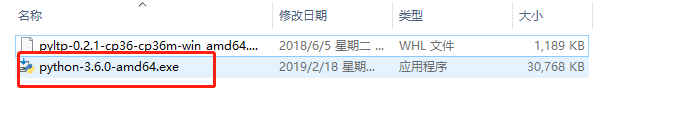
中文分词中间件v1.0的操作说明

中文分词中间件v1.0集成了jieba、lpt、nlpir三种分词模型。支持中文分词以及词性标注，支持用户自定义字典（暂时只支持jieba导入字典），所在位置为“middle\_fenci”文件夹下的“public\_dict”。中文分词中间件更易于操作，只需通过输入内容或者打开文件的方式就可以实现分词，而SDK模式更利于代码嵌套，可以将中间件完整的嵌套在程序中。下边分别介绍了中文分词中间件图形化系统的使用和中文分词SDK的使用，以及常见问题的处理。

环境准备

一、安装python3

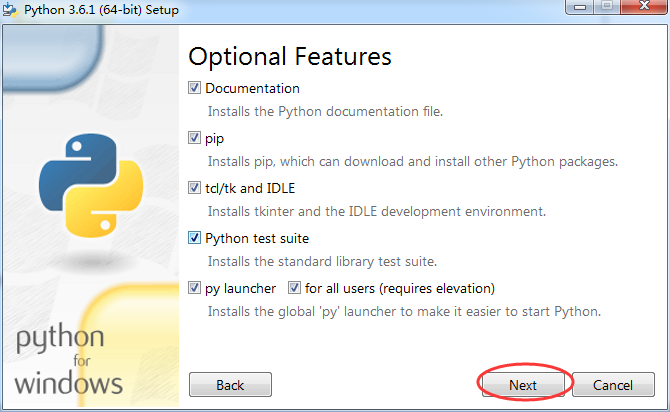
1. 打开env文件 双击“python-3.6.0-amd64.exe”文件安装python环境

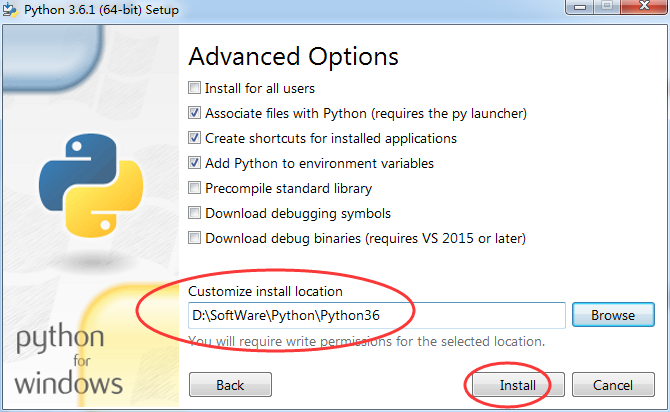


1. 安装python，注意添加环境变量

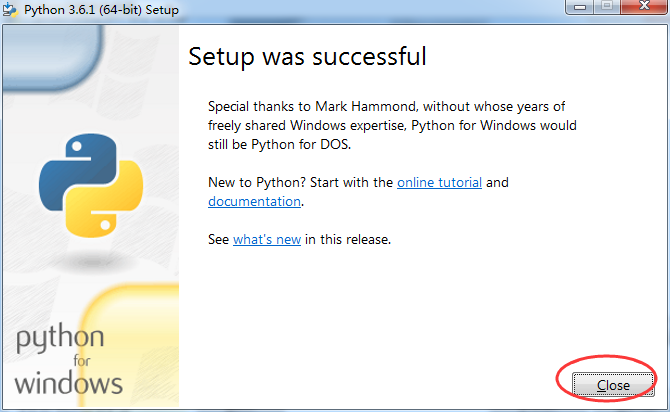


3、如果点击快速安装，程序会自动安装成功。如果为自定义安装，直接点“next”

  
4、上面的选项不用管理，但是可以选择安装路径，然后点击“Install”。比如，我们这里选择安装路径为D:\SoftWare\Python\Python36

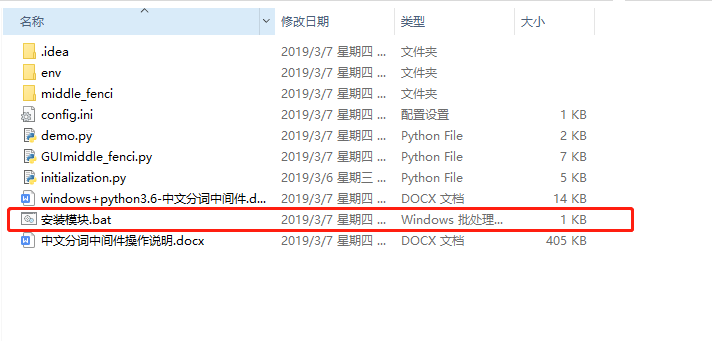


1. 安装成功，单机“Close”



二、安装所需要的模块

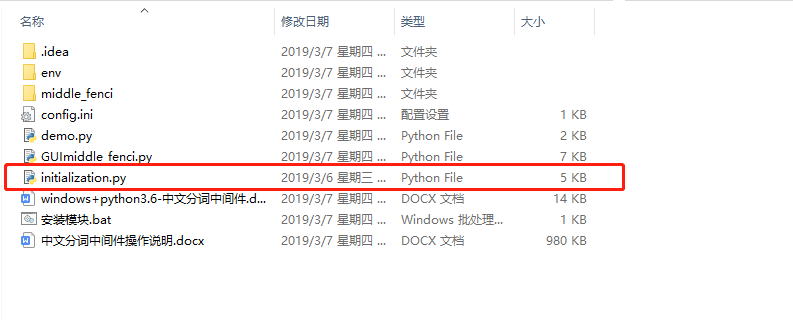
双击执行“安装模块.bat”批处理文件

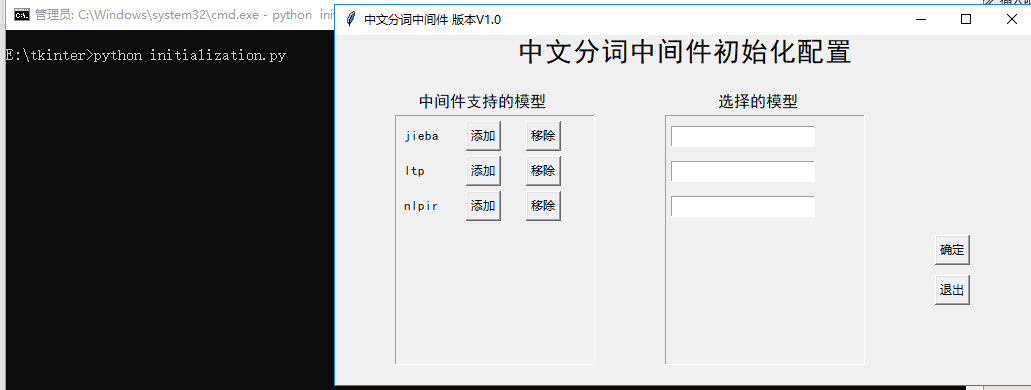


中文分词中间件图形化系统使用

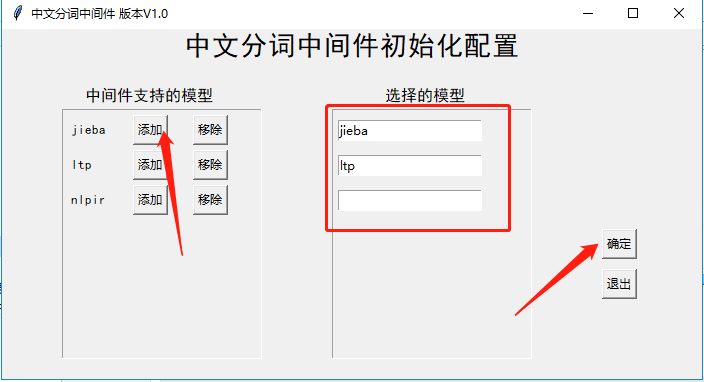
1. 配置管理器的使用

通过对配置管理器的可以选择中间件支持的模型，定义和管理中间件配备的分词系统信息，应用开发者也可以添加新的分词组件，并且可以设置和修改当前采用的分词系统管理中间件支持的分词组件。  
1、双击运行“initialization.py”或者在windows下打开命令终端，切换到中间件所在路径后，在命令行输入python initialization.py后回车。



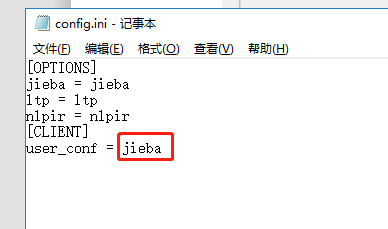


2、选择好需要添加的模型后点击确定按钮。

  
3、点击“是”按钮，模型被成功添加。



1. 双击打开文件夹中的config.ini文件，其中[OPTIONS]是当前程序支持的模型[CLIENT]为当前使用的模型，默认为jieba，可以在user\_info = jieba处进行修改



1. 中文分词操作

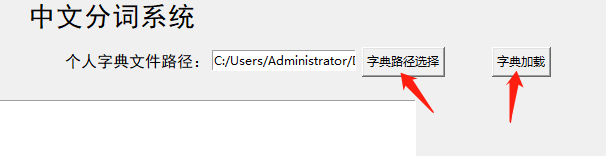
可以对字符串或者文件进行分词操作，支持的文件有txt，text，doc，docx。

1. 双击或者在命令行输入python GUImiddle\_fenci.py



“GUImiddle\_fenci.py”  
2、选择是否加载字典，只能为txt文件

点击“字典路径选择”后，添加想要加载的字典文件，之后点击字典加载，程序会自动将用户自定义的字典加载到当前程序中。

  
3、输入要进行分析的文本或者选择文件路径。

  
5、选择需要的模型以及输出格式并点击“分析”按钮

  
6、系统会自动分析并输出在下方的文本框中

  
  
  
  
SDK中间件模式间件的使用

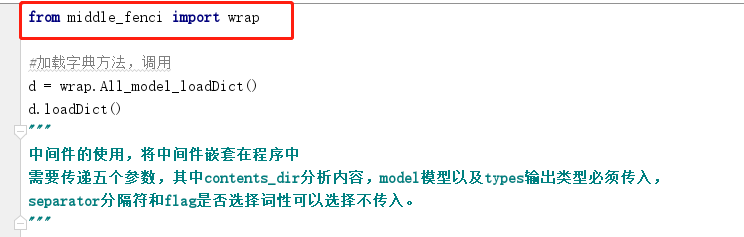
1. 配置管理器的使用

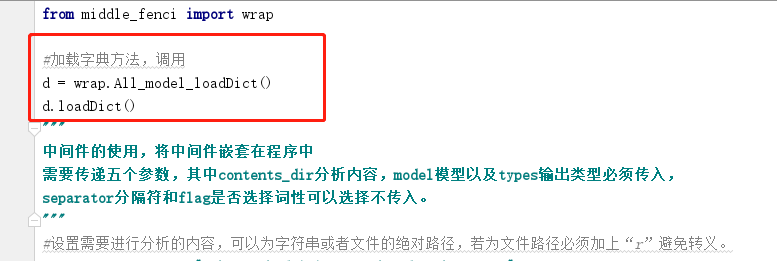
同上方的图形化界面使用  
二、分词代码嵌入

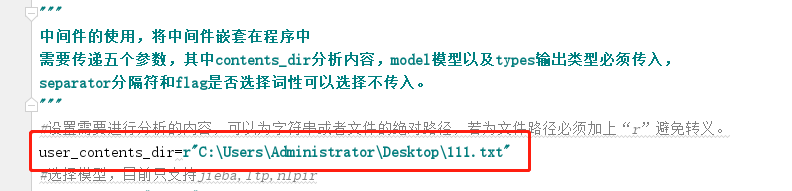
只需要导入中间件并调用方法就可以对字符串或者文件进行分词操作，支持的文件有txt，text，doc，docx。下边将以示例代码来详细解读如何进行代码嵌套



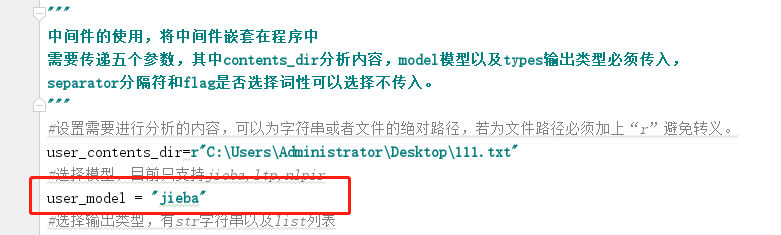
1. 使用开发工具打开demo.py文件，也可以使用记事本打开，输入**from** middle\_fenci **import** wrap

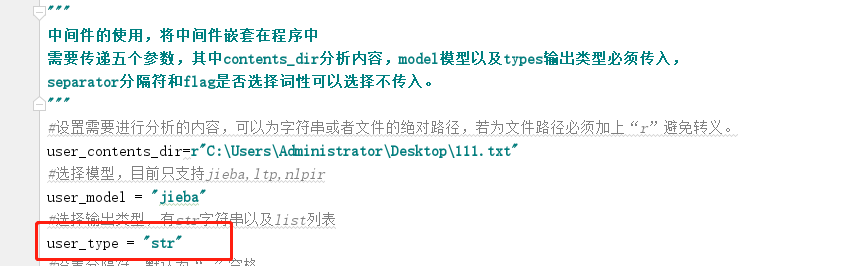
  
  
  
2、选择是否加载字典，将中间件的加载字典类实例化并调用加载方法，字典会自动加载，用户字典存放在middle\_fenci下的public\_dict文件夹下

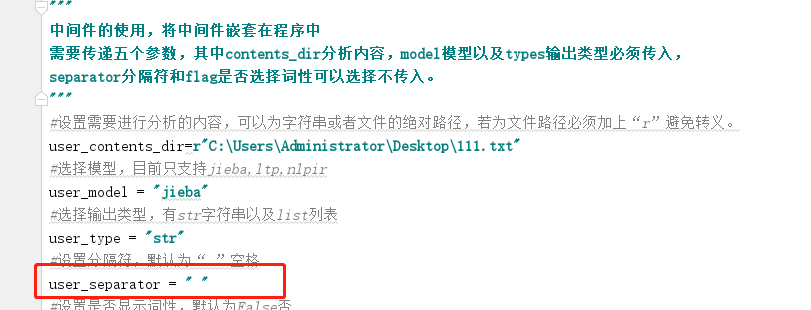
  
  
3、设置需要进行分析的内容，可以为字符串或者文件的绝对路径，若为文件路径必须加上“r”防止字符转义。

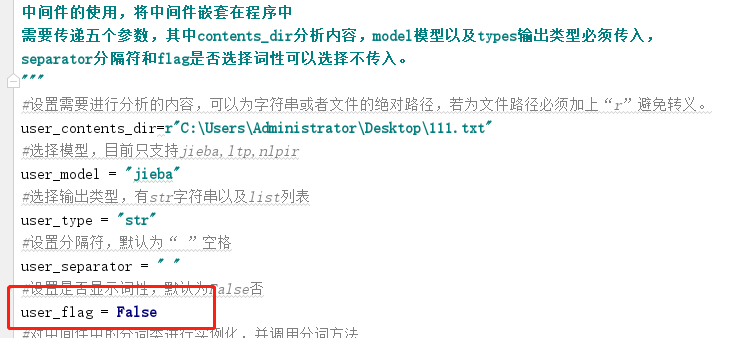


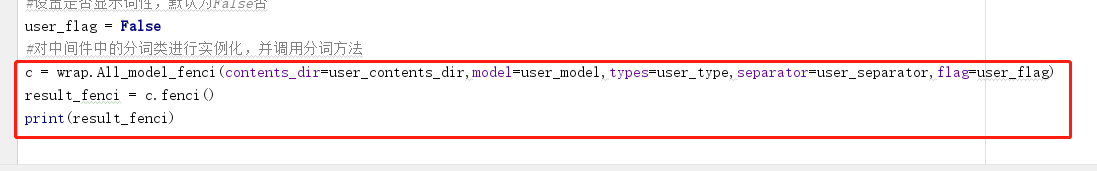
1. 选择模型，目前只支持jieba,ltp,nlpir



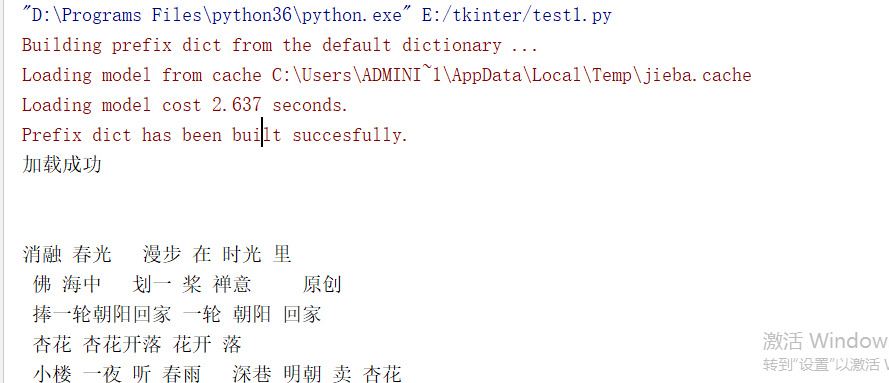
5、选择输出类型，有str字符串以及list列表

6、设置分隔符，默认为“ ”空格  
7、设置是否显示词性，默认为False否

  
8、对中间件中的分词类进行实例化，并调用分词方法，输出内容会展示在控制台上。



9、点击运行或者在命令行输入python demo.py，程序运行结果



附：运行环境需求：

Windows 64位

环境 python3.6  
安装库 在命令行输入

jiaba  
 pynlpir  
 pylpt

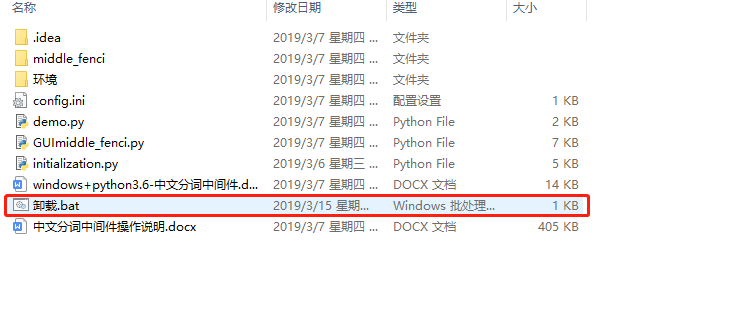
chardet

docx

python-docx

卸载程序

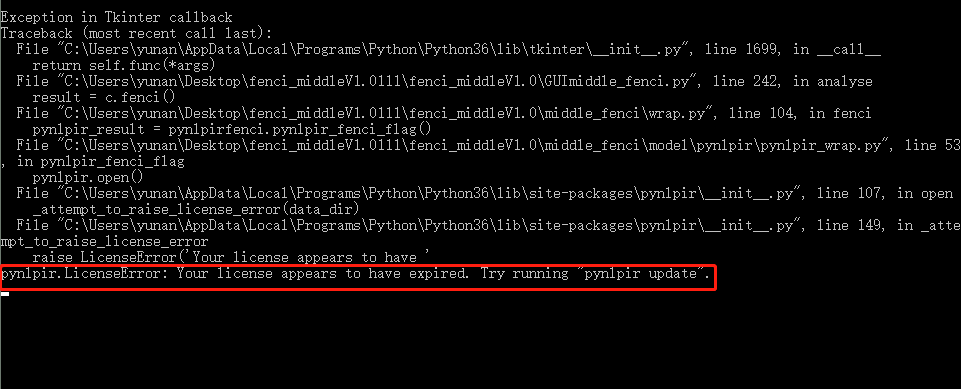
1. 双击运行“卸载.bat”文件，程序会自动卸载

2、删除空文件夹



常见问题

1. nlpir授权过期

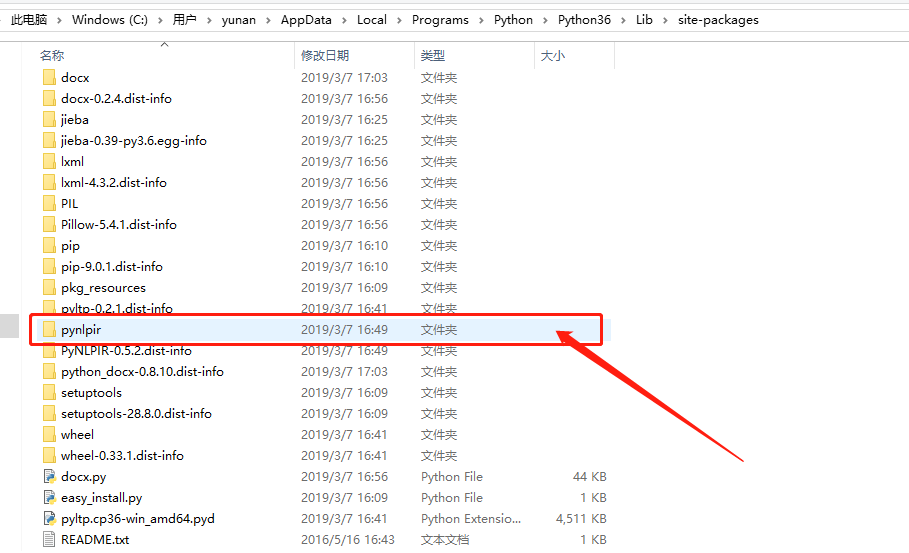


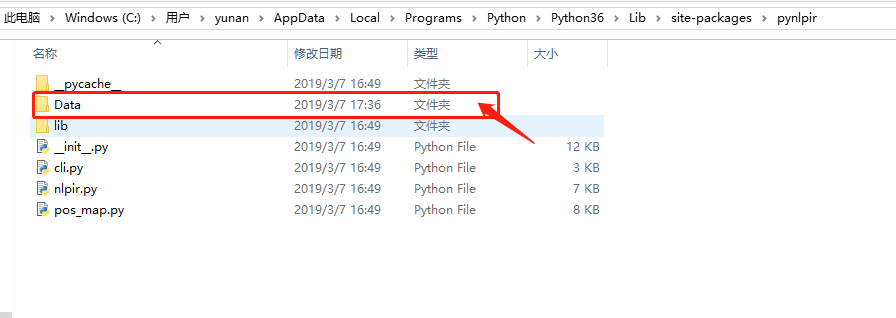
1、打开pynlpir文件夹所在的位置

Python默认安装路径C:\Users\Administrator\AppData\Local\Programs\Python\Python36\Scripts

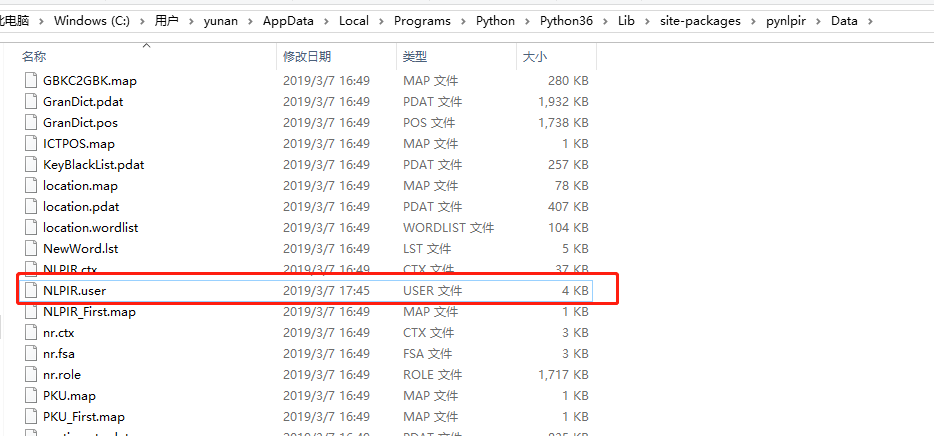
Pip默认安装路径

C:\Users\Administrator\AppData\Local\Programs\Python\Python36\Lib\site-packages





2、将env文件夹下的“NLPIR.user”替换Data文件夹中的NLPIT.user



3、重新运行程序