Capítulo 1

Lenguajes formales

1.1. Alfabetos y palabras

Un **alfabeto** es un conjunto finito no vacío cuyos elementos se llaman **símbolos**. Denotamos un alfabeto arbitrario con la letra Σ.

Una **palabra** o **cadena** sobre un alfabeto Σ es cualquier sucesión finita de elementos de Σ. Admitimos la existencia de una única palabra que no tiene símbolos, la cual se denomina **palabra vacía** y se denota con λ. La palabra vacía desempeña, en la teoría de lenguajes formales, un papel similar al que desempeña el conjunto vacío $\varnothing$ en la teoría de conjuntos.

Ejemplo Sea $\Sigma = \{a, b\}$ el alfabeto que consta de los dos símbolos a y b. Las siguientes son palabras sobre Σ:

aba
$ababaaa$
$aaaab$

Obsérvese que $aba \neq aab$. El orden de los símbolos en una palabra es significativo ya que las palabras se definen como *sucesiones*, es decir, conjuntos *secuencialmente ordenados*.

El conjunto de *todas* las palabras sobre un alfabeto Σ, incluyendo la palabra vacía, se denota por Σ^*.

Ejemplo Sea $\Sigma = \{a, b, c\}$, entonces

$$\Sigma^* = \{\lambda, a, b, c, aa, ab, ac, ba, bb, bc, ca, cb, cc, aaa, aab, abc, baa, \ldots\}.$$

Observaciones:

1. Si bien un alfabeto Σ es un conjunto finito, Σ^* es siempre un conjunto

infinito (enumerable). En el caso más simple, Σ contiene solo un símbolo, $\Sigma = \{a\}$, y $\Sigma^* = \{\lambda, a, aa, aaa, aaaa, aaaaa, \ldots\}$.

2. Hay que distinguir entre los siguientes objetos, que son todos diferentes entre sí:

$$\varnothing \qquad \lambda \qquad \{\varnothing\} \qquad \{\lambda\}$$

3. La teoría de lenguajes se hace con referencia a un alfabeto Σ fijo (pero arbitrario).

Notación usada en la teoría de lenguajes	
Σ	denota un alfabeto.
Σ^*	denota el conjunto de todas las palabras que se pueden formar con los símbolos de Σ.
$a, b, c, d, e, \ldots$	denotan símbolos del alfabeto Σ.
$u, v, w, x, y, z, \ldots$	denotan palabras, es decir, sucesiones finitas de símbolos de Σ.
λ	denota la palabra vacía, es decir, la única palabra en Σ^* que no tiene símbolos.
A,B,C,...,L,M,N,...	denotan lenguajes (definidos más adelante).

1.2. Concatenación de palabras

Dado un alfabeto Σ, y dos palabras $u, v \in \Sigma^*$, la **concatenación de** u **y** v se denota como $u \cdot v$ o simplemente uv y se define descriptivamente así:

1. Si $v = \lambda$, entonces $u \cdot \lambda = \lambda \cdot u = u$. Es decir, la concatenación de cualquier palabra u con la palabra vacía, a izquierda o a derecha, es igual a u.

2. Si $u = a_1 a_2 \cdots a_n$, $v = b_1 b_2 \cdots b_m$, entonces

$$u \cdot v = a_1 a_2 \cdots a_n b_1 b_2 \cdots b_m.$$

Es decir, $u \cdot v$ es la palabra formada escribiendo los símbolos de u y a continuación los símbolos de v.

La concatenación de palabras se puede definir inductiva o recursivamente de la siguiente manera. Si $u, v \in \Sigma^*$, $a \in \Sigma$, entonces

1. $u \cdot \lambda = \lambda \cdot u = u$.

2. $u \cdot (va) = (u \cdot v)a$.

Propiedad. *La concatenación de palabras es una operación asociativa. Es decir, si* $u, v, w \in \Sigma^*$*, entonces*

$$(uv)w = u(vw).$$

Demostración: Se puede hacer escribiendo explícitamente las palabras u, v, w y usando la definición descriptiva de concatenación. También se puede dar una demostración inductiva usando la definición recursiva de concatenación (ejercicio opcional).

1.3. Potencias de una palabra

Dada $u \in \Sigma^*$ y $n \in \mathbb{N}$, se define (descriptivamente) u^n en la siguiente forma

$$\begin{aligned} u^0 &= \lambda, \\ u^n &= \underbrace{uu\cdots u}_{n \text{ veces}} \end{aligned}$$

Ejercicio Dar una definición recursiva de u^n.

1.4. Longitud de una palabra

La **longitud** de una palabra $u \in \Sigma^*$ se denota $|u|$ y se define como el número de símbolos de u (contando los símbolos repetidos). Es decir,

$$|u| = \begin{cases} 0, & \text{si } u = \lambda, \\ n, & \text{si } u = a_1a_2\cdots a_n \end{cases}$$

Ejemplo $|aba| = 3$, $|baaa| = 4$.

Ejemplo Si $w \in \Sigma^*$, $n, m \in \mathbb{N}$, demostrar que

$$|w^{n+m}| = |w^n| + |w^m|$$

Solución:
Caso $n, m \geq 1$. $|w^{n+m}| = |\underbrace{ww\cdots w}_{n+m \text{ veces}}| = (n+m)|w|$. Por otro lado,

$$|w^n| + |w^m| = |\underbrace{ww\cdots w}_{n \text{ veces}}| + |\underbrace{ww\cdots w}_{m \text{ veces}}| = n|w| + m|w|.$$

Caso $n = 0, m \geq 1$. $|w^{n+m}| = |w^{0+m}| = |w^m|$. Por otro lado,

$$|w^n| + |w^m| = |w^0| + |w^m| = |\lambda| + |w^m| = 0 + |w^m| = |w^m|.$$

Caso $m = 0, n \geq 1$. Similar al caso anterior.
Caso $n = 0, m = 0$. $|w^{n+m}| = |w^{0+0}| = |\lambda| = 0$. Por otro lado,

$$|w^n| + |w^m| = |w^0| + |w^0| = |\lambda| + |\lambda| = 0 + 0 = 0.$$

1.5. Inversa de una palabra

La **inversa** o **transpuesta** de una palabra $u \in \Sigma^*$ se denota u^{-1} y se define descriptivamente así:

$$u^{-1} = \begin{cases} \lambda, & \text{si } u = \lambda, \\ a_n \cdots a_2 a_1, & \text{si } u = a_1 a_2 \cdots a_n. \end{cases}$$

De la definición se observa claramente que la inversa de la inversa de una palabra es la misma palabra, es decir,

$$(u^{-1})^{-1} = u \qquad \text{para } u, \in \Sigma^*.$$

Ejercicio Dar una definición recursiva de u^{-1}.

Ejercicio Si $u, v \in \Sigma^*$, demostrar que $(uv)^{-1} = v^{-1}u^{-1}$.

Ejercicio Generalizar la propiedad del ejercicio anterior a la concatenación de n palabras.

1.6. Subpalabras, prefijos y sufijos

Una palabra v es una **subpalabra** o **subcadena** de u si existen palabras x, y tales que $u = xvy$. Nótese que x o y pueden ser λ y, por lo tanto, la palabra vacía es una subpalabra de cualquier palabra.

Un **prefijo** de u es una palabra v tal que $u = vw$ para alguna palabra $w \in \Sigma^*$. Se dice que v es un **prefijo propio** si $v \neq u$.

Similarmente, un **sufijo** de u es una palabra v tal que $u = wv$ para alguna palabra $w \in \Sigma^*$. Se dice que v es un **sufijo propio** si $v \neq u$.

Obsérvese que λ es un prefijo y un sufijo de toda palabra u ya que $u\lambda = \lambda u = u$. Por la misma razón, toda palabra u es prefijo y sufijo de sí misma.

Ejemplo Sea $\Sigma = \{a, b, c, d\}$, $u = bcbaadb$.

Prefijos de u :	Sufijos de u :
λ	λ
b	b
bc	db
bcb	adb
$bcba$	$aadb$
$bcbaa$	$baadb$
$bcbaad$	$cbaadb$
$bcbaadb$	$bcbaadb$

1.7. Lenguajes

Un **lenguaje** L sobre un alfabeto Σ es un subconjunto de Σ^*, es decir $L \subseteq \Sigma^*$. Casos extremos:

$$\begin{array}{ll} L = \varnothing, & \text{(Lenguaje vacío)} \\ L = \Sigma^*, & \text{(Lenguaje de todas las palabras sobre } \Sigma\text{)} \end{array}$$

Todo lenguaje L satisface $\varnothing \subseteq L \subseteq \Sigma^*$, y puede ser finito o infinito. Los lenguajes se denotan con letras mayúsculas $A, B, C, \ldots, L, M, N, \ldots$.

Ejemplos Los siguientes son ejemplos de lenguajes sobre los alfabetos dados.

- $\Sigma = \{a, b, c\}$. $L = \{a, aba, aca\}$.
- $\Sigma = \{a, b, c\}$. $L = \{a, aa, aaa, \ldots\} = \{a^n : n \geq 1\}$.
- $\Sigma = \{a, b, c\}$. $L = \{\lambda, a, aba, ab^2a, ab^3a, \ldots\} = \{ab^na : n \geq 1\} \cup \{\lambda\}$.
- $\Sigma = \{a, b, c, \ldots, x, y, z\}$. $L = \{u \in \Sigma^* : u \text{ aparece en el diccionario español}\}$. L es un lenguaje finito.
- $\Sigma = \{a, b, c\}$. $L = \{u \in \Sigma^* : u \text{ no contiene el símbolo } c\}$. Por ejemplo, $abbaab \in L$ pero $abbcaa \notin L$.
- $\Sigma = \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$. El conjunto $\mathbb{N}$ de los números naturales se puede definir como un lenguaje sobre Σ, en la siguiente forma:

$$\mathbb{N} = \{u \in \Sigma^* : u = 0 \text{ o } 0 \text{ no es un prefijo de } u\}.$$

Ejercicio Definir el conjunto de los enteros $\mathbb{Z} = \{\ldots, -2, -1, 0, 1, 2, \ldots\}$ como un lenguaje sobre un alfabeto adecuado.

1.8. Operaciones entre lenguajes

Puesto que los lenguajes sobre Σ son subconjuntos de Σ^*, las operaciones usuales entre conjuntos son también operaciones válidas entre lenguajes. Así, si A y B son lenguajes sobre Σ (es decir $A, B \subseteq \Sigma^*$), entonces los siguientes también son lenguajes sobre Σ:

$A \cup B$	Unión
$A \cap B$	Intersección
$A - B$	Diferencia
$\overline{A} = \Sigma^* - A$	Complemento

Estas operaciones entre lenguajes se llaman **operaciones conjuntistas** para distinguirlas de las **operaciones lingüísticas** (concatenación, potencia, inverso, clausura) que son extensiones a los lenguajes de las ya mencionadas operaciones entre palabras.

1.9. Concatenación de lenguajes

La **concatenación** de dos lenguajes A y B sobre Σ, notada $A \cdot B$ o simplemente AB se define como

$$AB = \{uv : u \in A, \ \ v \in B\}$$

En general $AB \neq BA$.

Ejemplo Si $\Sigma = \{a, b, c\}$, $A = \{a, ab, ac\}$, $B = \{b, b^2\}$, entonces

$$\begin{aligned} AB &= \{ab, ab^2, ab^2, ab^3, acb, acb^2\}. \\ BA &= \{ba, bab, bac, b^2a, b^2ab, b^2ac\} \end{aligned}$$

Ejemplo Si $\Sigma = \{a, b, c\}$, $A = \{ba, bc\}$, $B = \{b^n : n \geq 0\}$, entonces

$$\begin{aligned} AB &= \{bab^n : n \geq 0\} \cup \{bcb^n : n \geq 0\}. \\ BA &= \{b^nba : n \geq 0\} \cup \{b^nbc : n \geq 0\} \\ &= \{b^{n+1}a : n \geq 0\} \cup \{b^{n+1}c : n \geq 0\} \\ &= \{b^na : n \geq 1\} \cup \{b^nc : n \geq 1\}. \end{aligned}$$

Ejercicio Dé un ejemplo de un alfabeto Σ y dos lenguajes A, B sobre Σ tales que $AB = BA$

Propiedades de la concatenación de lenguajes. *Sean A, B, C lenguajes sobre Σ, es decir $A, B, C \subseteq \Sigma^*$. Entonces*

1. $A \cdot \varnothing = \varnothing \cdot A = \varnothing$.

2. $A \cdot \{\lambda\} = \{\lambda\} \cdot A = A$.

3. Propiedad Asociativa.
$$A \cdot (B \cdot C) = (A \cdot B) \cdot C.$$

4. Distributividad de la concatenación con respecto a la unión.
$$\begin{array}{c} A \cdot (B \cup C) = A \cdot B \cup A \cdot C. \\ (B \cup C) \cdot A = B \cdot A \cup C \cdot A. \end{array}$$

5. Propiedad distributiva generalizada. Si $\{B_i\}_{i \in I}$ es una familia cualquiera de lenguajes sobre Σ, entonces
$$A \cdot \bigcup_{i \in I} B_i = \bigcup_{i \in I} (A \cdot B_i)$$
$$\bigcup_{i \in I} B_i \cdot A = \bigcup_{i \in I} (B_i \cdot A).$$

Demostración:

1. $A \cdot \varnothing = \{uv : u \in A,\ v \in \varnothing\} = \varnothing$.

2. $A \cdot \{\lambda\} = \{uv : u \in A,\ v \in \{\lambda\}\} = \{u : u \in A\} = A$.

3. Se sigue de la asociatividad de la concatenación de palabras.

4. Caso particular de la propiedad general, demostrada a continuación.

5. Demostración de la igualdad $A \cdot \bigcup\limits_{i \in I} B_i = \bigcup\limits_{i \in I} (A \cdot B_i)$:

$$\begin{array}{rcl} x \in A \cdot \bigcup_{i \in I} B_i & \Longleftrightarrow & x = u \cdot v, \quad \text{con } u \in A \ \&\ v \in \bigcup_{i \in I} B_i \\ & \Longleftrightarrow & x = u \cdot v, \quad \text{con } u \in A \ \&\ v \in B_j, \quad \text{para algún } j \in I \\ & \Longleftrightarrow & x \in A \cdot B_j, \quad \text{para algún } j \in I \\ & \Longleftrightarrow & x \in \bigcup_{i \in I} (A \cdot B_i). \end{array}$$

La igualdad $\bigcup\limits_{i \in I} B_i \cdot A = \bigcup\limits_{i \in I} (B_i \cdot A)$ se demuestra de forma similar.

Observaciones:

- La propiedad asociativa permite escribir concatenaciones de tres o más lenguajes sin necesidad de usar paréntesis.
- En general, no se cumple que $A \cdot (B \cap C) = A \cdot B \cap A \cdot C$. Es decir, la concatenación no es distributiva con respecto a la intersección. Contraejemplo: $A = \{a, \lambda\}$, $B = \{\lambda\}$, $C = \{a\}$.

  Se tiene: $A \cdot (B \cap C) = \{a, \lambda\} \cdot \varnothing = \varnothing$. Por otro lado,

  $$A \cdot B \cap A \cdot C = \{a, \lambda\} \cdot \{\lambda\} \cap \{a, \lambda\} \cdot \{a\} = \{a, \lambda\} \cap \{a^2, a\} = \{a\}.$$

Ejercicio Una de las dos contenencias siguientes es verdadera y la otra es falsa. Demostrar o refutar, según sea el caso:

1. $A \cdot (B \cap C) \subseteq A \cdot B \cap A \cdot C$.
2. $A \cdot B \cap A \cdot C \subseteq A \cdot (B \cap C)$.

1.10. Potencias de un lenguaje

Dado un lenguaje A sobre Σ ($A \subseteq \Sigma^*$), y $n \in \mathbb{N}$, se define A^n en la siguiente forma

$$\begin{aligned} A^0 &= \{\lambda\}, \\ A^n &= \underbrace{AA \cdots A}_{n \text{ veces}} \end{aligned}$$

Esta definición generaliza a lenguajes la definición de potenciación de palabras.

1.11. La clausura de Kleene de un lenguaje

La **clausura de Kleene** o **estrella de Kleene** de un lenguaje A, $A \subseteq \Sigma^*$, es la unión de todas las potencias de A y se denota por A^*.

$$\boxed{A^* = \bigcup_{i \geq 0} A^i = A^0 \cup A^1 \cup A^2 \cup \cdots \cup A^n \cdots} \qquad \text{(Descripción 1)}$$

A^* se puede describir de la siguiente manera

$$\boxed{\begin{aligned} A^* &= \text{conjunto de } \textit{todas} \text{ las concatenaciones} \\ & \text{ de palabras de } A \text{, incluyendo } \lambda \\ &= \{u_1 \cdots u_n : \quad u_i \in A, \quad n \geq 0\} \end{aligned}} \qquad \text{(Descripción 2)}$$

De manera similar se define la **clausura positiva** de un lenguaje A, $A \subseteq \Sigma^*$, la cual se denota por A^+.

$$\boxed{A^+ = \bigcup_{i\geq 1} A^i = A^1 \cup A^2 \cup \cdots \cup A^n \cdots}$$

A^+ se puede describir de la siguiente manera

$$\boxed{\begin{array}{rcl} A^+ & = & \text{conjunto de } \textit{todas} \text{ las concatenaciones de palabras de } A, \\ & = & \{u_1 \cdots u_n : \quad u_i \in A, \quad n \geq 1\} \end{array}}$$

Obsérvese que $A^* = A^+ \cup \{\lambda\}$ y que $A^* = A^+$ si y solamente si $\lambda \in A$.

Propiedades. Sea A un lenguaje sobre Σ, $(A \subseteq \Sigma^*)$.

1. $A^+ = A^* \cdot A = A \cdot A^*$
2. $A^* \cdot A^* = A^*$
3. $\big(A^*\big)^n = A^*$, para todo $n \geq 1$.
4. $\big(A^*\big)^* = A^*$.
5. $A^+ \cdot A^+ \subseteq A^+$.
6. $\big(A^*\big)^+ = A^*$.
7. $\big(A^+\big)^* = A^*$.
8. $\big(A^+\big)^+ = A^+$.

Demostración:

1. $$\begin{array}{rcl} A \cdot A^* & = & A \cdot (A^0 \cup A^1 \cup A^2 \cup \cdots) \\ & = & A^1 \cup A^2 \cup A^3 \cup \cdots \\ & = & A^+. \end{array}$$

   Similarmente se demuestra que $A^* \cdot A = A^+$.

2. Si $x \in A^* \cdot A^*$, entonces $x = u \cdot v$, con $u \in A^*$, $v \in A^*$. Entonces, $x = u \cdot v$, con $u = u_1 u_2 \cdots u_n, \quad u_i \in A, \quad n \geq 0$ y $v = v_1 v_2 \cdots v_m, \quad v_i \in A, \quad m \geq 0$.

   De donde
   $$x = u \cdot v = u_1 \cdot u_2 \cdots u_n \cdot v_1 \cdot v_2 \cdots v_m.$$

con $u_i \in A$, $v_i \in A$, $\quad n \geq 0$. Por lo tanto, x es una concatenación de $n+m$ palabras de A. Así que $x \in A^*$.

Recíprocamente, si $x \in A^*$, entonces $x = x \cdot \lambda \in A^* \cdot A^*$.

Esto prueba la igualdad de los conjuntos $A^* \cdot A^*$ y A^*.

3. Se sigue de la propiedad anterior.

4.
$$\begin{aligned} \big(A^*\big)^* &= \big(A^*\big)^0 \cup \big(A^*\big)^1 \cup \big(A^*\big)^2 \cup \cdots \\ &= \{\lambda\} \cup A^* \cup A^* \cup A^* \cup \cdots \\ &= A^*. \end{aligned}$$

5. La demostración de esta propiedad es similar a la de la propiedad 2, pero con la restricción $m, n \geq 1$.

   En general, no se tiene la igualdad $A^+ \cdot A^+ = A^+$; más adelante se mostrará un contraejemplo.

6.
$$\begin{aligned} \big(A^*\big)^+ &= \big(A^*\big)^1 \cup \big(A^*\big)^2 \cup \big(A^*\big)^3 \cup \cdots \\ &= A^* \cup A^* \cup A^* \cup \cdots \\ &= A^*. \end{aligned}$$

7.
$$\begin{aligned} \big(A^+\big)^* &= \big(A^+\big)^0 \cup \big(A^+\big)^1 \cup \big(A^+\big)^2 \cup \cdots \\ &= \{\lambda\} \cup A^+ \cup A^+A^+ \cup \cdots \\ &= A^* \cup \ (\text{conjuntos contenidos en } A^+) \\ &= A^* \end{aligned}$$

8.
$$\begin{aligned} \big(A^+\big)^+ &= \big(A^+\big)^1 \cup \big(A^+\big)^2 \cup \big(A^+\big)^3 \cup \cdots\,, \\ &= A^+ \cup \ (\text{conjuntos contenidos en } A^+) \\ &= A^+ \end{aligned}$$

Contraejemplo de $\boxed{A^+ \cdot A^+ = A^+}$.
Sea $\Sigma = \{a, b\}$, $A = \{a\}$. Se tiene

$$A^+ = A^1 \cup A^2 \cup \cdots = \{a\} \cup \{aa\} \cup \{aaa\} \cup \cdots = \{a^n : n \geq 1\}$$

Por otro lado,

$$A^+ \cdot A^+ = \{a, a^2, a^3, \ldots\} \cdot \{a, a^2, a^3, \ldots\} = \{a^2, a^3, a^4, \ldots\} = \{a^n : n \geq 2\}.$$

Observación: Según las definiciones dadas, Σ^* tiene dos significados:

Σ^* = conjunto de las palabras sobre el alfabeto Σ.
Σ^* = conjunto de todas las concatenaciones de palabras de Σ.

No hay conflicto de notaciones porque las dos definiciones anteriores de Σ^* dan lugar al mismo conjunto.

Ejercicio Sean $A, B \subseteq \Sigma^*$. Demostrar que

$$(A \cup B)^* = (A^* B^*)^* \tag{1.1}$$

Ayuda: tener en cuenta tanto la descripción 1 como la descripción 2 presentadas arriba.

1.12. Inverso de un lenguaje

Dado A un lenguaje sobre Σ, se define A^{-1} de la siguiente forma:

$$A^{-1} = \{u^{-1} : u \in A\}.$$

Propiedades. Sean A y B lenguajes sobre Σ (es decir, $A, B \subseteq \Sigma^*$).

1. $(A \cdot B)^{-1} = B^{-1} \cdot A^{-1}$.
2. $(A \cup B)^{-1} = A^{-1} \cup B^{-1}$.
3. $(A \cap B)^{-1} = A^{-1} \cap B^{-1}$.
4. $(A^{-1})^{-1} = A$.
5. $(A^*)^{-1} = (A^{-1})^*$.
6. $(A^+)^{-1} = (A^{-1})^+$.

Demostración:

1.
$$\begin{aligned} x \in (A \cdot B)^{-1} &\iff x = u^{-1}, \quad \text{donde } u \in A \cdot B. \\ &\iff x = u^{-1}, \quad \text{donde } u = vw,\ v \in A, w \in B. \\ &\iff x = (vw)^{-1}, \quad \text{donde } v \in A, w \in B. \\ &\iff x = w^{-1}v^{-1}, \quad \text{donde } v \in A, w \in B. \\ &\iff x \in B^{-1} \cdot A^{-1} \end{aligned}$$

2. ***Ejercicio***

3. ***Ejercicio***

4. ***Ejercicio***

5.
$$
\begin{aligned}
x \in (A^*)^{-1} &\iff x = u^{-1}, \text{ donde } u \in A^*. \\
&\iff x = (u_1 \cdot u_2 \cdots u_n)^{-1}, \text{ donde los } u_i \in A,\ n \geq 0. \\
&\iff x = u_n^{-1} \cdot u_2^{-1} \cdots u_1^{-1}, \text{ donde los } u_i \in A,\ n \geq 0. \\
&\iff x \in (A^{-1})^*.
\end{aligned}
$$

6. ***Ejercicio***

Ejercicio ¿Se pueden generalizar las propiedades 2 y 3 anteriores para uniones e intersecciones arbitrarias, respectivamente?

1.13. Lenguajes regulares

Los **lenguajes regulares** sobre un alfabeto dado Σ son todos los lenguajes que se pueden formar a partir de los lenguajes básicos $\varnothing$, $\{\lambda\}$, $\{a\}$, $a \in \Sigma$, por medio de las operaciones de unión, concatenación y estrella de Kleene.

Podemos dar una definición recursiva de los lenguajes regulares. Sea Σ un alfabeto.

1. $\varnothing$, $\{\lambda\}$ y $\{a\}$, para cada $a \in \Sigma$, son lenguajes regulares sobre Σ. Estos son los denominados lenguajes regulares básicos.

2. Si A y B son lenguajes regulares sobre Σ, también lo son

$$
\begin{array}{cl}
A \cup B & \text{(unión)} \\
A \cdot B & \text{(concatenación)} \\
A^* & \text{(estrella de Kleene)}
\end{array}
$$

Obsérvese que Σ y Σ^* son lenguajes regulares sobre Σ.

Ejemplos Sea $\Sigma = \{a, b\}$. Los siguientes son lenguajes regulares sobre Σ.

1. El lenguaje A de todas las palabras que tienen exactamente una a:

$$A = \{b\}^* \cdot \{a\} \cdot \{b\}^*.$$

2. El lenguaje B de todas las palabras que comienzan con b:

$$B = \{b\} \cdot \{a, b\}^*.$$

3. El lenguaje C de todas las palabras que contienen la cadena ba:

$$C = \{a, b\}^* \cdot \{ba\} \cdot \{a, b\}^*.$$

4. $(\{a\} \cup \{b\}^*) \cdot \{a\}$.
5. $\left[\left(\{a\}^* \cup \{b\}^*\right) \cdot \{b\}\right]^*$.

1.14. Expresiones regulares

Las expresiones regulares representan lenguajes regulares y su propósito es simplificar la escritura de los lenguajes regulares.

La siguiente es la definición recursiva de las **expresiones regulares** sobre un alfabeto Σ dado.

1. Expresiones regulares básicas:

   $\varnothing$ es una expresión regular que representa al lenguaje $\varnothing$.
   λ es una expresión regular que representa al lenguaje $\{\lambda\}$.
   a es una expresión regular que representa al lenguaje $\{a\}$, $a \in \Sigma$.

2. Si R y S son expresiones regulares sobre Σ, también lo son:

$$RS$$
$$R \cup S$$
$$R^*$$

   RS representa la concatenación de los lenguajes representados por R y S; $R \cup S$ representa su unión, y R^* representa la clausura de Kleene del lenguaje representado por R.

Ejemplo Dado el alfabeto $\Sigma = \{a, b, c\}$,

$$(a \cup b^*)a^*(bc)^*$$

es una expresión regular que representa al lenguaje

$$(\{a\} \cup \{b\}^*) \cdot \{a\}^* \cdot \{bc\}^*.$$

Ejemplo Dado el alfabeto $\Sigma = \{a, b\}$,

$$(\lambda \cup a)^*(a \cup b)^*(ba)^*$$

es una expresión regular que representa al lenguaje

$$(\{\lambda\} \cup \{a\})^* \cdot \{a, b\}^* \cdot \{ba\}^*.$$

Ejemplos Los tres primeros lenguajes de la sección 1.13 podemos representarlos con expresiones regulares:

1. El lenguaje A de todas las palabras que tienen exactamente una a:

$$A = b^*ab^*.$$

2. El lenguaje B de todas las palabras que comienzan con b:

$$B = b(a \cup b)^*.$$

3. El lenguaje C de todas las palabras que contienen la cadena ba:

$$C = (a \cup b)^*ba(a \cup b)^*.$$

Observación: La representación de lenguajes regulares por medio de expresiones regulares no es única. Es posible que haya varias expresiones regulares diferentes para el mismo lenguaje. Por ejemplo, $b(a \cup b)^*$ y $b(b \cup a)^*$ representan el mismo lenguaje. Otro ejemplo: las dos expresiones regulares

$$(a \cup b)^* \qquad (a^*b^*)^*$$

representan el mismo lenguaje en razón de la igualdad (1.1) de la sección 1.11.

Ejemplos Encontrar expresiones regulares que representen los siguientes lenguajes, definidos sobre el alfabeto $\Sigma = \{a, b\}$.

1. Lenguaje de todas las palabras que comienzan con b y terminan con a.

   Solución: $b(a \cup b)^*a$.

2. Lenguaje de todas las palabras que tienen exactamente dos a's.

   Solución: $b^*ab^*ab^*$.

3. Lenguaje de todas las palabras que tienen un número par de símbolos (palabras de longitud par).

   Solución: $(aa \cup ab \cup ba \cup bb)^*$.

4. Lenguaje de todas las palabras que tienen un número impar de símbolos (palabras de longitud impar).

   Solución: $a(aa \cup ab \cup ba \cup bb)^* \cup b(aa \cup ab \cup ba \cup bb)^*$.

5. Lenguaje de todas las palabras que tienen un número par de $a's$.

   Soluciones:

$$b^*(ab^*a)^*b^*.$$
$$(ab^*a \cup b)^*.$$
$$(b^*ab^*ab^*)^* \cup b^*.$$
$$b^*(b^*ab^*ab^*)^*b^*.$$

Ejemplo Encontrar una expresión regular que represente el lenguaje de todas las palabras que no contienen la cadena bc, definido sobre el alfabeto $\Sigma = \{a, b, c\}$.

Solución: $c^*(b \cup ac^*)^*$.

Ejercicio Encontrar expresiones regulares para los siguientes lenguajes:

1. $\Sigma = \{a, b\}$. Lenguaje de todas las palabras que tienen la cadena ab un número par de veces.

2. $\Sigma = \{a, b\}$. Lenguaje de todas las palabras que tienen un número impar de a's.

3. $\Sigma = \{a, b\}$. Lenguaje de todas las palabras que tienen un número par de a's o un número impar de b's.

4. $\Sigma = \{a, b, c\}$. Lenguaje de todas las palabras que tienen un número par de símbolos.

5. $\Sigma = \{a, b, c\}$. Lenguaje de todas las palabras que tienen un número impar de símbolos.

6. $\Sigma = \{a, b, c\}$. Lenguaje de todas las palabras que comienzan con c y terminan con b.

7. $\Sigma = \{a, b, c\}$. Lenguaje de todas las palabras que no contienen la cadena cc.

8. (Opcional, ¡Difícil!) $\Sigma = \{a, b\}$. Lenguaje de todas las palabras que tienen un número par de a's y un número impar de b's.

Observación:

No todos los lenguajes sobre un alfabeto dado Σ son regulares. Más adelante se mostrará que el lenguaje

$$L = \{\lambda, ab, aabb, aaabbb, \dots\} = \{a^n b^n : n \geq 0\}$$

sobre $\Sigma = \{a, b\}$ no se puede representar por medio de una expresión regular, y por lo tanto, no es un lenguaje regular.