

PRACTICA 9: Forma Normal de Chomsky de una Gramática Independiente del Contexto

9.1. Introducción

Una gramática independiente del contexto $G \equiv (V, \Sigma, S, P)$ se dice que está en Forma Normal de Chomsky si todas las producciones de P tienen una de las formas:

$A \rightarrow BC$

$A \rightarrow a$

$S \rightarrow \epsilon$

Si $\epsilon \in L(G)$ la única producción vacía ha de ser $S \rightarrow \epsilon$ y en este caso, el símbolo de arranque de la gramática, S no puede aparecer en la parte derecha de ninguna otra regla de producción.

El objetivo de esta práctica es convertir una gramática independiente del contexto sin producciones unitarias, inútiles ni vacías en su forma normal de Chomsky.

Un pseudocódigo para el algoritmo que se ha de utilizar para convertir G a su forma normal de Chomsky es el que se presenta en la figura 9.1.

Para cada producción de la forma $A \rightarrow \alpha_1\alpha_2 \dots \alpha_n$ ($\alpha_i \in (\Sigma \cup V), n \geq 2$)

Para cada α_i

Si $\alpha_i = a \in \Sigma$

Añadir a P la producción $C_a \rightarrow a$

Cambiar α_i por C_a en la producción $A \rightarrow \alpha_1\alpha_2 \dots \alpha_n$

Para cada producción de la forma $A \rightarrow B_1B_2 \dots B_m$ ($m \geq 3$)

Añadir $m-2$ variables D_1, D_2, \dots, D_{m-1} (distintas para cada producción)

Reemplazar la producción $A \rightarrow B_1B_2 \dots B_m$ por las producciones

$A \rightarrow B_1D_1, D_1 \rightarrow B_2D_2, \dots, D_{m-2} \rightarrow B_{m-1}B_m$

Figura 9.1: Conversión a Forma Normal de Chomsky

La entrada a este algoritmo es una gramática independiente del contexto sin producciones ni símbolos inútiles, sin producciones unitarias y sin producciones vacías salvo $S \rightarrow \epsilon$ en caso de que $\epsilon \in L(G)$.

9.2. 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 `~`.

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 9.1.

// 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

Figura 9.1: Fichero de representación de una gramática