

## Импорт данных из файлов MS Excel с помощью языка программирования python

*Кизянов Антон Олегович*

*Приамурский государственный университет имени Шолом-Алейхема  
Студент*

### **Аннотация**

Практическое создание программы на языке программирования Python для извлечения данных из электронных таблиц MS Excel.

**Ключевые слова:** Python, xlrd, Excel

### **Importing data from MS Excel files using the python programming language**

*Kizyanov Anton Olegovich*

*Sholom-Aleichem Priamursky State University  
student*

### **Abstract**

Practical creation of programs in the programming language Python for extracting data from MS Excel spreadsheets.

**Keywords:** Python, xlrd, Excel

MS Excel поддерживает диаграммы, иногда нужна более гибкая и мощная визуализация, и необходимо экспортировать данные из существующих электронных таблиц в Python для дальнейшего использования в программах визуализации.

Цель исследования – это написание программы для экспорта информации из электронных таблиц Microsoft Excel.

Ранее этим вопросом интересовались П.В. Арбузов, С.Б. Ревин, А.С. Рыльских, О.С. Ревина с темой «Экспорт графических данных из СУБД MS Access в электронную таблицу MS Excel» [1], а подробнее про то как передавать данные из базы данных в MS Access с подробным описанием возможных ошибок. Хорст-Дитер Радке развивал тему «Практическое использование Microsoft Excel для обобщения статистических данных и их презентации. подготовка и презентация статистических данных в Microsoft Excel» [2] в которой рассказывается как создавать отчеты и презентации из отчетов Excel. А. Левин опубликовал «Excel, включая Excel 2010 - это очень просто!» [3] рассказал насколько работа с MS Excel может быть простой когда их обучить ему.

Общим подходом к импорту данных из файлов Excel является экспорт данных из Excel в файлы в формате CSV. Это довольно простой процесс, если у нас есть один или два файла (и установлены Microsoft Excel или

OpenOffice.org), но, если мы автоматизируем канал данных для большого количества файлов мы не в состоянии вручную конвертировать каждый файл Excel в CSV. Итак, нам нужен способ прочитать любой файл Excel.

Python имеет достойную поддержку для чтения и записи файлов Excel через проект [www.python-excel.org](http://www.python-excel.org). Эта поддержка доступна в виде различных модулей для чтения и записи и не зависит от платформы. Другими словами, нам не нужно устанавливать Excel чтобы запускать его в Windows, чтобы читать файлы Excel.

Формат файлов MS Excel со временем изменился, а поддержка разных версий доступна в разных библиотеках Python. Последняя стабильная версия XLRD составляет 0,9 на момент написания этой статьи, и она поддерживает чтение файлов .xlsx.

Во-первых, нам нужно установить необходимый модуль. В этом примере мы будем использовать модуль xlrd и язык программирования Python.

```
$ pip install xlrd
```

Следующий пример кода демонстрирует, как читать данные из файла Excel.

Делается это в несколько шагов:

1. Открытие рабочей книги.
2. Найти лист по имени.
3. Прочитайте ячейки, используя количество строк (строк) и столбцов.

Здесь представлен полный текст программы:

```
import xlrd  
from pprint import pprint  
file = 'Файл.xlsx'  
wb = xlrd.open_workbook(filename=file)  
ws = wb.sheet_by_name('Таблица')  
dataset = []  
for r in xrange(ws.nrows):  
    col = []  
    for c in range(ws.ncols):  
        col.append(ws.cell(r, c).value)  
    dataset.append(col)  
pprint(dataset)
```

Попробуем объяснить простую объектную модель, используемую xlrd. На верхнем уровне есть книга (класс Python `xlrd.book.Book`), которая состоит из одного или нескольких листов (`xlrd.sheet.Sheet`), и на каждом листе есть ячейка (`xlrd.sheet.Cell`), из которой мы затем можем прочитать значение.

Мы загружаем книгу из файла с помощью `open_workbook ()`, которая возвращает `xlrd.book`. Экземпляр книги, содержащий всю информацию о книге, например, листы. Мы получаем доступ к листам с использованием `sheet_by_name ()`; Если нам нужны все листы, мы могли бы использовать `sheet ()`, который возвращает список экземпляров `xlrd.sheet.Sheet`. Класс `xlrd.sheet.Sheet` имеет несколько столбцов и строк в качестве атрибутов, которые мы можем использовать для вывода диапазонов для нашего цикла для доступа к каждой конкретной ячейке внутри листа с помощью метода `cell()`.

#### Вывод

Таким образом, задача, связанная с экспортом данных из файлов MS Excel решается с помощью применения вышенаписанного скрипта.

#### Библиографический список

1. Арбузов П.В., Ревин С.Б., Рыльских А.С., Ревина О.С. Экспорт графических данных из СУБД MS Access в электронную таблицу MS Excel // Математические методы и информационно-технические средства 2010. С. 13-16.
2. Радке Х.-Д. Практическое использование Microsoft Excel для обобщения статистических данных и их презентации. подготовка и презентация статистических данных в Microsoft Excel: учебное пособие. М.: NTPress, 2008. 271 с.
3. Левин А. Excel, включая Excel 2010 - это очень просто! СПб.: Питер, 2011. 109 с.