**Python实现定时爬取任务**

# 本节课使用的python模块:

**os模块的作用：**

os，语义为操作系统，所以肯定就是操作系统相关的功能了，可以处理文件和目录这些我们日常手动需要做的操作，就比如说：显示当前目录下所有文件/删除某个文件/获取文件大小……

**os模块的常用功能函数：**

os.name () #显示当前使用的平台

os.getcwd() #显示当前python脚本工作路径

os.listdir('dirname') #返回指定目录下的所有文件和目录名

os.remove('filename') #删除一个文件

os.system() #运行shell命令,打开一个新的shell，运行命令，当命令结束后，关闭shell

**time模块**:

时间处理模块, 时间处理有关的模块包括：time，datetime以及calendar。今天主要讲解time模块。

**time模块的常用功能函数：**

time.time()： 返回当前时间的时间戳

time.sleep(secs)： 线程推迟指定的时间运行。单位为秒

time.localtime([secs])：将一个时间戳转换为当前时区的struct\_time。secs参数未提供，则以当前时间为准。

time.strftime(format[, t]):把一个代表时间的元组转化为格式化的时间字符串。

**urllib2模块常用功能函数:**

urllib2.Request(): 向服务器发送请求并且接收响应

urllib2.urlopen(): 以类文件的形式将响应内容打开

urllib2.read(): 文件的读方法

**re正则模块常用函数:**

**compile(source, filename, mode[, flags[, dont\_inherit]])**

**参数source：字符串或者是Ast对象**

**参数filename：代码文件名称**

**参数mode：指定编译代码的种类**

findall():返回列表中每个元素包含的信息。

# python index()方法:

# index() 方法检测字符串中是否包含子字符串 str ，如果指定 beg（开始） 和 end（结束） 范围，则检查是否包含在指定范围内，该方法与 python find()方法一样，只不过如果str不在 string中会报一个异常

**sched模块是一种调度（延时处理机制）**

schedule.enter(参数)

第一个参数，整数或者浮点数，代表多少秒后执行任务

第二个参数，优先级，0代表优先级最高

第三参数，要执行的任务，要执行函数的函数名

第四个参数,传入执行任务函数名的参数，用小括号包起来,没有参数可以传一个空括号