



链滴

# Python 操作 Word 文档

作者: [mubai](#)

原文链接: <https://ld246.com/article/1484588449761>

来源网站: [链滴](#)

许可协议: [署名-相同方式共享 4.0 国际 \(CC BY-SA 4.0\)](#)

```

from zipfile import ZipFile
from urllib.request import urlopen
from io import BytesIO
from bs4 import BeautifulSoup

wordFile = urlopen("http://pythonscraping.com/pages/AWordDocument.docx").read()
wordFile = BytesIO(wordFile)
document = ZipFile(wordFile)
xml_content = document.read('word/document.xml')

# print(xml_content.decode('utf-8'))

wordObj = BeautifulSoup(xml_content.decode('utf-8'))
textStrings = wordObj.findAll("w:t")

for textElem in textStrings:
    closeTag = ""
    try:
        style = textElem.parent.previousSibling.find("w:pstyle")
        if style is not None and style["w:val"] == "Title":
            print("")
            closeTag = ""
    except AttributeError:
        #不打印标签
    pass
    print(textElem.text)
    print(closeTag)

# 参考：《Python网络数据采集》

import docx

# 加载Word文档
doc = docx.Document('test.docx')

# 统计段落数
print(len(doc.paragraphs))

# 段落
print(doc.paragraphs[0].text)
print(doc.paragraphs[0].style)

# 加入段落
doc.add_paragraph('Hello')

# Run对象
print(len(doc.paragraphs[0].runs))
print(doc.paragraphs[0].runs[0].text)
print(doc.paragraphs[0].runs[0].style)
doc.paragraphs[0].runs[0].underline = True

# 加入Run对象
doc.add_paragraph('Word').add_run(' hahaha')

```

```
doc.save('test.docx')
```