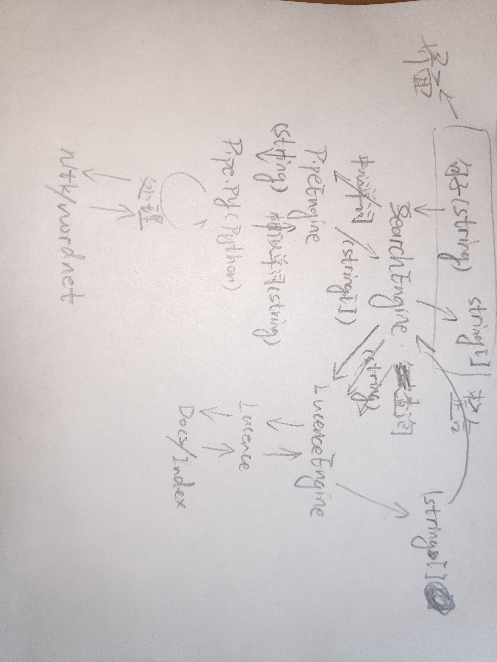
# 英文论文写作辅助程序

## 程序简介：

本程序是基于全文搜索引擎和wordnet实现的一个英文论文写作辅助程序，我们从<http://aclweb.org/>上爬取了13000份论文，整理得到了约130万个句子，组成数据集。用户输入一句还没写完的英文句子，程序将这前半句话作为query输入搜索引擎，得到数据集中最相似的几句话，返回给用户，使得用户可以依照这些论文集中的句子修改或完成自己句子，提高表达的地道性与流畅性；同时，wordnet将对用户输入句子中的最后一个单词做同义词检测，并将这个词的同义词也添加到query中，使得搜索得到的表达更加多样化，给用户更多的参考空间。

## 程序结构：



程序主要由如下几个部分组成：

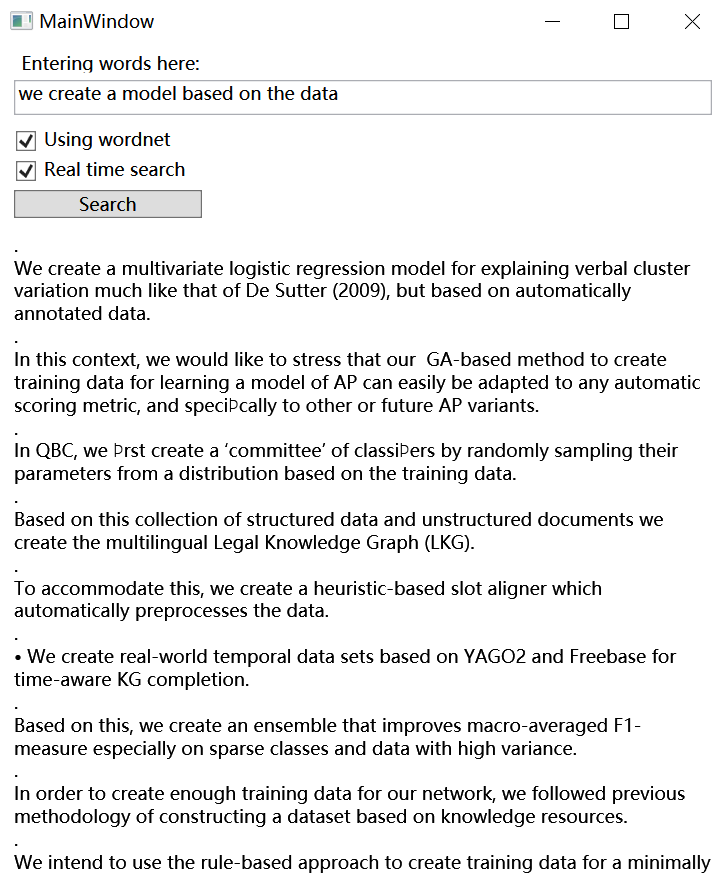
1. SearchEngine：用于和程序界面，lucence搜索引擎，管道交互，中转数据
2. LucenceEngie：调用lucence库，生成索引，接收query并返回搜索结果
3. PipeEngine：调用C#的管道库，负责启动相应的python程序并通过管道和其交换数据
4. pipe.py: python程序，代码在python文件夹下，通过调用python的管道库和C#进行数据交换，并且调用nltk中的wordnet来处理接收的数据

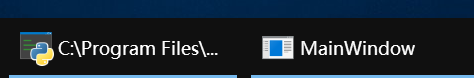
通常情况下，用户在输入一句话之后，SearchEngine将字符串交给PipeEngine；PipeEngine将字符串通过管道交给pipe.py处理，返回该字符串最后一个单词的同义词；SearchEngine将这些同义词附加到原字符串的末尾，将其作为query交给LucenceEngine，并将查询结果以字符串数组的形式返回给程序界面，分行显示在界面上。

但是由于在某些电脑的python环境下wordnet不能正常被调用及运行，这时候pipe.py和PipeEngine将不会运行，SearchEngine将直接把字符串作为query交给lucence查询，返回结果

## 程序界面

程序界面如下图所示



* 输入框是用户输入句子的地方，这里的示例输入是“we create a model based on the data”
* Using wordnet指的是使用wordnet进行联想，在python及nltk库不能正常运行的情况下这个选项不应被勾选
* Real time search表示实时搜索，输入框中的数据改变时就会进行搜索，这使得你可以不用点击“Search”按钮，但会增加一点系统资源的占用
* 点击“Search”进行一次搜索
* “Search”以下的部分都是用来显示搜索结果的
* 程序在调用python时可能会在任务栏形成一个窗口您可以打开它来查看python的相应输出信息，在您关闭主窗口后它会自动结束运行；注意：除非您发现它进入了一个无限输出错误信息的死循环或者在主窗口关闭之后它没有结束运行，否则建议您不要贸然关闭它，虽然这并不会导致任何数据的丢失。

## 程序运行注意事项

启动程序，请运行ProjectA.exe

程序第一次运行时可能需要联网来下载相应的python运行包（nltk，win32file，win32pipe）如果您有的话请忽略。

注意！程序会自动检测并配置您电脑的运行环境；一般情况下，这不会引发任何问题；但是为了严谨起见，请确保您电脑上的环境变量以及相应的python环境（如果有的话）能接受如下的更改：

1. 如果程序在全局环境变量中没有找到“python.exe”，那么程序将会运行一个python3.7.3 64位的安装文件
2. 如果程序在全局默认的python环境中没有找到nltk，win32file，win32pipe库，那么程序将调用cmd安装它们，相应的cmd命令如下：
   1. pip3 install --user nltk -i <https://pypi.tuna.tsinghua.edu.cn/simple/>
   2. pip3 install --user pywin32 -i <https://pypi.tuna.tsinghua.edu.cn/simple/>

这部分代码可以在python/pakcheck.py中找到

1. 如果程序在调用python的nltk包时遇到了没有加载的语料库，那么程序将会把python/nltk\_data文件夹中的数据拷贝在D:/nltk\_data中，如果电脑上没有找到D盘，需要您手动将上述数据拷贝在C:/nltk\_data或者其他有效的目录中

## 常见问题及解决方案

1. 电脑上没有安装python或者程序在全局环境中没有找到“python.exe”：

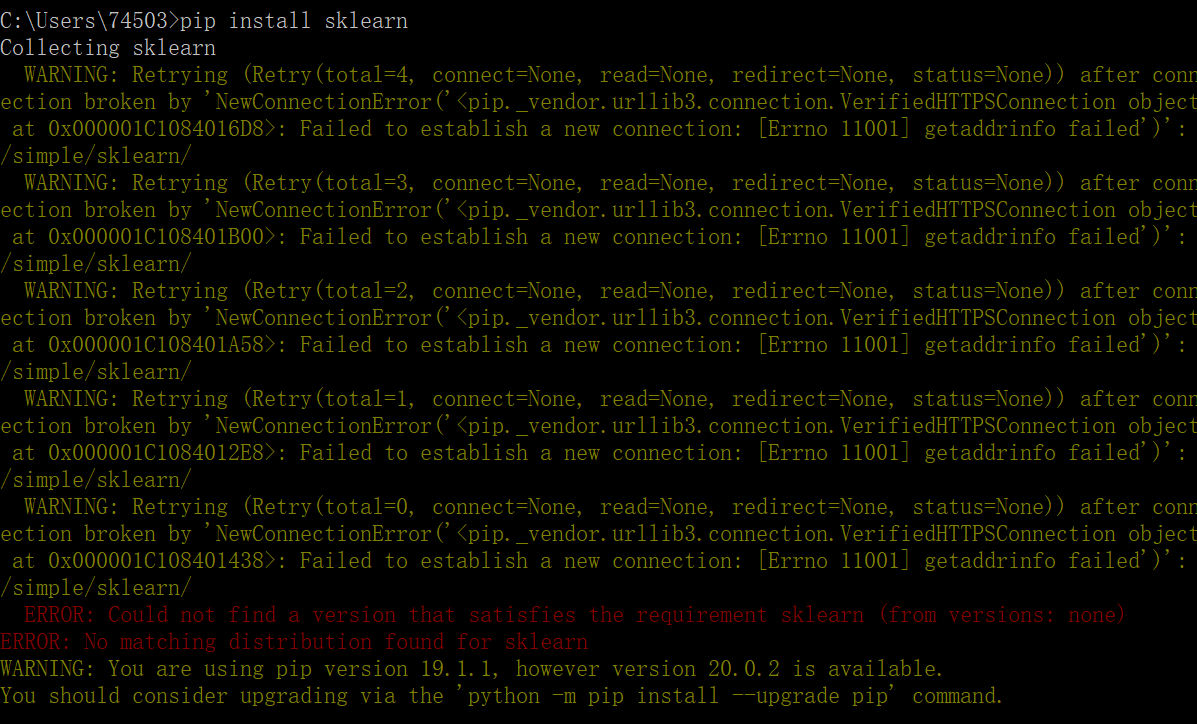
此时，程序将启动一个python3.7.3 64位的安装文件，请您注意如下事项

* 1. 如果您确认您的电脑上已经安装了python，您可以关闭这个安装程序，并请您电脑上的python添加到系统的环境变量中，确保在cmd中输入“python“时它能正常运行
  2. 如果您电脑上没有安装过python，请您按照安装程序的提示来做（一切默认即可）

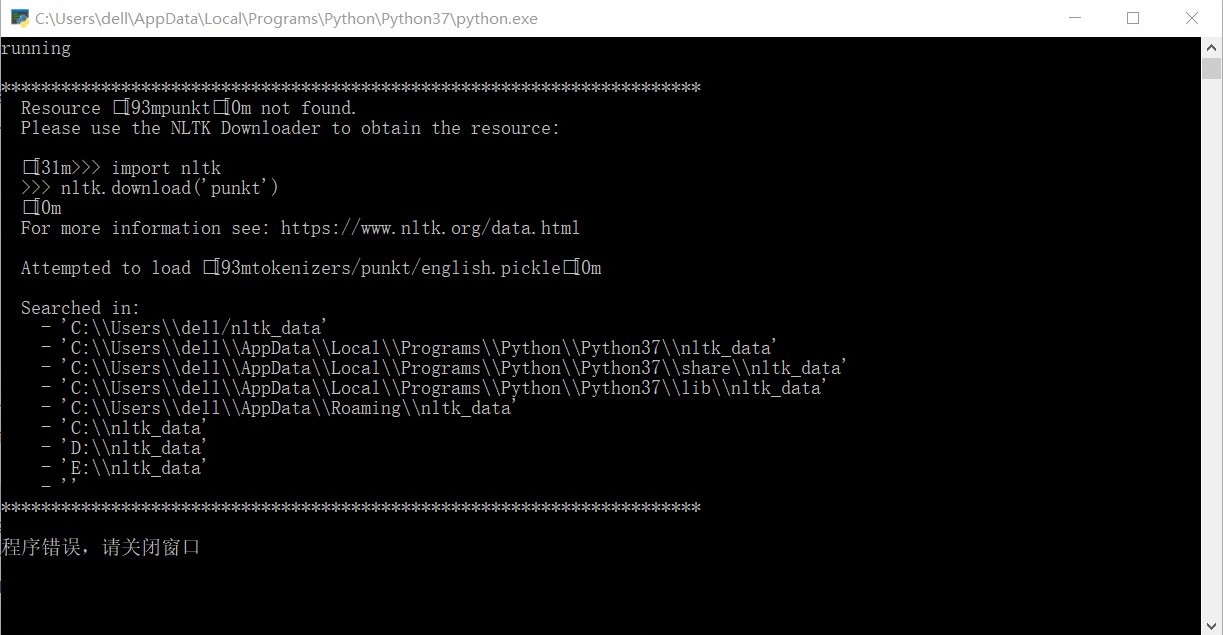
注意下面，Adding Python to PATH选项需要被勾选，否则您在安装之后需要按照a)的提示再进行一次操作。



* 1. 安装完成之后，第一次运行主程序可能会崩溃，关闭程序再运行一次即可

1. 程序多次运行、关闭、运行后仍不能正常运行，python界面出现类似如下状况：
   1. 

这是由于python没能从相应的网站上获取到相应的包造成的；请检查您的网络连接是否有问题，或者您可以自己安装nltk，win32file，win32pipe这三个包；如果上述方法均不能解决问题，您可以参考后文的“我不想使用python或wordnet“

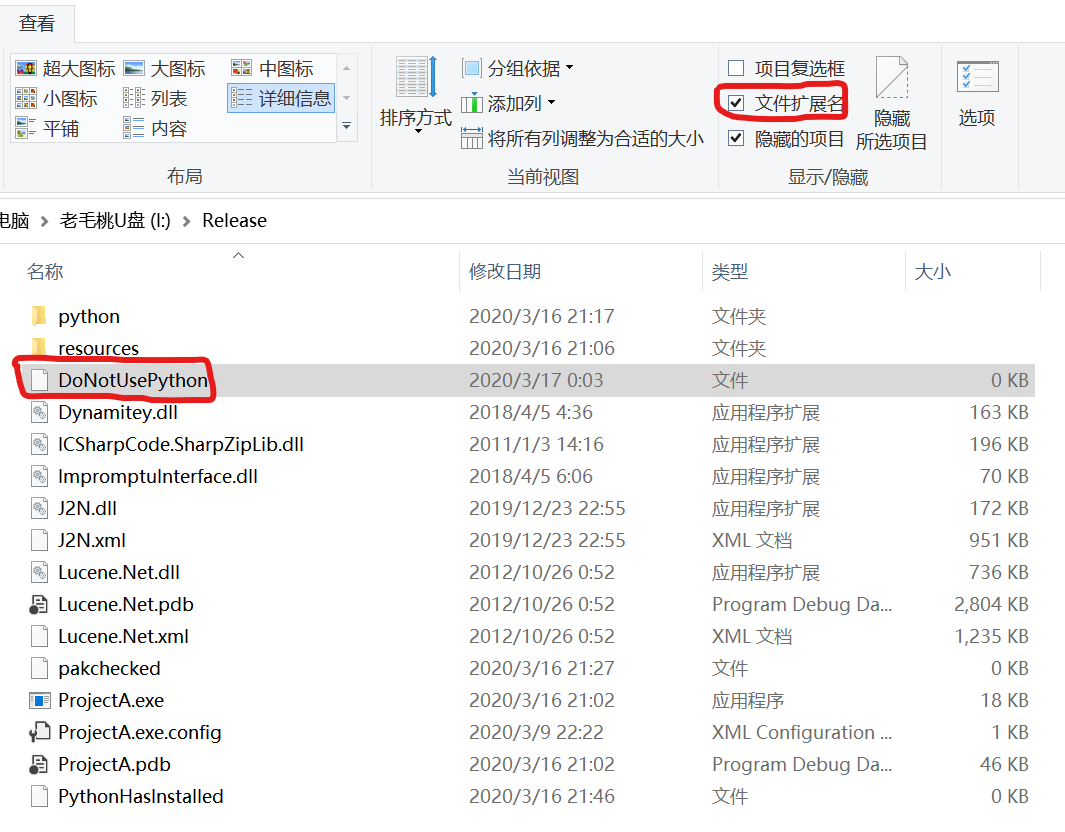


这是由于python在调用nltk时遇到了没有加载的语料库，并且自动拷贝也没有成功；请您将程序目录下的“python/nltk\_dat“文件夹中的数据拷贝在C:\nltk\_data或者上图中任意一个目录中即可；如果这不能解决问题，您可以参考后文的“我不想使用python或wordnet“

1. 我不想使用python或wordnet：

由于种种原因，这个程序的python代码在您的电脑上不能正确运行，那么您可以通过如下方法来消除程序对python的依赖：

在程序的目录下创建名字为” DoNotUsePython”的空文件如下，再次运行程序即可；



注意，它不是” DoNotUsePython.txt”文本文件，默认情况下不能被记事本打开，您可以创建一个txt文本文件后在查看文件的扩展名的情况下将“.txt”后缀去掉

而想要恢复运行python，删掉上述文件即可