Autómatas y Gramáticas

Grado de Ingeniería Informática

PEC3

Nombre estudiante

Apellidos, Nombre

Universitat Oberta de Catalunya

Presentación

En esta Prueba de Evaluación Continua se trabajan los conceptos básicos de la asignatura sobre alfabetos, palabras y lenguajes, así como distintos conceptos sobre lenguajes incontextuales y autómatas a pila.

Competencias

- Capacidad para utilizar los fundamentos matemáticos, estadísticos y físicos para comprender los sistemas TIC.
- Capacidad para analizar un problema en el nivel de abstracción adecuado a cada situación y aplicar las habilidades y conocimientos adquiridos para resolverlo.

Objetivos

- Conocer las operaciones sobre lenguajes y palabras (concatenación, clausuras) y saber utilizarlas para describir lenguajes complejos.
- Conocer la relación existente entre lenguajes incontextuales, gramáticas incontextuales y autómatas a pila.
- Saber construir autómatas a pila para reconocer las palabras de un lenguaje dado.

Recursos de aprendizaje

Los siguientes recursos son de utilidad para la realización de la PEC:

Básicos

- Módulo didáctico 1. Alfabetos, palabras y lenguajes.
- Módulo didáctico 2. Autómatas finitos y lenguajes regulares.
- Módulo didáctico 3. Gramáticas incontextuales.

Complementarios

- Documentos de ayuda, como exámenes, pruebas de validación y PECs de semestres anteriores publicados en el aula por el profesorado colaborador.
- Los recursos de aprendizaje que aparecen en la bibliografía de estos tres módulos.

Criterios de valoración

La ponderación de los ejercicios es la siguiente:

- Ejercicio 1: 40%
- Ejercicio 2: 20%
- Ejercicio 3: 20%
- Ejercicio 4: 20%

El uso de recursos externos (Internet, material bibliográfico, etc.) ha de estar referenciado en la bibliografía. https://biblioteca.uoc.edu/es/pagina/Como-citar/. La PEC se tiene que realizar de manera **individual** y el trabajo entregado tiene que ser original citando adecuadamente las fuentes bibliográficas utilizadas

La copia o plagio en la realización de la PEC comportará que la prueba sea evaluada con un suspenso (D). En caso de que se haga uso de algún tipo de herramienta que utilice inteligencia artificial, habrá que especificarlo explícitamente, y en el caso de que los contenidos introducidos en la práctica no hayan sido adecuadamente adaptados a la situación propuesta por el enunciado, la prueba será evaluada con una D.

De otra parte, y siempre a criterio de los Estudios de Informática, Multimedia y Telecomunicaciones, la reincidencia en el incumplimiento de este compromiso puede comportar que no se permita al estudiante superar ninguna otra asignatura mediante la evaluación continua ni en el semestre en curso ni en los siguientes.

Formato y fecha de entrega

- La PEC3 con todas las actividades claramente diferenciadas se tiene que entregar en un único documento Word, Open Office o PDF con las respuestas a las preguntas.

- El nombre del fichero tiene que ser: Apellido1-Apellido2-Nombre-PEC3.ext donde “ext” hace referencia a la extensión del fichero.

- Este documento se debe de entregar en el espacio de Entrega PEC3 del aula antes de las **23:59 del martes 27 de mayo de 2025**.

- **No se aceptarán entregas fuera de la fecha límite.**

Enunciado

Nombre y Apellidos: ____________________________________

Observación: Las soluciones de los ejercicios se deben desarrollar empleando el mismo convenio que los usados en los materiales, es decir, la notación matemática o lenguaje empleado en la resolución de los ejercicios tiene que ser el mismo que el empleado en los materiales de la asignatura (a no ser que en el enunciado del ejercicio se especifique lo contrario).

Tal y como se recuerda en todos los ejercicios, <u>las transiciones de los autómatas deben mostrarse en formato tabla</u>, tal y como aparece en la página 53 del módulo de gramáticas incontextuales y autómatas con pila.

<u>No se aceptarán grafos de autómatas con pila</u>, ya que no forman parte del contenido de esta asignatura.

Ejercicio 1 (40 %)

Dado el alfabeto ∑= {a, b, c}, sea L el lenguaje formado por todas aquellas cadenas que cumplen las siguientes condiciones:

- Las cadenas empiezan y terminan por el símbolo c.
- El número de símbolos a en las cadenas debe ser impar.
- El número de símbolos b en las cadenas es igual al doble de símbolos a.
- Todos los símbolos a’s de la cadena aparecen antes de los símbolos b’s.
- No hay condiciones en cuanto al número de símbolos c que puede haber en la cadena.

1.1 Determina cuáles de las palabras siguientes pertenecen al lenguaje:

{cabbc, ab, cacbbc, cbabc, caaabcbbbbbc, abb,cccabbccc, cac, ccaccbcbc, cc}

Solución:

Cadenas que pertenecen a L: cabbc, cacbbc, caaabcbbbbbc, cccabbccc, ccaccbcbc

Cadenas que no pertenecen a L: ab, abb, cbabc, cc, cac

1.2 Define el autómata a pila que reconoce el lenguaje por **estado final.**

Solución:

Se define el autómata a pila M = ({q0, q1, q2, q3, qf}, {a, b, c}, {z_0}, δ, q0, z_0, {qf}) donde la función de transición δ viene definida por la siguiente tabla de transiciones:

	c/z_0	c/c	a/c	c/a	a/a	b/a	λ/z_0
q0	$(q0,\ cz_0)$	$(q0,\ c_)$	(q1,aa)	∅	∅	∅	∅
q1	∅	∅	∅	(q1,a)	(q2,aaa)	(q3, λ)	∅
q2	∅	∅	∅	(q2,a)	(q1,aaa)	∅	∅
q3	$(q4,\ z_0)$	∅	∅	(q3,a)	∅	(q3, λ)	∅
q4	$(q4,\ z_0)$	∅	∅	∅	∅	∅	$(qf,\ z_0)$ $(q0,\ z_0)$
qf	∅	∅	∅	∅	∅	∅	∅

- Se comienza en el estado q0 reconociendo una c que se apila sobre el símbolo de fondo de pila. A continuación, se pueden reconocer más c´s en q0 sin necesidad de apilarlas al no existir condición numérica sobre el número de c´s que tiene la cadena (solo se pide que empiece por c y termine por c). En el mismo estado q0 se puede reconocer una a, en cuyo caso se apilan 2 a´s con el objetivo de cumplir la condición de que habrá el doble de bs que de á´s, y se transita al estado q1.

- El estado q1 representa que se ha reconocido hasta el momento un número impar de a´s. En este estado, se pueden reconocer un número indeterminado de c´s, en cuyo caso se permanece en q1 y no se produce cambios sobre la cima de la pila, o bien reconocer una a, debiendo apilarse en la cima 2 a´s más la a que ya estaba en la cima y transitar al estado q2. Por último, estando en q1 se podría reconocer una b, de manera que se desapilará una a de la cima y se transitará al estado q3.

- El estado q2 representa que se ha reconocido hasta el momento un número para de a´s. En este estado, se pueden reconocer un número indeterminado de c´s, en cuyo caso se permanece en q2 y no se produce cambios sobre la cima de la pila, o bien reconocer una a, debiendo apilarse en la cima 2 a´s más la a que ya estaba en la cima y transitar al estado q1.

- El estado q3 representa el estado donde se reconocen las b´s y c´s restantes de la cadena. Así se pueden reconocer un número indeterminado de c´s, en cuyo caso se permanece en q3 y no se producen cambios sobre la cima de la pila siempre que en la cima haya una a, pues si hubiera el símbolo de fondo de pila, entonces se transitará al estado q4. Así mismo, en el estado q3 se pueede reconocer b´s siempre que en la cima de la pila haya una a, de manera que se desapilará la a de la cima y se mantiene en el estado q3.

- El estado q4 representa el estado de aceptación donde se terminan de aceptar el resto de c´s que pudiera contener la cadena. No se apila ninguna c, y se mantiene el símbolo de pila. Una vez terminadas de reconocer todas las c´s se pasa al estado final qf.

Tal como se puede observar, el diseño del autómata impide que se puedan reconocer más a´s, una vez que se empiezan a reconocer b´s. Además también se impide que se pueda empezar a reconocer b´s si no se han reconocido antes un número impar de a´s, y no se puede alcanzar el estado final si no se han desapilado antes todas las a´s apiladas previamente (el doble de las reconocidas dado que por cada reconocida se apilan 2 a´s en vez de una). Por último, se pueden reconocer c´s en un número indeterminado en cualquier estado, con la única condición de que c es el primero y último símbolo que se reconoce en el autómata.

1.3 Encuentra las secuencias de las **descripciones instantáneas** de la palabra más larga del apartado 1.1 que pertenezca al lenguaje, hasta la aceptación de la palabra. Realizad la misma operación con la palabra más larga del apartado 1.1 que **no** pertenezca al lenguaje.

Solución:

caaabcbbbbbc $\in L$

(q0,caaabcbbbbbc,z_0) ⊦(q0, aaabcbbbbbc,cz_0) ⊦(q1, aabcbbbbbc,aaz_0) ⊦(q2, abcbbbbbc,aaaaz_0)

⊦(q1, bcbbbbbc,aaaaaaz_0) ⊦(q3, cbbbbbc, aaaaaz_0) ⊦(q3, bbbbbc, aaaaaz_0) ⊦(q3, bbbbc, aaaaz_0)

⊦(q3, bbbc, aaaz_0) ⊦(q3, bbc, aaz_0) ⊦(q3, bc, az_0) ⊦(q3, c, z_0) ⊦(qf, λ, z_0)

cbabc $\notin L$

(q0, cbabc,z_0) ⊦(q0, babc,cz_0)

Desde q0 no hay ninguna transición definida para el símbolo de entrada *b* y el símbolo de pila c. No podemos alcanzar el estado final y la palabra no es aceptada.

1.4 ¿Qué transiciones añadirías para que el autómata anterior pudiera reconocer L^+?

Solución:

Es suficiente con añadir la transición siguiente:

$\delta\left(q4, \lambda, z_0\right) = (q_0, z_0)$

Esta transición se indica en rojo en la tabla de transiciones.

1.5 ¿El autómata del apartado 1.2 es determinista? Después de añadir las transiciones del apartado 1.4, ¿el nuevo autómata es o no es determinista?

Solución:

Antes de añadir la transición, el autómata no era determinista porque no se satisface la condición:

$$\text{Si } \delta(q, \lambda, z) \neq \emptyset, \text{ entonces } \delta(q, a, z) = \emptyset\ \forall a \in \sum$$

Por ejemplo:

$\delta\left(q_4, \lambda, z_0\right) = (q_f, z_0)$ pero $\delta\left(q_4, c, z_0\right) \neq \emptyset$ ya que $\delta\left(q_4, c, z_0\right) = (q_4, z_0)$

La transición que hemos añadido es no determinista ya que para un mismo estado, un símbolo de entrada y un símbolo de pila, podemos llegar a dos estados distintos:

$$\delta\left(q_4, \lambda, z_0\right) = \{\left(q_0, z_0\right), \left(q_f, z_0\right)\}$$

$$\left|\delta(q_4, \lambda, z_0)\right| > 1$$

Ejercicio 2 [20%]

Considera los siguientes lenguajes:

a) L= { $a^{n+m} b^{m+t} a^t b^n$ | n, t > 0, m ≥ 0}

b) L={ ww^Rw | w ∈ $\{0, 1\}^*$}

Demuestra si son libres del contexto o no. En caso de ser libre del contexto proporciona un autómata a pila que lo reconozco y en caso de no serlo demuéstralo usando el lema del bombeo.

Solución:

a) Es libre del contexto y puede ser reconocido mediante el siguiente autómata a pila.

	a/z_0	a/A	b/A	b/B	a/B	λ/z_0
qo	(qo,Az_0)	(qo,AA) $(q1,AA)$	$(q3,BA)$	$\emptyset$	$\emptyset$	$\emptyset$
$q1$	$\emptyset$	$(q1,AA)$	$(q2,\lambda)$	$\emptyset$	$\emptyset$	$\emptyset$
$q2$	$\emptyset$	$\emptyset$	$(q2,\lambda)$ $(q3,BA)$	$\emptyset$	$\emptyset$	$\emptyset$
$q3$	$\emptyset$	$\emptyset$	$\emptyset$	$(q3,BB)$	$(q4,\lambda)$	$\emptyset$
$q4$	$\emptyset$	$\emptyset$	$\emptyset$	$(q5,\lambda)$	$(q4,\lambda)$	$\emptyset$
$q5$	$\emptyset$	$\emptyset$	$\emptyset$	$(q5,\lambda)$	$\emptyset$	$(q6,z_0)$
$q6$	$\emptyset$	$\emptyset$	$\emptyset$	$\emptyset$	$\emptyset$	$\emptyset$

b) No es libre del contexto y se va a demostrar con el lema del bombeo.

Supongamos que el lenguaje es libre del contexto y por tanto debe cumplir el lema del bombeo. Sea k la constante del lema y consideremos la cadena s=$0^k1^k1^k0^k0^k1^k$= $0^k1^{2k}0^{2k}1^k$, se tiene que |s|≥k y además sϵL. Por tanto se cumplen las condiciones del lema y s=uvxyz con:

|vy|≥0

|vxy|≤k

$\forall\ i\geq 0\ uv^ixy^iz\in L$

Se pueden dar los siguientes casos:

a) Si v y/o y contienen símbolos del primer bloque de 0´s.

Como |vxy|≤k entonces ni v ni y podrán contener 0´s del segundo bloque de 2k ceros. Si se considera la cadena $uv^0xy^{0i}z= uxz$ donde uxz tendrá la forma $0^{i_1}1^{i_2}0^{2k}1^k$, donde i_1 < k e i_2 ≤ 2k. Si $uxz \in L$ entonces debería ser de la forma ww^Rw.
Dado que uxz tiene al menos longitud 5k, la primera w debe comenzar con un bloque de i_1 < k ceros seguido de algunos unos. Entonces w^Rw contendría un bloque de a lo sumo $2i_1$ < 2k ceros, pero uxz contiene un bloque de 2k ceros → Contradicción.

b) Si v y/o y contienen símbolos del primer bloque de 2k 1´s.

Como |vxy|≤k entonces ni v ni y podrán contener 1´s del segundo bloque de k unos. Si se considera la cadena $uv^0xy^{0i}z= uxz$ donde uxz tendrá la forma $0^{i_1}1^{i_2}0^{i_3}1^k$**,** donde i_1 ≤ k**,** i_2 < 2k**,** i_3 ≤ 2k**.** Si $uxz \in L$ entonces debería ser de la forma ww^Rw.
Dado que uxz tiene al menos longitud 5k, la última w debe terminar con el bloque de k unos**,** precedido por ceros. Entonces ww^R contendría un bloque de 2k unos,
pero uxz contiene un bloque de $\mathbf{i_2}$ < 2k unos → Contradicción.

c) Si v y/o y contienen símbolos del segundo bloque de 2k 0´s.

Como |vxy|≤k entonces ni v ni y podrán contener 0´s del primer bloque de k ceros. Si se considera la cadena $uv^0xy^{0i}z= uxz$ donde uxz tendrá la forma $0^k1^{i_1}0^{i_2}1^{i_3}$, donde i_1 ≤ 2k**,** i_2 < 2k**,** i_3 ≤ k**.** Si $uxz \in L$ entonces debería ser de la forma ww^Rw.
Dado que uxz tiene al menos longitud 5k, la primera w debe comenzar con los k ceros, seguidos por unos. Entonces w^Rw contendría un bloque de 2k ceros,
pero uxz contiene un bloque de $\mathbf{i_2}$ < 2k ceros → Contradicción.

d) Si v y/o y contienen símbolos del segundo bloque de k 1´s.

Como |vxy|≤k entonces ni v ni y podrán contener 1´s del primer bloque de 2k unos. Si se considera la cadena $uv^0xy^{0i}z= uxz$ donde uxz tendrá la forma $0^k1^{2k}0^{i_1}1^{i_2}$, donde i_1 ≤ 2k y i_2 < k**.** Si $uxz \in L$ entonces debería ser de la forma ww^Rw.
Dado que uxz tiene al menos longitud 5k, la última w debe terminar i_2 < k unos, precedido de ceros. Entonces ww^R contendría un bloque de a lo sumo $2i_2$ < 2k unos,
pero uxz contiene un bloque de 2k unos → Contradicción**.**

Como se llega a contradicción en todos los casos, el lenguaje no puede ser libre de contexto

Ejercicio 3 [20%]

Considera la gramática siguiente, expresada en FNG (Forma Normal de Greibach):

$G = (V, T, S, P)$ donde $V = \{S, A, B, C, D, E\}$ y $T = \{x, y, z\}$

P:

$S \to xAC$

$A \to xAC | xDBDE | xDDE$

$B \to xDBDE | xDDE$

$C \to x$

$D \to y$

$E \to z$

3.1 ¿Cuál es el lenguaje generado por la gramática del enunciado?

Solución:

Para hallar la solución obtendremos todas las posibles derivaciones por producciones:

En cuanto a B:

$B \to xyyz$

$B \to xDBDE \to xyByz \to xyxy .... B .... yzyz \to (xy)^n (yz)^n$

En cuanto a A:

$A \to xAC \to xAx \to xx ... A ... xx \to x^k A x^k \to x^k (xy)^m (yz)^m x^k$

$A \to xDBDE \to xyByz \to (xy)^m (yz)^m$

$A \to xDDE \to xyyz$

En cuanto a S:

$S \to xAC \to xAx \to x^s (xy)^m (yz)^m x^s$

Por tanto, el lenguaje generado por la gramática G es el siguiente:

$$L(G) = \{x^s (xy)^m (yz)^m x^s : s > 0 , m > 0\}$$

3.2 Encontrad la tabla de transiciones de un autómata con pila equivalente a la gramática anterior y definid formalmente el autómata resultante:

Solución:

$$M = (Q = \{q\}, \ = \{x, y, z\}, Z = \{S, A, B, C, D, E\}, \sigma, q, S, F = \emptyset)$$

Aplicaremos, tal y como hemos visto en los materiales, el procedimiento de construcción de un autómata con pila a partir de una gramática incontextual en FNG.

$S \to xAC$

$A \to xAC|xDBDE|xDDE$

$B \to xDBDE|xDDE$

$C \to x$

$D \to y$

$E \to z$

δ	x/S	x/A	x/B	x/C	y/D	z/E
q	(q, AC)	(q, AC) $(q, DBDE)$ (q, DDE)	(q, DDE) $(q, DBDE)$	(q, λ)	(q, λ)	(q, λ)

3.3 Mostrad la secuencia de descripciones instantáneas de las palabras *xxxyyzxx* y *xxxyxyyzyzxx* a partir del autómata que habéis definido en el apartado 3.2 y que acaba con la aceptación de la palabra.

Solución:

$(q, xxxyyzxx,\ S) \vdash (q, xxyyzxx, AC) \vdash (q, xyyzxx, ACC) \vdash (q, yyzxx, DDECC) \vdash (q, yzxx, DECC) \vdash (q, zxx, ECC) \vdash$

$(q, xxxyxyyzyzxx,\ S) \vdash (q, xxyxyyzyzxx, AC) \vdash (q, xyxyyzyzxx, ACC) \vdash (q, yxyyzyzxx, DBDECC) \vdash (q, xyyzy$

Ejercicio 4 [20%]

Dado el siguiente autómata con pila, donde q0 representa el estado inicial, contestad las preguntas siguientes, justificando las respuestas:

	a/z_0	b/a	a/a	b/z_0	λ/z_0
q_0	(q_0, az_0)	(q_0, a)	(q_1, λ)	Ø	Ø
q_1	(q_1, z_0)	Ø	Ø	(q_1, z_0)	(q_2, λ)
q_2	Ø	Ø	Ø	Ø	Ø

4.1 ¿El autómata es determinista?

Solución:

No es determinista porque existe la transición $\delta(q_1, \lambda, z_0)$ y, para ser determinista no podría existir ninguna otra transición para el mismo símbolo de pila y estado, cosa que ocurre con la transición $\delta(q_1, a, z_0)$ y $\delta(q_1, b, z_0)$. En otras palabras, no se satisface la condición:

$$\text{Si } \delta(q, \lambda, z) \neq \emptyset, \text{ entonces } \delta(q, a, z) = \emptyset\ \forall a \in \Sigma$$

4.2 ¿Es verdad que el autómata acepta por pila vacía, un lenguaje incontextual que, en este caso, también es regular?

Solución:

Este autómata reconoce el lenguaje determinado por la expresión regular $ab^*a(a + b)^*$ y, por tanto, es regular y también incontextual. Efectivamente, acepta por pila vacía mediante la transición $\delta(q_1, \lambda, Z_0) = (q_1, \lambda)$, después de la lectura de $(a + b)^*$

4.3 Sea L_1 el lenguaje reconocido por el autómata del enunciado y L_2 el lenguaje representado por la expresión regular $(c + d)^* c(c + d)$. ¿Cómo modificas la tabla de transiciones para que el autómata reconozca el lenguaje $L_1 \cup L_2$? (podéis añadir columnas y filas a la tabla).

Solución:

Hemos señalado en rojo las transiciones que se deberían añadir. Para representar el lenguaje unión, el autómata deberá reconocer la expresión $(c + d)^* c(c + d)$.

	a/z_0	b/a	a/a	b/z_0	c/z_0	d/z_0	λ/z_0

q_0	(q_0, az_0)	(q_0, a)	(q_1, λ)	∅	(q_2, z_0) (q_3, z_0)	(q_2, z_0)	∅
q_1	(q_1, z_0)	∅	∅	(q_1, z_0)	∅	∅	(q_4, λ)
q_2	∅	∅	∅	∅	(q_2, z_0) (q_3, z_0)	(q_2, z_0)	∅
q_3	∅	∅	∅	∅	(q_4, λ)	(q_4, λ)	∅
q_4	∅	∅	∅	∅	∅	∅	∅

4.4 Como en el apartado anterior, consideremos L1 el lenguaje reconocido por el autómata del enunciado. Si $L_2 = \{c^n d^{2n} c^n | n > 0\}$, ¿qué transiciones añadiríais, si es posible, para que el autómata reconociera el lenguaje $L_1 \cup L_2$? Justificad la respuesta.

Solución:

El lenguaje L2 no es incontextual y, por tanto, no se puede representar mediante un autómata con pila. No existe ningún autómata con pila que pueda aceptar el lenguaje $L_1 \cup L_2$.