

# Python. Работа с текстовыми файлами

ноябрь, 2024

Обработка текстовых данных – яркий пример задачи, в которой заранее неизвестно, сколько данных нам предстоит считать, а именно, заранее не известно, сколько строк будет введено.

В Python для открытия файла используется функция `open()`, принимающая два аргумента: имя файла и режим открытия (“r” для чтения и “w” для записи). Эта функция возвращает ссылку на объект типа файл. Иногда в функции `open` используют аргумент `encoding` для указания кодировки символов, которая будет использоваться для чтения или записи файла. Файлы могут быть сохранены в различных кодировках, таких как UTF-8, ASCII и т.д. Если кодировка файла не совпадает с кодировкой, используемой вашей программой, то есть вероятность столкнуться с проблемами при чтении или записи данных. После окончания работы с файлами нужно вызвать для них методы `close()`.

```
in_file = open('input.txt', 'r', encoding='utf8')
...
in_file.close()
```

Конструкция `with open ... as ...` предлагает удобный способ работы с файлами, гарантируя их автоматическое закрытие после завершения операций. Внутри блока `with` файл доступен для чтения или записи, а по завершении блока Python автоматически закрывает его, избавляя от необходимости явного вызова метода `file.close()`.

```
with open('input.txt', 'r') as file:
    #Работа с данными
    #Работа с данными
#В этой точке файл уже закрыт
```

Методы для работы с файлами:

`read()` - Чтение содержимого файла и сохранение его в переменной

`read(size)` - Считывание из файла указанное количество символов

`readline()` - Построчное чтение

`readlines()` - Чтение всех строк в виде списка строк

`write()` - Запись строки в файл

**Подробную теорию для данной лабораторной вы можете найти в учебном пособии**

# Аудиторный практикум

## Задание 1

Дан текстовый файл file1.txt. Прочитайте данные из этого файла.

## Задание 2

Продолжаем работать с этим файлом. В этом задании необходимо:

1. Прочитать первую строку данного файла. (Для данного файла в первой строке хранится весь первый абзац)
2. Убьете из прочитанной строки все знаки препинания
3. Теперь разбейте строку на слова и храните их в списке words
4. Найдите первое самое длинное слово в этом списке
5. Найдите первое самое короткое слово в этом списке
6. Удалите слово "все" из всей строки (в ней их 3 шт.)

## Задание 3

Напишите программу, которая читает текстовый файл file2.txt и подсчитывает количество слов в нем.

## Задание 4

Довольно распространённая ошибка ошибка – это повтор слова. Вот в предыдущем предложении такая допущена. Необходимо исправить каждый такой повтор. Повтор это – слово , ТОЛЬКО один пробел, и снова то же слово.

ЧТО??? Где-то видели такое задание? Да, только теперь необходимо прочитать файл и исправить такие повторы. Входной файл file3.txt Результирующий текст необходимо записать в новый файл res\_task3.txt.

## Задание 5

Зачет в олимпиаде проводится без деления на классы. Выведите фамилию и имя победителя олимпиады. Гарантируется, что победитель будет один. Список человек и их количество баллов хранится в файле file4.txt

## Лабораторная работа 10

В данной лабораторной работе присутствуют задания повышенной сложности, поэтому оценивание выполненной лабораторной работы будет происходить следующим образом:

Критерии оценивания лабораторной:

- Чтобы получить оценку 4 (75%), необходимо сделать 6 заданий.
- Чтобы получить оценку 5 (90%), необходимо сделать 7 заданий.
- Чтобы получить оценку 5 (100%), необходимо сделать 8 заданий.

### Задание 1

Продолжение задания 4 из аудиторного практикума. Результаты олимпиады подводятся без деления на классы. Победителем олимпиады становятся те, кто набрал больше всего баллов. Призерами олимпиады становятся участники, следующие за победителями. Из файла file4.txt выведите на экран призера (Он один в данном файле)

Входные данные	Выходные данные
task4.txt	Лебедев Евгений (Набранное количество баллов: 98)

### Задание 2

Дано два файла file5.txt и file6.txt. Определите в каком файле находится слово Academy

### Задание 3

Истории литературы известен случай написания романа объемом около 50 тыс. слов, в котором ни разу не была употреблена самая популярная в английском алфавите буква E. Название его – «Gadsby».

Напишите программу, которая будет считывать список слов из файла и собирать статистику о том, в каком проценте слов используется 'e'. В качестве исходного файла можете использовать file6.txt.

В файле не нужно удалять знаки препинания! Просто прочитайте его и разделите на слова по пробелам.

Входные данные	Выходные данные
task_lab_4.txt	44.30...

### Задание 4

Набор данных, содержит 2 файла (file7.txt, file8.txt) с мужскими и женскими популярными именами. Каждая строка файлов - это <имя> <процент популярности> Напишите программу, которая будет выводить n популярных мужских или женских имен (в зависимости от того, что попросит пользователь)

!!! ПРОЦЕНТ ВЫВОДИТЬ НЕ НУЖНО !!!

\* данные два файла не являются официальной статистикой и были сгенерированы ChatGPT 4o

Входные данные	Выходные данные
5 м	Александр Максим Иван Дмитрий Андрей
6 ж	София Ева Мария Анастасия Анна Виктория

### Задание 5

Дан текстовый файл (можно самим создать). Пользователь вводит новую строку текста. Напишите программу для добавления в середину файла введенной пользователем строки, сохраняя при этом старое содержимое файла.

### Задание 6

Ты - криптограф, работающий над разоблачением заговора. Ты получил тайное сообщение, в котором каждое слово записано в обратном порядке.

Задача: 1. Напишите программу дешифрования, которая считывает текстовый файл, в котором каждое слово записано в обратном порядке, и выводит на экран расшифрованное сообщение. 2. Напишите программу шифрования текстовых файлов по логике написанной выше.

### Задание 7

Создание пароля посредством генерирования случайных символов может обернуться сложностью в запоминании полученной относительно надежной последовательности. Некоторые системы создания паролей рекомендуют сцеплять вместе два слова на английском языке, тем самым упрощая запоминание заветного ряда символов – правда, в ущерб его надежности.

Напишите программу, которая будет открывать файл со списком слов, случайным образом выбирать два из них и сцеплять вместе для получения итогового пароля. При создании пароля исходите из следующего требования: он должен состоять минимум из восьми символов и максимум из десяти, каждое из используемых слов должно быть длиной хотя бы в три буквы. Кроме того, сделайте заглавными первые буквы обоих слов, чтобы легко можно было понять, где заканчивается одно и начинается другое. По завершении процесса полученный пароль должен быть отображен на экране.

### Задание 8 (Отвлечемся от файлов)

Лиса Ciel учится программировать. Одним из её первых заданий было нарисовать лису! Однако для неё это показалось слишком сложным, так что вместо лисы она решила нарисовать змейку.

Змейка — это узор в прямоугольной таблице из  $n$  строк по  $m$  столбцов. Обозначим  $s$ -ю ячейку  $g$ -й строки как  $(g, s)$ . Хвост змеи располагается в ячейке  $(1, 1)$ , а её туловище простирается до  $(1, m)$ , затем спускается на 2 ряда вниз до  $(3, m)$ , затем простирается влево до  $(3, 1)$ , и так далее. Таким образом, голова змеи расположена в одном из углов  $n$ -й строки.

Ваша задача — изобразить такую змею для лисы Ciel: пустые ячейки должны быть обозначены точками ('.'), а ячейки, накрытые телом змеи, должны быть заполнены решетками ('#').

Для полного понимания требуемого узора обратите внимание на примеры выходных данных.

**Входные данные** В единственной строке записано два целых числа:  $n$  и  $m$  ( $3 \leq n, m \leq 50$ ).  
 $n$  — нечетное число.

**Выходные данные** Выведите  $n$  строк. В каждой строке должна быть записана последовательность из  $m$  символов.

Входные данные	Выходные данные
3 3	### ..# ###
3 4	#### ...# ####
5 3	### ..# ### #.. ###
9 9	##### .....# ##### #..... ##### .....# ##### #..... #####

# Решения аудиторного практикума

## Задание 1

```
# 1 вариант
f = open('/Users/anthonyagalakov/Documents/Informatics course/task1.txt', 'r')
text = f.read()
f.close()

print(text)

#2 вариант
with open('/Users/anthonyagalakov/Documents/Informatics course/task1.txt', 'r') as f:
    text1 = f.read()

print(text1)
```

## Задание 2

```
# 1 пункт задания
with open('/Users/anthonyagalakov/Documents/Informatics course/task1.txt', 'r') as f:
    l = f.readline()

# 2 пункт задания

    #1 вариант (решение в лоб)
if '.' in l:
    l = l.replace('.', '')
if ',' in l:
    l = l.replace(',', '')
if '!' in l:
    l = l.replace('-', '')
if '(' in l:
    l = l.replace('(', '')
if ')' in l:
    l = l.replace(')', '')
if '-' in l:
    l = l.replace('-', '')

    #2 вариант

symbols = ". , ! ( ) -"

for c in symbols:
    if c in l:
        l = l.replace(c, '')

# 3 пункт задания
```

```

words = l.split()
print(words)

# 4 пункт задания
long_word = words[0]
for word in words:
    if len(word) > len(long_word):
        long_word = word

print('Первое самое длинное слово: ', long_word)

# 5 пункт задания
shord_word = words[0]
for word in words:
    if len(word) < len(shord_word):
        shord_word = word

print('Первое самое короткое слово: ', shord_word)

# 6 пункт задания
delete_word = "Все"

# 1 способ (remove)
while delete_word in words:
    words.remove(delete_word)

while delete_word.title() in words:
    words.remove(delete_word.title())

# 2 способ (remove)

for word in words:
    if word == delete_word:
        words.remove(word)
    if word == delete_word.title():
        words.remove(word)

# 3 способ (pop)

i = 0
while i < len(words):
    if words[i] == delete_word or words[i] == delete_word.title():
        words.pop(i)
    else:
        i += 1

print(words)

```

### Задание 3

```
with open('task3.txt', 'r') as f:
    text = f.read()
```

```
words = text.split()
result = []
last_word = ""
for word in words:
    if word != last_word:
        result.append(word)
        last_word = word
```

```
res_text = " ".join(result)
```

```
with open('res.txt', 'w') as f:
    f.write(res_text)
```

### Задание 4

# 1 вариант

```
winner = ""
max_score = 0
```

```
with open('task4.txt', 'r') as file:
    record = file.readline()
    while record:
        human = record.split()
        if int(human[-1]) > max_score:
            winner = human[0] + " " + human[1]
            max_score = int(human[-1])
        record = file.readline()
```

```
print("Победитель:", winner, " ( Набранное количество баллов:", max_score, ')')
```

2 вариант

```
winner = ""
max_score = 0
```

```
with open('task4.txt', 'r') as file:

    for record in file:
        human = record.split()
        if int(human[-1]) > max_score:
            winner = human[0] + " " + human[1]
            max_score = int(human[-1])
```



```
print("Победитель:", winner, " ( Набранное количество баллов:", max_score, ')')
```