

Árboles de Derivación

Fabio Martínez Carrillo

Autómatas

Escuela de Ingeniería de Sistemas e Informatica
Universidad Industrial de Santander - UIS

24 de octubre de 2017



Árboles de Derivación

Representación para las derivaciones

- Muestra como se agrupan los simbolos terminales en subcadenas
- Las subcadenas pertenecen al lenguaje de una variable de la gramatica
- Es la estructura de datos que representa el programa fuente

Permite expresar el concepto de **Ambigüedad** en gramáticas.

- Propiedad de tener una unica estructura de arbol para cada cadena del lenguaje

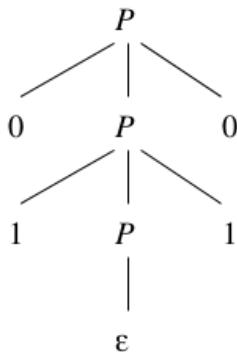
Arboles de Derivación

Para $G = (V, T, P, S)$

- Cada nodo esta etiquetado con la varaibale V
- Cada hoja con un simbolo, terminal o ε
 - Si es ε entonces tiene que ser el único hijo
 - Las variables son nodos interiores
- Si el nodo esta etiquetado con A y sus hijos como X_1, X_2, \dots, X_k
 - Entonces $A \rightarrow X_1, X_2, \dots, X_k$ es una producción
 - X puede reemplazarse por ε cuando es unico hijo y existe $A \rightarrow \varepsilon$

Ejemplo: “palindromos”

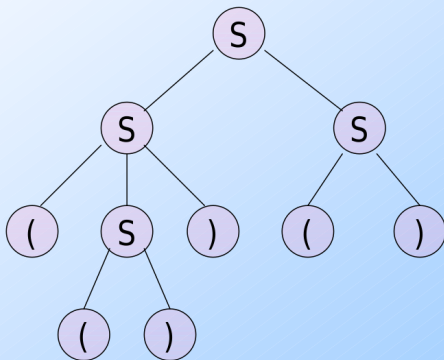
1. $P \rightarrow \varepsilon$
2. $P \rightarrow 0$
3. $P \rightarrow 1$
4. $P \rightarrow 0P0$
5. $P \rightarrow 1P1$



Arbol de derivación para la cadena 0110

Árboles de Derivación para parentesis

$S \rightarrow SS \mid (S) \mid ()$



Parentesis balanceados

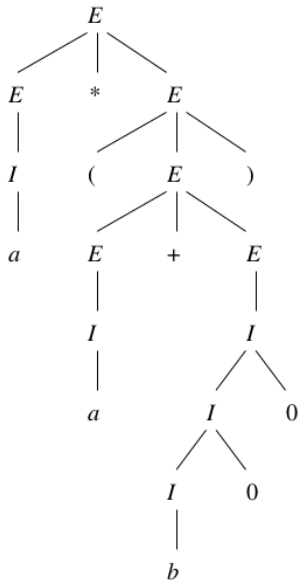
Observaciones

- la concatenación de las hojas de cualquier arbol obtenemos la cadena de resultado.
- El resultado es una cadena terminal
- La raiz esta etiquetada con el simbolo inicial

Cual es el arbol de derivación para la cadena $a * (a + b00)$

1. $E \rightarrow I$
2. $E \rightarrow E + E$
3. $E \rightarrow E * E$
4. $E \rightarrow (E)$
5. $I \rightarrow a$
6. $I \rightarrow b$
7. $I \rightarrow Ia$
8. $I \rightarrow Ib$
9. $I \rightarrow I0$
10. $I \rightarrow I1$

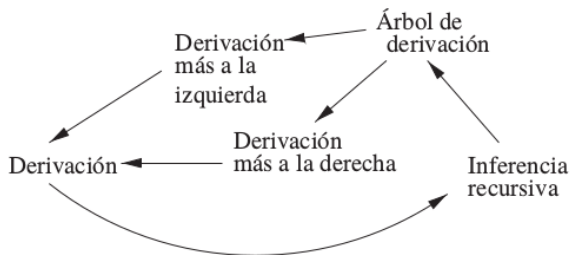
Cual es el arbol de derivación para la cadena $a * (a + b00)$



Inferencia, Derivaciones y Árboles de derivación

Descripción equivalente de cadenas para $G = (V, T, P, S)$

- 1 La inferencia recursiva determina que $w \in L(A)$
- 2 $A \Rightarrow^* w$
- 3 $A \Rightarrow_{lm}^* w$
- 4 $A \Rightarrow_{rm}^* w$
- 5 Existe un árbol de derivación de raíz A y resultado w



De los Árboles a las derivaciones

Propiedad de independencia del contexto

Podemos realizar sustituciones de cuerpos de producciones por sus cabezas de forma independiente.

- $E + (E) \Rightarrow E + (I) \Rightarrow E + (Ib) \Rightarrow E + (ab)$

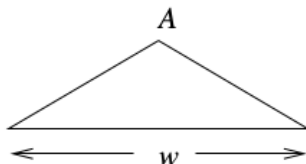
Teorema. Sea $G = (V, T, P, S)$

Existe un árbol de derivación con una raíz etiquetada con la variable A y resultado w . Entonces existe una derivación más a la izquierda $A \Rightarrow_{lm}^* w$ en la gramática G .

Demostración

Caso Base

La altura del árbol es 1. $A \rightarrow w$ tiene que ser una producción y por lo tanto $A \Rightarrow_{lm} w$ es una derivación

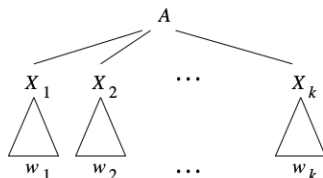


Demostración

Paso Inductivo

Si la altura del árbol es $n > 1$. Existe una raíz etiquetada con A e hijos $X_1 X_2 \dots X_N$ desde la izquierda. X_i puede ser terminal o variable:

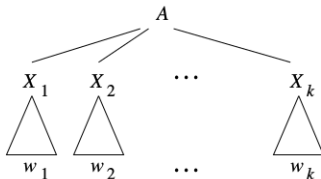
- Si X_i es un simbolo terminal, entonces w_i es la cadena formada unicamente por X_i
- Si X_i es una variable entonces tiene que ser la raíz de un sub-árbol.
 - EL árbol tiene altura menor que n
 - Por inducción existen entonces una derivación más a la izquierda:
 $X_i \Rightarrow_{lm}^* w_i$



Entonces construimos una derivación más a la izquierda para $w = w_1 w_2 \dots w_k$ como:

$$A \Rightarrow_{lm}^* w_1 w_2 \dots w_i X_{i+1} X_{i+2} \dots X_k$$

De forma análoga podemos construir un árbol a partir de una derivación más a la derecha



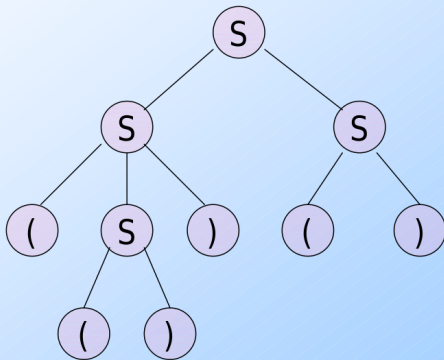
Agenda

1 Aplicaciones

2 Ambigüedad en gramáticas y lenguajes

Árboles de Derivación para parentesis

$S \rightarrow SS \mid (S) \mid ()$



Analizadores sintacticos

Parentesis balanceados

Evaluación de estructuras

Analizador para if - else C

- $S \rightarrow \varepsilon \mid SS \mid iS \mid iSe$
- Las secuencias *ieie*, *iie*, *iei* son posibles
- La secuencia *iee* no pertenece

iiieie

Evaluación de estructuras

Analizador para if - else C

- $S \rightarrow \varepsilon \mid SS \mid iS \mid iSe$
- Las secuencias *ieie*, *iie*, *iei* son posibles
- La secuencia *iee* no pertenece

ieie

```
if (Condición) {  
    ...  
    if (Condición) Instrucción;  
    else Instrucción;  
    ...  
    if (Condición) Instrucción;  
    else Instrucción;  
    ...  
}
```

Lenguaje de marcado

Es un meta-lenguaje que permite definir lenguajes de marcas. las cadenas estan determinados por determinadas marcas (etiquetas). Ej: HTML, XML, ...

HTML

- Crea vinculo entre documentos (cadenas terminales) y describe el formato.
- Existen una serie de clases de cadenas: documento, texto, item, otros.

Agenda

1 Aplicaciones

2 Ambigüedad en gramáticas y lenguajes

Eliminación Ambigüedad

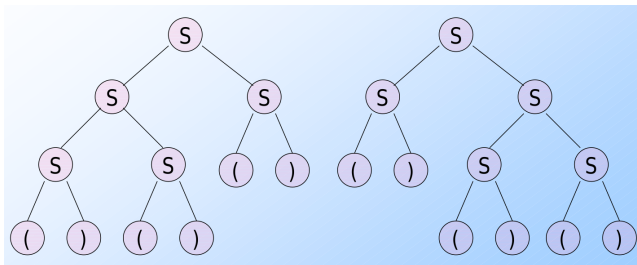
Una gramática libre de contexto es ambigua si existe una cadena en el lenguaje que puede ser representado por dos árboles de derivación.

parentesis balanceados

- $S \rightarrow SS \mid (S) \mid ()$

Existen dos árboles para generar: $()()()$?

parentesis balanceados para $()()()$



Ambigüedad: derivaciones por izq y der

- Si hay dos diferentes árboles, estos deben producir dos derivaciones diferentes por izquierda
- Dos diferentes derivaciones por izquierda deben producir dos diferentes árboles

Ambigüedad

Existe una cadena que tiene dos diferentes derivaciones por izquierda/derecha

Eliminación de la Ambigüedad

Ambigüedad es una propiedad de las gramáticas y no de los lenguajes

- Sin embargo, no existe un algoritmo que nos diga si una Gramática es ambigua

Re-definición para los parentesis balanceados

- $B \rightarrow (RB \mid \varepsilon$
- $R \rightarrow) \mid (RR$
- B es el simbolo de inicio, deriva cadenas balanceadas
- R genera una o más cadenas que tienen uno o mas parentesis a la derecha que a la izq

Eliminación de la Ambigüedad

- $B \rightarrow (RB \mid \varepsilon$
- $R \rightarrow) \mid (RR$
- Si se necesita expandir B se aplica $(RB$ en otro caso ε

Ejemplo: $(())()$

Pasos de derivación desde la izq.

- B

Eliminación de la Ambigüedad

- $B \rightarrow (RB \mid \varepsilon$
- $R \rightarrow) \mid (RR$
- Si se necesita expandir B se aplica $(RB$ en otro caso ε

Ejemplo: $((()))()$

Pasos de derivación desde la izq.

- B
- $(RB$

Eliminación de la Ambigüedad

- $B \rightarrow (RB \mid \varepsilon$
- $R \rightarrow) \mid (RR$
- Si se necesita expandir B se aplica $(RB$ en otro caso ε

Ejemplo: $((()))()$

Pasos de derivación desde la izq.

- B
- $(RB$
- $((RRB$

Eliminación de la Ambigüedad

- $B \rightarrow (RB \mid \varepsilon$
- $R \rightarrow) \mid (RR$
- Si se necesita expandir B se aplica $(RB$ en otro caso ε

Ejemplo: $((()))()$

Pasos de derivación desde la izq.

- B
- $(RB$
- $((RRB$
- $((())RB$

Eliminación de la Ambigüedad

- $B \rightarrow (RB \mid \varepsilon$
- $R \rightarrow) \mid (RR$
- Si se necesita expandir B se aplica $(RB$ en otro caso ε

Ejemplo: $((()))()$

Pasos de derivación desde la izq.

- B
- $(RB$
- $((RRB$
- $((()RB$
- $((()))B$

Eliminación de la Ambigüedad

- $B \rightarrow (RB \mid \varepsilon$
- $R \rightarrow) \mid (RR$
- Si se necesita expandir B se aplica $(RB$ en otro caso ε

Ejemplo: $((()))()$

Pasos de derivación desde la izq.

- B
- $(RB$
- $((RRB$
- $((()RB$
- $((())B$
- $((()))(RB$

Eliminación de la Ambigüedad

- $B \rightarrow (RB \mid \varepsilon$
- $R \rightarrow) \mid (RR$
- Si se necesita expandir B se aplica $(RB$ en otro caso ε

Ejemplo: $((()))()$

Pasos de derivación desde la izq.

- B
- $(RB$
- $((RRB$
- $((()RB$
- $((())B$
- $((())(RB$
- $((())()B$

Eliminación de la Ambigüedad

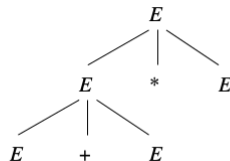
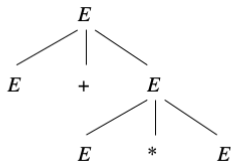
- $B \rightarrow (RB \mid \varepsilon$
- $R \rightarrow) \mid (RR$
- Si se necesita expandir B se aplica $(RB$ en otro caso ε

Ejemplo: $((()))()$

Pasos de derivación desde la izq.

- B
- $(RB$
- $((RRB$
- $((()RB$
- $((())B$
- $((())(RB$
- $((())()B$
- $((())()$

Otro ejemplo



1. $E \Rightarrow E + E \Rightarrow E + E * E$

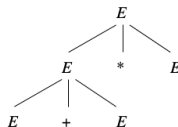
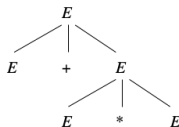
2. $E \Rightarrow E * E \Rightarrow E + E * E$

- $1 + (2 * 3) = 7$

- $(1 + 2) * 3 = 9$

Diferentes derivaciones para una cadena no implica que la gramática sea defectuosa

Causas de Ambigüedad



1. $E \Rightarrow E + E \Rightarrow E + E * E$

2. $E \Rightarrow E * E \Rightarrow E + E * E$

- 1 La precedencia de operadores no se respeta.
- 2 Una secuencia de operadores idénticos puede agruparse empezando por la izquierda o por la derecha

Fuerza de acoplamiento

- Se introducen algunas variables distintas que representan expresiones que comparten el mismo nivel
 - 1 **Factor:** expresión que no se puede separar mediante $*$, $+$
 - Identificadores
 - Cualquier expresión entre parentesis
 - 2 **Termino:** expresión que se puede separar mediante el operador $+$.
 - 3 **Una expresión:** cualquier posible expresión, incluyendo aquellas que pueden separarse mediante un signo $*$ adyacente o un signo $+$ adyacente

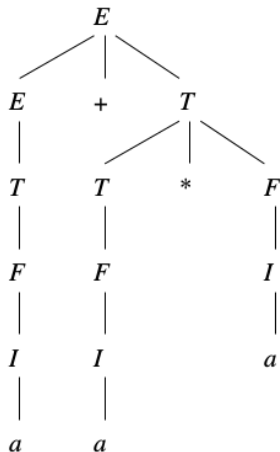
Gramática no ambigua

1. $E \rightarrow I$
2. $E \rightarrow E + E$
3. $E \rightarrow E * E$
4. $E \rightarrow (E)$

5. $I \rightarrow a$
6. $I \rightarrow b$
7. $I \rightarrow Ia$
8. $I \rightarrow Ib$
9. $I \rightarrow I0$
10. $I \rightarrow I1$

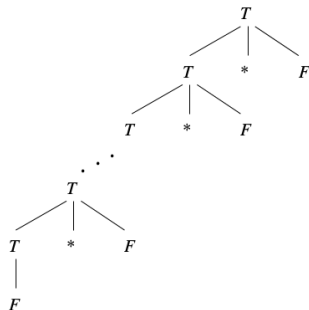
$$\begin{aligned} I &\rightarrow a \mid b \mid Ia \mid Ib \mid I0 \mid I1 \\ F &\rightarrow I \mid (E) \\ T &\rightarrow F \mid T * F \\ E &\rightarrow T \mid E + T \end{aligned}$$

Árbol de derivación: $a + a * a$



- Cualquier cadena derivada de T es factor de $*$. Cualquier identificador o expresión entre parentesis

Derivaciones de T



Árbol de derivación para la expresión $f_1 * f_1 * \dots * f_n$

- Una expresión es una secuencia de terminos conectados mediante el +

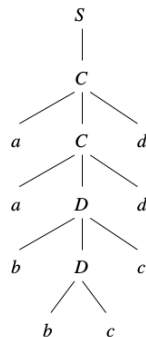
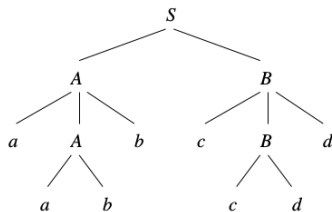
$$L = \{a^n b^n c^m d^m \mid n \geq 1, m \geq 1\} \cup \{a^n b^m c^m d^n \mid n \geq 1, m \geq 1\}$$

L esta conformado por **$a^+ b^+ c^+ d^+$**

- Existen tantos símbolos a como b y tantos símbolos c como d
- Existen tantos símbolos a como d y tantos símbolos b como c .

$$\begin{array}{ll} S & \rightarrow AB \mid C \\ A & \rightarrow aAb \mid ab \\ B & \rightarrow cBd \mid cd \\ C & \rightarrow aCd \mid aDd \\ D & \rightarrow bDc \mid bc \end{array}$$

Inherentemente ambiguos: *aabbccdd*



Considere la gramática:

$$S \rightarrow aS \mid aSbS \mid \varepsilon$$

- Muestre que aab tiene dos arboles de derivación
- Dos derivaciones a la izquierda
- Dos derivaciones a la derecha
- Determine una gramática no ambigua para el lenguaje

Muchas gracias por su atención

