1. 复习mongodb的聚合操作命令

1.首先新建数据库，数据库名称为：jobs

2.在jobs中新建一个collection名称为 jobdesc

3.在jobdesc中添加下列数据

jobname workyears tags needpople age

python 3 [python,scrapy,mongodb] 3 20

iOS 3 [iOS,mysql,mongodb] 2 25

java 2 [java, mysql] 1 23

python2 5 [python,scrapy] 5 24

python 2 [scrapy,mongodb] 2 30

1. 使用$group 根据jobname 进行分组，统计每种职位的总数

2. 使用$group 根据jobname 进行分组，统计每种职位需要的总人数

3. 使用$group将集合中的所有文档分为一组，统计职位总数，算出平均年纪age

4. 使用$project修改文档的结构，只展示jobname和needpople 两个文档字段

5. 使用$match函数筛选年纪在20～23岁的职位数量包含20和23，并且只返回jobname和age

6. 使用$match函数筛选年纪在20～23岁的职位数量包含20和23，并且只返回jobname和age

7. 使用$match函数筛选年纪在20～23岁的职位数量包含20和23后，通过jobname分组，并且显示每种职位的平均年龄

8. 使用limit和skip结合使用，返回集合中第二条开始的后2条数据

9. 使用$sort筛选出集合中职位年龄要求在23～30的职位，并根据职位需要的人数做降序排列

10. 使用$unwind 将文档中的tags字段拆分成多条后，返回拆分后的各职位的总数量。

作业一：

汽车之家：

目标url：<https://www.autohome.com.cn/all/>

注意：使用使用lxml解析库会出现乱码问题，这里使用html.parser来python自带的解析库



1.获取箭头指定的数据（图片连接、标题、时间、观看量、评论量（要获取评论量）、完整文章详情连接、简介内容）保存mongodb数据库，