采集器所采取的技术：

1. PHP（后端）
2. Bootstrap（前端静态页面）
3. js+jQuery（删除信息时）

首页截图：



采集结果截图：



选择入库条件截图：

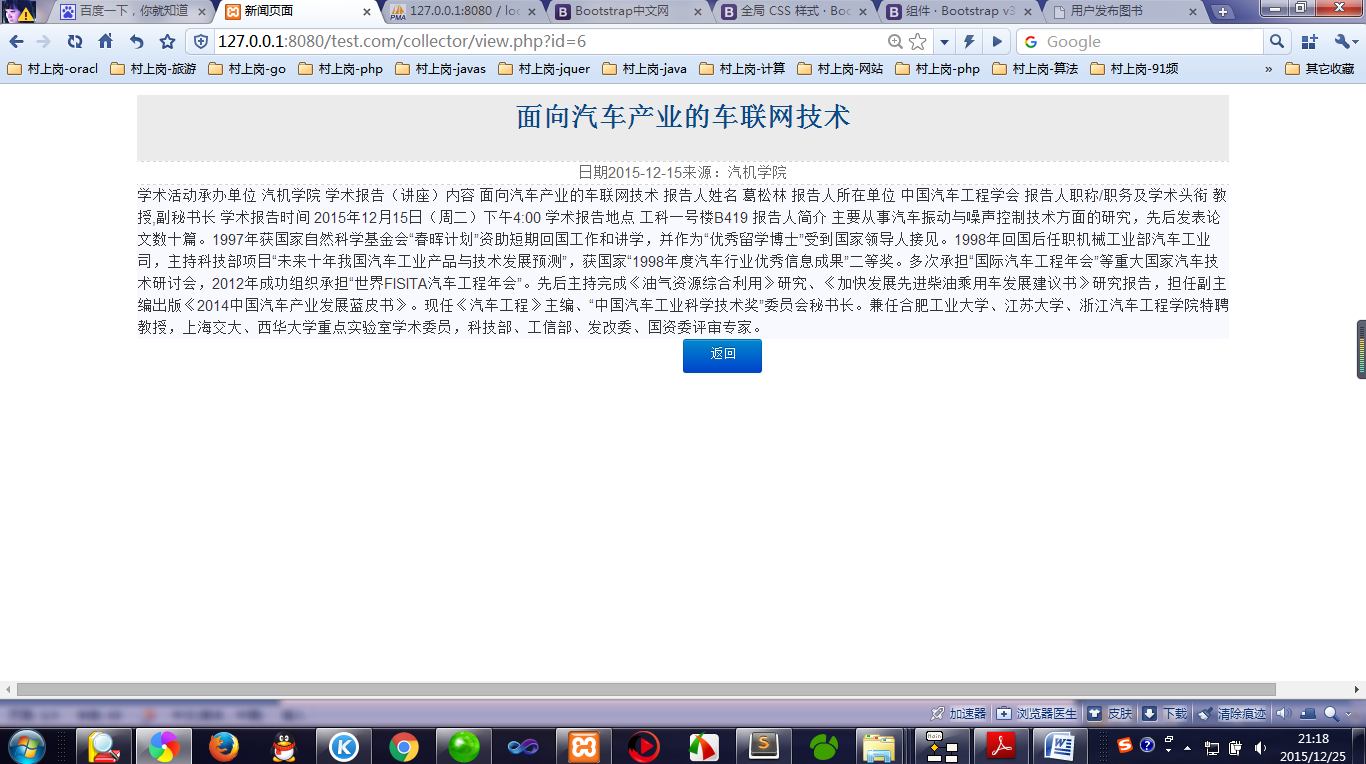




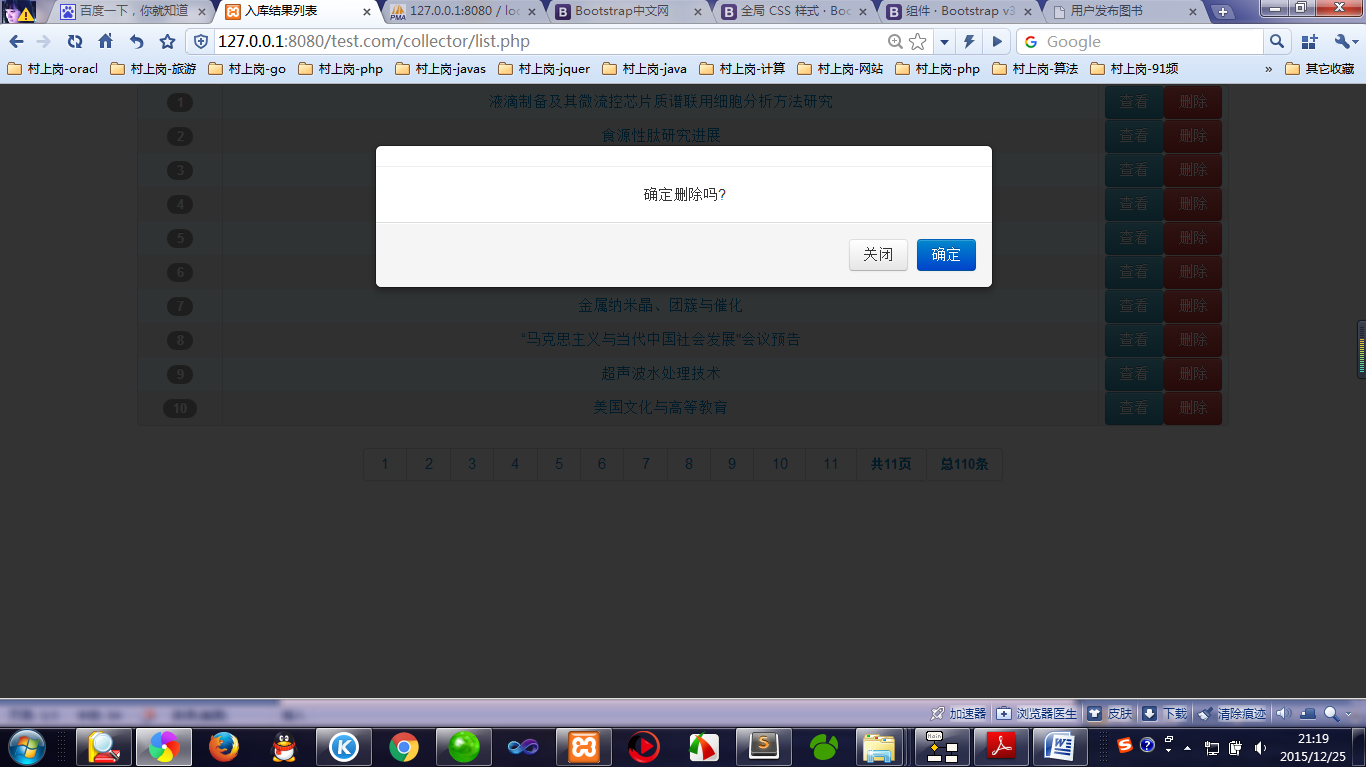
入库结果截图：



查看本地入库结果截图：



删除截图（用了一点js+jQuery的效果）:

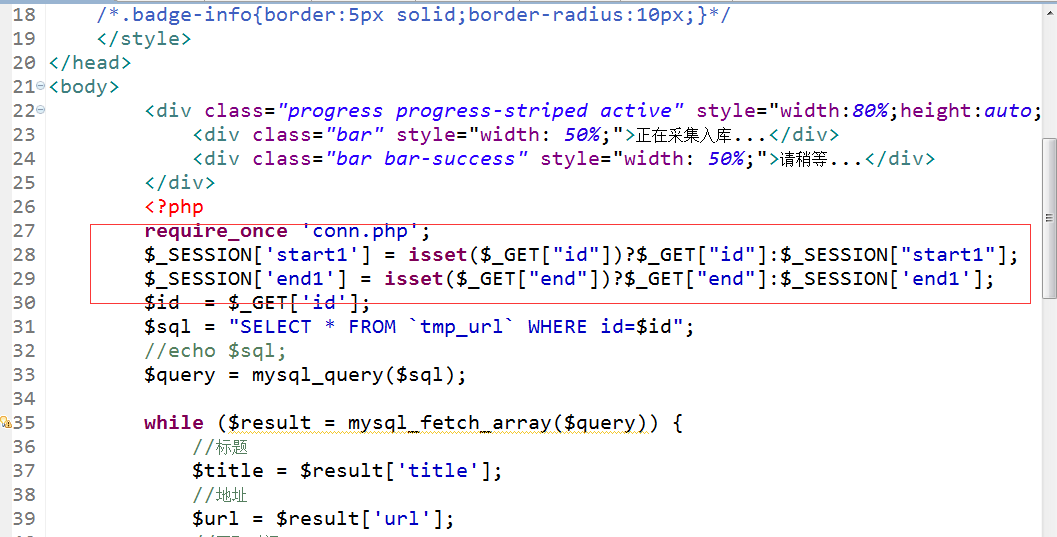


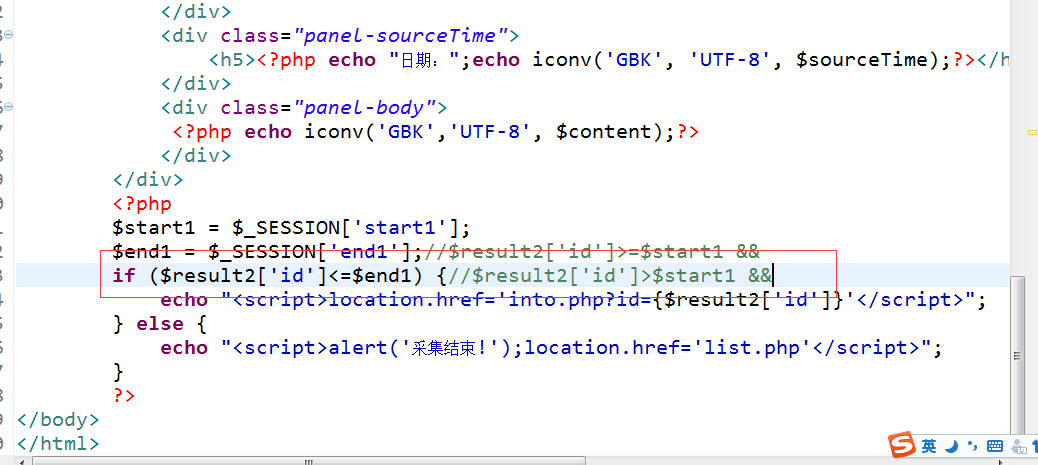
需要注意一点：

1. 数据库的编码问题一直都是看似不难但是做起来也是不是容易的。首先如果通过正则表达获取过来的信息，放进数据库后会造成乱码，但在页面不是显示乱码，入库是乱码。那么可以页面选择用UTF-8，而数据库保存数据编码可以是UTF-8，但是在连接数据库的时候要把编码改为：GBK。即：mysql\_query(“set names ‘utf8’”).但数据能正确入库时，即不乱码的时候，但是输出的时候就又会乱码了。因为数据库的编码格式是GBK，但是页面仍然采用UTF-8格式编码，能不乱码吗？这是也是一个头疼的问题，但是PHP早就有准备了。使用iconv()将编码强制转换输出。如图所示：

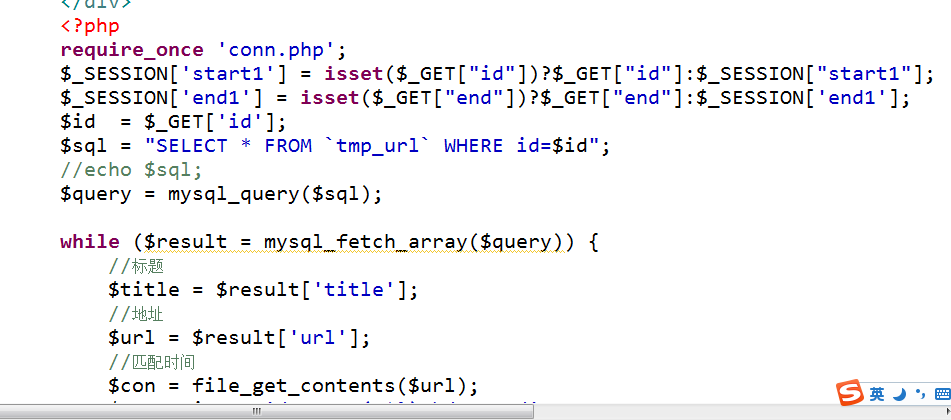


1. 采集入库页面(into.php)里面的不停跳转的判断条件值得考虑清楚。用户在提交了采集条件的时候，就用GET方法把开始条件作为id号发送出去，在第一次跳转到into.php页面时，就要把id号用sessio保存起来，用来做判断跳转的条件参数。如图所示：



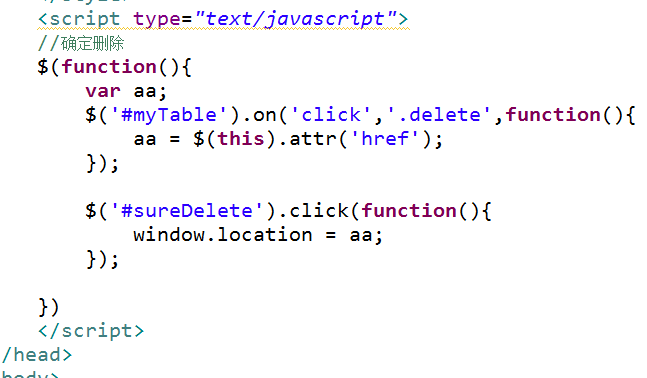


在此过程当中一定要弄清楚流程：页面不停跳转的原因是id号在一定的范围内运行，这个条件的id号获取，一是通过页面的GET方法，每次页面刷新session值如果为空的话，那么也会重新赋值为空。我就在这里停留了很久，session的值一次跳转后输出是没有错，但是第二次跳转后页面的session就重新获取GET方法传过来的$start，但$start的值只提交了一次，所以页面刷新后就为空了。这时候第二次跳转时，$start为空。此时的判断就变得有点难度了。三元符号在这里起到了作用了。如图所示，天衣无缝的解决了这个问题。

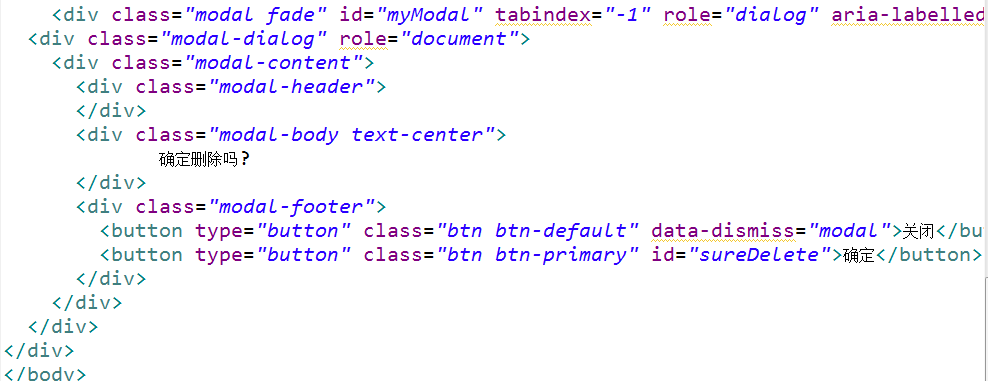


首先判断是否为空，不为空则赋新值，其实第二次跳转后一直都是为空的，所以当为空时，马上等于第一次提交过来的值，即用户提交时的起始条件和结束条件。这个问题算是这样解决了，基础知识的重要性在这里又一次的体现出来！！！

1. 正则表达很好，了解会用很重要，精通那是更好了。但是有点难。多加学习与联系，应该会很好。
2. 作为PHP的开发人员，会一点jQuery与js也是很重要的。如下图所示：









这是一个bootstrap的模态框，不是弹出框，值得思考它的方法提交。最后提醒使用者：把流程搞清楚了，做一个比这个更加完善的采集器出来。这个只限于某一个网站的新闻，局限性非常大，如果能把它完善好了，也是一种荣耀。搞技术的从来不要吝啬动手写代码，想出来，但做不出来也是不行的，没有那个技术大牛是轻而易举达到的。最重要的是要记住：问题总是能解决的，或许明天，或许下个月，只是你没有想明白罢了。需要不停地学习，加油！