Teoría de la Computación

Clase 4: Expresiones regulares

Mauro Artigiani

06 agosto 2021

Universidad del Rosario, Bogotá

Expresiones regulares

Expresiones regulares

Hemos visto que hay operaciones entre lenguajes regulares, como están las operaciones entre números. Entre números, se pueden concatenar las operaciones y obtener una *expresión* aritmética que sabemos evaluar. En otras palabras, el resultado de una expresión aritmética es un *número*, que el mismo objecto que estamos manipulando.

Expresiones regulares

Hemos visto que hay operaciones entre lenguajes regulares, como están las operaciones entre números. Entre números, se pueden concatenar las operaciones y obtener una *expresión* aritmética que sabemos evaluar. En otras palabras, el resultado de una expresión aritmética es un *número*, que el mismo objecto que estamos manipulando.

También las operaciones entre lenguajes se pueden concatenar, formando expresiones regulares como:

$$(0 \cup 1)0^*.$$

¿Qué significa esto? La expresión regular $(0 \cup 1)0^*$ tiene como valor un *lenguaje*, es decir el mismo objecto que estamos manipulando.

Expresiones regulares

Vamos a calcular la expresión regular $(0 \cup 1)0^*$.

Expresiones regulares

Vamos a calcular la expresión regular $(0 \cup 1)0^*$.

Primero, $(0 \cup 1) = (\{0\} \cup \{1\})$, entonces hasta ahora tenemos el lenguaje $\{0, 1\}$.

Expresiones regulares

Vamos a calcular la expresión regular $(0 \cup 1)0^*$.

Primero, $(0 \cup 1) = (\{0\} \cup \{1\})$, entonces hasta ahora tenemos el lenguaje $\{0, 1\}$. De manera similar $0^* = \{0\}^*$.

Expresiones regulares

Vamos a calcular la expresión regular $(0 \cup 1)0^*$.

Primero, $(0 \cup 1) = (\{0\} \cup \{1\})$, entonces hasta ahora tenemos el lenguaje $\{0, 1\}$. De manera similar $0^* = \{0\}^*$. Finalmente, cuando no hay signo de operación, se entiende que la operación sea $\circ$, la *concatenación* (como entre números se entiende la múltiplicación). Entonces el lenguaje que describe la expresión regular es:

$$L(R) = \{00^k, 10^k, k \geq 0\}.$$

Orden de las operaciones

Cómo en aritmética, hay un orden entre las operaciones que forman una expresión regular:

1. Potencia;
2. Concatenación;
3. Unión.

Claramente, las paréntesis alteran este orden.

Definición de una expresión regular

Definición de una expresión regular

Definición

Decimos que R es una expresión regular si R es

1. a para algún elemento del alfabeto Σ;
2. ε;
3. $\emptyset$;
4. $(R_1 \cup R_2)$, donde R_1 y R_2 son expresiones regulares;
5. $(R_1 \circ R_2)$, donde R_1 y R_2 son expresiones regulares;
6. R_1^*, donde R_1 es una expresión regular;

Definición de una expresión regular

Definición

Decimos que R es una expresión regular si R es

1. a para algún elemento del alfabeto Σ;
2. ε;
3. $\emptyset$;
4. $(R_1 \cup R_2)$, donde R_1 y R_2 son expresiones regulares;
5. $(R_1 \circ R_2)$, donde R_1 y R_2 son expresiones regulares;
6. R_1^*, donde R_1 es una expresión regular;

⚠ Cuidado! No confundan la expresión regular ε que es el lenguaje formado solo por la palabra vacía, con la expresión regular $\emptyset$ que es el lenguaje que no tiene palabras!

Definición de una expresión regular

Parece que nuestra definición sea *circular*. En realidad es una definición inductiva.

Definición de una expresión regular

Parece que nuestra definición sea *circular*. En realidad es una definición inductiva.

Tomamos la convención que $R^+ = RR^*$. Es decir: en R^+ están palabras obtenidas concatenando *por lo menos una* palabra de R. Con esta convención tenemos $R^+ \cup \varepsilon = R^*$. También convenimos que $R^k = \underbrace{RR\ldots R}_{k \text{ veces}}$.

Definición de una expresión regular

Parece que nuestra definición sea *circular*. En realidad es una definición inductiva.

Tomamos la convención que $R^+ = RR^*$. Es decir: en R^+ están palabras obtenidas concatenando *por lo menos una* palabra de R. Con esta convención tenemos $R^+ \cup \varepsilon = R^*$. También convenimos que $R^k = \underbrace{RR \ldots R}_{k \text{ veces}}$.

A veces es útil distinguir entre una expresión regular R y su lenguaje $L(R)$.

Ejemplos de expresiones regulares

Sea $\Sigma = \{0, 1\}$.

- $0^*10^* = \{w, w \text{ tiene un único } 1\}$;
- $1\Sigma^* = \{w, w \text{ tiene por lo meno un } 1\}$;
- $\Sigma^*010\Sigma^* = \{w, w \text{ contiene la subpalabra } 010\}$;
- $(10^+)^* = \{w, w \text{ cada } 1 \text{ está seguido de por lo menos un } 0\}$;
- $0\Sigma^*0 \cup 1\Sigma^*1 \cup 0 \cup 1 =$ $\{w, w \text{ empieza y termina con el mismo símbolo}\}$
- $(0 \cup \varepsilon)(1 \cup \varepsilon) = \{\varepsilon, 0, 1, 01\}$;
- $1^*\emptyset = \emptyset$;
- $\emptyset^* = \{\varepsilon\}$.

Diferencia entre ε y $\emptyset$

Dada una expresión regular R siempre tenemos:

$R \cup \emptyset = R$: añadir el lenguaje vacío al lenguaje R no cambia R.

Diferencia entre ε y $\emptyset$

Dada una expresión regular R siempre tenemos:

$R \cup \emptyset = R$: añadir el lenguaje vacío al lenguaje R no cambia R.

$R \circ \varepsilon = R$: concatenar la palabra vacía a cualquier palabra en R no le cambia.

Diferencia entre ε y $\emptyset$

Dada una expresión regular R siempre tenemos:

$R \cup \emptyset = R$: añadir el lenguaje vacío al lenguaje R no cambia R.

$R \circ \varepsilon = R$: concatenar la palabra vacía a cualquier palabra en R no le cambia.

⚠ Pero: $R \cup \varepsilon \neq R$ y tampoco $R \circ \emptyset \neq R$.

Aplicación a los compiladores

Las expresiones regulares sirven para definir la estructura aceptable de las expresiones usadas en nombres y valores de las variables y constantes de los lenguajes de programación.

Aplicación a los compiladores

Las expresiones regulares sirven para definir la estructura aceptable de las expresiones usadas en nombres y valores de las variables y constantes de los lenguajes de programación.

Por ejemplo, supongamos $\Sigma = \{+, -\} \cup D$, donde $D = \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$. Las expresiones numéricas pueden restringirse mediante la siguiente expresión regular:

$$(+ \cup - \cup \varepsilon)(D^+ \cup D^+.D^* \cup D^*.D^+)$$

Aplicación a los compiladores

Las expresiones regulares sirven para definir la estructura aceptable de las expresiones usadas en nombres y valores de las variables y constantes de los lenguajes de programación.

Por ejemplo, supongamos $\Sigma = \{+, -\} \cup D$, donde $D = \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$. Las expresiones numéricas pueden restringirse mediante la siguiente expresión regular:

$$(+ \cup - \cup \varepsilon)(D^+ \cup D^+.D^* \cup D^*.D^+)$$

Ejemplos

72

Aplicación a los compiladores

Las expresiones regulares sirven para definir la estructura aceptable de las expresiones usadas en nombres y valores de las variables y constantes de los lenguajes de programación.

Por ejemplo, supongamos $\Sigma = \{+, -\} \cup D$, donde $D = \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$. Las expresiones numéricas pueden restringirse mediante la siguiente expresión regular:

$$(+ \cup - \cup \varepsilon)(D^{+} \cup D^{+}.D^{*} \cup D^{*}.D^{+})$$

Ejemplos

72 , 3.14159

Aplicación a los compiladores

Las expresiones regulares sirven para definir la estructura aceptable de las expresiones usadas en nombres y valores de las variables y constantes de los lenguajes de programación.

Por ejemplo, supongamos $\Sigma = \{+, -\} \cup D$, donde $D = \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$. Las expresiones numéricas pueden restringirse mediante la siguiente expresión regular:

$$(+ \cup - \cup \varepsilon)(D^+ \cup D^+.D^* \cup D^*.D^+)$$

Ejemplos

72 , 3.14159 , +7.

Aplicación a los compiladores

Las expresiones regulares sirven para definir la estructura aceptable de las expresiones usadas en nombres y valores de las variables y constantes de los lenguajes de programación.

Por ejemplo, supongamos $\Sigma = \{+, -\} \cup D$, donde $D = \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$. Las expresiones numéricas pueden restringirse mediante la siguiente expresión regular:

$$(+ \cup - \cup \varepsilon)(D^+ \cup D^+.D^* \cup D^*.D^+)$$

Ejemplos

72 , 3.14159 , +7. , -.01

Aplicación a NLP

Considere el siguiente texto como una cadena en el alfabeto $\Sigma = \{$a,b,c,. . . ,x,y,z,A,B,C,. . . ,X,Y,Z,á,é,í,ó,ú,Á,É,Í,Ó,Ú,“,”,-,(,),,,.,␣$\}$:

> *A lo largo de la historia este concepto de mente ha sido concebido ontológicamente en diferentes categorías (como una sustancia distinta del cuerpo, una parte, un proceso o una propiedad). Sin embargo, las concepciones dominantes actuales, ambas materialistas, se engloban en la teoría de la identidad mente-cerebro y el funcionalismo. Esto sobre la mente. Un mentefacto es una representación gráfica que se utiliza para reflejar la estructura de los valores y pensamientos. El término está formado por la unión de “mente” y “facto”; el primero hace referencia al cerebro y el segundo a los hechos.*

Aplicación a NLP

Considere el siguiente texto como una cadena en el alfabeto $\Sigma = \{$a,b,c,...,x,y,z,A,B,C,...,X,Y,Z,á,é,í,ó,ú,Á,É,Í,Ó,Ú,“,”,-,(,),,,;,.,␣$\}$:

> *A lo largo de la historia este concepto de mente ha sido concebido ontológicamente en diferentes categorías (como una sustancia distinta del cuerpo, una parte, un proceso o una propiedad). Sin embargo, las concepciones dominantes actuales, ambas materialistas, se engloban en la teoría de la identidad mente-cerebro y el funcionalismo. Esto sobre la mente. Un mentefacto es una representación gráfica que se utiliza para reflejar la estructura de los valores y pensamientos. El término está formado por la unión de “mente” y “facto”; el primero hace referencia al cerebro y el segundo a los hechos.*

¿Qué expresión regular permite buscar las ocurrencias del sustantivo “mente”?

Aplicación a NLP

Considere el siguiente texto como una cadena en el alfabeto $\Sigma = \{$a,b,c,...,x,y,z,A,B,C,...,X,Y,Z,á,é,í,ó,ú,Á,É,Í,Ó,Ú,“,”,-,(,),,,.,␣$\}$:

> *A lo largo de la historia este concepto de mente ha sido concebido ontológicamente en diferentes categorías (como una sustancia distinta del cuerpo, una parte, un proceso o una propiedad). Sin embargo, las concepciones dominantes actuales, ambas materialistas, se engloban en la teoría de la identidad mente-cerebro y el funcionalismo. Esto sobre la mente. Un mentefacto es una representación gráfica que se utiliza para reflejar la estructura de los valores y pensamientos. El término está formado por la unión de “mente” y “facto”; el primero hace referencia al cerebro y el segundo a los hechos.*

¿Qué expresión regular permite buscar las ocurrencias del sustantivo “mente”?

Intento 1:

`mente`

Aplicación a NLP

Considere el siguiente texto como una cadena en el alfabeto $\Sigma = \{$a,b,c,...,x,y,z,A,B,C,...,X,Y,Z,á,é,í,ó,ú,Á,É,Í,Ó,Ú,“,”,-,(,),„,.,␣$\}$:

> *A lo largo de la historia este concepto de mente ha sido concebido ontológicamente en diferentes categorías (como una sustancia distinta del cuerpo, una parte, un proceso o una propiedad). Sin embargo, las concepciones dominantes actuales, ambas materialistas, se engloban en la teoría de la identidad mente-cerebro y el funcionalismo. Esto sobre la mente. Un mentefacto es una representación gráfica que se utiliza para reflejar la estructura de los valores y pensamientos. El término está formado por la unión de “mente” y “facto”; el primero hace referencia al cerebro y el segundo a los hechos.*

¿Qué expresión regular permite buscar las ocurrencias del sustantivo “mente”?

Intento 1:

`mente`	Correctas=4; Incorrectas=2; Faltan=0

Aplicación a NLP

Considere el siguiente texto como una cadena en el alfabeto $\Sigma = \{$a,b,c,...,x,y,z,A,B,C,...,X,Y,Z,á,é,í,ó,ú,Á,É,Í,Ó,Ú,“,”,-,(,),,,;,.,␣$\}$:

> *A lo largo de la historia este concepto de mente ha sido concebido ontológicamente en diferentes categorías (como una sustancia distinta del cuerpo, una parte, un proceso o una propiedad). Sin embargo, las concepciones dominantes actuales, ambas materialistas, se engloban en la teoría de la identidad mente-cerebro y el funcionalismo. Esto sobre la mente. Un mentefacto es una representación gráfica que se utiliza para reflejar la estructura de los valores y pensamientos. El término está formado por la unión de “mente” y “facto”; el primero hace referencia al cerebro y el segundo a los hechos.*

¿Qué expresión regular permite buscar las ocurrencias del sustantivo “mente”?

Intento 2:

`␣mente`

Aplicación a NLP

Considere el siguiente texto como una cadena en el alfabeto $\Sigma = \{$a,b,c,...,x,y,z,A,B,C,...,X,Y,Z,á,é,í,ó,ú,Á,É,Í,Ó,Ú,“,”,-,(,),,,;,.,␣$\}$:

> *A lo largo de la historia este concepto de mente ha sido concebido ontológicamente en diferentes categorías (como una sustancia distinta del cuerpo, una parte, un proceso o una propiedad). Sin embargo, las concepciones dominantes actuales, ambas materialistas, se engloban en la teoría de la identidad mente-cerebro y el funcionalismo. Esto sobre la mente. Un mentefacto es una representación gráfica que se utiliza para reflejar la estructura de los valores y pensamientos. El término está formado por la unión de “mente” y “facto”; el primero hace referencia al cerebro y el segundo a los hechos.*

¿Qué expresión regular permite buscar las ocurrencias del sustantivo “mente”?

Intento 2:

`␣mente` Correctas=3; Incorrectas=1; Faltan=1

Aplicación a NLP

Considere el siguiente texto como una cadena en el alfabeto $\Sigma = \{$a,b,c,...,x,y,z,A,B,C,...,X,Y,Z,á,é,í,ó,ú,Á,É,Í,Ó,Ú,“,”,-,(,),„,.,␣$\}$:

> *A lo largo de la historia este concepto de mente ha sido concebido ontológicamente en diferentes categorías (como una sustancia distinta del cuerpo, una parte, un proceso o una propiedad). Sin embargo, las concepciones dominantes actuales, ambas materialistas, se engloban en la teoría de la identidad mente-cerebro y el funcionalismo. Esto sobre la mente. Un mentefacto es una representación gráfica que se utiliza para reflejar la estructura de los valores y pensamientos. El término está formado por la unión de “mente” y “facto”; el primero hace referencia al cerebro y el segundo a los hechos.*

¿Qué expresión regular permite buscar las ocurrencias del sustantivo “mente”?

Intento 3:

(␣∪‘‘)`mente`

Aplicación a NLP

Considere el siguiente texto como una cadena en el alfabeto $\Sigma = \{$a,b,c,...,x,y,z,A,B,C,...,X,Y,Z,á,é,í,ó,ú,Á,É,Í,Ó,Ú,“,”,-,(,),,,;,.,␣$\}$:

> *A lo largo de la historia este concepto de mente ha sido concebido ontológicamente en diferentes categorías (como una sustancia distinta del cuerpo, una parte, un proceso o una propiedad). Sin embargo, las concepciones dominantes actuales, ambas materialistas, se engloban en la teoría de la identidad mente-cerebro y el funcionalismo. Esto sobre la mente. Un mentefacto es una representación gráfica que se utiliza para reflejar la estructura de los valores y pensamientos. El término está formado por la unión de “mente” y “facto”; el primero hace referencia al cerebro y el segundo a los hechos.*

¿Qué expresión regular permite buscar las ocurrencias del sustantivo “mente”?

Intento 3:

`(␣∪“)mente` Correctas=4; Incorrectas=1; Faltan=0

Aplicación a NLP

Considere el siguiente texto como una cadena en el alfabeto $\Sigma = \{$a,b,c,...,x,y,z,A,B,C,...,X,Y,Z,á,é,í,ó,ú,Á,É,Í,Ó,Ú,“,”,-,(,),,,;,.,␣$\}$:

> *A lo largo de la historia este concepto de mente ha sido concebido ontológicamente en diferentes categorías (como una sustancia distinta del cuerpo, una parte, un proceso o una propiedad). Sin embargo, las concepciones dominantes actuales, ambas materialistas, se engloban en la teoría de la identidad mente-cerebro y el funcionalismo. Esto sobre la mente. Un mentefacto es una representación gráfica que se utiliza para reflejar la estructura de los valores y pensamientos. El término está formado por la unión de “mente” y “facto”; el primero hace referencia al cerebro y el segundo a los hechos.*

¿Qué expresión regular permite buscar las ocurrencias del sustantivo “mente”?

Intento 4:

(␣∪‘)mente(␣∪’∪-∪.)

Aplicación a NLP

Considere el siguiente texto como una cadena en el alfabeto $\Sigma = \{$a,b,c,...,x,y,z,A,B,C,...,X,Y,Z,á,é,í,ó,ú,Á,É,Í,Ó,Ú,“,”,-,(,),,,;,.,␣$\}$:

> *A lo largo de la historia este concepto de mente ha sido concebido ontológicamente en diferentes categorías (como una sustancia distinta del cuerpo, una parte, un proceso o una propiedad). Sin embargo, las concepciones dominantes actuales, ambas materialistas, se engloban en la teoría de la identidad mente-cerebro y el funcionalismo. Esto sobre la mente. Un mentefacto es una representación gráfica que se utiliza para reflejar la estructura de los valores y pensamientos. El término está formado por la unión de “mente” y “facto”; el primero hace referencia al cerebro y el segundo a los hechos.*

¿Qué expresión regular permite buscar las ocurrencias del sustantivo “mente”?

Intento 4:

`(␣∪‘)mente(␣∪’∪-∪.)` Correctas=4; Incorrectas=0; Faltan=0

Regex de Python

```
import re

#Check if the string contains "ai"

txt = "The rain in Spain"
x = re.search(".*ai.*", txt)

if x:
  print("YES! We have a match!")
else:
  print("No match")
```

Metacaracteres

https://www.w3schools.com/python/python_regex.asp

Equivalencia entre expresiones regulares y lenguajes regulares

El teorema de Kleene

A un primer vistazo, los lenguajes regulares parecen bien diferentes de las expresiones regulares. Veremos que así no es:

Teorema de Kleene

Un lenguaje es regular si y solo si hay una expresión regular que lo describe.

El teorema de Kleene

En esta clase veremos una de las dos implicaciones:

Lema

Si un lenguaje está descrito por una expresión regular es regular.

El teorema de Kleene

En esta clase veremos una de las dos implicaciones:

Lema

Si un lenguaje está descrito por una expresión regular es regular.

Hemos visto que una expresión regular está definida en manera inductiva por varios bloques. Lo que haremos es construir un `NFA` N que reconozca el lenguaje de cada bloque.

Demostración

Vamos por casos:

- Si $R = a$ para algún $a \in \Sigma$, tenemos que $L(R) = \{a\}$.

  Consideramos el `NFA`: 

  Este `NFA` reconoce $L(R)$. Formalmente $N = (\{q_1, q_2\}, \Sigma, \delta, q_1, \{q_2\})$, donde $\delta(q_1, a) = q_2$ y todas las otras transiciones van a $\emptyset$;

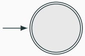
Demostración

Vamos por casos:

- Si $R = a$ para algún $a \in \Sigma$, tenemos que $L(R) = \{a\}$.

  Consideramos el `NFA`: 

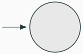
  Este `NFA` reconoce $L(R)$. Formalmente $N = (\{q_1, q_2\}, \Sigma, \delta, q_1, \{q_2\})$, donde $\delta(q_1, a) = q_2$ y todas las otras transiciones van a $\emptyset$;
- Si $R = \varepsilon$, tenemos que $L(R) = \{\varepsilon\}$.

  Consideramos el `NFA`: 

  Este `NFA` reconoce $L(R)$. Formalmente $N = (\{q_1\}, \Sigma, \delta, q_1, \{q_1\})$, donde todas las transiciones van a $\emptyset$;

Demostración

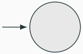
- Si $R = \emptyset$, tenemos que $L(R) = \emptyset$.

  Consideramos el `NFA`: 

  Este `NFA` reconoce $L(R)$. Formalmente $N = (\{q\}, \Sigma, \delta, q, \emptyset)$, donde todas las transiciones van a $\emptyset$;

Demostración

- Si $R = \emptyset$, tenemos que $L(R) = \emptyset$.

  Consideramos el `NFA`: 

  Este `NFA` reconoce $L(R)$. Formalmente $N = (\{q\}, \Sigma, \delta, q, \emptyset)$, donde todas las transiciones van a $\emptyset$;
- Si $R = R_1 \cup R_2$, $R = R_1 \circ R_2$, o $R = R_1^*$ podemos utilizar las construcciones que hemos visto en la clase pasada. $\square$

Ejemplos

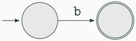
Vamos a aplicar el procedimiento que acabamos de explicar para construir un `NFA` que reconozca el lenguaje de la expresión regular $(\mathtt{ab} \cup \mathtt{a})^*$.

`a`: 

Ejemplos

Vamos a aplicar el procedimiento que acabamos de explicar para construir un NFA que reconozca el lenguaje de la expresión regular $(\mathtt{ab} \cup \mathtt{a})^*$.

a: 

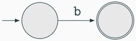
b: 

ab: 

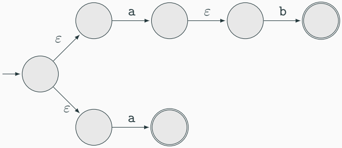
Ejemplos

Vamos a aplicar el procedimiento que acabamos de explicar para construir un `NFA` que reconozca el lenguaje de la expresión regular $(\mathtt{ab} \cup \mathtt{a})^*$.

$\mathtt{a}$: 

$\mathtt{b}$: 

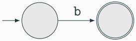
$\mathtt{ab}$: 

$\mathtt{ab} \cup \mathtt{a}$: 

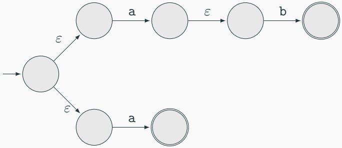
Ejemplos

Vamos a aplicar el procedimiento que acabamos de explicar para construir un `NFA` que reconozca el lenguaje de la expresión regular $(\mathtt{ab} \cup \mathtt{a})^*$.

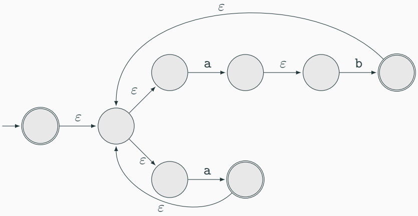
$\mathtt{a}$: 

$\mathtt{b}$: 

$\mathtt{ab}$: 



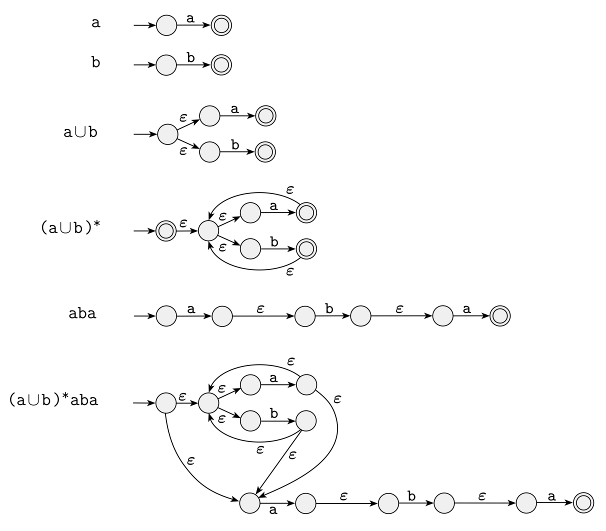
$\mathtt{ab} \cup \mathtt{a}$:



$(\mathtt{ab} \cup \mathtt{a})^*$:

Ejemplos

El siguiente es un $\texttt{NFA}$ que reconoce $(\texttt{a} \cup \texttt{b})^*\texttt{aba}$



Resumen

Resumen

Hoy aprendimos:

- Qué es una expresión regular;
- A construir un `NFA` que reconozca el lenguaje de una expresión regular.