



(12)发明专利申请

(10)申请公布号 CN 109561163 A

(43)申请公布日 2019. 04. 02

(21)申请号 201710892706.7

(22)申请日 2017.09.27

(71)申请人 阿里巴巴集团控股有限公司

地址 英属开曼群岛大开曼资本大厦一座四
层847号邮箱

(72)发明人 张旭俊

(74)专利代理机构 北京三友知识产权代理有限
公司 11127

代理人 李辉

(51) Int. Cl.

H04L 29/12(2006.01)

H04L 29/06(2006.01)

权利要求书3页 说明书15页 附图4页

(54)发明名称

统一资源定位符重写规则的生成方法及装
置

(57)摘要

本申请提供了统一资源定位符URL重写规则的生成方法及装置,其中,该方法包括:获取目标网站的目标URL集;所述目标网站为:待生成统一资源定位符URL重写规则的网站;获取所述目标URL集中相互对应的前缀参数和资源参数的参数集,其中,所述资源参数为所述前缀参数的子路径;依据所述参数集生成所述目标网站的URL重写规则集。采用本申请实施例,可以依据访问日志自动分析出网站的URL重写规则,无需人工参与。



1. 一种统一资源定位符URL重写规则的生成方法,其特征在于,该方法包括:
获取目标网站的目标URL集;所述目标网站为:待生成统一资源定位符URL重写规则的网站;
获取所述目标URL集中相互对应的前缀参数和资源参数的参数集,其中,所述资源参数为所述前缀参数的子路径;
依据所述参数集生成所述目标网站的URL重写规则集。
2. 根据权利要求1所述的方法,其特征在于,所述获取目标网站的目标URL集,包括:
对目标网站的访问日志中的初始URL集进行预处理,得到目标URL集。
3. 根据权利要求2所述的方法,其特征在于,所述对目标网站的访问日志中的初始URL集进行预处理,得到目标URL集,包括:
依据超文本传输协议HTTP状态码,从目标网站的访问日志中的初始URL集中过滤非法URL请求对应的非法URL;
针对过滤了非法URL后的初始URL集进行规范化处理,得到规范URL集,所述规范URL集中的规范URL包括:域名、路径和文件名;
对所述规范URL集进行去重处理,得到目标URL集。
4. 根据权利要求1所述的方法,其特征在于,所述获取所述目标URL集中的前缀参数和资源参数的参数集,包括:
基于预设分隔符对所述目标URL集中的各目标URL进行分割,分别得到各目标URL对应的字符数组;
依据所述字符数组中各字符串组成所述目标URL的顺序,分别确定所述各目标URL中对应的前缀参数和资源参数,以得到参数集。
5. 根据权利要求4所述的方法,其特征在于,所述依据所述字符数组中各字符串组成所述目标URL的顺序,分别确定各目标URL中对应的前缀参数和资源参数,包括:
获取任意一个字符数组作为当前数组,执行数组循环流程,所述数组循环流程包括:
按照从前到后的顺序,获取所述当前数组中的第一个字符串作为当前前缀参数;
将所述当前前缀参数与其后相邻的资源参数对应保存至初始参数集中;
判断所述当前前缀参数是否在初始URL重写规则集中,如果是,则将所述当前前缀参数与预设重写参数组合为更新前缀参数;如果否,则将所述当前前缀参数与其后相邻的资源参数组合为更新前缀参数;
以所述更新前缀参数为当前前缀参数,执行所述将所述当前前缀参数与其后相邻的资源参数对应保存至初始参数集中的步骤,直至当前目标数组的所有字符串都循环完毕;
判断所有字符数组是否都循环完毕,如果否,则将任意一个未循环的字符数组作为当前数组,触发执行所述数组循环流程;
如果是,则将所述初始参数集作为目标URL集对应的目标参数集。
6. 根据权利要求5所述的方法,其特征在于,所述依据所述路径参数和非路径参数生成所述目标网站的URL重写规则集,包括:
针对各前缀参数,分别判断所述前缀参数下资源参数的数量是否大于预设阈值,如果是,则将所述前缀参数更新至所述初始URL重写规则集中,得到更新后的URL重新规则集,直至所述初始URL重新规则集不再更新;

将更新后的URL重写规则集确定为所述目标URL重写规则集。

7. 根据权利要求1所述的方法,其特征在於,还包括:

依据所述目标网站的URL重写规则集,将待映射URL映射至重写后的URL。

8. 根据权利要求7所述的方法,其特征在於,所述依据所述目标网站的URL重写规则集,将待映射URL映射至重写后的URL,包括:

对所述待映射URL进行规范化,得到规范化后的URL;

基于预设分隔符对所述规范化后的URL进行分割,分别得到分割后的字符数组;

依据分割后的字符数组中各前缀参数在所述URL重写规则集中的匹配结果,将待映射URL映射至重写后的URL。

9. 根据权利要求8所述的方法,其特征在於,所述依据各分割后字符数组中各前缀参数在所述URL重写规则集中的匹配结果,将待映射URL映射至重写后的URL,包括:

按照从前到后的顺序,获取所述分割后的字符数组中的第一个字符串作为当前前缀参数;

判断所述当前前缀参数是否在所述URL重写规则集中,如果是,则将所述当前前缀参数与预设重写参数组合为更新前缀参数;如果否,则将所述当前前缀参数与其后相邻的资源参数组合为更新前缀参数;

以所述更新前缀参数为当前前缀参数,执行所述将所述当前前缀参数与其后相邻的资源参数对应保存至初始参数集中的步骤,直至所述分割后的字符数组中的所有字符串都循环完毕;

获取所述更新前缀参数作为重写后的URL。

10. 根据权利要求9所述的方法,其特征在於,在所述当前前缀参数在所述URL重写规则集中的情况下,还包括:

获取所述当前前缀参数对应的资源参数以及所述资源参数的值;

将所述资源参数、资源参数的值、查询字符串与所述重写后的URL对应保存。

11. 一种URL扫描方法,其特征在於,该方法包括:

获取预先生成的URL重写规则集,和,目标网站的待扫描的初始URL集;所述URL重写规则集采用如下方式生成:获取目标网站的目标URL集,所述目标网站为:待生成统一资源定位符URL重写规则的网站;获取所述目标URL集中相互对应的前缀参数和资源参数的参数集,并依据所述参数集生成所述目标网站的URL重写规则集;

依据所述URL重写规则集对所述初始URL集中的初始URL进行重写,得到重写后的初始URL集;

对所述重写后的初始URL集进行去重处理,得到目标URL集;

对所述目标URL集中的目标URL进行扫描。

12. 一种URL重写规则的生成装置,其特征在於,该装置包括:

获取URL集单元,用于获取目标网站的目标URL集;所述目标网站为:待生成统一资源定位符URL的重写规则的网站;

获取参数单元,用于获取所述目标URL集中相互对应的前缀参数和资源参数的参数集,其中,所述资源参数为所述前缀参数的子路径;

生成单元,用于依据所述参数集生成所述目标网站的URL重写规则集。

13. 一种扫描器,其特征在于,该扫描器包括:

获取URL单元,用于获取预先生成的URL重写规则集,和,目标网站的待扫描的初始URL集;所述URL重写规则集采用如下方式生成:获取目标网站的目标URL集,所述目标网站为:待生成统一资源定位符URL重写规则的网站;获取所述目标URL集中相互对应的前缀参数和资源参数的参数集,并依据所述参数集生成所述目标网站的URL重写规则集;

重写单元,用于依据所述URL重写规则集对所述初始URL集中的初始URL进行重写,得到重写后的初始URL集;

去重单元,用于对所述重写后的初始URL集进行去重处理,得到目标ULR集;

扫描单元,用于对所述目标URL集中的目标URL进行扫描。

14. 一种网络设备,其特征在于,该网络设备包括:处理器、存储器、网络接口和总线系统;

所述总线系统,用于将所述网络设备的各个硬件组件耦合在一起;

所述网络接口,用于实现所述网络设备与至少一个其它网络设备之间的通信连接;

所述存储器,用于存储程序指令和/或数据;

所述处理器,用于读取所述存储器中存储的指令和/或数据,执行以下操作:

获取目标网站的目标URL集;所述目标网站为:待生成统一资源定位符URL重写规则的网站;

获取所述目标URL集中相互对应的前缀参数和资源参数的参数集,其中,所述资源参数为所述前缀参数的子路径;

依据所述参数集生成所述目标网站的URL重写规则集。

统一资源定位符重写规则的生成方法及装置

技术领域

[0001] 本申请涉及互联网数据处理技术领域,特别涉及一种统一资源定位符(Uniform Resource Locator,URL)重写规则的生成方法及装置,一种基于URL重写规则的URL扫描方法及扫描器,以及一种网络设备。

背景技术

[0002] 为了保证网站的安全性,通常可以利用扫描器对网站进行安全扫描。扫描器可以利用网站的Web访问日志作为输入源,将URL去重后对各个URL下的各个参数进行扫描。因为在Web访问日志中,可能会有成千上万个的URL,但是有可能大量不同的URL都表示一个扫描路径,因为URL中还会包括无意义参数,这些无意义参数并不是路径的组成部分。在这种情况下,扫描器仍然需要对这大量的URL进行分别扫描,使得扫描的工作效率较低。

[0003] 现有技术中,为了提高扫描器的工作效率,可以为网站由开发人员人工配置一些URL重写规则,通过URL重写规则来标示出URL中的无意义参数有哪些,以便根据URL重写规则对原始URL进行映射后,能够将很多表示同一个扫描路径的URL过滤掉,只保留一个URL供扫描器扫描即可。

发明内容

[0004] 但是发明人在研究过程中发现,基于人工配置URL重写规则的方式,需要开发人员观察所有Web日志并基于观察结果手动配置,不仅存在由于Web日志的数据流较大而容易出错的情况,还使得配置URL重写规则的方式浪费了较大的人力物力成本。

[0005] 基于此,本申请提供了一种URL重写规则的生成方法及装置,一种基于URL重写规则的URL扫描方法和扫描器,以及一种网络设备,用以通过按照一定的规则自动分析网站的Web访问日志的URL,无需人工参与即可生成该网站的URL重写规则。

[0006] 为了解决上述问题,本申请公开了一种URL重写规则的生成方法,该方法包括:

[0007] 获取目标网站的目标URL集;所述目标网站为:待生成统一资源定位符URL重写规则的网站;

[0008] 获取所述目标URL集中相互对应的前缀参数和资源参数的参数集,其中,所述资源参数为所述前缀参数的子路径;

[0009] 依据所述参数集生成所述目标网站的URL重写规则集。

[0010] 其中,所述获取目标网站的目标URL集,包括:

[0011] 对目标网站的访问日志中的初始URL集进行预处理,得到目标URL集。

[0012] 其中,所述对目标网站的访问日志中的初始URL集进行预处理,得到目标URL集,包括:

[0013] 依据超文本传输协议HTTP状态码,从目标网站的访问日志中的初始URL集中过滤非法URL请求对应的非法URL;

[0014] 针对过滤了非法URL后的初始URL集进行规范化处理,得到规范URL集,所述规范

URL集包括:域名、路径和文件名;

[0015] 对所述规范URL集进行去重处理,得到目标URL集。

[0016] 其中,所述获取所述目标URL集中的前缀参数和资源参数的参数集,包括:

[0017] 基于预设分隔符对所述目标URL集中的各目标URL进行分割,分别得到各目标URL对应的字符数组;

[0018] 依据所述字符数组中各字符串组成所述目标URL的顺序,分别确定所述各目标URL中对应的前缀参数和资源参数,以得到参数集。

[0019] 其中,所述依据所述字符数组中各字符串组成所述目标URL的顺序,分别确定各目标URL中对应的前缀参数和资源参数,包括:

[0020] 获取任意一个字符数组作为当前数组,执行数组循环流程,所述数组循环流程包括:

[0021] 按照从前到后的顺序,获取所述当前数组中的第一个字符串作为当前前缀参数;

[0022] 将所述当前前缀参数与其后相邻的资源参数对应保存至初始参数集中;

[0023] 判断所述当前前缀参数是否在初始URL重写规则集中,如果是,则将所述当前前缀参数与预设重写参数组合为更新前缀参数;如果否,则将所述当前前缀参数与其后相邻的资源参数组合为更新前缀参数;

[0024] 以所述更新前缀参数为当前前缀参数,执行所述将所述当前前缀参数与其后相邻的资源参数对应保存至初始参数集中的步骤,直至当前目标数组的所有字符串都循环完毕;

[0025] 判断所有字符数组是否都循环完毕,如果否,则将任意一个未循环的字符数组作为当前数组,触发执行所述数组循环流程;

[0026] 如果是,则将所述初始参数集作为目标URL集对应的目标参数集。

[0027] 其中,所述依据所述路径参数和非路径参数生成所述目标网站的URL重写规则集,包括:

[0028] 针对各前缀参数,分别判断所述前缀参数下资源参数的数量是否大于预设阈值,如果是,则将所述前缀参数更新至所述初始URL重写规则集中,得到更新后的URL重新规则集,直至所述初始URL重新规则集不再更新;

[0029] 将更新后的URL重写规则集确定为所述目标URL重写规则集。

[0030] 其中,所述方法还包括:

[0031] 依据所述目标网站的URL重写规则集,将待映射URL映射至重写后的URL。

[0032] 其中,所述依据所述目标网站的URL重写规则集,将待映射URL映射至重写后的URL,包括:

[0033] 对所述待映射URL进行规范化,得到规范化后的URL;

[0034] 基于预设分隔符对所述规范化后的URL进行分割,分别得到分割后的字符数组;

[0035] 依据分割后的字符数组中各前缀参数在所述URL重写规则集中的匹配结果,将待映射URL映射至重写后的URL。

[0036] 其中,所述依据各分割后字符数组中各前缀参数在所述URL重写规则集中的匹配结果,将待映射URL映射至重写后的URL,包括:

[0037] 按照从前到后的顺序,获取所述分割后的字符数组中的第一个字符串作为当前前

缀参数；

[0038] 判断所述当前前缀参数是否在所述URL重写规则集中,如果是,则将所述当前前缀参数与预设重写参数组合为更新前缀参数;如果否,则将所述当前前缀参数与其后相邻的资源参数组合为更新前缀参数;

[0039] 以所述更新前缀参数为当前前缀参数,执行所述将所述当前前缀参数与其后相邻的资源参数对应保存至初始参数集中的步骤,直至所述分割后的字符数组中的所有字符串都循环完毕;

[0040] 获取所述更新前缀参数作为重写后的URL。

[0041] 其中,在所述当前前缀参数在所述URL重写规则集中的情况下,还包括:

[0042] 获取所述当前前缀参数对应的资源参数以及所述资源参数的值;

[0043] 将所述资源参数、资源参数的值、查询字符串与所述重写后的URL对应保存。

[0044] 本申请实施例还公开了基于URL重写规则的URL扫描方法,该方法包括:

[0045] 获取预先生成的URL重写规则集,和,目标网站的待扫描的初始URL集;所述URL重写规则集采用如下方式生成:获取目标网站的目标URL集;所述目标网站为:待生成统一资源定位符URL重写规则的网站;获取所述目标URL集中相互对应的前缀参数和资源参数的参数集,并依据所述参数集生成所述目标网站的URL重写规则集;

[0046] 依据所述URL重写规则集对所述初始URL集中的初始URL进行重写,得到重写后的初始URL集;

[0047] 对所述重写后的初始URL集进行去重处理,得到目标URL集;

[0048] 对所述目标URL集中的目标URL进行扫描。

[0049] 本申请实施例还公开了一种URL重写规则的生成装置,该装置包括:

[0050] 获取URL集单元,用于获取目标网站的目标URL集;所述目标网站为:待生成统一资源定位符URL的重写规则的网站;

[0051] 获取单元,用于获取所述目标URL集中相互对应的前缀参数和资源参数的参数集,其中,所述资源参数为所述前缀参数的子路径;

[0052] 生成单元,用于依据所述参数集生成所述目标网站的URL重写规则集。

[0053] 其中,所述获取URL集单元用于:对目标网站的访问日志中的初始URL集进行预处理,得到目标URL集。

[0054] 其中,所述获取URL集单元包括:

[0055] 过滤子单元,用于依据超文本传输协议HTTP状态码,从目标网站的访问日志中的初始URL集中过滤非法URL请求对应的非法URL;

[0056] 规范化子单元,用于针对过滤了非法URL后的初始URL集进行规范化处理,得到规范URL集,所述规范URL集中的规范URL包括:域名、路径和文件名;以及,

[0057] 去重子单元,用于对所述规范URL集进行去重处理,得到目标URL集。

[0058] 其中,所述获取参数单元,包括:

[0059] 分割子单元,用于基于预设分隔符对所述目标URL集中的各目标URL进行分割,分别得到各目标URL对应的字符数组;

[0060] 参数确定子单元,用于依据所述字符数组中各字符串组成所述目标URL的顺序,分别确定所述各目标URL中对应的前缀参数和资源参数,以得到参数集。

- [0061] 其中,所述参数确定子单元,具体用于:
- [0062] 获取任意一个字符数组作为当前数组,执行数组循环流程,所述数组循环流程包括:
- [0063] 按照从前到后的顺序,获取所述当前数组中的第一个字符串作为当前前缀参数;将所述当前前缀参数与其后相邻的资源参数对应保存至初始参数集中;判断所述当前前缀参数是否在初始URL重写规则集中,如果是,则将所述当前前缀参数与预设重写参数组合为更新前缀参数;如果否,则将所述当前前缀参数与其后相邻的资源参数组合为更新前缀参数;以所述更新前缀参数为当前前缀参数,执行所述将所述当前前缀参数与其后相邻的资源参数对应保存至初始参数集中的步骤,直至当前目标数组的所有字符串都循环完毕;
- [0064] 判断所有字符数组是否都循环完毕,如果否,则将任意一个未循环的字符数组作为当前数组,触发执行所述数组循环流程;
- [0065] 如果是,则将所述初始参数集作为目标URL集对应的目标参数集。
- [0066] 其中,所述生成单元包括:
- [0067] 判断子单元,用于针对各前缀参数,分别判断所述前缀参数下资源参数的数量是否大于预设阈值;
- [0068] 更新子单元,用于在所述判断子单元的结果为是的情况下,将所述前缀参数更新至所述初始URL重写规则集中,得到更新后的URL重新规则集,直至所述初始URL重新规则集不再更新;
- [0069] 规则确定子单元,用于将更新后的URL重写规则集确定为所述目标URL重写规则集。
- [0070] 其中,所述装置还包括:
- [0071] 映射单元,用于依据所述目标网站的URL重写规则集,将原始URL映射至重写后的URL。
- [0072] 其中,所述映射单元504可以包括:
- [0073] 规范化子单元,用于对所述待映射URL进行规范化,得到规范化后的URL;
- [0074] 分割子单元,用于基于预设分隔符对所述规范化后的URL进行分割,分别得到分割后的字符数组;
- [0075] 映射子单元,用于依据分割后的字符数组中各前缀参数在所述URL重写规则集中的匹配结果,将待映射URL映射至重写后的URL。
- [0076] 其中,所述映射子单元,具体用于:
- [0077] 按照从前到后的顺序,获取所述分割后的字符数组中的第一个字符串作为当前前缀参数;判断所述当前前缀参数是否在所述URL重写规则集中,如果是,则将所述当前前缀参数与预设重写参数组合为更新前缀参数;如果否,则将所述当前前缀参数与其后相邻的资源参数组合为更新前缀参数;以及,以所述更新前缀参数为当前前缀参数,执行所述将所述当前前缀参数与其后相邻的资源参数对应保存至初始参数集中的步骤,直至所述分割后的字符数组中的所有字符串都循环完毕;获取所述更新前缀参数作为重写后的URL。
- [0078] 其中,所述映射子单元,还用于:
- [0079] 获取所述当前前缀参数对应的资源参数以及所述资源参数的值;以及,并将所述资源参数、资源参数的值、查询字符串与所述重写后的URL对应保存。

[0080] 本申请实施例还公开了一种扫描器,该扫描器包括:

[0081] 获取URL单元,用于获取预先生成的URL重写规则集,和,目标网站的待扫描的初始URL集;所述URL重写规则集采用如下方式生成:如下方式生成:获取目标网站的目标URL集,所述目标网站为:待生成统一资源定位符URL重写规则的网站;获取所述目标URL集中相互对应的前缀参数和资源参数的参数集,并依据所述参数集生成所述目标网站的URL重写规则集;

[0082] 重写单元,用于依据所述URL重写规则集对所述初始URL集中的初始URL进行重写,得到重写后的初始URL集;

[0083] 去重单元,用于对所述重写后的初始URL集进行去重处理,得到目标URL集;

[0084] 扫描单元,用于对所述目标URL集中的目标URL进行扫描。

[0085] 本申请实施例还公开了一种该网络设备包括:处理器、存储器、网络接口和总线系统;

[0086] 所述总线系统,用于将所述网络设备的各个硬件组件耦合在一起;

[0087] 所述网络接口,用于实现所述网络设备与至少一个其它网络设备之间的通信连接;

[0088] 所述存储器,用于存储程序指令和/或数据;

[0089] 所述处理器,用于读取所述存储器中存储的指令和/或数据,执行以下操作:

[0090] 获取目标网站的目标URL集;所述目标网站为:待生成统一资源定位符URL的重写规则的网站;

[0091] 获取所述目标URL集中相互对应的前缀参数和资源参数的参数集,其中,所述资源参数为所述前缀参数的子路径;

[0092] 依据所述参数集生成所述目标网站的URL重写规则集。

[0093] 与现有技术相比,本申请实施例包括以下优点:

[0094] 在本申请实施例中,采用本申请实施例,可以基于一个网站的Web访问日志中的URL集,来对URL集中各URL的前缀参数和资源参数进行分析,从而确定出URL重写参数,并将URL重写参数前的前缀参数作为URL重写规则,最终得到该网站的目标URL重写规则集。因为本申请实施例无需人工分析Web访问日志,所以节省大量人力物力成本,并且还能减少手工配置URL规则时的错误,使得对于大量甚至海量网站的应用场景,也能较快地生成URL重写规则。

[0095] 进一步的,还可以依据URL重写规则集中的URL重写规则,将原始URL重写为与原始URL不同的另一个URL,供扫描器扫描。其中的URL重写参数“\$ {dynamic}”部分将不会被扫描器作为路径实施扫描,从而不仅减少了扫描器的扫描对象,还能保证扫描器不会被攻击者轻易攻击。

[0096] 当然,实施本申请的任一产品并不一定需要同时达到以上所述的所有优点。

附图说明

[0097] 为了更清楚地说明本申请实施例中的技术方案,下面将对实施例描述中所需要使用的附图作简单地介绍,显而易见地,下面描述中的附图仅仅是本申请的一些实施例,对于本领域普通技术人员来讲,在不付出创造性劳动性的前提下,还可以根据这些附图获得其

他的附图。

- [0098] 图1是本申请的URL重写规则的生成方法实施例的流程图；
- [0099] 图2是本申请的方法实施例中预处理URL的结果示意图；
- [0100] 图3是本申请的方法实施例中对URL进行分割得到参数集的流程图；
- [0101] 图4是本申请的方法实施例中对原始URL进行映射的流程图；
- [0102] 图5是本申请的基于URL重写规则的URL扫描方法实施例的流程图；
- [0103] 图6是本申请的URL重写规则的生成装置实施例的结构框图；
- [0104] 图7是本申请的扫描器实施例的结构框图；
- [0105] 图8是根据一示例性实施例示出的网络设备800的结构框图。

具体实施方式

[0106] 下面将结合本申请实施例中的附图,对本申请实施例中的技术方案进行清楚、完整地描述,显然,所描述的实施例仅仅是本申请一部分实施例,而不是全部的实施例。基于本申请中的实施例,本领域普通技术人员在没有做出创造性劳动前提下所获得的所有其他实施例,都属于本申请保护的范围。

[0107] 本申请的主要思想之一,针对获取到一个或多个网站的Web访问日志,根据Web访问日志中各URL生成一个或多个网站的URL重写规则集。具体的,可以先对访问日志中各URL进行预处理,例如,过滤、规范化或者去重等操作,得到预处理后的URL集作为目标URL集,再对目标URL集基于路径进行分割,得到对应该网站的一个数组。再利用该数组内,各前缀参数和该前缀参数下的路径参数的归属关系,例如,某个前缀参数下路径参数的个数是否大于预设阈值,来统计出需要添加至URL重写规则集中的所有前缀参数,得到该网站的URL重写规则集。

[0108] 参考图1,示出了本申请一种URL重写规则的生成方法实施例的流程图,本实施例可以包括以下步骤:

[0109] 步骤101:获取目标网站的目标URL集;所述目标网站为:待生成统一资源定位符URL的重写规则的网站。

[0110] 在本实施例中,目标网站可以是待生成URL重写规则集的Web站点,在实际应用中可以是一个也可以是多个,针对每个目标网站,分别生成各目标网站的URL重写规则集即可。目标网站的访问日志,可以从目标网站对应的服务器的数据库等中查询得到。

[0111] 在本步骤中,可以直接获取该目标网站的访问日志中的各初始URL作为目标URL集,当然,也可以对访问日志中的各初始URL进行预处理,并将预处理后的各初始URL作为目标URL集。其中,预处理过程主要是Web站点的访问日志中的URL进行过滤、规范化和去重等操作。具体的,步骤101可以包括以下步骤A1~步骤A3:

[0112] 步骤A1:依据HTTP状态码,从目标网站的访问日志中的初始URL集中过滤非法URL请求对应的非法URL。

[0113] 其中,HTTP状态码(HTTP Status Code)是用来表示网页服务器HTTP响应状态的3位数字代码。其中,HTTP状态码为200则说明HTTP请求已处理成功,HTTP请求所希望的响应头或数据体将随此响应返回。因此,在本步骤中,需要将HTTP状态码非200的URL过滤掉,以避免该类不存在或出错的URL在生成URL重写规则集的过程中进行干扰。

[0114] 步骤A2:针对过滤了非法URL后的初始URL集进行规范化处理,得到规范URL集,所述规范URL集中的规范URL包括:域名、路径和文件名。

[0115] 在过滤了HTTP状态码非200的非法URL之后,对过滤后的URL进行规范化,从而忽略掉所有URL中的协议、用户名和查询字符串等,从而得到规范URL,在该规范URL中包括域名、路径和文件名。可以理解的是,若端口为80或443,也可以同时忽略端口,只保留域名、路径和文件名。本领域技术人员还可以对域名中所有大写的英文字符转换为小写英文字符,最终得到规范URL。

[0116] 步骤A3:对所述规范URL集中的规范URL进行去重处理,得到目标URL集。

[0117] 在本步骤中,对规范化URL集中的规范URL进行去重,对于相互重复的各URL,只保留一个URL即可,最终得到目标URL集。

[0118] 参考图2所示,为对URL分别进行过滤、规范化和去重时各步骤结果的示意图。图2左上角处存在的HTTP状态码为“404”对应的URL“http://a.com/blog.php”将在步骤A1中进行过滤,得到图2右上角处的五个各URL。再对图2右上角处的五个URL进行规范化处理,删除“http://”等协议,以及“?id=211”的查询字符串,以及,将“A.COM”转化为小写的“a.com”,等等,得到图2右下角处的五个各规范URL。紧接着对图2右下角处的各URL进行去重操作,将两个重复的URL“a.com/blog.php”只保留一个即可,从而得到图2左下角处的四个URL。当然,图2的各URL仅仅是实际应用中的一个具体例子,本领域技术人员不应将其理解为本申请的限定。

[0119] 步骤102:获取所述目标URL集中相互对应的前缀参数和资源参数的参数集,其中,所述资源参数为所述前缀参数的子路径。

[0120] 在本步骤中,针对各预处理后的目标URL集中的目标URL,则需要进行切分,即依据路径来对预处理后的目标URL进行分割得到一个数组,该数组中包括了基于路径分割后的各字符串,第一个字符串是域名。接着,分别依据路径关系,依次将域名、域名与其子路径、域名与其子路径及子路径的子路径等作为前缀参数,来确定该前缀参数对应的资源参数,最终得到前缀参数与资源参数对应保存的参数集。

[0121] 在实际应用中,可以预先设置一个空的URL重写规则集,作为确定参数集的一个依据。则本步骤102首先基于预设分隔符对目标URL集中的各目标URL进行分割,分别得到各目标URL对应的字符数组,其中的预设分隔符可以是“/”,即路径分隔符。再依据各字符数组中各字符串组成所述目标URL的顺序,分别确定各目标URL中对应的前缀参数和资源参数,以得到参数集。

[0122] 具体的,确定对应的前缀参数和资源参数以得到参数集的过程可以包括:

[0123] 步骤B1:获取任意一个字符数组作为当前数组。

[0124] 在本实施例中,一个目标URL对应一个字符数组。则先将任意一个未处理过的字符数组作为当前数组,例如,针对目标URL“a.com/search/winter/2”得到的当前数组为: { ‘a.com’, ‘search’, ‘winter’, ‘2’ }。其中,‘a.com’为该数组的第1个字符串,‘2’为该数组的第4个字符串,该数组共有4个字符串。

[0125] 步骤B2:执行数组循环流程。

[0126] 针对得到的当前数组, { ‘a.com’, ‘search’, ‘winter’, ‘2’ },所述数组循环流程包括步骤21~步骤24:

[0127] 步骤B21:按照从前到后的顺序,获取所述当前数组中的第一个字符串作为当前前缀参数。

[0128] 在本步骤中,将该数组的第1个字符串‘a.com’作为当前前缀参数“prefix”,该当前前缀参数其后相邻的资源参数“resource”为数组中的第二个字符串‘search’。

[0129] 步骤B22:将所述当前前缀参数与其后相邻的资源参数对应保存至初始参数集中。

[0130] 将‘a.com’和‘search’对应保存至初始参数集中,在本实施方式中,初始参数集可以为空。

[0131] 步骤B23:判断所述当前前缀参数是否在初始URL重写规则集中,如果是,则将所述当前前缀参数与预设重写参数组合为更新前缀参数;如果否,则将所述当前前缀参数与其后相邻的资源参数组合为更新前缀参数。

[0132] 接着针对‘a.com’,判断是否在初始URL重写规则集中。其中,该初始URL重写规则集可以为空,随着目标URL分析的越来越多该初始URL重写规则集中的重写规则越来越多,直至不再更新为止。

[0133] 假设初始URL重写规则集为空,则‘a.com’并没有在该初始URL重写规则集中则可以将当前前缀参数‘a.com’与其后相邻的资源参数‘search’组合为更新前缀参数“a.com/search”。而假设‘a.com/search’在初始URL重写规则集中,则说明‘a.com/search’其后相邻的资源参数‘winter’为URL重写参数,则在这种情况下,将当前前缀参数‘a.com/search’与预设重写参数(例如“dynamic”)组合为更新前缀参数“a.com/search/{dynamic}”。

[0134] 步骤B24:以所述更新前缀参数为当前前缀参数,执行所述将所述当前前缀参数与其后相邻的资源参数对应保存至初始参数集中的步骤,直至当前目标数组的所有字符串都循环完毕。

[0135] 在本步骤中,再接着以步骤B23中得到的更新前缀参数,即,“a.com/search”或“a.com/search/{dynamic}”作为当前前缀参数,将“a.com/search”或“a.com/search/{dynamic}”与其后相邻的资源参数对应保存至初始参数集。例如,将前缀参数“a.com/search”与其资源参数“winter”对应保存至初始参数集。再例如,将前缀参数“a.com/search/{dynamic}”与其资源参数“2”对应保存至初始参数集,直至该目标数据中的所有字符串都循环完毕。

[0136] 步骤B3:判断所有字符数组是否都循环完毕,如果否,则将任意一个未循环的字符数组作为当前数组,触发步骤B2执行所述数组循环流程;如果是,则进入步骤B4。

[0137] 接着判断各目标网站对应的所有字符数组是否都循环完毕,如果否,以任意一个未循环过的字符数组作为当前数组,触发步骤B2执行该数组循环流程。

[0138] 步骤B4:获取所述初始参数集作为目标URL对应的目标参数集。

[0139] 在本步骤中,则可以将最终不再更新的初始参数集中的各组前缀参数和资源参数对应输出,作为更新URL重写规则集的依据。

[0140] 步骤103:依据所述参数集生成所述目标网站的URL重写规则集。

[0141] 在本步骤中,可以依据参数集中各前缀参数下对应了多少个资源参数,来确定该前缀参数是否应该添加至URL重写规则集。

[0142] 在实际应用中,可以同一个前缀参数下的不同的资源参数的数量,因为对于一个正常的URL路径而言,一个前缀参数下的资源参数的个数应该是有限的,对于需要作为URL

重写规则的前缀参数,其对应的资源参数相当于用户输入的任意参数,数量较大。所以,可预先设定一个阈值,若资源参数的数量大于该阈值,则将该前缀参数的值加入至URL重写规则集中,最后得到更新后的URL重写规则集,直至初始URL重新规则集不再更新,即可将更新后的URL重写规则集确定为所述目标URL重写规则集。

[0143] 其中,URL重写规则用于通过一个URL表示一条重写规则的方式,来表示哪些参数为重写参数,URL重写规则集中的URL后相邻的路径参数即为URL重写参数。具体的,URL重写规则集可以包括多条URL重写规则,其中,单条URL重写规则的格式为一URL,例如“a.com/s”,该URL重写规则的含义为:“a.com/s”后的一个路径参数(例如a.com/s/a1/a2中的“a1”)为URL重写参数。由于访问日志中的一个URL中可能有多个路径存在URL重写规则,例如“a.com/search/test/2”会被映射至“a.com/search.php?keyword=test&page=2”,因此需要对访问日志中的各URL进行迭代,直至URL重写规则集中的各条URL都不发生变化为止。

[0144] 可见,采用本申请实施例,可以基于一个网站的Web访问日志中的URL集,来对URL集中各URL的前缀参数和资源参数进行分析,从而确定出URL重写参数,并将URL重写参数前的前缀参数作为URL重写规则,从而得到该网站的目标URL重写规则集。因为本申请实施例无需人工分析Web访问日志,所以节省大量人力物力成本,并且还能减少手工配置URL规则时的错误,使得对于大量甚至海量网站的应用场景,也能较快地生成URL重写规则。

[0145] 在实际应用中,基于得到的URL重写规则集,还可以对原始的WEB日志进行处理,将原始URL映射至重写后的URL路径,因为URL重写规则可以表示出URL重写参数,因此,就可以将包括URL重写参数的各个URL作为同一个URL供扫描器调用,这样就减少了扫描器调用的URL个数。因此,在步骤103之后,还可以包括:

[0146] 步骤104:依据所述目标网站的URL重写规则集,将原始URL映射至重写后的URL。

[0147] 本步骤中可以基于URL重写规则集,将原始URL映射至重写后的URL,并提取查询字符串以及URL重写规则中的重写参数,作为扫描器的输入源。具体的,步骤104的实现过程可以包括步骤C1~步骤C5:

[0148] 步骤C1:对所述待映射URL进行规范化,得到规范化后的URL,并存储待映射URL的查询字符串。

[0149] 在本步骤中,需要对待映射的原始URL进行规范化,具体规范化的步骤可以参考步骤A2的描述,在此不再赘述。例如,原始URL为“a.com/search/winter/2?a=b”,则“?a=b”即为查询字符串,将其进行存储,本步骤中规范化后的URL为:a.com/search/winter/2。

[0150] 步骤C2:基于预设分隔符对所述规范化后的URL进行分割,分别得到分割后的字符数组。

[0151] 在本步骤中,也是基于预设的路径分隔符“/”来对规范后的URL进行分割,得到字符数组: { 'a.com', 'search', 'winter', '2' }。

[0152] 步骤C3:依据分割后的字符数组中各前缀参数在所述URL重写规则集中的匹配结果,将待映射URL映射至重写后的URL。

[0153] 再依据该字符数组中的各前缀参数,例如,“a.com”,“a.com/search”,“a.com/search/winter”,“a.com/search/winter/2”等,在URL重写规则集中的匹配结果,来将待映射URL映射至重写后的URL。

[0154] 具体的,步骤C3的映射过程可以包括步骤C31~步骤C35:

[0155] 步骤C31:按照从前到后的顺序,获取所述分割后的字符数组中的第一个字符串作为当前前缀参数。

[0156] 仍然获取字符数组中的“a.com”作为当前前缀参数。

[0157] 步骤C32:判断所述当前前缀参数是否在所述URL重写规则集中,如果是,则进入步骤C33,如果否,则进入步骤C34。

[0158] 如果当前前缀参数“a.com/search”在URL重写规则集中,则进入步骤C33,如果当前前缀参数为“a.com”不在URL重写规则集中,则进入步骤C34。

[0159] 步骤C33:将所述当前前缀参数与预设重写参数组合为更新前缀参数,获取所述当前前缀参数对应的资源参数以及所述资源参数的值,并将所述资源参数、资源参数的值、查询字符串与所述重写后的URL对应保存,进入步骤C35。

[0160] 在本步骤中,将当前前缀参数“a.com/search”与预设重写参数“\${dynamic}”组合为更新前缀参数,得到“a.com/search/\${dynamic}”。此外,还需要获取“a.com/search”的资源参数的值“winter”,以及,查询字符串“?a=b”。

[0161] 步骤C34:将所述当前前缀参数与其后相邻的资源参数组合为更新前缀参数,进入步骤C35。

[0162] 在本步骤中则将当前前缀参数“a.com/search”与其资源参数“winter”组合为更新前缀参数,得到“a.com/search/winter”。

[0163] 步骤C4:以所述更新前缀参数为当前前缀参数,执行所述将所述当前前缀参数与其后相邻的资源参数对应保存至初始参数集中的步骤,直至所述分割后的字符数组中的所有字符串都循环完毕。

[0164] 再将步骤C33中的“a.com/search/\${dynamic}”,或者C34中的“a.com/search/winter”,作为当前前缀参数,返回步骤33判断是否存在于URL重写规则集中,直至字符数组中的所有字符串都循环完毕,获取此时的URL重写参数,例如“winter”,或者“2”,以及查询字符串“?a=b”。

[0165] 步骤C5:获取所述更新前缀参数作为重写后的URL。

[0166] 最后获取不再更新的更新前缀参数作为重写后的URL。假设URL重写规则集中包括了“a.com/search”和“a.com/search/\${dynamic}”,则本步骤中获取到的重写后的URL包括:“a.com/search/\${dynamic}/\${dynamic}”;其中,对应的资源参数为dynamic_1时,资源参数值为winter;对应的资源参数为dynamic_2时,资源参数值为2;对应的资源参数为a时,资源参数值为b。

[0167] 可见,在本申请实施例中,还依据URL重写规则集中的URL重写规则,将多个包括相同的URL重写参数的原始URL重写为一个目标URL,供扫描器扫描。其中的URL重写参数“\${dynamic}”部分将不会被扫描器作为路径实施扫描,从而不仅减少了扫描器的扫描对象,还能保证扫描器不会被攻击者轻易攻击。

[0168] 参考图3,示出了本申请方法实施例中访问日志中的URL进行分割并得到前缀参数和资源参数的流程图,本流程可以包括以下步骤:

[0169] 步骤301:获取目标URL集。

[0170] 步骤302:对目标URL集中的目标URL进行分割,得到一个数组。

[0171] 步骤303:初始化参数,n等于1,取前缀参数prefix为数组中的第0个元素,即域名。

[0172] 在本步骤中,仍以得到的数组url_array为: { ‘a.com’, ‘search’, ‘winter’, ‘2’ } 为例。初始化参数,n=1,此时prefix为“a.com”。

[0173] 步骤304:触发步骤304~步骤307的循环;取resource为数组的第1个元素“search”。

[0174] 步骤305:存储此时的prefix和对应的resource的值。

[0175] 即将“a.com”和“search”对应存储。

[0176] 步骤306:判断此时的prefix的值是否在URL重写规则集中,如果在,令prefix为prefix+“\${dynamic}”,否则令prefix为prefix+resource;

[0177] 判断“a.com”是否在URL重写规则集中;假设对应的URL重写规则集包括由URL重写规则为:“a.com/search”;“a.com”则不在该URL重写规则集中,则prefix为prefix+resource,即a.com/search。

[0178] 步骤307:令n=n+1,判断n是否小于数组的长度,如果是,则继续步骤304~步骤307循环的第一步,否则进入步骤308。

[0179] 再令n=2,n小于数组的长度4,则继续取prefix的resource为:“winter”,进入步骤305进行存储,并依次执行步骤306,和307,直至n等于数组的长度。

[0180] 步骤308:输出在305步骤中存储的所有prefix及其对应的resource的值。

[0181] 例如,在本例子中,输出结果可以如表1所示:

[0182] 表1

输出	
<u>prefix</u>	<u>resource</u>
a.com	search
a.com/search	winter
a.com/search/\${dynamic}	2

[0184] 参考图4,示出了本申请方法实施例依据URL规则集映射原始URL的例子流程图,本实施例可以包括以下步骤:

[0185] 步骤401:将待映射URL进行规范化,并存储查询字符串中的参数名和参数值。

[0186] 假设待映射URL为:“a.com/search/winter/2?a=b”则本步骤中存储的查询字符串的参数名为a,参数值为b。

[0187] 步骤402:将规范化处理后URL基于路径分隔符“/”进行分割,得到一个数组。

[0188] 对规范后的URL“a.com/search/winter/2”基于分隔符“/”进行分割,得到一个数组。

[0189] 步骤403:初始化参数,n=1,取前缀参数prefix为数组第0个元素,即域名。

[0190] 本步骤中,初始化参数,n=1,此时prefix为“a.com”。

[0191] 步骤404:触发步骤404~步骤406循环体,取resource为数组的第1个元素“search”。

[0192] 步骤405:判断此时的prefix的值是否在URL重写规则集中,如果在,令prefix为

prefix+”\${dynamic}”,同时存储resource的值,否则令prefix为prefix+resource。

[0193] 步骤406:令 $n=n+1$,判断 n 是否小于数组的长度,如果是,则继续循环体的第一步即触发步骤404,否则进入步骤407。

[0194] 步骤407:输出此时的prefix为重写后的URL,同时输出步骤401中查询字符串的参数名和参数值,以及循环体步骤405中存储的所有资源参数名和资源参数的值。

[0195] 具体的,假设URL重写规则集中包括的重写规则有:“a.com/search”,和“a.com/search/\${dynamic}”,则步骤407的一种可能的输出可以参考表2所示。

[0196] 表2

输出		
重写 URL	key	value
a.com/search/\${dynamic}/\${dynamic}	dynamic_1	winter
a.com/search/\${dynamic}/\${dynamic}	dynamic_2	2
a.com/search/\${dynamic}/\${dynamic}	a	b

[0198] 参见图5,本申请还提供了一种基于URL重写规则的URL扫描方法实施例的流程图,本实施例可以包括以下步骤:

[0199] 步骤501:获取预先生成的URL重写规则集,和,目标网站的待扫描的初始URL集。

[0200] 在实际应用中,基于采用图1所示的方法得到URL重写规则集后,还可以对WEB日志中的URL集依据该URL重写规则集进行重写。具体的,扫描器可以预先保存采用图1所示的方法得到的URL重写规则集,例如保存至存储器中,还可以获取到目标网站的所有URL作为待扫描的初始URL集。因为URL重写规则可以表示出URL重写参数,因此,就可以将包括同一个URL重写参数的很多个初始URL后续经过重写和去重后,作为一个URL供扫描器调用,这样就减少了扫描器扫描的URL个数。

[0201] 步骤502:依据所述URL重写规则集对所述初始URL集中的初始URL进行重写,得到重写后的初始URL集。

[0202] 在获取到URL重写规则集和初始URL规则集后,利用URL重写规则集对初始URL规则集中的各个初始URL进行重写。具体的重写过程可以参考步骤C1~步骤C5的详细介绍,在此不再赘述。

[0203] 步骤503:对所述重写后的初始URL集进行去重处理,得到目标URL集。

[0204] 对重写后的初始URL进行去重处理,得到各不相同的、多条目标URL。因为对于初始URL来讲,很多个初始URL可能包括了相同的URL重写参数,则意味着这多个初始URL其实都指向同一个页面,则这多个初始URL重写后的地址是相同的,因此,针对这多个重写后的初始URL只保留一个即可。以此类推,可以得到比初始URL集中的个数小很多的目标URL集。

[0205] 步骤504:对所述目标URL集中的目标URL进行扫描。

[0206] 接着扫描器再对目标URL集中的各条目标URL进行扫描,因为目标URL的个数要比初始URL的个数小很多,因此,可以使得扫描器的扫描效率更高。

[0207] 对于前述的方法实施例,为了简单描述,故将其都表述为一系列的动作组合,但是

本领域技术人员应该知悉,本申请并不受所描述的动作顺序的限制,因为依据本申请,某些步骤可以采用其他顺序或者同时进行。其次,本领域技术人员也应该知悉,说明书中所描述的实施例均属于优选实施例,所涉及的动作和模块并不一定是本申请所必须的。

[0208] 与上述本申请一种URL重写规则的生成方法实施例所提供的方法相对应,参见图6,本申请还提供了一种URL重写规则的生成装置实施例,在本实施例中,该装置可以包括:

[0209] 获取URL集单元601,用于获取目标网站的目标URL集;所述目标网站为:待生成统一资源定位符URL的重写规则的网站。

[0210] 其中,所述获取URL集单元601可以用于:对目标网站的访问日志中的初始URL集进行预处理,得到目标URL集。

[0211] 其中,所述获取URL集单元601执行预处理时,可以包括:

[0212] 过滤子单元,用于依据超文本传输协议HTTP状态码,从目标网站的访问日志中的初始URL集中过滤非法URL请求对应的非法URL;

[0213] 规范化子单元,用于针对过滤了非法URL后的初始URL集进行规范化处理,得到规范URL集,所述规范URL集中的规范URL包括:域名、路径和文件名;以及,

[0214] 去重子单元,用于对所述规范URL集进行去重处理,得到目标URL。

[0215] 获取参数单元602,用于获取所述目标URL集中相互对应的前缀参数和资源参数的参数集,其中,所述资源参数为所述前缀参数的子路径。

[0216] 其中,所述获取参数单元602可以包括:

[0217] 分割子单元,用于基于预设分隔符对所述目标URL集中的各目标URL进行分割,分别得到各目标URL对应的字符数组;以及,

[0218] 参数确定子单元,用于依据所述字符数组中各字符串组成所述目标URL的顺序,分别确定所述各目标URL中对应的前缀参数和资源参数,以得到参数集。

[0219] 其中,所述参数确定子单元,具体用于:

[0220] 获取任意一个字符数组作为当前数组,执行数组循环流程,所述数组循环流程包括:

[0221] 按照从前到后的顺序,获取所述当前数组中的第一个字符串作为当前前缀参数;

[0222] 将所述当前前缀参数与其后相邻的资源参数对应保存至初始参数集中;

[0223] 判断所述当前前缀参数是否在初始URL重写规则集中,如果是,则将所述当前前缀参数与预设重写参数组合为更新前缀参数;如果否,则将所述当前前缀参数与其后相邻的资源参数组合为更新前缀参数;

[0224] 以所述更新前缀参数为当前前缀参数,执行所述将所述当前前缀参数与其后相邻的资源参数对应保存至初始参数集中的步骤,直至当前目标数组的所有字符串都循环完毕;

[0225] 判断所有字符数组是否都循环完毕,如果否,则将任意一个未循环的字符数组作为当前数组,触发执行所述数组循环流程;

[0226] 如果是,则将所述初始参数集作为目标URL集对应的目标参数集。

[0227] 生成单元603,用于依据所述参数集生成所述目标网站的URL重写规则集。

[0228] 其中,所述生成单元603具体可以包括:

[0229] 判断子单元,用于针对各前缀参数,分别判断所述前缀参数下资源参数的数量是

否大于预设阈值；

[0230] 更新子单元,用于在所述判断子单元的结果为是的情况下,将所述前缀参数更新至所述初始URL重写规则集中,得到更新后的URL重新规则集,直至所述初始URL重新规则集不再更新；

[0231] 规则确定子单元,用于将更新后的URL重写规则集确定为所述目标URL重写规则集。

[0232] 采用本申请实施例的装置,可以基于一个网站的Web访问日志中的URL,来对URL中的前缀参数和资源参数进行分析,从而确定出URL重写参数,并将URL重写参数前的前缀参数生成URL重写规则,最终得到目标URL重写规则集。因为无需人工分析Web访问日志,所以节省大量人力物力成本,并且还能减少手工配置URL规则时的错误,使得对于大量甚至海量网站的应用场景,也能较快地生成URL重写规则。

[0233] 其中,该装置还可以包括：

[0234] 映射单元604,用于依据所述目标网站的URL重写规则集,将原始URL映射至重写后的URL。

[0235] 其中,所述映射单元604可以包括：

[0236] 规范化子单元,用于对所述待映射URL进行规范化,得到规范化后的URL；分割子单元,用于基于预设分隔符对所述规范化后的URL进行分割,分别得到分割后的字符数组；以及,映射子单元,用于依据分割后的字符数组中各前缀参数在所述URL重写规则集中的匹配结果,将待映射URL映射至重写后的URL。

[0237] 其中,所述映射子单元,具体可以用于：

[0238] 按照从前到后的顺序,获取所述分割后的字符数组中的第一个字符串作为当前前缀参数；判断所述当前前缀参数是否在所述URL重写规则集中,如果是,则将所述当前前缀参数与预设重写参数组合为更新前缀参数；如果否,则将所述当前前缀参数与其后相邻的资源参数组合为更新前缀参数；以及,以所述更新前缀参数为当前前缀参数,执行所述将所述当前前缀参数与其后相邻的资源参数对应保存至初始参数集中的步骤,直至所述分割后的字符数组中的所有字符串都循环完毕；获取所述更新前缀参数作为重写后的URL。

[0239] 其中,所述映射子单元,还可以用于：获取所述当前前缀参数对应的资源参数以及所述资源参数的值；以及,并将所述资源参数、资源参数的值、查询字符串与所述重写后的URL对应保存。

[0240] 可见,所述映射单元604还依据URL重写规则集中的URL重写规则,将原始URL重写为另一个URL,供扫描器扫描。其中的URL重写参数“\$ {dynamic}”部分将不会被扫描器作为路径实施扫描,从而不仅减少了扫描器的扫描对象,还能保证扫描器不会被攻击者轻易攻击。

[0241] 与图5提供的扫描方法相对应,参考图7,本申请还提供了一种扫描器,该扫描器可以包括：

[0242] 获取URL单元701,用于获取预先生成的URL重写规则集,和,目标网站的待扫描的初始URL集；所述URL重写规则集如下方式生成：获取目标网站的目标URL集,所述目标网站为：待生成统一资源定位符URL重写规则的网站；获取所述目标URL集中相互对应的前缀参数和资源参数的参数集,并依据所述参数集生成所述目标网站的URL重写规则集。

[0243] 重写单元702,用于依据所述URL重写规则集对所述初始URL集中的初始URL进行重写,得到重写后的初始URL集。

[0244] 去重单元703,用于对所述重写后的初始URL集进行去重处理,得到目标URL集。

[0245] 扫描单元704,用于对所述目标URL集中的目标URL进行扫描。

[0246] 因为本实施例中目标URL的个数要比初始URL的个数小很多,因此,所以本实施例的扫描器的扫描效率更高。

[0247] 图8为本发明实施例中一种网络设备800的硬件结构示意图。网络设备800可用于实现8。即网络设备800可用于执行上述实施例中提供的方法。在本实施例中,网络设备800包括:处理器801、存储器802、网络接口803和总线系统804。

[0248] 所述总线系统804,用于将网络设备800的各个硬件组件耦合在一起。

[0249] 所述网络接口803,用于实现网络设备800与至少一个其它网络设备之间的通信连接,可以使用互联网,广域网,本地网,城域网等方式。

[0250] 所述存储器802,用于存储程序指令和/或数据。

[0251] 所述处理器801,用于读取存储器802中存储的指令和/或数据,执行以下操作:

[0252] 获取预先生成的URL重写规则集,和,目标网站的待扫描的初始URL集;所述URL重写规则集采用前述的URL重写规则的生成方法生成;

[0253] 依据所述URL重写规则集对所述初始URL集中的初始URL进行重写,得到重写后的初始URL;

[0254] 对所述重写后的初始URL集进行去重处理,得到目标URL集;

[0255] 对所述目标URL集中的目标URL进行扫描。

[0256] 需要说明的是,本说明书中的各个实施例均采用递进的方式描述,每个实施例重点说明的都是与其他实施例的不同之处,各个实施例之间相同相似的部分互相参见即可。对于装置类实施例而言,由于其与方法实施例基本相似,所以描述的比较简单,相关之处参见方法实施例的部分说明即可。

[0257] 最后,还需要说明的是,在本文中,诸如第一和第二等之类的关系术语仅仅用来将一个实体或者操作与另一个实体或操作区分开来,而不一定要求或者暗示这些实体或操作之间存在任何这种实际的关系或者顺序。而且,术语“包括”、“包含”或者其任何其他变体意在涵盖非排他性的包含,从而使得包括一系列要素的过程、方法、物品或者设备不仅包括那些要素,而且还包括没有明确列出的其他要素,或者是还包括为这种过程、方法、物品或者设备所固有的要素。在没有更多限制的情况下,由语句“包括一个……”限定的要素,并不排除在包括所述要素的过程、方法、物品或者设备中还存在另外的相同要素。

[0258] 以上对本申请所提供的URL重写规则的生成方法及装置、扫描方法及装置进行了详细介绍,本文中应用了具体个例对本申请的原理及实施方式进行了阐述,以上实施例的说明只是用于帮助理解本申请的方法及其核心思想;同时,对于本领域的一般技术人员,依据本申请的思想,在具体实施方式及应用范围上均会有改变之处,综上所述,本说明书内容不应理解为对本申请的限制。



图1

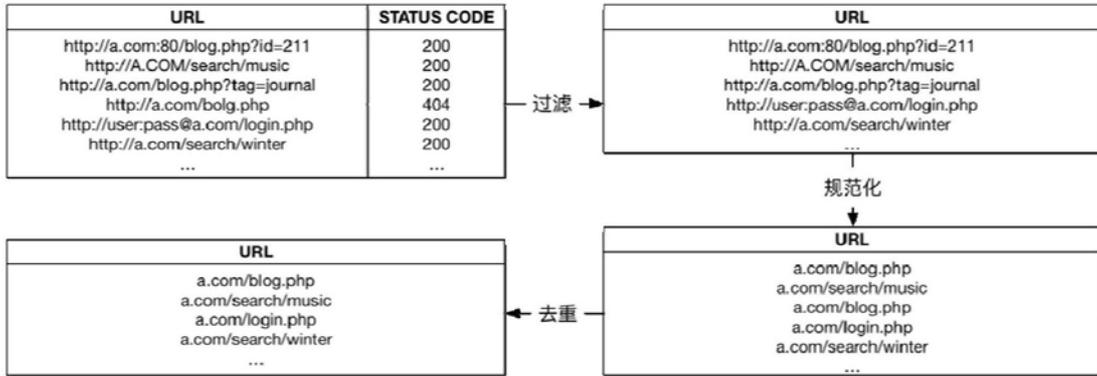


图2

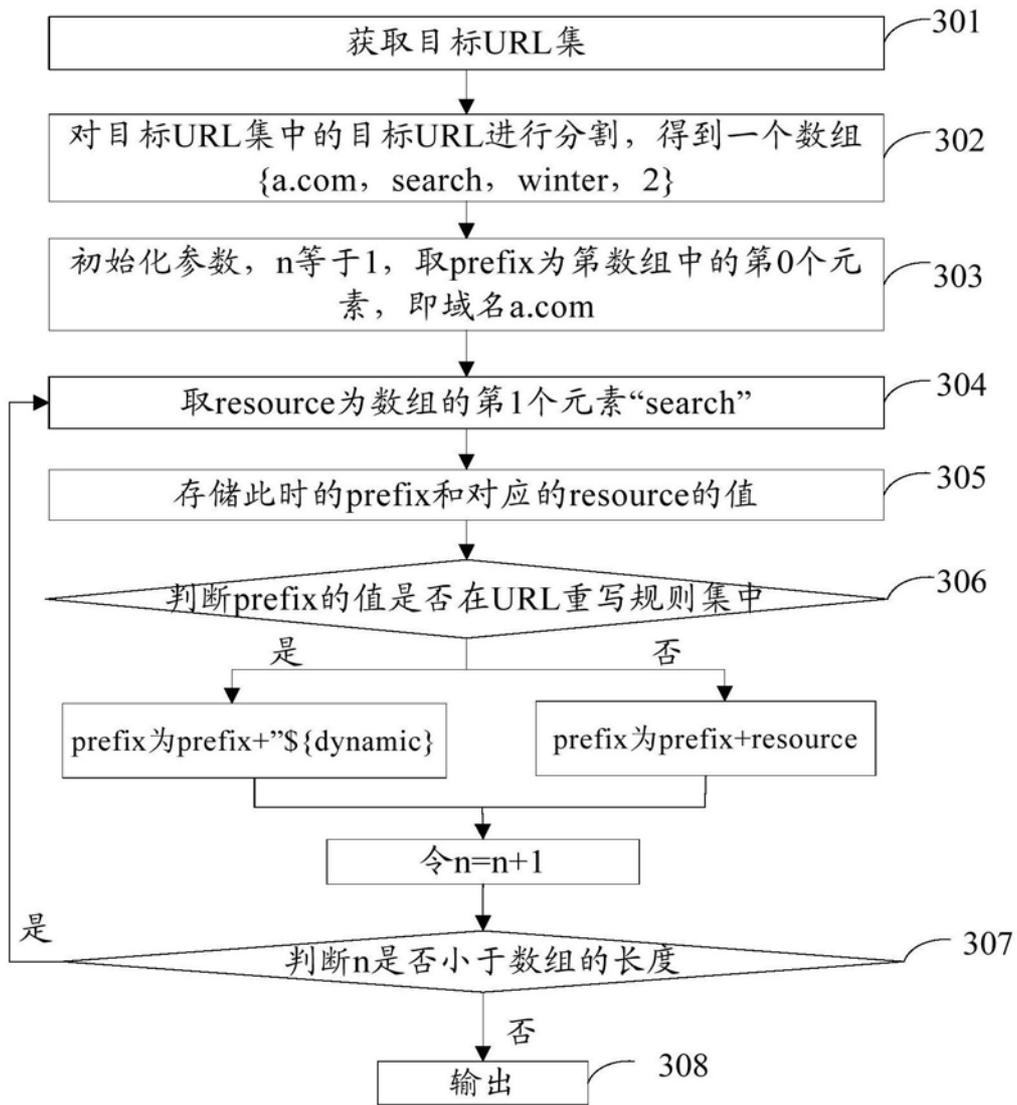


图3

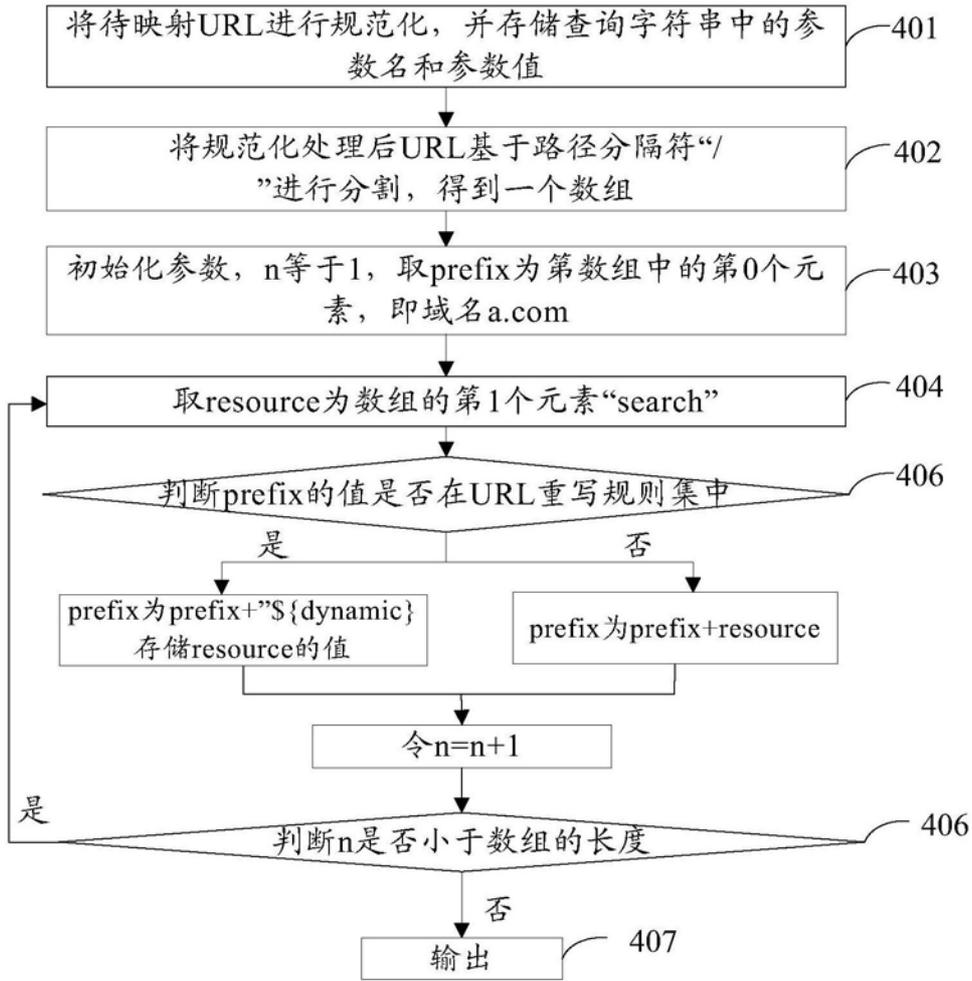


图4

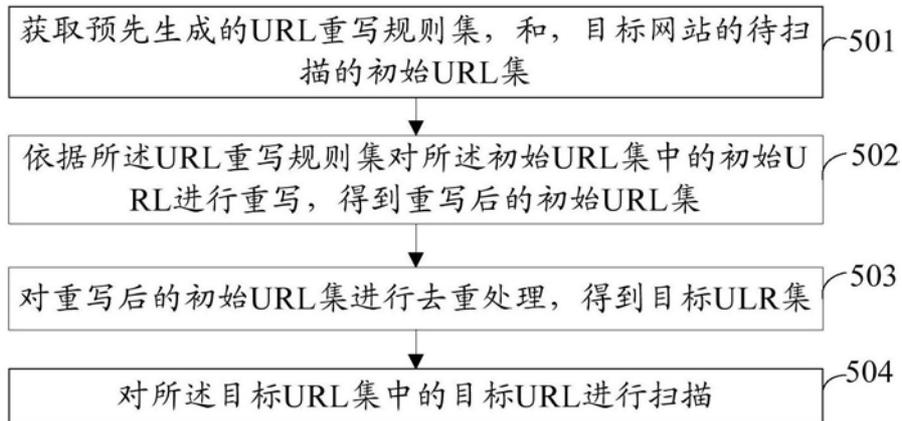


图5



图6



图7

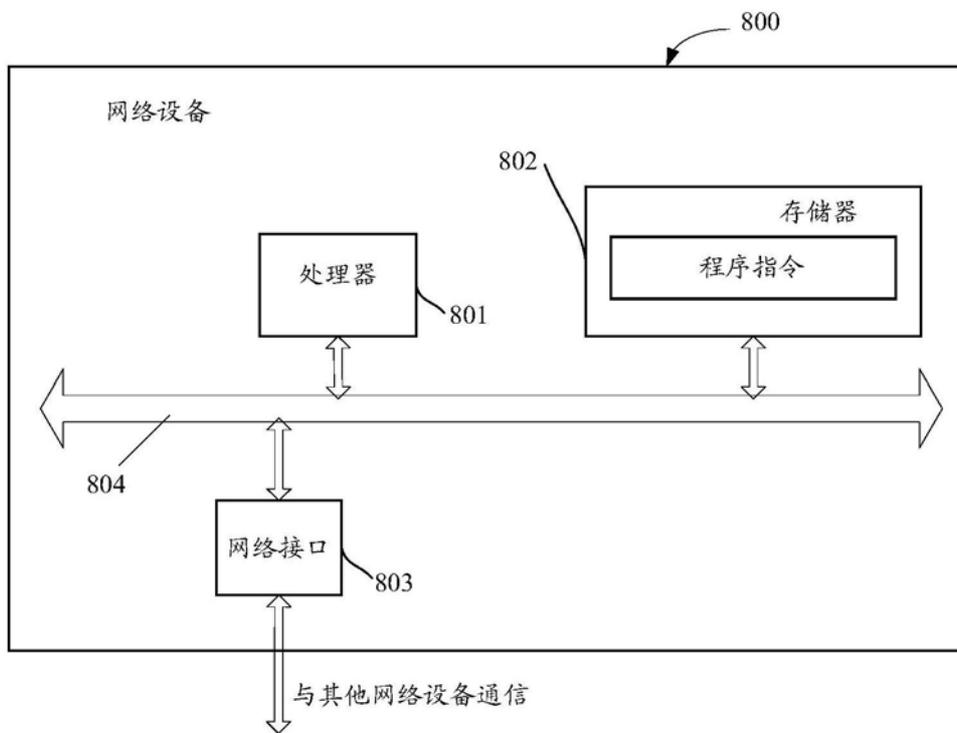


图8