

Разработка программы для склейки PDF файлов на основе языка программирования Python

Киселева Елизавета Александровна

Приамурский государственный университет им. Шолом-Алейхема

Студент

Аннотация

Зачастую сканируемые документы, письма, различные справки и важные бумаги отправлять по отдельности не очень удобно. К сожалению объединение невозможно произвести стандартными средствами системы и часто возникают проблемы при объединении файлов формата PDF. В ходе работы описана разработка программы для склейки файлов формата PDF на языке программирования Python.

Ключевые слова: Python, PDF, PyPDF2.

Development of a program for gluing PDF files based on the Python programming language

Kiseleva Elizaveta Alexandrovna

Sholom-Aleichem Priamursky State University

Student

Abstract

It is often not convenient to send scanned documents, letters, various references and important papers individually. Unfortunately, it is impossible to merge using standard system tools and often there are problems when combining PDF files. In the course of the work, the development of a program for gluing PDF files in the Python programming language is described.

Keywords: Python, PDF, PyPDF2.

В повседневной жизни либо при работе с компьютером нередко возникает потребность в использовании файлов формата PDF. Зачастую для отправки по почте мы сканируем документы, письма, различные справки и важные бумаги. К сожалению, не всегда удобно отправлять несколько файлов по отдельности и при склейке PDF документов часто возникают проблемы т.к. отсутствуют стандартные средства в операционной системе. Таким образом, в ходе работы планируется разработать программу для склейки файлов формата PDF.

Целью данной статьи: является разработка программы для склейки файлов формата PDF.

Многие исследователи занимались данной проблемой. В своей работе А.С. Маринчук [1] описал разработку системы для создания сборников научных статей. S. M. Velaga [2] описал методы извлечения изображения из PDF файлов. В.Г. Тронин и А.О. Куранов [3] описали способы извлечения страниц из PDF.

Для реализации данной проблемы было решено использовать язык программирования Python. Python - высокоуровневый язык программирования общего назначения, ориентированный на повышение производительности разработчика и читаемости кода, который включает в себя большой объём полезных функций и библиотек.

Для облегчения работы, было решено использовать библиотеку «PyPDF2» и «os». Библиотека «PyPDF2» используется для чтения, разделения и записи файлов формата PDF, библиотека «os» для работы с папками операционной системы, а также библиотека «webbrowser» для чтения файла после склейки.

В качестве примера реализации была выбрана курсовая работа за прошлые года обучения. В первом PDF файле хранится титульный лист, во втором хранится текст работы. На рисунке 1 изображены исходные PDF файлы.

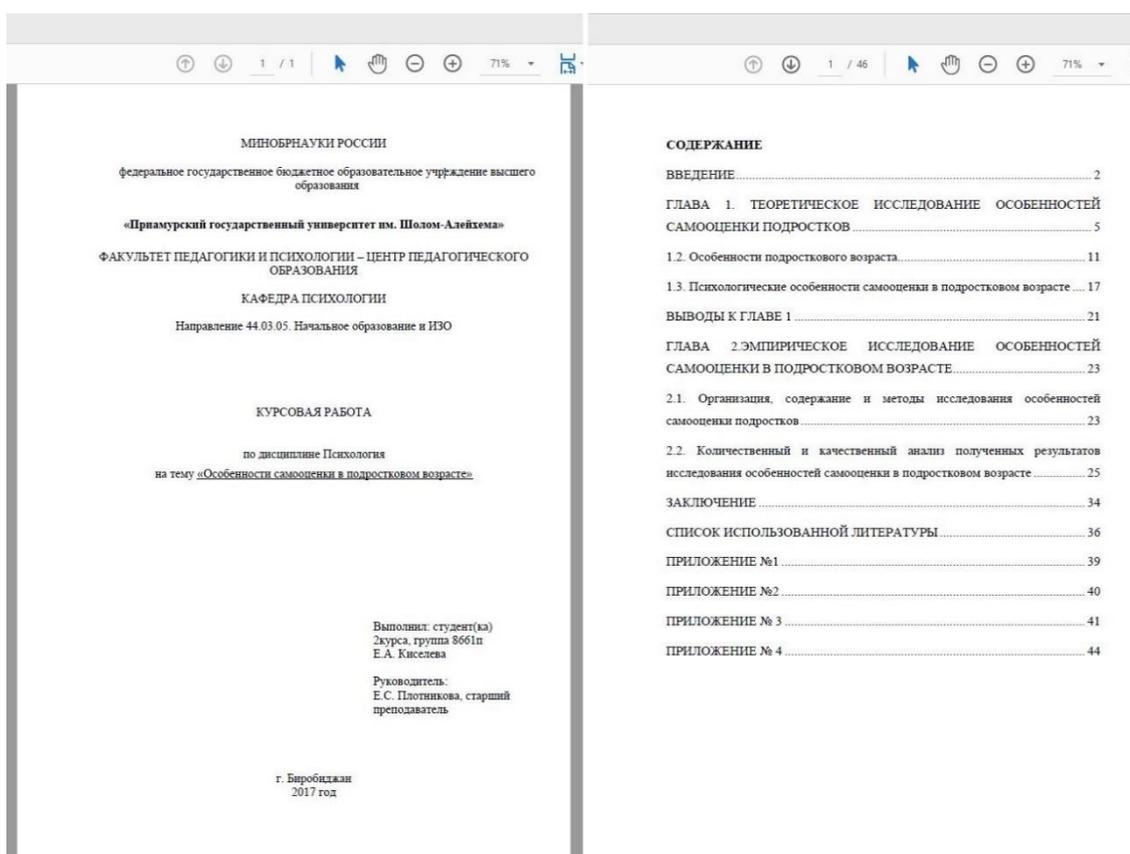


Рисунок 1 – Исходные PDF файлы

После того, как были выбраны исходные PDF файлы, приступаем к написанию кода. Оба файла помещены в одной директории с исполняемым

файлом, после чего с помощью библиотеки «os» обратились к текущим файлам для их открытия (Рисунок 2).

```
1  from PyPDF2 import PdfFileMerger
2      import webbrowser
3  import os
4
5      dir_path = os.path.dirname(os.path.realpath(__file__))
6
7  def list_files(directory, extension):
8      return (f for f in os.listdir(directory) if f.endswith('.' + extension))
9
10     pdfs = list_files(dir_path, "pdf")
11     merger = PdfFileMerger()
12     for pdf in pdfs:
13         merger.append(open(pdf, 'rb'))
14     with open('result.pdf', 'wb') as fout:
15         merger.write(fout)
16
17     webbrowser.open_new('file://' + dir_path + '/result.pdf')
18
```

Рисунок 2 – Код программы

Данный код с помощью цикла прогоняет все файлы в данной директории, после чего проверяет разрешение файла, в случае если файл заканчивается на «.pdf», то читается этот файл и записывается в переменную «pdfList». Таким образом, после запуска кода программа объединяет все имеющиеся файлы формата PDF в один и сохраняет в той же папке под названием «result.pdf» и сразу же запускает файл с результатом. На рисунке 3 изображен результат склеивания файлов.

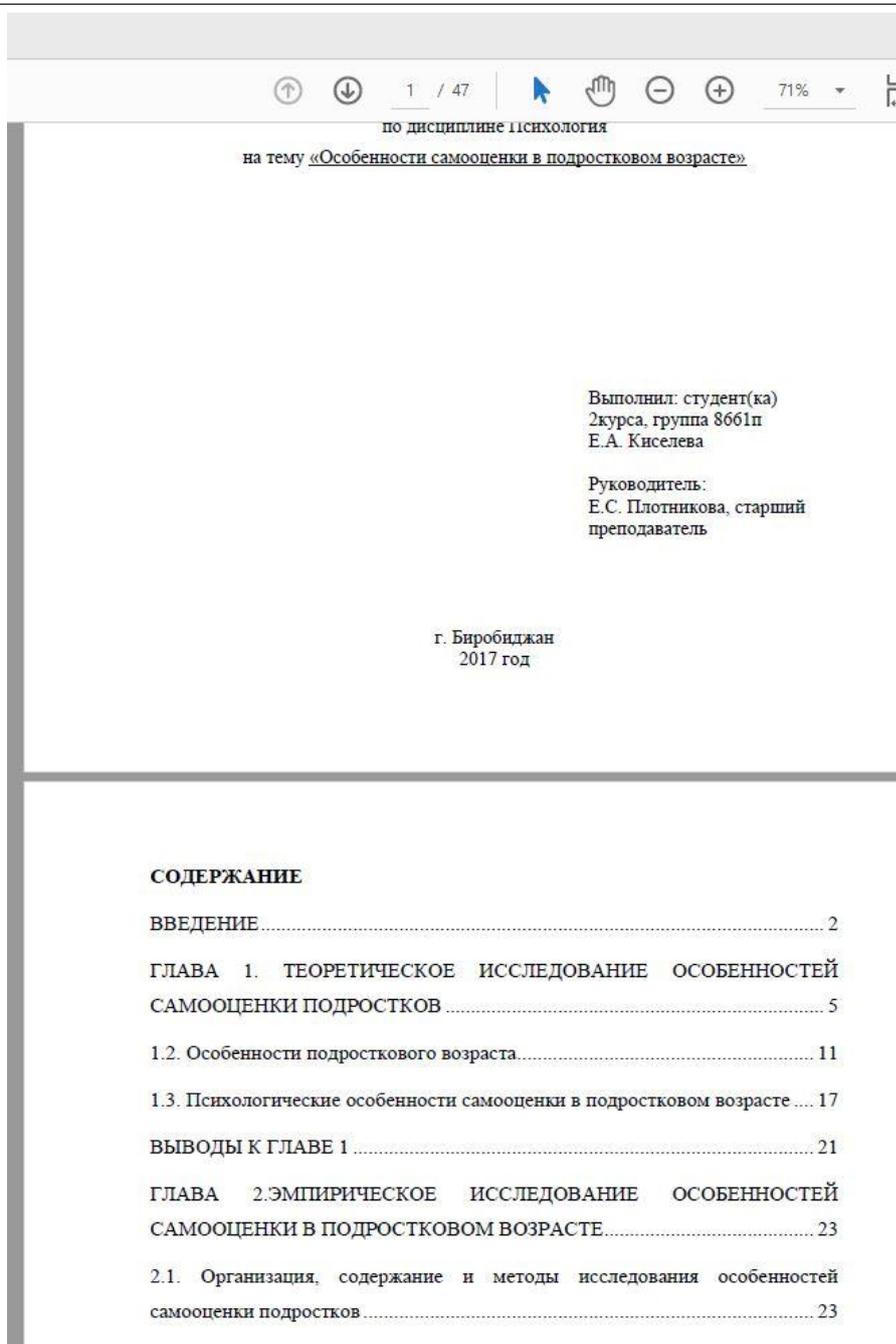


Рисунок 3 – Результат склеивания файлов

Заключение.

В ходе работы, была разработана программа на языке программирования Python для склейки файлов формата PDF. По ходу работы был продемонстрирован код программы и успешно протестирован в работе, а также показан результат. Благодаря данной программе всего за несколько секунд можно объединить файлы любого объема.

Библиографический список

1. Маринчук А. С. Разработка информационной системы «Верстка сборника материалов конференции» //Постулат. 2019. №. 6.

2. Velaga S. M. Prototype Device to Extract Text from Images, Videos and Document Files to Assist Visually Impaired People //Journal of Computational and Theoretical Nanoscience. 2018. Т. 15. №. 9-10. С. 2969-2978.
3. Тронин В. Г., Куранов А. О. Автоматизация процесса извлечения страниц из pdf-документа //Вестник Ульяновского государственного технического университета. 2018. №. 3 (83).