2 Lenguajes formales y autómatas

(Actividades Formativas)

Actividad formativa 1: Lema de bombeo

Ejemplo:

Sea el lenguaje L de *"a"* seguidas por el mismo número de *"b"*

$$L = \{\varepsilon, ab, aabb, aaabbb, aaaabbbb, \ldots\}$$

Proponemos una constante $n = k$

Proponemos una cadena $w = a^k b^k$, $w \in L$

Partimos w en $x = a^{(k-1)}$, $y = a$, $z = b^k$

Revisamos las restricciones $|xy| \leq n$ y $y \neq \varepsilon$

$$|xy| = n \leq n$$

$$y \neq \varepsilon$$

Actividad formativa 1: Lema de bombeo

Ejemplo:

Revisamos la última restricción $xy^mz \in L, \forall\ m \in N$

Con $x = a^{(k-1)}, y = a, z = b^k$ y m = 1, el lenguaje cumple. Sin embargo, se debe cumplir para todos los valores naturales de m.

Si m=2, vemos que se altera el número de “a”.

$$xy^2z$$
$$a^{(k-1)}a^2b^k$$

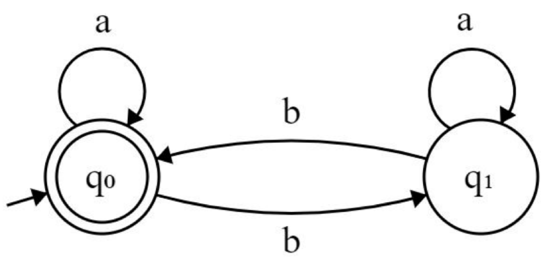
Con k = 2, se genera la cadena aaabb, la cual no pertenece al lenguaje L, por lo tanto, el lenguaje no es regular.

Actividad formativa 2

Leer el capítulo 3.2 “Autómatas finitos y expresiones regulares” de Hopcroft, Motwani & Ullman (2007)

Actividad formativa 3: Autómatas finitos

- *Ejemplo:* Identifiquemos los elementos del siguiente autómata finito:



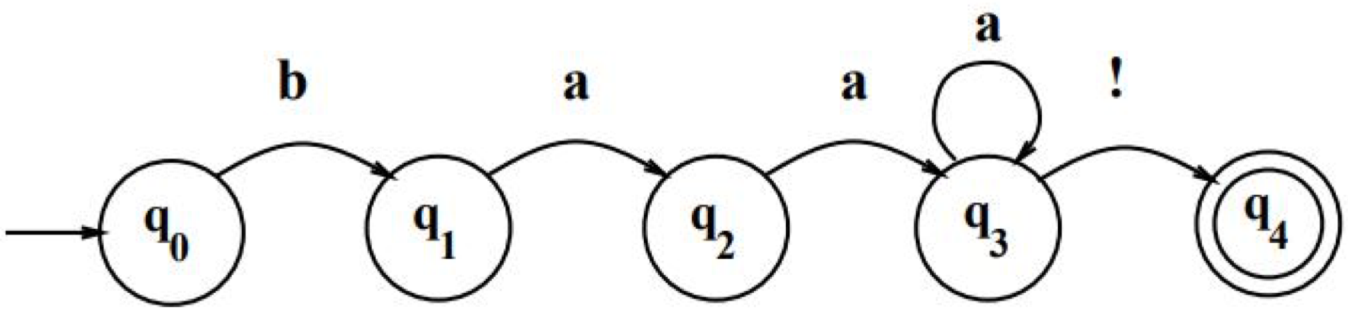
- $\delta = \begin{cases} (q_0, a) \rightarrow q_0 \\ (q_0, b) \rightarrow q_1 \\ (q_1, a) \rightarrow q_1 \\ (q_1, b) \rightarrow q_0 \end{cases}$

El conjunto de reglas o transiciones que se representan como arcos del grafo.

- Q = $\{q_0, q_1\}$, son los nodos de un grafo y representan los estados.
- Σ = {a,b}, el vocabulario, son los símbolos sobre los arcos.
- q_0 es el estado al que apunta la flecha sin símbolo, indicando que es el estado INICIAL.
- q_0 es el estado encerrado con dos círculos, lo que significa que es el estado FINAL. En este caso el estado inicial y el final son el mismo.

Actividad formativa 3: Autómatas finitos

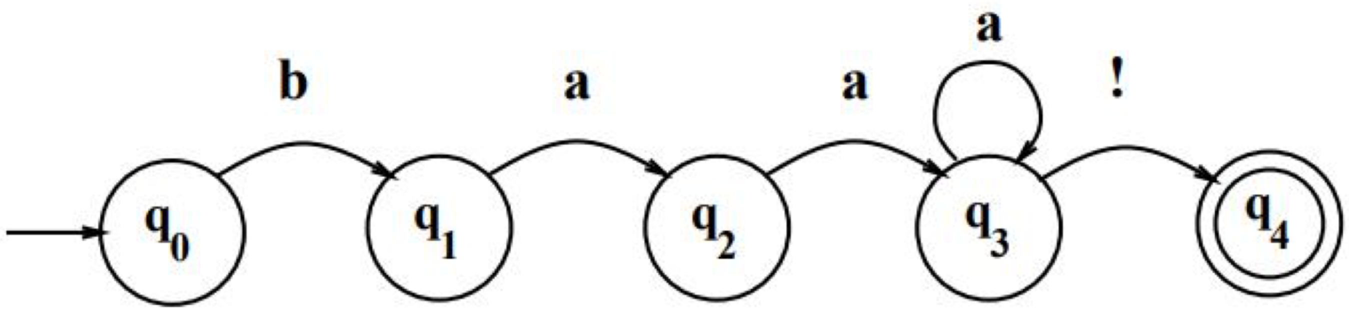
- *Ejemplo:* Ahora, dado un autómata, identifiquemos sus elementos, el lenguaje que representa, y la expresión regular equivalente.



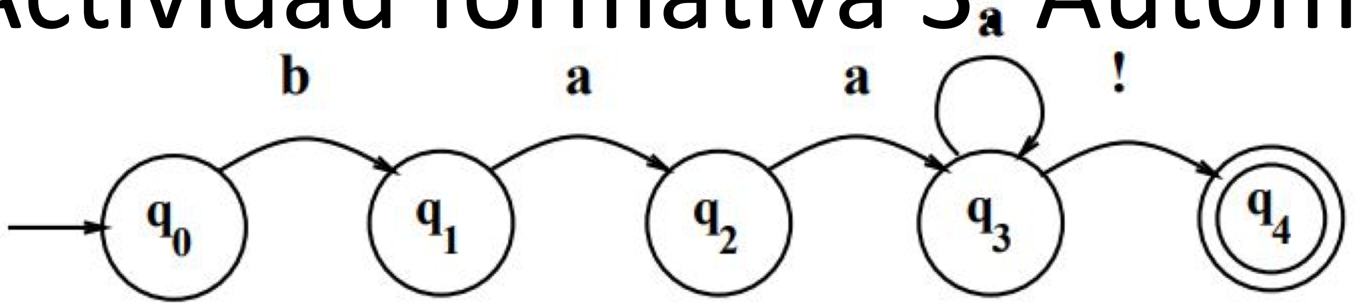
- Comencemos encontrando la 5-Tupla que define el autómata.

Actividad formativa 3: Autómatas finitos

- $Q = \{q_0, q_1, q_2, q_3, q_4\}$
- $\Sigma = \{a, b, !\}$
- $q_0 = q_0$
- $A = \{q_4\}$
- $\delta = \begin{cases} (q_0, b) \to q_1 \\ (q_1, a) \to q_2 \\ (q_2, a) \to q_3 \\ (q_3, a) \to q_3 \\ (q_3, !) \to q_4 \end{cases}$



Actividad formativa 3: Autómatas finitos



- Para obtener la expresión regular, proponemos cadenas que sean aceptadas y no aceptadas por el autómata:

Cadenas aceptadas	Cadenas rechazadas
baa!	aa
baaa!	baa
baaaaa!	aa!
baaaaaaaaaaaa!	aba!

- El lenguaje que describe el autómata es aquel con cadenas que empiezan con "b" seguido de dos o más "a", seguido de "!".
  La expresión regular sería: **baaa*!**

Actividad formativa 3: Autómatas finitos

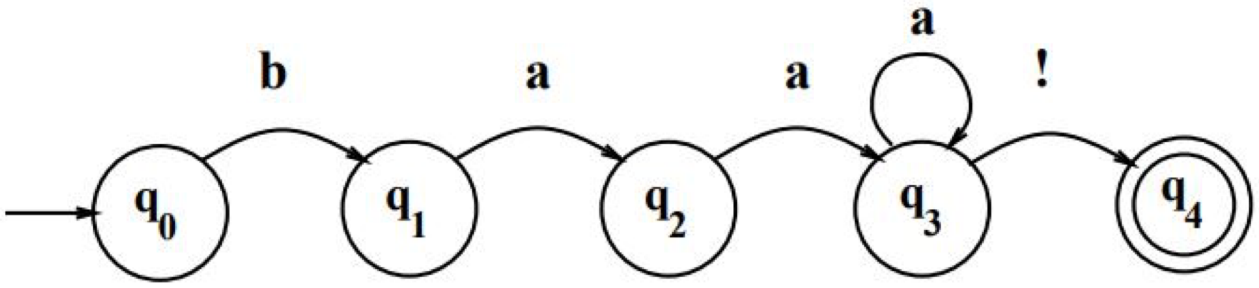
Encontremos ahora la gramática regular del autómata, siguiendo el algoritmo visto anteriormente.

1. Asociamos S al estado q_0
2. Asociamos a cada estado, un símbolo no terminal: q_1 con Q, q_2 con T, q_3 con U.
3. Obtenemos las producciones correspondientes: $S \rightarrow bQ$, $Q \rightarrow aT$, $T \rightarrow aU$, $U \rightarrow aU$, $U \rightarrow !$

El estado inicial no es final, por lo que terminamos el algoritmo. Así, nuestra gramática queda definida como:

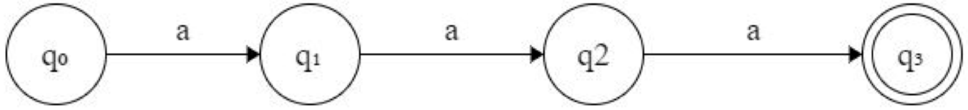
$G = (Q, \Sigma, R, q)$ donde
$Q = \{S,Q,T,U\}$
$\Sigma = \{a,b,!\}$
$q = \{S\}$

$R = \{$
$S \rightarrow bQ$
$Q \rightarrow aT$
$T \rightarrow aU$
$U \rightarrow aU$
$U \rightarrow !$
$\}$

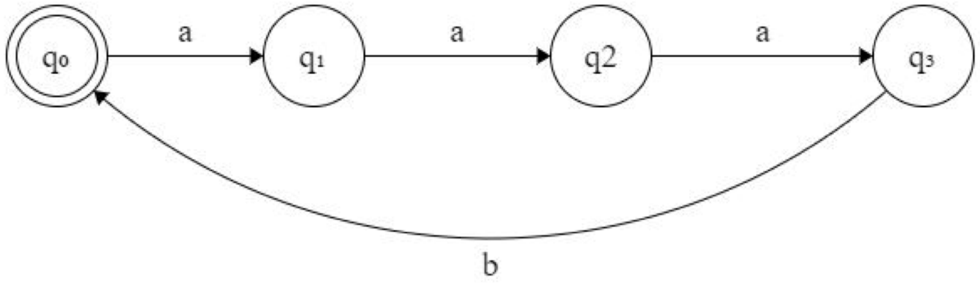


Actividad formativa 4: Autómatas finitos

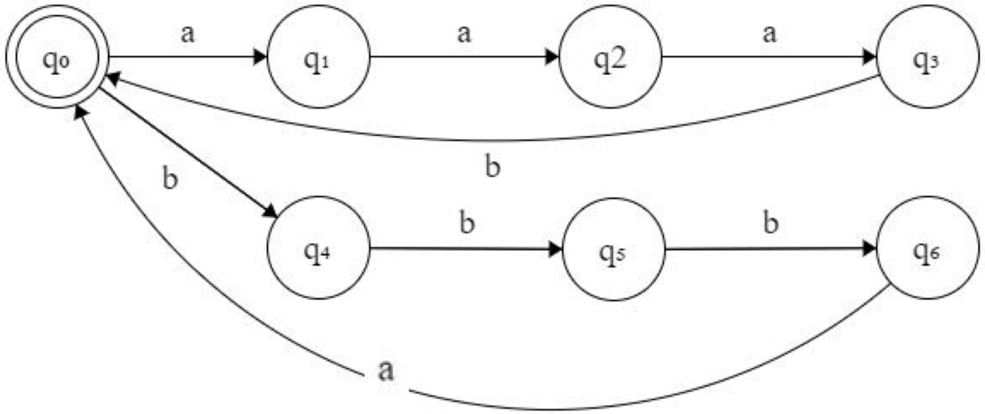
- A continuación se presentan ejemplos de autómatas con sus respectivas expresiones regulares. El estado inicial es q_0.



aaa



(aaab)*



(aaab|bbba)*

Actividad formativa 4: Máquinas de Turing

Ejemplo:

- Construir una máquina de Turing que verifique si el número de 1s en una palabra es par.

$$MT = (Q, \Sigma, \Gamma, q_0, \delta, F)$$

$Q = \{q_0, q_1\}$ $\Sigma = \{0,1\}$ $\Gamma = \{0,1,\vdash,B\}$ $F = \{q_0\}$

δ:

$$\begin{aligned}
\delta(q_0,1) &= (q_1, B, \rightarrow) \\
\delta(q_0,0) &= (q_0, B, \rightarrow) \\
\delta(q_1,1) &= (q_0, B, \rightarrow) \\
\delta(q_1,0) &= (q_1, B, \rightarrow)
\end{aligned}$$

Actividad formativa 4: Máquinas de Turing

Ejemplo:

- Probemos nuestra máquina de Turing con w = 11101

Inicio:

	1	1	1	0	1	B	B
	$\uparrow$ q_0						

Paso 1:

	B	1	1	0	1	B	B
		$\uparrow$ q_1					

Paso 2:

	B	B	1	0	1	B	B
			$\uparrow$ q_0				

Paso 3:

	B	B	B	0	1	B	B
				$\uparrow$ q_1			

Paso 4:

	B	B	B	B	1	B	B
					$\uparrow$ q_1		

Paso 5:

	B	B	B	B	B	B	B
						$\uparrow$ q_0	

La máquina de Turing ha terminado de leer la cadena en el estado de aceptación.
Por lo tanto la máquina acepta a w.

Actividad formativa 5: Síntesis

Responda las siguientes preguntas:

¿Qué es una expresión regular?

¿Para qué es útil una expresión regular?

¿Qué es un lenguaje regular, y cómo se comprueba que un lenguaje NO es regular?

¿Qué es un autómata finito?

¿Por qué todos los lenguajes finitos son regulares?

Mencione los elementos de la jerarquía de Chomsky.

¿Cuál es la diferencia entre lenguaje formal y lenguaje natural?

¿Qué mide la notación Landau?

Mencione un ejemplo no mencionado de un problema no computable.

¿Crees que el procesamiento de lenguaje natural es una tarea computable? ¿Por qué?

Bibliografía

- Jurafsky, Daniel & James H. Martin. (2008). Speech and Language Processing: *An Introduction to Natural Language Processing, Computational Linguistics and Speech Recognition* (2nd edition), Prentice Hall. (Chapter 1: Introduction and Chapter 13: Language and Complexity)
- Sipser M. (1996). Introduction to the Theory of Computation (1st ed.). International Thomson Publishing. Boston, MA. (Chapter 1: Regular Languages)
- Manning, C. & Schütze, H. (1999). *Foundations of Statistical Natural Language Processing*, MIT Press. Cambridge, MA. (Chapter 1, Section 1.2.3: Language and cognition as probabilistic phenomena & Chapter 1, Section 1.3: The Ambiguity of Language: Why NLP Is Difficult)
- Mijangos, V. (2017) Procesamiento de Lenguaje Natural. https://sites.google.com/site/victormijangoscruz/cursos/procesamiento-de-lenguaje-natural [5 de enero de 2018] (Capítulo 1: ¿Qué es la lingüística?)
- Polanco, D. (2000). *Evaluación y mejora de un sistema automático de análisis sintagmático*. Proyecto Fin de Carrera, UPM. Madrid. (Capítulo 2: Lenguajes naturales y lenguajes formales)
- Meza, Iván. Revisando la jerarquía de Chomsky. Disponible en línea: http://turing.iimas.unam.mx/~ivanvladimir/slides/lfya/chomsky.html#/20 [17 de enero del 2018]
- Jurado Málaga, E. (2008). *Teoría de autómatas y lenguajes formales*. Universidad de Extremadura. Servicio de Publicaciones. (Capítulo 8: Gramáticas de tipo 0 y 1, Capítulo 9: Computabilidad y Máquinas de Turing.)
- Arenas, M. (2009). *Complejidad Computacional.* Pontificia Universidad Católica de Chile Curso en línea disponible en: http://marenas.sitios.ing.uc.cl/iic3242-14/clases.html [20 de enero del 2018] (Capítulo 1: Máquinas de Turing)
- Longley, J. (2011). *Turing machines and linear bounded automata*. Universidad de Edinburgh. Disponible en línea: http://www.inf.ed.ac.uk/teaching/courses/inf2a/slides/2011_inf2a_L29_slides.pdf [20 de enero del 2018]
- Hopcroft, J. E., Motwani, R., & Ullman, J. D. (2007). Teoría de autómatas, lenguajes y computación. *Pearson educación*. (Capítulo 3: Lenguajes y expresiones regulares)
- Becerra-Bonache, L., Bel-Enguix, G., Jiménez-López, M.D., Martín-Vide, C. (2018), Mathematical Foundations: Formal Grammars and Languages. In Mitkov, R., Oxford Handbook of Computational Linguistics, 2nd Edition. Oxford, Oxford University Press. https://drive.google.com/file/d/1be6_EOm1WtUM9lAmwIQh5C8kliX2Wz4F