Gramáticas regulares

Universidad Nacional de Quilmes

Emanuel Alvaredo, Federico Lochbaum

Gramáticas

Como sabemos, las gramáticas especifican lenguajes (similar a AF y ER)

Ejemplo:

$$\langle oración \rangle \rightarrow \langle sujeto \rangle \ \langle predicado \rangle$$

$$\langle sujeto \rangle \rightarrow \langle artículo \rangle \ \langle nombre \rangle$$

$$\langle predicado \rangle \rightarrow \langle verbo \rangle$$

$$\langle artículo \rangle \rightarrow un$$

$$\langle artículo \rangle \rightarrow el$$

$$\langle nombre \rangle \rightarrow gato$$

$$\langle nombre \rangle \rightarrow perro$$

$$\langle verbo \rangle \rightarrow corre$$

$$\langle verbo \rangle \rightarrow salta$$

Una posible *derivación* de la gramática: “el perro corre”

$$
\begin{aligned}
\langle \textit{oración} \rangle &\Rightarrow \langle \textit{sujeto} \rangle \ \langle \textit{predicado} \rangle \\
&\Rightarrow \langle \textit{sujeto} \rangle \ \langle \textit{verbo} \rangle \\
&\Rightarrow \langle \textit{artículo} \rangle \ \langle \textit{nombre} \rangle \ \langle \textit{verbo} \rangle \\
&\Rightarrow \textit{el} \ \langle \textit{nombre} \rangle \ \langle \textit{verbo} \rangle \\
&\Rightarrow \textit{el} \ \ \textit{perro} \ \langle \textit{verbo} \rangle \\
&\Rightarrow \textit{el} \ \ \textit{perro} \ \ \textit{corre}
\end{aligned}
$$

“Se reescribe”

Una posible *derivación* de la gramática: “un gato salta”

$$
\begin{aligned}
\langle \textit{oración} \rangle &\Rightarrow \langle \textit{sujeto} \rangle \ \langle \textit{predicado} \rangle \\
&\Rightarrow \langle \textit{sujeto} \rangle \ \langle \textit{verbo} \rangle \\
&\Rightarrow \langle \textit{artículo} \rangle \ \langle \textit{nombre} \rangle \ \langle \textit{verbo} \rangle \\
&\Rightarrow \textit{un} \ \langle \textit{nombre} \rangle \ \langle \textit{verbo} \rangle \\
&\Rightarrow \textit{un} \ \ \textit{gato} \ \langle \textit{verbo} \rangle \\
&\Rightarrow \textit{un} \ \ \textit{gato} \ \ \textit{salta}
\end{aligned}
$$

“Se reescribe”

$$\langle oración \rangle \rightarrow \langle sujeto \rangle \ \langle predicado \rangle$$

$$\langle sujeto \rangle \rightarrow \langle artículo \rangle \ \langle nombre \rangle$$

$$\langle predicado \rangle \rightarrow \langle verbo \rangle$$

¿Cuál es su lenguaje?

$$\langle artículo \rangle \rightarrow un$$

$$\langle artículo \rangle \rightarrow el$$

$$\langle nombre \rangle \rightarrow gato$$

$$\langle nombre \rangle \rightarrow perro$$

$$\langle verbo \rangle \rightarrow corre$$

$$\langle verbo \rangle \rightarrow salta$$

Cada posible derivación de la gramática

L = {

“un gato salta”, “un gato corre”, “el gato salta”,

“el gato corre”, “un perro salta”, “un perro corre”,

“el perro salta”, “el perro corre”

}

Notación

Este conjunto de definiciones se llaman *reglas de producción/reducción*

$$\langle nombre \rangle \to gato$$
$$\langle nombre \rangle \to perro$$

No terminal/Variable ***Terminal***

$$\langle oración \rangle \to \langle sujeto \rangle\ \langle predicado \rangle$$

No terminal/Variable

Ejemplo

Gramática

$$S \to aSb$$

$$S \to \lambda$$

Lenguaje

$$L = \{a^n b^n : n \geq 0\}$$

Derivación de la cadena "ab"

$$S \Rightarrow aSb \Rightarrow ab$$

(S reduce en dos pasos a ab)

$$S \to aSb \qquad\qquad S \to \lambda$$

Derivación de la cadena "aabb"

$$S \Rightarrow aSb \Rightarrow aaSbb \Rightarrow aabb$$

$$S \to aSb \qquad\qquad S \to \lambda$$

Definición formal de una gramática

$$G = (V, \Sigma, S, P)$$

V: Conjunto de no terminales

Σ: Conjunto de terminales

S: No terminal/variable **inicial**

P: Conjunto de reglas de producción/reescritura

Qué gramática acepta el lenguaje L ?

$$L = \{a^n b^n : n \geq 0\}$$

Piensenlo 5 minutos :)

Qué gramática acepta el lenguaje L ?

$$L = \{a^n b^n : n \geq 0\}$$

Damos **un** G **= <*V*, *Σ*, *S*, *P* >**

***V*:** { S } ***Σ*:** { a, b } ***S*:** S ***P*:** { S → aSb, S → λ }

Detalles...

- Un string que contiene No terminales/variables y terminales lo llamamos **Forma sentencial**

- Un string que contiene sólo terminales lo llamamos **Sentencia**

$$S \Rightarrow aSb \Rightarrow aaSbb \Rightarrow aaaSbbb \Rightarrow aaabbb$$

- **Escribimos** $S \overset{*}{\Rightarrow} aaabbb$ Para denotar $S \Rightarrow aSb \Rightarrow aaSbb \Rightarrow aaaSbbb \Rightarrow aaabbb$

Una gramática:

$$S \to Ab$$

$$A \to aAb$$

$$A \to \lambda$$

Algunas derivaciones:

$$S \Rightarrow Ab \Rightarrow b$$

$$S \Rightarrow Ab \Rightarrow aAbb \Rightarrow abb$$

$$S \Rightarrow Ab \Rightarrow aAbb \Rightarrow aaAbbb \Rightarrow aabbb$$

Más derivaciones:

$$S \Rightarrow Ab \Rightarrow aAbb \Rightarrow aaAbbb \Rightarrow aaaAbbbb$$
$$\Rightarrow aaaaAbbbbb \Rightarrow aaaabbbbb$$

$$S \overset{*}{\Rightarrow} aaaabbbbb$$

$$S \overset{*}{\Rightarrow} aaaaaabbbbbbb$$

$$S \overset{*}{\Rightarrow} a^n b^n b$$

Lenguaje generado

Dada una gramática G con símbolo inicial S

$$L(G) = \{w \in \Sigma^* : S \overset{*}{\Rightarrow} w\}$$

Donde w es una cadena/string de terminales

Ejemplo

Sea G = S → Ab

A → aAb

A → λ

Notación conveniente

S → Ab

A → aAb | λ

$$L(G) = \{a^n b^n b : \quad n \geq 0\}$$

Ya que $S \overset{*}{\Rightarrow} a^n b^n b, \forall n \geq 0$

Gramáticas lineales

Una gramática es lineal si tiene **a lo sumo** una variable en el lado derecho o izquierdo de **CADA** producción

Vemos dos ejemplos:

$$S \to aSb$$
$$S \to \lambda$$

$$S \to Ab$$
$$A \to aAb$$
$$A \to \lambda$$

Una gramática **NO** lineal

$$S \to SS$$

$$S \to \lambda$$

$$S \to aSb$$

$$S \to bSa$$

$$L(G) = \{w: \ |w|_a = |w|_b\}$$

Otra gramática lineal

$$S \to A$$
$$A \to aB \mid \lambda$$
$$B \to Ab$$

$$L(G) = \{a^n b^n : n \geq 0\}$$

Gramática lineal a derecha

Una gramática donde todas producciones tienen la forma $A \to xB$ o bien $A \to y$

con *x* e *y* terminales

Ejemplo: **S → abS | a**

Notar que se puede leerse como $w \in A$ si w = *x*u y $u \in B$ **o bien** w = *y* (*Definición inductiva*)

Otra posible lectura puede ser $L(A) = \{x\}L(B) \cup \{y\}$ (*Definición algebraica*)

Gramática lineal a izquierda

Una gramática donde todas producciones tienen la forma $A \to Bx$ o bien $A \to y$

con *x* e *y* terminales

Ejemplo:

S → Aab

A → Aab | B

B → a

Preguntiña

Que lenguajes generan las siguientes gramáticas ?

G_1

$$S \rightarrow abS$$

$$S \rightarrow a$$

Lineal a derecha

G_2

$$S \rightarrow Aab$$

$$A \rightarrow Aab \mid B$$

$$B \rightarrow a$$

Lineal a izquierda

Preguntiña

Que lenguajes generan las siguientes gramáticas ?

G_1

$$S \rightarrow abS$$
$$S \rightarrow a$$

G_2

$$S \rightarrow Aab$$
$$A \rightarrow Aab \mid B$$
$$B \rightarrow a$$

Notar que son regulares!

$$L(G_1) = (ab)^* a$$

$$L(G_2) = aab(ab)^*$$

Definición

Una gramática es regular si es lineal a derecha o lineal a izquierda

Teorema: El conjunto de Lenguajes Generados por Gramáticas Regulares es exactamente el conjunto de Lenguajes Regulares

Demo? Yep!

Ida = *Toda gramática regular genera un lenguaje regular*

Vuelta = *Todo lenguaje regular es generado por una gramática regular*

Ida

"Toda gramática regular genera un lenguaje regular"

Sea G una gramática regular lineal a derecha, probamos que L(G) es regular (Construimos un AFND)

Dado una gramática G, cómo construimos una máquina M que acepte L(G) ?

"Dado que w se genera a partir de S, para reconocer w tengo que poder

reconocer un string que comienza con a y sigue con un string que se genera

a partir de A; o bien se genera a partir de B ..."

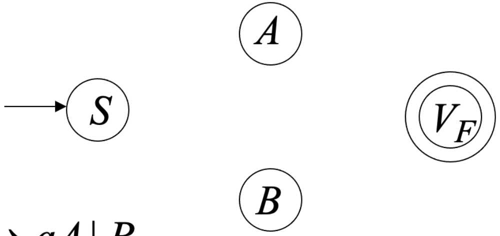
$$S \rightarrow aA \mid B$$
$$A \rightarrow aaB$$
$$B \rightarrow bB \mid a$$

$$L(G) = \{w \in \Sigma^* : S \overset{*}{\Rightarrow} w\}$$

Para reconocer **w** tengo que estar en el estado S y, si el símbolo observado es una a, hay que pasar al estado A; o bien, hay que pasar al estado B ..."

Construimos M

Cada estado es una variable de la gramática



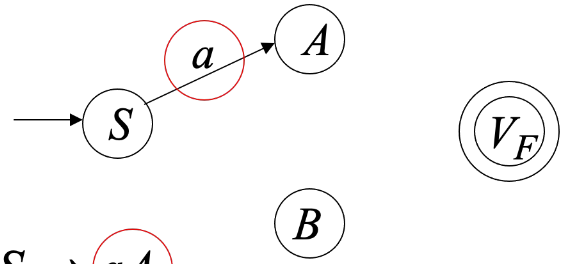
$$S \to aA \mid B$$

$$A \to aaB$$

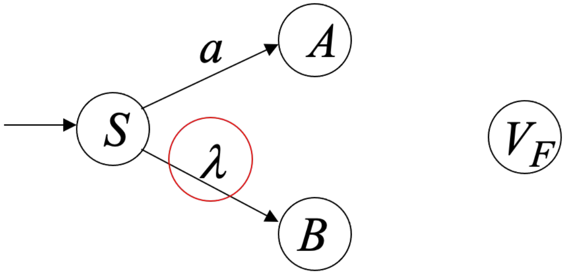
$$B \to bB \mid a$$

Agregamos una arista por cada producción

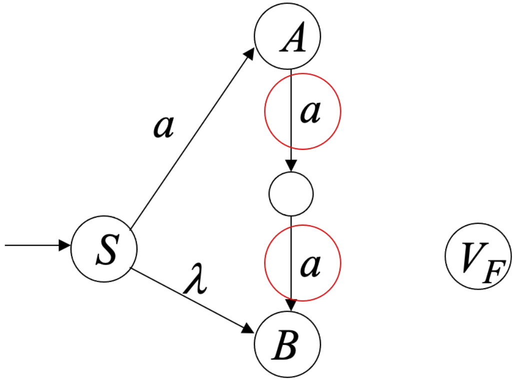
$$S \to aA \mid B$$
$$A \to aaB$$
$$B \to bB \mid a$$



$$S \to aA$$

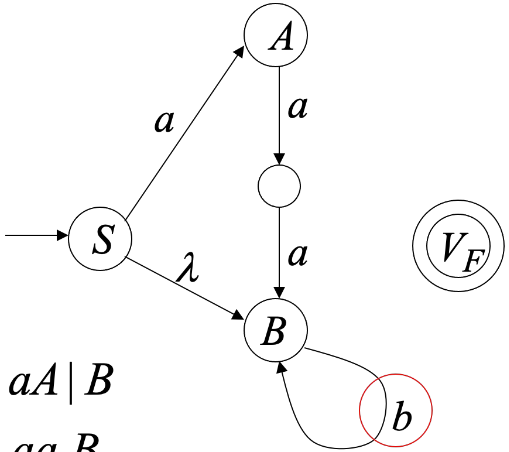


$$S \to aA \mid B$$



$$S \rightarrow aA \mid B$$

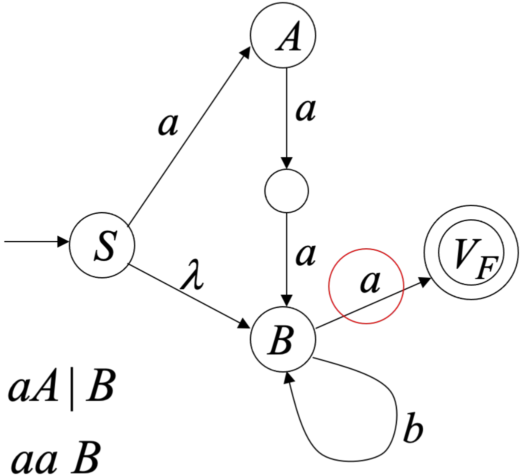
$$A \rightarrow aa\ B$$



$$S \rightarrow aA \mid B$$

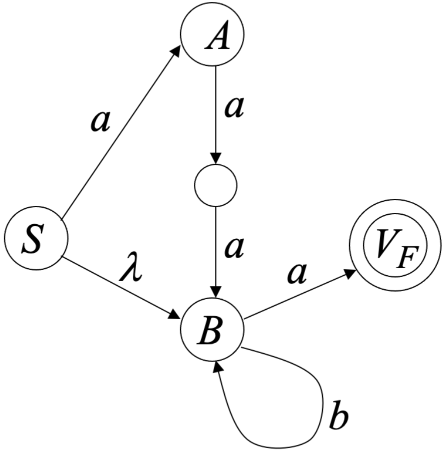
$$A \rightarrow aa\ B$$

$$B \rightarrow bB$$



$$S \to aA \mid B$$
$$A \to aa\ B$$
$$B \to bB \mid a$$

M



Gramática G

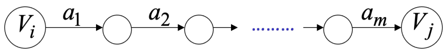
$$S \to aA \mid B$$
$$A \to aa\ B$$
$$B \to bB \mid a$$

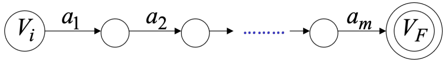
$$L(M) = L(G) = aaab^*a + b^*a = (aaa + \lambda)b^*a$$

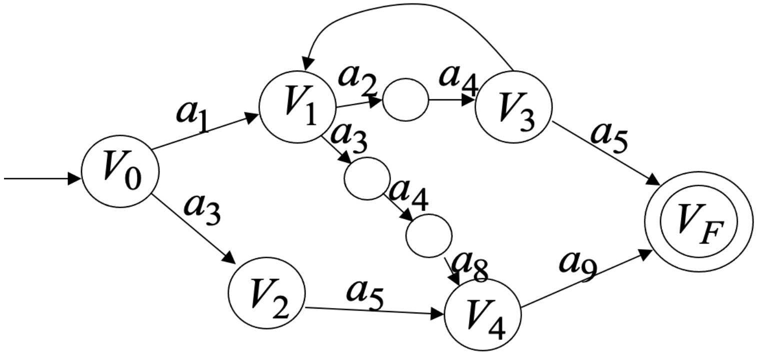
En general

Sea **G** una gramática lineal a derecha con variables $\mathbf{V_0, V_1, \ldots}$ y producciones $\mathbf{V_i \rightarrow a_1a_2\ldots a_mV_j}$ o bien $\mathbf{V_i \rightarrow a_1a_2\ldots a_m}$

Construimos un AFND

1) Cada variable $\mathbf{V_i}$ corresponde a un nodo y agregamos un estado final $\mathbf{V_f}$
2) Agregamos transiciones y nodos intermedios:
   a) $\mathbf{V_i \rightarrow a_1a_2\ldots a_mV_j}$ 

   b) $\mathbf{V_i \rightarrow a_1a_2\ldots a_m}$ 



Se cumple que L(G) = L(M)

Gramáticas lineales a izquierda

“Toda gramática regular genera un lenguaje regular”

Sea G una gramática regular lineal a izquierda, probamos que L(G) es regular (Construimos una GLD)

Dado una gramática G, cómo construimos una gramática lineal a derecha G’ tal que L(G) = L(G’)

Sabemos que G tiene la forma

$$A \to Ba_1a_2\cdots a_k$$

$$A \to a_1a_2\cdots a_k$$

G (Lineal a izquierda)

$$A \rightarrow Ba_1a_2 \cdots a_k$$

$$A \rightarrow v$$

G′ (Lineal a derecha)

$$A \rightarrow a_k \cdots a_2a_1B$$

$$A \rightarrow v^R$$

$L(G')$ ⟹ $L(G')^R$ ⟹ $L(G)$

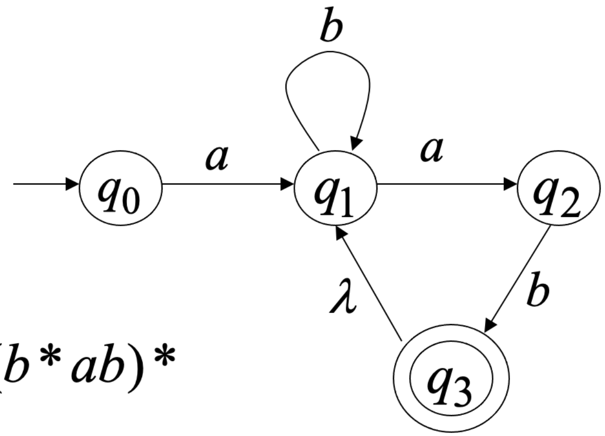
Lenguaje Regular | Lenguaje Regular | Lenguaje Regular

Vuelta

“Todo lenguaje L es generado por alguna gramática regular G”

Sea M el AFND con L = L(M), construimos una gramática regular G tal que L(M) = L(G)

M

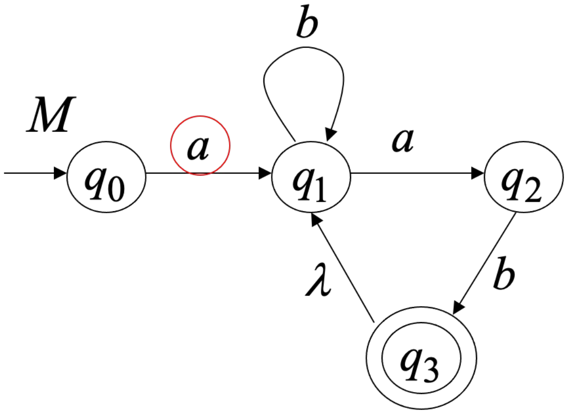


$$L = ab^*ab(b^*ab)^*$$

$$L = L(M)$$

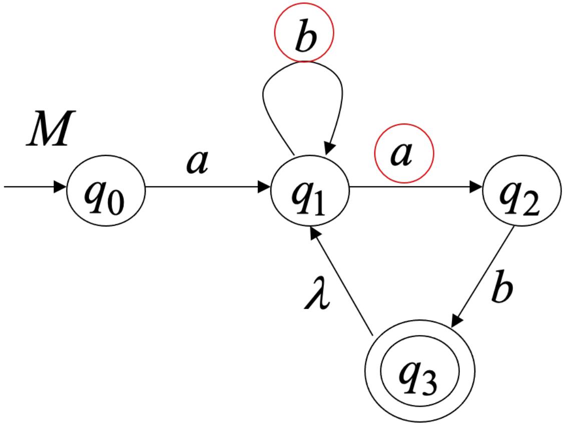
Construimos G

Convertimos cada transición en una producción (Donde cada Variable es un estado de M)



$$q_0 \to aq_1$$

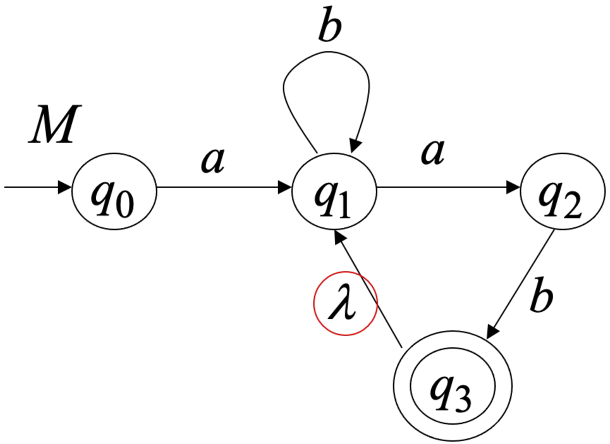
$$q_0 \to aq_1$$
$$q_1 \to bq_1$$
$$q_1 \to aq_2$$



$$L(G) = L(M) = L$$

G

$$q_0 \to aq_1$$
$$q_1 \to bq_1$$
$$q_1 \to aq_2$$
$$q_2 \to bq_3$$
$$q_3 \to q_1$$
$$q_3 \to \lambda$$



En general

Sea **M** un AFND

Construimos una gramática G

1) Cada transición $q \xrightarrow{a} p$ agregamos la producción $q \rightarrow ap$
2) Cada transición $q \xrightarrow{\lambda} p$ agregamos la producción $q \rightarrow p$
3) Para cada estado final q_f agregamos la producción $q_f \rightarrow \lambda$

Se cumple que G es una gramática regular con L(G) = L(M) = L