



IIC2115 – Programación como Herramienta para la Ingeniería (II/2017)

Actividad 7

Objetivos

- Aplicar contenidos para manejo de texto plano.
- Uso correcto de `format` para aplicar formato al texto.
- Introducción al manejo de bases de datos

Entrega

- **Lenguaje a utilizar:** Python
- **Lugar:** GitHub
- **Hora:** 16:55
- **Desarrollo** –

Advertencia: Esta actividad no está planeada para ser terminada solo en la hora de clases, sino para mostrar diferentes usos de I/O. Te sugerimos considerarla como una guía sobre este tema.

Problema 1: Histograma de palabras - Dificultad: Bastián (Lacayo)

Durante el 18, Bastián se encontraba navegando por internet cuando accidentalmente descargó un archivo llamado `tarea_2_herramientas.te_pillamos`. Sin saber qué hacer, Bastián entró en pánico y decidió pedirle a los alumnos del curso IIC2115 que lo ayudaran a ver el contenido del archivo.

Tu misión es crear una función llamada `frecuencia(nombre_archivo, palabras_vetadas)`, la cual recibe en forma de `string` el nombre de un archivo y una lista de palabras (lista de `string`). La función debe retornar un diccionario con las 10 palabras más repetidas en el texto contenido en el archivo, que no estén presentes en la lista “palabras_vetadas”. También debes incluir la cantidad de veces que estas palabras aparecen.

Ten en cuenta las siguientes restricciones:

1. Solo debes considerar palabras con más de 4 letras (5 o más caracteres).
2. No debes considerar números.

Problema 2: Uso de format - Dificultad: Laika¹ (Perro de Bastián)

Durante el 18, mientras Bastián hacía “sus cosas” en internet, Laika se encontraba analizando las cuentas de la casa y cuánto dinero le costaba mantener a Bastián. En eso, abre un archivo **cuentas.hans**, en donde estaban todos los gastos de la casa. Ella los visualiza en el computador de una forma elegante². Viendo que los alumnos del curso están aprendiendo sobre I/O, decide borrar todo lo que hizo y pedirle al curso que lo haga nuevamente como forma de estudio.

Problema 2.a: Para abrir el archivo *cuentas.hans* debes seguir los siguientes pasos:

1. Abrir el archivo con `open(archivo.hans)`.
2. Recorrer las líneas del archivo. La primera línea es el *header*.
3. Separar cada línea por “;”
4. Guardar los contenidos de las líneas, sin olvidar que cada elemento está asociado a su columna del *header* correspondiente.
5. Imprimir todo.

Problema 2.b; Ahora que aprendiste a leer un archivo *.hans*, además de observar el proceso que involucra hacerlo, Laika te solicita repetir lo realizado, pero usando la librería `csv`. Para esto debes:

1. Importar la librería `csv`.
2. Abrir el archivo en modo “rb”.
3. Usar la función `csv.reader(csvfile, delimiter=';')` para abrir el archivo
4. Revisar esta página por si quedan dudas.

Finalmente, Laika te pide visualizar el `csv` de forma elegante. Para esto, te pide usar `format` y cumplir las siguientes reglas:

- Los números enteros deben tener siempre 5 dígitos. Rellenar con 0 si es necesario.
- Los números decimales deben tener siempre 3 dígitos en la parte entera y 4 dígitos de decimales. Rellenar con 0 si es necesario.
- Las palabras deben estar centradas y siempre ocupar un espacio de 12 caracteres.

Con las reglas anteriores. una frase como “comida;42;12.5” quedará:

```
"  comida  ;00042;012.5000"
```

¹Laika demoró 5 minutos en hacer este ejercicio, Bastián, 10.

²Usando `format`, por supuesto.

Problema 3: Inner Join - Dificultad: Hugo (Ayudante)

La empresa “Antartica” se ha enterado que en el curso de Programación como Herramienta para la Ingeniería se está enseñando la manipulación de archivos y ha contratado a Hugo para obtener alguna información sobre una base de datos de libros. Como Hugo sabe muy bien que a los alumnos del curso les gusta programar, los ha subcontratado para realizar esta *pega*. La paga será el amor eterno de Hugo, para quien logre realizar estas consultas ³.

La empresa te ha mandado 3 archivos *.csv* los cuales tienen una estructura igual a los *.hans*. A continuación se presenta una descripción de cada archivo:

- *libros.csv*: este archivo tiene el nombre de todos los libros junto con su autor. Sus columnas son: *id*, *titulo*, *autor*, donde *id* es un número único para cada libro.
- *personas.csv*: este archivo tiene el registro de todas las personas que han entrado a una tienda de “Antartica”. Las columnas de este archivo son: *id*, *nombre*, *edad*, donde *id* es un número único para cada persona.
- *compras.csv*: este archivo tiene el registro de todas las compras de un libro vendido por “Antartica”. Las columnas que posee son las siguientes: *id_libro*, *id_persona*, donde *id_libro* es el número para identificar un libro y *id_persona* es el número para identificar a la persona.

La información a buscar es: para cada compra realizada, indicar el nombre del libro, el nombre del autor y el nombre y edad de la persona que realizó la compra.

Para esto recurriremos a un forma de unir tablas de datos llamada “inner join”. Así, se buscarán las personas que compraron cada libro de cada autor, **pero** si hay algún resultado que le falte información, por ejemplo, un libro que nunca fue vendido o el id de una persona que no existe, no se debe mostrar dicha fila, es decir, esta forma de unir tablas solo deja las filas que tengan toda la información presente. Por lo tanto, las columnas resultantes tendrán el nombre del libro, su autor, el nombre de su comprador y su edad.

El resultado debes imprimirlo en pantalla.

Problema 4: Group By - Dificultad: Hans (Deidad)

La siguiente consulta de la empresa “Antartica” fue solicitada a Hans, pero en reproche de la acción de Hugo, decidió derivarle el trabajo a Bastián, con la esperanza de que él no recurriera a la subcontratación. Si estás viendo este enunciado es por algo, así que puedes suponer el resto de la historia. La consulta a responder es “¿Cuál es el promedio de edad de los lectores para cada autor?”.

Utilizando la información obtenida en el problema anterior, deberás agrupar las filas por cada autor y calcular el promedio de edad de los compradores de sus libros.

El resultado debes imprimirlo en pantalla.

³por interno nos avisan que le pagarán 2 millones de pesos

Política de Integridad Académica

Los alumnos de la Escuela de Ingeniería deben mantener un comportamiento acorde al Código de Honor de la Universidad:

“Como miembro de la comunidad de la Pontificia Universidad Católica de Chile me comprometo a respetar los principios y normativas que la rigen. Asimismo, prometo actuar con rectitud y honestidad en las relaciones con los demás integrantes de la comunidad y en la realización de todo trabajo, particularmente en aquellas actividades vinculadas a la docencia, el aprendizaje y la creación, difusión y transferencia del conocimiento. Además, velaré por la integridad de las personas y cuidaré los bienes de la Universidad.”

En particular, se espera que mantengan altos estándares de honestidad académica. Cualquier acto deshonesto o fraude académico está prohibido; los alumnos que incurran en este tipo de acciones se exponen a un procedimiento sumario. Ejemplos de actos deshonestos son la copia, el uso de material o equipos no permitidos en las evaluaciones, el plagio, o la falsificación de identidad, entre otros. Específicamente, para los cursos del Departamento de Ciencia de la Computación, rige obligatoriamente la siguiente política de integridad académica en relación a copia y plagio: Todo trabajo presentado por un alumno (grupo) para los efectos de la evaluación de un curso debe ser hecho individualmente por el alumno (grupo), sin apoyo en material de terceros. Si un alumno (grupo) copia un trabajo, se le calificará con nota 1.0 en dicha evaluación y dependiendo de la gravedad de sus acciones podrá tener un 1.0 en todo ese ítem de evaluaciones o un 1.1 en el curso. Además, los antecedentes serán enviados a la Dirección de Docencia de la Escuela de Ingeniería para evaluar posteriores sanciones en conjunto con la Universidad, las que pueden incluir un procedimiento sumario. Por “copia” o “plagio” se entiende incluir en el trabajo presentado como propio, partes desarrolladas por otra persona. Está permitido usar material disponible públicamente, por ejemplo, libros o contenidos tomados de Internet, siempre y cuando se incluya la cita correspondiente.