

网页智能识别的使用

1. 网页智能识别的作用

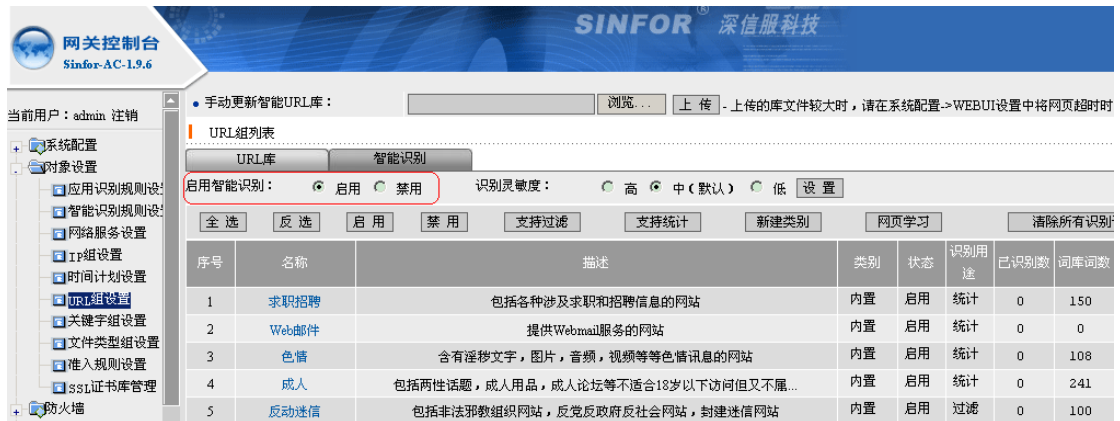
开启网页智能学习的功能，AC 可通过学习内网用户访问的网页内容，智能分辨出此网页的类型，做到对内置 URL 库中的补充，同时做到更全面的记录或封堵某类型的网站。

2. 网页智能识别的设置

2.1. 总体设置

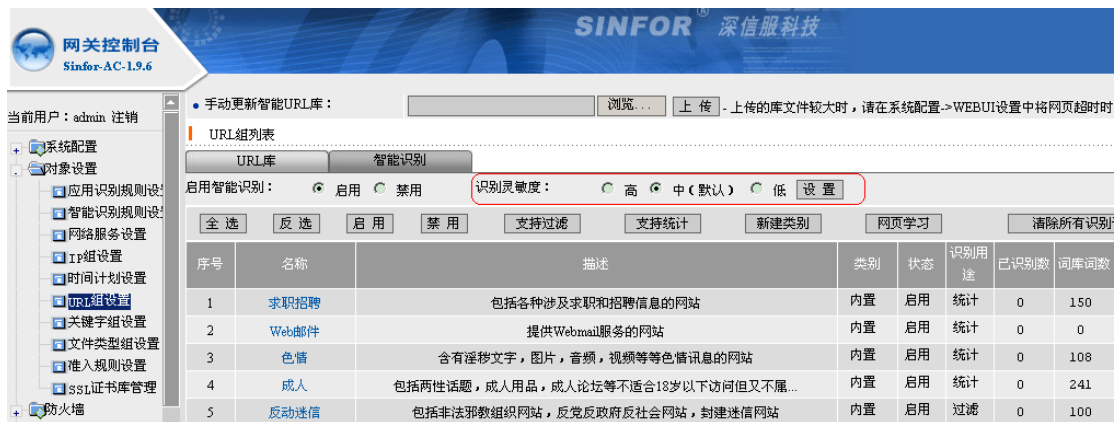
第一步：开启 URL 智能识别，设置如下图所示：

勾选“启用”智能识别的总开关。



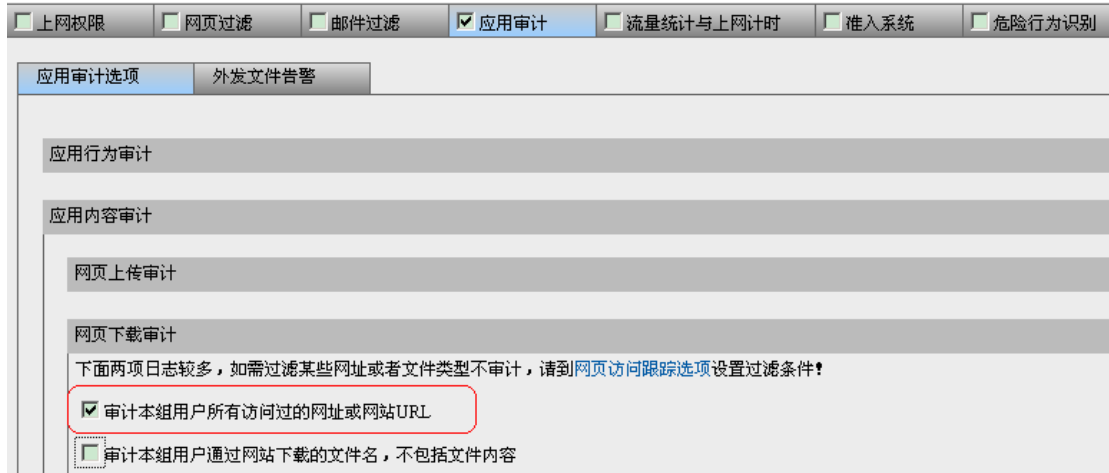
第二步：设置“识别灵敏度”，如下图所示：

建议灵敏度设置为“中（默认）”。（灵敏度越高越容易误判，灵敏度越低漏判的几率较大，也可根据实际情况适当调整灵敏度）。



第三步：开启 URL 审计：

如果需要对某用户组访问的网站进行智能识别,并使智能识别出的 URL 对此组用户生效,则必须设置一条策略,勾选“审计本组用户所有访问过的网址或网站 URL”,如下图所示:



2.2. 内置 URL 类型的设置

内置 URL 类别可以直接启用智能识别,因为用于智能学习网页的词库是内置就有的,并且内置词库和内置 URL 一样,会进行定时更新。

以“求职招聘”类网站为例:

第一步:开启“求职招聘”类别的智能识别,如下图:



第二步:设置“识别用途”,如下图:

“过滤”:智能识别出的 URL,在第二次被访问时会被封堵。(第一次访问时 AC 需要学习此网站,故不能在第一次访问就进行封堵。)

“统计”:识别出的 URL 只做记录,不会进行封堵。



第三步：设置“识别结果用于”，如下图所示：

“整个域名”：第一次访问 `xx.com/yy/old.htm` 时候，识别出该网页属于求职招聘网页，则 AC 记录“`xx.com`”（整个域名）为求职招聘网页，内网任意用户第二次访问地址 `xx.com/zz/new.htm`（相同域名不同目录）时则直接判断为求职招聘网页。

“完整目录”：第一次访问 `xx.com/yy/a.html` 时，识别出该网页为求职招聘类，则 AC 记录“`xx.com/yy/`”（完整目录）为求职招聘网页，内网任意用户下一次访问 `xx.com/yy/b.html` 网页时，会被判断为求职招聘类，但是访问 `xx.com/zz/c.html`（相同域名不同目录），不会直接判断成求职招聘，而是重新学习识别。**推荐使用“完整目录”**



第四步：如果需要过滤此类网站，还需要设置上网策略，设置封堵相应类别的 URL，并关联给用户，策略设置此处不再赘述。

第五步：查询识别结果，如下图所示：

点击此页面的“查询识别记录”，进入识别结果查询页面。



选择需要查询的 URL 类别，点击“查询”，即可查询到智能 URL 识别出的 URL。

如下图：



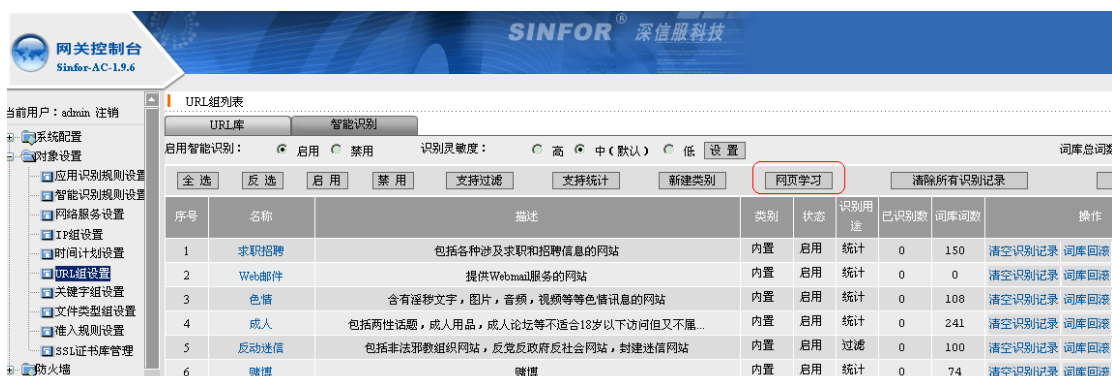
第六步：可以在数据中心查询到访问记录，并提示是否是智能识别识别的结果，如下图所示：



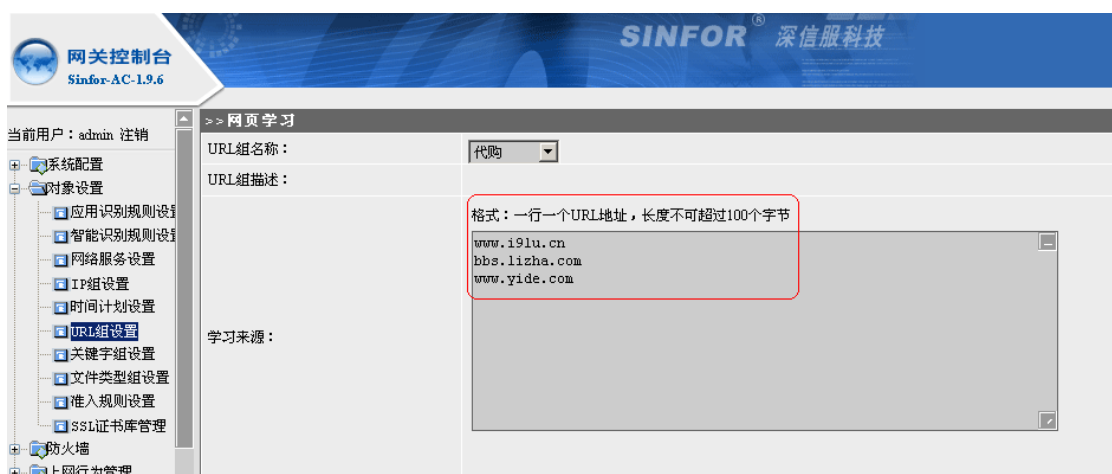
2.3. 自定义 URL 智能识别

客户自定义的 URL 类别，因为在默认情况下没有对应的词库，所以无法进行 URL 智能识别，如果需要对自定义的 URL 类别使用智能识别，需要进行的第一步操作是：学习网页，积累词库。

第一步：进行网页学习，积累词库，如下图：



例如要学习“代购”类型的网站，输入几个典型的代购类型的网站，提供学习的网站越典型越多，学习的词库就越精确。



学习完成后，可以看到对应的词库次数增加，如下图所示：



后续步骤请参考前一小节介绍，此处不再赘述。

注意：最多只能设置 10 个自定义 URL 组启用智能 URL 识别。

2.4. 排除关键字的使用

排除关键字的设置用于网页中包含此关键字时，则此网页不属于这类网站，对智能识别的 URL 生效。

例如：打击 赌博的网站并不属于赌博类别，但是可能含有赌博类别的关键字，如果有类误判，可将这些反义关键词填入“排除关键字”列表中，这样智能识别就不会将此网站判定为赌博类网站。

设置界面如下图：



3. URL 识别辅助选项

URL 识别辅助选项的功能指的是通过网页中的关键字识别网站类型，这种判断方法并不像智能识别那样智能和精确，只要网页上有此关键字就判断为某类型，比较容易误判，请慎用。

需求举例：客户要求网页中包含有“六合彩”的关键字判定为赌博网站，并进行封堵。设置步骤如下：

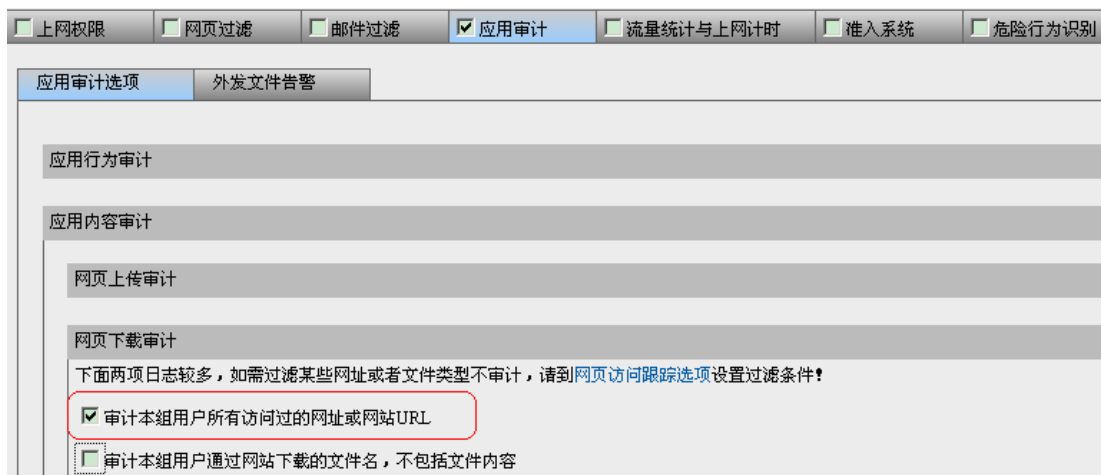
步骤一：开启 URL 智能识别，设置如下图所示：

勾选“启用”智能识别的总开关。



第二步：开启 URL 审计：

如果需要对某用户组访问的网站进行 URL 辅助识别，并使识别出的 URL 对此组用户生效，也必须设置一条策略，勾选“审计本组用户所有访问过的网址或网站 URL”，如下图所示：



第三步：设置辅助识别选项，定义关键字。

例如：凡是包含有“六合彩”的网页，即判断为“赌博”类别。



第四步：设置 URL 识别用途为“过滤”。



第五步：设置策略，拒绝“赌博”类 URL，并将此策略关联给用户。

4. 其他选项设置

4.1. 词库回滚

自定义的 URL 可以进行词库学习，当发现词库准确性不高时，可以将词库回滚至前次学习后的状态，避免误判。

如下图所示，点击“词库回滚”：



选择需要回滚的词库，点击“回滚”即可。



4.2. 删除或转移智能识别的 URL

如果发现有URL被误判了，可以将此URL转移到正确的URL组中，设置界面如下图所示：将识别成娱乐的url“www.tom51.com”转移到“电影音乐”类别。



如果需要删除www.tom51.com这个URL，则在进行上一步移动操作后，在传

统URL组“电影音乐”中进行删除，如下图所示：直接删除此URL，点击确定即可。

