Notas sobre tipos de gramáticas

(Tomado de "Matemática Discreta" de K. H. Rosen)

EJEMPLO 6 Obtén una gramática con estructura de frases que genere el conjunto $\{0^m1^n \mid m$ y n son enteros no negativos$\}$.

Solución: Daremos dos gramáticas G_1 y G_2 que generan este conjunto. Este ejemplo ilustra cómo dos gramáticas distintas pueden generar el mismo lenguaje.

La gramática G_1 consta del alfabeto $V = \{S, 0, 1\}$, conjunto de terminales $T = \{0, 1\}$ y las producciones $S \to 0S$, $S \to S1$ y $S \to \lambda$. G_1 genera el conjunto correcto, puesto que utilizando la primera producción m veces, se colocan m ceros al inicio de la cadena, y usando la segunda producción n veces, tenemos n unos al final de la cadena. Los detalles de esta verificación se dejan al lector.

La gramática G_2 consta del alfabeto $V = \{S, A, 0, 1\}$, conjunto de terminales $T = \{0, 1\}$ y las producciones $S \to 0S$, $S \to 1A$, $S \to 1$, $A \to 1A$, $A \to 1$ y $S \to \lambda$. Los detalles de que esta gramática genera el conjunto correcto se dejan al lector. ◀

En ocasiones, un conjunto de descripción sencilla sólo se puede generar mediante una gramática complicada. El Ejemplo 7 ilustra este hecho.

EJEMPLO 7 Una gramática que genera el conjunto $\{0^n1^n2^n \mid n = 0, 1, 2, ...\}$ es $G = (V, T, S, P)$ con $V = \{0, 1, 2, S, A, B\}$, $T = \{0, 1, 2\}$, símbolo inicial S y las producciones $S \to 0SAB$, $S \to \lambda$, $BA \to AB$, $0A \to 01$, $1A \to 11$, $1B \to 12$, $2B \to 22$. Dejamos como ejercicio para el lector demostrar que esta afirmación es correcta. La gramática dada es la más sencilla de las que generan este conjunto, en un sentido que quedará claro posteriormente en esta sección. El lector se preguntará por el origen de esta gramática, puesto que parece difícil obtenerla partiendo de cero. Puede ser reconfortante saber que esta gramática se puede construir de manera sistemática utilizando técnicas de teoría de la computación que exceden el nivel este libro. ◀

TIPOS DE GRAMÁTICA CON ESTRUCTURA DE FRASES

Enlaces

Las gramáticas con estructura de frases se pueden clasificar de acuerdo con el tipo de producciones que utilicen. Describiremos el esquema de clasificación introducido por Noam Chomsky. En la Sección 11.4 veremos que los diferentes tipos de lenguajes definidos con este esquema corresponden a las clases de lenguaje que pueden ser reconocidas utilizando diferentes modelos de máquinas de computación.

Una gramática de **tipo 0** no impone ninguna restricción a sus producciones. Una gramática de **tipo 1** puede tener producciones de la forma $w_1 \to w_2$, donde la longitud de w_2 es mayor que la de w_1, o de la forma $w_1 \to \lambda$. Una gramática de **tipo 2** sólo puede tener producciones de la forma $w_1 \to w_2$, donde w_1 es un único símbolo no terminal. Una gramática de **tipo 3** sólo puede tener producciones de la forma $w_1 \to w_2$, con $w_1 = A$, y bien $w_2 = aB$ o bien $w_2 = a$, siendo A y B símbolos no terminales y a un símbolo terminal, o con $w_1 = S$ y $w_2 = \lambda$.

De estas definiciones se sigue que toda gramática de tipo 3 es de tipo 2, toda gramática de tipo 2 es de tipo 1 y toda gramática de tipo 1 es de tipo 0. Las gramáticas de tipo 2 también se llaman **gramáticas libres de contexto** o **independientes del contexto,** puesto que un símbolo no terminal que esté en el lado izquierdo de una producción puede ser reemplazado en una cadena siempre que aparezca independientemente de lo que figure en la cadena. Un lenguaje generado por una gramática de tipo 2 se llama **lenguaje libre de contexto** o **independiente del contexto.** Cuando se tiene una producción de la forma $lw_1r \to lw_2r$ (pero no de la forma $w_1 \to w_2$), la gramática se denomina de tipo 1, **sensible al contexto** o **dependiente del contexto,** puesto, que w_1 puede ser reemplazado por w_2 sólo cuando está entre las cadenas l y r. Las gramáticas de tipo 3 se llaman también **regulares.** Un lenguaje generado por una gramática regular se llama **regular.** La Sección 11.4 trata de la relación existente entre las gramáticas regulares y las máquinas de estado finito. El diagrama de Venn de la Figura 1 muestra la relación que existe entre los diferentes tipos de gramáticas.

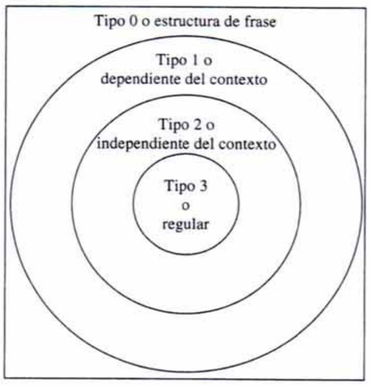

Figura 1. Tipos de gramáticas.

Las gramáticas libres de contexto y las regulares desempeñan un papel muy importante en los lenguajes de programación. Las gramáticas libres de contexto se utilizan para definir la sintaxis de casi todos los lenguajes de programación. Estas gramáticas son lo bastante potentes como para definir una amplia variedad de lenguajes. Además, se pueden diseñar algoritmos eficientes que determinen cuándo y cómo generar una cadena. Las gramáticas regulares se utilizan para buscar ciertos patrones dentro de un texto y en el análisis léxico, que es el proceso en el que un analizador sintáctico toma una secuencia de entrada y la transforma en componentes léxicos (en inglés, *tokens*).

EJEMPLO 9 Del Ejemplo 5 se sigue que $\{0^n1^n \mid n = 0, 1, 2, \ldots\}$ es un lenguaje libre de contexto, puesto que las producciones en esta gramática son $S \to 0S1$ y $S \to \lambda$. Sin embargo, éste no es un lenguaje regular. Esta afirmación se justificará en la Sección 11.4. ◀

EJEMPLO 10 El conjunto $\{0^n1^n2^n \mid n = 0, 1, 2, \ldots\}$ es un lenguaje dependiente del contexto, pues se puede generar mediante una gramática de tipo 1, como muestra el Ejemplo 7, pero no es un lenguaje de tipo 2. (Esto se muestra en el Problema 28 de los Problemas complementarios al final del capítulo). ◀

La Tabla 1 resume la terminología utilizada para clasificar las gramáticas con estructura de frase.

Tabla 1. Tipos de gramáticas.

Tipo	*Restricciones en las producciones* $w_1 \to w_2$
0	Sin restricciones
1	$l(w_1) < l(w_2)$, o $w_2 = \lambda$
2	$w_1 = A$, siendo A un símbolo no terminal
3	$w_1 = A$ y $w_2 = aB$ o $w_2 = a$, siendo $A \in N$, $B \in N$ y $a \in T$, o $S \to \lambda$

Enlaces

AVRAM NOAM CHOMSKY (nacido en 1928) Noam Chomsky, nacido en Filadelfia, es hijo de un profesor de hebreo. Se licenció y doctoró en lingüística por la Universidad de Pennsylvania. Fue profesor de la Universidad de Pennsylvania desde 1950 hasta 1951. En 1955 se incorporó al MIT, donde comenzó su carrera enseñando francés y alemán a estudiantes de ingeniería. Chomsky es actualmente el *Ferrari P. Ward Professor* de Lenguas Extranjeras y Lingüística en el MIT. Es conocido por sus muchas contribuciones fundamentales a la lingüística, incluido el estudio de las gramáticas. Chomsky es también muy conocido por su activismo político.