# 2 Autómatas

Empezaremos por introducir las definiciones más importantes para poder desarrollar la *teoría de autómatas*. En particular definiremos *alfabeto*, *palabra* y *lenguaje*, y veremos algunas operaciones dentro de los lenguajes.

Un *alfabeto* es un conjunto de símbolos finito y no vacío. Usaremos la letra $\Sigma$ para denotar a un alfabeto. Algunos ejemplos de alfabetos son los siguientes:

- $\Sigma = \{0, 1\}$ el alfabeto *binario*.
- $\Sigma = \{a, , b, c, \ldots, z\}$ el abc-dario con todas las letras minúsculas.
- El alfabeto de todos los caracteres ASCII (o el conjunto de todos los 126 caracteres ASCII imprimibles).

Una *palabra* (también llamada *cadena*) es una secuencia finita de símbolos de algún alfabeto. Por ejemplo 01101 es una palabra del alfabeto binario, y *holaaaa* es una palabra del abc-dario con todas las letras minúsculas.

Una palabra importante es *la palabra vacía* que se define como la palabra con cero ocurrencia de un símbolo, es decir, la palabra que no tiene ningún sı́mbolo. La palabra vacía se denota como $\epsilon$ y puede ser construida con símbolos (o mejor dicho sin ellos) de cualquier alfabeto.

Una propiedad que nos permite clasificar palabras es su *largo*. El *largo* de una palabra se define como el número de símbolos que ésta tiene, contando repeticiones. Por ejemplo, la palabra 01101 tiene largo 5. Es importante remarcar que para medir el largo de una palabra *se deben contar las repeticiones de cada símbolo*. Por ejemplo, si contáramos sólo el número de símbolos de una palabra, como la palabra 01101 tiene sólo dos símbolos, tendría largo dos, lo que es **incorrecto** porque ya vimos que su largo es 5. Si usamos $x$ para denotar una palabra, entonces el largo de $x$ se denota $|x|$. Por ejemplo $|011| = 3$. El largo de la palabra vacía es cero $|\epsilon| = 0$.

Si $\Sigma$ es un alfabeto, podemos denotar a todas las palabras con símbolos en $\Sigma$ de un cierto largo usando la notación de potencia. En particular, denotamos $\Sigma^k$ al conjunto de todas las palabras de largo $k$ definidas sobre el alfabeto $\Sigma$.

Ejemplos de potencia de alfabetos son los siguientes:

- $\Sigma^0 = \{\epsilon\}$, sin importar cual es el alfabeto $\Sigma$. Es decir, $\epsilon$ es la única palabra de largo 0.
- Si $\Sigma = \{0, 1\}$ entonces $\Sigma^1 = \{0, 1\}$. En este caso hay que poner atención ya que $\Sigma$ es un conjunto de símbolos, mientras que $\Sigma^1$ es un conjunto de palabras. Por lo tanto los símbolos 0 y 1 juegan un doble rol de palabra y símbolo.
- Si $\Sigma = \{0, 1\}$ entonces $\Sigma^2 = \{00, 01, 10, 11\}$.
- Si $\Sigma = \{0, 1\}$ entonces $\Sigma^3 = \{000, 001, 010, 011, 100, 101, 110, 111\}$.

El conjunto de todas las palabras sobre un alfabeto $\Sigma$ es denotado $\Sigma^*$. Por ejemplo: $\{0,1\}^* = \{\epsilon, 0, 1, 00, 01, 10, 10, 000, \ldots\}$. De forma equivalente, también podemos decir:

$$\Sigma^* = \Sigma^0 \cup \Sigma^1 \cup \Sigma^2 \cup \Sigma^3 \cup \cdots .$$

Algunas veces vamos a necesitar omitir la palabra vacía de $\Sigma^*$, para este conjunto usamos la notación $\Sigma^+$ que se define como sigue:

$$\Sigma^+ = \Sigma^1 \cup \Sigma^2 \cup \Sigma^3 \cup \cdots ,$$

de esta forma también tenemos que $\Sigma^* = \Sigma^+ \cup \{\epsilon\}$.

Si $x$ e $y$ son dos palabras, entonces definimos a la *concatenación* de $x$ e $y$, que denotamos $xy$, como la palabra formada por una copia de $x$ seguida de una copia de $y$. Equivalentemente, si $x$ es una palabra compuesta de $i$ símbolos $x = a_1a_2 \ldots a_i$ e $y$ es una palabra compuesta de $j$ símbolos $y = b_1b_2 \ldots b_j$, entonces $xy = a_1a_2 \ldots a_ib_1b_2 \ldots b_j$.

Por ejemplo, si $x = 01101$ e $y = 110$, entonces $xy = 01101110$ e $yx = 11001101$. Para cualquier palabra $w$ la composición con la palabra vacía no hace nada, es decir $w\epsilon = \epsilon w = w$.

**Definición 2.1.** Un *lenguaje* es un conjunto de palabras $L$ escogido sobre $\Sigma^*$, donde $\Sigma$ es algún alfabeto particular. Si $\Sigma$ es un alfabeto y $L \subseteq \Sigma^*$, entonces diremos que $L$ es un lenguaje sobre $\Sigma$.

Vale la pena mencionar que un lenguaje sobre $\Sigma$ no necesariamente incluye palabras de tal forma que todos los símbolos de $\Sigma$ sean usados, por lo tanto si $L$ es un lenguaje sobre $\Sigma$ también lo es sobre cualquier alfabeto que contenga a $\Sigma$.

Ejemplos de lenguajes:

- El lenguaje de todas las palabras en $\{0,1\}$ que consistan de $n$ ceros seguidos de $n$ unos. Es decir:

$$n \geq 0 : \{\epsilon, 01, 0011, 000111, 00001111, \ldots\}.$$

- El lenguajes de todas las palabras en $\{0,1\}$ que tengan el mismo número de ceros y de unos. Es decir:

$$\{\epsilon, 01, 10, 0011, 0101, 1001, \ldots\}.$$

- El lenguaje que contiene a todos los números primos escritos en binario:

$$\{10, 11, 101, 111, 1011 \ldots\}.$$

- $\Sigma^*$ es un lenguaje para cualquier alfabeto $\Sigma$.
- El lenguaje vacío, $\emptyset$, es un lenguaje sobre cualquier alfabeto.

- $L = \{\epsilon\}$, el lenguaje que contiene solamente a la palabra de largo cero es un lenguaje sobre cualquier alfabeto. Notar que $\{\epsilon\} \neq \emptyset$, ya que el primero contiene un elemento mientras que el segundo no contiene ninguno.

En *Teoría de Autómatas* nos interesaremos en un tipo de problemas.

**Definición 2.2.** Un *problema* consiste en decidir si dado una palabra $w$ y un lenguaje $L$:

$$\text{¿es cierto que } w \in L?$$

Es verdad que en un principio este tipo de problemas puede parecer restrictivo, sin embargo veremos que casi todos los problemas que conocemos se pueden expresar de esta manera.

Un ejemplo es el problema de saber si un número es primo o no. Si consideramos $L_p$ como el lenguaje que contiene a todos los números primos en binario y $w$ un número en binario que queremos saber si es primo, entonces el problema es decidir si $w \in L_p$.

A lo largo de esta parte de la asignatura veremos que existen problemas que no podemos responder y otros que sí. Nos centraremos en el estudio de un subconjunto de aquellos que sí podemos responder, llamados *Lenguajes Regulares*.

Los lenguajes son conjuntos de palabras, por lo tanto heredan la mayoría de las definiciones y operaciones de conjuntos. En particular, tenemos las siguientes definiciones.

**Definición 2.3.**
- Dado un lenguaje $L$ sobre $\Sigma$, el *complemento* de $L$, denotado $\bar{L}$, se define como:
$$\bar{L} := \Sigma^* \setminus L.$$
- Dados dos lenguajes $L_1$ y $L_2$ sobre $\Sigma$, la *concatenación* de ambos se define como:
$$L_1L_2 := \{xy : x \in L_1 \wedge y \in L_2\}.$$
- Dado un lenguaje $L$ sobre $\Sigma$, las potencias de $L$, denotadas $L^i$, se definen como:
$$L^0 := \{\epsilon\}, \quad L^{i+1} = L^iL.$$
- Dado un lenguaje $L$, definimos la *Estrella de Klein* de $L$, denotada $L^*$, como:
$$L^* := \bigcup_{n \in \mathbb{N} \cup \{0\}} L^n.$$
En este punto vale la pena mencionar que la estrella de Klein del lenguaje vacío es la palabra de largo cero: $\emptyset^* = \{\epsilon\}$.

Con estas definiciones podemos observar una serie de propiedades para las operaciones de lenguajes. Dados $L_1, L_2, L_3$ sobre $\Sigma$, se cumple que:

- $L_1\{\epsilon\} = \{\epsilon\}L_1 = L_1$.
- $L_1(L_2 \cup L_3) = L_1L_2 \cup L_1L_3$.

- $(L_1 \cup L_2)L_3 = L_1L_3 \cup L_2L_3$.
- $L_1(L_2 \cap L_3) \subseteq L_1L_2 \cap L_1L_3$.
- $(L_1 \cap L_2)L_3 \subseteq L_1L_3 \cap L_2L_3$.
- $L_1(L_2L_3) = (L_1L_2)L_3$.

Queda como ejercicio demostrar la veracidad de todas estas propiedades usando las definiciones antes dadas.