CS 130 Lab # 5: Formatos de archivo binario

Metas

Las aplicaciones informáticas más comunes almacenan datos de usuario en archivos compuestos por valores binarios. Normalmente, los datos se almacenan en estos archivos para que los datos descriptivos más amplios Viene primero, seguido por datos más específicos cuya forma o disposición depende de la descripción anterior: múltiples "capas" de dependencia surgen naturalmente en este camino.

Por ejemplo, la mayoría de los archivos tendrán algún tipo de código o "etiqueta" al principio que nombra o de otra manera señala el formato general del archivo. Para los archivos gráficos, Tag suele ser seguido por información sobre las dimensiones del gráfico: altura y ancho en píxeles y "profundidad" en bits. Qué orden esta información viene In y cuántos bits o bytes se utilizan, se especifica en el formato de archivo señalado por la etiqueta. Siguiendo la información dimensional, podríamos ver pixel "en bruto" Información (también llamada "datos ráster") y, en algunos casos, una tabla de colores. Tenga en cuenta que la interpretación correcta de la información de píxeles dependerá de la Dimensiones y, si está presente, la tabla de colores.

En formatos de archivo más complejos, este tipo de dependencias puede hacer bastante difícil "leer" el archivo directamente, a medida que las capas de contexto se acumulan. Además, la mayoría de los formatos de archivo utilizan representaciones binarias directas para los datos atómicos finales (por ejemplo, información de color). La mayoría de los programas editoriales están configurados para mostrar texto, suponiendo que algo como el código ASCII se utiliza para representar los caracteres. En el caso de los denominados archivos binarios (con mayor precisión, los interpretación no implica un código de carácter similar a ASCII), es posible que desee que el editor muestre los datos en su forma más accesible para el usuario, como colores, números, etc. Sin embargo, en muchos casos (como el laboratorio de hoy) queremos concentrarnos específicamente en cómo se utilizan las representaciones binarias, por lo que queremos ver los datos no como texto, y no directamente como colores, etc., sino en términos de los códigos binarios "en bruto". Para fines prácticos, esto significa que examinaremos los códigos

Hexadecimal, ya que preserva la estructura del binario subyacente, pero hace patrones más cortos y más fácilmente distinguibles.

En este laboratorio examinaremos "dentro" los detalles binarios de algunos formatos de archivo diferentes para ver cómo se representan los datos de varios tipos. Usted tendrá que copiar todo

De los archivos de datos de ejemplo de la siguiente carpeta en su disco duro, y también para localizar y ejecutar el programa "editor hexadecimal" que usaremos, llamado XVI32.

Acerca del editor hexadecimal

Con el fin de ver los archivos en su formato binario subyacente, vamos a utilizar un denominado editor hexadecimal: este es un programa que le permite abrir y ver un archivo no como lo

Se muestra normalmente por el programa en el que se creó, sino más bien como una serie de códigos hexadecimales.

El editor hexadecimal que usaremos se llama XVI32: se puede encontrar en los equipos de laboratorio en la sección "Mi PC", "Disco local (C :)", en "Archivos de programa" carpeta. (También puede descargar una copia de este programa gratuito para su propio ordenador desde el sitio web del autor, suponiendo que utiliza Windows (usuarios de Mac puede encontrar programas similares: pregúnteme si necesita ayuda para localizar uno).

Una vez que inicie el programa XVI, podrá abrir archivos y verlos y editarlos directamente en términos de los códigos hexadecimales que los componen:

Siempre debe trabajar en una copia de un archivo si desea conservar el original, ya que los cambios que realice a este nivel pueden dañar el archivo para que no lo sea válida o reconocible por sus programas "nativos" (Word, Excel, MS Paint, etc.). Para fines de laboratorio, por supuesto, copias frescas de los archivos estarán siempre disponibles en el sitio web.

Instalar el software y trabajar con archivos

Aunque normalmente no se deben almacenar archivos personales en la unidad "C" de una computadora de laboratorio, el software que estamos utilizando parece tener problemas para acceder a la red unidades. Por lo tanto, recomiendo el siguiente enfoque para configurar el programa y los archivos:

1. Busque y abra la carpeta Mis documentos en su máquina: debe estar en la carpeta Mi PC / Disco local (C:) / Documentos y Settings / WU\_User /);
2. Crear una nueva carpeta en la carpeta "Mis documentos" (se podría llamar "CS 130");
3. Cuando obtenga los archivos de ejemplo de la carpeta cs130 / docs, guárdelos en su carpeta CS 130 en Mis documentos;
4. Ahora debe ser capaz de utilizar el editor para trabajar en los archivos de muestra sin demasiados problemas (pero pregúnteme si usted está teniendo problemas).
5. Tratar de mantener un archivo abierto con sólo un programa a la vez: en otras palabras, abrir un archivo con XVI32 y modificarlo, luego cerrarlo y abrirlo con MSWord, MS Paint o MS Excel (luego vuelva a cerrarlo antes de volver a abrirlo en XVI32); Esto asegurará que está viendo (y guardando) una copia actual de el archivo en todo momento;
6. Asegúrese de copiar sus archivos de nuevo a su unidad H si desea guardarlos cuando haya terminado de trabajar;
7. Asegúrese de quitar los archivos de la carpeta Mi PC / CS 130 para que otras personas no "apropien" su trabajo duro.

Algunas notas sobre el uso del editor hexadecimal XVI32

El editor hexadecimal (XVI32) le permite mirar simultáneamente el contenido de un archivo desde varias perspectivas diferentes: tiene dos paneles principales, a la izquierda y a la derecha, cada una de las cuales muestra una "vista" diferente del contenido del archivo. El de la izquierda muestra los bytes del archivo en forma hexadecimal, mientras que el panel a la derecha intenta mostrarlos como caracteres ASCII (por supuesto, esto puede no funcionar si son valores de 8 bits completos, es decir, con un bit "1" principal). Tenga en cuenta que el dirección o número de índice del byte actualmente seleccionado se muestra en un pequeño campo en el lado izquierdo inferior del panel izquierdo, etiquetado como "Hex addr.": Este campo será importante para usted al completar algunos de los ejercicios a continuación.

En el lado hexadecimal, cada byte en el archivo (8 bits) está representado por dos dígitos hexadecimales sucesivos (por ejemplo, 1A o 23. Recuerde que incluso si los dígitos parecen decimales, probablemente estén en formato hexadecimal, de modo que un valor de 23 es realmente un valor decimal de (2 x 16) + (3 x 1) = 35.

Formatos de archivo de texto

Las primeras páginas que veremos son archivos de texto de varios tipos: en la carpeta docs mencionada anteriormente encontrarás estos archivos; Ábralos con MS Word (Utilice la opción "Abrir con ..." en el botón derecho del ratón si es necesario) y en el editor hexadecimal XVI.

* El archivo word.doc es un archivo estándar de MS Word con algún texto de ejemplo, incluyendo algunos caracteres especiales y algunos caracteres de "imagen / símbolo" escritos Utilizando la fuente WingDings.
* El archivo word.txt es un archivo de texto ASCII estándar: note que los caracteres especiales no aparecen correctamente dentro de MS Word, ya que no hay ASCII Equivalentes para estos caracteres. Observe también que no hay especificaciones de diferentes fuentes, por lo que los caracteres de imagen / símbolo no aparecen correctamente.
* El archivo word.rtf es un formato especial que utiliza "etiquetas" con formato ASCII para indicar la presencia de caracteres y fuentes especiales, etc.
* El archivo unicode.txt es un archivo que intenta grabar los caracteres especiales utilizados en formato Unicode (aunque no todos vienen a través de correctamente).

Ser capaz de responder a estas preguntas para su demo:

* Hay alguna información almacenada en el archivo word.doc sobre el equipo y las carpetas donde se originó el archivo. ¿Puedes encontrarlo usando el editor hexadecimal? (Muchos estudiantes se sorprenden al ver que esta información se almacena dentro de un documento de Word.)
* ¿Puedes modificar la palabra.rtf usando el editor hexadecimal para que aparezcan palabras diferentes? ¿Para qué se use una fuente diferente para el texto?

Dentro de una hoja de cálculo de Excel

El archivo excel.dif es un spreadhseet de Excel: ábralo tanto en Excel (doble clic debería hacer) y en el editor hexadecimal. Ser capaz de contestar estas preguntas para su demo:

* ¿Cómo se almacenan los números en el archivo Excel? En particular, ¿se almacenan como versiones binarias de los números? De no ser así, ¿qué forma se está utilizando?
* ¿Puede cambiar los números en el editor hexadecimal (digamos, el número "20") para que aparezcan de manera diferente en MS Excel?

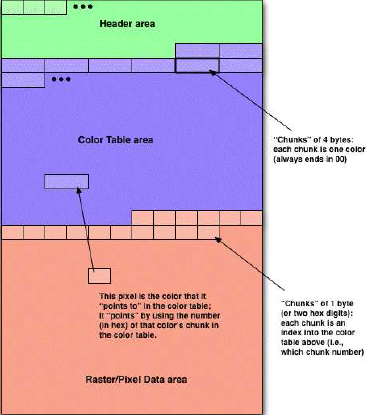
El formato de archivo gráfico BMP

Puede leer sobre el formato de archivo BMP en esta página web.CS

Lea algo de esta página y vea si puede darle sentido (esto es un ejercicio en "lectura técnica", en contra de "escritura técnica"). No te preocupes si tienes Dificultad para entenderlo: el punto de intentar es tener una idea del estilo de este tipo de documentos, no entender todo al principio.

Ahora copie el archivo helloSmall.bmp desde la carpeta docs y ábralo con el programa MS Paint (use "Open With ..." en el botón derecho del mouse si necesario).

Puede leer sobre el formato de archivo BMP y el diseño de los datos en la referencia on-line, pero este gráfico debe ayudar a visualizarlo:



(Nota: el color extraño y moteado en algunas partes de esta imagen se debe a las limitaciones en el modelo de color utilizado en los archivos GIF.)

La imagen muestra cómo se presentan las áreas de datos de cabecera, tabla de colores y píxeles. Tenga en cuenta especialmente que la tabla de colores se divide en "trozos" de 4 bytes (cada uno Byte está representado por dos dígitos hexadecimales) y que los datos de píxeles se dividen en bloques de 2 bytes, cada uno de los cuales es un índice numérico en la tabla de colores. Por último, tenga en cuenta que los índices numéricos de los datos de trama o píxel no son direcciones o ubicaciones de datos dentro del propio archivo, sino que cuentan cuántos de 4 bytes Trozos desde el inicio de la tabla de colores un cierto color es.

Cómo calcular la posición de un color en la tabla de colores

¿Cómo podemos determinar la posición dentro del archivo de un color en la tabla de colores, sobre la base de su número de índice dado en los datos de píxeles?

Necesitamos saber otras dos cosas: en primer lugar, la posición dentro del archivo del comienzo de la tabla de colores y, segundo, el tamaño de los "trozos" que son contados por el índice utilizado en los datos de píxeles.

La posición del comienzo de la tabla de colores se puede determinar leyendo la documentación acerca de los archivos BMP de cerca, o buscando la derecha patrones en los valores hexadecimales. Resulta que el primer color en la tabla es blanco, representado por hex FF FF FF 00 (la mayor cantidad posible (FF) de cada uno de rojo, verde y azul más el byte de arrastre siempre cero.

De la imagen y la descripción anteriores, sabemos que los colores de la tabla se dividen en trozos de 4 bytes. Por lo tanto, si un color está en un índice i, por ejemplo referido en los datos de trama / píxel, podemos determinar su dirección o ubicación a en el propio archivo como:

A = s + 4 x i

Donde s es el inicio de la tabla de colores.

Esta información también puede usarse para determinar el índice desde la dirección, usando un pequeño álgebra como sigue:

A = s + 4 x i

A - s = 4 x i

(A - s) / 4 = i

(Aquí hemos rescatado de ambos lados y divididos ambos lados por 4).

Recuerde, sin embargo, que todos estos números probablemente se expresarán en hexadecimal dentro del archivo: por lo tanto, debe convertir de hexágono a

Decimal y viceversa, si está calculando a mano, o hacer la aritmética usando una calculadora hexadecimal (la calculadora de utilidad WIndows se puede usar en un modo hexadecimal)

Ejercicios finales

Utilizando el archivo gráfico Mire el archivo como un gráfico, luego ábralo de nuevo usando el editor hexadecimal. Vea si puede responder a estas preguntas para su demo:

1. ¿Puedes encontrar la mesa de colores? ¿Qué ubicaciones / direcciones en el archivo ocupa (es decir, cuáles códigos hexagonales numerados)
2. ¿Puedes encontrar los datos ráster? ¿Qué ubicaciones / direcciones en el archivo ocupa (es decir, cuáles códigos hexagonales numerados)?
3. ¿Se puede cambiar un color en el gráfico mediante la modificación de los datos ráster mediante el editor hexadecimal? (Por ejemplo, hacer la fila inferior de píxeles naranja en lugar de azul.)
4. ¿Puede usted cambiar un color en el gráfico cambiando la tabla de color? (Por ejemplo, haga que cada píxel rojo sea verde en su lugar.) CS