**平台说明文档**

**一、平台综述**

本平台主要实现漏洞检测的功能，用爬虫作为中间功能获取网站资源，以及还有下载检测报告和推送报告到邮箱两个辅助功能，以上集成到一个Django架构搭建的网页平台上，实现了一个简便的前端可以和用户交互。



图表 1 平台主要界面

**二、爬虫模块介绍**

爬虫作为中间功能，通过调用python3的urllib3库，能够请求指定各种类型的URL资源，例如HTML文件、JS文件等，同时编写对应的正则表达式，使爬虫模块能够提取出HTML文件中指定类型资源的URL地址，实现深度爬取资源；此外，在将爬取的资源下载到本地时，还保留了原本资源的目录结构，方便后续的漏洞扫描和资源管理，在爬取资源并检测完毕之后，还会删除在本地下载的资源，不给服务器的存储造成压力。

**三、漏洞检测模块介绍**

用户给出代码本地路径，平台以路径扫描的方式，逐层深入，并将文件集以路径存下来以供之后扫描使用，对于漏洞检测，使用正则匹配，匹配的模式写在了pattern.xml，用户可通过修改pattern.xml修改匹配的模式，由于使用xml存储，可读性、可修改性高。现有的特征主要涵盖常用后端web开发语言php，js, java, python，前端开发语言，js,html等，根据不同语言，制定了不同的特征。

使用演示：



图表 2 输入url和线程数，点击enter，开始扫描



图表 3 点击Clean，即可清空扫描结果



图表 4 点击Download report，即可下载扫描结果的pdf文件到本地