_1.r_2$ và r* như sau :

a. Otomat ứng với $r_1 + r_2$, ký hiệu là $M_1 \cup M_2$, được xây dựng theo sơ đồ ở hình 3.29.

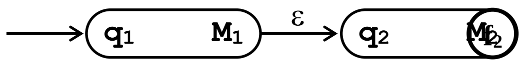
Trong sơ đồ này, M_1 là otomat tương ứng với biểu thức chính quy r_1 với $\mathbf{q_1}$ và $\mathbf{f_1}$ lần lượt là các trạng thái khởi đầu và kết thúc của M_1. Tương tự, M_2 là otomat tương ứng với biểu thức chính quy r_2 với $\mathbf{q_2}$ và $\mathbf{f_2}$ lần lượt là các trạng thái khởi đầu và kết thúc của M_2.

Để xây dựng otomat $M_1 \cup M_2$, ta thêm vào một trạng thái q_0 làm trạng thái xuất phát và trạng thái q_f làm trạng thái kết thúc duy nhất của otomat này, và thêm các cung rỗng nối các trạng thái mới với các otomat đã cho. Một xâu được đoán nhận bởi otomat hợp $M_1 \cup M_2$ thì hoặc được đoán nhận bởi M_1 hoặc được đoán nhận bởi M_2, tức là xâu đó thuộc $L(r_1) \cup L(r_2)$, tương ứng với biểu thức chính quy $r_1 + r_2$. Do vậy, $T(M_1 \cup M_2) = L(r_1 + r_2)$.



Hình 3.29. Đồ thị chuyển của otomat tương ứng với biểu thức chính quy $r = r_1 + r_2$

b. Otomat ứng với $r_1. r_2$, ký hiệu là $M_1.M_2$, được xây dựng theo sơ đồ sau:

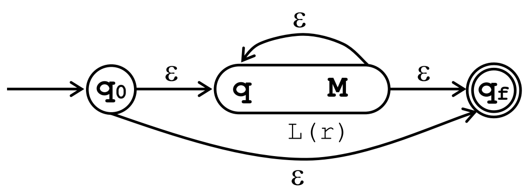


Hình 3.30. Đồ thị chuyển của otomat tương ứng với biểu thức chính quy $r = r_1 . r_2$

Trong sơ đồ trên, M_1 là otomat tương ứng với biểu thức chính quy r_1 với $\mathbf{q_1}$ và $\mathbf{f_1}$ lần lượt là các trạng thái khởi đầu và kết thúc của M_1. Tương tự, M_2 là otomat tương ứng với biểu thức chính quy r_2 với $\mathbf{q_2}$ và $\mathbf{f_2}$ lần lượt là các trạng thái khởi đầu và kết thúc của M_2.

Để xây dựng otomat $M_1.M_2$, ta chọn trạng thái xuất phát q_1 của M_1 làm trạng thái xuất phát của otomat tích ghép, chọn trạng thái kết thúc f_2 của otomat M_2 làm trạng thái kết thúc của otomat mới. Dễ thấy rằng $T(M_1.M_2) = L(r_1.r_2)$.

c. Otomat ứng với r*, ký hiệu là M* , được xây dựng theo sơ đồ sau:



Hình 3.31. Đồ thị chuyển của otomat tương ứng với biểu thức chính quy r = r*

Trong sơ đồ trên, M là otomat tương ứng với biểu thức chính quy r với **q** và **f** lần lượt là các trạng thái khởi đầu và kết thúc của M. Để xây dựng otomat M*, ta thêm vào một trạng thái q_0 làm trạng thái xuất phát và trạng thái q_f làm trạng thái kết thúc duy nhất của otomat mới, và thêm các cung rỗng nối các trạng thái mới với các otomat ban đầu, và tạo nên các ngôn ngữ lặp.

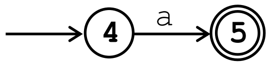
Trong otomat ở hình 3.31, nếu ta bỏ đi cung rỗng nối từ $\mathbf{q_0}$ đến $\mathbf{q_f}$, ta sẽ nhận được otomat M^+, tương ứng với biểu thức chính quy r^+.

Các thuật toán trong định lý trên còn được gọi là các thuật toán Thompson, cho phép xây dựng các otomat hữu hạn không đơn định từ các biểu thức chính quy *(Obtain NFA from Regular Expression)*

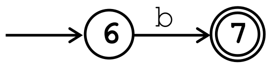
Thí dụ 3.14: Xây dựng otomat hữu hạn không đơn định (NFA) đoán nhận ngôn ngữ L gồm mọi xâu trên bảng chữ cái {a, b}, và được bắt đầu bởi 'ab'.

Dễ thấy rằng biểu thức chính quy tương ứng của L là : r = ab(a+b)*. Dựa trên biểu thức chính quy r , áp dụng thuật toán Thompson. ta sẽ xây dựng otomat tương ứng với r theo các bước dưới đây, chú ý rằng sau mỗi bước thì trạng thái xuất phát và trạng thái kết thúc được thay đổi. Để cho gọn, ta dùng các chữ số để ký hiệu các trạng thái.

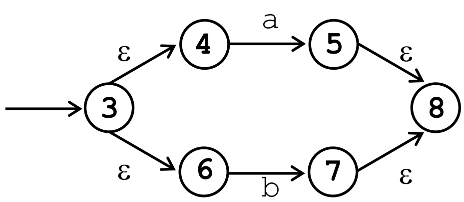
Bước 1: Xây dựng otomat tương ứng với biểu thức chính quy 'a':



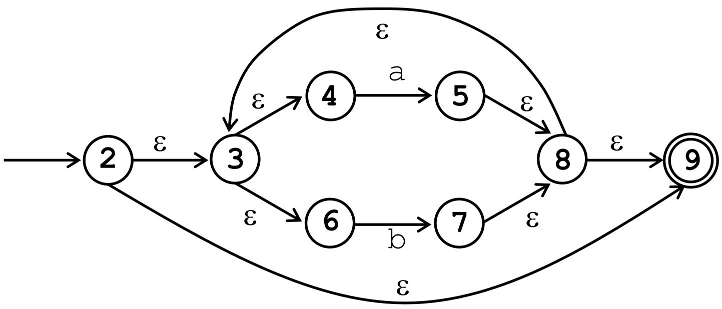
Bước 2: Xây dựng otomat tương ứng với biểu thức chính quy 'b':



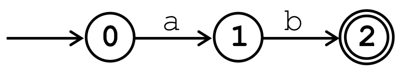
Bước 3: Xây dựng otomat tương ứng với biểu thức chính quy 'a + b':



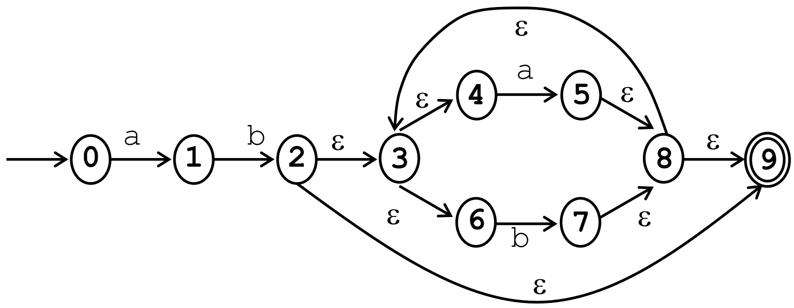
Bước 4: Xây dựng otomat tương ứng với biểu thức chính quy (a + b)*:



Bước 5: Xây dựng otomat tương ứng với biểu thức chính quy 'ab':



Bước 6: Nhân ghép otomat ở bước 5 với otomat được xây dựng ở bước 4, ta được otomat cuối cùng M, với trạng thái xuất phát là **0**, trạng thái kết thúc duy nhất là **9**.



Hình 3.32. Đồ thị chuyển của otomat M tương ứng với biểu thức chính quy **r = ab(a+b)***

Theo cách xây dựng trên, otomat M tương đương với biểu thức chính quy r = ab(a+b)*. Dễ dàng thấy rằng :

$$T(M) = \{ab\omega \mid \text{với } \omega \in \{a, b\}^*\}.$$

Chú ý. Ta có thể rút gọn các otomat để được otomat tương đương :

1. Khi một đỉnh q có cung rỗng (cung ghi ký hiệu ε) tới đỉnh kết thúc p thì đỉnh q được coi là đỉnh kết thúc, và xóa bỏ cung rỗng từ q tới p.

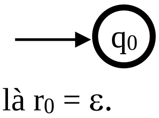
2. Khi một đỉnh trở thành đỉnh cô lập thì có thể xóa bỏ đỉnh đó khỏi đồ thị chuyển, và xóa bỏ trạng thái ở đỉnh đó khỏi otomat.

3. Một ngôn ngữ chính quy có chứa từ rỗng ε, thì trong otomat tương ứng, trạng thái xuất phát luôn được coi đồng thời là trạng thái kết thúc.

Thí dụ 3.15: Xây dựng otomat hữu hạn không đơn định (NFA), biểu thức chính quy và văn phạm chính quy xác định một số ngôn ngữ :

1. $L_0 = \{\varepsilon\}$, với $\Sigma = \{0, 1\}$:

▪ Otomat chấp nhận ngôn ngữ L_0 có đồ thị chuyển như trong định lý 3.7 (H.3.27), hoặc có thể gộp 2 trạng thái thành q_0, vừa là trạng thái khởi đầu, vừa là trạng thái kết thúc :

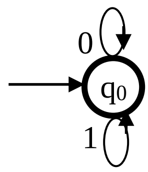


▪ Biểu thức chính quy xác định ngôn ngữ $L_0 = \{\varepsilon\}$ là $r_0 = \varepsilon$.

▪ Văn phạm chính quy sinh ngôn ngữ $L_0 = \{\varepsilon\}$ là : $G_0 = <\{0, 1\}, \{S\}, S, \{S \rightarrow \varepsilon\}>$

2. $L_1 = \Sigma^*$, với $\Sigma = \{0, 1\}$:

▪ Otomat chấp nhận ngôn ngữ L_1 có đồ thị chuyển như sau, ở đây q_0 vừa là trạng thái khởi đầu, vừa là trạng thái kết thúc :

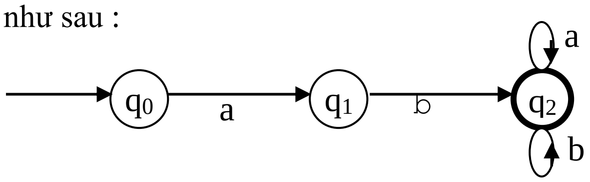


▪ Biểu thức chính quy xác định ngôn ngữ $L_1 = \Sigma^*$ là : $r_1 = (0+1)^*$.

▪ Văn phạm chính quy sinh ngôn ngữ $L_1 = \Sigma^*$ là:

$$G_1 = <\{0, 1, \{S\}, S, \{S \rightarrow 0S \mid 1S, S \rightarrow \varepsilon\} >$$

Theo các cách rút gọn trên đây, otomat M ở thí dụ 3.14 sẽ tương đương với otomat có đồ thị chuyển như sau :



Hình 3.33. Đồ thị chuyển rút gọn của otomat M trong thí dụ 3.14

Kết luận. Từ các định lý trên ta có kết luận về sự liên hệ giữa otomat hữu hạn và ngôn ngữ chính quy như sau:

1/. Gọi L_D là lớp các ngôn ngữ được đoán nhận bởi otomat hữu hạn đơn định, L_N là lớp các ngôn ngữ được đoán nhận bởi otomat hữu hạn không đơn định và L_R là lớp các ngôn ngữ chính quy.

Định lý 3.2 cho biết $L_D = L_N$,

Định lý 3.5 cho biết $L_R \subset L_N$,

Định lý 3.6 cho biết $L_D \subset L_R$,

Vậy $L_D = L_N = L_R$, tức là lớp các ngôn ngữ sinh bởi các otomat đơn định, lớp các ngôn ngữ sinh bởi các otomat không đơn định và lớp các ngôn ngữ chính quy là trùng nhau.

2/. Ngôn ngữ L là chính quy khi và chỉ khi:

a/. Tồn tại một biểu thức chính quy biểu diễn L,

b/. Tồn tại một văn phạm chính quy sinh ngôn ngữ L,

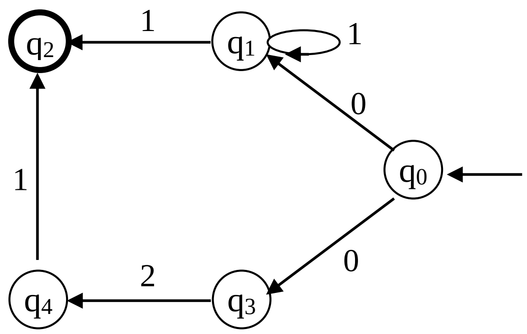
c/. Tồn tại một otomat hữu hạn đoán nhận L

Thí dụ 3.15: Cho ngôn ngữ chính quy L = {01^n, 021 | n ≥ 1}, khi đó ta L sẽ có các cách xác định như sau:

- Biểu thức chính quy biểu diễn L là: r = 01*1+021
- Văn phạm chính quy sinh ngôn ngữ L:

G = <{0, 1, 2}, {S, A, B, C}, S, {S→ 0A, A→ 1A, A→ 1, S→ 0B, B→ 2C, C→ 1}>

- Otomat hữu hạn A đoán nhận L có đồ thị chuyển là:



Hình 3.34. Đồ thị chuyển của otomat A trong thí dụ 3.15

3.4. ĐIỀU KIỆN CẦN CỦA NGÔN NGỮ CHÍNH QUY

Khi một ngôn ngữ được đoán nhận bởi otomat hữu hạn, hoặc được sinh bởi một văn phạm chính quy, hoặc được xác định bởi một biểu thức chính quy thì nó là ngôn ngữ chính quy. Như vậy việc chứng minh một ngôn ngữ là chính quy là khá dễ dàng bằng cách chỉ ra rằng nó được xác định bằng một trong những cách trên. Tuy nhiên, để khẳng định một ngôn ngữ L không phải là ngôn ngữ chính quy thì lại không hề đơn giản. Dù ta không xây dựng được otomat hữu hạn, văn phạm chính quy hay biểu thức chính quy để xác định L, nhưng ta vẫn không thể kết luận được ngôn ngữ này không phải là ngôn ngữ chính quy, bởi vì ta không thể khẳng định được rằng không tồn tại những văn phạm chính quy hay những otomat hữu hạn sinh ra L. Như vậy, cần có một tiêu chuẩn để căn cứ vào đó có thể kết luận một ngôn ngữ không phải là ngôn ngữ chính quy, tiêu chuẩn đó là *điều kiện cần* của ngôn ngữ chính quy.

3.4.1 Otomat tối tiểu

Cùng một ngôn ngữ chính quy L, có thể có nhiều otomat hữu hạn đoán nhận nó. Tuy nhiên, trong số đó, trước hết chúng ta quan tâm đến các otomat có số trạng thái ít nhất cùng đoán nhận ngôn ngữ L. Dưới đây chúng ta chỉ đưa ra định nghĩa và thí dụ về otomat tối tiểu, mà không trình bày thuật toán tìm otomat tối tiểu

Định nghĩa 3.10

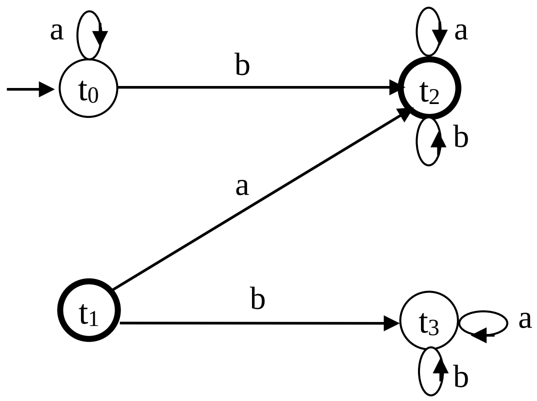
Otomat có số trạng thái ít nhất trong các otomat hữu hạn cùng đoán nhận ngôn ngữ L được gọi là otomat tối tiểu của ngôn ngữ L.

Nhận xét: Dễ thấy rằng với mỗi ngôn ngữ L, otomat tối tiểu của nó có thể không duy nhất.

Thí dụ 3.16: Giả sử ta có otomat M = <Q, {a, b}, δ, t_0, F>, với

- $Q = \{t_0, t_1, t_2, t_3\}$
- $\delta(t_0, a) = t_0$, $\delta(t_0, b) = t_2$, $\delta(t_1, a) = t_2$, $\delta(t_1, b) = t_3$, $\delta(t_2, a) = t_2$, $\delta(t_2, b) = t_2$, $\delta(t_3, a) = t_3$, $\delta(t_3, b) = t_3$.
- $F = \{t_1, t_2\}$.

otomat M là đơn định, có 4 trạng thái và có đồ thị chuyển như sau:

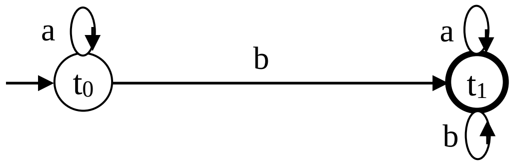


Hình 3.35. Đồ thị chuyển của otomat M trong thí dụ 3.16

Dễ thấy rằng otomat M đoán nhận ngôn ngữ:

$L = T(M) = \{a^n b\omega \mid n \geq 0, \omega \in \{a, b\}^*\}$.

Nhìn vào đồ thị chuyển của M, ta thấy ngay rằng không có đường đi nào từ t_0 đến được đỉnh kết thúc t_1, vì vậy otomat M sẽ tương đương với otomat M' có đồ thị chuyển như sau:



Hình 3.36. Đồ thị chuyển của otomat tối tiểu M' trong thí dụ 3.16

Rõ ràng là otomat M’ cũng đoán nhận ngôn ngữ L = T(M’) = $\{a^n b\omega \mid n \geq 0, \omega \in \{a, b\}^*\}$, M’ chỉ có hai trạng thái và là otomat tối tiểu của ngôn ngữ L = $\{a^n b\omega \mid n \geq 0, \omega \in \{a, b\}^*\}$.

3.4.2. Điều kiện cần của ngôn ngữ chính quy

Định lý 3.7

Nếu L là ngôn ngữ chính quy thì tồn tại số nguyên dương n sao cho với mọi $\omega \in L$ mà $|\omega| \geq n$ đều có thể phân tích được dưới dạng $\omega = uvw$, (với $|v| \geq 1$ hay $v \neq \varepsilon$) mà với mọi chỉ số i = 0, 1. 2,… ta có $uv^iw \in L$

Chứng minh: Vì L là một ngôn ngữ chính quy, khi đó tồn tại một otomat hữu hạn đoán nhận nó. Giả sử L = T(A) , với A = $<Q, \Sigma, \delta, q_0, F>$ là một otomat tối tiểu có n trạng thái, tức là $|Q| = n$. Ta chứng minh n là số tự nhiên cần tìm.

Giả sử $\omega = a_1a_2\ldots a_m \in L$ với $m \geq n$. Khi đó ta có $\delta(q_0, \omega) \in F$, tức là $\exists q_0, q_1\ldots, q_m \in Q$ sao cho $\delta(q_{i-1}, a_i) = q_i$, $1 \leq i \leq m$ và $q_m \in F$. Do $m \geq n$ nên trong dãy $q_0, q_1, \ldots, q_m$ có ít nhất hai trạng thái trùng nhau, giả sử đó là $q_i = q_k$, $i < k \leq n$, (với k là số nhỏ nhất mà ta có $q_i = q_k$)

Đặt $u = a_1\ldots a_i$, $v = a_{i+1}\ldots a_k$, $w = a_{k+1}\ldots a_m$. Ta có $\omega = uvw$, $|v| = |a_{i+1}\ldots a_k| \geq 1$ (do i<k).

Ngoài ra ta có: $\delta(q_0, u) = q_i = q_k = \delta(q_0, uv)$,

theo bổ đề 1.1: $\delta(q_0, u) = \delta(q_0, uv) = \delta(\delta(q_0, u), v) = \delta(\delta(q_0, uv), v) = \delta(q_0, uv^2)$,

Tương tự, ta có: $\delta(q_0, u) = \delta(q_0, uv^2) = \delta(\delta(q_0, u), v^2) = \delta(\delta(q_0, uv), v^2) = \delta(q_0, uv^3)$,

tiếp tục ta được: $\delta(q_0, u) = \delta(q_0, uv^i)$, $\forall i \in N$.

Cuối cùng ta có: $\delta(q_0, uv^iw) = \delta(\delta(q_0, uv^i), w) = \delta(\delta(q_0, uv), w) = \delta(q_0, uvw) \in F$.

Vậy $uv^iw \in L$, $\forall i = 0, 1, 2,\ldots$.

Định lý được chứng minh.

Định lý trên còn được gọi là “*Bổ đề bơm*” cho ngôn ngữ chính quy *(The Pumping Lemma for regular languages)* , là một bổ đề nổi tiếng trong lý thuyết ngôn ngữ hình thức.

Thí dụ 3.17 : Cho ngôn ngữ L = $\{ 0^{k^2} \mid$ với k = 0, 1, 2,...$\}$, (L gồm các xâu ký hiệu ‘0’ có độ dài là số chính phương). Chứng minh rằng L không phải là ngôn ngữ chính quy.

Thật vậy, giả sử ngược lại, L là ngôn ngữ chính quy và có số n như trong định lý 3.7. Xét xâu $\omega = 0^{k^2}$ với $p = n^2$ (p là số chính phương n^2). Theo định lý 3.7, có thể viết ω dưới dạng: ω = u.v.w, với $1 \le |v| \le n$, và $u.v^i.w \in L$, $\forall i = 0, 1, 2, \ldots$

Có thể lấy i = 2, theo định lý 3.7 ta phải có $u.v^2.w \in L$.

Do $n^2 \le n^2 + 2 = |u.v^2.w| \le n^2 + n < (n+1)^2$, mà n^2 và $(n+1)^2$ là hai số chính phương liên tiếp, nên $|u.v^2.w|$ không thể là số chính phương, vì vậy $u.v^2.w \notin L$, trái với định lý 3.7. Vì vậy, L không phải là ngôn ngữ chính quy.

Hệ quả 3.1

Cho A là otomat hữu hạn đơn định có n trạng thái và L là ngôn ngữ được đoán nhận bởi A. Khi đó $L \ne \emptyset$ khi và chỉ khi $\exists\ \omega \in L$ sao $|\omega| < n$.

Chứng minh: Điều kiện đủ là hiển nhiên. Bây giờ cho $L \ne \emptyset$. Giả sử mọi từ trong L đều có độ dài ≥ n. Gọi α là từ có độ dài nhỏ nhất trong L, mà $|\alpha| \ge n$. Theo định lý 4.1, ta có α = uvw, trong đó $|v| \ge 1$ và với mọi i = 0, 1, 2,… ta có $uv^iw \in L$. Với i = 0, $uw \in L$ mà $|uw| < |\alpha|$. Điều này mâu thuẫn với tính nhỏ nhất của $|\alpha|$. Vậy tồn tại $\omega \in L$ sao cho $|\omega| < n$.

Hệ quả 3.2

Tồn tại một ngôn ngữ phi ngữ cảnh mà không được đoán nhận bởi bất kỳ một otomat hữu hạn đơn định nào.

Chứng minh: Xét ngôn ngữ $L = \{a^nb^n \mid n \ge 1\}$ trên bảng chữ Σ = {a, b}. Ta có L = L(G), trong đó $G = \langle \Sigma, \{S\}, S, \{S \to aSb, S \to ab\}\rangle$ là văn phạm phi ngữ cảnh. Giả sử L = T(A) với $A = \langle Q, \Sigma, \delta, q_0, F\rangle$ là một otomat hữu hạn đơn định. Với n đủ lớn, $\alpha = a^nb^n$ có $|\alpha| \ge |Q|$. Theo định lý 1.1, ta có thể biểu diễn $a^nb^n = uvw$, trong đó $|v| \ge 1$, $uv^iw \in L$, $\forall i = 0, 1, 2, \ldots$ Ta hãy tập trung phân tích từ v và v^i:

– Nếu v chỉ chứa a ($|v|_a > 0$ và $|v|_b = 0$) thì với i đủ lớn $|uv^iw|_a > |uv^iw|_b$.

– Nếu v chỉ chứa b ($|v|_b > 0$ và $|v|_a = 0$) thì với i đủ lớn $|uv^iw|_b > |uv^iw|_a$

– Nếu $|v|_a > 0$ và $|v|_b > 0$ thì với i = 2 ta có $v = (ab)^2 = abab$, tức là a và b xen kẽ nhau trong uv^iw, khi đó uv^iw không thể có dạng a^nb^n.

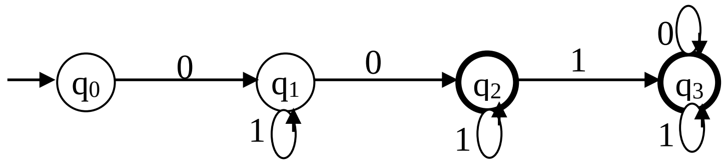
Cả ba trường hợp đều mâu thuẫn với $uv^iw \in L$. Vậy không tồn tại một otomat hữu hạn đơn định nào đoán nhận L, tức là L không phải là một ngôn ngữ chính quy.

Hệ quả này là một ứng dụng điều kiện cần của ngôn ngữ chính quy, để chứng minh một ngôn ngữ không phải là chính quy ta cần chỉ ra rằng nó không thỏa mãn điều kiện cần này.

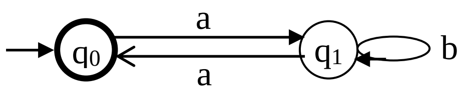
BÀI TẬP CHƯƠNG 3

1. Hãy xác định các thành phần của các otomat hữu hạn có đồ thị chuyển sau và xác định các ngôn ngữ được đoán nhận bởi chúng.

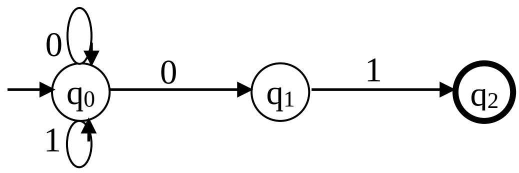
a).



b).



2. Cho otomat hữu hạn A có đồ thị chuyển như hình vẽ:



a). Biểu diễn hình thức otomat A dưới dạng một bộ 5 như định nghĩa.

b).Otomat A là otomat loại gì: đơn định/ không đơn định/ đầy đủ/ không đầy đủ.

c). Biểu diễn otomat A dưới dạng bảng chuyển trạng thái.

d). Xác định ngôn ngữ được chấp nhận bởi otomat A.

e). Xây dựng một otomat đơn định A' tương đương với A sao cho T(A') = T(A); vẽ dồ thị chuyển của otomat A'.

3. Hãy xây dựng các biểu thức chính quy và các otomat hữu hạn đơn định biểu diễn các ngôn ngữ sau:

a). $L = \{a^n b^m \mid n \geq 1, m \geq 1\}$.

b). $L = \{(01)^n, (010)^n \mid n \geq 0\}$.

c). $L = \{(aab)^n(baa)^m \mid n \geq 1, m \geq 1\}$.

4. Hãy xây dựng các biểu thức chính quy và các otomat hữu hạn đơn định biểu diễn các ngôn ngữ sau:

a). $L_1 = \{\omega \in \{0, 1\}^* \mid \omega$ bắt đầu bởi đúng 2 chữ số 0$\}$.

b). $L_2 = \{\omega \in \{0, 1\}^* \mid \omega$ bắt đầu bởi 2 chữ số 0$\}$

c). $L_3 = \{\omega \in \{0, 1\}^* \mid \omega$ kết thúc bởi đúng 2 chữ số 0$\}$.

d). $L_4 = \{\omega \in \{0, 1\}^* \mid \omega$ kết thúc bởi 2 chữ số 0$\}$.

e). $L_5 = \{\omega \in \{0, 1\}^* \mid \omega$ không bắt đầu bởi hai chữ số 1 liên tiếp$\}$.

5. Hãy xây dựng các otomat hữu hạn đoán nhận các ngôn ngữ phần bù của các ngôn ngữ trong bài tập 3.

Chú ý: Nếu A là otomat hữu hạn đơn định và đầy đủ thì otomat đoán nhận ngôn ngữ phần bù của T(A) (gọi là otomat bù của A, ký hiệu $\overline{A}$) là otomat nhận được từ A bằng cách đổi các trạng thái kết thúc thành trạng thái không kết thúc, và ngược lại..

6. Hãy xây dựng các otomat hữu hạn đoán nhận các ngôn ngữ ngược của các ngôn ngữ trong bài tập 3.

7. Hãy xây dựng các biểu thức chính quy và các otomat hữu hạn xác định các ngôn ngữ sau:

a). L_1 gồm các xâu có chẵn ký hiệu a.

b). L_2 gồm các xâu có lẻ ký hiệu a.

8. Hãy xây dựng các biểu thức chính quy và các otomat hữu hạn xác định các ngôn ngữ sau:

a). $L_1 = \{\omega \in \{a, b\}^* \mid$ với $|\omega|$ là một số chẵn$\}$.

b). $L_2 = \{\omega \in \{a, b\}^* \mid$ với $|\omega|$ là một số lẻ$\}$

9. Hãy xây dựng các otomat hữu hạn không đơn định đoán nhận các ngôn ngữ sau:

a). $L_1 = \{\omega_1 aba \omega_2 \mid \omega_1, \omega_2 \in \{a, b\}^*\}$.

b). $L_2 = \{\omega \in \{0, 1\}^* \mid \omega$ bắt đầu bằng luỹ thừa dương của 101$\}$.

c). $L_3 = \{(1111)^n \omega \mid \omega \in \{0, 1\}^*, n \geq 0\}$.

10. Hãy thành lập các văn phạm chính quy sinh ra các ngôn ngữ mà được đoán nhận bởi các otomat hữu hạn không đơn định sau:

a). $A = <\{q_0, q_1, q_2, q_3, q_4\}, \{a, b\}, \delta, q_0, \{q_2, q_4\}>$, trong đó:

$\delta(q_0, a) = \{q_0, q_1\}$, $\delta(q_0, b) = \{q_0, q_1\}$, $\delta(q_1, a) = \varnothing$, $\delta(q_1, b) = \{q_2\}$, $\delta(q_2, a) = \{q_2\}$, $\delta(q_2, b) = \{q_2\}$, $\delta(q_3, a) = \{q_4\}$, $\delta(q_3, b) = \varnothing$, $\delta(q_4, a) = \{q_4\}$, $\delta(q_4, b) = \{q_4\}$.

b). $A = <\{q_0, q_1\}, \{0, 1\}, \delta, q_0, \{q_1\}>$, trong đó:

$\delta(q_0, 0) = \{q_0, q_1\}$, $\delta(q_0, 1) = \{q_1\}$, $\delta(q_1, 0) = \varnothing$, $\delta(q_1, 1) = \{q_0, q_1\}$.

11. Hãy xây dựng các otomat hữu hạn đơn định đoán nhận các ngôn ngữ mà được sinh bởi các văn phạm chính quy sau:

a). G = <{a, b, c}, {S, A, B, C, D}, S, {$S\rightarrow aB$, $S\rightarrow aC$, $B\rightarrow aA$, $C\rightarrow bD$, $D\rightarrow bC$, $D\rightarrow bA$, $A\rightarrow cA$, $A\rightarrow c$}>.

b). G = <{0, 1}, {S, S_1 }, S, {$S\rightarrow 0S$,$S \rightarrow 0S_1$, $S_1\rightarrow 1S_1$, $S_1\rightarrow 1$}>

(hướng dẫn: có thể xây dựng otomat không đơn định từ các văn phạm trên theo định lý 3.5, sau đó dùng thuật toán đơn định hóa để nhận được các otomat đơn định, hoặc trước hết xác định biểu thức chính quy biểu diễn các ngôn ngữ sinh bởi các văn phạm đã cho, rồi dùng thuật toán Thompson để xây dựng các otomat từ các biểu thức chính quy).

12. Hãy xây dựng các ngôn ngữ được biểu diễn bởi các biểu thức chính quy sau:

a). $bba(a+b)^*$.

b). $(a+b)^*bab$.

c). $(bb+a)^*$

d). $(aa+b)^*$.

e). $(bb+a)^*(aa+b)^*$.

13. Hãy xây dựng các otomat hữu hạn đơn định đoán nhận các ngôn ngữ được biểu diễn bởi các biểu thức chính quy sau:

a). $bba(a+b)^*$.

b). $(a+b)^*bab$.

c). $(bb+a)^*$

d). $(aa+b)^*$.

e). $(bb+a)^*(aa+b)^*$.

14. Cho ôtomat A = (S, Σ , s_1, δ, F), với tập trạng thái S = {s_1, s_2, s_3, s_4}, bảng chữ cái Σ = {a, b}, F = {s_4}, và hàm chuyển trạng thái cho bởi bảng sau:

δ	s_1	s_2	s_3	s_4
a	{s_2, s_4}	s_3		s_3
b	{s_3, s_4}		s_4	

a). Vẽ đồ thị chuyển của ôtomat A.

b). Tìm otomat M đơn định và đầy đủ tương đương với otomat A

c). Vẽ đồ thị chuyển của ôtomat M.

15. Cho otomat A = (S, Σ, s_1, δ, F), với tập trạng thái S = {s_1, s_2, s_3}, Σ = {a, b, c}, F = {s_3}, và hàm chuyển trạng thái cho bởi bảng sau:

δ	s_1	s_2	s_3
a	s_2	s_3	{s_1, s_2}
b	s_3	s_1	
c	s_3	s_1	

a). Vẽ đồ thị chuyển của otomat A.

b). Tìm otomat M = (Q, Σ , q_1, f, P), đơn định và đầy đủ, tương đương otomat A

c). Vẽ đồ thị chuyển của otomat M.

16. Hãy chỉ ra rằng các ngôn ngữ dưới đây không phải là chính quy:

a). L = {$a^n b^m$ | n, m ≥ 0 và n < m }

b). L = {$a^n b^m$ | n, m ≥ 0 và n > m }

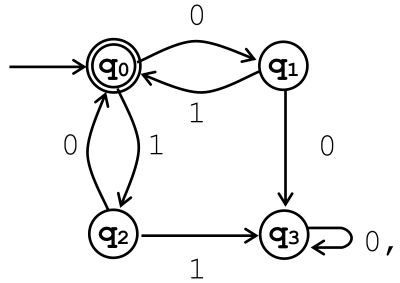
17. Áp dụng định lý về điều kiện cần của ngôn ngữ chính quy, chứng minh rằng:

a). Ngôn ngữ L = {a, aa, aaa, a^5, a^7, a^{11}, ,,, , a^p, …| với p là số nguyên tố} không phải là ngôn ngữ chính quy.

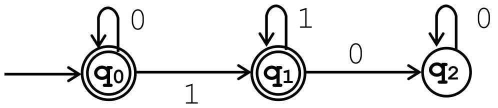
b). Ngôn ngữ L = {a, a^4, a^9, a^{16}, ,,, , a^k, …| với k là số chính phương} không phải là ngôn ngữ chính quy.

18. Hãy xác định biểu thức chính quy biểu diễn ngôn ngữ được đoán nhận bởi otomat hữu hạn có đồ thị chuyển như sau:

a).



b).



TÀI LIỆU THAM KHẢO

[1]. Đặng Huy Ruận [2002]. *"Lý thuyết ngôn ngữ hình thức và Otomat"* NXB Đại hoc Quốc gia Hà nội 2002.

[2]. Phan Đình Diệu, [1971]. *"Lý thuyết Otomat và Thuật toán",* NXB Đại hoc và trung hoc chuyên nghiệp, Hà nội 1971.

[3]. Đỗ Đức Giáo, [2000], "Toán rời rạc", NXB Đại học Quốc gia Hà nội, Hà nội, 2000.

[4]. Bài giảng "Ngôn ngữ Hình thức và Otomat", Đại học Khoa học, Đại học Huế.

[5]. Nguyễn Xuân Huy [1988], "*Thuật Toán*", NXB Thống Kê, Hà Nội, 1988.

[6]. Nguyễn Văn Định, [2000]. "*Thiết kế và phân tích thuật toán",* Bài giảng chuyên đề tại khoa Toán-Cơ-Tin học, ĐH Khoa học Tự nhiên, ĐHQG Hà nội.

[7]. J.E. Hopcroft, and J.D.Ullman [1979]. *"Introduction to Automata Theory, Languages and Computation"* Addison-Wesley, Reading, MA.1979.

[8]. Nguyen Van Dinh, [1998]. *"Solving Determinazation Problems of Automata on the computer",* VNU. Journal of Science, Nat. Sci., t.XIV, $n^0$1-1998, p: 27-32.

[9]. Nguyen Van Dinh [2001].*"The Automata complexity of the Language transformation schema that contains operations with restricted degree".* Journal of computer science and cybernetics, T.17, S.2 (2001), pages 39-44