



ISIS-1221

INTRODUCCIÓN A LA PROGRAMACIÓN

Nivel 3 - Proyecto

Análisis de colombianos en Wikipedia

Objetivo general

El objetivo general de este proyecto es que usted practique los conceptos estudiados en el nivel 3 del curso.

Recuerde que este proyecto debe realizarse de forma **completamente individual**.

Objetivos específicos

1. Practicar la lectura y escritura de archivos con formato CSV.
2. Ejercitar la implementación de algoritmos de recorrido y modificación de listas.
3. Familiarizarse con el manejo de estructuras de datos compuestas (Diccionario de Listas de Dictionaries).
4. Fomentar la habilidad de descomponer un problema en subproblemas y de implementar funciones que los resuelven, lo que se conoce comúnmente como la técnica de “Dividir y Conquistar”.

Contexto

Enciclopedia

Una enciclopedia es una obra de referencia que recopila una amplia variedad de conocimientos sobre distintos temas y disciplinas. Es una fuente de información que ofrece una visión general y completa de diferentes temas como ciencias, artes, historia, geografía, tecnología, literatura, etc. Las enciclopedias han existido desde la antigüedad y con el tiempo han evolucionado significativamente.

Inicialmente, en la Antigüedad, existían compilaciones de conocimiento que pueden considerarse precursores de enciclopedias modernas. Por ejemplo, en el antiguo Egipto, se crearon textos que reunían información sobre medicina, matemáticas, astronomía, entre otros. En la antigua Grecia y Roma, existían obras como la “Enciclopedia” de Plinio el Viejo. Luego, en la Edad Media, el conocimiento se preservó principalmente en monasterios y centros de aprendizaje religiosos. Las enciclopedias como la “Etymologiae” de Isidoro de Sevilla fueron compilaciones de temas teológicos, filosóficos y científicos que jugaron un rol importante en la preservación del conocimiento clásico. Por otra parte, durante los siglos XIX y XX, se publicaron numerosas enciclopedias temáticas y especializadas, que abarcaban campos como la ciencia, la historia, las artes, la tecnología, entre otros. Estas enciclopedias se

convirtieron en fuentes de referencia esenciales para el estudio y la investigación en diversas áreas del conocimiento.

Finalmente, con la llegada de Internet y la era digital, las enciclopedias en línea han cobrado una gran relevancia. Una enciclopedia que surgió en esta era es Wikipedia. Las enciclopedias en línea han revolucionado la percepción de las enciclopedias al proporcionar acceso instantáneo, actualización continua, participación colaborativa y una amplia gama de formatos multimedia. Sin embargo, también han planteado desafíos en términos de verificabilidad y confiabilidad de la información. En cualquier caso, Internet ha ampliado el acceso al conocimiento y ha transformado la forma en que las personas acceden, utilizan y perciben las enciclopedias en la era digital.

Wikipedia

Wikipedia es una enciclopedia en línea y de acceso libre creada y mantenida colaborativamente por voluntarios de todo el mundo. Wikipedia ha cambiado la forma en que se accede y se comparte el conocimiento, permitiendo la participación de la comunidad y el acceso instantáneo a una cantidad masiva de información. El contenido de Wikipedia es generado y editado por miles de personas, conocidas como "wikipedistas", que contribuyen en distintos idiomas y disciplinas. Los artículos de Wikipedia abarcan una amplia gama de temas, desde ciencia, historia, tecnología y cultura hasta temas cotidianos, biografías de personas famosas, eventos actuales y mucho más.

Para este proyecto, se quiere hacer un estudio sobre las personas de nacionalidad colombiana de las que haya información en Wikipedia. Estas personas se podrían considerar como las más conocidas y/o que han impactado más sobre la historia colombiana. Por tal motivo, será de interés quiénes son estas personas y qué han hecho. Los datos usados para este fin son adaptados de la base de datos de personas notables en Wikipedia de Kaggle¹. Estos datos contienen mucha información de una gran cantidad de personas, sin embargo, solo se usarán algunos de estos valores de acuerdo con los requerimientos de la aplicación. Es importante tener en cuenta que debido a que esta base de datos original es significativamente grande (más de 2 millones de personas), es susceptible a que existan ciertas imprecisiones o faltantes en datos tales como la fecha de nacimiento, la nacionalidad o incluso el género de algunos personajes. En particular, es seguro que algunos fallecimientos de colombianos no se encuentran debidamente registrados. Es por esto que les sugerimos no usar esta base de datos más allá de este proyecto y verificar con fuentes adicionales cualquier dato que encuentren y les pueda llamar la atención.

¹ <https://www.kaggle.com/datasets/konradb/wikipedia-notable-people>

Descripción de la aplicación

En este proyecto se va a implementar una estructura que nos permita analizar algunos datos interesantes sobre la población de colombianos que tienen un artículo en Wikipedia.

El programa debe ser capaz de recibir la información de los colombianos en Wikipedia a partir de un archivo en formato CSV (Comma-Separated Values) que tiene 7 columnas:

- nombre: Nombre completo de la persona.
- genero: Género con el que se identifica la persona. En este caso ("Male" o "Female").
- anio_nacimiento: Año en que nació la persona.
- anio_muerte: Año en que murió la persona. En caso, de no haber muerto aún, este campo tendrá el valor cero (0).
- ocupacion: Ocupación de la persona, a qué se dedica/dedicó.
- ciudadania: Nombre del país en el que la persona tiene nacionalidad. En este caso, todas las personas tienen nacionalidad colombiana, pero algunas podrían tener también una segunda nacionalidad en algún otro país. En este caso este campo contendrá los dos países separados por un guion y espacios. Es decir, que se verá de la manera "X – Colombia" o "Colombia – X", donde X es la otra nacionalidad con la que están relacionados.
- numero_lectores: Número de personas que han leído su información en Wikipedia.

Para el manejo de los datos, se debe leer el archivo .csv y desarrollar una estructura cuyo componente principal es un diccionario de ocupaciones, en el cual las llaves serán las diferentes ocupaciones existentes en el csv (ocupacion), y el valor una lista de diccionarios. Cada diccionario de la cada una de las listas contendrá la información de un colombiano con dicha ocupación. Los diccionarios de cada colombiano tendrán las llaves "nombre" con el nombre completo de la persona, "genero" con el género con el que se identifica la persona, "anio_nacimiento" con el año de nacimiento de la persona, "anio_muerte" con el año de muerte de la persona (cero (0) para las personas que aún sigan vivas), "ciudadania" con el nombre del país al que pertenece la persona y "numero_lectores" con el número de personas que han leído la información del colombiano en Wikipedia. Note que los diccionarios de cada colombiano no tienen el valor de la ocupación, en tanto dicha información ya se tiene presente en la llave de la lista que los contiene. A continuación, se muestra la estructura deseada para el manejo de los datos:

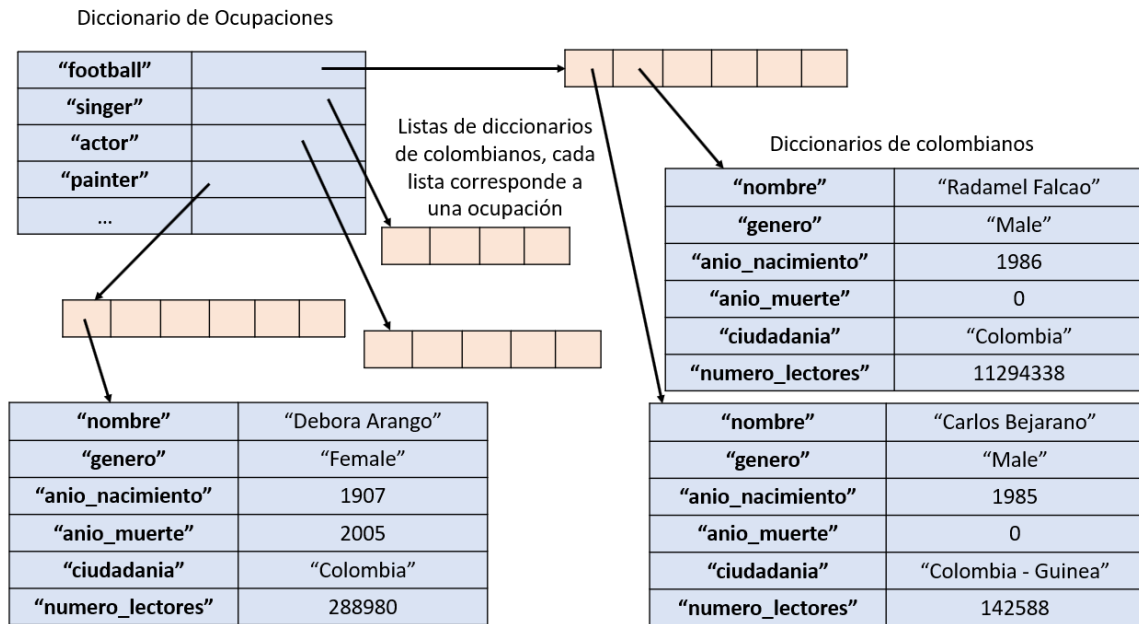


Figura 1. Ejemplo de la estructura deseada para el manejo de los datos del archivo .csv

La aplicación debe permitir al usuario ejecutar las siguientes acciones:

1. Cargar un archivo con la información de los colombianos en Wikipedia.
2. Encontrar el colombiano con mayor número de lectores.
3. Encontrar si al menos 3 colombianos de un género y ocupación específicos, superan un número de lectores.
4. Calcular el promedio de lectores de las personas de una ocupación específica.
5. Encontrar qué ocupación es la que tiene mayor rating, es decir, mayor número de lectores promedio.
6. Consultar colombianos con una ocupación específica que hayan nacido en un rango de años.
7. Contar cuantas personas hay de cada nacionalidad diferente.
8. Calcular la edad de cada persona.
9. Consultar los colombianos que han fallecido.

Actividad 1 | Preparación del ambiente de trabajo

1. Cree una carpeta para trabajar, poniéndole su nombre o login.
2. Descargue de Bloque Neón el archivo con el "esqueleto" del proyecto (**n3-esqueleto.zip**) y descomprímalo en su carpeta de trabajo. El esqueleto consiste en un conjunto de archivos que usted va a usar o a modificar.
3. Abra Spyder y cambie la carpeta de trabajo para que sea la carpeta con el esqueleto.

Actividad 2 | Construir el módulo de funciones

Usando Spyder, cree en su carpeta de trabajo un nuevo archivo con el nombre **"colombianos.py"**. En este archivo usted va a construir el módulo en el que va a implementar las funciones que responden a los requerimientos de la aplicación.

Defina, documente e implemente las funciones descritas a continuación en su nuevo archivo. **Lea cuidadosamente las descripciones de las funciones para determinar los parámetros de entrada junto con sus tipos y el valor de retorno con su respectivo tipo.**

Función 1:

Implemente una función que reciba como parámetro el nombre de un archivo que contiene la información de los colombianos en Wikipedia y la cargue en el programa bajo la forma de **un diccionario de listas de diccionarios**. Cada colombiano se va a representar utilizando un diccionario que tiene las siguientes llaves:

Llave	Tipo de dato	Descripción de la llave
<i>nombre</i>	str	Nombre del colombiano.
<i>genero</i>	str	Género con el que se identifica el colombiano. En este caso ("Male" o "Female").
<i>anio_nacimiento</i>	int	Año de nacimiento.
<i>anio_muerte</i>	int	Año de muerte. En caso, de no haber muerto aún, este campo tendrá el valor cero (0).
<i>ciudadania</i>	str	Nombre del país en el que la persona tiene nacionalidad. En este caso, todas las personas tienen nacionalidad colombiana, pero algunas podrían tener también una segunda nacionalidad en algún otro país.
<i>numero_lectores</i>	int	Número de personas que han leído la información del colombiano en Wikipedia.

Esta función debe retornar el diccionario que tiene como llave cada ocupación y como valor asociado su respectiva lista de colombianos (de diccionarios) creada. La estructura objetivo se muestra en la Figura 1.

Función 2:

Implemente una función que reciba por parámetro el diccionario completo de colombianos y retorne el nombre del colombiano con mayor número de lectores en Wikipedia.

Función 3:

Implemente una función que reciba por parámetro el diccionario completo de colombianos, el nombre de una ocupación, un género y un número de lectores, y retorne un booleano que indique si existen al menos 3 colombianos de dicha ocupación y género que superen ese número de lectores.

Función 4:

Implemente una función que reciba por parámetro el diccionario completo de colombianos y el nombre de una ocupación y retorne el promedio de número de lectores de los colombianos con la ocupación dada por parámetro redondeado a dos cifras decimales.

Función 5:

Implemente una función que reciba por parámetro el diccionario completo de colombianos y retorne qué ocupación tiene el mayor *número de lectores promedio*. Esta función debe hacer uso de la función 4.

Función 6:

Implemente una función que reciba por parámetro el diccionario completo de colombianos, una ocupación y un rango de años, es decir, un límite inferior y uno superior y retorne una lista con los colombianos que tienen la ocupación dada por parámetro y que nacieron en ese rango de años. La lista esperada debe tener como elementos los diccionarios con la información de los colombianos que nacieron en ese rango de años y tienen la ocupación dada.

Función 7:

Implemente una función que reciba por parámetro el diccionario completo de colombianos y retorne un diccionario con el número de personas de cada nacionalidad. El diccionario esperado debe tener como llaves las nacionalidades y como valor, el número de personas con esa nacionalidad.

Función 8:

Implemente una función que reciba por parámetro el diccionario completo de colombianos y retorne este mismo diccionario con una pareja llave-valor adicionada al diccionario de cada colombiano. La nueva llave debe ser "edad" y el valor de esta, es la edad del colombiano calculada. Para los colombianos que ya fallecieron, se calcula la edad que tenían cuando fallecieron.

Ayuda:

Para obtener el año actual de modo que se pueda calcular la edad de los colombianos que aún no han fallecido, pueden usar el siguiente fragmento de código:

```
import datetime

currentDateTime = datetime.datetime.now()
date = currentDateTime.date()
year = int(date.strftime("%Y"))
```

Función 9:

Implemente una función que reciba por parámetro el diccionario completo de colombianos y retorne un diccionario de listas de diccionarios (tal como la estructura del diccionario construido en la función 1) únicamente con los colombianos que ya fallecieron.

Actividad 3 | Completar la interfaz de usuario basada en consola

1. En esta actividad usted tiene que construir la interfaz basada en consola para que el usuario interactúe con la aplicación. Para construir esta interfaz usted debe completar el archivo **consola_colombianos.py**, la cual ya tiene una parte implementada que le facilitará su trabajo. Usted debe modificar los elementos marcados con la etiqueta **TODO**.
2. Pruebe la interfaz basada en consola ejecutando el archivo **"consola_colombianos.py"**. Verifique que las funcionalidades de su aplicación se comporten de acuerdo con lo esperado.
3. Pruebe cargando el archivo **colombianos.csv** o cree su propio archivo de prueba respetando el mismo formato.

Ejecute el programa y pruebe cada una de las funciones para asegurar que esté funcionando.

Entrega

1. Comprima la carpeta con su proyecto resuelto. El archivo debería llamarse **N3-PROY-login.zip**, donde login es su nombre de usuario de Uniandes.
2. Entregue el archivo comprimido a través de Bloque Neón en la tarea designada como **Proyecto del Nivel 3**.