

Санкт-Петербургский государственный политехнический университет
Институт машиностроения, материалов и транспорта
Кафедра «Мехатроника и Роботостроение» при ЦНИИ РТК

ОТЧЕТ ПО ЛАБОРАТОРНОЙ РАБОТЕ № 1

Язык программирования Python

по дисциплине «Математические методы интеллектуальных технологий»

Выполнил
студент гр.3341506/90401

<подпись>

М.А.Борискин

Проверил

<подпись>

С.Р.Орлова

«21» сентября 2019 г.

Санкт-Петербург

2019

1. Общие задачи

Скрипт приведен в репозитории https://github.com/makaryb/ml_1sem_5kurs/blob/master/lab1/firstTaskBoriskin.py

Пример вывода работы скрипта:

```
makary@pafeast:~/ml_1sem_5kurs/lab1$ time python3 firstTaskBoriskin.py
для x=15,y=5 операция + : 20
для x=15,y=5 операция - : 10
для x=15,y=5 операция * : 75
для x=15,y=5 операция / : 3.0
високосный 2019 : False
високосный 2020 : True
високосный 2021 : False
високосный 2022 : False
високосный 2023 : False
високосный 2024 : True
високосный 2025 : False
P, S, d для квадрата со стороной 4 : (16, 16, 5.656854249492381)
P, S, d для квадрата со стороной 8 : (32, 64, 11.313708498984761)
P, S, d для квадрата со стороной 9 : (36, 81, 12.727922061357855)
месяц 12 : зима
месяц 4 : весна
месяц 12 : зима
Сумма на счету будет: 4831.53000000000025
Наше число 598 простое? False
Сегодня 8 / 5 / 2022 - валидно? True
строка: В Берне идут переговоры между генералом Вольфом и американской стороной /// ключ: Юстас /
// шифровка: 3%Дахиа%нбцч%кахажлэлхю%йагбц%жаиахеол%Злощс%л%н%ейахн%п%и%ф%л%м%ф%ч%л%х%л%и%л%м%
шифровка: 3%Дахиа%нбцч%кахажлэлхю%йагбц%жаиахеол%Злощс%л%н%ейахн%п%и%ф%л%м%ф%ч%л%х%л%и%л%м% /// ключ: Юстас
/// дешифровка: В Берне идут переговоры между генералом Вольфом и американской стороной

real    0m0,034s
user    0m0,019s
sys     0m0,015s
makary@pafeast:~/ml_1sem_5kurs/lab1$
```

Explanation on `real`, `user` and `sys` (from `man time`):

- `real` : Elapsed real (wall clock) time used by the process, in seconds.
- `user` : Total number of CPU-seconds that the process used directly (in user mode), in seconds.
- `sys` : Total number of CPU-seconds used by the system on behalf of the process (in kernel mode), in seconds.

2. Создадим файлы

Скрипт приведен в репозитории https://github.com/makaryb/ml_1sem_5kurs/blob/master/lab1/SpamSMS/secondTaskBoriskin.py

Пример вывода работы скрипта:

```
makary@pafeast:~/ml_1sem_5kurs/lab1/SpamSMS$ time python3 secondTaskBoriskin.py

real    0m1,251s
user    0m0,325s
sys     0m0,337s
makary@pafeast:~/ml_1sem_5kurs/lab1/SpamSMS$
```

3. Проанализируем файлы

Скрипт приведен в репозитории https://github.com/makaryb/ml_1sem_5kurs/blob/master/lab1/SpamSMS/thirdTaskBoriskin.py

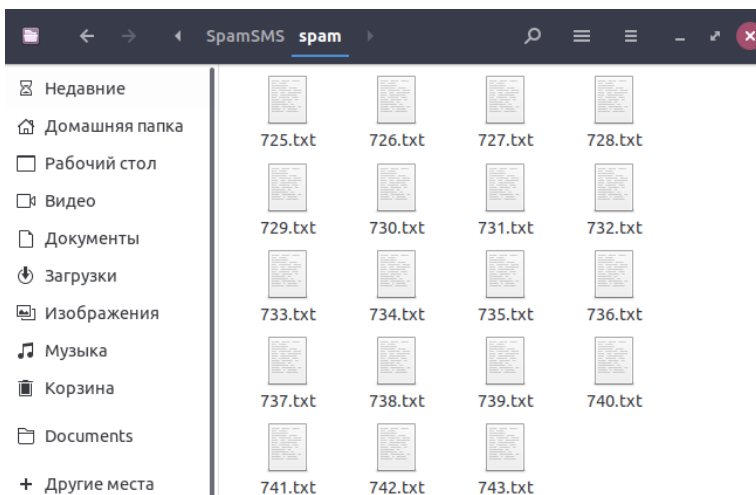
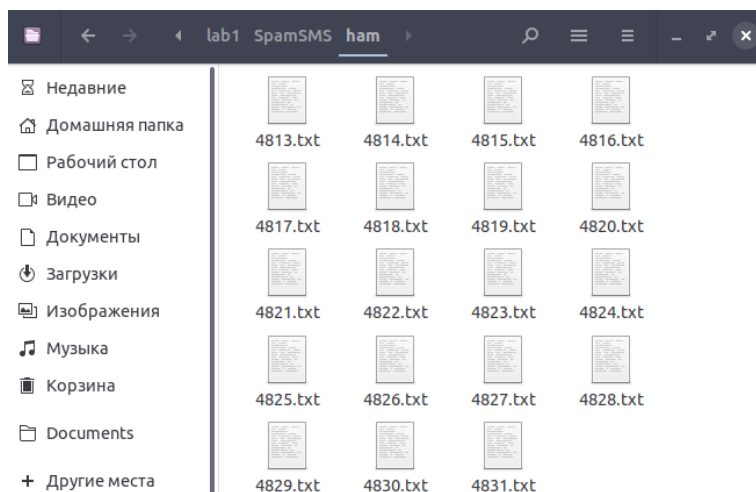
Результаты работы скрипта:

Найдите минимальную, максимальную и среднюю длину содержимого файлов

```
makary@pafeast:~/ml_1sem_5kurs/lab1/SpamSMS$ time python3 thirdTaskBoriskin.py
Минимальная длина содержимого ham файлов: 27
Максимальная длина содержимого ham файлов: 126
Средняя длина содержимого ham файлов: 72.53674187538812

Минимальная длина содержимого spam файлов: 148
Максимальная длина содержимого spam файлов: 161
Средняя длина содержимого spam файлов: 139.6164199192463

Минимальная длина содержимого всех файлов: 27
Максимальная длина содержимого всех файлов: 161
Средняя длина содержимого всех файлов: 81.47829207032652
```



Посчитайте, сколько раз каждый символ (включая пробел) встречается во всех файлах и выведите эти значения, отсортированные по величине ↓↓↓

Найдите самые популярные слова:

> 500 упоминаний

```
Для папки ham:  
( 'for', 505)  
( 'of', 521)  
( 'it', 524)  
( 'me', 710)  
( 'is', 725)  
( 'my', 746)  
( 'in', 802)  
( 'and', 850)  
( 'u', 931)  
( 'a', 1061)  
( 'the', 1124)  
( 'to', 1571)  
( 'you', 1755)  
( 'i', 2184)
```

>100 упоминаний

```
Для папки spam:  
( 'reply', 101)  
( 'stop', 102)  
( 'claim', 106)  
( 'with', 108)  
( 'mobile', 111)  
( 'text', 115)  
( 'now', 115)  
( 'and', 118)  
( 'from', 127)  
( 'txt', 134)  
( 'u', 135)  
( 'have', 135)  
( 'ur', 142)  
( 'on', 143)  
( 'is', 151)  
( '2', 169)  
( 'or', 187)  
( 'free', 191)  
( 'for', 201)  
( 'the', 204)  
( 'you', 258)  
( 'your', 262)  
( 'call', 343)  
( 'a', 373)  
( 'to', 674)
```

>300 упоминаний

```
В общем:  
( 'this', 324)  
( 'we', 346)  
( 'now', 356)  
( 'just', 368)  
( 'i'm', 377)  
( 'get', 378)  
( 'with', 379)  
( 'be', 380)  
( 'will', 380)  
( 'if', 382)  
( 'ur', 386)  
( 'do', 389)  
( 'can', 399)  
( 'at', 400)  
( 'or', 415)  
( 'not', 419)  
( 'but', 422)  
( 'so', 431)  
( '2', 469)  
( 'are', 489)  
( 'that', 495)  
( 'on', 527)  
( 'it', 550)  
( 'call', 571)  
( 'have', 573)  
( 'of', 616)  
( 'your', 677)  
( 'for', 706)  
( 'me', 738)  
( 'my', 756)  
( 'in', 872)  
( 'is', 876)  
( 'and', 968)  
( 'u', 1066)  
( 'the', 1328)  
( 'a', 1434)  
( 'you', 2013)  
( 'i', 2227)  
( 'to', 2245)
```

```
real    0m38,032s  
user    0m36,857s  
sys     0m0,593s  
makary@pafeast:~/ml_1sem_5kurs/lab1/SpamSMS$
```

Выводы:

В среднем длина строки в Spam-файлах \approx в 2 раза выше длины строки в ham-файлах.

Наиболее встречаемые символы в ham- и spam-файлах: первые пять полностью совпадают. Однако по спец. символам четкого вывода сделать нельзя, так как (как было показано выше на скриншотах) количество ham-файлов \approx в 6,5 раза больше количества рассмотренных spam-файлов.

При рассмотрении ham-файлов можно сделать вывод, что наиболее популярные слова: i, you, to, the, a, u, and, in, my, is, me, it, of, for ... - это слова, характерные для личной переписки.

В spam-файлах, напротив, наиболее популярные слова: to, a, call, your, you, the, for, free, or, 2, is, on, ur, have ... - это слова, характерные для рекламы.