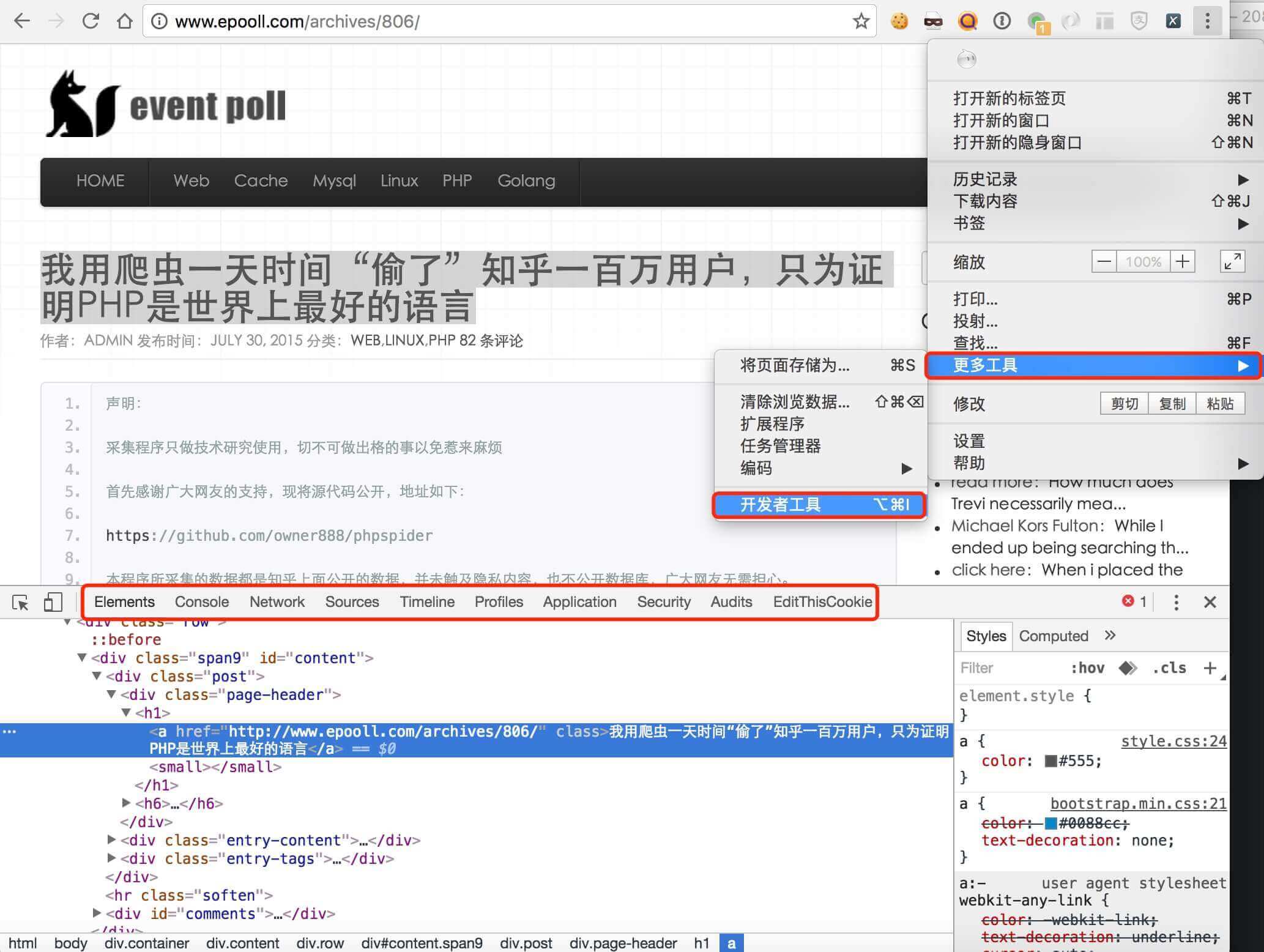
**开发PHPSpider爬虫的常用工具**

“工欲善其事，必先利其器”，开发PHPSpider爬虫，起码得有几件顺手的工具才行吧，接下来给你逐个介绍。

**谷歌Chrome浏览器**

说起谷歌的Chrome浏览器（以下简称Chrome），相信大家都耳熟能详了吧，不仅使用流畅，而且功能强大，对开发PHPSpider爬虫非常有帮助。

我们主要使用的是Chrome的开发者工具，如下图所示：



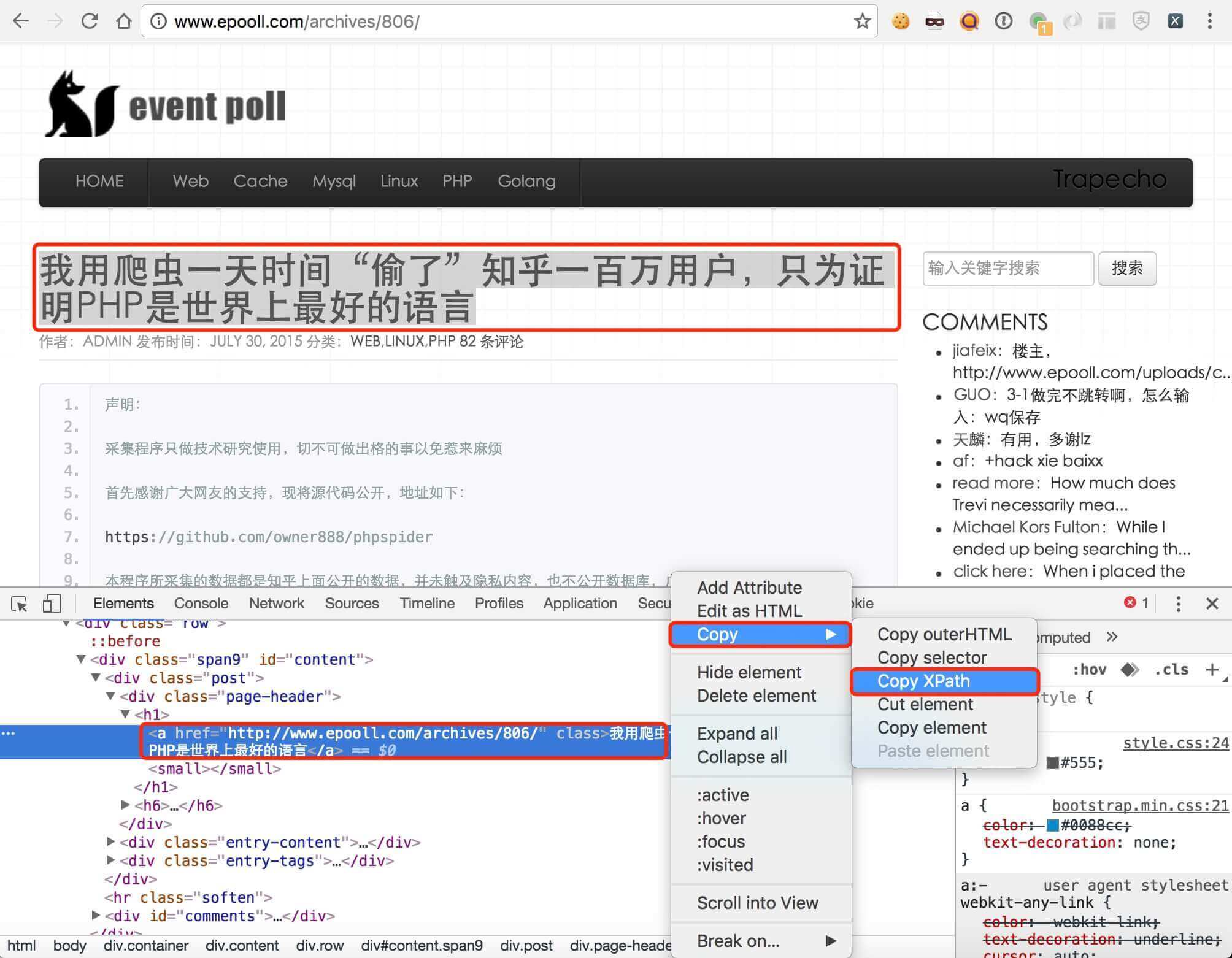
或者可以直接在网页上点击鼠标右键，选择“检查”，也可打开开发者工具。

开发者工具顶部有Elements、Console、Network等八个栏目。常用的有三个：Elements，用来查看需爬取字段的HTML标签信息；Console，可以检测你的JS代码；Network，用来分析HTTP请求。

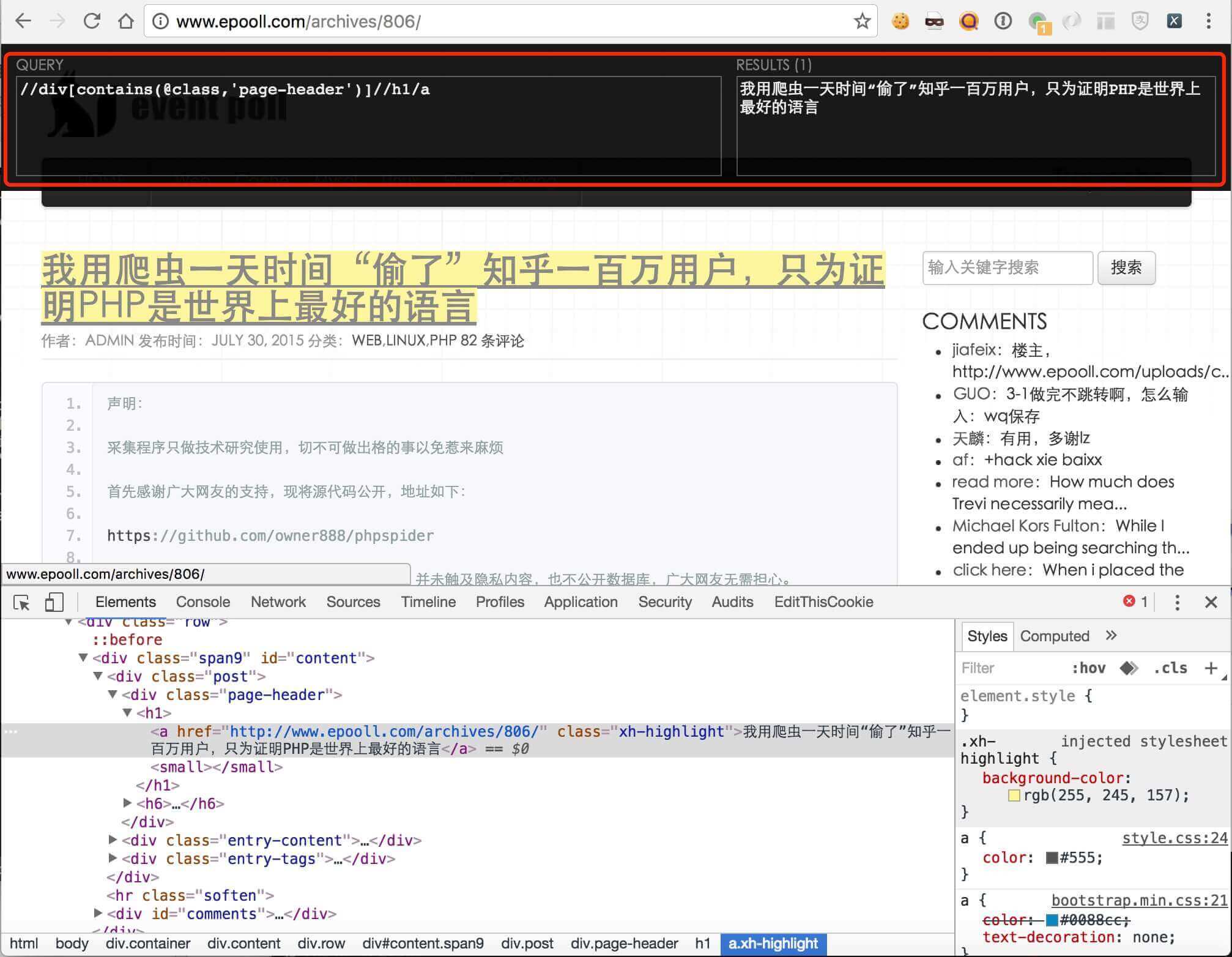
**XPath Helper**

XPath Helper是Chrome浏览器的插件，可以在Chrome应用商店安装下载，主要用来分析当前网页信息的XPath，并将其精简化。具体操作步骤如下：

1、在Chrome浏览器上，选择抽取的html字段并右击，点击“检查”，即可弹出开发者工具；右击已选字段，点击Copy XPath即可将该字段的XPath保存到浏览器剪贴板上，如下图所示：



2、打开XPath Helper插件，将得到的XPath复制进去，最好进行简化修改后再使用，如下图所示：



3、在XPath中，如果使用class属性来定位元素，最好使用contains函数，因为元素可能含有多个class：

(

"name" => "article\_title",

"selector" => "//div[contains(@class,'page-header')]//h1/a"

),

4、在XPath中，如果使用id属性来定位元素，因为理论上id是唯一的，可以直接使用\*[@id='']：

(

"name" => "article\_content",

"selector" => "//\*[@id='single-next-link']"

),

**DHC REST**

DHC REST也是Chrome浏览器的插件，可以在Chrome应用商店安装下载，主要用来模拟HTTP客户端发送测试数据到服务器。HTTP Get请求在开发中比较常用。

**正则表达式测试工具**

推荐使用站长工具中的正则表达式测试工具，链接如下： <http://tool.chinaz.com/regex/>