



Recordando

- Las **expresiones regulares** nos proporcionan una plantilla o patrón para las cadenas del lenguaje.
- Un **autómata finito** es un generador de cadenas del lenguaje. Especifica un lenguaje como el conjunto de todas las cadenas que lo hacen pasar del estado inicial a uno de aceptación.



Recordando

- Las expresiones regulares y los autómatas finitos nos proporcionan dos medios para especificar o definir lenguajes.
- Por lo tanto en ambos casos, todas las cadenas se corresponderán con un patrón en particular y dichas cadenas serán las únicas que formaran dicho lenguaje.



Gramática

- La **gramática** es el **estudio de las reglas y principios** que regulan el uso de las lenguas y la organización de las palabras dentro de una oración. También se denomina así al "**conjunto de reglas y principios que gobiernan el uso de un lenguaje**" así, cada lenguaje tiene su propia gramática.



Gramática

Una Gramática G se representa con una cuádrupla:

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

Donde:

- N es una colección finita de no terminales.
- Σ es un alfabeto (Conjunto de terminales).
- S es un no terminal llamado **Símbolo inicial**.
- P es una colección finita de **reglas de sustitución** llamadas **producciones**.



Elementos de una gramática

- **Símbolos terminales:** son elementos del alfabeto que no se pueden transformar, por eso se llaman terminales. Normalmente se denotan por letras minúsculas.
- **Variables o símbolos no terminales:** son elementos auxiliares que permiten poner restricciones sintácticas a un lenguaje. Las variables sí se pueden transformar, utilizando las reglas, en una cadena de variables y/o terminales. Por lo general se denotan por letras mayúsculas o por la notación $\langle \text{variable} \rangle$.



Elementos de una gramática

- **Símbolo inicial:** es el símbolo a partir del cual se generan todas las palabras válidas.
- **Producciones:** permiten reemplazar variables para generar oraciones válidas de un lenguaje. Puede haber varias reglas de producción para una misma variable, en algunos casos, las distintas opciones se colocan en una sola regla con los distintos reemplazos separados por “|”.
 - $\alpha \rightarrow \beta | \gamma | \delta$
 - abrevia las tres reglas
 - $\alpha \rightarrow \beta$
 - $\alpha \rightarrow \gamma$
 - $\alpha \rightarrow \delta$



Producciones

Cada producción consta de:

- Una variable a la que define (parcialmente) la producción. Esta variable a menudo se denomina cabeza de la producción.
- El símbolo de producción → ("puede ser", "se compone de" o "es sustituido por").



Producciones

- Es una cadena formada por cero o mas símbolos terminales y variables. Esta cadena, denominada *cuerpo* de la producción, representa una manera de formar cadenas pertenecientes al lenguaje de la variable cabeza de producción. De este modo, dejamos solo los símbolos terminales y sustituimos cada una de las variables del cuerpo por una cadena que sabemos que pertenece al lenguaje de dicha variable



Ejemplo 1: Gramática

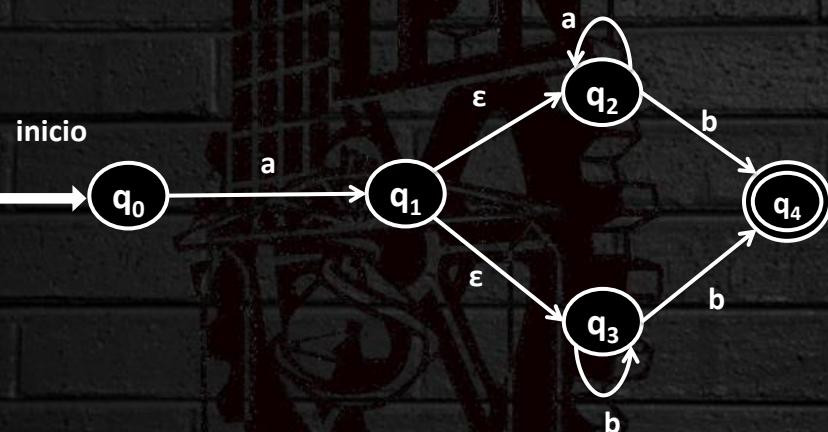
- $G = (N, \Sigma, S, P)$
 - $N = \{S, A, B\}$
 - $\Sigma = \{a, b, c\}$
 - $P: S \rightarrow AccA \quad A \rightarrow BA \mid \lambda \quad B \rightarrow a \mid b \mid c$

$w1 = abcc \in L(G)$

$w2 = acb \notin L(G)$



Ejemplo 2



- Este autómata finito acepta el lenguaje regular $a(a^* \mid b^*)b$.



Ejemplo 2

- Todas las cadenas del lenguaje precedente estarán formadas por una a seguida de alguna "parte final". Si representamos esa parte final con "E" y el símbolo inicial con "S" podemos crear la regla de producción:

$$S \rightarrow aE$$



Ejemplo 2

- La parte final (E) de una cadena estará formada por una de las dos listas de a's o b's. Por lo tanto, para indicar las múltiples posibilidades que hay para E se pueden crear 2 reglas:

$$E \rightarrow A$$

$$E \rightarrow B$$



Ejemplo 2

- Las 2 listas de a's y b's se pueden expresar de la siguiente manera:

$$\begin{array}{l} A \rightarrow aA \\ A \rightarrow b \end{array} \quad \left. \begin{array}{l} A \rightarrow aA \\ A \rightarrow b \end{array} \right\}$$

Para indicar que una cadena de a's va seguida de una b

$$\begin{array}{l} B \rightarrow bB \\ B \rightarrow b \end{array} \quad \left. \begin{array}{l} B \rightarrow bB \\ B \rightarrow b \end{array} \right\}$$

Para indicar que una cadena de b's va seguida de otra b



Ejemplo 2

Expresiones

$$\begin{array}{l} S \rightarrow aE \\ E \rightarrow A \\ E \rightarrow B \\ A \rightarrow aA \\ A \rightarrow b \\ B \rightarrow bB \\ B \rightarrow b \end{array}$$

Expresiones abreviadas

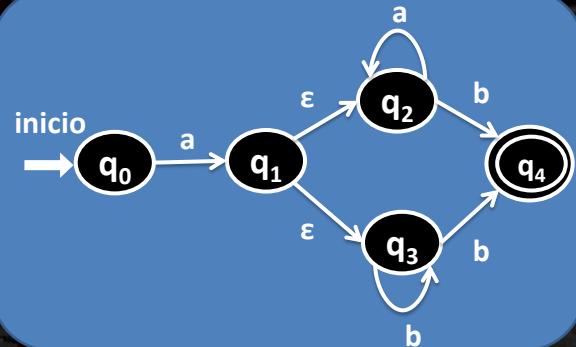
$$\begin{array}{l} S \rightarrow aE \\ E \rightarrow A \mid B \\ A \rightarrow aA \mid b \\ B \rightarrow bB \mid b \end{array}$$



Ejemplo 2

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

- $N = \{S, E, A, B\}$
- $\Sigma = \{a, b\}$
- $P:$
 - $S \rightarrow aE$
 - $E \rightarrow A \mid B$
 - $A \rightarrow aA \mid b$
 - $B \rightarrow bB \mid b$



$a(a^* \mid b^*)b$



Lenguaje generado por una gramática

- El lenguaje $L(G)$ generado por una gramática G es el conjunto de todas las **cadenas** que puede generar G .
- Una **cadena** pertenece a $L(G)$ si :
 - Está compuesta de símbolos terminales
 - La cadena puede derivarse del símbolo inicial S aplicando las reglas de producción de la gramática.



Gramáticas Formales

- Una gramática formal consta de un conjunto finito de *símbolos terminales* (las palabras en un lenguaje formal), un conjunto finito de *símbolos no terminales*, un conjunto de *reglas de producción* con un lado izquierdo y otro derecho, y un *símbolo inicial (No terminal "S")*.

Producción



Gramáticas Formales

- Las reglas se aplican sustituyendo la parte de la izquierda por la parte de la derecha. Una derivación es una secuencia de aplicaciones de reglas.
- Cada gramática define el lenguaje formal de todas las sentencias que están formadas exclusivamente por los símbolos terminales a los que se puede llegar mediante derivación a partir del símbolo inicial.



Derivaciones

- Una cadena $w \in \Sigma$ es **derivable** a partir de la gramática G si y solo si existe una secuencia de derivación iniciando en el símbolo inicial (no terminal "S") y terminando en la cadena w.
- La doble flecha \Rightarrow se interpreta como "deriva", "produce" o "genera".
- La notación $E \Rightarrow^* w$ indica que la cadena w se deriva a partir de S en 0 o mas etapas.



Derivaciones

- Cuando derivamos una cadena, los no terminales representan la parte de la cadena que todavía no se ha generado. Cuando la derivación se completa, todos los trozos no generados habrán sido sustituidos por cadenas de símbolos terminales (hay la posibilidad de que sean cadenas vacías).



Retomando Ejemplo 2

- $G = (N, \Sigma, S, P)$
 - $N = \{S, E, A, B\}$
 - $\Sigma = \{a, b\}$
 - $P: S \rightarrow aE \quad E \rightarrow A \mid B \quad A \rightarrow aA \mid b \quad B \rightarrow bB \mid b$

$w1 = aaab \in L(G)$



Retomando Ejemplo 2

- $G = (N, \Sigma, S, P)$
 - $N = \{S, E, A, B\}$
 - $\Sigma = \{a, b\}$
 - $P: S \rightarrow aE \quad E \rightarrow A \mid B \quad A \rightarrow aA \mid b \quad B \rightarrow bB \mid b$

$w1 = aaab \in L(G)$

- $S \Rightarrow aE \Rightarrow aA \Rightarrow aaA \Rightarrow aaaA \Rightarrow aaab$



Retomando Ejemplo 1: Gramática

- $\mathbf{G} = (\mathbf{N}, \Sigma, S, P)$
 - $\mathbf{N} = \{S, A, B\}$
 - $\Sigma = \{a, b, c\}$
 - $P: S \rightarrow AccA \quad A \rightarrow BA \mid \lambda \quad B \rightarrow a \mid b \mid c$

$w1 = abcc \in L(G)$

$w2 = acb \notin L(G)$



Retomando Ejemplo 1: Gramática

- $\mathbf{G} = (\mathbf{N}, \Sigma, S, P)$
 - $\mathbf{N} = \{S, A, B\}$
 - $\Sigma = \{a, b, c\}$
 - $P: S \rightarrow AccA \quad A \rightarrow BA \mid \lambda \quad B \rightarrow a \mid b \mid c$

$w1 = abcc \in L(G)$

$w2 = acb \notin L(G)$

- $S \Rightarrow AccA \Rightarrow BAccA \Rightarrow aAccA \Rightarrow aBAccA \Rightarrow$
 $abAccA \Rightarrow ab\lambda ccA \Rightarrow abcc\lambda \Rightarrow abcc$