Detecting duplicates

Project members

Qirui Chen

Yunkai Cui

Shaohua Du

# Introduction

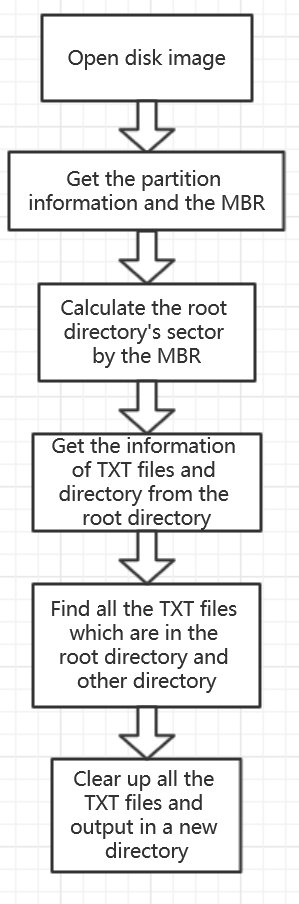
A tool that can find and sort out duplicate files and different files from several directory in a disk or disk image and generate a new directory set with every different files, making the directory clear and improving memory utilization.

Limitations

Detecting duplicates can find and sort out duplicate files and different files from several directory in a disk or disk image and generate a new directory set with every different files. The project will finish when we see the new directory with different files on the desktop of Windows. There are 3 limitations as following:

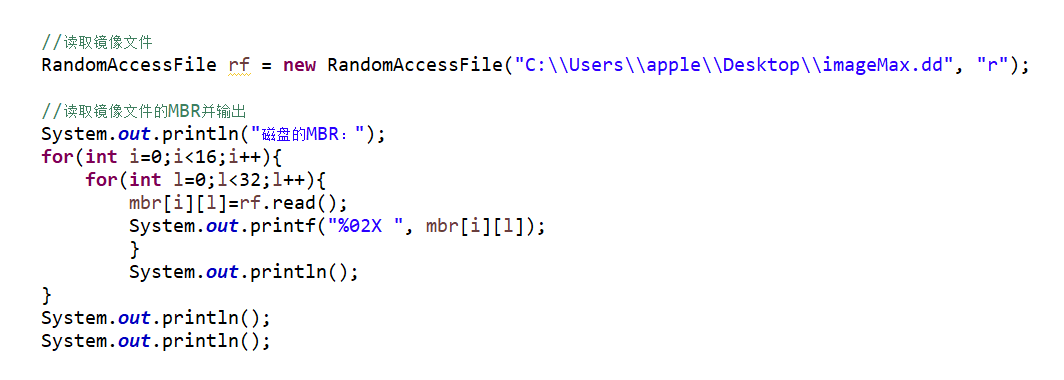
1. The tool will be limited to the TXT type of files.
2. The tool will be limited to the Fat32 type of partition.
3. The tool find the duplicate files only by their names. In other word, the tool judge the duplicate files when their names are the same. It can’t find by the validation of the data entered.

# Design



# Implementation

**First: find the MBR of the image**

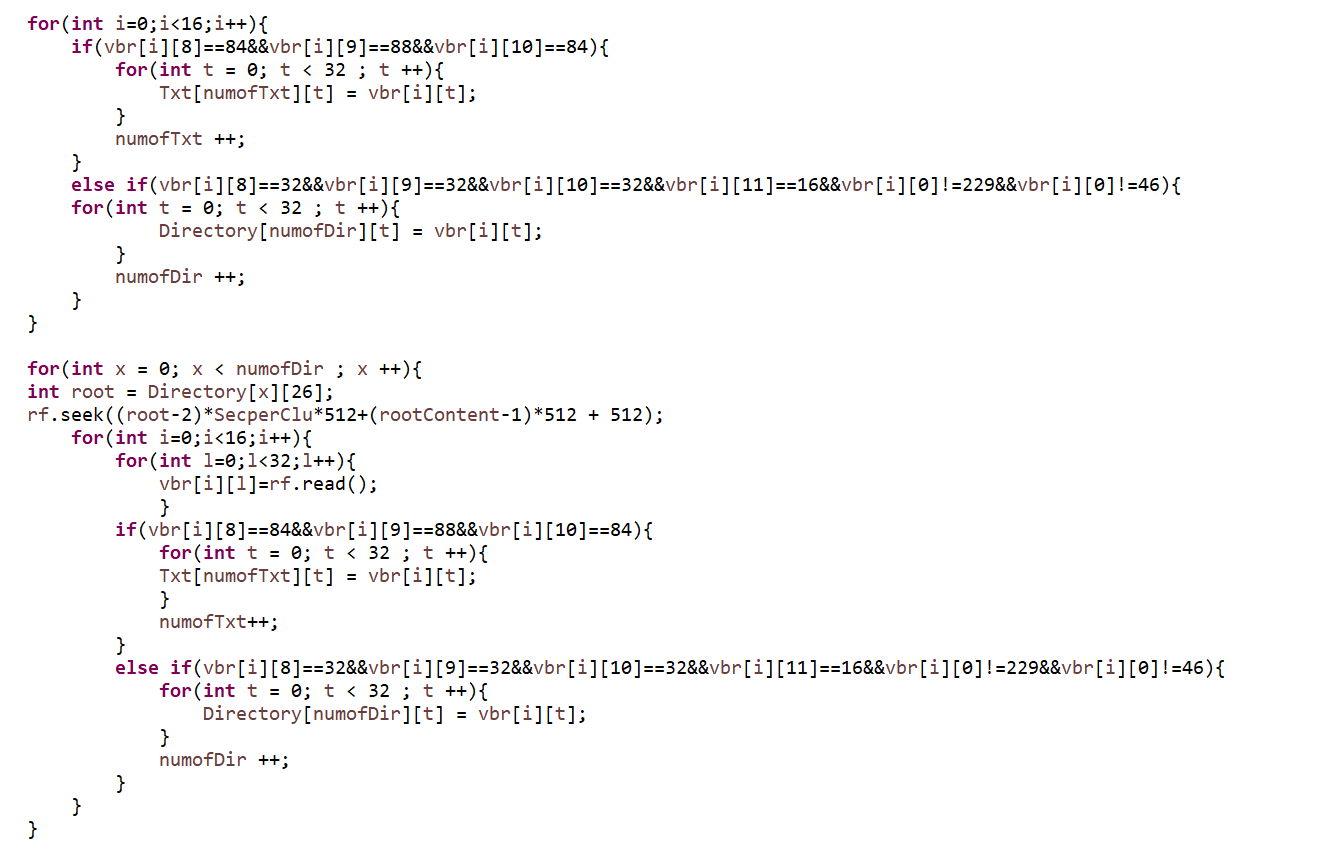
****

**• We find the MBR and establish a two-dimensional array named “mbr” to record it.**

**Second: find root directory and all files**

****

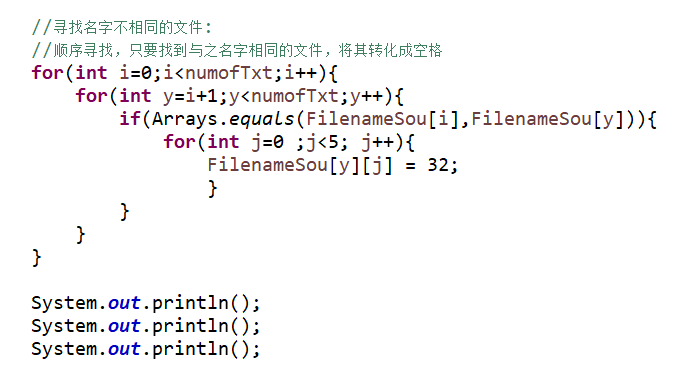
**• For this picture we find the root directory and create a two-dimensional array named “vbr” to record it**

****

**• For this picture we try to find and read all the files in the image**

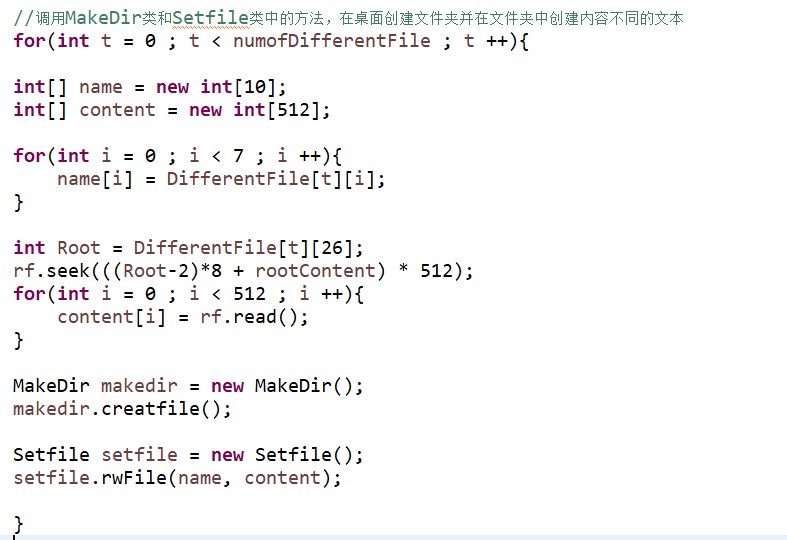
**• We use use “Txt” to record the TXT file and “Directory” to record the files**

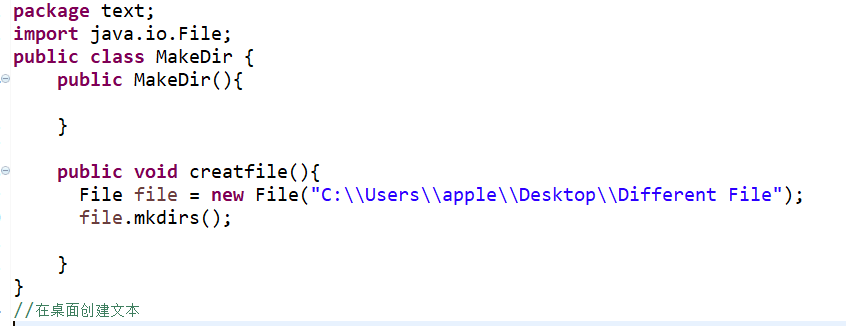
**Third: differentiate the same files**

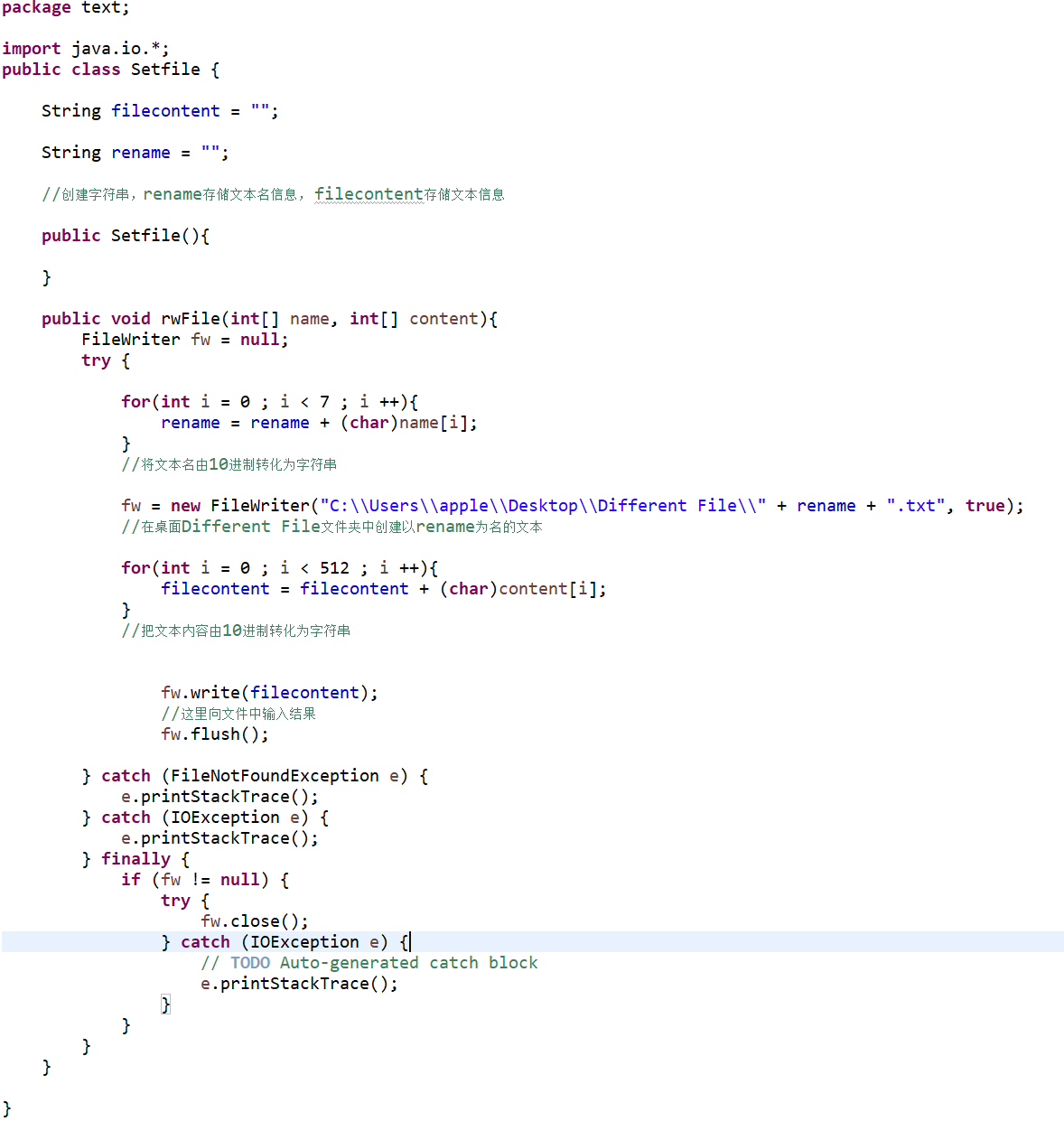
****

**• For this part we compare all the files and get the files which are in different**

**Last: create a file contains all the files which are different from each other**

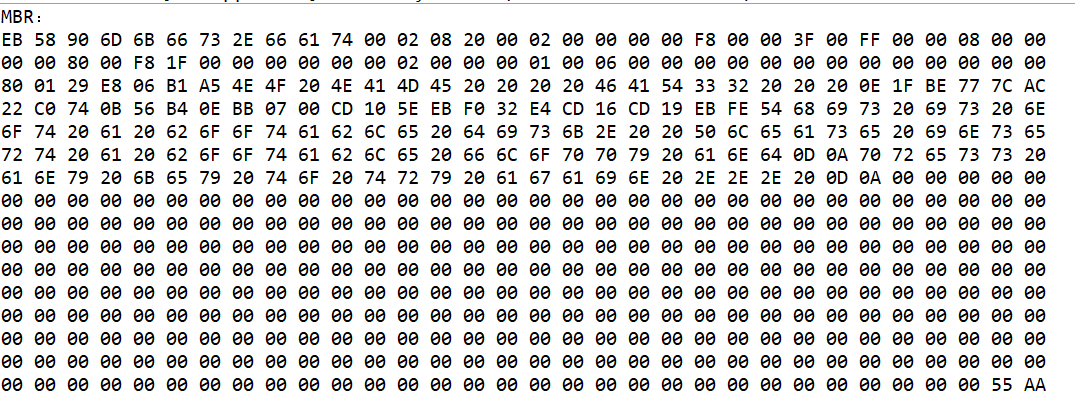
****

****

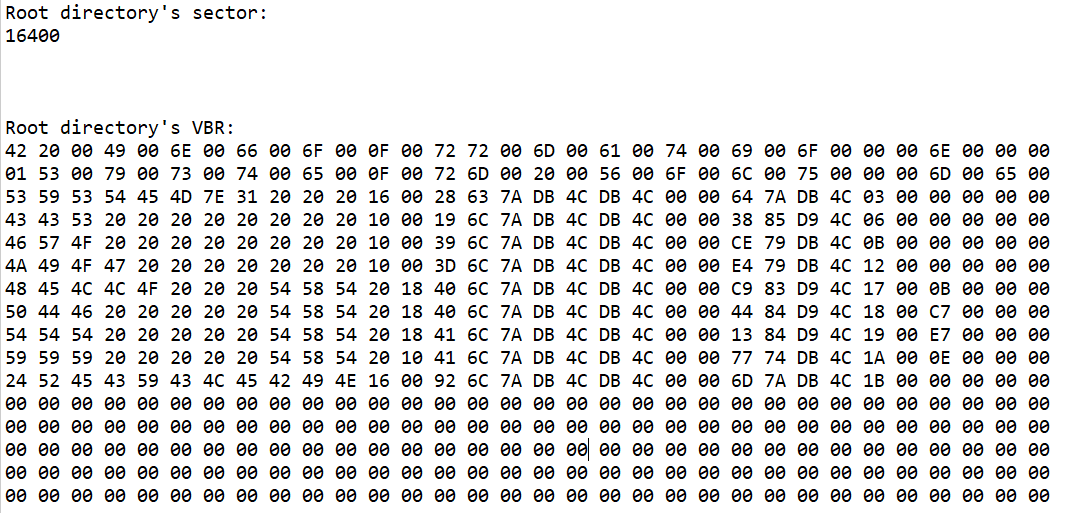
****

# Testing and validation

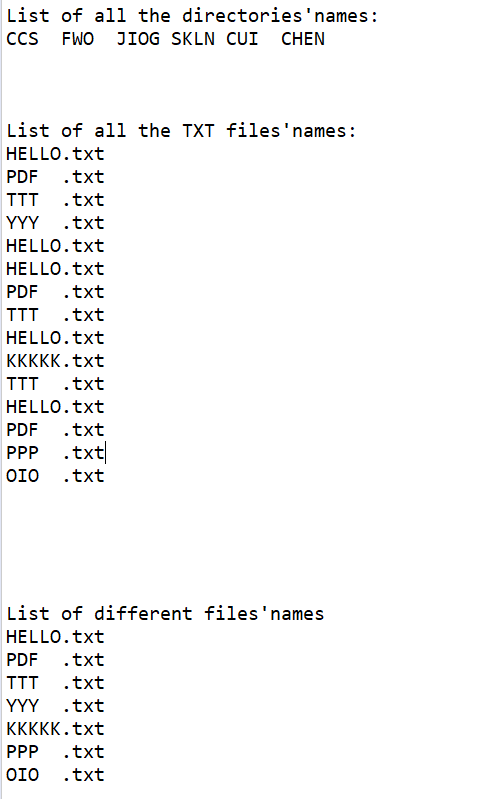
**•Output the MBR**

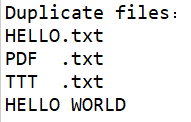


**•Calculate the root directory’s sector and output the VBR of the root directory**

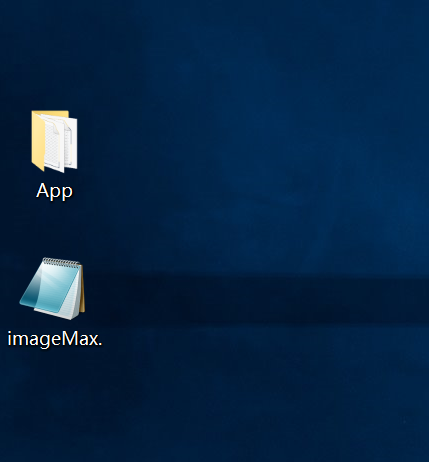


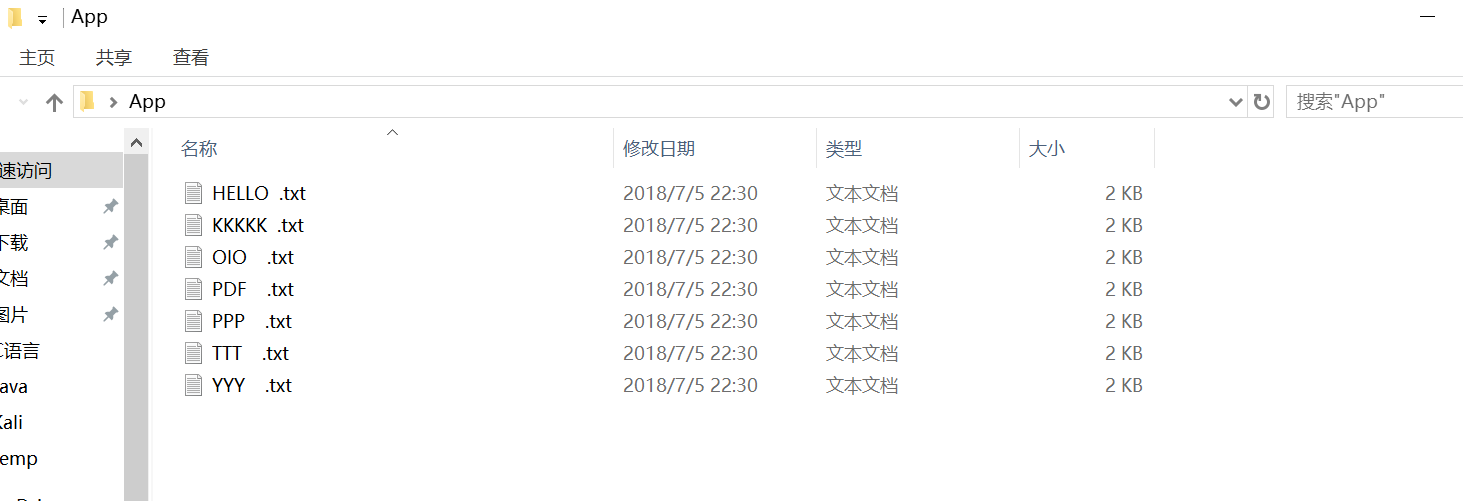
**•Output the file’s information**



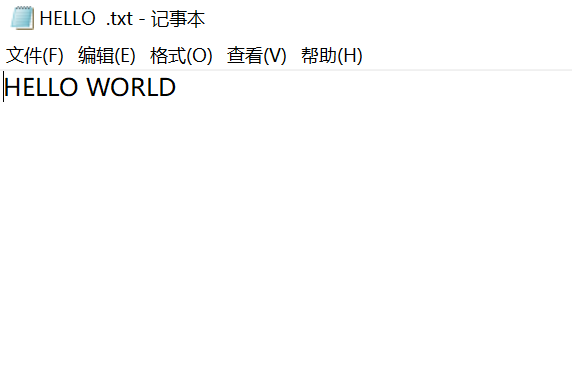


**•Successfully make a new directory with different TXT files on the desktop**





**•Open the file and match with each file in the disk image(“imageMax.”)**



# Summary

We have successfully found all the files and text files and differentiate the same files. It’s really a good experience for us to improve our capability in program. But there’s still some problem like our program can not deal with the file in other types. Also we didn’t separate the code in many section that means our code is hard to read.

# References

1. **Matt, “FAT32.pdf”, pages 0-35**
2. **Brian Carrier, “File System Forensic Analysis”, 2005-3**