

LAB 03

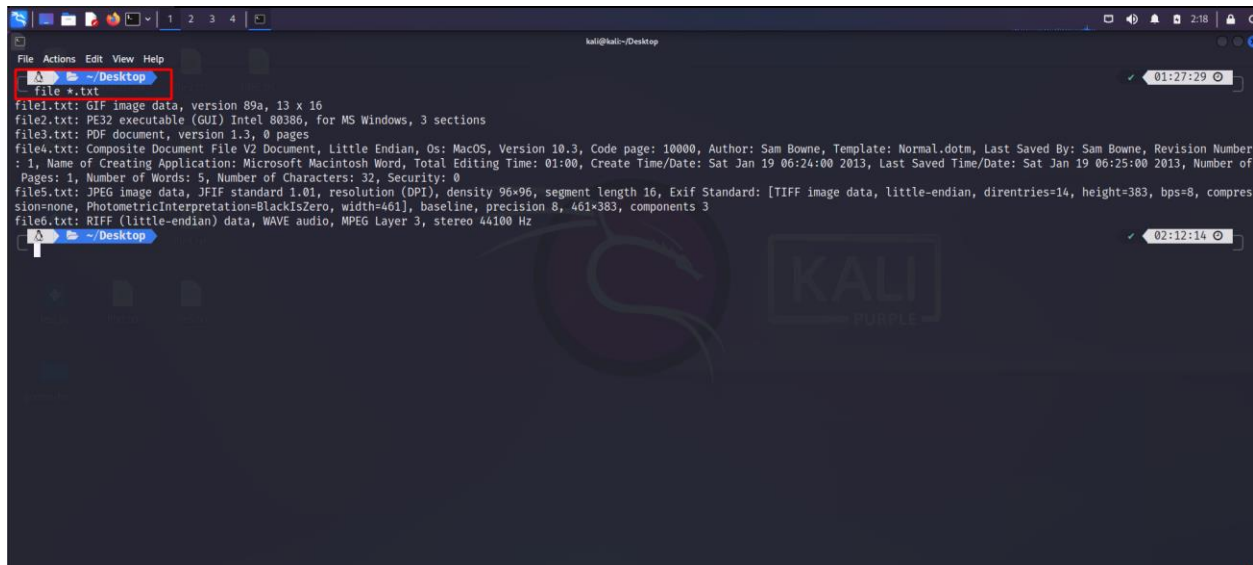
Thầy Mai Hoàng Đình
Trường đại học FPT

Người thực hiện

Đặng Hoàng Nguyên

Lab-Project 3: Identifying File Types

Bước đầu xác định loại file, cách nhanh nhất mà ta có thể làm đó chính là sử dụng command **file** <tên file> trong trường hợp này câu lệnh **file** có chức năng đọc được hết tất cả các header và trả về đúng file của chúng ta



```
file *.txt
file1.txt: GIF image data, version 89a, 13 x 16
file2.txt: PE32 executable (GUI) Intel 80386, for MS Windows, 3 sections
file3.txt: PDF document, version 1.3, 0 pages
file4.txt: Composite Document File V2 Document, Little Endian, Os: MacOS, Version 10.3, Code page: 10000, Author: Sam Bowne, Template: Normal.dotm, Last Saved By: Sam Bowne, Revision Number: 1, Name of Creating Application: Microsoft Macintosh Word, Total Editing Time: 01:00, Create Time/Date: Sat Jan 19 06:24:00 2013, Last Saved Time/Date: Sat Jan 19 06:25:00 2013, Number of Pages: 1, Number of Words: 5, Number of Characters: 32, Security: 0
file5.txt: JPEG image data, JFIF standard 1.01, resolution (DPI), density 96x96, segment length 16, Exif Standard: [TIFF image data, little-endian, direntries=14, height=383, bps=8, compression=none, PhotometricInterpretation=BlackIsZero, width=461], baseline, precision 8, 461x383, components 3
file6.txt: RIFF (little-endian) data, WAVE audio, MPEG Layer 3, stereo 44100 Hz
```

Theo như ta thấy được theo từng tệp nó đã lọc ra được các loại file như sau:

File1 TXT → Đây là file GIF

File2 TXT → Đây là file EXE

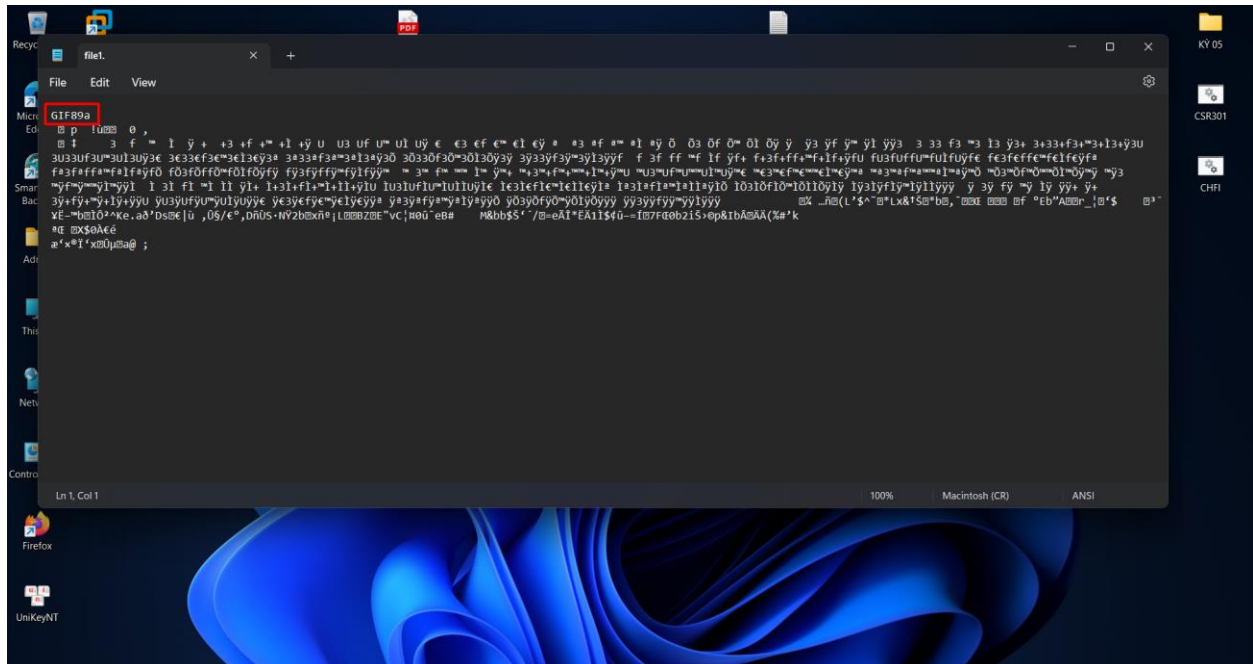
File3 TXT → Đây là file PDF

File4 TXT → đây là file Doc

File5 TXT → Đây là file jpeg

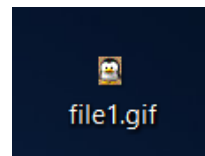
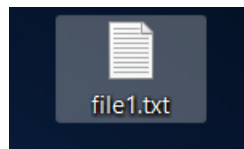
File6 TXT → Đây là file nhạc WAVE

Một cách khác mà chúng ta có thể làm là bật chúng trên notepad, bật trên notepad thì có thể ra những signature của một file. Lấy ví dụ như file File1 TXT



Như có thể thấy ở đây magic byte của nó hiện thị dưới định dạng Ascii là GIF89a, đó là file GIF.

Việc của chúng ta bây giờ chính cần chỉnh file extension của chúng về với đúng định dạng của nó. Đổi từ **FILE1.txt** thành **FILE1.gif**



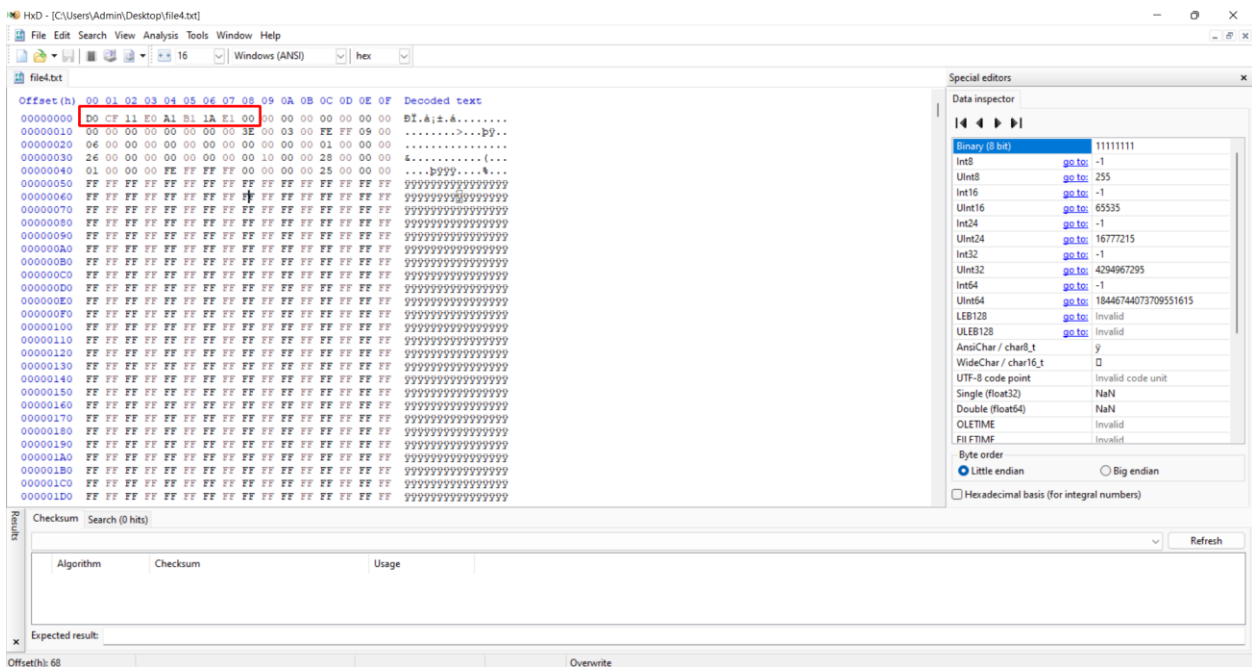
Làm tương tự với những file còn lại, bằng cách mở notepad, ta sẽ thấy được như sau:

FILE2 TXT:



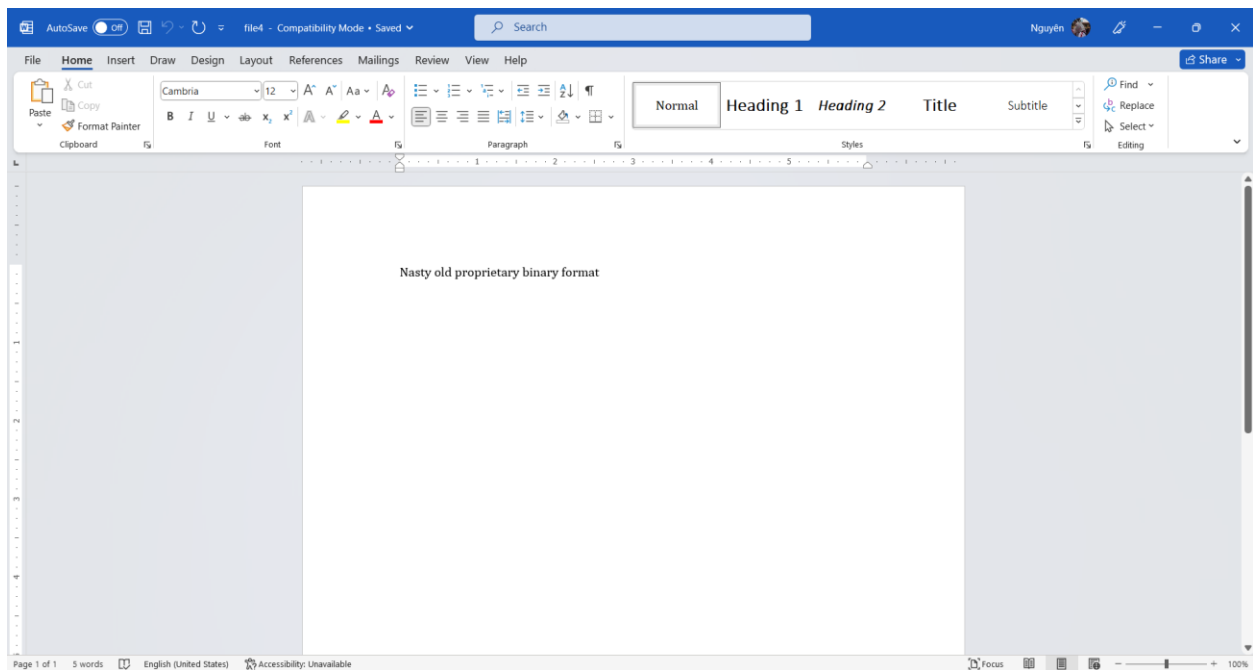
Thoạt nhìn qua file, ta có thể thấy rằng header của nó hơi lạ, tại vì nó để định dạng không phải định dạng Unicode, ở đây ngoài cách bỏ vào linux sử dụng command **file** nêu trên, ta có thể dùng **HXD** để phân tích đoạn byte hex của chúng.

Theo như trong HXD, ta có thể thấy rằng header của chúng bắt đầu là



Sau khi search ta thấy rằng header của nó chính là File office. Việc của chúng ta chỉ có việc là đổi sang những định dạng sau: doc, xls, ppt, msi, msg

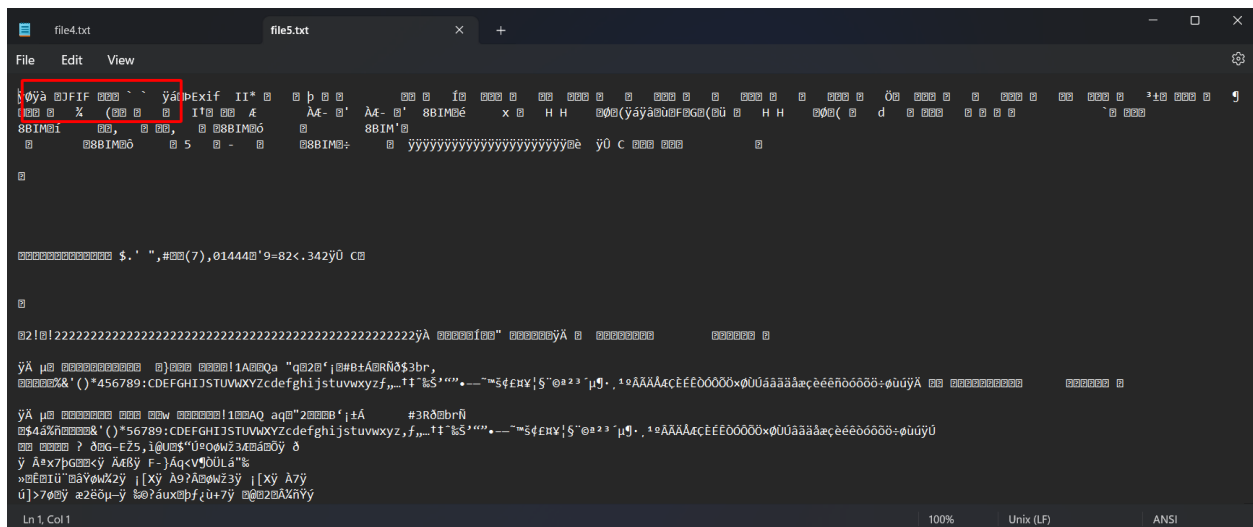
File Signature	File Type	File Format	File Description
69 60 61 67 65 20 66 69 6C 65	C64 tape image file	0	t84 Commodore 64 tape image
43 36 34 20 43 41 52 54 52 49 44 47 45 20 20 20	C64 CARTRIDGE	0	crt Commodore 64 cartridge image
53 49 40 50 4C 45 20 20 20 30 20 54	SIMPLE =~~~~~ ~~~~~ ~~~~~T	0	fits Flexible Image Transport System (FITS)^[34]
66 4C 61 43	flAc	0	flac Free Lossless Audio Codec^[35]
4D 54 68 64	MThd	0	mid midi MIDI sound file^[36]
50 CF 11 E0 A1 B1 1A E1	DL\à\z\á	0	doc xis ppt msl msg Compound File Binary Format , a container format defined by Microsoft COM. It can contain the equivalent of files and directories. It is used by Windows Installer and for documents in older versions of Microsoft Office . ^[37] It can be used by other programs as well that rely on the COM and OLE APIs.
64 65 78 0A 30 33 35 00	dex\035\	0	dex Dalvik Executable
4B 44 4D	KDM	0	vmdk VMDK files^[38]^[39]
23 20 44 69 73 68 20 44 65 73 63 72 69 70 74 6F	# Disk Descripto	0	vmdk VMware 4 Virtual Disk description file (split disk)
43 72 32 34	Cr24	0	crx Google Chrome extension^[40] or packaged app ^[41]
41 47 44 33	AGD3	0	fh8 FreeHand 8 document^[42]^[43]^[44]
05 07 00 00 42 4F 42 4F 05 07 00 00 00 00 00 00 00 00 00 00 00 01	'\''\BOBO '\''\~~~~~ '\''\~~~~~	0	cwk AppleWorks 5 document
06 07 E1 00 42 4F 42 4F	'\''\BOBO	0	



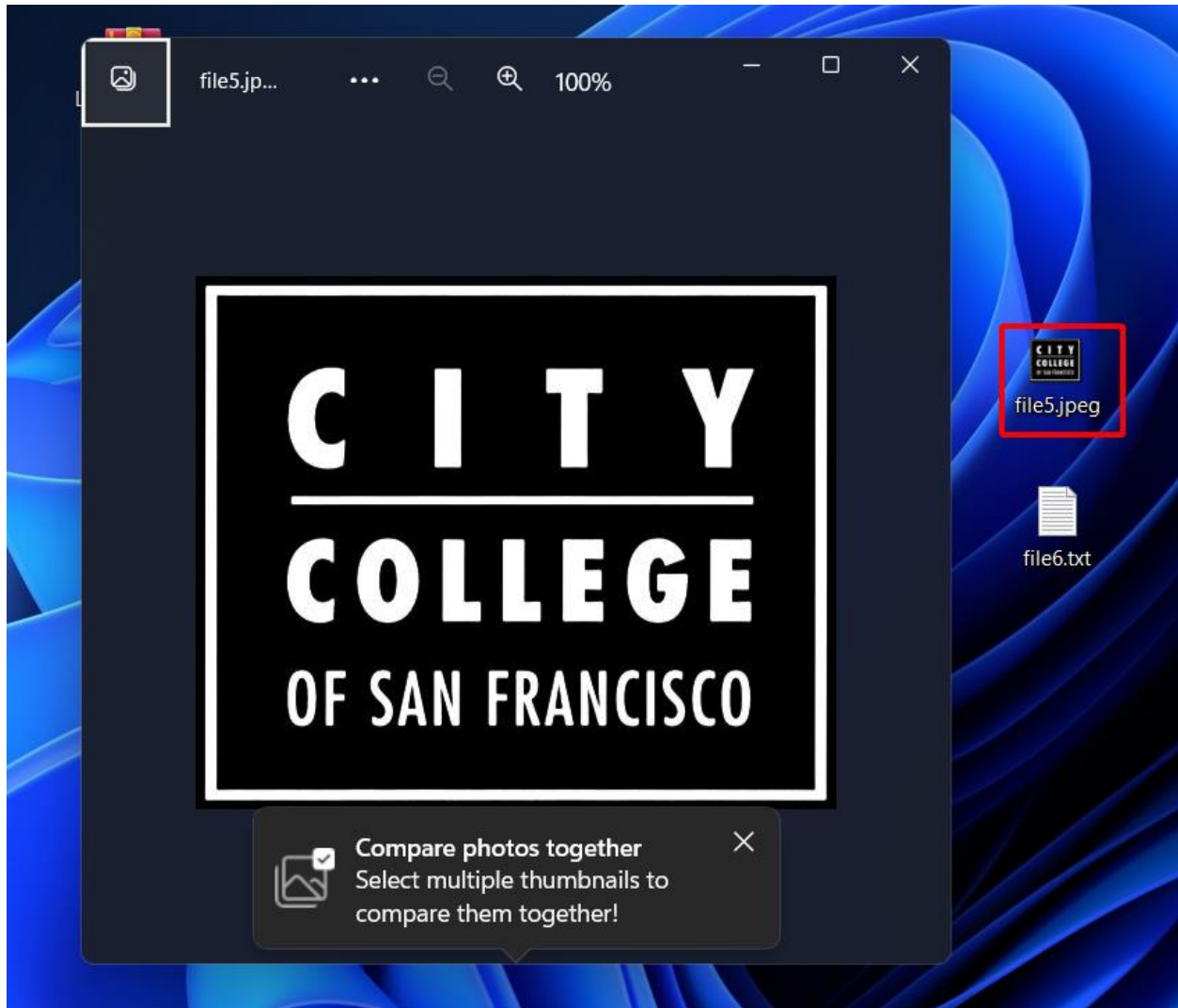
Đây là nội dung của file FILE4 TXT

File5 TXT

Vẫn tiếp tục bỏ vào notepad xem, ta có thể thấy đây là file jpeg với đầu file là jfif



Việc của chúng ta bây giờ chỉ cần chỉnh file extension về jpeg hoặc jpg là có thể mở được



Đây là bức hình sau khi hồi phục xong

File6 TXT

Vẫn như các file trước, ta sẽ bắt đầu phân tích bằng cách mở bằng notepad. Nhìn nóng theo header, chúng ta có thể thấy được rằng đây là một file WAV. Một gile âm thanh

