PHP中使用cURL操作网络资源

# 开发环境

1、环境搭建：Windows 7+Apache 2.4.18+MySQL 5.7.11+PHP 7.1.0。

2、文本编辑器：Sublime 3。

# 二、主要技术

本案例主要使用PHP 7中的数据传输神器cURL来抓取网络资源、动态获取WebService数据接口数据和发送HTTP请求，从而来实现简单网页爬虫程序的编写、调用WebService查询当前天气和访问HTTPS资源的功能。

# 三、效果图展示

实现效果图，如图1、2、3所示。



图1 效果图(网页爬虫)

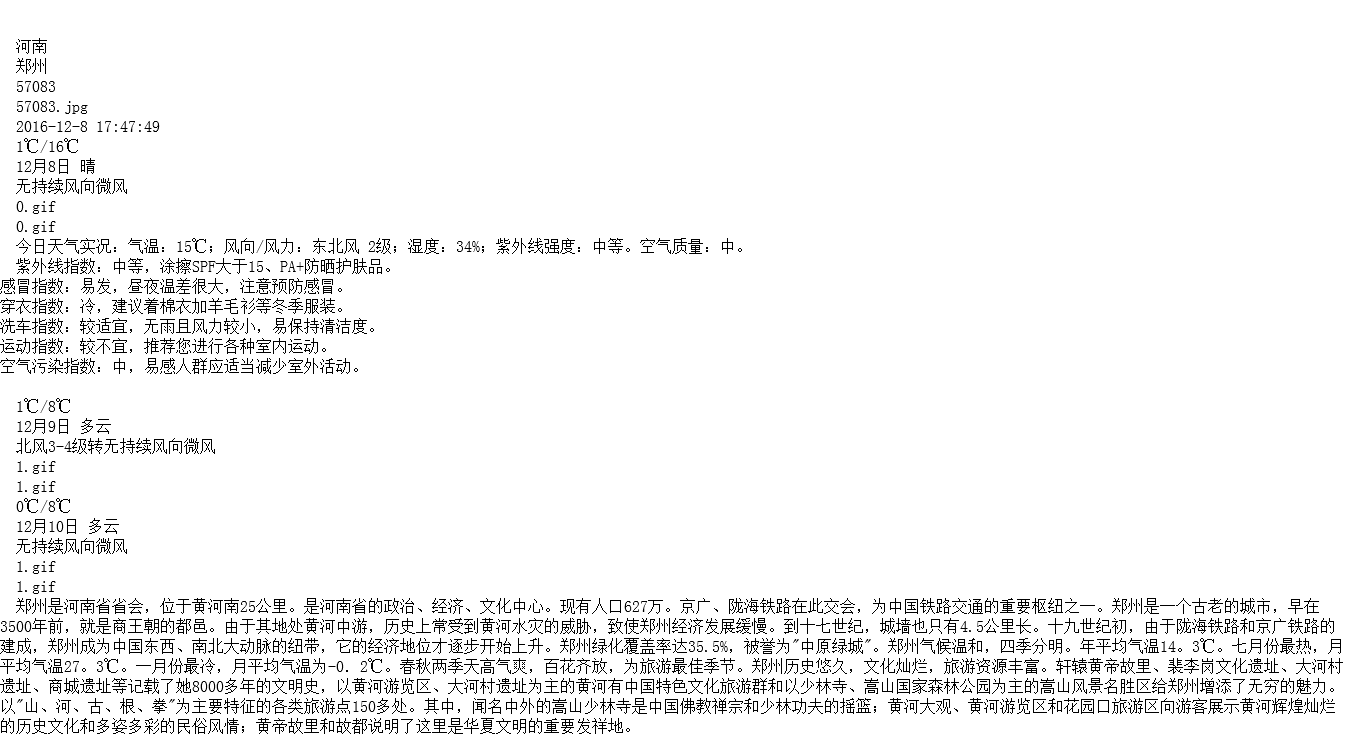


图2 效果图(查询当前郑州天气)

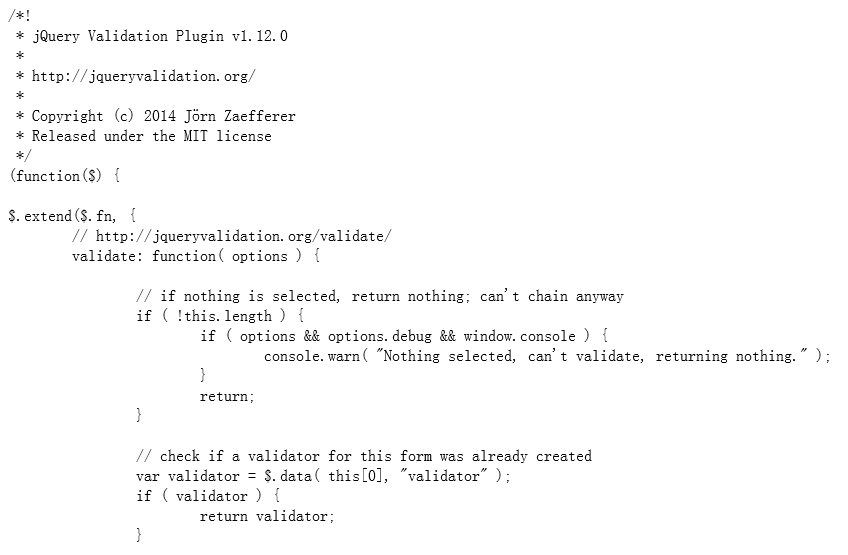


图3 效果图(访问HTTPS资源)

# 四、具体步骤

## 4.1、PHP使用cURL编写简单的网页爬虫

此网页爬虫程序主要用来获取百度的首页,并将其中的“百度”二字替换成“搜索引擎”。

### （1）创建php文件

创建一个名为“demo1.php”的文件。

### （2）主要代码

// 1.初始化

$curlobj=curl\_init();

// 2.设置访问的url

curl\_setopt($curlobj,CURLOPT\_URL,"http://www.baidu.com");

// 3.执行之后不直接打印出来（只是下载，不显示出来）

curl\_setopt($curlobj,CURLOPT\_RETURNTRANSFER,true);

// 4.执行

$output=curl\_exec($curlobj);

// 5.关闭

curl\_close($curlobj);

// 6.替换内容

echo str\_replace("百度","搜索引擎",$output);

## 4.2、PHP使用cURL通过调用WebService查询当前郑州天气

### （1）创建php文件

创建一个名为“demo2.php”的文件。

### （2）主要代码

<?php

/\*\*

\* curl通过调用WebService查询郑州的当前天气

\*/

header("Content-type: text/html; charset=utf-8");

$data = 'theCityName=郑州';//传递的参数

$curl\_obj = curl\_init(); //初始化一个cURL会话

//设置URL和相应的选项

curl\_setopt($curl\_obj, CURLOPT\_URL, "http://www.webxml.com.cn/WebServices/WeatherWebService.asmx/getWeatherbyCityName");//设置一个cURL传输选项

curl\_setopt($curl\_obj, CURLOPT\_USERAGENT, "user-agent:Mozilla/5.0 (Windows NT 5.1; rv:24.0) Gecko/20100101 Firefox/24.0");//在HTTP请求中包含一个”user-agent”头的字符串

curl\_setopt($curl\_obj, CURLOPT\_HEADER, 0); //关闭文件头输出,不可见,启用时会将头文件的信息作为数据流输出。

curl\_setopt($curl\_obj, CURLOPT\_RETURNTRANSFER, 1);//把获取到的内容输入到文件，而不是直接输出给浏览器,如果成功只将结果返回，不自动输出任何内容。如果失败返回FALSE

curl\_setopt($curl\_obj, CURLOPT\_POST, 1);//启用时会发送一个常规的POST请求

curl\_setopt($curl\_obj, CURLOPT\_POSTFIELDS, $data);//需要POST的数据

curl\_setopt($curl\_obj, CURLOPT\_HTTPHEADER, array("application/x-www-form-urlencoded; charset=utf-8",

"Content-length: ".strlen($data)

));

$rtn = curl\_exec($curl\_obj);//执行一个cURL会话

if(!curl\_errno($curl\_obj)){ //返回最后一次的错误号

$info = curl\_getinfo($curl\_obj); //获取一个cURL连接资源

句柄的信息

echo '<pre>';

echo($rtn);

echo '</pre>';

}else{

echo 'Curl error: ' . curl\_error($curl\_obj);//返回一个保

护当前会话最近一次错误的字符串

}

curl\_close($curl\_obj);//关闭一个cURL会话

?>

<!DOCTYPE html>

<html>

<head>

<style>

body{

cellpadding:0;

cellspacing:0;

padding:0;

margin:0;

}

pre {

white-space: pre-wrap; /\* css-3 \*/

white-space: -moz-pre-wrap; /\* Mozilla, since 1999 \*/

white-space: -pre-wrap; /\* Opera 4-6 \*/

white-space: -o-pre-wrap; /\* Opera 7 \*/

word-wrap: break-word; /\* Internet Explorer 5.5+ \*/

}

</style>

</head>

</html>

## 4.3、PHP使用cURL访问HTTPS资源。

### （1）创建php文件

创建一个名为“demo3.php”的文件。

### （2）主要代码

$curlobj = curl\_init(); // 初始化

//设置访问网页的URL

curl\_setopt($curlobj,CURLOPT\_URL,"https://ajax.aspnetcdn.com/ajax/jquery.validate/1.12.0/jquery.validate.js");

//执行之后不直接打印出来

curl\_setopt($curlobj,CURLOPT\_RETURNTRANSFER,true);

//设置HTTPS支持

//使用cookies的时候必须先设置时区

date\_default\_timezone\_set('PRC');

//终止从服务器端验证

curl\_setopt($curlobj,CURLOPT\_SSL\_VERIFYPEER,0);

$output = curl\_exec($curlobj);

curl\_close($curlobj);

echo "<pre>";

echo $output;

echo "</pre>";

# 五、网页测试

开启Apache服务器后，打开浏览器，在浏览器地址栏中输入文件地址进行测试。

### （1）网页爬虫程序测试

在浏览器地址栏中输入文件地址:

“localhost:8080/curl/demo1.php”，可以看到百度的首页，并且“百度”二字已经被替换成“搜索引擎”，如图4所示。



图4 案例结果图（网页爬虫）

### （1）调用WebService查询当前天气程序测试

在浏览器地址栏中输入文件地址:

“localhost:8080/curl/demo2.php”，可以看到页面中显示郑州当前的天气信息，如图5所示。

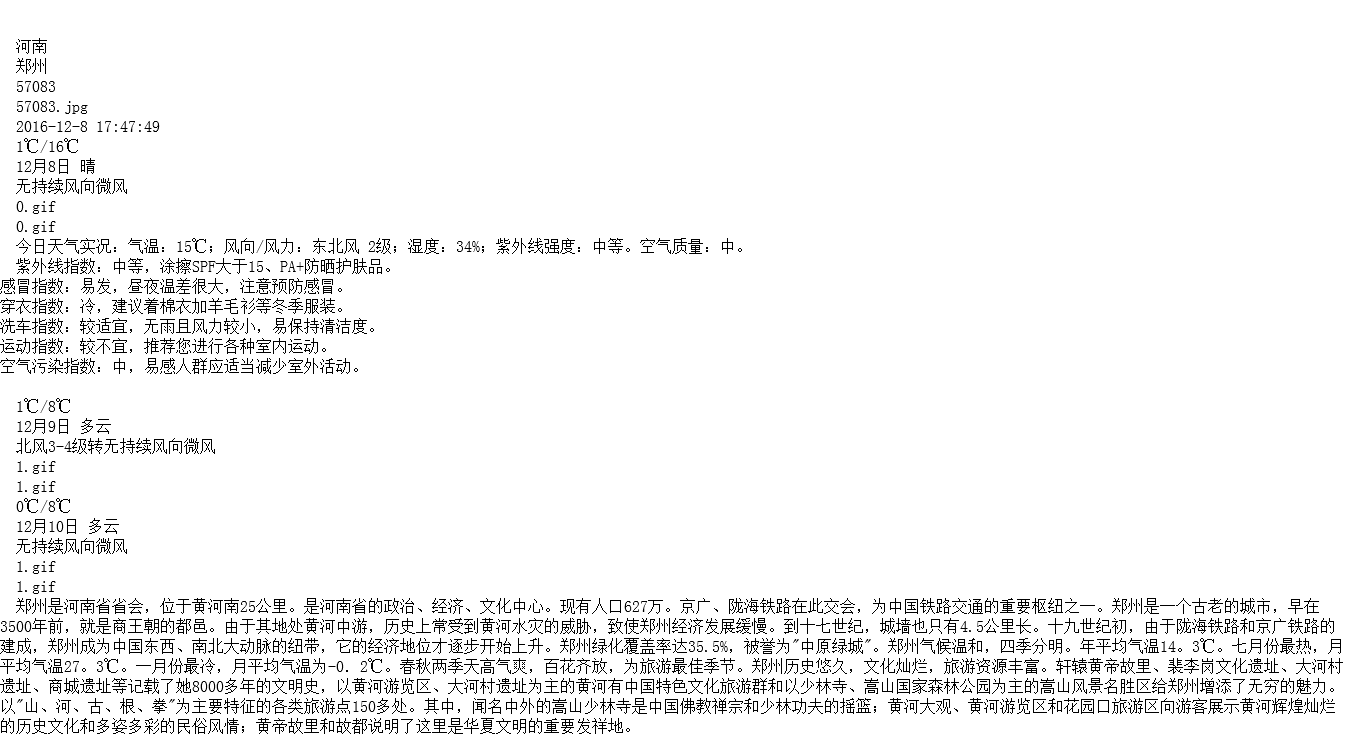


图5 案例结果图（调用WebService查询当前天气）

### （1）访问HTTPS资源程序测试

在浏览器地址栏中输入文件地址:

“localhost:8080/curl/demo3.php”，可以看到页面中获取到的HTTPS资源，如图6所示。

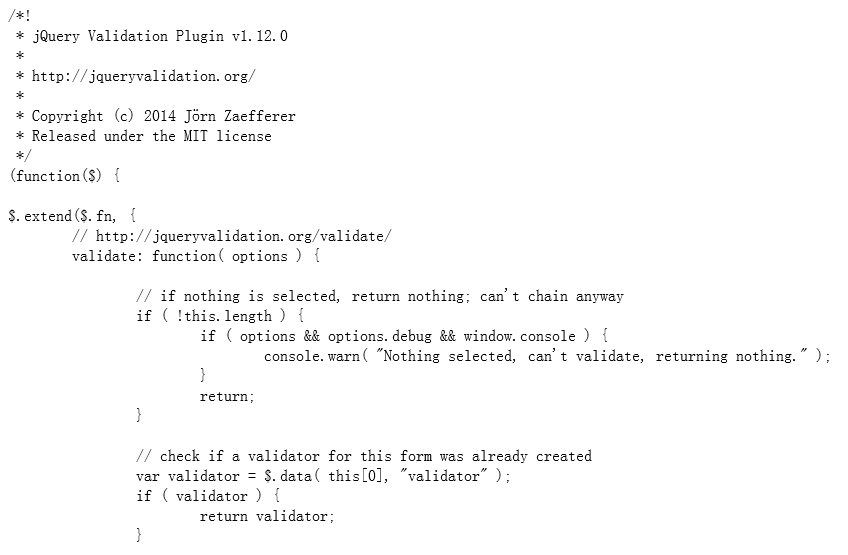


图6 案例结果图（访问HTTPS资源）