



empenoso

24 фев в 03:25

Как сканировать документы А3 формата, если под рукой только сканер А4 формата

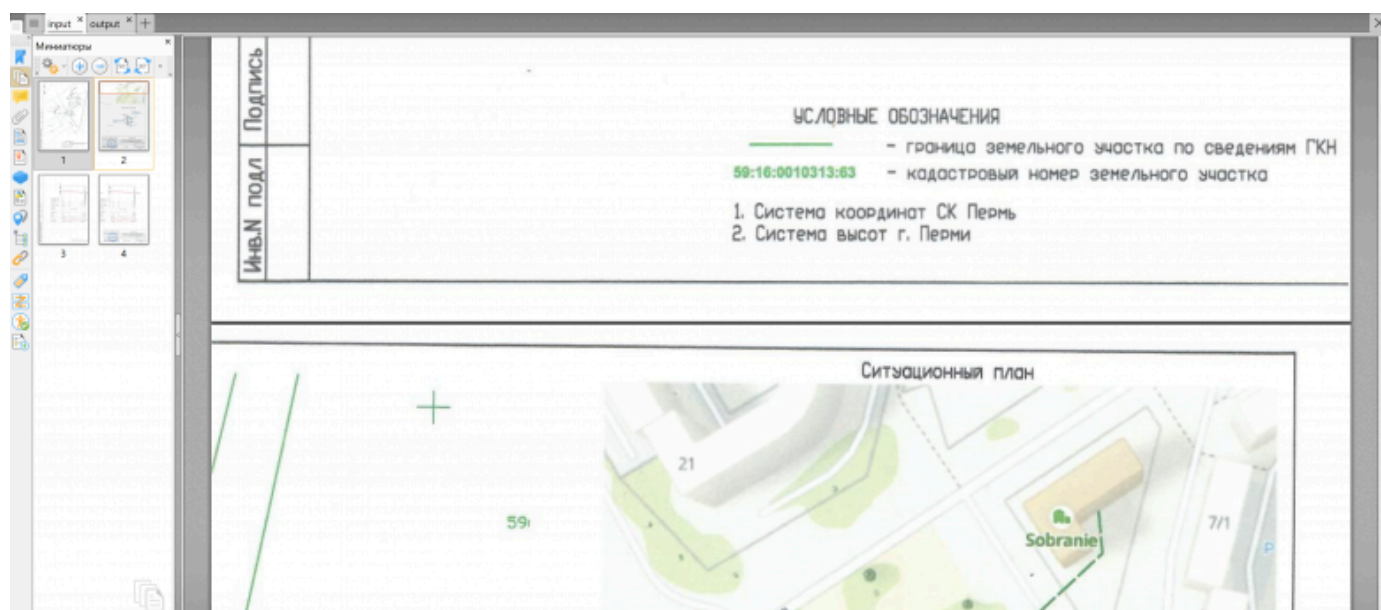
👉 Простой 🕒 5 мин 👁 4.7K

Open source*, Python*, Лайфхаки для гиков, PDF

Кейс

Оцифровка чертежей и документов формата А3 не представляет сложности, если у вас есть сканер соответствующего размера. Однако что делать, если под рукой лишь стандартное МФУ формата А4?

Формат А3 в два раза больше А4: это два стандартных листа, соединенных по длинной стороне. В результате, чтобы оцифровать А3, приходится сканировать его по частям - левую и правую половинки, а затем как-то их объединять.



Ручная склейка занимает много времени, особенно если документов много. А если тратить время на ручную склейку не хочется? Автоматизация такого процесса избавляет от рутинной работы. Актуально для чертежей и схем.

РЕКЛАМА



Горячая скидка 49%
на VPS в Малайзии

Для автоматизации склейки PDF-файлов потребуется всего несколько инструментов.

Первым делом понадобится **Python** — это простой в использовании язык программирования, идеально подходящий для подобных задач.

Для работы с PDF-файлами нам понадобится библиотека **PyPDF2**. Она позволяет извлекать страницы из PDF, объединять их, изменять порядок и выполнять другие операции.



Источник красивой картинки: dev.to

Чтобы склеить изображения из двух половинок, потребуется библиотека **Pillow**. С её помощью можно выполнять базовую обработку изображений: изменять размеры, соединять их и сохранять в нужном формате. Эта библиотека будет полезна, если ваши сканы нужно предварительно подогнать по размерам или качеству.

Установка Python

Если Python ещё не установлен на вашем компьютере, скачайте его с [официального сайта](#). Выберите последнюю стабильную версию для вашей операционной системы и установите её, следуя инструкциям. Во время установки не забудьте отметить пункт "Добавить Python



Горячая скидка 49%
на VPS в Малайзии

Установка библиотек

После установки Python, нужно добавить необходимые библиотеки. Сделать это просто: откройте терминал (или командную строку) и выполните следующие команды:

```
pip install PyPDF2 Pillow
```

Эти команды загрузят и установят всё необходимое. Теперь вы полностью готовы к написанию и запуску скрипта для автоматической склейки ваших PDF-документов.

Подготовка исходного файла

Перед началом автоматической склейки важно правильно подготовить исходный PDF-файл. Он должен содержать сканы двух половинок листа формата A4 каждая, расположенные в виде отдельных страниц. Это нужно для того, чтобы программа смогла корректно объединить их в одну страницу.

Важно, чтобы страницы PDF-файла следовали в правильном порядке:

1. **Первая страница** — это левая половина листа A3.
2. **Вторая страница** — правая половина.



Горячая скидка 49%
на VPS в Малайзии



Горячая скидка 49%
на VPS в Малайзии

Каждый лист A3 будет представлен двумя страницами. Если в вашем документе несколько листов, их половинки должны чередоваться так же: сначала левая часть первого листа, потом правая, затем левая часть второго листа, и так далее.

Эта структура поможет программе автоматически склеить половинки в правильном порядке.

Как работает процесс склейки

Автоматизация склейки документов A3 формата включает извлечение страниц, обработку изображений и преобразование в PDF. Вот как это работает:

1. Извлечение страниц из PDF

Каждая страница PDF, представляющая половину листа A3, сначала извлекается. Этот процесс позволяет получить изображения страниц, которые затем можно обработать.

2. Преобразование страниц в изображения

С помощью библиотеки Pillow страницы преобразуются в изображения. Это необходимо для удобной манипуляции — например, склеивания или вращения.

3. Объединение изображений

Изображения, соответствующие левой и правой половинам листа, склеиваются по горизонтали. Этот этап объединяет две части в единый файл, точно воспроизводящий лист A3.

4. Конвертация обратно в PDF

Склеенные изображения снова преобразуются в PDF с помощью PyPDF2. Это позволяет сохранить формат документа и сделать его удобным для дальнейшего использования.



Горячая скидка 49%
на VPS в Малайзии

Сам скрипт

Вот и сам скрипт: он автоматически склеивает половинки страниц формата A4 в один файл формата A3. Может использоваться на Windows, macOS и Linux:

```
# pip install PyPDF2 pillow

# Подробное описание: https://habr.com/ru/articles/875846/

import sys
sys.stdout.reconfigure(encoding='utf-8')

import os
from PyPDF2 import PdfReader, PdfWriter
from PIL import Image

# Выводим общую информацию о начале обработки PDF
print("[INFO] Начало обработки PDF...\n")

# Имя входного PDF-файла
input_pdf_file = "input.pdf"
# Имя выходного PDF-файла
output_pdf_file = "output.pdf"

# Углы поворота для левой и правой частей (в градусах)
# Укажите поворот для каждой части, например, 0, 90, 180 или 270
left_rotation = 0
right_rotation = 180

# Создадим временную папку для сохранения изображений
output_images_dir = "temp_images"
os.makedirs(output_images_dir, exist_ok=True)

# Считываем PDF и выгружаем каждую страницу как отдельное изображение
reader = PdfReader(input_pdf_file)
page_images = []

print(f"[INFO] Извлечение изображений из {input_pdf_file}...")
for page_number, page in enumerate(reader.pages):
```



Горячая скидка 49%
на VPS в Малайзии

```

if x_object[obj].get("/Subtype") == "/Image":
    data = x_object[obj].get_data()
    image_path = os.path.join(output_images_dir, f"page_{page_number + 1}.png")
    with open(image_path, "wb") as image_file:
        image_file.write(data)
    page_images.append(image_path)
    print(f"[INFO] Изображение страницы {page_number + 1} сохранено: {image_path}")

# Обрабатываем и склеиваем по две страницы
print("[INFO] Склеивание изображений в формат A3...")
glued_images = []
for i in range(0, len(page_images), 2):
    left_image_path = page_images[i]
    right_image_path = page_images[i + 1] if i + 1 < len(page_images) else None

    # Загружаем изображения
    left_image = Image.open(left_image_path)
    left_image = left_image.rotate(left_rotation, expand=True) # Применяем поворот лева

    if right_image_path:
        right_image = Image.open(right_image_path)
        right_image = right_image.rotate(right_rotation, expand=True) # Применяем поворос

    # Создаем новое изображение для A3 склеенного
    total_width = left_image.width + right_image.width
    max_height = max(left_image.height, right_image.height)
    new_image = Image.new("RGB", (total_width, max_height))

    # Вставляем левую и правую части
    new_image.paste(left_image, (0, 0))
    new_image.paste(right_image, (left_image.width, 0))
    glued_image_path = os.path.join(output_images_dir, f"glued_page_{i // 2 + 1}.png")
    new_image.save(glued_image_path)
    glued_images.append(glued_image_path)
    print(f"[INFO] Готово: склеенное изображение {glued_image_path}")

# Конвертируем склеенные страницы обратно в PDF
print("[INFO] Конвертация склеенных изображений обратно в PDF...")
writer = PdfWriter()
for glued_image_path in glued_images:

```



Горячая скидка 49%
на VPS в Малайзии

```
with open(pdf_page_path, "rb") as pdf_page_file:
    writer.add_page(PdfReader(pdf_page_file).pages[0])

# Сохраняем итоговый PDF
with open(output_pdf_file, "wb") as output_pdf:
    writer.write(output_pdf)

print(f"[INFO] Обработка завершена. Итоговый PDF сохранен как {output_pdf_file}")
```

Результат:

Заключение

После завершения обработки откройте итоговый PDF и убедитесь, что страницы формата



Горячая скидка 49%
на VPS в Малайзии

- Склейка нестандартных форматов.
- Использование компьютерного зрения для того, чтобы исключить небольшие повороты каждого из листов.

Автоматизация процесса оцифровки больших форматов избавляет от рутины и экономит время, особенно при работе с чертежами, планами и схемами.

Не забывайте сохранять резервные копии исходных файлов — это защитит вас от потери данных.

Автор: Михаил Шардин

 [Моя онлайн-визитка](#)

 [Telegram «Умный Дом Инвестора»](#)

24 февраля 2025 г.

Теги: [сканирование](#), [документы](#), [чертежи](#), [PyPDF2](#), [Pillow](#), [сезон open source](#)

Хабы: [Open source](#), [Python](#), [Лайфхаки для гиков](#), [PDF](#)

Редакторский дайджест



Присылаем лучшие статьи раз в месяц



189

17.5

Карма

Рейтинг

Михаил Шардин [@empenoso](#)

[Автоматизация](#) / [Данные](#) / [Финансы](#) / [Умные дома](#)

[Подписаться](#)



[Хабр Карьера](#) [Сайт](#) [Сайт](#) [Github](#)



Горячая скидка 49%
на VPS в Малайзии

Комментарии 9

Публикации

ЛУЧШИЕ ЗА СУТКИ

ПОХОЖИЕ



Exosphere

15 часов назад

Ещё 10 ошибок авторов Хабра



11 мин



3.2K



+93



34



64



Горячая скидка 49%

на VPS в Малайзии

22 часа назад

Больше нет входа в IT. Только выход

🕒 2 мин 👁 76K

📌 +49

🔖 78

💬 320



duran-duran

21 час назад

Трамплин в интернет: как мы ускорили запуск Яндекс Браузера

🕒 6 мин 👁 3.2K

📌 +42

🔖 9

💬 29



Nickmob

16 часов назад

Введение в Angie: краткая история и отличия от Nginx

👉 Простой 🕒 7 мин 👁 2.8K

Ретроспектива

📌 +37

🔖 29

💬 8



Myskat_90

21 час назад

Распределённый инференс и шардирование LLM. Часть 1: настройка GPU, проброс в Proxmox и настройка Kubernetes

👉 Сложный 🕒 14 мин 👁 1.8K

Тutorial

📌 +33

🔖 46

💬 0



slava_rumin

15 часов назад

Мое производство электроцитов приносит 40 млн в год. Спасибо нейросетям и СССР за конструкторскую школу



Горячая скидка 49%

на VPS в Малайзии

 +29 45 53**AlexeyNadezhin**

22 часа назад

Важное обновление BatteryTest 2

**Простой**

3 мин



2.5K

 +27 37 9**ru_vds**

16 часов назад

Как создавались вокальные эффекты Daft Punk

**Средний**

13 мин



832

[Обзор](#)[Перевод](#) +23 5 1**GordienkoAnd**

19 часов назад

Как настраивать сети: готовые решения Selectel для максимальной отказоустойчивости

**Средний**

20 мин



2.5K

[Обзор](#) +21 28 0**alizar**

20 часов назад

Почему из технологий делают культы

**Простой**

8 мин



1.3K

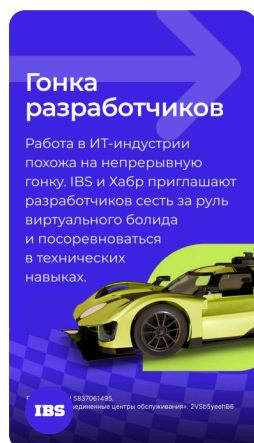
[Обзор](#) +21 7 5**Горячая скидка 49%**

на VPS в Малайзии

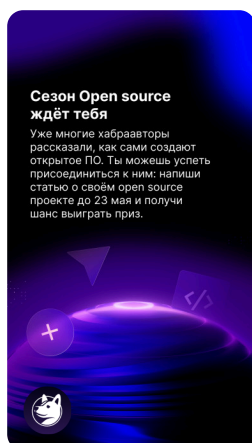
Турбо

Показать еще

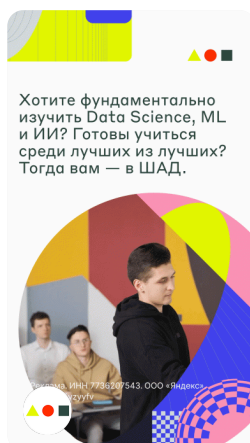
ИСТОРИИ



Старт гонки разработчиков



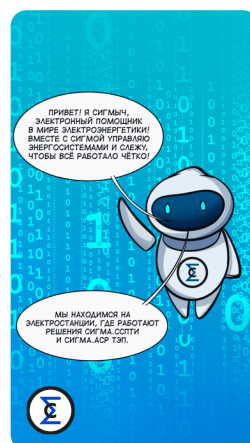
Торопись в сезон Open source



Открыт приём в Школу анализа данных



С Днём радио!



HELLO, WORLD — теперь на 110 киловольтах



Будущее

ВОПРОСЫ И ОТВЕТЫ

Как правильно настроить сканирование в папку на МФУ Ricoh Aficio MP 5002SP?

Сканирование · Простой · 1 ответ

Несколько вопросов по работе со сканами на андроиде?

Android · Простой · 0 ответов

Вылетает автоподатчик в Canon MF237w?

Canon · Средний · 3 ответа

МФУ Canon mf4018. Как сканировать с кнопки на ПК?

Сканирование · Средний · 0 ответов

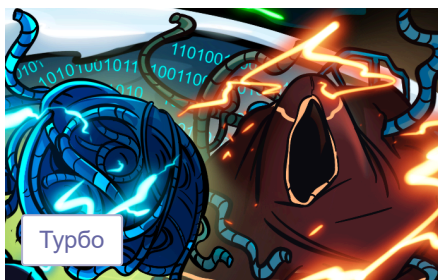
Почему программа для сканирования зависает?

HP · Простой · 1 ответ



Горячая скидка 49%
на VPS в Малайзии

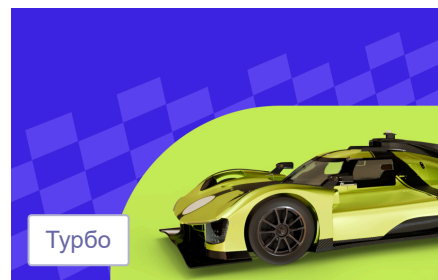
МИНУТОЧКУ ВНИМАНИЯ



Как упорядочить хаос тикетов на спринте



Весеннее пробуждение IT-мероприятий в Календаре



Гонка разработчиков backend, frontend, mobile уже началась! Займи место на старте

РАБОТА

Django разработчик

17 вакансий

Data Scientist

43 вакансии

Python разработчик

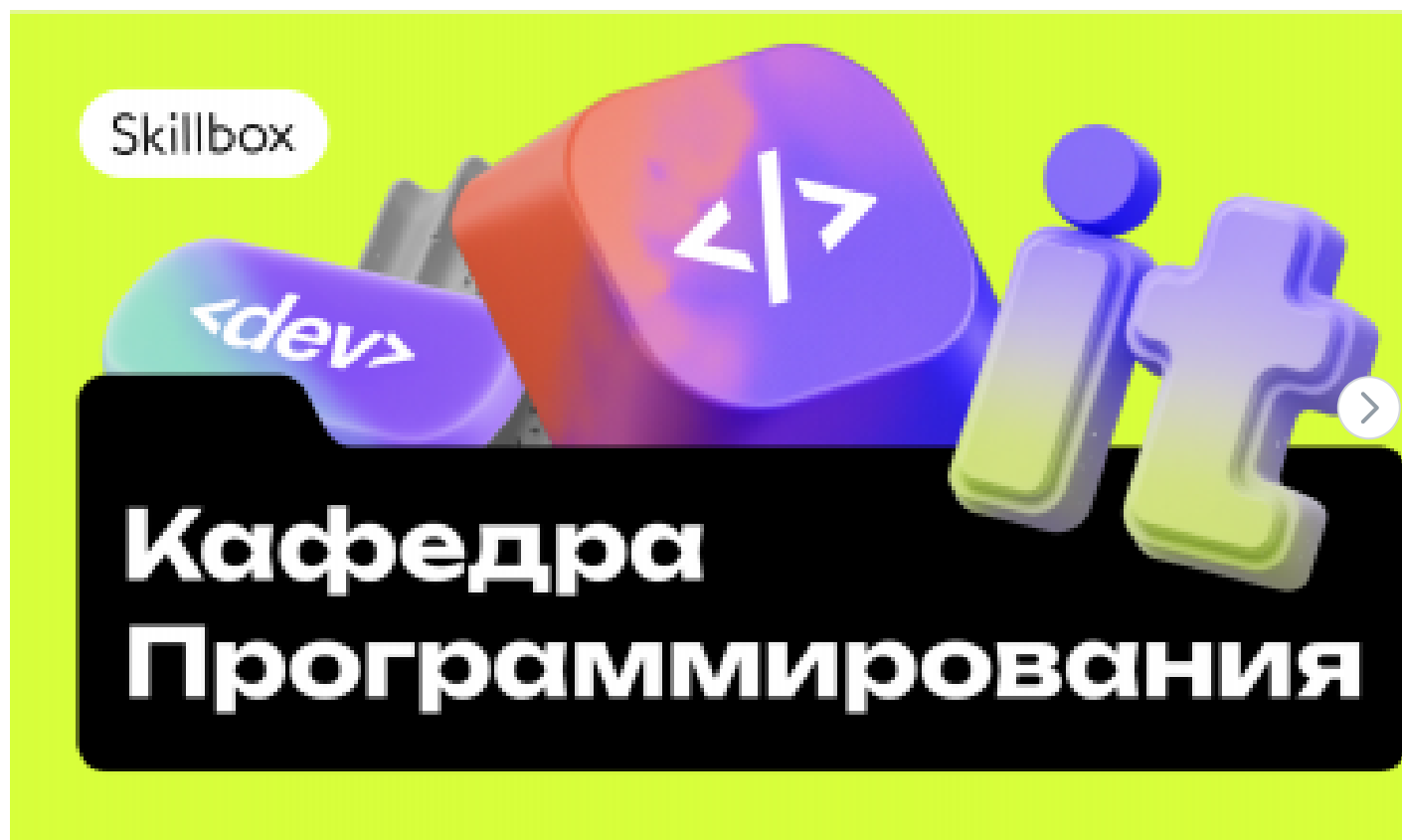
62 вакансии

[Все вакансии](#)

БЛИЖАЙШИЕ СОБЫТИЯ



Горячая скидка 49%
на VPS в Малайзии



17 апреля – 29 мая

Серия бесплатных офлайн-конференций «Кафедра Программирования» от Skillbox

Москва

Разработка

Больше событий в календаре

Хабр



Горячая скидка 49%
на VPS в Малайзии

техническая поддержка

© 2006–2025, Habr



Горячая скидка 49%
на VPS в Малайзии