204 周考题（第二周）

1.列举你目前知道的数据库，并根据关系型和非关系型进行分类,并且简单的介绍一下 。

2.说说HTTP协议有状态吗？是如何保持会话的？

3.列举下我们所学习到的urllib下的三大模块request、error、parse分别有什么作用？

4. 简单描述一下decode、encode的作用,假设有一个UTF-8编码的字符串为utf8Str，先将该字符串转换为Unicode编码，再将Unicode编码格式字符串转换成GBK，再将GBK编码格式字符串转化成Unicode，最后将Unicode编码格式字符串转换成UTF-8，用代码实现上述转换

５．使用requests 模块构造一个get请求，链接如下，并打印响应的状态和响应的结果，假设获取到的是一个json字符串，将结果转换为python对象。

<https://movie.douban.com/subject_search>

参数如下：[search\_text](https://movie.douban.com/subject_search?search_text)=美剧&cat=1002&page=1

(温馨提示request.get中的参数有url、params、headers、timeout、verify，proxies

,先解释每个参数的含义，再使用它构造一个请求,要求使用代理)

６．使用urllib发起请求时，我们如何处理ｇｅｔ请求后面的参数，如何处理ｐｏｓｔ表单中提交的数据？

７．python与mongodb交互我们要安装那个库、如何安装？

８.解释ｍｏｎｇｏｄｂ下列运算符的含义

　$lt $gt $gte $lte $in $ne $nin $or

９. mongodb与python交互部分

9.1手写两种链接方式（温馨提示：pymongo.MongoClient( )）

9.2.获取mongodb中的名为bole数据库

9.3.获取bole数据库中的article集合

9.4.向集合中插入文档数据格式如下，实现多条插入和单条插入

title： 　　　star： 　buy\_num: author：

向快乐出发 　２００ 　　５０ 　　春天里

　puython基础　　３０００ 　　１００　　　老翁

　爬虫入门　　　 ５００　　　　２００　　马云

9.5获取article集合中的第一条文档

9.6 将集合中爬虫入门的文档数据的buy\_num修改为１５０

9.7获取article集合中的所有文档ｓｔａｒ做升序排列，并打印出结果

9.8用mongodb数据库语句查询article集合实现一个分页的功能，要求第一页返回0～10条数据，第二页返回10～20条数据，第三页返回20～30条数据，以此类推

10. mongodb创建超级管理员在那个数据库下添加账户信息？

创建一个账户要分别要写入哪些内容？试着写出创建一个账户名为 201，密码为123456的超级管理员的语句。

11.如何手动备份ｍｏｎｇｏｄｂ的数据库？备份语句是什么？

12. 根据下面的HTML片段，使用正则表达式从中提取出a标签的href对应的链接，和a标签的text文本内容

<a href="https://movie.douban.com/subject/" data-moreurl="onclick',subject\_id:'6538606',from:'mv\_subject\_search',is\_tv:'0'})&quot;" class="title-text">

美剧大佬 Showrunners: The Art of Running a TV Show‎ (2013)

</a>

（提示，构建正则对象的方法为compile函数）

1３.解释xpath语法中的下列符号的含义

.

/

//

@

text()

[@class=””]

[@id=””]

/book/div[1]

1４.解释CSS选择器中下列符号的含义

#

.

.name

get\_text()

以下ａ表示a标签

a.string

a.attrs[＇href＇]

a.attrs

a.article

#item-1

１５．描述下列正则符号的含义?

　　．　？　＊　＋　［a-z］ {n,m} ^ | \d

1６.简单介绍一下ｓｃｒａｐｙ由哪大几部分(模块)组成？

１7．列举出你再写爬虫时所用到的第三方的库．

１8．列举ｒｅ模块下的常用的方法，并说明它们的作用．

19．列举一些常见的请求状态码，并说说含义．

20.利用自己的语言解释一下python的多线程与多进程，描述他们的优缺点以及使用的场景．