ТЕХНОАТОМ

Работа с файлами и каталогами

Лекция 10

Алена Елизарова

Лекция 10. Работа с файлами и каталогами

ТЕХНОАТОМ

1. Модуль os и его расширения
2. Операции с файлами
3. Файловый объект
4. Сериализация данных
5. Высокоуровневые инструменты для работы с файлами
6. Архивы, изображения, таблицы и другие форматы данных

Типы операции с файлами

ТЕХНОАТОМ

1) связанные с его открытием: открытие, закрытие файла, запись, чтение, перемещение по файлу и др.

2) выполняющиеся без его открытия: работа с файлом как элементом файловой системы - переименование, копирование, получение атрибутов и др.

Модуль os

ТЕХНОАТОМ

Bash vs. Python?

https://stackoverflow.com/questions/2424921/python-vs-bash-in-which-kind-of-tasks-each-one-outruns-the-other-performance-w

Обзор методов
https://docs.python.org/3/library/os.html

Модуль os

ТЕХНОАТОМ

```
os.environ

os.chdir(path)

os.listdir(path=".")

os.mkdir(path, mode=0o777, *, dir_fd=None)

os.walk(top, topdown=True, onerror=None, followlinks=False)

os.path # является вложенным модулем в модуль os

os.path.dirname(path)

os.path.getsize(path)

os.path.join(path1[, path2[, ...]])

os.stat()
```

Модуль os. Пример

ТЕХНОАТОМ

```
import os

updated_file_map = {}
path = '.'

for name in os.listdir(path):
     fullname = os.path.join(path, name)
     if os.path.isfile(fullname):
          t = datetime.fromtimestamp(os.path.getmtime(fullname))
          updated_file_map[fullname] = t
```

Файловый объект

При открытии файла операционная система возвращает специальный дескриптор файла (идентификатор), однозначно определяющий, с каким файлом далее будут выполняться операции.

В Python работа с файлами осуществляется через специальный абстрактный файловый объект. В зависимости от способа создания такого объекта, он может быть привязан как к физическому файлу на диске, так и другому устройству, поддерживающему схожие операции (стандартный ввод/вывод и пр.).

```
open(file, mode='r', buffering=-1, encoding=None, errors=None,
newline=None, closefd=True, opener=None)

# кодировка
import locale
locale.getpreferredencoding(False)
```

Файловый объект

ТЕХНОАТОМ

```
f = open('test.txt')
data = f.read()

print(*f)
print(f)
```

Файловый объект

ТЕХНОАТОМ

```
with open('test.txt') as f:
    f.read()
```

Обработка файлов

ТЕХНОАТОМ

Режим	Описание
r	Только для чтения.
w	Только для записи. Создаст новый файл, если не найдет с указанным именем.
rb	Только для чтения (бинарный).
wb	Только для записи (бинарный). Создаст новый файл, если не найдет с указанным именем.
r+	Для чтения и записи.
rb+	Для чтения и записи (бинарный).
w+	Для чтения и записи. Создаст новый файл для записи, если не найдет с указанным именем.
wb+	Для чтения и записи (бинарный). Создаст новый файл для записи, если не найдет с указанным именем.
a	Откроет для добавления нового содержимого. Создаст новый файл для записи, если не найдет с указанным именем.
a+	Откроет для добавления нового содержимого. Создаст новый файл для чтения записи, если не найдет с указанным именем.
ab	Откроет для добавления нового содержимого (бинарный). Создаст новый файл для записи, если не найдет с указанным именем.
ab+	Откроет для добавления нового содержимого (бинарный). Создаст новый файл для чтения записи, если не найдет с указанным именем.

Последовательный и произвольный доступ

ТЕХНОАТОМ

```
seek(offset[, whence])

offset - смещение в байтах относительно начала файла;
whence - по умолчанию равен нулю, указывает на то, что смещение
берется относительно начала файла.

f = open('my_file', 'w')
f.write('01234567890123456789')
f.seek(5)
f.write('Hello, World!')
f.close()
f = open('my_file')
f.read()
>>>'01234Hello, World!89'
```

Методы файлового объекта

ТЕХНОАТОМ

```
file.close()
file.fileno()
file.flush()
file.isatty()
file.next()
file.read(n)
file.readline()
file.readlines()
file.seekable()
file.tell()
file.truncate(n)
file.write(str)
file.writelines(sequence)
```

Pipe

TEXHOATOM

```
# test.py

import sys
text = sys.stdin.read()
words = text.split()
count = len(words)
print(f'There are {count} words')
```

```
cat some_file | python test.py
```

Сериализация и десериализация

ТЕХНОАТОМ

Сериализация — процесс перевода какой-либо структуры данных в последовательность битов. Десериализация — обратный процесс.

Чаще всего сериализация используется для сохранения объектов в файлы или передачи их по сети.

```
pickle.dump(obj, file, protocol=None, *, fix_imports=True)

pickle.load(file, *, fix_imports=True, encoding="ASCII", errors="strict")
```

Минусы формата pickle

ТЕХНОАТОМ

- специфичен для Python (не может быть использован, если файл будет читаться с использованием других языков программирования);

- небезопасен (десериализация готовых конструкций языка может привести к выполнению ненадежного кода).

Высокоуровневые библиотеки для работы с файлами

ТЕХНОАТОМ

shutil
https://docs.python.org/3/library/shutil.html

скопировать дерево, переместить/переименовать дерево, удалить дерево

glob

```
import glob
glob.glob('examples/*.xml')
['examples\\feed-broken.xml', 'examples\\feed-ns0.xml', 'examples\\feed.xml']
```

Wildcard	Matches	Example
*	any characters	**.txt* matches all files with the txt extension
?	any one character	*???* matches files with 3 characters long
[]	any character listed in the brackets	*[ABC]** matches files starting with A,B or C
[..]	any character in the range listed in brackets	*[A..Z]** matches files starting with capital letters
[!]	any character listed in the brackets	*[!ABC]** matches files that do not start with A,B or C

Высокоуровневые библиотеки для работы с файлами

ТЕХНОАТОМ

tempfile

создание временных файлов и каталогов

filecmp

сравнение файлов

filecmp.cmpfiles()

mimetypes
https://ru.wikipedia.org/wiki/%D0%A1%D0%BF%D0%B8%D1%81%D0%BE%D0%BA_MIME-%D1%82%D0%B8%D0%BF%D0%BE%D0%B2
работа с mime-типами

mimetypes.guess_type()

Архивы

ТЕХНОАТОМ

gzip (.gz)

tarfile (.tar, .tar.gz (а также .tgz) и .tar.bz2)

zipfile (.zip)

Работа с аудио

ТЕХНОАТОМ

aifc
поддержка формата AIFF (Audio Interchange File Format – формат файлов для обмена аудиоданными)

wave
возможность для работы с файлами .wav (несжатыми)

audioop
манипуляция некоторыми разновидностями аудиоданных

sndhdr
определить тип аудиоданных, хранящихся в файле, и некоторые характеристики этих данных, например, частота дискретизации

Работа с изображениями

ТЕХНОАТОМ

```
from PIL import Image
img = Image.open("some.jpg")

img.size
.format
.show()
.crop(...)
.tranpose(...)
```

Популярные форматы данных

ТЕХНОАТОМ

CSV (англ. Comma-Separated Values - значения, разделенные запятыми);

JSON (англ. JavaScript Object Notation) - текстовый формат обмена данными, основанный на JavaScript;

XML (англ. eXtensible Markup Language - расширяемый язык разметки);

YAML (англ. YAML Ain’t Markup Language - «YAML - Не язык разметки»);

Популярные форматы данных. CSV

ТЕХНОАТОМ

```
import csv

f_name = "my_file.csv"
# список покупок
shop_list = {"киви": [2, 100], "томат": [3, 250], "тыква": [1,
35]}

# Запись в файл
with open(f_name, "w", encoding="utf-8", newline="") as f:
    writer = csv.writer(f, quoting=csv.QUOTE_ALL)
    writer.writerow(["Название", "Вес", "Цена/кг."]) # заголовки
    for name, values in sorted(shop_list.items()):
        writer.writerow([name, *values])
    writer.writerow(["огурец", "4", "170"])  # добавим
```

Excel

TEXHOATOM

```
import xlwt

def generate_book(data, encoding='utf8'):
    book = xlwt.Workbook(encoding=encoding)
    for datasheet in data:
        data = datasheet['data']
        sheet = book.add_sheet(datasheet['name'])
        for rowx, row in enumerate(data):
            for colx, value in enumerate(row):
                sheet.write(rowx, colx, value)
    return book
```

Excel

TEXHOATOM

```
def save_excel_to_file(data, filename='data.xls'):
    res_file = open(filename, 'w')
    book = generate_book(data, encoding='utf8')
    book.save(res_file)
```

Домашнее задание

- Написать функцию для подсчета страниц в pdf-файле (используя стандартную библиотеку)
- Написать программу, куда вводится путь к каталогу, рекурсивно обрабатывается информация о содержимом (ссылки не учитываются) и на выходе создается xls-таблица с данными (название, директория или файл, размер, дата изменения, уровень вложенности, полный абсолютный путь)

Домашнее задание

Информация о занятости мест в театре хранится в текстовом файле:

0 0 1 1 0
0 0 0 1 1
1 1 0 1 0

где строка обозначает ряд, столбец - место (0 - свободно, 1 - занято).

- Написать программу, которая выводит количество свободных мест, а также, введя номер ряда и места, показывает, свободно оно или нет

ТЕХНОАТОМ

Спасибо за внимание!