1. json包：

Import json

用json.loads(),可以将字符串解析为json对象

然后可以由获取的json对象加上[‘xxx’]即可获取相应的值，类似dict

**（注意是loads不是load！！！！！！！！！！！！！）**

1. datetime包：

Import datetime

datetime.datetime表示日期时间，格式如：YY-MM-dd HH:mm:ss

datetime.date表示日期，格式如：YY-MM-dd

用datetime.datetime.now()可以获取当前时间的datetime对象

datetime.datetime.fromdatestamp()可以将括号内的时间戳转换为datetime对象

datetime.datetime.strptime(date\_string, format)可以将字符串根据格式转换为datetime 对象，而datetime.datetime.strftime()则正好相反

1. time包：

Import time

Time.time()为返回当前时间的时间戳(从1970年开始)

time.strptime(xxx,fmt)可以将一个str按照格式转换为一个时间元组对象 (time.struct\_time)

Int(time.mktime(xxx))可将一个time的时间元组转换为时间戳

而datetime想要转换为时间戳，可以调用datetime.timetuple()将其转换为时间元组，

再调用上述方法即可变为时间戳

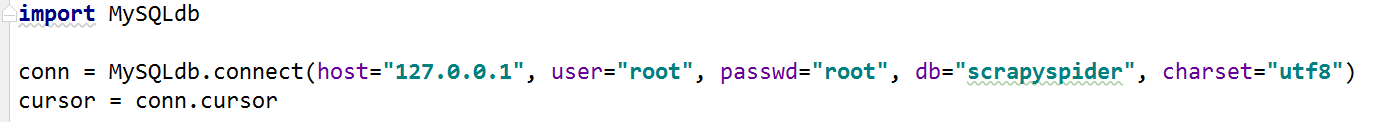
Time.sleep(xx)可以起到阻塞指定时间的作用

1. w3lib.html包：

**from** w3lib.html **import** remove\_tags

用remove\_tags方法就可以把一段爬下来的字符串中的标签给去掉

1. MySQLdb包:用于与mysql数据库的连接



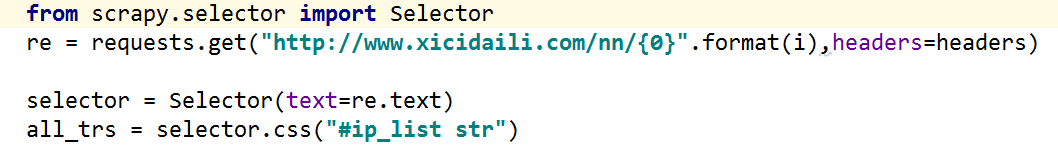
当要运行sql语句时调用cursor.execute(sql),并在之后调用conn.commit()用于提交

需要注意的是：若要将sql语句写活，可以结合str.format()来写活；commit方法是conn的，而不是cursor的

1. requests库：import requests

requests.get(url)即可返回一个访问了url的response，剩余详情见scrapy爬网页细节处的第2、3条

1. Selector：我们之前在spider中是直接使用的response的css和xpath方法来提取的，这是因为那里的response是scrapy.Request()发送给下载器后返回的响应response包装scrapy.Selector后而成的。当我们不在Spider中发送请求时，用的是requests库，用requests.get(url)返回了一个response，它当然是没有包装Selector的，我们可以:



等于是用selector包装了response后，直接用Selector来替代response进行操作

8.关于list和tuple:其中list是可改变对象，tuple是不可变对象。而list()在对zip对象进行强转时返回一个强转后的list对象，而且把这个zip对象内部也被改造了，即当我们再次对该zip对象强转时会返回一个[]