ĐẠI HỌC ĐÀ NẴNG
TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN VÀ TRUYỀN THÔNG VIỆT - HÀN
Vietnam - Korea University of Information and Communication Technology

# CHƯƠNG 1

# MỞ ĐẦU

http://vku.udn.vn/

# Nội dung

1. Khái niệm về ngôn ngữ
2. Khái niệm về văn phạm
3. Khái niệm về ô-tô-mát

# 1. Khái niệm về ngôn ngữ

## 1.1. Bảng chữ (alphabet)

- Bảng chữ *hay bảng chữ cái* là tập hợp không rỗng gồm hữu hạn các phần tử. Mỗi phần tử của bảng chữ được gọi là ký hiệu (symbol) *hay ký tự, chữ cái*.
  - Ví dụ
    - $V_1$ ={a, b, c, …, z} (bảng chữ cái Latinh)
    - $V_2$={0, 1, 2, 3, 4, 5, 6, 7, 8, 9} (bảng chữ số thập phân)
    - $V_3$={0, 1} (bảng chữ số nhị phân)

# 1. Khái niệm về ngôn ngữ

## 1.2. Xâu

- Cho bảng chữ V.
- Xâu *hay từ, chuỗi, câu* trên V là một dãy liên tiếp gồm hữu hạn các ký tự được chọn từ V.
  - Ví dụ
    - Cho $V_1$ ={a, b, c, …, z}. Lúc đó a, ab, otomat là các xâu trên $V_1$.
    - Cho $V_2$={0, 1, 2, 3, 4, 5, 6, 7, 8, 9}. Lúc đó 123, 5107901, 987654321, 2025 là các xâu trên $V_2$.
    - 0, 1, 01101, 10100101 là các xâu trên $V_3$={0, 1}.
- Xâu rỗng là xâu không chứa ký tự nào. Ký hiệu xâu rỗng là ε.

# 1. Khái niệm về ngôn ngữ

## 1.2. Xâu

- Độ dài (length) của xâu là tổng số vị trí của các ký tự xuất hiện trong xâu.
- Ký hiệu độ dài của xâu x là |x|.
  - Ví dụ
    - |a|=1, |ab|=2, |otomat|=6.
    - |123|=3, |5107901|=7, |987654321|=9, |2025|=4.
    - |01101|=5, |10100101|=8.
- Rõ ràng xâu rỗng ε có độ dài bằng 0, tức là |ε|=0.

# 1. Khái niệm về ngôn ngữ

## 1.2. Xâu

- Cho bảng chữ V.
- Ký hiệu các tập hợp:
  - $V^*$ là tập hợp gồm tất cả các xâu trên V.
  - $V^+$ là tập hợp gồm tất cả các xâu trên V nhưng không chứa xâu rỗng ε.
- Rõ ràng $\{\varepsilon\} \subset V^*$ và $V^+ = V^* \setminus \{\varepsilon\}$.
- Dễ dàng thấy rằng $V^*$ và $V^+$ là các tập hợp vô hạn đếm được.

# 1. Khái niệm về ngôn ngữ

## 1.2. Xâu

- Ví dụ
  - Cho bảng chữ V={a, b, c}.
  - Ta có:

  $V^*$={ε, a, b, c, aa, bb, cc, ab, ac, ba, bc, ca, cb, abc, acb, bac,bca,cab,cba,…}.

  $V^+$={a, b, c, aa, bb, cc, ab, ac, ba, bc, ca, cb, abc, acb, bac, bca, ab, cba,…}.

# 1. Khái niệm về ngôn ngữ

## 1.3. Các phép toán xử lý xâu

### 1.3.1. So sánh bằng nhau, khác nhau

- Cho bảng chữ V và x, y là các xâu trên V.
- Cho biểu diễn của x và y lần lượt là $x_1x_2\ldots x_n$ và $y_1y_2\ldots y_m$ với $n,m \in Z^+$ và $x_1,x_2,\ldots,x_n,y_1,y_2,\ldots,y_m \in V$.
- Xâu x được gọi là bằng xâu y, ký hiệu x=y, khi và chỉ khi n=m và $x_i$ trùng $y_i$ với mọi $i \in \{1, 2, \ldots, n\}$.
  - Ví dụ
    - Cho bảng chữ V={0,1} và x, y là các xâu trên V, trong đó x là 010111, y là 010111.
    - Ta có x=y.

# 1. Khái niệm về ngôn ngữ

## 1.3. Các phép toán xử lý xâu

### 1.3.1. So sánh bằng nhau, khác nhau

- Cho bảng chữ V và x,y là các xâu trên V. Rõ ràng nếu x=y thì y=x, do đó ta có thể nói *x bằng y* hay *y bằng x* hay *x và y là 2 xâu bằng nhau*.
- Trong trường hợp không phải x và y là 2 xâu bằng nhau, ta gọi *x khác y* hay *y khác x* hay *x và y là 2 xâu khác nhau*, ký hiệu x≠y hay y≠x.
  - Ví dụ
    - Cho bảng chữ V={0, 1} và x,y là các xâu trên V, trong đó x là 01111, y là 01011.
    - Ta có x≠y hay y≠x.

# 1. Khái niệm về ngôn ngữ

## 1.3. Các phép toán xử lý xâu

### 1.3.1. So sánh bằng nhau, khác nhau

- Rõ ràng mọi xâu rỗng đều bằng nhau và $x \neq \varepsilon$ với mọi xâu x sao cho $|x| \neq 0$.
- Cho 2 bảng chữ V, V' và $V \subseteq V'$. Rõ ràng nếu x là xâu trên V thì x cũng là xâu trên V'.
  - Ví dụ
    - Cho V={0, 1}, V'={0, 1, 2} và x=1101001 là xâu trên V.
    - Do $V \subseteq V'$ và x là xâu trên V nên x cũng là xâu trên V'.

# 1. Khái niệm về ngôn ngữ

## 1.3. Các phép toán xử lý xâu

## 1.3.2. Ghép nối (concatenation)

- Cho bảng chữ V và x, y là các xâu trên V.
- Ghép nối *hay kết nối* x với y, ký hiệu là xy *hay x.y*, tạo thành xâu mới từ việc ghép y vào bên phải của x, tức là viết x trước rồi đến y và không có dấu cách giữa x và y.
  - Ví dụ
    - Cho bảng chữ V={0, 1} và x,y là các xâu trên V, trong đó x=01, y=0110.
    - Ta có xy=010110, yx=011001. Rõ ràng xy≠yx.

# 1. Khái niệm về ngôn ngữ

## 1.3. Các phép toán trên xâu

### 1.3.2. Ghép nối

- Cho bảng chữ V và x, y, z là các xâu trên V.
- Lưu ý: Sử dụng cặp dấu () để quy định thứ tự thực hiện các phép toán.
- Một số tính chất:
  - $\varepsilon x = x\varepsilon = x$
  - $(xy)z = x(yz)$
  - $|xy| = |yx| = |x| + |y|$

# 1. Khái niệm về ngôn ngữ

## 1.3. Các phép toán trên xâu

## 1.3.2. Ghép nối

- Cho bảng chữ V và x, y, z là các xâu trên V.
- Nếu x=yz thì y được gọi là tiền tố (prefix) của x và z được gọi là hậu tố (postfix) của x.
  - Ví dụ
    - Cho bảng chữ V={a, b, c} và xâu x= abc trên V.
    - Rõ ràng x=abc=εabc=abcε.
    - Tiền tố của x là các xâu: ε, a, ab, abc.
    - Hậu tố của x là các xâu: ε, c, bc, abc.

# 1. Khái niệm về ngôn ngữ

## 1.3. Các phép toán trên xâu

### 1.3.3. Nghịch đảo (reverse)

- Cho bảng chữ V và x là xâu trên V.
- Nghịch đảo của xâu x, ký hiệu $x^r$, tạo thành xâu mới từ việc viết các ký tự của x theo thứ tự ngược lại.
  - Ví dụ
    - Cho bảng chữ V={0, 1} và xâu x=0101 trên V.
    - Ta có $x^r$ =1010.
  - Ví dụ
    - Cho xâu y=abcabd trên bảng chữ V={a, b, c, d}.
    - Ta có $y^r$ =dbacba.

# 1. Khái niệm về ngôn ngữ

## 1.3. Các phép toán trên xâu

## 1.3.3. Nghịch đảo

- Cho bảng chữ V và x, y là các xâu trên V.
- Một số tính chất:
  - Nếu $|x|=0$ hay $|x|=1$ thì $x^r = x$.
  - $(x^r)^r = x$.
  - $(xy)^r = y^r x^r$.
  - $|x^r| = |x|$.

# 1. Khái niệm về ngôn ngữ

## 1.3. Các phép toán trên xâu

### 1.3.3. Nghịch đảo

- Cho bảng chữ V và x là xâu trên V.
- Nếu $x^r = x$ thì x được gọi là xâu đối xứng (xâu hình tháp)
  - Ví dụ
    - Cho bảng chữ V={0, 1} và xâu x=01110 trên V.
    - x là xâu đối xứng vì $x^r$=01110=x.
- Như vậy xâu rỗng ε và mọi xâu x sao cho |x|=1 là các xâu đối xứng.

# 1. Khái niệm về ngôn ngữ

## 1.3. Các phép toán trên xâu

### 1.3.4. Lũy thừa (power)

- Cho bảng chữ V và x là xâu trên V.
- Lũy thừa n ($n \in Z^+$) của xâu x, ký hiệu $x^n$, tạo thành xâu mới bằng cách ghép nối xâu x liên tiếp n lần.
- Như vậy $x^n$=xxx...x (n lần x) và $x^n = (x^{n-1})x = x(x^{n-1})$.
- Chú ý rằng, nếu n=0 ta có $x^n = x^0$ và $x^0$ được định nghĩa: $x^0 = \varepsilon$ với mọi xâu x.
  - Ví dụ
    - Cho bảng chữ V={0, 1} và xâu x=011 trên V.
    - Ta có $x^0 = \varepsilon$, $x^1$ =011, $x^2$ =011011, $x^3$ =011011011.

# 1. Khái niệm về ngôn ngữ

## 1.3. Các phép toán trên xâu

### 1.3.5. Lũy thừa bảng chữ

- Cho bảng chữ V.
- Ta có thể biểu diễn tất cả các xâu trên V bằng phép toán lũy thừa bảng chữ V.
- Gọi $V^k$ ($k \in Z^+ \cup \{0\}$) là tập hợp các xâu có độ dài k trên V.
- Lúc đó $V^k$ được gọi là lũy thừa k của bảng chữ V.

# 1. Khái niệm về ngôn ngữ

## 1.3. Các phép toán trên xâu

## 1.3.5. Lũy thừa bảng chữ

- Chú ý rằng, $V^0 = \{\varepsilon\}$ và $V^1 = V$ với mọi bảng chữ V.
  - Ví dụ
    - Cho bảng chữ V={a, b}.
    - Ta có $V^0=\{\varepsilon\}$, $V^1=\{a,b\}$, $V^2=\{aa,ab,bb,ba\}$, $V^3=\{aaa, aab, aba, abb, baa, bab, bba, bbb\}$,….
- Cho bảng chữ V. Ta có:
  - $V^* = V^0 \cup V^1 \cup V^2 \cup V^3 \cup V^4 \cup \ldots$
  - $V^+ = V^1 \cup V^2 \cup V^3 \cup V^4 \cup \ldots$
  - $V^+ = V^* \setminus \{\varepsilon\}$ hay $V^+ = V^* \setminus V^0$.

# 1. Khái niệm về ngôn ngữ

## 1.4. Ngôn ngữ

- Cho bảng chữ V.
- Một cách tổng quát, ngôn ngữ là một tập hợp gồm các xâu trên V.
- Ví dụ:
  - Các ngôn ngữ tự nhiên: Tiếng Việt, Tiếng Anh,…
  - Các ngôn ngữ lập trình: PASCAL, C,…

# 1. Khái niệm về ngôn ngữ

## 1.4. Ngôn ngữ

- Cho bảng chữ V và L là một ngôn ngữ trên V. Rõ ràng $L \subseteq V^*$.
- Ta có
  - $L=V^*$: L là ngôn ngữ gồm tất cả các xâu trên bảng chữ V.
  - $L=V^+$: L là ngôn ngữ gồm tất cả các xâu trên bảng chữ V và không chứa xâu rỗng ε.
- Ngôn ngữ rỗng là ngôn ngữ không chứa bất kỳ xâu nào. Ký hiệu ngôn ngữ rỗng là ∅.
- Chú ý: ngôn ngữ {ε} khác ngôn ngữ rỗng ∅, tức là $\{\varepsilon\} \neq \emptyset$.

# 1. Khái niệm về ngôn ngữ

## 1.4. Ngôn ngữ

- Ví dụ
  - Cho bảng chữ V={0, 1}.
  - Gọi $L_1$ là ngôn ngữ trên V, trong đó $L_1$ gồm các xâu bắt đầu bằng ký tự 0 và kết thúc bằng ký tự 1. Ta có $L_1$ ={01, 001, 0101, 01101,…}.
  - Gọi $L_2$ là ngôn ngữ trên V, trong đó $L_2$ gồm các xâu có độ dài khác 0 và độ dài là số chẵn. Ta có $L_2$={00,11, 01, 10, 0000, 1111, 1001, 010110, …}.
  - $L_3=\{(010)^n \mid n \in Z^+\}$. Ta có $L_3$ là ngôn ngữ trên V gồm các xâu 010, 010010, 010010010,…

# 1. Khái niệm về ngôn ngữ

## 1.4. Ngôn ngữ

- Chú ý rằng trên bảng chữ đã cho, các ngôn ngữ có thể chứa hữu hạn xâu hay chứa vô hạn xâu, tương ứng gọi là ngôn ngữ hữu hạn hay ngôn ngữ vô hạn.
  - Ví dụ
    - Cho bảng chữ V={a, b}.
    - $L_4$={a, ab, ba, aaa, bbaaa, bbbba, aaaabb, abababb, bbaabba, bbbaaab} là ngôn ngữ hữu hạn trên V.
    - $L_5=\{(ab)^n \mid n \in Z^+\}$ là ngôn ngữ vô hạn trên V.

# 1. Khái niệm về ngôn ngữ

## 1.5. Các phép toán trên ngôn ngữ

### 1.5.1. Giao

- Cho bảng chữ V và $L_1$, $L_2$ là các ngôn ngữ trên V.
- $L_1$ giao $L_2$ là một ngôn ngữ trên V gồm các xâu thuộc đồng thời cả $L_1$ và $L_2$, ký hiệu $L_1 \cap L_2$ .
- Ta có $L_1 \cap L_2 = \{x \in V^* | \, x \in L_1$ và $x \in L_2\}$.
  - Ví dụ
    - Cho bảng chữ $V=\{a, b\}$ và $L_1$, $L_2$ là các ngôn ngữ trên V, trong đó $L_1=\{a, ab, abb, bba, bbb\}$, $L_2=\{(ab)^n \mid n \in Z^+\}$.
    - Ta có $L_1 \cap L_2=\{ab\}$, $L_2 \cap L_1=\{ab\}$.
    - Rõ ràng $L_1 \cap L_2 = L_2 \cap L_1$.

# 1. Khái niệm về ngôn ngữ

## 1.5. Các phép toán trên ngôn ngữ

### 1.5.1. Giao

- Định nghĩa phép toán giao có thể mở rộng cho hữu hạn ngôn ngữ $L_1$, $L_2$, …, $L_n$ ($n \in Z^+ \setminus \{1\}$) trên V.
- Ta có $L_1 \cap L_2 \cap \ldots \cap L_n = \{x \in V^* \mid x \in L_1$ và $x \in L_2$ và … và $x \in L_n\}$.
- Ký hiệu $L_1 \cap L_2 \cap \ldots \cap L_n$ bằng $\bigcap_{i=1}^{n} L_i$.
  - Ví dụ
    - Cho bảng chữ V={a, b, c} và $L_1$, $L_2$ , $L_3$ là các ngôn ngữ trên V, trong đó $L_1$={ac, cab, abc, bba, bbb}, $L_2$={ac, abc, bbb, bca, ccc}, $L_3$={abc, bbb, bba}.
    - Ta có $L_1 \cap L_2 \cap L_3$={abc, bbb}.

# 1. Khái niệm về ngôn ngữ

## 1.5. Các phép toán trên ngôn ngữ

### 1.5.1. Giao

- Cho bảng chữ V và L, $L_1$, $L_2$, $L_3$ là các ngôn ngữ trên V.
- Một số tính chất
  - $L_1 \cap L_2 = L_2 \cap L_1$.
  - $(L_1 \cap L_2) \cap L_3 = L_1 \cap (L_2 \cap L_3)$.
  - $L \cap \varnothing = \varnothing \cap L = \varnothing$.
  - Nếu $L_1 \subseteq L_2$ thì $L_1 \cap L_2 = L_1$ (do đó $L \cap V^* = L$).

# 1. Khái niệm về ngôn ngữ

## 1.5. Các phép toán trên ngôn ngữ

### 1.5.2. Hợp

- Cho bảng chữ V và $L_1$, $L_2$ là các ngôn ngữ trên V.
- $L_1$ hợp $L_2$ là một ngôn ngữ trên V gồm các xâu thuộc $L_1$ hoặc thuộc $L_2$, ký hiệu $L_1 \cup L_2$.
- Ta có $L_1 \cup L_2 = \{x \in V^* | \ x \in L_1$ hoặc $x \in L_2\}$.
  - Ví dụ
    - Cho bảng chữ $V=\{a, b\}$ và $L_1$, $L_2$ là các ngôn ngữ trên V, trong đó $L_1=\{a, bba, abb\}$, $L_2=\{bb, ab, bba\}$.
    - Ta có $L_1 \cup L_2=\{a, bba, abb, bb, ab\}$,

      $L_2 \cup L_1=\{a, bba, abb, bb, ab\}$.
    - Rõ ràng $L_1 \cup L_2 = L_2 \cup L_1$.

# 1. Khái niệm về ngôn ngữ

## 1.5. Các phép toán trên ngôn ngữ

### 1.5.2. Hợp

- Định nghĩa phép toán hợp có thể mở rộng cho hữu hạn ngôn ngữ $L_1, L_2, \ldots, L_n$ ($n \in Z^+ \setminus \{1\}$) trên V.
- Ta có $L_1 \cup L_2 \cup \ldots \cup L_n = \{x \in V^* \mid x \in L_1$ hoặc $x \in L_2$ hoặc … hoặc $x \in L_n\}$.
- Ký hiệu $L_1 \cup L_2 \cup \ldots \cup L_n$ bằng $\bigcup_{i=1}^{n} L_i$.
  - Ví dụ
    - Cho bảng chữ $V=\{a, b, c\}$ và $L_1$, $L_2$, $L_3$ là các ngôn ngữ trên V, trong đó $L_1=\{ac, cab, abc, bba, bbb\}$, $L_2=\{ac, abc, bbb, bca, ccc\}$, $L_3=\{abc, bbb, bba\}$.
    - Ta có $L_1 \cup L_2 \cup L_3 = \{ac, cab, abc, bba, bbb, bca, ccc\}$.

# 1. Khái niệm về ngôn ngữ

## 1.5. Các phép toán trên ngôn ngữ

### 1.5.2. Hợp

- Cho bảng chữ V và L, $L_1$, $L_2$, $L_3$ là các ngôn ngữ trên V.
- Một số tính chất
  - $L_1 \cup L_2 = L_2 \cup L_1$.
  - $(L_1 \cup L_2) \cup L_3 = L_1 \cup (L_2 \cup L_3)$.
  - $L \cup \varnothing = \varnothing \cup L = L$.
  - Nếu $L_1 \subseteq L_2$ thì $L_1 \cup L_2 = L_2$ (do đó $L \cup V^* = V^*$).
  - $L_1 \cap (L_2 \cup L_3) = (L_1 \cap L_2) \cup (L_1 \cap L_3)$.
  - $L_1 \cup (L_2 \cap L_3) = (L_1 \cup L_2) \cap (L_1 \cup L_3)$.

# 1. Khái niệm về ngôn ngữ

## 1.5. Các phép toán trên ngôn ngữ

### 1.5.3. Hiệu

- Cho bảng chữ V và $L_1$, $L_2$ là các ngôn ngữ trên V.
- $L_1$ hiệu $L_2$ là một ngôn ngữ trên V gồm các xâu thuộc $L_1$ nhưng không thuộc $L_2$, ký hiệu $L_1 \backslash L_2$ hay $L_1 - L_2$.
- Ta có $L_1 \backslash L_2 = \{x \in V^* \mid x \in L_1 \text{ và } x \notin L_2\}$,
- Hay $L_1 - L_2 = \{x \in V^* \mid x \in L_1 \text{ và } x \notin L_2\}$.

# 1. Khái niệm về ngôn ngữ

## 1.5. Các phép toán trên ngôn ngữ

### 1.5.3. Hiệu

- Ví dụ
  - Cho bảng chữ V={a, b} và $L_1$, $L_2$ là các ngôn ngữ trên V, trong đó $L_1$={aaa, aab, babb, baaaba, aabb}, $L_2$={babb, aab, bbaa}.
  - Ta có $L_1 \backslash L_2$={aaa, baaaba, aabb}, $L_2 \backslash L_1$={bbaa}.
  - Rõ ràng $L_1 \backslash L_2 \neq L_2 \backslash L_1$.

# 1. Khái niệm về ngôn ngữ

## 1.5. Các phép toán trên ngôn ngữ

### 1.5.3. Hiệu

- Cho bảng chữ V và L, $L_1$, $L_2$, $L_3$ là các ngôn ngữ trên V.
- Một số tính chất
  - $L \setminus \varnothing = L$.
  - Nếu $L_1 \subseteq L_2$ thì $L_1 \setminus L_2 = \varnothing$.
  - $L_1 \setminus (L_2 \cup L_3) = (L_1 \setminus L_2) \cap (L_1 \setminus L_3)$.
  - $L_1 \setminus (L_2 \cap L_3) = (L_1 \setminus L_2) \cup (L_1 \setminus L_3)$.

# 1. Khái niệm về ngôn ngữ

## 1.5. Các phép toán trên ngôn ngữ

### 1.5.4. Bù

- Cho bảng chữ V và L là ngôn ngữ trên V.
- Bù của L là một ngôn ngữ trên V gồm các xâu không thuộc L, ký hiệu $\overline{L}$.
- Ta có $\overline{L} = \{x \in V^* \mid x \notin L\}$ hay $\overline{L} = V^* \setminus L$ .
  - Ví dụ
    - Cho bảng chữ V={a, b} và $L_1$, $L_2$ là các ngôn ngữ trên V, trong đó $L_1$={aaa, aab, babb, baaaba, aabb}, $L_2$={babb, aab, bbaa}.
    - Ta có $\overline{L_1} = V^* \setminus L_1, \overline{L_2} = V^* \setminus L_2$ .
    - Cho ví dụ một số xâu của $\overline{L_1}$ , của $\overline{L_2}$

# 1. Khái niệm về ngôn ngữ

## 1.5. Các phép toán trên ngôn ngữ

### 1.5.4. Bù

- Cho bảng chữ V và $L_1$, $L_2$ là các ngôn ngữ trên V.
- Một số tính chất
  - $\overline{\varnothing} = V^*, \overline{V^*} = \varnothing$ .
  - $\overline{\{\varepsilon\}} = V^+, \overline{V^+} = \{\varepsilon\}$.
  - $\overline{L_1 \cap L_2} = \overline{L_1} \cup \overline{L_2}$ .
  - $\overline{L_1 \cup L_2} = \overline{L_1} \cap \overline{L_2}$ .

# 1. Khái niệm về ngôn ngữ

## 1.5. Các phép toán trên ngôn ngữ

### 1.5.5. Ghép nối

- Cho bảng chữ V và $L_1$, $L_2$ là các ngôn ngữ trên V.
- Ghép nối *hay nhân ghép* $L_1$ với $L_2$, ký hiệu $L_1L_2$ hay $L_1.L_2$, được định nghĩa:

$$L_1L_2=\{xy\in V^* \mid x\in L_1 \text{ và } y\in L_2\}.$$

  - Ví dụ
    - Cho bảng chữ V={a, b} và $L_1$, $L_2$ là các ngôn ngữ trên V, trong đó $L_1$={a, ab, bba}, $L_2$={bb, ab}.
    - Ta có $L_1L_2$={abb, aab, abbb, abab, bbabb, bbaab}, $L_2L_1$={bba, bbab, bbbba, aba, abab, abbba}.
    - Rõ ràng $L_1L_2 \neq L_2L_1$.

# 1. Khái niệm về ngôn ngữ

## 1.5. Các phép toán trên ngôn ngữ

### 1.5.5. Ghép nối

- Cho bảng chữ V và L, $L_1$, $L_2$ là các ngôn ngữ trên V.
- Một số tính chất:
  - $L\{\varepsilon\}=\{\varepsilon\}L=L$.
  - $L_1(L_2L_3)=(L_1L_2)L_3$.
  - $L_1(L_2\cup L_3)=(L_1L_2)\cup(L_1L_3)$.

# 1. Khái niệm về ngôn ngữ

## 1.5. Các phép toán trên ngôn ngữ

### 1.5.6. Nghịch đảo

- Cho bảng chữ V và L là ngôn ngữ trên V.
- Nghịch đảo của L, ký hiệu $L^r$, được định nghĩa:

$$L^r = \{x^r \in V^* \mid x \in L\}.$$

  - Ví dụ
    - Cho bảng chữ V={a,b} và L là ngôn ngữ trên V, trong đó L={a, ab, bba, abab}.
    - Ta có $L^r$ ={a, ba, abb, baba}.

# 1. Khái niệm về ngôn ngữ

**1.5. Các phép toán trên ngôn ngữ**

**1.5.6. Nghịch đảo**

- Cho bảng chữ V và L, $L_1$, $L_2$ là các ngôn ngữ trên V.
- Một số tính chất:
  - $\{\varepsilon\}^r=\{\varepsilon\}$.
  - $\varnothing^r=\varnothing$.
  - $(L^r)^r=L$.
  - $(L_1L_2)^r=L_2^rL_1^r$.
  - $(L_1\cup L_2)^r=L_1^r\cup L_2^r$.

# 1. Khái niệm về ngôn ngữ

## 1.5. Các phép toán trên ngôn ngữ

### 1.5.7. Lũy thừa

- Cho bảng chữ V và L là ngôn ngữ trên V.
- Lũy thừa n ($n \in Z^+$) của L, ký hiệu $L^n$, được định nghĩa: $L^n$ = LLL..L (n lần L)
- Lưu ý: $L^0=\{\varepsilon\}$, $L^i = L^{i-1}L$
  - Ví dụ
    - Cho bảng chữ V={a, b} và L là ngôn ngữ trên V, trong đó L={a, abb}.
    - Ta có $L^0=\{\varepsilon\}$, $L^1=L$, $L^2$={aa, aabb, abba, abbabb}, $L^3$ = {aaa, aaabb, aabba, aabbabb, abbaa, abbaabb, abbabba, abbabbabb}, …

# 1. Khái niệm về ngôn ngữ

## 1.5. Các phép toán trên ngôn ngữ

### 1.5.8. Bao đóng, bao đóng dương

- Cho bảng chữ V và L là ngôn ngữ trên V.
- Bao đóng của L, ký hiệu $L^*$, được định nghĩa:

$$L^*=L^0\cup L^1\cup L^2\cup\ldots\cup.$$

- Bao đóng dương của L, ký hiệu $L^+$, được định nghĩa:

$$L^+=L^1\cup L^2\cup\ldots\cup.$$

- Chú ý phân biệt $V^*$ và $L^*$, $V^+$ và $L^+$.

# 1. Khái niệm về ngôn ngữ

## 1.6. Biểu diễn (đặc tả) ngôn ngữ

- Đối với ngôn ngữ đơn giản
- Sử dụng biểu diễn của tập hợp:
  - Phương pháp liệt kê
    - Ví dụ: Cho bảng chữ $V_1=\{0, 1\}$ và ngôn ngữ $L_1=\{101, 11100, 1011, 11111\}$ trên $V_1$.
  - Phương pháp nêu tính chất
    - Ví dụ: Cho bảng chữ $V_2=\{a, b\}$ và ngôn ngữ $L_2=\{(ab)^n \mid n \in Z^+\}$ trên $V_2$

  hay $L_2=\{ab, abab, ababab, abababab, \dots\}$

# 1. Khái niệm về ngôn ngữ

## 1.6. Biểu diễn (đặc tả) ngôn ngữ

- Đối với ngôn ngữ phức tạp

☞ Sử dụng văn phạm: Văn phạm là quy tắc để sản sinh ra ngôn ngữ.

- Ví dụ
  - Đối với ngôn ngữ tự nhiên: văn phạm là hệ thống ngữ pháp.
  - Đối với ngôn ngữ lập trình: văn phạm là quy tắc cú pháp thể hiện chương trình.

# 1. Khái niệm về ngôn ngữ

**Bài tập**

1. Cho bảng chữ V={a, b, c} và x=abcab, y=bcba là 2 xâu trên V. Hãy tìm |x|, |y|, xy, yx, (xy)x, y(xx), $x^r$, $y^r$, $((xy)x)^r$, $x^2$, $y^3$, $(x^r)^3$, $(y(xx))^2$
2. Cho bảng chữ V={0, 1, 2, 3} và x=32102. Hãy tìm các tiền tố, hậu tố của x.
3. Cho bảng chữ V gồm n ký tự. Hãy cho biết có bao nhiêu xâu có độ dài k được tạo thành từ V, với n và k là các số nguyên dương.

# 1. Khái niệm về ngôn ngữ

**Bài tập**

4. Cho x, y là các xâu trên bảng chữ V. Hãy chứng tỏ $(xy)^r = y^r x^r$.

5. Cho bảng chữ V={a, b, c} và $L_1$, $L_2$ , $L_3$ là các ngôn ngữ trên V, trong đó $L_1$={ab,ac,cab,bba,ccc}, $L_2$={ac, ab, bc, bbb, ccc}, $L_3$={abc, ccc, cca}.

Hãy tìm: $L_1 \cap L_2$, $L_2 \cap L_3$, $L_1 \cap L_2 \cap L_3$, $L_1 \cup L_2$, $L_1 \cup L_2 \cup L_3$, $L_1 \backslash L_2$, $L_2 \backslash L_1$, $L_1 \backslash (L_2 \backslash L_3)$, $L_2L_3$, $L_3L_2$, $(L_1)^r$, $(L_1 \cup L_2)^r$, $(L_3)^2$

6. Cho $L_1$, $L_2$ là các ngôn ngữ trên V. Hãy chứng tỏ $(L_1L_2)^r = L_2^r L_1^r$, $(L_1 \cup L_2)^r = L_1^r \cup L_2^r$.

7. Cho L là ngôn ngữ trên bảng chữ V. Hãy chứng tỏ $L^+ = L^*$ khi và chỉ khi $\varepsilon \in L$.

# 2. Khái niệm về văn phạm

## 2.1. Các định nghĩa

- Định nghĩa 1: Văn phạm G được định nghĩa bằng bộ gồm 4 thành phần $\Sigma$, $\Delta$, S, R, ký hiệu $G=(\Sigma,\Delta,S,R)$, trong đó:
  - $\Sigma$ là tập hợp gồm hữu hạn các ký tự để tạo nên các xâu của ngôn ngữ, gọi là các ký tự kết thúc;
  - $\Delta$ là tập hợp gồm hữu hạn các ký tự chưa kết thúc (hay *biến*), trong đó $\Sigma\cap\Delta=\varnothing$;
  - S là ký tự bắt đầu (hay *biến đầu*), trong đó $S\in\Delta$;
  - R là tập hợp gồm hữu hạn các luật sản xuất (luật sinh) có dạng $\alpha\rightarrow\beta$ với $\alpha\in(\Sigma\cup\Delta)^+$ và $\beta\in(\Sigma\cup\Delta)^*$, trong đó $\alpha$ phải có chứa biến và có ít nhất một luật sản xuất sao cho $\alpha$ là S.

# 2. Khái niệm về văn phạm

## 2.1. Các định nghĩa

- Một số quy ước
  - Ký tự kết thúc: ký hiệu bằng chữ thường.
  - Ký tự chưa kết thúc: ký hiệu bằng chữ hoa.
- Ví dụ
  - Cho văn phạm $G_1=(\Sigma, \Delta, S, R)$, trong đó:
    - $\Sigma=\{0, 1\}$
    - $\Delta=\{S\}$
    - $S\in\Delta$
    - $R=\{S\rightarrow 0S1, S\rightarrow\varepsilon\}$ hay $R=\{S\rightarrow 0S1|\varepsilon\}$.
  - Văn phạm trên có thể ký hiệu $G_1=(\{0, 1\},\{S\},S, \{S\rightarrow 0S1|\varepsilon\})$.

# 2. Khái niệm về văn phạm

## 2.1. Các định nghĩa

- Ví dụ
  - Cho văn phạm $G_2=(\Sigma, \Delta, S, R)$, trong đó:
    - $\Sigma=\{a, b\}$
    - $\Delta=\{S, A, B\}$
    - $S \in \Delta$
    - $R=\{S\rightarrow aS, S\rightarrow bA, A\rightarrow bbB, B\rightarrow bB, B\rightarrow aA, B\rightarrow \varepsilon\}$ hay $R=\{S\rightarrow aS|bA, A\rightarrow bbB, B\rightarrow bB|aA|\varepsilon\}$.
  - Văn phạm trên có thể ký hiệu $G_2=(\{a, b\},\{S, A, B\}, S, \{S\rightarrow aS|bA, A\rightarrow bbB, B\rightarrow bB|aA|\varepsilon\})$.

# 2. Khái niệm về văn phạm

## 2.1. Các định nghĩa

- Định nghĩa 2: Cho văn phạm $G=(\Sigma, \Delta, S, R)$ và các xâu u,v, trong đó $u\in(\Sigma \cup \Delta)^{+}$, $v\in(\Sigma\cup\Delta)^{*}$.
  - Suy dẫn trực tiếp: ta nói u suy dẫn trực tiếp *hay dẫn xuất trực tiếp* sinh ra v, kí hiệu $u\Rightarrow v$, khi và chỉ khi: $u=x\alpha y$, $v=x\beta y$ và $\alpha\rightarrow\beta\in R$, trong đó $x,y\in(\Sigma\cup\Delta)^{*}$.
  - Suy dẫn gián tiếp: ta nói u suy dẫn gián tiếp *hay dẫn xuất gián tiếp* sinh ra v, kí hiệu $u\Rightarrow^{+}v$, khi và chỉ khi: $u\Rightarrow w_1\Rightarrow w_2\Rightarrow\ldots\ w_k\Rightarrow v$, trong đó $w_1,w_2,\ldots, w_k\in(\Sigma\cup\Delta)^{*}$.
- Độ dài suy dẫn từ u sinh ra v là số lần áp dụng các luật sản xuất để từ u có được v.
- Độ dài suy dẫn trực tiếp là 1.

# 2. Khái niệm về văn phạm

## 2.1. Các định nghĩa

- Định nghĩa 3: Cho văn phạm $G=(\Sigma, \Delta, S, R)$. Xâu $w \in \Sigma^*$ do văn phạm G sinh ra khi và chỉ khi $S \Rightarrow^+ w$, tức là tồn tại suy dẫn $S \Rightarrow w_1 \Rightarrow w_2 \Rightarrow \ldots\ w_k \Rightarrow w$.
- Các xâu $w_1, w_2, \ldots, w_k$ được gọi là các dạng xâu của suy dẫn, trong đó $w_1, w_2, \ldots, w_k \in (\Sigma \cup \Delta)^*$.
- Như vậy quá trình sinh ra một xâu bởi văn phạm luôn bắt đầu bởi biến đầu S và sẽ dừng lại khi dạng xâu chỉ chứa các ký tự kết thúc.

# 2. Khái niệm về văn phạm

## 2.1. Các định nghĩa

- Ví dụ
  - Cho văn phạm $G_2=(\Sigma, \Delta, S, R)$, trong đó:
    - $\Sigma=\{a, b\}$, $\Delta=\{S, A, B\}$, $S\in\Delta$
    - $R=\{S\rightarrow aS,\ S\rightarrow bA,\ A\rightarrow bbB,\ B\rightarrow bB,\ B\rightarrow aA,\ B\rightarrow\varepsilon\}$
  - Ta có các suy dẫn:
    - $S\Rightarrow bA \Rightarrow bbbB \Rightarrow bbbaA \Rightarrow bbbabbB \Rightarrow bbbabb$
    - $S\Rightarrow aS\Rightarrow aaS\Rightarrow aabA\Rightarrow aabbbB\Rightarrow aabbb$
    - $S\Rightarrow aS\Rightarrow abA\Rightarrow abbbB\Rightarrow abbbbB\Rightarrow abbbbaA$ $\Rightarrow abbbbabbB\Rightarrow abbbbabbbB \Rightarrow abbbbabbb$
    - $S\Rightarrow aS\Rightarrow abA\Rightarrow abbbB\Rightarrow abbb$

# 2. Khái niệm về văn phạm

## 2.1. Các định nghĩa

- Định nghĩa 4: Cho văn phạm G=(Σ, Δ, S, R). Ngôn ngữ L do văn phạm G sinh ra, ký hiệu L(G), được định nghĩa: $L(G)=\{w \in \Sigma^* | S \Rightarrow^+ w\}$.
- Như vậy ngôn ngữ được sinh ra bởi một văn phạm gồm mọi xâu được sinh ra bởi văn phạm đó.

# 2. Khái niệm về văn phạm

## 2.1. Các định nghĩa

- Ví dụ
  - Cho văn phạm $G_1=(\Sigma, \Delta, S, R)$, trong đó $\Sigma=\{0, 1\}$, $\Delta=\{S\}$, $S\in\Delta$, $R=\{S\rightarrow 0S1, S\rightarrow\varepsilon\}$.
  - Ta có các suy dẫn:
    - $S\Rightarrow\varepsilon$
    - $S\Rightarrow 0S1\Rightarrow 01$
    - $S\Rightarrow 0S1\Rightarrow 00S11\Rightarrow 0011$
    - $S\Rightarrow 0S1\Rightarrow 00S11\Rightarrow 000S111\Rightarrow 000111$
    - …
    - Có thể thấy $L(G_1)=\{0^n1^n \mid n$ là số nguyên không âm$\}$

# 2. Khái niệm về văn phạm

## 2.2. Phân cấp văn phạm của Chomsky

- Cho văn phạm G=(Σ, Δ, S, R).
- Nếu không có hạn chế gì trên tập các luật sản xuất R thì G được gọi là văn phạm tổng quát *hay văn phạm tự do* (văn phạm loại 0). Ngôn ngữ được sinh ra bởi văn phạm tự do được gọi là ngôn ngữ liệt kê đệ quy.
- Nếu các luật sản xuất có dạng $\alpha \rightarrow \beta$ với $\alpha \in (\Sigma \cup \Delta)^+$, $\beta \in (\Sigma \cup \Delta)^*$ và $|\alpha| <= |\beta|$ thì G thì được gọi là văn phạm cảm ngữ cảnh (văn phạm loại 1). Ngôn ngữ được sinh ra bởi văn phạm cảm ngữ cảnh được gọi là ngôn ngữ cảm ngữ cảnh.

# 2. Khái niệm về văn phạm

## 2.2. Phân cấp văn phạm của Chomsky

- Ví dụ
  - Cho văn phạm $G_1=(\Sigma, \Delta, S, R)$, trong đó:
    - $\Sigma=\{a, b, c\}$
    - $\Delta=\{S, A, B, C\}$
    - $S\in\Delta$
    - $R=\{S\rightarrow aSAC,\ S\rightarrow abC,\ CA\rightarrow BA,\ BA\rightarrow BC,\ BC\rightarrow AC,\ bA\rightarrow bb,\ C\rightarrow c\}$.
  - $G_1$ là văn phạm cảm ngữ cảnh.
  - $L(G_1)=\{a^nb^nc^n \mid n\in Z^+\}$.
  - $L(G_1)$ là ngôn ngữ cảm ngữ cảnh.

# 2. Khái niệm về văn phạm

## 2.2. Phân cấp văn phạm của Chomsky

- Nếu các luật sản xuất có dạng $A \longrightarrow \beta$ với $A \in \Delta$ và $\beta \in (\Sigma \cup \Delta)^*$ thì G được gọi là văn phạm phi ngữ cảnh (văn phạm loại 2). Ngôn ngữ được sinh ra bởi văn phạm phi ngữ cảnh được gọi là ngôn ngữ phi ngữ cảnh.
- Nếu các luật sản xuất có dạng $A \longrightarrow \alpha B$, $A \longrightarrow B\alpha$, $A \longrightarrow \alpha$ với $A, B \in \Delta$ và $\alpha \in \Sigma^*$ thì G được gọi là văn phạm chính quy (văn phạm loại 3). Ngôn ngữ được sinh ra bởi văn phạm chính quy được gọi là ngôn ngữ chính quy.

# 2. Khái niệm về văn phạm

## 2.2. Phân cấp văn phạm của Chomsky

- Ví dụ
  - Cho văn phạm $G_2=(\Sigma, \Delta, S, R)$, trong đó:
    - $\Sigma=\{a, b\}$
    - $\Delta=\{S, A\}$
    - $S \in \Delta$
    - $R=\{S\rightarrow Sa, S\rightarrow Aa, A\rightarrow aAb, A\rightarrow ab\}$
  - $G_2$ là văn phạm phi ngữ cảnh.
  - $L(G_2)=\{a^n b^n a^m \mid n,m \in Z^+\}$.
  - $L(G_2)$ là ngôn ngữ phi ngữ cảnh.

# 2. Khái niệm về văn phạm

## 2.2. Phân cấp văn phạm của Chomsky

- Ví dụ
  - Cho văn phạm $G_3=(\Sigma, \Delta, S, R)$, trong đó:
    - $\Sigma=\{1\}$
    - $\Delta=\{S, A, B\}$
    - $S\in\Delta$
    - $R=\{S\rightarrow\varepsilon, S\rightarrow 1A, A\rightarrow 1B, B\rightarrow 1A, A\rightarrow 1\}$
  - $G_3$ là văn phạm chính quy.
  - $L(G_3)=\{1^{2n} \mid n$ là số nguyên không âm$\}$.
  - $L(G_3)$ là ngôn ngữ chính quy.

# 2. Khái niệm về văn phạm

## 2.2. Phân cấp văn phạm của Chomsky

- Lưu ý:
  - Văn phạm loại 1 là trường hợp đặc biệt của văn phạm loại 0.
  - Văn phạm loại 2 là trường hợp đặc biệt của văn phạm loại 1.
  - Văn phạm loại 3 là trường hợp đặc biệt của văn phạm loại 2.

# 3. Khái niệm về ôtômat

- Một ôtômát bao gồm một tập hợp các trạng thái và các điều khiển dịch chuyển từ trạng thái này sang trạng thái khác khi nhận dữ liệu vào
  - Ví dụ:

- Ôtômát biểu diễn hoạt động của bóng điện


- Ôtômát đoán nhận từ khóa int