1. Introduction

The project is about crawling IMDB's TOP 250 movies list on http://www.imdb.com/chart/top?
ref = nv wl img 3 and then exporting it as a txt file. The information includes the ranks, the titles, the years and the rates of the movies.

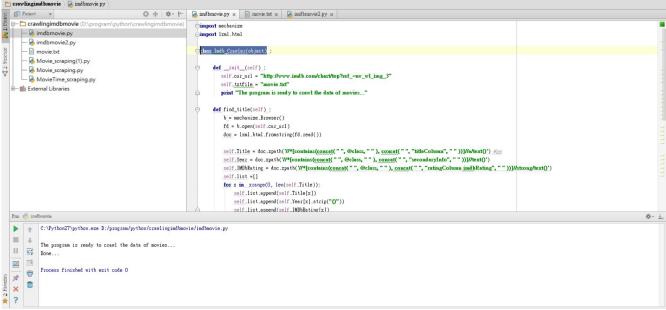
2. Requirements

- (1) mechanize: it needs to be run on Python 2.7.
- (2) lxml

3. Description of the Python program

- (1) class Imdb Crawler(object): crawling the top 250 movie list.
- (2) def init (self): setting intial values
- (3) def find_title(self): using the module mechanize to get the html of the website. Xpath helps find the informations I want for the project, which are lists. Put those data into one list called list.
- (4) def writefile(self): writing the information I want into a txt file called movie.txt
- (5) def main(): the main funcation

4. Screenshots of the program output





5. Conclusion

I used 6 Python main elements to write the program such as data structure list, classes, importing external modules , functions, list comprehension, and file input and output. The output is to check whether the program runs properly. The infomation I want to crawl is exported to the file movie.txt

6. Python program imdbmovie.py