

TC2020 – Matemáticas Computacionales

Examen 2

Nombre: _____
Nombre: _____
Nombre: _____

Este examen está pensado para resolverse con el nuevo equipo (*Season 2*). Revisen con calma lo que se pide. Es bastante trabajo, por lo que se sugiere que consideren bien sus tiempos. Al momento de contestar, intenten ser lo más explícitos posible: se calificará con base en lo que esté escrito, y se considerará el proceso aún cuando la respuesta final esté errada. Recuerden que pueden revisar material de la clase, el libro de texto o sus notas. Buena suerte.

1. Expresiones Regulares (36 %)

Describan cada uno de los siguientes lenguajes utilizando **expresiones regulares** genéricas o PCRE; *your choice*. En dado caso de que el lenguaje no se pueda describir con una ER, demuestren el porqué. Posteriormente, e independientemente de si tiene ER o no, escriban dos ejemplos de palabras aceptadas para cada lenguaje.

- a) El lenguaje de las palabras en $\{a, b\}$ que empiezan con a y terminan con b .
- b) El lenguaje de las palabras en $\{0, 1\}$ que son múltiplos de 100.
- c) El lenguaje de las palabras en $\{x, y\}$ que son de longitud impar o que terminan en xxx .
- d) El lenguaje de las palabras en $\{0, 1, 2\}$ que tengan un 1 en el centro, y sean simétricas.
- e) El lenguaje de las palabras en $\{z : z \in \mathbb{Z}, z \geq 0\}$ que son también denominaciones de billetes o monedas expedidas por el Banco de México y que están en circulación. Omitan centavos y series especiales (Onzas de plata, Centenarios, etc.).
- f) El lenguaje de las palabras en $\{a, b\}$ que tengan un número par de bs . (*Hint: consideren hacer un autómata y luego convertir si les parece muy complicado.*)

2. Gramáticas Regulares (14 % + 4 %)

En HTML, le llamamos “etiquetas” a las *keywords* para abrir o cerrar elementos de la página web. Una etiqueta siempre se escribe entre *brackets* para iniciar el elemento, y la misma etiqueta pero con una diagonal al principio para finalizar el elemento. Las etiquetas `<html>...</html>` sirven para delimitar el contenido de la página web.

- a) Escriban una expresión regular para seleccionar (o sea, que acepte) las *keywords* de inicio y fin de las listas ordenadas (`ol`) y no ordenadas (`ul`), incluyendo los brackets como parte de la selección (4 %)
- b) Conviertan a AF y escriban su definición formal (5 %)
- c) Conviertan a Gramática **Regular** (no es GLC) y escriban su definición formal (5 %)
- d) Escriban el árbol de derivación para dos palabras válidas (+ 4 %)

3. Ambigüedad en GLCs (15 %)

Sea $G = (\{S\}, \{\text{if}, c, \text{then}, \text{else}, x\}, \{S \rightarrow x, S \rightarrow \text{if } c \text{ then } S, S \rightarrow \text{if } c \text{ then } S \text{ else } S\}, S)$ una gramática libre de contexto.

- Demuestren que G es una gramática **ambigua** utilizando dos **árboles de derivación diferentes** para una misma *string* (10 %)
- Expresen cada árbol de derivación en forma secuencial (usando \Rightarrow) (5 %)

4. Lenguajes Libres de Contexto (35 % +2 %)

Identificando anomalías

El *Biologic Space Lab* (BSL) está en problemas nuevamente. Debido a una brecha en un ducto de ventilación, ciertos parásitos lograron entrar a uno de los laboratorios de acceso restringido (Sector 7), en el cual se almacenaban muestras de ADN de especies inteligentes para fines taxonómicos. Algunos de los bancos de datos están corruptos debido a la infección, pero otros aún están a salvo. HQ les ha asignado la tarea de diseñar un sistema de control para identificar aquellas muestras de ADN que aún pueden rescatarse, y les ha enviado la siguiente información:

- Los parásitos tienen la capacidad de alterar el genoma de las especies, copiando y replicando ciertas subcadenas de ADN mediante mutación.
- El parásito replica los genes pertenecientes al número de extremidades inferiores y de globos oculares de manera arbitraria.
- Gracias al análisis de los datos, se determinó que los parásitos sólo están afectando a ciertas especies que consideran 'aptas y balanceadas', que son aquellas que tienen el mismo número de ojos que de piernas.
- Sin embargo, se ha confirmado que el número de genes replicados pertenecientes a los ojos nunca es igual que el de las extremidades inferiores en el mismo individuo—en un individuo mutado, el número final de ojos nunca es igual al número final de piernas.

Además, dada la gran bio-diversidad del BSL, HQ les sugiere lo siguiente:

Tenemos 902,451 muestras en el databank del laboratorio, de las cuáles sólo 487,350 podrían estar infectadas. Nos interesa saber cuáles de esas muestras posiblemente infectadas están sanas para separarlas cuanto antes. Como son tantas especies distintas, creemos que es mejor enfocarse en encontrar diferencias en el número de genes en lugar de buscar una secuencia válida.

Diseñen un sistema de reglas para generar subcadenas de genes de ojos y piernas de individuos sanos.

- Generen el alfabeto de terminales (2 %)
- Describan (con notación de conjuntos o con palabras) el lenguaje aceptado por su gramática (5 %)
- Generen las reglas de su gramática (5 %)
- Describan formalmente su gramática (5 %)
- Usando su gramática, sus símbolos y reglas, escriban el árbol de derivación de un *Dachora* sano, que es una especie de ave que tiene dos piernas y dos ojos (8 %)
- Hagan un Autómata *adecuado* para su GLC y descríbanlo formalmente (5 %)
- Dibujen un *Dachora* (+2 %)

Uniendo mecanismos

HQ quiere tener acceso a las k especies sanas más rápidamente. La sugerencia de uno de los *interns* de algoritmia del BSL es agregar un prefijo a cada cadena sana y ordenar toda la base de datos para conseguir todas las cadenas de especies sanas juntas y cortar más fácilmente.

Generen alguna mecanismo (autómata o gramática; *your choice*) que incluya el sistema completo de tal manera que logre *reconocer* cadenas sanas con todo y el prefijo de ordenamiento. Sugerencias:

- a) Propongan un prefijo para ordenamiento (1 %)
- b) Describan con notación de conjuntos o palabras el lenguaje aceptado por el mecanismo completo (1 %)
- c) Generen las reglas, variables, estados o transiciones necesarias según sea el caso (1 %)
- d) Describan el mecanismo completo formalmente (1 %)
- e) Demuestren que su mecanismo es regular o libre de contexto, según consideren (1 %)

Extra Challenge (+8 %)

Determinen si llevar a cabo la sugerencia del *intern* es viable o no y expliquen por qué. Sugerencias:

- a) ¿Cuáles son los valores máximos y mínimos posibles de k ? (+2 %)
- b) Asuman que se utilizará un algoritmo **eficiente** de ordenamiento sobre las n muestras del laboratorio.
- c) ¿Cuál es el valor de n ? (+1 %)
- d) ¿Cuántas operaciones tomaría encontrar todas las k especies sanas si las buscas entre las n posibles de una por una? (+2 %)
- e) ¿Cuántas operaciones tomaría ordenar las n especies? (+2 %)
- f) ¿Qué es más rápido? ¿La opción d o la opción e? (+1 %)

De acuerdo con el Código de Ética del Tecnológico de Monterrey, nuestro desempeño en esta actividad estará guiado por la integridad académica.