



Programación – Certamen 2 - Jueves 14 de Noviembre de 2024

En dos tipos de archivos se han recopilado distintos datos acerca de las aves chilenas. El primer archivo, llamado `info.csv`, contiene características de cada especie, siguiendo el siguiente formato:

`id;dimorfismo;migratoria;origen;tamaño;riesgo`

El campo `id` es un identificador único para cada especie. Los campos `dimorfismo` y `migratoria` pueden ser "Sí" o "No", e indican si la especie tiene esas características. El campo `origen` señala si la especie es "Nativa" o "Endémica". Por otra parte, `tamaño` comprende dos valores enteros separados por un guión, y corresponde al rango de tamaños de los individuos de la especie, expresado en centímetros. Finalmente, `riesgo` se refiere al nivel de riesgo de extinción de la especie, y puede tomar como valor 6 etiquetas diferentes. Los campos están separados por un carácter de punto y coma. A continuación se muestra un extracto del archivo descrito:

`info.csv`

```
76;No;Sí;Nativa;38-48;Menor riesgo
75;No;Sí;Endémica;45-60;Vulnerable
8;No;Sí;Nativa;79-93;Menor riesgo
111;No;No;Nativa;26-30;Casi amenazado
112;Sí;No;Nativa;29-35;No evaluado
...
```

El segundo tipo de archivo tiene varias instancias, con la misma información pero en idiomas diferentes. Los nombres de estos archivos siguen el formato `nombres_IDIOMA.csv`, reemplazando `IDIOMA` con el nombre del idioma. Estos archivos contienen los nombres de las especies en el idioma correspondiente. Sus líneas siguen el formato:

`id;nombre`

El campo `id` es el identificador único de la especie, que permite hacer un vínculo con los datos del archivo `info.csv`. El campo `nombre` contiene el nombre de la especie en el idioma correspondiente al archivo. Los campos están separados por un carácter de punto y coma. A continuación se muestran extractos de dos de estos archivos, aunque podría haber más para otros idiomas:

`nombres_spanish.csv`

```
76;Aguilucho Chico
8;Albatros de ceja negra
68;Cóndor
183;Chercán
186;Zorzal
...
```

`nombres_latin.csv`

```
76;Buteo albigula
8;Thalassarche melanophrys
68;Vultur gryphus
183;Troglodytes musculus
186;Turdus falcklandii
...
```

Nota: Puede suponer que los datos de los archivos son correctos, y que toda especie que aparece en el archivo `info.csv` aparecerá también en los archivos de nombres, vinculada a través del identificador.

Definición

El tamaño de una especie se determina por el valor promedio entre el mínimo y el máximo especificados para la especie, según la siguiente tabla:

Tamaño	Criterio
Pequeña	$\text{promedio} < 25$
Mediana	$25 \leq \text{promedio} < 55$
Grande	$\text{promedio} \geq 55$

1. [40 %] Escriba la función **tamaño_aves(nombre_info)**, que recibe como parámetro el nombre del archivo que contiene los datos de las especies, siguiendo el formato descrito anteriormente.

La función debe crear y retornar un diccionario cuyas llaves deben ser los tamaños de las aves. Cada tamaño debe tener asociada como valor una lista de listas, en la que cada sublista corresponde a una especie y contiene el tamaño promedio redondeado a dos decimales, el identificador, y el nivel de riesgo de la especie. La lista debe estar ordenada en forma creciente, es decir, de menor a mayor, de acuerdo al tamaño promedio. Los identificadores deben ser números enteros.

Ejemplo:

```
>>> print(tamaño_aves('info.csv'))
{'Mediana': [[25.0, 186, 'Menor riesgo'], [25.5, 5, 'Menor riesgo'], ...],
'Grande': [[55.0, 37, 'Menor riesgo'], [56.5, 81, 'Menor riesgo'], ...],
'Pequeña': [[7.5, 148, 'En peligro de extinción'], [10.0, 170, 'Menor riesgo'], ...]}
```

2. [60 %] Escriba la función **reporte_por_idioma(nombre_info, idioma, origen)**, que recibe como parámetro el nombre del archivo que contiene los datos de las especies, además de un *string* que define un idioma y un *string* que señala el tipo de especies que se desea considerar ("*Nativa*" o "*Endémica*").

La función debe crear un archivo para cada nivel de riesgo encontrado, con las 5 especies de mayor tamaño promedio, ordenadas de mayor a menor según ese criterio, e identificadas con un correlativo al inicio de cada línea. Se debe considerar únicamente a las especies correspondientes al origen indicado.

El nombre de cada archivo debe ser igual al nivel de riesgo, seguido del sufijo *.txt*. Por ejemplo: "*Menor riesgo.txt*".

Las líneas de los archivos deben seguir el formato mostrado en los ejemplos. En particular, cada línea debe incluir el correlativo, el nombre de la especie, el tamaño promedio y alguno de los textos: "*migratoria*" o "*no migratoria*", según corresponda. Los promedios deben redondearse a dos decimales.

Además, la función debe retornar la cantidad de archivos que se crearon.

Ejemplo:

```
>>> print(reporte_por_idioma('info.csv', 'spanish', 'Nativa'))
5
```

A continuación, a modo de ejemplo y para mostrar el formato que se debe seguir, se muestran dos de los archivos que la función debería haber creado:

Menor riesgo.txt

```
1.- Albatros errante, tamaño promedio: 121.0cm, migratoria
2.- Cisne de cuello negro, tamaño promedio: 113.0cm, migratoria
3.- Cisne Coscoroba, tamaño promedio: 102.5cm, no migratoria
4.- Albatros de ceja negra, tamaño promedio: 86.0cm, migratoria
5.- Piuquén, tamaño promedio: 75.0cm, no migratoria
```

Casi amenazado.txt

```
1.- Cóndor, tamaño promedio: 115.0cm, no migratoria
2.- Flamenco Chileno, tamaño promedio: 102.5cm, no migratoria
3.- \Nandú - Suri, tamaño promedio: 96.0cm, no migratoria
4.- Becasina Común, tamaño promedio: 28.0cm, no migratoria
5.- Chorlo de Magallanes, tamaño promedio: 20.0cm, no migratoria
```

Nota: Es importante recordar que para que el ordenamiento de valores funcione adecuadamente debe trabajarse con tipos numéricos (*int* o *float*) y no con *strings*. Esto es aplicable a ambas preguntas.

Créditos: Este certamen fue propuesto por los profesores Hernán Sarmiento y Aníbal Silva, y fue adaptado por el profesor Federico Meza.