小明是 windtalker 系统的超级管理员，日常生活中他负责进行灰黑产网页的监测与管控，他将带我们详细地了解此系统。

小明输入自己的账号密码，成功进入到系统首页，首页上有许多子功能的快捷方式。

首先展示系统的大屏，它将给我们带来直观的数据感受。

左上角可以看到系统对检测网页的数量和其中的违规网页数量随时间做了一个统计。

下面有正常网页和违规网页的比例图，

以及我们在这个过程中发现的灰黑词的词云图。

中间是通过 whois 的反查信息，对违规网页进行了中国地区内的地域分类。我们可以直观地感受到，违规网站在不同省份的占比不同，其中北京、广东和江苏注册的灰黑产网页数量最为庞大。

下方滚动展示出系统最新检测出的网页网址，这些网页都是最近发现的。

右上角是系统中进行的任务以及进度的展示。当系统接收到新的网页检测列表时，它就会生成一个新的任务。检测过程分为浏览器访问、网页文字判别、图片信息提取、以及源码结构分析等子任务。

任务下方是系统利用网页结构特征对发现的违规网站进行了聚类分析，这里展示出了组织 001 的 5 个域名。

为了验证这些信息的准确性，小明决定对新检测到的违规网页进行实际访问。

我们发现这的确是个赌博类网站。

系统的检测与发现模块，由网页检测、whois 检测、灰黑词检测三个模块组成。

之前我们搜索的那个网页，系统是否能准确实时判别，又能否给出更详尽的反馈信息呢？

我们对它进行检测（http://www.dingyuejixie.com/），

在结果页面，我们可以看到该违规网页被准确判为了赌博网页。该页面同时还记录了历史网页的检测情况，我们可以通过网页的检测时间来完成对历史记录的快速定位。

点击查看详细信息按钮，

详细信息里包含了网页域名的反查信息、网页截图以及网页的 HTML 基本标签。这些信息是支撑起系统运作的核心数据。系统利用 HTML 基本标签内以及截图内的文本内容进行文本识别、利用截屏图像进行图像特征提取，此外还进行了 HTML 源码结构和 URL 特征的分析。

点击分析灰黑词按钮，

灰黑词信息包含了置信度最高的 4 个词和对应的置信度。（我们在浏览器上搜索第一个关键词：美狮贵宾会，点击搜索结果）。

可以看到，美狮贵宾会引出了大量的恶意赌博界面。即，我们成功发现灰黑词，并且借其发现了更多的赌博网站。

whois 检测可以给用户提供域名反查功能。（依旧查询这个网址 http://www.dingyuejixie.com/,）

可以看到反查的详细信息包括 ip 地址、域名注册商、注册人等信息。这些信息帮助到小明对灰黑产组织进行发现、定位和追踪，以进一步治理网络环境。

这里的灰黑词检测模块可以通过分析用户输入的文本，来得到置信度最高的四个灰黑词，以及相应的评分。这里我们粘贴一段文本进来，搜索置信度最高的第一个关键词：永利彩票，我们再次到浏览器中进行搜索，成功发现了更多的灰黑产赌博网站。

暂停

当然，小明不仅会关注系统能否成功分析我们所指定的网页，更会好奇这些灰黑产大数据所展示出来的详细统计信息。在数据分析模块，系统分析了灰黑产网页数量、位置及时间分布以及进一步进行了组织发现和灰黑词的统计。

灰黑词频统计，系统对发现的灰黑词和其出现的频次进行了统计，频次越高、色调越暖。（出现频次最高的是天豪棋牌和捕鱼天王）。

域名反查信息统计，系统对灰黑产网页域名的注册商进行了统计，这里展示出了拥有灰黑产域名数量 TOP10 的域名服务商，我们可以看到 GoDaddy 注册有最多的灰黑产恶意域名。此外，系统还展示出了注册有灰黑产 IP 数量前六的注册人姓名及其注册 IP。

数量分析：系统对所检测的不同类别的网站数量及其占比进行了统计。

位置分析模块分别从世界级的国家维度和国内的省份维度展示了内容威胁性灰黑产网站的地域分布特点。

时间分析展示了最近一周和近几月系统所分析到的不同类别的网页的数量变化情况。

组织发现，为了节约成本和突出组织的风格，同一灰黑产组织的灰黑产页面往往套用同一个网页源码模板（mu），因此其源码结构往往非常相似。系统通过计算不同网站之间的网页源码结构相似度，展示了突出的 26 个疑似组织。

这里对系统发现的灰黑词进行了词云统计和数量分析，可以看到昨日新增 top10 和增长最快的灰黑词。

小明是系统的超级管理员，他不会局限于只查看系统所反馈的恶意与否的结果，小明有时更喜欢手动验证与人工分析数据。为了保证系统的可利用性，系统还人性地提供了原始数据展示模块。

在灰黑词信息模块中，小明可以手动添加新灰黑词名单，不必局限于系统对网页的自动

拓展和进而对灰黑词的自动收集。便于打破局限，更快发现更多最新行话。

网页信息模块，小明可以方便地对网页原始的 HTML 标签等数据进行直接查询。(随意跳两个网页）

为了使规则更加灵活，以适应不同的应用场景及时间需求，系统配置模块提供了任务管理和爬虫管理的功能。

在任务管理部分，用户可以自己设定搜索种子词、指定爬虫所使用的浏览器、可以设置分类判别所考虑的特征维度以及手动添加黑白名单等。这里我们设置一个新任务：代号 007，设定种子词为威尼斯，等待任务开始。

暂停

在主界面上我们可以看到新设定的任务：代号 007，已成功加入到未开始任务栏。同时在此界面上我们还可以方便地查询到历史任务的执行情况。

一段时间过后……

暂停

现在代号 007 任务已执行完毕，我们点击它并查看检测结果。可以看到系统成功检测出了大量含有种子词“威尼斯”的恶意灰黑产网页。右侧我们依旧可以查看网页的详细信息以及查询对此网页进行灰黑词分析的结果。

紧接着，是爬虫管理，我们可以轻松获取爬虫的运行状态

在爬虫管理的子模块 最新内容 中，我们可以对爬虫最新收集到的网页信息进行信息检索。可以看到，针对种子词威尼斯，系统爬虫爬取到了很多条最新数据记录。

再接着是消息通知，可以将系统信息以微信和邮件的方式通知。例如告知小明某一任务已执行完毕。

核心功能模块就已经介绍完毕，作为超级管理员，小明很清楚自己的能力范围和责任。

他可以进行角色和用户管理、权限分配、　　他可以设置菜单、配置 api。

这里还审计着所有人员的对系统的操作记录。

此外，系统还有个人信息的展示，以及对服务器状态信息的监控。

以上就是我们系统展示的所有内容，谢谢大家。