if \_\_name\_\_ == '\_\_main\_\_'的意思是：当.py文件被直接运行时，if \_\_name\_\_ == '\_\_main\_\_'之下的代码块将被运行；当.py文件以模块形式被导入时，if \_\_name\_\_ == '\_\_main\_\_'之下的代码块不被运行。

#HTTP的端口号:80

#HTTPS的端口号:443

#邮箱的端口号:25

#数据库的端口号：3306

#微信端口号：80

#URL的完整格式：

#1.协议(http/https/ftp..)

#2.服务器的IP地址或者域名。

#3.服务器的端口(80,8080等等)

#4.访问资源的路径。

#5.参数，发送给http服务器的数据。

.通用网络爬虫

2.聚焦网络爬虫

3.增量式网络爬虫

4.DeepWeb爬虫

..........

+ def get\_name(self):

return self.\_\_name

def set\_name(self,name):

if isinstance(name,str):

self.\_\_name=name

else:

raise ValueError('name is not "str" type!')

connect=sqlite3.connect('qsbk.db')

cursor=connect.cursor()

create\_table="create table baike(id INTEGER PRIMARY KEY ,nick\_name VARCHAR ,age VARCHAR ,content VARCHAR ,smile\_num VARCHAR ,comment\_num VARCHAR )"

cursor.execute(create\_table)

for x in range(1,10):

download\_qsbk(x)

connect.commit()

cursor.close()

connect.close()

drop table tabname

3.查询：select \* from table1 where 范围

4.插入：insert into 表名(字段名1,字段名2) values(值1,值2)

5.删除：delete from table1 where 范围

6.修改：update 表名 set 修改的字段名=修改的字段值 where 范围

7.查找：select \* from table1 where 字段名 like ’%value1%’ N%匹配以N开头 %N匹配以N结尾 %N%匹配包含N

[a,b]% 以a或b开头 %[a,b]以a或b结尾

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

"objURL":"(.\*?)",

正则匹配的是列表

遍历正则匹配后后的是元组

字符串-列表-元组

result\_list=bs.find\_all(id=re.compile("two"))  
# print(result\_list)  
for i in result\_list:  
 print(i.get("href"))

res=bs.select("title")  
for i in res:  
 print(i.string)

HTTPDefaultErrorHandler 用于处理HTTP响应错误，错误都会抛出 HTTPError 类型的异常。

HTTPRedirectHandler 用于处理重定向。

HTTPCookieProcessor 用于处理 Cookie 。

ProxyHandler 用于设置代理，默认代理为空。

HTTPPasswordMgr 用于管理密码，它维护了用户名密码的表。

HTTPBasicAuthHandler 用于管理认证，如果一个链接打开时需要认证，那么可以用它来解决认证问题。 另外还有其他的 Handler ，可以参考官方文档。