Autómatas y Gramáticas

Grado de Ingeniería Informática

PEC2

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 y gramáticas regulares e incontextuales

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 utilizarlos para describir lenguajes complejos.
- Conocer la relación existente entre lenguajes regulares/incontextuales y gramáticas regulares/incontextuales.
- Saber construir gramáticas regulares e incontextuales que generen un lenguaje determinado.
- Saber expresar una gramática en cualquiera de las formas simplificadas y normales más habituales.

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 dos módulos.

Criterios de valoración

La ponderación de los ejercicios es la siguiente:

- Ejercicio 1: 20%
- Ejercicio 2: 20%
- Ejercicio 3: 20%
- Ejercicio 4: 10%
- Ejercicio 5: 10%
- Ejercicio 6: 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 PEC2 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-PEC2.ext donde “ext” hace referencia a la extensión del fichero.

- Este documento se debe de entregar en el espacio de Entrega PEC2 del aula antes de las **23:59 del miércoles 30 de abril de 2025**.

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

Enunciado

Nombre y Apellidos: ______________________________

Observación: Las definiciones y la notación formal empleadas en la resolución de los ejercicios tienen que ser los mismos que los empleados en los materiales de la asignatura (a no ser que el enunciado del ejercicio especifique lo contrario).

Ejercicio 1 (20%)

Define una gramática incontextual para cada uno de los siguientes lenguajes:

a) $L = \{ a^i b^j c^k \mid i,j,k >= 0, 2i = j \text{ o } 3j = k \}$
b) $L = \{ a^n b^{2n} c^{2m+1} d^m \mid n,m >= 0 \}$
c) $L = \{ w \mid w \in \{a, b\}^*, |w|_a = |w|_b + 1 \}$
d) $L = \{ w \mid w \in \{a, b\}^*, |w|_a \neq |w|_b \}$

Solución:

a) Sea la gramática G = (N, Σ, P, S) con N = {S, W, X, Y, Z}, S símbolo inicial de la gramática, Σ = {a, b, c} y P el conjunto de reglas siguiente:

$S \rightarrow W \mid X$

$W \rightarrow Wc \mid Y$

$X \rightarrow aX \mid Z$

$Y \rightarrow aYbb \mid \lambda$

$Z \rightarrow bZccc \mid \lambda$

b) Sea la gramática G = (N, Σ, P, S) con N = {S, X, Y}, S símbolo inicial de la gramática, Σ = {a, b, c, d} y P el conjunto de reglas siguiente:

$S \rightarrow XY$

$X \rightarrow aXbb \mid \lambda$

$Y \rightarrow ccYd \mid c$

c) Sea la gramática G = (N, Σ, P, S) con N = {S, X, Y}, S símbolo inicial de la gramática, Σ = {a, b} y P el conjunto de reglas siguiente:

$S \rightarrow aX \mid bSS$

$X \rightarrow bS \mid aY \mid \lambda$

$Y \rightarrow aYY \mid bX$

d) Sea la gramática G = (N, Σ, P, S) con N = {S, A, B, X, Y, Z}, S símbolo inicial de la gramática, Σ = {a, b} y P el conjunto de reglas siguiente:

$S \rightarrow A \mid B$

$A \rightarrow XAX \mid AA \mid a$

$B \rightarrow XBX \mid BB \mid b$

$X \rightarrow aY \mid bZ \mid \lambda$

$Y \rightarrow aYY \mid bX$

$Z \rightarrow aX \mid bZZ$

Ejercicio 2 (20%)

Define formalmente los lenguajes representados por las siguientes gramáticas y razona si el lenguaje generado por la gramática es un lenguaje regular o bien se trata de un lenguaje incontextual no regular. Encuentra una expresión regular que reconozca el lenguaje en caso que sea regular o usa el lema del bombeo para justificar que no existe ninguna expresión regular para el lenguaje.

a) Sea la gramática G = ({A, B, S}, {a, b, c, d}, S, P) con S el símbolo inicial y P el siguiente conjunto de producciones:

$S \rightarrow AB \mid A$

$A \rightarrow aAc \mid b$

$B \rightarrow cBd \mid cd$

b) Sea la gramática G = ({S, A, B, C}, {0, 1}, S, P) con S el símbolo inicial de la gramática y P el siguiente conjunto de producciones:

$S \rightarrow A1BC \mid SS$

$A \rightarrow 0B1 \mid 01$

$B \rightarrow 1A0 \mid 10$

$C \rightarrow 1C \mid 0$

Solución:

a) El lenguaje definido por la gramática es:

$L(G) = \{ a^n b c^{n+m} d^m \mid n,m >= 0 \}$

Podemos ver que no es regular ya que dado el valor de N escogemos la palabra $w = a^N b c^{2N} d^N$, que cumple $w \in L(G)$. Cualquier factorización $w = xyz$ con $|yz| <= N$ y $|y| >= 1$ cumplirá que $y = d^{|y|}$, y por lo tanto cualquier palabra $w' = xy^iz$ con $i > 1$ no pertenece a L(G).

b) El lenguaje definido por la gramática es:

$L(G) = ((01)^+1(10)^+1^*0)^+$

Al poderse representar con una expresión regular se trata de un lenguaje regular.

Ejercicio 3 (20%)

Indica razonadamente si L es regular o independiente del contexto no regular. Encuentra una expresión regular que reconozca el lenguaje en caso que sea regular o usa el lema del bombeo para justificar que no existe ninguna expresión regular para el lenguaje.

a) Sea el lenguaje $L = \{ a^i b^j c^k \mid i >= 2, j > 0, 3 >= k > 0\}$. Indicar razonadamente si el lenguaje L es regular o independiente del contexto no regular.
b) Sea L el lenguaje formado por las cadenas de símbolos del alfabeto $\sum = \{a, b, c\}$ que tienen el mismo número de a's que de b's. Indicar razonadamente si el lenguaje L es regular o independiente del contexto no regular.

Solución:

a) El lenguaje L es regular puesto que se puede generar con la siguiente expresión regular:

$L = aaa^*b^+(c+cc+ccc)$

b) El lenguaje L es independiente del contexto porque se puede expresar con la gramática G = (N, Σ, P, S) con N = {S, A, B}, S símbolo inicial de la gramática, Σ = {a, b, c} y P el conjunto de reglas siguiente:

$S \rightarrow aB \mid bA \mid cS \mid \lambda$

$A \rightarrow aS \mid bAA \mid cA$

$B \rightarrow aBB \mid bS \mid cB$

Por otra parte el lenguaje L no es regular puesto que para cualquier valor de N podemos tomar la palabra w = $a^N b^N$ que pertenece a L. Cualquier factorización w = xyz con |yz| <= N y |y| >= 1 cumplirá que y = $b^{|y|}$, y por lo tanto cualquier palabra w' = xy^iz con i > 1 no pertenece a L(G).

Ejercicio 4 (10%)

Analiza razonadamente si las siguientes gramáticas son ambiguas o no:

a) Gramática 1:

$S \rightarrow 0SS \mid A$

$A \rightarrow 1A \mid B$

$B \rightarrow 2B \mid 2$

b) Gramática 2:

$S \rightarrow aSd \mid AB$

$A \rightarrow aAb \mid ab$

$B \rightarrow Bc \mid c$

Solución:

a) Se trata de una gramática ambigua porque para la cadena "01222" se pueden encontrar dos derivaciones diferentes:

$S \rightarrow 0SS \rightarrow 0AS \rightarrow 01AS \rightarrow 01BS \rightarrow 012BS \rightarrow 0122S \rightarrow 0122A \rightarrow 0122B \rightarrow 01222$
$S \rightarrow 0SS \rightarrow 0AS \rightarrow 01AS \rightarrow 01BS \rightarrow 012S \rightarrow 012A \rightarrow 012B \rightarrow 0122B \rightarrow 01222$

b) Se trata de una gramática no ambigua. El lenguaje que reconoce es el siguiente:

$L(G) = \{ a^{i+j}b^jc^kd^i \mid i >= 0, j,k >= 1 \}$

Dada una palabra del lenguaje w = $a^{i+j}b^jc^kd^i$ con unos valores concretos de i,j,k solo existe una posible derivación para llegar de S a w:

$S \rightarrow (S \rightarrow aSd \text{ i veces}) \rightarrow a^iSd^i \rightarrow a^iABd^i \rightarrow (A \rightarrow aAb \text{ j-1 veces}) \rightarrow a^{i+j-1}Ab^{j-1}Bd^i \rightarrow a^{i+j}b^jBd^i \rightarrow (B \rightarrow Bc \text{ k-1 veces}) \rightarrow a^{i+j}b^jBc^{k-1}d^i \rightarrow a^{i+j}b^jc^kd^i$

Ejercicio 5 (10%)

Considera una gramática G = (N, T , S, P) donde S es el símbolo inicial de la gramática, P es el conjunto de producciones y T = { a, b, (,) }.

a) Se pide definir las producciones de la gramática que permitan generar el lenguaje de las expresiones correctamente parentizadas que contienen el mismo número de a's que de b's dentro de cada paréntesis.
b) Deriva la cadena "aa(b(ab)a)()".

Solución:

a) Se define la gramática G del enunciado con N = {S, X, Y, Z} y con el conjunto de producciones P siguiente:

S → (X)S | bS | aS | λ

X → bY | aZ | (X)X | λ

Y → bYY | aX | (X)Y

Z → aZZ | bX | (X)Z

b) La derivación que lleva a esta cadena es:

S → aS → aaS → aa(X)S → aa(X)(X)S → aa(X)()S → aa(X)() → aa(bY)() → aa(b(X)Y)() → aa(b(X)aX)() → aa(b(X)a)() → aa(b(aZ)a)() → aa(b(abX)a)() → aa(b(ab)a)()

Ejercicio 6 (20%)

a) 1. Aplicando los algoritmos en el orden correcto, encontrad una gramática limpia y pelada, equivalente a la gramática G = (V, T, S, P) con V = {S, A, B, C} y T = {a,b,c}. Definimos P:

S → SABAS | ABA

A → aACAa | C | λ

B → bB | b

C → cC | λ

2. A partir del resultado obtenido en el apartado anterior, encontrad una gramática simplificada en Forma Normal de Chomsky (FNC).

b) Encontrad una gramática simplificada en Forma Normal de Greibach (FNG) equivalente a la gramática G = (V, T, S, P) con V = {S, X, Y}, T = {(,), [,]}. Definimos P:

S → S(X) | S[Y] | λ

X → X[Y] | λ

Y → Y(X) | λ

Solución:

a) (1) Primero eliminamos las reglas vacías:

$S \rightarrow SABAS \mid ABA \mid SBAS \mid SABS \mid SBS \mid AB \mid BA \mid B$
$A \rightarrow aACAa \mid aCAa \mid aAAa \mid aACa \mid aAa \mid aCa \mid aa \mid C$
$B \rightarrow bB \mid b$
$C \rightarrow cC \mid c$

Ahora podemos eliminar las reglas unitarias:

S -> SABAS | ABA | SBAS | SABS | SBS | AB | BA | bB | b
A -> aACAa | aCAa | aAAa | aACa | aAa | aCa | aa | cC | c
B -> bB | b
C -> cC | c

(2) Añadimos las variables X_a, X_b y X_c para obtener la gramática sin reglas que contengan símbolos terminales y no sean unitarias.

S -> SABAS | ABA | SBAS | SABS | SBS | AB | BA | X_bB | b
A -> $X_aACA\ X_a$ | $X_aCA\ X_a$ | $X_aAA\ X_a$ | $X_aAC\ X_a$ | $X_aA\ X_a$ | $X_aC\ X_a$ | $X_a\ X_a$ | X_cC | c
B -> X_bB | b
C -> X_cC | c
$X_a \rightarrow a$
$X_b \rightarrow b$
$X_c \rightarrow c$

Añadimos las variables $Y_1,Y_2,Y_3,Y_4,Y_5,Z_1,Z_2,Z_3,Z_4$ para generar reglas con dos variables en la parte derecha.

S -> SY_1 | AY_4 | SY_2 | SY_6 | SY_5 | AB | BA | X_bB | b
A -> $X_a\ Z_1$ | $X_a\ Z_2$ | $X_a\ Z_4$ | $X_a\ Z_5$ | $X_a\ Z_3$ | $X_a\ Z_6$ | $X_a\ X_a$ | X_cC | c
B -> X_bB | b
C -> X_cC | c
$X_a \rightarrow a$
$X_b \rightarrow b$
$X_c \rightarrow c$
$Y_1 \rightarrow AY_2$
$Y_2 \rightarrow BY_3$
$Y_3 \rightarrow AS$
$Y_4 \rightarrow BA$
$Y_5 \rightarrow BS$
$Y_6 \rightarrow AY_5$
$Z_1 \rightarrow AZ_2$
$Z_2 \rightarrow CZ_3$
$Z_3 \rightarrow AX_a$
$Z_4 \rightarrow AZ_3$
$Z_5 \rightarrow AZ_6$
$Z_6 \rightarrow CX_a$

b) Encontramos primero la forma normal de Chomsky para la gramática, para ello primero limpiamos la gramática:

S → S(X) | S[Y] | S() | (X) | () | S[] | [Y] | []

$X \rightarrow X[Y] \mid X[] \mid [Y] \mid []$
$Y \rightarrow Y(X) \mid Y() \mid (X) \mid ()$

Añadimos las variables necesarias para generar la forma normal de Chomsky:

$S \rightarrow SC_{1X2} \mid SC_{3Y4} \mid SC_1 \mid P_1C_{X2} \mid P_1P_2 \mid SC_3 \mid P_3C_{Y4} \mid P_3P_4$
$X \rightarrow XC_{3Y4} \mid XC_3 \mid P_3C_{Y4} \mid P_3P_4$
$Y \rightarrow YC_{1X2} \mid YC_1 \mid P_1C_{X2} \mid P_1P_2$
$C_{1X2} \rightarrow P_1C_{X2}$
$C_{3Y4} \rightarrow P_3C_{Y4}$
$C_{X2} \rightarrow XP_2$
$C_{Y4} \rightarrow YP_4$
$C_1 \rightarrow P_1P_2$
$C_3 \rightarrow P_3P_4$
$P_1 \rightarrow ($
$P_2 \rightarrow)$
$P_3 \rightarrow [$
$P_4 \rightarrow]$

Empezamos ahora el proceso de crear la forma normal de Greibach. Crearemos las variables A_i necesarias:

$A_1 \rightarrow A_1A_4 \mid A_1A_5 \mid A_1A_8 \mid A_{10}A_6 \mid A_{10}A_{11} \mid A_1A_9 \mid A_{12}A_7 \mid A_{12}A_{13}$
$A_2 \rightarrow A_2A_5 \mid A_2A_9 \mid A_{12}A_7 \mid A_{12}A_{13}$
$A_3 \rightarrow A_3A_4 \mid A_3A_8 \mid A_{10}A_6 \mid A_{10}A_{11}$
$A_4 \rightarrow A_{10}A_6$
$A_5 \rightarrow A_{12}A_7$
$A_6 \rightarrow A_2A_{11}$
$A_7 \rightarrow A_3A_{13}$
$A_8 \rightarrow A_{10}A_{11}$
$A_9 \rightarrow A_{12}A_{13}$
$A_{10} \rightarrow ($
$A_{11} \rightarrow)$
$A_{12} \rightarrow [$
$A_{13} \rightarrow]$

Creamos la B_i auxiliares necesarias y modificamos las reglas para que siempre se cumpla que $A_i \rightarrow A_{j\alpha}$ $i < j$:

$A_1 \rightarrow A_{10}A_6 \mid A_{10}A_{11} \mid A_{12}A_7 \mid A_{12}A_{13} \mid A_{10}A_6B_1 \mid A_{10}A_{11}B_1 \mid A_{12}A_7B_1 \mid A_{12}A_{13}B_1$
$B_1 \rightarrow A_4 \mid A_5 \mid A_8 \mid A_9 \mid A_4B_1 \mid A_5B_1 \mid A_8B_1 \mid A_9B_1$
$A_2 \rightarrow A_{12}A_7 \mid A_{12}A_{13} \mid A_{12}A_7B_2 \mid A_{12}A_{13}B_2$
$B_2 \rightarrow A_5 \mid A_9 \mid A_5B_2 \mid A_9B_2$
$A_3 \rightarrow A_{10}A_6 \mid A_{10}A_{11} \mid A_{10}A_6B_3 \mid A_{10}A_{11}B_3$
$B_3 \rightarrow A_4 \mid A_8 \mid A_4B_3 \mid A_8B_3$
$A_4 \rightarrow A_{10}A_6$
$A_5 \rightarrow A_{12}A_7$
$A_6 \rightarrow A_{12}A_7A_{11} \mid A_{12}A_{13}A_{11} \mid A_{12}A_7B_2A_{11} \mid A_{12}A_{13}B_2A_{11}$
$A_7 \rightarrow A_{10}A_6A_{13} \mid A_{10}A_{11}A_{13} \mid A_{10}A_6B_3A_{13} \mid A_{10}A_{11}B_3A_{13}$
$A_8 \rightarrow A_{10}A_{11}$
$A_9 \rightarrow A_{12}A_{13}$
$A_{10} \rightarrow ($

$A_{11} \rightarrow)$
$A_{12} \rightarrow [$
$A_{13} \rightarrow]$

Ahora eliminamos la primera variable de la parte derecha de cada regla en la que la parte izquierda es A_i en orden inverso:

$A_1 \rightarrow (A_6 \mid (A_{11} \mid [A_7 \mid [A_{13} \mid (A_6B_1 \mid (A_{11}B_1 \mid [A_7B_1 \mid [A_{13}B_1$
$B_1 \rightarrow A_4 \mid A_5 \mid A_8 \mid A_9 \mid A_4B_1 \mid A_5B_1 \mid A_8B_1 \mid A_9B_1$
$A_2 \rightarrow [A_7 \mid [A_{13} \mid [A_7B_2 \mid [A_{13}B_2$
$B_2 \rightarrow A_5 \mid A_9 \mid A_5B_2 \mid A_9B_2$
$A_3 \rightarrow (A_6 \mid (A_{11} \mid (A_6B_3 \mid (A_{11}B_3$
$B_3 \rightarrow A_4 \mid A_8 \mid A_4B_3 \mid A_8B_3$
$A_4 \rightarrow (A_6$
$A_5 \rightarrow [A_7$
$A_6 \rightarrow [A_7A_{11} \mid [A_{13}A_{11} \mid [A_7B_2A_{11} \mid [A_{13}B_2A_{11}$
$A_7 \rightarrow (A_6A_{13} \mid (A_{11}A_{13} \mid (A_6B_3A_{13} \mid (A_{11}B_3A_{13}$
$A_8 \rightarrow (A_{11}$
$A_9 \rightarrow [A_{13}$
$A_{10} \rightarrow ($
$A_{11} \rightarrow)$
$A_{12} \rightarrow [$
$A_{13} \rightarrow]$

Para finalizar hacemos el mismo proceso en orden inverso pero ahora con las variables B_i:

$A_1 \rightarrow (A_6 \mid (A_{11} \mid [A_7 \mid [A_{13} \mid (A_6B_1 \mid (A_{11}B_1 \mid [A_7B_1 \mid [A_{13}B_1$
$B_1 \rightarrow (A_6 \mid [A_7 \mid (A_{11} \mid [A_{13} \mid (A_6B_1 \mid [A_7B_1 \mid (A_{11}B_1 \mid [A_{13}B_1$
$A_2 \rightarrow [A_7 \mid [A_{13} \mid [A_7B_2 \mid [A_{13}B_2$
$B_2 \rightarrow [A_7 \mid [A_{13} \mid [A_7B_2 \mid [A_{13}B_2$
$A_3 \rightarrow (A_6 \mid (A_{11} \mid (A_6B_3 \mid (A_{11}B_3$
$B_3 \rightarrow (A_6 \mid (A_{11} \mid (A_6B_3 \mid (A_{11}B_3$
$A_4 \rightarrow (A_6$
$A_5 \rightarrow [A_7$
$A_6 \rightarrow [A_7A_{11} \mid [A_{13}A_{11} \mid [A_7B_2A_{11} \mid [A_{13}B_2A_{11}$
$A_7 \rightarrow (A_6A_{13} \mid (A_{11}A_{13} \mid (A_6B_3A_{13} \mid (A_{11}B_3A_{13}$
$A_8 \rightarrow (A_{11}$
$A_9 \rightarrow [A_{13}$
$A_{10} \rightarrow ($
$A_{11} \rightarrow)$
$A_{12} \rightarrow [$
$A_{13} \rightarrow]$

Si analizamos la gramática resultante podemos ver que no está pelada ya que las variables, A_2,A_3,A_4,A_5,A_8,A_9 no son accesibles desde A_1 y por lo tanto pueden ser eliminadas. Además A_1 y B_1 son equivalentes por lo que podemos unirlas:

$A_1 \rightarrow (A_6 \mid (A_{11} \mid [A_7 \mid [A_{13} \mid (A_6A_1 \mid (A_{11}A_1 \mid [A_7A_1 \mid [A_{13}A_1$
$B_2 \rightarrow [A_7 \mid [A_{13} \mid [A_7B_2 \mid [A_{13}B_2$
$B_3 \rightarrow (A_6 \mid (A_{11} \mid (A_6B_3 \mid (A_{11}B_3$
$A_6 \rightarrow [A_7A_{11} \mid [A_{13}A_{11} \mid [A_7B_2A_{11} \mid [A_{13}B_2A_{11}$
$A_7 \rightarrow (A_6A_{13} \mid (A_{11}A_{13} \mid (A_6B_3A_{13} \mid (A_{11}B_3A_{13}$
$A_{10} \rightarrow ($
$A_{11} \rightarrow)$
$A_{12} \rightarrow [$

$A_{13} \rightarrow]$