

Python获取海贼王更新信息

2018-12-02

12月2日，晴，海贼王停更。

问题

做为一个合格的肥宅，海贼王和妖精的尾巴每周必追。这两部动漫播放源都在爱奇艺，都是VIP内容。每周末看动漫，都要在YouTube或其它网站上找资源。问题是，资源网站的更新往往不及时，常常需要Google“动漫名称 + 最新集数”，比如“海贼王 864”。

每个星期都精确的记住一部动漫应该更新的最新一集集数是多少，恐怕不是正常肥宅会做的事情，况且两部。这样，每次搜索资源前，都需要进入爱奇艺，搜索海贼王，看到最新的一集集数，关闭页面，进入Google搜索。妖精的尾巴也要同样的操作来一次。

而且，在爱奇艺里看到最新一集集数的瞬间，无法判断它是否停更，还需要在复杂的PC页面中找到“更新时间”这一标签，看更新状态是否正常，才可以做出判断。至于手机页面或APP，更是没有途径可以查看动漫的更新状态。

另一个获取动漫最新集数的方式是，百度直接搜索动漫名称，首页就倒序显示最新几集的列表（这一点百度好于谷歌），但也无法判断更新是否正常。再者，浏览器默认为Google，百度搜索需要先输入baidu，按TAB切换至百度搜索引擎，再输入要搜索的内容敲回车，步骤同样繁琐。

解决

写一段简单的Python脚本，从爱奇艺页面上抓取信息，自己直接访问程序便能知晓动漫的更新情况。引入工具包：

```
import urllib.request
from bs4 import BeautifulSoup
from wsgiref.simple_server import make_server

# 海贼王页面链接
url = "http://www.iqiyi.com/a_l9rrhb3xvl.html?vfm=2008_aldbd"

urllib用于发送http请求，并接收页面数据；bs4用于解析页面，更轻易获取内容；wsgiref用于建立http服务器，提供网络服务。url是全局变量，储存海贼王页面的链接地址。

# 从页面获取数据
def reciveData(url):
    # 获取页面内容
    response = urllib.request.urlopen(url)
    html = response.read()
    # 解析器
    soup = BeautifulSoup(html, "html.parser", from_encoding="utf-8")
    # 更新时间
    p = soup.find('p', class_="episodeIntro-update")
    # 最新集数
    i = soup.find('i', class_="title-update-num")
    return p, i
```

这几行代码发送了请求，并从页面中获取信息。这里更新时间和最新一集集数的信息就已经拿到了。接着要创建一个http服务器，让程序输出内容到页面：

```
# 服务器环境的处理函数
def application(environ, start_response):
    # 获取数据
    p, i = reciveData(url)
    # 拼接出页面内容
    start_response('200 OK', [('Content-Type', 'text/html')])
```

```
content = ('<h3>海贼王</h3>'
          + 'msg: ' + p.contents[2].get_text().strip()
          + '<br>'
          + 'num: ' + i.get_text())
return [bytes(content, encoding = "utf-8")]
```

最后启动一个本地服务器，访问8010端口即可看到页面。可将程序部署到服务器，之后直接访问服务器：

```
# 启动服务器
httpd = make_server('', 8010, application)
httpd.serve_forever()
```

运行结果如图：

海贼王

msg: 本周停播12月9日恢复播出
num: 863

妖精的尾巴

msg: 每周日07:30
num: 286

扩展

从这一想法出发，可以扩展程序。一种是从各大网站获取全面的动漫更新信息，主动提供服务；再一种是根据用户的输入，提供自定义的动漫更新信息；或者将两者结合，提供一种大而全的、可收藏、可定制的服务。虽然这种想法毫无意义。

更正

之前的代码犯了一个低级错误，程序只会在首次运行时发起网络请求，之后由于网络服务一直处于启动状态，返回网页的内容始终都是初始数据。解决这个问题也很容易，将请求网络的操作封装到一个函数中，再到application函数中调用该函数即可。（代码已更正，为保证简洁，去掉了妖尾部分的代码和控制台的日志输出）