



30 de Enero 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

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).

Punto 4:

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

El programa arroja "file header is corrupt".

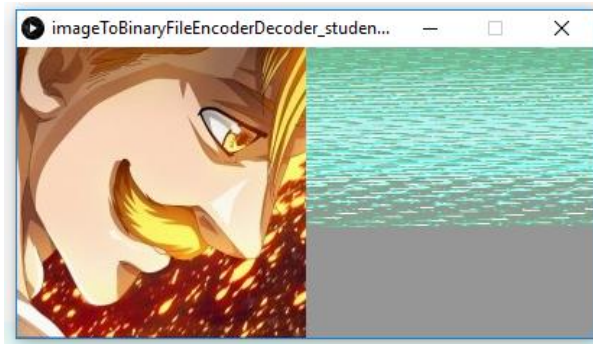
¿Por qué?

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. En nuestro caso el tamaño del archivo pasó de 160.002 bytes a 160.004 bytes.

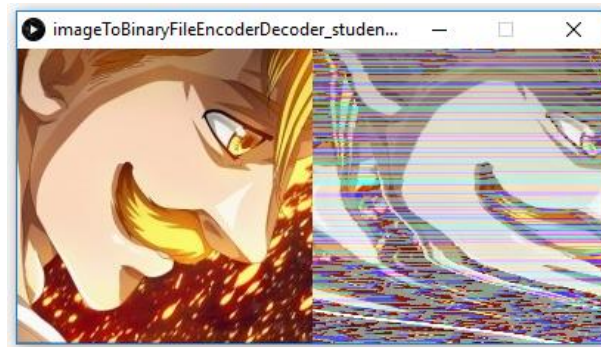
Ahora, si abrimos el archivo con WordPad también se corrompe, ya que cuando abrimos un archivo con dicho editor, que no tiene extensión ".docx" muchas de sus características no se mostrarán correctamente, es por esto que en este caso el tamaño del archivo pasa de 160.002 bytes a 161.383 bytes. (Peor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

¿Se corrompe el tamaño de la imagen?

El tamaño original de la imagen es 200x200 (Equivalente a 40.000 pixeles). 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)

5.2) Investigue el tamaño del arreglo resultante y relacione con el error que obtuvo.

Explicado en punto 4.

5.3) Determine la causa del error que obtiene.

Explicado en punto 4.

¿Qué relación tiene con la codificación de Windows?

Explicado en punto 4.

¿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, arréglo. 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).

BIBLIOGRAFIA:

1. Wikipedia La Enciclopedia Libre, 30 de Octubre de 2019,
https://en.wikipedia.org/wiki/Byte_order_mark#UTF-8
2. Huculak Mauro, Windows Central, 05 de Diciembre de 2019,
<https://www.windowscentral.com/notepad-vs-wordpad-and-when-use-them>