



01 de Febrero de 2020

Taller #1 Bases de Datos: Encoder & Decoder

Matemáticas Aplicadas y Ciencias de la Computación

Integrantes: Sebastián Martínez

Profesor: Camilo Rey

Jhan Carlos Celi

Victor Sicacha

Santiago Uribe

Oscar Velasco Ch.

Puntos 1 & 2:

Link Repositorio Github: <https://github.com/Oscarvch03/Bases-de-Datos>

Nota: El bloc de notas, WordPad y Notepad++ utilizan el estándar de codificación y transformación de caracteres Unicode de 8 bits (UTF-8) (Recuérdese pregunta 5.3b).

Punto 4:

¿Qué sucede cuando intenta volver a pasarlo por el encoder?

El programa arroja "file header is corrupt".

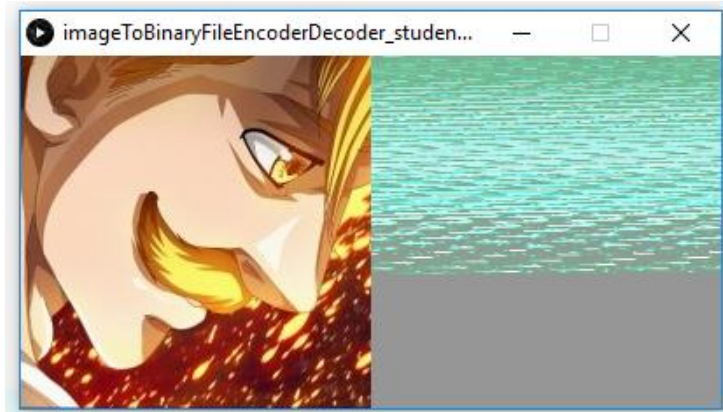
¿Por qué?

Según investigamos, el bloc de notas procesa con Byte Order Mark (BOM) -el cual es un uso particular o especial de Unicode- e interfiere con el uso de UTF-8 con un software que no espera el uso de bytes de caracteres que no son ASCII al inicio del archivo, lo que genera que el archivo se corrompa, y por ejemplo cambie su tamaño (Recuérdese pregunta 5.3b) (Wikipedia, 2019). En nuestro caso el tamaño del archivo pasó de 160.002 bytes a 160.004 bytes (Recuérdese pregunta 5.2).

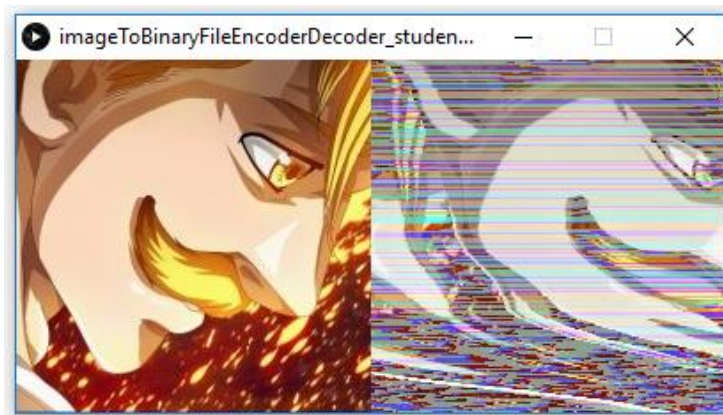
Ahora, si abrimos el archivo con WordPad también se corrompe, ya que cuando abrimos un archivo con dicho editor, que no tiene extensión ".docx" muchas de sus características no se mostrarán correctamente (Huculak, 2019), es por esto que en este caso el tamaño del archivo pasa de 160.002 bytes a 161.383 bytes (Recuérdese pregunta 5.2). (Peor aún).

¿Se corrompe el tamaño de la imagen?

El tamaño original de la imagen es 200x200 (Equivalente a 40.000 píxeles). Si corrompemos el archivo con bloc de notas la imagen que se obtuvo luego de "corregir" el archivo .bla, tenía tamaño 255x254.



Si corrompemos el archivo con WordPad, la imagen que se obtuvo luego de "corregir" el archivo .bla, tenía el tamaño de la imagen original, pero se distorsiona el color.



Punto 5:

¿Cómo arreglaría este problema?

Lo que se nos ocurrió para que se logre decodificar la imagen es borrarle al archivo .bla los bytes excedentes y así hacer el archivo decodeable. El problema es que no necesariamente va a generar la imagen original ya que se pierde información.

5.1) Diagnostique la situación con base en el error que obtiene

Suponemos que se alteró el tamaño del archivo. (Hipótesis comprobada)

5.3) Determine la causa del error que obtiene.

La causa es que cambia de tamaño el arreglo que contiene la información de la imagen, ya que el código proporcionado por el profesor, se encarga de verificar que el archivo no esté corrupto en dicho sentido.

¿Qué pasa con Processing?

Genera el error "file header is corrupt".

¿Si cambia el WordPad por Notepad++ sigue obteniendo el mismo error?

No, al abrir el archivo .bla con Notepad++ no se corrompe. Suponemos que es porque Notepad++ si logra procesar de manera correcta todos los caracteres de dicho archivo.

5.4) De ser posible, arréglole. Su Benchmark es que un archivo .bla abierto y guardado desde WordPad sea decodeable con su programa. Si no es posible, comente por qué no es posible.

Como habíamos dicho (en Punto 5) nuestra solución para que el archivo .bla sea decodeable de nuevo, es borrar los bytes excedentes (1381 bytes). Para ello diseñamos una función (*arreglarBla()*) que crea un nuevo archivo "*testarreglado.bla*" que contiene los primeros 160.002 bytes del archivo corrupto "*testcp.bla*". Este equivale al tamaño original del archivo .bla de la imagen original codificada (escanor.jpg).

Bibliografía

- [1] Huculak, M. (05 de Diciembre de 2019). *Windows Central*. Obtenido de Windows Central: <https://www.windowscentral.com/notepad-vs-wordpad-and-when-use-them>
- [2] Wikipedia, l. E. (30 de Octubre de 2019). *Wikipedia, la Enciclopedia Libre*. Obtenido de Wikipedia, la Enciclopedia Libre: https://en.wikipedia.org/wiki/Byte_order_mark#UTF-8