# 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 задачки

Скрипт приведен в репозитории <a href="https://github.com/makaryb/ml">https://github.com/makaryb/ml</a> 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
Р, S, d для квадрата со стороной 4 : (16, 16, 5.656854249492381)
P, S, d для квадрата со стороной 8 : (32, 64, 11.313708498984761)
Р, S, d для квадрата со стороной 9 : (36, 81, 12.727922061357855)
месяц 12 : зима
месяц 4 : весна
месяц 12 : зима
Сумма на счету будет: 4831.5300000000025
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%кахажлзлхю%йагбц%жаиахеолй%Злощслй%н%ейахнпеифплм%фчлхлилм
шифровка: 3%Дахиа%нбцч%кахажлзлхю%йагбц%жаиахеолй%Злощслй%н%ейахнпеифплм%фчлхлилм /// ключ: Юстас
/// дешифровка: В Берне идут переговоры между генералом Вольфом и американской стороной
       0m0,034s
real
user
        0m0,019s
       0m0,015s
sys
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 <a href="https://github.com/makaryb/ml">https://github.com/makaryb/ml</a> 1sem 5kurs/blob/master/lab1/SpamSMS/secondTaskBoriskin.py

Пример вывода работы скрипта:

# 3. Проанализируем файлы

Скрипт приведен в репозитории <a href="https://github.com/makaryb/ml">https://github.com/makaryb/ml</a> 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
Muнимальная длина содержимого ham файлов: 27
Maксимальная длина содержимого ham файлов: 126
Cpeдняя длина содержимого ham файлов: 72.53674187538812

Muнимальная длина содержимого spam файлов: 148
Maксимальная длина содержимого spam файлов: 161
Cpeдняя длина содержимого spam файлов: 139.6164199192463

Muнимальная длина содержимого всех файлов: 27
Maксимальная длина содержимого всех файлов: 161
Cpeдняя длина содержимого всех файлов: 161
```



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

```
('z', 298)
('"' 322)
                                                                                                            ('G', 324)
('D', 334)
('H', 335)
                                                                                                                                                                                ('7', 858)
('G', 880)
                                                                                                                                                  В общем:
Для папки ham:
                                                                        Для папки spam:
                                                                                                                                                                                (G', 880)
('B', 95
                                               , 322)
                                    (')', 433)
('2', 445)
                                                                                                            (H', 335)
                                                                                                                                                                               (B', 894)
('&', 922)
('P' 96
                                                                         '\t', 0)
'\x92', 0)
''', 0)
'ü', 0)
                                                                                                                                                   ('\t', 0)
('ú', 1)
('è', 1)
('\t', 0)
('ú', 0)
('<', 0)
                                                                                                            ('F', 338)
                                                                                                                                                                               ('P', 987)
('4', 1057)
('Y', 10
                                     (2', 449)
('P' 45
                                                                                                            ('/', 366)
 (, 0)
                                                                                                                                                    e', 1)
                                    ('P', 458)
('j', 550)
('G', 55
                                                                        ('Ü', 0)
                                                                                                                  (, 373)
                                                                                                                                                  ('»', 1)
                                                                                                                                                                                (Y', 1064)
('5', 1064)
                                                                                                                                                  ('-', 1)
                                                                                                            (',', 411)
('L', 421)
   '\x93'
                , 0)
                                     ( G', 556)
('B', 55
                                                                                                            ('I', 421)
                                                                                                                                                                                ('Y', 1064)
('5', 1073)
('L', 1125)
('U', 1138)
('R', 1154)
('D', 1157)
('M', 1229)
('x', 1298)
('8', 1315)
('W', 1374)
('!', 1397)
('¡', 0)
('>', 1)
                                                                         ()
()
                                     (B', 556)
                                                                                                            (1', 423)
('W', 423)
                                                                        ('-', 0)
                                                                                                                                                  ('i' 1)
                                                                                                           ('W', 434)
('M', 455)
('U', 455)
                                     (R', 557)
                                                                                                                                                 ( i', 1)
('鈥', 1)
('¦', 1)
('言', 1)
('^', 1)
                                     ('C', 566)
                                                                      ('\\', 0)
('\\', 0)
('è', 0)
('^', 0)
('»', 1)
                                    ('x', 615)
('U', 62
                                                                                                            (U', 464)
('P'
                                                                                                            (P', 529)
                                                                                                                                                    = ', 1)
         , 1)
                                     ('X', 615)
('U', 674)
('L', 704)
('&', 744)
(';', 766)
('Y', 769)
('M', 774)
('D', 823)
('!', 856)
('W', 940)
('i', 1)
                                                                                                            ('!', 541)
('A', 584)
                                                                                                                                                            , 2)
                                                                       ('\x94', 0)
('\x91', 0)
('\x96', 0)
('»', 0)
('-', 0)
                                                                                                                                                   ('\x91', 2)
('\x93', 2)
('鈥', 1)
                                                                                                            ('R', 597)
('9', 614)
('N', 650)
('鈥', 1)
('∱', 1)
('≐', 1)
('\x91',
                                                                                                                                                                                (!', 1397)
('C', 1
                                                                                                                                                   ('¡', 2)
('~', 3)
                                                                                                            ('N', 650)
('k', 659)
('k', 683)
('v', 684)
('0', 702)
('6', 720)
('3', 727)
('\n', 743)
('S', 779)
                                                                                                                                                                                ('C', 1405)
('H', 1549)
('?', 1550)
('N', 1672)
                                                                                                                                                           , 3)
                                                                        ('é', 0)
('É', 0)
'É', 0)
                   2)
                                                                                                                                                            , 3)
('[', 2)
(']', 2)
(''', 2)
                                                                                                                                                  ('\x96', 3)
('é', 3)
('\\', 4)
                                     ('!', 856)
('W', 940)
('N', 1022)
('E', 1066)
('O', 1173)
('S', 1204)
('H', 1214)
('T', 1280)
('A', 1310)
('?', 1376)
                                                                        ('让', 0)
('鈥', 0)
'¦', 0)
'言', 0)
                                                                                                                                                                                ('2', 1719)
('1', 1751)
('0', 1875)
("'", 1879)
         , 2)
                                                                                                                                                  ( \x94', 4)
('[', 5)
(']', 5)
         , 3)
( \x96', 3)
('é', 3)
('\\', 4)
('\x94'
                                                                                                                                                                                ('", 1879)
('A', 189
                                                                        =', 0)
''u', 1)
                                                                                                                                                  ('<' 5)
                                                                                                             ('S', 779)
('7', 796)
('4', 801)
('\x94',
('%', 7)
('£', 8)
                                                                                                                                                                                ('E', 1958)
(',', 1980)
('S', 1983)
                                                                                                                                                   '-')
                                                                                                                                                           , 9)
                                                                                 , 1)
                   4)
                                                                                                                                                  (' ', 10)
                                                                                , 2)
                                                                                                             ('4', 801)
('C', 827)
('b', 839)
                                                                                                                                                                                (',', 1980)
('S', 1983)
('T', 2339)
('v', 3175)
('0', 3189)
('I', 3503)
('k', 4156)
('b', 4460)
('f', 5137)
('\n', 5574)
('p', 5881)
('w', 7146)
('g', 7470)
('-', 8)
                                                                                                                                                    ,-,, 15)
                                                                                 , 2)
  ' , 9)
                                                                   ('\x93', 2)
('i', 2)
('i', 3)
('E', 3)
('E', 3)
('Y, 4)
('Y, 6)
('Y, 6)
('Y, 25)
('Y, 25)
('Y, 38)
('Y, 38)
('Y, 48)
('Y, 54)
('Y, 66)
('Y, 66)
(Y, 66)
(Y, 66)
(Y, 17)
(Y, 174)
                                                                                                                                                           , 16)
                                      (',', 1569)
("'"
                                                                                                            ('E', 892)
('f', 939)
('5', 965)
('g', 992)
('T', 1059)
('8', 1206)
                                                                                                                                                  ('$', 23)
('>', 31)
         , 12)
                                     (',', 1569)
("'", 1752)
('v', 2491)
('I', 3080)
('k', 3497)
('b', 3621)
('f', 4198)
('p', 4280)
('\n', 4831)
('c', 5425)
('w', 5586)
('g', 6478)
('m', 7678)
('y', 7901)
('@', 13)
('Z', 16)
('...', 16)
                                                                                                                                                  (Z', 33)
('|', 36)
         , 16)
  '$'
         , 19)
                                                                                                                                                           , 37)
('=', 22)
                                                                                                                                                   (ˈ@ˈ, 38)
                                                                                                             ('2', 1270)
('y', 1485)
('w', 1560)
('.', 1561)
('Q', 25)
('|', 36)
(''', 37)
                                                                                                                                                   ('\x92', 39)
                                                                                                                                                  ('=', 47)
                                                                                                                                                  ('Ü', 53)
          , 37)
                                                                                                                                                  ('Q', 73)
('(', 119
                                                                                                                                                                                ('g', 7470)
('c', 7511)
('y', 9386)
                                                                                                                     , 1561)
   '\x92', 39)
                                                                                                                                                  (', 119)
('ü', 129)
('+', 40)
('/', 46)
                                                                                                               'h', 1565)
('/', 46)
('9', 51)
('Ü', 53)
                                                                                                             ('1', 1565)
('p', 1601)
('m', 1726)
                                                                                                                                                                                         , 9386)
                                                                                                                                                                                ('m', 9404)
                                                                                                                                                   '+'ن
                                                                                                                                                           , 137)
                                                                                                                                                    'q', 163)
                                      ('y', 7901)
('d', 8415)
                                                                                                                                                                                 ('d', 10217)
                                                                                                                                                  ('#', 287)
 ('7', 62)
                                                                                                             ('d', 1802)
                                                                                                                                                                                         , 11214)
                                     ('d', 8415)
('u', 9240)
('.', 9653)
('h', 11186)
('l', 11591)
('r', 12664)
('s', 13757)
('i', 15681)
('n', 16540)
('(', 65)
('6', 86)
                                                                                                                                                           , 308)
                                                                                                                                                                                ('u'
                                                                                                             ('u', 2022)
                                                                                                                                                     יט'
                                                                                                                                                                                         , 11262)
                                                                                                             ('c', 2086)
                                                                                                                                                           , 311)
                                                                                                                                                                                ('h', 12751)
('X', 106)
                                                                                                             ('0', 3065)
                                                                                                                                                     'X'
                                                                                                                                                           , 319)
                                                                                                                                                                                ('l', 14728)
('r', 16677)
('5', 108)
                                                                                                             ('l', 3137)
                                                                                                                                                           , 329)
                                                                                                                                                                                         , 16677)
                                                                                                                                                     '£'
                                                                                                                                                                                ('s'
('8', 109)
                                                                                                             ('s', 3266)
                                                                                                                                                     1111
                                                                                                                                                            , 347)
                                                                                                                                                                                         , 17023)
                                                                                                                                                                                ('i', 19077)
('ü', 120)
                                                                                                             ('i', 3396)
                                                                                                                                                           , 367)
                                                                                                                                                     'v'
                                                                                                                                                                                ('n', 20227)
 ('0', 124)
                                                                                                             ('n', 3687)
                                                                                                                                                   Ċ'ĸ'
                                                                                                                                                           , 404)
('q', 125)
                                                                                                                                                           , 419)
                                                                                                                                                                                 ('a'
                                                                                                                                                                                         , 23478)
                                                                                                             ('r', 4013)
                                                                                                                                                                                 ('t'
('3', 125)
                                                                                                             ('a'
                                                                                                                     , 4394)
                                                                                                                                                     'z'
                                                                                                                                                           , 472)
                                                                                                                                                                                         , 25768)
                                      ('a', 19084)
  '1', 186)
                                      ('t', 20904)
                                                                                                             ('t', 4864)
                                                                                                                                                           , 499<sup>°</sup>)
                                                                                                                                                                                 ('o', 27278)
  'V', 193)
                                      ('o', 21484)
                                                                                                               'o', 5794)
                                                                                                                                                            , 585)
                                                                                                                                                                                ('e', 33204)
('', 81961)
         , 196)
                                      ('e', 26544)
                                                                                                             ('e', 6660)
                                                                                                                                                           , 640<sup>°</sup>)
                                                                                                                                                                                         , 81961)
                                                                                                                                                   ( J
('F'
('9'
         , 218)
                                               , 64916)
                                                                                                                      , 17045)
                                                                                                                                                           , 651)
  יכי
  '4', 256)
                                                                                 , 213)
                                                                                                                                                            , 665)
                                                                          'X'
         , 279<sup>°</sup>)
                                                                                 , 247)
                                                                                                                                                            , 745)
 ('K'
                                                                                 , 295)
  '#', 281)
                                                                          'Y'
                                                                                                                                                            , 768)
                                                                                                                                                           , 806)
                                                                                                                                                     '6'
   'F'
         , 285)
                                                                          '£'
                                                                                , 321)
                                                                                                                                                     '3'
                                                                                                                                                               852)
```

## Найдите самые популярные слова:

#### > 500 упоминаний

## >100 упоминаний

## >300 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)
```

```
Для папки spam:

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

```
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)
('acan', 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)
('in', 872)
('is', 876)
('the', 1328)
('the', 1328)
('the', 1328)
('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-фалах: первые пять полностью совпадают. Однако по спец. символам четкого вывода сделать нельзя, так как (как было показано выше на скриншотах) количество ham-файлов ≈ 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.