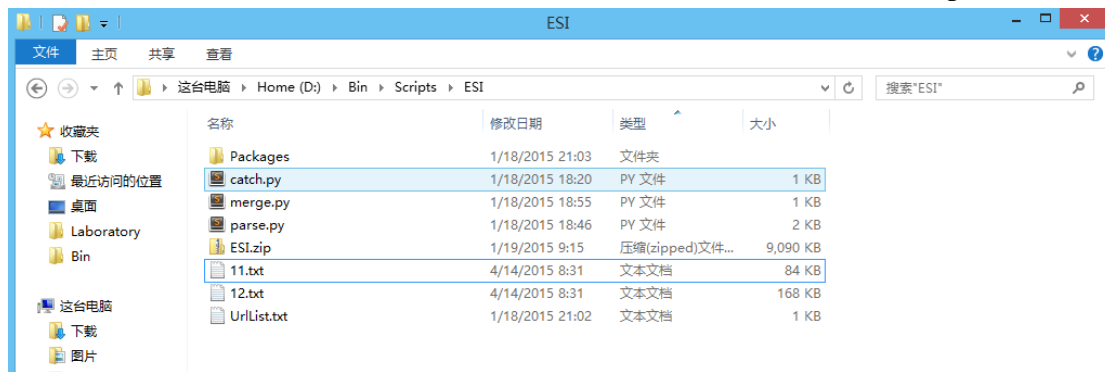
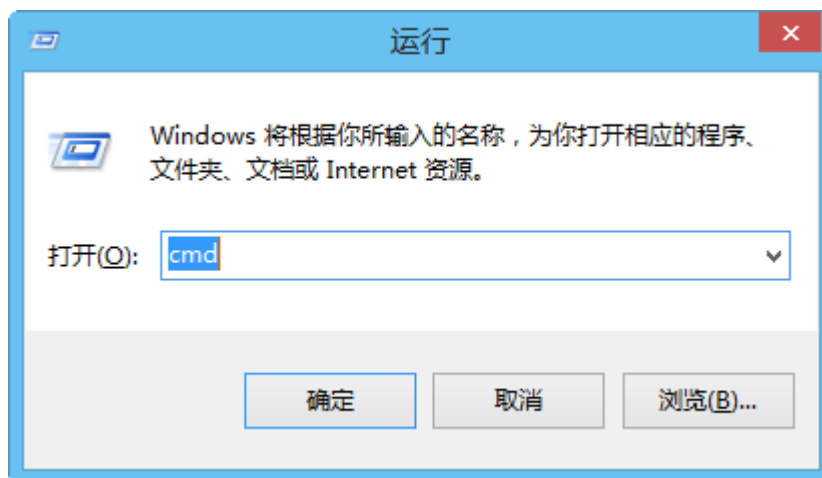


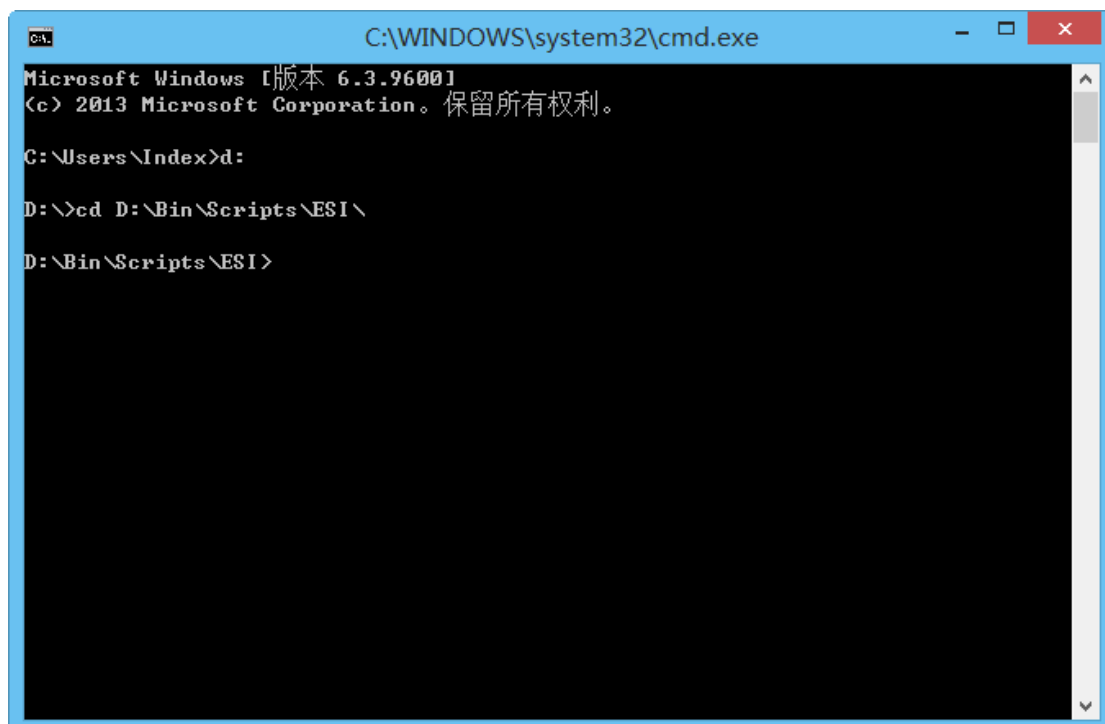
1, 首先确定一下程序所在目录, 如图所示, 我的目录在 D:\Bin\Scripts\ESI\



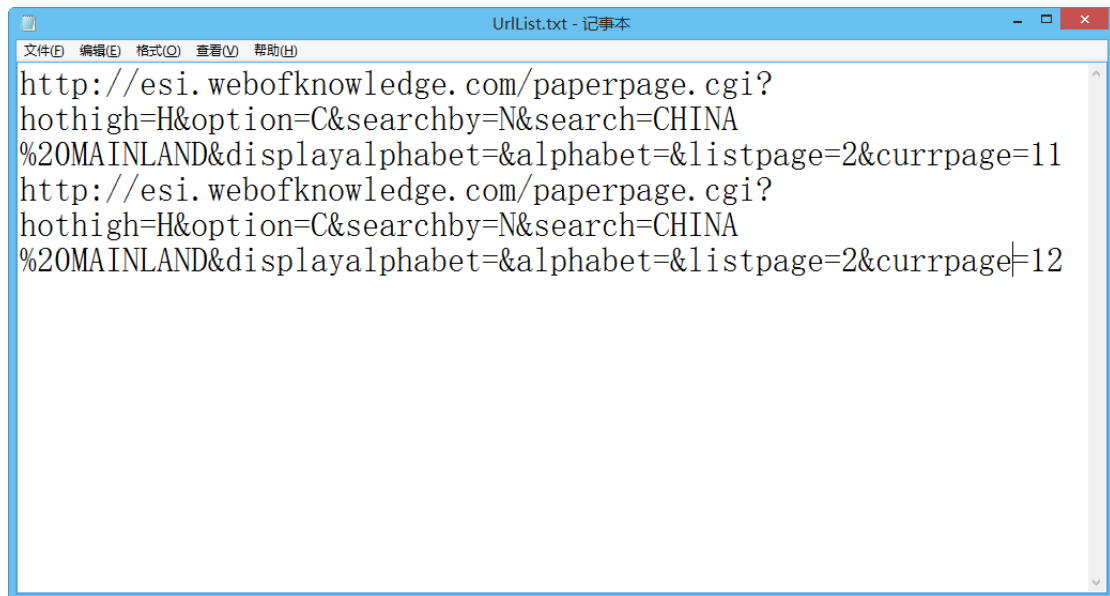
2, 通过 Ctrl+R 快捷键, 输入 cmd 启动命令程序



3, 启动后, 通过“d:”命令和“cd ”两个命令跳转到当前的工作目录

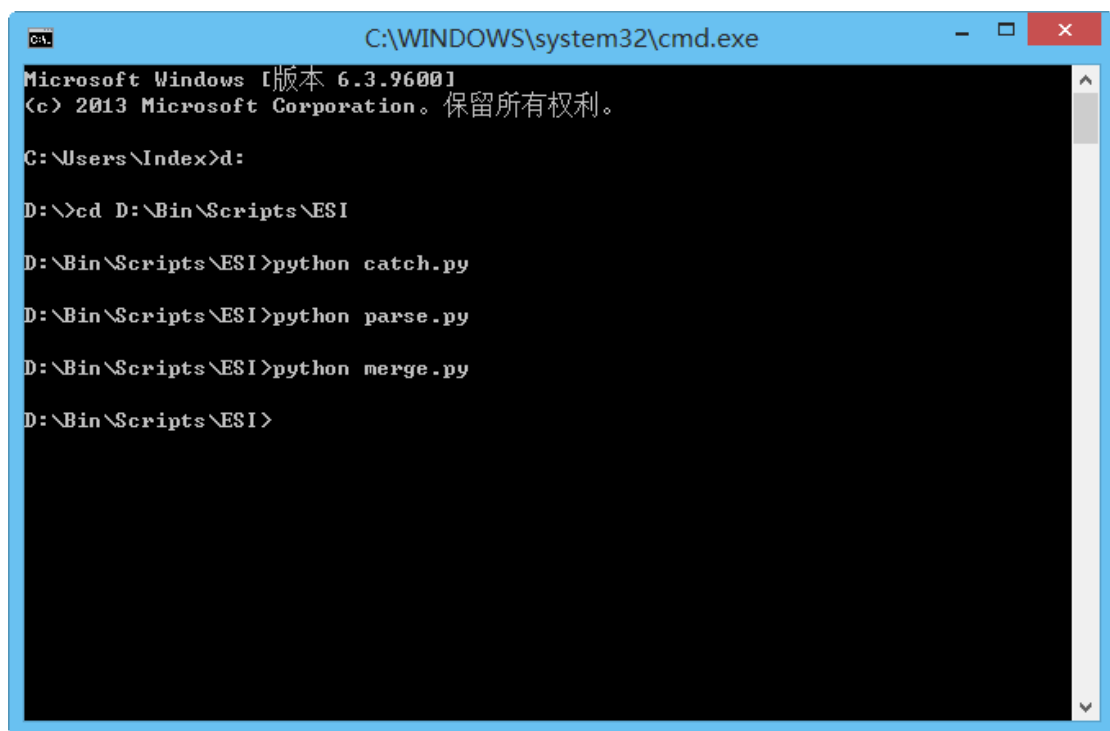


4, 编辑 UrlList.txt 文件, 将需要抓取的网页地址每行一个, 填写进去



```
http://esi.webofknowledge.com/paperpage.cgi?
hothigh=H&option=C&searchby=N&search=CHINA
%20MAINLAND&displayalphabet=&alphabet=&listpage=2&currpage=11
http://esi.webofknowledge.com/paperpage.cgi?
hothigh=H&option=C&searchby=N&search=CHINA
%20MAINLAND&displayalphabet=&alphabet=&listpage=2&currpage=12
```

5, 以此运行“python catch.py”, “python parse.py”, “python merge.py”三个命令即可完成抓取



```
C:\WINDOWS\system32\cmd.exe
Microsoft Windows [版本 6.3.9600]
(c) 2013 Microsoft Corporation。保留所有权利。

C:\Users\Index>d:

D:\>cd D:\Bin\Scripts\ESI

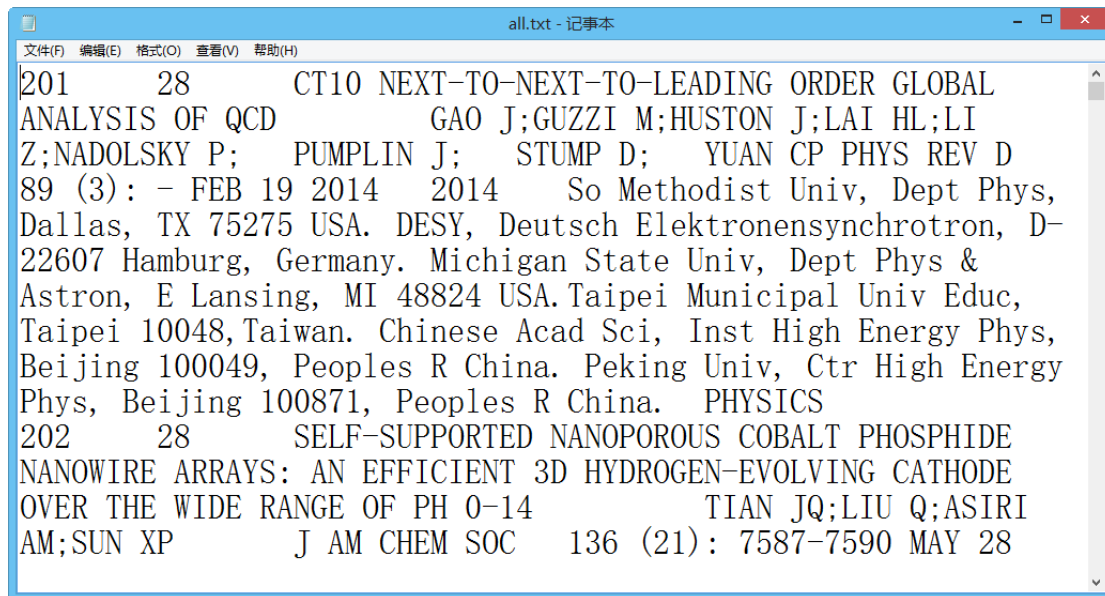
D:\Bin\Scripts\ESI>python catch.py

D:\Bin\Scripts\ESI>python parse.py

D:\Bin\Scripts\ESI>python merge.py

D:\Bin\Scripts\ESI>
```

6, 文件“all.txt”中就是抓取结果, 可以通过复制粘贴操作转移到 excel 文件中。



7, 删除除 UrlList.txt 之外的所有 txt 文件, 修改链接地址就可以抓取其他的网页。