Alfabetos, cadenas y lenguajes.

De manera muy amplia podría decirse que la computación es la manipulación de secuencias de símbolos. Pero el número de símbolos disponibles en cualquier mecanismo de cómputo es finito y todos los objetos usados como entradas o salidas (inputs/outputs) deben ser identificados en un tiempo finito. Desde el punto de vista teórico esto impone dos restricciones básicas: el conjunto de símbolos (alfabeto) debe ser finito y se deben considerar únicamente cadenas (secuencias de símbolos) de longitud finita. Surgen así los ingredientes esenciales de una teoría abstracta de la computación: ***alfabetos* y *cadenas***. Los conjuntos de cadenas (ya sean finitos o infinitos) se denominarán ***lenguajes***.

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 (o secuencia) 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 $\boldsymbol{\lambda}$. La cadena vacía desempeña, en la teoría de la computación, un papel similar al del conjunto vacío **ø** 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

$ababaa$

$aaaab$

Observe 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 $\Sigma = \{O, 1\}$ 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.

$\Sigma = \{a,b,c,\ldots,x,y,z,A,E,C,\ldots,X,Y,Z\}$, el alfabeto del idioma español. Las palabras oficiales del español (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.

Cerradura de Kleene

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

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

La siguiente tabla presenta la notación mas comúnmente utilizada en la teoría de la computación. De ser necesario, se pueden emplear subíndices.

Notación usada en la teoría de la computación	
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).

NOTAS IMPORTANTES:

- Algunos autores denotan la cadena vacía con la letra griega **ε.** En la medida de lo posible utilizaremos **λ** para evitar confusión con el símbolo **∈** 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 $\Sigma^*=\{\lambda, a, aa, aaa, aaaa, aaaaa, \ldots\}$
- La mayor parte de la teoría de la computación se hace con referencia a un alfabeto Σ fijo (pero arbitrario)
- Algunas características principales del uso con cadenas es que pueden: Concatenarse, Invertirse, obtener su longitud $|w|$.

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* $u = a1a2 \cdots an$, $v = b1b2 \cdots bm$, *entonces*

$$u \cdot v = a1a2 \cdots anb1b2 \cdots bm$$

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 $u, v \in \Sigma^*$, $a \in$ E, 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*

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

Potencias de una cadena

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

$u^0 = \lambda$,

$u^n = uu \cdots u.$ n veces

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,

$$|u| = \begin{cases} 0, & \text{si } u = \lambda \\ n & si \quad u = a_1 a_2 \cdots a_n. \end{cases}$$

Ejemplo

$|aba| = 3, \qquad |baaa| = 4.$

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:

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

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

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

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.

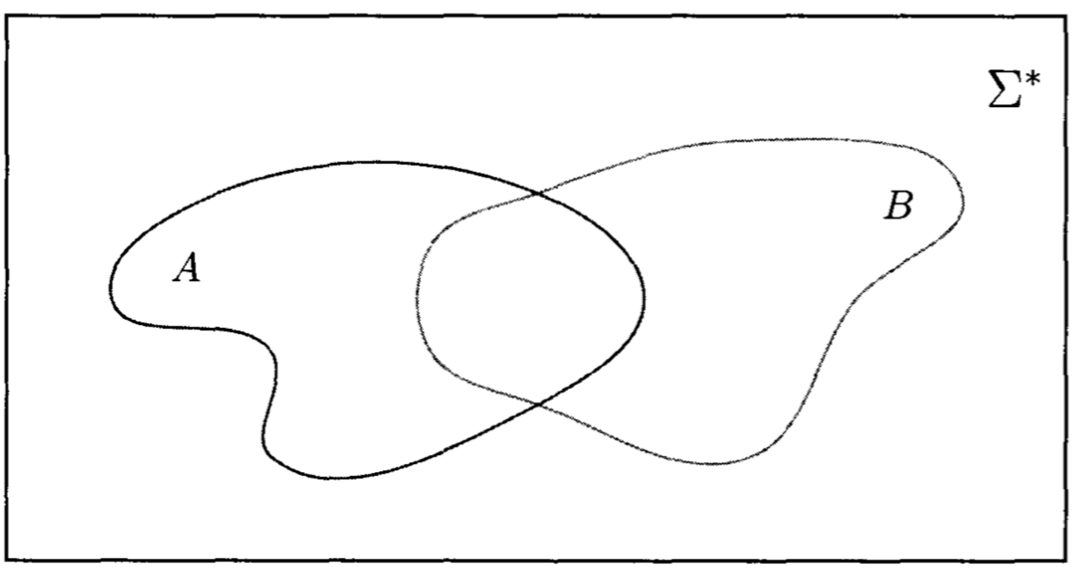
Todo lenguaje L satisface $\varnothing \subseteq L \subseteq \Sigma^*$, y puede ser finito o infinito. Los lenguajes se denotan con letras mayúsculas $A, B, C, \ldots, L, M, N, \ldots$. En la siguiente gráfica se visualizan dos lenguajes A y B sobre Σ.

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.



Ejemplos de lenguajes sobre los alfabetos especificados

- $\Sigma = \{a, b, c\}$. $L = \{a, aba, aca\}$.
- $\Sigma = \{a, b, c\}$. $L = \{a, aa, aaa, \ldots\} = \{a^n : n \geq 1\}$.
- $\Sigma = \{a, b, c\}$. $L = \{\lambda, aa, aba, ab^2a, ab^3a, \ldots\} = \{ab^na : n \geq 0\} \cup \{\lambda\}$.
- $\Sigma = \{a, b, c, \ldots, x, y, z, A, B, C, \ldots, X, Y, Z\}$. $L = \{u \in \Sigma^* : u$ 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:

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