1.首先打开redis-server

$redis-server

2.插入书籍链接：

$ cd book\_links

$ python insert.py xxx begin end

xxx 是自己分到的数据，为从1～11的数字。

begin 为起点

end为终点

因为本次爬取的书籍数据量大，一次如果插入过多数据容易出错。所以我们在插入链接时可以指定范围。

比如：

我要爬取第1个文件的前10000个链接：

python insert.py 1 0 10000

注意：0～9999会被插入，但10000不会，所以第二次可以使用：

python insert.py 1 10000 20000

来爬取接下来的10000个链接。

3.开启爬取书籍信息的爬虫：

scrapy crawl bookContent

4. 爬取完书籍信息后，可以导出书籍信息数据：  
python process\_content.py

会生成一个content.csv的文件。

5.上一步完成后，开启爬取评论区链接的爬虫reviewLinks

scrapy crawl reviewLinks

6.上一步完成后，开启爬取评论的爬虫review

scrapy crawl review

7.此时可以导出评论数据

$ python process\_review.py

会生成一个csvfile.csv