Autómatas y Lenguajes

3er curso
1er cuatrimestre

Alfonso Ortega: alfonso.ortega@uam.es

UNIDAD 3: Equivalencias y caracterización

TEMA Conjuntos independientes del contexto

Tema Propiedades

1 Bombeo y anidamiento

2 Propiedades de cierre

3 Equivalencias entre autómatas a pila y gramáticas independientes del contexto

1

Bombeo y anidamiento

Existencia de lenguajes no independientes del contexto

Propiedad de anidamiento doble-lema de bombeo: introducción

- La forma normal de Chomsky proporciona un criterio numérico para la identificación de repeticiones (subcadenas repetidas) en las derivaciones en el proceso de generación de lenguajes independientes del contexto infinitos que no incluyan la palabra vacía.
- Dada una gramática en forma normal de Chomsky:

$$G=(\Sigma_T,\Sigma_N,S,P)$$

- Se cumple que
  - Las reglas que no tienen terminales tienen exactamente dos símbolos no terminales en su parte derecha.
  - Las reglas que tienen terminales tienen exactamente uno en su parte derecha
  - Si se llama k al número de símbolos no terminales de la gramática.

$$k=|\Sigma_N|$$

  - La rama más larga posible sin repeticiones de símbolos no terminales en los árboles de derivación de la gramática tendrá longitud k.

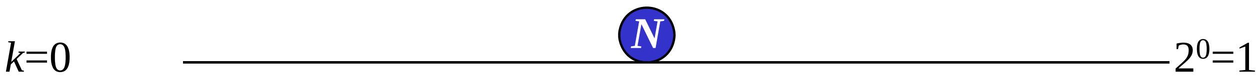
Existencia de lenguajes no independientes del contexto

Propiedad de anidamiento doble-lema de bombeo: introducción

- Como cada regla con no terminales tiene 2 en la parte derecha, y cada regla con terminales tiene sólo 1 en la parte derecha, la cadena más larga posible en un árbol de derivación de profundidad k cumplirá que
  - El número de hojas del árbol será 2^k
    - Con profundidad 0 hay 1 nodo
    - Con profundidad 1 hay 2 nodos
    - Con profundidad 2 hay (2 por cada nodo anterior) 4 nodos
    - Con profundidad 3 hay (2 por cada nodo anterior) 8 nodos, etc...
  - Como cada derivación final da sólo 1 terminal, la cadena generada tendrá una longitud igual a 2^k.

Existencia de lenguajes no independientes del contexto

Propiedad de anidamiento doble-lema de bombeo: introducción

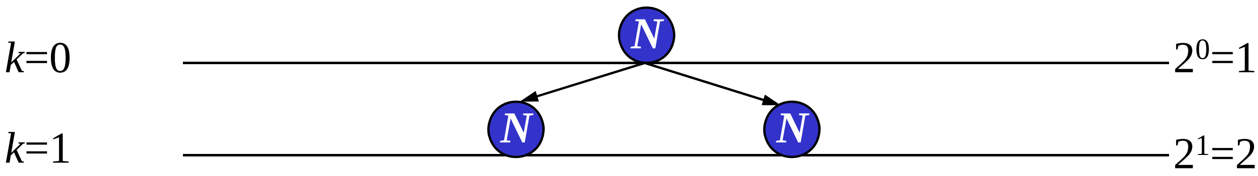


Profundidad

Nº nodos

Existencia de lenguajes no independientes del contexto

Propiedad de anidamiento doble-lema de bombeo: introducción

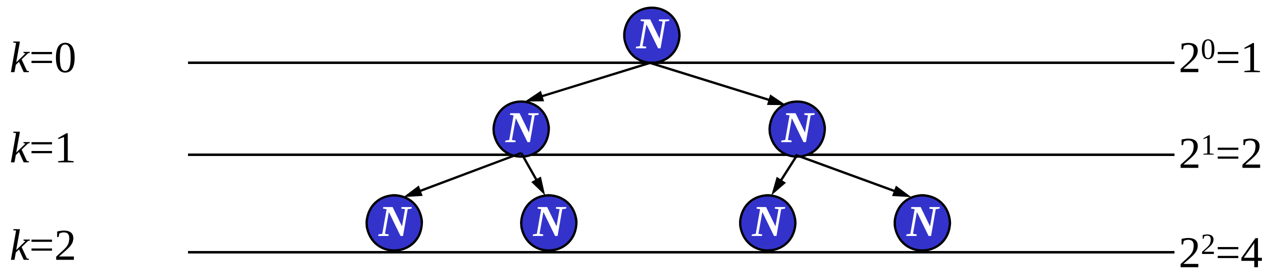


Profundidad

Nº nodos

Existencia de lenguajes no independientes del contexto

Propiedad de anidamiento doble-lema de bombeo: introducción

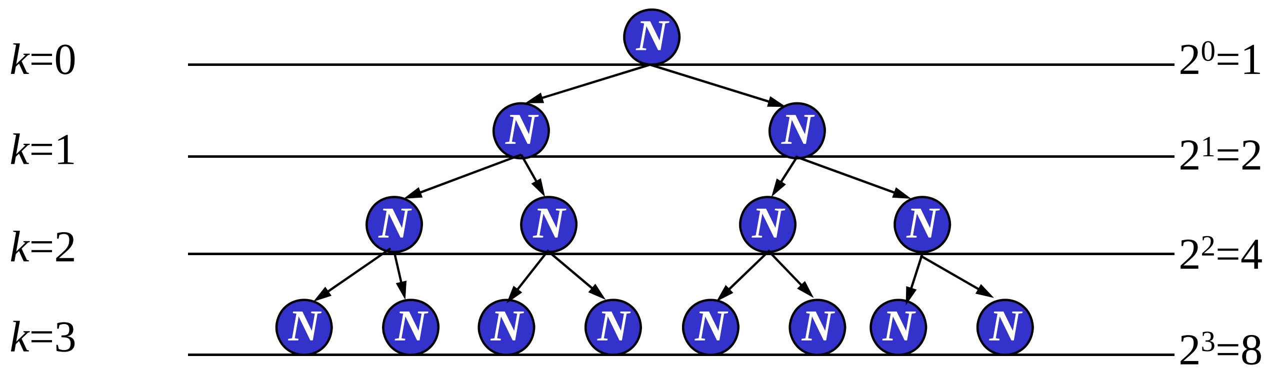


Profundidad

Nº nodos

Existencia de lenguajes no independientes del contexto

Propiedad de anidamiento doble-lema de bombeo: introducción

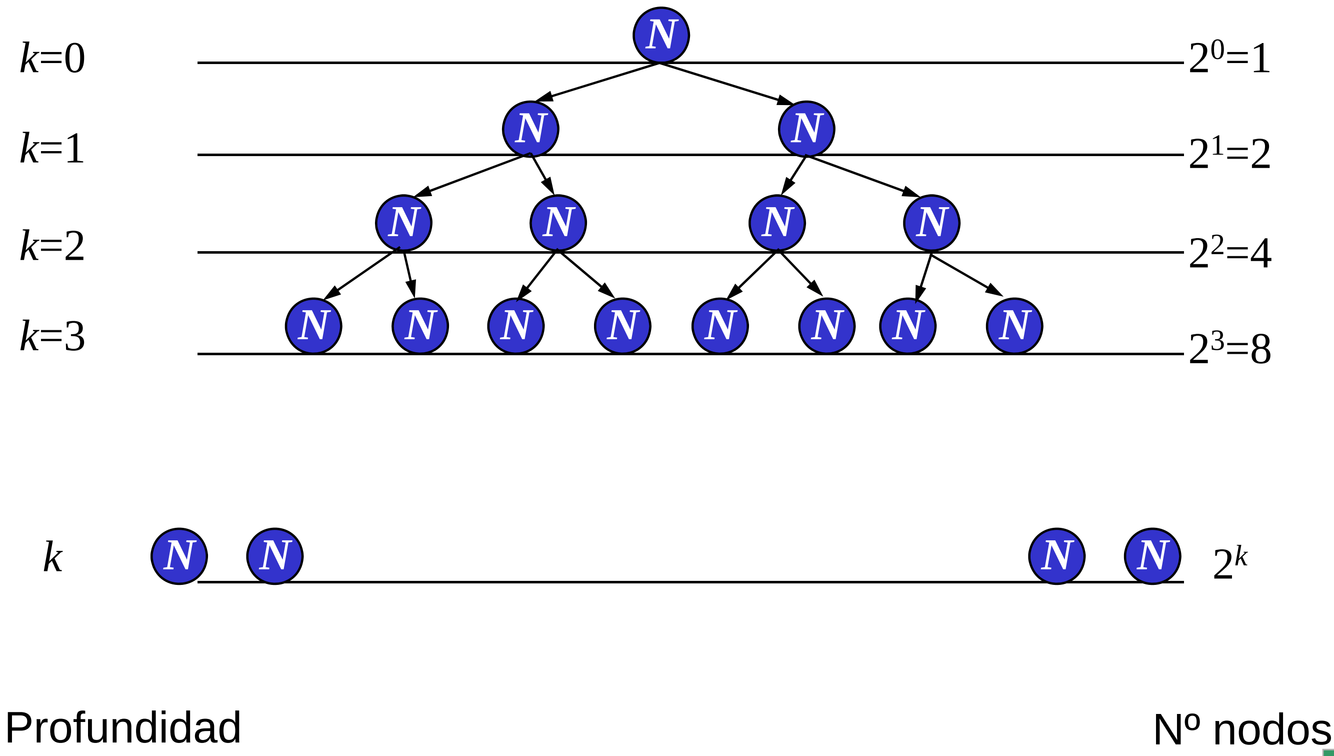


Profundidad

Nº nodos

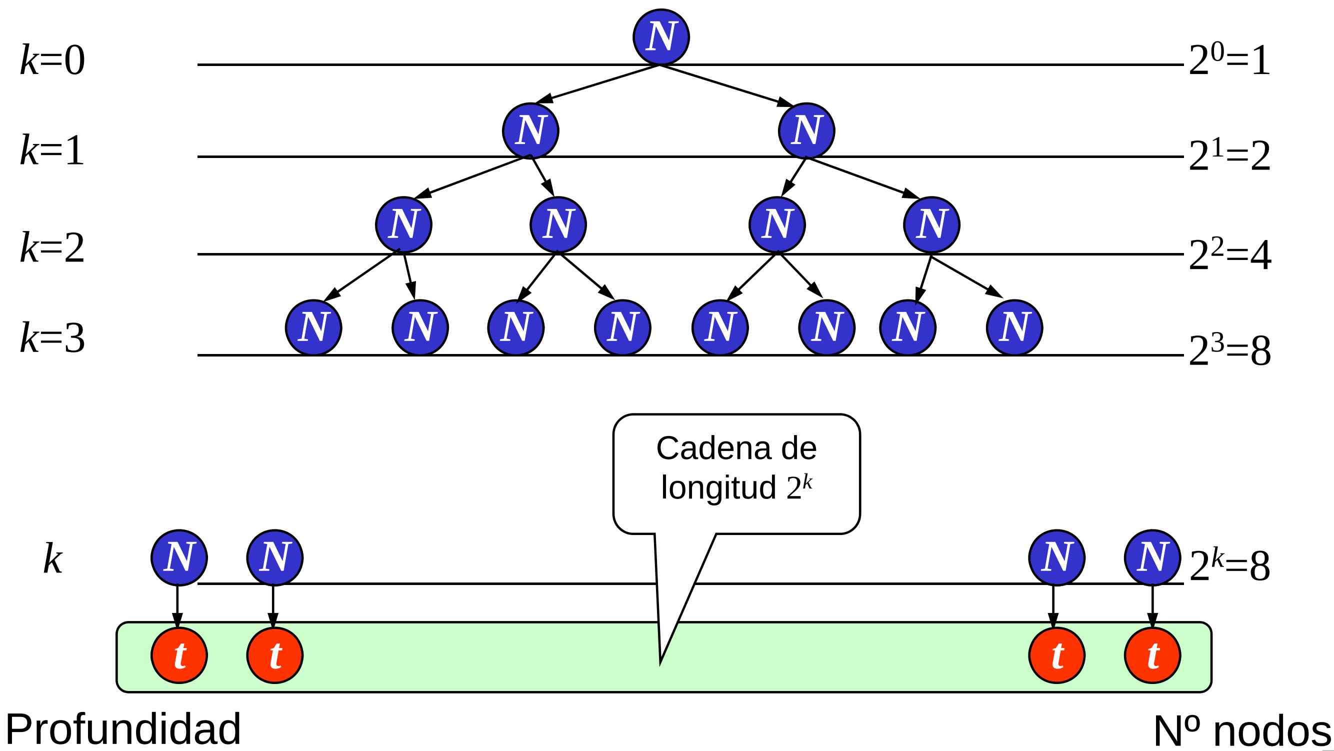
Existencia de lenguajes no independientes del contexto

Propiedad de anidamiento doble-lema de bombeo: introducción



Existencia de lenguajes no independientes del contexto

Propiedad de anidamiento doble-lema de bombeo: introducción



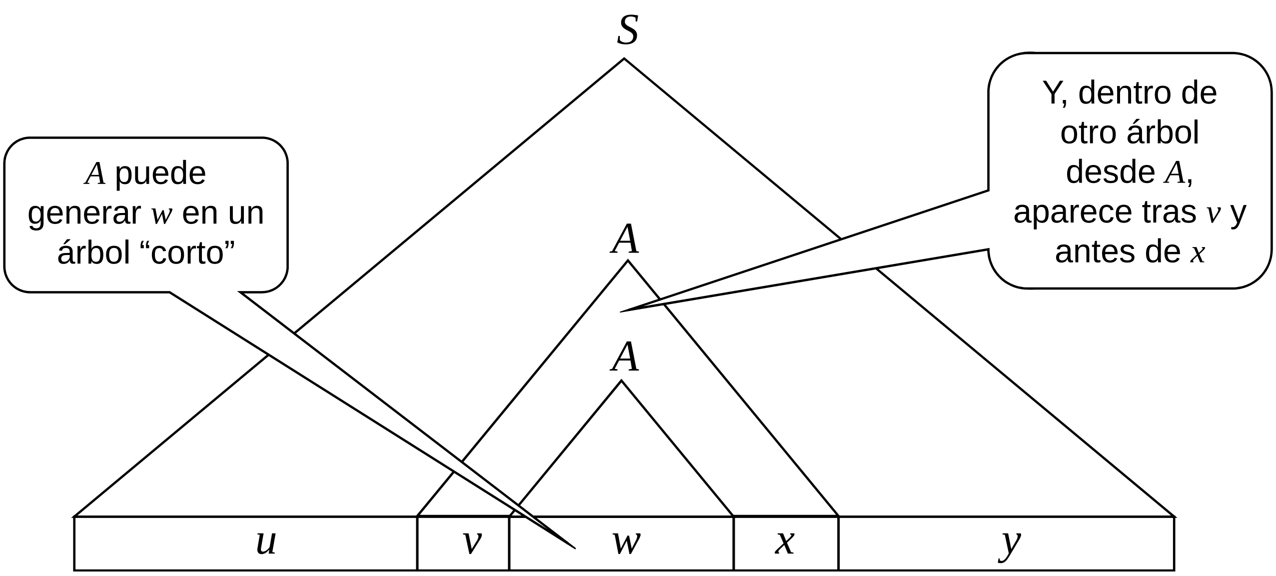
Existencia de lenguajes no independientes del contexto

Propiedad de anidamiento doble-lema de bombeo: introducción

- Esto significa que cualquier cadena de longitud mayor necesita un árbol de profundidad mayor y, por lo tanto, la rama más profunda tendrá algún nodo con el símbolo no terminal repetido.
- La siguiente figura muestra gráficamente este árbol

Existencia de lenguajes no independientes del contexto

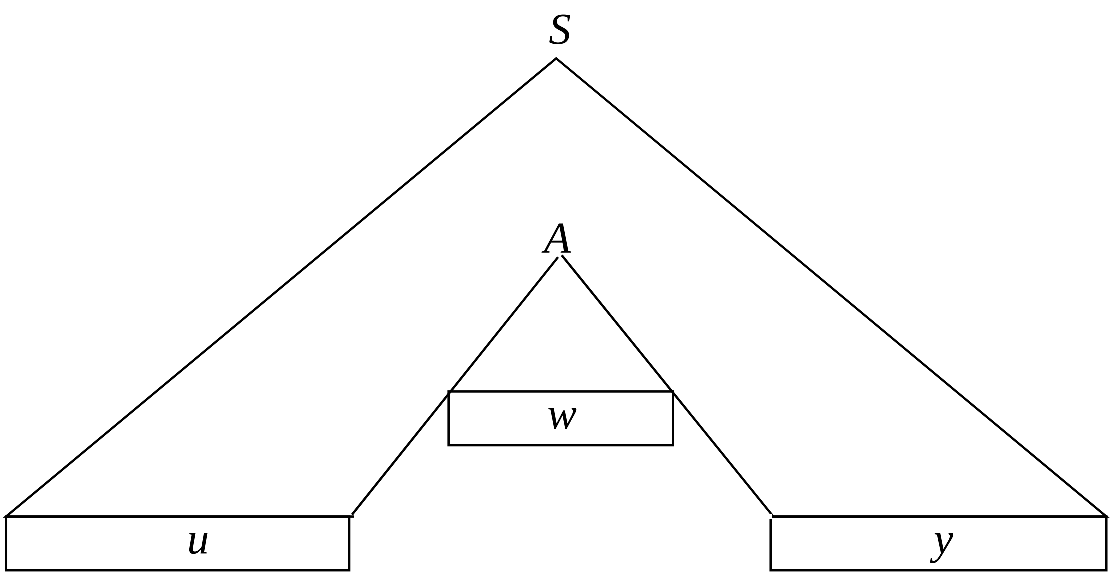
Propiedad de anidamiento doble-lema de bombeo: introducción



- Si se repite un no terminal (A) podremos identificar en la cadena correspondiente la estructura: $uvwxy$

Existencia de lenguajes no independientes del contexto

Propiedad de anidamiento doble-lema de bombeo: introducción

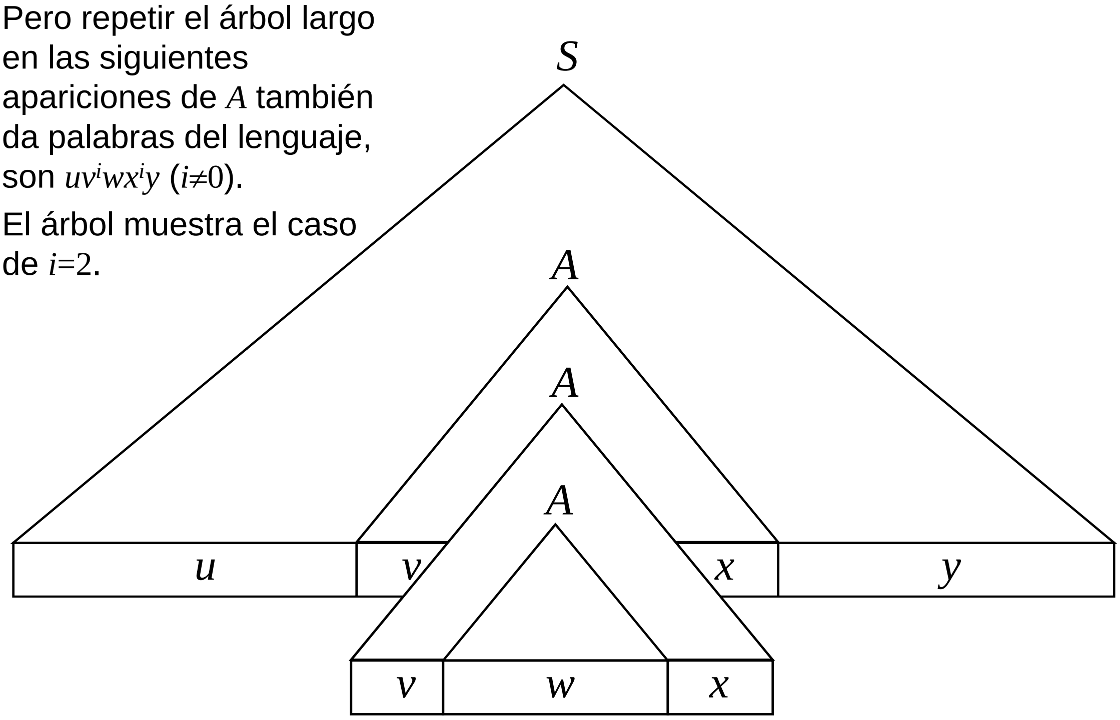


- Al sustituir el árbol "largo" por el "corto" a partir de la primera aparición de A se obtiene una palabra que también es del lenguaje y es uv^0wx^0y, es decir, el resultado de repetir simultáneamente 0 veces v y x.

Existencia de lenguajes no independientes del contexto

Propiedad de anidamiento doble-lema de bombeo: introducción

- Pero repetir el árbol largo en las siguientes apariciones de A también da palabras del lenguaje, son uv^iwx^iy ($i\neq 0$).
- El árbol muestra el caso de $i=2$.



Lema de bombeo

Lema de bombeo, propiedad de anidamiento doble: enunciado

- Lo anterior se puede formalizar en el **lema de bombeo**:

	$\forall$ L lenguaje ind. cont. sin λ e infinito $\Rightarrow$
(el nº asociado con $2^{\|\Sigma N\|}$)	$\exists\ n > 0 \in N \mid$
(todas las cadenas largas para tener no terminales “repetidos”)	$\forall\ z \in L$,, $\|z\| \geq n\ \wedge$
(identificación de una descom-posición donde hay no term. “repetido”: vwx trozo con repetidos long. $\leq n$; lo “repetido” no es λ)	$\exists\ u, v, w, x, y \mid$ $z = uvwxy\ \wedge$ $\|vwx\| \leq n\ \wedge$ $\|vx\| > 0\ \wedge$
(Cualquier número de repeticiones produce palabras del lenguaje)	$\forall i \in \{0,1,...\} \Rightarrow uv^i wx^i y \in L$

Lema de bombeo

Propiedad de anidamiento doble y lema de bombeo

- La propiedad expuesta antes y que cumplen los lenguajes independientes del contexto se llama **propiedad de anidamiento doble**.
- El **lema de bombeo** enuncia que los lenguajes independientes del contexto (infinitos y sin λ) cumplen la **propiedad de anidamiento doble.**

Lema de bombeo

Corolario: negación del lema de bombeo

- De igual forma a lo visto para lenguajes regulares, el interés de este resultado está en el significado de su negación.
- Si llamamos $p_a(L)$ a la propiedad de anidamiento doble sobre un lenguaje L, el enunciado visto afirma

$$\forall\ L \text{ lenguaje independiente del contexto} \Rightarrow p_a(L)$$

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

$$\neg p_a(L) \Rightarrow L \text{ no es lenguaje independiente del contexto}$$

- 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

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 derivaciones "largas")	$\forall\ n > 0 \in N \mid$
(hay alguna cadena larga como para no termin. "repetidos")	$\exists\ z \in L \text{ ,, } \lvert z \rvert \geq n\ \wedge$
(para todas las posibles descom-posiciones, para todas las posibles cadenas bombeables)	$\forall\ u, v, w, x, y \mid$ $z = uvwxy\ \wedge$ $\lvert vwx \rvert \leq n\ \wedge$ $\lvert vx \rvert > 0\ \wedge$
(Puede encontrarse una repetición de las subcadenas tal que la pa-labra completa no es del lenguaje)	$\exists\ i \in \{0,1,\ldots\} \Rightarrow uv^iwx^iy \notin L$
	$\Rightarrow L$ no es un lenguaje independiente del contexto

Lema de bombeo

Corolario: negación del lema de bombeo

- Observaciones (análogas a lo visto en lenguajes regulares):
  - 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

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

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

Lema de bombeo

Corolario: negación del lema de bombeo

- 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 ($z \in L$,, $|z| \geq n$) ya que es suficiente con que exista una
    - 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 las subcadenas bombeables produzcan una palabra que no esté en el lenguaje.

Lema de bombeo

Ejemplo 1

- Demostraremos que el siguiente lenguaje no es independiente del contexto:

$$L=\{0^n 1^n 2^n \mid n \geq 1\}$$

  - Sea $m \in N$ cualquiera,
  - Se considera la cadena $z=0^m 1^m 2^m$
    - Que pertenece al lenguaje
    - Y puede ser utilizada para la demostración ya que $|z| = 3m \geq m$
  - Se consideran todas las descomposiciones de interés $\forall\ u, v, w, x, y \mid z= uvwxy \wedge |vwx| \leq m \wedge |vx| > 0$. Se tienen que considerar los dos siguientes casos deducidos del hecho de que, ya que $|vwx| \leq m$ no puede abarcar por completo el bloque de 1's entero conteniendo simultáneamente 0's y 2's:

Lema de bombeo

Ejemplo 1

- **a)** vwx no contiene 2's.

$$z = uvwxy = \underbrace{\overbrace{00\ldots0}^{m \text{ veces}}\overbrace{11\ldots1}^{m \text{ veces}}}_{vwx}\overbrace{22\ldots22}^{m \text{ veces}}$$

- En este caso, como $|vx| > 0$ vx tiene algún 0 o algún 1 (supongamos que tiene $k \neq 0$).
- Veamos que para $i=0$ $uv^i wx^i y \notin L$
- Eso es cierto ya que $\#_2(uwy)=m$ y,
  - o bien $\#_1(uwy)=m\text{-}k \neq \#_2(uwy)$
  - o bien $\#_0(uwy)=m\text{-}k \neq \#_2(uwy)$
- Y los dos casos llevan a que $uv^i wx^i y \notin L$.

Ejemplo 1

- **b)** vwx no contiene 0's.

$$z = uvwxy = \overbrace{00\ldots0}^{m\text{ veces}}\underbrace{\overbrace{11\ldots1}^{m\text{ veces}}\overbrace{22\ldots2}^{m\text{ veces}}}_{vwx}$$

- En este caso, como $|vx| > 0$ vx tiene algún 1 o algún 2 (supongamos que tiene $k \neq 0$).
- Veamos que para $i=0$ $uv^iwx^iy \notin L$
- Eso es cierto ya que $\#_0(uwy)=m$ y,
  - o bien $\#_1(uwy)=m\text{-}k \neq \#_0(uwy)$
  - o bien $\#_2(uwy)=m\text{-}k \neq \#_0(uwy)$
- Y los dos casos llevan a que $uv^iwx^iy \notin L$.

11.2

Propiedades de cierre

Propiedades de cierre de los lenguajes independientes del contexto

Introducción

- De manera parecida a lo estudiado para lenguajes regulares, es interesante plantearse qué operaciones de lenguajes independientes del contexto aseguran que el resultado sigue siéndolo y cuáles no.
- A continuación se tratará con más detalle este aspecto

Propiedades de cierre de los lenguajes independientes del contexto

Cerrados respecto $\cup$, . , *

- Podemos afirmar la siguiente propiedad

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

- Es decir

$\forall\ L_1, L_2$ lenguajes independientes del contexto $\Rightarrow$

$L_1 \cup L_2$ es independiente del contexto

$L_1 . L_2$ es independiente del contexto

$\forall\ L$ lenguaje independiente del contexto $\Rightarrow$

L^* es independiente del contexto

Propiedades de cierre de los lenguajes independientes del contexto

No cerrados respecto $\cap$

- Los lenguajes independientes del contexto no son cerrados para la intersección.
- Eso significa que aunque L_1, L_2 sean lenguajes independientes del contexto $L_1 \cap L_2$ no necesariamente lo es.
- Para demostrarlo basta encontrar un contraejemplo.
- Se ha estudiado previamente que el siguiente lenguaje no es independiente del contexto.

$$\{0^n 1^n 2^n \mid n \geq 1\}$$

- Por otro lado, se puede ver que la gramática que se deduce de las siguientes reglas de producción

$S::=AB$

$A::=0A1 \mid 01$

$B::=2B \mid 2$

- Genera el lenguaje $\{0^n 1^n 2^j \mid n,j \geq 1\}$
- Y la que se deduce de las siguientes reglas de producción

$S::=AB$

$A::=0A \mid 0$

$B::=1B2 \mid 12$

- Genera el lenguaje $\{0^j 1^n 2^n \mid n,j \geq 1\}$

Propiedades de cierre de los lenguajes independientes del contexto

No cerrados respecto $\cap$

- Además, también se puede comprobar que se cumple

$$\{0^n 1^n 2^n \mid n \geq 1\} = \{0^n 1^n 2^j \mid n,j \geq 1\} \cap \{0^j 1^n 2^n \mid n,j \geq 1\}$$

- Es decir,
  - La existencia de las gramáticas independientes del contexto prueba que los dos *sublenguajes* son independientes del contexto.
  - Por lo tanto, el lenguaje de partida puede expresarse como la intersección de dos lenguajes independientes del contexto.
  - Si los lenguajes independientes del contexto fueran cerrados para la intersección $\{0^n 1^n 2^n \mid n \geq 1\}$ también lo sería.
  - Como no lo es, se puede concluir que el resultado enunciado es cierto.

Cerrados respecto $\cap$ con un lenguaje regular

- Podemos afirmar la siguiente propiedad

Los lenguajes independientes del contexto son cerrados respecto a la intersección con un lenguaje regular

  - Es decir

$\forall\ L_1$ lenguaje independiente del contexto, L_2 lenguaje regular $\Rightarrow$ $L_1 \cap L_2$ es independiente del contexto

No cerrados respecto - y complementario

- Los lenguajes independientes del contexto no son cerrados respecto al complementario.
  - Eso significa que aunque L sea lenguaje independiente del contexto su complementario no necesariamente lo es.
  - Podemos razonar esto teniendo en cuenta las propiedades de los conjuntos y el hecho visto previamente de que tampoco son cerrados respecto a la intersección.
  - Se conoce, de la teoría de conjuntos, el siguiente resultado (leyes de DeMorgan)

$$L_1 \cap L_2 = \overline{\overline{L_1} \cup \overline{L_2}}$$

  - Si el complementario fuera cerrado, como la unión de lenguajes independientes del contexto lo es, también lo sería la intersección.
  - Pero acabamos de estudiar que la intersección no lo es, por lo tanto, tampoco puede serlo el complementario.

Propiedades de cierre de los lenguajes independientes del contexto

No cerrados respecto - y complementario

- Vamos a presentar un resultado previo

Para cualquier alfabeto, su lenguaje universal es independiente del contexto

- Es decir

$\forall\Sigma,\ \ \Sigma^*$ es independiente del contexto

- Para ver esto es suficiente recordar que

$$\Sigma=\{a_1,\ldots,a_n\} \Rightarrow \Sigma^*=L((a_1+\ldots+a_n)^*)$$

  - Es decir, realmente es regular y, por tanto (como los regulares también son independientes del contexto) es independiente del contexto

Propiedades de cierre de los lenguajes independientes del contexto

No cerrados respecto - y complementario

- Apoyándonos en el resultado anterior, podemos afirmar también el siguiente:

Los lenguajes independientes del contexto no son cerrados respecto a la diferencia.

- Eso significa que aunque L_1, L_2 sean lenguajes independientes del contexto L_1-L_2 no necesariamente lo es.
  - Para ver esto es suficiente darse cuenta de que el complementario puede definirse en función de la diferencia de conjuntos

$$\forall L \text{ definido sobre } \Sigma \quad \overline{L}=\Sigma^*-L$$

  - Si la diferencia de lenguajes independientes del contexto siempre produjera un lenguaje independiente del contexto, en el caso concreto de esta diferencia

$$\Sigma^*-L=\overline{L}$$

  - También podríamos afirmar que sería independiente del contexto, pero acabamos de estudiar que el complementario no siempre lo es, ya que no es una operación cerrada en este conjunto, por lo que la diferencia tampoco lo es.

Propiedades de cierre de los lenguajes independientes del contexto

Cerrados respecto - de uno regular

- Podemos afirmar la siguiente propiedad

Cuando se quita de un lenguaje independiente del contexto uno regular, el resultado es independiente del contexto

- Es decir

$\forall\, L_1$ lenguaje independiente del contexto, L_2 lenguaje regular $\Rightarrow$ L_1-L_2 es independiente del contexto

Propiedades de cierre de los lenguajes independientes del contexto

Cerrados respecto reflexión

- Podemos afirmar la siguiente propiedad

La reflexión es una operación cerrada en el conjunto de los lenguajes independientes del contexto

- Es decir

$\forall\, L$ lenguaje independiente del contexto $\Rightarrow L^{-1}$ es independiente del contexto

2

Los autómatas a pila y las gramáticas independientes del contexto son equivalentes