LAB 03

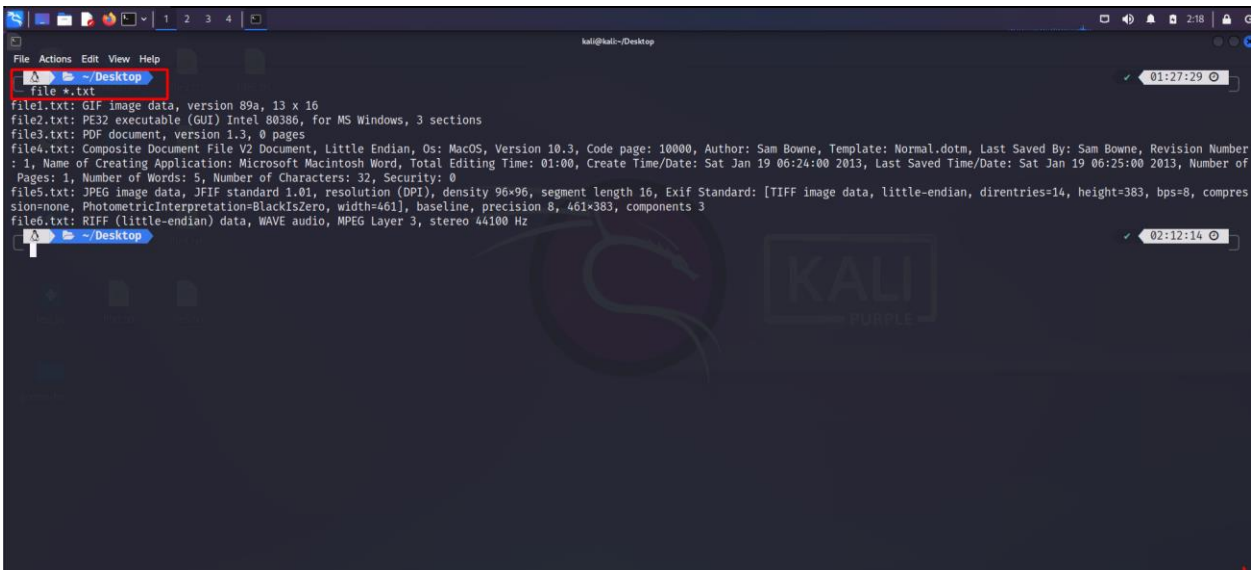
Thầy Mai Hoàng Đỉnh
Trường đại học FPT

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



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 fille EXE

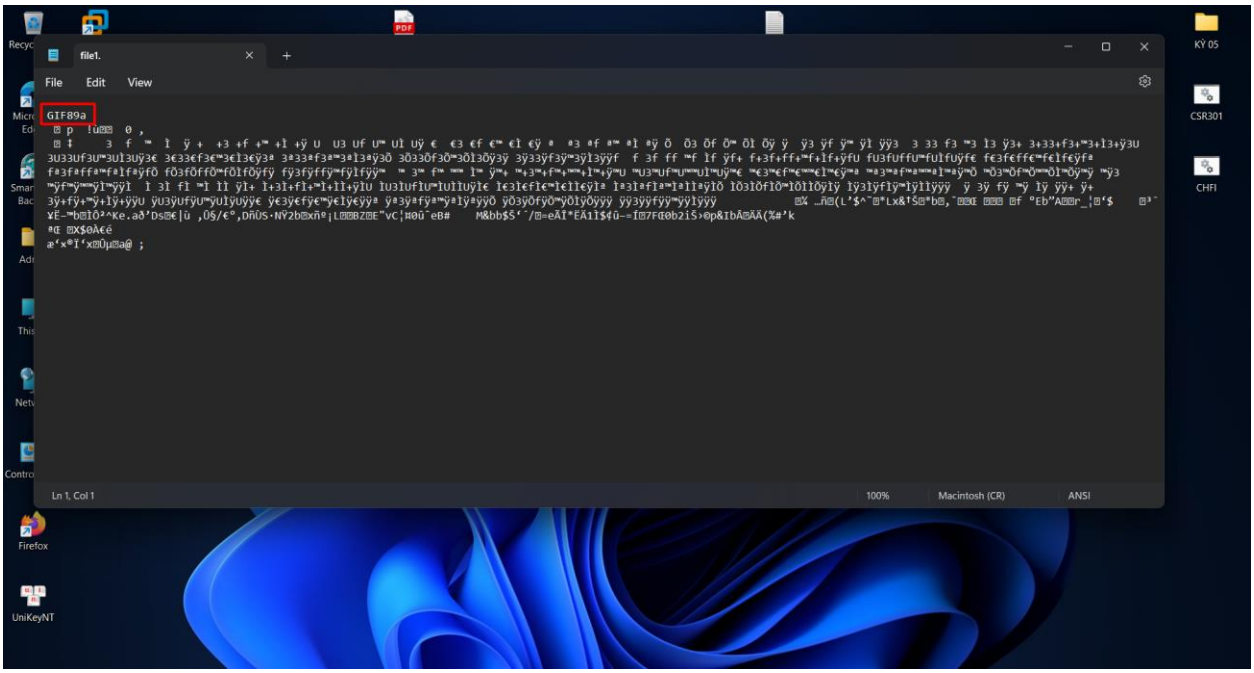
File3 TXT → Đây là file PDF

File4 TXT → đây là file Doc

File5 TXT → Đây là file jpeg

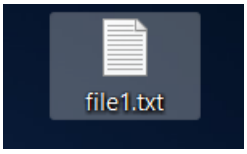
File6 TXT → Đây là file nhạc WAVE

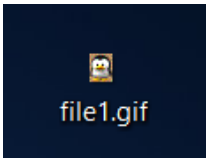
Một cách khác mà chúng ta có thể làm là bật chúng trên notepad, bật trên notepad thì có thể ra những signature của một file. Lấy ví dụ như file File1 TXT



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

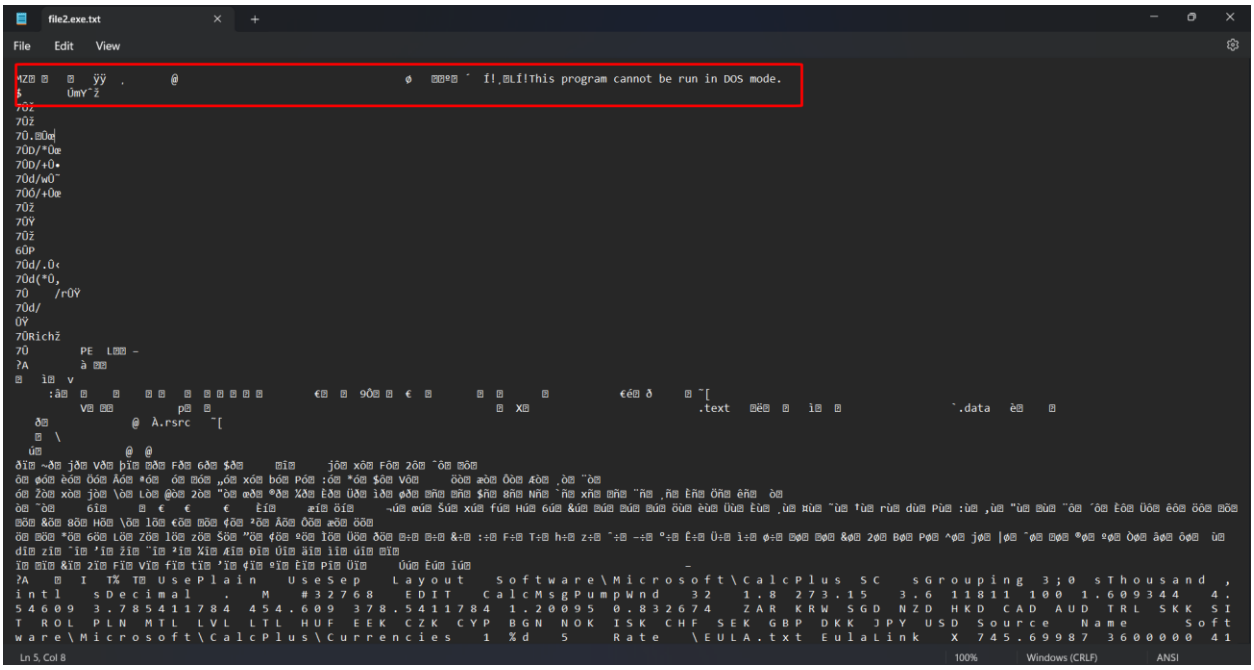
Việc của chúng ta bây giờ chỉnh cần chỉnh file extension của chúng về với đúng định dạng của nó. Đổi từ **FILE1.txt** thành **FILE1.gif**



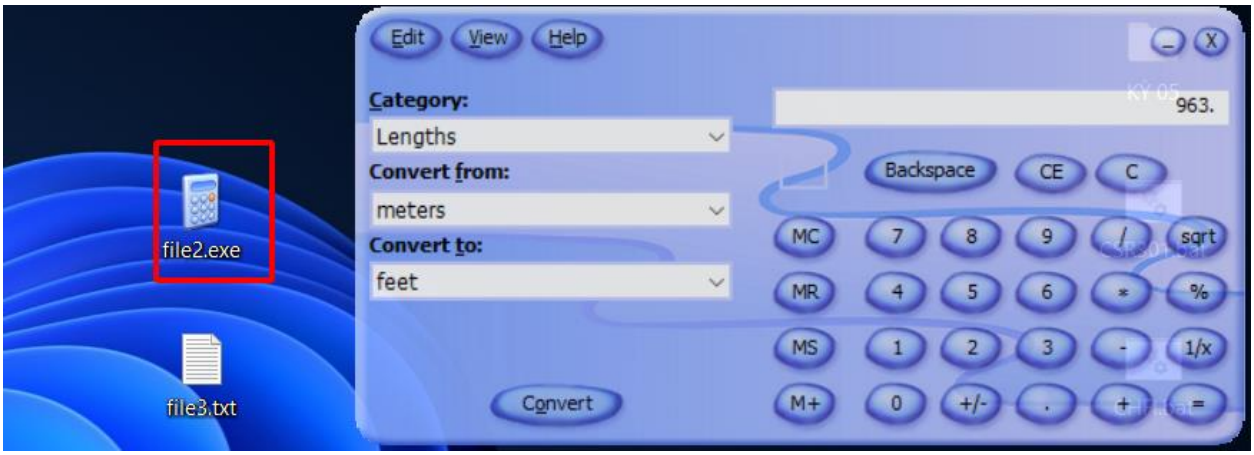


Làm tương tự với những file còn lại, bằng cách mở notepad, ta sẽ thấy được như sau:

FILE2 TXT:



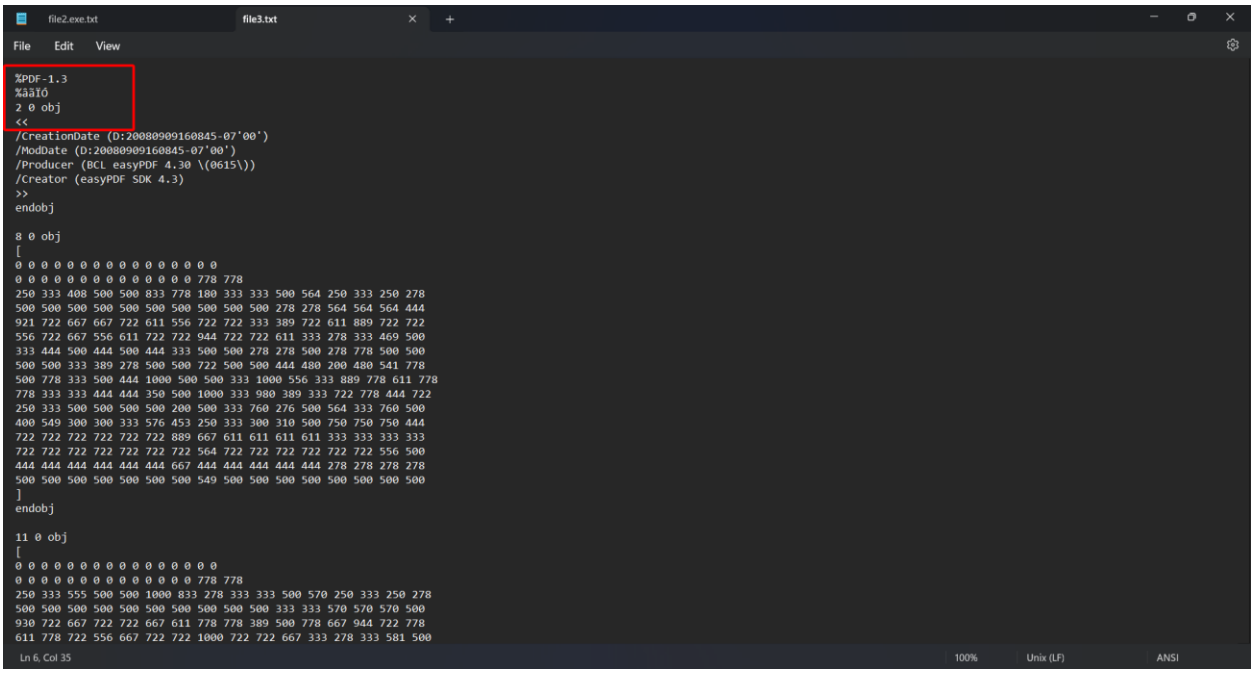
Đối với file này khi mở bằng notepad chúng ta sẽ thấy rằng ở đây nó có đầu MZ, đây là đầu file thực thi dành cho window, còn file thực thi của linux là ELF. Việc của chúng ta chỉ cần chỉnh về đuôi file .exe thì có thể thực thi file



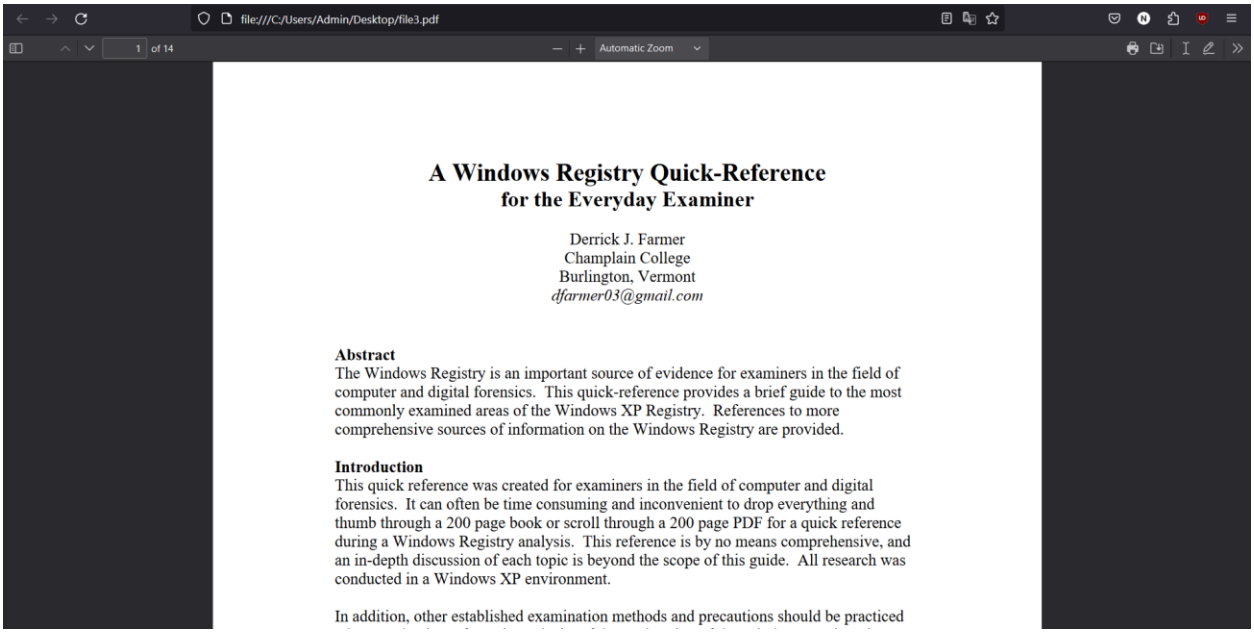
Sau khi đổi về đuôi file exe, ta có thể thấy rằng đây là một chương trình máy tính để tính toán

FILE3 TXT:

Mở bằng notepad **FILE3 TXT**



Dễ dàng nhận biết rằng đây chính là **File PDF** version 1.4 Ta chỉ cần đổi định dạng về đuôi .pdf là có thể mở được file đó lên



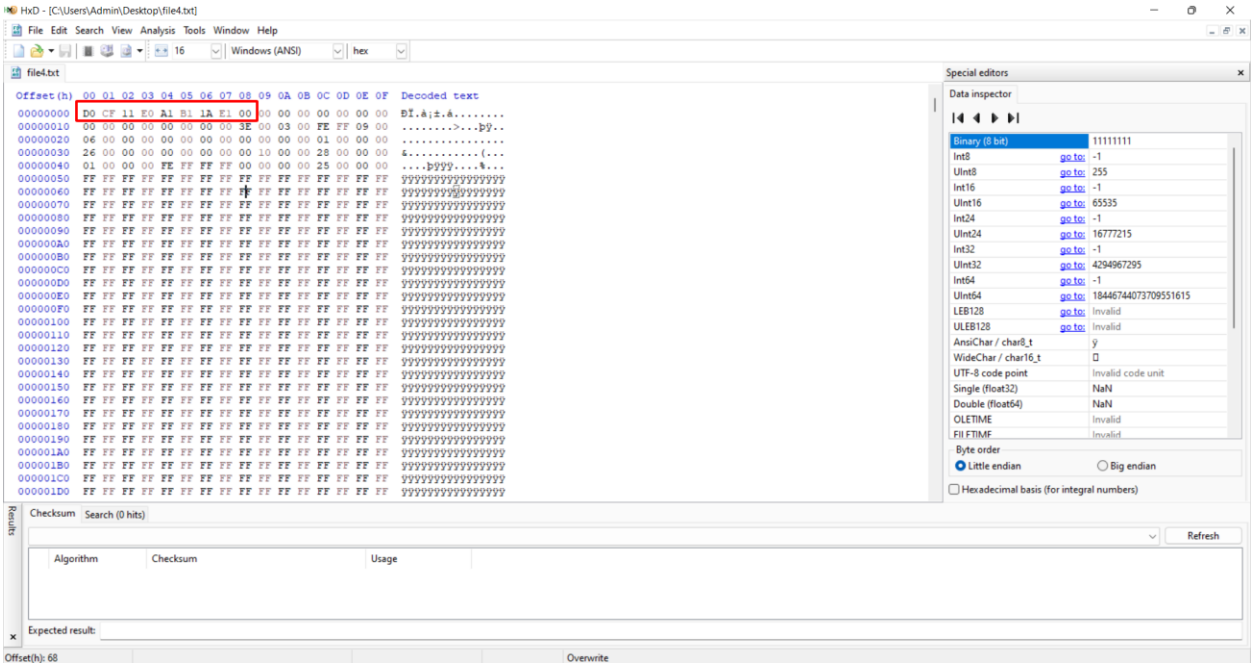
Đây là nội dung bên trong **file3.pdf**

FILE4 TXT

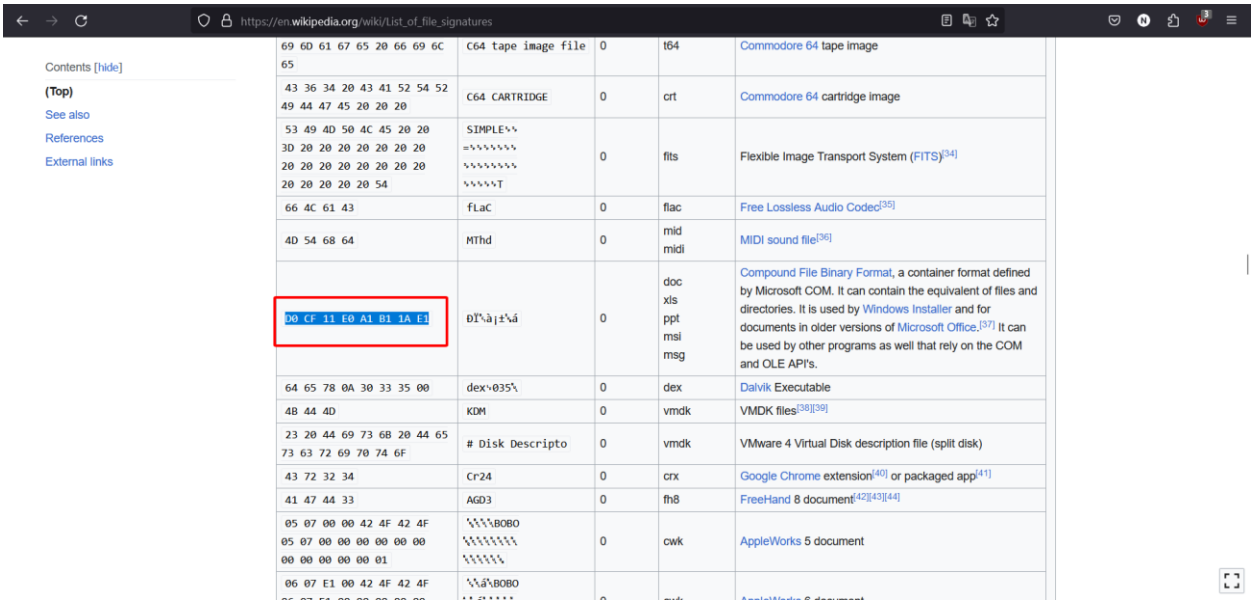
Vẫn tiếp tục mở bằng notepad, ta sẽ tiếp tục phân tích header của nó.

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 dung **HXD** để phân tích đoạn byte hex của chúng.

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

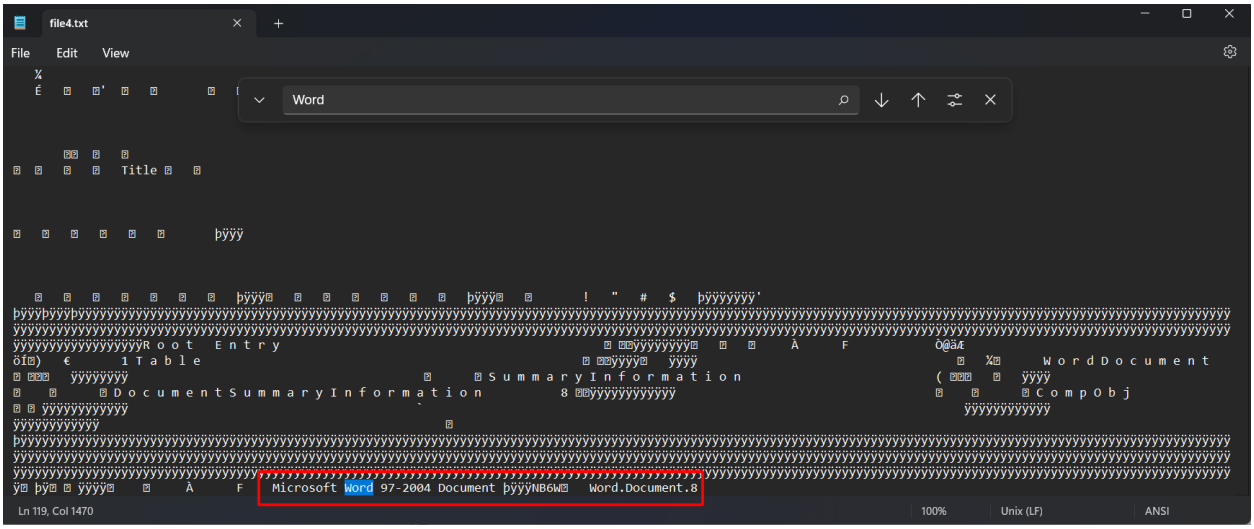


Sau khi search ta thấy rằng header của nó chính là File office. Việc của chúng ta chỉ có việc là đổi sang những định sạng sau: doc, xls, ppt, msi, msg

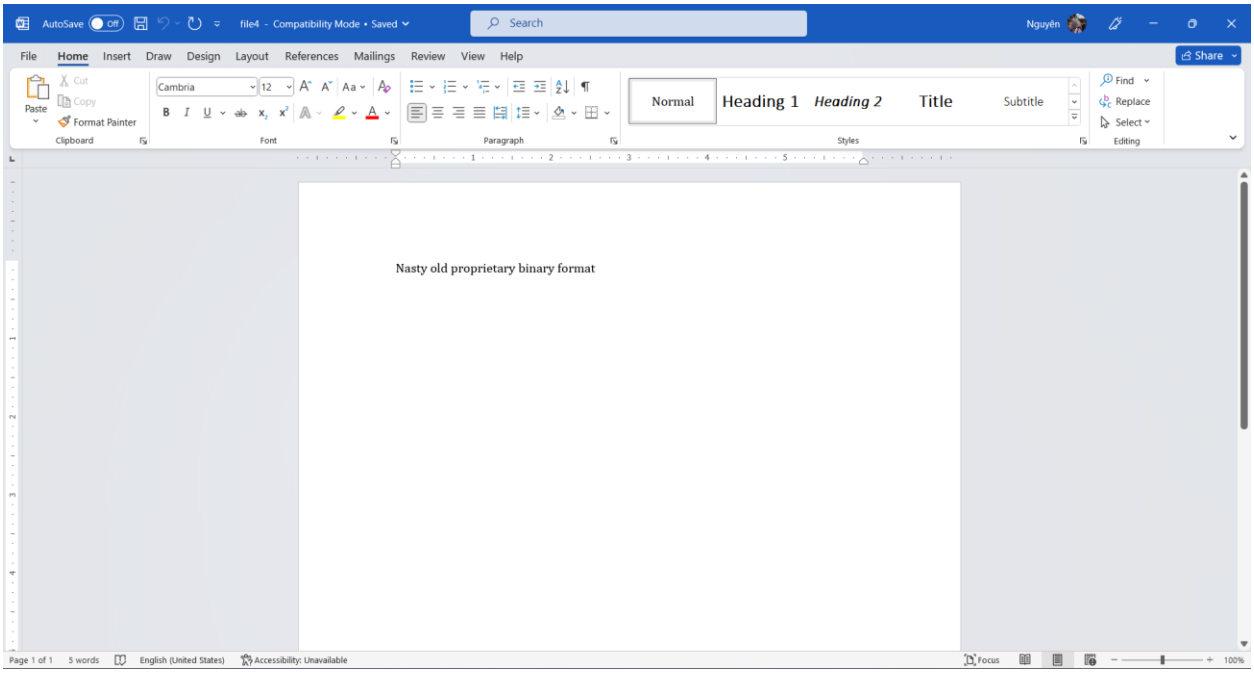


Dựa vào những kinh nghiệm của em, để có thể phân biệt nó là file word, excel, powerpoint, ta tiếp tục tìm bên trong notepad. Ví dụ, muốn biết nó là file word, ta sẽ search từ Word. Tại vì khi search kiểu này, như em đã nói từ trước, word hay excel gì cũng là file nén. Vì thế lúc search, nó sẽ search XML, thì trng các XML nó sẽ có một file chứa định dạng của file

Sau khi search một hồi, ta có thể thấy rằng đây là file word 2003



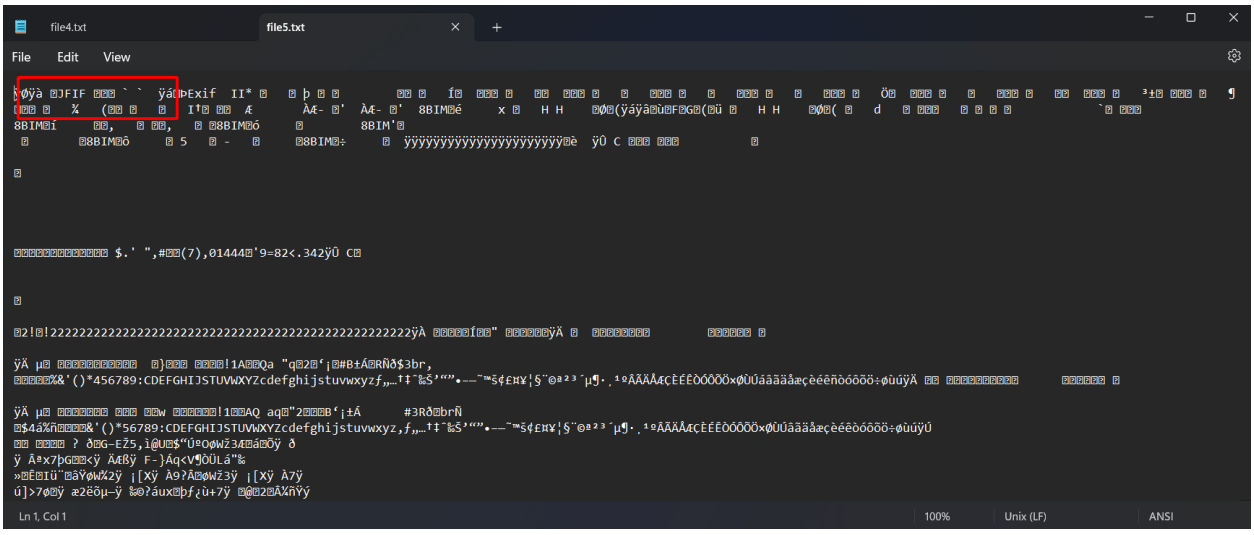
Chuyển file về doc ta có thể thấy được nội dung bên trong



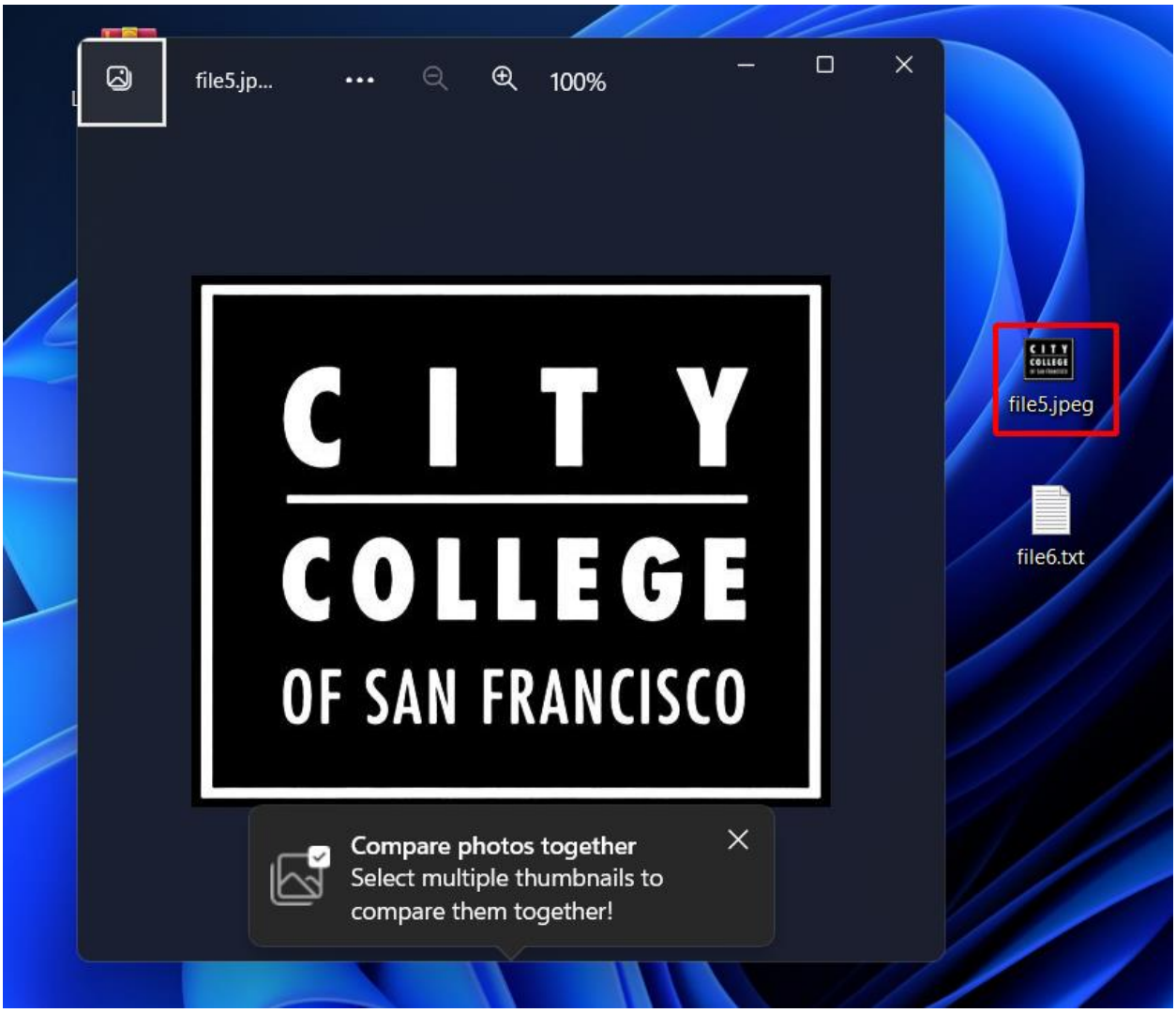
Đây là nội dung của file FILE4 TXT

File5 TXT

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



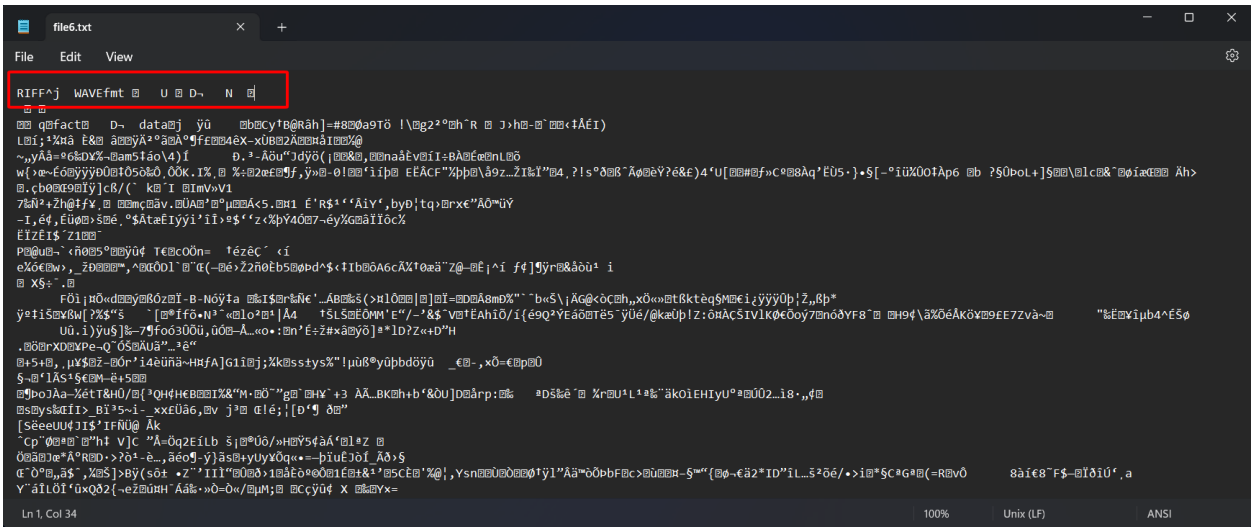
Việc của chúng ta bây giờ chỉ cần chỉnh file extension về jpeg hoặc jpg là có thể mở được



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



Nhìn ngay phần RIFF WAVE. Ta có thể dễ dàng nhận ra đây là file WAV. Chuyển đổi file extension thành WAV ta thu được kết quả sau: