

## Taller #2. Física Computacional / FISI 2025 Semestre 2013-I.

Profesor: Jaime E. Forero Romero

Febrero 7 2013

**Esta tarea debe estar en un repositorio de la cuenta de github de cada uno con un commit final hecho antes del medio día del jueves 14 de Febrero del 2013**

1. Visiten la página <http://www.gutenberg.org/catalog/>. Vamos a elegir algunos libros para trabajar.
2. Bajen la version de texto simple UTF-8 (no HTML, EPUB o Kindle) de un libro en cada uno de los cinco idiomas siguientes:
  - inglés: <http://www.gutenberg.org/browse/languages/en>
  - español: <http://www.gutenberg.org/browse/languages/es>
  - finlandés: <http://www.gutenberg.org/browse/languages/fi>
  - italiano: <http://www.gutenberg.org/browse/languages/it>
  - inglés antiguo: <http://www.gutenberg.org/browse/languages/ang>
3. Escriba un programa en Python que cuente cuántos caracteres diferentes hay en cada uno de los libros. Los caracteres con diéresis o tildes cuentan como caracteres diferentes. Los signos de puntuación no se deben incluir en la cuenta.
4. Por cada caracter diferente calcule su frecuencia en el texto. Es decir, el porcentaje de veces que aparece cada caracter con respecto al número total de caracteres en el texto.
5. En el caso de un libro llamado `DonQuijoteDeLaMancha.txt` el código se debe poder ejecutar de la siguiente manera:

```
python frecuencias.py DonQuijoteDeLaMancha.txt
```

6. El output del código en Python debe ser un archivo de texto por cada libro elegido. En el caso de este ejemplo, el archivo de salida se debe llamar `frecuencias_DonQuijoteDeLaMancha.txt`.

7. En el archivo de texto se deben escribir dos columnas: en la primera los caracteres del libro y en la segunda su frecuencia. Las filas deben estar ordenadas en orden decreciente de frecuencia.
8. El código fuente de Python y los archivos de los libros deben estar en un repositorio de GitHub. Dentro del repositorio, deben estar adentro de un directorio que se llame `python`.
9. Si tienen algún comentario que hacer sobre las diferentes distribuciones de caracteres en los idiomas estudiados lo pueden incluir en un archivo de texto que se llame `comentario.txt`
10. Enviar un email la dirección `j.e.forero.romero` en `gmail` con el subject **RESPUESTA TALLER 2 FISICA COMPUTACIONAL**. En el cuerpo del texto debe ir la dirección del repositorio donde está la tarea.
11. La calificación se hará en tres partes.
  - El código fuente y los libros están en un repositorio de github (20%).
  - El código funciona de manera adecuada con los cinco libros elegidos (60%).
  - El código funciona de manera adecuada con un libro en esperanto de mi elección. (20%)