本周作业

作业1

1. 首先根据自己本周所学的知识，进行总结，写一个makedown笔记，梳理爬虫的主要流程和一些步骤，笔记需要截图给我

作业2

1. 写一个爬虫爬取西刺免费代理网站<http://www.xicidaili.com/nn>

要求：



1. 提取上面箭头所指的字段
2. 将提取的免费代理写入数据库（数据库结构自己设计）
3. 存入前先测试该ip是否有效，只存储有效的ip（如何测试ip有效，自己搜集资料）
4. 写一个工具类实现从数据库提取ip

1）.在类中定义一个方法可以实现随机获取1个ip

2）.在类中定义一个方法可以实现根据输入的页码来获取某一页的ip，比如方法中传入1，获取前0～10个，传入2获取10～20

作业3:

https://news.baidu.com/

获取百度新闻首页下分类的文章内容（分类如下）

热门要闻、国内、国际、娱乐、体育、财经、科技、军事、互联网、探索、女人、健康的数据保存至本地文件夹

要求：

1.分析页面先根据每个分类创建一个文件夹（代码实现）

2.获取每一个分类下的文章、以文章的标题命名html文件，并将获取到的html文本放入对应分类的文件夹下。

作业4:

预习印象笔记课件2.2和2.2.1 Requests模块

1. 安装requests：pip3 install requests

作业在周末下午4点前交给组长，组长统一交给我。