



(12) 发明专利申请

(10) 申请公布号 CN 104834685 A

(43) 申请公布日 2015. 08. 12

(21) 申请号 201510184704. 3

(22) 申请日 2015. 04. 17

(71) 申请人 百度国际科技(深圳)有限公司

地址 518057 广东省深圳市南山区高新中二
道 5 号生产力大楼 D 单元三层 301

(72) 发明人 俞健

(74) 专利代理机构 北京汉昊知识产权代理事务
所(普通合伙) 11370

代理人 罗朋 赵晶

(51) Int. Cl.

G06F 17/30(2006. 01)

权利要求书2页 说明书16页 附图2页

(54) 发明名称

一种对评论类网页中的评论信息块进行处理
的方法与装置

(57) 摘要

本发明的目的是提供一种对评论类网页中的评论信息块进行处理的方法与装置。其中，本发明识别包含一个或多个评论信息块的评论类网页，并进一步基于无效内容规则库来识别并过滤这些评论信息块中的无效内容信息块，例如包含“顶”、“沙发”等无效的评论内容的帖子，进而将过滤后的评论类网页提供给用户。因此，本发明可以使用用户更方便、准确地获得有效的信息，节省用户的时间和精力，提高用户获取信息的效率。

开始

识别评论类网页，所述评论类网页包含一个或多个评论信息块 ↳ S1

基于无效内容规则库，识别所述评论信息块中的无效内
容信息块 ↳ S2

对所述无效内容信息块进行过滤，以获得过滤后的评论类网页 ↳ S3

将所述过滤后的评论类网页提供给用户 ↳ S4

结束

1. 一种对评论类网页中的评论信息块进行处理的方法，其中，该方法包括：
 - 识别评论类网页，所述评论类网页包含一个或多个评论信息块；
 - 基于无效内容规则库，识别所述评论信息块中的无效内容信息块；
 - 对所述无效内容信息块进行过滤，以获得过滤后的评论类网页；
 - 将所述过滤后的评论类网页提供给用户。
2. 根据权利要求 1 所述的方法，其中，所述识别评论类网页的步骤具体包括：
 - 根据评论类网页的网页结构特征信息，识别所述评论类网页。
3. 根据权利要求 1 或 2 所述的方法，其中，所述识别评论类网页的步骤具体包括：
 - 识别当前网页所属的评论类站点；
 - 根据所述评论类站点中评论类网页的网页结构特征信息，识别所述评论类网页。
4. 根据权利要求 1 所述的方法，其中，所述识别评论类网页的步骤具体包括：
 - 根据评论类网页的网页标识信息，识别所述评论类网页。
5. 根据权利要求 1 至 4 中任一项所述的方法，其中，所述无效内容规则库包括一个或多个无效内容规则，所述无效内容规则包括以下至少任一项：
 - 无效的评论内容；
 - 无效的评论者名称；
 - 无效的评论者等级；
 - 无效的评论时间。
6. 根据权利要求 5 所述的方法，其中，所述无效内容规则由所述用户配置。
7. 根据权利要求 6 所述的方法，其中，所述无效内容规则库为网络设备的网络数据库；其中，该方法还包括：
 - 根据多个用户所配置的无效内容规则，建立或更新所述网络数据库。
8. 根据权利要求 1 至 7 中任一项所述的方法，其中，所述无效内容规则库为浏览器的本地数据库；
其中，该方法还包括：
 - 从其他设备获取所述无效内容规则库的更新信息，以更新所述本地数据库。
9. 根据权利要求 1 至 8 中任一项所述的方法，其中，该方法还包括：
 - 统计所述无效内容信息块的类别以及其中每个类别所对应的无效内容信息块的数量；
其中，所述将所述过滤后的评论类网页提供给用户的步骤还包括：
 - 将所述类别以及所述每个类别所对应的无效内容信息块的数量一并提供给用户。
10. 一种对评论类网页中的评论信息块进行处理的装置，其中，该装置包括：
 - 用于识别评论类网页的装置，所述评论类网页包含一个或多个评论信息块；
 - 用于基于无效内容规则库，识别所述评论信息块中的无效内容信息块的装置；
 - 用于对所述无效内容信息块进行过滤，以获得过滤后的评论类网页的装置；
 - 用于将所述过滤后的评论类网页提供给用户的装置。
11. 根据权利要求 10 所述的装置，其中，所述识别评论类网页的装置具体用于：
 - 根据评论类网页的网页结构特征信息，识别所述评论类网页。
12. 根据权利要求 10 或 11 所述的装置，其中，所述识别评论类网页的装置具体用于：

- 识别当前网页所属的评论类站点；
- 根据所述评论类站点中评论类网页的网页结构特征信息，识别所述评论类网页。

13. 根据权利要求 10 所述的装置，其中，所述识别评论类网页的装置具体用于：
- 根据评论类网页的网页标识信息，识别所述评论类网页。

14. 根据权利要求 10 至 13 中任一项所述的装置，其中，所述无效内容规则库包括一个或多个无效内容规则，所述无效内容规则包括以下至少任一项：

- 无效的评论内容；
- 无效的评论者名称；
- 无效的评论者等级；
- 无效的评论时间。

15. 根据权利要求 14 所述的装置，其中，所述无效内容规则由所述用户配置。

16. 根据权利要求 15 所述的装置，其中，所述无效内容规则库为网络设备的网络数据库；
其中，该装置还包括：

- 用于根据多个用户所配置的无效内容规则，建立或更新所述网络数据库的装置。

17. 根据权利要求 10 至 16 中任一项所述的装置，其中，所述无效内容规则库为浏览器的本地数据库；
其中，该装置还包括：

- 用于从其他设备获取所述无效内容规则库的更新信息，以更新所述本地数据库的装置。

18. 根据权利要求 10 至 17 中任一项所述的装置，其中，该装置还包括：

- 用于统计所述无效内容信息块的类别以及其中每个类别所对应的无效内容信息块的数量的装置；
其中，所述将所述过滤后的评论类网页提供给用户的装置还用于：

- 将所述类别以及所述每个类别所对应的无效内容信息块的数量一并提供给用户。

19. 一种浏览器，包括如权利要求 10 至 18 中任一项所述的对评论类网页中的评论信息块进行处理的装置。

20. 一种网络设备，包括如权利要求 10 至 18 中任一项所述的对评论类网页中的评论信息块进行处理的装置。

一种对评论类网页中的评论信息块进行处理的方法与装置

技术领域

[0001] 本发明涉及互联网技术领域，尤其涉及一种对评论类网页中的评论信息块进行处理的技术。

背景技术

[0002] 当前，百度贴吧、天涯社区等评论网站非常流行。用户可以在这些网站中发表评论主题、回复评论内容。例如，一个用户在天涯社区发起一个评论主题，随后，该用户以及其他用户可以在该主题下的一个或多个网页中回复评论内容。在本文中，评论内容（即通常所称的帖子）所在的页面分块可以被称为评论信息块。并且，包含这些评论信息块的网页可以被称为评论类网页。

[0003] 然而，这些评论内容中通常存在大量无效的水帖，例如“顶”、“沙发”等。用户想在包含大量水帖的网页中找到有效帖子或自己需要的帖子通常需要花费较大的时间和精力。

[0004] 显然，现有技术仅简单地为用户提供评论类网页，但并未对其中无效的评论信息块进行任何特殊的处理。因此，用户很难准确地获得有效信息。

发明内容

[0005] 本发明的目的是提供一种对评论类网页中的评论信息块进行处理的方法与装置。

[0006] 根据本发明的一个方面，提供了一种对评论类网页中的评论信息块进行处理的方法，其中，该方法包括：

[0007] - 识别评论类网页，所述评论类网页包含一个或多个评论信息块；

[0008] - 基于无效内容规则库，识别所述评论信息块中的无效内容信息块；

[0009] - 对所述无效内容信息块进行过滤，以获得过滤后的评论类网页；

[0010] - 将所述过滤后的评论类网页提供给用户。

[0011] 根据本发明的另一个方面，还提供了一种对评论类网页中的评论信息块进行处理的装置，其中，该装置包括：

[0012] - 用于识别评论类网页的装置，所述评论类网页包含一个或多个评论信息块；

[0013] - 用于基于无效内容规则库，识别所述评论信息块中的无效内容信息块的装置；

[0014] - 用于对所述无效内容信息块进行过滤，以获得过滤后的评论类网页的装置；

[0015] - 用于将所述过滤后的评论类网页提供给用户的装置。

[0016] 与现有技术相比，本发明识别包含一个或多个评论信息块的评论类网页，并进一步基于无效内容规则库来识别并过滤这些评论信息块中的无效内容信息块，例如包含“顶”、“沙发”等无效的评论内容的帖子，进而将过滤后的评论类网页提供给用户。因此，本发明可以使用户更方便、准确地获得有效的信息，节省用户的时间和精力，提高用户获取信息的效率。

附图说明

[0017] 通过阅读参照以下附图所作的对非限制性实施例所作的详细描述，本发明的其它特征、目的和优点将会变得更明显：

[0018] 图 1 示出根据本发明一个实施例的一种对评论类网页中的评论信息块进行处理的方法流程图；

[0019] 图 2 示出根据本发明一个实施例的一种对评论类网页中的评论信息块进行处理的装置示意图。

[0020] 附图中相同或相似的附图标记代表相同或相似的部件。

具体实施方式

[0021] 在更加详细地讨论示例性实施例之前应当提到的是，一些示例性实施例被描述成作为流程图描绘的处理或方法。虽然流程图将各项操作描述成顺序的处理，但是其中的许多操作可以被并行地、并发地或者同时实施。此外，各项操作的顺序可以被重新安排。当其操作完成时所述处理可以被终止，但是还可以具有未包括在附图中的附加步骤。所述处理可以对应于方法、函数、规程、子例程、子程序等等。

[0022] 在上下文中所称“计算机设备”，也称为“电脑”，是指可以通过运行预定程序或指令来执行数值计算和 / 或逻辑计算等预定处理过程的智能电子设备，其可以包括处理器与存储器，由处理器执行在存储器中预存的程序指令来执行预定处理过程，或是由 ASIC、FPGA、DSP 等硬件执行预定处理过程，或是由上述二者组合来实现。计算机设备包括但不限于服务器、个人电脑、笔记本电脑、平板电脑、智能手机等。

[0023] 所述计算机设备例如包括用户设备与网络设备。其中，所述用户设备包括但不限于智能手机、PDA、PC、笔记本电脑等；所述网络设备包括但不限于单个网络服务器、多个网络服务器组成的服务器组或基于云计算 (Cloud Computing) 的由大量计算机或网络服务器构成的云，其中，云计算是分布式计算的一种，由一群松散耦合的计算机集组成的一个超级虚拟计算机。其中，所述计算机设备可单独运行来实现本发明，也可接入网络并通过与网络中的其他计算机设备的交互操作来实现本发明。其中，所述计算机设备所处的网络包括但不限于互联网、广域网、城域网、局域网、VPN 网络等。

[0024] 需要说明的是，所述用户设备、网络设备和网络等仅为举例，其他现有的或今后可能出现的计算机设备或网络如可适用于本发明，也应包含在本发明保护范围以内，并以引用方式包含于此。

[0025] 本文后面所讨论的方法（其中一些通过流程图示出）可以通过硬件、软件、固件、中间件、微代码、硬件描述语言或者其任意组合来实施。当用软件、固件、中间件或微代码来实施时，用以实施必要任务的程序代码或代码段可以被存储在机器或计算机可读介质（比如存储介质）中。（一个或多个）处理器可以实施必要的任务。

[0026] 这里所公开的具体结构和功能细节仅仅是代表性的，并且是用于描述本发明的示例性实施例的目的。但是本发明可以通过许多替换形式来具体实现，并且不应当被解释成仅仅受限于这里所阐述的实施例。

[0027] 应当理解的是，虽然在这里可能使用了术语“第一”、“第二”等等来描述各个单元，但是这些单元不应当受这些术语限制。使用这些术语仅仅是为了将一个单元与另一个单元进行区分。举例来说，在不背离示例性实施例的范围的情况下，第一单元可以被称为第二单

元，并且类似地第二单元可以被称为第一单元。这里所使用的术语“和 / 或”包括其中一个或更多所列出的相关联项目的任意和所有组合。

[0028] 这里所使用的术语仅仅是为了描述具体实施例而不意图限制示例性实施例。除非上下文明确地另有所指，否则这里所使用的单数形式“一个”、“一项”还意图包括复数。还应当理解的是，这里所使用的术语“包括”和 / 或“包含”规定所陈述的特征、整数、步骤、操作、单元和 / 或组件的存在，而不排除存在或添加一个或更多其他特征、整数、步骤、操作、单元、组件和 / 或其组合。

[0029] 还应当提到的是，在一些替换实现方式中，所提到的功能 / 动作可以按照不同于附图中标示的顺序发生。举例来说，取决于所涉及的功能 / 动作，相继示出的两幅图实际上可以基本上同时执行或者有时可以按照相反的顺序来执行。

[0030] 下面结合附图对本发明作进一步详细描述。

[0031] 本发明可由用户设备实现。具体地，用户设备识别评论类网页，所述评论类网页包含一个或多个评论信息块；接着，用户设备基于无效内容规则库，识别该（等）评论信息块中的无效内容信息块；随后，用户设备对所识别的无效内容信息块进行过滤，以获得过滤后的评论类网页；接着，用户设备将过滤后的评论类网页提供给用户。

[0032] 优选地，本发明的方案可以由一种安装并运行于用户设备中的具有浏览器性征的应用来实现。在此，所述具有浏览器性征的应用典型地如浏览器，其包括运行于 PC、笔记本电脑的各种浏览器应用程序，以及运行于移动终端的各种浏览器 APP 或具有浏览器功能的各种 APP。进一步地，本发明的方案可以作为一个功能模块集成在前述具有浏览器性征的应用中。

[0033] 其中，移动终端包括但不限于任何一种基于智能操作系统的手持式电子产品，其可与用户通过键盘、虚拟键盘、触摸板、触摸屏以及声控设备等输入设备来进行人机交互，诸如智能手机、平板电脑等。其中，智能操作系统包括但不限于任何通过向移动设备提供各种移动应用来丰富设备功能的操作系统，诸如安卓（Android）、IOS、Windows Phone 等。

[0034] 此外，本发明也可以由网络设备实现。具体地，网络设备识别评论类网页，所述评论类网页包含一个或多个评论信息块；接着，网络设备基于无效内容规则库，识别该（等）评论信息块中的无效内容信息块；随后，网络设备对所识别的无效内容信息块进行过滤，以获得过滤后的评论类网页；接着，网络设备将过滤后的评论类网页提供给用户。

[0035] 在此，网络设备包括但不限于各种在用户设备与网站服务器之间进行数据中转的中转网络设备。典型地，中转网络设备例如代理服务器、转码服务器等。

[0036] 为简单说明起见，本发明多以浏览器对评论类网页中的评论信息块进行处理进行举例。本领域技术人员应能理解，网络设备或其它设备对评论类网页中的评论信息块进行处理的方案，也应包含在本发明保护范围以内，并在此以引用方式包含于此。

[0037] 图 1 示出根据本发明一个实施例的方法流程图，其具体示出一种对评论类网页中的评论信息块进行处理的过程。在步骤 S1 中，浏览器识别评论类网页，所述评论类网页包含一个或多个评论信息块；在步骤 S2 中，浏览器基于无效内容规则库，识别该（等）评论信息块中的无效内容信息块；在步骤 S3 中，浏览器对所识别的无效内容信息块进行过滤，以获得过滤后的评论类网页；在步骤 S4 中，浏览器将过滤后的评论类网页提供给用户。

[0038] 具体地，在步骤 S1 中，浏览器识别评论类网页，所述评论类网页包含一个或多个

评论信息块。

[0039] 在此,应对评论类网页做广义理解,其不仅指通常所认为的就一评论主题回复评论内容的网页,诸如百度贴吧、天涯社区中各评论主题下的评论页面,还应包括其他任何具有该相同表现形式的交互类网页,只要该交互类网页满足以下形式要件:具有一个主题以及对应于该主题的一个或多个回复,具体如各种问答类网页,如百度知道中的各问答页面。

[0040] 评论信息块意指网页中包含有文字或图片等评论信息的分区或分块。评论信息块例如可以包括评论者名称(ID)、评论者头像、评论内容(帖子)等信息。

[0041] 浏览器至少可以基于以下3种方式识别评论类网页:

[0042] 1) 浏览器根据评论类网页的网页结构特征信息,识别评论类网页。

[0043] 在此,网页结构特征信息包括但不限于各种从结构或架构的角度体现网页特征的信息。

[0044] 评论类网页的网页结构特征信息主要通过块类别来体现,例如网页代码中<div class="atl-con-bd clearfix">、<div class="atl-reply">、<div class="core_reply_j_lz1_wrapper">、<div class="core_reply_tail">等块类别信息。评论类网页的网页结构特征信息可以通过预先分析各评论类网页的网页结构来确定。

[0045] 据此,浏览器可以基于预先分析确定的评论类网页的网页结构特征信息扫描网页的代码,如果一网页的网页代码中包含前述任一网页结构特征信息,则该网页属于评论类网页;如果一网页的网页代码中不包含前述任一网页结构特征信息,则该网页不属于评论类网页。

[0046] 预先分析确定的评论类网页的网页结构特征信息可以被保存在网页结构特征信息库中。该网页结构特征信息库可以从网络设备等其他设备获得或更新。

[0047] 2) 浏览器识别当前网页所属的评论类站点;接着,浏览器根据该评论类站点中评论类网页的网页结构特征信息,识别评论类网页。

[0048] 其中,评论类站点意指评论类网页所归属的站点。评论类站点例如百度贴吧、天涯论坛等。

[0049] 在此,属于评论类站点的网页并不一定属于评论类网页。例如,百度贴吧中“文科吧”、“樱桃小丸子吧”这些“吧”的首页以及百度贴吧的首页不一定包含评论信息块,因此,这些网页不一定属于评论类网页。

[0050] 当前网页所属的评论类站点可以由浏览器基于评论类站点的站点标识信息来识别。所述站点标识信息包括但不限于各种可以用于标识特定站点的信息。评论类站点的站点标识信息例如<http://bbs.tianya.cn>、<http://tieba.baidu.com>等网址,以及天涯论坛、百度贴吧等网页标题。评论类站点的站点标识信息可以通过预先分析评论类站点的网址、网页标题等站点标识信息来确定。

[0051] 接着,浏览器根据所识别的评论类站点中评论类网页的网页结构特征信息,识别评论类网页。

[0052] 在此,浏览器可以通过查询评论类站点特征信息关联表来确定评论类站点中评论类网页的网页结构特征信息。所述评论类站点特征信息关联表中保存有多个评论类站点以及每个评论类站点中评论类网页的网页结构特征信息。所述评论类站点特征信息关联表可以从网络设备等其他设备获得或更新。所述评论类站点特征信息关联表如以下表1所示:

[0053]

评论类站点	该评论类站点中评论类网页的网页结构特征信息
天涯论坛	<div class="atl-con-bd clearfix"> <div class="atl-reply">
百度贴吧	<div class="core_reply_j_lzl_wrapper"> <div class="core_reply_tail">
19楼

[0054] 表 1

[0055] 浏览器根据通过查询表 1 所确定的评论类站点中评论类网页的网页结构特征信息,扫描属于该评论类站点的网页的网页代码,如果该网页的网页代码中包含相应的网页结构特征信息,则该网页属于评论类网页;如果该网页的网页代码中不包含相应的网页结构特征信息,则该网页不属于评论类网页。

[0056] 根据本发明一个优选实施例,浏览器识别当前网页所属的评论类站点为天涯论坛;接着,浏览器根据该评论类站点中评论类网页的网页结构特征信息<div class = "atl-con-bd clearfix"> 和 <div class = "atl-reply">,确定当前网页是否属于评论类网页。

[0057] 浏览器先识别当前网页所属的评论类站点,再根据所识别的评论类站点中评论类网页的网页结构特征信息来识别评论类网页的过程可以独立实施,该过程也可以被认为是浏览器根据评论类网页的网页结构特征信息来识别评论类网页的一种优选实施方式。

[0058] 在此,对于不属于评论类站点的网页,浏览器将不再根据评论类网页的网页结构特征信息对该网页进行识别。因此,浏览器识别当前网页所属的评论类站点的步骤可以避免对所有网页进行其是否属于评论类网页的识别,从而提高了计算处理的效率。并且,浏览器可以根据所识别的评论类站点准确地确定相应的评论类网页的网页结构特征信息,从而可以避免根据预先分析确定的所有评论类网页的网页结构特征信息来识别评论类网页,这也可以提高计算处理的效率。

[0059] 3) 浏览器根据评论类网页的网页标识信息,识别评论类网页。

[0060] 其中,网页标识信息包括但不限于各种可以用于标识特定网页的信息。网页标识信息例如网址、网页标题等。

[0061] 由于评论类网页的网址、网页标题等网页标识信息具有一定的命名规则,因此评论类网页的网页标识信息可以通过预先分析相应的命名规则来确定。

[0062] 例如,百度贴吧中评论类网页的网页标识信息例如网址“<http://tieba.baidu.com/p/XXXXXXXXXX>”(XXXXXXXXXX 可以为任何数字字符)、网页标题“YYYY_ZZZZ 吧_百度贴吧”(YYYY、ZZZZ 均可以为任何数字或文字字符)等。

[0063] 据此,浏览器可以将预先分析确定的评论类网页的网页标识信息与当前网页的网

页标识信息相比较,如果相同,则该网页属于评论类网页;如果不同,则该网页不属于评论类网页。

[0064] 预先分析确定的评论类网页的网页标识信息可以被保存在网页标识信息库中。该网页标识信息库可以从网络设备等其他设备获得或更新。

[0065] 本领域技术人员应能理解,前述3种识别评论类网页的方式仅为举例,而不应被视为对本发明的任何限制,任何现有其它方案或今后可能出现的识别评论类网页的方式如可适用于本发明,也应包含在本发明保护范围以内,并在此以引用方式包含于此。

[0066] 对于不属于评论类网页的网页,浏览器并不做特殊处理,因此,该类网页不属于本文的讨论范围。

[0067] 在步骤S2中,浏览器基于无效内容规则库,识别评论类网页的一个或多个评论信息块中的无效内容信息块。

[0068] 其中,无效内容信息块意指包含无效的或被认为无效的信息的评论信息块。

[0069] 所述无效内容规则库包括一个或多个无效内容规则。浏览器可以对评论类网页的网页代码进行分析,以获得评论类网页的一个或多个评论信息块,进而识别该(等)评论信息块中符合无效内容规则的无效内容信息块。

[0070] 在此,无效内容规则包括但不限于以下至少任一项:

[0071] 1) 无效的评论内容。

[0072] 在此,无效的评论内容例如“沙发”、“顶”、“路过”等。

[0073] 如果一评论信息块的内容仅包括无效的评论内容而不包括任何其他信息,该评论信息块可以被识别为无效内容信息块。或者,浏览器可以按照更严厉地方式识别无效内容信息块。例如,只要一评论信息块的内容中包括无效的评论内容,则该评论信息块即被识别为无效内容信息块。

[0074] 2) 无效的评论者名称。

[0075] 在此,评论者名称意指评论者的ID。

[0076] 无效的评论者例如经常发布广告信息的评论者,又如被封禁的评论者,还如用户不希望看到其评论的评论者。

[0077] 如果一评论信息块的评论者名称与无效的评论者名称一致,该评论信息块可以被识别为无效内容信息块。

[0078] 3) 无效的评论者等级。

[0079] 在此,评论者等级包括但不限于评论者经验值、评论者头衔等表示评论者资历的信息。

[0080] 无效的评论者等级可以为任意等级。例如,无效的评论者等级可以被配置为评论者经验值低于500或菜鸟评论者等。

[0081] 如果一评论信息块的评论者等级对应于无效的评论者等级,则该评论信息块被识别为无效内容信息块。

[0082] 4) 无效的评论时间。

[0083] 在此,评论时间意指评论者发出评论的时间。

[0084] 无效的评论时间可以为任何时间。例如,无效的评论时间可以被配置为2015年3月24日10:00之前,或者被配置为2015年3月24日10:00至10:30的区间之内。

[0085] 如果一评论信息块的评论时间对应于无效的评论时间，则该评论信息块被识别为无效内容信息块。

[0086] 在此，无效内容规则库可以为浏览器的本地数据库，也可以为网络设备的网络数据库。进一步地，当无效内容规则库为浏览器的本地数据库，其也可以根据网络设备的网络数据库来建立和 / 或更新。

[0087] 例如，无效内容规则可以为开发人员预配置的。无效内容规则库可以在浏览器安装过程中被保存在用户设备中。或者，浏览器也可以向网络设备发送无效内容规则查询 / 获取请求，并获得该网络设备响应该请求为浏览器返回的一个或多个无效内容规则。

[0088] 优选地，无效内容规则可以由用户配置。

[0089] 例如，浏览器可以为用户呈现用于配置无效内容规则的图形用户界面。用户可以在该图形用户界面中对无效的评论内容、无效的评论者名称、无效的评论者等级、无效的评论时间等无效内容规则进行配置。浏览器可以保存用户所配置的无效内容规则。

[0090] 更优选地，当无效内容规则库还包括网络设备的网络数据库时，网络设备可以根据多个用户所配置的无效内容规则，建立或更新该网络数据库。

[0091] 例如，各浏览器可以将用户配置的无效内容规则发送给网络设备，相应地，网络设备获得多个用户配置的无效内容规则，以建立或更新网络数据库（无效内容规则库）。

[0092] 其中，所述建立网络数据库意指从无到有的存储无效内容规则，以构建网络数据库的过程。

[0093] 所述更新网络数据库意指在网络数据库建立后，对其中存储的无效内容规则进行更新的过程，该更新可以包括无效内容规则的新增、删除和更改。

[0094] 对于浏览器的本地数据库，浏览器可以从其他设备获取无效内容规则库的更新信息，以更新该本地数据库。

[0095] 其中，所述其他设备包括但不限于网络设备以及手机、ipad 等其他用户设备。

[0096] 例如，浏览器可以从网络设备获取无效内容规则库的更新信息，该更新信息包括新增的三个无效内容规则；接着，浏览器将这三个无效内容规则增加至浏览器的本地数据库（无效内容规则库）中。

[0097] 在此，浏览器从其他设备获取无效内容规则库的更新信息与浏览器获取用户配置的无效内容规则并非相互排斥的。例如，在基于用户配置的无效内容规则建立本地数据库后，浏览器再基于从其他设备获取的无效内容规则库的更新信息来更新本地数据库。

[0098] 从更新频率角度而言，浏览器可以按照多种方式更新本地数据库。例如，浏览器可以在每次被开启时均向其他设备请求无效内容规则库的更新信息并更新本地数据库。再如，浏览器可以在预定的时间（如每周五 20:00）向其他设备请求无效内容规则库的更新信息并更新本地数据库。

[0099] 在步骤 S3 中，浏览器对所识别的无效内容信息块进行过滤，以获得过滤后的评论类网页。在步骤 S4 中，浏览器将过滤后的评论类网页提供给用户。

[0100] 在此，浏览器可以按照以下两种方式对无效内容信息块进行过滤：

[0101] 1、浏览器对无效内容信息块进行过滤。

[0102] 在此，浏览器可以过滤掉整个无效内容信息块，也即，包含评论者名称（ID）、评论者头像、评论内容等信息的完整页面块被过滤。随后，被提供给用户的过滤后的评论类网页

中将不再包含关于无效内容信息块的任何信息。

[0103] 其中,浏览器对无效内容信息块进行过滤的方式包括但不限于以下 2 种:

[0104] 1) 浏览器可以将无效内容信息块的块 (div) 属性修改为隐藏 (hidden)。

[0105] 进一步地,浏览器还可以在评论类网页中被过滤的无效内容信息块的位置处,如右上角,增加一交互按钮,如隐藏 / 展开按钮。当用户触发该展开按钮时,被隐藏的无效内容信息块将被呈现给用户;该按钮随后被显示为隐藏,当用户再次触发该隐藏按钮时,该无效内容信息块将被隐藏。

[0106] 2) 浏览器可以删除无效内容信息块对应的全部网页代码。

[0107] 例如,一个评论类网页包括 20 个评论信息块。其中,第 5 个评论信息块被识别为无效内容信息块。在该无效内容信息块被过滤后,过滤后的评论类网页中仅 19 个评论信息块对用户可见。

[0108] 2、浏览器仅对无效内容信息块中无效的评论内容进行过滤。

[0109] 在此,浏览器也可以仅将对应于无效的评论内容的网页代码进行删除。随后,被提供给用户的过滤后的评论类网页中将不再包含无效的评论内容。而该无效的评论内容所属的无效内容信息块中的评论者名称 (ID)、评论者头像等其他信息仍然对用户可见。

[0110] 进一步地,浏览器还可以在无效内容信息块中增加一交互按钮,如隐藏 / 展开按钮。当用户触发该展开按钮时,被隐藏的无效内容信息将被呈现给用户;该按钮随后被显示为隐藏,当用户再次触发该隐藏按钮时,无效内容信息将被隐藏。

[0111] 本领域技术人员应能理解,前述两种对无效内容信息块进行过滤的方式仅为举例,而不应被视为对本发明的任何限制,任何现有或今后可能出现的其他对无效内容信息块进行过滤的方式如可适用于本发明,也应包含在本发明保护范围以内,并在此以引用方式包含于此。

[0112] 优选地,浏览器还可以统计无效内容信息块的类别以及其中每个类别所对应的无效内容信息块的数量。并且,浏览器还可以将所统计的无效内容信息块的类别以及每个类别所对应的无效内容信息块的数量与过滤后的评论类网页一并提供给用户。

[0113] 其中,无效内容信息块的类别可以基于多种方式确定。例如,无效内容信息块所对应的无效内容规则可以作为无效内容信息块的类别。

[0114] 例如,在包括 15 个评论信息块的评论类网页中,浏览器统计得到以下信息:内容为“沙发”的无效内容信息块有 2 个;内容为“顶”的无效内容信息块有 3 个。浏览器则将前述统计信息与过滤后的评论类网页一并提供给用户。

[0115] 在此,浏览器可以将无效内容信息块的类别以及每个类别所对应的无效内容信息块的数量与评论类网页同时提供给用户。或者,无效内容信息块的类别以及每个类别所对应的无效内容信息块的数量也可以在用户浏览到评论类网页中某一位置时通过弹框的方式提供给用户,如当页面滑动至邻近无效内容信息块的区域时,浏览器通过弹框向用户提示邻近无效内容信息块的类别及数量。

[0116] 图 2 示出根据本发明一个实施例的装置示意图,其具体示出一种对评论类网页中的评论信息块进行处理的装置,也即过滤装置 10。如图 2 所示,过滤装置 10 被装置于浏览器中,并具体包括装置 11-14。

[0117] 装置 11 识别评论类网页,所述评论类网页包含一个或多个评论信息块(为便于区

分,以下将装置 11 称为评论网页识别装置 11);装置 12 基于无效内容规则库,识别该(等)评论信息块中的无效内容信息块(为便于区分,以下将装置 12 称为无效内容识别装置 12);装置 13 对所识别的无效内容信息块进行过滤,以获得过滤后的评论类网页(为便于区分,以下将装置 13 称为无效内容过滤装置 13);装置 14 将过滤后的评论类网页提供给用户(为便于区分,以下将装置 14 称为评论网页提供装置 14)。

[0118] 具体地,评论网页识别装置 11 识别评论类网页,所述评论类网页包含一个或多个评论信息块。

[0119] 在此,应对评论类网页做广义理解,其不仅指通常所认为的就一评论主题回复评论内容的网页,诸如百度贴吧、天涯社区中各评论主题下的评论页面,还应包括其他任何具有该相同表现形式的交互类网页,只要该交互类网页满足以下形式要件:具有一个主题以及对应于该主题的一个或多个回复,具体如各种问答类网页,如百度知道中的各问答页面。

[0120] 评论信息块意指网页中包含有文字或图片等评论信息的分区或分块。评论信息块例如可以包括评论者名称(ID)、评论者头像、评论内容(帖子)等信息。

[0121] 评论网页识别装置 11 至少可以基于以下 3 种方式识别评论类网页:

[0122] 1) 评论网页识别装置 11 根据评论类网页的网页结构特征信息,识别评论类网页。

[0123] 在此,网页结构特征信息包括但不限于各种从结构或架构的角度体现网页特征的信息。

[0124] 评论类网页的网页结构特征信息主要通过块类别来体现,例如网页代码中<div class="atl-con-bd clearfix">、<div class="atl-reply">、<div class="core_reply_j_lz1_wrapper">、<div class="core_reply_tail">等块类别信息。评论类网页的网页结构特征信息可以通过预先分析各评论类网页的网页结构来确定。

[0125] 据此,评论网页识别装置 11 可以基于预先分析确定的评论类网页的网页结构特征信息扫描网页的代码,如果一网页的网页代码中包含前述任一网页结构特征信息,则该网页属于评论类网页;如果一网页的网页代码中不包含前述任一网页结构特征信息,则该网页不属于评论类网页。

[0126] 预先分析确定的评论类网页的网页结构特征信息可以被保存在网页结构特征信息库中。该网页结构特征信息库可以从网络设备等其他设备获得或更新。

[0127] 2) 评论网页识别装置 11 识别当前网页所属的评论类站点;接着,评论网页识别装置 11 根据该评论类站点中评论类网页的网页结构特征信息,识别评论类网页。

[0128] 其中,评论类站点意指评论类网页所归属的站点。评论类站点例如百度贴吧、天涯论坛等。

[0129] 在此,属于评论类站点的网页并不一定属于评论类网页。例如,百度贴吧中“文科吧”、“樱桃小丸子吧”这些“吧”的首页以及百度贴吧的首页不一定包含评论信息块,因此,这些网页不一定属于评论类网页。

[0130] 当前网页所属的评论类站点可以由评论网页识别装置 11 基于评论类站点的站点标识信息来识别。所述站点标识信息包括但不限于各种可以用于标识特定站点的信息。评论类站点的站点标识信息例如<http://bbs.tianya.cn>、<http://tieba.baidu.com> 等网址,以及天涯论坛、百度贴吧等网页标题。评论类站点的站点标识信息可以通过预先分析评论类站点的网址、网页标题等站点标识信息来确定。

[0131] 接着,评论网页识别装置 11 根据所识别的评论类站点中评论类网页的网页结构特征信息,识别评论类网页。

[0132] 在此,评论网页识别装置 11 可以通过查询评论类站点特征信息关联表来确定评论类站点中评论类网页的网页结构特征信息。所述评论类站点特征信息关联表中保存有多个评论类站点以及每个评论类站点中评论类网页的网页结构特征信息。所述评论类站点特征信息关联表可以从网络设备等其他设备获得或更新。所述评论类站点特征信息关联表如以上表 1 所示。

[0133] 评论网页识别装置 11 根据通过查询表 1 所确定的评论类站点中评论类网页的网页结构特征信息,扫描属于该评论类站点的网页的网页代码,如果该网页的网页代码中包含相应的网页结构特征信息,则该网页属于评论类网页;如果该网页的网页代码中不包含相应的网页结构特征信息,则该网页不属于评论类网页。

[0134] 根据本发明一个优选实施例,评论网页识别装置 11 识别当前网页所属的评论类站点为天涯论坛;接着,评论网页识别装置 11 根据该评论类站点中评论类网页的网页结构特征信息<div class = "atl-con-bd clearfix">和<div class = "atl-reply">,确定当前网页是否属于评论类网页。

[0135] 评论网页识别装置 11 先识别当前网页所属的评论类站点,再根据所识别的评论类站点中评论类网页的网页结构特征信息来识别评论类网页的操作可以独立实施,该操作也可以被认为是评论网页识别装置 11 根据评论类网页的网页结构特征信息来识别评论类网页的一种优选实施方式。

[0136] 在此,对于不属于评论类站点的网页,评论网页识别装置 11 将不再根据评论类网页的网页结构特征信息对该网页进行识别。因此,评论网页识别装置 11 识别当前网页所属的评论类站点的操作可以避免对所有网页进行其是否属于评论类网页的识别,从而提高了计算处理的效率。并且,评论网页识别装置 11 可以根据所识别的评论类站点准确地确定相应的评论类网页的网页结构特征信息,从而可以避免根据预先分析确定的所有评论类网页的网页结构特征信息来识别评论类网页,这也提高了计算处理的效率。

[0137] 3) 评论网页识别装置 11 根据评论类网页的网页标识信息,识别评论类网页。

[0138] 其中,网页标识信息包括但不限于各种可以用于标识特定网页的信息。网页标识信息例如网址、网页标题等。

[0139] 由于评论类网页的网址、网页标题等网页标识信息具有一定的命名规则,因此评论类网页的网页标识信息可以通过预先分析相应的命名规则来确定。

[0140] 例如,百度贴吧中评论类网页的网页标识信息例如网址“<http://tieba.baidu.com/p/XXXXXXXXXX>”(XXXXXXXXXX 可以为任何数字字符)、网页标题“YYYY_ZZZZ 吧_百度贴吧”(YYYY、ZZZZ 均可以为任何数字或文字字符)等。

[0141] 据此,评论网页识别装置 11 可以将预先分析确定的评论类网页的网页标识信息与当前网页的网页标识信息相比较,如果相同,则该网页属于评论类网页;如果不同,则该网页不属于评论类网页。

[0142] 预先分析确定的评论类网页的网页标识信息可以被保存在网页标识信息库中。该网页标识信息库可以从网络设备等其他设备获得或更新。

[0143] 本领域技术人员应能理解,前述 3 种识别评论类网页的方式仅为举例,而不应被

视为对本发明的任何限制,任何现有其它方案或今后可能出现的识别评论类网页的方式如可适用于本发明,也应包含在本发明保护范围以内,并在此以引用方式包含于此。

[0144] 对于不属于评论类网页的网页,过滤装置 10 并不做特殊处理,因此,该类网页不属于本文的讨论范围。

[0145] 随后,无效内容识别装置 12 基于无效内容规则库,识别评论类网页的一个或多个评论信息块中的无效内容信息块。

[0146] 其中,无效内容信息块意指包含无效的或被认为无效的信息的评论信息块。

[0147] 所述无效内容规则库包括一个或多个无效内容规则。无效内容识别装置 12 可以对评论类网页的网页代码进行分析,以获得评论类网页的一个或多个评论信息块,进而识别该(等)评论信息块中符合无效内容规则的无效内容信息块。

[0148] 在此,无效内容规则包括但不限于以下至少任一项:

[0149] 1) 无效的评论内容。

[0150] 在此,无效的评论内容例如“沙发”、“顶”、“路过”等。

[0151] 如果一评论信息块的内容仅包括无效的评论内容而不包括任何其他信息,该评论信息块可以被识别为无效内容信息块。或者,无效内容识别装置 12 可以按照更严厉地方式识别无效内容信息块。例如,只要一评论信息块的内容中包括无效的评论内容,则该评论信息块即被识别为无效内容信息块。

[0152] 2) 无效的评论者名称。

[0153] 在此,评论者名称意指评论者的 ID。

[0154] 无效的评论者例如经常发布广告信息的评论者,又如被封禁的评论者,还如用户不希望看到其评论的评论者。

[0155] 如果一评论信息块的评论者名称与无效的评论者名称一致,该评论信息块可以被识别为无效内容信息块。

[0156] 3) 无效的评论者等级。

[0157] 在此,评论者等级包括但不限于评论者经验值、评论者头衔等表示评论者资历的信息。

[0158] 无效的评论者等级可以为任意等级。例如,无效的评论者等级可以被配置为评论者经验值低于 500 或菜鸟评论者等。

[0159] 如果一评论信息块的评论者等级对应于无效的评论者等级,则该评论信息块被识别为无效内容信息块。

[0160] 4) 无效的评论时间。

[0161] 在此,评论时间意指评论者发出评论的时间。

[0162] 无效的评论时间可以为任何时间。例如,无效的评论时间可以被配置为 2015 年 3 月 24 日 10 :00 之前,或者被配置为 2015 年 3 月 24 日 10 :00 至 10 :30 的区间之内。

[0163] 如果一评论信息块的评论时间对应于无效的评论时间,则该评论信息块被识别为无效内容信息块。

[0164] 在此,无效内容规则库可以为浏览器的本地数据库,也可以为网络设备的网络数据库。进一步地,当无效内容规则库为浏览器的本地数据库,其也可以根据网络设备的网络数据库来建立和 / 或更新。

[0165] 例如,无效内容规则可以为开发人员预配置的。无效内容规则库可以在浏览器安装过程中被保存在用户设备中。或者,无效内容识别装置 12 或过滤装置 10 中的其他特定装置也可以向网络设备发送无效内容规则查询 / 获取请求,并获得该网络设备响应该请求返回的一个或多个无效内容规则。

[0166] 优选地,无效内容规则可以由用户配置。

[0167] 例如,过滤装置 10 可以为用户呈现用于配置无效内容规则的图形用户界面。用户可以在该图形用户界面中对无效的评论内容、无效的评论者名称、无效的评论者等级、无效的评论时间等无效内容规则进行配置。过滤装置 10 可以保存用户所配置的无效内容规则。

[0168] 更优选地,当无效内容规则库还包括网络设备的网络数据库时,网络设备还可以包括一网络数据库构建装置(网络设备、网络数据库构建装置均未在图 2 中示出),网络数据库构建装置可以根据多个用户所配置的无效内容规则,建立或更新该网络数据库。

[0169] 例如,各浏览器的过滤装置 10 可以将用户配置的无效内容规则发送给网络设备,相应地,网络设备的网络数据库构建装置获得多个用户配置的无效内容规则,以建立或更新网络数据库(无效内容规则库)。

[0170] 其中,所述建立网络数据库意指从无到有的存储无效内容规则,以构建网络数据库的过程。

[0171] 所述更新网络数据库意指在网络数据库建立后,对其中存储的无效内容规则进行更新的过程,该更新可以包括无效内容规则的新增、删除和更改。

[0172] 对于浏览器的本地数据库,浏览器还可以包括一本本地数据库更新装置(图 2 中未示出),本地数据库更新装置可以从其他设备获取无效内容规则库的更新信息,以更新该本地数据库。

[0173] 其中,所述其他设备包括但不限于网络设备以及手机、ipad 等其他用户设备。

[0174] 例如,本地数据库更新装置可以从网络设备获取无效内容规则库的更新信息,该更新信息包括新增的三个无效内容规则;接着,本地数据库更新装置将这三个无效内容规则增加至浏览器的本地数据库(无效内容规则库)中。

[0175] 在此,本地数据库更新装置从其他设备获取无效内容规则库的更新信息与过滤装置 10 获取用户配置的无效内容规则并非相互排斥的。例如,在过滤装置 10 基于用户配置的无效内容规则建立本地数据库后,本地数据库更新装置再基于从其他设备获取的无效内容规则库的更新信息来更新本地数据库。

[0176] 从更新频率角度而言,本地数据库更新装置可以按照多种方式更新本地数据库。例如,本地数据库更新装置可以在每次被开启时均向其他设备请求无效内容规则库的更新信息并更新本地数据库。再如,本地数据库更新装置可以在预定的时间(如每周五 20:00)向其他设备请求无效内容规则库的更新信息并更新本地数据库。

[0177] 随后,无效内容过滤装置 13 对所识别的无效内容信息块进行过滤,以获得过滤后的评论类网页。接着,评论网页提供装置 14 将过滤后的评论类网页提供给用户。

[0178] 在此,无效内容过滤装置 13 可以按照以下两种方式对无效内容信息块进行过滤:

[0179] 1、无效内容过滤装置 13 对无效内容信息块进行过滤。

[0180] 在此,无效内容过滤装置 13 可以过滤掉整个无效内容信息块,也即,包含评论者名称(ID)、评论者头像、评论内容等信息的完整页面块被过滤。随后,被提供给用户的过滤

后的评论类网页中将不再包含关于无效内容信息块的任何信息。

[0181] 其中,无效内容过滤装置 13 对无效内容信息块进行过滤的方式包括但不限于以下 2 种:

[0182] 1) 无效内容过滤装置 13 可以将无效内容信息块的块 (div) 属性修改为隐藏 (hidden)。

[0183] 进一步地,无效内容过滤装置 13 还可以在评论类网页中被过滤的无效内容信息块的位置处,如右上角,增加一交互按钮,如隐藏 / 展开按钮。当用户触发该展开按钮时,被隐藏的无效内容信息块将被呈现给用户;该按钮随后被显示为隐藏,当用户再次触发该隐藏按钮时,该无效内容信息块将被隐藏。

[0184] 2) 无效内容过滤装置 13 可以删除无效内容信息块对应的全部网页代码。

[0185] 例如,一个评论类网页包括 20 个评论信息块。其中,第 5 个评论信息块被识别为无效内容信息块。在该无效内容信息块被过滤后,过滤后的评论类网页中仅 19 个评论信息块对用户可见。

[0186] 2、无效内容过滤装置 13 仅对无效内容信息块中无效的评论内容进行过滤。

[0187] 在此,无效内容过滤装置 13 也可以仅将对应于无效的评论内容的网页代码进行删除。随后,被提供给用户的过滤后的评论类网页中将不再包含无效的评论内容。而该无效的评论内容所属的无效内容信息块中的评论者名称 (ID)、评论者头像等其他信息仍然对用户可见。

[0188] 进一步地,无效内容过滤装置 13 还可以在无效内容信息块中增加一交互按钮,如隐藏 / 展开按钮。当用户触发该展开按钮时,被隐藏的无效内容信息将被呈现给用户;该按钮随后被显示为隐藏,当用户再次触发该隐藏按钮时,无效内容信息将被隐藏。

[0189] 本领域技术人员应能理解,前述两种对无效内容信息块进行过滤的方式仅为举例,而不应被视为对本发明的任何限制,任何现有或今后可能出现的其他对无效内容信息块进行过滤的方式如可适用于本发明,也应包含在本发明保护范围以内,并在此以引用方式包含于此。

[0190] 优选地,过滤装置 10 还可以包括一无效内容统计装置(图 2 中未示出),无效内容统计装置统计无效内容信息块的类别以及其中每个类别所对应的无效内容信息块的数量。并且,评论网页提供装置 14 还可以将所统计的无效内容信息块的类别以及每个类别所对应的无效内容信息块的数量与过滤后的评论类网页一并提供给用户。

[0191] 其中,无效内容信息块的类别可以基于多种方式确定。例如,无效内容信息块所对应的无效内容规则可以作为无效内容信息块的类别。

[0192] 例如,在包括 15 个评论信息块的评论类网页中,无效内容统计装置统计得到以下信息:内容为“沙发”的无效内容信息块有 2 个;内容为“顶”的无效内容信息块有 3 个。评论网页提供装置 14 则将前述统计信息与过滤后的评论类网页一并提供给用户。

[0193] 在此,评论网页提供装置 14 可以将无效内容信息块的类别以及每个类别所对应的无效内容信息块的数量与评论类网页同时提供给用户。或者,无效内容信息块的类别以及每个类别所对应的无效内容信息块的数量也可以在用户浏览到评论类网页中某一位置时通过弹框的方式提供给用户,如当页面滑动至邻近无效内容信息块的区域时,评论网页提供装置 14 通过弹框向用户提示邻近无效内容信息块的类别及数量。

[0194] 需要注意的是，本发明可在软件和 / 或软件与硬件的组合体中被实施，例如，本发明的各个装置可采用专用集成电路 (ASIC) 或任何其他类似硬件设备来实现。在一个实施例中，本发明的软件程序可以通过处理器执行以实现上文所述步骤或功能。同样地，本发明的软件程序（包括相关的数据结构）可以被存储到计算机可读记录介质中，例如，RAM 存储器，磁或光驱动器或软磁盘及类似设备。另外，本发明的一些步骤或功能可采用硬件来实现，例如，作为与处理器配合从而执行各个步骤或功能的电路。

[0195] 对于本领域技术人员而言，显然本发明不限于上述示范性实施例的细节，而且在不背离本发明的精神或基本特征的情况下，能够以其他的具体形式实现本发明。因此，无论从哪一点来看，均应将实施例看作是示范性的，而且是非限制性的，本发明的范围由所附权利要求而不是上述说明限定，因此旨在将落在权利要求的等同要件的含义和范围内的所有变化涵括在本发明内。不应将权利要求中的任何附图标记视为限制所涉及的权利要求。此外，显然“包括”一词不排除其他单元或步骤，单数不排除复数。系统权利要求中陈述的多个单元或装置也可以由一个单元或装置通过软件或者硬件来实现。第一，第二等词语用来表示名称，而并不表示任何特定的顺序。

[0196] 虽然前面特别示出并且描述了示例性实施例，但是本领域技术人员将会理解的是，在不背离权利要求书的精神和范围的情况下，在其形式和细节方面可以有所变化。这里所寻求的保护在所附权利要求书中做了阐述。在下列编号条款中规定了各个实施例的这些和其他方面：

- [0197] 1. 一种对评论类网页中的评论信息块进行处理的方法，其中，该方法包括：
 - [0198] - 识别评论类网页，所述评论类网页包含一个或多个评论信息块；
 - [0199] - 基于无效内容规则库，识别所述评论信息块中的无效内容信息块；
 - [0200] - 对所述无效内容信息块进行过滤，以获得过滤后的评论类网页；
 - [0201] - 将所述过滤后的评论类网页提供给用户。
 - [0202] 2. 根据条款 1 所述的方法，其中，所述识别评论类网页的步骤具体包括：
 - [0203] - 根据评论类网页的网页结构特征信息，识别所述评论类网页。
 - [0204] 3. 根据条款 1 或 2 所述的方法，其中，所述识别评论类网页的步骤具体包括：
 - [0205] - 识别当前网页所属的评论类站点；
 - [0206] - 根据所述评论类站点中评论类网页的网页结构特征信息，识别所述评论类网页。
 - [0207] 4. 根据条款 1 所述的方法，其中，所述识别评论类网页的步骤具体包括：
 - [0208] - 根据评论类网页的网页标识信息，识别所述评论类网页。
 - [0209] 5. 根据条款 1 至 4 中任一项所述的方法，其中，所述无效内容规则库包括一个或多个无效内容规则，所述无效内容规则包括以下至少任一项：
 - [0210] - 无效的评论内容；
 - [0211] - 无效的评论者名称；
 - [0212] - 无效的评论者等级；
 - [0213] - 无效的评论时间。
 - [0214] 6. 根据条款 5 所述的方法，其中，所述无效内容规则由所述用户配置。
 - [0215] 7. 根据条款 6 所述的方法，其中，所述无效内容规则库为网络设备的网络数据库；
 - [0216] 其中，该方法还包括：

- [0217] - 根据多个用户所配置的无效内容规则,建立或更新所述网络数据库。
- [0218] 8. 根据条款 1 至 7 中任一项所述的方法,其中,所述无效内容规则库为浏览器的本地数据库;
- [0219] 其中,该方法还包括:
- [0220] - 从其他设备获取所述无效内容规则库的更新信息,以更新所述本地数据库。
- [0221] 9. 根据条款 1 至 8 中任一项所述的方法,其中,该方法还包括:
- [0222] - 统计所述无效内容信息块的类别以及其中每个类别所对应的无效内容信息块的数量;
- [0223] 其中,所述将所述过滤后的评论类网页提供给用户的步骤还包括:
- [0224] - 将所述类别以及所述每个类别所对应的无效内容信息块的数量一并提供给用户。
- [0225] 10. 一种对评论类网页中的评论信息块进行处理的装置,其中,该装置包括:
- [0226] - 用于识别评论类网页的装置,所述评论类网页包含一个或多个评论信息块;
- [0227] - 用于基于无效内容规则库,识别所述评论信息块中的无效内容信息块的装置;
- [0228] - 用于对所述无效内容信息块进行过滤,以获得过滤后的评论类网页的装置;
- [0229] - 用于将所述过滤后的评论类网页提供给用户的装置。
- [0230] 11. 根据条款 10 所述的装置,其中,所述识别评论类网页的装置具体用于:
- [0231] - 根据评论类网页的网页结构特征信息,识别所述评论类网页。
- [0232] 12. 根据条款 10 或 11 所述的装置,其中,所述识别评论类网页的装置具体用于:
- [0233] - 识别当前网页所属的评论类站点;
- [0234] - 根据所述评论类站点中评论类网页的网页结构特征信息,识别所述评论类网页。
- [0235] 13. 根据条款 10 所述的装置,其中,所述识别评论类网页的装置具体用于:
- [0236] - 根据评论类网页的网页标识信息,识别所述评论类网页。
- [0237] 14. 根据条款 10 至 13 中任一项所述的装置,其中,所述无效内容规则库包括一个或多个无效内容规则,所述无效内容规则包括以下至少任一项:
- [0238] - 无效的评论内容;
- [0239] - 无效的评论者名称;
- [0240] - 无效的评论者等级;
- [0241] - 无效的评论时间。
- [0242] 15. 根据条款 14 所述的装置,其中,所述无效内容规则由所述用户配置。
- [0243] 16. 根据条款 15 所述的装置,其中,所述无效内容规则库为网络设备的网络数据库;
- [0244] 其中,该装置还包括:
- [0245] - 用于根据多个用户所配置的无效内容规则,建立或更新所述网络数据库的装置。
- [0246] 17. 根据条款 10 至 16 中任一项所述的装置,其中,所述无效内容规则库为浏览器的本地数据库;
- [0247] 其中,该装置还包括:
- [0248] - 用于从其他设备获取所述无效内容规则库的更新信息,以更新所述本地数据库的装置。

[0249] 18. 根据条款 10 至 17 中任一项所述的装置, 其中, 该装置还包括 :

[0250] - 用于统计所述无效内容信息块的类别以及其中每个类别所对应的无效内容信息块的数量的装置 ;

[0251] 其中, 所述将所述过滤后的评论类网页提供给用户的装置还用于 :

[0252] - 将所述类别以及所述每个类别所对应的无效内容信息块的数量一并提供给用户。

[0253] 19. 一种浏览器, 包括如条款 10 至 18 中任一项所述的对评论类网页中的评论信息块进行处理的装置。

[0254] 20. 一种网络设备, 包括如条款 10 至 18 中任一项所述的对评论类网页中的评论信息块进行处理的装置。

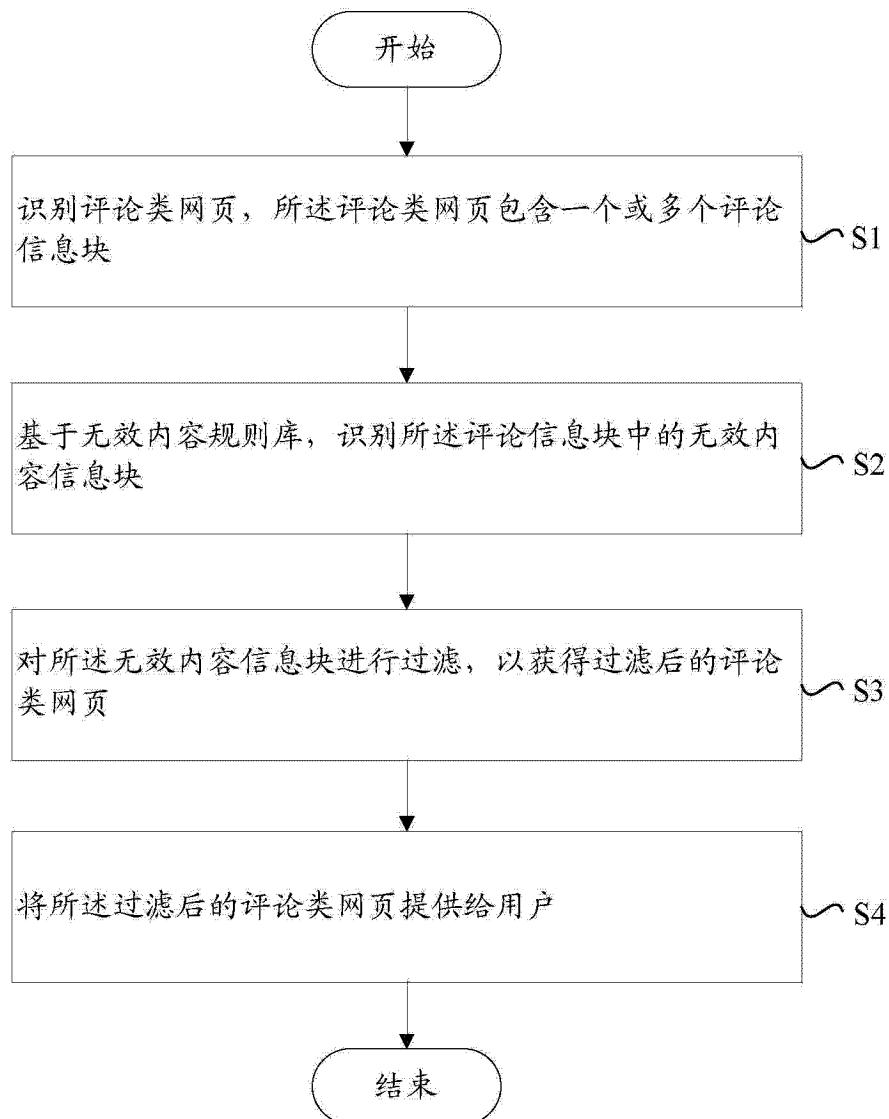


图 1

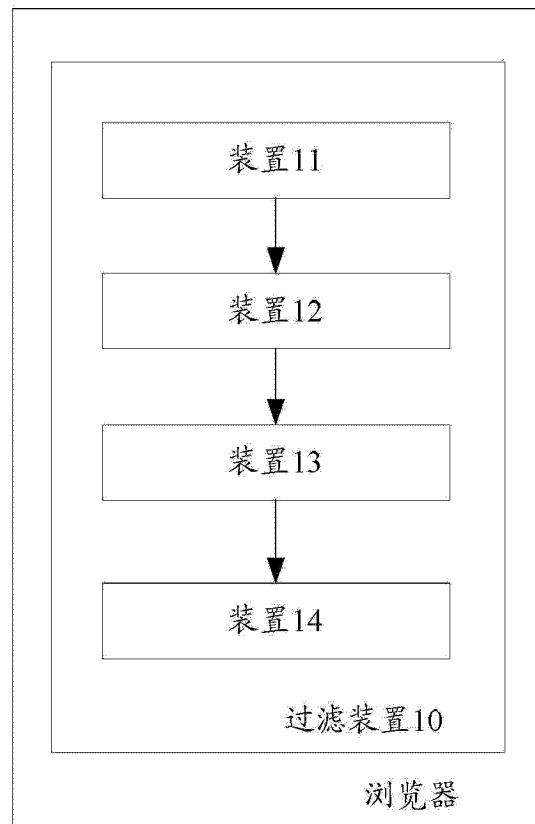


图 2