**云南大学软件学院**

**实 验 报 告**

课程： Python语言设计 任课教师： 何臻力 实验指导教师： 何臻力

姓名： 何志颖 学号： 20151120237专业： 软件工程 日期： 2018-6-24成绩：

**实验十 Python爬虫**

1. **实验目的**
2. 理解爬虫。
3. 学会使用python原生爬虫。
4. 学会简单使用使用python爬虫框架Scrapy。

**二、****实验要求**

（1）必须使用Python3.X版本。

（2）使用Python自带的IDLE或PyCharm为开发工具。（自行安装）

（3）请自行完成实验内容，切勿抄袭。

（4）代码要求有注释，注释量不低于20%。

（5）不能自行使用与课程内容无关的第三方库降低实验难度。

（6）提交实验报告：

a. 纸质版：实验报告完成后，打印出来，下次上实验课前提交。

b. 电子版：实验源码和电子版实验报告，一起打为一个压缩包，上传到FTP下对应目录。(<ftp://113.55.4.20> user:hezhenli\_std , password:std)

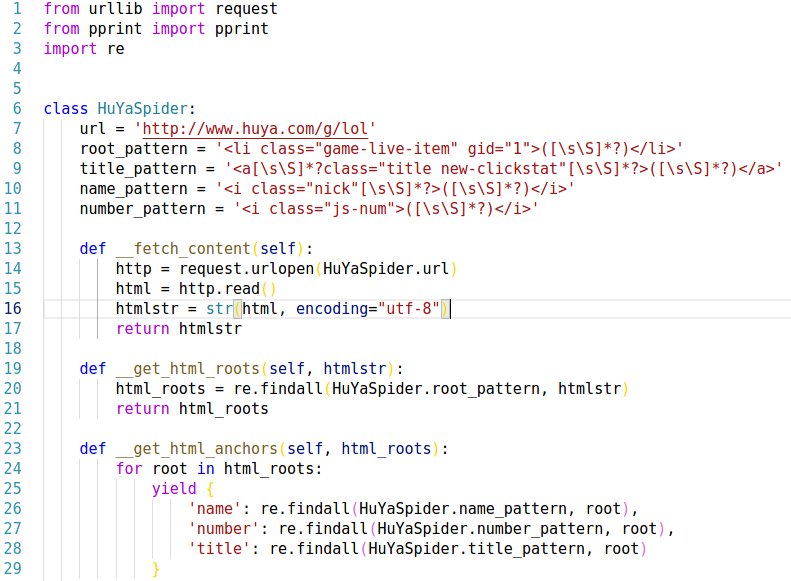
**三、实验内容**

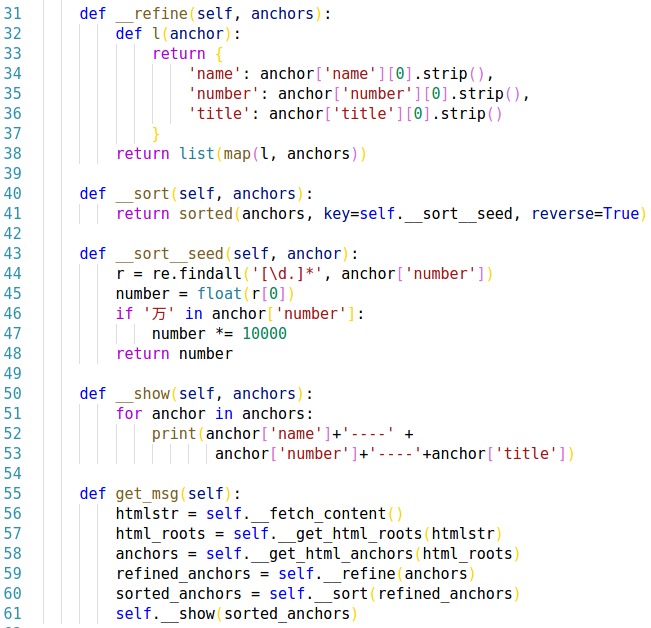
1、请任意选择一个你喜欢的网站（可以是新闻网站、直播网站、论坛等），使用Python原生urllib库，实现一个网络爬虫。要求可以爬抓网页的重点信息，并给出一定的数据分析（比如：最热的新闻头条、主播人气排行榜等等）。

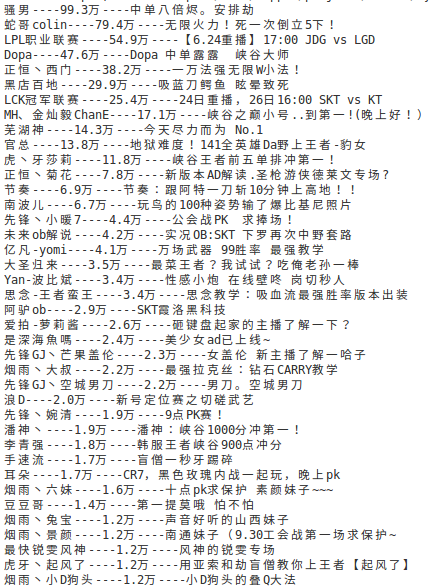
1. 请再使用python爬虫框架Scrapy实现第一题。要求：a.使用items.py定义待抓取域的模型。b. 使用pipelines.py定义爬抓数据的存储方式。

**四、关键实验步骤（请粘贴关键步骤、代码、实验结果）**

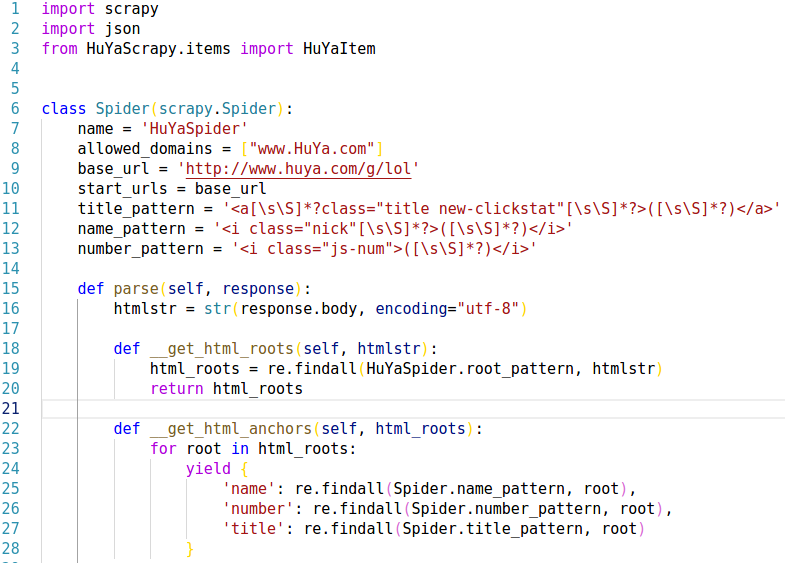
1、请任意选择一个你喜欢的网站（可以是新闻网站、直播网站、论坛等），使用Python原生urllib库，实现一个网络爬虫。要求可以爬抓网页的重点信息，并给出一定的数据分析（比如：最热的新闻头条、主播人气排行榜等等）。

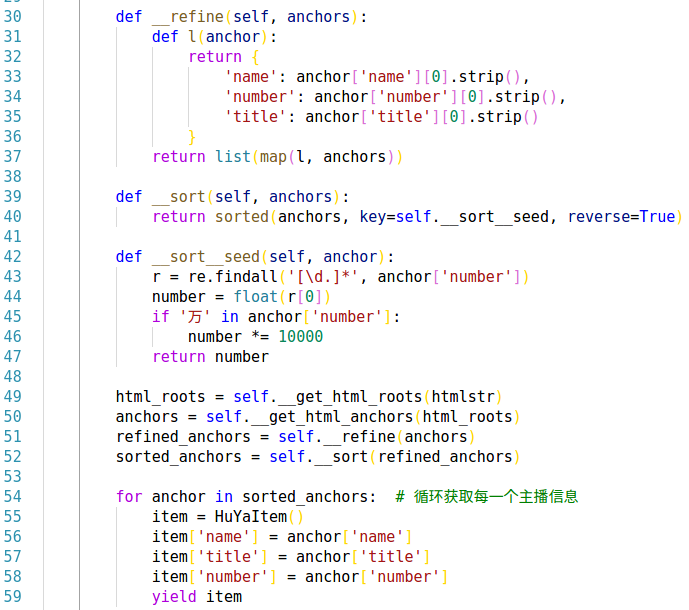




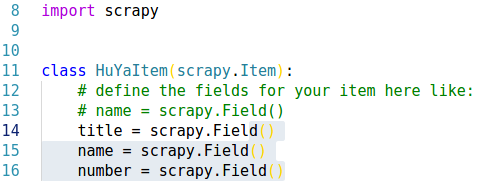


1. 请再使用python爬虫框架Scrapy实现第一题。要求：a.使用items.py定义待抓取域的模型。b. 使用pipelines.py定义爬抓数据的存储方式。

Spider:



Items:



Pipeline:



Lol.csv

