import requests  
import re  
from functools import reduce  
import os  
from hashlib import md5  
  
#用于创建目录  
image\_num = 1  
#获取网页内容  
def get\_page(offset):  
 params = {  
 "offset": offset,  
 "format": "json",  
 "keyword": "街拍",  
 "count": 10,  
 "autoload": True  
 }  
 response = requests.get("https://www.toutiao.com/api/search/content/", params=params)  
 return response.json()  
  
#提取标题和对应的图片链接组  
def extra\_info(json):  
 page\_dict = {}  
 data = json.get("data")  
 for single\_info in data:  
 url\_list = []  
 image\_list = single\_info.get("image\_list")  
 if image\_list:  
 full\_title = single\_info.get("emphasized").get("source")  
 title = extra\_title(full\_title)  
 for url in image\_list:  
 url\_list.append(url.get("url"))  
 #当标题重复时，重构标题  
 if page\_dict.get(title):  
 title = title + "1"  
 page\_dict[title] = url\_list  
 return page\_dict  
  
#去除标题的<em>标记  
def extra\_title(full\_title):  
 str\_list = []  
 compiler = re.compile(r'(.\*)<em>(.\*)</em>(.\*)')  
 content = re.search(compiler, full\_title)  
 if content:  
 for group in content.groups():  
 str\_list.append(group)  
 else:  
 return full\_title  
 #合并list内的字符串，reduce(lambda x,y:x+y,list)  
 return reduce(lambda x,y:x+y,str\_list)  
  
#分文件夹保存图片  
def save\_image(page\_dict):  
 for title,urls in page\_dict.items():  
 global image\_num  
 title = "picture" + str(image\_num)  
 image\_num += 1  
 if not os.path.exists(title):  
 os.mkdir(title)  
 cwd = os.getcwd()  
 os.chdir(title)  
 for url in urls:  
 response = requests.get(url)  
 file = "%s-%s.%s"%(title,md5(response.content).hexdigest(),"jpg")  
 if not os.path.isfile(file):  
 fp = open(file,'wb')  
 fp.write(response.content)  
 fp.close()  
 os.chdir(cwd)  
  
if \_\_name\_\_ == "\_\_main\_\_":  
 json = get\_page(0)  
 pic\_dict = extra\_info(json)  
 save\_image(pic\_dict)

使用Ajax技术，即分析xhr文件并获取链接