Capítulo 1

Alfabetos, cadenas y lenguajes

1.1. Alfabetos y cadenas

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

Una **cadena** o **palabra** sobre un alfabeto Σ es cualquier sucesión finita de elementos de Σ. Admitimos la existencia de una única cadena que no tiene símbolos, la cual se denomina **cadena vacía** y se denota con λ. La cadena 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 cadenas sobre Σ:

aba
$ababaaa$
$aaaab$.

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

Ejemplo El alfabeto $\Sigma = \{0, 1\}$ se conoce como *alfabeto binario*. Las cadenas sobre este alfabeto son secuencias finitas de ceros y unos, llamadas *secuencias binarias*, tales como

001
1011
001000001.

Ejemplo $\Sigma = \{a, b, c, \ldots, x, y, z\}$, el alfabeto del idioma castellano. Las palabras oficiales del castellano (las que aparecen en el diccionario DRA) son cadenas sobre Σ.

Ejemplo El alfabeto utilizado por muchos de los llamados *lenguajes de programación* (como Pascal o C) es el conjunto de caracteres ASCII (o un subconjunto de él) que incluye, por lo general, las letras mayúsculas y minúsculas, los símbolos de puntuación y los símbolos matemáticos disponibles en los teclados estándares.

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

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

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

En la siguiente tabla aparecen las convenciones de notación corrientemente utilizadas en la teoría de lenguajes formales. De ser necesario, se emplean subíndices.

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

- Algunos autores denotan la cadena vacía con la letra griega ε. Preferimos denotarla con λ porque ε tiende a confundirse con el símbolo $\in$ usado para la relación de pertenencia.
- 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 $\Sigma^* = \{\lambda, a, aa, aaa, aaaa, aaaaa, \ldots\}$.
- Hay que distinguir entre los siguientes cuatro objetos, que son todos diferentes entre sí: $\varnothing$, λ, $\{\varnothing\}$ y $\{\lambda\}$.
- La mayor parte de la teoría de lenguajes se hace con referencia a un alfabeto Σ fijo (pero arbitrario).

1.2. Concatenación de cadenas

Dado un alfabeto Σ y dos cadenas $u, v \in \Sigma^*$, 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 cadena u con la cadena 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 cadena formada escribiendo los símbolos de u y a continuación los símbolos de v.

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

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

Demostración: Se puede hacer escribiendo explícitamente las cadenas 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). $\square$

1.3. Potencias de una cadena

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

$$u^0 = \lambda,$$
$$u^n = \underbrace{uu \cdots u}_{n \text{ veces}}.$$

Ejercicio Dar una definición recursiva de u^n.

1.4. Longitud de una cadena

La **longitud** de una cadena $u \in \Sigma^*$ 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. Reflexión o inversa de una cadena

La **reflexión** o **inversa** de una cadena $u \in \Sigma^*$ se denota u^R y se define descriptivamente así:

$$u^R = \begin{cases} \lambda, & \text{si } u = \lambda, \\ a_n\cdots a_2a_1, & \text{si } u = a_1a_2\cdots a_n. \end{cases}$$

De la definición se observa claramente que la reflexión de la reflexión de una cadena es la misma cadena, es decir,

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

Ejercicio Dar una definición recursiva de u^R.

Ejercicio Si $u, v \in \Sigma^*$, demostrar que $(uv)^R = v^Ru^R$. Generalizar esta propiedad a la concatenación de n cadenas.

✎ Algunos autores escriben u^{-1} en lugar de u^R para denotar la reflexión de una cadena u.

1.6. Subcadenas, prefijos y sufijos

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

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

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

Obsérvese que λ es un prefijo y un sufijo de toda cadena u ya que $u\lambda = \lambda u = u$. Por la misma razón, toda cadena 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
$bcba$	$aadb$
$bcbaa$	$baadb$
$bcbaad$	$cbaadb$
$bcbaadb$	$bcbaadb$

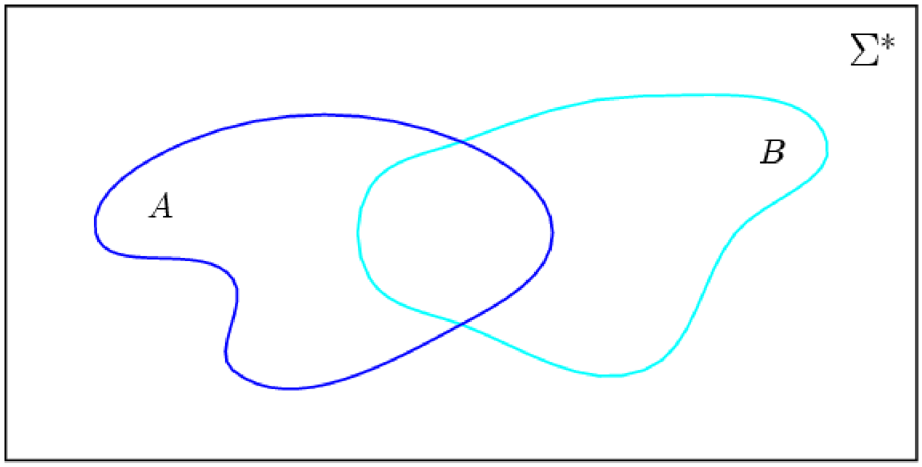
1.7. Lenguajes

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

Casos extremos:

$L = \varnothing$,	lenguaje vacío.
$L = \Sigma^*$,	lenguaje de todas las cadenas sobre Σ.

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



Ejemplo Los siguientes son ejemplos de lenguajes sobre los alfabetos especificados.

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

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

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

✎ El concepto abstracto de "lenguaje", tal como se ha definido, no es exactamente la misma noción utilizada en la expresión "lenguaje de programación". Para precisar la relación entre estos conceptos, consideremos el alfabeto Σ de los caracteres ASCII. Un programa en C o en Pascal, por ejemplo, es simplemente una cadena de símbolos de Σ y, por lo tanto, un conjunto de programas es un lenguaje (en el sentido formal definido en esta sección).

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* o *booleanas* para distinguirlas de las *operaciones lingüísticas* (concatenación, potencia, inverso, clausura) que son extensiones a los lenguajes de las operaciones entre cadenas.

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 diferentes 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.

$$\begin{aligned} A \cdot (B \cup C) &= A \cdot B \cup A \cdot C. \\ (B \cup C) \cdot A &= B \cdot A \cup C \cdot A. \end{aligned}$$

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

$$A \cdot \bigcup_{i\in I} B_i = \bigcup_{i\in I} (A \cdot B_i),$$
$$\Big(\bigcup_{i\in I} B_i\Big) \cdot A = \bigcup_{i\in I} (B_i \cdot A).$$

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 cadenas.

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

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

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

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

- 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 cadenas.

Ejercicio Dar una definición recursiva de A^n.

1.11. La clausura de Kleene de un lenguaje

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

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

A^* se puede describir de la siguiente manera

(Descripción 2)

A^*	$=$	conjunto de *todas* las concatenaciones de cadenas de A, incluyendo λ
	$=$	$\{u_1 \cdots u_n : \quad u_i \in A, \quad n \geq 0\}$

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

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

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

Propiedades de $*$ y $+$. Sea A un lenguaje sobre Σ, es decir, $A \subseteq \Sigma^*$.

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

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 $n + m$ cadenas 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.
$$\begin{aligned}(A^*)^* &= (A^*)^0 \cup (A^*)^1 \cup (A^*)^2 \cup \cdots \\ &= \{\lambda\} \cup A^* \cup A^* \cup A^* \cup \cdots \\ &= A^*.\end{aligned}$$

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

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

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

✎ Contraejemplo de $A^+ \cdot A^+ = A^+$. 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,

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

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

Σ^* = conjunto de las cadenas sobre el alfabeto Σ.

Σ^* = conjunto de todas las concatenaciones de cadenas de Σ.

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

$\boxed{\textit{Ejercicio}}$ Sean $A, B \subseteq \Sigma^*$. Demostrar que

$$(A \cup B)^* = (A^* B^*)^*.$$

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

1.12. Reflexión o inverso de un lenguaje

Dado A un lenguaje sobre Σ, se define A^R de la siguiente forma:

$$A^R = \{u^R : u \in A\}.$$

A^R se denomina la **reflexión** o el **inverso** de A.

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

1. $(A \cdot B)^R = B^R \cdot A^R$.
2. $(A \cup B)^R = A^R \cup B^R$.
3. $(A \cap B)^R = A^R \cap B^R$.
4. $\left(A^R\right)^R = A$.
5. $(A^*)^R = \left(A^R\right)^*$.
6. $(A^+)^R = \left(A^R\right)^+$.

Demostración:

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

2. $\boxed{\textit{Ejercicio}}$

3. $\boxed{\textit{Ejercicio}}$

4. Ejercicio

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

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

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

Obsérvese que tanto Σ como 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 cadenas que tienen exactamente una a:

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

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

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

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

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

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

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

Es importante observar que *todo lenguaje finito* $L = \{w_1, w_2, \ldots, w_n\}$ *es regular* ya que L es puede obtener con uniones y concatenaciones:

$$L = \{w_1\} \cup \{w_2\} \cup \cdots \cup \{w_n\},$$

y cada w_i es la concatenación de un número finito de símbolos, $w_i = a_1 a_2 \cdots a_k$; por lo tanto, $\{w_i\} = \{a_1\} \cdot \{a_2\} \cdots \{a_k\}$.

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:

$$(R)(S)$$
$$(R \cup S)$$
$$(R)^*$$

   $(R)(S)$ 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. Los paréntesis (y) son símbolos de agrupación y se pueden omitir si no hay peligro de ambigüedad.

Para una expresión regular R cualquiera se utiliza en ocasiones la siguiente notación:

$$L(R) := \text{lenguaje representado por } R.$$

Utilizando esta notación y la definición de expresión regular podemos escribir, para R y S expresiones regulares arbitrarias:

$$\begin{aligned}
&L(\varnothing) = \varnothing. \\
&L(\lambda) = \{\lambda\}. \\
&L(a) = \{a\}, \quad a \in \Sigma. \\
&L(RS) = L(R)L(S). \\
&L(R \cup S) = L(R) \cup L(S). \\
&L(R^*) = L(R)^*.
\end{aligned}$$

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

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

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

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

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

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

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

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

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

✎ 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 $(a \cup b)^*$ y $(a^*b^*)^*$ representan el mismo lenguaje por la igualdad establecida en el ejercicio final 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 cadenas que comienzan con b y terminan con a.
   Solución: $b(a \cup b)^*a$.

2. Lenguaje de todas las cadenas que tienen un número par de símbolos (cadenas de longitud par).
   Solución: $(aa \cup ab \cup ba \cup bb)^*$.

3. Lenguaje de todas las cadenas que tienen un número par de $a$es.
   Soluciones:

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

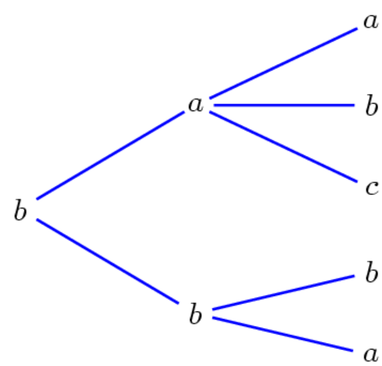
Ejemplos Encontrar expresiones regulares que representen los siguientes lenguajes, definidos sobre el alfabeto $\Sigma = \{0, 1\}$.

1. Lenguaje de todas las cadenas que tienen exactamente dos ceros.
   Solución: $1^*01^*01^*$.

2. Lenguaje de todas las cadenas cuyo penúltimo símbolo, de izquierda a derecha, es un 0.
   Solución: $(0 \cup 1)^*0(0 \cup 1)$.

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

Solución: Una b puede estar seguida solamente de otra b o de una a, mientras que las $a$es y las $c$es pueden estar seguidas de cualquier símbolo. Esto se puede visualizar por medio del siguiente diagrama:



Teniendo en cuenta todas las restricciones y posibilidades, arribamos a la siguiente expresión: $(a \cup c \cup b^+ a)^* b^*$. Una expresión regular más sencilla para este lenguaje es $c^*(b \cup ac^*)^*$.

Ejercicios Encontrar expresiones regulares para los lenguajes descritos a continuación:

1. $\Sigma = \{0, 1, 2\}$. Lenguaje de todas las cadenas que comienzan con 2 y terminan con 1.
2. $\Sigma = \{a, b, c\}$. Lenguaje de todas las cadenas que tienen un número par de símbolos.
3. $\Sigma = \{a, b\}$. Lenguaje de todas las cadenas que tienen un número impar de símbolos.
4. $\Sigma = \{a, b, c\}$. Lenguaje de todas las cadenas que tienen un número impar de símbolos.
5. $\Sigma = \{a, b\}$. Lenguaje de todas las cadenas que tienen un número impar de $a$es.
6. $\Sigma = \{0, 1\}$. Lenguaje de todas las cadenas que tienen por lo menos un 0 y por lo menos un 1.
7. $\Sigma = \{0, 1\}$. Lenguaje de todas las cadenas que tienen a lo sumo dos ceros consecutivos.
8. $\Sigma = \{0, 1\}$. Lenguaje de todas las cadenas cuyo quinto símbolo, de izquierda a derecha, es un 1.
9. $\Sigma = \{0, 1\}$. Lenguaje de todas las cadenas de longitud par ≥ 2 formadas por ceros y unos alternados.
10. $\Sigma = \{0, 1\}$. Lenguaje de todas las cadenas cuya longitud es ≥ 4.

11. $\Sigma = \{0, 1\}$. Lenguaje de todas las cadenas de longitud impar que tienen unos únicamente en las posiciones impares.

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

13. $\Sigma = \{a, b\}$. Lenguaje de todas las cadenas que tienen un número par de $a$es o un número impar de $b$es.

14. $\Sigma = \{0, 1\}$. Lenguaje de todas las cadenas cuya longitud es un múltiplo de tres.

15. $\Sigma = \{0, 1, 2\}$. Lenguaje de todas las cadenas que no contienen dos unos consecutivos.

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

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

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