## Vquery使用文档

在使用php进行网页抓取的时候你有没有感觉到用起来比较麻烦呢？目前我还没有发现php有这样针对网页抓取的类，每次用到这个的时候我总是觉得不顺手。每每想到jquery里面的find（）方法，获取网页上面的元素都是那么的方便，我在网上找过php相关的方法，但是好像并没有什么结果，于是和同事两个研究出来了一个类，专门针对php抓取网页的处理类，使用这个类你不需要再向以前那么麻烦了，你可以直接调用类似于jquery那种方法，直接对网页上面的数据进行抓取，例如$vp=$this->vq->find('a),你只需要这样操作你就可以获取该网页里面所有的a标签了，是不是觉得有点意思呢？

废话不多说，下面容我介绍下这个类的大致用法把！

为方便大家使用，Vquery类我们已经封装完成，你在项目里面直接引进Vquery类，再构造一个函数传入参数即可：

function \_initialize(){

import('ORG.Util.Vquery');//引用Vquery类，不同的框架或源生写法可能引入有差异。根据实际情况选择对应的引入方法。

$arr=array(

"url"=>”<http://www.baidu.com>”,

“method”=>”get”or”post”,

“data”=>array(“username”=>”admin”,”password”=>”admin”,……..参数自己根据需求补),

“header”=>array(“DESC:admin”,……同上),

);

// $arr=file\_get\_contents("http://www.jd.com/allSort.aspx");

$this->vq=new Vquery($arr);

}

说明：Vquery类内置curl函数，支持传参格式①数组，即依赖Vquery里的curl抓取网页进行分析取值。这种情况需要如上述的$arr数组中的格式进行配置；②Vquery也支持内容传参，例如你可以把已经获取到的网页content直接传入，例如使用file\_get\_contents获取网页直接传入。Vquery内部已经做了Url解析，支持各种Url传参。好了，引入工作到此就完成了，您可以试试手感了。

Vquery类共计使用的方法有6个：find()、attr()、html()、text()、getHtmlContent ()、deal();

**find()**：find()方法支持类似jquery的链式使用规则，例如$vq=$this->vq->find(“div”)->find(“a”);

**find()**传参规则: find(标签)，例如$vq=$this->vq->find(“div”)表示找文章中所有该标签的元素；

find(标签$属性=)例如$vq=$this->vq->find(div$class=\”nvai\”)表示找文章中所有的class=nvai的div

find(标签，标签)例如$vq=$this->vq->find(“div”,”span”)表示找出文章中多有的div及span标签

find(标签：数字)例如$vq=$this->vq->find(“a:2”)表示找出文章中的第二个a标签

find(属性)：例如$vq=$this->vq->find(“class=\”nvai\””)表示找出文章中所有class=”nvai”的标签。

**find()**方法支持链式调用，所以可以组合使用来达到不同的目的。系统检查到链式调用的结构之后，find()方法会自动让后面的链式结构中去上级的返回结果中匹配结果，例如，$vq=$this->vq->find(“div$class=\”nvai\””)->find(“a”);这样使用返回的结果是：找到所有class=nvai的div中的a标签

**attr() :**attr()方法适用于匹配属性的参数，例如我要找文中所有的herf，那么我们可以这样使用：attr(“herf”);返回结果会得到herf后面的地址参数，其他属性值的获取方式雷同，加上链式使用规则您可以这样使用函数：$vq=$this->vq->find(“a”)->attr(“herf”);

**html()**:html()方法适用于获取网页上面制定标签之间的文本类容，这个方法获取的是最原始的内容，不经过任何处理，所以可能包含其他标签或者字符，这个方法可以根据你自己的需求适用，适用方法跟jquery里面的一样，此方法不需要传递参数：$vq=$this->vq->find(“a”)->html();

**text()**:text()方法适用范围跟html()方法一样，但是唯一的区别就是test()方法是经过过滤处理了的，也就是说通过test()方法得到的就是纯文本了，里面不会包含任何标签。这个针对抓取网页文本非常有效，使用方法和html()一样，也不需要传参：$vq=$this->vq->find(“a”)->test();

**getHtmlContent ()**:getHtmlContent()方法适用于查看抓取目标网页的源代码。供开发者查看网页特征，本方法可分为传参和不传参，当有参数传递时，则返回目标网站的源码，开发者可以对返回值进行保存等处理，传参格式:getHtmlContent(true)；当不传参数的时候，则系统默认格式化打印网页源码：使用格式$vq=$this->vq->getHtmlContent()；

**deal()**：deal()方法是对前面的方法进行结果整理，将返回的多维数组整理成二维数组，方便数据的取用。不需要传递参数，一般使用方法如下：$vq=$this->vq->find(“a”)->deal();

方法介绍就大致入上文所述，Vquery类经过我们一段时间的测试，进本可以涵盖主流抓取网页的需求，当然肯定还有需要完善的地方，所以我们把这个文件发出来，希望更多的大神能使用，把不足的地方告知我们，相信在大家的智慧下，能让以后php抓取网页变的简易化。提高开发效率。