TEORÍA DE LA COMPUTACIÓN

Grado en Ingeniería Informática

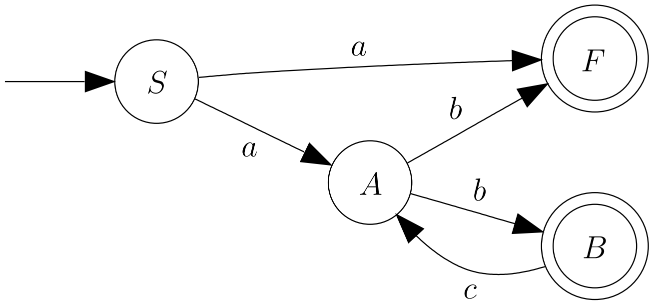
Soluciones del Boletín de Ejercicios nº 3

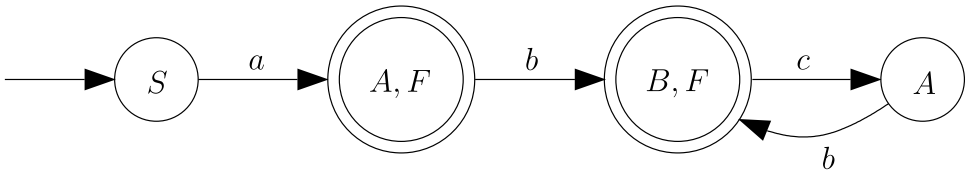
Gramáticas regulares y lenguajes regulares

41. Construya un autómata finito determinista que acepte el lenguaje generado por la siguiente gramática:

$$S \to a\,A \mid a$$
$$A \to b\,B \mid b$$
$$B \to c\,A \mid \epsilon$$

Solución:





Gramáticas independientes del contexto

42. Indique cuál de las siguientes gramáticas genera el lenguaje $\{a^n b^m \mid 0 \leq m \leq n \leq 3m\}$:

a) $S \to Saaab \mid aSaab \mid aaSab \mid aaaSb \mid aaabS \mid \epsilon$

b) $S \to AAASb \mid \epsilon \qquad A \to a \mid \epsilon$

c) $S \to aSb \mid aaSb \mid aaaSb \mid \epsilon$

Solución:

a) No es correcta, porque mezcla símbolos.

b) No es correcta, porque la regla $A \to \epsilon$ podría anular todas las Aes y generar cadenas con más $b$es que $a$es.

c) Es la respuesta correcta.

43. Escriba una gramática independiente del contexto que genere cadenas de la forma uv, donde $u \in (a \cup b)^*$, $v \in (c \cup d)^*$ y $|u| \geq |v|$.

Solución:

$$\begin{array}{l} S \to U\,S\,V \mid U\,S \mid \epsilon \\ U \to a \mid b \\ V \to c \mid d \end{array}$$

44. Escriba una gramática independiente del contexto que genere $\{a^i b^j \mid i \leq j \leq 2i\}$.

Solución:

$$S \to a\,S\,b \mid a\,S\,b\,b \mid \epsilon$$

45. Escriba una gramática independiente del contexto que genere $\{a^i b^j c^k \mid i + j = k\}$.

Solución:

$$\begin{array}{l} S \to a\,S\,c \mid B \mid \epsilon \\ B \to b\,B\,c \mid \epsilon \end{array}$$

La regla $S \to \epsilon$ se podría eliminar, ya que $S \to B$ y $B \to \epsilon$.

46. Encuentre una gramática independiente del contexto para el lenguaje sobre $\{a, b\}$ que consiste en las cadenas en las cuales la relación entre el número de $a$es y el de $b$es es de tres a dos (ejemplo: *abaab*).

Solución:

Una posibilidad es considerar toas las posibles combinaciones de tres $a$es y dos $b$es (*aaabb*, *aabab*, *aabba*, *abaab*, *ababa*, *abbaa*, *baaab*, *baaba*, *babaa*, *bbaaa*), introducir S entre todos los símbolos, y por último añadir la regla $S \to \epsilon$:

$$S \to SaSaSaSbSbS \mid SaSaSbSaSbS \mid SaSaSbSbSaS \mid \ldots \mid \epsilon$$

¿Existe una solución mejor?

47. Dado el alfabeto $\Sigma = \{p, q, r, \mathcal{V}, \mathcal{F}, \neg, \vee, \wedge, \Rightarrow, \Leftrightarrow, (,)\}$, sea L el conjunto de todas las expresiones lógicas válidas que se pueden construir con los símbolos de Σ (por ejemplo: $p \wedge (q \Rightarrow \mathcal{F}) \wedge \neg r$, $(p \Rightarrow q) \Leftrightarrow (\neg p \vee q)$, $\neg p \vee p \Leftrightarrow \mathcal{V}$, etc.). Escriba una gramática independiente del contexto G tal que $L(G) = L$.

Solución:

$$E \to \neg E \mid E \vee E \mid E \wedge E \mid E \Rightarrow E \mid E \Leftrightarrow E \mid (\ E\ ) \mid \mathcal{V} \mid \mathcal{F} \mid p \mid q \mid r$$

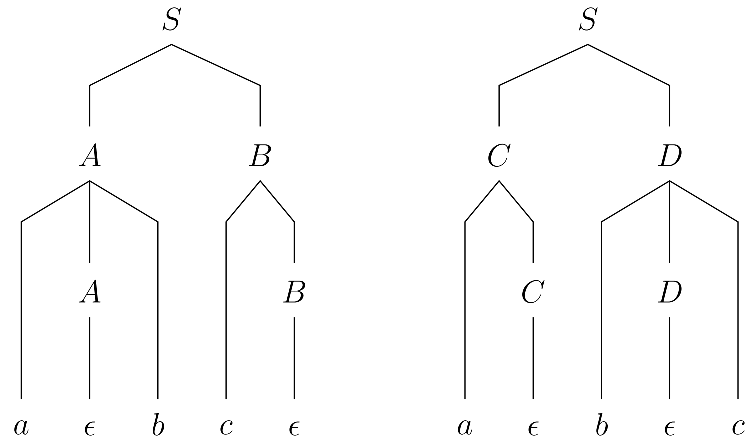
Árboles de derivación y ambigüedad

48. Escriba una gramática independiente del contexto que genere el lenguaje $\{a^i b^j c^k \mid i = j \text{ o bien } j = k\}$. Dicha gramática, ¿es ambigua? Demuéstrelo.

Solución:

$$\begin{array}{l} S \to AB \mid CD \\ A \to aAb \mid \epsilon \\ B \to cB \mid \epsilon \\ C \to aC \mid \epsilon \\ D \to bDc \mid \epsilon \end{array}$$

Esta gramática es ambigua. Cualquier cadena de la forma $a^n b^n c^n$ (por ejemplo, abc) tendrá dos árboles de derivación.



49. Considere el lenguaje $L = \{a^n b^n c^m d^m \mid n, m \geq 0\} \cup \{a^n b^m c^m d^n \mid n, m \geq 0\}$, es decir, L consta de las cadenas de $a^* b^* c^* d^*$ tales que:

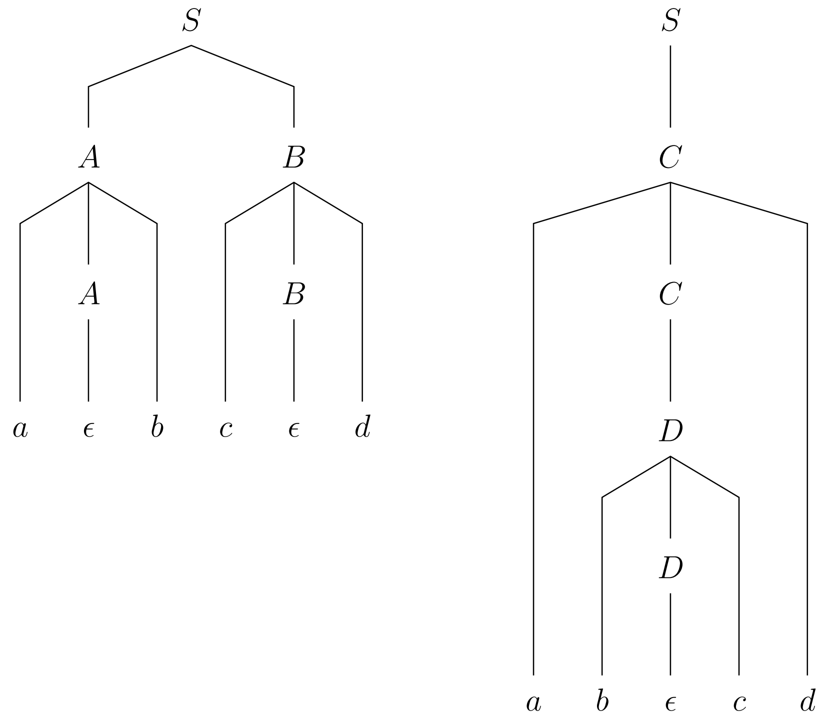
   - o bien $\#a$es $= \#b$es y $\#c$es $= \#d$es,
   - o bien $\#a$es $= \#d$es y $\#b$es $= \#c$es.

   Escriba una gramática independiente del contexto que genere el lenguaje L. Dicha gramática, ¿es ambigua? Demuéstrelo.

   Solución:

$$S \to AB \mid C \qquad A \to aAb \mid \epsilon \qquad B \to cBd \mid \epsilon \qquad C \to aCd \mid D \qquad D \to bDc \mid \epsilon$$

   Esta gramática es ambigua. Cualquier cadena de la forma $a^n b^n c^n d^n$ (por ejemplo, $abcd$) tendrá dos árboles de derivación.



Simplificación de gramáticas independientes del contexto

50. Simplifique tanto como sea posible la siguiente gramática:

$$\begin{aligned}&S \to A \mid aB \mid aCbB\\&A \to B \mid bC \mid bbD\\&B \to C \mid cD \mid ccE\\&C \to \epsilon\end{aligned}$$

<u>Solución</u>:

- Eliminación de ϵ-producciones:

$$\text{anulables} = \{S, A, B, C\}$$

$$\begin{aligned}&S \to A \mid aB \mid a \mid aCbB \mid abB \mid aCb \mid ab \mid \epsilon\\&A \to B \mid bC \mid b \mid bbD\\&B \to C \mid cD \mid ccE\end{aligned}$$

- Eliminación de producciones unitarias:

$$\begin{aligned}&\text{unitario}(S) = \{S, A, B, C\}\\&\text{unitario}(A) = \{A, B, C\}\\&\text{unitario}(B) = \{B, C\}\\&\text{unitario}(C) = \{C\}\\&\text{unitario}(D) = \{D\}\\&\text{unitario}(E) = \{E\}\end{aligned}$$

$$\begin{aligned}&S \to cD \mid ccE \mid bC \mid b \mid bbD \mid aB \mid a \mid aCbB \mid abB \mid aCb \mid ab \mid \epsilon\\&A \to cD \mid ccE \mid bC \mid b \mid bbD\\&B \to cD \mid ccE\end{aligned}$$

- Eliminación de símbolos inútiles:

$$\begin{aligned}&S \to b \mid a \mid ab \mid \epsilon\\&A \to b\end{aligned}$$

- Eliminación de símbolos no accesibles:

$$S \to b \mid a \mid ab \mid \epsilon$$

Propiedades de los lenguajes independientes del contexto

51. Demuestre mediante el lema del bombeo que el lenguaje $\{a^i b^j c^k \mid i \leq j \leq k\}$ no es independiente del contexto.

<u>Solución</u>:

Supongamos que L es un LIC y sea k su constante asociada según el lema del bombeo. Tomemos entonces una cadena $z \in L$ tal que $|z| \geq k$, por ejemplo, $z = a^k b^k c^k$. Las posibles descomposiciones de z en $uvwxy$, tales que $|vwx| \leq k$ y $|v| + |x| > 0$, son:

- La porción vwx está compuesta sólo por $a$es. En este caso, cualquier bombeo uv^iwx^iy, $i \geq 2$, produce cadenas con más $a$es que $b$es y $c$es, y que por tanto no pertenecen a L.
- La porción vwx está compuesta sólo por $b$es. En este caso, cualquier bombeo uv^iwx^iy, $i \geq 2$, produce cadenas con más $b$es que $a$es (lo cual estaría permitido), pero también con más $b$es que $c$es, y que por tanto no pertenecen a L.
- La porción vwx está compuesta sólo por $c$es. En este caso, cualquier bombeo uv^iwx^iy, $i \geq 2$, produce cadenas con más $c$es que $a$es y $b$es (lo cual estaría permitido). Pero el bombeo $i = 0$ produce una cadena con menos $c$es que $a$es y $b$es, y que por tanto no pertenece a L.
- La porción vwx está compuesta por $a$es y por $b$es. Detectamos entonces tres subcasos:
  * La subcadena v está formada sólo por $a$es, y la subcadena x sólo por $b$es. Cualquier bombeo $i \geq 2$ aumenta las $a$es o las $b$es o ambas, pero nunca las $c$es, produciendo cadenas que no pertenencen a L.
  * La subcadena v está formada por $a$es y por $b$es, y la subcadena x sólo por $b$es. Cualquier bombeo $i \geq 2$ hará que las $a$es y las $b$es se mezclen, y las $c$es no aumentan, produciendo cadenas que no pertenencen a L.
  * La subcadena v está formada sólo por $a$es, y la subcadena x por $a$es y por $b$es. Igual que antes, cualquier bombeo $i \geq 2$ mezcla las $a$es y las $b$es, y las $c$es no aumentan, produciendo cadenas que no pertenencen a L.
- La porción vwx está compuesta por $b$es y por $c$es. Detectamos también tres subcasos:
  * La subcadena v está formada sólo por $b$es, y la subcadena x sólo por $c$es. Cualquier bombeo $i \geq 2$ aumenta las $b$es o las $c$es o ambas, y podría hacerlo en una proporción que respetara la restricción del lenguaje. Pero el bombeo $i = 0$ elimina $b$es y/o $c$es, produciendo una cadena que no pertenence a L.
  * La subcadena v está formada por $b$es y por $c$es, y la subcadena x sólo por $c$es. Cualquier bombeo $i \geq 2$ hará que las $b$es y las $c$es se mezclen, produciendo cadenas que no pertenencen a L.
  * La subcadena v está formada sólo por $b$es, y la subcadena x por $b$es y por $c$es. Igual que antes, cualquier bombeo $i \geq 2$ mezcla las $b$es y las $c$es, produciendo cadenas que no pertenencen a L.
- La porción vwx no puede contener $a$es, $b$es y $c$es, ya que entonces tendría k $b$es y algún símbolo más, es decir, tendría longitud al menos $k + 2$, y no verificaría la condición $|vwx| \leq k$.

Así pues, hemos analizado todos los casos de descomposición de z y, en todos ellos, hemos detectado al menos un bombeo que produce cadenas fuera de L, con lo cual se puede concluir que L no es un LIC.

52. Demuestre mediante el lema del bombeo que el lenguaje $\{ww \mid w \in \{a, b\}^*\}$ no es independiente del contexto.

    <u>Solución</u>:

    Supongamos que L es un LIC y sea k su constante asociada según el lema del bombeo. Tomemos entonces una cadena $z \in L$ tal que $|z| \geq k$, por ejemplo, $z = a^kb^ka^kb^k$. Las posibles descomposiciones de z en $uvwxy$, tales que $|vwx| \leq k$ y $|v| + |x| > 0$, son:

    - La porción vwx está compuesta sólo por $a$es de la primera parte. En este caso, cualquier bombeo $i \geq 2$, produce cadenas con más $a$es en la primera parte que en la segunda, y que por tanto no pertenecen a L.

- La porción vwx está compuesta sólo por $b$es de la primera parte. En este caso, cualquier bombeo $i \geq 2$, produce cadenas con más $b$es en la primera parte que en la segunda, y que por tanto no pertenecen a L.
- La porción vwx está compuesta sólo por $a$es de la segunda parte. En este caso, cualquier bombeo $i \geq 2$, produce cadenas con más $a$es en la segunda parte que en la primera, y que por tanto no pertenecen a L.
- La porción vwx está compuesta sólo por $b$es de la segunda parte. En este caso, cualquier bombeo $i \geq 2$, produce cadenas con más $b$es en la segunda parte que en la primera, y que por tanto no pertenecen a L.
- La porción vwx está compuesta por $a$es de la primera parte y por $b$es de la primera parte. En este caso, es útil el bombeo $i = 0$, porque, si desaparecen símbolos, resulta muy sencillo ver que el resultado ya no obedecerá al formato ww. Existen otros bombeos igualmente válidos, pero quizás es más complejo razonar sobre la mezcla de símbolos presente en las cadenas que producen.
- La porción vwx está compuesta por $b$es de la primera parte y por $a$es de la segunda parte. Puede aplicarse el mismo razonamiento que en el caso anterior.
- La porción vwx está compuesta por $a$es de la segunda parte y por $b$es de la segunda parte. Puede aplicarse también el mismo razonamiento que en el caso anterior.
- Ninguna otra descomposición es posible.

Una vez más, hemos analizado todos los casos de descomposición de z y, en todos ellos, hemos detectado al menos un bombeo que produce cadenas fuera de L, con lo cual se puede concluir que L no es un LIC.

53. Demuestre mediante el lema del bombeo que el lenguaje

$$\{w \mid w \in \{a, b, c\}^* \text{ y } \#a\text{es} = \#b\text{es} = \#c\text{es, sin importar el orden}\}$$

no es independiente del contexto.

Solución:

Supongamos que L es un LIC y sea k su constante asociada según el lema del bombeo. Tomemos entonces una cadena $z \in L$ tal que $|z| \geq k$, por ejemplo, $z = a^k b^k c^k$. Las posibles descomposiciones de z en $uvwxy$, tales que $|vwx| \leq k$ y $|v| + |x| > 0$, son:

- La porción vwx está compuesta sólo por $a$es. En este caso, cualquier bombeo $i \geq 2$, produce cadenas con más $a$es que $b$es y $c$es, y que por tanto no pertenecen a L.
- La porción vwx está compuesta sólo por $b$es. En este caso, cualquier bombeo $i \geq 2$, produce cadenas con más $b$es que $a$es y $c$es, y que por tanto no pertenecen a L.
- La porción vwx está compuesta sólo por $c$es. En este caso, cualquier bombeo $i \geq 2$, produce cadenas con más $c$es que $a$es y $b$es, y que por tanto no pertenecen a L.
- La porción vwx está compuesta por $a$es y por $b$es. La mezcla de símbolos estaría permitida, pero cualquier bombeo $i \geq 2$ aumentaría el número de $a$es y $b$es, podría ser incluso en el mismo número, aunque no estaría garantizado, pero, en todo caso, no el de $c$es, produciendo cadenas que no estarían en L.
- La porción vwx está compuesta por $b$es y por $c$es. Nuevamente, la mezcla de símbolos estaría permitida, pero cualquier bombeo $i \geq 2$ aumentaría el número de $b$es y $c$es, podría ser incluso en el mismo número, aunque no estaría garantizado, pero, en todo caso, no el de $a$es, produciendo cadenas que no estarían en L.

– Ninguna otra descomposición es posible.

Tras analizar todos los casos de descomposición de z y, en todos ellos, detectar al menos un bombeo que produce cadenas fuera de L, se puede concluir que L no es un LIC.

54. Demuestre mediante el lema del bombeo que el lenguaje $\{a^{n^2} \,|\, n \geq 0\}$ no es independiente del contexto.

    Solución:

    Supongamos que L es un LIC y sea k su constante asociada según el lema del bombeo. Tomemos $z = a^{k^2}$. Se cumple entonces lo siguiente:

    $$k^2 = |uvwxy| < |uv^2wx^2y| \leq k^2 + k < (k+1)^2$$

    Es decir, el bombeo $i = 2$ produce una cadena cuya longitud está entre dos cuadrados perfectos, y que por tanto no pertenece a L. Así pues, L no es un LIC.

55. Demuestre que los lenguajes independientes del contexto no son cerrados para la operación de complementario.

    Solución:

    Los LIC,s no son cerrados para la operación de intersección. Para verlo, basta con proporcionar un contraejemplo. $L_1 = \{a^i b^j c^j \mid i, j \geq 0\}$ es un LIC, ya que puede ser generado por la GIC:

    $$S \rightarrow AC \qquad A \rightarrow aA \mid \epsilon \qquad C \rightarrow bCc \mid \epsilon$$

    Lo mismo ocurre con $L_2 = \{a^i b^i c^j \mid i, j \geq 0\}$, que puede ser generado por:

    $$S \rightarrow AC \qquad A \rightarrow aAb \mid \epsilon \qquad C \rightarrow cC \mid \epsilon$$

    Sin embargo, $L_1 \cap L_2 = \{a^k b^k c^k \mid k \geq 0\}$, que no es un LIC (puede demostrase mediante el lema del bombeo).

    A partir del resultado anterior, se deduce que el complementario de un LIC, en general, tampoco es un LIC, ya que $L_1 \cap L_2 = \overline{\overline{L_1} \cup \overline{L_2}}$.

56. Razone la verdad o falsedad de la siguiente afirmación: "*Todo subconjunto de un lenguaje independiente del contexto es independiente del contexto*".

    Solución:

    $L_1 = \{a^i b^i c^j \mid i, j \geq 0\}$ es un LIC.
    $L_2 = \{a^i b^i c^i \mid i \geq 0\}$ no es un LIC.
    Dado que $L_2 \subset L_1$, el enunciado es falso.

Autómatas de pila

57. Construya un autómata de pila que acepte el lenguaje $\{w \,|\, w \in \{a, b\}^*$ y w tiene el mismo número de $a$es que de $b$es, sin importar el orden$\}$.

    Solución:

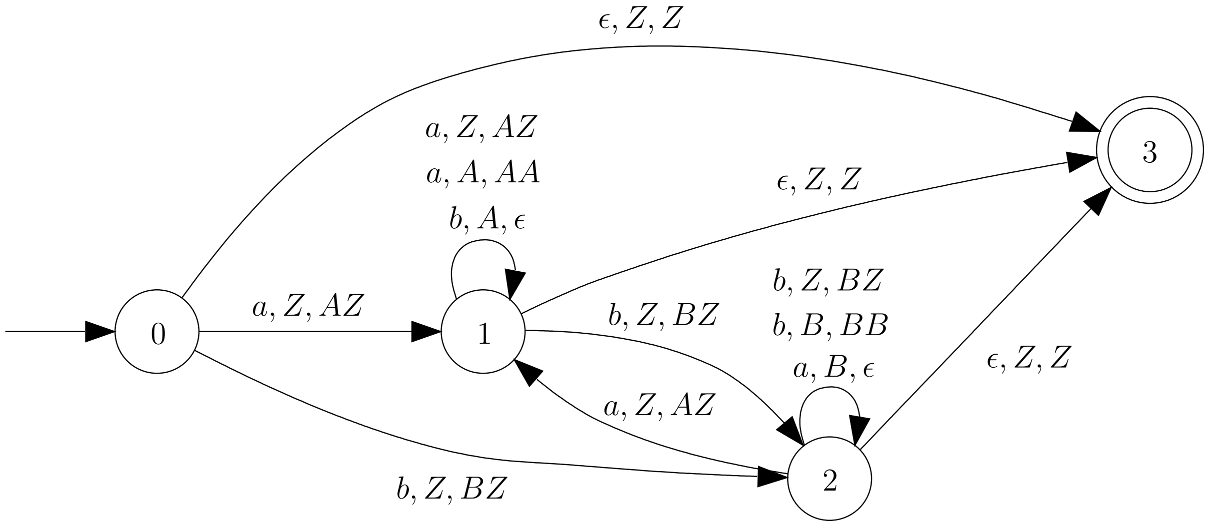
    En el autómta de pila que se muestra a continuación:

    – El estado 1 se corresponde con el estado en el que las $a$es apilan y las $b$es desapilan.
    – El estado 2 se corresponde con el estado en el que las $b$es apilan y las $a$es desapilan.

– Los arcos del estado 1 al 1, con etiquetas a, Z, AZ y a, A, AA podrían resumirse en un único arco con etiqueta a, ϵ, A.

– Los arcos del estado 2 al 2, con etiquetas b, Z, BZ y b, B, BB podrían resumirse en un único arco con etiqueta b, ϵ, B.

– En la pila se utilizan símbolos A y B por claridad, pero podría utilizarse un único símbolo (por ejemplo, X), ya que A y B nunca aparecerán en la pila al mismo tiempo.



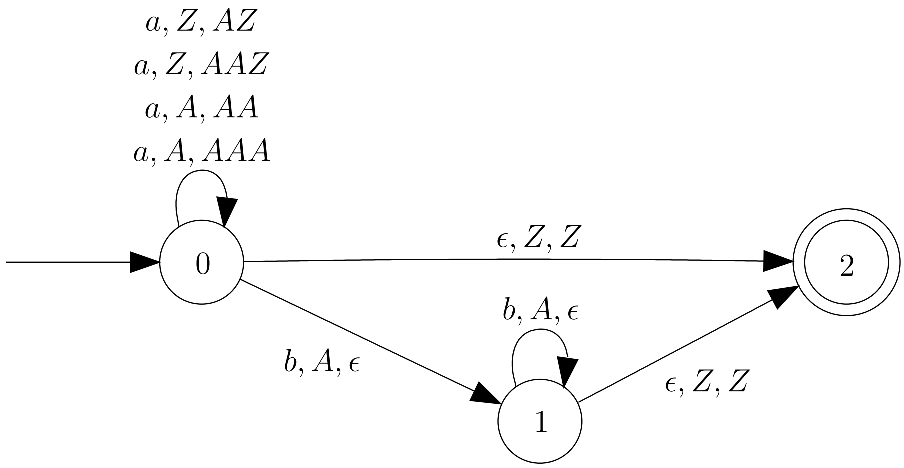
58. Construya un autómata de pila que acepte el lenguaje $\{a^i b^j \mid i \leq j \leq 2i\}$.

<u>Solución</u>:

En el autómta de pila que se muestra a continuación:

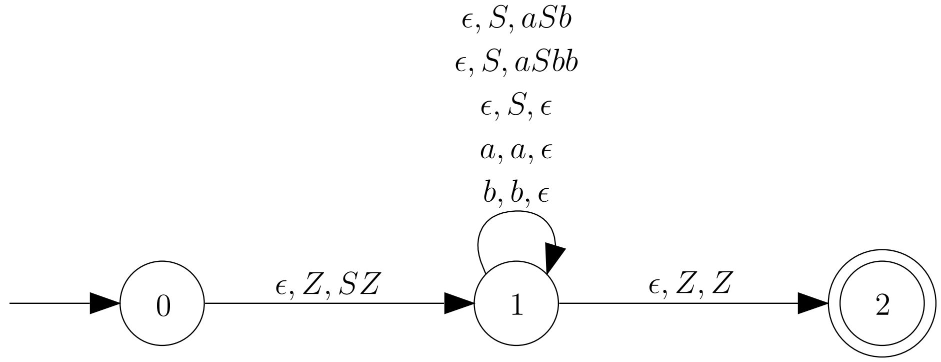
– Los arcos del estado 0 al 0, con etiquetas a, Z, AZ y a, A, AA podrían resumirse en un único arco con etiqueta a, ϵ, A.

– Y los arcos del estado 0 al 0, con etiquetas a, Z, AAZ y a, A, AAA podrían resumirse en un único arco con etiqueta a, ϵ, AA.



Otra solución a este ejercicio viene dada por la obtención de la gramática que genera este lenguaje y por la construcción del autómata de pila correspondiente a dicha gramática:

$$S \rightarrow a\,S\,b \mid a\,S\,b\,b \mid \epsilon$$



59. Construya un autómata de pila que acepte el conjunto de todas las cadenas de ceros y unos tales que ningún prefijo tenga más unos que ceros.

    Solución:

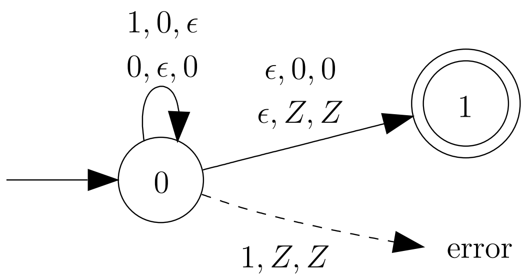
    Denominando L al lenguaje en cuestión, a continuación se muestran algunas secuencias de ceros y unos, indicando si pertenecen o no a dicho lenguaje:

    $$0 \in L \qquad 00 \in L \qquad 01 \in L \qquad 000 \in L \qquad 001 \in L \qquad 010 \in L \qquad \ldots$$

    $$1 \notin L \qquad 10 \notin L \qquad 11 \notin L \qquad 011 \notin L \qquad \ldots$$

    Puede observarse entonces que el autómata de pila que acepte L debe tener el siguiente comportamiento:

    - Los ceros siempre apilan un cero.
    - Los unos siempre desapilan un cero.
    - Los unos no se admiten si la cima de la pila es Z.



Formas normales

60. Obtenga la forma normal de Chomsky y la forma normal de Greibach de la siguiente gramática:

    $$S \to a\,S\,b \mid c\,A\,d$$
    $$A \to S \mid \epsilon$$

    Solución:

    A simple vista, sin necesidad de aplicar los algoritmos de limpieza, podríamos simplificarla en:

    $$S \to a\,S\,b \mid c\,S\,d \mid c\,d$$

La Formal de Normal de Chomsky sería:

$$
\begin{aligned}
&S \to C_a\, D_1 \mid C_c\, D_2 \mid C_c\, C_d \\
&D_1 \to S\, C_b \\
&D_2 \to S\, C_d \\
&C_a \to a \\
&C_b \to b \\
&C_c \to c \\
&C_d \to d
\end{aligned}
$$

Y la Formal de Normal de Greibach sería:

$$
\begin{aligned}
&S \to a\, S\, C_b \mid c\, S\, C_d \mid c\, C_d \\
&C_b \to b \\
&C_d \to d
\end{aligned}
$$