Teoría de la computación

Victor Melchor

Clase 17

Contenido de la clase

Capítulo 1

Árbol de Derivación

A veces es útil realizar un gráfico de la derivación que indique de qué manera ha contribuido cada símbolo no terminal a formar la cadena final de simbolos terminales. Tal gráfico tiene forma de árbol y se llama árbol de derivación (o árbol de análisis).

Reglas

1. La raíz se etiqueta con un símbolo inicial.
2. Los hijos de la raíz son aquellos símbolos que aparecen al lado derecho de la composición usada para reemplazar el símbolo inicial
3. Todo nodo etiquetado con un símbolo no terminal tiene unos nodos hijos etiquetados con los símbolos del lado derecho de la producción usada para sustituir ese símbolo no terminal.
4. Los nodos que no tienen hijos (hojas) deben ser etiquetados con símbolos terminales.

Definición: Sea $G = (\Sigma_N, \Sigma_T, S, P)$ una GLC. Un árbol de derivación (AD) es un árbol ordenado construido recursivamente como sigue:

1. Un árbol sin aristas cuyo único vértice tiene etiqueta S es un AD de S.

2. Si $X \in \Sigma_N$ es etiqueta de una hoja h de un árbol de derivación A, entonces:

   (a) Si $X \to \varepsilon \in \mathcal{P}$, entonces el árbol obtenido incrementando A a un vértice v con etiqueta ε y una arista $\{h, v\}$ es un AD.

   (b) Si $X \to X_1 X_2 \cdots X_n \in P$, donde $x_1 x_2, \ldots, x_n \in \Sigma_T \cup \Sigma_N$ entonces el árbol obtenido incrementando a A n vértices

$v_1, v_2, \ldots, v_n$ con etiquetas $x_1, x_2, \ldots, x_n$, en ese orden y con n aristas $\{h, v_1\}, \{h, v_2\}, \ldots, \{h, v_n\}$ es un AD.

Si la secuencia de los rótulos de la frontera del AD es la forma sentencial w, se dice que *AD* es un árbol de derivación de w.

Ejemplo: Sea G una GLC donde $\mathcal{P}$ está dado por:

$$E \to E + T|T$$

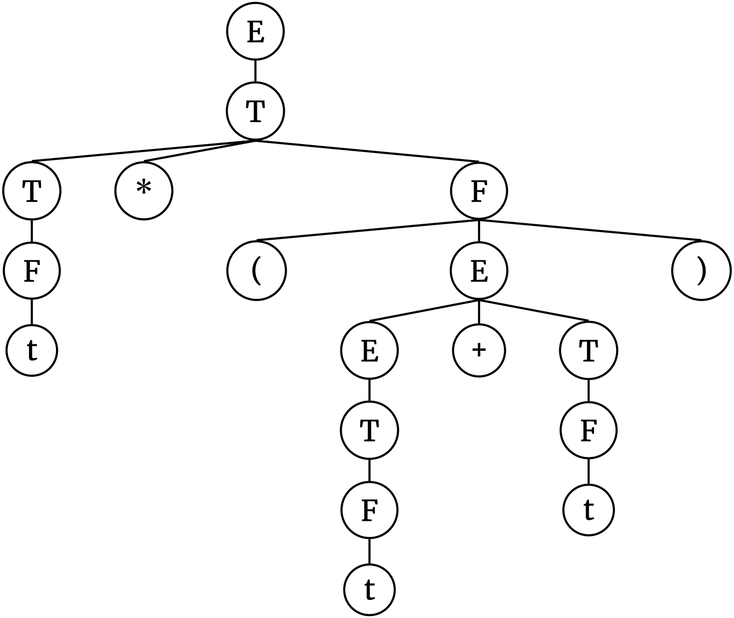
$$T \to T * F|F$$

$$F \to (E)|t$$

Obtener el árbol de derivación para $t * (t + t)$ partiendo de E.

Solución Se tiene

$$
\begin{aligned}
E &\to T && R2\\
&\to T * F && R3\\
&\to T * (E) && R5\\
&\to T * (E + T) && R1\\
&\to T * (T + T) && R2\\
&\to F * (T + T) && R4\\
&\to F * (F + T) && R4\\
&\to F * (F + F) && R4\\
&\to t * (F + F) && R6\\
&\to t * (t + F) && R6\\
&\to t * (t + t) && R6
\end{aligned}
$$


E
T
T
*
F
F
(
E
)
t
E
+
T
T
F
F
t
t

Derivación de 11 pasos.

El número de pasos de cualquier derivación que lleva a un AD $\mathcal{X}$ es el número de vértices internos de $\mathcal{X}$, ya que a cada vértice interno le corresponde la aplicación de una regla.

Cada nodo interno del árbol será un símbolo no terminal de la gramática mientras que las hojas serán los símbolos terminales.

Una regla $A ::= X_1 \dots X_n$ se representará como un subárbol cuyo nodo padre es A siendo sus hijos $X_1, \dots, X_n$

Ejemplo: Sea G la GLC dada por:

$$S \to AB$$
$$A \to aA|a$$
$$B \to bB|b$$

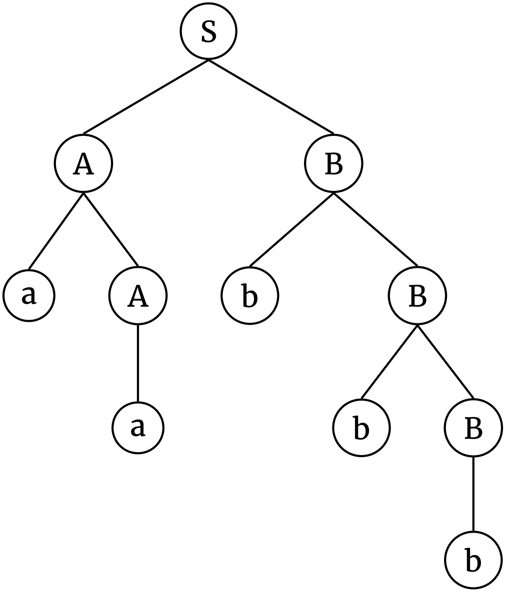
Obtener el AD para la cadena $w = aabbb$

Solución

La derivación de la cadena w es como sigue:

$$
\begin{aligned}
S &\rightarrow AB && R1 \\
&\rightarrow AbB && R4 \\
&\rightarrow AbbB && R4 \\
&\rightarrow Abbb && R5 \\
&\rightarrow aAbbb && R2 \\
&\rightarrow aabbb && R3
\end{aligned}
$$

Su árbol de derivación es:


S
A
B
a
A
b
B
a
b
B
b

Derivación de 6 pasos

Definición: Una GLC G se dirá ambigua cuando es posible construir dos o más AD diferentes para alguna sentencia que genere.

Pueden haber otras GLC equivalentes a una GLC ambigua que no sean ambiguas.

Ejemplo: Sea la GLC $G = (\Sigma_N, \Sigma_T, S, P)$ donde:

$$\Sigma_M = \{S\}$$
$$\Sigma_T = \{t, +, *, (,)\}$$

donde $\mathcal{P}$ está dada por las reglas:

$$S \to S + S|S * S|(S)|t$$

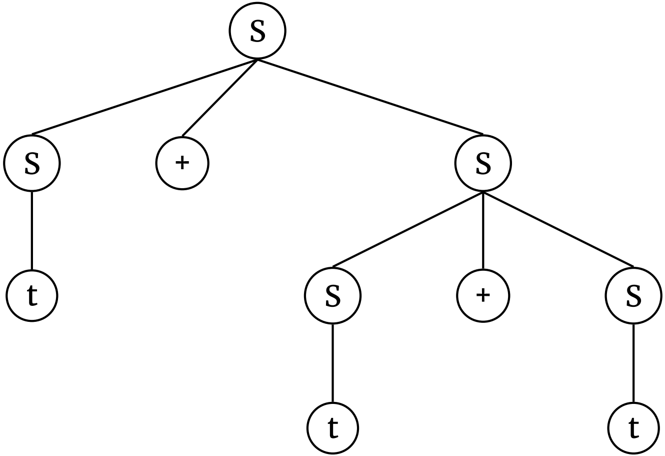
Obtener el AD para la cadena $w = t + t + t$

Solución

1. Tenemos la derivación

$$\begin{aligned} S &\to S+S && R1 \\ &\to t+S && R4 \\ &\to t+S+S && R1 \\ &\to t+t+S && R4 \\ &\to t+t+t && R4 \end{aligned}$$

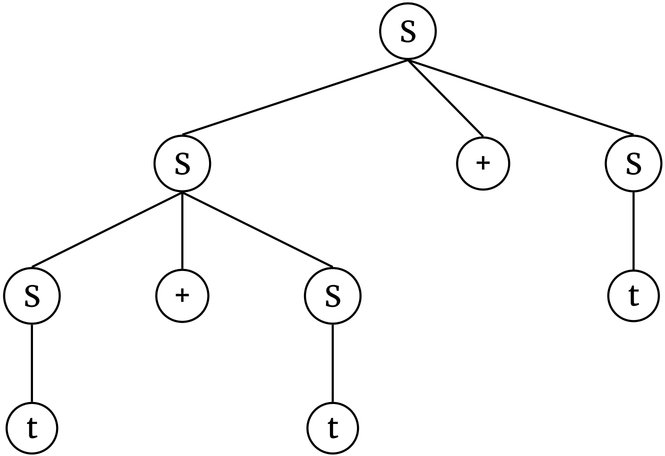
Su árbol de derivación será



2. Sin embargo tenemos la derivación

$$S \to S + S \qquad R1$$

$$
\begin{aligned}
&\rightarrow S + S + S && R1 \\
&\rightarrow t + S + S && R4 \\
&\rightarrow t + t + S && R4 \\
&\rightarrow t + t + t && R4
\end{aligned}
$$



El AD_1 nos conduce a interpretar $t + t + t$ como $t + (t + t)$

El AD_2 nos lleva a interpretar $t + t + t$ como $(t + t) + t$

Ejemplo: Sea la gramática

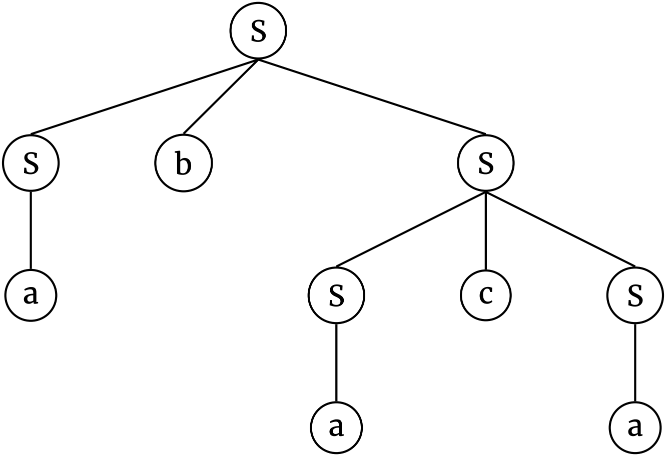
$$S \to SbS|ScS|a$$

Obtener el AD para la cadena $w = abaca$

Solución

1.

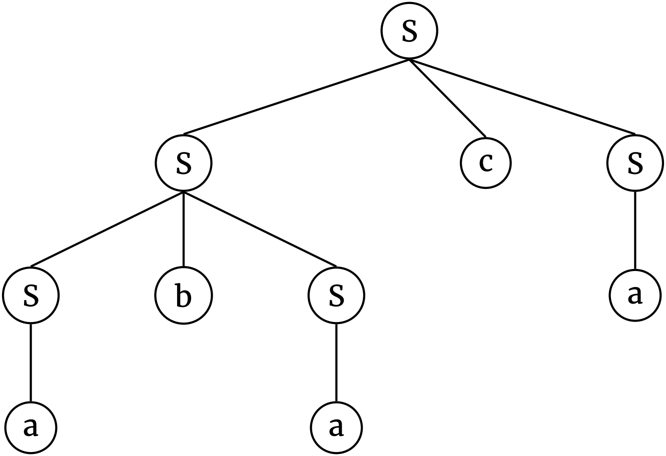
$$
\begin{aligned}
S &\to SbS \\
&\to SbScS \\
&\to SbSca \\
&\to Sbaca \\
&\to abaca
\end{aligned}
$$



2.

$$S \rightarrow ScS$$

$$
\begin{aligned}
&\rightarrow SbScS \\
&\rightarrow SbSca \\
&\rightarrow Sbaca \\
&\rightarrow abaca
\end{aligned}
$$



Ambos AD producen la misma cadena. LA cadena derivada corresponde a los nodos hoja y se llama producto del AD.

Ejemplo: Sea la gramática

$$A \to I := E$$
$$I \to a|b|c$$
$$E \to E + E|E * E|(E)|I$$

donde los símbolos terminales ha sido subrayados.

Obtenga el AD para la cadena

$$a := b + c * a$$

Solución

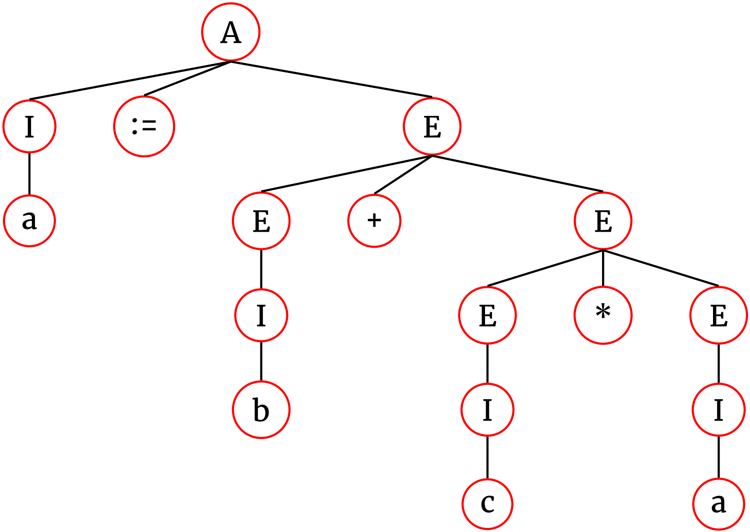
Se obtienen las derivaciones

1.

$$
\begin{aligned}
A &\to I := E \\
&\to I := E + E \\
&\to I := E + E * E \\
&\to c := E + E * E \\
&\to a := I + E * E \\
&\to a := b + E * E \\
&\to a := b + I * E \\
&\to a := b + I * I
\end{aligned}
$$

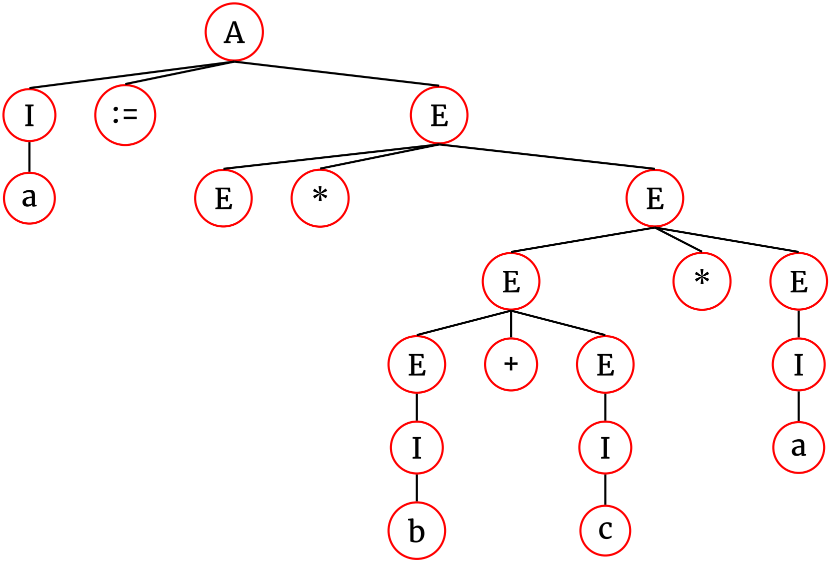
$$\rightarrow a := b + c * l$$

$$\rightarrow a := b + c * a$$


A
I
:=
E
a
E
+
E
I
E
*
E
b
I
I
c
a

2.

$$
\begin{aligned}
A \rightarrow & I := E \\
& I := E * E \\
& I :: E + E * E \\
& a := E + E * E \\
& a := I + E * E \\
& a := I + I * E \\
& a := I + I * I \\
& a := b + I * I \\
& a := b + c * I \\
& a := b + c * a
\end{aligned}
$$


A
I
:=
E
a
E
*
E
E
*
E
E
+
E
I
I
I
a
b
c

Se obtienen dos resultados posibles,

$$b + (c * a) \quad \text{ó} \quad (b + c) * a$$

En general, estos resultados no son iguales.

Definición: Un lenguaje libre de contexto se dice inherentemente ambiguo si todas las gramáticas libres de contexto para L son ambiguas.

Definición: Una derivación se dirá derivación a la izquierda si en cada paso se expande la variable más a la izquierda.

Definición: Una derivación se dirá derivación por la derecha si en cada paso se expande el no terminal más a la derecha.

Ejemplo: Sea la gramática:

$$S \rightarrow SbS|ScS|a$$

Para la cadena $w = abaca$ obtener:

1. Una derivación por la izquierda.
2. Una derivación por la derecha.
3. Árbol de derivación.

Solución

1. Derivación por la izquierda

$$
\begin{aligned}
S &\to ScS \\
&\to SbScS \\
&\to abScS \\
&\to abacS \\
&\to abaca
\end{aligned}
$$

2. Derivación por la derecha

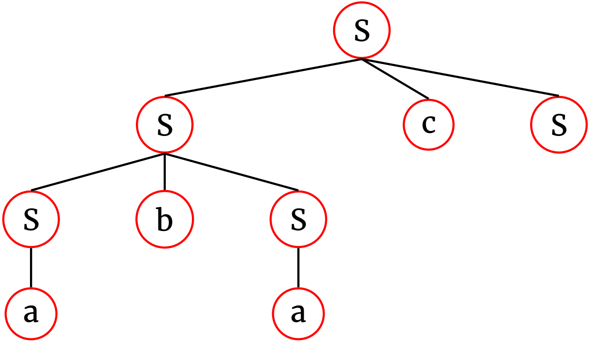
$$
\begin{aligned}
S &\to ScS \\
&\to Sca
\end{aligned}
$$

$$\rightarrow SbSca$$

$$\rightarrow Sbaca$$

$$\rightarrow abaca$$

3.



Una gramática ambigua se caracteriza por tener dos (o más) derivaciones por la izquierda para la misma cadena.