Tema 5 – Lenguajes independientes del contexto - Objetivos

Objetivos:

- Conocer los fundamentos de las gramáticas independientes del contexto
- Saber aplicar las distintas transformaciones en las gramáticas independientes del contexto (gramáticas de tipo 2)
- Conocer los problemas de decisión en los lenguajes libres de contexto
- Conocer el lema de bombeo y las propiedades de cerradura para lenguajes libres de contexto
- Conocer los fundamentos de los autómatas de pila
- Aplicar la transformación de gramática de tipo 2 a un autómata de pila

Tema 5 - Lenguajes independientes del contexto - Conceptos

Índice:

- Gramáticas independientes del contexto
- □ Transformación de gramáticas independientes del contexto
- Forma Normal de Chomsky
- Lema de bombeo para lenguajes independientes del contexto
- Propiedades de cerradura para los lenguajes independientes del contexto
- Problemas de decisión para lenguajes independientes del contexto
- Algoritmo CYK
- Autómatas de pila
- Obtención de un autómata de pila a partir de la gramática

Tema 5 Gramáticas independientes del contexto - Definición

Gramáticas independientes del contexto

Las gramáticas independientes del contexto (de tipo 2 o de contexto libre) son aquellas cuyas producciones se ajustan a la siguiente forma:

$$A \rightarrow \alpha$$
 en donde $A \in N$; $\alpha \in (\Sigma \mid N)^*$

El término independiente del contexto viene de que se puede cambiar A por α en cualquier contexto donde aparezca A.

La siguiente gramática es independiente del contexto:

G= (
$$\Sigma$$
={a, b}, N={S, A, B}, S, P} P: S \rightarrow aABAa, A \rightarrow ϵ |aA, B \rightarrow ϵ |bB

Es una gramática que genera las palabras del lenguaje::

L={
$$a^n b^m a^l | n, l \ge 1 \text{ y m} \ge 0$$
}

Tema 5 Gramáticas independientes del contexto - Derivación

Derivación por la izquierda

Consiste en aplicar un proceso de derivación al no terminal más a la izquierda

$$S \rightarrow aABAa \rightarrow aBAa \rightarrow aAa \rightarrow aa$$

Derivación por la derecha

Consiste en aplicar un proceso de derivación al no terminal más a la

$$S \rightarrow aABAa \rightarrow aABa \rightarrow aAa \rightarrow aa$$

Árbol de derivación

Es una representación gráfica (en forma de árbol invertido) de un proceso de derivación en una gramática

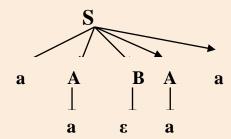
Se define el árbol de derivación como sigue:

- la raiz del árbol será el símbolo inicial de la gramática
- los nodo interiores del árbol están etiquetados por los símbolos no terminales
- las hojas están etiquetadas por símbolos terminales
- si un nodo interior etiquetado por A, posee como hijos los nodos etiquetados por $X_1,\,X_2,\,...X_n$, entonces A \to $X_1,\,X_2,\,...X_n$ es una producción de la gramática, en donde Xi representa un símbolo terminal o no terminal

Tema 5 Gramáticas independientes del contexto - Gramáticas ambiguas

Sea la siguiente gramática:

G = (Σ ={a, b}, N={S, A, B}, S, P} P: S \rightarrow aABAa; A \rightarrow ϵ |aA; B \rightarrow ϵ |bB árbol de derivación en el proceso de generación de la palabra x=*aaaa* es el siguiente:

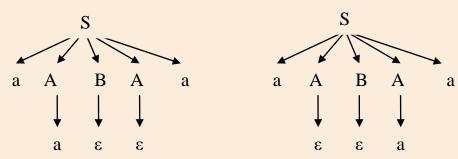


Gramáticas ambiguas

Son aquellas en las que se generan palabras por más de un árbol de derivación La siguiente gramática es una gramática ambigua, la palabra x=aaa, se genera por más de un árbol de derivación:

G = (
$$\Sigma$$
= {a, b}, N={S, A, B}, S, P} P: S \rightarrow aABAa; A $\rightarrow \varepsilon$ | aA; B $\rightarrow \varepsilon$ |bB

La generación de la palabra x=aaa se puede hacer por más de un árbol de derivación



Tema 5 Gramáticas independientes del contexto - Ambiguedad

La ambigüedad es un proceso indecidible, no existe un algoritmo para detectarlo y corregirlo

Lenguaje inherentemente ambiguo

Es aquel en que todas las gramáticas que lo generan son ambiguas

El siguiente lenguaje es inherentemente ambiguo:

L=
$$\{a^n b^m c' | n=m o m=1; n, m, l \ge 0\}$$

Una de las posibles gramáticas que lo genera es la que sigue:

$$G = (\Sigma = \{a, b, c\}, N = \{S, A, B, C, D\}, S, P\}$$

 $P: S \rightarrow AB|CD$

 $A \rightarrow aAb \mid \varepsilon$

 $B \rightarrow dB \mid \epsilon$

 $C \rightarrow aC \mid \epsilon$

 $D \rightarrow bDc \mid \epsilon$

Grado de ambigüedad - número de árboles que genera una misma palabra

Limpiar una gramática

Limpiar una gramática consiste en quitar los símbolos inútiles e inaccesibles

Símbolos inútiles

Símbolos no terminales a través de los cuales no se genera ninguna palabra perteneciente al lenguaje.

Son símbolos útiles los siguientes: $A \rightarrow^* w$; $w \in \Sigma^*$; A es útil

Símbolos inútiles=N - Símbolos útiles

Símbolos inaccesibles

Símbolos a los que no se llega a través del símbolo inicial en un proceso de derivación Son símbolos accesibles: $S \rightarrow^* \alpha$ a A β ; α , $\beta \in (\Sigma|N)^*$; a, A y S son accesibles Símbolos inaccesibles = $(\Sigma|N)$ - Símbolos accesibles

Proceso para el cálculo símbolos útiles:

- a) El símbolo de la parte izquierda de una derivación directa $A \to w$; $w \in \Sigma^*$; $A \in S$ útil
- b) Sí todos los símbolos de la parte derecha de una producción son útiles , el símbolo no terminal de la parte izquierda también lo es $B \to CcA$; si A y C son útiles, B también lo es
- c) Repetir proceso anterior hasta que se estén añadiendo símbolos útiles Quitar producciones con símbolos inútiles

Ejemplo: quitar los símbolos inútiles en la siguiente gramática:

$$G=(\Sigma=\{a, b, e\}, N=\{S, A, B, C, E\}, S, P\}$$

P: S \rightarrow ABaC| SB| ϵ

 $A \rightarrow aAB \mid bB$

 $B \rightarrow Ea \mid bB \mid \epsilon$

 $E \rightarrow Ee$

Aplicando el algoritmo, son símbolos útiles={S, B, A}, quitando las producciones con los símbolos inútiles queda la siguiente gramática: $S \rightarrow SB|_{\epsilon} A \rightarrow aAB| bB B \rightarrow bB|_{\epsilon}$

Proceso para el cálculo de símbolos accesibles:

- a) ∃ símbolo inicial es accesible (definido en la gramática) S→w; S es accesible
- b) Los símbolos que tengan como parte izquierda de la producción el símbolo inicial son accesibles S—AcB; A,c y B son accesibles
- c) Para los símbolos no terminales que son accesibles y son parte izquierda de una producción, los símbolos de su parte derecha son accesibles
 A→DeF; si A es accesible D, e y F son accesibles
- d) Repetir proceso anterior hasta que se estén añadiendo símbolos accesibles
- e) Quitar producciones con símbolos inaccesibles

Quitar los símbolos inaccesibles en la siguiente gramática: G = (Σ ={a, b}, N={S, A, B}, S, P} P: S \rightarrow SB| ϵ ; A \rightarrow aAB|bB; B \rightarrow bB| ϵ

Aplicando el algoritmo, son símbolos accesibles {S, B, b}, y quitando los símbolos inaccesibles queda la siguiente gramática:

G =
$$(\Sigma = \{b\}, N = \{S, B\}, S, P\}$$

P: $S \rightarrow SB|_{\epsilon}$,
B $\rightarrow bB|_{\epsilon}$
se trata de una gramática limpia (blanca)

Transformación de gramáticas de tipo 2

Quitar los prefijos comunes y la recursión inmediata por la izquierda

Prefijos comunes

Dos o más producciones tienen prefijos comunes cuando teniendo igual parte izquierda su parte derecha comienza por la misma subpalabra

$$A \rightarrow \delta \alpha_1 |\delta \alpha_2| \dots |\delta \alpha_n| \beta_1 |\beta_2| \dots |\beta_m; n \ge 2, m \ge 0, \delta \in (N|\Sigma)^+ \alpha_i, \beta_i \in (N|\Sigma)^*$$

Eliminación de prefijos comunes

$$A \rightarrow \delta A' | \beta_1 | \beta_2 | ... \beta m$$
, $A' \rightarrow \alpha_1 | \alpha_2 | | \alpha_n$

Ejemplo de gramáticas que generan el lenguaje L={ b an, n≥0}

Con prefijos comunes:

$$G_1$$
= (Σ = {a, b}, N={S, A}, S, P} P:

$$S \rightarrow bA \mid b$$
,

$$A \rightarrow aA \mid \epsilon$$

Sin prefijos comunes

$$G_2$$
= (Σ = {a, b}, N={S, R, A}, S, P}P:

$$S \rightarrow bR$$

$$R \rightarrow A|\epsilon$$
,

$$A \rightarrow aA \mid \epsilon$$

Recursión por la izquierda

Una o más producciones de una gramática tienen recursión por la izquierda, cuando existen uno o más símbolos de la parte derecha de la producción que coinciden con el símbolo no terminal de la parte izquierda

$$A \rightarrow \alpha_1 A \alpha_2 | \dots | \beta_1 | \beta_2 | \dots \beta_m \quad n > 0, m > 0 \quad \alpha_i \in (N|\Sigma)^+, \beta_i \in (N|\Sigma)^*$$

Recursión inmediata por la izquierda

Una o más producciones de una gramática tienen recursión inmediata por la izquierda, cuando el primer símbolo de la parte derecha de la producción coincide con el símbolo no terminal de la parte izquierda

$$A \rightarrow A\alpha_1 |A\alpha_2| \dots A\alpha_n |\beta_1|\beta_2| \dots \beta_m \quad \text{ n>0, m>0} \quad \alpha_i \in (N|\Sigma)^+, \ \beta_i \in (N|\Sigma)^*$$

Eliminación de la recursión inmediata por la izquierda

$$A \rightarrow \beta_1 A' | \beta_2 A' | \dots \beta_m A', A' \rightarrow \alpha_1 A' | \alpha_2 A' \dots \alpha_n A' | \epsilon$$

Ejemplo de gramáticas que generan el lenguaje L={ b a^{n,} n≥0}

Con recursión inmediata por la izquierda:

$$G_1=(\Sigma=\{a, b\}, N=\{A\}, A, P\} P: A \rightarrow Aa|b$$

Sin recursión inmediata por la izquierda:

$$G_2 = (\Sigma = \{a,b\}, N = \{A,A'\}, A, P)$$

 $A \rightarrow bA',$
 $A' \rightarrow aA' \mid \varepsilon$

Eliminación de producciones vacías

Los siguientes pasos eliminan las producciones vacías de gramáticas cuya parte izquierda es un símbolo no terminal.

- Determinar los símbolos anulables: un símbolo A es anulable si A→*ε
- Incluir en P' todas las producciones de P menos las producciones vacías
- Si A→αBβ está en P´y B es anulable, entonces añadir a P´ la producción A→αβ (excepto si es una producción vacía)
- Repetir el apartado anterior, hasta que no se pueden añadir más producciones.
- Si el símbolo inicial es anulable se añade una producción de la forma $S' \rightarrow \varepsilon |S|$

Ejemplo: Sean las producciones correspondiente a una gramática P:

$$S\!\to aSAb|\epsilon$$

$$A \rightarrow aAB|\epsilon$$

$$B \rightarrow bB|\epsilon$$

- P':
$$\{ S \rightarrow aSAb, A \rightarrow aAB, B \rightarrow bB \}$$

Eliminación de producciones unitarias (método de expansión)

Son producciones de la siguiente forma $A \rightarrow B$

Para su eliminación, habrá que realizar los siguientes pasos:

- Eliminar producciones vacías
- Para toda producción unitaria $A \rightarrow B$, con $B \rightarrow \alpha_1 | \alpha_2 | \dots | \alpha_n$; n > 0, $\alpha_i \in (N|\Sigma)^+$
 - Sustituir A \rightarrow B por A $\rightarrow \alpha_1 | \alpha_2 | \dots | \alpha_n$; n>0, $\alpha_i \in (N|\Sigma)^+$
- Repetir el paso anterior, mientras queden producciones unitarias

Ejemplo: Sea la siguiente gramática:

$$G = (\Sigma = \{a, b, c\}, N = \{S, A, B\}, S, P) \Leftrightarrow G = (\Sigma = \{a, b, c\}, N = \{S, A, B\}, S, P'')$$

$$P: S \rightarrow SaA|A$$

$$A \rightarrow AbB|B$$

$$B \rightarrow cB|c$$

$$A \rightarrow AbB|B$$

$$B \rightarrow cB|c$$
 $B \rightarrow cB|c$

$$P: S \rightarrow SaA|A$$
 $P: S \rightarrow SaA|AbB|B$ $P: S \rightarrow SaA|AbB|B|C$

$$A \rightarrow AbB|B$$
 $A \rightarrow AbB|B$ $A \rightarrow AbB|B|c$

$$B \rightarrow cB|c$$

Ejemplo 2: G = (
$$\Sigma$$
={a, b, c}, N={S, A, B}, S, P) S \rightarrow A | cS | ϵ A \rightarrow B | aA B \rightarrow S | bB

El método de expansión trabaja siempre y cuando no haya ciclos en las reglas, por ejemplo en: $S \rightarrow A$, $A \rightarrow B$, $B \rightarrow S$.

Para resolverlo, primero calculamos el conjunto de todos los **pares unitarios** de G, u(G) utilizando el siguiente algoritmo:

- Base: (A,A) es un par unitario para cualquier variable A. Esto es, A ⇒* A en cero pasos. u(G)={(A,A): A∈N}.
- Inducción: Suponemos que (A,B) ∈ u(G) y que B→C ∈ P donde C es un no terminal. Entonces añadimos (A,C) a u(G).

Calcular
$$u(G)$$
 en la gramática del ejemplo 2:
 $u(G) = \{(S,S),(S,A),(S,B),(A,A),(A,B),(A,S),(B,B),(B,S),(B,A)\}$

Eliminación de producciones unitarias (con recursividad)

Dada G = (Σ, N, S, P) , podemos construir G1 = (Σ, N, S, P') :

- 1. Encontrando el conjunto u(G) de todos los pares unitarios de G.
- 2. Para cada par unitario (A, B), añadimos a P1 todas las producciones $A \rightarrow \alpha$, donde $B \rightarrow \alpha$ es una producción no unitaria en P.
- 3. Note que es posible tener A = B; de esta manera, P1 contiene todas las producciones unitarias en P
- 4. $P1 = \{A \rightarrow \alpha : \alpha \in \mathbb{N}, B \rightarrow \alpha \in \mathbb{P}, (A, B) \in u(G)\}$

G = (
$$\Sigma$$
={a, b, c}, N={S, A, B}, S, P)
S \rightarrow A | cS | ϵ
A \rightarrow B | aA

 $B \rightarrow S \mid bB$



G1 = (
$$\Sigma$$
={a, b, c}, N={S, A, B}, S, P1)
S \rightarrow cS | aA | bB | ϵ
A \rightarrow aA | bB | cS | ϵ
B \rightarrow bB | cS | ϵ | aA

Forma Normal de Chomsky (FNC)

Todo LCL (sin ϵ) se genera por una gramática donde todas las producciones son de la forma: A \rightarrow BC o A \rightarrow a, donde A, B y C son no terminales, y a es un terminal.

A esto se le conoce como Forma Normal de Chomsky (FNC), y para llegar a ella debemos:

- 1. Limpiar la gramática, siguiendo los siguientes pasos (en este orden)
 - Eliminar las producciones- ϵ , es decir, producciones de la forma $A \rightarrow \epsilon$.
 - Eliminar producciones unitarias, es decir, producciones de la forma $A \rightarrow B$
 - Eliminar símbolos inútiles, aquellos que no aparecen en ninguna derivación
 - $S \rightarrow^* w$, para el símbolo de inicio S y la cadena de símbolos terminales w.
- 2. Hacemos que todos los cuerpos de longitud 2 o más consistan solo de no terminales
- 3. Dividimos los cuerpos de longitud 3 o más en una cascada de producciones con cuerpos de dos variables

Para el paso 2, por cada terminal a que aparece en una parte derecha de longitud \geq 2, creamos un nuevo no terminal, A, y reemplazamos la a por A en todos las partes derechas. Después añadimos una nueva regla A \rightarrow a.

Para el paso 3, por cada regla de la forma

$$A \rightarrow B^1B^2...B^k, k \ge 3,$$

introducimos variables nuevas C1, C2, ... Ck-2, y reemplazamos la regla con:

$$A \rightarrow B^1C^1$$

$$C^1 \rightarrow B^2C^2$$

. . .

$$C^{k-3} \rightarrow B^{k-2}C^{k-2}$$

$$C^{k-2} \rightarrow B^{k-1}B^k$$

Ejemplo

Con la gramática:

 $S \to ASB|\epsilon$

 $A \rightarrow aAS|a$

 $B \rightarrow SbS|A|bb$

(a) Eliminar producciones ε

(b) Eliminar producciones unitarias

(c) Eliminar símbolos inútiles

(d) ponerla en CNF

a) Solo S es anulable, por lo que tenemos que ponerla y quitarla en cada lugar donde ocurre:

 $\mathsf{S} \to \mathsf{ASB}|\mathsf{AB}$

 $A \rightarrow aAS|aA|a$

 $B \rightarrow SbS|Sb|bS|b|A|bb$

b) La única producción unitaria es: B → A por lo que la reemplazamos directamente:

 $S \rightarrow ASB|AB$

 $A \rightarrow aAS|aA|a$

B → SbS|Sb|bS|b|aAS|aA|a|bb

- c) A y B generan símbolos terminales, y por lo tanto también S, por lo que no hay símbolos inútiles
- d) Introducir variables y producciones $C \rightarrow a$ y $D \rightarrow b$ y ponerla en todas las partes derechas que no tienen un solo símbolo terminal

 $S \rightarrow ASB \mid AB$

 $A \rightarrow CAS \mid CA \mid a$

 $B \rightarrow SDS|SD|DS|b|CAS|CA|a|DD$

 $C \rightarrow a, D \rightarrow b$

Para las producciones con más de 3 símbolos se introducen nuevas variables:

 $S \rightarrow AE \mid AB$

 $A \rightarrow CF | CA | a$

 $B \rightarrow SG|SD|DS|b|CF|CA|a|DD$

 $C \rightarrow a, D \rightarrow b$

 $E \rightarrow SB, F \rightarrow AS$

 $G \rightarrow DS$

5.4 Lenguajes independientes del contexto – Lema de bombeo

Lema de bombeo para lenguajes independientes del contexto

Sea L un LIC. Existe una constante n tal que si z es cualquier cadena de L tal que |z| es al menos |n|, podemos escribir z = uvwxy tal que:

- 1. $|vwx| \leq n$
- 2. $vx \neq \varepsilon$
- 3. $\forall i \geq 0$, $uv^i w x^i y \in L$

5.4 Lenguajes independientes del contexto – Lema de bombeo

- Es muy parecido al de lenguajes regulares solo que ahora tenemos que encontrar dos subcadenas que podamos repetir (bombear) indefinidamente.
- La idea de demostración viene de que todo LLC lo podemos representar como FNC y generar árboles de parseo binarios
- Mientras que los LLC pueden servir para dos grupos de caracteres, no sirven para tres (e.g., $\{0^n1^n2^n \mid n \ge 1\}$)
- Los LLC tampoco sirven si tenemos pares de símbolos intercalados (e.g., $\{0^i1^i2^i3^j\mid i\geq 1, j\geq 1\}$)
- Tampoco pueden verificar la igualdad de dos cadenas arbitrariamente largas (e.g., $ww \mid w \in \{0,1\}^*$)

5.4 Lenguajes independientes del contexto – Lema de bombeo

Ejemplos

• $L = \{0^n 1^n 2^n | n \ge 1\}$

Idea: z = uvwxy, $|vwx| \le n$ (lema), vwx no puede tener al mismo tiempo 0's y 2's porque el último 0 y el primer 2 están separados por n+1 posiciones. Si vwx no tiene 2's (o 0's) con el bombeo, por ejemplo, de i=0 lo desbalanceamos.

• $L = \{0^i 1^i 2^i 3^j | i \ge 1, j \ge 1\}$

Con un argumento similar podemos tener que la cadena no tiene todos los símbolos y desbalancear la cadena.

• $L = \{ww \mid w \in \{0, 1\}^*\}$

En este caso, podemos poner como ejemplo: $0^n 1^n 0^n 1^n$, la cual es una cadena ww y mostrar, tomando las ideas de los ejemplos anteriores, que podemos generar una cadena que no está en L.

5.5 Lenguajes independientes del contexto – Propiedades de cerradura

Propiedades de cerradura de los LLCs

- Existen varias propiedades de los LLC, una de las más importantes es la de sustitución.
- Sustitución: Sea Σ un alfabeto y supongamos que para cada símbolo a en Σ definimos un lenguaje arbitrario L_a
- Estos lenguajes definen una función s o una sustitución.
- Si $w = a_1 a_2 \dots a_n$ es una cadena en Σ^* , s(w) es la concatenación de los lenguajes $s(a_1)s(a_2) \dots s(a_n)$.

Ejemplo

$$\Sigma = \{0, 1\} \text{ y } s(0) = \{a^n b^n : n \ge 1\}, s(1) = \{aa, bb\}$$

Sea w = 01. Calcula s(w)

•
$$s(w) = s(0)s(1) = \{a^nb^naa : n \ge 1\} \cup \{a^nb^{n+2} : n \ge 1\}.$$

Si $L = L(0^*)$, calcula s(L)

•
$$s(L) = (s(0))^* = \{a^{n_1}b^{n_1}a^{n_2}b^{n_2} \dots a^{n_k}b^{n_k}: k \ge 0, n_i \ge 1\}$$

5.5 Lenguajes independientes del contexto – Propiedades de cerradura

Teoremas

Sea L un LLC sobre Σ , y s una sustitución, tal que s(a) sea un LLC, $\forall a \in \Sigma$. Entonces s(L) es un LLC.

Si tenemos uno o más LLC's, también son LLC el resultado de hacer:

- (i) Unión: La unión $L = L_1 \cup L_2$ es un lenguaje de contexto libre, ya que la gramática $G = (\Sigma_i, N_i, S_i, P_i)$ con $\Sigma = \Sigma_1 \cup \Sigma_2$, $N = N_1 \cup N_2$, $P = P_1 \cup P_2 \cup \{S \to S_1 | S_2\}$, siendo S una nueva variable, que satisface que L = L(G)
- (ii) Concatenación: La concatenación $L = L_1L_2$ es un LLC. Se prueba igual que antes, siendo ahora $P = P_1 \cup P_2 \cup \{S \to S_1S_2\}$.
- (iii) Cerradura de Kleene: $L = L_1^*$ es un LLC. Basta considerar el conjunto de producciones $P = P_1 \cup \{S \to SS_1 \mid \epsilon\}$, $donde S \notin N1$.
- (iv) cerradura positiva +
- (v) Inversión
- (vi) homomorfismo, y homomorfismo inverso.

5.5 Lenguajes independientes del contexto – Propiedades de cerradura

Intersección:

La intersección $L=L_1\cap L_2$ no es, en general, un LLC. Por ejemplo, los lenguajes $L_1=\{a^nb^nc^m:n,m\geq 0\}$ y $L_2=\{a^nb^mc^m:n,m\geq 0\}$ son ambos de contexto libre, pero $L=L_1\cap L_2=\{a^nb^nc^n:n\geq 0\}$ no lo es

Complementario:

El complemento $L=L^c$ no es, en general, un LLC. Esto se sigue teniendo en cuenta que $L_1\cap L_2=(L_1^c\cup L_2^c)^c$, ya que si el complementario fuera siempre LLC, también lo sería la intersección de cualesquiera dos lenguajes de contexto libre.

Diferencia:

La diferencia $L=L_1-L_2$ no es en general un LCL. Esto es debido a que $L=L_1\cap L_2^c$.

Intersección con lenguaje regular

Si R es un lenguaje regular, entonces $L = L_1 \cap R$ es siempre un LLC.

5.5 Lenguajes independientes del contexto – Problemas de decisión

Problemas de decisión para LLC

Problema de infinitud. Dada una gramática G, resolver la pregunta ¿es L(G) infinito? *Ejemplo 1*

$$S \to AB$$

$$A \to BC \mid a$$

$$B \to CC \mid b$$

$$C \to a$$

El lenguaje generado por la gramática es {ab, aaa, bab, baaa, aaab, aaaaa} que es finito Ejemplo 2

$$S \rightarrow BA \mid BC$$

$$A \rightarrow SC$$

$$B \rightarrow a$$

$$C \rightarrow b$$

El lenguaje generado por esta gramática es infinito

Decisión: Si el árbol de parseo de la gramática, en forma FNC, tiene un ciclo, el lenguaje es infinito

5.5 Lenguajes independientes del contexto – Problemas de decisión

Lenguaje vacío

Dado un LCL L, ¿es L vacío? Si L=L(G) basta con simplificar la gramática eliminando los símbolos inútiles,. Si el símbolo inicial es inútil, entonces $L=\emptyset$, en otro caso $L\neq 0$

Problemas no-decidibles

Hay problemas que no los podemos resolver con una computadora (no-decidibles) Por ejemplo:

- ¿Es una GLC dada ambigua?
- Es un LLC dado inheremente ambiguo?
- ¿Es la intersección de dos GLCs vacía?
- ¿Son dos GLCs iguales?
- ¿Es un LLC dado igual a Σ^* donde Σ es el alfabeto de ese lenguaje?

Probar membresía en un LLC: $\lambda w \in L$?

- Para probar si una cadena w es parte del lenguaje (L) de una LLC (i.e., $w \in L$), usar el algoritmo CYK (Cocke, Younger, Kasami), que soluciona este problema en $O(|w|^3)$
- $lue{}$ Se construye una matriz triangular inferior de $n \times n$
- ☐ La gramática debe estar en FNC
- \square El eje x corresponde a la cadena a probar $w=a_1,a_2,a_3,\ldots,a_n$
- \Box En el primer renglón se ponen producciones tipo $A \rightarrow a$
- En el resto (e.g., X_{ij}) se ponen los no terminales (e.g., A) tales que tengan una producción (e.g., $A \to BC$) donde B está en X_{ik} (parte del prefijo) y C está en X_{kj} (resto de la cadena), i < k < j.

$$X_{15}$$
 X_{14} X_{25}
 X_{13} X_{24} X_{35}
 X_{12} X_{23} X_{34} X_{45}
 X_{11} X_{22} X_{33} X_{44} X_{55}

Ejemplo

Probar que la siguiente gramática genera: baaba

$$S \rightarrow AB|BC$$

$$A \rightarrow BA|a$$

$$B \rightarrow CC|b$$

$$C \rightarrow AB|a$$

$$b: \{B\}$$
$$a: \{A,C\}$$

$$ba: \{B\}\{A,C\} = \{BA,BC\}: \{S,A\}$$

 $aa: \{A,C\}\{A,C\} = \{AA,AC,CA,CC\}: \{B\}$
 $ab: \{A,C\}\{B\} = \{AB,CB\}: \{S,C\}$
 $ba: \{B\}\{A,C\} = \{BA,BC\}: \{S,A\}$

```
bab : b/ab \ o \ ba/b : \{B\}\{B\} \ \cup \ \{S,A\}\{A,C\} = \{BB,SA,SC,AA,AC\} : \emptyset
aab : a/ab \ o \ aa/b : \{A,C\}\{S,C\} \ \cup \ \{B\}\{B\} = \{AS,AC,CS,CC,BB\} : \{B\}
aba : a/ba \ o \ ab/a : \{A,C\}\{S,A\} \ \cup \ \{S,C\}\{A,C\} = \{AS,AA,CS,CA,SA,SC,CA,CC\} : \{B\}
baab : b/aab \ o \ ba/ab \ o \ baa/b : \{B\}\{B\} \ \cup \ \{S,A\}\{S,C\} \ \cup \ \emptyset\{B\} = \{BB,SS,SC,AS,AC\}
: \emptyset
aaba : a/aba \ o \ aa/ba \ o \ aab/a : \{A,C\}\{B\} \ \cup \ \{B\}\{S,A\} \ \cup \ \{B\}\{A,C\}
= \{AB,AC,BS,BA,BA,BC\} : \{S,A,C\}
```

$$baaba: b/aaba o ba/aba o baa/ba o baab/a:$$
 $\{B\}\{S,A,C\} \cup \{S,A\}\{B\} \cup \emptyset \cup \emptyset = \{BS,BA,BC,SB,AB\}: \{S,A,C\}$

$$\{S,A,C\}$$

$$- \{S,A,C\}$$

$$- \{B\} \{B\}$$

$$\{S,A\} \{B\} \{S,C\} \{S,A\}$$

$$\{S,A\} \{A,C\} \{A,C\} \{B\} \{A,C\}$$

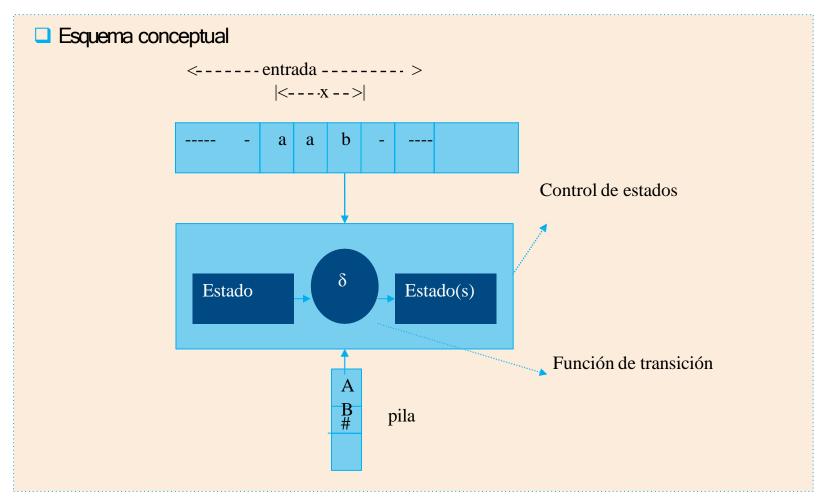
$$S \to AB|BC$$

$$A \to BA|a$$

$$B \to CC|b$$

$$C \to AB|a$$

Tema 5 - Autómata de pila



Lenguajes independientes del contexto

Tema 5 Definición formal de un AP

Definición formal AP

Un AP se define mediante una septupla AP=(Σ , Q, Γ , δ , q0, #, F), donde:

- Σ : alfabeto de las palabras a reconocer por el autómata
- Q: conjunto de estados, es un conjunto finito no vacío
- Γ: alfabeto de los símbolos de la pila
- δ : Q x Σ U {ε ,\$} x Γ → Px Γ*, en donde P⊆ Q
 función de transición, que transforma la terna ordenada (q_i, a, A) → (q_i, α)
- $q_0 \in Q$, estado inicial (estado de partida)
- # : símbolo de la base de la pila
- F ⊆ Q: conjunto de estados finales o de aceptación
- Significación de una función de transición (q_i, a, A) → (q_i, BB)
 - q_i estado actual del autómata
 - a símbolo leído de la entrada, puede que no se lea de la entrada ε
 - A elemento que se quita de la pila, puede que no se quite nada de la pila ε
 - qi estado destino del autómata
 - BB elementos a poner en la pila. Puede que no se ponga nada en la pila ϵ

Tema 5 Construcción de un AP

☐ Ejemplo de construcción de un AP

Dado el lenguaje $L = \{a^n b^n \mid n > 0\}$, especificar un AP que lo reconozca:

- Hay que buscar una lógica que simule la entrada a través de la pila:

Apilar las aes mientras que vengan, desapilar las aes cuando llegan las bes, cuando llegue el final de palabra no habrá elementos en la pila

AP= (
$$\Sigma$$
={a, b}, Q={ q₀, q₁, q₂}, Γ ={A, #}, δ , q₀, #, F={ q₂})

- $\delta(q_0, a, \#) \rightarrow (q_0, A\#)$ lee a , desapila # y apila A#, no se cambia de estado
- δ (q₀, a, A) \rightarrow (q₀, AA) lee a, desapila A y apila AA, no se cambia de estado
- $\delta (q_0, b, A) \rightarrow (q_1, \epsilon)$ lee b, desapila A y no apila nada, se cambia de estado
- δ (q₁, b, A) \rightarrow (q₁, ϵ) lee b, desapila A y no apila nada, no se cambia de estado
- $\delta (q_1, \$, \#) \rightarrow (q_2, \epsilon)$ lee \$, desapila # y no apila nada, se cambia de estado

La palabra ha finalizado - la pila está vacía y estamos en un estado final

☐ Representación gráfica de los AP

Para mejorar la legibilidad y visibilidad de los autómatas se pueden utilizar dos tipos de representaciones gráficas: diagramas o las tablas de transición:

Tema 5 Diagrama de transición de un AP

□ Diagramas de transición

La representación de diagrama de transiciones, es un grafo dirigido, en el que:

- Los nodos se representa por un círculo, que se etiquetan con los estados $q \in Q$
- Los arcos son etiquetados con el símbolo de lectura, una coma que lo separa del símbolo que se desapila, un punto y coma que lo separa de los símbolos que se quieren apilar
- así desde el nodo qi al qj, si existe la función $\delta(qi, a, A) \rightarrow (qj, BB)$. El arco será a, A; BB
- □ estado inicial tiene un arco entrante no etiquetado →
- Los estados finales $q_f \in F$ se señalan por doble círculo, en los AP puede que no existan estados finales
- □ Ejemplo de AP representado por un diagrama de transición, que reconoce el lenguaje :



Tema 5 Tabla de transición de un AP

□ Tablas de transición

Representación tabular que relaciona los componentes: entradas, estados y pila

- Cada fila corresponde a un estado q ∈ Q
- Cada columna corresponde símbolo de la entrada $a \in \Sigma$ y símbolo de la cima de pila $A \in \Gamma$
- La fila asociada al estado inicial tiene una →
- Las filas asociadas a los estados finales de existir tienen el símbolo * En la posición (q, (a, A)), $q \in Q$, $a \in \Sigma$ y $A \in \Gamma$ está el estado y los elementos a apilar que determine la función $\delta(q, a, A) \to (p, \alpha)$ en este caso (p, α) $p \in Q$ y $\alpha \in \Gamma^*$

AFD representado por la tabla de transición que reconoce el lenguaje: L={a nbm | n,m>0}

estados	a, A	a,#	b,A	b,#	\$,#	ε,Β	Α, ε	Β, ε
$\rightarrow q_0$	$q_{0,AA}$	q _{0, A#}	q _{1,} ε					
q_1			q_1 , ϵ		q _{2,} ε			
*q ₂								

Tema 5 Lenguaje reconocido por un AP

Lenguaje reconocido por un autómata de pila

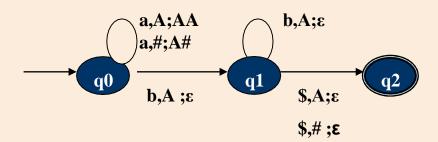
El lenguaje que reconoce un AP será: el conjunto de palabras que partiendo del una situación inicial, se llega a una situación final en cero o más movimientos.

En los AP existen dos formas de reconocimiento de los lenguajes:

- Reconocimiento por llegada a estado final

$$\label{eq:lenguage} \begin{split} &\text{Lenguaje=}\{\,x\mid x\,\in\!\Sigma^*,\quad (q_i\,,\,x,\,\#)\!\!\to^*\!\!(q_f,\,\epsilon,\,\alpha),\ q_f\in\!\!F,\alpha\in\Gamma^*\!\!\}\quad AP\\ &\text{Ejemplo: L=}\,\{a^n\,b^m\mid\ n\!\!>=\!\!m;\,n,\,m\!\!>\!\!0\} \end{split}$$

$$\begin{aligned} \mathsf{AF} &= (\Sigma = \{\mathsf{a},\,\mathsf{b}\},\,\mathsf{Q} = \{\,\mathsf{q}_0,\,\mathsf{q}_1,\,\mathsf{q}_2\}\,\,\Gamma = \{\mathsf{A},\,\#\},\,\delta,\,\mathsf{q}_0,\,\#,\,\,\mathsf{F} = \{\,\mathsf{q}_2\}),\\ \delta(\mathsf{q}_0,\,\mathsf{a},\,\#) &\to (\mathsf{q}_0,\,\mathsf{A}\#)\,,\,\delta(\mathsf{q}_0,\,\mathsf{a},\,\mathsf{A}) \to (\mathsf{q}_0,\,\mathsf{AA}),\,\,\delta(\mathsf{q}_0,\,\mathsf{b},\,\mathsf{A}) \to (\mathsf{q}_1,\,\epsilon)\\ \delta(\mathsf{q}_1,\,\mathsf{b},\,\mathsf{A}) &\to (\mathsf{q}_1,\,\epsilon),\,\,\delta(\mathsf{q}_1,\,\$,\,\mathsf{a}) \to (\mathsf{q}_2,\,\epsilon) \end{aligned}$$



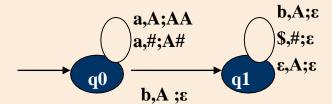
Tema 5 Lenguaje reconocido por AP por vaciado de pila

Reconocimiento por llegada vaciado de pila

$$\begin{split} & \text{Lenguaje=}\{\;x\;|\;\;x\in \Sigma^*,\quad (q_i\;,\,x,\,\#) {\longrightarrow}^* (q_j,\,\epsilon,\,\epsilon)\;\;\} \quad \text{AP} \\ & \text{L=}\; \{a^nb^m\,|\;\;n{>}\text{=}m;\;n,\,m{>}0\} \end{split}$$

$$\begin{split} AF &= (\Sigma = \{a, b\}, \, Q = \{ \, q_0, \, q_1 \} \, \, \Gamma = \{A, \#\}, \, \delta, \, q_0, \#, \, F = \{q_1\}), \\ \delta \, (q_0, \, a, \#) \, &\rightarrow (q_0, \, A\#) \,, \, \delta \, (q_0, \, a, \, A) \rightarrow (q_0, \, AA), \, \, \delta (q_0, \, b, \, A) \rightarrow (q_1, \, \epsilon) \\ \delta \, (q_1, \, b, \, A) \rightarrow (q_1, \, \epsilon), \, \, \delta \, (q_1, \, \$, \, \#) \, &\rightarrow (q_1, \, \epsilon) \end{split}$$

$$\begin{aligned} (q_0, x = aab\$, \#) &\rightarrow (q_0, ab\$, A\#) \rightarrow (q_0, b\$, AA\#) \rightarrow (q_1,\$, A\#) \\ &\rightarrow (q_1, \epsilon, \#) \rightarrow (q_1, \epsilon, \epsilon) \end{aligned}$$



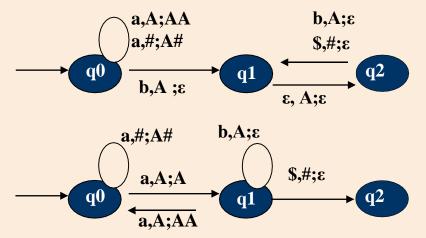
Tema 5 Autómatas de pila deterministas - APD

Autómatas de pila deterministas APD

Un autómata de pila es determinista APD cuando para cualquier situación sólo puede darse como mucho un movimiento, para ello debe cumplir las siguientes condiciones:

$$\forall$$
 q \in Q , \forall A \in Γ y a \in Σ | (qi, a, A)| <2
 \forall q \in Q y \forall A \in Γ | (qi, ϵ , A)| <0
 \forall q \in Q y \forall a \in Σ | (qi, a, ϵ)| <0

Ejemplo: L= $\{a^{2n} b^n | n>0\}$, APD por vaciado de pila que reconozca el anterior



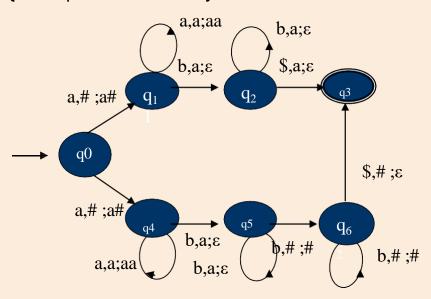
Tema 5 Autómatas de pila no deterministas - APND

Autómatas de pila no deterministas APND

Un autómata de pila es no determinista APD cuando existe alguna situación que puede darse mas de un movimiento.

Se pueden dar tres situaciones de no determinismo que veremos a continuación:

1) $(qi, a, a) \rightarrow \{(qj, \alpha 1) (qk, \alpha 2),\}$ Ejemplo: L= $\{a^n b^m \mid n <> m; n, m > 0\}$



Lenguajes independientes del contexto

Tema 5 Autómatas de pila no determinista - APND

APND

2)
$$(q_i, a, a) \rightarrow (...)$$

 $(q_i, \epsilon, a) \rightarrow (...)$

Ejemplo: $L = \{a^nb^m \mid n > = m; n, m > = 1\}$

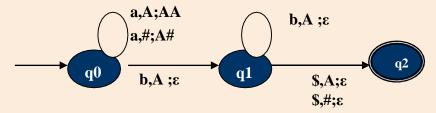
a) reconocimiento por vaciado de pila



$$(q0, aab\$, \#) \rightarrow (q0, ab\$, a\#) \rightarrow (q0, b\$, aa\#) \rightarrow (q1,\$, a\#) \rightarrow (q1, \epsilon, \#) \rightarrow (q1, \epsilon, \epsilon)$$

 $(q0, aaabb\$, \#) \rightarrow (q0, aabb\$, a\#) \rightarrow (q0, bb\$, aaa\#) \rightarrow (q1, b\$, aa\#) \rightarrow (q1, b\$, \#)$

b) reconocimiento por llegada a estado final



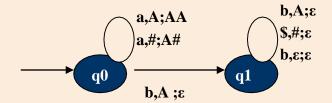
Lenguajes independientes del contexto

Tema 5 Autómatas de pila no deterministas - APND

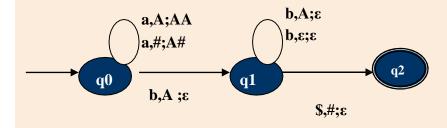
3)
$$(qi, a, a) \rightarrow (....)$$

 $(qi, a, \epsilon) \rightarrow (....)$
Ejemplo: L= $\{a^n b^m \mid n \le m, n, m \ge 1\}$

a) reconocimiento por vaciado de pila



b) reconocimiento por llegada a estado final



Tema 5 Gramáticas independientes del contexto transformación GIC->AP

Transformación de gramática independiente del contexto a AP

Dada la gramática
$$G = (\Sigma = \{a, b\}, N = \{S, A\}, S, P)$$

 $P: S \rightarrow aA A \rightarrow Sb|b$

construir un AP que reconozca el lenguaje que genera la gramática G.

Proceso de transformación:

- 1) Poner el símbolo inicial en la pila (ε,#;S#)
- 2) Para cada producción $A\rightarrow\alpha$ sustituir parte izquierda por parte derecha (ϵ,A ; α)
- 3) Para cada símbolo del alfabeto de la gramática $\forall a \in \Sigma$ poner $(a,a;\epsilon)$
- 4) Poner la transición para que el reconocimiento sea por vaciado de pila (\$,#;ε)

$$q_0,aabb\$,\#) \rightarrow (q_1,aabb\$,S\#) \rightarrow (q_1,aabb\$,aA\#) \rightarrow (q_1,abb\$,A\#) \rightarrow (q_1,abb\$,AB\#) \rightarrow (q_1,abb\$,aAb\#) \rightarrow (q_1,bb\$,Ab\#) \rightarrow (q_1,bb\$,bb\#) \rightarrow (q_1,b\$,b\#) \rightarrow (q_1,s,\#) \rightarrow (q_1,s,E)$$

Tema 5 – Ejercicios de gramáticas independientes del contexto

Obtener una gramática y autómata de pila para cada uno de los siguientes lenguajes:

- 1. $\Sigma = \{a, b, c\}, L = \{a^n b^m c^n | n, m > = 1\}$
- 2. $\Sigma = \{a, b, c\}, L = \{a^n b^m c^{n+2m} | n \ge 0, m \ge 1\}$
- 3. $\Sigma = \{a, b, c\}, L = \{a^n b^{n+m} c^m | n > = 1, m > = 0\}$
- 4. $\Sigma = \{a, b\}, L = \{a^n b^m \mid n > = m > = 1\}$
- 5. $\Sigma = \{a, b\}, L = \{a^n b^m \mid n > m > = 0\}$
- 6. $\Sigma = \{a, b, c\}, L = \{a^m b^n c^k \mid m > n + k; n, k > = 0\}$
- 7. $\Sigma = \{a, b\}, L = \{a^nb^m \mid m \le n \le 2m; n, m \ge 0\}$
- 8. $\Sigma = \{a, b\}, L = \{a^nb^m \mid n <> m; n, m > 0\}$
- 9 L= $\{v \cdot w \mid v \in (ab)^+, con w \in (a \mid b)^+ y con |v| \le |w| \}$
- 10. $\Sigma = \{a, b, c\}$, L= $\{a^n b^m c^k \mid n \in S \text{ un número par, posiblemente el } 0, k \in S \text{ un número impar y } m > k+n \}$
- 11. $\Sigma = \{a, b\}, L = \{a^{2n} b^n | n > 0\} U \{a^n b^{2n} | n > 0\}.$
- 12. $\Sigma = \{a, b, c\}, L = \{x c y \mid x, y \in \{a, b\}^* \text{ en donde } |x| = 2 |y| \}$

Tema 5 - Gramáticas independientes del contexto - Ejercicios

Sea el alfabeto Σ = {a, b} se considera el lenguaje de las palabras que tienen al menos dos símbolos "b" y que cumplen la propiedad de que el número de símbolos "a" que hay antes de la primera "b" es estrictamente mayor que el número de símbolos "a" que hay después de la última "b".

Por ejemplo, pertenecen a este lenguaje: aaabbaa, aabb, abbb, aabababaaba.

No pertenecen: abaabaa, baaba, bab.

- 14. Dado el alfabeto Σ = {a, b} sea el lenguaje: L = { vbbw | v, w \in {a,b}* y |v| >= |w| (la longitud de v \in mayor o igual que la de w)}
- 15. Dado el alfabeto $\Sigma = \{a, b\}$ y, sea el lenguaje $L = \{vw \mid en donde v, w \in \{a,b\}^* y \mid v \mid = \mid w \mid \}$.
- 16. Con el alfabeto $\Sigma = \{a,b,c\}$ se define el lenguaje $L = \{w \cdot c^n \mid w \text{ pertenece a } \{a,b\}^+, n>0\}$ en el que además se cumple que:
 - a. si w termina en a, entonces la longitud de w es mayor o igual que n
 - b. si w termina en b, entonces la longitud de w es menor o igual que n
 - Ejemplos de palabras del lenguaje: abbacc, aaaccc, abccc, bbbabccccc
 - Ejemplos de palabras que <u>no</u> son del lenguaje: abbacccc, aaacccccc, abc, bbbabccc

Tema 5 - Lenguajes independientes del contexto - Ejercicios

- Dado el alfabeto $\Sigma = \{a, b, 0, 1\}$, sea el lenguaje $L = \{w \cdot (0|1) \cdot v \mid v, w \text{ pertenecen}\}$
 - a $\{a,b\}^+$ y $|w|=|v|\}$, en el que además se cumplen las siguientes restricciones:
 - a. Si el dígito central es 0, w deberá empezar por a
 - b. Pero si el dígito central es 1, entonces es v quien empezará por a.
 - Ejemplos de palabras válidas: aba0bba, a0a, a0b, a1a, b1a, abb1abb
 - Ejemplos de algunas palabras **no** válidas: b0a, ba0aa, a1b, ba1ba
- 18. Dado el alfabeto $\Sigma = \{a, b, c\}$, sea el lenguaje $L = \{wc^n \mid w \in \{a, b\}^* ; n = número de subcadenas ab de w ; n>= 0\}$
 - Es decir, el número de símbolos c del final coincide con el número de subcadenas ab que tiene la palabra al principio, es decir, en w (teniendo en cuenta que en w puede haber cualquier combinación de símbolos a y b)

Tema 5 - Lenguajes independientes del contexto - Ejercicios

Convierte la siguiente gramática en Forma Normal de Chomsky (FNC) $G = (\Sigma = \{a, b, 0, 1, *, +, (,)\}, N = \{E, T, F, I\}, E, P), donde P:$

$$E \rightarrow T \mid E+T$$

 $T \rightarrow F \mid T^*F$
 $F \rightarrow I \mid (E)$
 $I \rightarrow a \mid b \mid Ia \mid Ib \mid I0 \mid I1$

- Mostrar que $\{a^i b^j c^k \mid i < j < k\}$ no es LCL
- 19. Probar que: $\{a^nb^nc^i \mid i \leq n\}$ no es LCL.
- Probar que: $\{a^{n^2} | n \ge 1\}$ no es LCL.

Tema 5 - Lenguajes independientes del contexto - Ejercicios

^{21.} Dada la siguiente gramática libre de contexto en FNC:

$$S \rightarrow AB \mid BC$$

$$A \rightarrow BA \mid a$$

$$B \rightarrow CC \mid b$$

$$C \rightarrow AB \mid a$$

Comprueba si las siguientes palabras pertenecen al lenguaje utilizando el algoritmo CYK

- w = abb
- w = aaba
- w = baba
- w = abaa

Tema 5 - Ejercicios de autómatas de pila

1. Justificar si las siguientes transiciones de un AP suponen no determinismo: $(q_0, a, b) = (q_2, ab), (q_0, a, \varepsilon) = (q_1, a)$

2. Sea el siguiente autómata a pila con reconocimiento por vaciado de pila:

$$\delta(q_0, a, \#) = (q_0, a\#), \delta(q_0, a, a) = (q_0, aa), \delta(q_0, b, a) = \{ (q_1, \epsilon), (q_2, \epsilon) \}$$

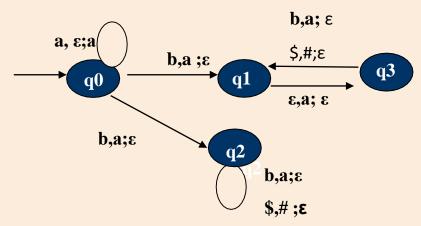
$$\delta(q_1, b, a) = (q_1, \epsilon), \delta(q_1, \$, \#) = (q_1, \epsilon)$$

$$\delta(q_2, \varepsilon, a) = (q_3, \varepsilon)$$

$$\delta(q_3, b, a) = (q_2, \epsilon), \delta(q_3, \$, \#) = (q_3, \epsilon)$$

Determinar el lenguaje reconocido por el mismo.

3. Dado el siguiente AP₁ que realiza el reconocimiento por vaciado de pila:



Se pide: Obtener el lenguaje reconocido por cada uno de ellos

Tema 5 – Ejercicios de autómatas de pila

4. Dado el alfabeto $\Sigma = \{a, b\}$, sea el lenguaje $L = \{a^n b^n \mid n > = 0\}$.

Encontrar un autómata de pila que reconozca el lenguaje <u>complementario de L</u>. Considerar que todas las palabras vienen finalizadas por el símbolo \$.

5. Dado el alfabeto Σ = {a, b}, sea el lenguaje L= { aⁿb^m | n > m; n, m > 0 }. Encontrar un autómata a pila (AP) que reconozca L* (L estrella).

Se pide:

Sin aplicar ningún algoritmo, es decir, partiendo de la definición del lenguaje, obtener el diagrama de transiciones de un autómata a pila (AP) que lo reconozca.