Autómatas y Lenguajes Formales
Nota 02. Conceptos y definiciones básicas*

Noé Salomón Hernández S.

1. Problemas de decisión

Un problema de decisión es una función con una salida binaria: 'si' o 'no'. Para especificar un problema de decisión, debemos indicar:

- el conjunto A de posibles entradas, y
- el subconjunto $B \subseteq A$ de instancias aceptadas.

Por ejemplo, para decidir si un número dado es primo, el conjunto de posibles entradas es el conjunto de todas las codificaciones de números enteros, y las instancias aceptadas con un 'si' son número primos.

En este curso vamos a considerar en su mayoría problemas de decisión. Realizamos esto por simplicidad matemática y porque el comportamiento que deseamos estudiar está presente a este nivel.

2. Cadenas

Siempre tomaremos el conjunto de posibles entradas a un problema de decisión como un conjunto de cadenas de longitud finita sobre un alfabeto. Elegimos este proceder por simplicidad y uniformidad. Otros tipos de datos –gráficas, números naturales, programas– pueden ser codificados como cadenas. Al realizar esta abstracción, tenemos que lidiar únicamente con un tipo de datos y unas cuantas operaciones básicas.

Definición 2.1

- Un *alfabeto* es cualquier conjunto finito. Por ejemplo, podríamos usar el alfabeto $\{0, 1, \ldots, 9\}$ si queremos hablar acerca de número decimales; el conjunto de todos los caracteres ASCII; $\{0, 1\}$ si hablamos de cadenas binarias. Usualmente denotamos a un alfabeto finito arbitrario por Σ. Llamamos a los elementos de Σ *letras* o *símbolos* y los denotamos como $a, b, c, \ldots$
- Una *cadena* sobre Σ es una secuencia de longitud finita de elementos de Σ. Por ejemplo: si $\Sigma = \{a, b\}$, entonces $aabababa$ es una cadena sobre Σ. Usamos $w, x, y, z, \ldots$ para referirnos a cadenas.

*Esta nota se basa en el libro: D. C. Kozen. *Automata and Computability*, Springer-Verlag, Inc.

- La *longitud* de una cadena x es el número de elementos en x. La longitud de x se denota $|x|$. Por ejemplo, $|aabababa| = 8$.

- Hay una única cadena de longitud cero sobre Σ llamada la *cadena vacía*, denotada por ε. Así, $|\varepsilon| = 0$.

- Escribimos a^n como la cadena de a's de longitud n. Por ejemplo, $a^4 = aaaa$. Formalmente, a^n se define inductivamente como:

$$\begin{aligned} a^0 &= \varepsilon, \\ a^{n+1} &= a^n a. \end{aligned}$$

- El conjunto de todas las cadenas sobre el alfabeto Σ se denota Σ^*. Por ejemplo:

$$\{a, b\}^* = \{\varepsilon, a, b, aa, ab, ba, bb, aaa, aab, \ldots\}$$

$$\begin{aligned} \{a\}^* &= \{\varepsilon, a, aa, aaa, aaaa, \ldots\} \\ &= \{a^n \mid n \geq 0\}. \end{aligned}$$

$\dashv$

Por convención, tomamos

$$\varnothing^* = \{\varepsilon\}$$

Si Σ no es vacío, entonces Σ^* es un conjunto infinito de cadenas de longitud finita. Notemos que $\varnothing$, $\{\varepsilon\}$ y ε son tres objetos distintos. El primero es un conjunto sin elementos; el segundo es un conjunto con un elemento, a saber ε; y el último es una cadena, no un conjunto.

3. Operaciones sobre cadenas

La operación de *concatenación* toma dos cadenas x y y y genera una nueva cadena xy al juntarlas escribiendo primero x y enseguida y. La cadena xy es llamada la *concatenación* de x y y. Observemos que xy y yx son en general diferentes. Algunas propiedades útiles de la concatenación son:

- la concatenación es *asociativa*: $(xy)z = x(yz)$;

- la cadena vacía es una *identidad* para la concatenación: $\varepsilon x = x\varepsilon = x$;

- $|xy| = |x| + |y|$.

Un caso especial de la última ecuación es $a^m a^n = a^{m+n}$, para toda $m, n \geq 0$.

Definición 3.1

- Escribimos x^n para representar la cadena que se obtiene al concatenar n copias de x. Por ejemplo, $(baa)^5 = baabaabaabaabaa$, $(baa)^1 = baa$, y $(baa)^0 = \varepsilon$. Formalmente, x^n de define inductivamente como:

$$\begin{aligned} x^0 &= \varepsilon, \\ x^{n+1} &= x^n x. \end{aligned}$$

- Si $a \in \Sigma$ y $x \in \Sigma^*$, escribimos $n_a(x)$ para denotar el número de a's en x. Por ejemplo, $n_a(abbbabaa) = 4$ y $n_b(aaaaa) = 0$.

- Un *prefijo* de una cadena x es una subcadena inicial de x, es decir, una cadena y para la cual existe una cadena z tal que $x = yz$. Por ejemplo, $abbbaa$ es un prefijo $abbbaabab$. La cadena vacía es un prefijo de toda cadena, y toda cadena es un prefijo de sí misma. Un prefijo y de x es un *prefijo propio* de x si $y \neq \varepsilon$ y $y \neq x$.

$\dashv$

4. Operaciones sobre conjuntos

Usualmente denotamos conjuntos de cadenas (subconjuntos de Σ^*) por $A, B, C, \ldots$. La *cardinalidad* del conjunto A se denota como $|A|$. El conjunto vacío $\varnothing$ es el único conjunto de cardinalidad 0. Definimos ahora algunas operaciones sobre conjuntos que nos serán útiles a lo largo del curso.

- *Concatenación de conjuntos*:

$$AB = \{xy \,|\, x \in A \text{ y } y \in B\}.$$

  En otras palabras, z es una cadena de AB syss z puede ser escrita como la concatenación de dos cadenas x y y, donde $x \in A$ y $y \in B$. Por ejemplo, $\{b, ba\}\{aa, ab\} = \{baa, bab, baaa, baab\}$. Al encontrar la concatenación de conjuntos, se tienen que incluir *todas* las cadenas que pueden ser obtenidas de esta forma. Observemos que AB y BA son en general diferentes conjuntos. Por ejemplo, $\{aa, ab\}\{b, ba\} = \{aab, aaba, abb, abba\}$.

- Las *potencias* A^n de un conjunto A se definen de manera inductiva como sigue:

$$\begin{aligned} A^0 &= \{\varepsilon\}, \\ A^{n+1} &= A^n A. \end{aligned}$$

  En otras palabras, A^n se genera concatenando n copias de A juntas. Por ejemplo,

$$\begin{aligned} \{ba, bab\}^0 &= \{\varepsilon\}, \\ \{ba, bab\}^1 &= \{ba, bab\}, \\ \{ba, bab\}^2 &= \{baba, babab, babba, babbab\}, \\ \{ba, bab\}^2 &= \{bababa, bababab, bababba, bababbab, babbaba, \\ &\quad\; babbabab, babbabba, babbabbab\} \end{aligned}$$

  También,

$$\begin{aligned} \{a, b\}^n &= \{x \in \{a, b\}^* \,|\, \ |x| = n\} \\ &= \{\text{cadenas sobre } \{a, b\} \text{ de longitud } n\}. \end{aligned}$$

- La *cerradura de Klenne* A^* de un conjunto A es la unión de todas las potencias de A:

$$\begin{aligned} A^* &= \bigcup_{n\geq 0} A^n \\ &= A^0 \cup A^1 \cup A^2 \cup A^3 \cup \cdots . \end{aligned}$$

  Otra forma de definir esto es

$$A^* = \{x_1 x_2 \cdots x_n \,|\, n \geq 0 \text{ y } x_i \in A, 1 \leq i \leq n\}.$$

  Observemos que n puede ser 0; así la cadena vacía ε está en A^* para toda A.

  Definir a Σ^* como el conjunto de todas las cadenas de longitud finita sobre el alfabeto Σ, es exactamente la estrella de Klenne del conjunto Σ.

- Definimos A^+ como la unión de todas las potencias *no cero* de A:

$$A^+ = A^* A = \bigcup_{n\geq 1} A^n.$$

La operación $*$ satisface las siguientes propiedades:

$$\begin{aligned} A^* A^* &= A^*, \\ A^{**} &= A^*, \\ A^* &= \{\varepsilon\} \cup A^* A = \{\varepsilon\} \cup A A^*, \\ \varnothing^* &= \{\varepsilon\}. \end{aligned}$$

Ejercicio 4.1 Para un lenguaje finito L, sea $|L|$ el número de elementos de L. Por ejemplo, $|\{\varepsilon, a, ababb\}| = 3$. El enunciado $|L_1 L_2| = |L_1||L_2|$ dice que el número de cadenas en la concatenación $L_1 L_2$ es el mismo que el producto de los dos números $|L_1|$ y $|L_2|$. ¿Es esto siempre verdad? Si es así, justifique este hecho; si no es el caso, encuentre dos lenguajes finitos $L_1, L_2 \subseteq \{a, b\}^*$ tal que $|L_1 L_2| \neq |L_1||L_2|$.

Ejercicios 4.2

a. Considere el lenguaje L de todas las cadenas de $a's$ y $b's$ que no terminan en b y no contienen a la subcadena bb. Encuentre un lenguaje finito S tal que $L = S^*$.

b. Muestre que no hay lenguaje finito S tal que S^* es el lenguaje de todas las cadenas de $a's$ y $b's$ que no contienen a la subcadena bb.