技术交底书

|  |
| --- |
| 发明名称：基于目标网站的域名仿冒检测系统 |
| 申请类型：发明 实用新型 |
| 发明人： |
| 第一发明人姓名及身份证号： |
| **中文关键字：**域名仿冒，钓鱼网站监测，域名相似性，FUZZ  **英文关键字：** |
| **现有技术的说明及其缺点：**  域名仿冒是域名抢注的一种形式。域名仿冒者批量注册错别字域名，以便从将其卖回品牌所有者、从用户的错误点击中获得流量收入、将错误点击请求重定向到竞争对手、部署钓鱼站点拦截密码或安装木马等恶意软件来获利。域名仿冒已为互联网用户和流行网站的品牌所有者带来严重威胁。关于域名仿冒的以往工作大多数侧重于衡量仿冒现象，而不是打击域名仿冒滥用。 |
| **具体实施方案：**  本发明提出了一种用于企业用户的相似域名识别系统，利用排列组合算法和同义词替换方法生成相似域名Fuzzer。接着利用信息熵等算法进行数据提炼，剔除无效或重复的域名。最后采用request/whois/ping等技术手段进行域名探测存活，得到有效的域名列表。   1. **业务流程**   **未命名文件2**   1. .域名分解：将用户输入的URL分解，提取出主机部分，以左边第一个“.”为分界线，左侧部分称为域名前缀，右侧部分称为域名后缀。 2. 相似域名前缀生成：采用插值法、删除法、同义词替换法等多种基于字符的混淆域名生成手段，对域名前缀进行变形，生成一系列相似的域名前缀。 3. 相似域名后缀生成：采用百度索引值指数作为筛选条件，利用selenium爬虫，从互联网中收集到的域名后缀中，选出网站数量超过10万的作为系统内置后缀库。 4. 域名过滤：用于对第2步和第3步中生成的文本进行提炼，计算原始字段与生成字段的信息熵之差，设定阈值将大于阈值的字段过滤。系统还内置了difflib、simhash、Levenshtein distance、Levenshtein ratio、Levenshtein jaro、Levenshtein jaro\_winkler等文本相似度算法作为插件使用。 5. 域名探测：系统提供三种内置函数用于域名存活探测，默认采用whois探测。 6. **系统架构**   根据上述业务流程，本系统对相似域名的检测流程如下图所示：  D:\Desktop\上班\域名仿冒专利\未命名文件.png未命名文件   1. 用户输入待检测URL进行Fuzzer子系统。Fuzzer子系统根据得到的URL进行域名分解。 2. 将分解得到的域名分为域名前缀与域名后缀两部分。 3. 对域名前准采用插值法、删除法、同义词替换法等多种基于字符的混淆域名生成手段进行域名生成，并计算原始域名与生成域名的信息熵差值，过滤大于阈值的不合格内容。 4. 对域名后缀计算信息熵，对系统内置的高频域名库（采用百度索引值指数作为筛选条件，从众多域名后缀中选出网站数量超过10万的作为系统内置库）进行相似域名后缀过滤。使用信息熵差值对大于阈值的后缀进行一次过滤，将常见TLD域名回补。 5. 将相似域名前缀与后缀组合，生成相似域名，根据信息熵进行排序，筛选出前3w条数据。 6. 提供三种域名过滤策略：whois探测、request探测、ping探测。检测精确度为：ping≥request≥whois。检测速度为：whois≥request≥ping。 7. 最后将结果以列表的方式返回给用户。 |
| **技术效果：**  本发明可以辅助企业用户判断是否存在被域名仿冒的可能，做出威胁预警，维护企业安全。相比旧的实现方法，本发明：   1. 支持任意域名检测 2. 实现高并发高精确度的仿冒域名检测，有效地提高了仿冒域名的检测效率。 3. 可以根据不同的用户需求，个性化检测策略，提高检测准确性和覆盖率。 |

附件、技术交底书—撰写示例

|  |
| --- |
| 发明名称：一种基于虹膜跟踪与OCR图像识别技术的信息推送方法及系统 |
| 申请类型：发明 实用新型 |
| 发明人： |
| 第一发明人姓名及身份证号： |
| **中文关键字：**虹膜识别、信息推送  **英文关键字：** |
| **现有技术的说明及其缺点：**  目前，信息推送主要是基于历史消费记录为主，通过分析用户的历史消费数据，找到用户消费的偏爱喜好，对用户进行信息推送。在线下的购物场景里，这种方式存在两个问题：一是不能发觉用户的即时需求，而当注意力分散后，这个营销的时机就可能错过。二是目前LBS的定位距离还无法做到短距离，推送优惠的商户可能与用户之间有一定距离，这个距离会影响用户的购买决定。 |
| **具体实施方案：**  本发明提出了一种基于虹膜跟踪与图像识别技术的信息推送方法及系统，抓取用户的即时消费需求，快速推送优惠提醒。关于业务流程和系统架构，具体如下：   1. **业务流程** 2. 用户虹膜焦点抓取：采集用户的虹膜信息，勾画用户的焦点范围，实际业务场景中，可对焦点距离做设置。 3. 视觉图像抓取：根据用户的焦点范围，并结合用户的视觉停顿时间，对视觉图像进行抓取。 4. OCR图像识别：根据图像抓取的内容，进行OCR文字识别，包括字符识别、汉子识别等，解析出文字内容。 5. 判定用户兴趣点：如发生用户焦点发生静止状态超过一定的时间范围，如5s，则判定该用户可能对商家感兴趣。 6. 判定是否属于相关信息商户：将视线聚焦的商户与后台优惠商户清单做对比，判断是否属于优惠商户。 7. 推送信息提醒：如属于相关信息商户，则向用户推送信息提醒，提醒既可直接发至智能眼镜，也可发至用户手机终端。 8. **系统架构**   根据上述业务流程，本系统分为虹膜焦点识别模块、视觉图像抓取模块、OCR图像解析模块、聚焦定时器、优惠商户数据库、商户对比模块、优惠发送模块。具体如下图所示：    各模块的主要功能与实现如下：   1. **虹膜焦点识别模块**：采集用户的虹膜焦点，实现对视觉焦点的跟踪。 2. **视觉图像抓取模块**：根据用户的焦点范围，采集对外的视觉画面，可对视觉画面设置距离限制。对图像范围内的信息进行二次过滤，抓取商户的logo、名称等重要信息。 3. **聚焦定时器**：判定用户的焦点的变化定时器，当用户焦点静止超过默认值时，将触发图像抓取与图像识别处理。 4. **优惠商户数据库**：存储优惠商户名单，以及对应的优惠活动信息。 5. **商户对比模块**：将图像识别后的商户信息与存储数据库做对比，如对比数据有重叠，则触发优惠发送操作。 6. **优惠发送模块**：向用户的手机终端或智能眼镜下发优惠推送通知。 7. **图像解析模块**：采用光学传感技术，对抓取图像进行内容识别，通过获得图片上的字符，经二值化、噪声去除、倾斜校正、字符切割等过程，完成字符识别，成功解析出图像中的汉字、字母、数字的内容。 |
| **技术效果：**  相对目前方案，本系统有如下优势：  **1、抓取用户的即时需求，快速推送优惠信息**：解决了线下无法即时抓取用户的实时需求的困难，帮助商家挖掘有潜在消费倾向的用户，并推送优惠信息。  **2、了解用户的潜在需求点，为消费行为分析提供支持**：通过搜集用户的关注点，了解用户的需求热点，为消费行为分析提供数据支持。 |