Лабораторная работа №3

Тема: «Работа с вводом и выводом данных, чтение и запись файлов».

Открытие и закрытие файлов

Python поддерживает множество различных типов файлов, но условно их можно разделить на два виде: текстовые и бинарные. Текстовые файлы – это, к примеру, файлы с расширением cvs, txt, html, в общем любые файлы, которые сохраняют информацию в текстовом виде. Бинарные файлы – это изображения, аудио и видеофайлы и т.д. В зависимости от типа файла работа с ним может немного отличаться.

При работе с файлами необходимо соблюдать некоторую последовательность операций:

1. Открытие файла с помощью метода ***open()***
2. Чтение файла с помощью метода ***read()*** или запись в файл посредством метода ***write()***
3. Закрытие файла методом ***close()***

Чтобы начать работу с файлом, его надо открыть с помощью функции ***open()***, которая имеет следующее формальное определение:

```
open(file, mode)
```

Первый параметр функции представляет путь к файлу. Путь файла может быть абсолютным, то есть начинаться с буквы диска, например, *C://somedir/somefile.txt*. Либо он может быть относительным, например, *somedir/somefile.txt* - в этом случае поиск файла будет идти относительно расположения запущенного скрипта Python.

Второй передаваемый аргумент (***mode***) устанавливает режим открытия файла в зависимости от того, что мы собираемся с ним делать. Существует 4 общих режима:

- **r** (Read). Файл открывается для чтения. Если файл не найден, то генерируется исключение FileNotFoundError
- **w** (Write). Файл открывается для записи. Если файл отсутствует, то он создается. Если подобный файл уже есть, то он создается заново, и соответственно старые данные в нем стираются.
- **a** (Append). Файл открывается для дозаписи. Если файл отсутствует, то он создается. Если подобный файл уже есть, то данные записываются в его конец.

- **b** (Binary). Используется для работы с бинарными файлами. Применяется вместе с другими режимами - w или r.

После завершения работы с файлом его обязательно нужно закрыть методом ***close()***. Данный метод освободит все связанные с файлом используемые ресурсы.

Например, откроем для записи текстовый файл `"hello.txt"`:

```
myfile = open("hello.txt", "w")
myfile.close()
```

При открытии файла или в процессе работы с ним мы можем столкнуться с различными исключениями, например, к нему нет доступа и т.д. В этом случае программа выдаст ошибку, а ее выполнение не дойдет до вызова метода ***close***, и соответственно файл не будет закрыт.

В этом случае мы можем обрабатывать исключения:

```
try:
    somefile = open("hello.txt", "w")
    try:
        somefile.write("hello world")
    except Exception as e:
        print(e)
    finally:
        somefile.close()
except Exception as ex:
    print(ex)
```

В данном случае вся работа с файлом идет во вложенном блоке ***try***. И если вдруг возникнет какое-либо исключение, то в любом случае в блоке ***finally*** файл будет закрыт.

Однако есть и более удобная конструкция - конструкция ***with***:

```
with open(file, mode) as file_obj:
    инструкции
```

Эта конструкция определяет для открытого файла переменную ***file_obj*** и выполняет набор инструкций. После их выполнения файл автоматически закрывается. Даже если при выполнении инструкций в блоке ***with*** возникнут какие-либо исключения, то файл все равно закрывается.

Так, перепишем предыдущий пример:

```
with open("hello.txt", "w") as somefile:
    somefile.write("hello world")
```

Запись в текстовый файл

Чтобы открыть текстовый файл на запись, необходимо применить режим **w** (перезапись) или **a** (дозапись). Затем для записи применяется метод ***write(str)***, в который передается записываемая строка. Стоит отметить, что записывается именно строка, поэтому, если нужно записать числа, данные других типов, то их предварительно нужно конвертировать в строку.

Запишем некоторую информацию в файл `"hello.txt"`:

```
with open("hello.txt", "w") as file:
    file.write("hello world")
```

Если мы откроем папку, в которой находится текущий скрипт ***Python***, то увидим там файл hello.txt. Этот файл можно открыть в любом текстовом редакторе и при желании изменить.

Теперь допишем в этот файл еще одну строку:

```
with open("hello.txt", "a") as file:
    file.write("\ngood bye, world")
```

Дозапись выглядит как добавление строку к последнему символу в файле, поэтому, если необходимо сделать запись с новой строки, то можно использовать эскейп-последовательность `"\n"`. В итоге файл `hello.txt` будет иметь следующее содержимое:

```
hello world
good bye, world
```

Еще один способ записи в файл представляет стандартный метод ***print()***, который применяется для вывода данных на консоль:

```
with open("hello.txt", "a") as hello_file:
    print("Hello, world", file=hello_file)
```

Для вывода данных в файл в метод ***print*** в качестве второго параметра передается название файла через параметр ***file***. А первый параметр представляет записываемую в файл строку.

Чтение файла

Для чтения файла он открывается с режимом **r** (Read), и затем мы можем считать его содержимое различными методами:

- ***readline()***: считывает одну строку из файла
- ***read()***: считывает все содержимое файла в одну строку
- ***readlines()***: считывает все строки файла в список

Например, считаем выше записанный файл построчно:

```
with open("hello.txt", "r") as file:
    for line in file:
        print(line, end="")
```

Несмотря на то, что мы явно не применяем метод ***readline()*** для чтения каждой строки, но в при переборе файла этот метод автоматически вызывается для получения каждой новой строки. Поэтому в цикле вручную нет смысла вызывать метод ***readline***. И поскольку строки разделяются символом перевода строки `"\n"`, то, чтобы исключить излишнего переноса на другую строку, в функцию ***print*** передается значение `end=""`.

Теперь явным образом вызовем метод ***readline()*** для чтения отдельных строк:

```
with open("hello.txt", "r") as file:
    str1 = file.readline()
    print(str1, end="")
    str2 = file.readline()
    print(str2)
```

Консольный вывод:

```
hello world
good bye, world
```

Метод ***readline*** можно использовать для построчного считывания файла в цикле ***while***:

```
with open("hello.txt", "r") as file:
    line = file.readline()
    while line:
        print(line, end="")
        line = file.readline()
```

Если файл небольшой, то его можно разом считать с помощью метода ***read()***:

```
with open("hello.txt", "r") as file:
    content = file.read()
    print(content)
```

И также применим метод ***readlines()*** для считывания всего файла в список строк:

```
with open("hello.txt", "r") as file:
    contents = file.readlines()
    str1 = contents[0]
    str2 = contents[1]
    print(str1, end="")
    print(str2)
```

При чтении файла мы можем столкнуться с тем, что его кодировка не совпадает с ASCII. В этом случае мы явным образом можем указать кодировку с помощью параметра ***encoding***:

```
filename = "hello.txt"
with open(filename, encoding="utf8") as file:
    text = file.read()
```

Теперь напишем небольшой скрипт, который будет записывать введенный пользователем массив строк и считывать его обратно из файла на консоль:

```
# имя файла
FILENAME = "messages.txt"
# определяем пустой список
messages = list()

for i in range(4):
    message = input("Введите строку " + str(i+1) + ": ")
    messages.append(message + "\n")

# запись списка в файл
with open(FILENAME, "a") as file:
    for message in messages:
        file.write(message)

# считываем сообщения из файла
print("Считанные сообщения")
with open(FILENAME, "r") as file:
    for message in file:
        print(message, end="")
```

Пример работы программы:

```
Введите строку 1: hello
Введите строку 2: world peace
Введите строку 3: great job
Введите строку 4: Python
Считанные сообщения
hello
world peace
great job
Python
```

Файлы CSV

Одним из распространенных файловых форматов, которые хранят в удобном виде информацию, является формат **csv**. Каждая строка в файле csv представляет отдельную запись или строку, которая состоит из отдельных столбцов, разделенных запятыми. Собственно, поэтому формат и называется ***Comma Separated Values***. Но хотя формат **csv** – это формат текстовых файлов, Python для упрощения работы с ним предоставляет специальный встроенный модуль ***csv***.

Рассмотрим работу модуля на примере:

```
import csv

FILENAME = "users.csv"

users = [
    ["Tom", 28],
    ["Alice", 23],
    ["Bob", 34]
]

with open(FILENAME, "w", newline="") as file:
    writer = csv.writer(file)
    writer.writerows(users)


with open(FILENAME, "a", newline="") as file:
    user = ["Sam", 31]
    writer = csv.writer(file)
    writer.writerow(user)
```

В файл записывается двумерный список – фактически таблица, где каждая строка представляет одного пользователя. А каждый пользователь содержит два поля: имя и возраст. То есть фактически таблица состоит из трех строк и двух столбцов.

При открытии файла на запись в качестве третьего параметра указывается значение `newline=""` – пустая строка позволяет корректно считывать строки из файла вне зависимости от операционной системы.

Для записи нам надо получить объект ***writer***, который возвращается функцией `csv.writer(file)`. В эту функцию передается открытый файл. А собственно запись производится с помощью метода `writer.writerows(users)`. Этот метод принимает набор строк. В нашем случае это двумерный список.

Если необходимо добавить одну запись, которая представляет собой одномерный список, например, `["Sam", 31]`, то в этом случае можно вызвать метод ***writer.writerow(user)***.

В итоге, после выполнения скрипта в той же папке окажется файл `users.csv`, который будет иметь следующее содержимое:

```
Tom,28
Alice,23
Bob,34
Sam,31
```

Для чтения из файла нам нужно создать объект ***reader***:

```
import csv

FILENAME = "users.csv"

with open(FILENAME, "r", newline="") as file:
    reader = csv.reader(file)
    for row in reader:
        print(row[0], " - ", row[1])
```

При получении объекта reader мы можем в цикле перебрать все его строки:

```
Tom  -  28
Alice  -  23
Bob  -  34
Sam - 31
```

Работа со словарями

В примере выше каждая запись или строка представляла собой отдельный список, например, `["Sam", 31]`. Но кроме того, модуль ***csv*** имеет специальные дополнительные возможности для работы со словарями. В частности, функция ***csv.DictWriter()*** возвращает объект ***writer***, который позволяет записывать в файл. А функция ***csv.DictReader()*** возвращает объект ***reader*** для чтения из файла. Например:

```
import csv

FILENAME = "users.csv"

users = [
    {"name": "Tom", "age": 28},
    {"name": "Alice", "age": 23},
    {"name": "Bob", "age": 34}
]

with open(FILENAME, "w", newline="") as file:
    columns = ["name", "age"]
    writer = csv.DictWriter(file, fieldnames=columns)
    writer.writeheader()

    # запись нескольких строк
    writer.writerows(users)

    user = {"name" : "Sam", "age": 41}
    # запись одной строки
    writer.writerow(user)
```

```
    with open(FILENAME, "r", newline="") as file:
        reader = csv.DictReader(file)
        for row in reader:
            print(row["name"], "-", row["age"])
```

Запись строк также производится с помощью методов ***writerow()*** и ***writerows()***. Но теперь каждая строка представляет собой отдельный словарь, и кроме того, производится запись и заголовков столбцов с помощью метода ***writeheader()***, а в метод ***csv.DictWriter*** в качестве второго параметра передается набор столбцов.

При чтении строк, используя названия столбцов, мы можем обратиться к отдельным значениям внутри строки: `row["name"]`.

Бинарные файлы

Бинарные файлы, в отличие от текстовых, хранят информацию в виде набора байт. Для работы с ними в ***Python*** необходим встроенный модуль ***pickle***. Этот модуль предоставляет два метода:

- ***dump(obj, file)***: записывает объект **obj** в бинарный файл **file**
- ***load(file)***: считывает данные из бинарного файла в объект

При открытии бинарного файла на чтение или запись также надо учитывать, что нам нужно применять режим `"b"` в дополнение к режиму записи (`"w"`) или чтения (`"r"`). Допустим, надо сохранить два объекта:

```
    import pickle

    FILENAME = "user.dat"

    name = "Tom"
    age = 19

    with open(FILENAME, "wb") as file:
        pickle.dump(name, file)
        pickle.dump(age, file)

    with open(FILENAME, "rb") as file:
        name = pickle.load(file)
        age = pickle.load(file)
        print("Имя:", name, "\tВозраст:", age)
```

С помощью функции ***dump*** последовательно записываются два объекта. Поэтому при чтении файла также последовательно посредством функции ***load*** мы можем считать эти объекты. Консольный вывод программы:

```
Имя: Tom        Возраст: 28
```

Подобным образом мы можем сохранять и извлекать из файла наборы объектов:

```
import pickle

FILENAME = "users.dat"

users = [
    ["Tom", 28, True],
    ["Alice", 23, False],
    ["Bob", 34, False]
]

with open(FILENAME, "wb") as file:
    pickle.dump(users, file)


with open(FILENAME, "rb") as file:
    users_from_file = pickle.load(file)
    for user in users_from_file:
        print("Имя:",    user[0],    "\tВозраст:",    user[1],
"\tЖенат(замужем):", user[2])
```

В зависимости от того, какой объект мы записывали функцией ***dump***, тот же объект будет возвращен функцией ***load*** при считывании файла.

Консольный вывод:

```
Имя: Tom        Возраст: 28    Женат(замужем): True
Имя: Alice      Возраст: 23    Женат(замужем): False
Имя: Bob        Возраст: 34    Женат(замужем): False
```

Модуль shelve

Для работы с бинарными файлами в ***Python*** может применяться еще один модуль – ***shelve***. Он сохраняет объекты в файл с определенным ключом. Затем по этому ключу может извлечь ранее сохраненный объект из файла. Процесс работы с данными через модуль ***shelve*** напоминает работу со словарями, которые также используют ключи для сохранения и извлечения объектов.

Для открытия файла модуль ***shelve*** использует функцию ***open()***:

```
open(путь_к_файлу[,flag="c"[,protocol=None[,writeback=False]]])
```

Параметр ***flag*** может принимать следующие значения:

- c: файл открывается для чтения и записи (значение по умолчанию). Если файл не существует, то он создается.
- r: файл открывается только для чтения.
- w: файл открывается для записи.
- n: файл открывается для записи. Если файл не существует, то он создается. Если он существует, то он перезаписывается

Для закрытия подключения к файлу вызывается метод ***close()***:

```
import shelve
d = shelve.open(filename)
d.close()
```

Также можно открывать файл с помощью оператора ***with***. Сохраним и считаем в файл несколько объектов:

```
import shelve

FILENAME = "states2"
with shelve.open(FILENAME) as states:
    states["London"] = "Great Britain"
    states["Paris"] = "France"
    states["Berlin"] = "Germany"
    states["Madrid"] = "Spain"

with shelve.open(FILENAME) as states:
    print(states["London"])
    print(states["Madrid"])
```

Запись данных предполагает установку значения для определенного ключа:

```
states["London"] = "Great Britain"
```

А чтение из файла эквивалентно получению значения по ключу:

```
print(states["London"])
```

В качестве ключей используются строковые значения.

При чтении данных, если запрашиваемый ключ отсутствует, то генерируется исключение. В этом случае перед получением мы можем проверять на наличие ключа с помощью оператора ***in***:

```
    with shelve.open(FILENAME) as states:
        key = "Brussels"
        if key in states:
            print(states[key])
```

Также мы можем использовать метод ***get()***. Первый параметр метода – ключ, по которому следует получить значение, а второй – значение по умолчанию, которое возвращается, если ключ не найден.

```
    with shelve.open(FILENAME) as states:
        state = states.get("Brussels", "Undefined")
        print(state)
```

Используя цикл ***for***, можно перебрать все значения из файла:

```
    with shelve.open(FILENAME) as states:
        for key in states:
            print(key," - ", states[key])
```

Метод ***keys()*** возвращает все ключи из файла, а метод ***values()*** – все значения:

```
    with shelve.open(FILENAME) as states:

        for city in states.keys():
            print(city, end=" ")        # London Paris Berlin Madrid
        print()
        for country in states.values():
            print(country, end=" ")    # Great Britain France Germany
Spain
```

Еще один метод ***items()*** возвращает набор кортежей. Каждый кортеж содержит ключ и значение.

```
    with shelve.open(FILENAME) as states:

        for state in states.items():
            print(state)
```

Консольный вывод:

```
    ("London", "Great Britain")
    ("Paris", "France")
    ("Berlin", "Germany")
    ("Madrid", "Spain")
```

Обновление данных

Для изменения данных достаточно присвоить по ключу новое значение, а для добавления данных – определить новый ключ:

```
import shelve

FILENAME = "states2"
with shelve.open(FILENAME) as states:
    states["London"] = "Great Britain"
    states["Paris"] = "France"
    states["Berlin"] = "Germany"
    states["Madrid"] = "Spain"

with shelve.open(FILENAME) as states:

    states["London"] = "United Kingdom"
    states["Brussels"] = "Belgium"
    for key in states:
        print(key, " - ", states[key])
```

Удаление данных

Для удаления с одновременным получением можно использовать функцию ***pop()***, в которую передается ключ элемента и значение по умолчанию, если ключ не найден:

```
with shelve.open(FILENAME) as states:

    state = states.pop("London", "NotFound")
    print(state)
```

Также для удаления может применяться оператор ***del***:

```
with shelve.open(FILENAME) as states:
    del states["Madrid"]    # удаляем объект с ключом Madrid
```

Для удаления всех элементов можно использовать метод ***clear()***:

```
with shelve.open(FILENAME) as states:

    states.clear()
```

Модуль OS и работа с файловой системой

Ряд возможностей по работе с каталогами и файлами предоставляет встроенный модуль ***os***. Хотя он содержит много функций, рассмотрим только основные из них:

- mkdir(): создает новую папку
- rmdir(): удаляет папку
- rename(): переименовывает файл
- remove(): удаляет файл

Создание и удаление папки

Для создания папки применяется функция ***mkdir()***, в которую передается путь к создаваемой папке:

```
import os

# путь относительно текущего скрипта
os.mkdir("hello")
# абсолютный путь
os.mkdir("c://somedir")
os.mkdir("c://somedir/hello")
```

Для удаления папки используется функция ***rmdir()***, в которую передается путь к удаляемой папке:

```
import os

# путь относительно текущего скрипта
os.rmdir("hello")
# абсолютный путь
os.rmdir("c://somedir/hello")
```

Переименование файла

Для переименования вызывается функция ***rename(source, target)***, первый параметр которой – путь к исходному файлу, а второй – новое имя файла. В качестве путей могут использоваться как абсолютные, так и относительные. Например, пусть в папке *C://SomeDir/* располагается файл ***somefile.txt***. Переименуем его в файл ***"hello.txt"***:

```
import os

os.rename("C://SomeDir/somefile.txt", "C://SomeDir/hello.txt")
```

Удаление файла

Для удаления вызывается функция ***remove()***, в которую передается путь к файлу:

```
import os

os.remove("C://SomeDir/hello.txt")
```

Существование файла

Если мы попытаемся открыть файл, который не существует, то ***Python*** выбросит исключение ***FileNotFoundError***. Для отлова исключения мы можем использовать конструкцию `try...except` (*см. теорию Л.Р. №6*). Однако можно уже до открытия файла проверить, существует ли он или нет с помощью метода ***os.path.exists(path)***. В этот метод передается путь, который необходимо проверить:

```
filename = input("Введите путь к файлу: ")
if os.path.exists(filename):
    print("Указанный файл существует")
else:
    print("Файл не существует")
```

Задания

Общее задание

1. Изучите теорию и ответьте на вопросы по теме лабораторной работы.
2. Во всех заданиях необходимо проверять корректность вводимых данных, возможность чтения и записи файлов и выводить соответствующие сообщения об ошибках.

№ Варианта = (**№ПК** + 1) % 2 + 1

Вариант №1

Задание №1. Считать из файла ***input.txt*** 10 чисел (числа записаны через пробел). Затем записать их произведение в файл ***output.txt***.

Задание №2. В файле записаны сведения о студентах в формате:

```
Фамилия;Имя;Год г.р.
Иванов;Иван;2001 г.р.
...
Сидоров;Пётр;2000 г.р.
```

Необходимо записать в текстовый файл фамилии самого старшего и самого младшего студента (если их несколько, вывести ту фамилию, которая находится выше).

Задание №3. Вывести все файлы (*полный путь до файла*) с расширением ***.py*** внутри текущего проекта (*включая вложенные директории*).

Задание №4. Даны два текстовых файла (***input_1.txt***, ***input_2.txt***), необходимо записать в файл ***output.txt*** все слова, которые встречаются в обоих файлах.

Вариант №2

Задание №1. В файле записаны целые числа. Найти максимальное и минимальное число и записать в другой файл.

Задание №2. В файле записаны сведения о товарах в формате:

```
Артикул;Наименование;Цена
420250;SSD накопитель KINGSTON;1700
...
1154016; Ноутбук ACER Aspire;22140
```

Необходимо записать в текстовый файл артикул самого дорогого и самого дешевого товара (если их несколько, вывести тот артикул, который находится ниже).

Задание №3. Вывести все файлы (*полный путь до файла*) размером менее 1МБ, внутри текущего проекта (*включая вложенные директории*).

Задание №4. Даны два текстовых файла (***input_1.txt***, ***input_2.txt***), необходимо записать в файл ***output.txt*** все слова, которые встречаются в файле ***input_2.txt***, но не встречаются в файле ***input_1.txt***.

Дополнительное задание

Дан текстовый файл ***input.txt***. Определить частоту повторяемости каждой буквы в тексте, отсортировать в порядке убывания частоты, результат записать в файл ***output.txt***. Продемонстрировать на разных примерах (русскоязычные и англоязычные текстовые файлы различной длины).