ИТМО

Лекция №3.
Форензика. Форматы файлов

Форматы файлов

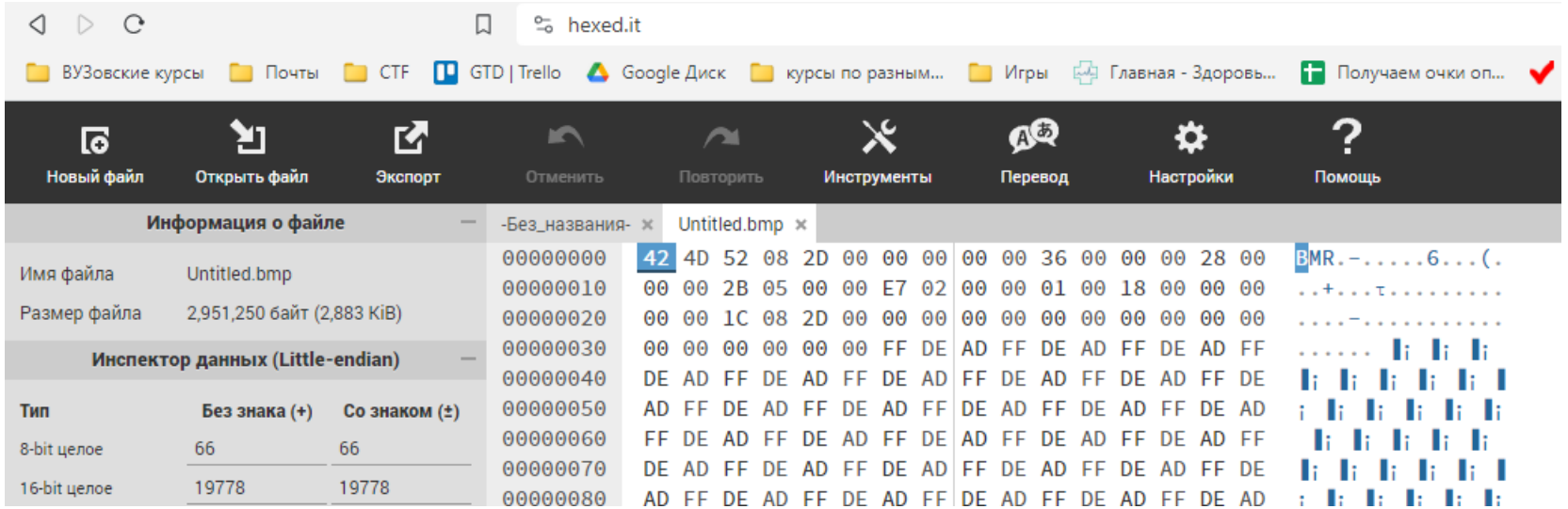
- Archive files (ZIP, TGZ)
- Image file formats (JPG, GIF, BMP, PNG)
- Filesystem images (especially EXT4)
- Packet captures (PCAP, PCAPNG)
- Memory dumps
- PDF
- Video (especially MP4) or Audio (especially WAV, MP3)
- Microsoft's Office formats (RTF, OLE, OOXML)

Как определить типа файла?

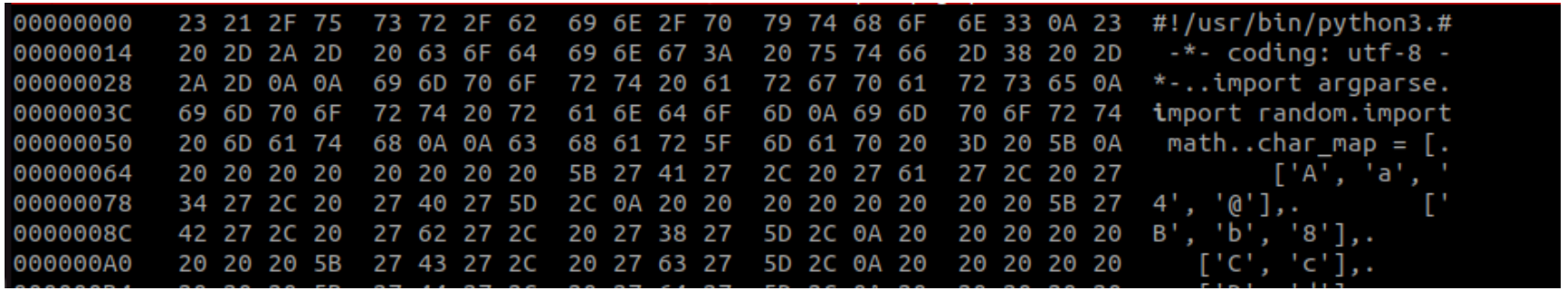
- По расширению имени (н-р: «.png»)
- По магическому числу – особой последовательности байт, характерная для конкретных форматов
- По шебангам – строке, содержащей то, каким интерпретатором должен обрабатываться код

Спецификация – подробное описание структуры файла конкретного формата.

Нех-редактор.

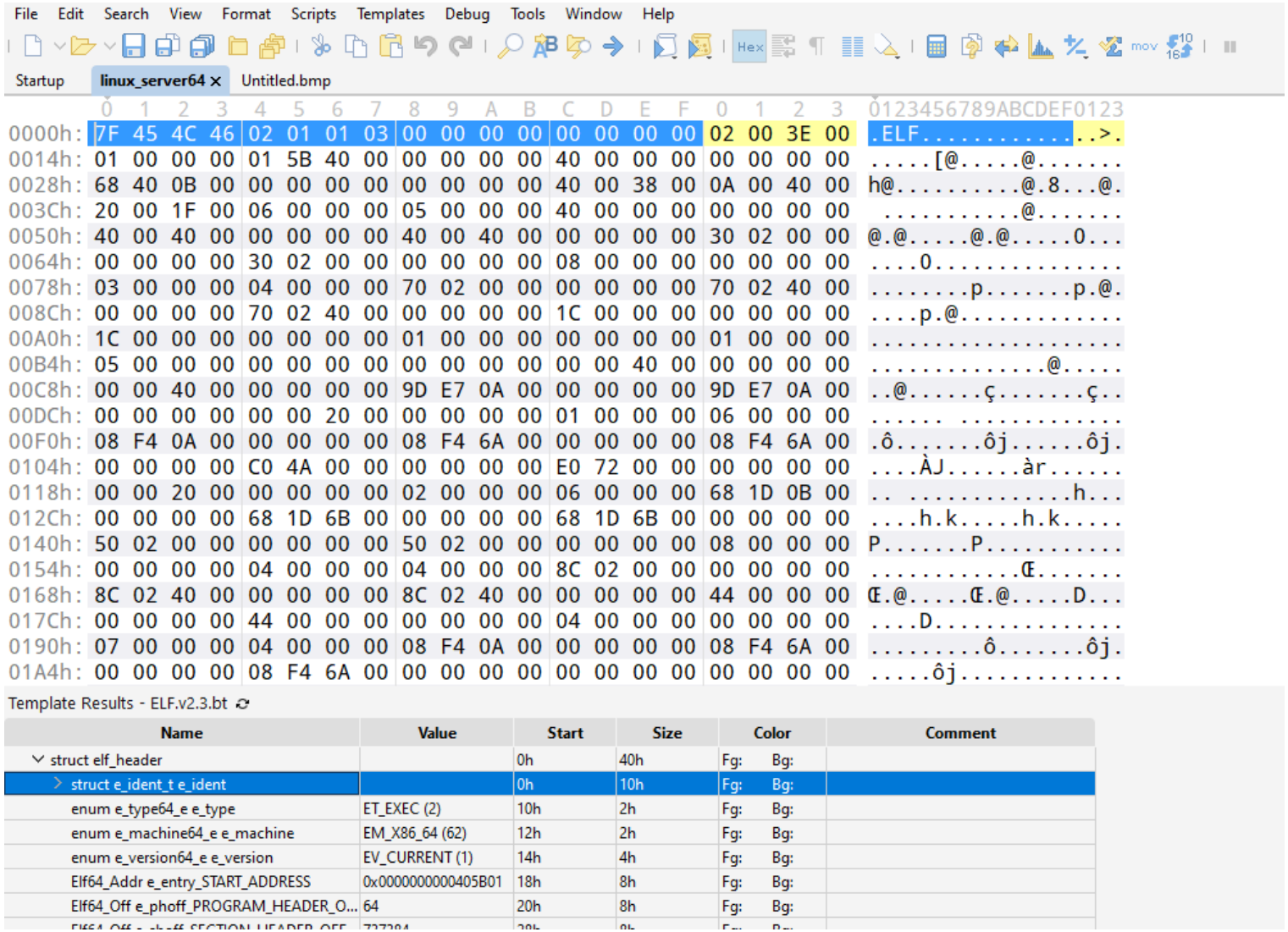


- hexed.it



-hexedit

010 editor - база.



- 010 editor

Hex-редактор.

- Очевидно

```
89 50 4E 47 0D 0A 1A 0A 00 00 00 0D 49 48 44 52 00 00 02 72  %PNG........IHDR...r
00 00 03 42 08 02 00 00 00 6D 41 B6 DB 00 00 25 F9 69 54 58  ...B.....mA¶Û..%ùiTX
```

```
25 50 44 46 2D 31 2E 35 0A 25 E4 F0 ED F8 0A 37 20 30 20 6F  %PDF-1.5.%äðíø.7 0 o
62 6A 0A 3C 3C 2F 46 69 6C 74 65 72 2F 46 6C 61 74 65 44 65  bj.<</Filter/FlateDe
```

```
7F 45 4C 46 02 01 01 03 00 00 00 00 00 00 00 00 02 00 3E 00  .ELF..............>.
01 00 00 00 01 5B 40 00 00 00 00 00 40 00 00 00 00 00 00 00  .....[@.....@.......
```

Hex-редактор.

- Очевидно

```
89 50 4E 47 0D 0A 1A 0A 00 00 00 0D 49 48 44 52 00 00 02 72  %PNG........IHDR...r
00 00 03 42 08 02 00 00 00 6D 41 B6 DB 00 00 25 F9 69 54 58  ...B.....mA¶Û..%ùiTX
```

- .png file

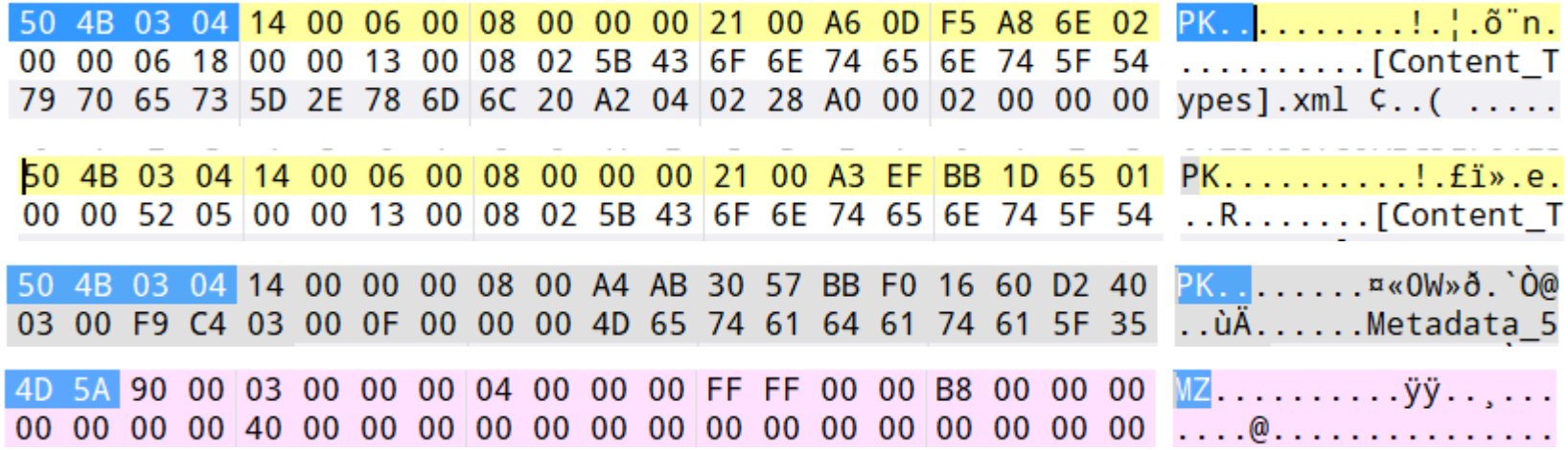
```
25 50 44 46 2D 31 2E 35 0A 25 E4 F0 ED F8 0A 37 20 30 20 6F  %PDF-1.5.%äðíø.7 0 o
62 6A 0A 3C 3C 2F 46 69 6C 74 65 72 2F 46 6C 61 74 65 44 65  bj.<</Filter/FlateDe
```

- .pdf file

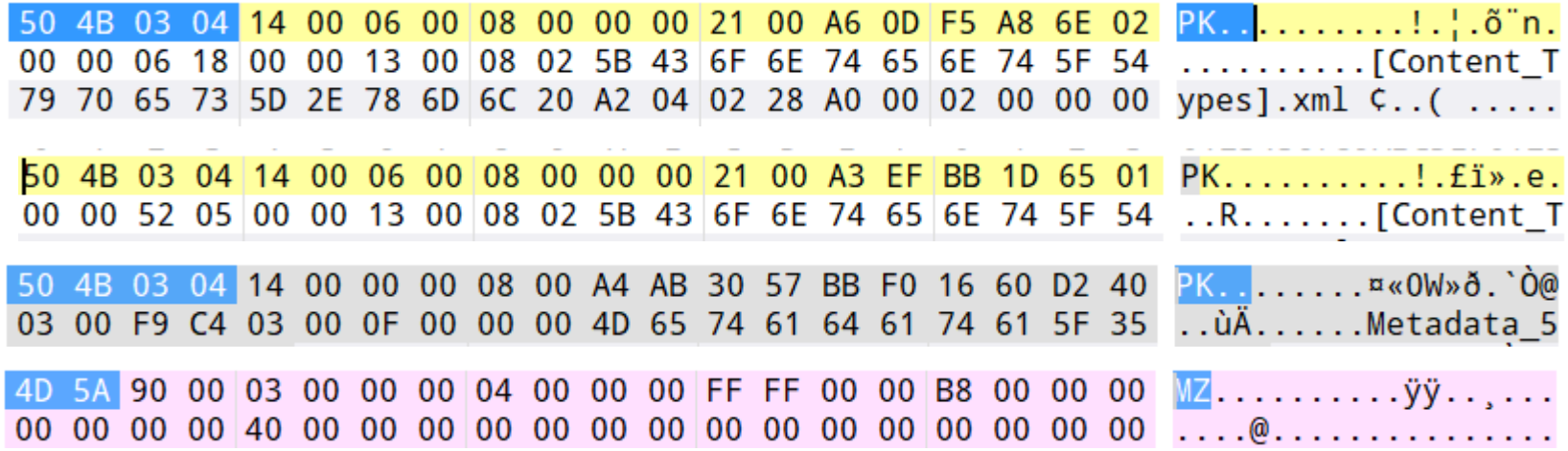
```
7F 45 4C 46 02 01 01 03 00 00 00 00 00 00 00 00 02 00 3E 00  .ELF..............>.
01 00 00 00 01 5B 40 00 00 00 00 00 40 00 00 00 00 00 00 00  .....[@.....@.......
```

- .elf file

- Чуть сложнее



- Чуть сложнее



- .pptx file

- .docx file

- .zip file

- .exe file

- Чуть сложнее

```
50 4B 03 04 14 00 06 00 08 00 00 00 21 00 A6 0D F5 A8 6E 02  PK..........!.¦.õ¨n.
00 00 06 18 00 00 13 00 08 02 5B 43 6F 6E 74 65 6E 74 5F 54  ..........[Content_T
79 70 65 73 5D 2E 78 6D 6C 20 A2 04 02 28 A0 00 02 00 00 00  ypes].xml ¢..( .....
```

- .pptx file

```
50 4B 03 04 14 00 06 00 08 00 00 00 21 00 A3 EF BB 1D 65 01  PK..........!.£ï».e.
00 00 52 05 00 00 13 00 08 02 5B 43 6F 6E 74 65 6E 74 5F 54  ..R.......[Content_T
```

- .docx file

```
50 4B 03 04 14 00 00 00 08 00 A4 AB 30 57 BB F0 16 60 D2 40  PK........¤«0W»ð.`Ò@
03 00 F9 C4 03 00 0F 00 00 00 4D 65 74 61 64 61 74 61 5F 35  ..ùÄ......Metadata_5
```

- .zip file

```
4D 5A 90 00 03 00 00 00 04 00 00 00 FF FF 00 00 B8 00 00 00  MZ..........ÿÿ..¸...
00 00 00 00 40 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00  ....@...............
```

- .exe file

- Что это вообще?

```
00 00 00 20 66 74 79 70 71 74 20 20 20 05 03 00 71 74 20 20  ... ftypqt    ...qt
00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 08 77 69 64 65  ................wide
```

- .mov audio file

Strings – выводим печатаемые символы

Минимальная длина строки

Сама утилита

Файл, который мы просматриваем

```
amogus@AMOGUS:/mnt/h/Универ/CTFClub/лекция первокурсникам$ strings -n 2 zip_for_strings.zip
PK
VGW
/m
qwertyuiopasdfghjkl.txt+NM.J-
0.2
0LqI
w3
PK
VGW
/m
qwertyuiopasdfghjkl.txt
PK
```

```
PK.......ðVGW‡/mê..
..........qwertyuiop
asdfghjkl.txt+NM.J-©
ö0.2ŽÏ,ŽÏ0LqIÍ‹w3ÌI-
..PK.........ðVGW‡/
mê.........$.......
 .......qwertyuiopas
dfghjkl.txt.. ......
...GÀ×¥óøÙ.GÀ×¥óøÙ.d
„°ŠóøÙ.PK.........i
...R.....
```

Binwalk - идентификации типов файлов

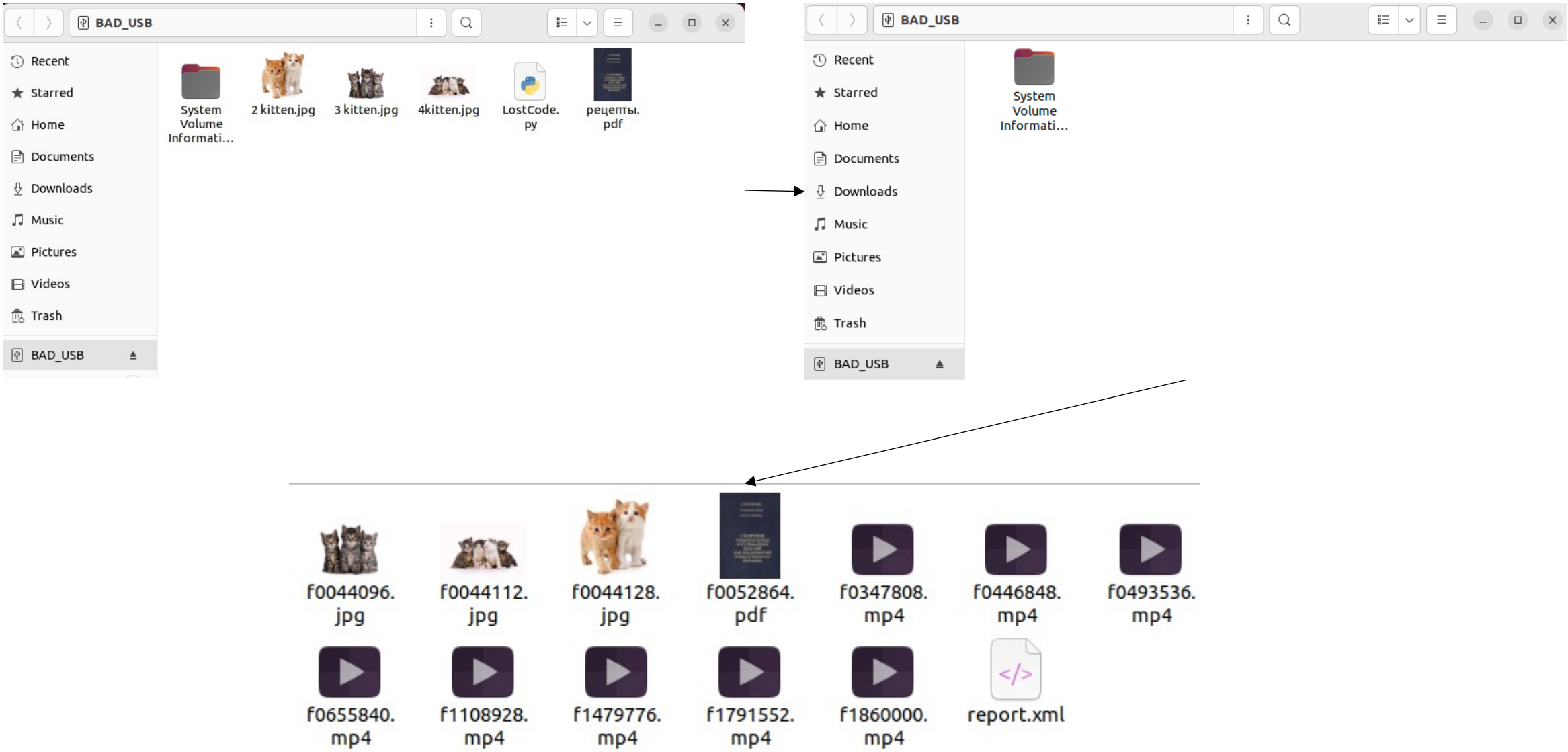
Посмотрим, какие сигнатуры найдутся в exe файле

```
amogus@AMOGUS:/mnt/c/Users/Mefistofele/Desktop/Tasks/forensic_for_itmo$ binwalk executable.3720.exe

DECIMAL       HEXADECIMAL     DESCRIPTION
--------------------------------------------------------------------------------
0             0x0             Microsoft executable, portable (PE)
9178          0x23DA          Copyright string: "CopyrightAttribute"
116288        0x1C640         PNG image, 4800 x 1454, 8-bit/color RGBA, non-interlaced
116416        0x1C6C0         Zlib compressed data, compressed
344098        0x54022         PNG image, 800 x 600, 8-bit colormap, non-interlaced
344511        0x541BF         Zlib compressed data, best compression
420575        0x66ADF         XML document, version: "1.0"
```

Извлечём все сигнатуры, а изображения сохраним как png файлы

```
amogus@AMOGUS:/mnt/c/Users/Mefistofele/Desktop/Tasks/forensic_for_itmo$ binwalk --dd="image:png" executable.3720.exe

DECIMAL       HEXADECIMAL     DESCRIPTION
--------------------------------------------------------------------------------
0             0x0             Microsoft executable, portable (PE)
9178          0x23DA          Copyright string: "CopyrightAttribute"
116288        0x1C640         PNG image, 4800 x 1454, 8-bit/color RGBA, non-interlaced
116416        0x1C6C0         Zlib compressed data, compressed
344098        0x54022         PNG image, 800 x 600, 8-bit colormap, non-interlaced
344511        0x541BF         Zlib compressed data, best compression
420575        0x66ADF         XML document, version: "1.0"
```

Photorec – программа для восстановления данных



</>

Демонстрация

BMP или как оно есть

```
42 4D 52 08 2D 00 00 00 00 00 36 00 00 00 28 00 00 00 2B 05  BMR.-.....6...(...+.
00 00 E7 02 00 00 01 00 18 00 00 00 00 00 1C 08 2D 00 00 00  ..ç.............-...
00 00 00 00 00 00 00 00 00 00 00 00 00 00 FF DE AD FF DE AD  ..............ÿÞ-ÿÞ-
FF DE AD FF DE AD FF DE AD FF DE AD FF DE AD FF DE AD FF DE  ÿÞ-ÿÞ-ÿÞ-ÿÞ-ÿÞ-ÿÞ-ÿÞ
AD FF DE AD FF DE AD FF DE AD FF DE AD FF DE AD FF DE AD FF  -ÿÞ-ÿÞ-ÿÞ-ÿÞ-ÿÞ-ÿÞ-ÿ
DE AD FF DE AD FF DE AD FF DE AD FF DE AD FF DE AD FF DE AD  Þ-ÿÞ-ÿÞ-ÿÞ-ÿÞ-ÿÞ-ÿÞ-
FF DE AD FF DE AD FF DE AD FF DE AD FF DE AD FF DE AD FF DE  ÿÞ-ÿÞ-ÿÞ-ÿÞ-ÿÞ-ÿÞ-ÿÞ
AD FF DE AD FF DE AD FF DE AD FF DE AD FF DE AD FF DE AD FF  -ÿÞ-ÿÞ-ÿÞ-ÿÞ-ÿÞ-ÿÞ-ÿ
```

В hex-редакторе

Картинка

Как мы её видим

JPG

```
A2 8A 28 00 A2 8A 28 00 A2 8A 28 00 A2 8A 28 00 A2 8A 28 00  ¢Š(.¢Š(.¢Š(.¢Š(.¢Š(.
A2 8A 28 00 A2 8A 28 00 A2 8A 28 00 A2 8A 28 00 A2 8A 28 00  ¢Š(.¢Š(.¢Š(.¢Š(.¢Š(.
A2 8A 28 00 A2 8A 28 00 A2 8A 28 00 A2 8A 28 00 A2 8A 28 00  ¢Š(.¢Š(.¢Š(.¢Š(.¢Š(.
A2 8A 28 00 A2 8A 28 00 A2 8A 28 00 A2 8A 28 00 A2 8A 28 00  ¢Š(.¢Š(.¢Š(.¢Š(.¢Š(.
A2 8A 28 00 A2 8A 28 00 A2 8A 28 00 A2 8A 28 00 A2 8A 28 00  ¢Š(.¢Š(.¢Š(.¢Š(.¢Š(.
A2 8A 28 00 A2 8A 28 00 A2 8A 28 00 A2 8A 28 00 A2 8A 28 00  ¢Š(.¢Š(.¢Š(.¢Š(.¢Š(.
A2 8A 28 00 A2 8A 28 00 A2 8A 28 00 A2 8A 28 00 A2 8A 28 00  ¢Š(.¢Š(.¢Š(.¢Š(.¢Š(.
A2 8A 28 00 A2 8A 28 00 A2 8A 28 00 A2 8A 28 00 A2 8A 28 00  ¢Š(.¢Š(.¢Š(.¢Š(.¢Š(.
```

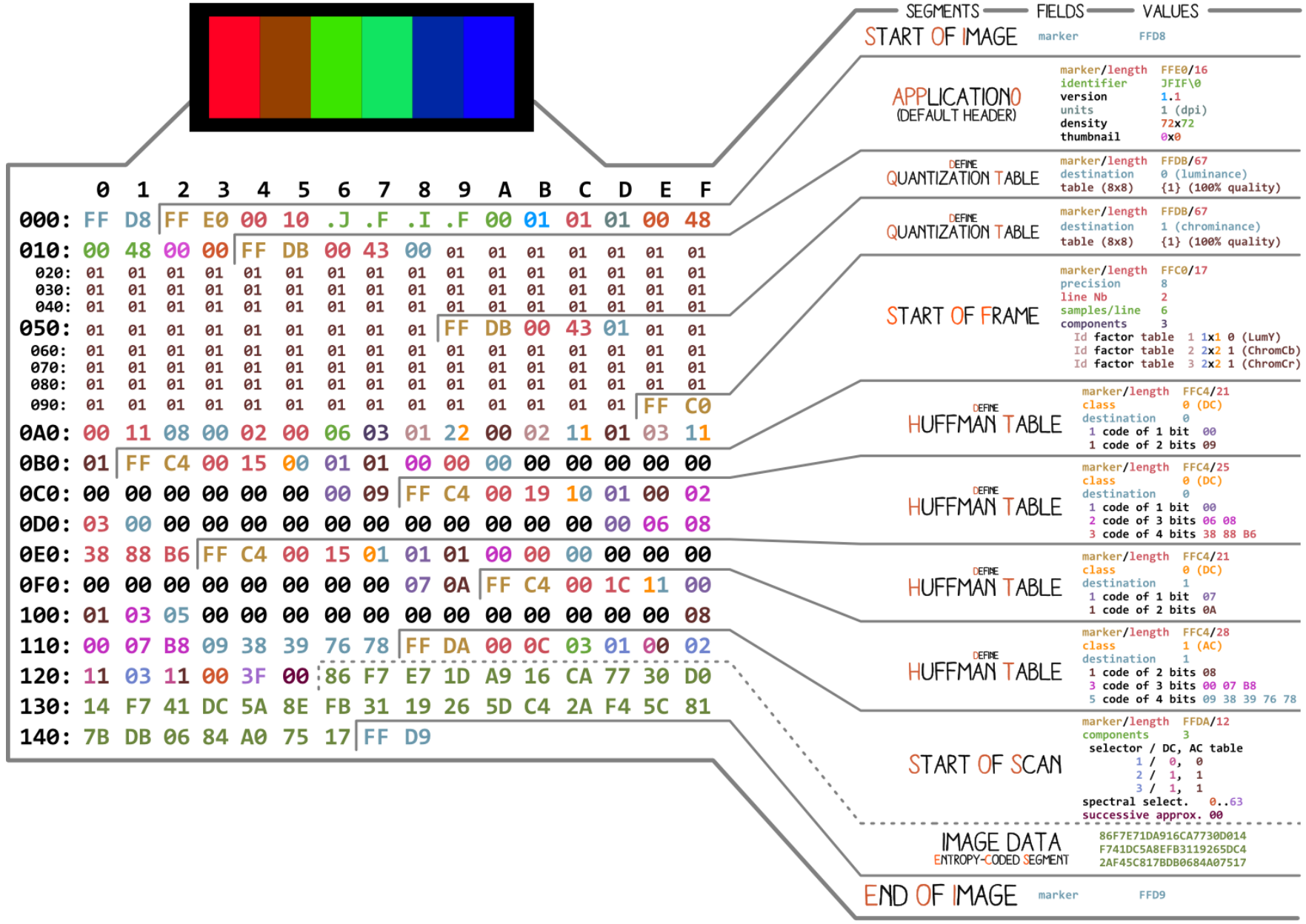
В hex-редакторе

Картинка

Как мы её видим

Лучше этой статьи о jpg я ещё не видел: https://parametric.press/issue-01/unraveling-the-jpeg/

JPG

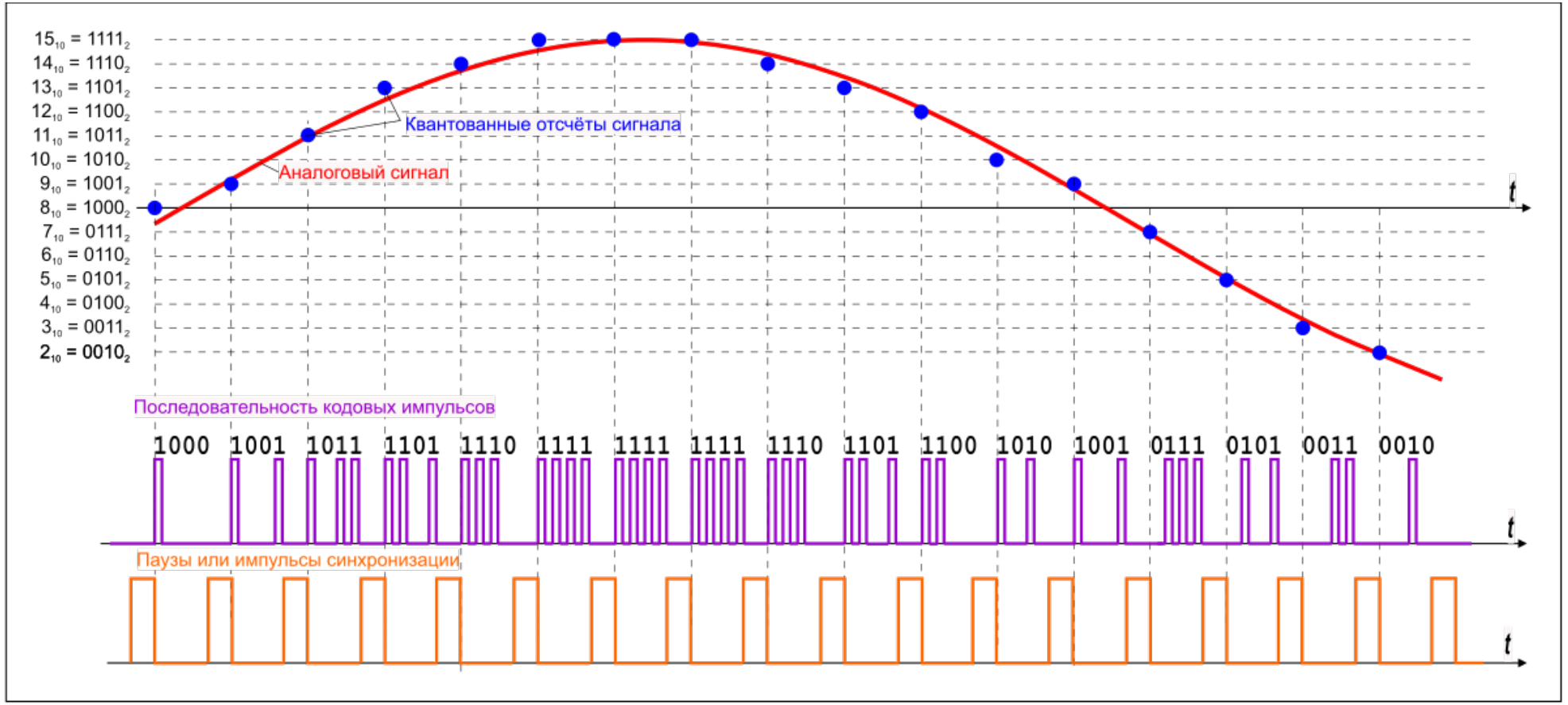


	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
000:	FF	D8	FF	E0	00	10	.J	.F	.I	.F	00	01	01	01	00	48
010:	00	48	00	00	FF	DB	00	43	00	01	01	01	01	01	01	01
020:	01	01	01	01	01	01	01	01	01	01	01	01	01	01	01	01
030:	01	01	01	01	01	01	01	01	01	01	01	01	01	01	01	01
040:	01	01	01	01	01	01	01	01	01	01	01	01	01	01	01	01
050:	01	01	01	01	01	01	01	01	01	FF	DB	00	43	01	01	01
060:	01	01	01	01	01	01	01	01	01	01	01	01	01	01	01	01
070:	01	01	01	01	01	01	01	01	01	01	01	01	01	01	01	01
080:	01	01	01	01	01	01	01	01	01	01	01	01	01	01	01	01
090:	01	01	01	01	01	01	01	01	01	01	01	01	01	01	FF	C0
0A0:	00	11	08	00	02	00	06	03	01	22	00	02	11	01	03	11
0B0:	01	FF	C4	00	15	00	01	01	00	00	00	00	00	00	00	00
0C0:	00	00	00	00	00	00	00	09	FF	C4	00	19	10	01	00	02
0D0:	03	00	00	00	00	00	00	00	00	00	00	00	00	00	06	08
0E0:	38	88	B6	FF	C4	00	15	01	01	01	00	00	00	00	00	00
0F0:	00	00	00	00	00	00	00	00	07	0A	FF	C4	00	1C	11	00
100:	01	03	05	00	00	00	00	00	00	00	00	00	00	00	00	08
110:	00	07	B8	09	38	39	76	78	FF	DA	00	0C	03	01	00	02
120:	11	03	11	00	3F	00	86	F7	E7	1D	A9	16	CA	77	30	D0
130:	14	F7	41	DC	5A	8E	FB	31	19	26	5D	C4	2A	F4	5C	81
140:	7B	DB	06	84	A0	75	17	FF	D9							

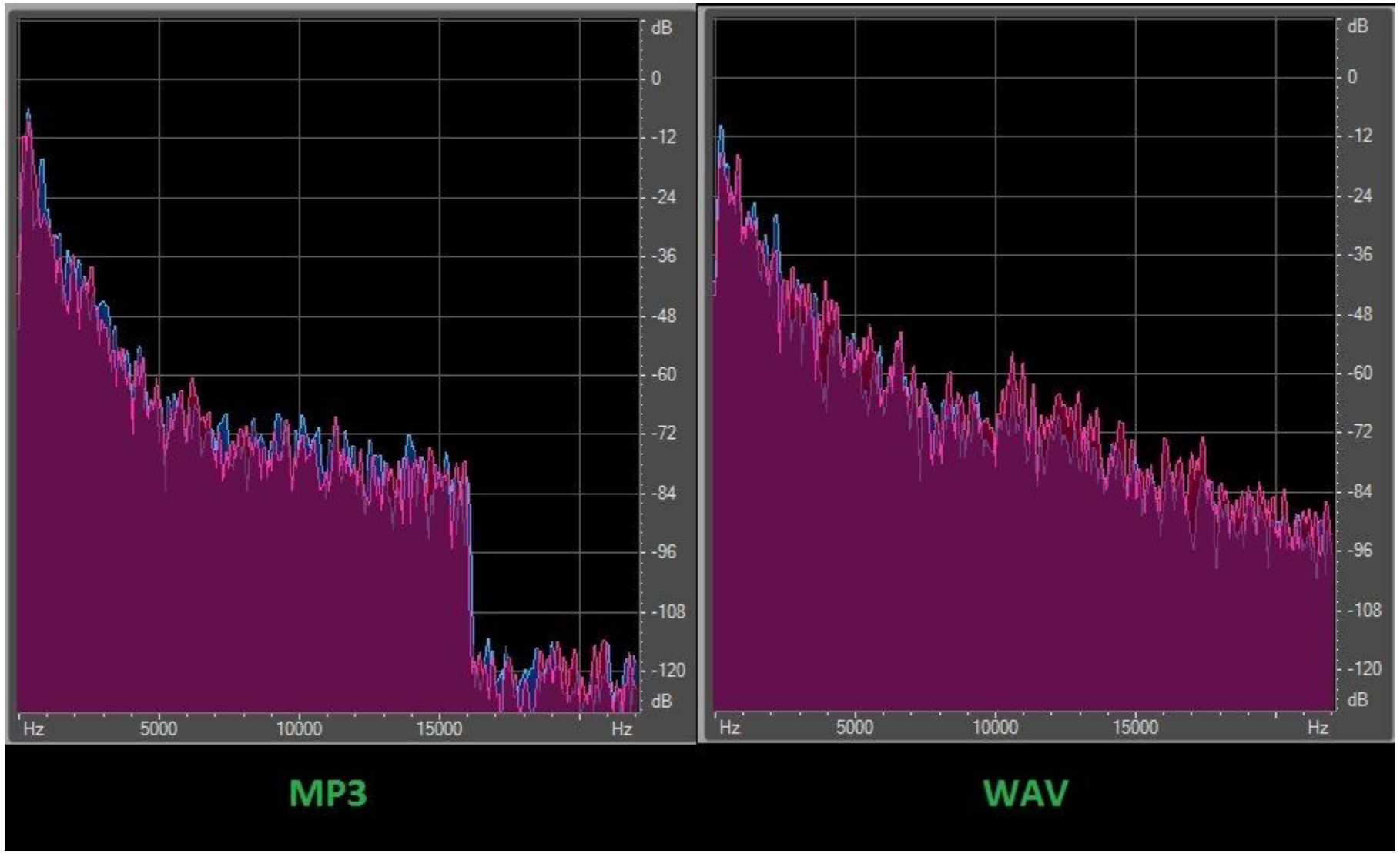
Waveform Audio File Format (WAV)

- **Без сжатия нет потерь**

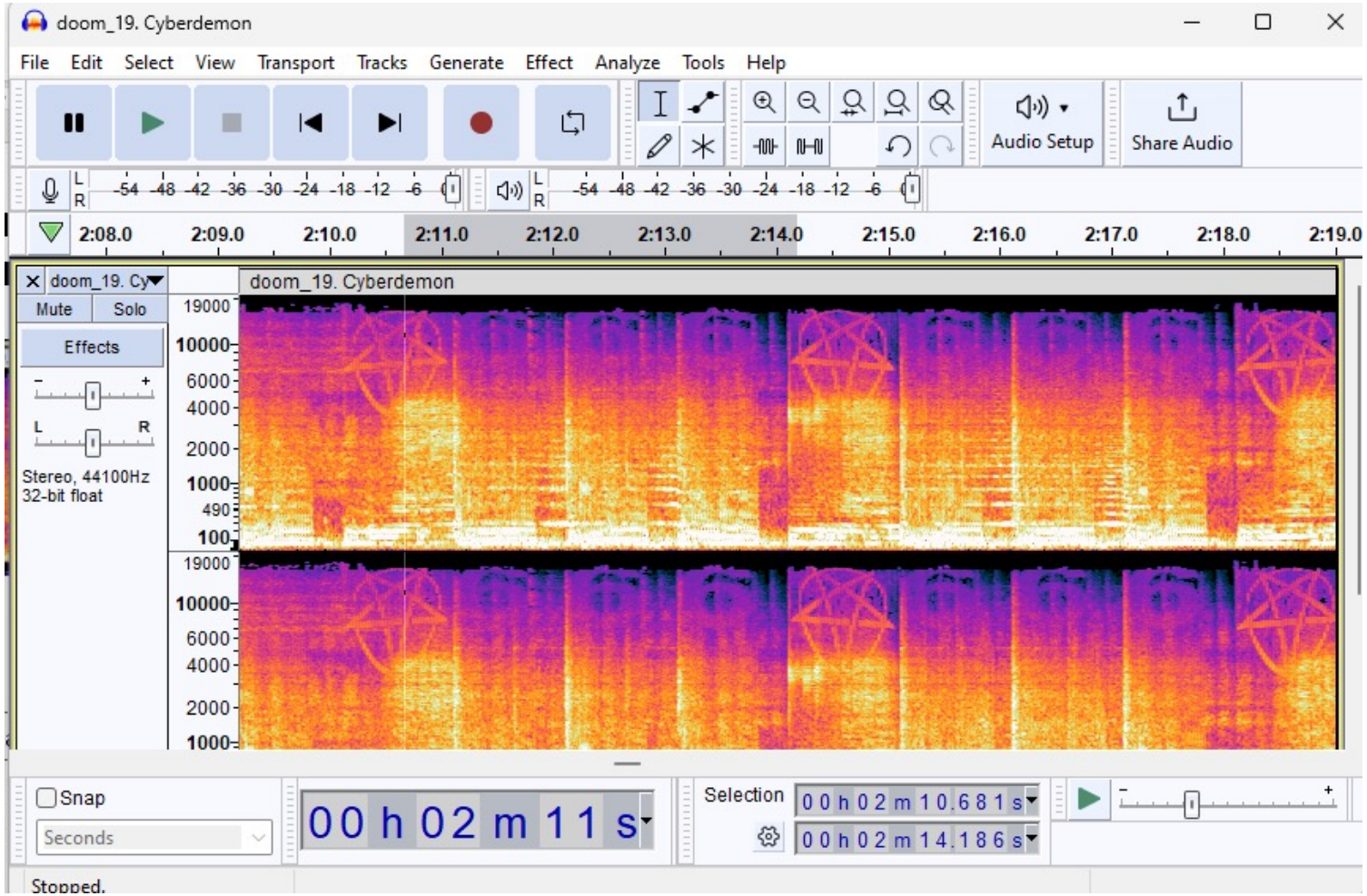
Как оно работает?
1. поток разбивается на малейшие отрезки
2. каждый такой отрезок времени пишется текущее значение аналогового сигнала в двоичной форме



WAV


dB
0
-12
-24
-36
-48
-60
-72
-84
-96
-108
-120
dB
Hz
5000
10000
15000
Hz
MP3
WAV

Знаменитая спектрограмма в треке из игры Doom 2016



Как я посмотрел спектрограмму? – В программе “Audacity”

Metadata.

Метаданные файла – это данные о данных (об их составе, содержании, статусе, происхождении, местонахождении, качестве, форматах, объёме, условиях доступа, авторских правах и т. п.

Тулзы для поиска метаданных в файле:

1) exif
2) exfitool
3) Identif (для изображений)
4) …

Метаданные должны отвечать на вопросы:

Почему?	Кто?	Что?	Где?	Когда?	Как?
Почему мы храним эти данные?	Кто создал эти данные?	Каковы бизнес-определения элементов данных?	Откуда взялись эти данные?	Когда были созданы эти данные?	Как эти данные форматируются?
Каково назначение и использование этих данных?	Кто использует эти данные?	Каковы бизнес-правила для этих данных?	Где хранятся эти данные?	Когда последний раз обновлялись эти данные?	Как много баз данных / источников хранят эти данные?
Каковы бизнес-драйверы для использования этих данных?	Кто является распорядителем этих данных?	Какие аббревиатуры / акронимы элементов данных?	Где используются переиспользуются эти данные?	Как долго должны храниться эти данные?	
	Кто владеет этими данными?	Каковы технические стандарты именования для реализации базы данных?	Где находится резервная копия этих данных?	Когда эти данные нужно удалять?	
	Кто регулирует / аудирует эти данные?	Каков уровень безопасности / конфиденциальности этих данных?	Есть ли региональные политики конфиденциальности / безопасности по регулированию этих данных?		

Сжатие данных – как это работает?

Главный признак данных, которые можно сжать - избыточность

Основной принцип алгоритмов сжатия базируется на том, что в любом файле, содержащем неслучайные данные, информация частично повторяется. Дальше идут математика с её моделями, что позволяет нам сжимать данные с потерями и без.

Gzip x bzip2 x XZ

Всё это разные утилиты, использующие разные алгоритмы сжатия данных.

К каждому типу архивирования используется своя тулза для разархивирования

Замечательная статья о где, сравниваются эти форматы сжатия в различных областях.

https://www.rootusers.com/gzip-vs-bzip2-vs-xz-performance-comparison/

Утилиты для сжатых данных без разархивации

- for 7z: 7z x -so example.7z
- for bz2: bzcat example.bz2
- for gz: zcat example.gz
- for xz: xzcat example.xz
- for zip: zcat example.zip
- for **bz2**, **gz** and **xz:** less example.bz2/gz/xz