

# python利用win32com读取doc和pdf内容，并保存到文件

原创

vincent\_hahaha

🕒

于 2021-05-03 15:34:02 发布

👁

1209

★

收藏 6

版权

分类专栏:

python



python 专栏收录该内容

0 订阅

49 篇文章

订阅专栏

将使用win32com包进行处理

## 读取doc文件

登录后复制

```
1 # coding=utf-8
2 import os, fnmatch
3 from win32com import client as wc
4 from win32com.client import Dispatch
5
6
7 def word2txt(filePath, savePath = ''):
8     dirs, filename = os.path.split(filePath)
9     print(dirs, '\n', filename)
10    new_name = ''
11    if fnmatch.fnmatch(filename, "*.docx"):
12        new_name = filename[:-5] + '.txt'
13    if fnmatch.fnmatch(filename, "*.doc"):
14        new_name = filename[:-4] + '.txt'
15    if savePath == '':
16        savePath = dirs
17    else:
18        savePath = savePath
19    word2txtPath = os.path.join(savePath, new_name)
20    print(word2txtPath)
21    wordapp = wc.Dispatch('Word.Application')
22    mytxt = wordapp.Documents.Open(filePath)
23    mytxt.SaveAs(word2txtPath, 4) # 4代表抽取结果保存为文本
24    mytxt.Close()
25
26
27 if __name__ == '__main__':
28     filePath = os.path.abspath(r'./专业课.docx')
29     word2txt(filePath)
```

## 读取pdf

```
1 # coding=utf-8
2 import os, fnmatch
3 from win32com import client as wc
4 from win32com.client import Dispatch
5
6
7 def pdf2txt(filePath, savePath=''):
8     dirs, filename = os.path.split(filePath)
9     print(dirs, '\n', filename)
10    new_name = ''
11    if fnmatch.fnmatch(filename, '*.pdf') or fnmatch.fnmatch(filename, '*.PDF'):
12        new_name = filename[:-4] + '.txt'
13    else:
14        print('格式不正确，仅支持pdf格式')
15        return
16    if savePath == '':
17        savePath = dirs
18    else:
19        savePath = savePath
```



vincent\_hahaha

关注

```
19 pdf2txtPath = os.path.join(savePath, new_name)
20 print(pdf2txtPath)
21 wordapp = wc.Dispatch('Word.Application')
22 mytxt = wordapp.Documents.Open(filePath)
23 mytxt.SaveAs(pdf2txtPath, 4) # 4代表抽取文本
24 mytxt.Close()
25
26
27 if __name__ == '__main__':
28     filePath = os.path.abspath(r'./论文.pdf')
29     pdf2txt(filePath)
```

文章知识点与官方知识档案匹配，可进一步学习相关知识

Python入门技能树 进阶语法 文件 190267 人正在系统学习中

- Python解析并读取PDF文件内容的方法

09-20

主要介绍了Python解析并读取PDF文件内容的方法,结合实例形式分别描述了Python2.7在win32与win64环境下...
- win32com excel转pdf

Zhong\_\_\_\_的博客 2782

需要用到库 from win32com.client import DispatchEx import os from PyPDF2 import PdfFileReader, PdfFileW...
- Python使用win32com获取Word文档总页码/页数\_小白救小白的博客-CSDN博 ...

11-22

不在前台显示 w=win32com.client.Dispatch("Word.Application")w.Visible=0w.DisplayAlerts=0#打开一个word...
- python win32com在读取word文档时,遇到的问题\_acfcv00088的博客-CSDN博...

11-22

在网址http://docs.activestate.com/activepython/2.4/pywin32/pythoncom.html中,查询pythoncom的Colnitalize(...
- python处理word文件:win32com用法详解

12-01

目标: 用python处理 doc 文件 方法: 引入 win32com 首先要下载安装 win32com模块 下载地址: 下载地址: ...
- pywin32文档\_Word文档批量转PDF实现

weixin\_39548972的博客 173

本期推文将向大家介绍如何批量把word 文档转换成PDF，以下注重实现过程的分析。这里主要使用了pywin32...
- python win32中文手册-python win32com 操作word总结\_weixin\_39936134的...

11-19

打开word文档 读取表格 表格插入行 向表格中填写内容 文档另存为 环境 python 2.7 文件保存的编码格式是utf...
- python win32com、docx 操控word\_qyfxl001的博客

11-16

import win32com.client 有时候直接引用win32com 找不到client 最常用的模块其实是win32com.client 打开与...
- 【python基础】——python读写doc/docx/txt/xls文件 热门推荐

汤庆 4万+

一、处理文件 1、读取txt文件 #!usr/bin/python # python3.6 file = open('E:\个人文件\6-desktop\new.txt','r') f...
- 使用Python通过win32 COM实现Word文档的写入与保存

小灰笔记 1万+

通过win32 COM接口实现软件的操作本质上来跟直接操作软件一致，这跟我之前经常用的通过各种扩展的...
- python win32com word\_利用python中的win32com模块操作Word、Excel文件...

11-27

from win32com.client import Dispatch, DispatchEx, constants def doc\_to\_docx(f\_path): w = Dispatch("Word....
- win32com 操作Excel

Zhong\_\_\_\_的博客 5800

目录 打开表格，读取数据 win32的写入操作 刷格式 数据透视表 小贴士 win32com是一个VBA转py的库（双厨...
- python自动化系列之使用win32com操作Excel

weixin\_42551921的博客 807

win32com模块主要为Python提供调用windows底层组件对word、Excel、PPT等进行操作的功能；不过有其局...
- Python pywin32实现word和Excel的处理 最新发布

Xy-Huang的博客 650

Python pywin32实现office（word和excel）开发
- Python 使用 win32com 模块对 word 文件进行操作

Python热爱者的博客 6884

what's the win32com 模块 win32com 模块主要为 Python 提供调用 windows 底层组件对 word、Excel、PPT ...
- pythonwin32com 操作word excl

qq\_31772351的博客 511


一、导入库 import win32com from win32com.client import Dispatch, constants 二、打开进程 w = win32com....
- python打开word内对象\_使用pywin32打开Word文档并读取现有的te...weixin\_32795125的博客

714

我试图用pywin32打开Word文档，从一些文本框中复制值并打印出来。代码和错误如下。非常感谢任何帮助。...
- python利用win32com实现doc文档转为pdf文档的功能

葇夜 2686

1、首先，必须去安装相关的库，使用命令：pip install pypiwin32，接下来是三
- python win32com、docx 操控word

vincent\_hahaha

关注

- win32com模块 import win32com import win32com.client 有时候直接引用win32com 找不到client 最常用的模...

win32com 读取excel

huiyinimen的博客 4523

使用 win32com 读取excel def open\_excel(path,open\_password,write\_password): #载入Excel xlApp=win32c...
- 使用win32com操作word的方法记录（一）

Wave\_Alex的博客 1万+

基于工作需要，最近需要对word进行一些批量操作，但考虑到word中VBA学习成本较高（多是VBA教程针对E...
- win32com 读doc,doc表格

哈哈儿 博客 3216

#coding:utf-8import osimport win32comimport win32com.client as win32from Ltp import Ltp\_dbfrom win32co...
- python读doc文档-win32com

qq\_40006760的博客 515

上一篇提到用python-docx读写.docx文章，但该模块并不能读取.doc文档,这里使用win32com模块，更高级，...
- 使用PYTHON win32com处理word，例如-自动读取页码范围

sinat\_35758792的博客 1010

# -\*- coding: utf-8 -\*- """Created on Fri Dec 13 17:57:45 2019 @author: johwang """ import win32com#引|入wi...
- Python操作Word（Win32com）

ldgk3ekkd的博客 952

https://zhuanlan.zhihu.com/p/67543981
- 办公自动化：Python-win32com自动将word文档转换成pdf格式！

梦在硅谷的博客 473

在平常的办公中需要将word文件调整为pdf的格式，然后没有更好可用的免费软件可以使用，python为我们提...

“相关推荐”对你有帮助么？

- 非常没帮助
- 没帮助
- 一般
- 有帮助
- 非常有帮助

©2022 CSDN 皮肤主题：撸撸猫 设计师：马嘞嘞 返回首页

关于我们 招贤纳士 商务合作 寻求报道 400-660-0108 kefu@csdn.net 在线客服 工作时间 8:30-22:00

公安备案号11010502030143 京ICP备19004658号 京网文〔2020〕1039-165号 经营性网站备案信息

北京互联网违法和不良信息举报中心 家长监护 网络110报警服务 中国互联网举报中心 Chrome商店下载 账号管理规范 版权与免责声明

版权申诉 出版物许可证 营业执照 ©1999-2022北京创新乐知网络技术有限公司

vincent\_hahaha  
码龄10年 暂无认证

778

1万+

3407

54万+

原创

周排名

总排名

访问

等级

6775

197

339

143

1437

积分

粉丝

获赞

评论

收藏

私信

关注

搜博主文章

热门文章

word2vec介绍 17631

Thread|

CallerR

vincent\_hahaha

关注

Apache Flink 零基础入门（七）Flink中  
keyBy三种方式指定key 14087

linux 根据进程号查看对应的进程 12285

docker 修改容器的主机名 10808

分类专栏

|   |        |     |
|---|--------|-----|
|  | 工具     | 3篇  |
|  | 机器学习   | 40篇 |
|  | python | 49篇 |
|  | pyg    | 7篇  |
|  | Latex  | 2篇  |
|  | GNN    | 10篇 |

最新评论

word2vec介绍  
夏木阁下: 写得真好，第二次看又是不同的  
感觉

word2vec介绍  
刘昊然老婆来了: 好厉害啊啊啊👀👀👀

word2vec介绍  
Inbbbb: 好的好的

word2vec介绍  
vincent\_hahaha: 一开始需要初始化，随着  
多次迭代，最终训练得到

word2vec介绍  
Inbbbb: W1到底是怎么得到的呀

您愿意向朋友推荐“博客详情页”吗？

强烈不推荐 不推荐 一般般 推荐 强烈推荐


最新文章

Python time和datetime

Python JSON数据序列化

Python 内置模块

|            |            |
|------------|------------|
| 2022年 29篇  | 2021年 119篇 |
| 2020年 140篇 | 2019年 214篇 |
| 2018年 35篇  | 2017年 81篇  |
| 2016年 93篇  | 2015年 82篇  |

 **vincent\_hahaha**

关注