



МИНОБРНАУКИ РОССИИ

Федеральное государственное бюджетное образовательное  
учреждение  
высшего образования  
**«МИРЭА – Российский технологический университет»**

**РТУ МИРЭА**

Институт информационных технологий (ИТ)  
Кафедра прикладной математики

**ОТЧЁТ ПО ПРАКТИЧЕСКОЙ РАБОТЕ №1**  
**по дисциплине «Большие данные»**

Выполнил студент группы \_\_\_\_\_

Проверил ассистент кафедры ПМ ИИТ

\_\_\_\_\_

Есипов И.В.

Москва 2023

## Задание 1

Развернуть операционную систему Ubuntu Linux Server 18.04 с помощью средства для виртуализации ОС на персональном компьютере. Разрешено как использовать готовый образ ОС по ссылке, так и вручную настроить серверную операционную систему с нуля. Войти в ОС Linux под созданной учетной записью пользователя (в случае с готовым образом системы пользователь/логин: user, пароль: user). (2 балла)

Поскольку архитектура \*NIX систем и их команды оболочки схожи, а так же из-за невозможности установки Oracle VirtualBox на мою домашнюю машину я выполняю работу в системе Mac OS Ventura. В качестве эксперимента я устанавливал второй системой Asahi Linux, единственный дистрибутив Linux, доступный для такой установки для компьютеров Macintosh под управлением процессорами серии M, основанными на архитектуре ARM, но поскольку он находится в альфа стадии разработки, работа на нем не очень удобна.

```

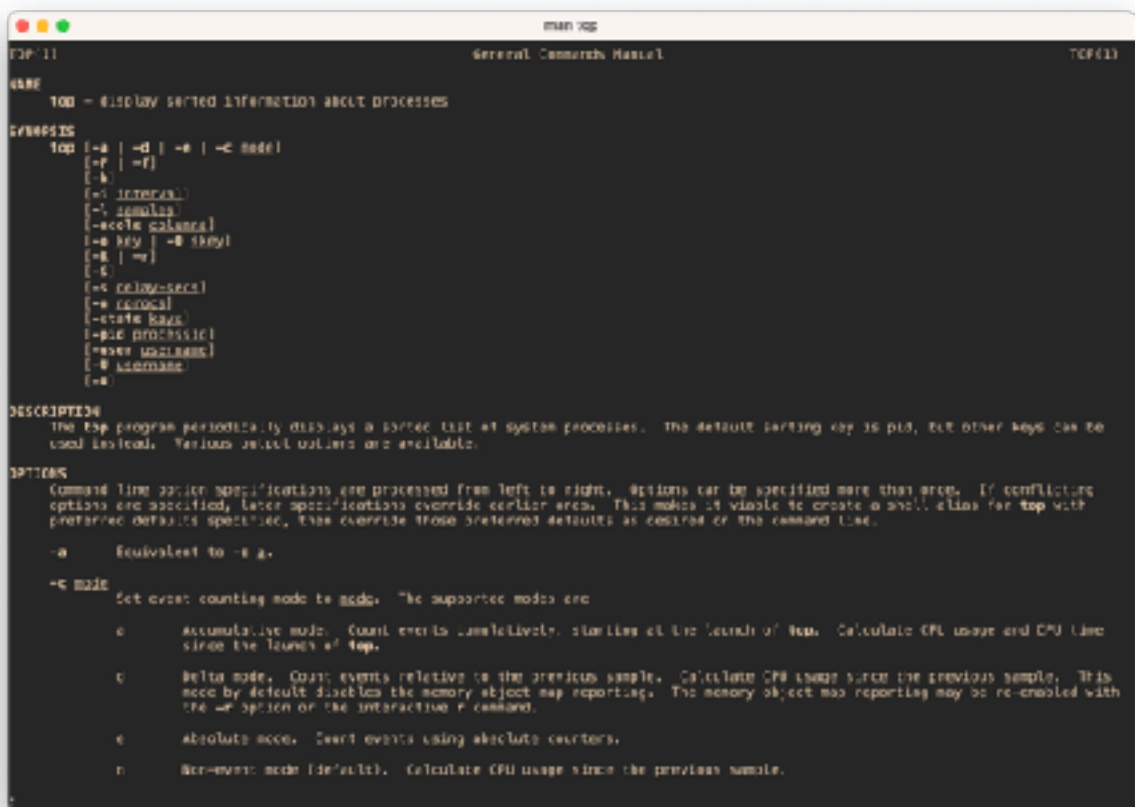
nl25@nuc-mlt:~
n1: Flags=8003<UP,BROADCAST,SMART,RUNNING,SIMPLEX,MULTICAST> mtu 1500
    options=6453<RXCSUM,TXCSUM,TX34,TSOO,CHANNEL_ID,PARTIAL_CSUM,ZEROINVERT_CSUM>
    ether 50:00:13:c0:0b:1a:45
    link 16801:1410102E1J54619C184E0B prefixlen 34 scopeid 300
    link 152:153:161:227 actnum 3<1ffff3 broadcast 152:153:161:253
    nd5 options=281<PRF30MMID,DAD>
    media: aut30e<act
    status: active
bridge0: Flags=8003<UP,BROADCAST,SMART,RUNNING,SIMPLEX,MULTICAST> mtu 1500
    ndirndmcccl50M,16<sm,100a,100a,100a
    ether 16:00:18:c4:d3:00
    Configuration
        id 0 0:0:0:0:0 priority 0 hellotime 0 holdtime 0
        maxage 0 holdcn 0 prio slp maxsize 100 liminal 1000
        root id 00:00:00:00:00 priority 0 ifcost 0 port 0
        spri 100 member flags 000
    member: en1 16cc03<L,CAPIING,3<COVER-
        ifmaxsize 0 port 0 priority 0 path cost 0
    member: en2 16c0dcLAPING,3<COVER-
        ifmaxsize 0 port 0 priority 0 path cost 0
    nd5 options=281<PRF30MMID,DAD>
    media: aut30e<act
    status: inactive
en0: Flags=6643<UP,BROADCAST,RUNNING,SIMPLEX,MULTICAST> mtu 1500
    options=6453<RXCSUM,TXCSUM,TX34,TSOO,CHANNEL_ID,PARTIAL_CSUM,ZEROINVERT_CSUM>
    ether 50:00:13:c0:0b:2f:00
    link 16801:1805102FF1ef0e27a84e0D prefixlen 34 scopeid 600
    nd5 options=281<PRF30MMID,DAD>
    media: aut30e<act
    status: active
en1: Flags=8003<UP,BROADCAST,SMART,RUNNING,SIMPLEX,MULTICAST> mtu 1500
    options=803<CHANNEL_ID>
    ether 50:00:13:c0:0b:2f:00
    link 16801:1805102FF1ef0e27a84e0D prefixlen 34 scopeid 800
    nd5 options=281<PRF30MMID,DAD>
    media: aut30e<act
    status: inactive
en2: Flags=6651<UP,POINTOPOINT,RUNNING,MULTICAST> mtu 1500
    link 16801:1805102FF1ef0e27a84e0D prefixlen 34 scopeid 801
    nd5 options=281<PRF30MMID,DAD>
en3: Flags=6651<UP,POINTOPOINT,RUNNING,MULTICAST> mtu 2000
    link 16801:168016135 ec2107844actum prefixlen 64 scopeid 610
    nd5 options=281<PRF30MMID,DAD>
en4: Flags=6651<UP,POINTOPOINT,RUNNING,MULTICAST> mtu 1000
    link 16801:cc01:01:142c:09a0a0a2 prefixlen 58 scopeid 611
    nd5 options=281<PRF30MMID,DAD>
[save] => ~ wrcm1
nl25@
[save] => ~

```

## Задание 2

Получить справку по команде `top`. С помощью команды `top` просмотреть занимаемое операционной системой место в оперативной памяти. Выйдите из выполнения команды `top` в терминале. (2 балла)

Справка получается с помощью команды `man`, отсортировать процессы по занимаемой оперативной памяти можно с помощью флага `-o` и `secondary key mem`.



```
man top
TOP(1)                                General Commands Manual                                TOP(1)

NAME
    top - display sorted information about processes

SYNOPSIS
    top [-a] [-d] [-s] [-C mode]
        [-P] [-f]
        [-b]
        [-i interval]
        [-t samples]
        [-mode column]
        [-k key] [-K key1]
        [-k] [-u]
        [-S]
        [-s column-sect]
        [-e column]
        [-state keys]
        [-pid processes]
        [-user users]
        [-u username]
        [-u]

DESCRIPTION
    The top program periodically displays a sorted list of system processes. The default sorting key is pid, but other keys can be used instead. Various useful options are available.

OPTIONS
    Command line option specifications are processed from left to right. Options can be specified more than once. If conflicting options are specified, later specifications override earlier ones. This makes it viable to create a shell alias for top with preferred defaults specified, then override those preferred defaults as desired on the command line.

    -a          Equivalent to -s a.

    -C mode     Set event counting mode to mode. The supported modes are:

        a       Accumulative mode. Count events cumulatively, starting at the launch of top. Calculate CPU usage and CPU time since the launch of top.

        d       Delta mode. Count events relative to the previous sample. Calculate CPU usage since the previous sample. This mode by default disables the memory object map reporting. The memory object map reporting may be re-enabled with the -M option or the interactive F command.

        e       Absolute mode. Count events using absolute counters.

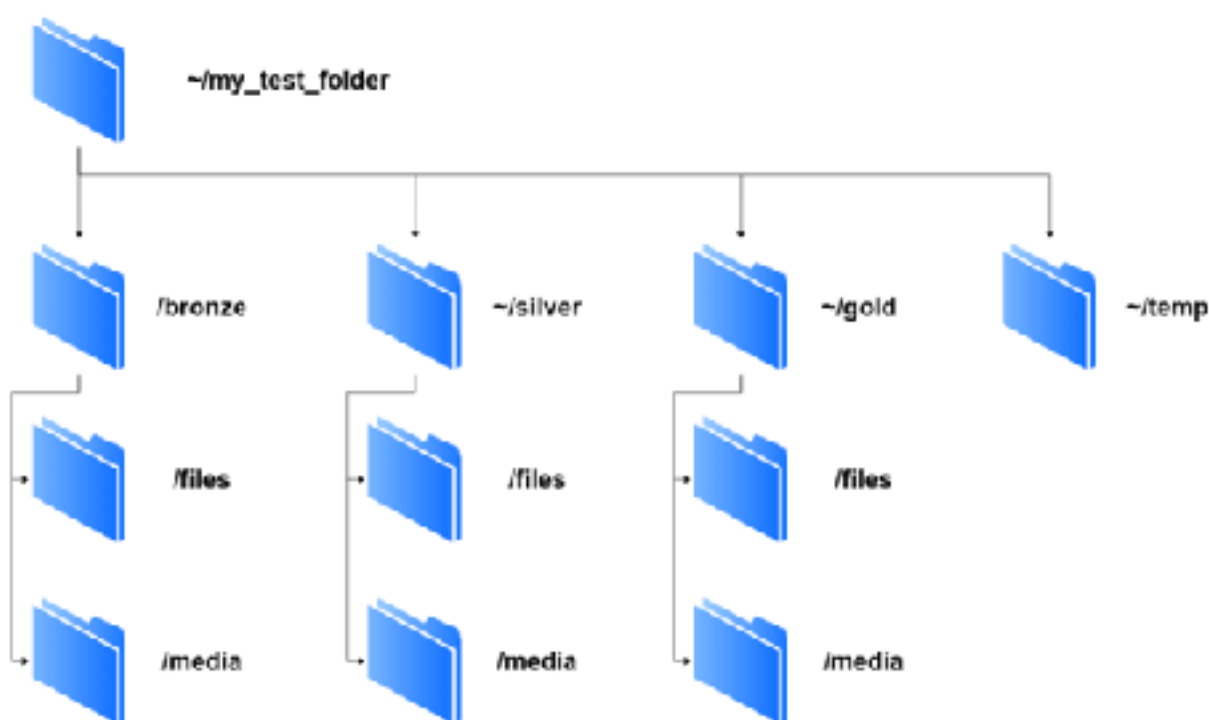
        n       Non-event mode (default). Calculate CPU usage since the previous sample.
```



Для выхода из выполнения команды `top` можно нажать `q` или `ctrl + c`.

### Задание 3

Просмотреть путь к текущей директории. Создать в пользовательской директории папку для самостоятельной работы цифры\_шифра/`my_test_folder`. Внутри нее создать директории в связи со следующей структурой:



Вывести на экран всю созданную древовидную структуру в виде списка папок и подпапок в терминал командной строки.

Перевести вывод созданной структуры в файл `~/my_test_folder/temp/dirs.txt`. (2 балла)

Папки создаются с помощью команды `mkdir`, вывести файловую структуру на экран можно с помощью команды `ls` с флагом `-R`.

```
my_test_folder - m125@mac-af - ..g_test_folder - -zsh - 154x62
(base) ~ % ls
Calendar Library      Public               music                RWD
Desktop               Applications         my_test_folder      tmp
Documents             bin                 my_test_folder      vim
Downloads             books               my_test_folder      xcode
Library               docker-compose      apt                  xcodeproj
MicrosoftWindowsStore Services Fish               parallels            programs
Music                 gha                 projects
Pictures              inst.exe
(base) ~ % pwd
/Users/m125
(base) ~ % cd my_test_folder
(base) ~ % cd my_test_folder RWD
(base) ~ % cd my_test_folder
(base) ~ % cd my_test_folder
(base) ~ % cd my_test_folder
```

```
my_test_folder - m125@mac-af - ..g_test_folder - -zsh - 154x62
-- gold
-- silver
  -- files
    -- media
  -- temp

9 directories, 8 files
(base) ~ % cd my_test_folder; mkdir -p gold/Films
(base) ~ % cd my_test_folder; mkdir -p gold/Films/media
(base) ~ % cd my_test_folder; touch

-- bronze
  -- files
    -- media
-- gold
  -- files
    -- media
-- silver
  -- files
    -- media
-- temp

11 directories, 8 files
(base) ~ % cd my_test_folder; touch
bronze gold silver temp

./bronze:
files

./bronze/files:
media

./bronze/files/media:

./gold:
files

./gold/files:
media

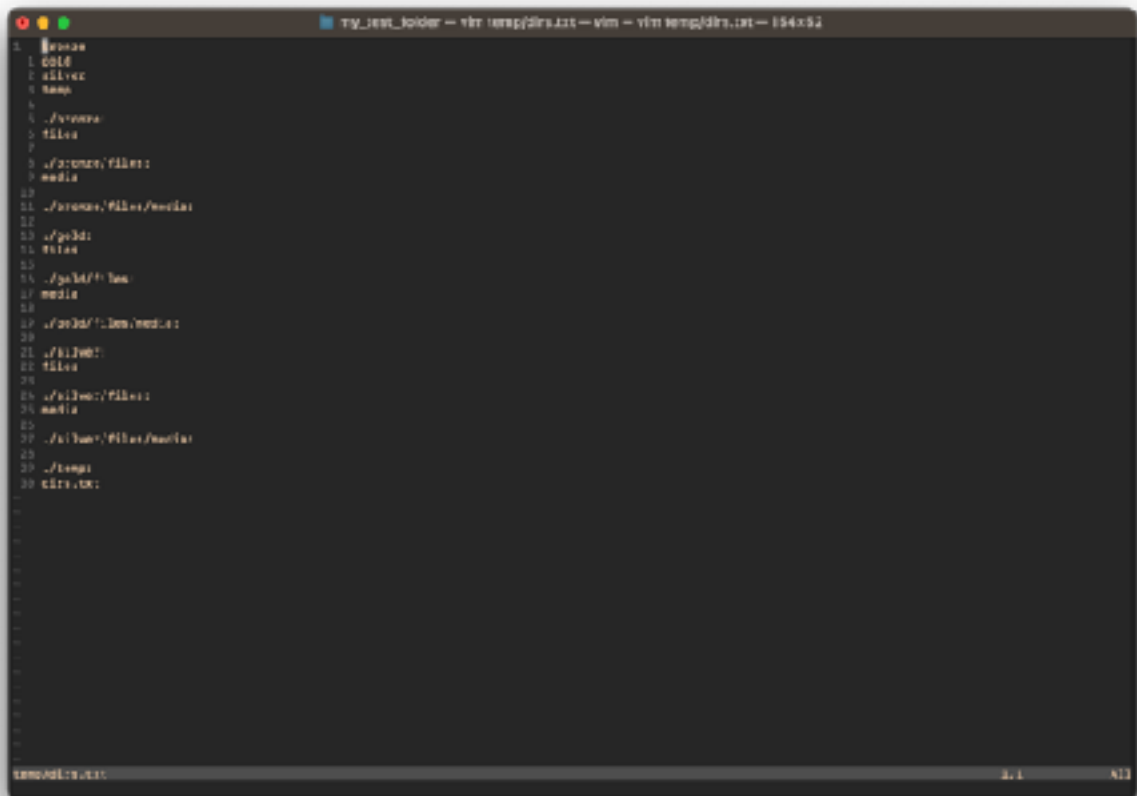
./gold/files/media:

./silver:
files

./silver/files:
media

./silver/files/media:

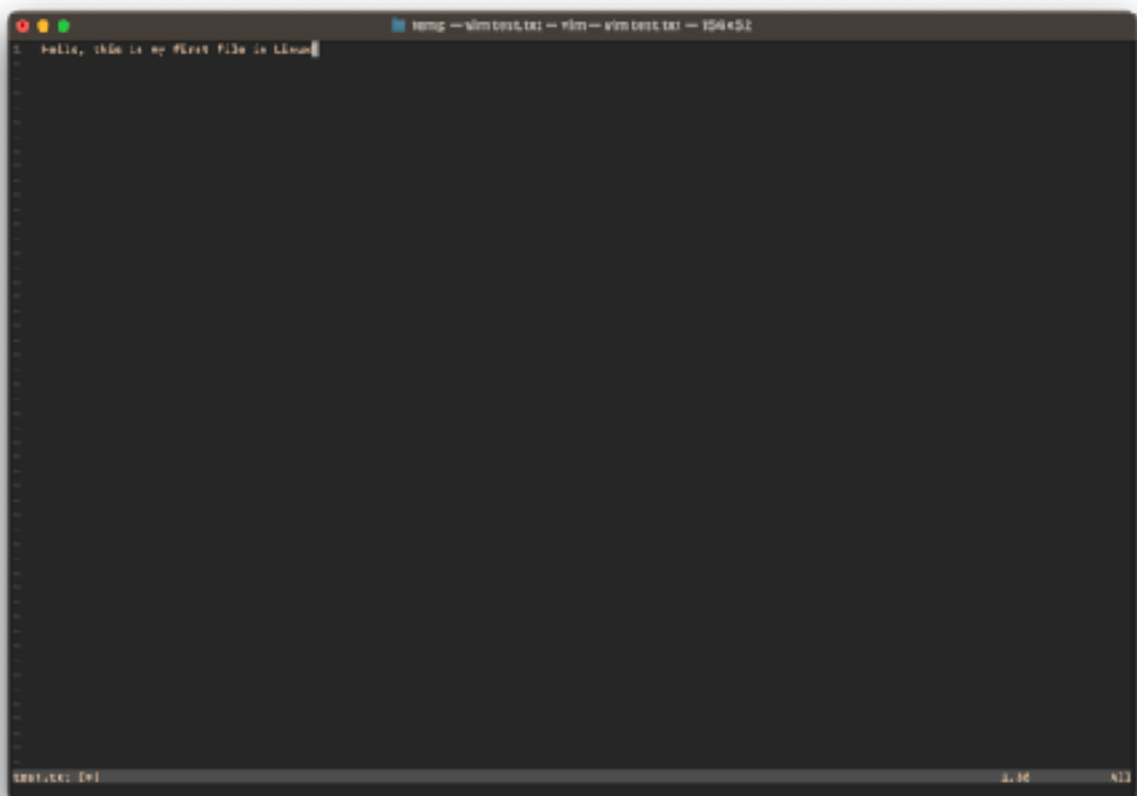
./temp:
(base) ~ % cd my_test_folder
```



```
1 #words
2 gold
3 silver
4 temp
5
6 ./temp
7 files
8
9 ./temp/files:
10 media
11
12 ./temp/files/media:
13
14 ./gold:
15 files
16
17 ./gold/files:
18 media
19
20 ./gold/files/media:
21
22 ./silver:
23 files
24
25 ./silver/files:
26 media
27
28 ./silver/files/media:
29
30 ./temp:
31 test.txt
```

#### Задание 4

Перейти в папку `/my_test_folder/temp`. Создайте в папке пустой файл `test.txt`. Ввести в файл информацию «Hello, its my first file in Linux!». Вывести сообщение из файла в консоль. Просмотрите размер созданного файла. (2 балла)



```
1 Hello, this is my first file in Linux
```

```
temp --al28c@temp:~/.ssh_folder/temp -- ssh -- 154x32

├── gold
│   └── files
│       └── media
├── silver
│   └── files
│       └── media
└── temp

ls -ldircrecursive,* files
(base) ~ my.test_folder ls -ld
because gold silver temp

./findme!
files

./findme!files!
media

./findme!files/media
./gold
files

./gold/files
media

./gold/files/media
./follow!
files

./follow/files
media

./follow/files/media

./temp
(base) ~ my.test_folder ls -ld my.test_folder
(base) ~ my.test_folder cd temp/files
(base) ~ my.test_folder cd temp
(base) ~ temp touch test.txt
(base) ~ temp cd test.txt
(base) ~ temp cd test.txt
(base) ~ temp cd test.txt
(base) ~ temp cd test.txt
public, its my first file in linux!
(base) ~ temp ls -ld
total 16
-rw-r--r-- 1 al28c staff 1280 Feb 24 22:28 test.txt
-rw-r--r-- 1 al28c staff 256 Feb 24 22:27 test.txt
(base) ~ temp
```

## Задание 5

С помощью команды `wget` загрузить в папку `temp` файл архива [https://github.com/qwerty29544/BigDataEssentials/raw/main/Practice1\\_LinuxCommands/data.tar.gz](https://github.com/qwerty29544/BigDataEssentials/raw/main/Practice1_LinuxCommands/data.tar.gz) с данными для выполнения дальнейших пунктов практической работы. Разархивировать файлы в папку `temp`, не создавая новых каталогов. Просмотрите список файлов и их уровней доступа. (2 балла)

Исправление: в первый раз я создал в процессе разархивации папку `data`. Для того, чтобы избежать создания папки, используется ключ `--strip-component=1`

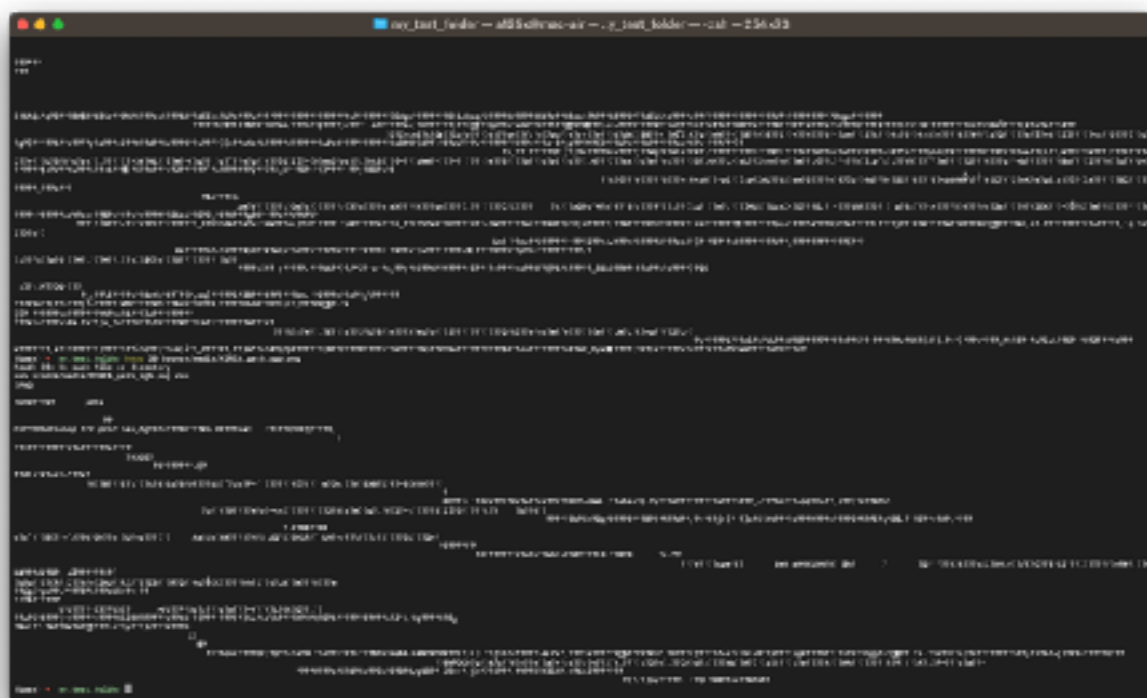






## Задание 7

Выведите в терминал первые строки из файлов изображений. Покажите какой формат изображений выводит информацию в структурированном виде. Чем визуально различаются префиксы во всех форматах файлов изображений. (2 балла)



Поскольку .jpg и .png файлы не просматриваются в структурированном виде, можно сделать вывод, что нужным свойством обладает .eps файл. К сожалению, он не открывается стандартными приложениями и по-видимому является проприетарным форматом компании Adobe.

## Задание 8

Перенести файлы форматов .txt, .TXT, .csv, .db в папку /bronze/files. Для файла TNVED1.TXT выведите первые строки. Проверьте кодировку файла TNVED1.TXT с помощью команды file. Создайте новый файл с именем tnved1\_utf.txt переводя кодировку файла TNVED1.TXT из кодировки CP866 в кодировку UTF-8. Выведите ещё раз содержимое файла в терминал. Прodelайте то же самое с файлами TNVED2 и TNVED3. Перенесите файлы в читаемой новой кодировке в папку /silver/files. (4 балла)

Команда смены кодировки и записи в новый файл: `iconv -f CP866 -t UTF-8//TRANSLIT TNVED1.TXT > tnved1_utf.txt`

```
my_test_folder - 1125x@mac-air - /_test_folder - curl - 15:40:49
├── files
│   ├── media
│   ├── silver
│   │   ├── files
│   │   └── media
│   └── temp
│       ├── CAR.csv
│       ├── INV003.TXT
│       ├── INV003.TXT
│       ├── INV003.TXT
│       ├── data.txt.gz
│       ├── data.txt
│       ├── data.csv
│       ├── data.csv
│       ├── data_orders.db
│       └── test.txt
└── 33 directories, 54 files
(base) ~ my_test_folder on macos:home$tree
(base) ~ my_test_folder on linux:home$tree
(base) ~ my_test_folder$tree
├── files
│   ├── files
│   │   ├── 0002.csv
│   │   ├── INV003.TXT
│   │   ├── INV003.TXT
│   │   ├── INV003.TXT
│   │   ├── data.txt
│   │   ├── data.csv
│   │   ├── data.csv
│   │   ├── data_orders.db
│   │   └── test.txt
│   └── media
│       ├── 0000a_garb_rgb.png
│       ├── 0000a_garb_rgb.jpg
│       ├── 0000a_garb_rgb.png
│       └── 0000a_logo.jpg
├── gold
│   ├── files
│   └── media
├── silver
│   ├── files
│   └── media
└── temp
    ├── data.txt.gz
```

```
files - virt tnved1_utf.txt - virt - nbt tnved1_utf.txt - 15:41:22
1. [51.0001133.0000]
2. [51.0001133.0000]
3. [51.0001133.0000]
4. [51.0001133.0000]
5. [51.0001133.0000]
6. [51.0001133.0000]
7. [51.0001133.0000]
8. [51.0001133.0000]
9. [51.0001133.0000]
10. [51.0001133.0000]
11. [51.0001133.0000]
12. [51.0001133.0000]
13. [51.0001133.0000]
14. [51.0001133.0000]
15. [51.0001133.0000]
16. [51.0001133.0000]
17. [51.0001133.0000]
18. [51.0001133.0000]
19. [51.0001133.0000]
20. [51.0001133.0000]
21. [51.0001133.0000]
22. [51.0001133.0000]
23. [51.0001133.0000]
24. [51.0001133.0000]
25. [51.0001133.0000]
26. [51.0001133.0000]
27. [51.0001133.0000]
28. [51.0001133.0000]
29. [51.0001133.0000]
30. [51.0001133.0000]
31. [51.0001133.0000]
32. [51.0001133.0000]
33. [51.0001133.0000]
34. [51.0001133.0000]
35. [51.0001133.0000]
36. [51.0001133.0000]
37. [51.0001133.0000]
38. [51.0001133.0000]
39. [51.0001133.0000]
40. [51.0001133.0000]
41. [51.0001133.0000]
42. [51.0001133.0000]
43. [51.0001133.0000]
44. [51.0001133.0000]
45. [51.0001133.0000]
46. [51.0001133.0000]
47. [51.0001133.0000]
48. [51.0001133.0000]
49. [51.0001133.0000]
50. [51.0001133.0000]
51. [51.0001133.0000]
52. [51.0001133.0000]
53. [51.0001133.0000]
54. [51.0001133.0000]
55. [51.0001133.0000]
56. [51.0001133.0000]
57. [51.0001133.0000]
58. [51.0001133.0000]
59. [51.0001133.0000]
60. [51.0001133.0000]
61. [51.0001133.0000]
62. [51.0001133.0000]
63. [51.0001133.0000]
64. [51.0001133.0000]
65. [51.0001133.0000]
66. [51.0001133.0000]
67. [51.0001133.0000]
68. [51.0001133.0000]
69. [51.0001133.0000]
70. [51.0001133.0000]
71. [51.0001133.0000]
72. [51.0001133.0000]
73. [51.0001133.0000]
74. [51.0001133.0000]
75. [51.0001133.0000]
76. [51.0001133.0000]
77. [51.0001133.0000]
78. [51.0001133.0000]
79. [51.0001133.0000]
80. [51.0001133.0000]
81. [51.0001133.0000]
82. [51.0001133.0000]
83. [51.0001133.0000]
84. [51.0001133.0000]
85. [51.0001133.0000]
86. [51.0001133.0000]
87. [51.0001133.0000]
88. [51.0001133.0000]
89. [51.0001133.0000]
90. [51.0001133.0000]
91. [51.0001133.0000]
92. [51.0001133.0000]
93. [51.0001133.0000]
94. [51.0001133.0000]
95. [51.0001133.0000]
96. [51.0001133.0000]
97. [51.0001133.0000]
98. [51.0001133.0000]
99. [51.0001133.0000]
100. [51.0001133.0000]
```



