

1. 多线程实现批量 PDF 转换器工具

参考资料:

- 1). pdfminer3k 是 pdfminer 的 python3 版本，主要用于读取 pdf 中的文本。
- 2). python-docx 模块是读取和写入 word 文档的工具。



部分参考代码:

```
def read_from_pdf(file_path):  
    with open(file_path, 'rb') as file:  
        resource_manager = PDFResourceManager()  
        return_str = StringIO()  
        lap_params = LAParams()  
  
        device = TextConverter(  
            resource_manager, return_str, laparams=lap_params)  
        process_pdf(resource_manager, device, file)  
        device.close()  
  
        content = return_str.getvalue()  
        return_str.close()  
        return content
```



```
def save_text_to_word(content, file_path):  
    doc = Document()  
    for line in content.split('\n'):  
        paragraph = doc.add_paragraph()  
        paragraph.add_run(remove_control_characters(line))  
    doc.save(file_path)
```

```
def remove_control_characters(content):  
    mpa = dict.fromkeys(range(32))  
    return content.translate(mpa)  
  
def pdf_to_word(pdf_file_path, word_file_path):  
    content = read_from_pdf(pdf_file_path)  
    save_text_to_word(content, word_file_path)
```



2. IP 地址归属地批量查询任务

注意: 使用创建子类的方式实现多线程任务



3. 基于多线程的批量主机存活探测

注意：使用实例化对象的方式实现多线程任务

项目描述： 如果要在本地网络中确定哪些地址处于活动状态或哪些计算机处于活动状态，则可以使用此脚本。我们将依次 ping 地址，每次都要等几秒钟才能返回值。这可以在 Python 中编程，在 IP 地址的地址范围内有一个 for 循环和一个 `os.popen (“ping -q -c2” + ip)`。

项目瓶颈： 没有线程的解决方案效率非常低，因为脚本必须等待每次 ping。

