(12) 按照专利合作条约所公布的国际申请

(19) 世界知识产权组织
国 际 局

(43) 国际公布日
2016 年 10 月 27 日 (27.10.2016)

WIPO | PCT

(10) 国际公布号
WO 2016/169193 A1

(51) 国际专利分类号:
*G06F 17/30* (2006.01)

(21) 国际申请号: PCT/CN2015/089545

(22) 国际申请日: 2015 年 9 月 14 日 (14.09.2015)

(25) 申请语言: 中文

(26) 公布语言: 中文

(30) 优先权:
201510202474.9 2015 年 4 月 24 日 (24.04.2015) CN

(71) 申请人: **百度在线网络技术（北京）有限公司 (BAIDU ONLINE NETWORK TECHNOLOGY (BEIJING) CO., LTD.)** [CN/CN]; 中国北京市海淀区上地十街 10 号百度大厦三层, Beijing 100085 (CN)。

(72) 发明人: **庄馨 (ZHUANG, Xin)**; 中国北京市海淀区上地十街 10 号百度大厦, Beijing 100085 (CN)。 **田天 (TIAN, Tian)**; 中国北京市海淀区上地十街 10 号百度大厦, Beijing 100085 (CN)。 **朱军 (ZHU, Jun)**; 中国北京市海淀区上地十街 10 号百度大厦, Beijing 100085 (CN)。 **夏粉 (XIA, Fen)**; 中国北京市海淀区上地十街 10 号百度大厦, Beijing 100085 (CN)。 **张潼 (ZHANG, Tong)**; 中国北京市海淀区上地十街 10 号百度大厦, Beijing 100085 (CN)。

(74) 代理人: **北京英赛嘉华知识产权代理有限责任公司 (INSIGHT INTELLECTUAL PROPERTY LIMITED)**; 中国北京市海淀区知春路甲 48 号盈都大厦 A 座 19A, Beijing 100098 (CN)。

(81) 指定国 (除另有指明，要求每一种可提供的国家保护): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW。

(84) 指定国 (除另有指明，要求每一种可提供的地区保护): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), 欧亚 (AM, AZ, BY, KG, KZ, RU, TJ, TM), 欧洲 (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG)。

**本国际公布**:

— 包括国际检索报告(条约第 21 条(3))。

(54) **Title**: METHOD AND APPARATUS FOR DETECTING CHEATED CLICKS

(54) **发明名称** : 用于检测点击作弊的方法及装置

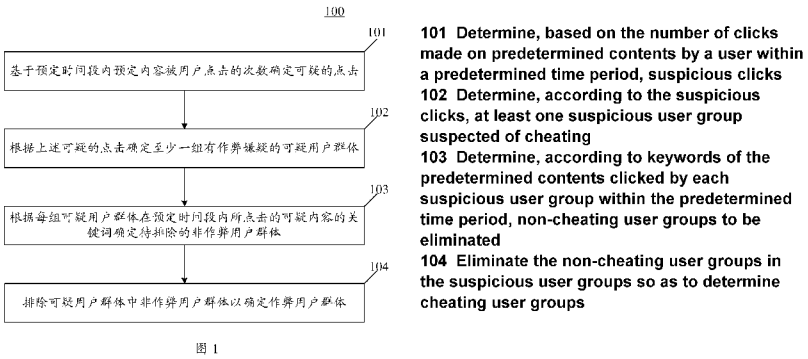



图 1

(57) **Abstract**: A method and apparatus for detecting cheated clicks. The method comprises: determining, based on the number of clicks made on predetermined contents by a user within a predetermined time period, suspicious clicks (101); determining, according to the suspicious clicks, at least one suspicious user group suspected of cheating (102); determining, according to keywords of the predetermined contents clicked by each suspicious user group within the predetermined time period, non-cheating user groups to be eliminated (103); and eliminating the non-cheating user groups in the suspicious user groups so as to determine cheating user groups (104). Clicks on predetermined contents can be monitored, thereby improving the detection efficiency and detection accuracy of cheating clicks, and reducing the waste of time and resources.

(57) **摘要**: 用于检测点击作弊的方法及装置。所述方法包括：基于预定时间段内预定内容被用户点击的次数确定可疑的点击(101)；根据所述可疑的点击确定至少一组有作弊嫌疑的可疑用户群体(102)；根据每组所述可疑用户群体在所述预定时间段内所点击的可疑内容的关键词确定待排除的非作弊用户群体(103)；以及排除所述可疑用户群体中非作弊用户群体以确定作弊用户群体(104)。实现了对预定内容的点击进行监控，提高了检测作弊点击的效率和检测精度，减少了时间和资源的浪费。

WO 2016/169193 A1

# 说 明 书

## 用于检测点击作弊的方法及装置

### 相关申请的交叉引用

本申请要求申请日为 2015 年 4 月 24 日，申请号为 201510202474.9，发明名称为“用于检测点击作弊的方法及装置”的中国专利申请的优先权，其全部内容作为整体并入本申请中。

### 技术领域

本申请涉及网络技术领域，尤其涉及用于检测点击作弊的方法及装置。

### 背景技术

随着移动互联网众包技术的迅猛发展，一类通过众包网站发布任务，组织网民人工点击特定广告，并给予一定费用的作弊方式开始兴起。这类作弊行为均是由人工真实的触发，具有很强的隐蔽性，很难用传统的方法进行检测。

目前，现有的对广告众包作弊的检测方法包括：①通过建立描述单个点击者行为特征的规则，来发现可疑点击者，进而判断作弊行为。这种技术的缺陷在于：因为众包作弊来源为真实的用户点击而非来自机器，因此行为有很大的随机性，很难用基于单个用户或广告主建立的规则去判断。②通过观察一个广告主的点击流量来判断是否有针对它的作弊行为。这种技术的缺陷在于：由于众包作弊来自真实用户，当他们感受到作弊被检出后，可以迅速调整自身行为方式，从而使之前的规则失效。③通过寻找点击日志中行为一致的一批点击者，来判断作弊行为。这种技术的缺陷在于：该方法针对人工众包作弊检测实用性不大。

### 发明内容

为了解决上述问题，本申请提供了一种用于检测点击作弊的方法及装置，解决了现有技术中对作弊点击的检测效率不高，检测精度低的技术问题。

第一方面，本申请提供了一种用于检测点击作弊的方法，所述方法包括：基于预定时间段内预定内容被用户点击的次数确定可疑的点击；根据所述可疑的点击确定至少一组有作弊嫌疑的可疑用户群体；根据每组所述可疑用户群体在所述预定时间段内所点击的可疑内容的关键词确定待排除的非作弊用户群体；以及排除所述可疑用户群体中非作弊用户群体以确定作弊用户群体。

在某些实施方式中，所述基于预定时间段内预定内容被用户点击的次数确定可疑的点击，包括：获取所述预定时间段内每个预定内容被用户点击的次数；判断所述每个预定内容被点击的次数是否满足预定条件；将被点击的次数满足预定条件的预定内容所对应的点击确定为可疑的点击。

在某些实施方式中，判断所述被点击的次数是否满足预定条件，包括：判断所述被点击的次数是否大于等于第一预定阈值，且小于等于第二预定阈值；如果是，确定所述被点击的次数满足预定条件。

在某些实施方式中，所述根据所述可疑的点击确定至少一组有作弊嫌疑的可疑用户群体，包括：获取所述可疑的点击的相关信息；基于所述相关信息确定至少一组可疑用户群体，其中，每组可疑用户群体在相同时间段内点击同一组内容。

在某些实施方式中，所述可疑的点击的相关信息包括以下至少一项：可疑的点击所对应的用户的标识信息；可疑的点击所对应的内容的标识信息；以及可疑的点击所对应的时刻。

在某些实施方式中，所述基于所述相关信息确定至少一组可疑用户群体，包括：基于所述相关信息对所述可疑的点击进行聚类，使得每个聚类中心对应的用户群体在相同时间段内点击同一组内容；将所述每个聚类中心对应的用户群体确定为一组可疑用户群体。

在某些实施方式中，所述根据每组所述可疑用户群体在所述预定时间段内所点击的可疑内容的关键词确定待排除的非作弊用户群体，

包括：获取每组所述可疑用户群体在所述预定时间段内所点击的可疑内容的关键词；基于所述关键词判断所述可疑内容是否为同类内容；如果是，将所述可疑内容对应的可疑用户群体确定为待排除的非作弊用户群体。

在某些实施方式中，所述基于所述关键词判断所述可疑内容是否为同类内容，包括：判断所述关键词中同类关键词所占比例是否大于等于预定比例；如果是，确定所述可疑内容为同类内容。

第二方面，本申请提供了一种用于检测点击作弊的装置，所述装置包括：第一确定单元，用于基于预定时间段内预定内容被用户点击的次数确定可疑的点击；第二确定单元，用于根据所述可疑的点击确定至少一组有作弊嫌疑的可疑用户群体；第三确定单元，用于根据每组所述可疑用户群体在所述预定时间段内所点击的可疑内容的关键词确定待排除的非作弊用户群体；以及第四确定单元，用于排除所述可疑用户群体中非作弊用户群体以确定作弊用户群体。

在某些实施方式中，所述第一确定单元包括：获取子单元，用于获取所述预定时间段内每个预定内容被用户点击的次数；判断子单元，用于判断所述每个预定内容被点击的次数是否满足预定条件；确定子单元，用于将被点击的次数满足预定条件的预定内容所对应的点击确定为可疑的点击。

在某些实施方式中，所述判断子单元配置用于：判断所述被点击的次数是否大于等于第一预定阈值，且小于等于第二预定阈值；如果是，确定所述被点击的次数满足预定条件。

在某些实施方式中，所述第二确定单元包括：信息获取子单元，用于获取所述可疑的点击的相关信息；用户群体确定子单元，用于基于所述相关信息确定至少一组可疑用户群体，其中，每组可疑用户群体在相同时间段内点击同一组内容。

在某些实施方式中，所述可疑的点击的相关信息包括以下至少一项：可疑的点击所对应的用户的标识信息；可疑的点击所对应的内容的标识信息；以及可疑的点击所对应的时刻。

在某些实施方式中，所述用户群体确定子单元配置用于：基于所

述相关信息对所述可疑的点击进行聚类，使得每个聚类中心对应的用户群体在相同时间段内点击同一组内容；将所述每个聚类中心对应的用户群体确定为一组可疑用户群体。

在某些实施方式中，所述第三确定单元包括：关键词获取子单元，用于获取每组所述可疑用户群体在所述预定时间段内所点击的可疑内容的关键词；类别判断子单元，用于基于所述关键词判断所述可疑内容是否为同类内容；待排除群体确定子单元，用于响应于可疑内容为同类内容，将所述可疑内容对应的可疑用户群体确定为待排除的非作弊用户群体。

在某些实施方式中，所述类别判断子单元配置用于：判断所述关键词中同类关键词所占比例是否大于等于预定比例；如果是，确定所述可疑内容为同类内容。

本申请提供的用于检测点击作弊的方法及装置，通过缩小检测范围，确定有作弊嫌疑的可疑用户群体，并根据可疑用户群体所点击的可疑内容的关键词将可疑用户群体中的非作弊用户群体排除，从而实现了对预定内容的点击进行监控，提高了检测作弊点击的效率和检测精度，减少了时间和资源的浪费。

## 附图说明

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，本申请的其它特征、目的和优点将会变得更明显：

图 1 是本申请实施例提供的用于检测点击作弊的方法的一个实施例的流程图；

图 2 是本申请实施例提供的确定可疑的点击的方法的一个实施例的流程图；

图 3 是本申请实施例提供的根据可疑的点击确定至少一组有作弊嫌疑的可疑用户群体的方法的一个实施例的流程图；

图 4 是本申请实施例提供的根据每组可疑用户群体在上述预定时间段内所点击的可疑内容的关键词确定待排除的非作弊用户群体的方法的一个实施例的流程图；

图 5 是本申请实施例提供的用于检测点击作弊的装置的一个实施例的结构示意图;

图 6 适于用来实现本申请实施例的终端设备或服务器的计算机系统的结构示意图。

## 具体实施方式

下面结合附图和实施例对本申请作进一步的详细说明。可以理解的是，此处所描述的具体实施例仅仅用于解释相关发明，而非对该发明的限定。另外还需要说明的是，为了便于描述，附图中仅示出了与有关发明相关的部分。

需要说明的是，在不冲突的情况下，本申请中的实施例及实施例中的特征可以相互组合。下面将参考附图并结合实施例来详细说明本申请。

请参考图 1，其示出了用于检测点击作弊的方法的一个实施例的流程 100。

如图 1 所示，在步骤 101 中，基于预定时间段内预定内容被用户点击的次数确定可疑的点击。

在本实施例中，预定内容为有可能涉及作弊点击的内容，诸如一些能够靠点击量获得利益的内容（如广告、投票以及社交网站等等），这些内容被点击的点击量越大，和其相关的受益人或单位获得的收益就越大。一般来说，利用众包平台进行作弊，通常是与预定内容相关的受益者通过任务发布者，在众包平台发布点击预定内容（如广告等）的任务，然后由任务发布者组织大批网民领取任务，网民通过点击预定内容完成任务，从而获得完成任务的奖励。

由此可见，通过众包平台进行作弊点击，点击时间一般都比较集中，通常发生在发布点击任务之后的一定时间段内。所以基本可以排除点击预定内容不够集中的时间段内包含作弊点击的可能。因此，预定时间段可以选取点击预定内容比较集中的一段时间段，具体来说，在一种实现方式中，可以获取对应于预定内容的点击在时间轴上的分布，取分布密度大于预定阈值的一段时间段作为预定时间段。在另一

种实现方式中，也可以获取每个时刻对预定内容的点击量，取连续每个时刻点击量均大于预定阈值的一段时间段作为预定时间段。可以理解，还可以有其它的选取预定时间段的实现方式，本申请对此方面不限定。

在本实施例中，为了能够提高检测作弊点击的效率，减少时间和资源的浪费，首先，可以先确定一些可能与作弊相关的点击作为可疑的点击，排除一些作弊的可能性较小的点击，从而减小了检测的范围。在后续的检测过程中，只在可疑点击的范围中进行检测。具体来说，可以基于预定时间段内预定内容被用户点击的次数确定可疑的点击。

接着，在步骤 102 中，根据上述可疑的点击确定至少一组有作弊嫌疑的可疑用户群体。

一般来说，进行作弊的任务发布者可能一次发布一个包含多个预定内容的点击任务到众包平台，例如，向众包平台发布一个包含 10 个广告的点击任务。进行作弊的网民在领取作弊点击任务后，通常会在一定时间段内将任务中的一组预定内容全部进行点击。因此，可以通过对可疑点击的相关信息（如点击时间，点击的内容等等）进行分析，确定执行该任务（点击对应于该任务的一组预定内容）的用户群体。将点击同一组预定内容的用户群体确定为具有作弊嫌疑的可疑用户群体。

因为，在一定时间段内，可能有多个任务发布者发布点击任务，所以，上述确定的可疑点击中可能包含多组对应于不同作弊点击任务的预定内容。其中，对应于同一个作弊点击任务的预定内容为一组预定内容。同时，每组预定内容也可以对应一组有作弊嫌疑的可疑用户群体。因此，上述确定的可疑点击对应至少一组有作弊嫌疑的可疑用户群体。

继而，在步骤 103 中，根据每组可疑用户群体在预定时间段内所点击的可疑内容的关键词确定待排除的非作弊用户群体。

一般来说，有些预定内容可能在一定时期较为热门，具有很大的用户点击量，或者某些用户在某个时间段内会对相关的一批预定内容感兴趣。例如，某个时期某区域爆发流感，那么该区域可能有大批居

民在网上搜索抗流感或者预防流感的药物等，这些居民点击的广告或者网站可能会有部分重叠，并且均与流感相关。又例如，某个季节比较适合去某地区旅游，那么在该季节可能有大量用户在网上搜索并点击与去该地区旅游相关的广告或者网站，这些用户点击的内容也可能有部分重叠，并且均与该地区的旅游相关。

上述这些预定内容并没有涉及作弊点击，而点击上述这些预定内容的用户也并非作弊用户群体，但有可能被确定为有作弊嫌疑的作弊用户群体，因此，需要将这些非作弊用户群体从有作弊嫌疑的可疑用户群体中找出并排除。

在本实施例中，可疑内容为可疑用户群体在上述预定时间段内所点击的预定内容。可以根据每组可疑用户群体所点击的可疑内容的关键词确定待排除的非作弊用户群体。如果某组可疑用户群体所点击的所有可疑内容的关键词近似度较高，则可以确定该用户群体为非作弊用户群体。

最后，在步骤 104 中，排除可疑用户群体中非作弊用户群体以确定作弊用户群体。

在本实施例中，将可疑用户群体中的非作弊用户群体加以排除，则将剩下的可疑用户群体确定为作弊用户群体。

本申请的上述实施例提供的方法，通过缩小检测范围，确定有作弊嫌疑的可疑用户群体，并根据可疑用户群体所点击的可疑内容的关键词将可疑用户群体中的非作弊用户群体排除，从而实现了对预定内容的点击进行监控，提高了检测作弊点击的效率和检测精度，减少了时间和资源的浪费。

进一步参考图 2，其示出了确定可疑的点击的方法的一个实施例的流程 200。

如图 2 所示，在步骤 201 中，获取上述预定时间段内每个预定内容被用户点击的次数。

在本实施例中，可以通过预定内容的点击日志来获取上述预定时间段内每个预定内容被用户点击的次数，可以理解，还可以通过其它的方式获取上述预定时间段内每个预定内容被用户点击的次数。本申

请对此方面不限定。

接着，在步骤 202 中，判断每个预定内容被点击的次数是否满足预定条件。

一般来说，如果一段时间内预定内容被点击的次数过小，如，小于一个预定阈值 a，那么该预定内容涉及作弊的可能性就很小。因为，作弊的目的就是为了增加点击量，如果该预定内容涉及了作弊，其被点击的次数一定不会太小。

而如果一段时间内预定内容被点击的次数过大，如，大于一个预定阈值 b，那么该预定内容涉及作弊的可能性也会很小。因为，作弊虽然能够增加点击量，但组织作弊的规模通常会比较有限，不可能达到一个过高的量级。例如，假设作弊点击能够增加点击量 1000，如果某预定内容被点击的次数为 10000，则可以判断该预定内容一定没有涉及作弊。因为，即使该预定内容涉及作弊，那么其对应的正常点击量也比作弊点击所能增加的点击量大接近一个数量级，所以，通过作弊增加点击量的意义并不大。

在本实施例中，预定条件为可能涉及作弊的预定内容被点击的次数所满足的条件，首先判断预定时间段内预定内容被点击的次数是否大于等于第一预定阈值，且小于等于第二预定阈值。如果该次数大于等于第一预定阈值且小于等于第二预定阈值，则确定该次数满足预定条件。

最后，在步骤 203 中，将被点击的次数满足预定条件的预定内容所对应的点击确定为可疑的点击。

在本实施例中，如果某预定内容被点击的次数满足预定条件，则该预定内容很有可能涉及作弊，将其对应的点击确定为可疑的点击。需要说明的是，可疑的点击并不意味着一定是作弊点击，因为，即使某内容涉及作弊，该内容也同样会被非作弊用户正常点击。

进一步参考图 3，其示出了根据可疑的点击确定至少一组有作弊嫌疑的可疑用户群体的方法的一个实施例的流程 300。

如图 3 所示，在步骤 301 中，获取可疑的点击的相关信息。

在本实施例中，可疑的点击的相关信息可以包括以下至少一项：

可疑的点击所对应的用户的标识信息；可疑的点击所对应的内容的标识信息；以及可疑的点击所对应的时刻。

具体来说，可疑的点击所对应的用户的标识信息可以是执行该可疑点击的用户的 MAC 地址，或者 IP 地址，或者终端设备（如手机、电脑等）的序列号等，本申请对可疑点击对应的用户标识信息的具体内容和形式不限定。可疑的点击所对应的内容标识信息可以是被可疑的点击所点击过的内容的名称，或者编号等用于标识或区分内容的信息，本申请对可疑点击对应的内容标识信息的具体内容和形式不限定。可疑的点击所对应的时刻可以是用户执行上述可疑的点击时对应的时刻。在本实施例中，可疑的点击的相关信息可以从点击日志中获得。可以理解，可疑的点击的相关信息也可以通过其它的方式获得，本申请对获得可疑的点击的相关信息的方式不限定。

接着，在步骤 302 中，基于相关信息确定至少一组可疑用户群体，其中，每组可疑用户群体在相同时间段内点击同一组内容。

在本实施例中，可以基于上述相关信息确定可疑用户群体，其中，可以有一组或多组可疑用户群体，每组可疑用户群体在相同时间段内点击同一组内容。

在本实施例的一种实现中，可以采用非参数化的聚类算法确定可疑用户群体，具体来说，首先基于上述相关信息对所有的可疑的点击进行聚类分析，使得每个聚类中心对应的用户群体在相同时间段内点击同一组内容。然后将每个聚类中心对应的用户群体确定为一组可疑用户群体。

请参考图 4，其示出了根据每组可疑用户群体在上述预定时间段内所点击的可疑内容的关键词确定待排除的非作弊用户群体的方法的一个实施例的流程 400。

如图 4 所示，在步骤 401 中，获取每组可疑用户群体在上述预定时间段内所点击的可疑内容的关键词。

在本实施例中，可疑内容为可疑用户群体在上述预定时间段内所点击的预定内容。需要说明的是，可疑用户群体中的用户在上述预定时间段内还可能点击其它不涉及作弊的内容，但这些不涉及作弊的内

容与可疑的点击无关，因此，不会被判定为可疑内容。

在本实施例中，可疑内容的关键词为最能够体现可疑内容各种特征的词。例如，对于一种药品的广告，其关键词可以是广告产品的类别（药品），该药品能够治疗的疾病名称，生产该药品的制药厂的厂名，该药品所含最重要的化学成分的名称等等。

在本实施例的一种实现中，可以对上述可疑内容进行内容解析，以获取其相关的关键词。在另一种实现中，还可以从对上述可疑内容的名称或者标识信息中获取其相关的关键词。可以理解，还可以有其它的获取可疑内容相关的关键词的方式，本申请对获取可疑内容相关的关键词的方式不限定。

接着，在步骤 402 中，基于上述关键词判断上述可疑内容是否为同类内容。

在本实施例中，可以根据不同可疑内容对应的关键词确定不同可疑内容是否同类。具体来说，首先判断每组可疑用户群体在上述预定时间段内所点击的一组可疑内容的关键词中，同类关键词所占比例是否大于等于预定比例。如果同类关键词所占比例大于等于预定比例，则确定上述可疑内容为同类内容。

最后，在步骤 403 中，将上述可疑内容对应的可疑用户群体确定为待排除的非作弊用户群体。

应当注意，尽管在附图中以特定顺序描述了本发明方法的操作，但是，这并非要求或者暗示必须按照该特定顺序来执行这些操作，或是必须执行全部所示的操作才能实现期望的结果。相反，流程图中描绘的步骤可以改变执行顺序。附加地或备选地，可以省略某些步骤，将多个步骤合并为一个步骤执行，和/或将一个步骤分解为多个步骤执行。

进一步参考图 5，其示出了根据本申请的用于检测点击作弊的装置的一个实施例的结构示意图。

如图 5 所示，本实施例的装置 500 包括：第一确定单元 501，第二确定单元 502，第三确定单元 503 和第四确定单元 504。其中，第一确定单元 501 用于基于预定时间段内预定内容被用户点击的次数确定

可疑的点击。第二确定单元 502 用于根据所述可疑的点击确定至少一组有作弊嫌疑的可疑用户群体。第三确定单元 503 用于根据每组所述可疑用户群体在所述预定时间段内所点击的可疑内容的关键词确定待排除的非作弊用户群体。第四确定单元 504 用于排除所述可疑用户群体中非作弊用户群体以确定作弊用户群体。

在一些可选实施方式中，第一确定单元 501 包括获取子单元，判断子单元和确定子单元（未示出）。其中，获取子单元用于获取所述预定时间段内每个预定内容被用户点击的次数。判断子单元用于判断所述每个预定内容被点击的次数是否满足预定条件。确定子单元用于将被点击的次数满足预定条件的预定内容所对应的点击确定为可疑的点击。

在一些可选实施方式中，判断子单元配置用于：判断所述被点击的次数是否大于等于第一预定阈值，且小于等于第二预定阈值。如果是，确定所述被点击的次数满足预定条件。

在一些可选实施方式中，所述第二确定单元 502 包括信息获取子单元和用户群体确定子单元（未示出）。其中，信息获取子单元用于获取所述可疑的点击的相关信息。用户群体确定子单元用于基于所述相关信息确定至少一组可疑用户群体，其中，每组可疑用户群体在相同时间段内点击同一组内容。

在一些可选实施方式中，所述可疑的点击的相关信息包括以下至少一项：可疑的点击所对应的用户的标识信息；可疑的点击所对应的内容的标识信息；以及可疑的点击所对应的时刻。

在一些可选实施方式中，所述用户群体确定子单元配置用于：基于所述相关信息对所述可疑的点击进行聚类，使得每个聚类中心对应的用户群体在相同时间段内点击同一组内容；将所述每个聚类中心对应的用户群体确定为一组可疑用户群体。

在一些可选实施方式中，第三确定单元 503 包括关键词获取子单元，类别判断子单元和待排除群体确定子单元（未示出）。关键词获取子单元用于获取每组所述可疑用户群体在所述预定时间段内所点击的可疑内容的关键词。类别判断子单元用于基于所述关键词判断所述可

疑内容是否为同类内容。待排除群体确定子单元用于响应于可疑内容为同类内容，将所述可疑内容对应的可疑用户群体确定为待排除的非作弊用户群体。

在一些可选实施方式中，所述类别判断子单元配置用于：判断所述关键词中同类关键词所占比例是否大于等于预定比例；如果是，确定所述可疑内容为同类内容。

应当理解，装置 500 中记载的诸单元或模块与参考图 1-4 描述的方法中的各个步骤相对应。由此，上文针对方法描述的操作和特征同样适用于装置 500 及其中包含的单元，在此不再赘述。装置 500 可以预先设置在服务器中，也可以通过下载等方式而加载到服务器中。装置 500 中的相应单元可以与服务器中的单元相互配合以实现的用于检测点击作弊的方案。

下面参考图 6，其示出了适于用来实现本申请实施例的终端设备或服务器的计算机系统 600 的结构示意图。

如图 6 所示，计算机系统 600 包括中央处理单元（CPU）601，其可以根据存储在只读存储器（ROM）602 中的程序或者从存储部分 608 加载到随机访问存储器（RAM）603 中的程序而执行各种适当的动作和处理。在 RAM 603 中，还存储有系统 600 操作所需的各种程序和数据。CPU 601、ROM 602 以及 RAM 603 通过总线 604 彼此相连。输入/输出（I/O）接口 605 也连接至总线 604。

以下部件连接至 I/O 接口 605：包括键盘、鼠标等的输入部分 606；包括诸如阴极射线管（CRT）、液晶显示器（LCD）等以及扬声器等的输出部分 607；包括硬盘等的存储部分 608；以及包括诸如 LAN 卡、调制解调器等的网络接口卡的通信部分 609。通信部分 609 经由诸如因特网的网络执行通信处理。驱动器 610 也根据需要连接至 I/O 接口 605。可拆卸介质 611，诸如磁盘、光盘、磁光盘、半导体存储器等等，根据需要安装在驱动器 610 上，以便于从其上读出的计算机程序根据需要被安装入存储部分 608。

特别地，根据本公开的实施例，上文参考流程图描述的过程可以被实现为计算机软件程序。例如，本公开的实施例包括一种计算机程

序产品，其包括有形地包含在机器可读介质上的计算机程序，所述计算机程序包含用于执行流程图所示的方法的程序代码。在这样的实施例中，该计算机程序可以通过通信部分 609 从网络上被下载和安装，和/或从可拆卸介质 611 被安装。

附图中的流程图和框图，图示了按照本发明各种实施例的系统、方法和计算机程序产品的可能实现的体系架构、功能和操作。在这点上，流程图或框图中的每个方框可以代表一个模块、程序段、或代码的一部分，所述模块、程序段、或代码的一部分包含一个或多个用于实现规定的逻辑功能的可执行指令。也应当注意，在有些作为替换的实现中，方框中所标注的功能也可以以不同于附图中所标注的顺序发生。例如，两个接连地表示的方框实际上可以基本并行地执行，它们有时也可以按相反的顺序执行，这依所涉及的功能而定。也要注意的是，框图和/或流程图中的每个方框、以及框图和/或流程图中的方框的组合，可以用执行规定的功能或操作的专用的基于硬件的系统来实现，或者可以用专用硬件与计算机指令的组合来实现。

描述于本申请实施例中所涉及到的单元模块可以通过软件的方式实现，也可以通过硬件的方式来实现。所描述的单元模块也可以设置在处理器中，例如，可以描述为：一种处理器包括操作第一确定单元，第二确定单元，第三确定单元和第四确定单元。其中，这些单元模块的名称在某种情况下并不构成对该单元模块本身的限定，例如，第一确定单元还可以被描述为“用于基于预定时间段内预定内容被用户点击的次数确定可疑的点击的单元”。

作为另一方面，本申请还提供了一种计算机可读存储介质，该计算机可读存储介质可以是上述实施例中所述装置中所包含的计算机可读存储介质；也可以是单独存在，未装配入终端中的计算机可读存储介质。所述计算机可读存储介质存储有一个或者一个以上程序，所述程序被一个或者一个以上的处理器用来执行描述于本申请的用于检测点击作弊的方法。

以上描述仅为本申请的较佳实施例以及对所运用技术原理的说明。本领域技术人员应当理解，本申请中所涉及的发明范围，并不限

于上述技术特征的特定组合而成的技术方案，同时也应涵盖在不脱离所述发明构思的情况下，由上述技术特征或其等同特征进行任意组合而形成的其它技术方案。例如上述特征与本申请中公开的(但不限于)具有类似功能的技术特征进行互相替换而形成的技术方案。

# 权 利 要 求 书

1、一种用于检测点击作弊的方法，其特征在于，所述方法包括：

基于预定时间段内预定内容被用户点击的次数确定可疑的点击；

根据所述可疑的点击确定至少一组有作弊嫌疑的可疑用户群体；

根据每组所述可疑用户群体在所述预定时间段内所点击的可疑内容的关键词确定待排除的非作弊用户群体；以及

排除所述可疑用户群体中非作弊用户群体以确定作弊用户群体。

2、根据权利要求 1 所述的方法，其特征在于，所述基于预定时间段内预定内容被用户点击的次数确定可疑的点击，包括：

获取所述预定时间段内每个预定内容被用户点击的次数；

判断所述每个预定内容被点击的次数是否满足预定条件；

将被点击的次数满足预定条件的预定内容所对应的点击确定为可疑的点击。

3、根据权利要求 2 所述的方法，其特征在于，判断所述被点击的次数是否满足预定条件，包括：

判断所述被点击的次数是否大于等于第一预定阈值，且小于等于第二预定阈值；

如果是，确定所述被点击的次数满足预定条件。

4、根据权利要求 1-3 任意一项所述的方法，其特征在于，所述根据所述可疑的点击确定至少一组有作弊嫌疑的可疑用户群体，包括：

获取所述可疑的点击的相关信息；

基于所述相关信息确定至少一组可疑用户群体，其中，每组可疑用户群体在相同时间段内点击同一组内容。

5、根据权利要求 4 所述的方法，其特征在于，所述可疑的点击的相关信息包括以下至少一项：

可疑的点击所对应的用户的标识信息;

可疑的点击所对应的内容的标识信息;以及

可疑的点击所对应的时刻。

6、根据权利要求 4-5 任意一项所述的方法，其特征在于，所述基于所述相关信息确定至少一组可疑用户群体，包括:

基于所述相关信息对所述可疑的点击进行聚类，使得每个聚类中心对应的用户群体在相同时间段内点击同一组内容;

将所述每个聚类中心对应的用户群体确定为一组可疑用户群体。

7、根据权利要求 1-6 任意一项所述的方法，其特征在于，所述根据每组所述可疑用户群体在所述预定时间段内所点击的可疑内容的关键词确定待排除的非作弊用户群体，包括:

获取每组所述可疑用户群体在所述预定时间段内所点击的可疑内容的关键词;

基于所述关键词判断所述可疑内容是否为同类内容;

如果是，将所述可疑内容对应的可疑用户群体确定为待排除的非作弊用户群体。

8、根据权利要求 7 所述的方法，其特征在于，所述基于所述关键词判断所述可疑内容是否为同类内容，包括:

判断所述关键词中同类关键词所占比例是否大于等于预定比例;

如果是，确定所述可疑内容为同类内容。

9、一种用于检测点击作弊的装置，其特征在于，所述装置包括:

第一确定单元，用于基于预定时间段内预定内容被用户点击的次数确定可疑的点击;

第二确定单元，用于根据所述可疑的点击确定至少一组有作弊嫌疑的可疑用户群体;

第三确定单元，用于根据每组所述可疑用户群体在所述预定时间

段内所点击的可疑内容的关键词确定待排除的非作弊用户群体；以及

第四确定单元，用于排除所述可疑用户群体中非作弊用户群体以确定作弊用户群体。

10、根据权利要求 9 所述的装置，其特征在于，所述第一确定单元包括：

获取子单元，用于获取所述预定时间段内每个预定内容被用户点击的次数；

判断子单元，用于判断所述每个预定内容被点击的次数是否满足预定条件；

确定子单元，用于将被点击的次数满足预定条件的预定内容所对应的点击确定为可疑的点击。

11、根据权利要求 10 所述的装置，其特征在于，所述判断子单元配置用于：

判断所述被点击的次数是否大于等于第一预定阈值，且小于等于第二预定阈值；

如果是，确定所述被点击的次数满足预定条件。

12、根据权利要求 9-11 任意一项所述的装置，其特征在于，所述第二确定单元包括：

信息获取子单元，用于获取所述可疑的点击的相关信息；

用户群体确定子单元，用于基于所述相关信息确定至少一组可疑用户群体，其中，每组可疑用户群体在相同时间段内点击同一组内容。

13、根据权利要求 12 所述的装置，其特征在于，所述可疑的点击的相关信息包括以下至少一项：

可疑的点击所对应的用户的标识信息；

可疑的点击所对应的内容的标识信息；以及

可疑的点击所对应的时刻。

14、根据权利要求 12-13 任意一项所述的装置，其特征在于，所述用户群体确定子单元配置用于：

基于所述相关信息对所述可疑的点击进行聚类，使得每个聚类中心对应的用户群体在相同时间段内点击同一组内容；

将所述每个聚类中心对应的用户群体确定为一组可疑用户群体。

15、根据权利要求 9-14 任意一项所述的装置，其特征在于，所述第三确定单元包括：

关键词获取子单元，用于获取每组所述可疑用户群体在所述预定时间段内所点击的可疑内容的关键词；

类别判断子单元，用于基于所述关键词判断所述可疑内容是否为同类内容；

待排除群体确定子单元，用于响应于可疑内容为同类内容，将所述可疑内容对应的可疑用户群体确定为待排除的非作弊用户群体。

16、根据权利要求 15 所述的装置，其特征在于，所述类别判断子单元配置用于：

判断所述关键词中同类关键词所占比例是否大于等于预定比例；

如果是，确定所述可疑内容为同类内容。

17、一种设备，其特征在于，包括：

一个或者多个处理器；

存储器；

一个或者多个程序，所述一个或者多个程序存储在所述存储器中，当被所述一个或多个处理器执行时：

基于预定时间段内预定内容被用户点击的次数确定可疑的点击；

根据所述可疑的点击确定至少一组有作弊嫌疑的可疑用户群体；

根据每组所述可疑用户群体在所述预定时间段内所点击的可疑内容的关键词确定待排除的非作弊用户群体；以及

排除所述可疑用户群体中非作弊用户群体以确定作弊用户群体。

18、一种非易失性计算机存储介质，所述计算机存储介质存储有一个或多个程序，当所述一个或者多个程序被一个设备执行时，使得所述设备：

基于预定时间段内预定内容被用户点击的次数确定可疑的点击；

根据所述可疑的点击确定至少一组有作弊嫌疑的可疑用户群体；

根据每组所述可疑用户群体在所述预定时间段内所点击的可疑内容的关键词确定待排除的非作弊用户群体；以及

排除所述可疑用户群体中非作弊用户群体以确定作弊用户群体。

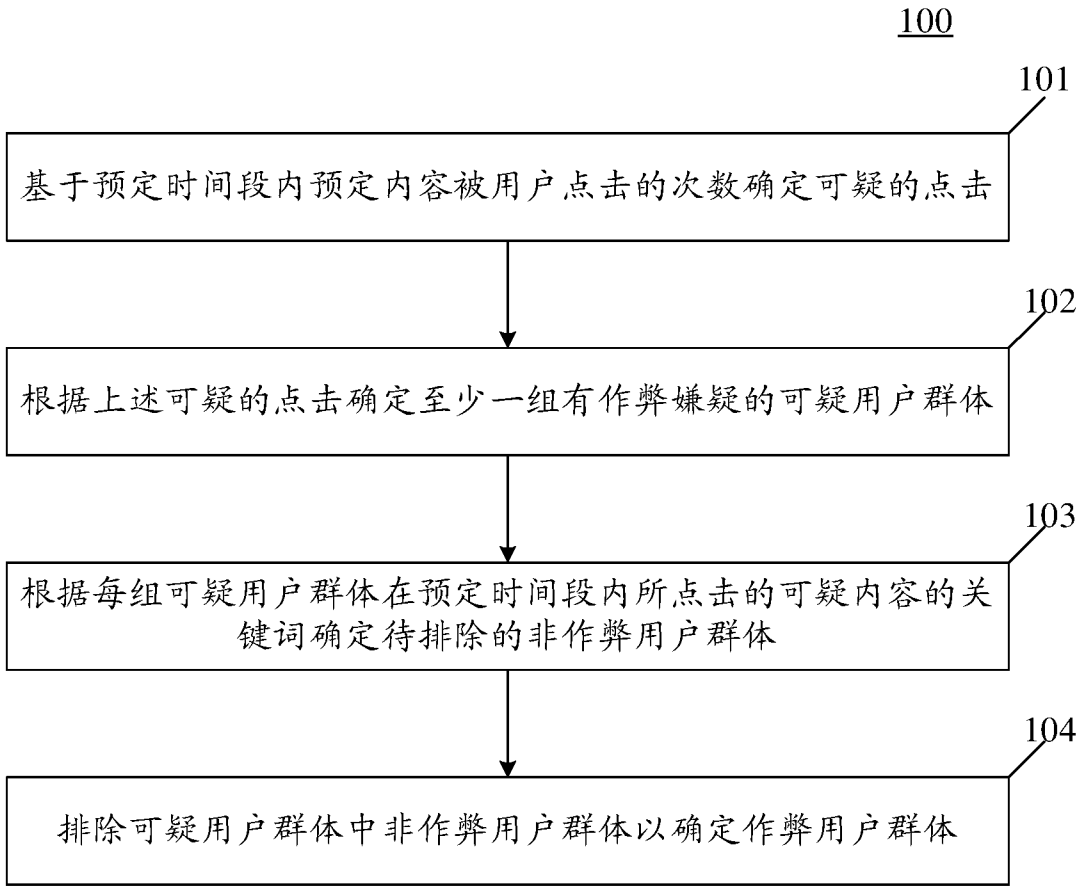

100
101
基于预定时间段内预定内容被用户点击的次数确定可疑的点击
102
根据上述可疑的点击确定至少一组有作弊嫌疑的可疑用户群体
103
根据每组可疑用户群体在预定时间段内所点击的可疑内容的关键词确定待排除的非作弊用户群体
104
排除可疑用户群体中非作弊用户群体以确定作弊用户群体

图 1

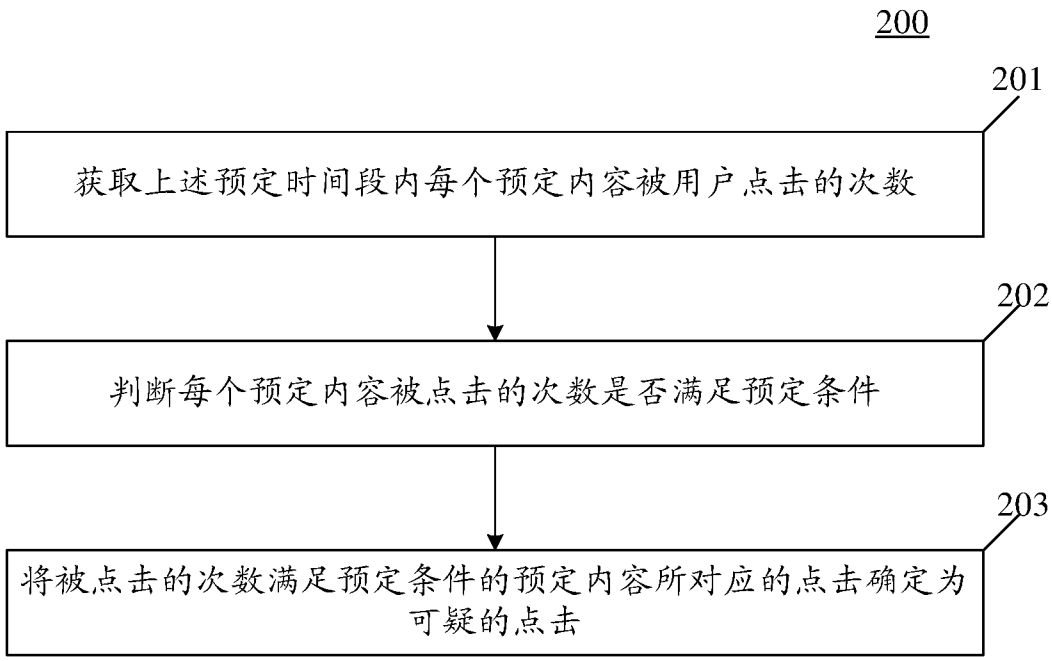

200
201
获取上述预定时间段内每个预定内容被用户点击的次数
202
判断每个预定内容被点击的次数是否满足预定条件
203
将被点击的次数满足预定条件的预定内容所对应的点击确定为可疑的点击

图 2

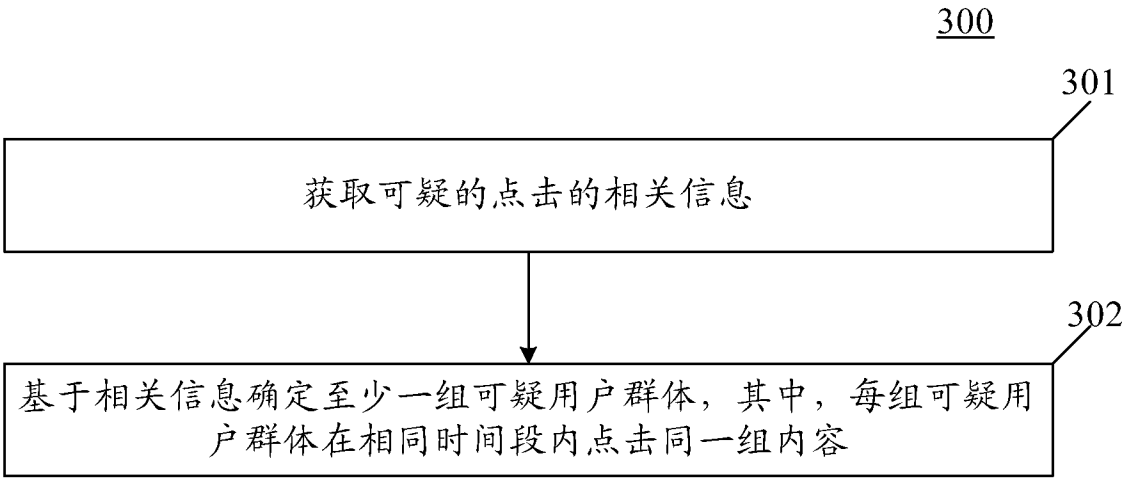

300
301
获取可疑的点击的相关信息
302
基于相关信息确定至少一组可疑用户群体，其中，每组可疑用户群体在相同时间段内点击同一组内容

图 3

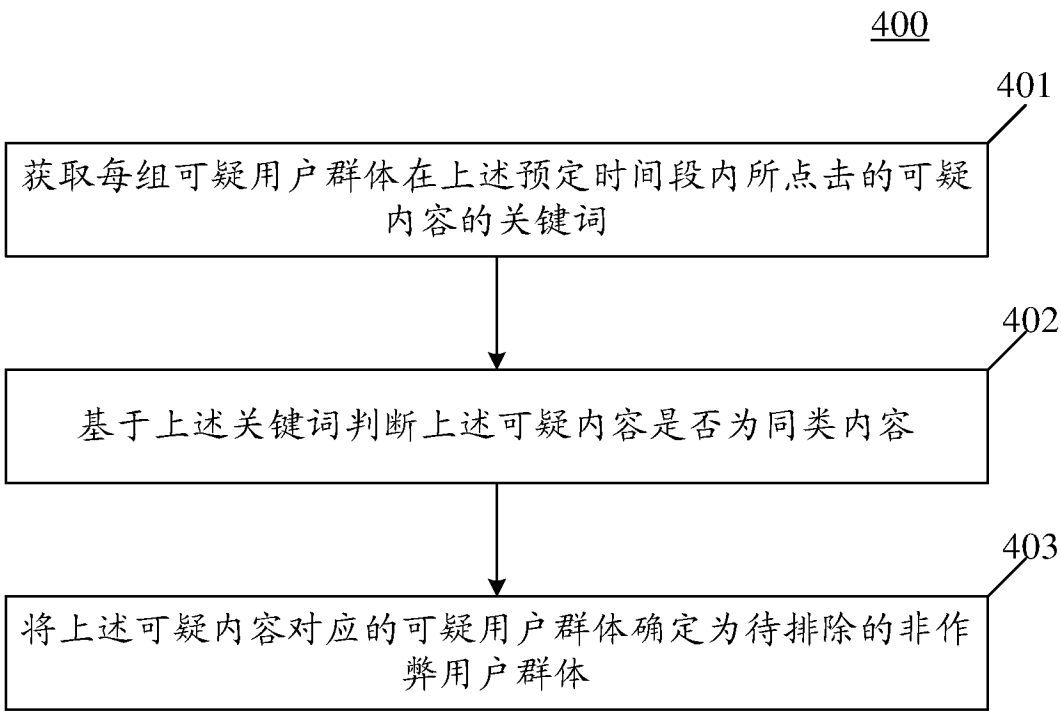

400
401
获取每组可疑用户群体在上述预定时间段内所点击的可疑内容的关键词
402
基于上述关键词判断上述可疑内容是否为同类内容
403
将上述可疑内容对应的可疑用户群体确定为待排除的非作弊用户群体

图 4

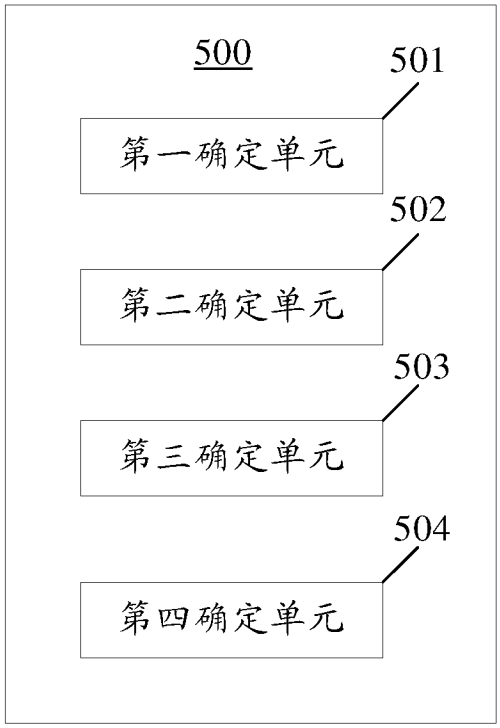

500
501
第一确定单元
502
第二确定单元
503
第三确定单元
504
第四确定单元

图 5

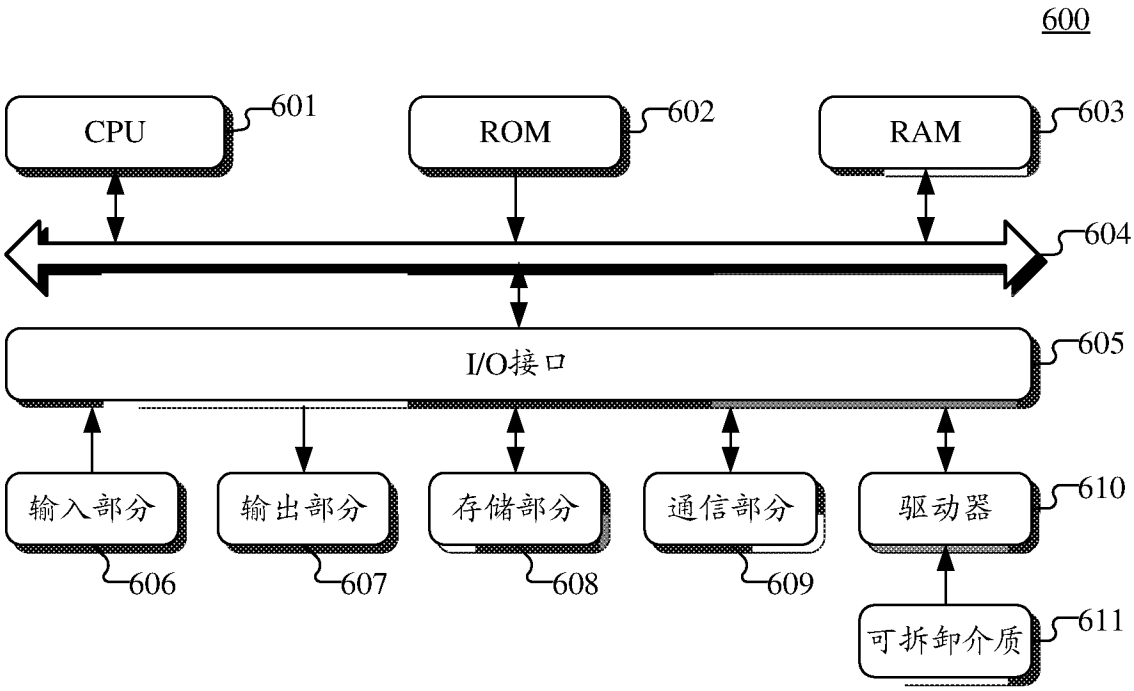

600
CPU
601
ROM
602
RAM
603
604
I/O接口
605
输入部分
606
输出部分
607
存储部分
608
通信部分
609
驱动器
610
可拆卸介质
611

图 6

# INTERNATIONAL SEARCH REPORT

International application No.
PCT/CN2015/089545

**A. CLASSIFICATION OF SUBJECT MATTER**

G06F 17/30 (2006.01) i

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

G06F

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

WPI, EPODOC, CNPAT, CNKI: cheat+, click, user, group, malicious click, paid listing, detect+, non-cheating, normal, exclude, suspicious, keyword?, real, fraud, check+, website, content?, remov+

C. DOCUMENTS CONSIDERED TO BE RELEVANT

| Category* | Citation of document, with indication, where appropriate, of the relevant passages | Relevant to claim No. |
|---|---|---|
| PX | CN 104765874 A (BAIDU ONLINE NETWORK TECHNOLOGY BEIJING CO., LTD.) 08 July 2015 (08.07.2015) claims 1 to 16, description, page 6, paragraph [0072] to page 8, paragraph [0087] | 1-18 |
| A | CN 103390027 A (IZP BEIJING TECHNOLOGY CO., LTD.) 13 November 2013 (13.11.2013) description, page 5, paragraph [0087] to page 7, paragraph [0122] | 1-18 |
| A | CN 101093510 A (BEIJING SOGO TECH DEV CO., LTD.) 26 December 2007 (26.12.2007) the whole document | 1-18 |
| A | US 2009299967 A1 (MICROSOFT CORPORATION) 03 December 2009 (03.12.2009) the whole document | 1-18 |

☐ Further documents are listed in the continuation of Box C. ☒ See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&"document member of the same patent family

| Date of the actual completion of the international search | Date of mailing of the international search report |
|---|---|
| 12 January 2016 | 02 February 2016 |
| Name and mailing address of the ISA<br>State Intellectual Property Office of the P. R. China<br>No. 6, Xitucheng Road, Jimenqiao<br>Haidian District, Beijing 100088, China<br>Facsimile No. (86-10) 62019451 | Authorized officer<br>TIAN, Jing<br>Telephone No. (86-10) 62413701 |

Form PCT/ISA /210 (second sheet) (July 2009)

**INTERNATIONAL SEARCH REPORT**
Information on patent family members

International application No.
PCT/CN2015/089545

| Patent Documents referred in the Report | Publication Date | Patent Family | Publication Date |
|---|---|---|---|
| CN 104765874 A | 08 July 2015 | None | |
| CN 103390027 A | 13 November 2013 | None | |
| CN 101093510 A | 26 December 2007 | None | |
| US 2009299967 A1 | 03 December 2009 | None | |

# 国际检索报告

国际申请号
PCT/CN2015/089545

A. 主题的分类

G06F 17/30(2006.01)i

按照国际专利分类(IPC)或者同时按照国家分类和IPC两种分类

B. 检索领域

检索的最低限度文献(标明分类系统和分类号)

G06F

包含在检索领域中的除最低限度文献以外的检索文献

在国际检索时查阅的电子数据库(数据库的名称，和使用的检索词（如使用））

WPI, EPODOC, CNPAT, CNKI: 作弊, 点击, 用户, 群, 恶意点击, 竞价排名, 检测, 非作弊, 正常, 排除, 可疑, 关键词, 真实, user, click, cheat+, fraud, check+, keyword?, website, detect+, content?, group, remov+

C. 相关文件

| 类 型* | 引用文件，必要时，指明相关段落 | 相关的权利要求 |
|---|---|---|
| PX | CN 104765874 A (百度在线网络技术北京有限公司) 2015年 7月 8日 (2015-07-08)<br>权利要求1-16，说明书第6页第72段-第8页87段 | 1-18 |
| A | CN 103390027 A (亿赞普北京科技有限公司) 2013年 11月 13日 (2013-11-13)<br>说明书第5页第87段-第7页第122段 | 1-18 |
| A | CN 101093510 A (北京搜狗科技发展有限公司) 2007年 12月 26日 (2007-12-26)<br>全文 | 1-18 |
| A | US 2009299967 A1 (MICROSOFT CORPORATION) 2009年 12月 3日 (2009-12-03)<br>全文 | 1-18 |

☐ 其余文件在C栏的续页中列出。

☑ 见同族专利附件。

* 引用文件的具体类型：

"A" 认为不特别相关的表示了现有技术一般状态的文件

"E" 在国际申请日的当天或之后公布的在先申请或专利

"L" 可能对优先权要求构成怀疑的文件，或为确定另一篇引用文件的公布日而引用的或者因其他特殊理由而引用的文件（如具体说明的）

"O" 涉及口头公开、使用、展览或其他方式公开的文件

"P" 公布日先于国际申请日但迟于所要求的优先权日的文件

"T" 在申请日或优先权日之后公布，与申请不相抵触，但为了理解发明之理论或原理的在后文件

"X" 特别相关的文件，单独考虑该文件，认定要求保护的发明不是新颖的或不具有创造性

"Y" 特别相关的文件，当该文件与另一篇或者多篇该类文件结合并且这种结合对于本领域技术人员为显而易见时，要求保护的发明不具有创造性

"&" 同族专利的文件

| 国际检索实际完成的日期 | 国际检索报告邮寄日期 |
|---|---|
| 2016年 1月 12日 | 2016年 2月 2日 |
| ISA/CN的名称和邮寄地址<br>中华人民共和国国家知识产权局(ISA/CN)<br>中国北京市海淀区蓟门桥西土城路6号 100088<br>传真号 (86-10)62019451 | 受权官员<br>田晶<br>电话号码 (86-10)62413701 |

国际检索报告
关于同族专利的信息

国际申请号
PCT/CN2015/089545

| 检索报告引用的专利文件 | 公布日<br>(年/月/日) | 同族专利 | 公布日<br>(年/月/日) |
|---|---|---|---|
| CN 104765874 A | 2015年 7月 8日 | 无 | |
| CN 103390027 A | 2013年 11月 13日 | 无 | |
| CN 101093510 A | 2007年 12月 26日 | 无 | |
| US 2009299967 A1 | 2009年 12月 3日 | 无 | |