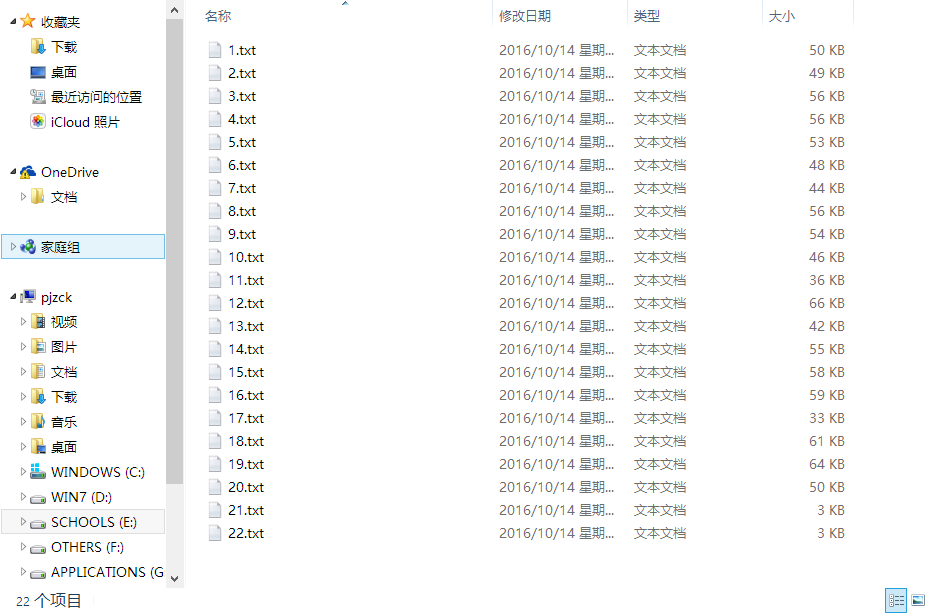
第二周作业 计算机学院网人事处网络公告节选程序

周晨恺 3140102277

程序说明：

能够从 [www.readers365.com](http://www.readers365.com) 下载一本书的所有章节，比如 <http://www.readers365.com/jinyong/08/> 对应的是金庸的小说《书剑恩仇录》。下载之后，我们可以在 java 文件所在的根目录下的 novel 文件夹中找到对应章节的 .txt 文件，按照章节顺序从 1.txt 开始编号，如图所示：

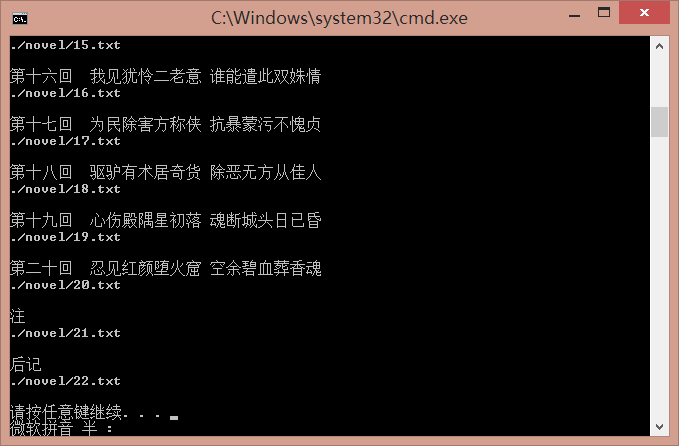
然后点开其中的文件就可以开始阅读小说。

其中，在使用 txt 阅读小说的时候需要打开自动换行的功能，否则会出现单行有非常多的文字的情况。这是因为在下载小说的时候没有进行自动换行处理的结果，而是将所有的一段文字都放在同一行的原因。

周晨恺 3140102277

2016.10.16

程序运行实例：



程序使用说明：

将压缩包解压后的根目录下，修改 Test.java 源代码中第 9 行的网址，就可以实现对不同通知的解析。

现在暂时只支持对网站 [www.readers365.com](http://www.readers365.com) 中对应的小说和文章进行解析和下载。实例网站地址如下：<http://www.readers365.com/jinyong/08/> (书剑恩仇录)

程序实现说明：

整个程序总共实现了三个类，FileManager 类、 GetHtml 类和 ParseHtml 类。此外，使用 Test 类作为整个程序的主入口。

FileManager 类主要负责后台对文件的管理，在这里主要的功能就是将解析得到的字符串写到对应的文件中。之后的扩展优化可以实现对多本小说的本地文件的维护，包括对文件的创建删除，按照作者分类管理小说等。

GetHtml 类主要负责从对应的网站获取相关的 Html 文本，然后将其转换成方便处理的 String 对象类型传递到后面的操作中。

ParseHtml 类主要负责对得到的 Html 文本进行解析处理。整个程序的解析通过正则表达式实现，通过对网页源代码的分析得到相关的文本处理标记，然后使用 Java 自带的正则表达式包中的 Pattern 和 Matcher 进行正则表达式的处理和截取，将所得到的结果放到 Vector 容器中并传递出去。现在实现了对小说目录中所有章节目录网址以及小说章节名的获取，对小说文本的去广告处理和内容整理。

优化展望：

一个是希望能够支持更多的小说网站的转码支持。

第二个是希望同一个网站不同小说书本之间的文件管理。

第三个是希望能够实现对小说的搜索以及从不同小说网站下载小说。