3.3 Gramáticas Regulares

José de Jesús Lavalle Martínez

Benemérita Universidad Autónoma de Puebla
Facultad de Ciencias de la Computación
Lenguajes Formales y Autómatas CCOS 014

Contenido

Conceptos previos

- La base de nuestro trabajo se llama **alfabeto** (se denota mediante Σ) y es un conjunto finito de símbolos (denotados mediante $a, b, c, 0, 1$, por ejemplo).

Conceptos previos

- La base de nuestro trabajo se llama **alfabeto** (se denota mediante Σ) y es un conjunto finito de símbolos (denotados mediante $a, b, c, 0, 1$, por ejemplo).
- Al yuxtaponer símbolos del alfabeto formamos **palabras** (también llamadas **cadenas**) (denotadas mediante u, v, w, x, y, z, por ejemplo) de cualquier longitud, pero finitas.

Conceptos previos ...

- El conjunto infinito de todas las palabras finitas basadas en el alfabeto Σ lo denotamos mediante Σ^*.

Conceptos previos ...

- El conjunto infinito de todas las palabras finitas basadas en el alfabeto Σ lo denotamos mediante Σ^*.
- Así, un **lenguaje** L es un conjunto (finito o infinito) de palabras tomadas de Σ^*, en decir, $L \subseteq \Sigma^*$.

Conceptos previos ...

1. Decimos que un Lenguaje L es **regular** si es **reconocido** por un autómata finito.

Conceptos previos ...

1. Decimos que un Lenguaje L es **regular** si es **reconocido** por un autómata finito.
2. Sabemos que una **expresión regular** denota a un conjunto regular y que para todo conjunto regular existe una expresión regular que lo denota.

Conceptos previos ...

1. Decimos que un Lenguaje L es **regular** si es **reconocido** por un autómata finito.
2. Sabemos que una **expresión regular** denota a un conjunto regular y que para todo conjunto regular existe una expresión regular que lo denota.
3. Una tercera manera de tratar con los lenguajes regulares es mediante las **Gramáticas Regulares**.

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

- V es un conjunto finito, a cada uno de sus elementos se le llama **símbolo no terminal** o **variable**, sus elementos se denotan mediante $S, A, B, C, \cdots, V_0, V_1, \cdots$.

Gramática

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

- V es un conjunto finito, a cada uno de sus elementos se le llama **símbolo no terminal** o **variable**, sus elementos se denotan mediante $S, A, B, C, \cdots, V_0, V_1, \cdots$.
- T es un conjunto finito disjunto de V, a cada uno de sus elementos se le llama **símbolo terminal**, sus elementos se denotan mediante $a, b, c, 0, 1, \cdots$.

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

- V es un conjunto finito, a cada uno de sus elementos se le llama **símbolo no terminal** o **variable**, sus elementos se denotan mediante $S, A, B, C, \cdots, V_0, V_1, \cdots$.
- T es un conjunto finito disjunto de V, a cada uno de sus elementos se le llama **símbolo terminal**, sus elementos se denotan mediante $a, b, c, 0, 1, \cdots$.
- $S \in V$ es el **símbolo no terminal inicial** o **variable inicial**.

Gramática

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

- V es un conjunto finito, a cada uno de sus elementos se le llama **símbolo no terminal** o **variable**, sus elementos se denotan mediante $S, A, B, C, \cdots, V_0, V_1, \cdots$.
- T es un conjunto finito disjunto de V, a cada uno de sus elementos se le llama **símbolo terminal**, sus elementos se denotan mediante $a, b, c, 0, 1, \cdots$.
- $S \in V$ es el **símbolo no terminal inicial** o **variable inicial**.
- P es una relación finita entre V y $(V \cup T)^*$. A los miembros de la relación P se les llama **producciones de la gramática**, **reglas de reescritura**, **producciones** o **reglas**.

Notación para las producciones

- Si bien una regla es un par $(\alpha, \beta) \in P$, con $\alpha \in V$ y $\beta \in (V \cup T)^*$, se acostubra usar la notación $\alpha \rightarrow \beta$, se dice que α es la **parte izquierda** de la regla y que β es la **parte derecha** de la regla. También se dice que α **produce** β.

Notación para las producciones

- Si bien una regla es un par $(\alpha, \beta) \in P$, con $\alpha \in V$ y $\beta \in (V \cup T)^*$, se acostubra usar la notación $\alpha \to \beta$, se dice que α es la **parte izquierda** de la regla y que β es la **parte derecha** de la regla. También se dice que α **produce** β.
- Es usual listar, separadas por el carácter $|$, las partes derechas que tienen en común su parte izquierda, es decir, suponga que tiene las producciones $\alpha \to \beta_1, \alpha \to \beta_2, \cdots, \alpha \to \beta_n$, dichas reglas se escriben como $\alpha \to \beta_1|\beta_2|, \cdots, |\beta_n$ y se lee α produce β_1, β_2 hasta β_n.

Derivaciones

- Sea la cadena $u_1\alpha u_2$ con $u_1, u_2 \in (V \cup T)^*$ y $\alpha \in V$, decimos que la palabra $u_1\alpha u_2$ **deriva en un paso a** a la palabra $u_1\beta u_2$, lo que se escribe $u_1\alpha u_2 \Rightarrow u_1\beta u_2$, si existe $\alpha \to \beta \in P$ con $\alpha \in V$ y $\beta \in (V \cup T)^*$.

Derivaciones

- Sea la cadena $u_1 \alpha u_2$ con $u_1, u_2 \in (V \cup T)^*$ y $\alpha \in V$, decimos que la palabra $u_1 \alpha u_2$ **deriva en un paso a** a la palabra $u_1 \beta u_2$, lo que se escribe $u_1 \alpha u_2 \Rightarrow u_1 \beta u_2$, si existe $\alpha \to \beta \in P$ con $\alpha \in V$ y $\beta \in (V \cup T)^*$.
- Note que $\Rightarrow$ es una relación binaria entre cadenas en $(V \cup T)^*$, su cerradura reflexiva y transitiva la denotaremos mediante $\stackrel{*}{\Rightarrow}$. Si tenemos que $u \stackrel{*}{\Rightarrow} v$, decimos que la palabra u **deriva en cero o más pasos** a la palabra v.

Lenguaje generado por una gramática

El lenguaje **generado** por la gramática $G = (V, T, S, P)$ es el conjunto

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

es decir, todas las palabras que constan sólo de símbolos terminales que son derivables en uno o más pasos a partir de la variable inicial S.

En una derivación multipasos, se dice que todas las cadenas en $(V \cup T)^*$ derivadas de S, pero exceptuando a $w \in T^*$, están en **forma sentencial**. Dicho de otra manera, todas las cadenas derivadas de S a las que se les puede aplicar una producción.

Gramática Lineal Derecha y Gramática Lineal Izquierda

Hay diferentes tipos de gramáticas, la diferencia está dada por la forma que tienen sus producciones. El tipo más restrictivo (con respecto a la forma de sus producciones) es el que ahora nos ocupa y se define a continuación.

Definición 1

Una gramática $G = (V, T, S, P)$ se dice que es **lineal derecha** si todas sus producciones son de alguna de las siguientes dos formas

$$A \to xB$$
$$A \to x$$

donde $A, B \in V$ y $x \in T^*$.

Definición 1

Una gramática $G = (V, T, S, P)$ se dice que es **lineal derecha** si todas sus producciones son de alguna de las siguientes dos formas

$$A \to xB$$
$$A \to x$$

donde $A, B \in V$ y $x \in T^*$. Una gramática se dice que es **lineal izquierda** si todas sus producciones son de alguna de las siguientes dos formas

$$A \to Bx$$
$$A \to x$$

Definición 1

Una gramática $G = (V, T, S, P)$ se dice que es **lineal derecha** si todas sus producciones son de alguna de las siguientes dos formas

$$A \rightarrow xB$$
$$A \rightarrow x$$

donde $A, B \in V$ y $x \in T^*$. Una gramática se dice que es **lineal izquierda** si todas sus producciones son de alguna de las siguientes dos formas

$$A \rightarrow Bx$$
$$A \rightarrow x$$

Una **gramática regular** es una que es lineal derecha o lineal izquierda.

Ejemplo gramática lineal derecha

Ejemplo 2

La gramática $G = (\{S\}, \{a, b\}, S, P)$ donde P está dado por las producciones

$$S \to abS|a$$

Ejemplo gramática lineal derecha

Ejemplo 2

La gramática $G = (\{S\}, \{a, b\}, S, P)$ donde P está dado por las producciones

$$S \to abS|a$$

es lineal derecha.

Ejemplo gramática lineal derecha

Ejemplo 2

La gramática $G = (\{S\}, \{a, b\}, S, P)$ donde P está dado por las producciones

$$S \rightarrow abS|a$$

$$S \overset{1}{\Rightarrow} a$$

Ejemplo gramática lineal derecha

Ejemplo 2

La gramática $G = (\{S\}, \{a, b\}, S, P)$ donde P está dado por las producciones

$$S \rightarrow abS|a$$

$$S \overset{1}{\Rightarrow} a$$

$$S \overset{1}{\Rightarrow} abS \overset{2}{\Rightarrow} aba$$

Ejemplo gramática lineal derecha

Ejemplo 2

La gramática $G = (\{S\}, \{a, b\}, S, P)$ donde P está dado por las producciones

$$S \to abS|a$$

$$S \overset{1}{\Rightarrow} a$$

$$S \overset{1}{\Rightarrow} abS \overset{2}{\Rightarrow} aba$$

$$S \overset{1}{\Rightarrow} abS \overset{2}{\Rightarrow} ababS \overset{3}{\Rightarrow} abababS \overset{4}{\Rightarrow} \cdots \overset{i-1}{\Rightarrow} (ab)^{i-1}S \overset{i}{\Rightarrow} (ab)^{i-1}a$$

Ejemplo gramática lineal derecha

Ejemplo 2

La gramática $G = (\{S\}, \{a, b\}, S, P)$ donde P está dado por las producciones

$$S \rightarrow abS|a$$

$$S \overset{1}{\Rightarrow} a$$

$$S \overset{1}{\Rightarrow} abS \overset{2}{\Rightarrow} aba$$

$$S \overset{1}{\Rightarrow} abS \overset{2}{\Rightarrow} ababS \overset{3}{\Rightarrow} abababS \overset{4}{\Rightarrow} \cdots \overset{i-1}{\Rightarrow} (ab)^{i-1}S \overset{i}{\Rightarrow} (ab)^{i-1}a$$

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

Ejemplo gramática lineal derecha

Ejemplo 2

La gramática $G = (\{S\}, \{a, b\}, S, P)$ donde P está dado por las producciones

$$S \to abS|a$$

$$S \overset{1}{\Rightarrow} a$$

$$S \overset{1}{\Rightarrow} abS \overset{2}{\Rightarrow} aba$$

$$S \overset{1}{\Rightarrow} abS \overset{2}{\Rightarrow} ababS \overset{3}{\Rightarrow} abababS \overset{4}{\Rightarrow} \cdots \overset{i-1}{\Rightarrow} (ab)^{i-1}S \overset{i}{\Rightarrow} (ab)^{i-1}a$$

$$L(G) = \{(ab)^n a : n \geq 0\} \text{ o } (ab)^* a$$

Demostración por inducción

$$S \rightarrow abS|a$$

Por demostrar que $L(G) = \{(ab)^n a : n \geq 0\}$, demostración por inducción sobre n.

$$S \to abS|a$$

Por demostrar que $L(G) = \{(ab)^n a : n \geq 0\}$, demostración por inducción sobre n.

Demostración 3

Hipótesis de inducción: $S \overset{i+1}{\Rightarrow} (ab)^i a = S \overset{i}{\Rightarrow} (ab)^i S \overset{1}{\Rightarrow} (ab)^i a$.

$$S \rightarrow abS|a$$

Por demostrar que $L(G) = \{(ab)^n a : n \geq 0\}$, demostración por inducción sobre n.

Demostración 3

Hipótesis de inducción: $S \overset{i+1}{\Rightarrow} (ab)^i a = S \overset{i}{\Rightarrow} (ab)^i S \overset{1}{\Rightarrow} (ab)^i a$.

Caso base: $n = 0$ $S \overset{1}{\Rightarrow} a = (ab)^0 a$.

$$S \to abS|a$$

Por demostrar que $L(G) = \{(ab)^n a : n \geq 0\}$, demostración por inducción sobre n.

Demostración 3

Hipótesis de inducción: $S \overset{i+1}{\Rightarrow} (ab)^i a = S \overset{i}{\Rightarrow} (ab)^i S \overset{1}{\Rightarrow} (ab)^i a$.

Caso base: $n = 0$ $S \overset{1}{\Rightarrow} a = (ab)^0 a$.

Caso inductivo: $n = i + 1$ Por hipótesis de inducción:

$$S \overset{i}{\Rightarrow} (ab)^i S \overset{1}{\Rightarrow} (ab)^i abS = (ab)^{i+1} S \overset{1}{\Rightarrow} (ab)^{i+1} a.$$

Demostración por inducción

$$S \rightarrow abS|a$$

Por demostrar que $L(G) = \{(ab)^n a : n \geq 0\}$, demostración por inducción sobre n.

Demostración 3

Hipótesis de inducción: $S \overset{i+1}{\Rightarrow} (ab)^i a = S \overset{i}{\Rightarrow} (ab)^i S \overset{1}{\Rightarrow} (ab)^i a$.

Caso base: $n = 0$ $S \overset{1}{\Rightarrow} a = (ab)^0 a$.

Caso inductivo: $n = i + 1$ Por hipótesis de inducción:

$$S \overset{i}{\Rightarrow} (ab)^i S \overset{1}{\Rightarrow} (ab)^i abS = (ab)^{i+1} S \overset{1}{\Rightarrow} (ab)^{i+1} a.$$

Por lo tanto se cumple para todo $n \in \mathbb{N}$.

Teorema 4

Si $G = (V, T, S, P)$ una gramática lineal derecha, entonces $L(G)$ es un lenguaje regular.

Demostración: Asuma que $V = \{V_0, V_1, \cdots\}, S = V_0$ y que tenemos producciones de la forma $V_0 \rightarrow v_1 V_i, V_i \rightarrow v_2 V_j, \cdots$ o de la forma $V_n \rightarrow v_l, \cdots$. Si $w \in L(G)$, entonces dada la forma de las producciones tenemos.

$$\begin{aligned} V_0 &\Rightarrow v_1 V_i \\ &\Rightarrow v_1 v_2 V_j \\ &\overset{*}{\Rightarrow} v_1 v_2 \cdots v_k V_n \\ &\Rightarrow v_1 v_2 \cdots v_k v_l = w. \end{aligned} \tag{1}$$

$L(G)$ es un lenguaje regular II

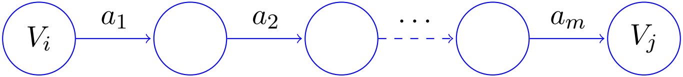
El autómata M que construiremos reproducirá las derivaciones reconociendo cada una de las $v's$ en (1).
El estado inicial del autómata se etiqueta como V_0 y a cada variable V_i le corresponde un estado no final etiquetado V_i.

Para cada producción de la forma

$$V_i \to a_1 a_2 \cdots a_m V_j$$

tenemos que definir δ de tal forma que

$$\delta^*(V_i, a_1 a_2 \cdots a_m) = V_j$$



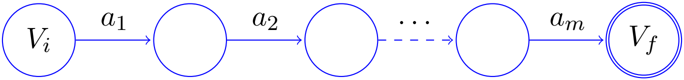
$L(G)$ es un lenguaje regular III

Para cada producción de la forma

$$V_i \to a_1 a_2 \cdots a_m$$

tenemos que definir δ de tal forma que

$$\delta^*(V_i, a_1 a_2 \cdots a_m) = V_f$$



Si $w \in L(G)$, entonces se satisface (1) y por construcción en el NFA existe un camino de V_0 a V_i etiquetado v_1, un camino de V_i a V_j etiquetado v_2, etc. Por lo tanto

$$V_f \in \delta^*(V_0, w)$$

y w es aceptado por M.

Para ver que M sólo acepta las palabras generadas por G, note que para que el autómata acepte a w tiene que pasar a través de una secuencia de estados $V_0, V_i, \cdots, V_f$ usando caminos con etiquetas $v_1, v_2, \cdots$. Por lo tanto, w debe tener la forma

$$w = v_1 v_2 \cdots v_k v_l$$

y en G se daría la derivación

$$V_0 \Rightarrow v_1 V_i \Rightarrow v_1 v_2 V_j \stackrel{*}{\Rightarrow} v_1 v_2 \cdots v_k V_n \Rightarrow v_1 v_2 \cdots v_k v_l = w$$

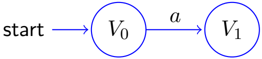
Así $w \in L(G)$ y el teorema queda demostrado.

Ejemplo de gramática lineal derecha a AFN

$$
\begin{aligned}
V_0 &\to aV_1, \\
V_1 &\to abV_0|b.
\end{aligned}
$$

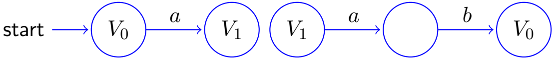
Ejemplo de gramática lineal derecha a AFN

$$V_0 \to aV_1,$$
$$V_1 \to abV_0|b.$$



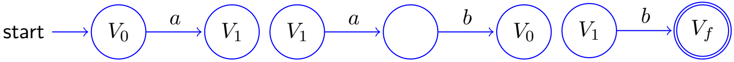
Ejemplo de gramática lineal derecha a AFN

$$V_0 \to aV_1,$$
$$V_1 \to abV_0|b.$$



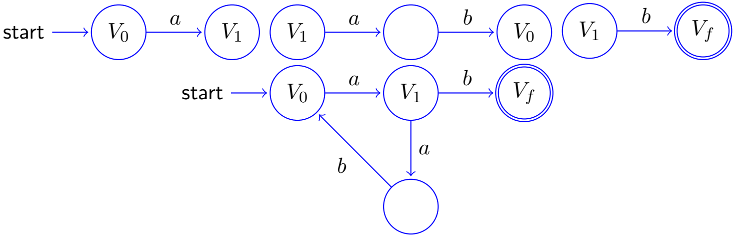
Ejemplo de gramática lineal derecha a AFN

$$V_0 \rightarrow aV_1,$$
$$V_1 \rightarrow abV_0|b.$$



Ejemplo de gramática lineal derecha a AFN

$$V_0 \rightarrow aV_1,$$
$$V_1 \rightarrow abV_0|b.$$



Teorema 5

Si L es un lenguaje regular sobre el alfabeto Σ, entonces existe una gramática lineal derecha $G = (V, \Sigma, S, P)$ tal que $L = L(G)$.

Demostración: Sea $M = (Q, \Sigma, \delta, q_0, F)$ un AFD que acepta L. Asumimos que $Q = \{q_0, q_1, \cdots, q_n\}$ y $\Sigma = \{a_1, a_2, \cdots, a_m\}$. Construya la gramática lineal derecha $G = (V, \Sigma, S, P)$ con

$$V = \{q_0, q_1, \cdots, q_n\} \text{ y } S = q_0.$$

Para cada transición

$$\delta(q_i, a_j) = q_k$$

de M, ponemos en P la producción

$$q_i \to a_j q_k.$$

Si L es regular entonces $L = L(G)$ II

Para terminar, si $q_k \in F$, añada a P la producción

$$q_k \to \lambda.$$

Primero demostremos que G así definida puede generar cualquier palabra de L. Considere que $w \in L$ tiene la forma

$$w = a_i a_j \cdots a_k a_l$$

Para que M acepte esta palabra debe hacer las siguientes transiciones

$$\begin{aligned}
\delta(q_0, a_i) &= q_p \\
\delta(q_p, a_j) &= q_r \\
&\vdots \\
\delta(q_s, a_k) &= q_t \\
\delta(q_t, a_l) &= q_f \in F
\end{aligned}$$

Por construcción, la gramática tendrá una producción para cada una de estas transiciones. Por lo tanto, se pueden hacer las siguientes derivaciones con la gramática G

$$\begin{aligned} q_0 &\Rightarrow a_i q_p \Rightarrow a_i a_j q_r \stackrel{*}{\Rightarrow} a_i a_j \cdots a_k q_t \\ &\Rightarrow a_i a_j \cdots a_k a_l q_f \Rightarrow a_i a_j \cdots a_k a_l \end{aligned} \tag{2}$$

por lo tanto $w \in L(G)$.
Para ver que G sólo genera las palabras de L, note que si $w \in L(G)$, entonces su derivación debe ser como en (2). Pero esto implica que

$$\delta^*(q_0, a_i a_j \cdots a_k a_l) = q_f,$$

con lo que se completa la prueba.

Ejemplo 6

Construya una gramática lineal derecha que genere el lenguaje $\{aab^na : n \geq 0\}$.

El siguiente AFN reconoce el lenguaje $\{aab^na : n \geq 0\}$

De AFN a GLD

Ejemplo 6

Construya una gramática lineal derecha que genere el lenguaje $\{aab^n a : n \geq 0\}$.

El siguiente AFN reconoce el lenguaje $\{aab^n a : n \geq 0\}$

$$\begin{array}{ccc}
\delta(q_0, a) = \{q_1\} & \qquad & q_0 \to aq_1 \\
\delta(q_1, a) = \{q_2\} & & q_1 \to aq_2 \\
\delta(q_2, b) = \{q_2\} & & q_2 \to bq_2 \\
\delta(q_2, a) = \{q_f\} & & q_2 \to aq_f \\
q_f \in F & & q_f \to \lambda
\end{array}$$

Ejercicios I

Los ejercicios los deben enviar en pdf a jlavallenator@gmail.com, pueden usar cualquier herramienta (de preferencia LaTeX), en el peor de los casos, si no tienen alternativa, lo hacen con papel y lapiz con letra legible y visible, lo escanean y me lo envían en pdf.

1. Demuestre que el AFN propuesto en el ejemplo 6 reconoce el lenguaje $\{aab^n a : n \geq 0\}$.
2. Construya un autómata finito que acepte el lenguaje generado por la gramática

$$\begin{aligned} S &\rightarrow abA, \\ A &\rightarrow baB, \\ B &\rightarrow aA|bb. \end{aligned}$$

Ejercicios II

3. Construya un autómata finito que acepte el lenguaje generado por la gramática

$$
\begin{aligned}
S &\to abS|A, \\
A &\to baB, \\
B &\to aA|bb.
\end{aligned}
$$

4. Construya una gramática lineal derecha para el lenguaje $L = \{a^n b^m : n \geq 3, m \geq 2\}$.
5. Construya una gramática lineal derecha para el lenguaje sobre $\Sigma = \{a, b\}$ que consiste de todas las cadenas con no más de dos $a's$.
6. Muestre que para toda gramática lineal derecha que no genere λ existe una gramática lineal derecha cuyas producciones están restringidas a las siguientes dos formas

$$A \to aB \text{ o } A \to a, A, B \in V, a \in T.$$