

编号：\_\_\_\_\_

实验	一	二	三	四	五	六	七	八	总评	教师签名
成绩										

武汉大学国家网络安全学院

# 课程实验（设计）报告

课程名称：\_\_\_\_\_ 计算机病毒

实验名称：\_\_\_\_\_ 磁盘结构与文件系统

专业（班）：\_\_\_\_\_ 网络安全专业

队长学号：\_\_\_\_\_ 2020302181081

队长姓名：\_\_\_\_\_ 陈曦

队员学号：\_\_\_\_\_ 2020302181202

队员姓名：\_\_\_\_\_ 梁刘琪

任课教师：\_\_\_\_\_ 陈泽茂

2023 年 3 月 20 日

# 目 录

一. 实验目的 .....	3
二. 实验内容 .....	3
三. 实验关键过程、数据及其分析 .....	4
四. 实验总结 .....	15
五. 实验心得（个人贡献） .....	18

## 一. 实验目的

1. 加深对 FAT32 分区及文件系统格式的理解
2. 掌握借助 WinHex 等工具手工定位磁盘文件数据的技能
3. 通过开发一个磁盘文件数据提取工具，强化编程实践能力。

## 二. 实验内容

### 1. 手工定位和提取 FAT32 分区中的文件数据

在 FAT32 分区下创建一个不小于 10K 的 Word 文档（文件名为本组组长名字的拼音，该名字必须在实验过程的截图中明确标出。），根据课上介绍的 FAT32 分区及文件系统知识，必要时自行上网查阅相关资料，借助 WinHex 或其它十六进制工具，以手工方式从磁盘中逐一找到该文件的各个存储扇区，复制其中的有效内容，并拼接组合成一个与原文档内容相同的完整文档。

### 2. 编程实现“内容一”的全过程

输入：任一文件（A）的路径

输出：

- （1）该文件的短文件名目录项信息
- （2）该文件的簇链
- （3）根据上述的文件簇链，从磁盘上提取数据并拼接而得的新文件（B）
- （4）文件 A 与文件 B 内容的比较结果（要求二者完全一致）

### 三. 实验关键过程、数据及其分析

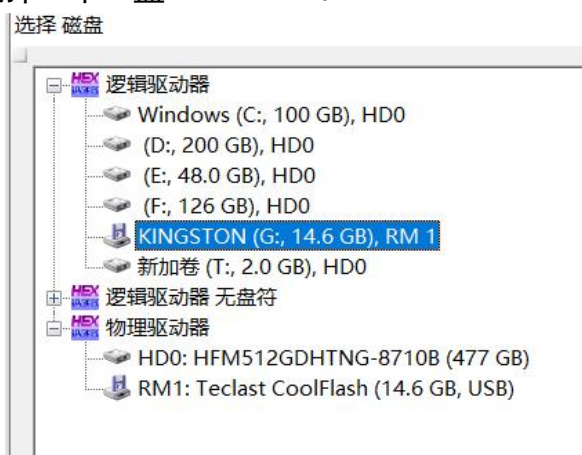
#### 1. 手工定位和提取 FAT32 分区中的文件数据

##### 【步骤思路】

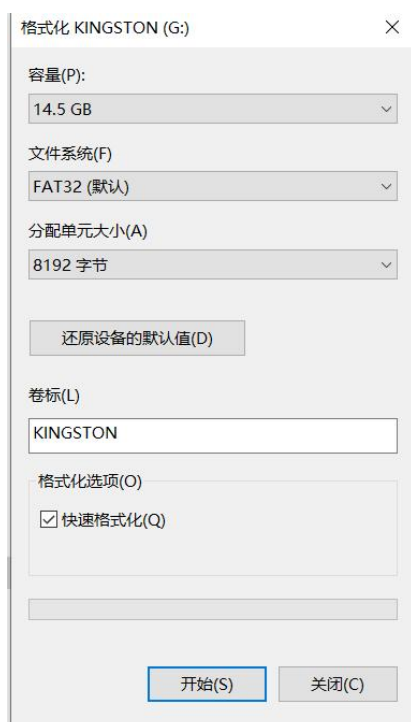
- (1) 首先从 0 扇区中读取出磁盘的基本信息，得到根目录的位置。
- (2) 将根目录所在扇区的位置读取出来，找到目标文件的起始簇号。
- (3) 找到文件对应的 FAT 表项对文件进行顺序读取。

##### 【环境准备】

选择一个 U 盘 KINGSTON。



格式化 U 盘。



创建一个名为 cx. txt 的文件。

> KINGSTON (G:)



安装 WinHex。



## 【使用 WinHex 分析】

这里要求的是直接从磁盘，文件系统的层面来提取数据。需要用到的工具是 winHex 。我们将 U 盘格式化为 FAT32 格式，创建一个文件，然后使用 winHex 将打开磁盘。

(1) 查看第一个扇区——引导扇区

得到的内容如下，可以提取出一些基本信息：

Offset	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
00000000	EB	58	90	4D	53	44	4F	53	35	2E	30	00	02	10	6E	0B
00000010	02	00	00	00	00	F8	00	00	3F	00	FF	00	30	00	00	00
00000020	40	C0	D2	01	49	3A	00	00	00	00	00	00	02	00	00	00
00000030	01	00	06	00	00	00	00	00	00	00	00	00	00	00	00	00
00000040	80	00	29	B0	45	BA	D6	4E	4F	20	4E	41	4D	45	20	20
00000050	20	20	46	41	54	33	32	20	20	20	33	C9	8E	D1	BC	F4
00000060	7B	8E	C1	8E	D9	BD	00	7C	88	56	40	88	4E	02	8A	56
00000070	40	B4	41	BB	AA	55	CD	13	72	10	81	FB	55	AA	75	0A
00000080	F6	C1	01	74	05	FE	46	02	EB	2D	8A	56	40	B4	08	CD
00000090	13	73	05	B9	FF	FF	8A	F1	66	0F	B6	C6	40	66	0F	B6
000000A0	D1	80	E2	3F	F7	E2	86	CD	C0	ED	06	41	66	0F	B7	C9
000000B0	66	F7	E1	66	89	46	F8	83	7E	16	00	75	39	83	7E	2A
000000C0	00	77	33	66	8B	46	1C	66	83	C0	0C	BB	00	80	B9	01
000000D0	00	E8	2C	00	E9	A8	03	A1	F8	7D	80	C4	7C	8B	F0	AC
000000E0	84	C0	74	17	3C	FF	74	09	B4	0E	BB	07	00	CD	10	EB
000000F0	EE	A1	FA	7D	EB	E4	A1	7D	80	EB	DF	98	CD	16	CD	19
00000100	66	60	80	7E	02	00	0F	84	20	00	66	6A	00	66	50	06
00000110	53	66	68	10	00	01	00	B4	42	8A	56	40	8B	F4	CD	13
00000120	66	58	66	58	66	58	66	58	EB	33	66	3B	46	F8	72	03
00000130	F9	EB	2A	66	33	D2	66	0F	B7	4E	18	66	F7	F1	FE	C2
00000140	8A	CA	66	8B	D0	66	C1	EA	10	F7	76	1A	86	D6	8A	56
00000150	40	8A	E8	C0	E4	06	0A	CC	B8	01	02	CD	13	66	61	0F
00000160	82	74	FF	81	C3	00	02	66	40	49	75	94	C3	42	4F	4F
00000170	54	4D	47	52	20	20	20	20	00	00	00	00	00	00	00	00
00000180	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00
00000190	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00
000001A0	00	00	00	00	00	00	00	00	00	00	00	00	0D	0A	44	69

卷	文件	预览	详细	缩略图	时间轴	图例说明											
Offset	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F	UTF-8
0000001B0	73	6B	20	65	72	72	6F	72	FF	0D	0A	50	72	65	73	73	sk errorress
0000001C0	20	61	6E	79	20	6B	65	79	20	74	6F	20	72	65	73	74	any key to rest
0000001D0	61	72	74	0D	0A	00	00	00	00	00	00	00	00	00	00	00	art
0000001E0	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	
0000001F0	00	00	00	00	00	00	00	00	AC	01	B9	01	00	00	55	AA	

字节位移对应的定义表如图：

字节位移	字段长度	字段名和定义	第1个分区项
0x01BE	1B	分区状态	
0x01BF	1B	开始磁头号	
0x01C0	6位	低0~5位为开始扇区号 高10位为开始柱面号	
0x01C1	10位		
0x01C2	1B	分区的类型	
0x01C3	1B	结束磁头号	
0x01C4	6位	低0~5位为结束扇区号 高10位为结束柱面号	
0x01C5	10位		
0x01C6	4B	起始扇区的LBA	
0x01CA	4B	分区包含的扇区总数	

提取出的信息结果如下：

含义	数据内容	偏移量
每扇区字节数	0x0200	第 11-12 处
每簇扇区数	0x10	第 13 处
保留扇区数	0x0B6E	第 14-15 处
FAT 表的个数	0x02	第 16 处
该卷总扇区数	0x01D2C040	第 32-25 处
每个 FAT 占用的扇区数	0x00003A49	第 36-69 处
BPB_RootClus	0x00000002	第 44-47 处

计算公式：

数据区起始号 (FirstDataSecto)  $r = \text{保留扇区数} + (\text{FAT 个数} * \text{每个 FAT 所占扇区数})$

n 号簇的第一个扇区号  $\text{FirstSectorofCluster} = ((N-2) * \text{每簇扇区数}) + \text{FirstDataSector}$

计算结果：

我们通过上面可以求出根目录的起始扇区为：  $0x0B6E + 0x02 * 0x3A49 = 0x8000 = 32768$ 。

根目录是 2 号簇。

## (2) 查看根目录扇区

根目录扇区存储的内容就是根目录里的文件（夹）。

Offset	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F	UTF-8
001000000	4E	49	4E	47	53	54	4F	4E	20	20	20	08	00	00	00	00	KINGSTON
001000010	00	00	00	00	00	00	AF	5A	6C	56	00	00	00	00	00	00	
001000020	42	20	00	49	00	6E	00	66	00	6F	00	0F	00	72	72	00	B
001000030	6D	00	61	00	74	00	69	00	6F	00	00	00	6E	00	00	00	mCaCtCiCoCnCn
001000040	01	53	00	79	00	73	00	74	00	65	00	0F	00	72	6D	00	CSCyCsCtCeCrmf
001000050	20	00	56	00	6F	00	6C	00	75	00	00	00	6D	00	65	00	QVCoClCuCmCef
001000060	53	59	53	54	45	4D	7E	31	20	20	20	16	00	7C	AE	5A	SYSTEM~1
001000070	6C	56	6C	56	00	00	AF	5A	6C	56	03	00	00	00	00	00	1VCC
001000080	43	58	20	20	20	20	20	20	54	58	54	20	18	9B	CF	5A	CX
001000090	6C	56	73	56	00	00	D9	59	69	56	05	00	A7	1E	00	00	VsVCC
0010000A0	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	
0010000B0	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	
0010000C0	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	
0010000D0	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	
0010000E0	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	

001000070	6C	56	6C	56	00	00	AF	5A	6C	56	03	00	00	00	00	00	1VCC
001000080	43	58	20	20	20	20	20	20	54	58	54	20	18	9B	CF	5A	CX
001000090	6C	56	73	56	00	00	D9	59	69	56	05	00	A7	1E	00	00	VsVCC
0010000A0	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	
0010000B0	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	
0010000C0	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	

组长姓名cx的文件

字节位移对应的定义表如表：

字段名称	长度	含义	偏移量
jumpBoot	3	跳转指令	0
OEMName	8	标记格式化该分区的操作系统名称与版本号	13
BytesPerClus	2	每扇区字节数	11
SecPerClus	1	每簇扇区数	13
RsvdSecCnt	2	保留扇区数目	14
NumFATs	1	此卷中 FAT 数	16
RootEntCnt	2	FAT32 为 0	17



字段名称	长度	含义	偏移量
TotSec16	2	FAT32 为 0	19
Media	1	存储介质	21
FATSz16	2	FAT32 为 0	22
SecPerTrk	2	磁道扇区数	24
NumHeads	2	磁头数	26
HiddSec	4	FAT 区前隐扇区数	28
TotSec32	4	该卷总扇区数	32
FATSz32	4	FAT 表扇区数	36
ExtFlags	2	FAT32 特有	40
FSVer	2	FAT32 特有	42
RootClus	4	根目录簇号	44
FSInfo	2	文件系统信息	48
BKBootSec	2	通常为 6	50
Reserved	12	扩展用	52

计算文件大小：

可以得到我们需要的文件（前八个字节是文件名）的起始簇号为 0x00000005，文件大小为  $0x1EA7 = 7847$ 。

### 【查看 FAT1 表】

（扇区为 0x0B6E），FAT2 是 FAT1 的一个备份，这里面存储的是簇链：

Offset	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
00016DC00	F8	FF	FF	0F	FF	FF	FF	FF	FF	FF	FF	0F	FF	FF	FF	0F
00016DC10	FF	FF	FF	0F	06	00	00	00	FF	FF	FF	0F	00	00	00	00
00016DC20	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00
00016DC30	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00
00016DC40	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00
00016DC50	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00
00016DC60	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00

我们可以看到第 7 项就是一个结束项，说明一共有 7 个簇，说明我们的文件有两个簇（ $7-2=5$ ）。

计算过程：

文件第 6 簇的偏移 = 根目录首簇偏移 + （文件某簇号 - 根目录首簇号）\* 每簇扇区数  
 $= 32768 + (5 - 2) * 10 = 32798$ 。



查看第 32800 到 32808 个扇区（一簇是 10 个扇区），就是文件内容。

## 【查看文件内容】

Offset	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F	UTF-8
001006000	E6	AF	8F	E4	B8	AA	E4	BA	BA	E9	83	BD	E6	9C	89	E4	每个人都有一
001006010	B8	80	E4	BB	B6	E9	9A	BE	E5	BF	98	E7	9A	84	E4	BA	件难忘的事
001006020	8B	E6	83	85	2C	E8	80	8C	E9	9A	BE	E5	BF	98	E7	9A	情,而难忘的
001006030	84	E4	BA	8B	E6	83	85	E4	B9	9F	E5	B0	B1	E5	83	8F	事情也就像
001006040	E6	B5	B7	E8	BE	B9	E9	82	A3	E5	90	84	E8	89	B2	E5	海边那各色各
001006050	90	84	E6	A0	B7	E7	9A	84	E8	B4	9D	E5	A3	B3	2C	E5	样的贝壳,在
001006060	9C	A8	E6	B5	B7	E6	B0	B4	E7	9A	84	E9	99	AA	E4	BC	海水的陪伴
001006070	B4	E4	B8	8B	E9	97	AA	E9	97	AA	E5	8F	91	E5	85	89	下闪闪发光
001006080	2C	E6	95	A3	E5	8F	91	E7	9D	80	E4	BA	94	E9	A2	9C	,散发着五颜
001006090	E5	85	AD	E8	89	B2	E7	9A	84	E5	85	89	E5	BD	A9	2C	六色的光彩,
0010060A0	E6	95	B0	E9	83	BD	E6	95	B0	E4	B8	8D	E5	AE	8C	E3	数都数不完。
0010060B0	80	82	E4	BB	A4	E4	BD	A0	E6	9C	80	E9	9A	BE	E5	BF	令你难忘
0010060C0	98	E7	9A	84	E4	B8	80	E5	A4	A9	E6	98	AF	E5	93	AA	的一天是哪
0010060D0	E4	B8	80	E5	A4	A9	E5	91	A2	3F	0D	0A	0D	0A	E9	9A	一天呢? 难
0010060E0	BE	E5	BF	98	E7	9A	84	E4	B8	80	E5	A4	A9	E4	B8	80	忘的一天一
0010060F0	EF	BC	9A	0D	0A	E5	8E	BB	E5	B9	B4	E6	94	BE	E6	9A	: 去年放暑
001006100	91	E5	81	87	E7	9A	84	E6	97	B6	E5	80	99	EF	BC	8C	假的时候,
001006110	E6	88	91	E6	9A	82	E6	97	B6	E4	BD	8F	E5	9C	A8	E5	我暂时住在
001006120	A4	96	E5	85	AC	E5	AE	B6	EF	BC	8C	E9	A9	AC	E4	B8	外公家,马
001006130	8A	E5	B0	B1	E8	A6	81	E4	B8	AD	E5	8D	88	31	32	E7	就要中午12
001006140	82	B9	E4	BA	86	EF	BC	8C	E6	88	91	E5	92	8C	E8	A1	了,我和表
001006150	A8	E5	A7	90	E8	BF	98	E6	9C	89	E8	A1	A8	E5	BC	9F	姐还有表弟
001006160	E6	AD	A3	E4	BB	8E	E5	A4	96	E9	9D	A2	E5	9B	9E	E6	正从外面回
001006170	9D	A5	EF	BC	8C	E8	82	9A	E5	AD	90	E6	AD	A3	E9	A5	,肚子正饿
001006180	BF	E7	9A	84	E2	80	9C	E5	92	95	E5	92	95	E2	80	9D	的“咕咕”
001006190	E5	8F	AB	E7	9A	84	E6	97	B6	E5	80	99	EF	BC	8C	E5	叫的时候,却
0010061A0	8D	B4	E5	8F	91	E7	8E	B0	E6	89	80	E6	9C	89	E7	9A	发现所有的
0010061B0	84	E4	BA	BA	E7	AB	9F	E7	84	B6	E9	83	BD	E7	9D	A1	人竟然都睡
0010061C0	E7	9D	80	E4	BA	86	E3	80	82	E6	88	91	E5	A4	96	E5	着了。我外公
0010061D0	85	AC	E5	AE	B6	E7	9A	84	E4	BA	BA	E9	83	BD	E6	9C	家的人都有
0010061E0	89	E7	9D	A1	E5	8D	88	E8	A7	89	E7	9A	84	E4	B9	A0	睡午觉的习
0010061F0	E6	83	AF	EF	BC	8C	E9	99	A4	E4	BA	86	E8	A1	A8	E5	惯,除了表姐
001006200	A7	90	E5	92	8C	E8	A1	A8	E5	BC	9F	EF	BC	8C	E8	80	和表弟,而

Offset	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F	UTF-8
001007DD0	BD	A0	E4	BB	AC	E3	80	82	E2	80	9D	E8	AF	B4	E5	BF	“说心
001007DE0	83	E9	87	8C	E8	AF	9D	E6	88	91	E4	B9	9F	E9	9D	9E	里话我也非
001007DF0	E5	B8	B8	E6	83	B3	E7	88	B8	E7	88	B8	E3	80	82	0D	常想爸爸。
001007E00	0A	E5	B0	B1	E4	BB	8E	E4	BB	8A	E5	A4	A9	E8	B5	B7	就从今天起
001007E10	EF	BC	8C	E6	88	91	E4	B8	8D	E5	86	8D	E5	92	8C	E7	,我不再和爸
001007E20	88	B8	E7	88	B8	E9	82	A3	E4	B9	88	E9	99	8C	E7	94	爸那么陌生
001007E30	9F	E4	BA	86	EF	BC	8C	E5	9B	A0	E4	B8	BA	E6	88	91	了,因为我
001007E40	E7	9F	A5	E9	81	93	E7	88	B8	E7	88	B8	E4	B9	9F	E5	知道爸爸也喜
001007E50	96	9C	E6	AC	A2	E6	88	91	EF	BC	8C	E8	B7	9F	E6	88	欢我,跟我也
001007E60	91	E9	9D	9E	E5	B8	B8	E4	BA	B2	E5	88	87	E3	80	82	非常亲切。也
001007E70	E6	88	91	E7	9C	9F	E6	83	B3	E4	BA	B2	E5	88	87	E5	我真想亲切地
001007E80	9C	B0	E5	8F	AB	E4	B8	80	E5	A3	B0	EF	BC	9A	E2	80	叫一声:“
001007E90	9C	E7	88	B8	E7	88	B8	EF	BC	8C	E6	82	A8	E7	9C	9F	爸爸,您真
001007EA0	E5	A5	BD	21	E2	80	9D	00	00	00	00	00	00	00	00	00	好!”
001007EB0	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	CCCCCCCC
001007EC0	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	
001007ED0	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	
001007EE0	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	
001007EF0	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	



## 【手工拼接一个新文档】

### (1) 改写根目录扇区

建一个文件，文件名为 CX2.txt，创建日期为 2023.03.28，访问日期为 2.09.8.8，修改日期为 22019.8.8；创建时间为 23:25:10，修改时间为 23:25:20，文件大小为 1000 字节。

偏移 00-07 存放文件名：C 对应 ASCII 值为 43，X 对应 ASCII 值为 58，2 对应 ASCII 值为。

偏移 08-0A 存放后缀：后缀中 TXT 的 ASCII 值为 84, 88, 84. 因此是十六进制是 54, 58, 54。

偏移字节 0B 中存放的是该文件的属性：

0x01——只读；0x02——隐藏；0x04——系统文件；0x08——卷标；0x10——目录；0x20——存档；我们新建文件后，是要在里面存放数据的，因此是存档文件 0x20。

偏移字节 0C 是保留字节，不做修改，保留原来值 00。

偏移字节 0D 中存放创建时间，精确到十分之一秒，这里写 00，意为精确到 0 毫秒。

偏移字节 0E-0F 这两个字节存放文件的创建时间，16bit 中要存放时、分、秒三种信息，其中时占 5bit，分占 6bit，秒占 5bit，其中秒是以 2s 为一个单位的。创建时间为 23:25:10（单位是 5），组成 16bit，即 0xBB25。

偏移字节 10-11 中存放文件创建日期，16bit 中要存放年、月、日三种信息，年占 7bit，月占 4，日占 5bit。创建日期为 2023.3.12，转化为 0x6C56。

偏移字节 12-13 中存放文件的最后访问日期，最后访问时间为 2023.3.29，转化为 0x7064

偏移字节 14-15 中存放的是文件起始簇号的高两个字节，0x0000

偏移字节 16-17 存放文件的最后修改时间，为 22:05:20，转化为 0xADB0

偏移字节 18-19 存放文件最后被修改的日期，为 2023.3.28，转化为 7C56

偏移字节 1A-1B 存放文件起始簇号低两个字节，0x0500

偏移字节 1C-1F 存放文件的字节大小，占 4 个字节，我们设置为 16KB，转化为 0x00003E2D

43 58 20 20 20 20 20 20	54 58 54 20 18 9B CF 5A	CX	TXT	00
6C 56 7C 56 00 00 AD B0	7C 56 05 00 2D 3E 00 00	V V	00	00->00
43 58 32 20 20 20 20 20	54 58 54 20 00 00 25 BB	CX2	TXT	00%0
6C 56 70 64 00 00 AD B0	7C 56 05 00 2D 3E 00 00	1Vpd	00	00->00

### (2) 改写文件内容区

复制源文件内容偏移地址处的内容，拼接成功一个与原文件内容完全相同的文件 CX2.txt

Offset	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F	UTF-8
001006000	E6	AF	8F	E4	B8	AA	E4	BA	BA	E9	83	BD	E6	9C	89	E4	每个人都有一
001006010	B8	80	E4	BB	B6	E9	9A	BE	E5	BF	98	E7	9A	84	E4	BA	件难忘的事
001006020	8B	E6	83	85	2C	E8	80	8C	E9	9A	BE	E5	BF	98	E7	9A	情，而难忘的
001006030	84	E4	BA	8B	E6	83	85	E4	B9	9F	E5	B0	B1	E5	83	8F	事情也就像
001006040	E6	B5	B7	E8	BE	B9	E9	82	A3	E5	90	84	E8	89	B2	E5	海边那各式各
001006050	90	84	E6	A0	B7	E7	9A	84	E8	B4	9D	E5	A3	B3	2C	E5	样的贝壳，在
001006060	9C	A8	E6	B5	B7	E6	B0	B4	E7	9A	84	E9	99	AA	E4	BC	海水的陪伴
001006070	B4	E4	B8	8B	E9	97	AA	E9	97	AA	E5	8F	91	E5	85	89	下闪闪发光
001006080	2C	E6	95	A3	E5	8F	91	E7	9D	80	E4	BA	94	E9	A2	9C	，散发着五颜
001006090	E5	85	AD	E8	89	B2	E7	9A	84	E5	85	89	E5	BD	A9	2C	六色的光彩，
0010060A0	E6	95	B0	E9	83	BD	E6	95	B0	E4	B8	8D	E5	AE	8C	E3	数都数不完。
0010060B0	80	82	E4	BB	A4	E4	BD	A0	E6	9C	80	E9	9A	BE	E5	BF	令你最难忘
0010060C0	98	E7	9A	84	E4	B8	80	E5	A4	A9	E6	98	AF	E5	93	AA	的一天是哪
0010060D0	E4	B8	80	E5	A4	A9	E5	91	A2	3F	0D	0A	0D	0A	E9	9A	一天呢？难
0010060E0	BE	E5	BF	98	E7	9A	84	E4	B8	80	E5	A4	A9	E4	B8	80	忘的一天一
0010060F0	EF	BC	9A	0D	0A	E5	8E	BB	E5	B9	B4	E6	94	BE	E6	9A	：去年放暑
001006100	91	E5	81	87	E7	9A	84	E6	97	B6	E5	80	99	EF	BC	8C	假的时候，
001006110	E6	88	91	E6	9A	82	E6	97	B6	E4	BD	8F	E5	9C	A8	E5	我暂时住在
001006120	A4	96	E5	85	AC	E5	AE	B6	EF	BC	8C	E9	A9	AC	E4	B8	公家，马上
001006130	8A	E5	B0	B1	E8	A6	81	E4	B8	AD	E5	8D	88	31	32	E7	就要中午 12点

## 2. 编程实现“内容一”的全过程

### 【工具导入】

导入 binascii 工具包，可以用来在二进制和 ASCII 码中转换。

```
1. import binascii
```

### 【函数声明】

声明一个 hexa\_to\_dec 函数，该函数可以把十六进制字符串，转化为十进制整数输出。该函数将每两个十六进制数字提取，并转化为十进制整型，再通过叠加每两个十六进制数字得到最终结果（每提取完一对十六进制数，下一对的真实值就要乘两个 16，也就是 256）。res 是要作为结果输出的整型变量，c 是需要乘的倍数（每次循环都要增长为原来的 256），a 是被转化为整型的高位十六进制数，b 是被转化为整型的低十六进制数。关于如何将十六进制字符转换为十进制整型，则是利用了 ASCII 码。当检测到该十六进制字符是 1-9 时，只将其变为 int 类型，不做其它改变，储存在 a 或 b 变量中；当检测到十六进制字符时 a-f 时，将其 ASCII 码减 48，结果 a 字符转化为 1，b 字符转化为 2……再加上数字 9，并将其变为 int 类型，则为十六进制字符对应的十进制数。

十六进制	十六进制 ASCII 码	数字字符	数字字符 ASCII 码	ASCII 码差值	int (十六进制 ASC-差值)+9
a	97	1	49	48	10
b	98	2	50	48	11
c	99	3	51	48	12
d	100	4	52	48	13
e	101	5	53	48	14
f	102	6	54	48	15

```
1. def hexa_to_dec(s: str) -> int:
2.     res = 0
3.     c = 1
4.     a = 0
5.     b = 0
6.     for i in range(0, len(s), 2):
7.         if ord(s[i]) > 60:
8.             a = 9 + 1 * int(chr(int(ord(s[i])) - 48))
9.         if ord(s[i]) < 60:
10.            a = int(s[i])
11.        if ord(s[i + 1]) < 60:
12.            b = int(s[i + 1])
```

```

13.     if ord(s[i + 1]) > 60:
14.         b = 9 + 1 * int(chr(int(ord(s[i + 1])) - 48))
15.         res += c * (a * 16 + b)
16.         c *= 256
17.     return res

```

声明一个计算簇链的函数 cal。

思路是得到首簇后，将首簇和后面的簇放到簇链列表中。后面簇的计算方式为上一个簇的结尾到下一个簇的结尾。如果到结尾则跳出循环，并返回簇链列表。

```

1.  def cal(fat, fir_clust):
2.      res = []
3.      while True:
4.          res.append(fir_clust)
5.          fir_clust = hexa_to_deci(fat[fir_clust * 8: (fir_clust + 1) * 8])
6.          if fir_clust == int('0xffffffff', 16):
7.              break
8.      return res

```

## 【读取文件】

输入文件路径，用“/”相隔，并将文件路径切分，将文件夹名和文件名都转化为对应的 16 进制字符串，文件名不考虑后缀。

```

1.  if __name__ == '__main__':
2.      file_get_path = input("请输入文件路径:").split("/")
3.      file_hexadecimal = [str(binascii.hexlify(bytes(file_get_path[i], 'utf-8'))
4.                             [2:-1]) for i in range(len(file_get_path))]
5.      file_get_path[-1] = file_get_path[-1].split(".")[0]

```

打开文件系统，读取 DBR 获得数据，如根目录的扇区号，FAT1 的扇区号，每个扇区的字节数，根目录的簇号和每簇的扇区数。再将根目录簇号转化为列表，以便后续分析。最开始的簇链就是根目录。

```

1.  file_open = open(r'\\.\.' + file_get_path[0], 'rb')
2.  dbr = str(binascii.hexlify(file_open.read(512)))[2:-1]
3.  byte_per_sector = hexa_to_deci(dbr[22: 26])
4.  sector_per_clust = hexa_to_deci(dbr[26: 28])
5.  reserve_sector = hexa_to_deci(dbr[28: 32])
6.  fat_num = hexa_to_deci(dbr[32: 34])
7.  count_sector = hexa_to_deci(dbr[64: 72])
8.  sector_per_fat = hexa_to_deci(dbr[72: 80])
9.  rootclust = hexa_to_deci(dbr[88: 96])
10. fir_sector = reserve_sector + fat_num * sector_per_fat
11. clust = [rootclust]

```

## 【输出文件名目录项信息】

计算目录项信息。

对应每个文件夹名，通过簇链定位它的上一级目录的内容从而得到它的首簇，再得到簇链。扇区号要乘每个扇区的字节数才是偏移量。

文件名查找。如果没找到这个文件名则报错，提示输入正确的文件地址。

```
1. print("\n 该文件名目录项信息:")
2. for i in range(1, len(file_get_path)):
3.     name = str(binascii.hexlify(bytes(file_get_path[i].upper(), 'utf-8')))[2:
-1]
4.     file_open.seek((fir_sector + (clust[0] - rootclust) * sector_per_clust) *
byte_per_sector)
5.     print("数据起始扇区
号:", (fir_sector + (clust[0] - rootclust) * sector_per_clust))
6.     s_sector = str(binascii.hexlify(file_open.read(sector_per_clust * byte_p
er_sector)))[2:-1]
7.     print("当前扇区:", s_sector[0:20], '... (此处省略后段)')
8.     print("文件名:", name)
9.     file_str = -1
10.    file_str = s_sector.find(name)
11.    print("文件名对应的字符串:", file_str)
12.    if file_str == -1:
13.        print("Not Found. Check your path. ")
14.        exit(-1)
```

## 【输出文件簇链】

计算首簇并查找簇链。簇链通过上面的函数计算得出并打印。

```
1. fir_clust = hexa_to_deci(s_sector[file_str + 40:file_str + 44]) * pow(16, 4) + h
exa_to_deci(s_sector[file_str + 52:file_str + 56])
2. print("\n 该文件的簇链:")
3. print("首簇:", fir_clust)
4. file_open.seek(reserve_sector * byte_per_sector)
5. fat = str(binascii.hexlify(file_open.read(sector_per_clust * byte_per_se
ctor)))[2:-1]
6. clust = cal(fat, fir_clust)
7. print("簇链:", clust)
```

## 【拼接新文件】

得到最后的文件簇链，全部读取，然后写入到文件之中。

新建 cx2.txt 文件，并根据簇链读取原始文件内容，写入 cx2.txt 文件。

拼接成功之后会打印 Success 字符。

```
1.  print("\n 拼接新文件:")
2.  with open("cx2.txt", mode="w+", encoding="utf-8", newline="") as f:
3.      for i in clust:
4.          file_open.seek((fir_sector + (i - rootclust) * sector_per_clust) * byte_per_sector)
5.          content = str(binascii.hexlify(file_open.read(sector_per_clust * byte_per_sector)))[2:-1]
6.          v = binascii.unhexlify(content.encode()).decode().rstrip('\0')
7.          f.write(v)
8.  print("Success.")
```

## 【比较新旧文件内容】

判断生成的 cx2.txt 文件和原始文件是否完全相同。将原始文件路径作为字符串形式打开按行读取内容，并将内容存放在 pre\_content 中；再将生成文件按行读取并存放再 copy\_content 中。将两者相比较，一致则打印“文件内容一致”，不一致则打印“文件内容不一致”。

```
1.  print("\n 判断生成文件与原文件内容是否一致:")
2.  file_get_path_str = '/'.join(file_get_path) + '.txt'
3.  with open(file_get_path_str, 'r', encoding='utf-8') as g:
4.      pre_content = g.readlines()
5.  with open('cx2.txt', 'r', encoding='utf-8') as f:
6.      copy_content = f.readlines()
7.  if pre_content == copy_content:
8.      print("文件内容一致")
9.  else:
10.     print("文件内容不一致")
```



# 四. 实验结果

## 1. 实验任务一

原文件内容如下：





## 手工创建文件内容如下：



经过对比，内容一致。

## 2. 实验任务二

代码运行结果：

运行成功，得到文件名目录项信息，簇链，拼接新文件成功，并判断文件内容一致。

为了使簇链更长一些，在任务一的基础上加长了文件的内容长度。

```
F:\pythonProject1\python3.77\Scripts\python.exe F:/pythonProject1/venv/319bd/pcbd1main.py
请输入文件路径: G:/ex.txt

该文件名目录项信息:
数据起始扇区号: 32768
当前扇区: 4b494e4753544f4e2020 ... (此处省略后段)
文件名: 4358
文件名对应的字符串: 256

该文件的簇链:
首簇: 5
簇链: [5, 6]

拼接新文件:
Success.

判断生成文件与原文件内容是否一致:
文件内容一致

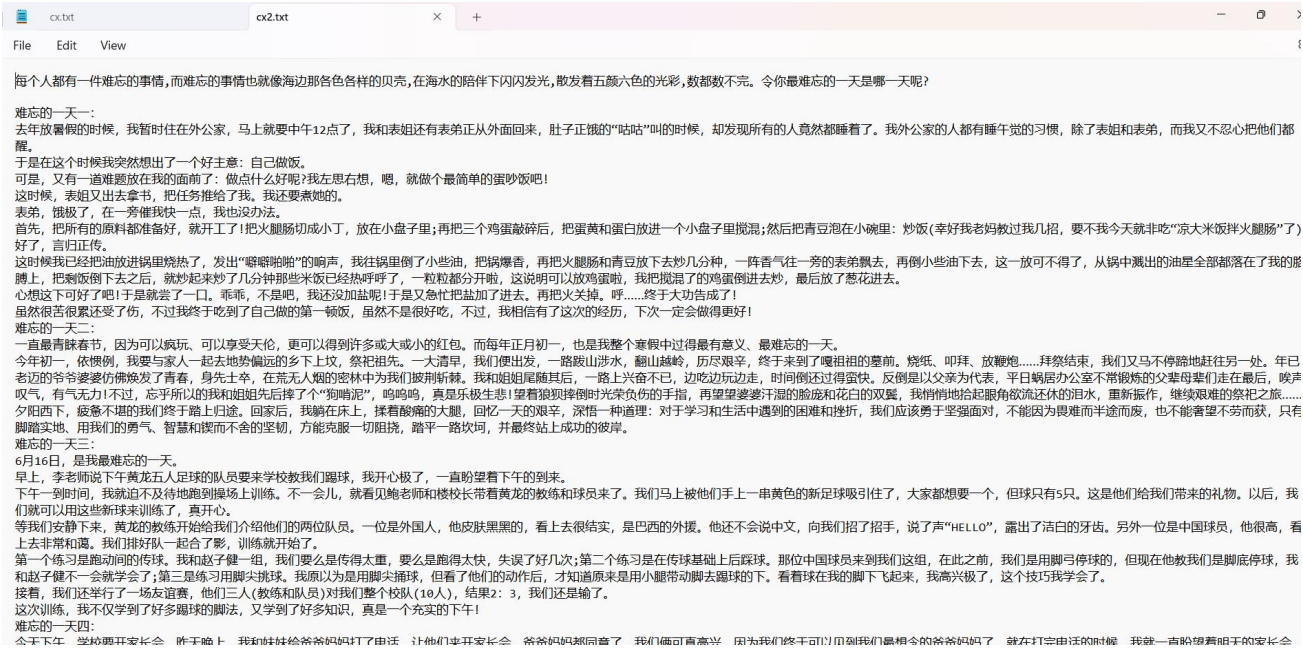
Process finished with exit code 0
```

原文件内容预览：





## 新生成文件预览：



## 五．实验心得（个人贡献）

陈曦：

个人贡献：在此次实验中我负责的是内容二，撰写代码部分。以及编写内容二的实验报告，并安排报告的排版，代码说明 Readme 的撰写。

实验心得：我学习了 FAT32 文件系统是如何安排文件数据的，以及扇区，簇等概念，学习一个文件在 FAT32 文件系统中是如何分布的。在此基础上，我撰写了输出文件信息以及拼接相同文件的代码，对 FAT32 文件系统分区以及文件系统格式有了基础的认识。

梁刘琪：

个人贡献：在此次实验中负责内容一，手工定位和提取 FAT32 的文件内容，并手工创建与原文件内容一致的文件；帮助组员调试 debug 内容二的代码，撰写内容一的实验报告。

实验心得：学会了手工在 Winhex 中提取，定位，创建文件，了解熟悉了 FAT32 文件系统分区以及文件系统格式了解了引导扇区，根目录，FAT1 表的字节位移的含义，更好的理解 FAT32 文件系统的工作原理。