2020 年全国大学生信息安全竞赛

测试报告

作品名称：慧眼——基于客户端蜜罐和机器学习的风险网站检测系统

电子邮箱：824342218@qq.com

提交日期：2020 年 6 月 15 号

作品测试与分析

1.1 引言

我们设计并完成了“慧眼——基于客户端蜜罐和机器学习的风险网站检测系统”，可以针对性地对目前市面上系统地相关不足提出针对性的改进。该系统实现了黑名单法＋蜜罐检测法+机器学习三种方法多维度结合，并且为不同类型的恶意网站设计了不同的识别方案。从而大大提高了检测的准确性和针对性。下面我们针对检测的准确性、检测速度及成品的稳健性进行检测。

1.1.1 编写目的

编写测试文档的目的是通过整个测试过程，包括测试计划、测试设计、案例编写、测试总结等步骤去了解基于客户端蜜罐和机器学习的多维恶意风险网站识别系统的质量如何。是否存在缺陷，若存在缺陷则其原因是什么以及该如何修复。希望在本系统发布之前能将缺陷修复。同时向用户呈现本系统的具体测试细节，使用户能够明白系统的软件缺陷以及具体的操作方法，这将便于用户理解和使用本系统。本文档面向的读者主要是项目的开发人员和测试人员。

1.1.2 测试范围及方法

本测试对系统的各个模块先单独进行测试，再对产品整体进行测试，以保证开发人员既能将测试结果用于改进局部细节，又能对产品整体的稳健性进行完善。

- **系统功能**，我们采取模拟客户使用的动态黑盒模拟测试，从使用者体验角度出发，综合衡量系统性能和使用效果。
- **对比测试**，我们还与市面上的主流的产品进行横向对比，以直观显示出慧眼系统的优势与瓶颈，为我们在突破瓶颈时提供方向。

1.1.3 测试环境

测试环境一：云服务器

CPU	4 核
内存	8GiB
实例类型	I/O 优化
操作系统	Ubuntu 18.04 64 位
弹性网卡	eni-bp1810qgg86oxuu999g
系统盘大小	40GiB
带宽	3Mbps
Python 环境	ver3.6.9

测试环境二：PC

CPU	Intel(R) Core(TM)i5-8250U
内存	8.00GB
操作系统	Windows 10 家庭版 64 位
Web 浏览器	Chrome ver 83.0.4103.97
Python 环境	3.732
网络环境	10Mbps

1.1.4 系统可能风险

序号	终止条件	解决措施
1	系统崩溃、卡死	找出系统不稳定的原因，以及对相应模块进行修改，再重新进行测试。
2	模块集成测试中出现错误，导致有功能无法正常运行	对未正常工作的模块进行纠错、修改再重新进行集成测试以确保系统正常。
3	客户端蜜罐模块误判率过高	分析原因并修正蜜罐模块代码。
4	机器学习模块误判率过高	调整参数或适当增加特定数据集重新训练分类器。
5	前端页面展示错误	修改前端代码并重新测试。
6	服务器不稳定或出现错误	调整服务器配置或转用其他可靠服务器

1.1.5 测试结束条件

软件系统经过单元、集成、系统测试，分别达到单元、集成、系统测试的测试标准。

1. 单元测试标准

单元	结束标准
客户端蜜罐模块	正常运行，检测效率较高，检测准确率能达到 85%以上。
机器学习模块	正常运行，检测效率较高，检测准确率达 85%以上。
后端程序	流畅运行，能及时处理用户请求并运算后返回结果。
前端页面	流畅运行，提供方便且具有美感的前端交互和检测结果展示。

2. 集成测试标准

集成部位	结束标准
前端、后端之间接口	前端能将用户请求快速、准确地传递给后端处理后，后端能快速、准确地传递给前端进行展示。
后端、机器学习部分接口	机器学习能准确获得 url，在利用已训练好的分类器判断后能准确、快速返回给后端判断结果。
后端、客户端蜜罐部分接	蜜罐能准确获得 url，能利用爬虫技术

口	和引擎判断网页中各文件是否非法，然后快速返回给后端判断结果。

3. 系统测试标准

经过动态黑箱测试，使用者体验良好。系统能够结合客户端蜜罐及机器学习算法快速、准确判断 90%以上的 URL 并返回给前端进行展示。

1.2 系统功能测试过程

1.2.1 客户端功能测试

1.2.1.1 查询功能

测试编号	User-1.1
测试用例	用户正常输入正常格式 URL 进行检测
测试用例说明	验证前端能正常传递给后端用户查询的 URL，并验证能返回准确的结果
预置条件	1. 服务器处于打开状态 2. 进入检测服务主页
输入	1. 输入正常格式 URL 2. 点击确认
预期结果	正确返回判断结果
实际结果	a. 正常网页 诈骗博彩、资产敏感判断返回 false，黄色判断返回 false b. 色情网页 诈骗博彩、资产敏感判断返回 false，黄色判断返

	回 true c. 博彩诈骗等资产敏感网页 诈骗博彩、资产敏感判断返回 true，黄色判断返回 false

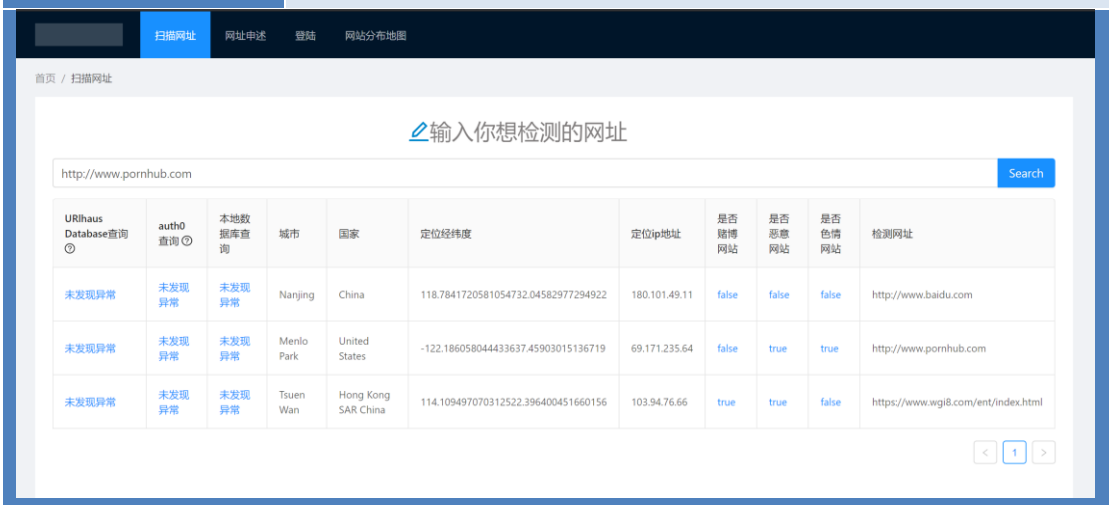

图 3.2.1.1a 查询功能正常测试

结果分析	实际结果与预期结果一致，网站查询功能运行正常

测试编号	User-1.2
测试用例	用户输入非法格式 URL 进行检测
测试用例说明	验证前、后端代码的健壮性与异常处理
预置条件	1. 服务器处于打开状态 2. 进入检测服务主页
输入	1. 输入非法格式 URL 2. 点击确认
预期结果	网站提示 URL 格式错误，且不干扰系统运行
实际结果	网站提示 URL 格式错误，且不干扰系统运行

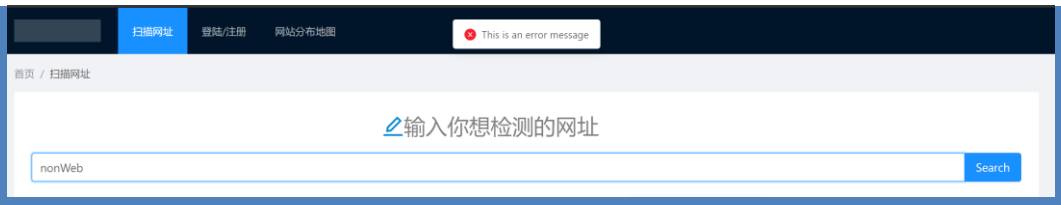

图 3.2.1.1b 查询功能非法输入测试

结果分析	实际结果与预期结果一致，在非法格式 URL 处理上的代码具有好的健壮性

1.2.1.2 用户注册

测试编号	User-2.1
测试用例	用户输入账号密码进行注册
测试用例说明	验证用户注册功能
预置条件	1.服务器处于打开状态 2.进入用户注册页面
输入	1.输入账号、密码 2.提交
预期结果	注册成功，能再次利用账号密码登录
实际结果	注册成功，能再次利用账号密码登录

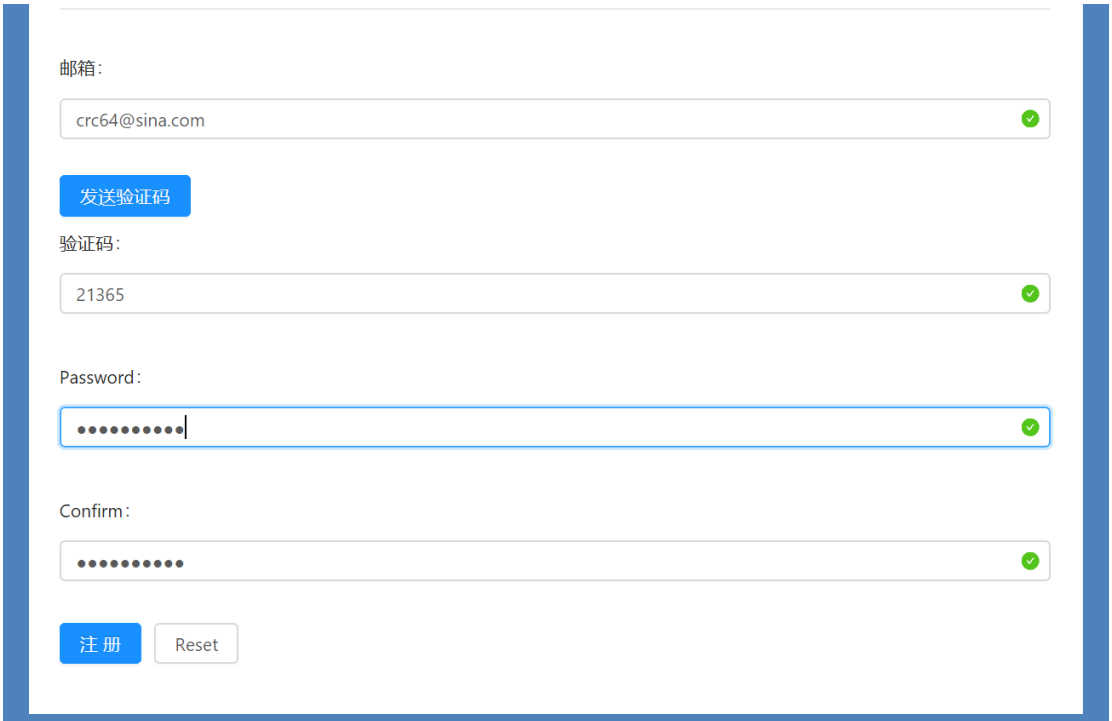

图 3.2.1.2 用户注册功能测试

结果分析	实际结果与预期结果一致，注册功能正常

1.2.1.3 网站位置分布与危险等级显示

测试编号	User-3.1
测试用例	检视网站地图页面
测试用例说明	验证网站位置查询与其前端展示的正确性
预置条件	1.服务器处于打开状态 2.进入检测服务主页查询 3.查询后进入网站地图页面检视
输入	查询几个合法与恶意网站，再点击网站地图检视页面
预期结果	显示查询过的网站位置，并且用颜色标注出其危险等级
实际结果	正确显示查询过网站的大致地理位置，并有颜色

	标注其危险等级

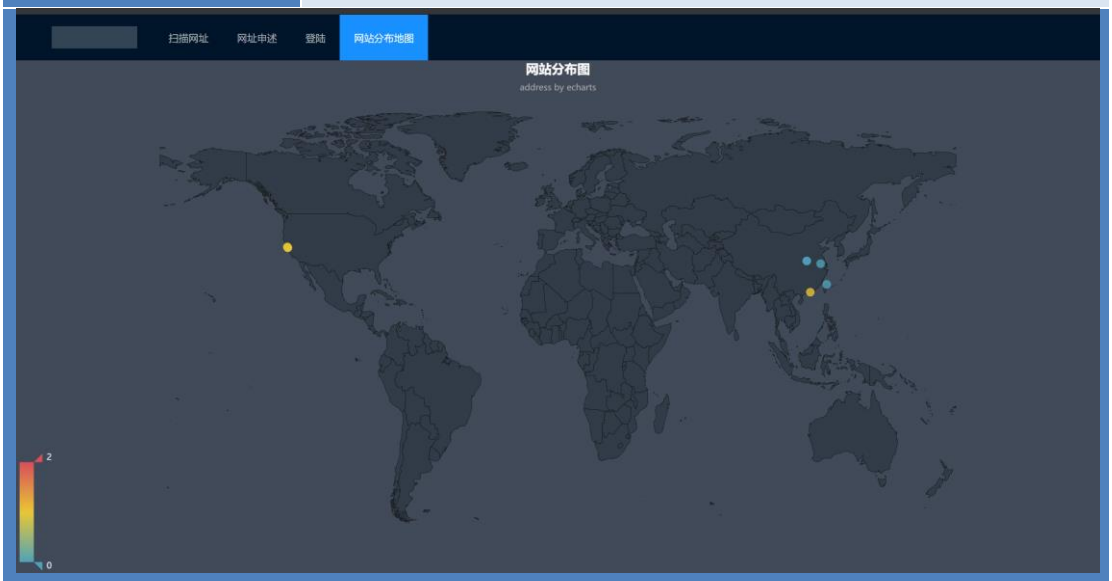

图 3.2.1.3 网站分布地图测试

结果分析	实际结果与预期结果一致，网站地图功能正常

1.2.1.4 用户申诉

测试编号	User-4.1
测试用例	申报网站分类或地理位置错误信息
测试用例说明	进入网址申报界面，填写申报信息，点击提交
预置条件	1.服务器处于打开状态 2.进入网站申诉页面
输入	填写申报信息后提交
预期结果	后台返回信息，提示已提交申诉信息
实际结果	后台返回信息，提示已提交申诉信息

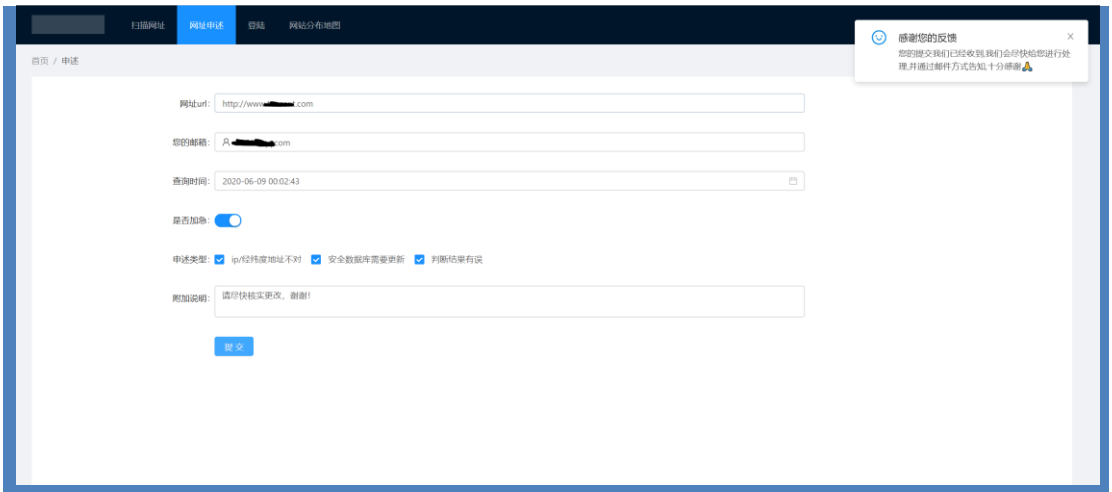

图 3.2.1.4 申诉功能测试

结果分析	实际结果与预期结果一致，申诉功能可用

1.2.1.5 管理员登录后台

测试编号	Admin-1.1
测试用例	管理员登录后台
测试用例说明	进入管理员登陆页面，输入账户和密码
预置条件	1.服务器处于打开状态 2.进入网站管理员登录页面 3.网站有存储管理员的账户
输入	按要求输入管理员账户与密码后点击提交
预期结果	导航栏出现“审批申诉”模块
实际结果	导航栏出现“审批申诉”模块

图 3.2.1.5 管理员登录后台功能测试

结果分析	实际结果与预期结果一致，管理员通过可登录进入审批申诉页面

1.2.1.6 管理员裁决

测试编号	Admin-2.1
测试用例	管理员裁决用户申诉
测试用例说明	登录管理员后进入审批申诉页面进行申诉处理
预置条件	1.服务器处于打开状态 2.进入网站管理员登录页面 3.网站有存储管理员的账户
输入	进入申诉处理页面，点击按钮进行操作
预期结果	后台成功显示出申诉内容，点击按钮可进行处理
实际结果	后台成功显示出申诉内容，点击按钮可进行处理

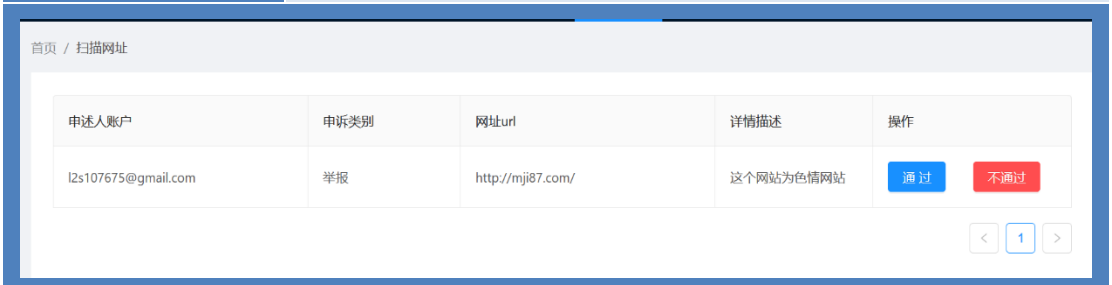

图 3.2.1.6 管理员裁决测试

结果分析	实际结果与预测结果一致，管理员裁决申诉功能正常，

1.2.1.7 产品信息安全测试

尽管我们在设计该系统时就考虑到了产品本身的信息安全，但出于客观、严谨考虑，且为了提供更稳定可靠的服务，我们还需要另外对产品本身进行一定的安全性测试，以保证系统使用者和系统的信息安全。

1.2.1.7.1 测试工具与方法

我们采用了 Acunetix Web Vulnerability Scanner 作为我们的测试工具。Acunetix 作为一款先进的 Web 漏洞扫描程序，可以独立使用，也可以作为复杂环境的一部分使用。它提供内置的漏洞评估和漏洞管理，以及许多与市场领先的软件开发工具集成的选项。

它具有以下强大的功能：

a)、自动的客本分析器,允许对 Ajax 和 web2.0 应用程序进行安全性测试

b)、业内最先进的深入的 SQL 注入和跨站脚本测试

c)、高级渗透测试工具,例如 Http Editor 和 HTTP Fuzzer

使用 Acunetix Web Vulnerability Scanner 作为测试工具，能够客观有力地检测我们产品的安全性。

1.2.1.7.2 测试结果

测试编号	Sec-1.1
测试用例	AWVS 自动化扫描漏洞
测试用例说明	使用 AWVS 测试系统是否存在安全漏洞
预置条件	1.服务器处于开启状态 2.拥有一台装有 AWVS 的个人电脑
输入	实用另一台装有 AWVS 的电脑对域名进行扫描
预期结果	扫描结果为低危
实际结果	扫描结果为低危

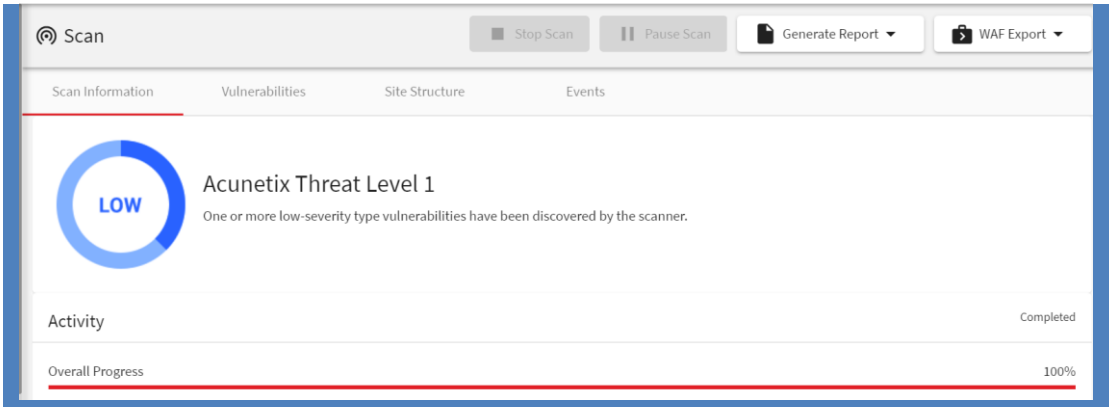

图 4.4.2 AWVS 测试结果

结果分析	实际结果与预期结果一致，系统安全性强，证明了网站本身的信息安全。

1.2.2 服务端功能测试

1.2.2.1 生成管理员账户

测试编号	Back-1.1
测试用例	生成管理员账户
测试用例说明	测试能否正常生成管理员账户
预置条件	服务器处于开启状态
输入	1. 进入服务器命令行模式 2. 执行代码文件夹中的 admin.py 文件 3. 进入 mysql 命令行 4. 执行相关数据库指令
预期结果	成功生成管理员账户
实际结果	成功生成管理员账户

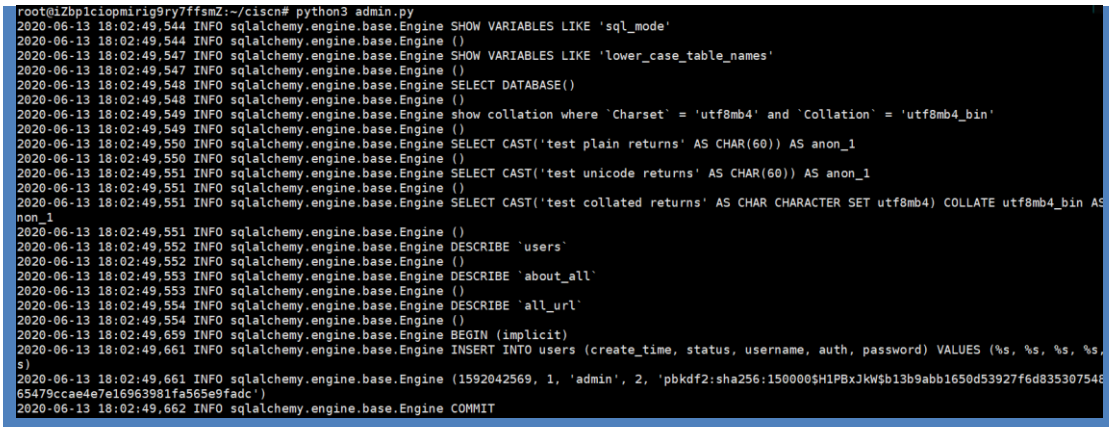

图 3.2.2.1 生成管理员账户测试

结果分析	实际结果与预期结果一致，数据库管理系统功能正常。

1.2.2.2 删除数据库中的网站

测试编号	Back-1.2
测试用例	删除数据库中的网站
测试用例说明	测试能否正常删除数据库中的网站
预置条件	服务器处于开启状态
输入	1.打开服务器端 2.进入 mysql 命令行 3.执行相关数据库指令
预期结果	成功删除该条记录

实际结果	成功删除该条记录

```
root@iZbp1ciopmirig9ry7ffsmZ:~# mysql -uroot -p
Enter password:
Welcome to the MySQL monitor.  Commands end with ; or \g.
Your MySQL connection id is 145
Server version: 5.7.30-0ubuntu0.18.04.1 (Ubuntu)

Copyright (c) 2000, 2020, Oracle and/or its affiliates. All rights reserved.

Oracle is a registered trademark of Oracle Corporation and/or its
affiliates. Other names may be trademarks of their respective
owners.

Type 'help;' or '\h' for help. Type '\c' to clear the current input statement

mysql> use ciscn;
Reading table information for completion of table and column names
You can turn off this feature to get a quicker startup with -A

Database changed
mysql> delete from all_url where url = 'www.baidu.com';
Query OK, 1 row affected (0.00 sec)

mysql> select * from all_url where url = 'www.baidu.com';
Empty set (0.00 sec)
```

图 3.2.2.2 删除数据库中的网站

结果分析	实际结果与预期结果一致，数据库中网站可被正常删除

1.3 对比测试

1.3.1 测试说明

我们从互联网上分别搜集了 100 个恶意软件下载网站，100 个钓鱼网址，100 个色情网站，100 个博彩网站，100 个正常网站。将这 500 个样本分别送入我们的**“基于客户端蜜罐和机器学习的风险网站识别系统”，腾讯网址安全中心，百度网址安全中心，** VirusTotals **四个不同的网站安全检测系统**中。我们将从总体检出率，假阳性率（实际为正常网站但检出为风险网站），假阴性率（实际为风险网站但检出为正常网站），分项检出率这四个不同的指标来分别衡量不同系统的优势和不足，也为接下来系统的改进提供思路。

1.3.2 总体检出率对比测试与分析

表 1.3.2 总体检出率测试

	慧眼	腾讯网址安全中心	百度网址安全中心	VirusTotals
风险网站样本数	400	400	400	400
检出数量	357	268	279	129
未检出数量	43	132	121	271
检出率	89.25%	67%	69.75%	32.25%

从表 1.3.2 可以看出，我们的系统在检出率上远超其余的三个系统，对其他几个系统均有显著性的优势。腾讯网址安全中心和百度网址安全中心检出率相近，均为 70%左右，这主要是由于他们两家拥有广大的用户，可以根据广大的用户反馈来添加黑名单的结果。在四种不同的系统中，VirusTotals 的检出率最低，这主要是因为 VirusTotals 主要针对的是恶意软件下载这种类型的风险网站进行检测，对其他类型的风险网站几乎没有检测能力。

1.2.3 恶意软件下载网站检出对比测试与分析

表 1.2.3 恶意软件检出率测试

	慧眼	腾讯网址安全中心	百度网址安全中心	VirusTotals
风险网站样本数	100	100	100	100
检出数量	80	24	28	89

未检出数量	20	76	72	11
检出率	80%	24%	28%	89%

在恶意软件检测方面，可以看到 VirusTotals 的检出率最高，主要原因是 VirusTotals 集成了 60 多种不同的引擎，在恶意软件检测方面较为擅长。而腾讯网址安全中心和百度网址安全中心在这一方面表现较差，检出率均仅为 30%不到，主要原因是大多数恶意软件下载网站生命周期较短，仅为几个小时不到，此时腾讯网址安全中心和百度网址安全中心所广泛采用的黑名单法则表现较差。

而我们使用的客户端蜜罐+机器学习检测恶意软件下载网站，虽然检出率达不到 VirusTotals 的 89%，但也达到了较为优秀的 80%。如果要进一步提升检出率，可以考虑在以后添加除了 ClamAV 的病毒扫描引擎，运用多引擎解析来提高检出率。

1.2.4 钓鱼网站检出对比测试与分析

表 1.2.4 钓鱼网站检测率测试

	慧眼	腾讯网址安全中心	百度网址安全中心	VirusTotals
风险网站样本数	100	100	100	100
检出数量	91	85	77	11
未检出数量	9	15	23	89
检出率	91%	85%	77%	11%

从表 1.2.4，可以看到我们为钓鱼网站所专门使用的基于预取的钓鱼网站检测法取得了优秀的效果，检出率达到了 91%。而腾讯网址安全中心在这方面表现

得也不错，值得注意的是，在 100 个钓鱼网站样本中，有 30 个是对腾讯产品的伪冒，而腾讯网址安全中心对这 30 个样本实现了 100%检出，由此可以看出，腾讯对自家产品的相关检测能力较为优秀。而 VirusTotals 由于只能针对恶意软件下载网站进行分析，在这里表现不佳，检测出的 11 个样本为恰好拥有恶意软件的样本。

1.2.5 色情网站检出对比测试与分析

表 1.2.5 色情网站检测率测试

	慧眼	腾讯网址安全中心	百度网址安全中心	VirusTotals
风险网站样本数	100	100	100	100
检出数量	97	73	85	20
未检出数量	3	27	15	80
检出率	97%	73%	85%	20%

色情网站检测是慧眼系统表现得最好的一个方面，检出率达到了 97%，我们通过爬取外网色情平台上的关键词利用文本聚类等相关技术获取相关的关键词获取色情关键词列表，再去除一些生活中的常用分词来使得分词算法更加真实，从而建立起一个最真实的语料库。由于使用了这一技术，使得我们系统对色情网站有着极高的检出率。

VirusTotals 系统对色情网站的检测是其在非恶意软件下载类网页中表现最好的，原因可能是色情网站很大概率也是携带各种恶意跳转，恶意下载。

1.2.6 博彩赌博网站检出对比测试与分析

表 1.2.6 博彩赌博类网站检测率测试

	慧眼	腾讯网址安全中心	百度网址安全中心	VirusTotals
风险网站样本数	100	100	100	100
检出数量	89	86	89	9
未检出数量	11	14	11	91
检出率	89%	86%	89%	9%

从表 1.2.6 可以看出，对于博彩网站，慧眼系统，腾讯网址安全中心和百度网址安全中心表现均较为优秀。这主要是因为博彩网站相当大一部分都带有色情信息，而容易同时被色情网站的识别算法所识别出来。但比起腾讯网址安全中心和百度网址安全中心，慧眼系统拥有独特的博彩网站识别算法。算法中主要加入了对正常网站的数据测验集。由于金融系统的特殊性，很多小众但合法的金融网站容易被误判为风险网站。这一点在腾讯网址安全中心和百度网址安全中心上均没有被妥善解决，而我们的系统可以根据正常的数据集不断学习适当的修改模型，这在很大程度上减少了误报。

1.2.7 假阳性率对比测试与分析

表 1.2.7 假阳性率测试

	慧眼	腾讯网址安全中心	百度网址安全中心	VirusTotals

正常网站样本数	100	100	100	100
正确判断次数	92	82	86	98
误判次数	5	18	14	2
假阳性率	5%	18%	14%	2%

从表 1.2.7 可以看出，假阳性方面，表现得最优秀的是 VirusTotals 系统，这是由于其特异性针对恶意软件下载网页，覆盖面较小的原因。在其他三种检测系统中，慧眼系统的表现也要明显优于其他两款系统。造成此种区别的主要就是在博彩赌博的检测上，腾讯网址安全中心和百度网址安全中心由于缺乏针对性的算法且一味追求效率，造成对网页检测的误报率较高。这在生活中也屡见不鲜，互联网上很多个人站长就一直在反映被腾讯或百度误报的问题，而我们的系统通过设计针对性的算法，有效的解决了这一问题，把误报率降低到 5%，适合于大规模使用。

1.4 测试总结与改进空间

1.4.1 总结

经过我们数次缜密的测试，我们将 bug 出现的概率降到最低，并通过测试验证了我们预想的设计已完全实现：

- 前后端分离，可插拔性高，bug 概率被降到最低
- UI 界面宜人、友好，运行流畅，结果展示直观
- 结合客户端蜜罐、机器学习、黑名单等方法有机结合综合判断网站安全性的效果显著。检测综合全面，不留短板；检测速度快，方便实用
- 用户登录、后台管理、申诉以及申诉审批过程、邮件回复审批结果等十分流畅，体验良好

- 检测、防止非正常输入的机制起效，系统在容错性和代码健壮性良好
- 系统本身有较高信息安全的水准
- 与同类产品相比，具有总体检出率、恶意软件下载网站检出率、钓鱼网站检出率、色情网站检出率、博彩赌博网站检出率高等显著优势

1.4.2 改进空间

- 表现形式上还可以更加丰富，由于前后端分离、可插拔的优良特性，之后我们还可以方便、快捷地进行完善
- 该系统的部分功能还可以提供 API 给其他网站或平台（如不良信息识别模块可以用在论坛网站对不良信息出现进行警告），从而进一步增加我们慧眼系统的适用性和使用范围
- 我们选择的算法模型还可以根据具体领域应用需求调整参数和数据集，来达到动态适应特定场景的效果