



PRACTICA 5: Gramáticas Regulares y Autómatas Finitos

5.1. Introducción

Una Gramática $G \equiv (\Sigma, V, P, S)$ es lineal por la derecha si todas sus producciones tienen la forma:

$$A \rightarrow uB|v$$

Si todas las producciones son de la forma

$$A \rightarrow Bu|v$$

La gramática es lineal por la izquierda.

Una gramática es regular si es lineal por la izquierda o lineal por la derecha. Nótese que la restricción consiste en que las partes derechas de las reglas de producción han de contener un único símbolo no terminal y éste debe ser el primero (lineal por la izquierda) o el último símbolo (lineal por la derecha) de la regla.

Veamos ejemplos:

$G_1 \equiv (\Sigma, V, P, S)$ definida por: $\Sigma = \{a, b\}$ $V = \{S, A\}$ y las producciones:

$$S \rightarrow bA$$

$$A \rightarrow aaA|b|\epsilon$$

Es una gramática regular (lineal por la derecha) que genera el lenguaje $L(G_1) = L(b(aa)^*b^?)$ (Cadenas que comienzan con b, seguidas de un número par de aes y que pueden acabar también en b.

$G_2 \equiv (\Sigma, V, P, S)$ definida por: $\Sigma = \{0, 1\}$ $V = \{S, A\}$ y las producciones

$$S \rightarrow 0A$$

$$A \rightarrow 10A|\epsilon$$

Es una gramática regular (lineal por la derecha) que genera el lenguaje $L(G_2) = L(0(01)^*)$.

Una gramática regular, lineal por la izquierda que genera el mismo lenguaje es: $G_3 \equiv (\Sigma, V, P, S)$ definida por: $\Sigma = \{0, 1\}$ $V = \{S\}$ y las producciones

$$S \rightarrow S10|0$$

$$L(G_3) = L(G_2)$$

El objetivo de esta práctica consiste en generar la gramática regular correspondiente a un determinado DFA y por otra parte el paso inverso, es decir, obtener el NFA que reconoce el lenguaje generado por una gramática regular. Para obtener el Autómata Finito Determinista (DFA) correspondiente a una gramática regular, lineal por la derecha se utilizará el siguiente algoritmo:

Entrada: Un DFA $M \equiv (Q, \Sigma, q_0, \delta, F)$

Salida: Una gramática regular $G \equiv (V, \Sigma', S, P)$ lineal por la derecha tal que $L(G) = L(M)$

La Gramática viene definida del siguiente modo:

$$V = Q$$

$$\Sigma' = \Sigma$$

$$S = q_0$$

$$P = \{(q, ap) \mid \delta(q, a) = p\} \cup \{(q, \epsilon) \mid q \in F\}$$

(Habrá una producción $q \rightarrow ap$ si $\delta(q, a) = p$ y una producción $q \rightarrow \epsilon$ si $q \in F$)

Para obtener el NFA correspondiente a la gramática, el algoritmo a utilizar será el siguiente (supondremos que la gramática será lineal por la derecha):

Entrada: Una gramática regular $G \equiv (V, \Sigma, S, P)$ lineal por la derecha

Salida: Un NFA $M \equiv (Q, \Sigma', q_0, \delta, F)$ tal que $L(M) = L(G)$

El NFA viene definido del siguiente modo:

$$Q = V \cup \{f\} \text{ (donde } f \text{ es un nuevo símbolo)}$$

$$F = \{f\}$$

$$q_0 = S$$

$$\Sigma' = \Sigma$$

Y la función de transición se define:

- Si $(A \rightarrow a_1 a_2 \dots a_n B) \in P$ (es una producción de la gramática) entonces se añaden a Q los estados q_1, q_2, \dots, q_{n-1} y las transiciones siguientes: $\delta(A, a_1 a_2 \dots a_n) = \delta(q_1, a_2 \dots a_n) = \dots = \delta(q_{n-1}, a_n) = B$
- Si $(A \rightarrow a_1 a_2 \dots a_n) \in P$ (es una producción de la gramática) entonces se añaden a Q los estados q_1, q_2, \dots, q_{n-1} y las transiciones siguientes: $\delta(A, a_1 a_2 \dots a_n) = \delta(q_1, a_2 \dots a_n) = \dots = \delta(q_{n-1}, a_n) = f$

5.2. Ficheros de descripción de los autómatas

Se utilizará el mismo formato que hasta ahora hemos utilizado para representar los DFAs y NFAs.

5.3. Ficheros de descripción de las gramáticas

Los ficheros de texto que describen gramáticas tienen la extensión `.gra` y en ellos se listan las producciones de la gramática. Los símbolos no terminales de la gramática se representan mediante letras mayúsculas, y los terminales mediante letras minúsculas y caracteres no alfabéticos (a, b, x, z, +, -, *, %, 0, etc.)

Para distinguir los elementos de una cadena $\alpha \in (\Sigma \cup V)^*$, es necesario separarlos mediante espacios o tabulaciones. El fichero contendrá una producción de la forma

$A \rightarrow \alpha$

en cada línea del fichero.

La cadena vacía (ϵ) se representará mediante el carácter \sim .

Se supondrá que el símbolo de arranque de la gramática será el símbolo no terminal que aparece en la parte izquierda de la primera producción listada.

Tanto en los ficheros de descripción de autómatas como de gramáticas, cualquier línea que comience con los caracteres $//$ será ignorada, puesto que representará un comentario. Los comentarios se utilizarán para anotar en los ficheros características significativas de los autómatas o gramáticas correspondientes.

Así para representar la siguiente gramática:

$S \rightarrow SabT | Ta | \epsilon | b$

$T \rightarrow \epsilon | b$

en la que $\Sigma = \{a, b\}$, y $V = \{S, T\}$ es el conjunto de símbolos no terminales, siendo S el símbolo de arranque de la gramática, el fichero tendrá el contenido que muestra el Listado 5.1.

// Fichero que representa a la gramática:

//

// S -> SabT | Ta | ~ | b

// T -> ~ | b

//

S -> S a b T

S -> T a

S -> ~

S -> b

T -> ~

T -> b

Listado 5.1: Fichero de representación de una gramática