Teoría de Lenguajes

Simplificación de Gramáticas Libres de Contexto

Ejemplo

S → aaA | aS | E | bA | BD

A → aaA | D

B → Bbb | bb

C → bC | b

D → Db | aA

E → aEb | aF

F → aFb | ε

Definiciones

Variable Positiva

$$A \overset{*}{\Rightarrow} w \qquad A \in V \quad w \in T^*$$

Variable Alcanzable

$$S \overset{*}{\Rightarrow} \alpha A \beta \qquad A \in V \quad \alpha, \beta \in (V \cup T)^*$$

Variable Útil

$$S \overset{*}{\Rightarrow} \alpha A \beta \overset{*}{\Rightarrow} w \quad A \in V \quad w \in T^* \quad \alpha, \beta \in (V \cup T)^*$$

Útil $\Longmapsto$ Positiva y Alcanzable √

Positiva y Alcanzable $\Longmapsto$ Útil ?

Definiciones

Variable Positiva

$$A \overset{*}{\Rightarrow} w \qquad A \in V \quad w \in T^*$$

Variable Alcanzable

$$S \overset{*}{\Rightarrow} \alpha A \beta \qquad A \in V \quad \alpha,\beta \in (V \cup T)^*$$

Variable Útil

$$S \overset{*}{\Rightarrow} \alpha A \beta \overset{*}{\Rightarrow} w \quad A \in V \quad w \in T^* \quad \alpha,\beta \in (V \cup T)^*$$

Útil $\Longrightarrow$ Positiva y Alcanzable √

Positiva y Alcanzable $\Longrightarrow$ Útil ?

$S \rightarrow aaA \mid aS \mid E \mid bA \mid BD$
$A \rightarrow aaA \mid D$
$B \rightarrow Bbb \mid bb$
$C \rightarrow bC \mid b$
$D \rightarrow Db \mid aA$
$E \rightarrow aEb \mid aF$
$F \rightarrow aFb \mid \varepsilon$

B?

Definiciones

Producción Unitaria

$$A \rightarrow B \quad A, B \in V$$

Producción épsilon

$$A \rightarrow \varepsilon \quad A \in V$$

Variable Nulable

$$A \stackrel{*}{\Rightarrow} \varepsilon \quad A \in V$$

Definiciones

Gramática Simplificada

es una gramática G = (V,T,P,S) donde:

- ***todas*** sus variables son ***útiles***
- ***no tiene producciones unitarias***
- ***no tiene producciones épsilon***

Algoritmos

(1) Eliminación de producciones épsilon

$A \rightarrow X_1X_2\ldots X_{i-1}\,\mathbf{X_i}\,X_{i+1}\ldots X_{n-1}X_n \in \mathbf{P}$

mientras existan variables nulables hacer

para cada producción $X_i \rightarrow \varepsilon \in P$ hacer

para cada producción que contenga X_i del lado derecho hacer

$$P \leftarrow P \cup \{A \rightarrow X_1X_2\ldots X_{i-1}\ X_{i+1}\ldots X_{n-1}X_n\}$$

fin para

$P \leftarrow P - \{X_i \rightarrow \varepsilon\}$

fin para

fin mientras

Algoritmos

(2) Eliminación de producciones unitarias

$B \rightarrow \boldsymbol{\alpha} \in P \quad \boldsymbol{\alpha} \in (V \cup T)^*$

mientras existan producciones unitarias hacer

para cada producción $A \rightarrow B \in P$ hacer

para cada producción que contenga B del lado izquierdo hacer

$P \leftarrow P \cup \{A \rightarrow \boldsymbol{\alpha}\}$

fin para

$P \leftarrow P - \{A \rightarrow B\}$

fin para

fin mientras

Algoritmos

(3) Obtención de variables positivas

$POS \leftarrow \{ A \,/\, A \rightarrow w \in P \quad w \in T^{*} \}$

mientras POS cambie o existan producciones por chequear hacer

    si $A \notin POS \;\wedge\; A \rightarrow \boldsymbol{\alpha} \in P \;\wedge\; \boldsymbol{\alpha} \in (POS \cup T)^{*}$ entonces

        $POS \leftarrow POS \cup \{ A \}$

    fin si

fin mientras

Algoritmos

(4) Obtención de variables alcanzables

ALC $\leftarrow \{ S \}$

mientras ALC cambie o existan producciones por chequear hacer

si $A \in ALC \wedge A \rightarrow \alpha B \beta \in P \wedge B \in V \wedge \alpha, \beta \in (V \cup T)^*$ entonces

$ALC \leftarrow ALC \cup \{ B \}$

fin si

fin mientras

Teorema General de Simplificación

Todo lenguaje libre de contexto distinto de vacío y a menos de la tira épsilon (ε), puede ser generado por una ***gramática libre de contexto simplificada***.

Demostración:

Sea L = L(G), donde G=(V,T,P,S) es una gramática que genera L – {ε} y además L ≠ ∅

$$G \underset{(1)}{\overset{*}{\Rightarrow}} G_1 \underset{(2)}{\overset{*}{\Rightarrow}} G_2 \underset{(3)}{\overset{*}{\Rightarrow}} G_3 \underset{(4)}{\overset{*}{\Rightarrow}} G_4$$

Teorema General de Simplificación (cont.)

Ver que alcanza con aplicar en ese orden los algoritmos, sobre las gramáticas que van resultando en cada caso.

Los primeros 2 algoritmos (eliminación de producciones épsilon y unitarias) agregan nuevas producciones.

Puede quedar la duda del orden de aplicar los otros 2 algoritmos – que debieran aplicarse una sola vez cada uno para poder obtener todos símbolos útiles.

$$G \overset{*}{\underset{(1)}{\Longrightarrow}} G_1 \overset{*}{\underset{(2)}{\Longrightarrow}} G_2 \overset{*}{\underset{(3)}{\Longrightarrow}} G_3 \overset{*}{\underset{(4)}{\Longrightarrow}} G_4$$

Teorema General de Simplificación (cont.)

Supongamos una tira $\boldsymbol{w} \in L(G)$ y que quede en $\mathbf{G_4}$ una variable X que **no** es útil (inútil)

Como el último algoritmo aplicado es el (4) $S \overset{*}{\Rightarrow} \boldsymbol{\alpha} X \boldsymbol{\beta}$

pero a su vez, X viene de una gramática G_3 dónde todas las variables eran positivas (al igual que las de $\boldsymbol{\alpha}$ y $\boldsymbol{\beta}$)

Con lo cual podemos decir que $\boldsymbol{\alpha} X \boldsymbol{\beta} \overset{*}{\Rightarrow} \boldsymbol{w}$

De donde se tiene que $S \overset{*}{\Rightarrow} \boldsymbol{\alpha} X \boldsymbol{\beta} \overset{*}{\Rightarrow} \boldsymbol{w}$

O sea que **X es útil**

Aplicación

$S \rightarrow aaA \mid aS \mid E \mid bA \mid BD$

$A \rightarrow aaA \mid D$

$B \rightarrow Bbb \mid bb$

$C \rightarrow bC \mid b$

$D \rightarrow Db \mid aA$

$E \rightarrow aEb \mid aF$

$F \rightarrow aFb \mid \varepsilon$

Aplicación

(1) Elimino producciones-ε

Las producciones afectadas son:

E → aEb | aF
F → aFb | ε

quedando….

E → aEb | aF | a
F → aFb | ab

Aplicación

S → aaA | aS | E | bA | BD

A → aaA | D

B → Bbb | bb

C → bC | b

D → Db | aA

E → aEb | aF | a

F → aFb | ab

Aplicación

(2) Elimino producciones unitarias

Las producciones afectadas son:

S → E
A → D

quedando en su lugar ....

S → aEb | aF | a
A → Db | aA

Aplicación

S → aaA | aS | bA | BD | aEb | aF | a

A → aaA | Db | aA

B → Bbb | bb

C → bC | b

D → Db | aA

E → aEb | aF | a

F → aFb | ab

Aplicación

(3) Elimino Variables NO Positivas

S → aaA | aS | bA | BD | aEb | aF | a

A → aaA | Db | aA

B → Bbb | bb

C → bC | b

D → Db | aA

E → aEb | aF | a

F → aFb | ab

POS = {F,E,C,B,S}

Aplicación

(3) Elimino Variables NO Positivas

S → aaA | aS | bA | BD | aEb | aF | a

A → aaA | Db | aA

B → Bbb | bb

C → bC | b

D → Db | aA

E → aEb | aF | a

F → aFb | ab

POS = {F,E,C,B,S}

S → aS | aEb | aF | a

B → Bbb | bb

C → bC | b

E → aEb | aF | a

F → aFb | ab

Aplicación

(4) Elimino variables NO Alcanzables

S → aS | aEb | aF | a

B → Bbb | bb

C → bC | b

E → aEb | aF | a

F → aFb | ab

ALC = {S}

ALC= {S, E, F}

S → aS | aEb | aF | a

E → aEb | aF | a

F → aFb | ab

Aplicación

$S \rightarrow aS \mid aEb \mid aF \mid a$

$E \rightarrow aEb \mid aF \mid a$

$F \rightarrow aFb \mid ab$

¿qué lenguaje genera esta gramática? $\mathbf{a^k b^p}$ $k>p\geq 0$

$$\boxed{S \rightarrow aS \mid aSb \mid a}$$

Aplicación

Supongamos que se aplica primero el 4 – Alcanzables y luego el 3 – Positivas y veamos con un ejemplo que ese orden NO nos asegura que se obtenga una gramática simplificada.

Tomemos la gramática que veníamos trabajando ya habiendo eliminado las producciones-ε y las producciones unitarias

Aplicación

S → aaA | aS | bA | BD | aEb | aF | a

A → aaA | Db | aA

B → Bbb | bb

C → bC | b

D → Db | aA

E → aEb | aF | a

F → aFb | ab

ALC = {S}
ALC = {S,A,B,D,E,F}

S → aaA | aS | bA | BD | aEb | aF | a

A → aaA | Db | aA

B → Bbb | bb

D → Db | aA

E → aEb | aF | a

F → aFb | ab

Aplicación

$S \rightarrow aaA \mid aS \mid bA \mid BD \mid aEb \mid aF \mid a$

$A \rightarrow aaA \mid Db \mid aA$

$B \rightarrow Bbb \mid bb$

$D \rightarrow Db \mid aA$

$E \rightarrow aEb \mid aF \mid a$

$F \rightarrow aFb \mid ab$

POS = {S,B,E,F}

$S \rightarrow aS \mid aEb \mid aF \mid a$

$B \rightarrow Bbb \mid bb$

$E \rightarrow aEb \mid aF \mid a$

$F \rightarrow aFb \mid ab$

Finalmente entonces...

Gramática Simplificada

es una gramática $G = (V,T,P,S)$ donde:

- ***todas*** sus variables son ***útiles***
- ***no tiene producciones unitarias***
- ***no tiene producciones epsilon***

Todo Lenguaje Libre de Contexto distinto de vacío ($L \neq \varnothing$) y que no contenga la tira épsilon $L - \{\varepsilon\}$, puede ser generado por una ***gramática simplificada***.

Normalización - FNC

Sea una GLC $G = (V,T,P,S)$

- *Forma Normal de Chomsky*

$A \rightarrow BC \qquad \forall P \quad \text{con } A,B,C \in V$

$A \rightarrow a \qquad\qquad\qquad a \in T$

Ejemplo

$S \rightarrow AB$
$A \rightarrow 0$
$B \rightarrow 1 \mid SC$
$C \rightarrow 1$

Normalización - FNG

Sea una GLC $G = (V,T,P,S)$

- *Forma Normal de Greibach*

$A \rightarrow a\alpha \qquad \forall P \text{ con } a \in T,\ A \in V \text{ y } \alpha \in V^*$

Ejemplo

$S \rightarrow 0SB$
$S \rightarrow 0B$
$B \rightarrow 1$