

Arboles de Derivacion

Un árbol de derivación permite mostrar gráficamente cómo se puede derivar cualquier cadena de un lenguaje a partir del símbolo distinguido de una gramática que genera ese lenguaje.

Un árbol es un conjunto de puntos, llamados nodos, unidos por líneas, llamadas arcos. Un arco conecta dos nodos distintos. Para ser un árbol un conjunto de nodos y arcos debe satisfacer ciertas propiedades:

- hay un único nodo distinguido, llamado raíz (se dibuja en la parte superior) que no tiene arcos incidentes.
- todo nodo c excepto el nodo raíz está conectado con un arco a otro nodo k , llamado el padre de c (c es el hijo de k). El padre de un nodo, se dibuja por encima del nodo.
- todos los nodos están conectados al nodo raíz mediante un único camino.
- los nodos que no tienen hijos se denominan hojas, el resto de los nodos se denominan nodos interiores.

El árbol de derivación tiene las siguientes propiedades:

- el nodo raíz está rotulado con el símbolo distinguido de la gramática;
- cada hoja corresponde a un símbolo terminal o un símbolo no terminal;
- cada nodo interior corresponde a un símbolo no terminal.



Para cada cadena del lenguaje generado por una gramática es posible construir (al menos) un árbol de derivación, en el cual cada hoja tiene como rótulo uno de los símbolos de la cadena.

Ejemplo 1

Árbol de derivación para $a * 0 + b$

$$\Sigma = \{a, b, 0, 1, +, *, -, /, (,)\}$$

$$E \rightarrow E + E \mid E * E \mid E - E \mid E / E \mid (E) \mid I \mid N$$

$$N \rightarrow 0 \mid 1 \mid N0 \mid N1$$

$$I \rightarrow a \mid b \mid Ia \mid Ib \mid I0 \mid I1$$

Figura 5: $a * 0 + b$ interpretado como $(a * 0) + b$

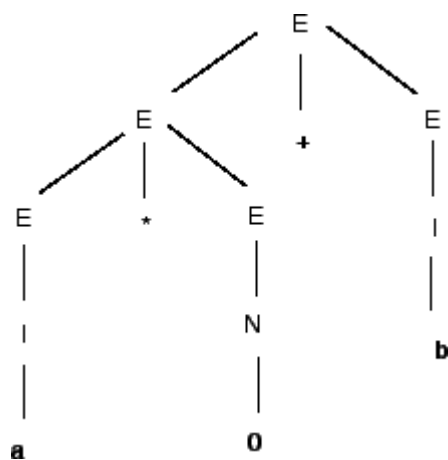


Figura 6: $a * 0 + b$ interpretado como $a * (0 + b)$

Ambigüedades

- Una gramática G es ambigua si existe $x \in L(G)$ con al menos dos árboles de derivación diferentes. Ej: La gramática $S \rightarrow SbS \mid ScS \mid a$ es ambigua
- Equivalentemente, una gramática G es ambigua si existe $x \in L(G)$ con al menos dos derivaciones a la izquierda.
- Un lenguaje L es inherentemente ambiguo si todas las gramáticas para dicho lenguaje son ambiguas. Ej: El lenguaje siguiente es inherentemente ambiguo: $L = \{a^i b^j c^k \mid i=j \text{ o } j=k\}$

Ejemplo 2

El siguiente lenguaje es inherentemente ambiguo:

$$L = \{a^n b^n c^m d^m \mid n \geq 1, m \geq 1\} \cup \{a^n b^m c^m d^n \mid n \geq 1, m \geq 1\}$$

$$\begin{aligned} I &\rightarrow A \mid B \\ A &\rightarrow CD \\ C &\rightarrow ab \mid aCb \\ D &\rightarrow ad \mid cDd \\ B &\rightarrow aBd \mid aEd \\ E &\rightarrow bc \mid bEc \end{aligned}$$

Esta gramática es ambigua:

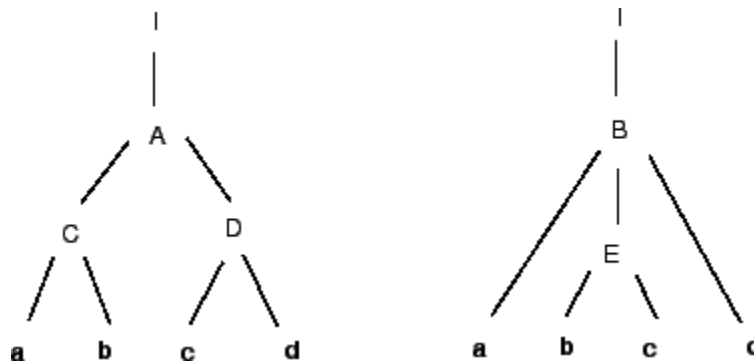


Figura 7: $abcd$

Eso no garantiza que L sea ambiguo. Por eso debemos comprobar que toda GI que lo represente es ambigua. ¿Cómo? Es bastante complejo, no lo veremos.

Intuición

Siempre encontraremos cadenas de la forma $a^n b^n c^n d^n$ que se puedan generar de dos formas distintas (pertenecen a L por dos "motivos" distintos no disjuntos que no podemos "mezclar" en uno).