Lab 11: Binary Reader/File Signature Analysis

Details

Aim: To provide a foundation in analysing file formats

Activities

1. .NET provides an excellent interface in reading from files, and viewing them as ASCII characters or in a hexadecimal format. For this lab download the solution from:

↳http://www.dcs.napier.ac.uk/~bill/sigAnalysis.zip

It has a Windows interface, such as:

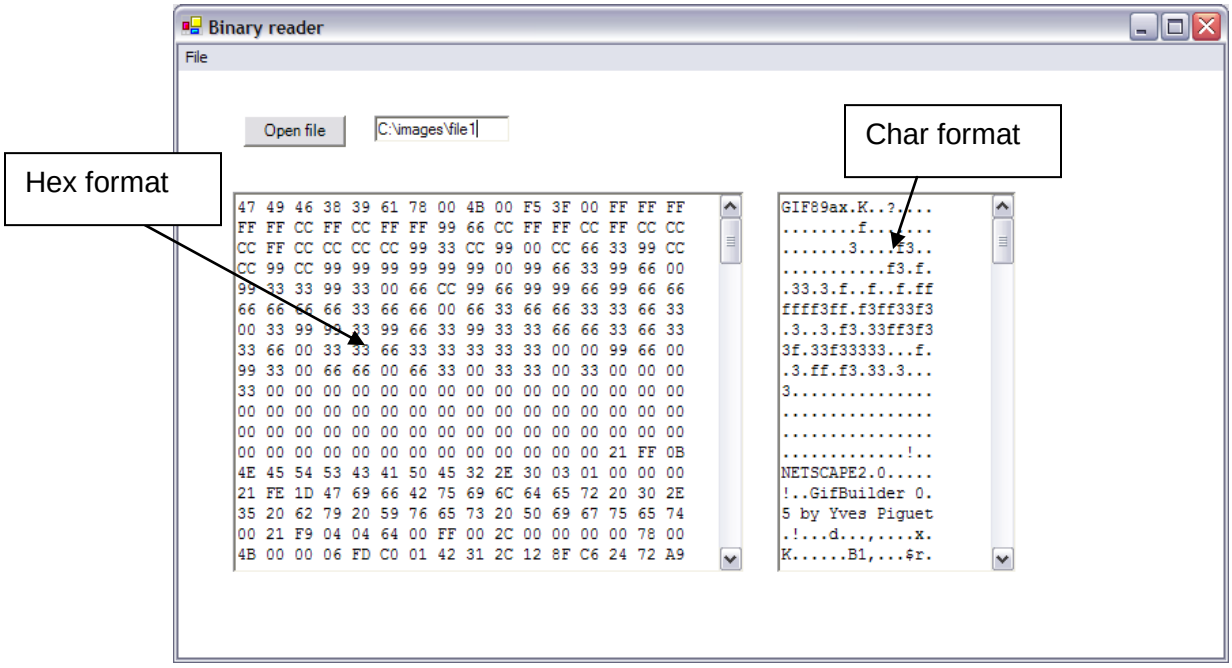

Figure 1: Interface

2. Open the solution, and for the **Open** button add the following code:

```
textBox1.Text="";
textBox2.Text="";
DialogResult result = this.openFileDialog1.ShowDialog();

textBox3.Text=openFileDialog1.FileName;

byte [] buff= getBytes(openFileDialog1.FileName);

for (int i=0;i<buff.Length;i++)
{
```

```
            char c = (char) buff[i];

            if (c>=' ' && c<='z')   textBox1.Text+=(char)buff[i];
            else textBox1.Text+=".";
            textBox2.Text+=buff[i].ToString("X2")+" "; // hex format
            if ((i+1)%16==0)  // add a new line very 16 characters
            {
               textBox1.Text+="\r\n";
               textBox2.Text+="\r\n";
            }
         }
```

and also add the following (which reads the file into a byte array):

```
public byte [] getBytes(string  f)
{
   FileStream fsIn = new FileStream(f,FileMode.Open, FileAccess.Read);
   byte [] b = new byte[2048];
   int bytesRead = fsIn.Read(b,0,2048);
   fsIn.Close();
   return(b);
}
```

The following tutorial uses files which are in a ZIP file:

http://www.dcs.napier.ac.uk/~bill/files.zip

3. Download this file, and extract them to a folder.

4. Now run the file and open the first file (**file1**). The output should be something like in Figure 1.

Refer to the Appendix given, and determine the format of the file. **What is the format of the file (such as GIF, JPEG, ZIP, etc):**

Now repeat for files 2 to 10, and complete the following table:

Name	**File format (circle correct one)**	**Is there any copyright information in the file (or associated information that is readable)?**
File2	DOC/PPT/XLS/JPEG/GIF/WMF/ZIP	
File3	DOC/PPT/XLS/JPEG/GIF/WMF/ZIP	

File4	DOC/PPT/XLS/JPEG/GIF/WMF/ZIP	
File5	DOC/PPT/XLS/JPEG/GIF/WMF/ZIP	
File6	DOC/PPT/XLS/JPEG/GIF/WMF/ZIP	
File7	DOC/PPT/XLS/JPEG/GIF/WMF/ZIP	
File8	DOC/PPT/XLS/JPEG/GIF/WMF/ZIP	
File9	DOC/PPT/XLS/JPEG/GIF/WMF/ZIP	
File10	DOC/PPT/XLS/JPEG/GIF/WMF/ZIP	

5. For the ZIP file:

Identify the file name contained within the ZIP file:

What is the termination character used to terminate the file name:

Can you tell the date and time that it was last modified?

6. Now add a new button and give it the text of **Identify File**, and use it to read in a file, and to try and determine the file type from the basic header signature. For example, the following shows some of the code required to identify a ZIP file and a JPEG file:

0x identifies a hex format

```
textBox1.Text="";
textBox2.Text="";

DialogResult result = this.openFileDialog1.ShowDialog();

textBox3.Text=openFileDialog1.FileName;

byte [] buff= getBytes(openFileDialog1.FileName);

if (buff[0]==0x50 && buff[1]==0x4B) textBox1.Text="ZIP file";
```

```
else if (buff[0]==0xff && buff[1]==0xD8) textBox1.Text="JPEG file";
else textBox1.Text="Not known";
```

7. For other binary file formats, determine their signature (if possible).

PDF file signature:

SWF (Flash) file signature:

DLL file signature:

RTF file signature (open up a Word document, and save it in an RTF file format):

XML file signature (open up a Word document, and save it in an XML file format):
[or use: http://www.dcs.napier.ac.uk/~bill/1.xml]

8. Modify the program in 6 to identify these files.

Appendix

JPEG file format:
FFD8 – start of image
length -- two bytes
identifier -- five bytes: 4A, 46, 49, 46, 00 (the ASCII code equivalent of a zero terminated "JFIF" string)
version -- two bytes: often 01, 02

ZIP file format:

```
00    ZIPLOCSIG     HEX  504B0304    ;Local File Header Signature
04    ZIPVER        DW   0000        ;Version needed to extract
06    ZIPGENFLG     DW   0000        ;General purpose bit flag
08    ZIPMTHD       DW   0000        ;Compression method
0A    ZIPTIME       DW   0000        ;Last mod file time (MS-DOS)
0C    ZIPDATE       DW   0000        ;Last mod file date (MS-DOS)
0E    ZIPCRC        HEX  00000000    ;CRC-32
12    ZIPSIZE       HEX  00000000    ;Compressed size
16    ZIPUNCMP      HEX  00000000    ;Uncompressed size
1A    ZIPFNLN       DW   0000        ;Filename length
1C    ZIPXTRALN     DW   0000        ;Extra field length
1E    ZIPNAME       DS   ZIPFNLN     ;filename
```

GIF file format:
The header is 6 bytes long and identifies the GIF signature and the version number of the chosen GIF specification. Its format is:

- 3 bytes with the characters 'G', 'I' and 'F'.
- 3 bytes with the version number (such as 87a or 89a). Version numbers are ordered

with two digits for the year, followed by a letter ('a', 'b', and so on).

WMF file format:

Standard header of: `d7 cd c6`

Excel file format:

Standard header of: `d0 cf 11 e0 a1 b1 1a`

Byte position 40(hex): `00`

Word file format:

Standard header of: `d0 cf 11 e0 a1 b1 1a`

Byte position 40(hex): `01`

PPT file format:

Standard header of: `d0 cf 11 e0 a1 b1 1a`

Byte position 40(hex): `01`