

就成敦煌种业拼接好的链接地址了

内容导航：

2021.11.01 20-30038729-063Ae8m1x5Fdhz0fi069s2UA9q7k40Hfiq5H42

对于做seo的人员来说，有一项技能不会的就是网站日志分析了。因为我们只能从网站日志和根据百度账号管理后台的数据才能大体上了解我们网站seo优化做得如何？整体趋势走向该去怎么调整，而不是所谓网站一拿到手乱优化。

【要安装下载的工具具有 大族激光股票Cygwin 和 Notepad++600大宗论坛 和 Excel表格 这三款工具加一个log日志文件。】600855股票网上也有一款算比较不错的网站日志分析工具，网页端在线分析的，具体使用方法也挺简单，就是输入你的网站网址（这里是随便填写的），然后导入你的网站日志文件，然后“点击分析”就OK了。这里面能看到你的站点被哪些搜索引擎收录，蜘蛛的抓取次数频率，哪个时间段？以及判断有哪些低高权重的蜘蛛过来抓取等都可以看得一清二楚。

002340

但是唯一的缺点就是不能通过一键导出数据的功能，对于大型站点，内容较多的是比较麻烦的，如果是博客等小型的网站，内容少的，手工操作复制粘贴也很快。以及不能保证所检测的是否百分百准确性问题，总体来看，cygwin分析出来的数据要更准确且专业些。抓取到的网站404错误页面这里同花顺300033说一下什么是Cygwin，按照百度百科的解释：Cygwin是一个在windows平台上运行的类UNIX模拟环境，是cygnus solutions公司开发的自由软件。它对于学习UNIX/Linux操作环境，或者从UNIX到Windows的应用程序移植，或者进行某些特殊的开发工作，尤其是使用GNU工具集在Windows上进行嵌入式系统开发方面非常有用。随着嵌入式系统开发在国内日渐流行，越来越多的开发者对Cygwin产生了兴趣。嗯，听太多官方没什么用处，我们直接拿它用来进行网站日志分析，如何怎么分析的就行了。在对这款工具使用之前，先要进行下载安装，好像目前不支持苹果电脑的安装，允许windows系统。具体的安装步骤在这里不作具体赘述。安装步骤就是下一步，具体可以搜索引擎一下安装教程，这方面教程挺多的，这里直接讲怎么利用这款工具的使用。

股票000672

当下载好工具之后，有一个注意的地方，会在你下载保存的某个路径下有以下几个文件，缺一不可，就这几个文件：和顺电气股票当打开cygwin工具后，不知道这几个文件在哪里，可以通过输入“pwd”命令查看存放路径：我放的是在系统盘里好，知道这些之后，咱们的目的是分析网站日志，首先要下载你网站的log日志，先命好名，我这里的命名为1.log，然后将下载好的1.log日志文件一起和那5个文件放在一起：这是第一步，下载好了log文件之后，接着是我们要通过它提取网站404错误页面，什么搜索引擎的蜘蛛抓取到的404错误页面我们以“百度”为例，写法为：“Baiduspider”。提取过程如下所示，这里有几个步骤记下即可：

先分离出来百度的蜘蛛，先分出5段出来看下效果，看看是否可以提取成功？输入命令为：

```
cat 1.log|grep 'Baiduspider'|head -5
```

发现，正常输出，说明百度的搜索引擎蜘蛛是有来爬取我们的网站的，好，接着来仅提取百度搜索引擎蜘蛛爬取的所有信息，将提取到的信息命名后缀为.txt的文件，为了方便确认，就命名为“baidu.txt”，输入以下命令：

```
cat 1.log|grep 'Baiduspider'>>baidu.txt
```

输入命令敲回车之后，1.log存放的那个目录文件里就会多了一个baidu.txt的文件，如下图所示：

OK，咱们把百度搜索引擎蜘蛛爬取的网页都提取出来了，包括404，200，301状态码的页面都有，因为我们只要提取404状态码的页面就行。这时候，为了找到对应状态码的位置，要知道它在baidu.txt文件中的哪一段，提取前，要先数一下，咱们一起分析一下吧。首先利用Notepad++工具打开这个baidu.txt文件，结果密密麻麻的数字：先不用管看得懂看不懂，一起来“数羊”吧。这个段数是每空一格算一段，我们是要找到所有状态码的在一行中的第几个，刚好在第9段数中。知道在第几段数之后，知道状态码有什么，以及数量多少，也更了解当前网站的具体情况了。咱们就通过这串命令提取所有状态码页面的信息，命令如下所示：

```
cat baidu.txt|awk '{print $9}'|sort|uniq -c
```

回车之后，终端就会显示网站所有的状态码，以及对应的个数，如下图所示：

所以再次判断404情况：只统计状态码的输入指令即可：

```
cat baidu.txt|grep '404'>>404.txt
```

这里提醒一下，直接分离404状态码即可，为了好确认，保存为404.txt文件，同样，文件一样多出了这个404.txt的文件。当然，这一步还没完，因为提取的还是404状态码整行的数据，我们的目的是提取网站404页面的链接，所以要提取这一块的链接，找到对应链接在一行中的第几个？发现在一行中第7个：

接着输入提取链接的命令：

```
cat 404.txt|awk '{print $7}'>>baidu404.txt
```

注意：因为要对哪个文件提取，就要对哪个文件进行提取，这里是对404.txt文件的提取。敲回车之后，发现文件里有多出来了一个baidu404.txt的文件，为了知道是百度抓取的404页面，方便辨认，所以才命名为baidu404.txt。打开是这几个链接是404状态码的：我们一般把404错误页面提交到百度网站后台是把所有404页面一起放到某一个文件，比如404.txt文件中，然后提交给百度。这时候如果你的网站很大，经常修修改改，避免不了404页面，甚至是几万，十几万条的数据量，在开头之前所说的利用loghao的网页版显得不符合要求，一个是不能导出的功能，也不能确定提取是否百分百准。而使用这款cygwin工具，虽然使用过程中复杂了一点，但是靠谱，对于一些小白SEOer来说，会使用，是不是有觉得很有成就感的样子？如果404页面多，我们再配合Excel表格拼接，网址我就以虚拟的为好，作为例子讲，具体实现过程如下：首先把提取到的404链接放在Excel的B列数据格里，在A列数据格里写入你的网站真实网址，做拼接处理，拼接命令就是先在C列里输入“=”号，接着Ctrl+A 然后输入“&”再输入Ctrl+B 最后回车，就成拼接好的链接地址了。最后在C列中拼接好的数据格中，鼠标停留一小会儿，有一个“+”，双击它，就可以把所有的都拼接出来。如果上万条数据，这种方法是不是更能提高效率呢？然后其他的就可以交给复制粘贴了，具体怎么提交百度后台，相信做seo的人都懂。

最后补充一点：做seo优化，它是活来活去的，如果是要查看一个链接被搜索引擎蜘蛛重复抓取的次数，因为知道了哪些页面被重复抓取，可以给网站做一些相应的网站链接中添加nofollow属性操作，避免多重抓取，更多的机会和数量让它去抓取其他页面，输入如下命令：

```
cat baidu.txt|awk '{print $7}'|sort|uniq -c
```

注：博晖创新仅有 “/” 代表首页，前面的数字代表被抓取的次数。

300030阳普医疗

如果想通过查看自己的某个栏目被抓取哪些页面，比如通过“hlw”这个栏目下的内容被百度搜索引擎蜘蛛抓取了哪些页面，可以通过输入命令：

```
cat baidu.txt|awk '/hlw/'|sort|uniq -c
```

然后按回车键，结果会出现有多少条记录的。002065东华软件如果往后网站出现不收录的问题，都可以通过用这样的查询方法，去解决你网站问题的根源在哪里，而不要再去盲目的优化站点，利用这块cygwin分析网站日志和百度网站后台的索引量一个定制规则那里，看看哪些栏目的内容为什么不被建立索引哪些索引下降严重？网站做了什么？把每天做了的事情进行统计数据进行分析，才能针对性地去解决问题，就像知道病因在哪，才能对症下药，药到病除的效果。

举报/反馈期指持仓经典90坦克大战dn0LkP80TY7T2jILDNJ77uh9nw49Lnp87F6
Hdr80jzs53MJweB0n4wjoh2KyUSWjVhVmKT5s5do3vt

新上市的股票 000690资金流向 002623股票 600770 601689 080002
002294股票 603000人民网 300293 嘉寓股份股票 金融界 300112股票
股票600332