**《Python程序设计》课程设计要求：**

基本要求：自己选择一个网站，用Python语言编写一个爬虫程序，将其网站中的文字、图像或视频等信息抓取到本地文件或文件夹中保存。

说明：

1. 强调：抓取的是网站不是网站的首页。抓取的内容一定要分布在整个网站的多个页面和多个链接中。
2. 以抓取小说网站为例，必须抓取一系列小说（不是一部小说）的篇名、作者、出版单位（或首发网站）、出版时间（或网上发布时间）、内容简介、小说封面图画、价格、读者评论或评分等多项信息，并将上述信息组织成表格形式（可以是csv、json、excel等）加以保存。另外，还可以深度抓取某部小说的多个章节或全部章节进行分词和词频统计。
3. 抓取内容的命名要与抓取内容相衔接。
4. 一个行政班的同学所选网站不能重复，选好的网站域名要到学委那里申报并获批，只有获批同意的网站才能认可其作业的合规性。审批原则是：如有重复，先报先得，因申报晚了而导致重复的要更换网站，直到不重复为止。只要是主域名相同，则被界定为同一网站。
5. 完成基本要求只能获得60-75分左右，加分需要完成下列更多任务。
6. 在完成基本功能后，如果程序中加入了反爬技术的可有进一步的加分（5分左右，具体分数视所用技术情况和编写的程序质量，后面同）。
7. 在完成（5）（6）两项的基础上，如果采用正则表达式技术的再有进一步加分（10-15分）
8. 在完成（5）（6）（7）三项的基础上，如果能将获取的有意义的文字文档进行分词并进行词频统计则再获加分（5-10分）
9. 文字、表格、图像等各种数据可以自行选择不同文档格式来保存，这些格式可以是记事本、csv、json、xml、excel等。
10. 上述（7）（8）（9）所涉及的知识，因为课时少的原因，将由同学们自学获取。
11. 提交的作业包含程序、存放程序执行结果的文件或文件夹、相关说明文档等三份材料。程序执行除能获得所需结果外，清晰明白的程序注释、格式漂亮的程序结果数据组织形式和高质量的说明文档将是获取区间分数中高分的依据之一。
12. 提交的课程设计大作业（包含程序、执行结果和说明文档）的压缩文件命名要有学号、姓名、班级三要素。压缩文件请发送到老师的ftp://10.5.1.5，用户名：dch；密码：dch。
13. 上面要提交的材料最早可在寒假开学前提交，开学后，任何时候都可提交，最迟提交时间不能超过课程结束后的第二周周末。第二周周末后再提交的，每延迟一天，扣除平时成绩1分（指整个课程成绩100分中从平时成绩中获得的1分）。提交时间越早，平时成绩会有加分。
14. 如果在开学前提交并能获得满分的同学，可以向老师申请免听本课。如果获得接近满分的同学可以申请免听理论课，但要上机。
15. 在课程结束前提交的同学可以获得较早答辩机会，以便获得更好的成绩。
16. 如还有不明确的地方，可反馈给班级学委，老师将会在开课后回复。