Pontificia Universidad Católica de Chile
Departamento de Ciencia de la Computación
IIC2223 – Teoría de Autómatas y Lenguajes Formales
Segundo semestre de 2024
Profesor: Cristian Riveros
Ayudante: Amaranta Salas

Ayudantia 12

Lema de Bombeo y Algoritmo CKY

Problema 1

1. Demuestre que el siguiente lenguaje NO es libre de contexto:

$$S = \{a^i b^j c^k \mid i < j \wedge j < k\}$$

2. Considere el siguiente lenguaje sobre el alfabeto $\Sigma = \{a, b, c, d\}$:

$$L = \{a^i b^j c^k d^l \mid 2i = k \wedge 3j = l\}$$

   Demuestre que el lenguaje anterior **NO** es libre de contexto.

3. Para todo $u, v \in \{0,1\}^*$ se define $u * v = \bigvee_{i-1}^{n}(a_i \wedge b_i)$ tal que $u = a_1...a_n$ y $v = b_1...b_n$. Demuestre que el siguiente lenguaje no es libre de contexto:

$$L = \{u\#v \mid u, v \in \{0,1\}^+ \wedge |u| = |v| \wedge u * v = 0\}$$

Solución

1. Por contrapositivo Lema de Bombeo para CFG. Sea para todo $N > 0$ la palabra $z = a^N b^{N+1} c^{N+2}$. Sean $i = N$, $j = N + 1$ y $k = N + 2$ se tiene que $i < j < k$, esto es, z está en S y $|z| \geq N$.

   Sea $z = uvwxy$ una descomposición cualquiera tal que $vx \neq \epsilon$ y $|vwx| \leq N$. Sea $z' = uv^i wx^i y$ para algún $i \geq 0$. Luego, se tienen los siguientes casos (no necesariamente excluyentes entre sí):

   (a) Si v o x son combinaciones de 2 letras. Entonces, con $i = 2$, z' ya no está en $\mathcal{L}(a^*b^*c^*)$ ni en S.
   (b) Si $x \in \mathcal{L}(a^*)$. Entonces, con $i = 2$, se tienen mayor o igual letras a que b y $z' \notin S$.
   (c) Si $x \in \mathcal{L}(b^*)$. Entonces, con $i = 2$, se tienen mayor o igual letras b que c y $z' \notin S$.
   (d) Si $v \in \mathcal{L}(b^*)$. Entonces, con $i = 0$, se tienen mayor o igual letras a que b y $z' \notin S$.
   (e) Si $v \in \mathcal{L}(c^*)$. Entonces, con $i = 0$, se tienen mayor o igual letras b que c y $z' \notin S$.

   Por lo tanto, S no es un Lenguaje Libre de Contexto.

2. Para demostrar que L no es libre de contexto, usaremos el contrapositivo del lema de bombeo para lenguajes libres de contexto. Sea un $N > 0$, definimos la siguiente palabra z perteneciente al lenguaje:

$$z = a^N b^N c^{2N} d^{3N}$$

   Y sea la siguiente descomposición cualquiera para z:

$$z = uvwxy$$

   Con $vx \neq \epsilon$ y $|vwx| \leq N$. Según la estructura que tome dicha descomposición, podemos tener la siguientes tres posibilidades:

(a) $vwx \in a^*b^*$. En este caso, sea $i = 0$ de forma que la palabra queda como:

$$uv^0wx^0y = a^{N_1}b^{N_2}c^{2N}d^{3N} \tag{1}$$

Lo que fuerza a que $N_1 < N$ ó $N_2 < N$, por lo que la palabra nueva no pertenece a L.

(b) $vwx \in b^*c^*$. Es este caso, sea $i = 0$:

$$uv^0wx^0y = a^{N}b^{N_2}c^{N_3}d^{3N} \tag{2}$$

Lo que fuerza a que $N_2 < N$ ó $N_3 < 2N$, por lo que la palabra nueva no pertenece a L.

(c) $vwx \in c^*d^*$. Es este caso, sea $i = 0$:

$$uv^0wx^0y = a^{N}b^{N}c^{N_3}d^{N_4} \tag{3}$$

Lo que fuerza a que $N_3 < 2N$ ó $N_4 < 3N$, por lo que la palabra nueva no pertenece a L.

(d) $vwx \in a^* \vee vwx \in b^* \vee vwx \in c^* \vee vwx \in d^*$. Es este caso, sea $i = 0$, por argumento análogo a los anteriores, la palabra nueva no pertenece a L.

Al ser analizadas todas las posibilidades de descomposición para la palabra z, se demuestra que L no es un lenguaje libre de contexto.

3. Demostraremos que $L = \{u\#v \mid u, v \in \{0,1\}^+, |u| = |v| \wedge u * v = 0\}$ no es libre de contexto usando el contrapositivo del lema de bombeo. Sea $z = 0^N1^N\#1^N0^N$, con $N > 0$. Es claro que $z \in L \wedge |z| \geq N$. Luego, tomando $z = u \cdot v \cdot w \cdot x \cdot y$, con $v \cdot x \neq \varepsilon \wedge |v \cdot w \cdot x| \leq N$, nos encontraremos con los siguientes casos posibles, de acuerdo a dónde quede el símbolo #:

   (a) Si # está en v o # está en x. Entonces $|u \cdot v^2 \cdot w \cdot x^2 \cdot y|_\# = 2$ y por lo tanto $u \cdot v^2 \cdot w \cdot x^2 \cdot y \notin L$.

   (b) Si # está en u o # está en y. Acá sabemos que $a\#b = u \cdot v^2 \cdot w \cdot x^2 \cdot y$ con $|a| \neq |b|$. Por lo tanto, $u \cdot v^2 \cdot w \cdot x^2 \cdot y \notin L$.

   (c) Si # está en w, entonces $v, x \in 1^+$. Este caso se divide en dos:

      - Si $|v| \neq |x|$, entonces $a\#b = u \cdot v^2 \cdot w \cdot x^2 \cdot y$ y $|a| \neq |b|$. Por lo tanto, $u \cdot v^2 \cdot w \cdot x^2 \cdot y \notin L$.
      - Si $|v| = |x|$ tenemos que $u \cdot v^2 \cdot w \cdot x^2 \cdot y = 0^N1^{N+k}\#1^{N+k}0^N$ para algún $k > 0$. Acá es claro que $0^N1^{N+k} * 1^{N+k}0^N = 1$. De esto concluimos que $u \cdot v^2 \cdot w \cdot x^2 \cdot y \notin L$.

Por tanto, para toda descomposición $z = u \cdot v \cdot w \cdot x \cdot y$ existe un $i \geq 0$ tal que $u \cdot v^i \cdot w \cdot x^i \cdot y \notin L$, lo que quiere decir que L no es libre de contexto.

Problema 2

Considere la gramática libre de contexto $\mathcal{G}$:

$$\begin{aligned} S &\to ABS \mid AB \\ A &\to aA \mid a \\ B &\to BAb \mid \epsilon \end{aligned}$$

Convierta $\mathcal{G}$ en una gramática libre de contexto en forma normal de Chomsky y verifique si la palabra *aabab* esta en el lenguaje $L(\mathcal{G})$ usando el algoritmo CKY. Explique los pasos en cada algoritmo que utilice.

Solución

Para pasar la gramática a forma normal de Chomsky, primero debemos eliminar las producciones en vacío y unitarias.

1. Simplificar la gramática

   (a) Eliminar producciones en vacío
   Agregamos las producciones que obtendríamos al no considerar la variable que da ϵ.

$$\begin{aligned} S &\to ABS \mid AB \mid AS \mid A \\ A &\to aA \mid a \\ B &\to BAb \end{aligned}$$

   (b) Eliminar producciones unitarias
   Reemplazamos la variable unitaria por sus producciones.

$$\begin{aligned} S &\to ABS \mid AB \mid AS \mid aA \mid a \\ A &\to aA \mid a \\ B &\to BAb \end{aligned}$$

2. Pasar la gramática obtenida a CNF

   (a) Para cada $a \in \Sigma$, agregar un nueva variable X_a y una regla $X_a \to a$. Luego, reemplazar todas las ocurrencias antiguas de a por X_a.
   En este caso creamos las reglas $X_a \to a$ y $X_b \to b$ y reemplazamos donde sea necesario, cuidando de no generar variables unitarias.

$$\begin{aligned} S &\to ABS \mid AB \mid AS \mid X_aA \mid a \\ A &\to X_aA \mid a \\ B &\to BAX_b \\ X_a &\to a \\ X_b &\to b \end{aligned}$$

   (b) Para cada regla $p : X \to Y_1Y_2 \ldots Y_k$ con $k \geq 3$ agregamos una nueva variable Z y reemplazamos la regla p por dos reglas $X \to Y_1Z$ y $Z \to Y_2 \ldots Y_k$. Repetimos este paso hasta llegar a la forma normal de Chomsky.
   En este caso solo tenemos reglas que derivan en a lo más 3 variables, por lo que creamos $Y = BS$ y $Z = AX_b$ y reemplazamos donde corresponde.

$$\begin{aligned} S &\to AY \mid AB \mid AS \mid X_aA \mid a \\ A &\to X_aA \mid a \\ B &\to BZ \\ X_a &\to a \\ X_b &\to b \\ Y &\to BS \\ Z &\to AX_b \end{aligned}$$

3. Aplicar el algoritmo CKY para verificar si $aabab$ está en $\mathcal{L}(\mathcal{G})$.

   Para esto consideremos el siguiente cuadro, donde cada C_{ij} corresponde a un 'camino' entre i y j.

C_{15}				
C_{14}	C_{25}			
C_{13}	C_{24}	C_{35}		
C_{12}	C_{23}	C_{34}	C_{45}	
C_{11}	C_{22}	C_{33}	C_{44}	C_{55}
a_1	a_2	a_3	a_4	a_5

   (a) Para cada i, construimos el conjunto $C_{ii} \in V$ tal que $C_{ii} = \{X \in V \mid X \to a_i \in P\}$. Es decir, C_{ii} va a ser el conjunto de variables que tienen como derivación la letra correspondiente.

$\{S, A, X_a\}$	$\{S, A, X_a\}$	$\{X_b\}$	$\{S, A, X_a\}$	$\{X_b\}$
a	a	b	a	b

   (b) Para cada i, construimos el conjunto $C_{ii+1} \in V$ tal que $C_{ii+1} = \{X \in V \mid X \to YZ \in P$ para algún $Y \in C_{ii} \wedge Z \in C_{i+1i+1}\}$. Es decir, C_{ii+1} va a ser el conjunto de variables tales que su derivación tiene como primera variable una del conjunto C_{ii} y la segunda en C_{i+1i+1}.

$\{S, A\}$	$\{Z\}$	$\varnothing$	$\{Z\}$	
$\{S, A, X_a\}$	$\{S, A, X_a\}$	$\{X_b\}$	$\{S, A, X_a\}$	$\{X_b\}$
a	a	b	a	b

   (c) Para cada i, construimos el conjunto $C_{ii+2} \in V$ tal que $C_{ii+2} = \{X \in V \mid \exists j \in [i, i+2).\ X \to YZ \in P$ para algún $Y \in C_{ij} \wedge Z \in C_{j+1i+2}\}$. Es decir, C_{ii+2} va a ser la unión de los conjuntos de variables, de cada j, tales que su derivación tiene como primera variable una del conjunto del camino de i a j y de $j+1$ a $i+2$. Por ejemplo, C_{13} va a ser la union de los conjuntos de C_{11} a C_{23} y de C_{12} a C_{33}.

$\{Z\}$	$\varnothing$	$\varnothing$		
$\{S, A\}$	$\{Z\}$	$\varnothing$	$\{Z\}$	
$\{S, A, X_a\}$	$\{S, A, X_a\}$	$\{X_b\}$	$\{S, A, X_a\}$	$\{X_b\}$
a	a	b	a	b

(d) Para cada i, construimos el conjunto $C_{ii+3} \in V$ tal que $C_{ii+3} = \{X \in V \mid \exists j \in [i, i+3).\ X \to YZ \in P \text{ para algún } Y \in C_{ij} \wedge Z \in C_{j+1i+3}\}$. Es decir, C_{ii+3} va a ser la unión de los conjuntos de variables, de cada j, tales que su derivación tiene como primera variable una del conjunto del camino de i a j y de $j+1$ a $i+3$.

$\varnothing$	$\varnothing$			
$\{Z\}$	$\varnothing$	$\varnothing$		
$\{S, A\}$	$\{Z\}$	$\varnothing$	$\{Z\}$	
$\{S, A, X_a\}$	$\{S, A, X_a\}$	$\{X_b\}$	$\{S, A, X_a\}$	$\{X_b\}$
a	a	b	a	b

(e) Para cada i, construimos el conjunto $C_{ii+4} \in V$ tal que $C_{ii+4} = \{X \in V \mid \exists j \in [i, i+4).\ X \to YZ \in P \text{ para algún } Y \in C_{ij} \wedge Z \in C_{j+1i+4}\}$. Es decir, C_{ii+4} va a ser la unión de los conjuntos de variables, de cada j, tales que su derivación tiene como primera variable una del conjunto del camino de i a j y de $j+1$ a $i+4$.

$\varnothing$				
$\varnothing$	$\varnothing$			
$\{Z\}$	$\varnothing$	$\varnothing$		
$\{S, A\}$	$\{Z\}$	$\varnothing$	$\{Z\}$	
$\{S, A, X_a\}$	$\{S, A, X_a\}$	$\{X_b\}$	$\{S, A, X_a\}$	$\{X_b\}$
a	a	b	a	b

Finalmente como llegamos al conjunto vacío, la palabra *aabab* no pertenece a $\mathcal{L}(\mathcal{G})$.