Autómatas y Lenguajes

3er curso
1er cuatrimestre

Alfonso Ortega: alfonso.ortega@uam.es

UNIDAD 3: Equivalencias y caracterización

TEMA : Conjuntos regulares

c) Propiedades de los conjuntos regulares

Propiedades

1

Bombeo y anidamiento

Propiedades de los conjuntos regulares

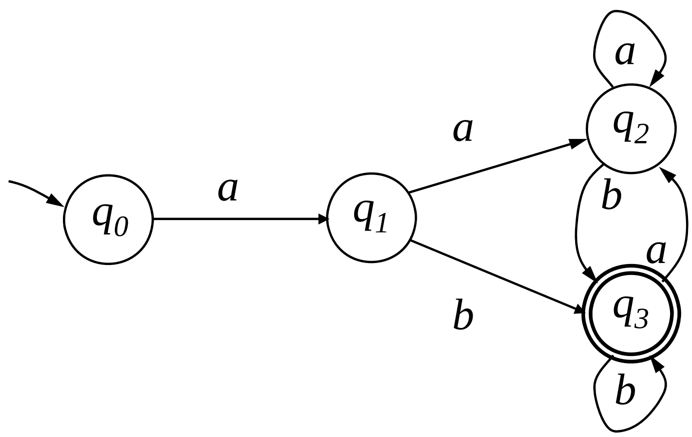
Introducción

- Se pueden agrupar las propiedades de los lenguajes regulares de la siguiente manera:
    - Determinación de lenguajes no regulares: lema de bombeo-anidamiento simple
    - Propiedades de cierre: determinación de qué operaciones de lenguajes, aplicadas a lenguajes regulares, generan con seguridad lenguajes regulares.
    - Algoritmos para problemas de decisión sobre lenguajes regulares. Estos no son objetivo de este curso.
    - Propiedades sobre la existencia de los autómatas más pequeños posibles para reconocer un lenguaje dado. Estas propiedades se estudiaron al describir el algoritmo de minimización de autómatas finitos deterministas.

Propiedad de anidamiento simple

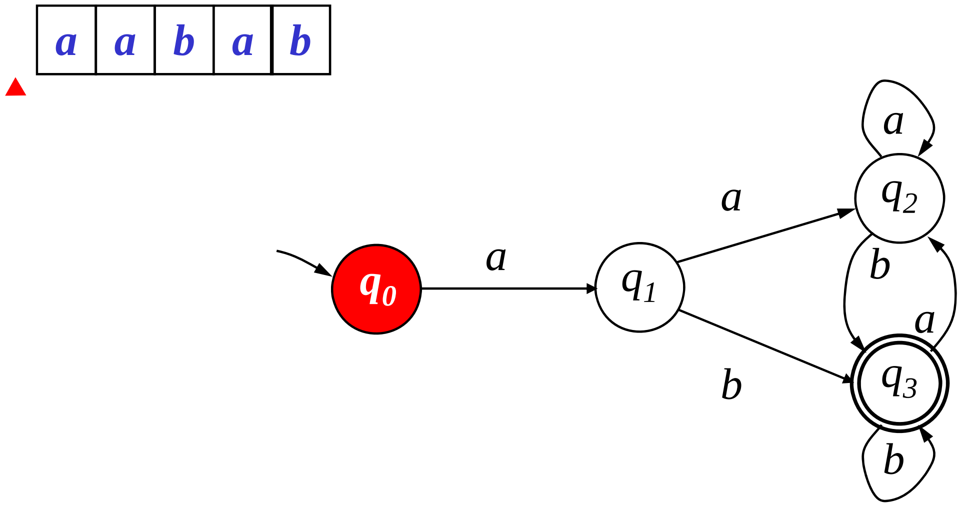
Ejemplo preliminar

- Considérese el siguiente autómata finito determinista (se ha omitido el sumidero).
- Es fácil comprobar que reconoce el lenguaje $\{awb \mid w \in \{a,b\}^*\}$.



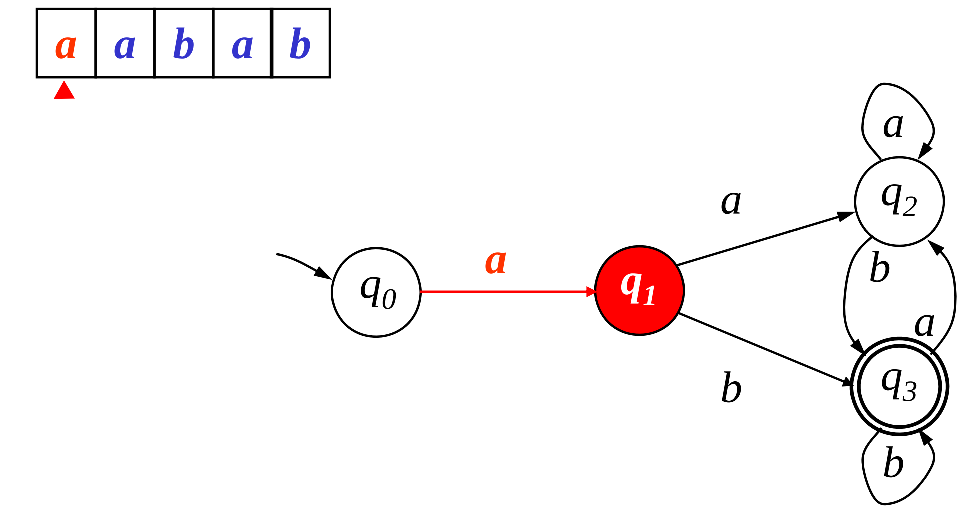
Ejemplo preliminar

- Por lo que aceptará la cadena *aabab*:



Ejemplo preliminar

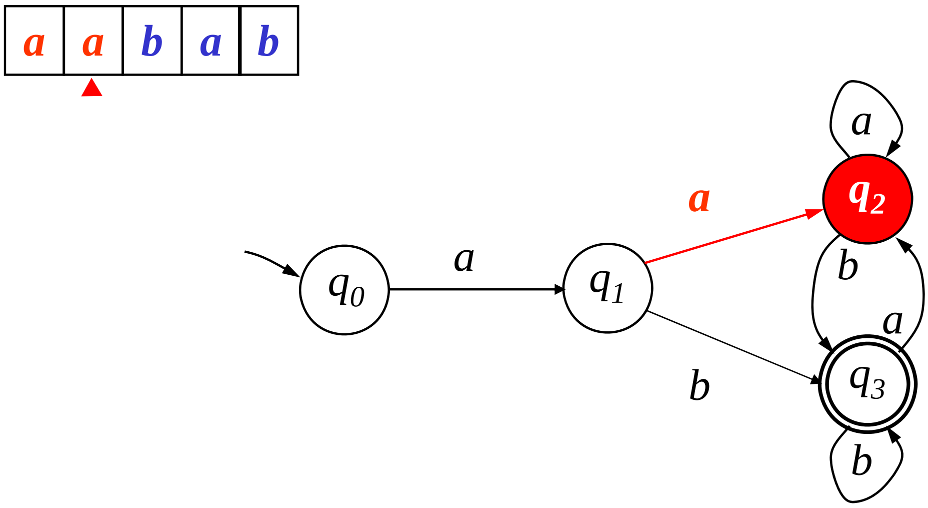
- Por lo que aceptará la cadena



Propiedad de anidamiento simple

Ejemplo preliminar

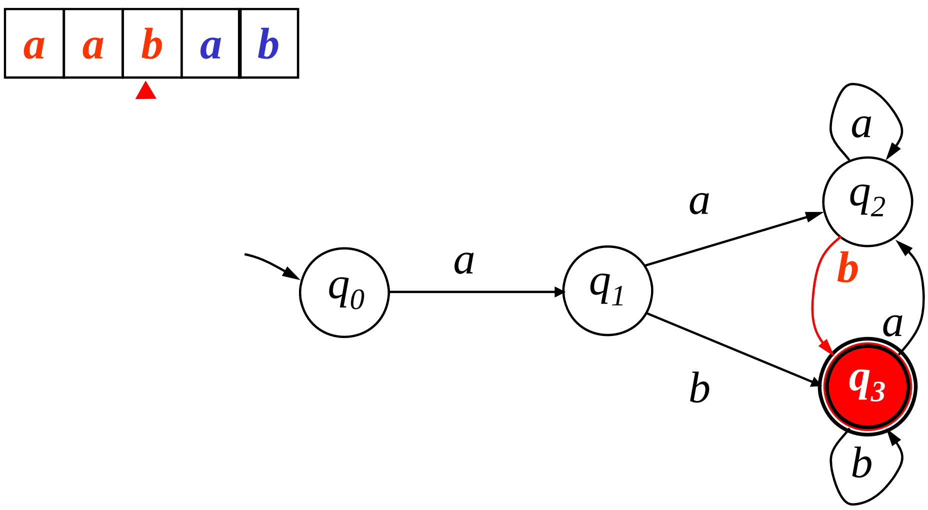
- Por lo que aceptará la cadena



Propiedad de anidamiento simple

Ejemplo preliminar

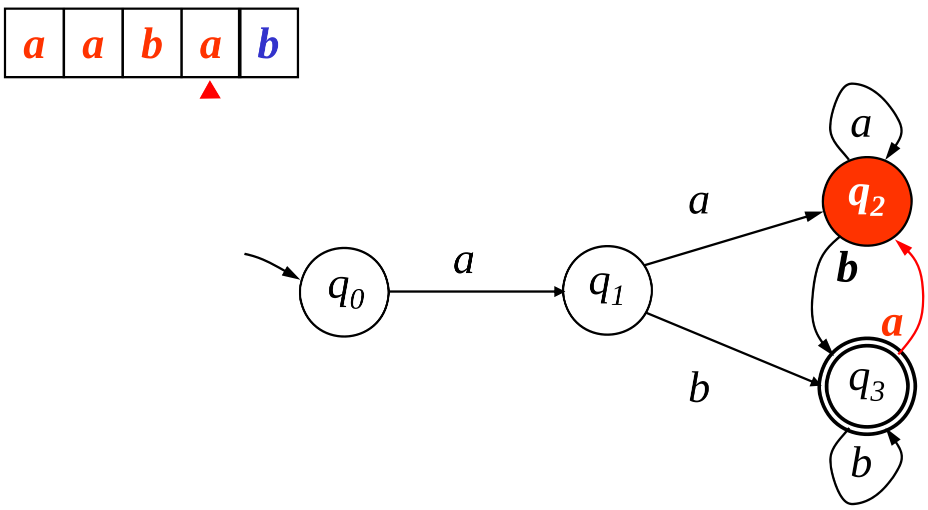
- Por lo que aceptará la cadena



Propiedad de anidamiento simple

Ejemplo preliminar

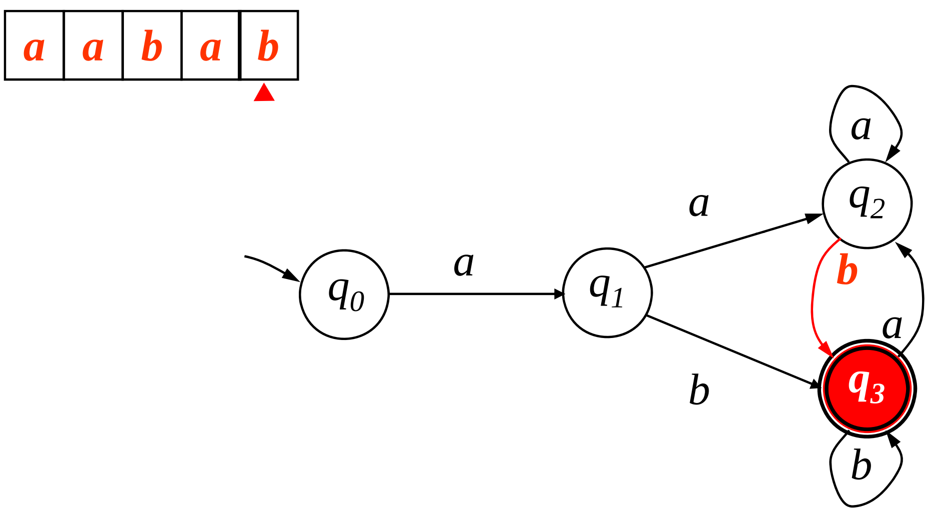
- Por lo que aceptará la cadena



Propiedad de anidamiento simple

Ejemplo preliminar

- Por lo que aceptará la cadena

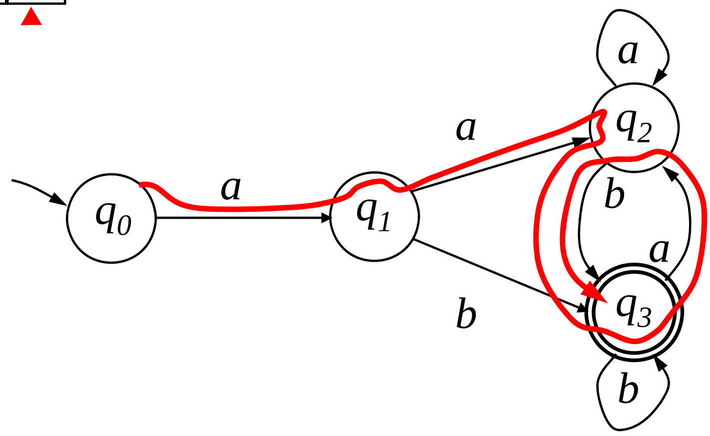


Propiedad de anidamiento simple

Ejemplo preliminar

- Se puede observar que los dos últimos símbolos recorren un bucle en el diagrama de transiciones $q_3 \to q_2 \to q_3$

a	*a*	*b*	*a*	*b*

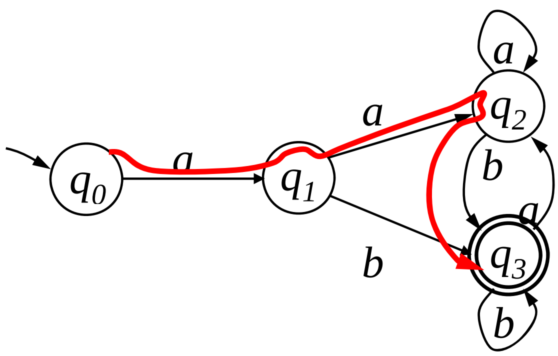


Propiedad de anidamiento simple

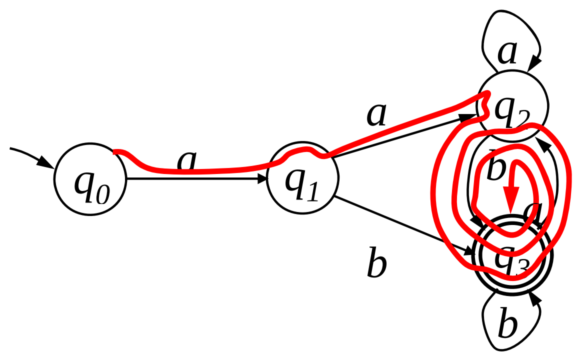
Ejemplo preliminar

- Por lo que, por la naturaleza del bucle, eliminarlo no modifica el destino del camino (la cadena resultante sigue siendo aceptada por el autómata -figura de la izquierda-) y repetirlo i veces tampoco (figura de la derecha)

aab



$aab(ab)^i$



Propiedad de anidamiento simple

Descripción informal

- El ejemplo anterior permite extraer las siguientes conclusiones:
  - Al estudiar un lenguaje regular podemos tener en cuenta un valor numérico interesante relacionado con la posibilidad de encontrar "ciclos":
    - El número es el de estados del autómata finito que reconoce el lenguaje.
    - Los ciclos se encontrarán en el diagrama de transiciones del autómata al procesar las diferentes palabras
  - Si se estudia una cadena suficientemente grande, se encontrarán ciclos en el diagrama de transiciones del autómata al procesar la cadena.
    - En nuestro ejemplo, la cadena aab es "demasiado corta" para que se produzcan ciclos; sin embargo, $aabab$ tiene un ciclo y $aab(ab)^i$ tiene i ciclos.
  - En estas condiciones
    - Se ha localizado una cadena, por ejemplo $aa\boldsymbol{ab}$, y una subcadena dentro de ella, aparece resaltada, que genera un ciclo en el funcionamiento del autómata
    - Se puede afirmar que todas las cadenas pertenecientes a la familia $\{aab(ab)^i, \forall i \in \{0,1,...\}\}$ necesariamente tienen que ser reconocidas por el autómata.
  - Las próximas páginas formalizan este resultado como las **propiedades de anidamiento simple** y **el lema de bombeo**.

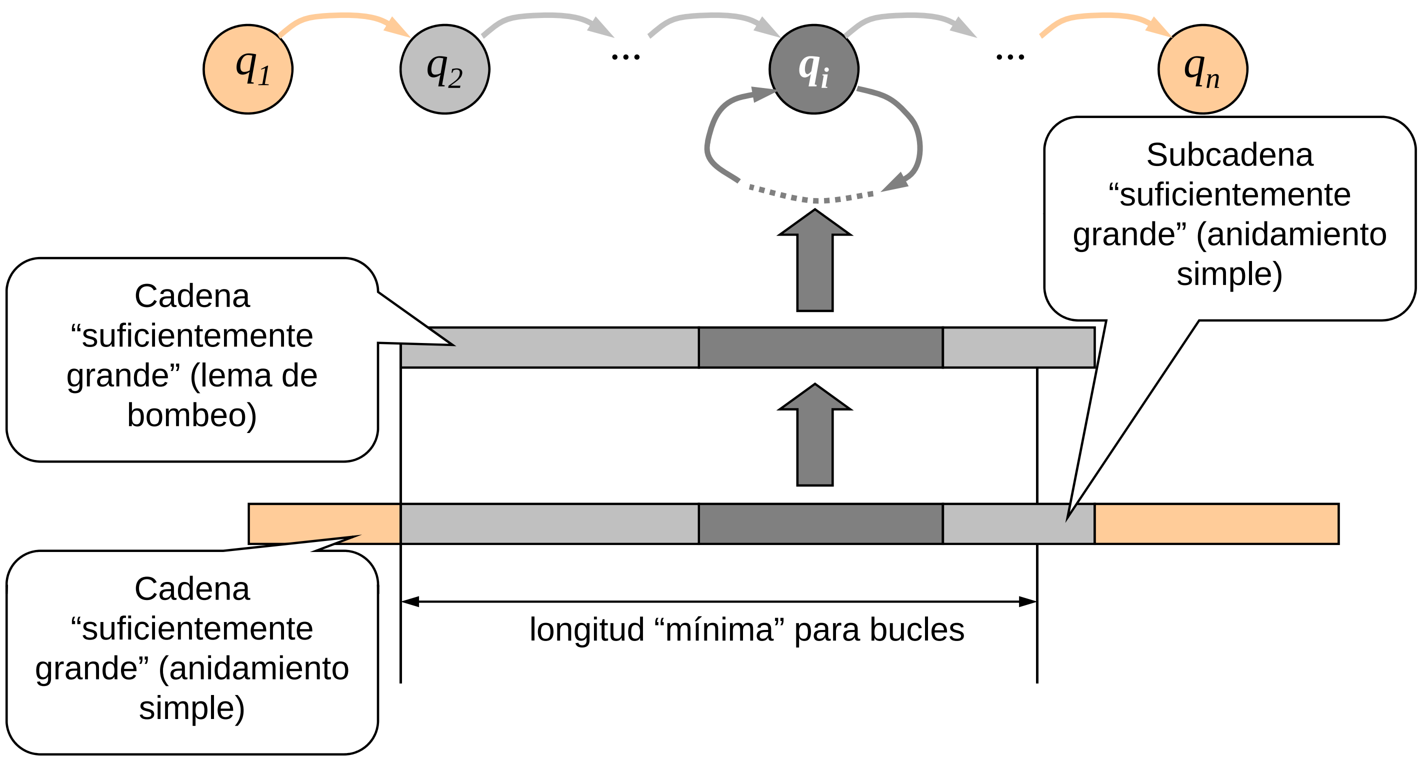
Propiedad de anidamiento simple

Descripción informal

- Hay dos resultados importantes relacionados con este aspecto.
- La explicación anterior serviría realmente para el que se conoce como **lema de bombeo de lenguajes regulares**.
- El otro resultado es más general y se llama **propiedad de anidamiento simple de los lenguajes regulares**.
- La diferencia entre los dos consiste en lo siguiente:
  - El lema de bombeo se refiere a cadenas suficientemente largas en las que se pueden encontrar ciclos.
  - La propiedad de anidamiento simple se refiere a cadenas suficientemente largas en las que se pueden encontrar subcadenas suficientemente largas en las que se pueden encontrar ciclos.
- La siguiente página muestra gráficamente esa diferencia

Propiedad de anidamiento simple

Descripción informal



Propiedad de anidamiento simple

Enunciado

- Los lenguajes regulares cumplen la siguiente propiedad:

$\forall\ L$ lenguaje regular infinito $\Rightarrow$

(el nº asociado con $\|Q\|$)	$\exists\ n>0\in N\ \|$
(todas las cadenas largas como para que haya “ciclos”)	$\forall\ x\in L\ ,,\ \|x\|\geq n\ ,$
(todas las subcadenas largas como para que haya “ciclos”)	$\forall z\ \|\ x=\alpha z\beta$ (subcadena de x) $,,\ \|z\|\geq n$
(identificación de una descomposición donde se encuentran “ciclos”: uv -hasta el “ciclo”- la long. es $\leq n$; el “ciclo” no puede ser λ)	$\exists\ u, v, w\ \|$ $z=uvw\ \wedge$ $\|uv\|\leq n\ \wedge$ $\|v\|>0\ \wedge$
(Cualquier número de repeticiones produce palabras del lenguaje)	$\forall i\in\{0,1,\ldots\}\Rightarrow \alpha uv^i w\beta\in L$

Propiedad de anidamiento simple

Corolario: negación de la propiedad de anidamiento simple

- El interés de este resultado está en el significado de su negación.
- Si llamamos $p_a(L)$ a la propiedad de anidamiento simple sobre un lenguaje L, el enunciado visto afirma

$$\forall\ L \text{ lenguaje regular} \Rightarrow p_a(L)$$

- ¿Qué ocurre si de un lenguaje concreto L' se puede demostrar que no se cumple la propiedad de anidamiento simple ($\neg p_a(L')$)?. Se pueden aplicar las propiedades de la lógica formal para afirmar:

$$\neg p_a(L) \Rightarrow L \text{ no es lenguaje regular}$$

- Es aplicación de la **ley del contrarrecíproco**

Si $P \Rightarrow Q$ es verdadera, entonces $\neg Q \Rightarrow \neg P$ es verdadera

- A continuación se enuncia formalmente la negación de la propiedad

Propiedad de anidamiento simple

Corolario: negación de la propiedad de anidamiento simple

- Formalmente, la negación de la propiedad de anidamiento simple se describe así:

(para todos los posibles tamaños de conjuntos de estados)	$\forall\ n > 0 \in N \mid$
(hay alguna cadena larga como para que haya “ciclos”)	$\exists\ x \in L$,, $\lvert x \rvert \geq n$,
(con alguna subcadena larga como para que haya “ciclos”)	$\exists\ z \mid\ x = \alpha z \beta$ (subcadena de x) ,, $\lvert z \rvert \geq n$
(para todas las posibles descom-posiciones, para todos los posibles candidatos a ciclos)	$\forall\ u, v, w \mid$ $z = uvw \wedge$ $\lvert uv \rvert \leq n \wedge$ $\lvert v \rvert > 0 \wedge$
(Puede encontrarse una repetición de la subcadena para la que la pa-labra completa no es del lenguaje)	$\exists\ i \in \{0,1,...\} \Rightarrow \alpha u v^i w \beta \notin L$
	$\Rightarrow L$ no es un lenguaje regular

Propiedad de anidamiento simple

Corolario: negación de la propiedad de anidamiento simple

- Observaciones:
  - Es importante darse cuenta de la alternancia de los cuantificadores y de su significado. Obsérvese también que, al aplicar las propiedades de la lógica de primer orden, la negación de la propiedad tiene el efecto de intercambiar los cuantificadores universales y existenciales.
  - Es una aplicación de la relación de los cuantificadores con la negación expresada en las Leyes de DeMorgan

$$\neg\forall x A(x) \Leftrightarrow \exists x \neg A(x)$$

$$\neg\exists x A(x) \Leftrightarrow \forall x \neg A(x)$$

Propiedad de anidamiento simple

Corolario: negación de la propiedad de anidamiento simple

- Observaciones:
  - Para aplicar la negación de la propiedad será necesario:
    - No realizar suposiciones sobre ningún valor entero (n) inicial, es decir, considerar uno cualquiera.
    - Para el que se tiene libertad en la elección de la cadena de longitud suficiente ($x \in L$,, $|x| \geq n$) ya que es suficiente con que exista una
    - Y la subcadena suya de longitud suficiente (z | $x=\alpha z\beta$ (subcadena de x) ,, $|z| \geq n$)
    - Para la cadena elegida, no se puede hacer ningún supuesto respecto a la descomposición considerada, es decir, hay que analizar todos los casos posibles.
    - Para cada posible descomposición es suficiente encontrar un valor particular (de i) para el que la repetición i veces de la subcadena candidata a ciclo produzca una palabra que no esté en el lenguaje.

Negación de la propiedad de anidamiento simple

Ejemplo 1

- Demostraremos que el siguiente lenguaje no es regular:

$$L=\{ww^{-1} \mid w\in\{a,b\}^*\}$$

  - Sea $m\in N$ cualquiera,
  - Se considera la cadena $x = \underbrace{a\ldots a}_{m \text{ veces}}\underbrace{b\ldots bb\ldots b}_{2m \text{ veces}}\underbrace{a\ldots a}_{m \text{ veces}}$
    - Que pertenece al lenguaje ya que $b^m a^m = (a^m b^m)^{-1}$
    - Y puede ser utilizada para la demostración ya que $|x| = |a^m b^m b^m a^m| = m+m+m+m = 4m \geq m$
  - Y la subcadena resaltada $z=a^m$ de $a^m b^m b^m \boldsymbol{a^m}$
    - Que también puede ser utilizada para la demostración ya que $|z| = |a^m| = m \geq m$
  - Se consideran todas las descomposiciones de interés $\forall\ u, v, w \mid z=uvw \wedge |uv|\leq n \wedge |v|>0$
    - Resulta claro que ya que $z=a^m$, en todos los casos ocurrirá que $v=a^j$,, $j>0$.
  - Se puede demostrar que para cualquier valor de $i\neq 1$ la cadena resultante $\alpha uv^i w\beta$ no pertenece al lenguaje, en particular para $i=0$, $\alpha uv^0 w\beta= a^m b^m b^m a^{m-j}$ y no es posible identificar una w tal que $a^m b^m b^m a^{m-j}=ww^{-1}$

Negación de la propiedad de anidamiento simple

Ejemplo 2

- Demostraremos que el siguiente lenguaje no es regular:

$$L=\{w\in\{a,b\}^* \mid \#_a(w)< \#_b(w)\}$$

  - Sea $m\in N$ cualquiera,
  - Se considera la cadena $x=a^m b^{m+1}$:
    - Que pertenece al lenguaje ya que $\#_a(a^m b^{m+1})=m<m+1=\#_b(a^m b^{m+1})$
    - Y puede ser utilizada para la demostración ya que $|x| = |a^m b^{m+1}| = m+m+1 = 2m+1 \geq m$
  - Y la subcadena, en este caso, es la misma $z=x$
  - Se consideran todas las descomposiciones de interés $\forall\ u, v, w \mid z=uvw \wedge |uv|\leq n \wedge |v|>0$

$$x = \underbrace{a\ldots a}_{m \text{ veces}}\overbrace{b\ldots b}^{m \text{ veces}}\, b$$

como mucho uv llega ahí ($|uv|\leq n$)

    - Resulta claro que ya que $uv=a^k$, en todos los casos ocurrirá que $v=a^j$,, $j>0$.
  - Demostraremos que para $i=3$ la cadena $\alpha uv^i w\beta$ no pertenece al lenguaje,
    - ya que $\alpha uv^3 w\beta= a^{m+2j} b^{m+1}$ y como $j>0\Rightarrow$(j es 1 o más, por tanto $2j$ es 2 o más, es decir, mayor que 1) $2j>1 \Rightarrow$(sumando m) $m+2j>m+1$
    - Por tanto $\#_a(\alpha uv^3 w\beta)= m+2j >m+1=\#_b(\alpha uv^3 w\beta)$ y no pertenece al lenguaje.

Propiedad de anidamiento simple / lema de bombeo

Observaciones

- Se ha explicado ya que la diferencia entre los dos resultados es la especificación de una subcadena adicional (llamada z en los resultados) sobre la que se identifica la subcadena posible ciclo.
- En el lema de bombeo se utiliza directamente la cadena "total" (llamada x en los resultados)
- Eso quiere decir que los ejemplos 2, 4 y 6 (en los que se elegía $z=x$) se corresponden con aplicaciones del lema de bombeo.
- Obsérvese que, en la práctica, el lema de bombeo necesita que la identificación del posible ciclo (la ubicación de la subcadena uv) se inicie al comienzo de la cadena mientras que en la propiedad de anidamiento simple esta cadena puede comenzar en cualquier posición de x.
- Es frecuente encontrar en la literatura especializada sólo este lema en lugar de la propiedad de anidamiento simple.
- Exclusivamente por esa razón (se ha visto que es un caso particular de la propiedad de anidamiento simple) se menciona a continuación el enunciado del lema de bombeo.

Lema de bombeo

Enunciado

- Los lenguajes regulares cumplen la siguiente propiedad:

	$\forall\ L$ lenguaje regular infinito $\Rightarrow$
(el nº asociado con $\|Q\|$)	$\exists\ n > 0 \in N \mid$
(todas las cadenas largas como para que haya "ciclos")	$\forall\ x \in L$,, $\|x\| \geq n\ \wedge$
(identificación de una descom-posición donde se encuentran "ciclos": uv -hasta el "ciclo"- la long. es $\leq n$; el "ciclo" no puede ser λ)	$\exists\ u, v, w \mid$ $x=uvw\ \wedge$ $\|uv\| \leq n\ \wedge$ $\|v\| > 0\ \wedge$
(Cualquier número de repeticiones produce palabras del lenguaje)	$\forall i \in \{0,1,\ldots\} \Rightarrow uv^i w \in L$

Lema de bombeo

Corolario: negación del lema de bombeo

- Formalmente, la negación del lema de bombeo se describe así:

(para todos los posibles tamaños de conjuntos de estados)	$\forall\ n > 0 \in N \mid$
(hay alguna cadena larga como para que haya "ciclos")	$\exists\ x \in L$,, $\lvert x \rvert \geq n\ \wedge$
(para todas las posibles descom-posiciones, para todos los posibles candidatos a ciclos)	$\forall\ u, v, w \mid$ $x = uvw \wedge$ $\lvert uv \rvert \leq n \wedge$ $\lvert v \rvert > 0 \wedge$
(Puede encontrarse una repetición de la subcadena para la que la pa-labra completa no es del lenguaje)	$\exists\ i \in \{0,1,...\} \Rightarrow uv^i w \notin L$
	$\Rightarrow L$ no es un lenguaje regular

2

Propiedades de cierre

Propiedades de los conjuntos regulares

Observaciones: regularidad de lenguajes finitos

- La propiedad de anidamiento simple (junto con el lema de bombeo) proporciona un mecanismo eficaz para el análisis de lenguajes infinitos.
- Los lenguajes finitos no presentan problemas ya que es fácil comprobar que siempre es posible construir un autómata finito que reconozca todas las posibles palabras del lenguaje.

Propiedades de cierre

Observaciones previas

- **Objetivo de las propiedades**
  - Las propiedades de cierre estudian si el resultado de las operaciones de lenguajes regulares siguen siendo regulares

Propiedades de cierre

Cerrados para $\cup$, ., *

- Estos tres resultados pueden presentarse juntos por su relación con las expresiones regulares:

Los lenguajes regulares son cerrados respecto a las operaciones de $\cup$, . y *

- Es decir

$\forall\ L_1, L_2$ lenguajes regulares $\Rightarrow$
$L_1 \cup L_2$ es lenguaje regular
$L_1 . L_2$ es lenguaje regular
$\forall\ L$ lenguaje regular $\Rightarrow$
L^* es lenguaje regular

Propiedades de cierre

Cerrados para -: observaciones previas a la demostración

- Podemos enunciar esta propiedad de la siguiente manera:

Los lenguajes regulares son cerrados respecto al complementario respecto al lenguaje universal

- Es decir

$\forall\ L$ lenguaje regular $\Rightarrow \Sigma^* - L$ es lenguaje regular

Propiedades de cierre

Cerrados para ∩

- Podemos enunciar esta propiedad de la siguiente manera:

Los lenguajes regulares son cerrados respecto a la intersección de conjuntos

- Es decir

$\forall\ L_1, L_2$ lenguajes regulares $\Rightarrow\ L_1 \cap L_2$ es lenguaje regular

Introducción

Bibliografía

[Alf] "*Teoría de Autómatas y lenguajes formales*" M. Alfonseca y otros

[Hop] "*Introducción a la teoría de autómatas, lenguajes y computación*" Hopcroft, J.; Motwani, R.; Ullman, J.