甲方：孙佳音

乙方：刘洪清

现甲方向乙方购买用以在互联网上搜集数据的 Python 程序。经双方协商同意，订立如下协议：

一、乙方在北京时间 2006 年 11 月 6 日零时之前向甲方提交第一版程序。程序应符合本协议第二、三页所提出的所有要求。

二、甲方在北京时间 2006 年 11 月 4 日零时之前向乙方提供一个电影列表文件，供乙方测试第一版程序。

三、甲方于 2006 年 11 月 6 日零时之前收到乙方提供的第一版程序，测试无误之后，向甲方支付人民币壹千二百元整。

四、甲方于北京时间 2006 年 11 月 9 日零时之前根据本协议第二、三页提出程序调试意见。

五、乙方于北京时间 2006 年 11 月 12 日零时之前向甲方提交调试后的程序。

六、甲方于 2006 年 11 月 12 日零时收到调试好的程序，测试满意后，向甲方支付人民币八百元整。

七、本协议一式二份，由甲、乙双方各执一份。

甲　方：孙佳音　　　　　　　乙　方：刘洪清

代表人：　王彦龙　　　　　　代表人：

甲方代表人签字：＿＿＿＿＿＿＿＿＿　　乙方代表人签字：＿＿＿＿＿＿＿＿＿

甲方签字日期：＿＿＿＿＿＿＿＿＿　　　乙方签字日期：＿＿＿＿＿＿＿＿＿

Given a list of movie titles in a text file (the file name is "c:\movies.txt"), each line consisting of one title, the program, which should be written in Python and run on WindowsXP, should read the text file. For each movie title, it should perform the following operations:

1. Go to Boxofficemojo.com and find the movie (using the site's search engine)
2. Retrieve information as listed in the attached Excel sheet BO (this requires one click away from the movie's page)
3. Go to Metacritic.com and find the movie (using the site's searching engine)
4. Retrieve information as listed in the attached Excel sheet REVIEW (this requires many clicks away from the movie's page)
5. Go to http://www.blogpulse.com/trend and key in the movie title in the first box, and then click on the button "Get Trend"
6. Retrieve information as listed in the attached Excel sheet BLOG

There are several things to notice.

1. It is possible that Review Time (date) is not available on the link provided by Metacritic.com. In this case, use the date on http://www.rottentomatoes.com/ . For example, TV Guide does not provide review date. But the reviewer Ken Fox can be found on Rottentomatoes.com at http://www.rottentomatoes.com/author-3260/; all the dates of his reviews are given on this page.

2. Nevertheless, the program should read information from the review link provided on Metacritic.com whenever possible.

3. For Boxofficemojo.com, if the First Day BO and Second Day BO can not be found by clicking the "Daily Box Office" tab, then they can be found on http://www.boxofficemojo.com/daily/chart/?sortdate=yyyy-mm-dd&p=.htm, where yyyy-mm-dd is replaced by the movie's release date.

4. It is also possible that some movie titles are not available at either or both websites. In that case, just skip that movie. If some information is not available for a particular movie, just fill "N/A" in the corresponding cells.

甲　方：孙佳音　　　　　　　　乙　方：刘洪清

代表人：　王彦龙　　　　　　　代表人：

甲方代表人签字：_____　乙方代表人签字：_____

甲方签字日期：_____　　乙方签字日期：_____

5. I have provided an example (without counting the Review Length, which is the number of words contained in each review) in the Excel sheet. Let me know if you have any questions after reading the Excel sheet. The due date for delivery is in two weeks.

6. As Excel has a limit on the number of rows each sheet can hold, the output of the program should be in three tab-delimited text files, each corresponding to a sheet in the excel file (as a result, please remove tabs in the strings before entering into these fields so it can be opened properly).

甲　方：孙佳音   乙　方：刘洪清

代表人：　王彦龙   代表人：

甲方代表人签字：_____   乙方代表人签字：_____

甲方签字日期：_____   乙方签字日期：_____