



---

**МОСКОВСКИЙ АВТОМОБИЛЬНО-ДОРОЖНЫЙ ГОСУДАРСТВЕННЫЙ ТЕХНИЧЕСКИЙ  
УНИВЕРСИТЕТ (МАДИ)**

Кафедра "Высшая математика"

**Лабораторная работа №1**

По дисциплине КОМПЬЮТЕРНАЯ ГРАФИКА

на тему:

"Форматы данных и задача парсинга"

**Выполнил:**

Серов Павел Геннадьевич

Группа 36ПМ

Вариант 14

**Преподаватель:**

Кутейников И.А

**1. Цель**

Изучить библиотеки и инструменты для парсинга страниц, обработки html файлов и данных.

**2. Исходные данные**

html-файл с кафедральным расписанием

|               |                   |              |        |                       |                                |
|---------------|-------------------|--------------|--------|-----------------------|--------------------------------|
| Алероев М.Т.  |                   |              |        |                       |                                |
| Вторник       |                   |              |        |                       |                                |
| Время         | Занятия на неделе | Аудитория    | Группа | Дисциплина            | Тип занятий                    |
| 09:55 - 11:25 | Еженедельно       | Дистанционно | 1лДМ   | Математика            | Лекции                         |
| 11:35 - 13:05 | Знаменатель       | Дистанционно | 1лДМ   | Математика            | Практические занятия /семинар/ |
| Четверг       |                   |              |        |                       |                                |
| Время         | Занятия на неделе | Аудитория    | Группа | Дисциплина            | Тип занятий                    |
| 09:55 - 11:25 | Еженедельно       | 536          | 1лДМ   | Математика            | Практические занятия /семинар/ |
| Пятница       |                   |              |        |                       |                                |
| Время         | Занятия на неделе | Аудитория    | Группа | Дисциплина            | Тип занятий                    |
| 09:55 - 11:25 | Еженедельно       | 539          | 3бПМ   | Функциональный анализ | Практические занятия /семинар/ |
| Буслаев Д.А.  |                   |              |        |                       |                                |
| Четверг       |                   |              |        |                       |                                |
| Время         | Занятия на неделе | Аудитория    | Группа | Дисциплина            | Тип занятий                    |
| 08:15 - 09:45 | Числитель         | 443          | 2бАМТ1 | Математика            | Практические занятия /семинар/ |
| 09:55 - 11:25 | Числитель         | 443          | 2бАМТ2 | Математика            | Практические занятия /семинар/ |

3. Задание

Считать данные и сформировать текстовый файл по ключу:

|                   |             |       |            |        |             |           |                   |
|-------------------|-------------|-------|------------|--------|-------------|-----------|-------------------|
| ФИО преподавателя | День недели | Время | Дисциплина | Группа | Тип занятия | Аудитория | Занятия на неделе |
|-------------------|-------------|-------|------------|--------|-------------|-----------|-------------------|

С помощью собранных данных заполнить шаблон расписания в формате .xlsx.

|         |                |       |                               |  |                                |                              |                                |                          |                      |                       |                              |
|---------|----------------|-------|-------------------------------|--|--------------------------------|------------------------------|--------------------------------|--------------------------|----------------------|-----------------------|------------------------------|
|         |                |       |                               |  |                                |                              |                                |                          |                      |                       |                              |
| ПОДПИСЬ | Дата           | Время | Зав. кафедрой<br>Дашков И.В.  | Старпр.<br>Венгачев В.В.                         | Старпр.<br>Григорьев С.А.      | Асс.<br>Буслаев Д.А.         | Старпр.<br>Девя Н.В.           | Старпр.<br>Долгушин А.С. | Докент<br>Жуков Д.А. | Докент<br>Зинков А.А. | Докент<br>Шаров С.А.         |
|         | 08:15<br>09:45 |       |                               |  |                                |                              |                                |                          |                      |                       |                              |
|         | 09:55<br>11:25 |       |                               |  |                                |                              |                                |                          |                      |                       |                              |
|         | 11:35<br>13:05 |       |                               |  |                                |                              |                                |                          |                      |                       |                              |
|         | 13:10<br>13:50 |       |                               | Высшая математика<br>14ДВС, 14АМ,<br>14КС, 14АК1 |                                |                              | 1114/1111<br>14ДМ1, 2, 4, 441  |                          |                      |                       |                              |
|         | 13:55<br>15:25 |       |                               |  |                                | Математика<br>14ДВ, 444, 443 | 1114/1111<br>14ДМ1, 444, 441   |                          | Математика<br>14Д1-3 |                       |                              |
|         | 15:30<br>17:00 |       |                               |  |                                | Математика<br>14СА, 444, 443 |                                |                          | Математика<br>14Д1   |                       |                              |
|         | 17:15<br>18:45 |       |                               |  |                                |                              |                                |                          |                      |                       |                              |
|         | 18:50<br>20:20 |       |                               |  |                                |                              |                                |                          |                      |                       |                              |
|         | 20:30<br>22:00 |       |                               |  |                                |                              |                                |                          |                      |                       |                              |
|         | 08:15<br>09:45 |       |                               |  | Математика<br>14АМТ1, 444, 443 |                              |                                |                          |                      |                       | Математика<br>14КС12         |
|         | 09:55<br>11:25 |       | Математика<br>14ДМ1, 14ДМ2    | Высшая математика<br>14АК1, 444, 443             |                                | Математика<br>14Д4, 444, 114 |                                |                          |                      |                       | Математика<br>14КС15         |
|         | 11:35<br>13:05 |       | Математика<br>14ДМ1, 444, 443 |  |                                | Математика<br>14Д1, 444, 114 | Математика<br>14АК1, 444, 9014 |                          |                      |                       | Математика<br>14ДВ, 444, 111 |

4. Ход решения

Импортируем библиотеки

```
import pandas as pd
from openpyxl import load_workbook
from copy import copy
import codecs
```

Так же заранее подготовим шаблон excel файла со всеми названиями строк и столбцов, но без данных.

```
answerFileName = 'test.xlsx'
templateFileName = 'kek.xlsx'
htmlFileName = ""
```

| День  | Время          | Зав. кафедрой<br>Яшина М.В. | Ст.преп.<br>Величкин В.Н. | Ст.преп.<br>Гильдерман С.А. | Асс.<br>Буслаев Д.А. | Ст.преп.<br>Деза Н.В. | Ст.преп.<br>Доткулова А.С. | Доцент<br>Жуков Д.А. | Доцент<br>Зленко А.А. | Доцент<br>Изотова С.А. |  |
|---|----------------|-----------------------------|---------------------------|-----------------------------|----------------------|-----------------------|----------------------------|----------------------|-----------------------|------------------------|--|
| П<br>О<br>Н<br>Е<br>Д<br>Е<br>Л<br>Ь<br>Н<br>И<br>К | 12:20<br>13:50 |                             |                           |                             |                      |                       |                            |                      |                       |                        |  |
|   | 13:55<br>15:25 |                             |                           |                             |                      |                       |                            |                      |                       |                        |  |
|   | 15:35<br>17:05 |                             |                           |                             |                      |                       |                            |                      |                       |                        |  |
|   | 17:15<br>18:45 |                             |                           |                             |                      |                       |                            |                      |                       |                        |  |
|   | 18:50<br>20:20 |                             |                           |                             |                      |                       |                            |                      |                       |                        |  |
|   |                |                             |                           |                             |                      |                       |                            |                      |                       |                        |  |
|   |                |                             |                           |                             |                      |                       |                            |                      |                       |                        |  |
|   |                |                             |                           |                             |                      |                       |                            |                      |                       |                        |  |
|   |                |                             |                           |                             |                      |                       |                            |                      |                       |                        |  |
|   |                |                             |                           |                             |                      |                       |                            |                      |                       |                        |  |
|   |                |                             |                           |                             |                      |                       |                            |                      |                       |                        |  |

Создадим классы:

"Lesson" - для обработки информации о предметах

"Person" - для обработки информации о преподавателях

```
class Lesson(object):
    def __init__(self):
        self.dayOfTheWeak = ""
        self.time = ""
        self.lessonName = ""
        self.frequency = ""
        self.group = ""
        self.type = ""
        self.classNumber = ""
```

```
class Person(object):
    def __init__(self, name):
        self.name = name
        self.lessons = []
```

## Импортируем таблицу из файла "html"

```
table = pd.read_html('2.html', encoding="utf8")[1]
#print(table[0])
persons = []
numberOfRowsWith1Column = 0
numberOfRowsWithSomeColumns = 0
lastPerson = Person("name")
lastDayOfTheWeek = ""
```

table

|     | 0             | 1                 | 2            | 3            | 4                          |    |
|-----|---------------|-------------------|--------------|--------------|----------------------------|----|
| 0   | Алероев М.Т.  | Алероев М.Т.      | Алероев М.Т. | Алероев М.Т. | Алероев М.Т.               |    |
| 1   | Вторник       | Вторник           | Вторник      | Вторник      | Вторник                    |    |
| 2   | Время         | Занятия на неделе | Аудитория    | Группа       | Дисциплина                 |    |
| 3   | 09:55 - 11:25 | Еженедельно       | Дистанционно | 1пДМ         | Математика                 |    |
| 4   | 11:35 - 13:05 | Знаменатель       | Дистанционно | 1пДМ         | Математика                 | Пр |
| ... | ...           | ...               | ...          | ...          | ...                        |    |
| 439 | 11:35 - 13:05 | Числитель         | 518          | 26ПМ2        | Дифференциальные уравнения |    |
| 440 | Четверг       | Четверг           | Четверг      | Четверг      | Четверг                    |    |
| 441 | Время         | Занятия на неделе | Аудитория    | Группа       | Дисциплина                 |    |

## Выбираем лист в файле эксель, куда сохранять таблицу

```
wb = load_workbook(templateFileName)
ws = wb['Лист1']
wb.save(answerFileName)
```

## Парсим из html в список persons с lessons

```
for rowIndex in range(0, len(table[0])):
    if table[0][rowIndex] == table[1][rowIndex]:
        numberOfRowsWith1Column += 1
        numberOfRowsWithSomeColumns = 0
    else:
        if numberOfRowsWith1Column == 1:
            lastDayOfTheWeek = table[0][rowIndex - 1]
        if numberOfRowsWith1Column == 2:
```

```

        lastPerson = Person(table[0][rowIndex - 2])
        persons.append(lastPerson)
        lastDayOfTheWeak = table[0][rowIndex - 1]
        numbersOfRowsWith1Column = 0

    numbersOfRowsWithSomeColumns += 1
    if numbersOfRowsWithSomeColumns > 1:
        newLesson = Lesson()
        newLesson.time = table[0][rowIndex]
        newLesson.frequency = table[1][rowIndex]
        newLesson.classNumber = table[2][rowIndex]
        newLesson.group = table[3][rowIndex]
        newLesson.lessonName = table[4][rowIndex]
        newLesson.type = table[5][rowIndex]
        newLesson.dayOfTheWeak = lastDayOfTheWeak
        lastPerson.lessons.append(newLesson)

```

Создадим метод, который позволяет разбить ячейки в зависимости от частоты занятий:  
Еженедельно, Числитель, Знаменатель

```

def merged_cells(row, column, sheet, typeMerged):
    cell1 = sheet.cell(row, column)
    cell2 = sheet.cell(row + 1, column)
    unmergedAll = False
    for mergedCell in sheet.merged_cells.ranges:
        if cell1.coordinate in mergedCell and cell2.coordinate in mergedCell:
            if typeMerged == 1:
                return
            else:
                unmergedAll = True
                sheet.unmerge_cells(start_row=row, start_column=column, end_row=row + 1, end_c

    if not unmergedAll:
        cell1 = sheet.cell(row, column)
        cell2 = sheet.cell(row, column + 1)
        for mergedCell in sheet.merged_cells.ranges:
            if cell1.coordinate in mergedCell and cell2.coordinate in mergedCell:
                sheet.unmerge_cells(start_row=row, start_column=column, end_row=row, end_c
                break

        cell1 = sheet.cell(row + 1, column)
        cell2 = sheet.cell(row + 1, column + 1)
        for mergedCell in sheet.merged_cells.ranges:
            if cell1.coordinate in mergedCell and cell2.coordinate in mergedCell:
                sheet.unmerge_cells(start_row=row + 1, start_column=column, end_row=row +
                break

    if typeMerged == 1:
        sheet.merge_cells(start_row=row, start_column=column, end_row=row + 1, end_column=
    else:
        sheet.cell(row + 1, column).fill = copy(sheet.cell(row, column).fill)
        sheet.cell(row + 1, column).font = copy(sheet.cell(row, column).font)

```

```

sheet.cell(row + 1, column).border = copy(sheet.cell(row, column).border)
sheet.merge_cells(start_row=row, start_column=column, end_row=row, end_column=column)
sheet.merge_cells(start_row=row + 1, start_column=column, end_row=row + 1, end_column=column)

```

сохраняем в excel файл

```

for person in persons:
    #номер колонки с нужным преподавателем
    columnNumberPerson = -1
    for column in range(1, ws.max_column):
        if not ws.cell(row=3, column=column).value is None and person.name in ws.cell(row=
            columnNumberPerson = column

    if columnNumberPerson == -1:
        personNameWithoutIO = person.name
        for i in personNameWithoutIO:
            if not i.isalpha() and i != ' ':
                personNameWithoutIO = personNameWithoutIO.replace(i, '')

        personNameWithoutIO = personNameWithoutIO.split()[0].strip() + ' '
        for column in range(1, ws.max_column):
            if not ws.cell(row=3, column=column).value is None and personNameWithoutIO in
                columnNumberPerson = column

    if columnNumberPerson == -1:
        #print(person.name + " NOT FOUND")
        continue

    # сохраняем пары для преподавателя
    for lesson in person.lessons:
        rowNumberDayOfWeak = -1
        # находим строку с днем недели
        for rowNumber in range(1, ws.max_row):
            rowValue = ws.cell(row=rowNumber, column=1).value
            if not rowValue is None:
                rowValue = rowValue.replace('\n', "")
                if rowValue.lower() == lesson.dayOfTheWeak.replace('\n', "").lower():
                    rowNumberDayOfWeak = rowNumber

        if rowNumberDayOfWeak == -1:
            print("DAY NOT FOUND")

        # находим строку со временем начала занятий
        rowNumberTime = -1
        for rowNumber in range(rowNumberDayOfWeak, ws.max_row):
            rowValue = ws.cell(row=rowNumber, column=2).value
            if not rowValue is None:
                if rowValue[:5] == lesson.time[:5]:
                    rowNumberTime = rowNumber
                    break

        if rowNumberTime == -1:
            print("TIME NOT FOUND")

```

```

# разбиваем или объединяем ячейки
if lesson.frequency == "Еженедельно":
    merged_cells(rowNumberTime, columnNumberPerson, ws, 1)
else:
    merged_cells(rowNumberTime, columnNumberPerson, ws, 2)

if lesson.frequency == "Знаменатель":
    rowNumberTime += 1

ws.cell(row=rowNumberTime, column=columnNumberPerson).value = lesson.lessonName +
if not lesson.classNumber is None:
    ws.cell(row=rowNumberTime, column=columnNumberPerson).value = ws.cell(row=rowN
column=c

```

Сохраняем изменения в эксель файле

```
wb.save(templateFileName)
```

| День  | Время          | Зав. кафедрой<br>Яшина М.В. | Ст.преп.<br>Величкин В.Н. | Ст.преп.<br>Гильдерман С.А.                         | Асс.<br>Буслаев Д.А. | Ст.преп.<br>Дежа Н.В.        | Ст.преп.<br>Доткулова А.С.   | Доцент<br>Жуков Д.А. | Доцент<br>Зленко А.А.                    | Доцент<br>Изотова С.А. | Кл |
|---|----------------|-----------------------------|---------------------------|---|----------------------|------------------------------|--|----------------------|--|------------------------|----|
| П<br>О<br>Н<br>Е<br>Д<br>Е<br>Л<br>Ь<br>Н<br>И<br>К | 08:15<br>09:45 |                             |                           |   |                      |                              |  |                      |  |                        |    |
|   | 09:55<br>11:25 |                             |                           |   |                      |                              |  |                      |  |                        |    |
|   | 11:35<br>13:05 |                             |                           |   |                      |                              |  |                      |  |                        |    |
|   | 12:20<br>13:50 |                             |                           | Высшая<br>математика<br>16АЭ2, ауд.<br>Дистанционно |                      |                              | Прикладное<br>программирование   |                      |  |                        |    |
|   | 13:55<br>15:25 |                             |                           |   |                      | Математика<br>16Д6, ауд. 443 | Прикладное<br>программирование<br>и пакеты программ<br>16ПМ1, ауд. 441 |                      | Математика<br>16Д3, ауд.<br>Дистанционно |                        |    |
|   | 15:35<br>17:05 |                             |                           |   |                      | Математика<br>16СА, ауд. 443 |  |                      | Математика<br>16Д1, ауд.<br>Дистанционно |                        |    |
|   |                |                             |                           |   |                      |                              |  |                      |  |                        |    |

## 5. Вывод

Изучили библиотеки и инструменты для парсинга страниц, обработки html файлов и данных. Научились работать с таблицами в python.