Unidad 1
Introducción

Contenido

1.1 La importancia de estudiar los autómatas y lenguajes formales

1.2 Símbolos, alfabetos y cadenas

1.3 Operaciones sobre cadenas

1.4 Definición de lenguaje y operaciones sobre lenguajes

1.5 La jerarquía de Chomsky: Clasificación de gramáticas y lenguajes

Conceptos básicos

- **Símbolo** es una representación distinguible de cualquier información.
  - Los símbolos pueden ser cualesquiera, como w, 9, #, etc.,
  - Un símbolo es una entidad indivisible.
- **Alfabeto**
  - Un alfabeto es un conjunto finito no vacío de símbolos.
  - Ejemplos:

$$
\begin{aligned}
\Sigma_1 &= \{0, 1\} \\
\Sigma_2 &= \{a, b\} \\
\Sigma_3 &= \{\mathrm{na}, \mathrm{pa}, \mathrm{bra}, \mathrm{la}\} \\
\Sigma_4 &= \{\texttt{<HTML>}, \texttt{</HTML>}, \texttt{<BODY>}, \texttt{</BODY>}, \ldots\} \\
\Sigma_5 &= \{|\} \\
\Sigma_6 &= \{a, ab, aab\}
\end{aligned}
$$

Conceptos básicos

- Cardinalidad del alfabeto (número de elementos del alfabeto): $|\Sigma|>0$ y $|\Sigma|<\infty$
- Palabras
  - Una secuencia finita de símbolos de un alfabeto es una palabra sobre dicho alfabeto.

$$\begin{aligned}
\Sigma_1 &: \quad 0, 1, 00, 01, 11, 000, 1001101 \\
\Sigma_2 &: \quad a, aa, abb, ababa \\
\Sigma_3 &: \quad \mathrm{napa}, \mathrm{palabra} \\
\Sigma_6 &: \quad a, ab, aab, aaab, abab
\end{aligned}$$

  - Escribimos la palabra vacía, es decir, la palabra que no contiene ningún símbolo, como ε.
  - Definición, $|\varepsilon|=0$

- La longitud de una palabra sobre un alfabeto es el número de símbolos que contiene.

$$\begin{aligned}
\Sigma_1 &: \quad w = 0 \Longrightarrow |w| = 1, \quad w = 1001101 \Longrightarrow |w| = 7 \\
\Sigma_2 &: \quad w = a \Longrightarrow |w| = 1, \quad w = ababa \Longrightarrow |w| = 5 \\
\Sigma_3 &: \quad w = \text{napa} \Longrightarrow |w| = 2, \quad w = \text{palabra} \Longrightarrow |w| = 3 \\
\Sigma_6 &: \quad w = ab \Longrightarrow |w| = 1, \quad w = aab \Longrightarrow |w| = 1 \; o \; |w| = 2 \;\; ??
\end{aligned}$$

- Σ^* denota el conjunto de todas las cadenas que se pueden construir a partir del alfabeto Σ. Por ejemplo,
  - $\{a\}^* = \{\varepsilon, a, aa, aaa, aaaa, \ldots\}$
  - $\{a,b\}^* = \{\varepsilon, a, b, aa, ab, ba, bb, aaa, aab, \ldots\}$
- Notación $\{a\}^* = \{a^n \mid n >= 0\}$

- Convención
  - $\emptyset^* = \{\varepsilon\}$, ø denota al conjunto vacío.
  - Si $\sum \neq \emptyset$ entonces $\sum^*$ es un conjunto infinito de cadenas cada una de ellas de longitud finita.
- Observación:
  - $\{a, b\} = \{b, a\}$ pero $ab \neq ba$
  - $\{a, a, b\} = \{a, b\}$ pero $aab \neq ab$
  - $\emptyset \neq \{\varepsilon\} \neq \varepsilon$
    - $\emptyset$ conjunto sin elementos
    - $\{\varepsilon\}$ conjunto con un elemento
    - ε cadena, no es un conjunto

Operaciones sobre cadenas

- Concatenación
- *Si u*, $v \in A^*$, $u=a_1\ldots a_n$, $v=b_1\ldots b_n$, *se llama concatenación de u y v a la cadena uv dada por* $a_1\ldots a_n b_1\ldots b_m$.
- *Propiedades*
  - $|uv| = |u| + |v|, \ \forall\ u, v \in A^*$
  - $u(vw) = (uv)w, \ \forall\ u, v, w \in A^*$
  - $u\varepsilon = \varepsilon u = u, \ \forall\ u \in A^*$

Ejemplos

- Sean u=aab, v=abab, z=bba, entonces
  - uv=aababab, mientras que vu=ababaab
  - (uv)z= (aababab)bba = aabababbba
  - u(vz) = aab(ababbba) = aabababbba
  - $(a^m)(a^n) = a^{(n+m)}$ para cualquier, m,n >=0.
    - Ejemplos:
      - a^5=aaaaa, a^1=a, a^0=ε

- Definición
  - Si $u \in A^*$ entonces
    - $u^0 = \varepsilon$
    - $u^{i+1} = u^i u \quad \forall\ i >= 0$
- Si $u=a_1\ldots a_n \in A^*$, entonces, la cadena inversa de $u^{-1} = a_n\ldots a_1 \in A^*$.

Operaciones sobre conjuntos

- A, B, C denotan conjuntos de cadenas (Subconjuntos de Σ^*).
  - Complemento en Σ^*
    - $\sim A = \{x \in \Sigma^* \mid x \notin A\}$
  - Concatenación
    - $AB = \{ xy \mid x \in A \text{ y } y \in B\}$
  - Potencia n-ésima A^n de un conjunto A:
    - $A^0 = \{\varepsilon\}$
    - $A^{(n+1)} = AA^n$
    - Es decir, A^n se forma concatenando n copias de A consigo misma. $A^m A^n = A^{m+n}$

- Ejemplos
  - $\{ab, aab\}^0 = \{\varepsilon\}$
  - $\{ab, aab\}^1 = \{ab, aab\}$
  - $\{ab, aab\}^2 = \{abab, abaab, aabab, aabaab\}$
  - $\{ab, aab\}^3 = \{ababab, ababaab, abaabab, abaabaab, aababab, aababaab, aabaabab, aabaabaab\}$
- De igual forma:
  - $\{a,b\}^n = \{x \in \{a,b\}^* \mid |x|=n\}$
  - Que son todas las cadenas tomadas del alfabeto {a,b} de longitud n.

- Cerradura de Kleene: A* es la unión de todas las posibles potencias finitas de A:

$$\begin{aligned} A^* &\stackrel{def}{=} \bigcup_{n\geq 0} A^n \\ &= A^0 \cup A^1 \cup A^2 \cup A^3 \ldots \end{aligned}$$

- También se define como:

$$A^* = \{a_1 a_2 a_3 \ldots a_n | n \geq 0, a_i \in A, 1 \leq i \leq n\}$$

- Cuando n=0 entonces está siempre en A* para cualquier A. Ejemplo:
  - {aab, baa}*= {ε, aabbaa, aabaab, …}

- Cerradura positiva: A+ para un conjunto A es la unión de todas las potencias de A distintas de cero:

$$A^{+} \stackrel{def}{=} AA^{*} = \bigcup_{n \geq 1} A^{n}$$

Ejercicios

- Sea el alfabeto $\Sigma = \{a, b, c\}$ y los conjuntos $A=\{aab, bba\}$, $B=\{aa,aaaa,ab\}$; obtenga:
  - Σ^*
  - Σ^+
  - $A \cup B$
  - $A \cap B$
  - $\sim A$
  - AB
  - BA
  - A^2
  - B^2

- La unión, intersección y concatenación son asociativas:
  - $(A \cup B) \cup C = A \cup (B \cup C)$
  - $(A \cap B) \cup C = (A \cap B) \cap C)$
  - $(AB)C = A(BC)$
- Conmutatividad en unión e intersección
  - $A \cup B = B \cup A$
  - $A \cap B = B \cap A$
- El conjunto ø es aniquilador para concatenación:
  - $A ø = øA = ø$

- La concatenación se distribuye sobre la unión:

$$A(B \cup C) = AB \cup AC$$
$$(A \cup B)C = AC \cup BC$$
$$A(\bigcup_{i \in I} B_i) = \bigcup_{i \in I} AB_i$$

- La concatenación de conjuntos no se distribuye con respecto a la intersección.
  - Ejercicio: A={a, ab}, B={b}, C={ε}, calcule A(B∩C) y AB ∩ AC.

- La cerradura Kleene satisface las siguientes propiedades:

$$A^*A^* = A^*$$
$$A^{**} = A^*$$
$$A^* = \{\epsilon\} \cup AA^* = \{\epsilon\} \cup A^*A$$
$$\emptyset^* = \{\epsilon\}$$

Ejercicios

1. Dado el alfabeto $\Sigma = \{a, b, c\}$ y los conjuntos $A = \{a, aab, aaaab, aaaba, aaabba\}$, $B = \{\epsilon, aab, bbc, cca, bbc, aac\}$ y $C = \{cba, cccb, ccbcc, ccacc, aacbb\}$, obtenga los siguientes:

   (a) $A \cup B$, $A \cup C$, $B \cup C$.

   (b) $A \cap B, B \cap C$

   (c) AB, BC

   (d) B^3

   (e) $AB \cap AC$.

1.4 Definición de lenguaje y operaciones sobre lenguajes

- Llamamos lenguaje sobre el alfabeto $\sum$ a cualquier subconjunto de $\sum^*$.
- $\sum^*$, ø, y $\sum$ pueden considerarse como lenguajes.
- Un lenguaje es tan sólo una clase especial de conjunto, podemos especificar un lenguaje finito **por extensión** enumerando sus elementos entre llaves.
  - {aba, czr, d, f} es un lenguaje sobre el alfabeto {a, b, c, ..., z}
- Pero la mayoría de los lenguajes de interés son infinitos. En este caso podemos especificar un lenguaje **por comprensión** de la siguiente forma:

  - $L = \{w \in \sum^* \mid w \text{ cumple la propiedad } P\}$

Ejemplo

- $L = \{w \in \{0, 1\}^* \mid ceros(w) = unos(w)\}$, palabras que tienen el mismo número de ceros que de unos.

Operaciones del algebra de conjuntos

- Sean L_1 y L_2 dos lenguajes definidos sobre el alfabeto $\sum$.
  - Se define la **unión** de estos dos lenguajes como el lenguaje L sobre $\sum$ que se especifica como:
  - $L = L_1 \cup L_2 = \{w \in \sum^* \mid (w \in L_1) \vee (w \in L_2)\}$
- De forma análoga a la unión se pueden definir otras operaciones del álgebra de conjuntos como la *intersección, diferencia, y complemento* de lenguajes.
  - Por ejemplo, el complemento del lenguaje L sobre el alfabeto V será:
    - $\overline{L} = V^* - L$

Concatenación, potencia e inversión de lenguajes

- Sean L_1 y L_2 dos lenguajes definidos sobre el alfabeto $\sum$, la concatenación de estos dos lenguajes es otro lenguaje L definido como:
  - $L_1 \cdot L_2 = \{x \cdot y \in \sum^* \mid (x \in L_1) \wedge (y \in L_2)\}$
  - La definición anterior sólo es válida si L_1 y L_2 contienen al menos un elemento.
  - $\varnothing \cdot L = L \cdot \varnothing = \varnothing$

Concatenación de lenguajes

- Dados los lenguajes A,B,C sobre un alfabeto V , la concatenación de lenguajes es distributiva con respecto a la unión, esto es, se cumple que:

1. $A \cdot (B \cup C) = (A \cdot B) \cup (A \cdot C)$
2. $(B \cup C) \cdot A = (B \cdot A) \cup (C \cdot A)$

Potencia n-ésima

- Consiste en concatenar el lenguaje consigo mismo n veces. La definición inductiva es:

  1. $L^0 = \{\lambda\} = \{\varepsilon\}$
  2. $L^n = L \cdot L^{n-1}, \forall\, n > 0$
- Ejemplo:
- Si $L = \{ab, c\}$ es un lenguaje sobre el alfabeto $\{a, b, c\}$ entonces,
  - $L^0 = \{\lambda\}$
  - $L^1 = L = \{ab, c\}$
  - $L^2 = L \cdot L^1 = \{abab, abc, cab, cc\}$
  - $L^3 = L \cdot L^2 = \{ababab, ababc, abcab, abcc, cabab, cabc, ccab, ccc\}$

- Lenguaje inverso de L:
  - $L^R = \{w^R \mid w \in L\}$
- Cerradura positiva

$$L^+ = \bigcup_{n=1}^{\infty} L^n$$

- Cerradura

$$L^* = \bigcup_{n=0}^{\infty} L^n$$

1.5 La jerarquía de Chomsky, 1956

Gramática	Lenguaje	Modelo matemático
Tipo 0: sin restricción	Recursivamente enumerable (Nivel pragmático)	Máquina de Turing (MT)
Tipo 1: dependiente del contexto	Dependiente del contexto (Nivel semántico)	Autómata linealmente limitado (ALL)
Tipo 2: independiente del contexto	Independiente del contexto (Nivel sintáctico)	Autómata de pila (AP)
Tipo 3: regular	Regular (Nivel léxico)	Autómata finito (AF)

1.5 La jerarquía de Chomsky