



(12)发明专利

(10)授权公告号 CN 104268289 B

(45)授权公告日 2017.12.12

(21)申请号 201410564162.8

(56)对比文件

(22)申请日 2014.10.21

CN 101242336 A, 2008.08.13,

(65)同一申请的已公布的文献号

CN 101854404 A, 2010.10.06,

申请公布号 CN 104268289 A

CN 103546830 A, 2014.01.29,

(43)申请公布日 2015.01.07

审查员 刘可

(73)专利权人 中国建设银行股份有限公司

地址 100033 北京市西城区金融大街25号

(72)发明人 胡建波 何明杰 张兴强 王海潮
章梦 邱群业

(74)专利代理机构 北京集佳知识产权代理有限
公司 11227

代理人 王宝筠

(51)Int.Cl.

G06F 17/30(2006.01)

G06Q 30/02(2012.01)

权利要求书3页 说明书15页 附图7页

(54)发明名称

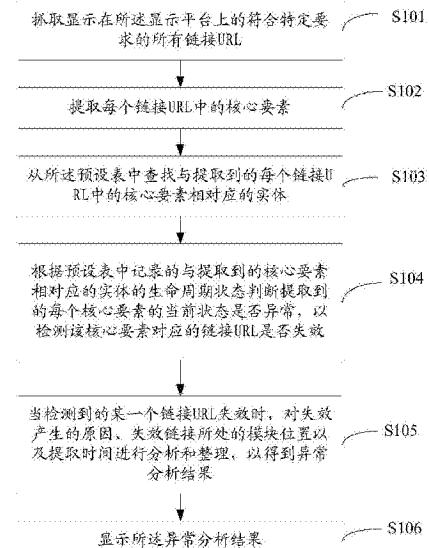
链接URL的失效检测方法和装置

(57)摘要

本发明提供了一种链接URL的失效检测方法及装置，所述失效检测方法包括：步骤A、抓取显示在所述显示平台上的符合特定要求的所有链接URL；步骤B、提取每个链接URL中的核心要素；步骤C、从所述预设表中查找与提取到的每个链接URL中的核心要素相对应的实体；步骤D、根据所述预设表中记录的与提取到的核心要素相对应的实体的生命周期状态判断提取的每个核心要素的当前状态是否异常，以检测该核心要素对应的链接URL是否失效。该失效检测方法可以通过计算机自动实现，相较于现有技术中通过人工进行检测的方法，提高了检测效率，同时也使检测效果的误差大大减小。

B

CN 104268289



1. 一种链接URL的失效检测方法,所述链接URL显示在显示平台上,每个所述链接URL至少包括一个核心要素,所述核心要素包括参数名及参数值;其特征在于,所述显示平台上设置有预设表,所述预设表中记录有与每个核心要素相对应的实体的生命周期状态,所述失效检测方法包括:

步骤A、抓取显示在所述显示平台上的符合特定要求的所有链接URL;

步骤B、提取每个链接URL中的核心要素;

步骤C、从所述预设表中查找与提取到的每个链接URL中的核心要素相对应的实体;

步骤D、根据所述预设表中记录的与提取到的核心要素相对应的实体的生命周期状态判断提取的每个核心要素的当前状态是否异常,以检测该核心要素对应的链接URL是否失效。

2. 根据权利要求1所述的失效检测方法,其特征在于,各个所述链接URL具有统一的数据结构格式,所述预设表包括多个第一表格,一个所述第一表格中记录有一个类别的各个参数值对应的实体的生命周期状态;每个所述类别中的各个链接URL具有相同的参数名;

所述步骤A之后,还包括:

步骤E、根据每个链接URL中的参数名对各个链接URL进行分类,将各个链接URL归属到相应的类别;在每个类别中包含有各个核心要素的记录;

所述步骤B具体为:

步骤B'、从每个所述类别中的每个记录提取参数值;

所述步骤C具体包括:

步骤C1、从显示平台上查找与每个类别分别对应的第一表格;

步骤C2、从所述第一表格中查找与类别中的每个参数值相对应的主键值,根据所述主键值查找与所述主键值相对应的实体;

所述步骤D具体包括:

步骤D'、根据类别对应的异常判断规则和所述第一表格中记录的每个实体的生命周期状态判断类别中的每个参数值对应的每个记录的当前状态是否异常,以检测该记录对应的链接URL是否失效。

3. 根据权利要求2所述的失效检测方法,其特征在于,所述步骤E之后还包括:

步骤F、当归属于同一个类别中的链接URL中有N个具有相同的参数值时,从该类别中移除(N-1)个具有该相同参数值的记录,使得在类别中只保留一个该相同参数值对应的记录,其中,N≥2,N为整数;

执行所述步骤F的同时或之后还包括:

步骤G、建立被移除的记录与其包含的参数值的映射关系,以根据参数值能够找到包含该参数值的所有链接URL;

所述步骤D'之后,还包括:

步骤H、当提取的某一类别下的某一个参数值对应的记录的当前状态异常时,根据所述映射关系查找包含该参数值的所有被移除的链接URL。

4. 根据权利要求1-3任一项所述的失效检测方法,其特征在于,检测完链接URL是否失效之后还包括:

步骤I、当提取到的某一参数名的某一个参数值的当前状态异常时,对异常产生的原

因、失效链接所处的模块位置以及提取时间进行分析和整理,以得到异常分析结果;

步骤J、显示所述异常分析结果。

5.根据权利要求2所述的失效检测方法,其特征在于,所述步骤B'具体包括:

确定链接URL的类别;

检验链接URL中是否存在与该类别相对应的参数名:如果是,检验所述参数名是否位于所述链接URL的参数域,如果是,提取参数值,所述参数值以链接URL中的“=”开始,且“=”之前为所述参数名;

校验所述参数值是否满足预设条件,如果是,记录该提取的参数值。

6.根据权利要求3所述的失效检测方法,其特征在于,所述映射关系包括链接类别、参数名、参数值以及链接URL的对应关系。

7.根据权利要求2所述的失效检测方法,其特征在于,当一个链接URL中包括多个参数名时,所述步骤E具体包括:

根据该链接URL中的多个参数名将该链接URL分别归属到与其参数名相对应的多个类别中。

8.一种链接URL的失效检测装置,所述链接URL显示在显示平台上,每个所述链接URL至少包括一个核心要素,所述核心要素包括参数名及参数值;其特征在于,所述显示平台上设置有预设表,所述预设表中记录有与每个核心要素相对应的实体的生命周期状态,所述失效检测装置包括:

抓取单元,用于抓取显示在所述显示平台上的符合特定要求的所有链接URL;

提取单元,用于提取每个链接URL中的核心要素;

第一查找单元,用于从所述预设表中查找与提取到的每个链接URL中的核心要素相对应的实体;

判断单元,用于根据所述预设表中记录的与提取到的核心要素相对应的实体的生命周期状态判断提取的每个核心要素的当前状态是否异常,以检测该核心要素对应的链接URL是否失效。

9.根据权利要求8所述的失效检测装置,其特征在于,所有各个所述链接URL具有统一的数据结构格式,所述预设表包括多个第一表格,一个所述第一表格中记录有一个类别的各个参数值对应的实体的生命周期状态;每个所述类别中的所有各个链接URL具有相同的参数名;

所述失效检测装置还包括:

分类单元,用于在接收到所述抓取单元抓取的链接URL后,根据每个链接URL中的参数名对所有各个链接URL进行分类,将所有各个链接URL归属到相应的类别;在每个类别中包含有各个核心要素的记录;

所述提取单元包括从每个所述类别中的每个记录中提取参数值的子单元;

所述第一查找单元具体包括:

第一查找子单元,用于从显示平台上查找与每个类别分别对应的第一表格;

第二查找子单元,用于从所述第一表格中查找与类别中的每个参数值相对应的主键值,根据所述主键值查找与所述主键值相对应的实体;

所述判断单元具体包括:

判断子单元，所述判断子单元用于根据类别对应的异常判断规则和所述第一表格中记录的每个实体的生命周期状态判断类别中的每个参数值对应的每个记录的当前状态是否异常，以检测该记录对应的链接URL是否失效。

10. 根据权利要求9所述的失效检测装置，其特征在于，所述失效检测装置还包括：移除单元，用于当归属于同一个类别中的链接URL中有N个具有相同的参数值时，从该类别中移除(N-1)个具有该相同参数值的记录，使得在类别中只保留一个该相同参数值对应的记录，其中， $N \geq 2$ ，N为整数。

11. 根据权利要求10所述的失效检测装置，其特征在于，还包括：

建立映射关系单元，用于在接收到所述移除单元移除记录的信号的同时或之后，建立被移除的记录与其包含的参数值的映射关系，以根据参数值能够找到包含该参数值的所有链接URL；

第二查找单元，用于当所述判断子单元判断的提取的某一类别下的某一个参数值对应的记录的当前状态异常时，根据所述映射关系查找包含该参数值的所有被移除的链接URL。

12. 根据权利要求8所述的失效检测装置，其特征在于，还包括：

异常结果分析单元，用于当提取到的某一参数名的某一个参数值对应的记录的当前状态异常时，对导致异常产生的原因、失效链接所处的模块位置以及提取时间进行分析和整理，以得到异常分析结果。

13. 根据权利要求12所述的失效检测装置，其特征在于，还包括：

显示单元，用于显示所述异常分析结果。

链接URL的失效检测方法和装置

技术领域

[0001] 本发明涉及互联网技术领域，尤其涉及一种链接URL的失效检测方法和装置。

背景技术

[0002] 在互联网领域，信息的发布者通常通过链接的方式将信息发布给浏览者。例如，在电子商务领域，商家通常通过向浏览者发布带有商品信息的链接来实现商品的展销。

[0003] 所谓链接也称超级链接，超级链接是指从一个网页指向一个目标的连接关系，而在一个网页中用来超链接的对象，可以是一段文本或者是一个图片。当浏览者单击已经链接的文字或图片后，链接目标将显示在浏览器上，并且根据目标的类型来打开或运行。

[0004] 然而，当链接中包含的信息发生异常后如信息过期或信息删除有可能导致浏览者无法浏览该链接，从而使得链接失效，成为失效链接。

[0005] 目前，在发布的链接中，有相当一部分链接尤其是人工维护的链接，其只通过人工检测链接的状态。这种通过人工检测链接状态的方法，效率较低且检测结果的误差较大。

发明内容

[0006] 有鉴于此，本发明提供了一种链接URL的失效检测方法和装置，以提高检测的效率并减少检测误差。

[0007] 为了解决上述技术问题，本发明采用了如下技术方案：

[0008] 一种链接URL的失效检测方法，所述链接URL显示在显示平台上，每个所述链接URL至少包括一个核心要素，所述核心要素包括参数名及参数值；所述显示平台上设置有预设表，所述预设表中记录有与每个核心要素相对应的实体的生命周期状态，所述失效检测方法包括：

[0009] 步骤A、抓取显示在所述显示平台上的符合特定要求的所有链接URL；

[0010] 步骤B、提取每个链接URL中的核心要素；

[0011] 步骤C、从所述预设表中查找与提取到的每个链接URL中的核心要素相对应的实体；

[0012] 步骤D、根据所述预设表中记录的与提取到的核心要素相对应的实体的生命周期状态判断提取的每个核心要素的当前状态是否异常，以检测该核心要素对应的链接URL是否失效。

[0013] 可选地，所有各个所述链接URL具有统一的数据结构格式，所述预设表包括多个第一表格，一个所述第一表格中记录有一个类别的各个参数值对应的实体的生命周期状态；每个所述类别中的所有各个链接URL具有相同的参数名；

[0014] 所述步骤A之后，还包括：

[0015] 步骤E、根据每个链接URL中的参数名对所有各个链接URL进行分类，将所有各个链接URL归属到相应的类别；在每个类别中包含有各个核心要素的记录；

[0016] 所述步骤B具体为：

- [0017] 步骤B'、从每个所述类别中的每个记录提取参数值；
- [0018] 所述步骤C具体包括：
 - [0019] 步骤C1、从显示平台上查找与每个类别分别对应的第一表格；
 - [0020] 步骤C2、从所述第一表格中查找与类别中的每个参数值相对应的主键值，根据所述主键值查找与所述主键值相对应的实体；
- [0021] 所述步骤D具体包括：
 - [0022] 步骤D'、根据类别对应的异常判断规则和所述第一表格中记录的每个实体的生命周期状态判断类别中的每个参数值对应的每个记录的当前状态是否异常，以检测该记录对应的链接URL是否失效。
- [0023] 可选地，所述步骤E之后还包括：
 - [0024] 步骤F、当归属于同一个类别中的链接URL中有N个具有相同的参数值时，从该类别中移除(N-1)个具有该相同参数值的记录，使得在类别中只保留一个该相同参数值对应的记录，其中，N≥2，N为整数。
- [0025] 执行所述步骤F的同时或之后还包括：
 - [0026] 步骤G、建立被移除的记录与其包含的参数值的映射关系，以根据参数值能够找到包含该参数值的所有链接URL；
- [0027] 所述步骤D'之后，还包括：
 - [0028] H、当提取的某一类别下的某一个参数值对应的记录的当前状态异常时，根据所述映射关系查找包含该参数值的所有被移除的链接URL。
- [0029] 可选地，检测完链接URL是否失效之后还包括：
 - [0030] 步骤I、当提取到的某一参数名的某一个参数值的当前状态异常时，对异常产生的原因、失效链接所处的模块位置以及提取时间进行分析和整理，以得到异常分析结果；
- [0031] 步骤J、显示所述异常分析结果。
- [0032] 可选地，所述步骤B'具体包括：
 - [0033] 确定链接URL的类别；
 - [0034] 检验链接URL中是否存在与该类别相对应的参数名：如果是，检验所述参数名是否位于所述链接URL的参数域，如果是，提取参数值，所述参数值以链接URL中的“=”开始，且“=”之前为所述参数名；
 - [0035] 校验所述参数值是否满足预设条件，如果是，记录该提取的参数值。
 - [0036] 可选地，所述映射关系包括链接类别、参数名、参数值以及链接URL的对应关系。
 - [0037] 可选地，当一个链接URL中包括多个参数名时，所述步骤E具体包括：
 - [0038] 根据该链接URL中的多个参数名将该链接URL分别归属到与其参数名相对应的多个类别中。
 - [0039] 一种链接URL的失效检测装置，所述链接URL显示在显示平台上，每个所述链接URL至少包括一个核心要素，所述核心要素包括参数名及参数值；所述显示平台上设置有预设表，所述预设表中记录有与每个核心要素相对应的实体的生命周期状态，所述失效检测装置包括：
 - [0040] 抓取单元，用于抓取显示在所述显示平台上的符合特定要求的所有链接URL；
 - [0041] 提取单元，用于提取每个链接URL中的核心要素；

[0042] 第一查找单元,用于从所述预设表中查找与提取到的每个链接URL中的核心要素相对应的实体;

[0043] 判断单元,用于根据所述预设表中记录的与提取到的核心要素相对应的实体的生命周期状态判断提取的每个核心要素的当前状态是否异常,以检测该核心要素对应的链接URL是否失效。

[0044] 可选地,所有各个所述链接URL具有统一的数据结构格式,所述预设表包括多个第一表格,一个所述第一表格中记录有一个类别的各个参数值对应的实体的生命周期状态;每个所述类别中的所有各个链接URL具有相同的参数名;

[0045] 所述失效检测装置还包括:

[0046] 分类单元,用于在接收到所述抓取单元抓取的链接URL后,根据每个链接URL中的参数名对所有各个链接URL进行分类,将所有各个链接URL归属到相应的类别;在每个类别中包含有各个核心要素的记录;

[0047] 所述提取单元包括从每个所述类别中的每个记录中提取参数值的子单元;

[0048] 所述第一查找单元具体包括:

[0049] 第一查找子单元,用于从显示平台上查找与每个类别分别对应的第一表格;

[0050] 第二查找子单元,用于从所述第一表格中查找与类别中的每个参数值相对应的主键值,根据所述主键值查找与所述主键值相对应的实体;

[0051] 所述判断单元具体包括:

[0052] 判断子单元,所述判断子单元用于根据类别对应的异常判断规则和所述第一表格中记录的每个实体的生命周期状态判断类别中的每个参数值对应的每个记录的当前状态是否异常,以检测该记录对应的链接URL是否失效。

[0053] 可选地,所述失效检测装置还包括:移除单元,用于当归属于同一个类别中的链接URL中有N个具有相同的参数值时,从该类别中移除(N-1)个具有该相同参数值的记录,使得在类别中只保留一个该相同参数值对应的记录,其中,N≥2,N为整数。

[0054] 可选地,还包括:

[0055] 建立映射关系单元,用于在接收到所述移除单元移除记录的信号的同时或之后,建立被移除的记录与其包含的参数值的映射关系,以根据参数值能够找到包含该参数值的所有链接URL;

[0056] 第二查找单元,用于当所述判断子单元判断的提取的某一类别下的某一个参数值对应的记录的当前状态异常时,根据所述映射关系查找包含该参数值的所有被移除的链接URL。

[0057] 可选地,还包括:

[0058] 异常结果分析单元,用于当提取到的某一参数名的某一个参数值对应的记录的当前状态异常时,对导致异常产生的原因、失效链接所处的模块位置以及提取时间进行分析和整理,以得到异常分析结果。

[0059] 可选地,还包括:

[0060] 显示单元,用于显示所述异常分析结果。

[0061] 相较于现有技术,本发明具有以下有益效果:

[0062] 本发明提供的链接URL的失效检测方法,通过根据预设表中记录的与提取到的核

心要素相对应的实体的生命周期状态判断提取的每个核心要素的当前状态是否异常,来判断链接URL是否失效。这是因为一个链接URL是否失效与其包含的核心要素的生命周期状态是否异常相对应。当核心要素的生命周期状态正常时,链接URL有效,当生命周期状态异常时,链接URL失效。

[0063] 该失效检测方法可以通过计算机自动实现,相较于现有技术中通过人工进行检测的方法,提高了检测效率,同时也使检测效果的误差大大减小。

附图说明

[0064] 为了清楚地理解本发明的技术方案,下面将在描述本发明的具体实施方式时用到的附图作一简要说明。显而易见地,这些附图仅是本发明实施例的部分附图,本领域技术人员在不付出创造性劳动的前提下还可以获得其它的附图。

[0065] 图1是本发明实施例一提供的链接URL的失效检测方法流程示意图;

[0066] 图2是本发明实施例二提供的链接URL的失效检测方法流程示意图;

[0067] 图3是本发明实施例提供的每个类别下的每个链接的参数值的提取方法流程示意图;

[0068] 图4是本发明实施例三提供的链接URL的失效检测方法流程示意图;

[0069] 图5是本发明实施例四提供的链接URL的失效检测方法流程示意图;

[0070] 图6是本发明实施例五提供的链接URL的失效检测装置的结构示意图;

[0071] 图7是本发明实施例六提供的链接URL的失效检测装置的结构示意图。

具体实施方式

[0072] 为使本发明的技术方案更加清楚、完整,效果更加突出,下面结合附图对本发明的具体实施方式进行详细描述。

[0073] 正如背景技术部分所述,在显示平台上有一部分链接只能通过人工来检测其状态。这种通过人工来检测链接状态的方法对于电商运营信息的发布平台来说更为普遍。这是因为:

[0074] 1、电商平台商品种类的多样性,营销手段的复杂性和时效性,导致展示给用户的链接中蕴含的信息(即核心要素)具有复杂的生命周期。而相当一部分链接,尤其是人工维护的营销和广告链接,不由程序或应用服务器控制,无法形成有效的通信机制。因此并不能即时感应其核心要素生命周期的变化,从而导致链接失效。

[0075] 例如一个链接中可能包含一个核心要素商品id,该商品id唯一标示一个商品。该商品在某时刻可能过期,或者在某时刻被修改甚至下架,而包含该信息的链接,除非是运营人员主动测试,可能无法自动感应到包含的商品信息生命周期的变化,从而导致该链接失效了仍能被用户看到。

[0076] 2、电子商务独特的交易模式导致了链接的发布者和链接所蕴含核心要素的发布者不一定来自同一机构和人员。当链接中蕴含的信息出现异常时,发布链接的一方往往不能够及时收到反馈信息。

[0077] 3、无论是手工链接,还是自动生成的链接,当电商平台信息达到一定数量级时,通过人工来逐一检测失效链接效率极低且不精确。

[0078] 基于上述原因,在电商营销信息的发布平台中,当发布的链接中蕴含的核心要素(如商品id、咨询id或者公告id等)出现异常而导致链接失效时,一般无法及时地检测到,因此,就会出现一些失效链接会显示在电商平台中,而当用户试图打开该链接时却不能打开。

[0079] 基于此,有必要提供一种方法来自动检测显示在显示平台上的链接URL是否能够正常显示。

[0080] 有鉴于此,本发明提供了一种链接URL的失效检测方法。需要说明的是,该失效检测方法不仅适用于电商平台,还适用于其它显示平台。

[0081] 本发明提供的链接URL的失效检测方法的具体实现方式请参见以下实施例。

[0082] 实施例一

[0083] 需要说明的是,在本发明实施例中,用于显示链接URL的显示平台上设置有一个预设表,所述预设表中记录有与每个核心要素相对应的实体的生命周期状态。所述实体是数据库中对现实世界中的对象或概念的描述。

[0084] 每个链接URL至少包括一个核心要素,所述核心要素是指链接URL中所包含的参数名以及参数值。在预设表中包括主键名和主键值。其中,主键名与链接URL中的参数名相对应,主键值与链接URL中的参数值相对应。在一个链接URL中至少包括一个核心要素。也就是说,在一个链接URL中可以包括多个核心要素。该核心要素的参数名一般位于URL中的参数域的位置。所述参数名为链接URL中符号“?”或“&&”后面的关键词。参数值是以符号“=”开始,且“=”之前为参数名,并且以空符号或者“&&”结束。在本发明实施例中,参数名可以认为是核心要素的key值。参数值可以认为是核心要素的value值。

[0085] 例如,在链接“a.b.c.com/d.jhtml?advId=123”中,参数名为“advId”,参数值为“123”。

[0086] 对于不同的核心要素来说,生命周期状态也不尽相同。不同核心要素的生命周期状态可以包括:审核是否通过的状态、是否开放的状态、有效期等等。

[0087] 图1是本发明实施例一提供的链接URL的失效检测方法的流程示意图。如图1所示,实施例一提供的链接URL的失效检测方法包括以下步骤:

[0088] S101、抓取显示在所述显示平台上的符合特定要求的所有链接URL:

[0089] 显示平台对某些抓包软件进行授权,然后利用这些授权的抓包软件抓取显示在所述显示平台上的符合特定要求的所有链接URL。

[0090] 抓取到的所有各个链接URL统一存储在数据库或者特定格式的文档中,以便进行后续工作。

[0091] 需要说明的是,一个显示平台上的链接URL一般会多种多样,其格式也会千差万别。而当用户需要检测显示平台上的链接是否失效时,一般只对某些种类的链接进行检测。所以,一般情况下,不需要抓取显示平台上的所有链接URL,只需要根据用户需求抓取显示平台上的符合特定要求的所有链接URL。

[0092] S102、提取每个链接URL中的核心要素:

[0093] 在HTTP协议中,要获取一个响应结果集,一般需要在请求的链接URL中加入参数域,参数域上的字段即为该链接的核心要素。如上所述,核心要素包括该链接的参数名和参数值。

[0094] 为了保证提取到的核心要素的准确性,优选按照预设的提取规则提取每个链接

URL中的核心要素。需要说明的是,在本发明实施例中,可以将链接URL理解为一个字符串,按照这种理解方式,从链接URL中提取核心要素的过程可以理解为从一个字符串中提取关键词的过程。

[0095] 例如,如果想从以下广告链接中提取每个链接的核心要素即参数名和参数值。

[0096] 1、<http://a.b.c.com/d.jhtml?advId=123&&goodId=1234>;

[0097] 2、<http://a.b.advId.com/d.jhtml?advId=123>;

[0098] 即提取key = “advId”,value = “123”。那么首先应查找关键词“advId”,但由于一条链接中可能存在多个相同关键词,也可能同时包含多个核心要素,因此预设提取规则可以具体为:

[0099] 1、核心要素的参数名存在于链接的参数域,即符号“?”或“&&”后面的关键字才是我们所需要的。

[0100] 2、核心要素的参数值以符号“=”开始,且“=”之前为参数名,同时以空符号或者“&&”符号结束。

[0101] 3、核心要素的参数值满足预设条件,该预设条件包括数值大小和位数要求。由于不同核心要素的起始点、参数值的位数一般有所区别,因此在链接中同时存在多个核心要素时,可以以此作为进一步判断和提取核心要素参数值的依据。

[0102] S103、从所述预设表中查找与提取到的每个链接URL中的核心要素相对应的实体:

[0103] 由于预设表中记录有每个实体的生命周期状态,所以,根据预设表的实体与核心要素的对应关系,从预设表中可以查找到与提取到的每个链接URL中的核心要素相对应的实体。其中,需要说明的是,预设表中的实体与链接URL的核心要素的对应关系是预先设置好的。具体地说,核心要素中的参数名对应预设表中的实体的主键名,参数值对应预设表中的主键值。因此,本步骤可以根据该预先设置好的对应关系从所述预设表中查找与提取到的每个链接URL中的核心要素相对应的实体。

[0104] S104、根据预设表中记录的与提取到的核心要素相对应的实体的生命周期状态判断提取到的每个核心要素的当前状态是否异常,以检测该核心要素对应的链接URL是否失效:

[0105] 由于预设表中记录有每个实体的生命周期状态,所以可以根据预设表中记录的与提取到的核心要素相对应的实体的生命周期状态判断提取到的每个核心要素的当前状态是否异常,从而检测该核心要素对应的每个链接URL是否失效。当核心要素的当前状态异常时,该核心要素对应的每个链接URL失效,当核心要素的当前状态正常时,该核心要素对应的每个链接URL有效。

[0106] 需要说明的是,判断核心要素的当前状态是否异常的判断方法随着不同类型的链接URL的不同而变化。

[0107] 步骤S104所述的过程也可以认为是异常匹配的过程。即对提取到的核心要素的生命周期的当前状态与预设表中记录的与其相对应的实体的正常生命周期状态进行匹配,如果提取到的核心要素的当前状态与正常的生命周期状态相匹配,则认为提取到的核心要素的当前状态正常,该核心要素所在的链接URL为有效链接,反之,如果提取到的核心要素的当前状态与正常的生命周期状态不相匹配,则认为提取到的核心要素的当前状态异常,该核心要素所在的链接URL为失效链接。需要说明的是,在异常匹配的过程中,需要根据链接

URL的不同类型选择不同的异常匹配规则。

[0108] 以广告链接类型为例,其匹配规则可以如下:

[0109] 1)、该广告的审核状态为:通过;

[0110] 2)、该广告的开放状态为:开放;

[0111] 3)、该广告的生命周期状态为:投放中;

[0112] 4)、检索时间点处于开放时间和过期时间之间。

[0113] 没有同时满足以上条件的核心要素的状态为异常状态。

[0114] 通过以上步骤S101至步骤S104就实现了对链接URL的失效的自动检测。相较于现有技术中通过人工来对链接是否失效进行检测的方法,提高了检测效率,而且提高了检测的准确性,减少了检测误差。

[0115] 需要说明的是,这种失效检测方法尤其适用于电商平台。

[0116] 为了能够了解失效链接产生的原因、失效链接所处的模块位置以及提取时间等信息,方便运行维护人员的处理,本发明实施例一提供的链接URL的失效检测方法还可以包括以下步骤:

[0117] S105、当检测到的某一个链接URL失效时,对失效产生的原因、失效链接所处的模块位置以及提取时间进行分析和整理,以得到异常分析结果;

[0118] 具体地,在选择相应的异常匹配规则对生命周期状态进行异常匹配时,通常需要对该链接URL类型的所有各个项目的状态进行匹配。当有一个项目的状态发生异常后,就认为包含有该异常状态核心要素的链接URL为失效链接。

[0119] 所以,在匹配过程中,记录状态发生异常的项目,该发生异常的项目即为链接失效产生的原因。

[0120] 异常结果分析中,链接URL在网站平台中投放的位置尤其重要。因此在抓取链接和对链接进行分类的过程中,需要记录每个链接所处的位置。

[0121] 另外,一个链接URL的时间有效性很重要,所以,在提取每个链接的核心要素时,还可以记录每个核心要素的提取时间,从而在可以分析出核心要素的提取时间。

[0122] 为了清楚地了解链接失效的产生原因、失效链接所处的模块位置以及提取时间,本步骤对失效链接的失效产生的原因、失效链接所处的模块位置以及提取时间进行分析和整理,从而得到异常分析结果。

[0123] S106、显示所述异常分析结果;

[0124] 显示上述得到的异常分析结果,以供显示平台的运行维护人员对失效链接进行处理。

[0125] 通过步骤S105和步骤S106将异常分析结果展示给运行维护人员,方便运行维护人员对失效链接的处理,例如,将那些失效链接从显示平台上删除,这样使得显示在显示平台上的链接URL均为可打开的链接,消除了显示在显示平台上的某些链接URL无法打开的现象,有利于提高用户体验。

[0126] 以上为本发明实施例一提供的链接URL的失效检测方法。通过以上方法能够代替人工实现对链接URL是否失效的自动检测,因此,上述实施例一提供的失效检测方法能够提高检测效率,减小检测误差。

[0127] 需要说明的是,上述实施例一提供的失效检测方法可以适用于显示在各种平台的

链接URL。例如：显示在电商平台上的商品链接、广告链接等等。

[0128] 一般情况下，显示在同一显示平台上的链接URL包括多种类型的链接，如用户链接、商品链接、资讯链接、广告链接等等。然而相同类型的链接在进行判断核心要素的生命周期状态是否异常的判断规则相同。所以，为了方便对预设表中记录的实体的生命周期状态与提取的核心要素的生命周期状态进行匹配和判断，优选对所有各个链接URL根据其核心要素的参数名进行分类，然后根据类型对各个参数值进行异常匹配，具体参见实施例二。

[0129] 实施例二

[0130] 需要说明的是，能够对显示在同一显示平台上的链接URL根据其核心要素的参数名进行分类得以进行的条件是显示在同一显示平台上的链接URL具有统一的数据结构格式。只有具有统一的数据结构格式，才能按照预设的分类规则对链接URL进行分类，否则容易出现分类错误。而大部分使用HTTP协议的网站平台一般都有相对规范的命名格式，尤其是对于电商网站，因此一般能满足该条件。所以实施例二提供的链接URL的失效检测方法特别适用于电商平台上的链接URL。

[0131] 另外，由于是根据链接URL中的参数名对链接URL进行分类，所以位于同一类别下的各个链接URL具有相同的参数名。

[0132] 另外，在实施例二中，为了便于判断每个类别中的每个参数值对应的核心要素的当前状态是否异常，在显示平台上设置有多个第一表格。每个第一表格与每个类别的对应关系是预设的。因此，可以根据类别查找到与其对应的第一表格。

[0133] 一个第一表格记录有一个类别的各个参数值对应的实体的生命周期状态。

[0134] 结合附图2对实施例二提供的链接URL的失效检测方法进行描述。如图2所示，实施例二提供的链接URL的失效检测方法包括以下步骤：

[0135] S201、抓取显示在所述显示平台上的符合特定要求的所有链接URL；

[0136] 该步骤与实施例一中的步骤S101相同，为了简要起见，请参见实施例一的描述。

[0137] S202、根据每个链接URL的参数名对所有各个链接URL进行分类，以将所有各个链接URL归属到相应的类别中：

[0138] 每个链接URL的参数名表示该链接所属的类别，所以根据每个链接URL的参数名对所有各个链接URL进行分类，以将所有各个链接URL归属到相应的类别中。

[0139] 在每个类别中包括多条具有预设数据结构的记录，在将各个链接URL归属到相应的类别的过程就是将链接URL中的信息填充在记录的预设数据结构中。

[0140] 换句话说，该预设数据结构的记录用于记录链接URL的完整信息。所以一条记录包含有一个链接URL的完整信息。并且，在每个类别中，一条记录对应一个链接URL。

[0141] 所述预设数据结构的记录是指记录有链接URL对应的链接类别、对应的参数名、以及该参数值的提取时间等信息。其中，提取时间可以精确到天。

[0142] 在本发明实施例中，每条预设数据结构的记录具体可以如下表所示：

| 字段名 | | 字段描述 |
|--------|--------|-----------------------------|
| [0143] | 编号 | 序列号, 唯一标示一条核心要素信息。 |
| | 所属链接类别 | 对应核心要素的参数名, 例如商品、广告、资讯、公告等。 |
| | 参数值 | 对应参数值, 可重复。 |
| | 所属链接 | 该核心要素的参数名 |
| | 提取时间 | 精确到某天 yyyy-mm-dd |

[0144] 在本发明实施例中, 根据链接所主要包含的参数名不同, 可将链接按多个类别进行分类, 例如可以分为广告链接、用户链接、商品链接、资讯链接和公告链接等类别。相应的参数名即为链接中的广告id、用户id、商品id、资讯id以及公告id等。

[0145] 对于只包含一个参数名的链接URL, 根据其参数名将该链接归属到一个类别中。对于包含多个参数名的链接URL, 则将该链接URL同时归属到不同的类别中。也就是说, 对于包含多个参数名的链接URL, 该一个链接归属到多个不同的类别中。

[0146] 此外, 不同类别下链接中包含的参数名也有可能重名, 这时通过对链接格式、参数值的位数的进一步分析来确定其类别。另外, 对于某些特殊的链接中不包含任何参数名, 则无法识别其类别。

[0147] 链接分类如下表所示:

[0148]

| 链接URL | 参数名 | 归类的类别 |
|---|--------------|---------|
| a.b.c.com/d.jhtml?advId=123 | advId | 广告链接 |
| a.b.c.com/d.jhtml?userId=1234 | userId | 用户链接 |
| a.b.c.com/d.jhtml?goodId=12345 | goodId | 商品链接 |
| a.b.c.com/d.jhtml?infoId=123456 | infoId | 资讯链接 |
| a.b.c.com/d.jhtml?noticId=1234567 | noticId | 公告链接 |
| a.b.c.com/d.jhtml?advId=123&&goodId=12345 | advId、goodId | 广告和商品链接 |
| a.b.c.com/d/ef/ee.html | 无法识别 | 其它 |

[0149] 需要说明的是, 链接URL能够根据以上规则进行归类优选具有几个前提条件, 具体如下:

[0150] 1) 链接具有固定格式和规律性: 如一般以http开头, 链接中所带参数(即核心要素的参数名)使用'?'或'\$'符号分隔。

[0151] 2) 链接中核心要素的参数名的命名固定: 如商品链接中的核心要素使用某个固定的词汇命名, 例如goodId。链接中包含多个核心要素的参数名时, 各核心要素的参数名的先后顺序最好也相对固定。

[0152] 3) 链接中核心要素的参数名都有值。

[0153] 由于电商平台对商品链接以及其各要素都有相对规范的格式, 因此基本满足以上

条件,所以,本发明提供的失效检测方法特别适用于电商平台中的链接。

[0154] 通过以上规则生成链接的识别和归类算法,通过程序自动完成链接的归类工作。对于无法识别其类目的链接,则统一归类到“其它”类别。

[0155] S203、提取每个类别下的每个记录的参数值:

[0156] 该步骤的具体实现过程如图3所示,其包括以下步骤:

[0157] S2031、确定链接URL的类别。

[0158] S2032、检验链接URL中是否存在与该类别相对应的参数名,如果是,执行步骤S2033。

[0159] S2033、检验所述参数名是否位于所述链接URL的参数域,如果是,执行步骤S2034。

[0160] S2034、提取参数值,所述参数值以链接URL中的“=”开始,且“=”之前为所述参数名;

[0161] S2035、校验所述参数值是否满足预设条件,如果是,执行步骤S2036。

[0162] 该预设条件包括数值大小和位数要求。由于不同核心要素的起始点、参数值的位数一般有所区别,因此在链接中同时存在多个核心要素时,可以以此作为进一步判断和提取核心要素参数值的依据。

[0163] S2036、记录该提取的参数值。

[0164] S204、从显示平台上查找与每个类别分别对应的第一表格:

[0165] 根据第一表格和类别的预设对应关系,从显示平台上查找与每个类别分别对应的第一表格。

[0166] S205、从所述第一表格中查找与类别中的每个参数值相对应的主键值,根据该主键值查找到与该主键值相对应的实体。

[0167] S206、根据第一表格中记录的每个实体的生命周期状态和异常判断规则判断类别中的每个参数值对应的每个记录的当前状态是否异常,以检测该记录对应的链接URL是否失效:

[0168] 在判断每个记录的当前状态是否异常之前,需要确定记录所处的类别。因为,每个链接类别的异常判断规则不同,在确定了记录所处的类别后,方便选择所对应的异常判断规则。

[0169] 具体判断过程如下:第一表格中记录的每个实体的生命周期状态包括每个实体的各个数据的生命周期状态,因此,根据该类别对应的异常判断规则和第一表格中记录的每个实体的生命周期状态判断类别中的每个记录的当前状态是否异常,从而检测该记录对应的链接URL是否失效。

[0170] 步骤S207至步骤S208与实施例一中的步骤S105至步骤S106相同,为了简要起见,在此不再赘述,具体参见实施例一的相关描述。

[0171] 通过实施例二提供的失效检测方法,在异常匹配之前将不同类型的链接URL进行了分类确定了链接URL所属的类别,从而能够在异常匹配前就选择好了异常判断规则,这种失效检测方法有利于提高异常判断效率,进而有利于提高链接的失效检测效率。

[0172] 实施例二所述的失效检测方法通过将链接URL进行分类提高了链接的失效检测效率,但是有可能存在这样一种情况:多个链接URL中可能包含同一个核心要素,即在同一类别下的多个记录可能具有相同的参数值,因此,步骤S203提取出的参数值可能存在大量重

复的值。然而,在进行异常匹配时,每个核心要素都要进行匹配,因而针对每个类别来说,其包括的每条记录都要进行异常匹配,而重复的记录的异常匹配会带来大量的不必要的消耗,所以,为了避免对重复的参数值的异常匹配,本发明还提供了实施例三。

[0173] 实施例三

[0174] 实施例三所述的方法与实施例二所述的方法有诸多相似之处,为了突出两个实施例之间的区别点,本实施例仅对其不同之处进行着重描述,其相似之处请参见实施例二的描述。

[0175] 参见图4,实施例三提供的链接URL的失效检测方法包括以下步骤:

[0176] 步骤S401至步骤S402与实施例二中的步骤S201至S202相同,为了简要起见,在此不再赘述,详细参见实施例二的描述。

[0177] S403、当归属于同一个类别中的链接URL中有N个具有相同的参数值时,从该类别中移除(N-1)个具有该相同参数值的记录使得在类别中只保留一个该相同参数值对应的记录,其中,N≥2,N为整数:

[0178] 举例说明:在goodID(商品)类别中有5个链接URL,在这5个链接URL中,有3个链接URL的参数值均为“123”,则从该商品类别中移除2个参数值为“123”的记录,这样在该商品类别中只留有一个参数值为“123”的记录。

[0179] 步骤S404至步骤S409与实施例二中的步骤S203至步骤S208相同,为了简要起见,在此不再赘述,详细信息参见实施例二的描述。

[0180] 实施例三提供的失效检测方法能够避免同一核心要素重复匹配的问题,但是当一个核心要素的当前状态异常时,其对应的所有各个链接均为失效链接。为了能够找到状态异常的核心要素对应的所有各个链接,本发明还提供了实施例四。

[0181] 实施例四

[0182] 实施例四所述的方法与实施例三所述的方法有诸多相似之处,为了简要起见,本实施例仅对其不同之处进行描述,请相似之处请参见实施例三的描述。

[0183] 如图5所示,实施例四提供的失效检测方法包括以下步骤:

[0184] 步骤S501至步骤S502与实施例三中的步骤S401至S402相同,为了简要起见,在此不再赘述,详细参见实施例三的描述。

[0185] S503、提取每个类别下的每个记录的参数值。

[0186] 由于多个链接URL中可能包含同一个核心要素,即在同一类别下的多个记录可能具有相同的参数值,因此,步骤S503提取出的参数值可能存在大量重复的值。然而,在进行异常匹配时,每个核心要素都要进行匹配,因而针对每个类别来说,其包括的每个记录都要进行异常匹配,而重复的记录的异常匹配会带来大量的不必要的消耗,所以,为了避免对重复记录的异常匹配,在进行异常匹配前,还可以包括步骤S504。

[0187] S504、当归属于同一个类别中的链接URL中有N个具有相同的参数值时,从该类别中移除(N-1)个具有该相同参数值的记录,使得在类别中只保留一个该相同参数值对应的记录,其中,N≥2,N为整数:

[0188] 举例说明:在goodID(商品)类别中有5个记录,在这5个记录中,有3个记录的参数值均为“123”,则从该商品类别中移除2个参数值为“123”的记录,这样在该商品类别中只留有一个参数值为“123”的记录。

[0189] S505、建立被移除的记录与其包含的参数值的映射关系,以根据参数值能够找到包含该参数值的所有链接URL:

[0190] 如果上述步骤S504移除的记录的核心要素的状态异常时,则说明包含该核心要素的所有链接URL均为失效链接。为了能够找到所有失效链接,在从类别中移除具有相同参数值的(N-1)个记录的同时或之后,建立被移除的记录与其包含的参数值的映射关系,以根据参数值能够找到包含该参数值的所有链接URL。在本发明实施例中,是在移除之后,建立被移除的记录与其包含的参数值的映射关系。

[0191] 需要说明的是,建立的映射关系可以为链接类别、参数名、参数值以及链接URL的对应关系。

[0192] 具体地说,建立映射关系的过程可以为将移除的记录进行拆分,将这些拆分的信息存储在另一数据结构中的过程。

[0193] 作为本发明的一个具体实施例,拆分后的信息存储的数据结构如下表所示:

[0194]

| 字段名 | 字段描述 |
|-------|--|
| 编号 | 唯一标示一条链接。 |
| 链接URL | 展示给用户的链接,如 http://a.b.c.com |
| 所属类别名 | 如商品链接、广告链接、公告链接等。 |
| 链接位置 | 当前链接所处的模块位置。 |

[0195] 2链接类目表

[0196]

| 字段名 | 字段描述 |
|-------|--|
| 链接类别名 | 展示给用户的链接,如 http://a.b.c.com |
| 核心要素名 | 即核心要素的key值,如goodId advId等 |

[0197] 3核心要素表

[0198]

| 字段名 | 字段描述 |
|-------|------------------------|
| 核心要素值 | 对应核心要素value值,唯一标示一条记录。 |
| 核心要素名 | 对应核心要素key值, |
| 关联 | 该核心要素 |
| 提取时间 | 精确到某天yyyy-mm-dd |

[0199] 4链接和核心要素关联表

[0200]

| 字段名 | 字段描述 |
|---------|-------------------|
| 编号 | 唯一标示一条记录。 |
| 关联链接编号组 | 编号组每一个记录对应一个链接编号。 |

[0201] 步骤S506至步骤S508与实施例二中的步骤S205至步骤S207相同,为了简要起见,在此不再赘述,具体参见实施例一的相关描述。

[0202] S509、当提取的某一类别下的某一个参数值对应的记录的当前状态异常时,根据所述映射关系查找包含该参数值的所有被移除的链接URL:

[0203] 如果包含状态异常的参数值的链接URL因为参数值相同,在步骤S505中被移除,为了查找到所有包含该状态异常的参数值的链接URL,本发明实施例需要根据上述建立的映射关系查找包含当前状态异常的参数值的所有被移除的链接URL。

[0204] 步骤S510至步骤S511与实施例一中的步骤S105至步骤S106相同,为了简要起见,在此不再赘述,具体参见实施例一的相关描述。

[0205] 实施例三提供的失效检测方法除了具有实施例一和实施例二所述的有益效果外,还能够查找到被移除的所有链接URL。

[0206] 基于上述实施例一至实施例四所述的失效检测方法,本发明实施例还提供了失效检测装置的具体实施例。具体参见如下实施例。

[0207] 实施例五

[0208] 需要说明的是,实施例五提供的链接URL的失效检测装置适用于显示在显示平台上的链接URL,并且每个所述链接URL至少包括一个核心要素,所述核心要素包括参数名及参数值。另外,在显示平台上设置有预设表,所述预设表记录有每个核心要素的生命周期状态。

[0209] 如图6所示,实施例五提供的链接URL的失效检测装置包括以下单元:

[0210] 抓取单元601,用于抓取显示在所述显示平台上的符合特定要求的所有链接URL;

[0211] 提取单元602,用于提取每个链接URL中的核心要素;

[0212] 第一查找单元603,用于从所述预设表中查找与提取到的每个链接URL中的核心要素相对应的实体;

[0213] 判断单元604,用于根据所述预设表中记录的与提取到的核心要素相对应的实体的生命周期状态判断提取的每个核心要素的当前状态是否异常,以检测该核心要素对应的链接URL是否失效。

[0214] 通过实施例四提供的链接URL的失效检测装置能够实现对链接是否失效的自动检测,有利于提高检测效率和准确性。

[0215] 另外,为了能够查找到异常原因、异常核心要素对应的链接等等,本发明实施例提供的失效检测装置还可以包括:

[0216] 异常结果分析单元605,用于当提取到的某一参数名的某一个参数值对应的记录的当前状态异常时,对导致异常产生的原因、失效链接所处的模块位置以及提取时间进行分析和整理,以得到异常分析结果。

[0217] 为了能够将异常分析结果显示出来,以供运行维护人员处理,本发明实施例提供的失效检测装置还可以进一步包括:

[0218] 显示单元606,用于显示所述异常分析结果。

[0219] 以上为本发明实施例五提供的链接URL的失效检测装置。该失效检测装置能够实现对链接URL的自动失效检测,并且能够分析异常原因、失效链接所在的模块位置等等。但是,每个类型的链接URL的异常判断规则不同,为了方便进行异常判断,优选在异常判断前,对链接URL的类型进行分类处理。基于此,本发明还提供了实施例六。

[0220] 实施例六

[0221] 需要说明的是,要想对链接URL进行分类处理,所有各个链接URL具有统一的数据结构格式。而且,在显示平台上设置有多个第一表格,每个所述第一表格中记录有一个类别

的各个参数值对应的实体的生命周期状态；每个所述类别中的所有各个链接URL具有相同的参数名。

- [0222] 如图7所示，实施例六所述的失效检测装置包括以下单元：
 - [0223] 抓取单元701，用于抓取显示在所述显示平台上的符合特定要求的所有链接URL；
 - [0224] 分类单元702，用于在接收到所述抓取单元抓取的链接URL后，根据每个链接URL中的参数名对所有各个链接URL进行分类，将所有各个链接URL归属到相应的类别；在每个类别中包含有各个核心要素的记录；
 - [0225] 提取单元703，所述提取单元703包括从每个所述类别中的每个记录中提取参数值的子单元；
 - [0226] 第一查找单元706，用于从所述预设表中查找与提取到的每个链接URL中的核心要素相对应的实体；其具体包括：
 - [0227] 第一查找子单元7061，用于从显示平台上查找与每个类别分别对应的第一表格；
 - [0228] 第二查找子单元7062，用于从所述第一表格中查找与类别中的每个参数值相对应的主键值，根据所述主键值查找与所述主键值相对应的实体；；
 - [0229] 判断单元707，用于根据所述预设表中记录的与提取到的核心要素相对应的实体的生命周期状态判断提取的每个记录的当前状态是否异常，以检测该记录对应的链接URL是否失效，其具体包括：
 - [0230] 判断子单元7071，所述判断子单元7071用于根据类别对应的异常判断规则和所述第一表格中记录的每个实体的生命周期状态判断类别中的每个参数值对应的每个记录的当前状态是否异常，以检测该记录对应的链接URL是否失效；
 - [0231] 异常结果分析单元709，用于当提取到的某一参数名的某一个参数值对应的记录的当前状态异常时，对导致异常产生的原因、失效链接所处的模块位置以及提取时间进行分析和整理，以得到异常分析结果；
 - [0232] 显示单元710，用于显示所述异常分析结果。
 - [0233] 为了避免对同一个核心要素的重复匹配，本发明提供的失效检测装置还可以包括：
 - [0234] 移除单元704，用于当归属于同一个类别中的链接URL中有N个具有相同的参数值时，从该类别中移除(N-1)个具有该相同参数值的记录，使得在类别中只保留一个该相同参数值对应的记录，其中，N≥2，N为整数。
 - [0235] 进一步地，当被移除的链接URL中包括的核心要素为状态异常的核心要素时，为了能够查找到包含异常状态的所有链接URL，所述失效检测装置还可以进一步包括：
 - [0236] 建立映射关系单元705，用于在接收到所述移除单元移除记录的信号的同时或之后，建立被移除的记录与其包含的参数值的映射关系，以根据参数值能够找到包含该参数值的所有链接URL；
 - [0237] 第二查找单元708，用于当所述判断单元判断的提取的某一类别下的某一个参数值对应的记录的当前状态异常时，根据所述映射关系查找包含当前状态异常的参数值的所有被移除的链接URL。
 - [0238] 以上所述，仅是本发明的较佳实施例而已，并非对本发明作任何形式上的限制。
 - [0239] 虽然本发明已以较佳实施例披露如上，然而并非用以限定本发明。任何熟悉本领

域的技术人员,在不脱离本发明技术方案范围情况下,都可利用上述揭示的方法和技术内容对本发明技术方案作出许多可能的变动和修饰,或修改为等同变化的等效实施例。因此,凡是未脱离本发明技术方案的内容,依据本发明的技术实质对以上实施例所做的任何简单修改、等同变化及修饰,均仍属于本发明技术方案保护的范围内。

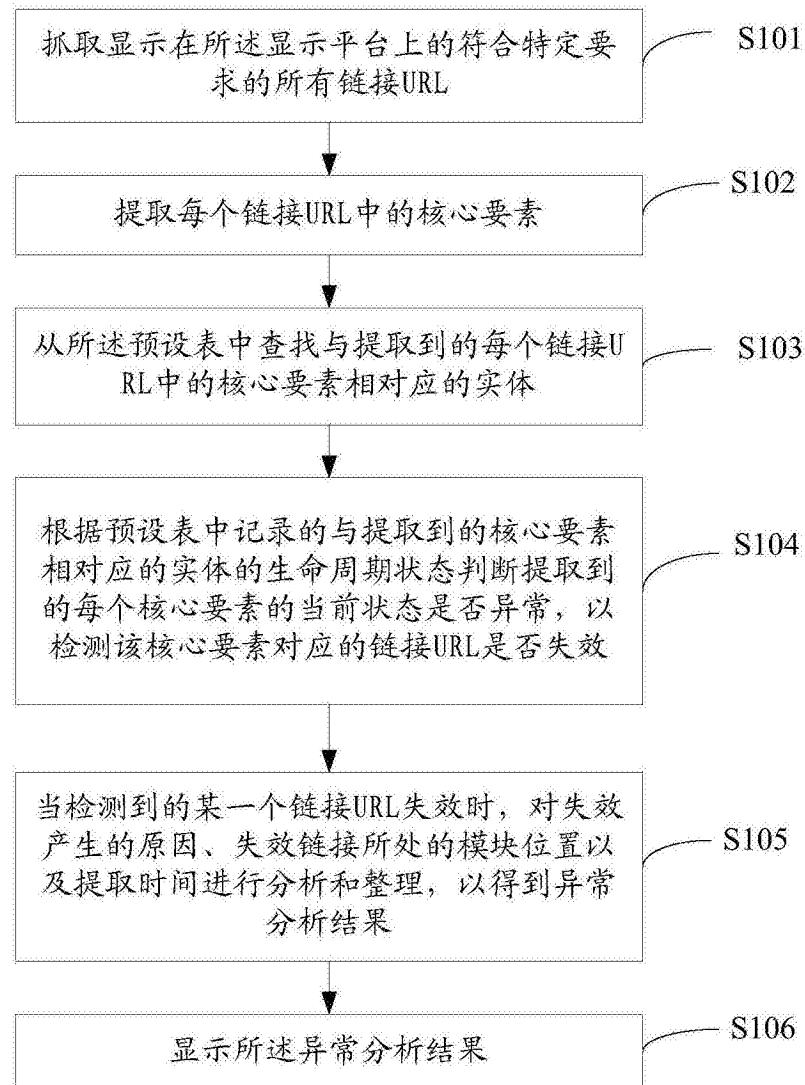


图1

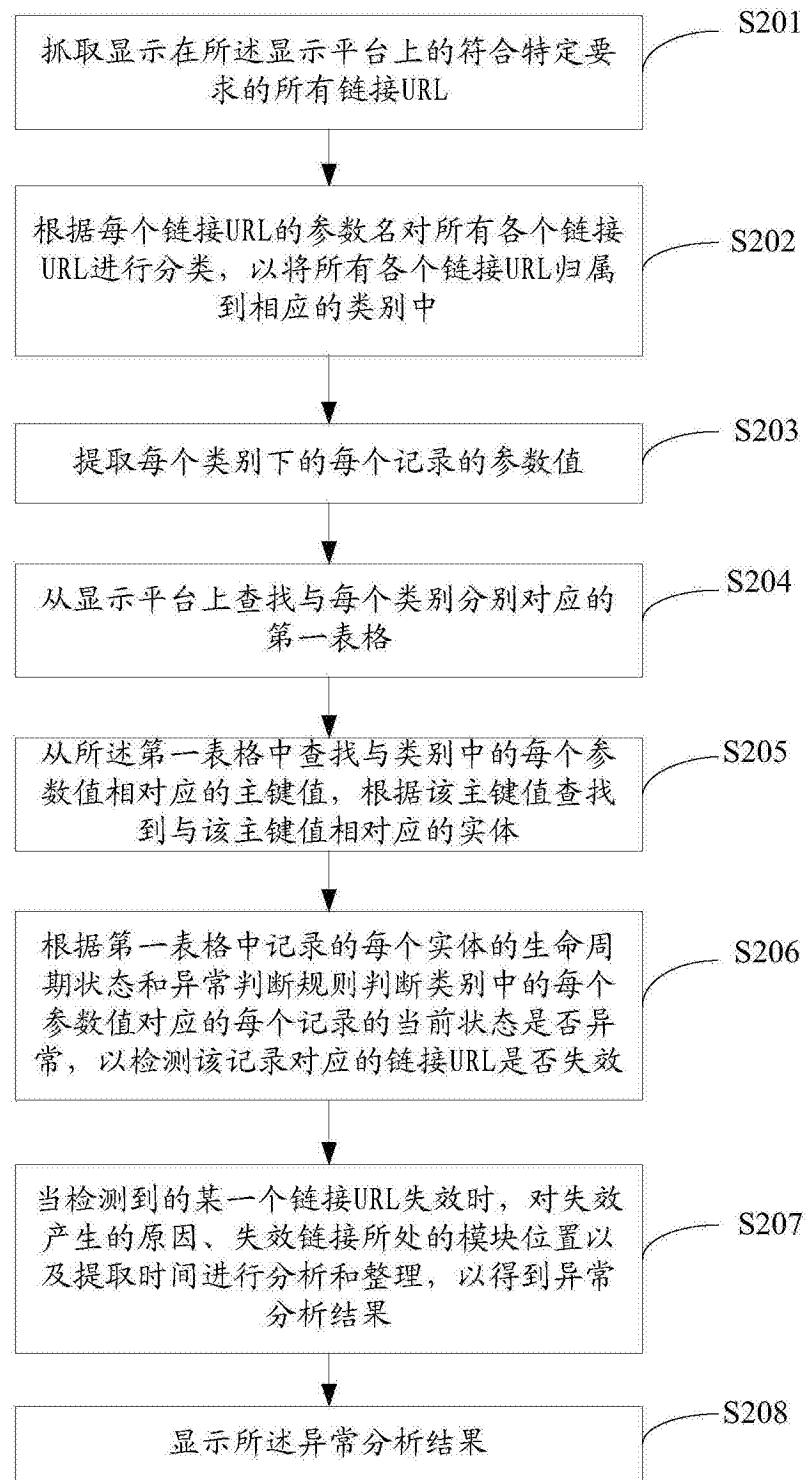


图2

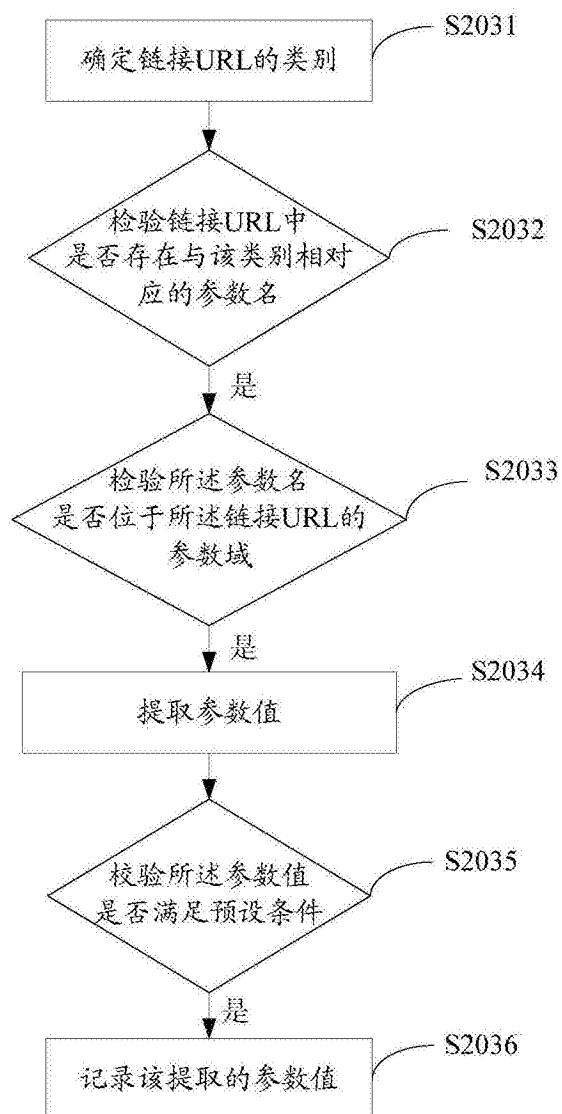


图3

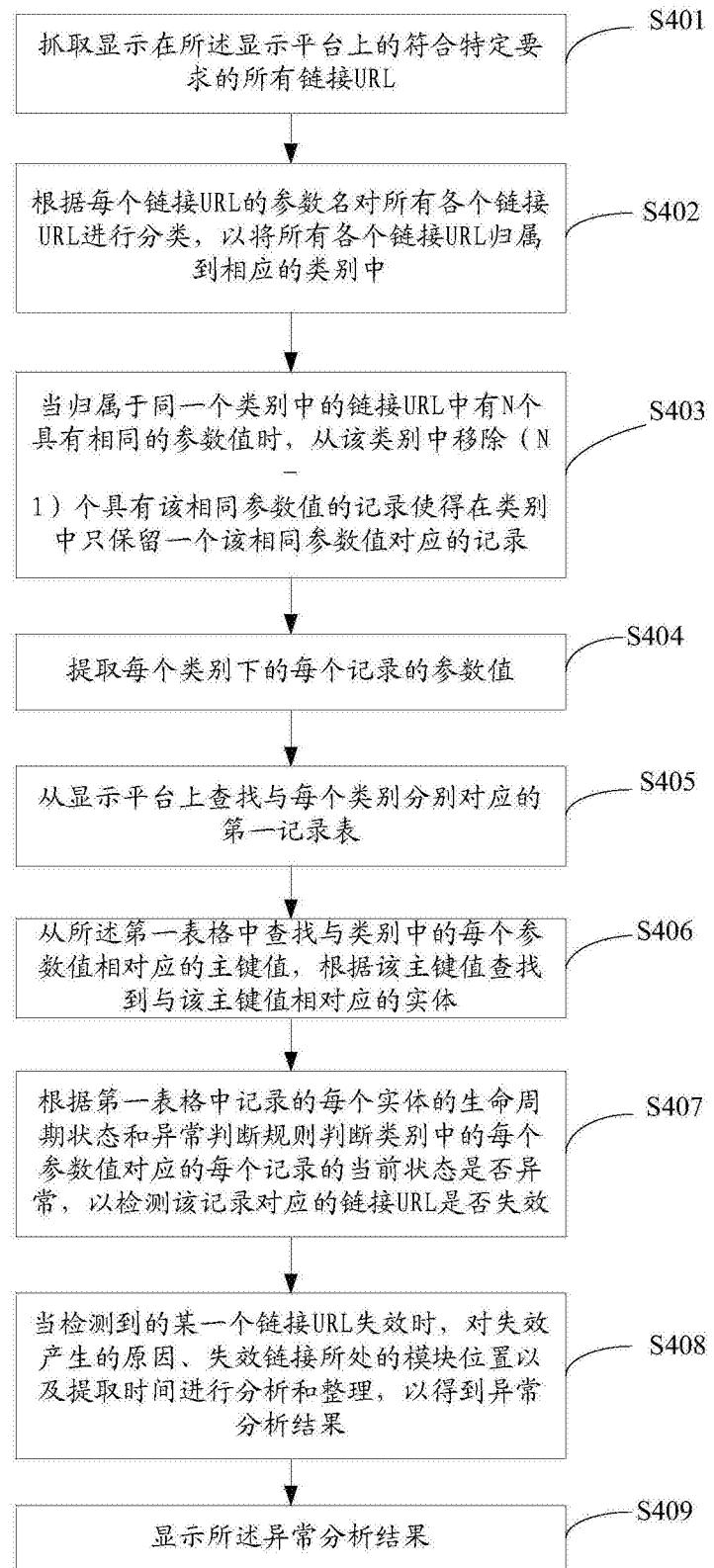


图4



图5



图6

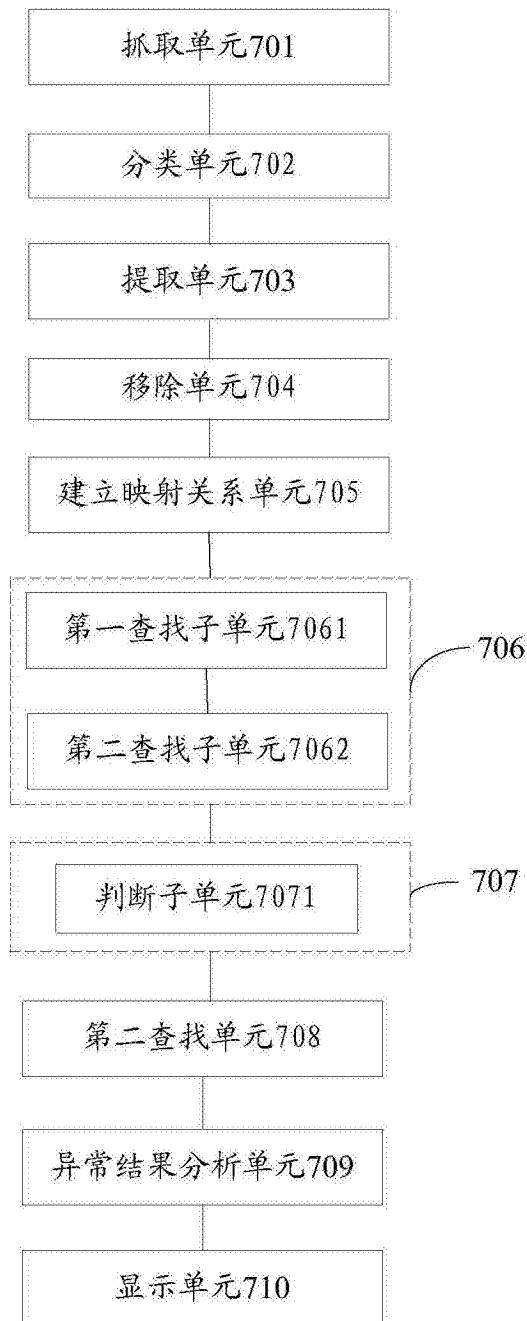


图7