

Problema 1

Se necesita extraer algunas estadísticas de un árbol estructurado de información contenido en un archivo de texto en formato JSON.

La estructura está formada por 2 tipos de nodos, los nodos padre son aquéllos que contienen un conjunto de nodos hijos y los nodos terminales que ya no contienen más nodos y que poseen un conjunto de variables en un diccionario.

Ejemplo de un árbol estructurado de 2 niveles					Tipos de nodos	
Nodo padre	Nodo terminal	var1	var2	Objeto	Nodo padre	Nodo terminal
A	a1	10	20	[{ "name": "A", "children": [{ "name": "a1", "values": { "var1": 10, "var2": 20 } }, { "name": "a2", "values": { "var1": 15, "var2": 25 } }], "name": "B", "children": [{ "name": "b1", "values": { "var1": 100, "var2": 200 } }, { "name": "b2", "values": { "var1": 150, "var2": 250 } }] }]	{ "name": "nombre del nodo", "children": [<conjunto de nodos>] }	{ "name": "nombre del nodo", "values": { "var1": 10, "var2": 50 } }
	a2	15	25			
B	b1	100	200			
	b2	150	250			

Set de datos: <https://drive.google.com/file/d/1D4B7DZPVfq4kENyprY3njdlCJwanPeLO>

Determinar para la estructura de 3 niveles: Región - Provincia - Ciudad

- 1) Extraer el promedio de la variable var1
- 2) Extraer la suma de la variable var2 para la provincia 2
- 3) Extraer el máximo de la variable var1 de la región 4

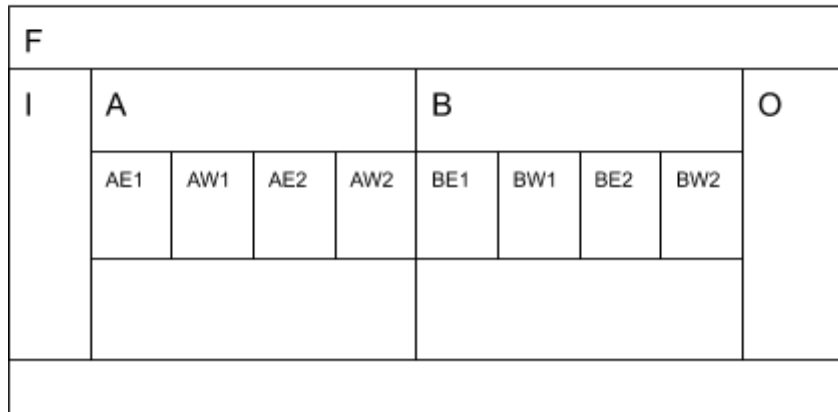
Problema 2

Se necesita determinar estadísticos de operación de camiones que recorren una faena geográfica según la figura.

La faena geográfica (F) posee 4 áreas funcionales: entrada (I), A, B y salida (O).

El área funcional A posee 2 áreas de espera AE1 y AE2, y 2 áreas de trabajo AW1 y AW2.

El área funcional B posee la misma estructura que el área funcional A.



Importante: un camión no necesariamente debe permanecer en todas las áreas descritas, por ejemplo, un ciclo de faena puede ser: I,AE1,AW2,BE1,BW2,O, omitiendo estadía en AW1 y BW1.

Se poseen registros de la fecha y hora de entrada y salida de cada camión que opera en la faena de alrededor de un mes, contenidos en un archivo de datos en formato JSON. Cada registro contiene los siguientes campos:

Set de datos:

<https://drive.google.com/file/d/1N9thMzo09HqeutmmhU2bgcyqePRFM2Dw>

asset : nombre del camión que realiza un ciclo de faena

cycle : número de ciclo de faena para ese camión

zone : nombre de la zona en la que permanece (F,I,A,AE1,etc..)

dt_in : fecha y hora en la que entra en la zona (YYYY-mm-ddTHH:MM:SS) (24h)

dt_out : fecha y hora en la que sale de la zona f(YYYY-mm-ddTHH:MM:SS) (24h)

Determinar:

1. El promedio de tiempo de espera en zonas A y B
2. El porcentaje de ciclos de faena que incluyeron alguna área de trabajo tipo 2
3. Proponer otro análisis de los datos. Se valorará la creatividad y complejidad del análisis.