**Bài Tập Thực Hành Tuần 1**

**Bài 1 : Phân biệt các văn phạm theo tiêu chuẩn của Chomsky?**

 Phân loại văn phạm theo Chomsky Dựa vào đặc điểm của tập quy tắc mà người ta chia các văn phạm thành các nhóm khác nhau. Noam Chomsky (Institute Professor, Massachusetts Institute of Technology. Born December 7, 1928 Philadelphia, Pennsylvania, USA) đã phân loại văn phạm thành bốn nhóm: - Nhóm 0: Văn phạm không hạn chế (hay văn phạm ngữ cấu, văn phạm tổng quát);

- Nhóm 1: Văn phạm cảm ngữ cảnh;

- Nhóm 2: Văn phạm phi ngữ cảnh;

- Nhóm 3: Văn phạm chính quy.

Văn phạm G = < Σ, , S, P > mà không có một ràng buộc nào đối với các quy tắc của nó được gọi là văn phạm tổng quát hay văn phạm không hạn chế. Như vậy, các quy tắc trong văn phạm nhóm 0 có dạng: α→β, với α = α’Aα’’, A ∈ , α’, α’’, β ∈ (Σ ∪ ) \* . Các quy tắc của văn phạm nhóm 0 được gọi là quy tắc không hạn chế. Ngôn ngữ do văn phạm nhóm 0 sinh ra được gọi là ngôn ngữ tổng quát.

Văn phạm G = < Σ, Δ, S, P > mà các quy tắc của nó đều có dạng: α→β, với α = α’Aα’’, A ∈ Δ, α’, α’’, β ∈ (Σ ∪ Δ)\* và | α | ≤ | β |, được gọi là văn phạm nhóm 1 hay văn phạm cảm ngữ cảnh. Các quy tắc trong văn phạm nhóm 1 được gọi là quy tắc cảm ngữ cảnh. Ngôn ngữ do văn phạm cảm ngữ cảnh sinh ra được gọi là ngôn ngữ cảm ngữ cảnh. Các văn phạm mà các quy tắc của chúng có dạng trên, đồng thời chứa thêm quy tắc rỗng S→ε, cũng được xếp vào lớp văn phạm nhóm 1

Văn phạm G = < Σ, Δ, S, P > mà các quy tắc của nó có dạng A→ω, trong đó A∈, ω∈(Σ ∪ Δ ) \* , được gọi là văn phạm nhóm 2 hay văn phạm phi ngữ cảnh. Như vậy, các quy tắc trong văn phạm phi ngữ cảnh có vế trái chỉ chứa một ký hiệu phụ còn vế phải là tùy ý, và được gọi là quy tắc phi ngữ cảnh. Ngôn ngữ do văn phạm phi ngữ cảnh sinh ra được gọi là ngôn ngữ phi ngữ cảnh.

Văn phạm G = < Σ, Δ, S, P > mà các quy tắc của nó chỉ có dạng A→aB, A→a (hoặc chỉ có dạng A→Ba, A→a ), trong đó A, B∈, a∈Σ, được gọi là văn phạm nhóm 3 hay văn phạm chính quy. Các văn phạm mà các quy tắc của chúng có dạng trên, đồng thời chứa thêm quy tắc rỗng S→ε cũng được gọi là văn phạm chính quy (hay còn gọi là văn phạm chính quy suy rộng). Các quy tắc trong văn phạm chính quy được gọi là quy tắc chính quy. Ngôn ngữ do văn phạm chính quy sinh ra được gọi là ngôn ngữ chính quy.

Nhận xét: Từ các định nghĩa trên, ta thấy lớp văn phạm không hạn chế là rộng nhất, nó chứa đựng các văn phạm cảm ngữ cảnh, lớp văn phạm cảm ngữ cảnh chứa các văn phạm phi ngữ cảnh và lớp văn phạm phi ngữ cảnh chứa các văn phạm chính quy. Ngôn ngữ hình thức được gọi là ngôn ngữ tổng quát (hay cảm ngữ cảnh, phi ngữ cảnh, chính quy) nếu tồn tại văn phạm loại tương ứng sinh ra nó. Vì vậy, đối với các lớp ngôn ngữ, nếu ký hiệu L0, L1, L2, L3 lần lượt là các lớp ngôn ngữ tổng quát, cảm ngữ cảnh, phi ngữ cảnh và chính quy thì ta có bao hàm thức: L3 ⊂ L2 ⊂ L1 ⊂ L0

 Hình vẽ dưới đây cho một sự so sánh về độ lớn của các lớp ngôn ngữ theo phân loại của Chomsky, cho thấy lớp ngôn ngữ chính quy L3 là nhỏ nhất, nó bị chứa thực sụ trong lớp ngôn ngữ phi ngữ cảnh L2, lớp ngôn ngữ phi ngữ cảnh lại bị chứa thực sự trong lớp ngôn ngữ cảm ngữ cảnh L1 và cuối cùng lớp ngôn ngữ tổng quát L0 (ngôn ngữ ngữ cấu) là rộng nhất.

Diagram

Description automatically generated

      Hình 1.2. So sánh các lớp ngôn ngữ Ta cũng thấy về mặt cấu trúc ngữ pháp thì các quy tắc của các văn phạm phi ngữ cảnh và văn phạm chính quy là đơn giản hơn cả và chúng có nhiều ứng dụng trong việc thiết kế các ngôn ngữ lập trình và trong nghiên cứu về chương trình dịch… Vì vậy, trong các phần tiếp theo chúng ta dành thêm sự quan tâm tới hai lớp văn phạm đó

**Bài 2 : Sự tương đương của văn phạm và Ôtômát?**

**Sự tương đương của ngữ pháp thông thường và Automata hữu hạn**

**Ngữ pháp thông thường / ngữ pháp loại -3**

Văn phạm chính quy hoặc loại 3 xác định ngôn ngữ được gọi là ngôn ngữ chính quy được chấp nhận bởi Automata hữu hạn. Văn phạm chính quy G bao gồm 4 bộ (V, T, P, S).

**Ngữ pháp tuyến tính:** Ngữ pháp được gọi là tuyến tính trong đó nhiều nhất một điểm không đầu cuối có thể xảy ra ở phía bên phải của bất kỳ quy tắc sản xuất nào. Sau đây là các loại Ngữ pháp tuyến tính:

* **Đúng tuyến tính ngữ pháp**
* **Ngữ pháp tuyến tính trái**

**Văn phạm tuyến tính đúng:** Văn phạm tuyến tính đúng là văn phạm G = (V, T, P, S) sao cho tất cả các luật sản sinh P đều có dạng sau:

một -> một

Một -> một B

Trong đó A và B là các biến trong V tức là A và B thuộc biến V và a là terminal. Phía bên trái của quy tắc sản xuất trong ngữ pháp tuyến tính bên phải chỉ bao gồm một ký hiệu từ tập hợp các biến và phía bên phải chứa các chuỗi đầu cuối hoặc chỉ một biến có ở vị trí ngoài cùng bên phải.

**Văn phạm tuyến tính trái:** Văn phạm tuyến tính trái là văn phạm G = (V, T, P, S) sao cho tất cả các luật sản sinh P đều có dạng sau:

một -> một

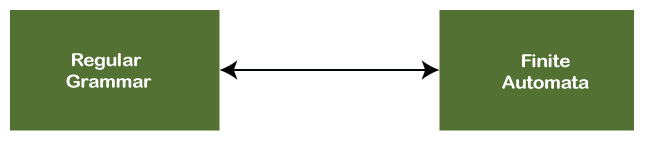
Một -> Ba

Trong này A và B là biến trong V tức là A và B thuộc biến còn a là đầu cuối.

**Máy tự động hữu hạn**

Máy tự động hữu hạn là mô hình đơn giản nhất được chấp nhận ngôn ngữ gọi là ngôn ngữ thông thường. Thuật ngữ hữu hạn trong automat hữu hạn là nó có số lượng trạng thái giới hạn và số lượng bảng chữ cái giới hạn trong chuỗi. Automata hữu hạn bao gồm 5 bộ (Q, ?, q 0 , F, ?).

Mối quan hệ của ngữ pháp thông thường và automata hữu hạn được hiển thị bên dưới:



**Định lý:**

Nếu G là văn phạm chính quy thì L(G) là ngôn ngữ chính quy.

**Bằng chứng:**

Các ngôn ngữ thông thường được nhận dạng bởi automata hữu hạn. Vì vậy, trước hết chúng ta sẽ xây dựng một NFA tương đương với văn phạm tuyến tính đúng đã cho chấp nhận ngôn ngữ được xác định bởi văn phạm thông thường G đã cho.

Cho G = (V, T, P, S) là một văn phạm tuyến tính đúng. Đặt V = {A0A1…..An}, trong đó A0 là ký hiệu bắt đầu S. Chúng tôi xác định một NFA N **= ({q0q1….qnqf}, ?, ?, q0, qf})** trong đó ? được định nghĩa là:

* Với mỗi lần sản xuất Ai -> bAj. Trong **N** này có quá trình chuyển đổi từ q i sang q j với nhãn b.
* Với mỗi lần sản xuất Ak -> b. Trong N này có quá trình chuyển đổi từ q k sang q f với nhãn b.

Từ cách dựng, rõ ràng A0 => b1A1 => b1b2A2 => b1b2b3A3 => …..=>b1bn-1 => b1….bn, nếu và chỉ khi có một đường đi trong N bắt đầu từ trạng thái ban đầu q0 và kết thúc ở trạng thái cuối cùng qf với giá trị đường dẫn b1b2….bn.

Do đó L(G) = T(N).

**Bài 3 :  So sánh Ôtômát đơn định (DFA) và Ô tô mát không đơn định (NFA)?**

1. "DFA" là viết tắt của "Automatic Finite Xác định" trong khi "NFA" là viết tắt của "Automat hữu hạn không xác định. "

2. Cả hai đều là các chức năng chuyển đổi của automata. Trong DFA, trạng thái có thể tiếp theo được đặt rõ ràng trong khi ở mỗi cặp NFA của biểu tượng nhà nước và ký tự đầu vào có thể có nhiều trạng thái kế tiếp có thể xảy ra.

3. NFA có thể sử dụng quá trình chuyển đổi chuỗi rỗng trong khi DFA không thể sử dụng chuỗi chuyển đổi rỗng.

4. NFA được dễ dàng hơn để xây dựng trong khi đó là khó khăn hơn để xây dựng DFA.

5. Theo dõi được cho phép trong DFA, trong khi ở NFA, nó có thể hoặc không được cho phép.

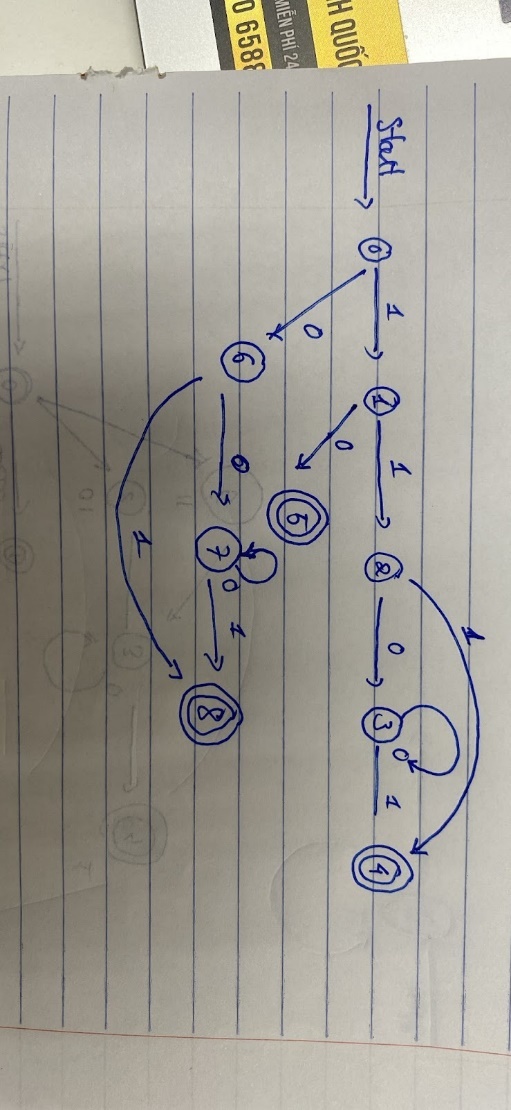
6. DFA đòi hỏi nhiều không gian hơn trong khi NFA cần ít không gian hơn.

7. Trong khi DFA có thể được hiểu là một máy và một máy DFA có thể được xây dựng cho mọi đầu vào và đầu ra,

 8. NFA có thể được hiểu là một số ít máy tính tính cùng nhau, và không có khả năng xây dựng một máy NFA cho mỗi đầu vào và đầu ra

**Bài 4:. Cho các Biểu thức chính quy (BTCQ) sau. Hãy xây dựng Các Ô tô mát đơn định đoán nhận chúng :**

a, 10 ∪ (0 ∪ 11) 0 ∗1

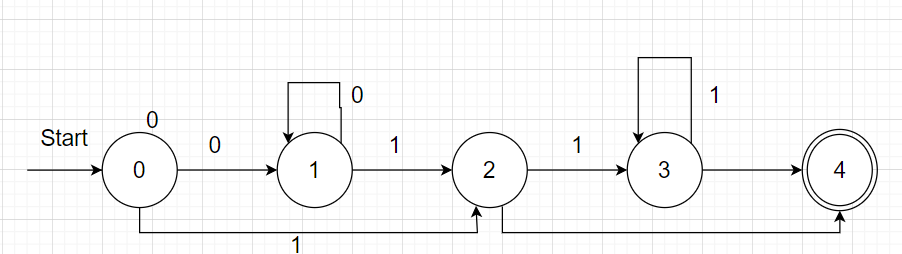


b, 1 ∗00∗

Diagram

Description automatically generated

 00∗11∗



**Bài 5. Mô tả các ngôn ngữ được biểu thị bằng các biểu thức chính quy sau:**

a. a(a|b)\*a

=> chuỗi trên được biểu thị như sau :

ký tự đầu tiên : a

ký tự trung gian : a or b or không có ký tự nào

ký tự kết thúc  :  a

b. ((ε|a)b\*)\*

=> chuỗi trên được biểu thị như sau :

ký tự đầu tiên : BTCQ or a

ký tự trung gian : không có ký tự hoặc ký tự b được lặp lại nhiều lần

ký tự kết thúc  :

c. (a|b)\*a(a|b)(a|b)

=> chuỗi trên được biểu thị như sau :

ký tự đầu tiên : a hoặc b có thể lập lại nhiều lần hoặc rỗng

ký tự trung gian 1: a

ký tự trung gian 2: a hoặc b

ký tự kết thúc  :  a hoặc  b

d. a\*ba\*ba\*ba\*

=> chuỗi trên được biểu thị như sau :

ký tự đầu tiên : không có ký tự nào hoặc a được lặp lại nhiều lần

ký tự trung gian 1 : b

ký tự trung gian 2 : không có ký tự nào hoặc a được lặp lại nhiều lần

ký tự trung gian 3 : b

ký tự trung gian 4 : không có ký tự nào hoặc a được lặp lại nhiều lần

ký tự trung gian 5 : b

ký tự kết thúc : không có ký tự nào hoặc a được lặp lại nhiều lần

e. !! (aa|bb)\*((ab|ba)(aa|bb)\*(ab|ba)(aa|bb)\*)\*

=> chuỗi trên được biểu thị như sau :

ký tự đầu tiên : aa hoặc bb hoặc không có ký tự nào

ký tự trung gian 1:   ab hoặc ba

ký tự trung gian 2:    aa hoặc bb  hoặc không có ký tự nào

ký tự trung gian 3: ab hoặc ba

ký tự kết thúc  :  aa hoặc bb hoặc không có ký tự nào sau đó lặp lại hoặc không có ký tự nào

**Bài 6. Viết định nghĩa thông thường cho các ngôn ngữ sau:**

a. Tất cả các chuỗi chữ thường chứa năm nguyên âm theo thứ tự.

ký tự nguyên âm là **“a”, ă, â, “e”, ê, “i”, “o”, ô, ơ, “u”, ư và “y”**.

chuỗi : aa\* ee\* ii\* oo\* uu\*

b. Tất cả các chuỗi chữ thường trong đó các chữ cái được sắp xếp theo thứ tự từ điển tăng dần.

chuỗi : a\*b\*c\*d\*e\*....z\*

c. Chú thích, bao gồm một chuỗi được bao quanh bởi / và /, không có \*/ ở giữa, trừ khi nó nằm trong dấu ngoặc kép (")!!

chuỗi : //\* (\*/ |”\*/”)/

d. Tất cả các chuỗi chữ số không có chữ số lặp lại. Gợi ý: Hãy thử bài toán này trước với một vài chữ số, chẳng hạn như {O, 1, 2}.!! Tất cả các chuỗi chữ số có nhiều nhất một chữ số lặp lại.

e. !! Tất cả các chuỗi a và b với số a chẵn và số b lẻ.Tập hợp các nước đi của Cờ vua, theo ký hiệu không chính thức, chẳng hạn như p-k4 hoặc kbp\*qn.

f. !! Tất cả các xâu a và b không chứa xâu con abb.Tất cả các chuỗi a và b không chứa dãy con abb.

chuỗi : a\*bbbb\*