

Capítulo 1

Conceptos previos

A lo largo del texto, se verán los lenguajes como conjuntos de cadenas que acepta una determinada máquina teórica. Por ello, se hará uso de conceptos básicos de teoría de conjuntos. Para ayudar al lector a recordar dichos conceptos, se ha introducido en este capítulo alguno de los conceptos más relevantes que se utilizarán a lo largo del texto. Así mismo, y a modo de glosario, se incluye una presentación de los conceptos fundamentales de la materia de autómatas, gramáticas y lenguajes.

1.1. Conceptos fundamentales de la teoría de conjuntos

■ Conjunto

Un conjunto es una colección de objetos que se denominan elementos o miembros. Normalmente, los elementos de un conjunto se representan separados por comas y encerrados entre corchetes ($\{$ y $\}$).

Se dice que un conjunto es finito cuando contiene un número finito de elementos. Algunos ejemplos de conjuntos finitos son:

- El conjunto de letras vocales del alfabeto: $V = \{a, e, i, o, u\}$.
- El conjunto de días de la semana: $S = \{\text{lunes, martes, miércoles, jueves, viernes}\}$.
- El conjunto formado por los números menores de 100:
 $N = \{1, 2, 3, \dots, 97, 98, 99, 100\}$.

Se dice que un conjunto es infinito cuando contiene un número infinito de elementos. Algunos ejemplos de conjuntos infinitos son:

- El conjunto de números naturales: $N = \{1, 2, 3, \dots\}$.
- El conjunto de números primos: $P = \{1, 3, 5, 7, \dots\}$.
- El conjunto formado por los números mayores de 100:
 $M = \{101, 102, 103, 104, \dots\}$.

La cardinalidad de un conjunto determina el tamaño de dicho conjunto, esto es, el número de elementos que contiene. Se representa mediante dos líneas verticales, así, si se considera el conjunto finito $V = \{a, e, i, o, u\}$, entonces su cardinalidad se representa de la siguiente manera: $|V| = 5$.

■ Conjunto vacío

El conjunto vacío es aquel que no contiene ningún elemento. El conjunto vacío se representa mediante el siguiente símbolo: \emptyset .

■ Pertenencia

Se dice que un elemento pertenece a un conjunto cuando cumple las condiciones que lo definen. El operador de pertenencia se representa mediante el símbolo \in y en el ejemplo del conjunto $V = \{a, e, i, o, u\}$, se cumple que $a \in V$.

■ Subconjunto

Se dice que A es un subconjunto de B ($A \subseteq B$), si todos los elementos A pertenecen también al conjunto B . Esto es, para todo elemento $w \in A$ se cumple que $w \in B$. El conjunto A es un **subconjunto propio** de B si $A \subseteq B$ y existen elementos de B que no pertenecen a A . En este caso se representa la relación entre los dos conjuntos de la siguiente forma: $A \subset B$. Por ejemplo, dados los conjuntos:

$$A = \{a, b, c, d\} \quad B = \{c, d\} \quad C = \{c, d\} \quad D = \{e, f\}$$

B es un subconjunto propio de A : $B \subset A$.

■ Igualdad y desigualdad entre conjuntos

Dos conjuntos A y B son iguales ($A = B$) si se cumple que $A \subseteq B$ y $B \subseteq A$.

Dos conjuntos A y B , son diferentes ($A \neq B$) si existen elementos de A que no pertenecen a B y viceversa. Esto es, existen $w \in A$ tales que $w \notin B$ y existen $z \in B$ tales que $z \notin A$.

Así, siguiendo el ejemplo anterior, se cumple que: $B = C$ y $A \neq D$.

■ **Conjunto potencia**

Dado un conjunto A , el conjunto potencia $P(A)$ es la colección de todos los subconjuntos que se pueden formar con los elementos de A . Por tanto, los elementos del conjunto potencia son a su vez conjuntos.

Por ejemplo, considerando el conjunto $A = \{a, b, c\}$ entonces el conjunto potencia de A es: $P(A) = \{\{a\}, \{b\}, \{c\}, \{a, b\}, \{a, c\}, \{b, c\}, \{a, b, c\}\}$.

Dado un conjunto A , la cardinalidad de su conjunto potencia se define: $P(A) = 2^{|A|}$, donde $|A|$ es la cardinalidad del conjunto A . El conjunto potencia de un conjunto infinito es un conjunto **incontable**.

■ **Operaciones entre conjuntos**

Las principales operaciones entre conjuntos que se van a considerar son:

– **Unión**

La unión de dos conjuntos A y B , representada por $A \cup B$, es la colección de todos los elementos que se encuentran en A o en B :

$$A \cup B = \{x : x \in A \text{ o } x \in B\}$$

Por ejemplo, dados los conjuntos:

$$A = \{a, b, c, d\} \quad B = \{c, d\} \quad C = \{c, d\} \quad D = \{e, f\}$$

entonces:

$$\begin{aligned} A \cup B &= \{a, b, c, d\} \\ B \cup C &= \{c, d\} = B = C \\ C \cup D &= \{c, d, e, f\} \end{aligned}$$

– **Intersección**

La intersección de dos conjuntos A y B , representada por $A \cap B$, es la colección de objetos que son elementos tanto de A como de B . Por consiguiente:

$$A \cap B = \{x : x \in A \text{ y } x \in B\}$$

Por ejemplo, dados los conjuntos $A = \{a, b, c\}$ y $B = \{b, c, d\}$, entonces:

$$A \cap B = \{b, c\}$$

– **Diferencia**

La diferencia entre dos conjuntos A y B se representa mediante el signo “–”, y es el conjunto que resulta de eliminar del conjunto A los elementos del conjunto B . Por ejemplo, dados los conjuntos $A = \{a, b, c\}$, $B = \{a, c\}$, $C = \{a, b\}$ y $D = \{d, e\}$, entonces:

$$\begin{aligned} A - B &= \{b\} \\ C - D &= \{a, b\} \end{aligned}$$

El conjunto $A - B$ se denomina conjunto **complemento** de B con respecto a A . En ocasiones, se da por sentado que los elementos de todos los conjuntos considerados, pertenecen a un conjunto universal de mayor tamaño. En estos casos, el complemento de un conjunto X con relación a este conjunto universal, recibe el nombre de complemento de X . Por ejemplo, si se considera el conjunto universal como el conjunto de números naturales, entonces el complemento del conjunto de números naturales pares, será el conjunto formado por todos los números naturales impares.

– Producto cartesiano

El producto cartesiano de dos conjuntos A y B , representado por $A \times B$, es el conjunto de todos los pares ordenados de la forma (a, b) , donde $a \in A$ y $b \in B$. Por lo general, $A \times B \neq B \times A$. Por ejemplo, dados los conjuntos:

$$A = \{a, b, c\} \quad B = \{1, 2\}$$

entonces:

$$A \times B = \{(a, 1), (a, 2), (b, 1), (b, 2), (c, 1), (c, 2)\}$$

Es posible generalizar el concepto del producto de conjuntos para obtener el producto de más de dos conjuntos. Así, dados los conjuntos: $A = \{a, b\}$, $B = \{1, 2\}$ y $C = \{x, y\}$, entonces:

$$A \times B \times C = \{(a, 1, x), (a, 1, y), (a, 2, x), (a, 2, y), (b, 1, x), (b, 1, y), (b, 2, x), (b, 2, y)\}$$

1.2. Conceptos fundamentales de la teoría de autómatas

A lo largo de este texto e independientemente del tipo de máquina que se considere, se van a utilizar una serie de conceptos comunes a todas las máquinas. Con el fin de servir de guía de consulta, se recogen en esta sección un listado de términos comunes y su definición.

Alfabeto

Un *alfabeto* es un conjunto de símbolos finito no vacío. Normalmente se utiliza el símbolo Σ para denotar un alfabeto. A continuación se muestran algunos ejemplos:

- $\Sigma = \{a, b\}$ un alfabeto compuesto por dos símbolos.

- $\Sigma = \{0, 1, 2\}$ un alfabeto compuesto por tres símbolos.
- $\Sigma = \{a, b, c, d\}$ un alfabeto compuesto por cuatro símbolos.

Cadena de caracteres

Una cadena de caracteres es una secuencia finita de símbolos de un alfabeto. Por ejemplo, *aba* es una cadena del alfabeto $\Sigma = \{a, b, c\}$.

Una cadena especial a considerar es la *cadena vacía*. La cadena vacía es aquella cadena que no contiene ningún símbolo. Esta cadena se representa con el símbolo ϵ o con el símbolo λ y se puede definir con cualquier alfabeto.

La longitud de una cadena es igual al número de símbolos que contiene la cadena. Así, la cadena 001 tiene una longitud igual a 3.

Para indicar la longitud de una cadena w se utiliza la notación $|w|$. Así, por ejemplo, $|001| = 3$ mientras que $|\epsilon| = 0$.

Concatenación de cadenas

La concatenación de cadenas es una operación que permite construir nuevas cadenas a partir de otras más simples. Así, dado el alfabeto $\Sigma = \{a, b\}$, la concatenación de los símbolos a y b da como resultado la cadena ab .

De igual forma, la concatenación de las cadenas abb y bb da como resultado la cadena $abbbb$.

Potencias de un alfabeto

Intuitivamente, esta operación es similar a la potencia de números. Así, por ejemplo, para calcular la potencia 2 del número 2 hay que multiplicar dos veces el número 2, mientras que para calcular la potencia 3 del número 2 hay que multiplicar 3 veces el número 2.

Cuando se trata de símbolos, la operación similar a la multiplicación es la concatenación. De igual manera, se puede considerar la potencia de un alfabeto. Así, dado un alfabeto Σ , entonces se define Σ^k como el conjunto de las cadenas de longitud k , tales que cada uno de los símbolos de las cadenas pertenece a Σ .

Algunas definiciones importantes:

- Σ^* es el conjunto de todas las cadenas que se pueden formar con los símbolos de un alfabeto.
- Σ^+ es el conjunto de todas las cadenas que se pueden formar con los símbolos de un alfabeto excepto la cadena vacía.

- La cadena vacía ϵ es el elemento neutro de la concatenación. Esto es, para cualquier alfabeto Σ , la cadena vacía concatenada con cualquier cadena perteneciente a Σ^* da como resultado, esa misma cadena.

Lenguaje

Dado un alfabeto Σ , un lenguaje L , es un subconjunto de cadenas que se pueden formar con los símbolos del alfabeto. Por tanto:

- Σ^* es un lenguaje, y está compuesto por todas las cadenas que se forman con los símbolos del alfabeto.
- Cualquier lenguaje L es un subconjunto de Σ^* , esto es, $L \subseteq \Sigma^*$.
- Se dice que un lenguaje es finito, cuando contiene un número finito de cadenas. Los lenguajes $L = \{a, aa, aaa, aaaa\}$ y $M = \{b, ab\}$ son dos ejemplos de lenguajes finitos.
- Se dice que un lenguaje es infinito, cuando contiene un número infinito de cadenas. Por ejemplo, el lenguaje formado por todas las cadenas con un número par de símbolos, es un lenguaje infinito.
- El lenguaje vacío es aquel que no contiene ninguna cadena. Se representa mediante el símbolo \emptyset . El lenguaje vacío es un lenguaje que se puede definir para cualquier alfabeto.
- El lenguaje formado únicamente por la cadena vacía ($L = \{\epsilon\}$). Es un lenguaje que se puede definir para cualquier alfabeto. No obstante, es importante insistir que $\emptyset \neq \{\epsilon\}$ ya que el primero no contiene ninguna cadena y el segundo contiene una única cadena que es la cadena vacía.
- Un lenguaje finito se puede describir enumerando cada una de sus cadenas. No obstante, esto no es posible cuando el lenguaje es infinito. En este caso, es útil utilizar la definición de lenguajes mediante descripciones de conjuntos. Así por ejemplo, dado el alfabeto $\Sigma = \{a, b\}$, se pueden definir los siguientes lenguajes:

- El lenguaje formado por todas las cadenas que tienen el mismo número de a 's y b 's:

$$L = \{w \mid w \text{ contiene el mismo número de } a\text{'s y } b\text{'s}\}.$$

- El lenguaje formado por todas las cadenas que empiezan por el símbolo a :

$$L = \{ax : x \in \{a, b\}^*\}$$

donde la expresión $x \in \{a, b\}^*$ representa que la subcadena x puede contener cualquier combinación de símbolos a y b .

- El lenguaje formado por todas las cadenas que terminan por el símbolo b :

$$L = \{xb : x \in \{a, b\}^*\}$$

donde, de nuevo, la expresión $x \in \{a, b\}^*$ representa que la subcadena x puede contener cualquier combinación de símbolos a y b .

Operaciones entre lenguajes

Considerando los lenguajes como conjunto de cadenas, es claro que se pueden definir para los lenguajes las siguientes operaciones:

- **Unión:** La unión de dos lenguajes L y M , designada como $L \cup M$, es el conjunto de cadenas que pertenecen a L , a M o a ambos.
- **Concatenación:** La concatenación de los lenguajes L y M es el conjunto de cadenas que se puede formar tomando cualquier cadena de L y concatenándola con cualquier cadena de M . Esta operación se puede denotar con un símbolo “.” ($L \cdot M$) o simplemente sin operador (LM).
- **Complementario:** El complementario de un lenguaje L , se representa de la siguiente manera: \bar{L} , y es el conjunto de cadenas que se pueden formar con el alfabeto Σ , y que no están contenidas en L . Por tanto, $\bar{L} = \Sigma^* - L$.
- **Clausura de Kleene o clausura:** La clausura de un lenguaje, representa el conjunto de cadenas que se pueden formar concatenando cualquier número de veces las cadenas del lenguaje. Para ver un ejemplo sencillo de cómo se realiza la clausura de Kleene de un lenguaje, se considera el lenguaje formado por una única cadena $L = \{a\}$. En este caso, la clausura de Kleene contendría las cadenas que se forman al concatenar cero o más veces el símbolo a . Así, $L^* = \{\epsilon, a, aa, aaa, aaaa, aaaaa, \dots\}$.

Como caso más general, sea ahora el lenguaje $L = \{0, 1, 01, 10\}$, entonces:

- L^0 es el resultado de concatenar cero veces las cadenas de un lenguaje únicamente da como resultado la cadena vacía. Así, $L^0 = \{\epsilon\}$.
- L^1 es el resultado de concatenar una vez las cadenas del lenguaje, por tanto, el resultado son las mismas cadenas del lenguaje original. Así, $L^1 = L = \{0, 1, 01, 10\}$.

- L^2 es el resultado de concatenar dos veces las cadenas del lenguaje. En la siguiente tabla se muestra cómo se obtiene cada una de las cadenas del lenguaje L^2 . Es importante recordar que la concatenación no es una operación conmutativa. En la tabla se considera que se concatena la cadena que etiqueta la fila con la cadena que etiqueta la columna:

	0	1	01	10
0	00	01	001	010
1	10	11	101	110
01	010	011	0101	0110
10	100	101	1001	1010

Así:

$$L^2 = \{00, 01, 001, 010, 10, 11, 101, 110, 010, 011, 0101, 0110, 100, 101, 1001, 1010\}.$$

- L^3 es el resultado de concatenar tres veces las cadenas del lenguaje. Se puede obtener concatenando las cadenas de L con las cadenas de L^2 vistas en el punto anterior. Así:

$$L^3 = \{000, 001, 0001, 0010, 010, 011, 0101, 0110, 0010, 0011, 00101, 00110, 0100, 0101, 01001, 01010, 100, 101, 1001, 1010, 110, 111, 1101, 1110, 1010, 1011, 10101, 10110, 1100, 1101, 11001, 11010, 0100, 0101, 01001, 01010, 0110, 0111, 01101, 01110, 01010, 01011, 010101, 010110, 01100, 01101, 011001, 011010, 1000, 1001, 10001, 10010, 1010, 1011, 10101, 10110, 10010, 10011, 100101, 100110, 10100, 10101, 101001, 101010\}.$$

Puesto que la clausura de Kleene es el resultado de concatenar cero o más veces las cadenas del lenguaje se puede definir L^* como una unión de lenguajes de la siguiente manera:

$$L^* = L^0 \cup L^1 \cup L^2 \cup L^3 \cup \dots$$

Por tanto, si $L = \{0, 1, 01, 10\}$ entonces $L^* = \{\epsilon, 0, 1, 01, 10, 010, 100, \dots\}$.

La estrella de Kleene de cualquier lenguaje cumple las siguientes dos propiedades:

1. Para cualquier lenguaje L , se cumple que $\epsilon \in L^*$.
2. Para cualquier lenguaje L que cumpla que $L \neq \{\epsilon\}$ y $L \neq \emptyset$, L^* es un lenguaje que contiene un número infinito de cadenas.

Además:

1. Si $L = \{\epsilon\}$ entonces $L^* = \{\epsilon\}$.
2. Si $L = \emptyset$ entonces $L^* = \{\epsilon\}$.

Problemas

En teoría de autómatas, un *problema* es la cuestión de decidir si una determinada cadena pertenece a un determinado lenguaje [Hopcroft *et al.*, 2008]. Esto es, dado un alfabeto Σ y un lenguaje L sobre este alfabeto, entonces dada una cadena w de Σ^* , el problema es decidir si w pertenece o no a L .

1.3. Ejercicios

- (1) Dado el alfabeto $\Sigma = \{0, 1\}$, sea L el lenguaje formado por todas las cadenas con un número par de símbolos. Dar cinco ejemplos de cadenas pertenecientes al lenguaje L y cinco cadenas que no pertenezcan al lenguaje.
- (2) Dado el alfabeto $\Sigma = \{0, 1\}$, sea L el lenguaje definido de la siguiente forma:

$$L = \{w \mid n_0(w) = n_1(w)\}$$

donde $n_0(x)$ es el número de 0's de la cadena w y $n_1(w)$ es el número de 1's de la cadena w . Dar cinco ejemplos de cadenas que pertenezcan al lenguaje y cinco ejemplos de cadenas que no pertenezcan al lenguaje.

- (3) Dado el alfabeto $\Sigma = \{0, 1\}$, sea L el lenguaje definido de la siguiente forma:

$$M = \{0^n 1^n : n \geq 1\}$$

Dar cinco ejemplos de cadenas que pertenezcan al lenguaje y cinco ejemplos de cadenas que no pertenezcan al lenguaje. Indicar las diferencias con el lenguaje del ejercicio anterior y dar un ejemplo de cadena que pertenezca al lenguaje L y no pertenezca a M . Indicar si es posible encontrar una cadena que esté en M y no en L .

- (4) Dado un alfabeto Σ , indique si es verdadera o falsa la siguiente afirmación: $P(\Sigma) = \Sigma^*$.