本代码以向量空间模型为核心，基于安然公司150位用户50万封电子邮件，成功构建了邮件信息检索系统。

代码实现了对邮件作者（From）、标题（subject）以及内容（subject）三方面的邮件检索功能。**在此声明，索引构建核心环节、向量空间模型核心环节均未借助工具包，所有代码均为个人独立完成。**

以下是对代码运行环境等相关介绍。

编程语言：python

使用到的包：string, nltk, nltk.corpus, nltk.stem, collections, numpy, math, pickle, codecs, os, email.parser, tkinter

运行设备：Inter(R) Core(TM) i5-8250U CPU @ 1.60GHz 1.80GHz

内存大小：8GB

代码文件说明：

allen-p：安然公司150位用户50万封邮件数据集

run.py：程序运行脚本

clean\_text.py：利用nltk等工具清洗读取到的相应文本

email\_handling.py：邮件信息提取、清洗与处理

create\_InvertedIndex.py：构建倒排索引

inquire.py：向量空间模型构建

train.py：训练脚本，用于训练邮件的倒排索引，计算tf\_idf权重

author.pkl：邮件作者索引信息

subject.pkl：邮件标题索引信息

body.pkl：邮件内容索引信息

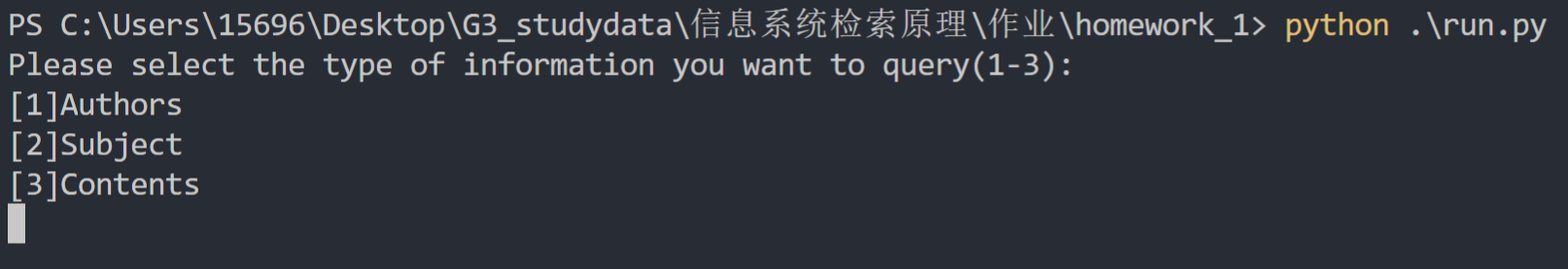
doc\_index.pkl：文件索引列表

doc\_num.pkl：文件总数

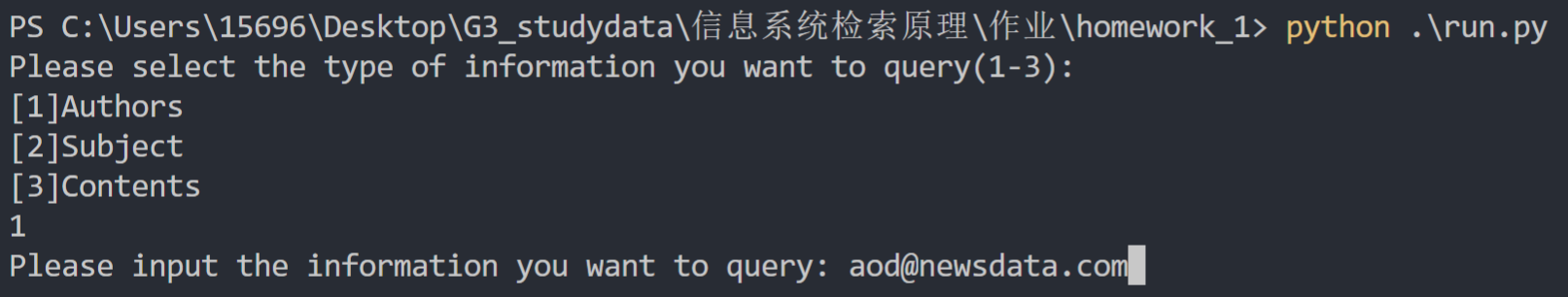
接下来是功能运行截图：

打开终端，输入python .\run.py

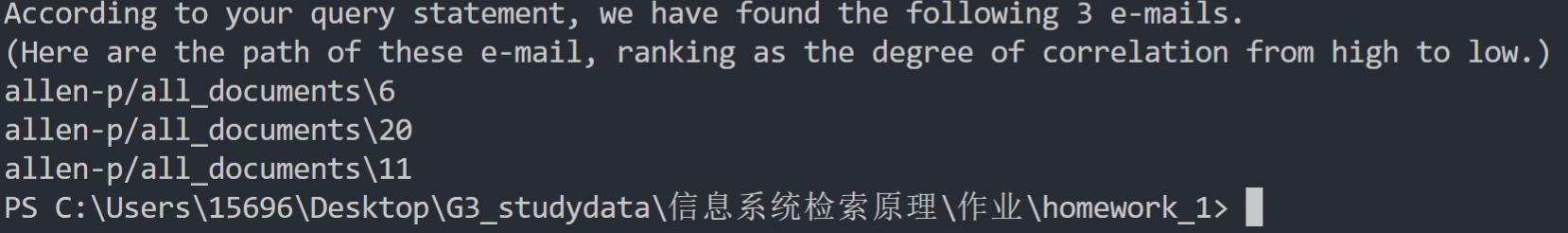
出现界面，首先选择要检索的邮件内容，1到3依次为邮件作者，邮件标题和邮件内容。



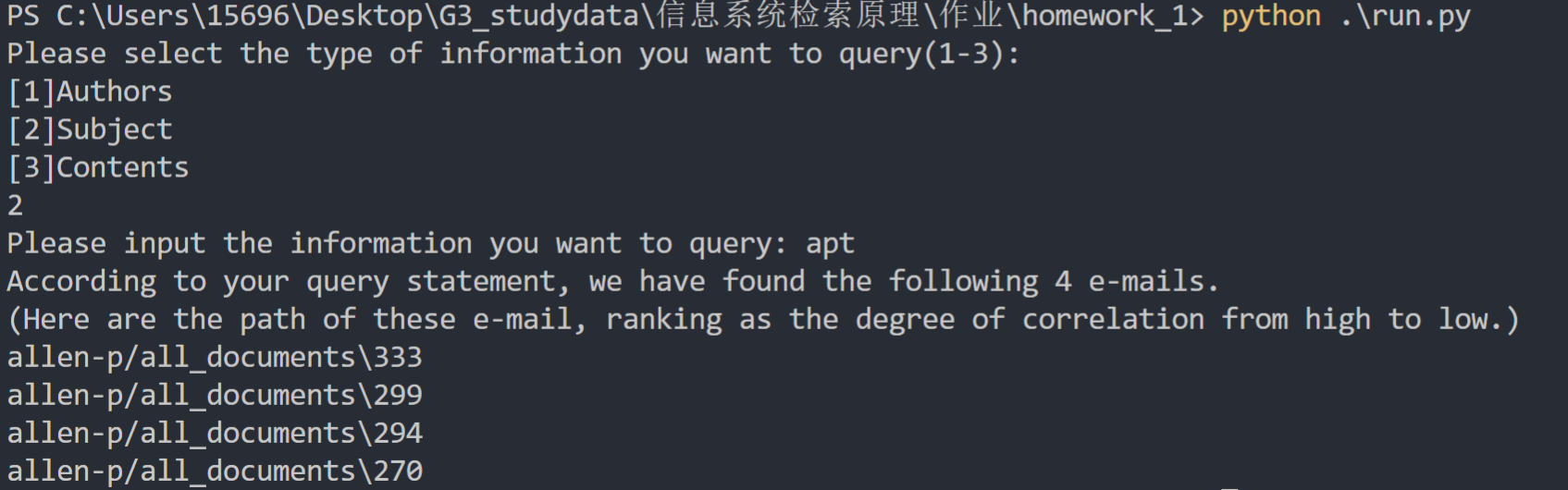
接下来，选择1，输入要检索的邮件作者：



成功得到运行结果：

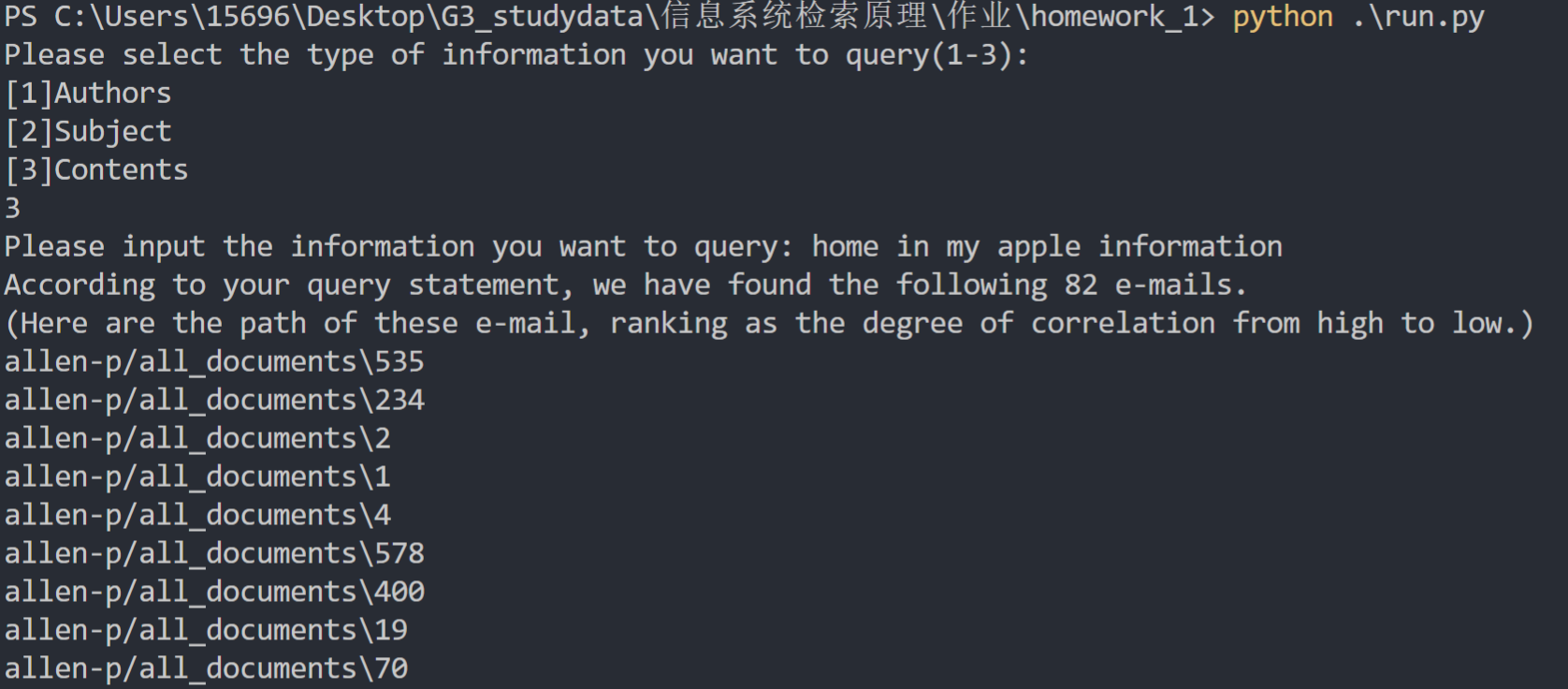


接下来，再测试邮件标题检索：



成功运行。

最后测试邮件内容检索：



成功运行。

综上，即为代码运行实例及环境说明。